**Abstract:**

Ongoing brain activity constitutes a structuring context for the stimulus processing and has been shown to predict both neural and behavioural responses. In the present study, we used EEG to investigate the relationship between brain activity during a speech-shaped noise-filled pre-target time window, and the recognition of digit-words embedded in noise. We evaluated both objective (accuracy) and subjective (clarity rating) measures of comprehension at three signal-to-noise ratios calibrated to individual participants' performance. We find that in correctly reported trials, EEG power averaged over participants and channels during the noise-filled pre-target period is systematically elevated in the alpha (8.5-12Hz) frequency band compared to incorrectly reported trials. The same pattern is observed for subjectively clearer as compared to less clear trials. Linear mixed effects logistic regression analysis of source-reconstructed alpha power on behavioural outcomes revealed that alpha power in a left posterior temporal area in a time window extending from at least 300ms to 50ms prior to digit onset predicted target recognition. Alpha power interacts with stimulus difficulty, such that the pattern of activity revealed benefits for performance for easier stimuli but has a deleterious impact on recognition of more heavily masked words. We propose that the mechanism we uncover reflects a gain modulation mechanism that can suppress noise in speech-relevant perceptual channels. This is a first step in identifying pre-stimulus activity patterns that may be exploited in the future as targets for neurofeedback or neurostimulation to assist with masked speech comprehension, for instance for individuals suffering from mild hearing impairment.

**Introduction:**

Ongoing brain activity as reflected in EEG recording is characterised by a hierarchical system of metastable brain rhythms producing an “ever-changing landscape” of neural activity (Buzsáki, 2019). Such brain rhythms are thought to reflect ongoing neural processes that can affect the behavioural outcomes of stimulus processing (Busch et al., 2009; Kayser et al., 2016; Mathewson et al., 2009; VanRullen et al., 2011), which parallels changes in the neural responses evoked by the stimuli (Becker et al., 2008, 2011; De Blasio et al., 2013; De Blasio & Barry, 2013, 2018; Rajagovindan & Ding, 2011; van Es & Schoffelen, 2019; Zazio et al., 2021). The relation between ongoing EEG and stimulus processing has been assessed from two complementary perspectives: on the one hand, studies aimed at elucidating the relation between properties of the EEG and individual traits, such as speech reception thresholds. On the other, studies aimed at assessing how within-subject variation over time (e.g., from trial to trial) in properties of the ongoing EEG predicts variation over time in behaviour, such as the probability of perceiving an ambiguous syllable as either of two (categorical) precepts.

It has been previously demonstrated an association between individual variability in word in noise comprehension and resting state (RS) power recorded in MEG (Houweling et al., 2020), whereby higher resting state high-β (21-29Hz) power in a left-lateralised temporal region and low-γ (30-40Hz) power in a smaller right posterior superior temporal gyrus (pSTG) predicted better performance on a Word-in-Noise task (WiN; Wilson et al., 2003, 2007). Interestingly, RS EEG data were collected at least one full day before the WiN experiment took place. Because of the high test-retest reliability of both RS MEG power (Martín-Buro et al., 2016) and WiN test scores (Wilson & Mcardle, 2007), these results suggest that WiN performance is associated with stable (at least over the time scale of days) brain states that are not contingent upon fluctuations in attention or arousal. A recent study by Alhanbali and colleagues (Alhanbali et al., 2021) provides further evidence for the role of ongoing EEG in the processing of speech in noise. Participants’ α (individual centre frequencies within the 8-13Hz range) power recorded at a centro-parietal electrode (but spatially unresolved) prior to the presentation of spoken digits presented in a noise background, was analysed for an association with their ability to recognise the digits. It was found that mean α power averaged over trials positively predicted performance, such that higher α power was associated with improved performance in the digit recognition task. One secondary aim of the current study was to assess the reproducibility of this result in a new dataset.

Whilst studying the relation between ongoing EEG and individual differences in speech recognition is highly valuable, in this study we aim at assessing the association between ongoing EEG power and the recognition of words in noise within participant. Understanding this relation may provide further insights on the neurobiological basis of the role of ongoing EEG activity in speech perception. Studies using intracranial recordings have demonstrated that α activity is inversely related to cortical excitability (Haegens et al., 2011; Iemi et al., 2021). Thus, reduced pre-stimulus α power results in the cortex being generally more susceptible to input, determining more frequent reports of stimulus presence (including false alarms) and higher levels of confidence in the perceptual decision (Lange et al., 2013; Limbach & Corballis, 2016; Samaha et al., 2017). In studies of auditory perception, properties of ongoing EEG have been found not only to predict stimulus detection, but also intensity- and frequency- discrimination of tones in noise (Kayser et al., 2016; McNair et al., 2019). Once more, the α rhythm plays a prominent role in these studies too. Properties of ongoing, pre-stimulus EEG activity also predict performance in higher-level speech perception tasks such as syllable discrimination (ten Oever & Sack, 2015) and lexical decision in noise (Strauß et al., 2015).

In sum, EEG power has been shown to predict both inter-individual differences in performance in speech-in-noise comprehension tasks and within-subject performance variability in a number of auditory detection and discrimination tasks. However, whether within-subject variability in EEG activity (e.g., on a trial-to-trial basis) modulates the probability of correctly recognising individual speech items embedded in noise, is yet to be demonstrated. By attempting to fill this knowledge gap, we aim to elucidate whether modulating EEG activity has the potential to affect speech recognition. Shall this be the case, this study may pave the way for the development and evaluation of non-invasive interventions, such as transcranial electrical stimulation or neurofeedback, aimed at restoring sub-optimal speech perception through externally- or internally-driven regulation of neural activity. In the current exploratory study, we used a Digit-in-Noise (DiN) recognition task to probe the role of pre-target EEG activity in predicting speech in noise recognition.

**Materials and methods:**

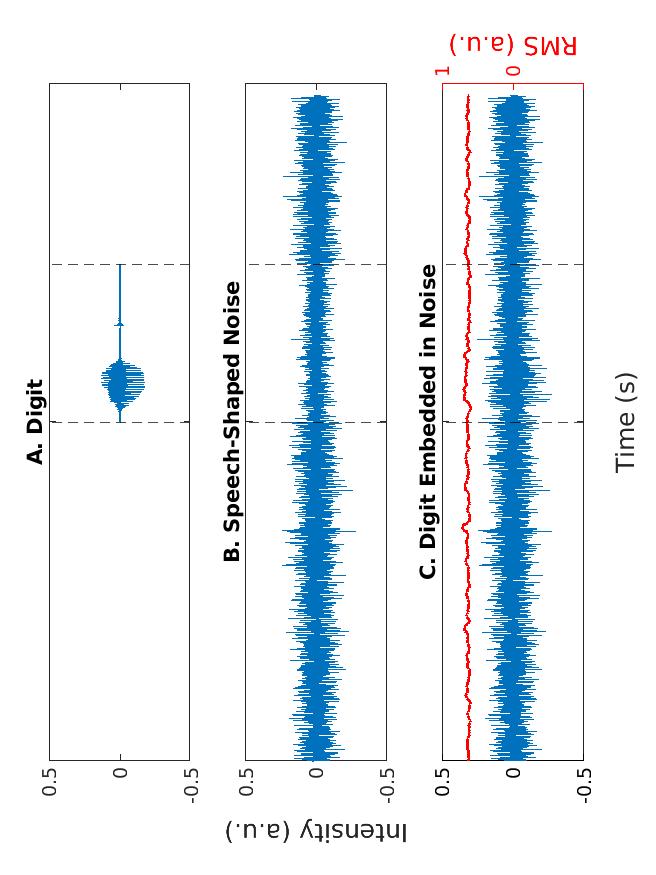
**Participants and experimental procedures.** Thirty-two healthy right-handed native speakers of German (mean age 23.5, SD=3.9 years; 29 female) took part in the experiment. Participants reported normal hearing abilities, and reported no history of neurological, psychiatric, or language-related problems. They were recruited through online and offline advertisement and received financial compensation for their participation. All procedures had ethical approval from the local ethical review board (anonymised for review).

Participants were seated in front of a computer screen and were asked to listen to the auditory stimuli and to respond by means of a numeric keyboard using the index or middle fingers of the right hand. Stimuli were sampled at 16kHz and delivered via a Focusrite™ Scarlett 2i4 Soundcard through earphones (Etymotic™ ER4-SR Studio) using PsychToolbox version 3 (PTB-3; Brainard, 1997) running under Matlab 2018b (the MathWorks®).

**Stimulus Material.** Stimuli consisted of monosyllabic German digits – zero to nine excluding seven – spoken in Standard German by ten speakers (five female), available as part of the EUROM corpus (Chan et al., 1995). Digits were chosen as target stimuli due to their high familiarity in order to minimise potential influence of individual differences in linguistic competence on performance. This choice was motivated by findings that have shown lexical access to have a substantial impact on speech in noise comprehension (Kaandorp et al., 2016). Digit durations were equalised to 465ms (the mean of the original recordings) using the WSOLA algorithm implemented in Matlab (Verhelst & Roelands, 1993). Digits were equalised for subjective loudness, based on the EMBSD loudness model (Timoney et al., 2004). Speech-shaped noise (SSN) was generated by convolving white noise with the long-term average spectrum of the concatenated digits. Both speech and noise were band-pass filtered (6th order Butterworth filter) between 50 and 5000Hz.

