Neural attention filters do not predict behavioral success in a large cohort of aging listeners

Sarah Tune*, Mohsen Alavash, Lorenz Fiedler, & Jonas Obleser*

Department of Psychology, University of Lübeck, 23562 Lübeck, Germany

* *Author correspondence*:
Sarah Tune, Jonas Obleser
Department of Psychology
University of Lübeck
Maria-Goeppert-Str. 9a
23562 Lübeck
Email: sarah.tune@uni-luebeck.de; jonas.obleser@uni-luebeck.de

Running title: Modelling neural filters and behavioral outcome in attentive listening

Number of words in Abstract: 193
Number of words in Introduction: 791
Number of words in Discussion: 1882

41 pages, 6 Figures, 0 Tables, Includes Supplemental Information

Conflict of Interest: The authors declare no competing financial interests.

Keywords: EEG, alpha power lateralization, neural speech tracking, speech processing, auditory selective attention

Author contributions: S.T., M.A., and J.O. designed the experiment; S.T. and M.A. oversaw data collection and preprocessing of the data; S.T., M.A. and L.F. analyzed the data; S.T., L.F., M.A., and J.O. wrote the paper.

Acknowledgments: Research was funded by the European Research Council (grant no. ERC-CoG-2014-646696 "Audadapt" awarded to J.O.). The authors are grateful for the help of Franziska Scharata in acquiring the data.

Abstract

Successful speech comprehension requires the listener to differentiate relevant from irrelevant sounds. Most studies on this problem address one of two candidate neural auditory filter solutions: Selective neural tracking of attended versus ignored speech in auditory cortex ("speech tracking"), and the lateralized modulation of alpha power in wider temporo–parietal cortex. Using an unprecedentedly large, age-varying sample (N=155; age=39–80 years) in a difficult listening task, we here demonstrate their limited potency to predict behavioral listening success at the single-trial level. Within auditory cortex, we observed attentional–cue-driven modulation of both, speech tracking and alpha power. Both filter mechanisms clearly operate functionally independent of each other and appear undiminished with chronological age. Importantly, single-trial models and cross-validated predictive analyses in this large sample challenge the immediate functional relevance of these neural attentional filters to overt behavior: The direct impact of spatial attentional cues as well as typical confounders age and hearing loss on behavior reliably outweighed the relatively minor predictive influence of speech tracking and alpha power. Our results emphasize the importance of large-scale, trial-by-trial analyses and caution against simplified accounts of neural filtering strategies for behavior often derived from younger, smaller samples.

Introduction

Real-life listening is characterized by the concurrence of sound sources that compete for our attention [1]. Successful speech comprehension thus relies on the selective filtering of relevant and irrelevant inputs. How does the brain instantiate attentional filter mechanisms that allow for the controlled inhibition and amplification of speech? Recent neuroscientific approaches to this question have focused on two potential neural filter strategies originating from distinct research traditions:

From the visual domain stems an influential line of research that support a role of alpha-band (~8–12 Hz) oscillatory activity in the implementation of controlled, top-down suppression of behaviorally-irrelevant information [2-5]. Importantly, across modalities, it was shown that tasks requiring spatially-directed attention are neurally supported by a hemispheric lateralization of alpha power over occipital, parietal but also the respective sensory cortices [6-15]. This suggests that asymmetric alpha modulation could act as a filter mechanism by modulating sensory gain already in early processing stages.

In addition, a prominent line of research focuses on the role of low-frequency (1–8 Hz) neural activity in auditory and, broadly speaking, perisylvian cortex in the selective representation of speech input ("speech tracking"). It is assumed that slow cortical dynamics temporally align with (or "track'') auditory input signals to allow for a prioritized neural representation of behaviorally-relevant sensory information [16-19]. In human speech comprehension, a key finding is the preferential neural tracking of attended compared to ignored speech in superior temporal brain areas close to auditory cortex [20-24].

However, with few exceptions [6], these two proposed auditory filter strategies have been studied independently of one another [but see 25,26 for recent results on visual attention]. Also, they have often been studied using tasks that are difficult to relate to natural, conversation-related listening situations [27,28].

We thus lack understanding whether or how modulations in lateralized alpha power and the neural tracking of attended and ignored speech in wider auditory cortex interact in the service of successful listening behavior. At the same time, few studies using more real-life listening and speech-tracking measures were able to explicitly address the functional relevance of the discussed neural filter strategies, that is, their potency to explain and predict behavioral listening success [22].

In the present EEG study, we aim at closing these gaps by leveraging the statistical power and representativeness from a large, age-varying participant sample. We use a novel dichotic listening paradigm to enable a synoptic look at concurrent changes in auditory alpha power and neural speech tracking at the single-trial level. More specifically, our linguistic variant of a classic Posner paradigm [29] emulates a challenging dual-talker listening situation in which speech comprehension is supported by two different listening cues [30]. These cues

encourage the use of two complementary cognitive strategies to improve comprehension: A spatial-attention cue guides auditory attention in space, whereas a semantic cue affords more specific semantic predictions of upcoming speech. Previous research has shown that the sensory analysis of speech and, to a lesser degree, the modulation of alpha power are influenced by the availability of higher-order linguistic information [31-35].

Varying from trial to trial, both cues were presented either in an informative or uninformative version, allowing us to examine relative changes in listening success and in the modulation of neural measures thought to enable auditory selective attention. Based on the neural and behavioral results, we focused on four research questions (see Fig. 1). Note that in pursuing these, we take into account additional notorious influences on listening success and its supporting neural strategies. These include age, hearing loss, or hemispheric asymmetries in speech processing due to the well-known right-ear advantage [36,37].

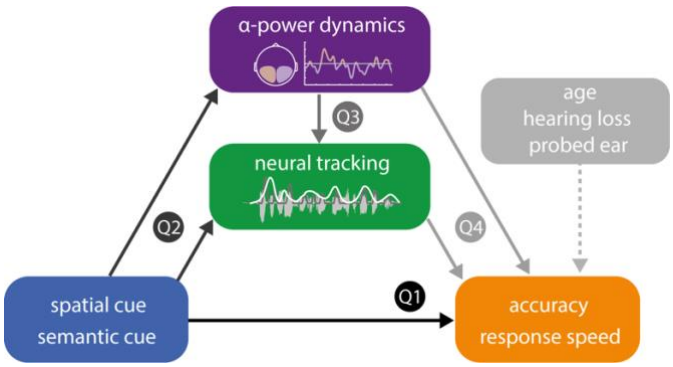

Figure 1. Schematic illustration of the research questions addressed in the present study. The dichotic listening task manipulated the attentional focus and semantic predictability of upcoming input using two separate visual cues. We investigated whether informative cues would enhance behavioral performance (Q1). In line with (Q2), we also examined the degree to which listening cues modulated the two auditory neural measures of interest: neural tracking and lateralization of auditory alpha power. Finally, we assessed (Q3) the co-variation of neural measures, and (Q4) their potency in predicting behavioral performance. Furthermore, we controlled for

First, we predicted that informative listening cues should increase listening success: These cues allow the listener to deploy auditory selective attention (compared to divided attention), and to generate more specific (compared to only general) semantic predictions, respectively.

Second, we asked how the different cue–cue combinations would modulate the two key neural measures—alpha power lateralization and neural speech tracking. We expected to replicate previous findings of increased alpha power lateralization and stronger tracking of the to-be-attended speech signal under selective (compared to divided) spatial attention.

Third, an important and often neglected research question pertains to a direct, trial-by-trial relationship of these two candidate neural measures: Do changes in alpha power lateralization impact the degree to which attended and ignored speech signals are neurally tracked by low-frequency cortical responses?

Our final and overarching research question is the most arguably the most relevant one for all translational aspects of auditory attention: Would alpha power and neural speech tracking of attended and ignored speech allow us at all to explain and predict single-trial behavioral success in this challenging listening situation? While tacitly assumed in most studies that deem these filter mechanisms "attentional", this has remained a surprisingly open empirical question.

Results

We recorded EEG from an age-varying sample (N=155) of healthy middle-aged and older adults (age=39–80 years) who performed a challenging dichotic listening task[30]. In this linguistic variant of a classic Posner paradigm, participants listened to two competing sentences spoken by the same female talker, and were asked to identify the final word in one of the two sentences. Importantly, sentence presentation was preceded by two visual cues. First, a spatial-attention cue encouraged the use of either selective or divided attention by providing informative or uninformative instructions about the to-be-attended, and thus later probed, ear. The second cue indicated the semantic category that applied to both final target words. The provided category could represent a general or specific level, thus allowing for more or less precise prediction of the upcoming speech signal (Fig. 2a, b). While this listening task does not tap into the most naturalistic forms of speech comprehension, it provides an ideal scenario to probe the neural underpinnings of successful selective listening.

To address our research questions outlined above, we employed two complementary analysis strategies: First, we aimed at *explaining* behavioral task performance in a challenging listening situation, and the degree to which it is leveraged by two key neural measures of auditory attention: the lateralization of 8–12 Hz alpha power, and the neural tracking of attended and ignored speech by slow cortical responses. Using generalized linear mixed-effects models on single-trial data, we investigate the cue-driven modulation of behavior and neural measures, the interaction of neural measures, and their (joint) influence on selective listening success.

Second, we went beyond common explanatory statistical analyses and used cross-validated regularized regression to explicitly probe the potency of cue-driven modulation of behavior and neural measures to *predict* single-trial task performance [38].

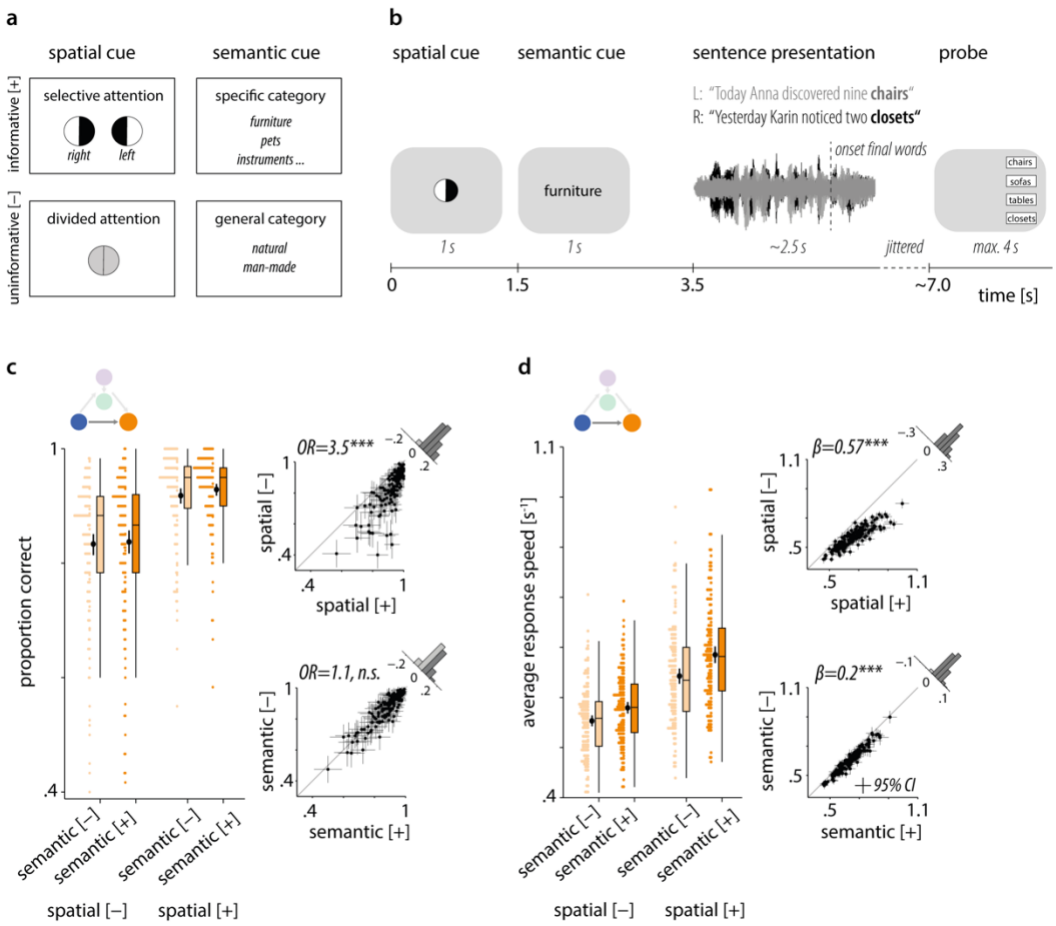

Figure 2. Experimental design and behavioral benefit from informative cues.
(a) Visualization of the employed 2×2 design [39]. The level of the spatial and semantic cue differed on a trial-by-trial basis. Top row shows the informative [+] cue levels, bottom row the uninformative [–] cue levels.
(b) Schematic representation of the temporal order of events for a given trial. Successive display of the two visual cues precedes the dichotic presentation of two sentences spoken by the same female talker. After sentence presentation, participants had to select the final word from four alternative words.
(c) Grand average and individual results of accuracy shown per cue-cue combination. Colored dots are individual (N=155) trial-averages, black dots and vertical lines show group means ± bootstrapped 95% confidence intervals (left side). Individual cue benefits displayed separately for the two cues (top: spatial cue, bottom: semantic cue). Black dots indicate individual trial-averages ± bootstrapped 95 % confidence intervals. Histograms show the distribution of the difference for informative vs. uninformative levels. OR: Odds ratio parameter estimate from generalized linear mixed-effects models (right side); *** = $p < .001$.
(d) As in (c) but shown for response speed; β: slope parameter estimate from general linear mixed-effects models.

