Degradation levels of continuous speech affect neural speech tracking and alpha power differently

Hauswald, Anne; Keitel, Anne; Chen, Ya-Ping; Rösch, Sebastian; Weisz, Nathan

Published in:
European Journal of Neuroscience

DOI:
10.1111/ejn.14912

Publication date:
2020

Document Version
Publisher's PDF, also known as Version of record

Received: 14 April 2020 | Revised: 12 July 2020 | Accepted: 13 July 2020

DOI: 10.1111/ejn.14912

SPECIAL ISSUE

EJN European Journal of Neuroscience FENS WILEY

Degradation levels of continuous speech affect neural speech tracking and alpha power differently

Anne Hauswald[1,2] | **Anne Keitel**[3,4] | **Ya-Ping Chen**[1,2] | **Sebastian Rösch**[5] | **Nathan Weisz**[1,2]

[1]Center of Cognitive Neuroscience, University of Salzburg, Salzburg, Austria

[2]Department of Psychology, University of Salzburg, Salzburg, Austria

[3]Psychology, School of Social Sciences, University of Dundee, Dundee, UK

[4]Centre for Cognitive Neuroimaging, University of Glasgow, Glasgow, UK

[5]Department of Otorhinolaryngology, Paracelsus Medical University, Salzburg, Austria

Correspondence
Anne Hauswald, Centre for Cognitive Neuroscience, University of Salzburg, Hellbrunnerstraße 34, 5020 Salzburg, Austria.
Email: anne.hauswald@sbg.ac.at

Funding information
Austrian Science Fund, Grant/Award Number: P 31230; Wellcome Trust, Grant/Award Number: 204820/Z/16/Z

Abstract

Making sense of a poor auditory signal can pose a challenge. Previous attempts to quantify speech intelligibility in neural terms have usually focused on one of two measures, namely low-frequency speech-brain synchronization or alpha power modulations. However, reports have been mixed concerning the modulation of these measures, an issue aggravated by the fact that they have normally been studied separately. We present two MEG studies analyzing both measures. In study 1, participants listened to unimodal auditory speech with three different levels of degradation (original, 7-channel and 3-channel vocoding). Intelligibility declined with declining clarity, but speech was still intelligible to some extent even for the lowest clarity level (3-channel vocoding). Low-frequency (1–7 Hz) speech tracking suggested a U-shaped relationship with strongest effects for the medium-degraded speech (7-channel) in bilateral auditory and left frontal regions. To follow up on this finding, we implemented three additional vocoding levels (5-channel, 2-channel and 1-channel) in a second MEG study. Using this wider range of degradation, the speech-brain synchronization showed a similar pattern as in study 1, but further showed that when speech becomes unintelligible, synchronization declines again. The relationship differed for alpha power, which continued to decrease across vocoding levels reaching a floor effect for 5-channel vocoding. Predicting subjective intelligibility based on models either combining both measures or each measure alone showed superiority of the combined model. Our findings underline that speech tracking and alpha power are modified differently by the degree of degradation of continuous speech but together contribute to the subjective speech understanding.

KEYWORDS

alpha power, continuous speech, degraded speech, low-frequency speech tracking, MEG

Abbreviations: ANOVA, analysis of variance; CI, cochlear implant; EEG, electroencephalography; FDR, false discovery rate; fMRI, functional magnetic resonance imaging; MEG, magnetoencephalography; MNI, Montreal Neurological Institute; MRI, magnetic resonance image; *SD*, standard deviation; TRF, temporal response function.

Edited by Dr. Niko Busch

1 | INTRODUCTION

Understanding speech can be challenging in normal acoustic environments (e.g. background noise) or due to hearing damage. To compensate for the inferior signal-to-noise ratio of the acoustic information reaching the central auditory system, an effortful process is necessary. Indeed, subjective listening effort has been shown to increase with concurrent background noise or competing speakers for speech sounds with fewer acoustic details or lower predictiveness (Wöstmann, Herrmann, Wilsch, & Obleser, 2015). "Listening effort" however, from a conceptual perspective is not so straightforward, often describing a combination of cognitive demand (usually due to challenging listening situations) and affective–motivational aspects (Peelle, 2018). A related compensatory process may be the allocation of increased attentional resources to the incoming (behaviorally relevant) sounds (Wild et al., 2012). Interestingly, both the broader concept of listening effort and the selective attention have been linked to neural oscillations in the alpha (8–12 Hz) frequency range (e.g., Dimitrijevic, Smith, Kadis, & Moore, 2019; Frey et al., 2014; Obleser, Wöstmann, Hellbernd, Wilsch, & Maess, 2012). Most studies have reported an alpha power increase in response to degraded speech. This modulation occurs when short degraded stimuli are used (Obleser & Weisz, 2012; Obleser et al., 2012). However, in the rare—albeit more naturalistic—situation where sentences have been used, alpha power seems to show the opposite pattern (McMahon et al., 2016; Miles et al., 2017).

Besides the induced neuronal responses broadly linked to the task demands, listening to speech also elicits temporal synchronization of auditory cortical activity to the speech sound. Different frequency bands have been assigned to carry different information with regard to speech signal, with a dominance of delta (1–4 Hz) and theta frequencies (4–7 Hz) capturing phrasal and syllable structure, respectively (Greenberg, 1998; Poeppel, 2003). Synchronization between speech and brain signals is often called neural speech entrainment or speech tracking (see, however, Alexandrou, Saarinen, Kujala, & Salmelin, 2018). Different measures can be used for quantification such as coherence (Hauswald, Lithari, Collignon, Leonardelli, & Weisz, 2018), mutual information (Gross et al., 2013; Keitel, Gross, & Kayser, 2018), inter-trial correlation (Ding, Chatterjee, & Simon, 2014), dissimilarity index (Luo & Poeppel, 2007) or temporal response functions (TRF, Crosse, Di Liberto, Bednar, & Lalor, 2016; Ding & Simon, 2011). Just as alpha power, low-frequency speech tracking is modulated by degradation of the speech signal with studies providing mixed findings: Reduced synchronization in this frequency range is linked to reduced intelligibility either operationalized through vocoding (Ding et al., 2014; Luo & Poeppel, 2007), time-reversed presentation (Gross et al., 2013; Howard & Poeppel, 2010), speech in noise (Dimitrijevic et al., 2019) or transcranial electrical stimulation (Riecke, Formisano, Sorger, Başkent, & Gaudrain, 2018; Zoefel, Archer-Boyd, & Davis, 2018). However, using other measures or experimental procedures the opposite pattern has also been shown: For example, using TRF yields higher M50 and delta synchronization is enhanced for degraded stimuli compared with unaltered stimuli in quiet environment (Ding et al., 2014) and non-native speakers show higher speech entrainment than native speakers (Song & Iverson, 2018). Interestingly, the latter observation has also been linked to an increase of listening effort. To complicate things further, multi-speaker and auditory spatial attention studies using sentences or narratives have repeatedly found stronger speech tracking (delta and theta band) for attended compared with unattended speech (Ding & Simon, 2012; Rimmele, Zion Golumbic, Schröger, & Poeppel, 2015; Viswanathan et al. 2019) in auditory cortices and areas in the vicinity thereof (Horton, D'Zmura, & Srinivasan, 2013; Zion Golumbic et al., 2013).

Thus, both relevant measures—that is, speech tracking and alpha power—are frequently linked with similar concepts such as listening effort (Dimitrijevic et al., 2019; Song & Iverson, 2018), selective attention (Frey et al., 2014; Rimmele et al., 2015) or intelligibility of the stimuli (Vanthornhout, Decruy, & Francart, 2019). Despite this conceptual overlap, very few studies have investigated these measures simultaneously. One study, using a speech-in-noise task, reported decreasing speech tracking and increasing alpha power in response to increasing listening effort in cochlear implant users (Dimitrijevic et al., 2019). However, here again, short stimuli (digits) were presented, which is a paradigm rather remote from real-life listening situations and still leaves open the question how degraded continuous speech affects speech tracking and alpha power. We report findings from two MEG studies that together aim at answering this question. Therefore, we presented continuous speech with a wide range of degradation levels and analyzed both speech tracking and alpha power. Derived from the same data set, we show a differential modulation pattern of both measures: Speech tracking increases the stronger stimuli are degraded as long as some intelligibility is still warranted, to then decrease beyond this critical point. Alpha power on the other hand decreases with increased degradation and stays low even when unintelligible. Using linear mixed-effects models, we show that combining speech tracking and alpha power is superior in predicting subjective intelligibility of degraded speech, as compared to models based on one of the neural measures alone.

2 | Study 1

2.1 | Materials and methods

2.1.1 | Participants

Twenty-eight individuals participated in the study (female = 17, male = 11). Mean age was 23.82 years (standard deviation, *SD* = 3.712), with a range between 19 and 37 years. We recruited only German native speakers and people who were eligible for MEG recordings, that is, without nonremovable ferromagnetic metals in or close to the body. Participants provided informed consent and were compensated monetarily or via course credit. Participation was voluntary and in line with the declaration of Helsinki and the statutes of the University of Salzburg. The study was preregistered at OSF (https://osf.io/dpt34/). In the preregistration, we aimed at a sample size of 30–34 instead of the 28 we ended up with. This was due to a technical problem that occurred after an upgrade of the Vpixx stimulation software, after which the visual stimulation would freeze during presentation. We could not run the whole experiment in the same way as for the initial 28 subjects and therefore stopped after those 28. We also hoped to compare degradation across auditory and visual modalities, but realized that the degradation of auditory and visual modality was not comparable. Therefore, we focused on the auditory modality and added a follow-up study for this modality.

The study was approved by the ethical committee of the University of Salzburg.