Digit in noise stimuli were generated such that they had an approximately constant RMS in the pre-, peri-, and post-target windows. This was achieved by reducing the amplitude of the SSN in the target window prior to mixing with the target at the required amplitude to achieve the desired SNR (calculated based on the mean RMS of the speech signal and the mean RMS of the SSN in the target window). Levels of SSN and the target digits were faded using a 1ms linear ramp to avoid clicks due to discontinuities in the signal. A summary of the stimulus processing steps is shown in Fig.1.

For the main experiment, duration of the pre-target segment was drawn from a uniform distribution between 3400 and 6000ms in 200ms steps, whilst that of the post-target segment was fixed to 500ms. For the pre-test stimulus calibration phase (see ‘Calibration’ section), duration of the pre-target segment was drawn from a uniform distribution between 600 and 2400ms in 200ms steps, with no post-target noise. Performance on the first block at the SRT50 level was 48.180 ± 13.646 (SD over subjects). Processing of the stimuli was carried out in Matlab 2018b (the MathWorks®).



**Figure 1. Stimulus processing steps.** **A.** Digit waveform. **B.** SSN waveform. **C.** SiN waveform. SSN amplitude in the target window is reduced before being mixed with the digit at the required amplitude to achieve the desired SNR such that the RMS is constant throughout the snippet. The moving RMS calculated over sliding windows of 320 samples (50ms) with an overlap corresponding to window length – 1 is represented by the red line on top of the waveform in panel C.

**Calibration of intelligibility.**  As part of pilot testing (on a sample of N=24) we observed a large variability in performance associated with both listeners and stimuli. In an attempt to minimise these differences, before the main experiment, stimulus difficulty was calibrated for each participant and digit individually, using an interleaved staircase procedure to find the Signal-to-Noise Ratio (SNR) of the DiN stimuli yielding a close to 50% correct performance level (SRT50). In other words, in a similar fashion to perceptual detection or discrimination studies, as part of the main test we presented DiN stimuli at a level close to the comprehension threshold, such that each item would be associated with the same probability of being comprehended both within and across participants. We reasoned that both listener- and digit-related sources of variability might obscure smaller effects associated with dynamics of the EEG.

On the first presentation, a DiN item was presented at a digit-specific SNR, established based on pilot testing. If the participant responded correctly, the digit was made less intelligible by decreasing its SNR by 5dB. The procedure was repeated until the participant provided an incorrect response (negative reversal). On the following presentations, the digit was made easier by increasing its SNR by 5dB, until a correct response was provided (positive reversal). At this point, SNR was decreased, albeit this time by 4dB SNR. Following two reversals, the SNR step size was progressively decreased, using the following values: 5, 4, 3, 2, 1, 0.5, and 0.25dB SNR. The test ended after every digit reached 14 reversals (7 positive and 7 negative) or after 270 trials in total were presented (30 per digit). In order to present a similar number of tokens per each digit, stimuli were drawn from random permutations of the nine possible digits with no replacement. The final dB SNR the test converged approximated the SRT50 of the digit.

**Task.** Digits were presented at three SNRs, centred on the SRT50, lower SNR at SRT50 -1.5dB, and higher SNR at SRT50 +1.5dB. Fifty-four control stimuli in quiet (6 per digit), and 108 DiN stimuli for each of the three SNRs (12 per digit per level) were presented. Participants were instructed to listen carefully to the stimuli whilst staring at a fixation cross in the centre of the screen. Following stimulus offset, they were prompted to try to identify them by a message on the screen. They reported the digit they heard using a numeric keypad, with onscreen display of their response, which they were required to confirm after entry, by pressing the "enter" key. There was no time limit for providing the response. Participants were then prompted by a second message on screen to rate their subjective impression of digit clarity on a scale ranging from ‘1 = totally unclear’ to ‘9 = totally clear’ (DiN stimuli only, to avoid anchoring of the ratings to the control stimuli). Again, they were required to confirm the response, for which they had no time limit, by pressing the "enter" key, in order to proceed to the next trial. We used both accuracy and clarity because subjective indices of comprehension have been shown to reveal subtle differences in listeners’ subjective experience in spite of similar comprehension performance (Sohoglu et al., 2014). Moreover, the subjective experience of listening to speech in noise is also of substantial import to the real-world experience of dealing with speech in noisy conditions.

**EEG data acquisition and pre-processing.** Participants' cerebral activity was recorded using a 128-channel EEG system (EGI HydroCel Geodesic Sensor Net™) at a 2kHz sampling rate. EEG data were pre-processed in EEGLAB12 (ver. 2020.0) toolbox (Delorme & Makeig, 2004) running under Matlab 2018b. Continuous data were down-sampled to 200Hz, high-pass filtered (0.1 Hz) to remove baseline drifts, and low-pass filtered at 48Hz using zero-phase non-causal Hamming windowed FIR filters. Residual line noise and low-frequency drifts were removed by adaptively estimating and removing significant sinusoidal artifacts from scalp channels using a frequency-domain (multi-taper) regression technique (Mitra & Bokil, 2007). Automated Artifcat Subspace Reconstruction (ASR) (Mullen et al., 2015, 2013) was then applied to detect and remove high-amplitude non-brain artifactual activity (produced by, e.g., eye blinks, muscle activity, and sensor motion) by comparing its structure to that of known artifact-free reference data, while recovering EEG background activity that lies in the subspace spanned by the artifact components when possible. Flatline and noisy channels were removed and interpolated, and robust average re-reference was applied. Following epoching (from -3395 to 1960 ms with reference to target onset), noisy epochs and no-response epochs were removed. Epochs were considered noisy if samples were removed following, if they showed a voltage difference greater than 200μV between the maximum and minimum sample in one or more channels between -500 and -5ms, or if they showed a voltage difference between contiguous samples greater than 75μV. Participants with fewer than 100 DiN trials after artefact rejection (n = 7) were excluded from further analyses. Thus, 25 individual data sets were analysed, with a mean of 270 trials (SD = 40, range = 199-312) per participant out of 324 trials presented, of which 92±15 (mean±SD over participants) at lower SNRs, 82±19 at medium SNRs, and 96±18 at higher SNRs, resulting in 123±25 correct and 147±27 incorrect trials per participant**.** *χ2-*test revealed that the number of trials included was not biased by difficulty level (*χ2*(2) = 1.169, *p* = .558), and that the ratio of correct to incorrect responses did not deviate significantly from 1:1 (*χ2*(1) = 2.136, *p* = .144).Independent component analysis was then applied to the segmented data (Makeig et al., 1996), and non-brain components were removed semi-automatically by scrutiny of their topography and spectral profile. On average, 115.760 linearly independent components (ICs) were estimated per each subject dataset (SD = 4.824, range = 105-122), of which 35.760 components (SD = 9.714, range = 21-52), explaining 48.150% (SD = 15.664%, range = 25.561-82.337%) of total scalp variance, were retained. In order to quantify the change produced by the removal of non-brain components to the signal of interest (i.e., 8.5-12Hz α activity, see below), we computed the Pearson’s correlation between trial-wise α activity before and after IC removal. Mean Person’s correlation coefficient, averaged over trials, channels, and subject was 0.781 (± 0.172 SD over subjects), indicating that, in spite of the rather conservative retention rate of ICs, the cleaned signal explains more than 60% of the variance in the original signal.