The following figure supplements are available for figure 2:
Figure supplement 1. (a) Histogram showing age distribution of N=155 participants across 2-year age bins. (b) Individual (N=155) and mean air-conduction thresholds (PTA) averaged across the left and right ear, and four age groups.
Figure supplement 2. Analysis of response types split into correct responses, spatial stream confusions and random errors.

Informative spatial cue improves listening success

For behavioral performance, we tested the impact of informative versus uninformative listening cues on listening success. Overall, participants achieved a mean accuracy of 87.8 % ± 9.1 % with a mean reaction time of 1742 ms ± 525 ms; as response speed: 0.62 s^{-1} ± 0.17.

As shown in Figure 2c/d, the behavioral results varied between the different combinations of listening cues. As expected, the analyses revealed a strong behavioral benefit of informative compared to uninformative spatial-attention cues. In selective-attention trials, participants responded more accurately and faster (accuracy: generalized linear mixed-effects model (GLMM); *odds ratio (OR)*=3.45, *std. error (SE)* =.12, $p<.001$; response speed: linear mixed-effects model (LMM); β=.57, SE=.04, $p<.001$). That is, when cued to one of the two sides, participants responded on average 261 ms faster and their probability of giving a correct answer increased by 6 %.

Participants also responded generally faster in trials in which they were given a specific and thus informative semantic cue (LMM; β=.2, *SE*=.03, $p<.001$), most likely reflecting a semantic priming effect that led to faster word recognition.

As in a previous fMRI implementation of this task [30], we did not find evidence for any interactive effects of the two listening cues on either accuracy (GLMM; *OR*=1.3, *SE*=.22, p=.67) or response speed (LMM; β=.09, *SE*=.06, p=.41). Moreover, the breakdown of error trials revealed a significantly higher proportion of spatial stream confusions (6 % ±8.3 %) compared to random errors (3 % ±3.4 %; paired t-test on logit-transformed proportions: t_{155} = 6.53, $p<.001$; see also Fig. 2–supplement 2). The increased rate of spatial stream confusions (i.e., responses in which the last word of the to-be-ignored sentence was chosen) attests to the distracting nature of dichotic sentence presentation and thus heightened task difficulty.

Spatial attention modulates both alpha power lateralization and neural speech tracking in auditory cortex

In line with our second research question, following source projection of EEG data, we tested the hypothesis that the presence of an informative spatial attention cue would lead to reliable modulation of both alpha power and neural speech tracking in and around auditory cortex. To this end, we focused on our analyses to an *a priori* defined auditory region of interest (ROI; see Fig. 3–supplement1 and Methods for details).

For alpha power, we expected an attention-modulated lateralization caused by the concurrent decrease in power contralateral and increase in power ipsilateral to the focus of attention. For the selective neural tracking of speech, we expected an increase in the strength of neural tracking of attended but potentially also of ignored speech[cf. 40] in selective-attention compared to divided-attention trials.

For the lateralization of alpha power we relied on an established neural index that compares alpha power ipsi- and contralateral to the to-be-attended ear (in divided-attention trials: probed ear) to provide a temporally-resolved single-trial measure of alpha power lateralization [ALI = (α-power$_{ipsi}$ − α-power$_{contra}$) / (α-power$_{ipsi}$ + α-power$_{contra}$)] [12,22].

As shown in Figure 3a, the instruction to pay attention to one of the two sides elicited a pronounced lateralization of 8–12 Hz alpha power within the auditory ROI: We found a strong but transient increase in lateralization in response to an informative spatial-attention cue. After a subsequent break-down of the lateralization during semantic-cue presentation, it re-emerged in time for the start of dichotic sentence presentation and peaked again during the presentation of the task-relevant sentence-final words.

As expected, the statistical analysis of alpha power lateralization during sentence presentation (time window: 3.5–6.5 s) revealed a significant modulation by attention that was additionally influenced by the probed ear (LMM; spatial cue x probed ear: β=.13, SE=.02, p<.001). Follow-up analysis showed a significant increase in alpha power lateralization in selective-attention compared to divided attention-trials when participants were instructed to pay attention to the right ear (LMM, simple effect spatial cue: β=.12, SE=.01, p<.001). No such difference was found for attend-left/ignore-right trials (LMM, simple effect spatial cue: β=–.01, SE=.01, p=.63; see Table S3 for full model details).

Notably, we did not find any evidence for a modulation of alpha lateralization during sentence presentation by the semantic cue nor any joint influence of the spatial and semantic cue (LMM; semantic cue main effect: β=–.007, SE=.01, p=.57, interaction spatial x semantic cue: β=–.02, SE=.02, p=.42).

We ran additional control analyses for the time window covering the spatial cue (0–1 s) and the sentence-final target word (5.5–6.5 s), respectively. For the target word, results mirrored those observed for the entire duration of sentence presentation (see Table S8). Alpha power lateralization during spatial-cue presentation was modulated by independent effects of the spatial-attention cue and probed ear (See Table S9 for details). None our analyses revealed an effect of age or hearing ability on the strength of alpha power lateralization (all p-values>.14).

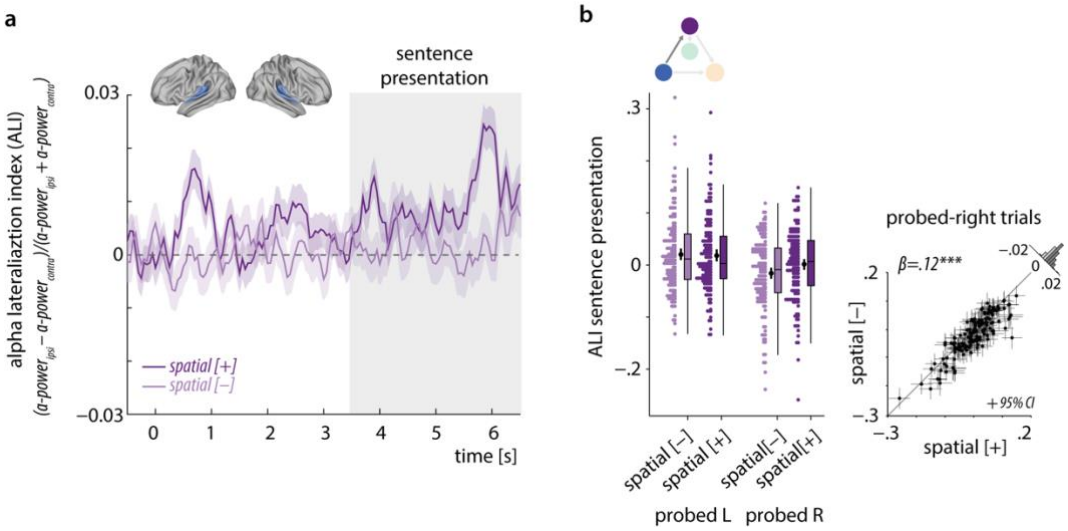

Figure 3. Informative spatial cue elicits increased alpha power lateralization before and during speech presentation.

(a) Attentional modulation of 8–12 Hz auditory alpha power shown throughout the trial for all N=155 participants. Brain models indicate the spatial extent of the auditory region of interest (shown in blue). Purple traces show the temporally resolved alpha lateralization index (ALI) for the informative (dark color) and the uninformative spatial cue (light color), each collapsed across the semantic cue levels. Positive values indicate relatively higher alpha power in the hemisphere ipsilateral to the attended/-probed sentence compared to the contralateral hemisphere. Shaded grey area show the temporal window of sentence presentation.

(b) Strength of the alpha lateralization index (ALI) during sentence presentation (3.5–6.5 s) shown separately per spatial cue condition and probed ear as indicated by a significant interaction in the corresponding mixed-effects model (left plot). Colored dots represent trial-averaged individual results, black dots and error bars indicate the grand-average and bootstrapped 95% confidence intervals. For attend-right/ignore-left trials we observed a significant increase in alpha power lateralization (right plot). Black dots represent individual mean ALI values ± bootstrapped 95 % confidence intervals. Histogram shows distribution of differences in ALI for informative vs. uninformative spatial-cue levels. β: slope parameter estimate from the corresponding general linear mixed-effects model; ***=p<.001.

In close correspondence to the alpha-power analysis, we investigated whether changes in attention or semantic predictability would modulate the neural tracking of attended and ignored speech. We used linear backward ('decoding') models to reconstruct the onset envelopes of the to-be-attended and ignored sentences (for simplicity hereafter referred to as *attended* and *ignored*) from neural activity in the auditory ROI. We compared the reconstructed envelopes to the envelopes of the actually presented sentences and used the resulting Pearson's correlation coefficients as a fine-grained, single-trial measure of neural tracking strength. Reconstruction models were trained on selective-attention trials, only, but then utilized to reconstruct attended (probed) and ignored (unprobed) envelopes for both attention conditions (see Methods and Fig. 4a along with supplement 1 for details).

In line with previous studies, but here observed particularly for right-ear inputs processed in the left auditory ROI, we found increased encoding of attended compared to ignored speech in the time window covering the $N1_{TRF}$ and $P2_{TRF}$

component (see Fig. 4b). Notably, within both left and right auditory cortex, for later processing stages (>300 ms), deflections in the forward-transformed temporal response functions (TRFs) for attended and ignored speech follow trajectories of opposite polarity with overall stronger deflections for the ignored sentence.

Further attesting to the validity of our reconstruction models, we found that envelopes reconstructed by the attended decoder model were more similar to the envelope of the actually presented to-be-attended sentence than to that of the to-be-ignored sentence (see Fig. 4b, bottom left plot). As reported in previous studies, we also found significantly stronger neural tracking of attended (mean r=.06, range: –.02–.15) compared to ignored speech (ignored: mean r=.03, range: –.03–.1; LMM, β=.022, SE=.002, p<.001; see Fig. 4b, bottom right plot).

As shown in Figure 4c/d, the neural tracking of attended and ignored envelopes was modulated by attention. Note that for sake of consistency, for divided-attention trials, we likewise refer to the reconstruction of the *probed* envelope as the *attended* envelope, and to the reconstruction of the *unprobed* envelope as the *ignored* envelope despite the absence of corresponding instructions.

Giving credence to the suggested role of selective neural speech tracking as a neural filter strategy, we found stronger neural tracking of the attended envelope following an informative spatial-attention cue as compared to an uninformative one (LMM; β=.06, SE=.01, p<.001; see Table S4 for full model details). The semantic predictability of the sentence-final words, however, did not modulate the neural tracking of the attended envelope (LMM; β=.02, SE=.01, p=.26). We also found stronger tracking in trials where later probed sentence started ahead of the distractor sentence (LMM: β=.13, SE=.01, p<.001; see Methods and Supplements for details). The latter effect may indicate that the earlier onset of a sentence increased its bottom-up salience and attentional capture. The observed effect may also suggest that participants focused their attention to one of the two sentences already during early parts of auditory presentation.

The analysis of ignored-speech tracking revealed a qualitatively similar pattern of results: we found stronger tracking of ignored speech following an informative compared to an uninformative spatial-attention cue (LMM; spatial cue: β=.04, SE=.01, p=.005). We additionally observed stronger tracking of the ignored (under divided attention: unprobed) sentence when it was played to the left ear (LMM; main effect probed ear: β=.17, SE=.02, p<.001; see Table S5 for full model details).