2.1.2 | Stimuli

For the MEG recording, 12 audio files were created from audio–visual recordings of a female speaker reading Goethe's "Das Märchen" (1795). Stimulus lengths varied between approximately 30 s and 3 min, with two stimuli of 15, 30, 60, 90, 120 and 150 s, and 12 of 180 s. Each stimulus ended with a two-syllable noun within the last four words. In order to keep participants' attention on the stimulation, we asked participants after each stimulus to choose from two presented two-syllable nouns the one that had occurred within the last four words of a sentence. The syllable rate of the stimuli varied between 4.1 and 4.5 Hz with a mean of 4.3 Hz.

Noise-vocoding of all audio stimuli was done using the vocoder toolbox for MATLAB (Gaudrain, 2016), and we created conditions with 7 and 3 channels (Figure 1a). For the vocoding, the waveform of each audio stimulus was passed through two Butterworth analysis filters (for 7 and 3 channels) with a range of 200–7,000 Hz, representing equal distances along the basilar membrane. Amplitude envelope extraction was done with half-wave rectification and low-pass filtering at 250 Hz. The envelopes were then normalized in each channel and multiplied with the carrier. Then, they were filtered in the band, and the RMS of the resulting signal was adjusted to that of the original signal filtered in that same band. Auditory stimuli were presented binaurally using MEG-compatible pneumatic in-ear headphones (SOUNDPixx, VPixx technologies). The trigger-sound delay of 16 ms was measured (The Black Box Toolkit v2) and corrected for during preprocessing.

In the experiment, in addition to the unimodal auditory stimuli also unimodal visual stimuli were presented, which will not be reported or discussed here as the visual degradation manipulation was not comparable to the acoustic one. The unimodal stimuli were presented to the participants in three consecutive audio-only blocks and three consecutive video-only blocks via in-ear-phones and a projector system, respectively. The order of video and audio blocks was balanced. Each block contained 4 stimuli, which were presented either in an unaltered version or in one of the two degraded versions. The order of the stimuli was random and did not follow the order of the original story. The assignment of stimuli to conditions was controlled in order to obtain similar overall length of stimulus presentation (approx. 400 s) for each modality and degradation levels. We instructed participants to attend to the speech which they would either see or hear. In order to keep participants' attention on the stimulation, a behavioral response was required after each stimulus. At the end of each stimulus, a target and a distractor word would appear next to each other. The participants were asked to decide which of the words was presented as the last noun and within the last four words by pressing the button on the side of the response pad that matched the presentation side of the word they chose (Figure 1b). Presentation side of target and distractor words was random. Following the response, they could self-initiate the next trial via a button press. Each block was followed by a short self-determined break. This procedure resulted in only four responses per condition, and therefore, we added a behavioral experiment following all six blocks, to assess performance. Responses were acquired via a response pad (TOUCHPixx response box by VPixx Technologies).

For this additional behavioral experiment, we used a total of 24 unimodal audio stimuli of a different female speaker reading Antoiné St. Exupery's "The little prince" (1943). Each stimulus contained a single sentence (length between 2 and 15 s) with a two-syllable noun (target word) within the last four words. We created a list of different two-syllable nouns (distractor words), which we also drew from "The little prince" but were not presented during the stimulation. Similar to the main experiment, participants had to choose between two alternatives and the chance level was 50%. The behavioral stimuli were manipulated in the same way as the stimuli for the MEG experiment. Stimulus presentation was

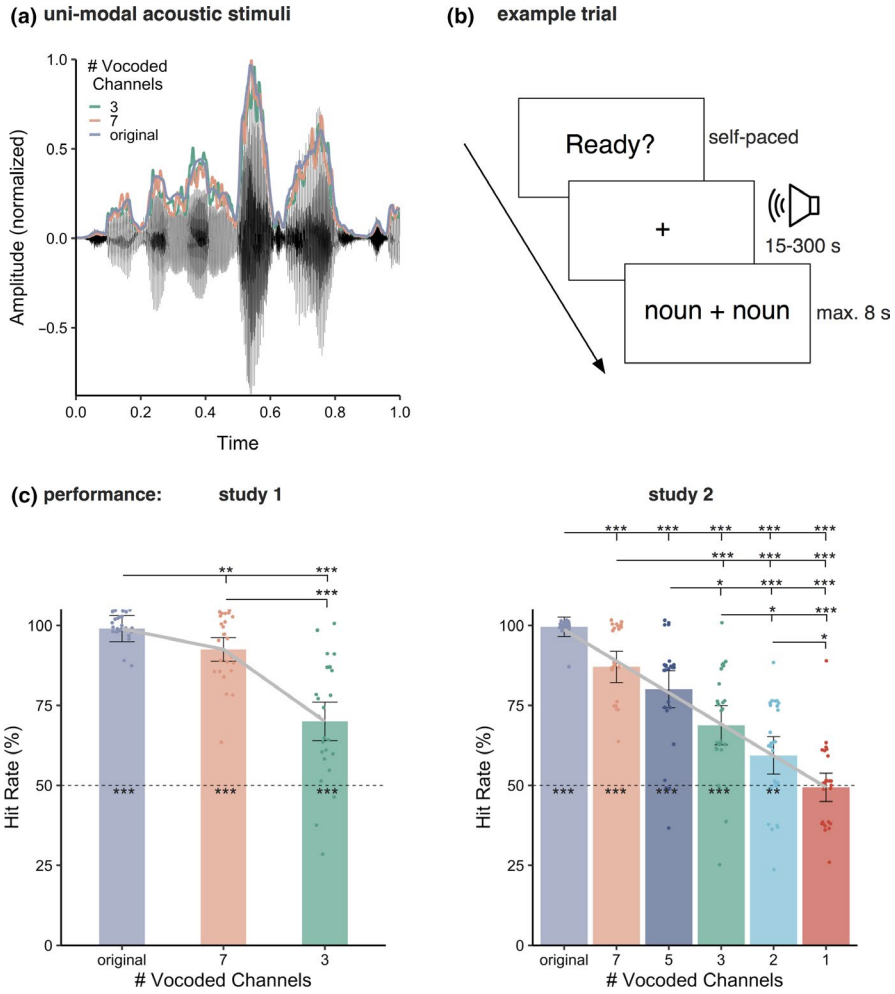

FIGURE 1 (a) an exemplary audio file with the corresponding envelope and with the envelopes of the vocoded audio stimuli presenting either 7 or 3 channels as used in study 1. (b) Example trial of unimodal acoustic stimulation. Participants started the presentation self-paced and listened to the stimulus during the visual presentation of a fixation cross. When the stimulus ended, participants were presented with two nouns of which they had to pick the one they perceived in the sentence before. (c) Hit rates in the behavioral experiment in studies 1 and 2 using acoustic stimuli of single sentences (range of 2–15 s). The gray curves represent the model-based predicted behavioral response (left: model combining linear and quadratic term; right: linear model). Bars represent 95% confidence intervals, $p_{fdr} < .05$*, $p_{fdr} < .01$**, $p_{fdr} < .001$***

controlled using in-house wrapper (https://gitlab.com/thht/o_ptb) for the MATLAB-based Psychtoolbox (Brainard, 1997; Kleiner et al., 2007; Pelli, 1997).

2.1.3 | Data acquisition and analyses

Extraction of acoustic speech envelope

For calculation of the coherence between speech envelope and brain activity, we extracted the acoustic speech envelope from all acoustic stimuli using the Chimera toolbox by Delguette and colleagues (http://research.meei.harvard.edu/chimera/More.html) where nine frequency bands in the range of 100 to 10,000 Hz were constructed as equidistant on the cochlear map (Chandrasekaran, Trubanova, Stillittano, Caplier, & Ghazanfar, 2009; Gross et al., 2013; Smith, Delgutte, & Oxenham, 2002). Sound stimuli were band-pass-filtered (forward and reverse) in these bands using a 4th-order Butterworth filter. For each band, envelopes were calculated as absolute values of the Hilbert transform and were averaged across bands to obtain the full-band envelope that was used for coherence analysis. We did this for all three conditions (original, 7-chan and 3-chan) resulting in virtually identical envelopes for those conditions (Figure 1a).

MEG acquisition and preprocessing

Data acquisition and analyses closely resemble with minor exceptions the one described in Hauswald et al. (2018). MEG was recorded at a sampling rate of 1 kHz using a 306-channel (204 first-order planar gradiometers) Triux MEG system (Elekta-Neuromag Ltd.) in a magnetically shielded room (AK3B, Vacuumschmelze). The MEG signal was online high-pass- and low-pass-filtered at 0.1 Hz and 330 Hz, respectively. Prior to the experiment, individual head shapes were digitized for each participant including fiducials (nasion, pre-auricular points) and around 300 points on the scalp using a Polhemus Fastrak Digitizer (Polhemus). We use a signal space separation algorithm provided by the MEG manufacturer and implemented in the Maxfilter program (version 2.2.15) to remove external noise from the MEG signal (mainly 16.6, and 50 Hz plus harmonics) and realign data to a common standard head position (across different blocks based on the measured head position at the beginning of each block).

Data were analyzed offline using the Fieldtrip toolbox (Oostenveld et al. 2011). First, a high-pass filter at 1 Hz (6th-order Butterworth IIR) was applied to continuous MEG data. Then, trials were defined according to the duration of each stimulus and cut into segments of 2 seconds to increase signal-to-noise ratio. As we were interested in frequency bands below 20 Hz and in order to save computational power, we resampled the data to 150 Hz. Independent component analysis was applied separately for visual and auditory blocks, and we then identified components corresponding to blinks and eye movements and cardiac activity and removed them. On average, 3.25 (*SD*: 1.143) components were removed for auditory blocks. Sensor space data were projected to source space using linearly constrained minimum variance beamformer filters (Van Veen, van Drongelen, Yuchtman, & Suzuki, 1997), and further analysis was performed on the obtained time series of each brain voxel (http://www.fieldtriptoolbox.org/tutorial/shared/virtual_sensors in FieldTrip). To transform the data into source space, we used a template structural magnetic resonance image (MRI) from Montreal Neurological Institute (MNI) and warped it to the subject's head shape (Polhemus points) to optimally match the individual fiducials and head shape landmarks. This procedure is part of the standard SPM (http://www.fil.ion.ucl.ac.uk/spm/) procedure of canonical brain localization (Mattout, Henson, & Friston, 2007).