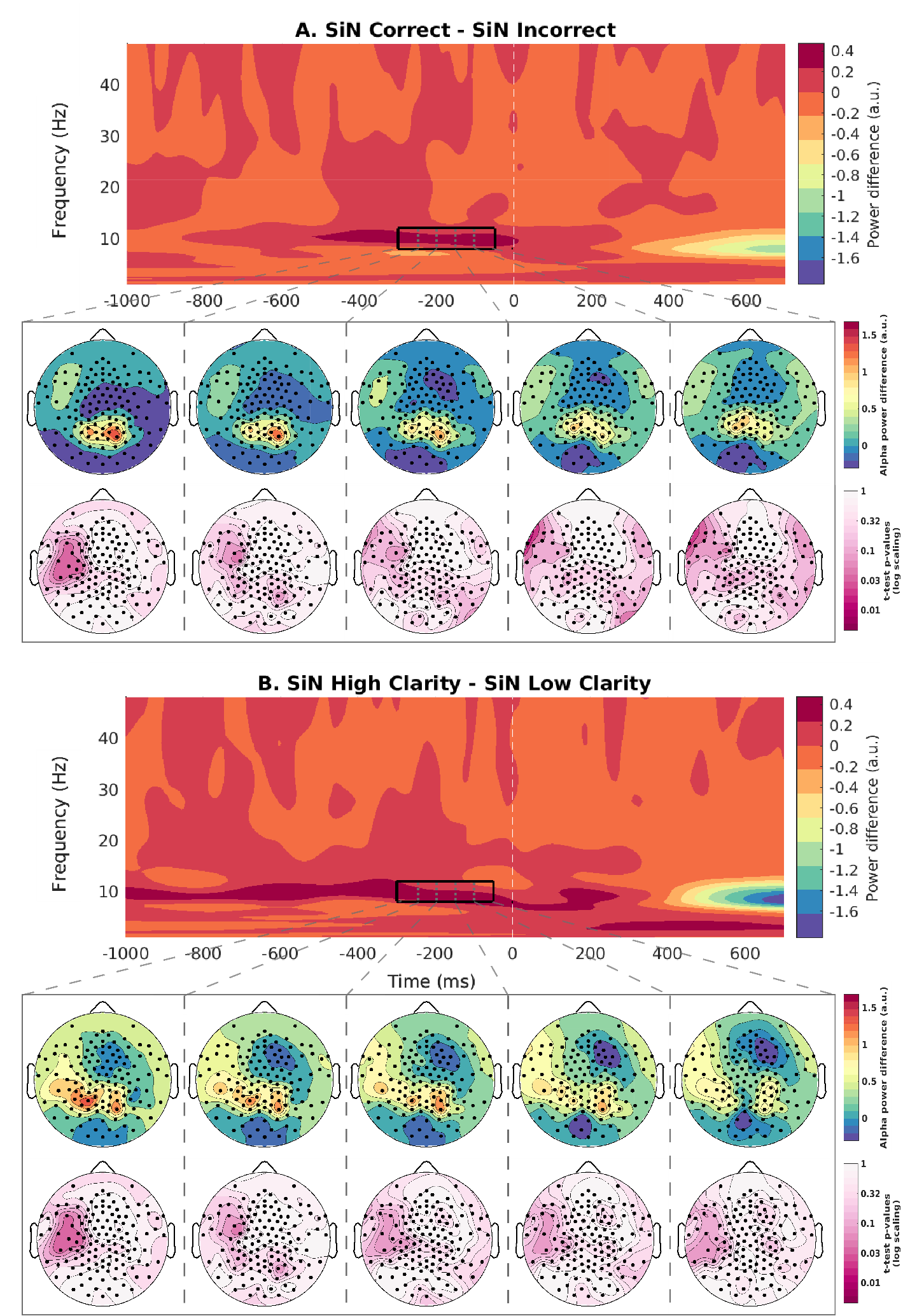
**Replication of previous findings.** One secondary aim of the study was to test whether the previously-reported association between subject means of pre-stimulus α power and individual SRT scores (Alhanbali et al., 2021) could be replicated in the current dataset. To do so, we aimed at reproducing, with the smallest differences possible, the setup, pre-processing and analysis steps described in Alhanbali et al. (2021). First, we re-referenced epoched voltage time series data to the linked mastoids (electrode TP9-TP10 in the 10-20 system). We then computed time-frequency power at Pz over the whole -3.4 to 1 sec epoch through convolution of voltage with a family of 1Hz-wide linearly-spaced wavelets spanning the 5-20Hz range, and averaged them between -600 and -100ms. We then estimated the peak α power in the 8-13Hz range and averaged peak α power over trials, separately for clear (~ pre-STIM in Alhanbali et al., 2021) and noisy (~pre-TARG) trials. We then correlated the log of this subject-wise means with subjects’ mean SRT over digits. Spearman’s correlation, revealed that neither pre-STIM (*r*=-0.117, *p*=0.576), nor pre-TARG (*r*=0.140, *p*=0.505) α power at Pz were predictive of individual SRTs. Having failed to replicate this finding, we then aimed at expanding the scope of this inquiry to the brain as a whole (as compared to one electrode), and to more discrete points in time. Further, we aimed at complementing the investigation of the association between (trait-like) individual measures of EEG activity and SiN comprehension ability with that of the association between (state-like) within-individual EEG measures and comprehension probability, by means of a trial-by-trial analysis. Finally, it has to be noticed that our SiN test was different from that used in Alhanbali et al. (2021) in various regards and that the failed replication might be due to such differences.

**EEG source reconstruction.** A volume conduction model of the head was created from a 3-shell triangulated source model with 8196 source points using the boundary element method (BEM) as implemented by OpenMEEG (Gramfort et al., 2010; Kybic et al., 2006). The volume conduction model was used to compute the lead field matrix at each source point. A spatial filter was then computed for each participant using linear constrained minimum variance (LCMV) beamformer (Van Veen et al., 1997) to project trial data in source space. Data were averaged over a number of source points, as defined by the cortical Conte69 surface-based atlas (Van Essen et al., 2012) composed of 386 parcels, including 382 cortical parcels (http://brainvis.wustl.edu/wiki/index.php/Caret:Atlases:Conte69\_Atlas). Cortical parcels consisted of a mean of 18.78 source points (SD = 4.44). All the steps for source reconstruction were carried out using FieldTrip (Oostenveld et al., 2011) version 20210614.

**Definition of time-frequency region-of-interest (TFROI).** Time-frequency power was calculated at each source point by means of convolution of the EEG signal with a family of 56 log-spaced analytic Morse wavelets between 1 and 48Hz. We then computed grand mean (over participants, channels, and trials) differences between correct and incorrect trials, and between perceptually clear and unclear (defined, per participant, as clarity ratings between the 67th and 100th, and between the 1st and 33rd percentiles of the individual clarity ratings distribution, respectively. Differences in power between correct and incorrect, as well as between perceptually clear and unclear trials were most prominently observed in the α band (Fig.2). Further analyses that aimed to better characterise the relationship between brain activity and performance were therefore subsequently restricted to this frequency band, by averaging parcel power in the range 8.49 and 12.00Hz, in 50-ms time bins between -300ms and -50ms.

We chose to run the analysis within this time window since the studies most closely related to the current (i.e., studies assessing the effect of pre-stimulus EEG on tasks involving verbal material) one observed effects within this time frame. Specifically, the predictive effect of α phase in distinguishing between words and pseudo-words presented in noise was observed between 100ms preceding target onset. Further, the effect of a 6Hz oscillation in biasing the perception of ambiguous syllables was observed between 250ms and 100ms prior to sound onset (ten Oever & Sack, 2015). We opted for analysing data in 50ms-wide time bins for reasons of computational tractability. We did not analyse EEG data between -50ms and digit onset because data within this time frame is possibly contaminated by digit-evoked activity. Specifically, some TF points within this time bin would fall within the boundaries of a hypothetical Cone of Influence (the region where the wavelet power spectra are distorted because of the influence of data points outside of the analysis window) resulting from the application of the wavelet analysis shifted at earlier points in time such that the right edge of the analysis windows is set at digit onset.

We also notice that correct and high clarity SiN trials show increased α event-related desynchronization from about 400ms (in the post-target period) compared to incorrect and low clarity trials, respectively. However, discussion of event-related activity goes beyond the scope of this paper. Further, preliminary analyses indicated that the strongest differences in subject means of channel-wise α power (pairwise *t-*test) were localised to left temporal electrodes between -150 and -50ms (Fig.2). Therefore, although we could not replicate the association between α power and SRT at Pz observed by Alhanbali and colleagues (2021), we observe a significant association between left temporal α power and proportion of correct responses. Subsequent analyses were aimed at characterising the relation between source-reconstructed α activity and SiN comprehension, and the effect that SNR has on this relation.



**Figure 2.** Spectrograms show the difference in power, averaged over participants and sensors, between trial types. The black solid rectangle represents the time-frequency region of interest over which subsequent analyses were carried. Topographic maps show the difference in α power (8.5-12Hz), averaged over participants and 50ms time bins (dashed grey lines in spectrograms) and the p-values associated with the pairwise t-test comparing subject means. **A.** SiN correct – SiN incorrect. **B.** High clarity SiN – low clarity SiN.

**Association between pre-target EEG power and DiN recognition accuracy and subjective clarity.** In order to test whether fluctuations in pre-target α power predict variability in comprehension or perceived clarity of DiN – and whether they do so differently depending on the SNR of the stimuli – whilst accounting for variability in the outcome associated with different participants and digits (see behavioural results), we conducted a univariate regression analysis at each cortical parcel and time bin, separately for comprehension and clarity.

Specifically, the effect of comprehension was assessed using the following mixed-effect logistic regression model: *comprehension ~ α power \* SNR + (1|subject:digit)*, where *comprehension* is dichotomous variable representing the behavioural outcome (correct vs incorrect), *α power* is a continuous variable representing pre-target α power at one parcel and time bin (50ms each, from -300 to -50ms) standardised within participants over trials to mean = 0 and SD = 1, *SNR* is a 3-level mean-centred ordinal variable describing change from digit SRT50, and *subject* and *digit* are categorical random effect grouping variables. The rationale for standardising EEG data within trials is two-fold. First, to remove baseline individual differences in the expression of α power and second to ensure that power changes over time are relative to the mean expression of power over the course of the trial. Second, we aimed at removing potential drifts in the expression of α power over the course of the experiment related to fatigue and learning which may confound the association between EEG measures and comprehension. Thus, the variable *α power*, standardized in this manner, does not reflect absolute power expression, but rather fluctuations around its own within-trial mean. Therefore, the regression models test whether relatively enhanced or decreased α power prior to the onset of the digit, as compared to its mean expression over the whole SSN-filled pre-stimulus period analysed (i.e., between -3395 and 0ms), in each channel and time bin, predicts a higher or lower probability of correctly comprehending the forthcoming DiN item. Similarly, in order to test for effects of pre-target α power on the perceived clarity of DiN items, we ran the following mixed-effect ordinal regression model (cumulative link mixed model) at each cortical parcel and each time-bin: *clarity ~ α power \* SNR + (1|subject:digit)*, where *clarity* represents the mean-centred subjective clarity rating.