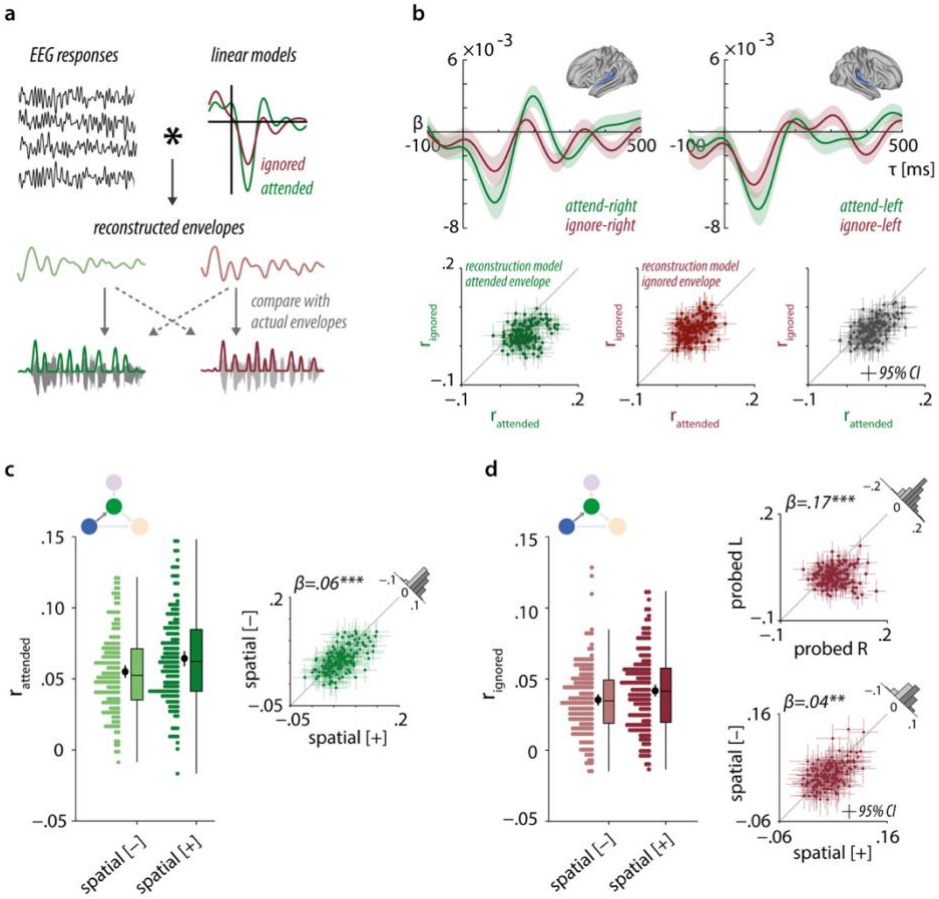

Figure 4. Neural speech tracking of attended and ignored sentences.
(a) Schematic representation of the linear backward model approach used to reconstruct onset envelopes on the basis of recorded EEG responses. Following the estimation of linear backward models using only selective-attention trials (see Fig. 4–supplement 1 for details), envelopes of attended and ignored sentences were reconstructed via convolution of EEG responses with the estimated linear backward models in all functional parcels within the auditory ROI. Reconstructed envelopes were compared to the envelopes of presented sentences to assess neural tracking strength and reconstruction accuracy (see Supplements).
(b) Forward-transformed temporal response functions (TRFs) for attended (green) and ignored (red) speech in the selective-attention condition in the auditory ROI. Models are averaged across all N=155 participants, and shown separately for per hemisphere. Left panel compares the TRFs in the left auditory ROI to right-ear input, right panel compares the TRFs in the right auditory ROI left-ear input. Error bands indicate 95 % confidence intervals. Bottom row: Left plot shows the Pearson's correlation of the reconstructed attended envelope to the envelopes of the actually presented sentences, middle plot analogously for the reconstructed ignored envelope. Right plot compares the neural tracking strength of the attended and ignored speech.
(c) Attentional modulation in the neural tracking of attended (under divided attention: probed) speech. Colored dots represent trial-averaged individual results, black dots and error bars indicate the grand-average and bootstrapped 95 % confidence intervals. Single-trial analysis reveals that neural tracking of attended speech is stronger under selective attention (left plot and bottom right plot). Colored dots in 45° plot represent trial-averaged correlation coefficients per participant along with bootstrapped 95 % confidence intervals; ***= p<.001.
(d) as in (c) but for the neural tracking of ignored/unprobed speech; **=p<.01.

The following figure supplements are available for figure 4:
Figure supplement 1. Training and testing of envelope reconstruction models.
Figure supplement 2. Neural tracking strength per reconstruction model for divided attention condition (N=155).

Figure supplement 3. Decoding accuracy of attended and ignored speech shown per attention condition (N=155).

Changes in neural speech tracking are independent of concurrent changes in auditory alpha power lateralization

Our third major goal was to investigate whether the modulation of alpha power and neural tracking strength reflect two dependent neural strategies of auditory attention at all. To this end, using two single-trial linear mixed-effects models, we tested whether the neural tracking of the attended and ignored sentence could be explained by auditory alpha power dynamics (here, alpha power ipsi- and contralateral to the focus of attention/probed ear, that is, the relative signal modulations that constitute the single-trial alpha lateralization index). If modulations of alpha power over auditory cortices indeed act as a neural filter mechanism to selectively gate processing during early stages of sensory analysis then this would be reflected as follows: decreases in contralateral alpha power should lead to stronger neural tracking of the attended envelope whereas increases in ipsilateral alpha power should lead to weaker tracking of the ignored envelope (cf. Fig. 5a).

However, neither the tracking of the attended envelope (LMM; main effect alpha ipsi: β=−.007, *SE*=.006, *p*=.63, *Bayes factor (BF)* = .009; main effect alpha contra: β=−.01, *SE*=.006, *p*=.43, *BF*=.02; interaction ipsi × contra: β=−.003, *SE*=.004, *p*=.72, *BF*=.006) nor the tracking of the ignored envelope were correlated to changes in ipsi- or contralateral alpha power or their interaction (LMM; main effect alpha ipsi: β=.01, SE=.006, *p*=.35; *BF*=.02; main effect alpha contra: β=.003, SE=.006, *p*=.85, *BF*=.006; interaction ipsi × contra: β=−.006, SE=.004, *p*=.53, *BF*=.01; see Tables S6 and S7 for full model details). This notable absence of an alpha lateralization–speech tracking relationship was independent of the respective spatial-attention condition (LMM; all corresponding interaction terms with p-value >.38, BFs <.0076).

A control analysis relating alpha power modulation during spatial-cue presentation to neural speech tracking showed qualitatively similar results (see Table S10 for details).

Alternatively, selective neural tracking of speech might be driven not primarily by alpha-power modulations emerging in auditory cortices, but rather by those generated in domain-general attention networks in parietal cortex [41]. Additional control models included changes in alpha power within the inferior parietal lobule (see Table S11 and Fig. 5–supplement 1 for details) to test this hypothesis: We found a trend towards a negative relationship between contralateral alpha power and the neural tracking of the attended envelope (LMM; β=−.017, *SE*=.007, *p*=.082, *BF*=.1). In addition, we observed a significant joint effect of ipsi- and contralateral alpha power in inferior parietal cortex on the tracking of the ignored envelope (LMM; interaction alpha power ipsi x alpha power contra: β=−.011, *SE*=.004, *p*=.022, *BF*=*.17*): the strength of the positive correlation of

ipsilateral alpha power and ignored speech tracking increased with decreasing levels of contralateral alpha power.

A final control analysis included changes in visual alpha power and did not find any significant effects of alpha power on neural speech tracking (see Table S12).

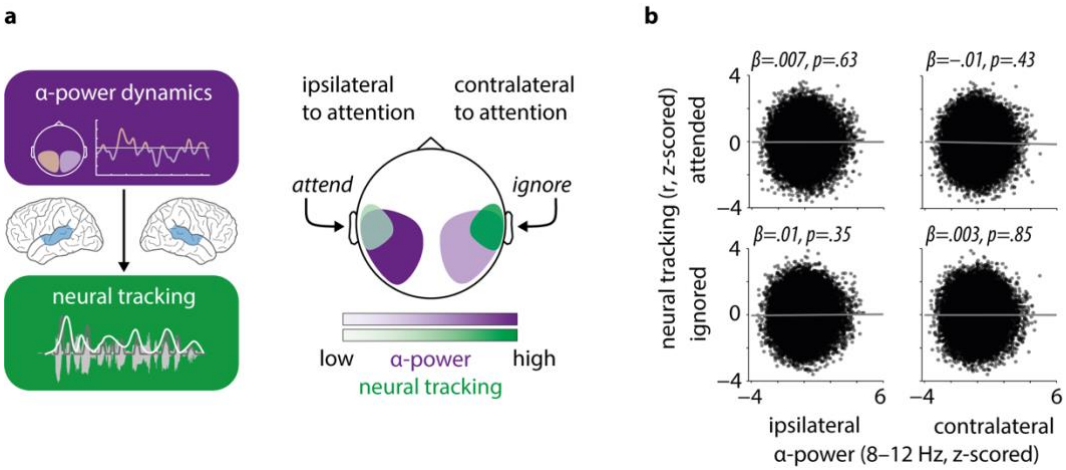

Figure 5. Independence of auditory alpha power dynamics and neural tracking.
(a) Hypothesized inverse relationship of changes in alpha power and neural tracking within the auditory ROI. Changes in alpha power are assumed to drive changes in neural tracking strength. Schematic representation of expected co-variation in the two neural measures during an attend-left trial.
(b) Independence of changes in neural tracking and alpha power in the auditory ROI during sentence processing as revealed by two separate linear mixed-effects models predicting the neural tracking strength of attended and ignored speech, respectively. Error band denotes 95 % confidence interval. Black dots represent single-trial observed values from N=155 participants (i.e., a total of k = 35675 trials).

Alpha power lateralization and neural tracking fail to explain single-trial listening success

Having established the functional independence of alpha power lateralization and speech tracking, the final and most important piece of our investigation becomes in fact statistically more tractable: If alpha power dynamics in auditory cortex (but also in inferior parietal and visual cortex), and neural speech tracking essentially act as two largely independent neural filter strategies, we can proceed to probe their relative functional relevance for behavior in a factorial-design fashion.

We employed two complementary analysis strategies to quantify the relative importance of neural filters for listening success: (i) using the same (generalized) linear mixed-effects models as in testing our first research question (Q1 in Fig.1), we investigated whether changes in task performance could be explained by the independent (i.e., as main effects) or joint influence (i.e., as an interaction) of neural measures; (ii) using cross-validated regularized regression we directly assessed the ability of neural filter dynamics to predict out-of-sample single-trial listening success.

With respect to response accuracy, our most important indicator of listening success, we did not find evidence for any direct effects of alpha power lateralization and the neural tracking of attended or ignored speech (GLMM; main effect alpha power lateralization (ALI): *OR*=1.0, *SE*=.02, *p*=.93; main effect attended speech tracking: *OR*=1.01, *SE*=.02, *p*=.93; main effect ignored speech tracking: *OR*=1.01, *SE*=.02, *p*=.82, see Fig. 6) nor for any joint effects (GLMM; ALI x attended speech tracking: *OR*=1.01, *SE*=.02, *p*=.82; ALI x ignored speech tracking: *OR*=.98, *SE*=.02, *p*=.72, see Fig. 6b for comparison of Bayes factors). Importantly, the absence of an effect did not hinge on differences across spatial-cue, semantic-cue or probed-ear levels as relevant interactions of the neural measures with these predictors were included in the model (all p-values > .05, see Table S1 for full details).

The analysis of response speed did not reveal any meaningful brain-behavior relationships, either. We did not find evidence for any direct effects of neural measures on response speed (LMM; alpha power lateralization (ALI): β=0.002, SE=.005, *p*=.9; attended speech tracking: β=.003, SE=.005, *p*=.73; ignored speech tracking: β=.00001, SE=.005, *p*=.99) nor for their joint influence (LMM; ALI x attended speech tracking: β=–0.005, SE=.004, *p*=.61; ALI x ignored speech tracking: β=.004, SE=.004, *p*=.61; see Table S2 for full details).

We ran two sets of control analyses: first, we tested for the influence of changes in auditory alpha power during the presentation of the sentence-final target word and the spatial-attention cue, respectively (see Tables S13–16). Next, we investigated whether modulation of alpha power lateralization during sentence presentation extracted from inferior parietal or visual cortex would influence accuracy or response speed (see Tables S17-20). However, all results were qualitatively in line with those of our main analyses.

In contrast to the missing brain-behavior link, but in line with the literature on listening behavior in aging adults [e.g. 42-44], the behavioral outcome was instead reliably predicted by age, hearing loss, and probed ear. We observed that participants' performance varied in line with the well-attested *right-ear advantage* (REA, also referred to as *left-ear disadvantage*) in the processing of linguistic materials [36,37]. More specifically, participants responded both faster and more accurately (response speed: LMM; β=.08, *SE*=.013, *p*<.001; accuracy: GLMM; *OR*=1.23, *SE*=.07, *p*=.017; see also Fig. 6 supplement 2) when they were probed on the last word presented to the right compared to the left ear.

Increased age led to less accurate and slower response (accuracy: GLMM; *OR*=.81, *SE*=.08, *p*=.033; response speed: LMM; β=–.15, *SE*=.03, *p*<.001). In contrast, increased hearing loss led to less accurate (GMM; *OR*=.75, *SE*=.08, *p*=.003) but not slower responses (LMM; β=–.05, *SE*=.08, *p*=.28).

To test whether alpha power lateralization and neural speech tracking would be related to behavior on a between-subjects level where their effects could also be more readily compared to the effects of age or PTA, we included separate

within- and between-subjects effects in simplified control models ([45]; see Fig. 6b and Methods for details).

We did not observe any between-subject effects of alpha power lateralization (LMM; OR=1.14, *SE*=.07, *p*=.11) or neural speech tracking (LMM; attended speech tracking: OR=1.12, *SE*=.08, *p*=.35; ignored speech tracking: OR=1.05, *SE*=.09, *p*=.74) on accuracy (see Table S21 and Fig. 6–supplement 1). For response speed, we found a significant between-subject effect of attended speech tracking (LMM; attended speech tracking: β=.013, *SE*=.005, *p*=.039) but no comparable effects for the other two neural measures (LMM; ignored speech tracking: β=.002, *SE*=.006, *p*= .89; ALI: β=.001, *SE*=.005, *p*=.89; see Table S22).