A 3D grid covering the entire brain volume (resolution of 1 cm) was created based on the standard MNI template MRI. The MNI space equidistantly placed grid was then morphed to individual headspace. Finally, we used a mask to keep only the voxels corresponding to the gray matter (1,457 voxels). Using a grid derived from the MNI template allowed us to average and compute statistics as each grid point in the warped grid belongs to the same brain region across participants, despite different head coordinates. The aligned brain volumes were further used to create single-sphere head models and lead field matrices (Nolte, 2003). The average covariance matrix, the head model and the lead field matrix were used to calculate beamformer filters (regularization factor of 10%). The filters were subsequently multiplied with the sensor space trials resulting in single-trial time series in source space. The number of epochs across conditions was equalized.

We applied a frequency analysis to the 2-s segments of all three conditions (original, 7-chan and 3-chan) calculating multi-taper frequency transformation (dpss taper: 1–25 Hz in 1 Hz steps, 3 Hz smoothing, no baseline correction). These values were used for the analyses of alpha and for the coherence calculation between each virtual sensor and the acoustic speech envelope. For all three conditions, we used the envelopes of the original, nonvocoded acoustic signal. Then, the coherence between activity at each virtual sensor and the acoustic speech envelope during acoustic stimulation in the frequency spectrum was calculated and averaged across trials. We refer to the coherence between acoustic speech envelope and brain activity as speech tracking. As a sanity check, we calculated grand averages of the speech tracking of the three conditions to see whether they show the expected peak around 4 Hz (Figure 2a).

2.1.4 | Statistical analyses

We analyzed the responses from the behavioral experiment. Due to technical problems, behavioral measures are missing for 3 participants and the responses of the remaining 25 participants were analyzed. We used repeated-measures ANOVA to compare across the conditions and then dependent-samples *t* tests to compare hit rates between conditions and against chance level (50%), which were corrected for multiple comparisons by using the false discovery rate method (FDR, Benjamini & Hochberg, 1995).

Most studies on speech-brain entrainment report findings of frequencies below 7 Hz; therefore, we analyzed frequencies between 1 and 7 Hz. For alpha power, we analyze 8–12 Hz. For both MEG alpha power and 1–7 Hz coherence data, we applied repeated-measures ANOVA for each frequency within the range (ft_statfun_depsamplesFunivariate in FieldTrip, no averaging over frequency band) to test modulations of neural measures across the different degradation levels. To control for multiple comparisons, a nonparametric Monte Carlo randomization test was undertaken (Maris & Oostenveld, 2007). The test statistic was repeated 5,000 times on data shuffled across conditions, and the largest statistical value of a cluster coherent in source space was kept in memory. The observed clusters were compared against the distribution obtained from the randomization procedure and were considered significant when their probability was below 5%. Effects were identified in source space. All voxels within the cluster and the corresponding individual coherence and power values were extracted and averaged. Post hoc *t* tests between conditions were corrected for multiple comparisons by using the FDR method (Benjamini & Hochberg, 1995). For visualization, source localizations were averaged across the 1–7 Hz and respectively 8–12 Hz frequency bands and mapped onto inflated surfaces as implemented in FieldTrip.

We used linear mixed models to further test how our data (i.e., behavioral response, speech tracking and alpha power) are influenced by the vocoding levels. At the outset, we tested a simple linear model [*recorded measure* = (*vocoding levels*)] and compared it with a more complex (combined) by adding a quadratic term [*recorded measure* = (*vocoding levels* + (*vocoding levels*)^2)]. These two models were compared using an ANOVA test. The respective best model was subsequently reapplied to the data for each individual, and the average for these predicted model outcomes is displayed alongside the

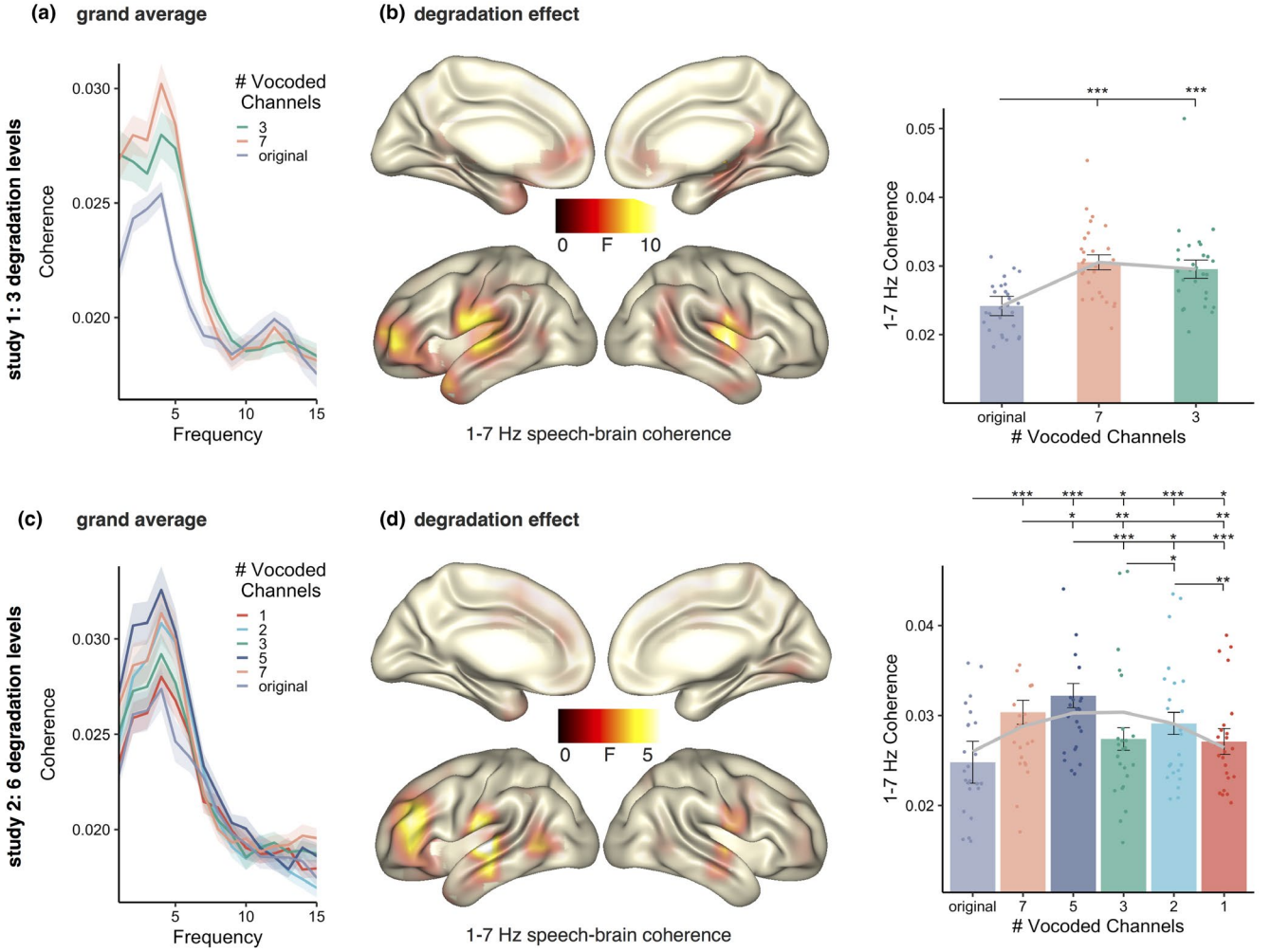

FIGURE 2 (a) Frequency spectrum of the speech tracking (coherence) for the three conditions averaged across all voxels. (b) Left: source localizations of degradation effects on speech tracking (1–7 Hz) during acoustic stimulation across three conditions (original, 7-chan and 3-chan) in bilateral temporal and left frontal regions. Right: individual speech tracking values of the three conditions extracted at voxels showing a significant effect contrasted with each other. The gray curve represents the predicted tracking values by the model combining linear and quadratic terms. (c) Frequency spectrum of the speech tracking for the six conditions averaged across all voxels. (d) Left: source localizations of degradation effects on speech tracking (1–7 Hz) during acoustic stimulation across six conditions (original, 7-chan, 5-chan, 3-chan, 2-chan 1-chan) in bilateral temporal and left frontal regions. Right: individual speech tracking values of the six conditions extracted at voxels showing a significant effect contrasted with each other. The gray curve represents the predicted tracking values by the model that combines linear and quadratic terms. Bars represent 95% confidence intervals, $p_{fdr} < .05$*, $p_{fdr} < .01$**, $p_{fdr} < .001$***

actual (grand) average results in the relevant bar graphs (gray curves).