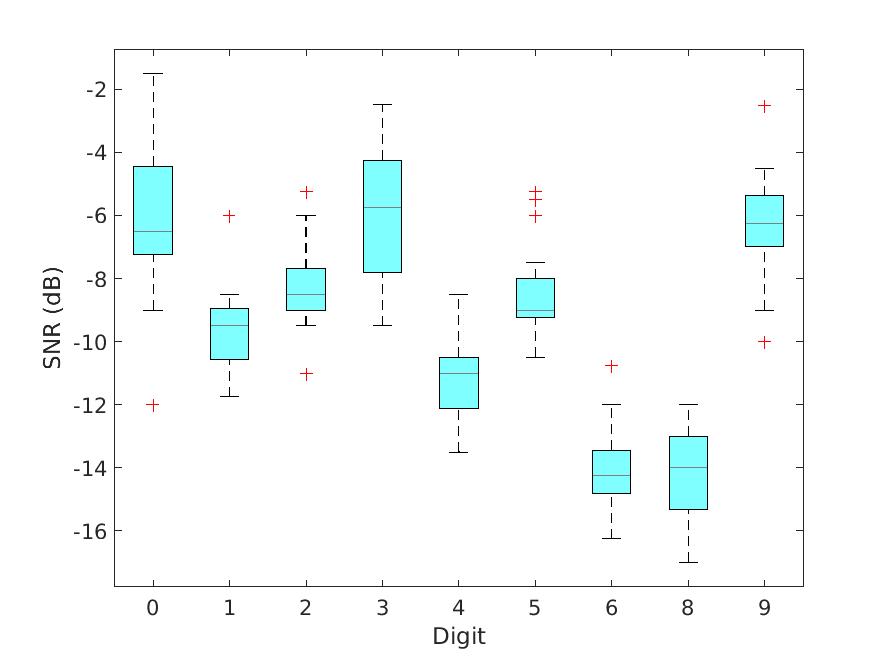
In order to select regression models, we ran several models in sensor space, averaged AIC values over sensors and time bins, and selected the one with the lowest mean Akaike Information Criterion (AIC; Table ST1). Regression analyses were carried on R version 4.1.0. Logistic regression models were run with the *lme4* package (Bates et al., 2015) version 1.1-27. Ordinal regression models without random effects were run with the *MASS* (Venables & Ripley, 2002)package version 7.3-54, whilst ordinal regression models including random effects were run with the *ordinal* package (Christensen, 2018) version 2019.12-10.

To assess the significance of each fixed factor, in addition to the full models, we ran restricted models thereof in which one model term at a time (*α power,* *SNR,* and *α power* by *SNR* interaction) is removed. The log-likelihood of each model is then computed and the test statistic describing the effect of each predictor is determined through the likelihood ratio test (LRT; Birkes, 2005). The test statistic follows a *χ2* distribution with degrees of freedom equal to the difference in the number of free parameters between the complex model and the nested model. This allows for the possibility to calculate the *p*-value describing the significance level of each predictor. Post-hoc comparison of the estimated marginal trends was run with the *emmeans* package (Lenth, 2021) version 1.6.2-1.

**Cluster-based permutation testing** was carried out following the procedure outlined by Maris & Oostenveld (2007). For each time point and cortical parcel, we fitted the regression models selected. For the predictor α power and its interaction with SNR, a set of “cluster statistics” was determined on the basis of all parcels whose p-value for the term under investigation was below 0.01. The cluster statistic was computed as the sum of the predictor’s *χ2* values of parcels that were contiguous in both time and space. Parcels were considered as neighbours if they shared at least one connection among their respective source points (determining spatial clusters). Contiguity in time was defined as spatial clusters in adjacent time bins sharing at least one parcel. Parcels were grouped separately for the positive and negative main effects of α power. Similarly, they were grouped separately for α power \* SNR interaction effects showing positive and negative differences among the marginal means of low and high degradation trials, indicating a relatively beneficial and detrimental effect of α power on the (objective or subjective) comprehension of less compared to more degraded stimuli, respectively. The cluster statistics were then tested against a null distribution of cluster statistics calculated in exactly the same way over 500 permutations of the data. We report clusters whose cluster statistic fell in bottom or top 2.5% of the null distribution (yielding a two-sided test with α = 0.05).

**Results:**

**Results of calibration procedure.** The dB SNR levels associated with the digits’ SRT50 are listed in Table ST2 and displayed in Fig.3.



**Figure 3. SRT50 of the digits.** Boxplots display the distribution of individual digit SRT50s. The red line represents the median, the lower limit of the body represents the 1st quartile. The upper limit of the body represents the 3rd quartile. Whiskers represent 1.5\* the interquartile range, red crosses represent outliers, defined as data points whose absolute values exceed 1.5\*IQR.

**Behavioural results.** We assessed the effect of a number of factors on behavioural outcomes by means of a series of repeated-measures Analysis of Variance (ANOVAs), described in Tables 1 and 2. Both individual mean comprehension and clarity significantly differed among SNR levels (Tables 1-2, Fig.4A-B). To ensure that the calibration procedure resulted in the same performance levels across participants and digits, we also evaluated their effect on individual comprehension and clarity scores. Despite the extensive effort to calibrate stimuli such that performance across digits would be equal, some significant differences remained: whilst comprehension did not differ significantly among digits, the digit\*SNR interaction on comprehension was found to be significant (Fig.S1A). In other words, a positive or negative change in SNR by 1.5dB with respect to the SRT50 did not result in an equal performance change across digits. This may relate to the highly different SNRs characterising the digits’ SRT50 (Fig.3). Perceived clarity, instead, differed significantly depending on both digit and digit\*SNR (Fig.S1B).

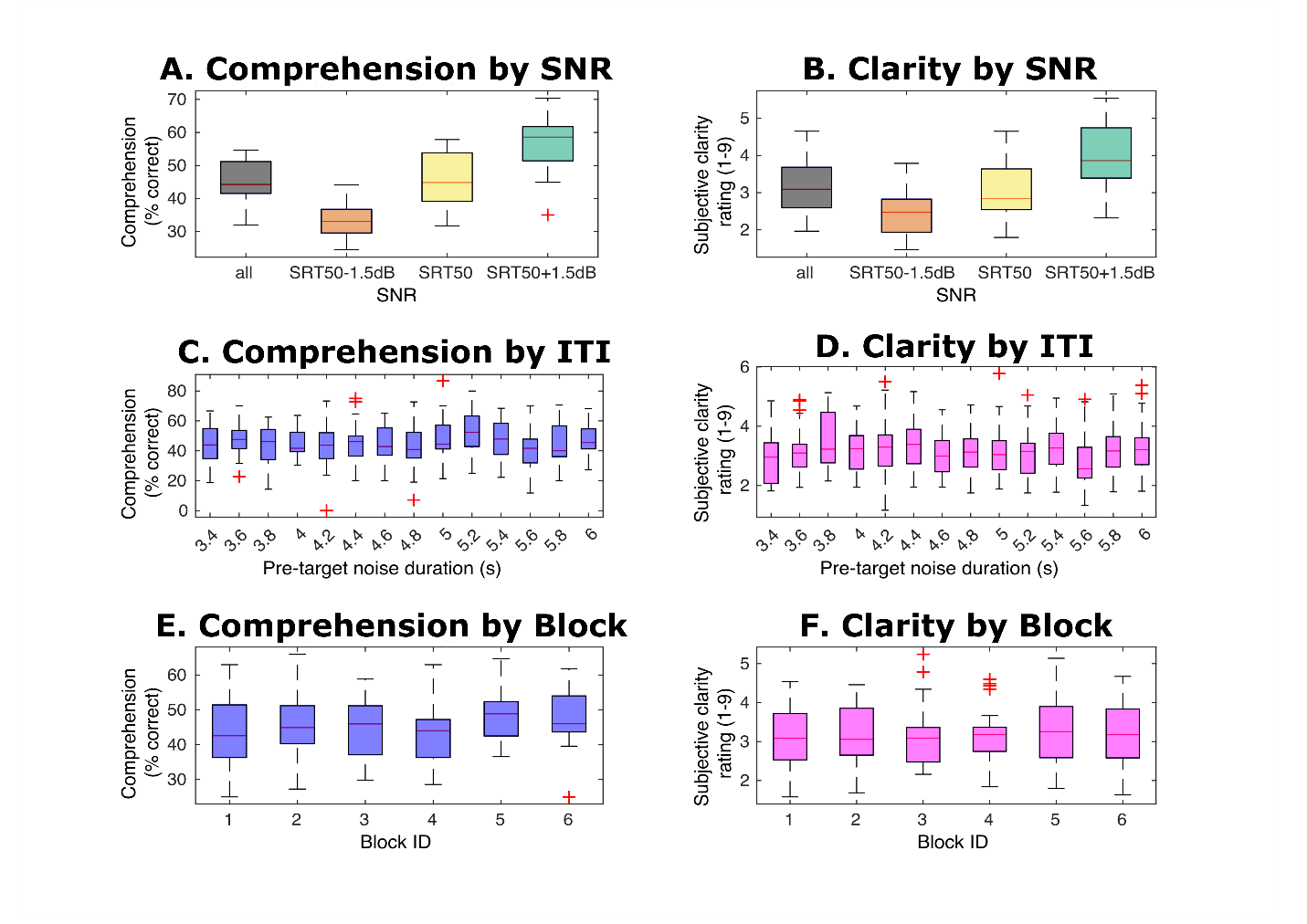
We further tested whether the differences varied over participants. For both comprehension and clarity, the strongest effect size was associated with the subject by digit interaction, in line with the results of the model selection (Table ST2). In order to rule out any influence of anticipation on performance, we further compared subject mean comprehension and clarity over pre-target noise duration. The analysis indicated that comprehension did not depend upon pre-target noise duration, but clarity did (Fig.4C-D). Further, experimental block did not predict comprehension nor clarity, thereby ruling out any significant learning taking place over blocks (Fig.4E-F). Finally, we assessed the association between the outcome measures *comprehension* and *clarity*. The probability of correctly comprehending the digit is 22.56%, 32.17%, 43.57%, 55.70%, 67.19%, 76.93%, 84.45%, 89.84%, and 93.51%, when it is given a subjective clarity rating score of 1, 2, 3, 4, 5, 6, 7, 8, and 9, respectively (Fig.S1C). The Wald test confirmed that the positive association between clarity and comprehension is significant (*χ2(8)* = 1260.9, p < .001).