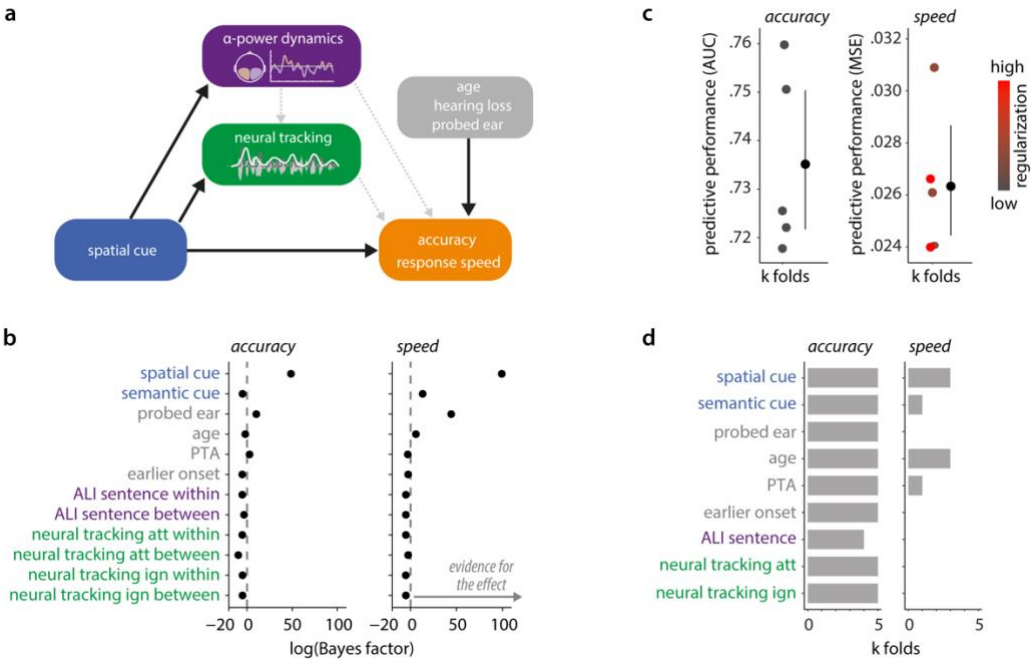

Figure 6. Results of explanatory and predictive analysis.

(a) Path diagram summarizing the results of our four research questions. Solid black arrows indicate significant effects, grey dashed lines the absence of an effect. An informative spatial cue led to a boost in both accuracy and response speed, as well as in alpha power lateralization and neural speech tracking. Neural dynamics varied independently of one another and of changes in behavior. Age, hearing loss and probed ear additionally influenced behavioral listening success.

(b) Bayes factors (BF) for individual predictors of accuracy (left) and response speed (right). BFs are derived by comparing the full model with a reduced model in which only the respective regressor was removed. Log-transformed BFs larger than zero provide evidence for the presence of a particular effect whereas BFs below zero provide evidence for the absence of an effect.

(c) Predictive performance of models of accuracy (left) and response speed (right) as estimated by the area under the receiver operating characteristic curve (AUC) and mean-squared error (MSE) metric, respectively. Colored dots show the performance in each of k=5 outer folds in nested cross-validation along with the mean and 95 % confidence interval shown on the right side of each panel. Color gradient indicates the amount of regularization applied.

(d) Variable selection in the k=5 final model shown here only for main effects. Horizontal bars indicate in how many of the final models a given effect was included (i.e., it had a non-zero coefficient). All main effects were included in the winning model of accuracy; for response speed only the main effects of spatial and semantic cue, age and PTA were included (see figure supplement 4 for the full model).

Neural filter mechanisms make limited contributions to overall good model performance in predicting behavior

Lastly, to complement the analysis of the explanatory power of neural filters for listening success, we explicitly scrutinized their potency to predict out-of-sample single-trial behavioral outcomes. To this end, based on the final explanatory model for accuracy and speed, we performed variable selection via cross-validated LASSO regression to test which subset of variables would allow for optimal prediction of single-trial task performance (see Fig. 6–supplement 3 and Methods for details).

This approach yields two important results: On the one hand, based on the k=5 cross-validated final models, it quantifies how well on average we can predict single-trial accuracy (expressed via the area under the receiver operating characteristic curve, AUC) and response speed (expressed via the mean-squared error, MSE). On the other hand, the comparison of terms included in the final models allows to assess how consistently a given effect was deemed important for prediction.

As shown in Figure 6c, the models of single-trial accuracy and response speed yielded overall high levels of predictive performance. We achieved a classification of accuracy with a mean AUC of 73.5 ± 0.019 %. This means that in classifying correct and incorrect responses a random trial with a correct response has a 73.5 % chance of being ranked higher by the prediction algorithm than a random incorrect trial. Single-trial prediction of response speed had an average prediction error (MSE) of 0.027 s^{-1} ± 0.003 that was thus approximately within one (squared) standard deviation of the average response speed.

For each predicted behavioral outcome, we estimated one winning model based on the full dataset by performing LASSO regression with the regularization parameter λ value of the best-performing final model. Across the k=5 final models, overall higher degrees of regularization were applied to the regression models predicting response speed than to those predicting accuracy, indicating that predictive performance hinged on the effects of fewer variables (see Fig. 6c along with supplement 4 and Table S23 and S24 for details).

As expected, the applied regularization resulted in the removal of several regressors from the winning predictive model. In line with the results of our explanatory analyses, we observed that predictors strongly correlated with changes in behavior were also consistently included in the final predictive models. As shown in Figure 6d for all direct effects (i.e., main effects), the spatial and semantic cue, as well as age and PTA contributed most consistently to the prediction of behavioral outcomes. While the direct effects of neural filter dynamics were retained in almost all final models of accuracy, many of their associated higher-order interactive terms were removed (see Table S23). However, even though the direct (and some of the joint) effects of neural measures were included by the variable selection procedure,

in the winning model fit to all available data, they were still not significantly correlated with the behavioral outcome.

For response speed, the direct effects of neural filter dynamics were not included in any of the final models, that is, they did not contribute at all to the overall good predictive performance.

In sum, we conclude that in this large, age-varying sample of middle-aged to older adults, single-trial fluctuations in neural filtering make limited contributions to the prediction of accuracy and do not have any meaningful role at all for the prediction of single-trial response speed. Instead, the behavioral outcome was reliably predicted by the effects of the spatial cue and age, and to a lesser degree by the semantic cue, probed ear, and hearing loss.

Discussion

The present study has utilized the power of a large, representative sample of middle-aged and older listeners to explicitly address the question of how two different neural filter strategies, typically studied in isolation, jointly shape listening success.

In our dichotic listening task, we source-localized the electroencephalogram, and observed systematic cue-driven changes within auditory cortex in alpha power lateralization and the neural tracking of attended and ignored speech.

These results do provide unprecedented large-sample support for their suggested roles as neural signatures of selective attention. Additionally, they address two important but still insufficiently answered questions: How do these two neural filter strategies relate to one another, and how do they impact listening success in a demanding real-life listening situation?

First, when related at the single-trial, single-subject level, we found the modulation of the two neural attentional filter mechanisms to be statistically independent, underlining their functional segregation and speaking to two distinct neurobiological implementations.

Second, an informative spatial-attention cue not only boosted both neural measures but also consistently boosted behavioral performance. However, we found changes at the neural and at the behavioral level to be generally unrelated. In fact, the effect of neural measures on single-trial task performance was reliably outweighed by the impact of age, hearing loss, or probed ear.

Finally, a cross-validated predictive model showed that, despite their demonstrated validity as measures of deployed attention, trial-by-trial neural filter states have a surprisingly limited role to play in predicting out-of-sample listening success.

What does the missing direct link from brain to behavior teach us?

With the present study, we have explicitly addressed the often overlooked question of how trial-by-trial neural dynamics impact behavior, here single-trial listening success [46-49]. The observed cue-driven modulation of behavior, alpha lateralization, and the tracking of attended and ignored speech allowed us investigate this question.

Notably, despite an unprecedentedly large sample of almost 160 listeners; an adequate number of within-subject trials; and a sophisticated linear-model approach, we did not find any evidence for a direct or indirect (i.e., moderating) influence of alpha power lateralization or of the neural tracking of the attended or ignored sentence on behavior (see Fig. 6).

At first glance, the glaring absence of any brain–behavior relationship may seem puzzling given the reliable spatial-cue effects on both behavior and neural measures. However, previous studies have provided only tentative support for the

behavioral relevance of the two neural filter solutions, and the observed brain–behavior relations do not always converge.

The available evidence for the behavioral relevance of selective speech tracking is generally sparse given the emphasis on more complex naturalistic language stimuli [50,51]. While these stimuli provide a window onto the most natural forms of speech comprehension, they do not afford fine-grained measures of listening behavior. This makes it particularly challenging to establish a direct link of differential speech tracking with behavior [20,21,23,52-55]. Nevertheless, there is preliminary evidence for the behavioral benefit of a more pronounced cortical representation of to-be-attended speech [22].

Previous studies on lateralized alpha oscillations suggest that their functional relevance is complicated by a number of different factors.

First of all, most of the evidence associating increased levels of alpha power lateralization with better task performance in spatial-attention tasks stems from studies on younger adults [9,12,56,57]. By contrast, the establishment of such a consistent relationship in middle-age and older adults is obscured by considerable variability in age-related changes at the neural (and to some extent also behavioral) level [11,58-61]. We here found the fidelity of alpha power lateralization unchanged with age (see also [11]. Other studies on auditory attention have observed diminished and less sustained alpha power lateralization for older compared to younger adults that were to some extend predictive of behavior [58,62].

Second, previous findings and their potential interpretation differ along (at least) two dimensions: (i) whether the functional role of alpha power lateralization during attention cueing, stimulus anticipation, or stimulus presentation is studied [6,58,63], and (ii) whether the overall strength of power lateralization or its stimulus-driven rhythmic modulation is related to behavior [cf. 9,11]. Depending on these factors, the observed brain–behavior relations may relate to different top-down and bottom-up processes of selective auditory attention.

Third, as shown in a recent study by Wöstmann et al. [63] on younger adults, the neural processing of target and distractor are supported by two uncorrelated lateralized alpha responses emerging from different neural networks. Notably, their results provide initial evidence for the differential behavioral relevance of neural responses related to target selection and distractor suppression.

Lastly, the present study deliberately aimed at establishing a brain–behavior relationship at the most fine-grained trial-by-trial, single-subject level. However, many of the previous studies relied on coarsely binned data (e.g. correct vs. incorrect trials) or examined relationships at the between-subjects level, that is, from one individual to another. By contrast, effects observed at the single-trial level are generally considered the most compelling evidence that cognitive neuroscience is able to provide [64].

In sum, the current large sample with a demonstrably good signal-to-noise ratio also at the trial-to-trial or within-subjects level is unable to discern either form

of relationship. This result should give the translational ambition of manipulating neural filter mechanisms as a way to improve speech comprehension some pause.

How important to auditory attention is the neural representation of ignored speech?

Early studies of neural speech tracking in multi-talker scenarios have particularly emphasized the prioritized cortical representation of attended speech as a means to selectively encode behaviorally relevant information [20-23,65]. Based on this view, selective attention to one of our two dichotically presented sentences should primarily lead to an increase in the neural tracking of attended, and if anything, a decrease in the neural tracking of ignored speech.

However, our findings only partially agree with this hypothesis: In line with previous results, under selective attention, we observed a significantly stronger tracking of attended compared to ignored speech. At the same time, however, cortical responses during later processing stages (> 300 ms) suggest a heightened and diverging neural representation of to-be-ignored speech (see Fig. 4b). This was mirrored by an increase in the neural tracking strength of both attended and ignored speech in selective-attention (compared to divided-attention) trials.

How might the increased neural representation of ignored sounds impact the fidelity of auditory selective attention? At least two scenarios seem possible: on the one hand, stronger neural responses to irrelevant sounds could reflect a "leaky" attentional filter that fails to sufficiently suppress distracting information. On the other hand, it may reflect the implementation of a better calibrated, "late-selection" filter for ignored speech that selectively gates information flow [66,67].

Largely compatible with our findings, a recent, source-localized EEG study [40] provided evidence for active suppression of to-be-ignored speech via its enhanced cortical representation. Importantly, this amplification in the neural representation of to-be-ignored speech was found for the most challenging condition in which the distracting talker was presented at a higher sound level than the attended and was thus perceptually more dominant.

This finding allows for the testable prediction that the engagement of additional neural filter strategies supporting the suppression of irrelevant sounds is triggered by particularly challenging listening situation. Several key characteristics of our study may have instantiated such a scenario: (i) both sentences were spoken by the same female talker, (ii) had a similar sentence structure and thus prosodic contour, (iii) were presented against the background of speech-shaped noise at 0 dB SNR, and (iv) were of relatively short duration (~2.5 s). It thus seems plausible that the successful deployment of auditory selective attention in our task may have relied on the recruitment of additional filter strategies that depend on the neural fate of both attended and ignored speech [68-72].

Is the modulation of lateralized alpha power functionally connected to the selective neural tracking of speech?