2.2 | Results

2.2.1 | Behavioral results

The mean hit rate for original stimuli was 99% (*SD*: 3.43%) for the original sound files, 92.5% (*SD*: 10.21%) for 7-chan vocoded stimuli and 69.44% (*SD*: 18.75%) for 3-chan vocoded stimuli. A one-way ANOVA across the three conditions revealed a main effect ($F(72) = 37.14$, $p = 8.28\text{e-}12$).Comparing the different vocoding levels with each other showed higher hit rates for non-vocoded stimuli than for 7-chan ($t(24) = 3.376$, $p_{fdr} = .0025$) or 3-chan vocoded stimuli ($t(24) = 7.632$, $p_{fdr} = 1.437\text{e-}7$). 7-chan had higher hit rates than 3-chan vocoded stimuli ($t(24) = 6.2354$, $p_{fdr} = 2.8733\text{e-}6$). All conditions also showed significant above-chance (50%) hit rates (Figure 1c): for nonvocoded stimuli, $t(24) = 70.787$, $p_{fdr} = 1.3341\text{e-}28$, for 7-chan, $t(24) = 20.821$, $p_{fdr} = 2.1531\text{e-}16$, and for 3-chan vocoded, $t(24) = 5.333$, $p_{fdr} = 2.1494\text{e-}5$. The linear mixed models revealed significant linear decrease across conditions ($\chi^2 = 72.003$, $p < 2.2\text{e-}16$). Adding a quadratic term to the model benefitted the data prediction (*model comparison*: $\chi^2 = 7.8982$, $p < .004949$; gray curve in Figure 1c left).

2.2.2 | MEG data

Degradation-related effects

To investigate the effects of reducing the acoustic information, we ran a cluster-corrected repeated-measures ANOVA for the speech tracking (1–7 Hz coherence; see spectral distribution in Figure 2a) of the 3 conditions (original, 7-chan and 3-chan). An effect of degradation between 1 and 7 Hz (p = .0009) was yielded with maxima in bilateral middle temporal and left frontal regions and right thalamus and insula (Figure 2b, left). In these areas, the original audio stimuli lead to the weakest speech tracking, while the stimuli with the medium degradation (7-chan) elicited the strongest speech tracking (Figure 2b, right). Listening to the original audio files elicited lower tracking than listening to the 7-chan ($t(27) = -7.798$, p_{fdr} = 6.58e-8) or 3-chan version ($t(27) = -5.593$, p_{fdr} = 9.33e-6). The two vocoded stimulus classes did not differ significantly ($t(27) = 1.139$, $p_{fdr} = .264$). The linear mixed models revealed a significant linear pattern across conditions ($\chi^2 = 26.868$, p = 2.179e-07). Adding a quadratic term to the model benefitted the data prediction (*model comparison*: $\chi^2 = 19.998$, p = 7.751e-06; gray curve in Figure 2b right).

The same statistical analysis applied to alpha power (8–12 Hz, spectral distribution in Figure 3a) over original, 7-chan and 3-chan revealed an effect of degradation (p = .0009, Figure 3b), with alpha power during unaltered stimuli being higher during than 7-chan vocoding ($t(27) = 3.095$, p_{fdr} = .0045) and 3-chan vocoding ($t(27) = 4.09$, p_{fdr} = .001). Compared with 7-chan

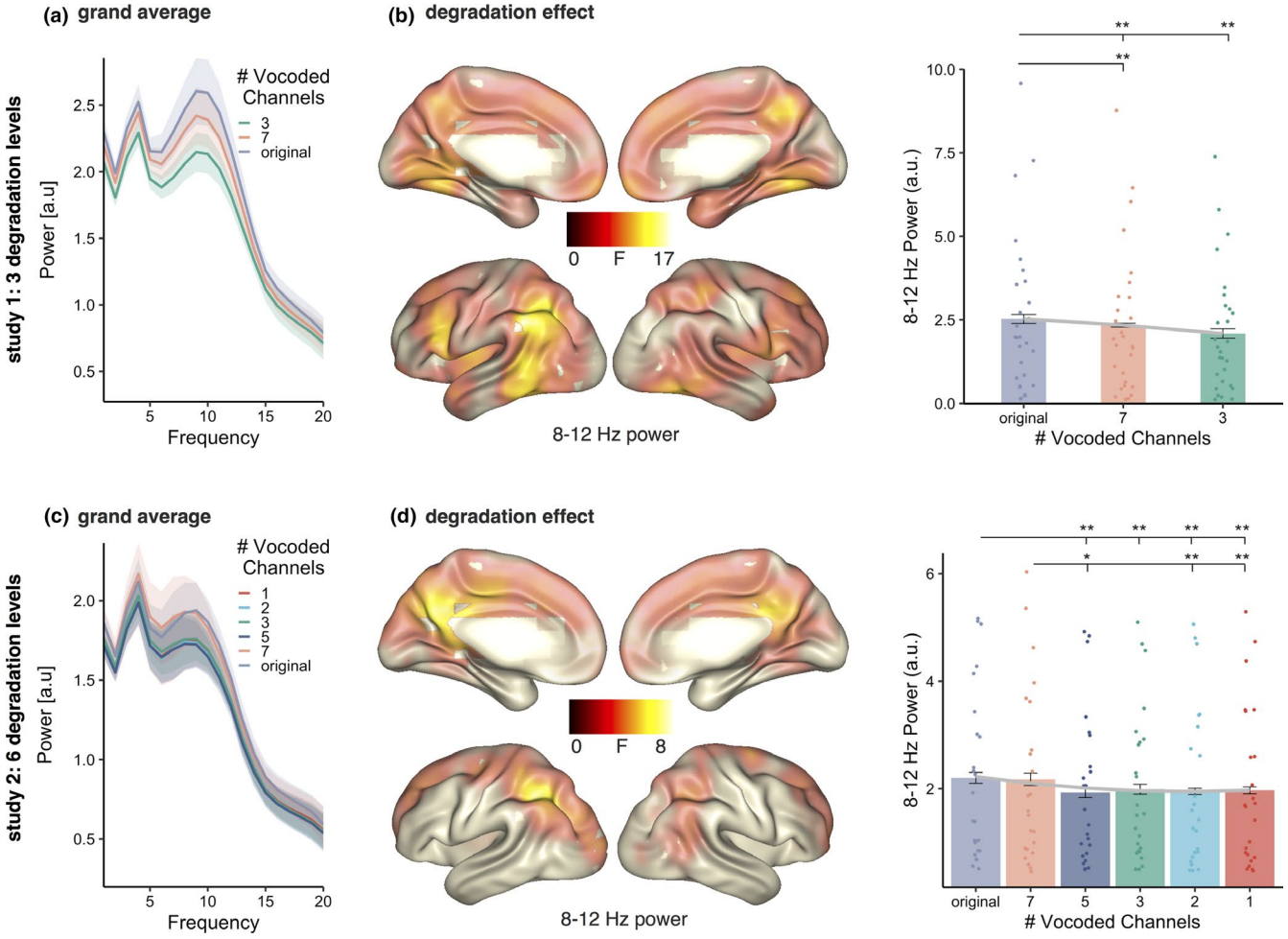

FIGURE 3 (a) Frequency spectrum of the power for the three conditions averaged across all voxels. (b) Left: source localizations of degradation effects on alpha power (8–12 Hz) across three conditions (original, 7-chan and 3-chan) with maxima in left angular gyrus and inferior parietal lobe, left frontal and inferior temporal regions. Right: individual 8–12 Hz power values of the three conditions extracted at voxels showing a significant effect contrasted with each other. The gray curve represents the predicted tracking values by the linear model. (c) Frequency spectrum of the power for the six conditions averaged across all voxels. (d) Left: source localizations of degradation effects on alpha power (8–12 Hz) across six conditions (original, 7-chan, 5-chan, 3-chan, 2-chan and 1-chan) with maxima in left angular gyrus and inferior parietal lobe. Right: individual 8–12 Hz power values of the three conditions extracted at voxels showing a significant effect contrasted with each other. The gray curve represents the predicted alpha power values by the model that combines linear and quadratic terms. Bars represent 95% confidence intervals, $p_{fdr} < .05$*, $p_{fdr} < .01$**, $p_{fdr} < .01$***

vocoding, alpha power during 3-chan vocoding decreased even further ($t(27) = 3.738$, $p_{fdr} = .0013$). The effect was widespread and covered most of the brain (present in 1,357 of 1,457 voxel) with a clear maximum in the left angular/parietal inferior cortex. The linear mixed models revealed a significant linear pattern across conditions ($\chi^2 = 30.292$, $p = 3.716e\text{-}08$). Adding a quadratic term to the model did not benefit the data prediction (*model comparison*: $\chi^2 = 0.2185$, $p = .6402$; gray curve Figure 3b right).

3 | Study 2

The findings from study 1 offer two important insights: First, the increase in speech-brain coherence and the decrease in alpha power with decline in acoustic detail are at odds with several previous studies (e.g. Dimitrijevic et al., 2019; Obleser et al., 2012). However, those studies have usually employed very brief stimuli, which is uncommon in natural listening situations. Second, the findings suggest that the relationship between degradation and speech tracking might not be linear and possibly behave differently than the relationship between degradation and alpha power. Therefore, we conducted a second study, to replicate the first by using again the previous vocoding levels and further extend it by adding three more vocoding levels: First, we added 5-channel vocoding to fill the gap between 7- and 3-channel vocoding, where comprehension is challenging but still possible. Furthermore, we also added 2- and 1-channel vocoding to make sure we also present unintelligible material.

3.1 | Materials and methods

3.1.1 | Participants

Twenty-four individuals participated in the second MEG study (female = 11, male = 13). Mean age was 26.37 years ($SD = 5.648$), with a range between 18 and 45 years. We recruited only German native speakers and people who were eligible for MEG recordings, that is, without nonremovable ferromagnetic metals in or close to the body. Seventeen of these also provided behavioral data (female = 8, male = 9, mean age = 27.2, $SD = 6.4$, age range = 18–45 years). Ten additional individuals participated in the behavioral part only (female = 6, male = 4, mean age = 23.2, $SD = 3.5$, age range = 20–33 years). Participants provided informed consent and were compensated monetarily or via course credit. Participation was voluntary and in line with the declaration of Helsinki and the statutes of the University of Salzburg.