**Table 1.** ANOVA table comparing comprehension across SNRs, participants, and digits, and two-way interactions thereof.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Factor** | ***df*** | **Sum of squares** | **Mean square** | ***F*** | ***p*** | ***η*2*p*** |
| **SNR** | 1,2 | 39466.068 | 19733.034 | 88.190 | < 0.001 | 0.838 |
| **Digit** | 1,8 | 2934.493 | 366.812 | 0.402 | 0.918 | 0.023 |
| **SNR\*digit** | 2,16 | 7830.961 | 489.435 | 2.421 | < 0.01 | 0.125 |
| **Noise duration** | 1,13 | 2943.784 | 226.445 | 1.620 | 0.078 | 0.063 |
| **Block** | 1,5 | 468.022 | 93.604 | 1.854 | 0.108 | 0.075 |

**Table 2.** ANOVA table comparing subjective clarity ratings across SNRs, participants, and digits, and two-way interactions thereof.

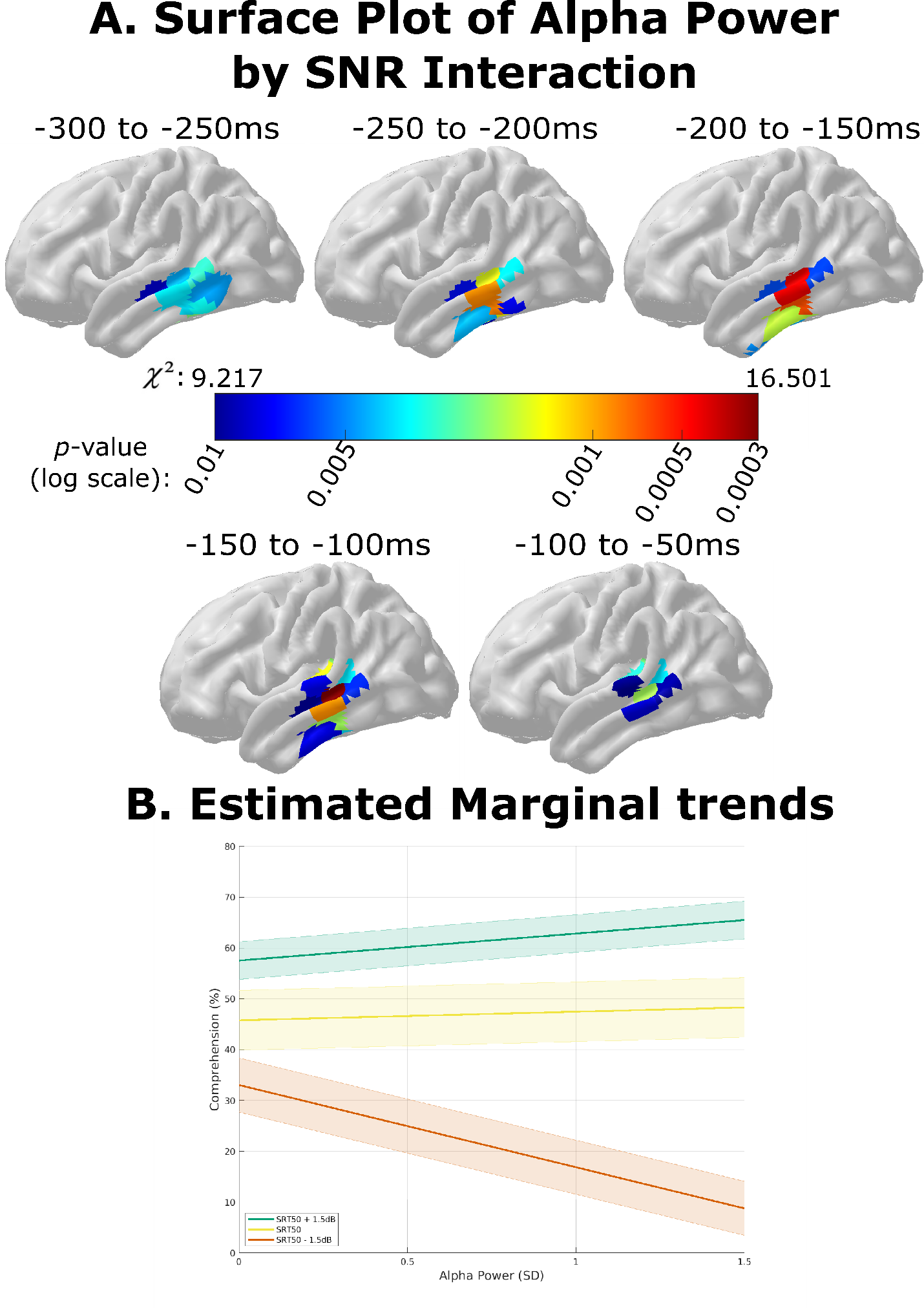
|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Factor** | ***df*** | **Sum of squares** | **Mean square** | ***F*** | ***p*** | ***η*2*p*** |
| **SNR** | 1,2 | 166.477 | 83.239 | 193.820 | < 0.001 | 0.919 |
| **Digit** | 1,8 | 147.685 | 18.461 | 7.441 | < 0.001 | 0.304 |
| **SNR\*digit** | 2,16 | 15.849 | 0.991 | 5.115 | < 0.001 | 0.169 |
| **Noise duration** | 1,13 | 7.424 | 0.571 | 2.324 | < 0.01 | 0.029 |
| **Block** | 1,5 | 0.574 | 0.115 | 0.838 | 0.526 | 0.089 |



**Figure 4.** Summary of behaviour: effect of SNR, pre-target noise duration (ITI), and block on comprehension and clarity.Boxplots display the distribution of individual behavioural scores. The red line represents the median, the lower limit of the body represents the 1st quartile (Q1), or 25th percentile. The upper limit of the body represents the 3rd quartile (Q3), or 75th percentile. The upper limit of the body represents the 3rd quartile. Whiskers represent 1.5\* the interquartile range, red crosses represent outliers, defined as data points whose absolute values exceed 1.5\*IQR. **(A)** Comprehension at the different SNRs. **(B)** Clarity at the different SNRs. **(C)** Comprehension at each target onset with respect to noise onset (ITI). **(D)** Clarity at each target onset. **(E)** Comprehension at each block. **(F)** Clarity at each block.

**Relationship between α power, SNR, and subjective clarity rating.** We found no significant main effect of α power on subjective clarity rating nor interaction with SNR.