In our study, we investigated attention-related changes in alpha power dynamics and in the neural tracking of speech signals that (i) involve neurophysiological signals operating at different frequency regimes, (ii) are assumed to support auditory attention by different neural strategies, and (iii) are typically studied in isolation [3,19]. Here, when focusing on changes in neural activity within auditory areas, we found both neural filter strategies to be impacted by the same spatial-attention cue.

This begs the questions of whether changes in alpha power, assumed to reflect a neural mechanisms of controlled inhibition, and changes in neural speech tracking, assumed to reflect a sensory-gain mechanism, are functionally connected signals [16,17]. Indeed, there is preliminary evidence suggesting that these two neural filter strategies may exhibit a systematic relationship [6,11,27,28,73]. However, our fine-grained trial-by-trial analysis revealed independent modulation of alpha power and neural speech tracking. Therefore, the present results speak against a consistent, linear relationship of neural filter strategies. We see instead compelling evidence for the coexistence of two complementary but independent neural solutions to the implementation of auditory selective attention for the purpose of speech comprehension.

Our results are well in line with recent reports of independent variation in alpha band activity and steady-state (SSR) or frequency following responses (FFR) in studies of visual spatial attention [25,26,74]. Additionally, the inclusion of single-trial alpha power lateralization as an additional training feature in a recent speech-tracking study failed to improve the decoding of attention [75].

The observed independence of neural filters may be explained by our focus on their covariation within auditory cortex while the controlled inhibition of irrelevant information via alpha power modulation typically engages areas in the parietal-occipital cortex. However, our control analyses relating neural speech tracking to modulation in alpha power in those areas did not find strong support for the interplay of neural filters, either.

These findings also agree with results from our recent fMRI study on the same listening task and a subsample of the present participant cohort [30]. The graph-theoretic analysis of cortex-wide networks revealed a network-level adaption to the listening task that was characterized by increased local processing and higher segregation. Crucially, we observed a reconfigured composition of brain networks that involved brain areas in auditory and superior temporal areas generally associated with the neural tracking of speech but not those typically engaged in the controlled inhibition of irrelevant information via top-down regulation of alpha power [76-79].

Taken together with findings from previous electrophysiological studies [6,27,28], our results provide strong evidence in favor of a generally existent functional tradeoff between attentional filter mechanisms [80]. However, we

suggest that the precise nature of this interplay of attentional filter mechanisms hinges on a number of factors such as the particular task demands, and potentially on the level of temporal and/or spatial resolution at which the two neural signatures are examined [28].

Conclusion

In a large, representative sample of adult listeners, we have provided evidence that single-trial listening success in a challenging, dual-talker acoustic environment cannot be easily explained by the individual (or joint) influence of two independent neurobiological filter solutions—alpha power lateralization and neural speech tracking. Supporting their interpretation as neural signatures of spatial attention, we find both neural filter strategies engaged following a spatial-attention cue. However, there was no direct link between neural and ensuing behavioral effects. Our findings highlight the level of complexity associated with establishing a stable brain–behavior relationship in the deployment of auditory attention. They should also temper simplified account on the predictive power of neural filter strategies for behavior as well as translational neurotechnological advances that build on them.

Materials and Methods

Participants and procedure

A total of *N* = 155 right-handed German native speakers (median age 61 years; age range 39–80 years; 62 males; see Fig. 1–supplement 1 for age distribution) were included in the analysis. All participants are part of a large-scale study on the neural and cognitive mechanisms supporting adaptive listening behavior in healthy middle-aged and older adults ("The listening challenge: How ageing brains adapt (AUDADAPT)"; https://cordis.europa.eu/project/rcn/197855_en.html). Handedness was assessed using a translated version of the Edinburgh Handedness Inventory[81]. All participants had normal or corrected-to-normal vision, did not report any neurological, psychiatric, or other disorders and were screened for mild cognitive impairment using the German version of the 6-Item Cognitive Impairment Test (6CIT [82]).

As part of our large-scale study on adaptive listening behavior in healthy aging adults, participants also underwent a session consisting of a general screening procedure, detailed audiometric measurements, and a battery of cognitive tests and personality profiling (see ref. [11] for details). This session always preceded the EEG recording session. Only participants with normal hearing or age-adequate mild-to-moderate hearing loss were included (see Fig. 1–supplement 2 for individual audiograms). As part of this screening procedure an additional 17 participants were excluded prior to EEG recording due to non-age related hearing loss or a medical history. Three participants dropped out of the study prior to EEG recording and an additional 9 participants were excluded from analyses after EEG recording: three due to incidental findings after structural MR acquisition, and six due to technical problems during EEG recording or overall poor EEG data quality. Participants gave written informed consent and received financial compensation (8€ per hour). Procedures were approved by the ethics committee of the University of Lübeck and were in accordance with the Declaration of Helsinki.

Dichotic listening task

In a recently established [30] linguistic variant of a classic Posner paradigm[29], participants listened to two competing, dichotically presented sentences. They were probed on the sentence-final word in one of the two sentences. Critically, two visual cues preceded auditory presentation. First, a spatial-attention cue either indicated the to-be-probed ear, thus invoking selective attention, or did not provide any information about the to-be-probed ear, thus invoking divided attention. Second, a semantic cue specified a general or a specific semantic category for the final word of both sentences, thus allowing to utilize a semantic prediction. Cue levels were fully crossed in a 2×2 design and presentation of cue combinations varied on a trial-by-trial level (Fig. 2a). The trial structure is

exemplified in Figure 2b. Details on stimulus construction and recording can be found in the Supplemental Information.

Each trial started with the presentation of a fixation cross in the middle of the screen (jittered duration: mean 1.5 s, range 0.5–3.5 s). Next, a blank screen was shown for 500 ms followed by the presentation of the spatial cue in the form of a circle segmented equally into two lateral halves. In selective-attention trials, one half was black, indicating the to-be-attended side, while the other half was white, indicating the to-be-ignored side. In divided-attention trials, both halves appeared in grey. After a blank screen of 500 ms duration, the semantic cue was presented in the form of a single word that specified the semantic category of both sentence-final words. The semantic category could either be given at a general (natural vs. man-made) or specific level (e.g. instruments, fruits, furniture) and thus provided different degrees of semantic predictability. Each cue was presented for 1000 ms.

After a 500 ms blank-screen period, the two sentences were presented dichotically along with a fixation cross displayed in the middle of the screen. Finally, after a jittered retention period, a visual response array appeared on the left or right side of the screen, presenting four word choices. The location of the response array indicated which ear (left or right) was probed. Participants were instructed to select the final word presented on the to-be-attended side using the touch screen. Among the four alternatives were the two actually presented nouns as well as two distractor nouns from the same cued semantic category. Note that because the semantic cue applied to all four alternative verbs, it could not be used to post-hoc infer the to-be-attended sentence-final word.

Stimulus presentation was controlled by PsychoPy [83]. The visual scene was displayed using a 24" touch screen (ViewSonic TD2420) positioned within an arm's length. Auditory stimulation was delivered using in-ear headphones (EARTONE 3A) at sampling rate of 44.1 kHz. Following instructions, participants performed a few practice trials to familiarize themselves with the listening task. To account for differences in hearing acuity within our group of participants, individual hearing thresholds for a 500-ms fragment of the dichotic stimuli were measured using the method of limits. All stimuli were presented 50 dB above the individual sensation level. During the experiment, each participant completed 60 trials per cue-cue condition, resulting in 240 trials in total. The cue conditions were equally distributed across six blocks of 40 trials each (~ 10 min) and were presented in random order. Participants took short breaks between blocks.

Behavioral data analysis

We evaluated participants' behavioral performance in the listening task with respect to accuracy and response speed. For the binary measure of accuracy, we excluded trials in which participants failed to answer within the given 4-s response window ('timeouts'). Spatial stream confusions, that is trials in which the sentence-final word of the to-be-ignored speech stream were selected, and random errors

were jointly classified as incorrect answers. The analysis of response speed, defined as the inverse of reaction time, was based on correct trials only. Single-trial behavioral measures were subjected to (generalized) linear mixed-effects analysis and regularized regression (see Statistical analysis).

EEG data acquisition and preprocessing

Participants were seated comfortably in a dimly-lit, sound-attenuated recording booth where we recorded their EEG from 64 active electrodes mounted to an elastic cap (Ag/AgCl; ActiCap / ActiChamp, Brain Products, Gilching, Germany). Electrode impedances were kept below 30 kΩ. The signal was digitized at a sampling rate of 1000 Hz and referenced on-line to the left mastoid electrode (TP9, ground: AFz). Before task instruction, 5-min eyes-open and 5-min eyes-closed resting state EEG was recorded from each participant.

For subsequent off-line EEG data analyses, we used the EEGlab [84] (version 14_1_1b) and Fieldtrip toolboxes [85] (version 2016-06-13), together with customized Matlab scripts. Independent component analysis (ICA) using EEGlab's default runica algorithm was used to remove all non-brain signal components including eye blinks and lateral eye movements, muscle activity, heartbeats and single-channel noise. Prior to ICA, EEG data were re-referenced to the average of all EEG channels (average reference). Following ICA, trials during which the amplitude of any individual EEG channel exceeded a range of 200 microvolts were removed.

EEG data analysis

Sensor-level analysis

The preprocessed continuous EEG data were high-pass-filtered at 0.3 Hz (finite impulse response (FIR) filter, zero-phase lag, order 5574, Hann window) and low-pass-filtered at 180 Hz (FIR filter, zero-phase lag, order 100, Hamming window). The EEG was cut into epochs of –2 to 8 s relative to the onset of the spatial-attention cue to capture cue presentation as well as the entire auditory stimulation interval.

For the analysis of changes in alpha power, EEG data were down-sampled to f_s=250 Hz. SpectrotemporalSpectro-temporal estimates of the single-trial data were then obtained for a time window of −0.5 to 6.5 s (relative to the onset of the spatial-attention cue) at frequencies ranging from 8 to 12 Hz (Morlet's wavelets; number of cycles = 6).

For the analysis of the neural encoding of speech by low-frequency activity, the continuous preprocessed EEG were down-sampled to f_s=125 Hz and filtered between f_c=1 and 8 Hz (FIR filters, zero-phase lag, order: $8f_s/f_c$ and $2f_s/f_c$, Hamming window). The EEG was cut to yield individual epochs covering the presentation of auditory stimuli, beginning at noise onset until the end of auditory presentation.

Individual epochs were z-scored prior to submitting them to regression analysis (see below).

EEG source and forward model construction

Twenty-seven participants had completed functional and structural magnetic resonance imaging (MRI) from performing the same experiment on a separate session. For these participants individual EEG source and forward models were created on the basis of each participant's T1-weighted MRI image (Siemens MAGNETOM Skyra 3T scanner; 1-mm isotropic voxel). The T1 image of one female and one male participant were used as templates for the remaining participants.

First, anatomical images were resliced (256x256x256 voxels) and the CTF convention was imposed as their coordinate system. Next, the cortical surface of each individual (or template) T1 image was constructed using the FreeSurfer function *recon-all*. The result was used to generate a cortical mesh in accordance with the Human Connectome Project (HCP) standard atlas template (available at https://github.com/Washington-University/HCPpipelines). We used FieldTrip (*ft_postfreesurferscript.sh*) to create the individual cortical mesh encompassing 4002 grid points per hemisphere. This representation provided the individual EEG source model geometry. To generate individual head and forward models, we segmented T1 images into three tissue types (skull, scalp, brain; Fieldtrip function *ft_volumesegment*).

Subsequently, the forward model (volume conduction) was estimated using the boundary element method in FieldTrip ('dipoli' implementation). Next, we optimized the fit of digitized EEG channel locations (xensor digitizer, ANT Neuro) to each individual's head model in the CTF coordinate system using rigid-body transformation and additional interactive alignment. Finally, the lead-field matrix for each channel x source pair was computed using the source and forward model per individual.

Beamforming

Source reconstruction (inverse solution) was achieved using a frequency-domain beamforming approach, namely partial and canonical coherence (PCC) [86]. To this end, we concatenated 5-min segments of eyes-open resting state and a random 5-min segment of task data and calculated the cross-spectral density using a broad-band filter centered at 15 Hz (bandwidth = 14 Hz). The result was then used together with the source and forward model to obtain a common spatial filter per individual in FieldTrip (regularization parameter: 5 %, dipole orientation: axis of most variance using singular value decomposition).

Source projection of sensor data

For the source-level analysis, sensor-level single-trial data in each of our two analysis routines were projected to source space by matrix-multiplication of the

spatial filter weights. To increase signal-to-noise ratio in source estimates and computationally facilitate source-level analyses, source-projected data were averaged across grid-points per cortical area defined according to the HCP functional parcellation template ([87,similar to 88]. This parcellation provides a symmetrical delineation of each hemisphere into 180 parcels for a total of 360 parcels.