3.1.2 | Stimuli

We used the same auditory stimulus material and experimental design as in study 1, but expanded the degradation levels to include additionally 5-channel, 2-channel and 1-channel vocoding. Overall, we had six levels of degradation: original, 7-channel, 5-channel, 3-channel, 2-channel and 1-channel.

3.1.3 | Data acquisition and analyses and statistical analyses

All steps of data acquisition, analysis and statistics were identical to study 1.

3.2 | Results

3.2.1 | Behavioral results

The mean hit rate was 99.46% (*SD*: 2.61%) for the original sound files, 86.415% (*SD*: 13.54%) for 7-chan vocoded stimuli, 78.8% (*SD*: 17.45) for 5-chan, 67.39% (*SD*: 17.57%) for 3-chan, 56.52% (*SD*: 15.01) for 2-chan and 50% (*SD*: 13.06%) for 1-chan vocoded stimuli. A one-way ANOVA across the three conditions revealed a main effect ($F(156) = 47.83$, $p = 8.28e\text{-}30$). Comparing the different vocoding levels with each other showed higher hit rates for nonvocoded stimuli than any of the other conditions (all $t > 4.83$, all $p_{fdr} < .000051$). 7-chan vocoded had higher hit rates than 3-, 2- and 1-chan vocoded stimuli (all $t > 4.96$, all $p_{fdr} < .000055$). 5-chan vocoding had higher hit rates than 3-, 2- and 1-chan vocoded stimuli (all $t > 2.62$, all $p_{fdr} < .05$). 3-chan vocoding had higher hit rates than 2- and 1-chan vocoded stimuli (all $t > 2.24$, all $p_{fdr} < .05$). 2-chan had higher hit rates than 1-chan vocoded stimuli ($t(26) = 2.74$, $p_{fdr} = .013$). The nonvocoded stimuli and the 7-chan, 5-chan, 3-chan and 2-chan vocoded conditions showed significant above-chance (50%) hit rates (all $t > 3.1$, all $p_{fdr} < .01$, Figure 1c, right). The contrast with 1-chan vocoded stimuli did not show a difference ($t(26) = -0.25$, $p_{fdr} = .801$). The linear mixed models revealed significant linear decrease across conditions ($\chi^2 = 282.09$, $p < 2.2e\text{-}16$). Adding a quadratic term to the model did not result in better prediction of the data (*model comparison*: $\chi^2 = 0.012$, $p = .9126$; gray curve in Figure 1c right).

3.2.2 | MEG data

Degradation-related effects

To investigate the effects of reducing the acoustic information, we ran a cluster-corrected repeated-measures ANOVA

for the speech tracking (1–7 Hz coherence; see spectral distribution in Figure 2c) of the 6 conditions (original, 7-chan, 5-chan, 3-chan, 2-chan and 1-chan). An effect of degradation between 1 and 7 Hz ($p = .0009$) was located in virtually identical regions as in study 1 (bilateral middle temporal and left frontal regions). In these areas, the original audio stimuli and the most strongly degraded (1-chan) led to the weakest speech tracking, while the stimuli with 5-chan degradation elicited strongest speech tracking (Figure 2d). Listening to the 5-chan vocoded audio files elicited higher tracking than listening to any of the other conditions (all $t > 2.5$, all $p_{fdr} < .05$). Listening to the original nonvocoded audio files elicited lower tracking than listening to any of the other conditions (all $t > -2.21$, all $p_{fdr} < .05$). Similarly, listening to 1-chan vocoded audio elicited lower tracking than listening to the 7-chan, 5-chan and 2-chan versions (all $t > -3.88$, all $p_{fdr} < .01$). Further, 3-chan vocoding yielded lower tracking than 7-chan ($t(23) = -3.64$, $p_{fdr} = .0026$) and 2-chan version ($t(23) = -2.72$, $p_{fdr} = .02$). The linear mixed models did not reveal a linear pattern ($\chi^2 = 0.1588$, $p = .6903$). Adding a quadratic term to the model significantly benefited data prediction (*model* comparison: $\chi^2 = 23.642$, $p = 1.16e\text{-}06$; gray curve in Figure 2d right).

Calculating cluster-corrected repeated-measures ANOVA for alpha power (8–12 Hz; see spectral distribution in Figure 3c) over the six conditions revealed an effect of degradation ($p = .0009$, Figure 3d) with maxima analogous to study 1, that is, in left angular gyrus and inferior parietal lobe. Nonvocoded and 7-chan vocoded stimuli eliciting higher alpha power in any of the other conditions (all $t > 2.904$, all $p_{fdr} < .05$) except 7-chan and 3-chan did not show a conclusive difference ($t(23) = 2.243$, $p_{fdr} = .0653$). The linear mixed models did reveal a significant linear pattern ($\chi^2 = 22.206$, $p = 2.449e\text{-}06$). Adding a quadratic term to the model significantly benefited data prediction (*model comparison*: $\chi^2 = 6.6019$, $p = .01019$, gray curve in Figure 3d right).

3.2.3 | Using neural measures to predict speech intelligibility

Our MEG data, especially using the richer set of degradation levels in study 2, indicate a differential impact on our neural measures. This should serve as a precaution against simplistically equating the neural measures to such abstract concepts as listening effort. In order to be functionally relevant, one would expect that these neural measures predict speech intelligibility. However, based on the previous analysis this is not clear. In a last hypothesis generating step of this study, with the aim of guiding future research, we postulate *alpha* to be an "activation" proxy of neural ensembles. However, such an "activation" may not necessarily lead to activation of veridical (i.e. intelligible) representations (Griffiths et al., 2019) especially when the sound becomes increasingly degraded. We speculate that speech tracking may reflect the outcome of this combination between "activation" and "veridicability." As no continuous time-varying quantification of the latter concept is available, behaviorally assessed "intelligibility" can serve as a proxy. The basic assumption of this *combined model* can thus be expressed as:

1. Speech Tracking = Activation × Intelligibility.

Thus by reordering (1), we obtain a simple model to predict intelligibility of speech from neural data:

2. Intelligibility = Speech Tracking/Activation.

The parameters of the model can be estimated using a linear mixed model (using lme4 library implemented in R; Bates, Mächler, Bolker, & Walker, 2015), and the model can be compared with competing models (see below). Models were fit using random intercepts. We used the speech-brain coherence and the alpha power of all significant voxels during the nonvocoded "effortless" condition to normalize the other five challenging (i.e., vocoded) listening conditions. For each participant (17 participants who contributed MEG and behavioral data), we then used the model to estimate *intelligibility* values for the vocoded conditions. This *predicted intelligibility* was then compared with the *observed intelligibility* (behavioral response; $\chi^2 = 8.3457$, $p = .003866$).

In order to evaluate whether a combination between speech tracking and activation yields a benefit, we compared predicted intelligibility with two simpler models either using only speech tracking (*tracking model*).

3. Intelligibility = Speech Tracking ($\chi^2 = 4.6476$, $p = .0311$).

or only activation (*activation model*).

4. Intelligibility = Activation ($\chi^2 = 2.7638$, $p = .09642$).

Directly comparing the *combined model* with the *tracking model* and in a separate step with the *activation model* shows superiority of the *combined model* (*combined model* vs. *tracking model*: $\chi^2 = 3.436$, $p < 2.2e\text{-}16$; *combined model* vs. *activation model*: $\chi^2 = 5.2411$, $p < 2.2e\text{-}16$). This means that speech tracking and alpha power together can better predict the behavioral response than either of the factors alone.

4 | DISCUSSION

As shown in previous studies (e.g. Luo & Poeppel, 2007; Obleser & Weisz, 2012; Obleser et al., 2012), listening to

degraded speech modulates speech tracking and alpha power. The pattern of this modulation varies across studies, suggesting that it might depend on experimental implementation and the two measures are not commonly reported together in the field of degraded speech. We advance these previous findings by investigating the effects of degraded speech stimuli on speech tracking and on alpha power in two studies using continuous speech and various degradation levels. In the first study, we used three levels of vocoding. Based on the behavioral results and the MEG findings, we conducted a second study expanding the degradation levels with one additional intermediate vocoding level (5-channel) and two very low vocoding levels (1- and 2-channel). As both studies yield very similar results in terms of behavior, speech tracking and alpha power, we will discuss them together.

4.1 | Behaviorally assessed intelligibility

To be sure that our manipulation actually affects intelligibility, participants performed a behavioral experiment after the MEG experiments. These were in both cases similar to the MEG experiment (with identical degradation levels) but with shorter stimuli, enabling us to assess more trials. The stimuli varied between 2 and 10 s instead of 15 s and 3 min as during the MEG recording. The data showed that participants decline in performance when the stimuli are degraded, which is in line with other studies showing a linear decline in performance (McGettigan et al., 2012; Strelnikov, Massida, Rouger, Belin, & Barone, 2011). The exact number of channels needed for high-performance understanding depends on the stimulus material and the specific experimental setup (Dorman, Loizou, & Rainey, 1997; Loizou, Dorman, & Tu, 1999). For our first study, we conclude that even the 3-channel condition was challenging yet not completely unintelligible given that performance is still higher than expected by chance. Therefore, we added the two lower vocoding conditions (2-channel, 1-channel) in the second study. Results of study 2 showed again that performance declines with degradation and that complete unintelligibility is reached with 1-channel vocoding.