**Relationship between α power, SNR, and comprehension.** We observed a significant (*p* < .05) interaction between α power and SNR in in a left temporo-parietal cluster lasting from at least -300ms to -50ms such that higher α power is associated with an increased probability of comprehending less difficult items and with a decreased probability of comprehending more difficult ones (Fig.5A). Post hoc comparisons revealed that the estimated marginal trends (EMT) for the effect of α power was .075 ± .0351 on the comprehension of lower SNR items, -.013 ± .060 on the comprehension of mid difficulty items, and -.122 ± .058 (*M* ± *SD* over parcels constituting the cluster) on the comprehension of high difficulty ones (Fig.5B). In other words, 1 *SD* increase in α power over the cluster is associated with a 7.5% increase in the probability of correctly comprehending a higher SNR item, and a -1.3% and 12.2% decrease in the probability of correctly comprehending a mid and lower SNR item, respectively. The peak effect is found between -150ms and -100ms in a posterior section of the left middle temporal gyrus. To test whether the set of time bin- and parcel-wise trends at each degradation level as a whole is significant, we ran two-sided one-sample nonparametric Wilcoxon signed rank tests. The tests revealed that none of the set of trends has zero median, representing an absence of an effect of α power on comprehension (low degradation: *Z* = 8.3739, *p* < .001; mid degradation: *Z* = -2.109, *p* = .035, high degradation: *Z* = -8.1326, *p* < .001).



**Figure 5.** EEG results. **(A).** Surface plot (left hemisphere) of parcel *p*-value and *χ2* statistic associated with the α power \* SNR interaction effect within the significant cluster. **(B)** Estimated marginal trends for the effect of α power on the comprehension at each of the three SNR levels (*M*±*SD* over cluster’s parcels).

**Discussion:**

The current study shows that pre-stimulus brain states, reflected in EEG power, affect the recognition of words presented in noise. The results of this study add to previous observations that pre-stimulus brain states influence the perception of speech material, including discriminating among ambiguous syllables (ten Oever & Sack, 2015) and deciding whether verbal stimuli presented in noise are words or pseudo-words (Strauß et al., 2015). To the best of our knowledge, this is the first study in which a property of the ongoing EEG, namely α power, was shown to be able to predict word recognition in noise. Interestingly, a clear interaction of α power with the degradation level of the stimuli was observed: increased left temporal α (8.5-12Hz) power during the noisy pre-target period increases the probability of correctly recognising a forthcoming Digit-in-Noise (DiN) stimulus at higher SNRs, and decreases the probability of recognising the stimulus at lower SNRs. Specifically, 1 *SD* increase in α power with respect to mean power over the duration of the prestimulus period (-3395 to 0ms) within the significant cluster is associated with a 5.3% increased probability of correctly comprehending a high SNR item, and a 16.2% decrease in the probability of comprehending a low SNR item. Our results further underscore the importance of α activity in the processing of forthcoming stimuli, even for complex stimuli such as spoken words embedded in noise, and at intensity levels far away from sensory thresholds.

Mean differences over participants and EEG sensors between correct and incorrect and between clear and unclear DiN trials, were associated with α increase during the pre-target, noise-present period. This observation is in line with a report indicating that higher pre-stimulus α predicts improved performance in a DiN recognition task (Alhanbali et al., 2021). In this study, α power was computed at two points in time: during a silent period before sound onset, and at a later point in time (several seconds thereafter) following the onset of noise (but before the onset of the target digits). Both EEG measures were significantly associated with performance and with each other, although only the first measure - α power during the silent baseline - was predictive of individual differences in performance after accounting for the (strong) association between participants’ hearing thresholds (pure tone audiometry) and DiN recognition. Our results parallel these findings in that individual differences in spatially-unresolved α power predict performance in DiN recognition such that higher α is associated with improved performance. Further, this effect is sustained over time, lasting hundreds of milliseconds up to seconds. Given the results reported by Alhanbali and colleagues, this effect might not be specific to the processing of noise but may rather reflect general attentional processes. This interpretation would be in line with a study showing that, compared to passive listening, attentive listening is related to α synchronisation during a DiN recognition task (Dimitrijevic et al., 2017). Nevertheless, while at this coarse and spatially unresolved level, our results are in line with these previous observations, by taking into account individual differences in performance and in the expression of α power, our study adds a whole new perspective to the role of α power: if at low SNRs it is beneficial, at high SNRs the functional benefit flips and becomes detrimental to performance.

The effect of within-subject trial-to-trial α power on comprehension was observed in left-lateralised posterior temporal cortical regions crucial to auditory processing and speech perception. These regions include the posterior superior temporal gyrus, involved in the earliest stages of cortical speech processing (spectrotemporal analysis), and the posterior superior temporal sulcus (pSTS), involved in phonological processing. However, the peak effect was found in the left posterior middle temporal gyrus (pMTG) between 150ms and 100ms preceding target onset. According to the ‘dual-stream model’ of speech processing (Hickok & Poeppel, 2007), the pMTG, receiving input from the pSTS, is involved in lexical access, or the mapping from sounds (phonological representations) to interpretable (semantic) representations. Collectively, these regions are responsible for the cascade of processes from the encoding of speech sounds to semantic access. Fluctuations of α power over time may thus represent variability in a brain state that reflects changes in sensitivity to the set of conceptually simpler (spectral) to more complex (spectrotemporal modulation) features of speech that constitute the basis for auditory object formation, referred to as “channels of auditory information” by Strauß, Wöstmann, & Obleser (2014)*.* Some of the features that channels are sensitive to, particularly frequency content, are common between speech-shaped noise (SSN) and speech. Instead, spectrotemporal features such as coherent spectral changes in various frequency bands over time are typical of the phonological content of speech and are not found in SSN. The model by Strauß and colleagues suggests that features of noise and speech are processed as much as possible in separate processing channels. However, the features cannot be easily separated at the low SNRs used in this study. Therefore, one attempted solution is to pre-emptively suppress all the channels in which there is noise.

According to the ‘gating by inhibition’ hypothesis, α activity reflects the suppression of information flow to task-irrelevant regions, allowing individuals to stay on task (Foxe & Snyder, 2011; Jensen & Mazaheri, 2010). Neurophysiologically, the gating by inhibition hypothesis has found support in studies showing that both α power and phase are associated with modulation of cortical excitability (Haegens et al., 2011; Iemi et al., 2021). Lower α power is associated with increased neuronal firing rate. It is thought that α activity’s neurophysiological role is consistent across different brain areas and sensory modalities, although the functional consequences of such α power-associated modulation of cortical excitability depends on the specific sources of α activity in the brain (Schneider et al., 2021).

Further support for the inhibitory role of α activity in auditory processing comes from studies on auditory spatial attention. An extensive series of studies has shown that focusing of attention to one location in space increases α activity in parietal and temporal (sensory) regions in the hemisphere contralateral to the attended hemispace and, vice versa, decreases α activity in the ipsilateral hemisphere (Tune et al., 2018; Wöstmann et al., 2016, 2019, 2021). Increased ipsilateral α power is thought to reflect inhibition of cortical activity in the hemisphere processing the unattended hemispace. By contrast, decreased contralateral α power is thought to represent disinhibition of cortical activity in the hemisphere processing the attended hemispace. Moreover, growing body of evidence indicates that, in addition to operating a spatial filter, α activity is also temporally modulated as a function of task demands, rendering it a potential implementation of spatio-temporal filtering in the auditory system. In one study (Wöstmann et al., 2016), participants were presented with a to-be-attended stream of spoken numbers to one ear and with a tightly synchronized to-be-ignored stream of numbers to the other ear. Spatial attention induced hemispheric lateralization of MEG-assessed α power in parietal and auditory regions, in line with previous findings. Interestingly, however, it was noticed that such power lateralization is temporally-specific: rather than being steadily present, it fluctuates synchronously with speech rate (the presentation rate of the numbers), and stronger modulation at the speech rate predicts increased comprehension. Further evidence for the idea that α activity operates as a spatio-temporal filter comes from a study involving a similar spatial attention paradigm described above (Wöstmann et al., 2021). This time, however, a visual cue was presented before the digit streams indicating the position of the to-be-identified number along the to-be-attended stream (e.g., 4th number). The cue was either instructive and indicated the position of the to-be-identified number (the cue was ‘valid’ 70% of the times, as it corresponded to the correct position, and invalid 30% of the times) or neutral and did not provide information regarding the position of the to-be-identified number. Once again, spatial attention was found to induce MEG α power lateralization, and this result underscores the sensitivity of α to spatial attention. Further, temporal modulation of α lateralization at the speech rate was stronger following instructive compared to neutral temporal cues, and this result underscores the sensitivity of α to temporal attention. Finally, the modulation was strongest specifically at the onsets of temporally cued numbers, and this result highlights the temporal specificity of α activity. All in all, these results speak for the idea that α power can act as a spatio-temporal filter susceptible to top-down information.

We hypothesize that high α activity in the context of this experiment indexes inhibition (potentially manifest as down-modulation of sensory gain) in auditory channels. Such a mechanism can serve to suppress noise in the pre-stimulus period, and presumably also during the (temporally non-predictable) target. We speculate that such a suppression mechanism can account for worsened performance under low signal to noise ratios, and improved performance at higher SNRs. We propose that inhibition of auditory channels results in improved performance when enough sensory evidence remains after input suppression (i.e., at high SNRs), but is systematically deleterious for performance when suppression results in insufficient residual sensory evidence (i.e., at low SNRs) to reliably activate target representations. We speculate that the addition of noise and speech without RMS-equalisation, resulting in substantially higher sound levels during target presentation as compared to before target presentation, might reduce, if not completely eliminate, the negative impact of high α activity on the recognition of low SNR items. Further studies are needed to compare the effects of α power under these different stationary noise conditions. It would be interesting to further probe the generalisability of the effect observed here to different SNRs, different kinds of noise, such as babble or amplitude-modulated noise, and noise characterised by different degrees of overlap in its frequency content with respect to that of speech. Comparing the effects of pre-target α activity under these different circumstances would allow investigation of whether its modulation is a universally useful mechanism or just one tool of many.