Regions of interest

We constrained the analysis of neural measures to an a priori defined auditory region of interest (ROI) as well as two control ROIs in the inferior parietal lobule and visual cortex, respectively. Regions of interest were constructed by selecting a bilaterally symmetric subset of functional parcels (see Figure 3 – supplement 1). Following the notation used in ref. [87], the auditory ROI encompassed eight parcels per hemisphere covering the primary auditory cortex, lateral, medial, and parabelt complex, A4 and A5 complex that extended laterally into the posterior portion of the superior temporal gyrus (Brodmann area (BA) 22). It also included parts of the retroinsular and parainsular cortex (BA52) that lie at the boundary of temporal lobe and insula. The inferior parietal (IPL) ROI consisted of six parcels per hemisphere that included the angular gyrus (BA39), supramarginal gyrus (BA40), as well as the anterior lateral bank of the intraparietal sulcus. The visual ROI consisted of four parcels that encompassed the areas V1 (BA17), V2 (BA18), as well as V3 and V4 (BA19).

Attentional modulation of alpha power

Absolute source power was calculated as the square-amplitude of the spectro-temporal estimates. Since oscillatory power values typically follow a highly skewed, non-normal distribution, we applied a nonlinear transformation of the Box-Cox family ($power_{trans} = (power_p - 1)/p$ with $p=0.5$) to minimize skewness and to satisfy the assumption of normality for parametric statistical tests involving oscillatory power values [89]. To quantify attention-related changes in 8–12 Hz alpha power, per ROI, we calculated the single-trial, temporally resolved alpha lateralization index as follows [9,11,12]:

$$ALI = (\alpha\text{-power}_{ipsi} - \alpha\text{-power}_{contra})/(\alpha\text{-power}_{ipsi} + \alpha\text{-power}_{contra}) \quad (1)$$

To account for overall hemispheric power differences that were independent of attention modulation, we first normalized single-trial power by calculating per parcel and frequency the whole-trial (–0.5–6.5 s) power averaged across all trials and subtracted it from single trials. We then used a robust variant of the index that applies the inverse logit transform [(1 / 1+ exp(−x)] to both inputs to scale them into a common, positive-only [0;1]-bound space prior to index calculation.

For visualization and statistical analysis of cue-driven neural modulation, we then averaged the ALI across all parcels within the auditory ROI and extracted single-trial mean values for the time window of sentence presentation (3.5–6.5 s), and treated them as the dependent measure in linear mixed-effects analysis. In addition, they served as continuous predictors in the statistical analysis of brain-behavior relationships (see below). We performed additional control analyses that focused on the ALI during sentence presentation in the inferior parietal and visual ROI, as well as on the auditory ALI during the presentation of the spatial cue (0–1 s) and the sentence-final word (5.5–6.5 s).

Extraction of envelope onsets

From the presented speech signals, we derived a temporal representation of the acoustic onsets in the form of the onset envelope [40]. To this end, using the NSL toolbox [90], we first extracted an auditory spectrogram of the auditory stimuli (128 spectrally resolved sub-band envelopes logarithmically spaced between 90–4000 Hz), which were then summed across frequencies to yield a broad-band temporal envelope. Next, the output was down-sampled and low-pass filtered to match the specifics of the EEG. To derive the final onset envelope to be used in linear regression, we first obtained the first derivative of the envelope and set negative values to zero (half-wave rectification) to yield a temporal signal with positive-only values reflecting the acoustic onsets (see Fig. 4A and figure supplements).

Estimation of envelope reconstruction models

To investigate how low-frequency (i.e., < 8 Hz) fluctuations in EEG activity relate to the encoding of attended and ignored speech, we used a linear regression approach to describe the mapping from the presented speech signal to the resulting neural response [91,92]. More specifically, we trained stimulus reconstruction models (also termed decoding or backward models) to predict the onset envelope of the attended and ignored speech stream from EEG. In this analysis framework, a linear reconstruction model g is assumed to represent the linear mapping from the recorded EEG signal, $r(t,n)$, to the stimulus features, $s(t)$:

$$\hat{s}(t) = \sum_n \sum_\tau g(\tau, n) r(t + \tau, n) \quad (2)$$

where $\hat{s}(t)$ is the reconstructed onset envelope at time point t. We used all parcels within the bilateral auditory ROI and time lags τ in the range of –100 ms to 500 ms to compute envelope reconstruction models using ridge regression [93]:

$$g = (R^T R + \lambda m I)^{-1} R^T s \quad (3)$$

where R is a matrix containing the sample-wise time-lagged replication of the neural response matrix r, λ is the ridge parameter for regularization, m is a scalar

representing the mean of the trace of RTR [94], and I is the identity matrix. We followed the procedures described in ref. [95] to estimate the optimal ridge parameter.

Compared to linear forward ('encoding') models that derive temporal response functions (TRFs) independently for each EEG channel or source, stimulus reconstruction models represent multivariate impulse response functions that exploit information from all time lags and EEG channels/sources simultaneously. To allow for a neurophysiological interpretation of backward model coefficients, we additionally transformed them into linear forward model coefficients following the inversion procedure described in ref. [96]. All analyses were performed using the multivariate temporal response function (mTRF) toolbox [91] (version 1.5) for Matlab.

Prior to model estimation, we split the data based on the two spatial attention conditions (selective vs. divided), resulting in 120 trials per condition. Envelope reconstruction models were trained on selective-attention single-trial data, only. Two different backward models were estimated for a given trial, an envelope reconstruction model for the-be-attended speech stream (short: attended reconstruction model), and one for the to-be-ignored speech stream (short: ignored reconstruction model). Reconstruction models for attended and ignored speech signals were trained separately for attend-left and attend-right trials which yielded 120 single-trial decoders (60 attended, 60 ignored) per attentional setting. For illustrative purposes, we averaged the forward transformations of attended and ignored speech per side across all participants (Fig. 4B).

Evaluation of envelope reconstruction accuracy

To quantify how strongly the attended and ignored sentences were tracked by slow cortical dynamics, at the single-subject level, we reconstructed the attended and ignored envelope of a given trial using a leave-one-out cross-validation procedure. Following this approach, the envelopes of each trial were reconstructed using the averaged reconstruction models trained on all but the tested trial. For a given trial, we only used the trained models that corresponded to the respective cue condition (i.e., for an attend-left/ignore-right trial we only used the reconstruction models trained on the respective trials). The reconstructed onset envelope obtained from each model was then compared to the two onset envelopes of the actually presented speech signals. The resulting Pearson-correlation coefficients, $r_{attended}$ and $r_{ignored}$, reflect the single-trial neural tracking strength or reconstruction accuracy [23] (see Figure 4–figure supplement 1 and 2).

We proceeded in a similar fashion for divided-attention trials. Since these trials could not be categorized based on the to-be-attended and -ignored side, we split them based on the ear that was probed at the end the trial. Given that even in the absence of a valid attention cue, participants might still (randomly) focus their attention to one of the two streams, we wanted to quantify how strongly the

probed and unprobed envelopes were tracked neurally. To this end, we used the reconstruction models trained on selective-attention trials to reconstruct the onset envelopes of divided-attention trials. Sentences presented in probed-left/unprobed-right trials were reconstructed using the attend-left/ignore-right reconstruction models while probed-right/unprobed-left trials used the reconstruction models trained on attend-right/ignore-left trials.

One of the goals of the present study was to investigate the degree to which the neural tracking of the attended and ignored envelopes would be influenced by spatial attention and semantic predictability, and how these changes would relate to behavior. We thus used the neural tracking strength of the attended envelope (when reconstructed with the attended reconstruction model) and that of the ignored envelope (when reconstructed with the ignored reconstruction model) as variables in our linear mixed-effects analyses (see below).

Decoding accuracy

To evaluate how accurately we could decode an individual's focus of attention, we separately evaluated the performance of the reconstruction models for the attended or the ignored envelope. When the reconstruction was performed with the attended reconstruction models, the focus of attention was correctly decoded if the correlation with the attended envelope was greater than that with the ignored envelope. Conversely, for reconstructions performed with the ignore reconstruction models, a correct identification of attention required the correlation with the ignored envelope to be greater than that with the attended envelope (see Fig. 4–supplement 1). To quantify whether the single-subject decoding accuracy was significantly above chance, we used a binomial test with $\alpha=0.05$.

Statistical analysis

We used (generalized) linear mixed-effect models to investigate the influence of the experimental cue conditions (spatial cue: divided vs. selective; semantic cue: general vs. specific) on behavior, see Q1 in Fig. 1) as well as on our neural measures of interest (Q2). Finally, we investigated the relationship of neural measures (Q3) and their (joint) influence of the different cue conditions and neural measures on behavior (Q4). Using linear mixed-effects models allowed us to model and control for the impact of various additional covariates known to influence behavior and/or neural measures. These included the probed ear (left/right), whether the later-probed sentence had the earlier onset (yes/no), as well as participants' age and hearing acuity (pure-tone average across both ears).

Model selection

To avoid known problems associated with a largely data-driven stepwise model selection that include the overestimation of coefficients [97,98] or the selection of irrelevant predictors [99], our selection procedure was largely constrained by *a*

priori defined statistical hypotheses. The influence of visual cues and of neural measures were tested in same brain-behavior model.

At the fixed-effects level, these models include the main effects of both cue types, auditory alpha lateralization during sentence presentation, separate regressors for the neural tracking of attended and ignored speech, as well as of probed ear, participants' age, PTA, and difference in the onset of attended and ignored sentence. In line with the specifics of the experimental design and derived hypotheses, we included additional 2- and 3-way interactions to model how behavioral performance was affected by the joint influence of visual cues, and by cue-driven changes in neural measures (see Tables S1 and S2 for the full list of included parameters). The brain-behavior model of accuracy and response speed included random intercepts by subject and item. In a data-driven manner, we then tested whether model fit could be improved by the inclusion of by-subject random slopes for the effect of the spatial-attention cue, semantic cue, or probed ear. The change in model fit was assessed using likelihood ratio tests on nested models.

The set of models that analyzed how neural measures changed as a function of visual cue information included main effects of visual cues, probed ear, as well as the same set of additional covariates. We also included 2-way interaction of spatial cue and semantic cue, as well as spatial cue and probed ear (see Tables S3-5 for full list of included parameters). Due to convergence issue, we did not include a random intercept by item but otherwise followed the same procedure for the inclusion of random slopes as described above.

Lastly, models that probed the interdependence of neural filter mechanisms included the neural tracking of attended or ignored speech as dependent variable and alpha power, spatial cue, and their interactions as predictors (see Tables S6 and S7 for details). The selection of the random structure followed the same principles as for the cue-driven neural modulation models described in the paragraph above).

Deviation coding was used for categorical predictors. All continuous variables were z-scored. For the dependent measure of accuracy, we used generalized linear mixed-effects model (binomial distribution, logit link function). For response speed, we used general linear mixed-effects model (gaussian distribution, identity link function). *P*-values for individual model terms in the linear models are based on *t*-values and use the Satterthwaite approximation for degrees of freedom[100]. *P*-values for generalized linear models are based on z-values and asymptotic Wald tests. In lieu of a standardized measure of effect size for mixed-effects models, we report odds ratios (OR) for generalized linear models and standardized regression coefficients (β) for linear models along with their respective standard errors (SE). Given the large number of individual terms included in the models, all reported p-values corrected to control for the false discovery rate [101].

All analyses were performed in R ([102]; version 3.6.1) using the packages lme4 [103], and sjPlot [104].

Bayes factor estimation

To facilitate the interpretability of significant and non-significant effects, we also calculated the Bayes Factor (BF) based on the comparison Bayesian information criterion (BIC) values as proposed by Wagenmakers [105].

For the brain–behavior models, BF calculation was based on simplified brain-behavior models that included all main effects but no interaction terms to avoid problems associated with the dependence of lower- and higher-order effects. These models included separate regressors for each of our key neural measures to arbitrate between within- and between-subject effects on behavior. Between-subject effect regressors consisted of neural measures that were averaged across all trials at the single-subject level, whereas the within-subject effect was modelled by the trial-by-trial deviation from the subject-level mean [cf. 45].

To calculate the BF for a given term, we compared the BIC values of the full model to that of a reduced model in which the respective term was removed [$BF10 = \exp((BIC(H_0)-BIC(H_1))/2)$]. When log-transformed, BFs larger than 0 provide evidence for the presence of an effect (i.e., the observed data are more likely under the more complex model) whereas BFs smaller than 0 provide evidence for the absence of an effect (i.e., the observed data are more likely under the simpler model).

Predictive analysis

To quantify how well single-trial behavioral performance of new participants could be predicted on the basis of on our final explanatory brain–behavior models, we performed a predictive analyses including variables selection using LASSO regression [98]. All computations were carried out using the glmmLasso package in R [106,107]. As part of the implemented procedure, the design matrix was centered and orthonormalized prior to model fitting. The resulting coefficients were then transformed back to the original units.

Given the imbalance of correct and incorrect trials, for accuracy, we relied on the area under the receiver operating characteristic curve (AUC) as measure of classification accuracy, and on the mean squared error (MSE) as a measure of prediction error for the analysis of response speed.

We used a nested cross-validation scheme in which hyperparameter tuning and the evaluation of predictive performance were carried out in separate inner and outer loops (see Figure 6-supplement 3). For the outer loop, we split the full dataset at random into five folds for training and testing. Each set of training data was subsequently split into further 5 folds for the purpose of hyperparameter tuning. For both outer and inner loop, the data was split at the subject level, that is, all trials from one participants were assigned to either training or testing, hence avoiding data leakage and associated overfitting [108]. Lasso regression models for accuracy and response speed relied on the same data splits.