4.2 | Speech tracking across degradation level follows an inverted U shape

To elucidate whether the intelligibility, measured by degradation level, affects the speech tracking, measured by speech-brain coherence, we calculated a repeated-measures ANOVA of the low-frequency speech-brain coherence (1–7 Hz) across the three (study 1), respectively six (study 2) conditions. For both studies, this revealed bilateral sources in temporal—including auditory—cortex and left frontal regions in which higher tracking was associated with a medium level of degradation. The linear mixed models using the individual coherence values of the sources identified by the ANOVA, suggest with both three and six conditions that the relationship between degradation levels and speech tracking follows an inverted U shape. These results nicely fit with fMRI findings of increased activation of (left) temporal and frontal inferior regions for degraded but yet intelligible stimuli compared with unaltered and completely unintelligible speech as reported by Davis and Johnsrude (2003) and interpreted as indicating recruitment of compensatory attentional resources. The authors showed that the effect in temporal areas was further depending on other acoustic features, while the frontal regions did not respond to those suggesting that the frontal regions serve a more general executive function (Davis & Johnsrude, 2003). Interestingly, those two regions (left inferior frontal gyrus and temporal region) exhibited enhanced fMRI responses to degraded but intelligible speech when attention was directed to the speech again interpreted as a marker of effortful listening (Wild et al., 2012) and left inferior cortex further plays a role in perceptual learning (Eisner, McGettigan, Faulkner, Rosen, & Scott, 2010). This is also consistent with a study showing non-native speakers produce higher delta/theta speech entrainment than native speakers and the authors have also proposed this as reflecting the higher effort (Song & Iverson, 2018). Similarly, speech tracking is increased during active compared with passive listening only for low levels of intelligibility (Vanthornhout et al., 2019). Further, the M50 of TRF is enhanced for degraded stimuli compared with unaltered ones in quiet environments as is delta entrainment, the latter again suggested to reflect listening efforts (Ding et al., 2014). Although studies have also reported decreased theta entrainment for degraded speech (Ding et al., 2014; Peelle, 2018; Rimmele et al., 2015), synchronization with the speech signal in both frequency bands is enhanced when attended to: Multi-speaker and auditory spatial attention studies using sentences or narratives have repeatedly found stronger low-frequency (1–7 Hz) speech tracking for attended compared with unattended speech (Ding & Simon, 2012; Horton et al., 2013; Rimmele et al., 2015; Zion Golumbic et al., 2013).

4.3 | Alpha power decreases across degradation levels

Another commonly used measure in studies of degraded speech—a common operationalization for listening effort—is the alpha rhythm (McMahon et al., 2016; Miles et al., 2017; Obleser & Weisz, 2012; Obleser et al., 2012). Interestingly, we found that alpha power followed a different pattern than coherence, which became most obvious in study 2. While speech tracking seems to have a U-shaped relationship with

degradation level, alpha power shows a widespread decrease for the stimuli with less acoustic information compared with clear speech. Study 2 suggests that this decrease reaches a floor effect already with 5-channel vocoding. Both studies show the maximum of this decrease in left angular and parietal inferior gyrus. This is a region that has been reported to play a crucial role in complex speech comprehension (Van Ettinger-Veenstra, McAllister, Lundberg, Karlsson, & Engström, 2016), especially important in successful comprehension of degraded but predictable speech (Hartwigsen, Golombek, & Obleser, 2015; Obleser, Wise, Dresner, & Scott, 2007) and in perceptual learning of degraded speech (Eisner et al., 2010). The pattern of decreasing alpha power is further consistent with other studies using degradation of complex speech material as for example sentences (McMahon et al., 2016; Miles et al., 2017). However, studies using short and simple speech stimuli such as single words (Becker, Pefkou, Michel, & Hervais-Adelman, 2013; Obleser & Weisz, 2012) or digits (Obleser et al., 2012; Wöstmann et al., 2015) report enhanced alpha for stimuli with more acoustic detail compared with degraded sounds. The source localizations of the enhanced alpha in those studies show overlapping regions and distinct regions compared with our studies, offering the possibility that alpha power reflects at least partly different processes being recruited in the different studies. However, based on the consistent differences regarding the length of the stimulus material, one compelling explanation for the enhanced versus reduced alpha power might be linked to the linguistically more complex nature of the longer speech stimuli as also suggested by Miles et al. (2017).

To the best of our knowledge, so far no study investigated the influences of vocoded continuous speech on both alpha power and speech tracking. A study on a related topic found that cochlear implant (CI) users show alpha power to be positively correlated with subjective listening effort, while speech-brain coherence showed a negative relationship (Dimitrijevic et al., 2019). Besides the differences in study groups (participants with normal hearing vs. CI users) and operationalization of listening effort (vocoded speech vs. speech-in-noise tasks) between our study and the one of Dimitrijevicet al. (2019), they used short auditory stimuli (digits) as many of the studies (Becker et al., 2013; Obleser & Weisz, 2012; Obleser et al., 2012; Wöstmann et al., 2015) reporting the opposite pattern in alpha power than us.

4.4 | Influences of stimulus material

Degrading the speech by vocoding as we did in the present study and as done by many other studies (e.g. Miles et al., 2017; Obleser et al., 2007) reduces the phonetic fine structure while temporal information, for example, segmentation of syllables, is preserved (Shannon, Zeng, Kamath, Wygonski, & Ekelid, 1995). Based on our results, it seems that for challenging speech (reduced fine structure) people have to rely more on the temporal structure of speech leading to enhanced tracking in this frequency range (delta, theta) and that this process reverses beyond a critical point of degradation.

This effect might be amplified by the choice of long continuous speech stimuli. Unlike several other studies that used degraded single words (Becker et al., 2013; Obleser & Weisz, 2012; Obleser et al., 2012; Wöstmann et al., 2015), we implemented long stimuli in the MEG studies, between 30 s and 3 min. The long duration of our stimuli might affect the perception of the different degradation levels differently via "warming-up" to the stimuli (Dorman et al., 1997). Experimental investigation of this warm-up or perceptual learning effect shows that indeed speech understanding increased over time for degraded stimuli (e.g. 4-channel vocoding: Rosen, Faulkner, & Wilkinson, 1999; 6-channel vocoding: Davis, Johnsrude, Hervais-Adelman, Taylor, & McGettigan, 2005) and that this increase was bigger for sentences than for single words (Hervais-Adelman, Davis, Johnsrude, & Carlyon, 2008) and smallest for very strong (1-channel) or very little (24-channel) vocoding (Sohoglu & Davis, 2016). Similar nonlinear patterns have been reported for dual-task measures of listening effort. Reaction times (Wu, Stangl, Zhang, Perkins, & Eilers, 2016) and pupil sizes (Zekveld & Kramer, 2014) were enhanced for the middle range of speech intelligibility. Based on these findings, we speculate that the processes underlying listening to degraded speech dynamically vary depending on the stimulus length.

4.5 | Can listening effort explain results?

Intuitively, listening effort seems like an easy-to-understand concept, and individuals usually can answer without difficulty whether listening to a stimulus was effortful. Stimulus degradation (e.g. vocoding) is a common operationalization for listening effort (e.g. Obleser & Weisz, 2012). However, listening effort combines many dimensions. Peelle (2018) proposed a model comprising person-related characteristics (e.g. motivation) and stimulus-related characteristics (e.g. signal-to-noise ratio). Various measures exist for capturing listening effort, alpha power being one of them (e.g. Dimitrijevic et al., 2019; Miles et al., 2017). Speech tracking is not classically viewed as a measure of listening effort; nevertheless, its modulations when listening to challenging speech have been interpreted as increased effort (Song & Iverson, 2018) and increased attentional demands (Rimmele et al., 2015). Importantly, it has been shown that listening effort is of multidimensional nature with the different dimensions being captured by different measures that do not necessarily correlate (Alhanbali, Dawes,

Millman, & Munro, 2019; McGarrigle et al., 2014). This fits with our findings of degradation levels (across a wide range) affecting alpha power and speech tracking differently and suggests that such measures are not ideal to explain abstract concepts as listening effort independent of circumstances.

4.6 | Beyond listening effort, towards intelligibility

Results of the linear mixed models suggest that subjective intelligibility (behavioral response) can best be predicted by a combination of speech coherence and alpha power: We propose that for continuous degraded speech, understanding speech depends on the activation of veridical representations. Along the lines of a recent framework by Griffiths et al. (2019), we propose that this activation is reflected by alpha decrease. This process will however only support listening (e.g. reflected in the ability to track specific features) up to a specific (breaking) point, when speech becomes too degraded so that no veridical information is activated. This interpretation integrates well with the frameworks on alpha oscillations in the context of working memory as proposed, for example, by van Ede (2018), but also for auditory perception by Griffiths et al. (2019), and for auditory memory by Kraft, Demarchi, and Weisz (2019). Van Ede (2018) puts the idea forward that alpha power increases for tasks with sensory disengagement, while it decreases for tasks, which recruit the sensory representation. Our task of asking participants to identify which of two presented words did occur within the just heard four last words of a speech stimulus will most likely recruit the sensory representation of words, thereby leading to a relative alpha decrease. For our results, this would imply that the sensory representation is activated for all conditions of challenging speech as reflected by alpha decrease. For challenging conditions, this increased engagement is accompanied by increased tracking, which decreases again when speech becomes unintelligible even though neural activation per se remains high. This fits nicely with the ease of language understanding model (Rönnberg et al., 2013), which puts the idea forward that the perceived phonological signals are tested against the stored phonological representation in memory, and when they do not match, explicit working memory processes are elicited that aim at reconstructing the signal content. Several studies support the direct relationship between working memory and speech processing (e.g., Eisner et al., 2010; Rönnberg et al. 2010; Rudner, Lunner, Behrens, Thorén, & Rönnberg, 2012). Within these frameworks, also different findings in the literature concerning alpha can be unified by taking the specific task and the resulting demands into account.