Despite best efforts, the current study has a number of inevitable limitations. First, we have opted to use a task involving only isolated (monosyllabic) words, thereby reducing the effect of individually-variable linguistic competence on comprehension (Kaandorp et al., 2016), but consequently eliminating any potential insight into individual differences stemming from higher-level linguistic processes, including contextual effects allowing for prediction, which are highly relevant to comprehension of acoustically challenging speech (Obleser & Kotz, 2010; Rysop et al., 2021)**.** Whilst the use of a low-perplexity stimulus set allowed participants to strategically exploit top-down generated expectations, this implies that we cannot establish the specificity of our findings with respect to: (i) speech, since we have no non-speech auditory stimuli; (ii) non-digit words, or open stimulus sets; (iii) connected speech. Further studies using more elaborate sets of stimuli are required to determine whether the effect of α power holds true under more ecological conditions. Additionally, we could not achieve calibration of the intelligibility of the stimuli at the individual participant and digit level. Despite having capitalized on the possibility to model these effects in the current study, simpler models may still be preferable. Thus, future studies should improve on the calibration procedure used here to establish more stable performance levels.

In sum, this study provides further evidence that pre-stimulus brain states provide a behaviourally relevant structuring context for the processing of stimuli. Although pre-stimulus modulation of behaviour constitutes an established finding within the psychophysical literature, especially within the context of detection or discrimination of near-threshold stimuli, to the best of our knowledge, this is the first study demonstrating effects of pre-target brain states on a speech recognition task. The results of this study underscore the importance of α activity in the processing of forthcoming stimuli. The findings reported above represent a stepping-stone in the development of non-invasive interventions based on modulation of ongoing EEG activity, such as transcranial electrical stimulation or neurofeedback, aimed at restoring sub-optimal speech perception in people suffering from mild speech perception deficits. Such application should take into account our observation that α enhancement (or suppression) may not be always beneficial for speech recognition. Rather, interventions should aim at having listeners express just the right amount α activity given the listening conditions and their ability to cope with them. In other words, the results of this study highlight the importance of tailoring any such intervention to both the beneficiary and the environment.

**References:**

Alhanbali, S., Munro, K. J., Dawes, P., Perugia, E., & Millman, R. E. (2021). Associations between pre-stimulus alpha power, hearing level and performance in a digits-in-noise task. *International Journal of Audiology*, *0*(0), 1–8. https://doi.org/10.1080/14992027.2021.1899314

Bates, D., Mächler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1). https://doi.org/10.18637/jss.v067.i01

Becker, R., Reinacher, M., Freyer, F., Villringer, A., & Ritter, P. (2011). How Ongoing Neuronal Oscillations Account for Evoked fMRI Variability. *Journal of Neuroscience*, *31*(30), 11016–11027. https://doi.org/10.1523/JNEUROSCI.0210-11.2011

Becker, R., Ritter, P., & Villringer, A. (2008). Influence of ongoing alpha rhythm on the visual evoked potential. *NeuroImage*, *39*(2), 707–716. https://doi.org/10.1016/j.neuroimage.2007.09.016

Birkes, D. (2005). Likelihood Ratio Tests. In *Encyclopedia of Biostatistics*. https://doi.org/https://doi.org/10.1002/0470011815.b2a15074

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*(4), 433–436.

Busch, N. A., Dubois, J., & VanRullen, R. (2009). The phase of ongoing EEG oscillations predicts visual perception. *The Journal of Neuroscience*, *29*(24), 7869–7876. https://doi.org/10.1523/JNEUROSCI.0113-09.2009

Buzsáki, G. (2019). Brain Rhythms Provide a Framework for Neural Syntax. In *The Brain from Inside Out*. Oxford University Press. https://doi.org/10.1093/oso/9780190905385.003.0006

Chan, D., Fourcin, A., Gibbon, D., Grandstrom, B., Huckvale, M., Kokkinakis, G., Kvale, K., Lamel, L., Lindberg, B., Moreno, A., Mouropoulos, J., Senia, F., Trancoso, I., Veld, C. in’T, & Zeiliger, J. (1995). EUROM - A spoken language resource for the EU. In *European Conference on Speech Communication and Technology* (pp. 867–870). http://www.phon.ucl.ac.uk/resource/eurom1/eurospeech95eurom.pdf

Christensen, R. (2018). *Cumulative Link Models for Ordinal Regression with the R Package ordinal*.

De Blasio, F. M., & Barry, R. J. (2013). Prestimulus alpha and beta determinants of ERP responses in the Go/NoGo task. *International Journal of Psychophysiology*, *89*(1), 9–17. https://doi.org/10.1016/j.ijpsycho.2013.04.018

De Blasio, F. M., & Barry, R. J. (2018). Prestimulus delta and theta contributions to equiprobable Go/NoGo processing in healthy ageing. *International Journal of Psychophysiology*, *130*(April), 40–52. https://doi.org/10.1016/j.ijpsycho.2018.05.005

De Blasio, F. M., Barry, R. J., & Steiner, G. Z. (2013). Prestimulus EEG amplitude determinants of ERP responses in a habituation paradigm. *International Journal of Psychophysiology*, *89*(3), 444–450. https://doi.org/10.1016/j.ijpsycho.2013.05.015

Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, *134*, 9–21.

Dimitrijevic, A., Smith, M. L., Kadis, D. S., & Moore, D. R. (2017). Cortical alpha oscillations predict speech intelligibility. *Frontiers in Human Neuroscience*, *11*(February), 1–10. https://doi.org/10.3389/fnhum.2017.00088

Foxe, J. J., & Snyder, A. C. (2011). The role of alpha-band brain oscillations as a sensory suppression mechanism during selective attention. *Frontiers in Psychology*, *2*(JUL), 1–13. https://doi.org/10.3389/fpsyg.2011.00154

Gramfort, A., Papadopoulo, T., Olivi, E., & Clerc, M. (2010). OpenMEEG: opensource software for quasistatic bioelectromagnetics. *BioMedical Engineering OnLine*, *9*(1), 45. https://doi.org/10.1186/1475-925X-9-45

Haegens, S., Nácher, V., Luna, R., Romo, R., & Jensen, O. (2011). α-Oscillations in the monkey sensorimotor network influence discrimination performance by rhythmical inhibition of neuronal spiking. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(48), 19377–19382. https://doi.org/10.1073/pnas.1117190108

Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nat Rev Neurosci*, *8*(393), 393–402. https://doi.org/10.7554/eLife.14521

Houweling, T., Becker, R., & Hervais-Adelman, A. (2020). The noise-resilient brain: Resting-state oscillatory activity predicts words-in-noise recognition. *Brain and Language*, *202*(104727), 1–9. https://doi.org/10.1016/j.bandl.2019.104727

Iemi, L., Gwilliams, L., Samaha, J., Auksztulewicz, R., Cycowicz, Y. M., King, J.-R., Nikulin, V. V, Thesen, T., Doyle, W., Devinsky, O., Schroeder, C. E., Melloni, L., & Haegens, S. (2021). Ongoing neural oscillations influence behavior and sensory representations by suppressing neuronal excitability. *NeuroImage*, 118746. https://doi.org/10.1016/j.neuroimage.2021.118746

Jensen, O., & Mazaheri, A. (2010). Shaping functional architecture by oscillatory alpha activity: Gating by inhibition. *Frontiers in Human Neuroscience*, *4*(November), 1–8. https://doi.org/10.3389/fnhum.2010.00186

Kaandorp, M. W., De Groot, A. M. B., Festen, J. M., Smits, C., & Goverts, S. T. (2016). The influence of lexical-access ability and vocabulary knowledge on measures of speech recognition in noise. *International Journal of Audiology*, *55*(3), 157–167. https://doi.org/10.3109/14992027.2015.1104735

Kayser, S. J., McNair, S. W., & Kayser, C. (2016). Prestimulus influences on auditory perception from sensory representations and decision processes. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(17), 4842–4847. https://doi.org/10.1073/pnas.1524087113

Kybic, J., Clerc, M., Faugeras, O., Keriven, R., & Papadopoulo, T. (2006). Generalized head models for MEG/EEG: Boundary element method beyond nested volumes. *Physics in Medicine and Biology*, *51*(5), 1333–1346. https://doi.org/10.1088/0031-9155/51/5/021

Lange, J., Oostenveld, R., & Fries, P. (2013). Reduced occipital alpha power indexes enhanced excitability rather than improved visual perception. *Journal of Neuroscience*, *33*(7), 3212–3220. https://doi.org/10.1523/JNEUROSCI.3755-12.2013