For the selection of the optimal tuning parameter λ that controls the amount of shrinkage applied to the coefficients, we used a predefined grid of 100 logarithmically-spaced λ values. The highest λ value (λ_{max}) caused all coefficients to be shrunk to zero, while λ_{min} was defined as $\lambda_{min} = \lambda_{max} * 0.001$. To determine the regression path, we computed models with decreasing λ values and passed the resulting coefficients to the next iteration thus improving the efficiency of the optimization algorithm. For each inner loop cross-validation cycle, we determined the optimal λ value (λ_{opt}) according to the 1-SE rule [109]. These values were then passed onto the corresponding outer cross-validation folds for training and testing. We quantified the average predictive performance of our variable selection procedure by averaging the respective performance metric (AUC / MSE) across all k=5 outer fold models.

Finally, we determined which of these final models yielded the predictive performance and used the associated λ value to fit this winning model to all of the available data [110]. As implemented in glmmLasso, we computed p-values by re-estimating the model including only terms with non-zero coefficients and the application of Fisher scoring.

Data availability

The complete dataset associated with this work including raw data, EEG data analysis results, as well as corresponding code will be publicly available under https://osf.io/28r57/.

References

1. Cherry EC. Some experiments on the recognition of speech, with one and with two ears. J Acoust Soc Am. 1953;25: 975–979. doi:10.1121/1.1907229

2. Jensen O, Mazaheri A. Shaping functional architecture by oscillatory alpha activity: gating by inhibition. Front Hum Neurosci. 2010;4: 186. doi:10.3389/fnhum.2010.00186

3. Foxe JJ, Snyder AC. The Role of Alpha-Band Brain Oscillations as a Sensory Suppression Mechanism during Selective Attention. Front Psychology. 2011;2: 154. doi:10.3389/fpsyg.2011.00154

4. Händel BF, Haarmeier T, Jensen O. Alpha oscillations correlate with the successful inhibition of unattended stimuli. J Cogn Neurosci. 2011;23: 2494–2502. doi:10.1162/jocn.2010.21557

5. Rihs TA, Michel CM, Thut G. Mechanisms of selective inhibition in visual spatial attention are indexed by α-band EEG synchronization. Eur J Neurosci. 2007;25: 603–610. doi:10.1111/j.1460-9568.2007.05278.x

6. Kerlin JR, Shahin AJ, Miller LM. Attentional gain control of ongoing cortical speech representations in a "cocktail party." J Neurosci. 2010;30: 620–628. doi:10.1523/JNEUROSCI.3631-09.2010

7. Ahveninen J, Huang S, Belliveau JW, Chang W-T, Hämäläinen M. Dynamic oscillatory processes governing cued orienting and allocation of auditory attention. J Cogn Neurosci. 2013;25: 1926–1943. doi:10.1162/jocn_a_00452

8. Müller N, Weisz N. Lateralized auditory cortical alpha band activity and interregional connectivity pattern reflect anticipation of target sounds. Cereb Cortex. 2011;22: 1604–1613. doi:10.1093/cercor/bhr232

9. Wöstmann M, Herrmann B, Maess B, Obleser J. Spatiotemporal dynamics of auditory attention synchronize with speech. Proc Natl Acad Sci USA. 2016;113: 3873–3878. doi:10.1073/pnas.1523357113

10. Wöstmann M, Vosskuhl J, Obleser J, Herrmann CS. Opposite effects of lateralised transcranial alpha versus gamma stimulation on auditory spatial attention. Brain Stimulation. 2018;11: 752–758. doi:10.1016/j.brs.2018.04.006

11. Tune S, Wöstmann M, Obleser J. Probing the limits of alpha power lateralisation as a neural marker of selective attention in middle-aged and older listeners. Eur J Neurosci. 2018;48: 2537–2550. doi:10.1111/ejn.13862

12. Haegens S, Handel BF, Jensen O. Top-Down Controlled Alpha Band Activity in Somatosensory Areas Determines Behavioral Performance in a Discrimination Task. J Neurosci. 2011;31: 5197–5204. doi:10.1523/JNEUROSCI.5199-10.2011

13. Bauer M, Kennett S, Driver J. Attentional selection of location and modality in vision and touch modulates low-frequency activity in associated sensory cortices. J Neurophysiol. 2012;107: 2342–2351. doi:10.1152/jn.00973.2011

14. Worden MS, Foxe JJ, Wang N, Simpson GV. Anticipatory biasing of visuospatial attention indexed by retinotopically specific alpha-band electroencephalography increases over occipital cortex. J Neurosci. 2000;20: 1–6. doi:https://doi.org/10.1523/JNEUROSCI.20-06-j0002.2000

15. Kelly SP, Lalor EC, Reilly RB, Foxe JJ. Increases in alpha oscillatory power reflect an active retinotopic mechanism for distracter suppression during sustained visuospatial attention. J Neurophysiol. 2006;95: 3844–3851. doi:10.1152/jn.01234.2005

16. Schroeder CE, Lakatos P. Low-frequency neuronal oscillations as instruments of sensory selection. Trends Neurosci. 2009;32: 9–18. doi:10.1016/j.tins.2008.09.012

17. Schroeder CE, Wilson DA, Radman T, Scharfman H, Lakatos P. Dynamics of Active Sensing and perceptual selection. Current Opinion in Neurobiology. 2010;20: 172–176. doi:10.1016/j.conb.2010.02.010

18. Henry MJ, Obleser J. Frequency modulation entrains slow neural oscillations and optimizes human listening behavior. Proc Natl Acad Sci USA. 2012;109: 20095–20100. doi:10.1073/pnas.1213390109

19. Obleser J, Kayser C. Neural Entrainment and Attentional Selection in the Listening Brain. Trends Cogn Sci. 2019;23: 913–926. doi:10.1016/j.tics.2019.08.004

20. Zion Golumbic EMZ, Ding N, Bickel S, Lakatos P, Schevon CA, McKhann GM, et al. Mechanisms Underlying Selective Neuronal Tracking of Attended Speech at a ''Cocktail Party''. Neuron. 2013;77: 980–991. doi:10.1016/j.neuron.2012.12.037

21. Ding N, Simon JZ. Emergence of neural encoding of auditory objects while listening to competing speakers. Proc Natl Acad Sci USA. 2012;109: 11854–11859. doi:10.1073/pnas.1205381109

22. Mesgarani N, Chang EF. Selective cortical representation of attended speaker in multi-talker speech perception. Nature. 2012;485: 233–236. doi:10.1038/nature11020

23. O'Sullivan JA, Power AJ, Mesgarani N, Rajaram S, Foxe JJ, Shinn-Cunningham BG, et al. Attentional Selection in a Cocktail Party Environment Can Be Decoded from Single-Trial EEG. Cereb Cortex. 2014;25: 1697–1706. doi:10.1093/cercor/bht355

24. Horton C, D'Zmura M, Srinivasan R. Suppression of competing speech through entrainment of cortical oscillations. J Neurophysiol. 2013;109: 3082–3093. doi:10.1152/jn.01026.2012

25. Keitel C, Keitel A, Benwell CSY, Daube C, Thut G, Gross J. Stimulus-Driven Brain Rhythms within the Alpha Band: The Attentional-Modulation Conundrum. J Neurosci. 2019;39: 3119–3129. doi:10.1523/JNEUROSCI.1633-18.2019

26. Gundlach C, Moratti S, Forschack N, Müller MM. Spatial Attentional Selection Modulates Early Visual Stimulus Processing Independently of Visual Alpha Modulations. Cereb Cortex. 2020;16: 637–18. doi:10.1093/cercor/bhz335

27. Henry MJ, Herrmann B, Kunke D, Obleser J. Aging affects the balance of neural entrainment and top-down neural modulation in the listening brain. Nat Commun. 2017;8: 15801. doi:10.1038/ncomms15801

28. Lakatos P, Barczak A, Neymotin SA, McGinnis T, Ross D, Javitt DC, et al. Global dynamics of selective attention and its lapses in primary auditory cortex. Nat Neurosci. 2016;19: 1707–1717. doi:10.1038/nn.4386

29. Posner MI. Orienting of attention. Quarterly Journal of Experimental Psychology. 1980;32: 3–25. doi: 10.1080/00335558008248231

30. Alavash M, Tune S, Obleser J. Modular reconfiguration of an auditory control brain network supports adaptive listening behavior. Proc Natl Acad Sci USA. 2019;116: 660–669. doi:10.1073/pnas.1815321116

31. Presacco A, Simon JZ, Anderson S. Effect of informational content of noise on speech representation in the aging midbrain and cortex. J Neurophysiol. 2016;116: 2356–2367. doi:10.1152/jn.00373.2016

32. Sohoglu E, Peelle JE, Carlyon RP, Davis MH. Predictive Top-Down Integration of Prior Knowledge during Speech Perception. J Neurosci. 2012;32: 8443–8453. doi:10.1523/JNEUROSCI.5069-11.2012

33. Peelle JE, Gross J, Davis MH. Phase-Locked Responses to Speech in Human Auditory Cortex are Enhanced During Comprehension. Cereb Cortex. 2013;23: 1378–1387. doi:10.1093/cercor/bhs118

34. Obleser J, Weisz N. Suppressed Alpha Oscillations Predict Intelligibility of Speech and its Acoustic Details. Cereb Cortex. 2012;22: 2466–2477. doi:10.1093/cercor/bhr325

35. Wöstmann M, Lim S-J, Obleser J. The human neural alpha response to speech is a proxy of attentional control. Cereb Cortex. 2017;27: 3307–3317. doi:10.1093/cercor/bhx074

36. Broadbent DE, Gregory M. Accuracy of recognition for speech presented to the right and left ears. Quarterly Journal of Experimental Psychology. 1964;16: 359–360. doi:10.1080/17470216408416392

37. Kimura D. Cerebral dominance and the perception of verbal stimuli. Canadian Journal of Psychology/Revue canadienne de psychologie. 1961;15: 166–171. doi:10.1037/h0083219

38. Shmueli G. To Explain or to Predict? Statist Sci. 2010;25: 289–310. doi:10.1214/10-STS330

39. Alavash M, Tune S, Obleser J. Modular reconfiguration of an auditory control brain network supports adaptive listening behavior. Proc Natl Acad Sci USA. 2018;48: 201815321–10. doi:10.1073/pnas.1815321116

40. Fiedler L, Wöstmann M, Herbst SK, Obleser J. Late cortical tracking of ignored speech facilitates neural selectivity in acoustically challenging conditions. NeuroImage. 2019;186: 33–42. doi:10.1016/j.neuroimage.2018.10.057

41. Banerjee S, Snyder AC, Molholm S, Foxe JJ. Oscillatory alpha-band mechanisms and the deployment of spatial attention to anticipated auditory and visual target locations: supramodal or sensory-specific control mechanisms? J Neurosci. 2011;31: 9923–9932. doi:10.1523/JNEUROSCI.4660-10.2011

42. Schneider BA, Daneman M, Pichora-Fuller MK. Listening in aging adults: from discourse comprehension to psychoacoustics. Can J Exp Psychol. 2002;56: 139–152. doi:10.1037/h0087392

43. Passow S, Westerhausen R, Wartenburger I, Hugdahl K, Heekeren HR, Lindenberger U, et al. Human aging compromises attentional control of auditory perception. Psychol Aging. 2012;27: 99–105. doi:10.1037/a0025667

44. Anderson S, White-Schwoch T, Parbery-Clark A, Kraus N. A dynamic auditory-cognitive system supports speech-in-noise perception in

older adults. Hearing Research. 2013;300: 18–32. doi:10.1016/j.heares.2013.03.006

45. Bell A, Fairbrother M, Jones K. Fixed and random effects models: making an informed choice. Qual Quant. 2018;53: 1051–1074. doi:10.1007/s11135-018-0802-x

46. Waschke L, Tune S, Obleser J. Local cortical desynchronization and pupil-linked arousal differentially shape brain states for optimal sensory performance. eLife. 2019;8: 1868–27. doi:10.7554/eLife.51501

47. Krakauer JW, Ghazanfar AA, Gomez-Marin A, MacIver MA, Poeppel D. Neuroscience needs behavior: correcting a reductionist bias. Neuron. 2017;93: 480–490. doi:10.1016/j.neuron.2016.12.041

48. van Ede F, Köster M, Maris E. Beyond establishing involvement: quantifying the contribution of anticipatory α- and β-band suppression to perceptual improvement with attention. J Neurophysiol. 2012;108: 2352–2362. doi:10.1152/jn.00347.2012

49. Ding N, Simon JZ. Cortical entrainment to continuous speech: functional roles and interpretations. Front Hum Neurosci. 2014;8: 13367. doi:10.3389/fnhum.2014.00311

50. Hamilton LS, Huth AG. The revolution will not be controlled: natural stimuli in speech neuroscience. Language, Cognition and Neuroscience. 2018;27: 1–10. doi:10.1080/23273798.2018.1499946

51. Sassenhagen J. How to analyse electrophysiological responses to naturalistic language with time-resolved multiple regression. Language, Cognition and Neuroscience. 2018;35: 474–490. doi:10.1080/23273798.2018.1502458