5 | CONCLUSIONS

In sum, prior research reports mixed results concerning the link between degradation and speech-brain coherence and alpha power. We conducted two experiments with different levels of degradation, importantly of continuous speech. The results of these two studies show that the level of degradation affects speech tracking and alpha power differently: Speech tracking shows a U-shaped pattern with the easiest (original) and hardest (1-channel) to understand producing the lowest tracking values and the middle degradation level (5-channel) eliciting the highest tracking values. On the other hand, alpha power seems to overall decline with the declining clarity of speech. As study 2 shows, this decline likely reaches a floor effect also with 5-channel vocoding. Use of EEG signals is gaining momentum in the discussion about hearing aids improvement (Bech Christensen, Hietkamp, Harte, Lunner, & Kidmose, 2018; Fiedler, Obleser, Lunner, & Graversen, 2016). In this context, our findings have wider implications as they provide insights into more naturalistic, that is, continuous speech compared with single words and digits. Importantly, our results indicate that taking into account alpha modulations (interpreted in terms of neural activation) and neural speech tracking in a combined manner may open up avenues to monitor the (subjective) intelligibility of speech sounds. This perspective goes beyond simplistic listening effort accounts and could have important applied implications.

COMPETING INTERESTS
The authors declare no competing financial interests.

ACKNOWLEDGEMENTS
This study was supported by a FWF Einzelprojekt (P 31230). AK was supported by the Wellcome Trust [204820/Z/16/Z]. We would like to thank Joachim Gross and Hyojin Park for providing their original MATLAB script for extracting the lip area. We would also like to thank Siri Ebert and Jonas Heilig for help with data acquisition.

AUTHOR CONTRIBUTIONS
A.H., A.K. and N.W. designed the study. A.H. analyzed data. A.H., A.K., Y.C., S.R. and N.W. drafted paper.

DATA AVAILABILITY STATEMENT
Behavioral and processed MEG data are stored on https://osf.io/pm8xg/files/

PEER REVIEW
The peer review history for this article is available at https://publons.com/publon/10.1111/ejn.14912

ORCID

Anne Hauswald https://orcid.org/0000-0002-3754-0807
Anne Keitel https://orcid.org/0000-0003-4498-0146
Ya-Ping Chen https://orcid.org/0000-0001-5171-4255
Sebastian Rösch https://orcid.org/0000-0003-3579-6435
Nathan Weisz https://orcid.org/0000-0001-7816-0037

REFERENCES

Alexandrou, A. M., Saarinen, T., Kujala, J., & Salmelin, R. (2018). Cortical entrainment: What we can learn from studying naturalistic speech perception. *Language, Cognition and Neuroscience*, *35*(6), 1–13. https://doi.org/10.1080/23273798.2018.1518534

Alhanbali, S., Dawes, P., Millman, R. E., & Munro, K. J. (2019). Measures of listening effort are multidimensional. *Ear and Hearing*, *40*, 1084–1097. https://doi.org/10.1097/AUD.0000000000000697

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*, 1–48.

Bech Christensen, C., Hietkamp, R. K., Harte, J. M., Lunner, T., & Kidmose, P. (2018). Toward EEG-assisted hearing aids: Objective threshold estimation based on Ear-EEG in subjects with sensorineural hearing loss. *Trends in Hearing*, *22*, 233121651881620. https://doi.org/10.1177/2331216518816203

Becker, R., Pefkou, M., Michel, C. M., & Hervais-Adelman, A. G. (2013). Left temporal alpha-band activity reflects single word intelligibility. *Frontiers in Systems Neuroscience*, *7*. https://doi.org/10.3389/fnsys.2013.00121

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, *57*, 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*, 433–436. https://doi.org/10.1163/156856897X00357

Chandrasekaran, C., Trubanova, A., Stillittano, S., Caplier, A., & Ghazanfar, A. A. (2009). The natural statistics of audiovisual speech. *PLoS Computational Biology*, *5*, e1000436. https://doi.org/10.1371/journal.pcbi.1000436

Crosse, M. J., Di Liberto, G. M., Bednar, A., & Lalor, E. C. (2016). The multivariate temporal response function (mTRF) toolbox: A MATLAB toolbox for relating neural signals to continuous stimuli. *Frontiers in Human Neuroscience*, *10*. https://doi.org/10.3389/fnhum.2016.00604

Davis, M. H., & Johnsrude, I. S. (2003). Hierarchical processing in spoken language comprehension. *The Journal of Neuroscience*, *23*, 3423–3431. https://doi.org/10.1523/JNEUROSCI.23-08-03423.2003

Davis, M. H., Johnsrude, I. S., Hervais-Adelman, A., Taylor, K., & McGettigan, C. (2005). Lexical information drives perceptual learning of distorted speech: Evidence from the comprehension of noise-vocoded sentences. *Journal of Experimental Psychology: General*, *134*, 222–241. https://doi.org/10.1037/0096-3445.134.2.222

Dimitrijevic, A., Smith, M. L., Kadis, D. S., & Moore, D. R. (2019). Neural indices of listening effort in noisy environments. *Scientific Reports*, *9*, 1–10. https://doi.org/10.1038/s41598-019-47643-1

Ding, N., Chatterjee, M., & Simon, J. Z. (2014). Robust cortical entrainment to the speech envelope relies on the spectro-temporal fine structure. *NeuroImage*, *88*, 41–46. https://doi.org/10.1016/j.neuroimage.2013.10.054

Ding, N., & Simon, J. Z. (2011). Neural coding of continuous speech in auditory cortex during monaural and dichotic listening. *Journal of Neurophysiology*, *107*, 78–89. https://doi.org/10.1152/jn.00297.2011

Ding, N., & Simon, J. Z. (2012). Emergence of neural encoding of auditory objects while listening to competing speakers. *Proceedings of the National Academy of Sciences*, *109*, 11854–11859. https://doi.org/10.1073/pnas.1205381109

Dorman, M. F., Loizou, P. C., & Rainey, D. (1997). Speech intelligibility as a function of the number of channels of stimulation for signal processors using sine-wave and noise-band outputs. *Journal of the Acoustical Society of America*, *102*, 2403–2411. https://doi.org/10.1121/1.419603

Eisner, F., McGettigan, C., Faulkner, A., Rosen, S., & Scott, S. K. (2010). Inferior frontal gyrus activation predicts individual differences in perceptual learning of cochlear-implant simulations. *Journal of Neuroscience*, *30*, 7179–7186. https://doi.org/10.1523/JNEUROSCI.4040-09.2010

Fiedler, L., Obleser, J., Lunner, T., & Graversen, C. (2016). Ear-EEG allows extraction of neural responses in challenging listening scenarios—A future technology for hearing aids? 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Orlando, FL, 2016. *IEEE Engineering in Medicine and Biology Society. Conference* (pp. 5697–5700). doi: 10.1109/EMBC.2016.7592020.

Frey, J. N., Mainy, N., Lachaux, J.-P., Muller, N., Bertrand, O., & Weisz, N. (2014). Selective modulation of auditory cortical alpha activity in an audiovisual spatial attention task. *Journal of Neuroscience*, *34*, 6634–6639. https://doi.org/10.1523/JNEUROSCI.4813-13.2014

Gaudrain, E. (2016). Vocoder, v1.0 Online code at https://github.com/egaudrain/vocoder, doi:10.5281/zenodo.48120.

Greenberg, S. (1998). A syllable-centric framework for the evolution of spoken language. *The Behavioral and Brain Sciences*, *518*. https://doi.org/10.1017/S0140525X98301260

Griffiths, B. J., Mayhew, S. D., Mullinger, K. J., Jorge, J., Charest, I., Wimber, M., & Hanslmayr, S. (2019). Alpha/beta power decreases track the fidelity of stimulus-specific information. *eLife*, *8*, e49562. https://doi.org/10.7554/eLife.49562

Gross, J., Hoogenboom, N., Thut, G., Schyns, P., Panzeri, S., Belin, P., & Garrod, S. (2013). Speech rhythms and multiplexed oscillatory sensory coding in the human brain. *PLoS Biology*, *11*, e1001752. https://doi.org/10.1371/journal.pbio.1001752

Hartwigsen, G., Golombek, T., & Obleser, J. (2015). Repetitive transcranial magnetic stimulation over left angular gyrus modulates the predictability gain in degraded speech comprehension. *Cortex, Special Issue: Prediction in Speech and Language Processing*, *68*, 100–110. https://doi.org/10.1016/j.cortex.2014.08.027

Hauswald, A., Lithari, C., Collignon, O., Leonardelli, E., & Weisz, N. (2018). A visual cortical network for deriving phonological information from intelligible lip movements. *Current Biology*, *28*, 1453–1459.e3. https://doi.org/10.1016/j.cub.2018.03.044

Hervais-Adelman, A., Davis, M. H., Johnsrude, I. S., & Carlyon, R. P. (2008). Perceptual learning of noise vocoded words: Effects of feedback and lexicality. *Journal of Experimental Psychology: Human Perception and Performance*, *34*, 460–474. https://doi.org/10.1037/0096-1523.34.2.460

Horton, C., D'Zmura, M., & Srinivasan, R. (2013). Suppression of competing speech through entrainment of cortical oscillations. *Journal of Neurophysiology*, *109*, 3082–3093. https://doi.org/10.1152/jn.01026.2012

Howard, M. F., & Poeppel, D. (2010). Discrimination of speech stimuli based on neuronal response phase patterns depends on acoustics but

not comprehension. *Journal of Neurophysiology*, *104*, 2500–2511. https://doi.org/10.1152/jn.00251.2010

Keitel, A., Gross, J., & Kayser, C. (2018). Perceptually relevant speech tracking in auditory and motor cortex reflects distinct linguistic features. *PLoS Biology*, *16*, e2004473. https://doi.org/10.1371/journal.pbio.2004473

Kleiner, M., Brainard, D., Pelli, D., Ingling, A., Murray, R., & Broussard, C. (2007). What's new in psychtoolbox-3. *Perception*, *36*, 1–16.