Lenth, R. V. (2021). *emmeans: Estimated Marginal Means, aka Least-Squares Means*. https://cran.r-project.org/package=emmeans

Limbach, K., & Corballis, P. M. (2016). Prestimulus alpha power influences response criterion in a detection task. *Psychophysiology*, *53*(8), 1154–1164. https://doi.org/10.1111/psyp.12666

Makeig, S., Bell, A. J., Jung, T.-P., & Sejnowski, T. J. (1996). Independent Component Analysis of Electroencephalographic Data. *Advances in Neural Information Processing Systems*, *8*, 145–151. https://papers.nips.cc/paper/1091-independent-component-analysis-of-electroencephalographic-data.pdf

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, *164*(1), 177–190. https://doi.org/10.1016/j.jneumeth.2007.03.024

Martín-Buro, M. C., Garcés, P., & Maestú, F. (2016). Test-retest reliability of resting-state magnetoencephalography power in sensor and source space. *Human Brain Mapping*, *37*(1), 179–190. https://doi.org/10.1002/hbm.23027

Mathewson, K. E., Gratton, G., Fabiani, M., Beck, D. M., & Ro, T. (2009). To See or Not to See: Prestimulus   Phase Predicts Visual Awareness. *Journal of Neuroscience*, *29*(9), 2725–2732. https://doi.org/10.1523/JNEUROSCI.3963-08.2009

McNair, S. W., Kayser, S. J., & Kayser, C. (2019). Consistent pre-stimulus influences on auditory perception across the lifespan. *NeuroImage*, *186*(November 2018), 22–32. https://doi.org/10.1016/j.neuroimage.2018.10.085

Mitra, P., & Bokil, H. (2007). *Observed Brain Dynamics*. Oxford University Press.

Mullen, T. R., Kothe, C. A. E., Chi, Y. M., Ojeda, A., Kerth, T., Makeig, S., Jung, T.-P., & Cauwenberghs, G. (2015). Real-time neuroimaging and cognitive monitoring using wearable dry EEG. *IEEE Transactions on Biomedical Engineering*, *62*(11), 2553–2567. https://doi.org/10.1109/TBME.2015.2481482

Mullen, T. R., Kothe, C., Chi, Y. M., Ojeda, A., Kerth, T., Makeig, S., Cauwenberghs, G., & Tzyy-Ping Jung. (2013). Real-time modeling and 3D visualization of source dynamics and connectivity using wearable EEG. *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, *23*(1), 2184–2187. https://doi.org/10.1109/EMBC.2013.6609968

Obleser, J., & Kotz, S. A. (2010). Expectancy constraints in degraded speech modulate the language comprehension network. *Cerebral Cortex*, *20*(3), 633–640. https://doi.org/10.1093/cercor/bhp128

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J. M. (2011). FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational Intelligence and Neuroscience*, *2011*. https://doi.org/10.1155/2011/156869

Rajagovindan, R., & Ding, M. (2011). From prestimulus alpha oscillation to visual-evoked response: An inverted-U function and its attentional modulation. *Journal of Cognitive Neuroscience*, *23*(6), 1379–1394. https://doi.org/10.1162/jocn.2010.21478

Rysop, A. U., Schmitt, L. M., Obleser, J., & Hartwigsen, G. (2021). Neural modelling of the semantic predictability gain under challenging listening conditions. *Human Brain Mapping*, *42*(1), 110–127. https://doi.org/10.1002/hbm.25208

Samaha, J., Iemi, L., & Postle, B. R. (2017). Prestimulus alpha-band power biases visual discrimination confidence, but not accuracy. *Consciousness and Cognition*, *54*, 47–55. https://doi.org/10.1016/j.concog.2017.02.005

Schneider, D., Herbst, S. K., Klatt, L. I., & Wöstmann, M. (2021). Target enhancement or distractor suppression? Functionally distinct alpha oscillations form the basis of attention. *European Journal of Neuroscience*, *April*, 1–10. https://doi.org/10.1111/ejn.15309

Sohoglu, E., Peelle, J. E., Carlyon, R. P., & Davis, M. H. (2014). Top-down influences of written text on perceived clarity of degraded speech. *Journal of Experimental Psychology: Human Perception and Performance*, *40*(1), 186–199. https://doi.org/10.1037/a0033206

Strauß, A., Henry, M. J., Scharinger, M., & Obleser, J. (2015). Alpha phase determines successful lexical decision in noise. *Journal of Neuroscience*, *35*(7), 3256–3262. https://doi.org/10.1523/JNEUROSCI.3357-14.2015

Strauß, A., Wöstmann, M., & Obleser, J. (2014). Cortical alpha oscillations as a tool for auditory selective inhibition. *Frontiers in Human Neuroscience*, *8*(MAY), 1–7. https://doi.org/10.3389/fnhum.2014.00350

ten Oever, S., & Sack, A. T. (2015). Oscillatory phase shapes syllable perception. *Proceedings of the National Academy of Sciences*, *112*(52), 15833–15837. https://doi.org/10.1073/pnas.1517519112

Timoney, J., Lysaght, T., & Schoenwiesner, M. (2004). Implementing loudness models in matlab. *Proc. of the 7th Int. Conference on Digital Audio Effects (DAFX-04)*, *1*, 5–9. http://eprints.nuim.ie/420/

Tune, S., Wöstmann, M., & Obleser, J. (2018). Probing the limits of alpha power lateralisation as a neural marker of selective attention in middle-aged and older listeners. *European Journal of Neuroscience*, *48*(7), 2537–2550. https://doi.org/10.1111/ejn.13862

van Es, M. W. J., & Schoffelen, J. M. (2019). Stimulus-induced gamma power predicts the amplitude of the subsequent visual evoked response. *NeuroImage*, *186*(September 2018), 703–712. https://doi.org/10.1016/j.neuroimage.2018.11.029

Van Essen, D. C., Glasser, M. F., Dierker, D. L., Harwell, J., & Coalson, T. (2012). Parcellations and hemispheric asymmetries of human cerebral cortex analyzed on surface-based atlases. *Cerebral Cortex*, *22*(10), 2241–2262. https://doi.org/10.1093/cercor/bhr291

Van Veen, B. D., Van Drongelen, W., Yuchtman, M., & Suzuki, A. (1997). Localization of brain electrical activity via linearly constrained minimum variance spatial filtering. *IEEE Transactions on Biomedical Engineering*, *44*(9), 867–880. https://doi.org/10.1109/10.623056

VanRullen, R., Busch, N. A., Drewes, J., & Dubois, J. (2011). Ongoing EEG phase as a trial-by-trial predictor of perceptual and attentional variability. *Frontiers in Psychology*, *2*(APR), 1–9. https://doi.org/10.3389/fpsyg.2011.00060

Venables, W. N., & Ripley, B. D. (2002). *Modern Applied Statistics with S* (Fourth). Springer. https://www.stats.ox.ac.uk/pub/MASS4/

Verhelst, W., & Roelands, M. (1993). Overlap-add technique based on waveform similarity (WSOLA) for high quality time-scale modification of speech. *Proceedings - ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing*, *2*, 554–557. https://doi.org/10.1109/icassp.1993.319366

Wilson, R. H., Abrams, H. B., & Pillion, A. L. (2003). A word-recognition task in multitalker babble using a descending presentation mode from 24 dB to 0 dB signal to babble. *Journal of Rehabilitation Research and Development*, *40*(4), 321–327.

Wilson, R. H., Carnell, C. S., & Cleghorn, A. L. (2007). The Words-in-Noise (WIN) test with multitalker babble and speech-spectrum noise maskers. *Journal of the American Academy of Audiology*, *18*(6), 522–529. https://doi.org/10.3766/jaaa.18.6.7

Wilson, R. H., & Mcardle, R. (2007). Intra- and Inter-session Test, Retest Reliability of the Words-in-Noise (WIN) Test. *Journal of the American Academy of Audiology*, *825*, 813–825.

Wöstmann, M., Alavash, M., & Obleser, J. (2019). Alpha oscillations in the human brain implement distractor suppression independent of target selection. *The Journal of Neuroscience*, *39*(49), 9797–9805. https://doi.org/10.1101/721019

Wöstmann, M., Herrmann, B., Maess, B., & Obleser, J. (2016). Spatiotemporal dynamics of auditory attention synchronize with speech. *Proceedings of the National Academy of Sciences*, *113*(14), 3873–3878. https://doi.org/10.1073/pnas.1523357113

Wöstmann, M., Maess, B., & Obleser, J. (2021). Orienting auditory attention in time: Lateralized alpha power reflects spatio-temporal filtering. *NeuroImage*, *228*(July 2020). https://doi.org/10.1016/j.neuroimage.2020.117711

Zazio, A., Ruhnau, P., Weisz, N., & Wutz, A. (2021). Pre-stimulus alpha-band power and phase fluctuations originate from different neural sources and exert distinct impact on stimulus-evoked responses. *European Journal of Neuroscience*, *July 2020*, 1–13. https://doi.org/10.1111/ejn.15138