52. Ding N, Simon JZ. Neural coding of continuous speech in auditory cortex during monaural and dichotic listening. J Neurophysiol. 2012;107: 78–89. doi:10.1152/jn.00297.2011

53. Broderick MP, Anderson AJ, Di Liberto GM, Crosse MJ, Lalor EC. Electrophysiological Correlates of Semantic Dissimilarity Reflect the Comprehension of Natural, Narrative Speech. Curr Biol. 2018;28: 1–11. doi:10.1016/j.cub.2018.01.080

54. Brodbeck C, Hong LE, Simon JZ. Rapid Transformation from Auditory to Linguistic Representations of Continuous Speech. Curr Biol. 2018;28: 3976–3982.e6. doi:10.1016/j.cub.2018.10.042

55. Fiedler L, Woestmann M, Herbst SK, Obleser J. Late cortical tracking of ignored speech facilitates neural selectivity in acoustically

challenging conditions. NeuroImage. 2018;186: 33–42. doi:https://doi.org/10.1016/j.neuroimage.2018.10.057

56. Thut G. α-band electroencephalographic activity over occipital cortex indexes visuospatial attention bias and predicts visual target detection. J Neurosci. 2006;26: 9494–9502. doi:10.1523/JNEUROSCI.0875-06.2006

57. Bengson JJ, Mangun GR, Mazaheri A. The neural markers of an imminent failure of response inhibition. NeuroImage. 2012;59: 1534–1539. doi:10.1016/j.neuroimage.2011.08.034

58. Dahl MJ, Ilg L, Li S-C, Passow S, Werkle-Bergner M. Diminished pre-stimulus alpha-lateralization suggests compromised self-initiated attentional control of auditory processing in old age. NeuroImage. 2019;197: 414–424. doi:10.1016/j.neuroimage.2019.04.080

59. Hong X, Sun J, Bengson JJ, Mangun GR, Tong S. Normal aging selectively diminishes alpha lateralization in visual spatial attention. NeuroImage. 2015;106: 353–363. doi:10.1016/j.neuroimage.2014.11.019

60. Leenders MP, Lozano-Soldevilla D, Roberts MJ, Jensen O, De Weerd P. Diminished Alpha Lateralization During Working Memory but Not During Attentional Cueing in Older Adults. Cereb Cortex. 2018;28: 21–32. doi:10.1093/cercor/bhw345

61. Mok RM, Myers NE, Wallis G, Nobre AC. Behavioral and neural markers of flexible attention over working memory in aging. Cereb Cortex. 2016;26: 1831–1842. doi:10.1093/cercor/bhw011

62. Rogers CS, Payne L, Maharjan S, Wingfield A, Sekuler R. Older adults show impaired modulation of attentional alpha oscillations: Evidence from dichotic listening. Psychol Aging. 2018;33: 246–258. doi:10.1037/pag0000238

63. Wöstmann M, Alavash M, Obleser J. Alpha Oscillations in the Human Brain Implement Distractor Suppression Independent of Target Selection. J Neurosci. 2019;39: 9797–9805. doi:10.1523/JNEUROSCI.1954-19.2019

64. Pernet CR, Sajda P, Rousselet GA. Single-trial analyses: why bother? Front Psychology. 2011;2: 322. doi:10.3389/fpsyg.2011.00322

65. O'Sullivan J, Herrero J, Smith E, Schevon C, McKhann GM, Sheth SA, et al. Hierarchical Encoding of Attended Auditory Objects in Multi-talker Speech Perception. Neuron. 2019;104:1195–1204. doi:10.1016/j.neuron.2019.09.007

66. Serences JT, Kastner S. A multi-level account of selective attention. In: Nobre AC, Kastner S (Eds.). The Oxford Handbook of Attention. Oxford University Press. 2014; pp. 76–104. doi:10.1093/oxfordhb/9780199675111.013.022

67. Puvvada KC, Simon JZ. Cortical Representations of Speech in a Multitalker Auditory Scene. J Neurosci. 2017;37: 9189–9196. doi:10.1523/JNEUROSCI.0938-17.2017

68. Melara RD, Rao A, Tong Y. The duality of selection: Excitatory and inhibitory processes in auditory selection attention. Journal of Experimental Psychology: Human Perception and Performance. 2002;28: 279–306. doi:10.1037//0096-1523.28.2.279

69. Chait M, de Cheveigné A, Poeppel D, Simon JZ. Neural dynamics of attending and ignoring in human auditory cortex. Neuropsychologia. 2010;48: 3262–3271. doi:10.1016/j.neuropsychologia.2010.07.007

70. Shinn-Cunningham BG, Best V. Selective Attention in Normal and Impaired Hearing. Trends in Amplification. 2008;12: 283–299. doi:10.1177/1084713808325306

71. Shinn-Cunningham BG. Object-based auditory and visual attention. Trends Cogn Sci. 2008;12: 182–186. doi:10.1016/j.tics.2008.02.003

72. Kaya EM, Elhilali M. Modelling auditory attention. Philos Trans R Soc Lond, B, Biol Sci. 2017;372. doi:10.1098/rstb.2016.0101

73. Wöstmann M, Obleser J. Acoustic detail but not predictability of task-irrelevant speech disrupts working memory. Front Hum Neurosci. 2016;10: 201–9. doi:10.3389/fnhum.2016.00538

74. Zhigalov A, Herring JD, Herpers J, Bergmann TO, Jensen O. Probing cortical excitability using rapid frequency tagging. NeuroImage. 2019;195: 59–66. doi:10.1016/j.neuroimage.2019.03.056

75. Teoh ES, Lalor EC. EEG decoding of the target speaker in a cocktail party scenario: considerations regarding dynamic switching of talker location. J Neural Eng. 2019;16: 036017–30. doi:10.1088/1741-2552/ab0cf1

76. Nourski KV, Brugge JF. Representation of temporal sound features in the human auditory cortex. Rev Neurosci. 2011;22: 187–203. doi:10.1515/RNS.2011.016

77. Giraud AL, Poeppel D. Cortical oscillations and speech processing: emerging computational principles and operations. Nat Neurosci. 2012;15: 511–517. doi:10.1038/nn.3063

78. Sadaghiani S, Kleinschmidt A. Brain Networks and α-Oscillations: Structural and Functional Foundations of Cognitive Control. Trends Cogn Sci. 2016;20: 805–817. doi:10.1016/j.tics.2016.09.004

79. Banerjee S, Snyder AC, Molholm S, Foxe JJ. Oscillatory Alpha-Band Mechanisms and the Deployment of Spatial Attention to Anticipated Auditory and Visual Target Locations: Supramodal or Sensory-Specific Control Mechanisms? J Neurosci. 2011;31: 9923–9932. doi:10.1523/JNEUROSCI.4660-10.2011

80. Zoefel B, VanRullen R. Oscillatory Mechanisms of Stimulus Processing and Selection in the Visual and Auditory Systems: State-of-the-Art, Speculations and Suggestions. Front Neurosci. 2017;11: 225–13. doi:10.3389/fnins.2017.00296

81. Oldfield RC. The assessment and analysis of handedness: The Edinburgh inventory. 1971;9: 97–113. doi:10.1016/0028-3932(71)90067-4

82. Jefferies K, Gale TM. 6-CIT: Six-Item Cognitive Impairment Test. Cognitive Screening Instruments. London: Springer, London; 2013. pp. 209–218. doi:10.1007/978-1-4471-2452-8_11

83. Peirce JW. PsychoPy—Psychophysics software in Python. Journal of Neuroscience Methods. 2007;162: 8–13. doi:10.1016/j.jneumeth.2006.11.017

84. Delorme A, Makeig S. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. Journal of Neuroscience Methods. 2004;134: 9–21. doi:10.1016/j.jneumeth.2003.10.009

85. Oostenveld R, Fries P, Maris E, Schoffelen J-M. FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. Comput Intell Neurosci. 2011;2011: 1–9. doi:10.1155/2011/156869

86. Schoffelen J-M, Gross J. Source connectivity analysis with MEG and EEG. Salmelin R, Baillet S, editors. Hum Brain Mapp. 2009;30: 1857–1865. doi:10.1002/hbm.20745

87. Glasser MF, Coalson TS, Robinson EC, Hacker CD, Harwell J, Yacoub E, et al. A multi-modal parcellation of human cerebral cortex. Nature. 2016;536: 171–178. doi:10.1038/nature18933

88. Keitel A, Gross J. Individual Human Brain Areas Can Be Identified from Their Characteristic Spectral Activation Fingerprints. PLoS Biol. 2016;14: e1002498–22. doi:10.1371/journal.pbio.1002498

89. Smulders FTY, Oever ten S, Donkers FCL, Quaedflieg CWEM, van de Ven V. Single-trial log transformation is optimal in frequency analysis of resting EEG alpha. Eur J Neurosci. 2018;44: 94–14. doi:10.1111/ejn.13854

90. Chi T, Ru P, Shamma SA. Multiresolution spectrotemporal analysis of complex sounds. J Acoust Soc Am. 2005;118: 887–906. doi:10.1121/1.1945807

91. Crosse MJ, Di Liberto GM, Bednar A, Lalor EC. The Multivariate Temporal Response Function (mTRF) Toolbox: A MATLAB Toolbox for Relating Neural Signals to Continuous Stimuli. Front Hum Neurosci. 2016;10: 604. doi:10.3389/fnhum.2016.00604

92. Lalor EC, Foxe JJ. Neural responses to uninterrupted natural speech can be extracted with precise temporal resolution. Eur J Neurosci. 2010;31: 189–193. doi:10.1111/j.1460-9568.2009.07055.x

93. Hoerl AE, Kennard RW. Ridge Regression: Biased Estimation for Nonorthogonal Problems. Technometrics. 1970;12: 55–67. doi:10.1080/00401706.1970.10488634

94. Biesmans W, Das N, Francart T, Bertrand A. Auditory-Inspired Speech Envelope Extraction Methods for Improved EEG-Based Auditory Attention Detection in a Cocktail Party Scenario. IEEE Trans Neural Syst Rehabil Eng. 2017;25: 402–412. doi:10.1109/TNSRE.2016.2571900

95. Fiedler L, Wöstmann M, Graversen C, Brandmeyer A, Lunner T, Obleser J. Single-channel in-ear-EEG detects the focus of auditory attention to concurrent tone streams and mixed speech. J Neural Eng. 2017;14: 036020. doi:10.1088/1741-2552/aa66dd

96. Haufe S, Meinecke F, Görgen K, Dähne S, Haynes J-D, Blankertz B, et al. On the interpretation of weight vectors of linear models in multivariate neuroimaging. NeuroImage. 2014;87: 96–110. doi:10.1016/j.neuroimage.2013.10.067

97. Chatfield C. Model Uncertainty, Data Mining and Statistical Inference. Journal of the Royal Statistical Society: Series A (Statistics in Society). John Wiley & Sons, Ltd; 1995;158: 419–444. doi:10.2307/2983440

98. Tibshirani R. Regression shrinkage and selection via the lasso. J R Stat Soc Series B Stat Methodol. 1996;56: 267–288.

99. Derksen S, Keselman HJ. Backward, forward and stepwise automated subset selection algorithms: Frequency of obtaining authentic and noise variables. British Journal of Mathematical and Statistical

Psychology. 1992;45: 265–282. doi:10.1111/j.2044-8317.1992.tb00992.x

100. Luke SG. Evaluating significance in linear mixed-effects models in R. Behav Res. 2017;49: 1494–1502. doi:10.3758/s13428-016-0809-y

101. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Series B Stat Methodol. 1995;57: 289–300. doi:10.2307/2346101

102. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria. 2019. doi:https://www.R-project.org/

103. Bates D, Mächler M, Bolker B, Walker S. Fitting Linear Mixed-Effects Models Using lme4. J Stat Soft. 2015;67: 1–48. doi:10.18637/jss.v067.i01

104. Lüdecke D, Lüdecke D. Data Visualization for Statistics in Social Science [R package sjPlot version 2.6.1]. Comprehensive R Archive Network (CRAN); 2018. doi:10.5281/zenodo.1308157

105. Wagenmakers E-J. A practical solution to the pervasive problems of p values. Psychonomic Bulletin & Review. 2007;14: 779–804. doi:10.3758/bf03194105

106. Groll A. glmmLasso: Variable selection for generalized linear mixed models by L1-penalized estimation. https://CRAN.R-project.org/package=glmmLasso

107. Groll A, Tutz G. Variable selection for generalized linear mixed models by L 1-penalized estimation. Stat Comput. 2012;24: 137–154. doi:10.1007/s11222-012-9359-z

108. Poldrack RA, Huckins G, Varoquaux G. Establishment of Best Practices for Evidence for Prediction. JAMA Psychiatry. 2020;77: 534–540. doi:10.1001/jamapsychiatry.2019.3671

109. Hastie T, Tibshirani R, Friedman J. The elements of statistical learning: data mining, inference, and prediction. The Elements of Statistical Learning. New York, NY: Springer; 2009. pp. 1–51. doi:10.1007/b94608_5

110. James G, Witten D, Hastie T, Tibshirani R. An Introduction to Statistical Learning. New York, NY: Springer Science & Business Media; 2013