Kraft, N., Demarchi, G., & Weisz, N. (2019). Auditory cortical alpha desynchronization prioritizes the representation of memory items during a retention period, *bioRxiv*, 626929.

Loizou, P. C., Dorman, M., & Tu, Z. (1999). On the number of channels needed to understand speech. *Journal of the Acoustical Society of America*, *106*, 2097–2103. https://doi.org/10.1121/1.427954

Luo, H., & Poeppel, D. (2007). Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. *Neuron*, *54*, 1001–1010. https://doi.org/10.1016/j.neuron.2007.06.004

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

Mattout, J., Henson, R. N., & Friston, K. J. (2007). Canonical source reconstruction for MEG. *Computational Intelligence and Neuroscience*, *2007*, 1–10. https://doi.org/10.1155/2007/67613

McGarrigle, R., Munro, K. J., Dawes, P., Stewart, A. J., Moore, D. R., Barry, J. G., & Amitay, S. (2014). Listening effort and fatigue: What exactly are we measuring? A British Society of Audiology Cognition in Hearing Special Interest Group "white paper". *International Journal of Audiology*, *53*, 433–440. https://doi.org/10.3109/14992027.2014.890296

McGettigan, C., Faulkner, A., Altarelli, I., Obleser, J., Baverstock, H., & Scott, S. K. (2012). Speech comprehension aided by multiple modalities: Behavioural and neural interactions. *Neuropsychologia*, *50*, 762–776. https://doi.org/10.1016/j.neuropsychologia.2012.01.010

McMahon, C. M., Boisvert, I., de Lissa, P., Granger, L., Ibrahim, R., Lo, C. Y., … Graham, P. L. (2016). Monitoring alpha oscillations and pupil dilation across a performance-intensity function. *Frontiers in Psychology*, *7*. https://doi.org/10.3389/fpsyg.2016.00745

Miles, K., McMahon, C., Boisvert, I., Ibrahim, R., de Lissa, P., Graham, P., & Lyxell, B. (2017). Objective assessment of listening effort: Coregistration of pupillometry and EEG. *Trends in Hearing*, *21*, 2331216517706396. https://doi.org/10.1177/2331216517706396

Nolte, G. (2003). The magnetic lead field theorem in the quasi-static approximation and its use for magnetoencephalography forward calculation in realistic volume conductors. *Physics in Medicine & Biology*, *48*, 3637–3652. https://doi.org/10.1088/0031-9155/48/22/002

Obleser, J., & Weisz, N. (2012). Suppressed alpha oscillations predict intelligibility of speech and its acoustic details. *Cerebral Cortex*, *22*, 2466–2477. https://doi.org/10.1093/cercor/bhr325

Obleser, J., Wise, R. J. S., Dresner, M. A., & Scott, S. K. (2007). Functional Integration across brain regions improves speech perception under adverse listening conditions. *Journal of Neuroscience*, *27*, 2283–2289. https://doi.org/10.1523/JNEUROSCI.4663-06.2007

Obleser, J., Wöstmann, M., Hellbernd, N., Wilsch, A., & Maess, B. (2012). Adverse listening conditions and memory load drive a common alpha oscillatory network. *Journal of Neuroscience*, *32*, 12376–12383. https://doi.org/10.1523/JNEUROSCI.4908-11.2012

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*,1–9.

Peelle, J. E. (2018). Listening effort: how the cognitive consequences of acoustic challenge are reflected in brain and behavior. *Ear and Hearing*, *39*, 204. https://doi.org/10.1097/AUD.0000000000000494

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, *10*, 437–442. https://doi.org/10.1163/156856897X00366

Poeppel, D. (2003). The analysis of speech in different temporal integration windows: Cerebral lateralization as 'asymmetric sampling in time'. *Speech Communication, The Nature of Speech Perception*, *41*, 245–255. https://doi.org/10.1016/S0167-6393(02)00107-3

Riecke, L., Formisano, E., Sorger, B., Başkent, D., & Gaudrain, E. (2018). Neural entrainment to speech modulates speech intelligibility. *Current Biology*, *28*, 161–169.e5. https://doi.org/10.1016/j.cub.2017.11.033

Rimmele, J. M., Zion Golumbic, E., Schröger, E., & Poeppel, D. (2015). The effects of selective attention and speech acoustics on neural speech-tracking in a multi-talker scene. *Cortex, Special Issue: Prediction in Speech and Language Processing*, *68*, 144–154. https://doi.org/10.1016/j.cortex.2014.12.014

Rönnberg, J., Lunner, T., Zekveld, A., Sörqvist, P., Danielsson, H., Lyxell, B., … Rudner, M. (2013). The Ease of Language Understanding (ELU) model: Theoretical, empirical, and clinical advances. *Frontiers in Systems Neuroscience*, *7*, 1–17.

Rönnberg, J., Rudner, M., Lunner, T., & Zekveld, A. A. (2010). When cognition kicks in: Working memory and speech understanding in noise. *Noise and Health*, *12*(49), 263–269. https://doi.org/10.4103/1463-1741.70505

Rosen, S., Faulkner, A., & Wilkinson, L. (1999). Adaptation by normal listeners to upward spectral shifts of speech: Implications for cochlear implants. *Journal of the Acoustical Society of America*, *106*, 3629–3636. https://doi.org/10.1121/1.428215

Rudner, M., Lunner, T., Behrens, T., Thorén, E., & Rönnberg, J. (2012). Working memory capacity may influence perceived effort during aided speech recognition in noise. *Journal of the American Academy of Audiology*, *23*, 577–589. https://doi.org/10.3766/jaaa.23.7.7

Shannon, R. V., Zeng, F.-G., Kamath, V., Wygonski, J., & Ekelid, M. (1995). Speech recognition with primarily temporal cues. *Science*, *270*, 303–304. https://doi.org/10.1126/science.270.5234.303

Smith, Z. M., Delgutte, B., & Oxenham, A. J. (2002). Chimaeric sounds reveal dichotomies in auditory perception. *Nature*, *416*, 87–90. https://doi.org/10.1038/416087a

Sohoglu, E., & Davis, M. H. (2016). Perceptual learning of degraded speech by minimizing prediction error. *Proceedings of the National Academy of Sciences*, *113*, E1747–E1756. https://doi.org/10.1073/pnas.1523266113

Song, J., & Iverson, P. (2018). Listening effort during speech perception enhances auditory and lexical processing for non-native listeners and accents. *Cognition*, *179*, 163–170. https://doi.org/10.1016/j.cognition.2018.06.001

Strelnikov, K., Massida, Z., Rouger, J., Belin, P., & Barone, P. (2011). Effects of vocoding and intelligibility on the cerebral response to speech. *BMC Neuroscience*, *12*. https://doi.org/10.1186/1471-2202-12-122

van Ede, F. (2018). Mnemonic and attentional roles for states of attenuated alpha oscillations in perceptual working memory: A review. *European Journal of Neuroscience*, *48*, 2509–2515.

Van Ettinger-Veenstra, H., McAllister, A., Lundberg, P., Karlsson, T., & Engström, M. (2016). Higher language ability is related to angular

gyrus activation increase during semantic processing, independent of sentence incongruency. *Frontiers in Human Neuroscience*, *10*, 1–9.

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*, 867–880. https://doi.org/10.1109/10.623056

Vanthornhout, J., Decruy, L., & Francart, T. (2019). Effect of task and attention on neural tracking of speech. *Frontiers in Neuroscience*, *13*, 977. https://doi.org/10.3389/fnins.2019.00977

Viswanathan, V., Bharadwaj, H. M., & Shinn-Cunningham, B. G. (2019). Electroencephalographic signatures of the neural representation of speech during selective attention. *Eneuro*, *6*(5), ENEURO.0057-19.2019. https://doi.org/10.1523/ENEURO.0057-19.2019

Wild, C. J., Yusuf, A., Wilson, D. E., Peelle, J. E., Davis, M. H., & Johnsrude, I. S. (2012). Effortful listening: The processing of degraded speech depends critically on attention. *Journal of Neuroscience*, *32*, 14010–14021. https://doi.org/10.1523/JNEUROSCI.1528-12.2012

Wöstmann, M., Herrmann, B., Wilsch, A., & Obleser, J. (2015). Neural alpha dynamics in younger and older listeners reflect acoustic challenges and predictive benefits. *Journal of Neuroscience*, *35*, 1458–1467. https://doi.org/10.1523/JNEUROSCI.3250-14.2015

Wu, Y.-H., Stangl, E., Zhang, X., Perkins, J., & Eilers, E. (2016). Psychometric functions of dual-task paradigms for measuring listening effort. *Ear and Hearing*, *37*, 660–670. https://doi.org/10.1097/AUD.0000000000000335

Zekveld, A. A., & Kramer, S. E. (2014). Cognitive processing load across a wide range of listening conditions: Insights from pupillometry. *Psychophysiology*, *51*, 277–284. https://doi.org/10.1111/psyp.12151

Zion Golumbic, E. M., Ding, N., Bickel, S., Lakatos, P., Schevon, C. A., McKhann, G. M., … Schroeder, C. E. (2013). Mechanisms underlying selective neuronal tracking of attended speech at a "Cocktail Party". *Neuron*, *77*, 980–991. https://doi.org/10.1016/j.neuron.2012.12.037

Zoefel, B., Archer-Boyd, A., & Davis, M. H. (2018). Phase entrainment of brain oscillations causally modulates neural responses to intelligible speech. *Current Biology*, *28*, 401–408.e5. https://doi.org/10.1016/j.cub.2017.11.071

How to cite this article: Hauswald A, Keitel A, Chen Y-P, Rösch S, Weisz N. Degradation levels of continuous speech affect neural speech tracking and alpha power differently. *Eur J Neurosci*. 2020;00:1–15. https://doi.org/10.1111/ejn.14912