Neural attentional filters and behavioural outcome follow independent individual trajectories over the adult life span

Sarah Tune[1,2] & Jonas Obleser[1,2]

1 – Department of Psychology, University of Lübeck, 23562 Lübeck, Germany
2 – Center of Brain, Behavior, and Metabolism, University of Lübeck, 23562 Lübeck, Germany

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

Running title: Modelling trajectories of individual listening behaviour

Number of words in Abstract: 144
Number of words in Introduction: 755
Number of words in Discussion: 1683

Keywords: EEG, aging, speech processing, auditory attention, neural speech tracking, longitudinal analysis, latent change score models

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

Author contributions:
S.T.: Conceptualization, data curation, formal analysis, investigation, visualization, methodology, writing–original draft, writing–review and editing
J.O.: Conceptualization, investigation, methodology, writing–review and editing, supervision, project administration, resources, funding acquisition

Abstract

Preserved communication abilities promote healthy aging. To this end, the age-typical loss of sensory acuity might in part be compensated for by an individual's preserved attentional neural filtering. Is such a compensatory brain–behaviour link longitudinally stable? Can it predict individual change in listening behaviour? We here show that individual listening behaviour and neural filtering ability follow largely independent developmental trajectories modelling electroencephalographic and behavioural data of N = 105 aging individuals (39–82 yrs). First, despite the expected decline in hearing-threshold–derived sensory acuity, listening-task performance proved stable over 2 years. Second, neural filtering and behaviour were correlated only within each separate measurement timepoint (T1, T2). Longitudinally, however, our results raise caution on attention-guided neural filtering metrics as predictors of individual trajectories in listening behaviour: Neither neural filtering at T1 nor its two-year change could predict individual two-year behavioural change, under a combination of modelling strategies.

Introduction

Speech comprehension is an essential aspect of human communication, enabling us to understand and interact with others effectively. Preserved communication is therefore critical to social well-being and healthy aging[1]. Any translational advance aimed at maintaining and restoring successful cognitive aging crucially relies on understanding the factors that explain and predict individual trajectories of listening performance. However, the evidence on these potential factors is astonishingly limited.

As we age, our ability to comprehend speech can decline due to age-related changes in the auditory system (i.e., sensory acuity) and in cognitive resources. Age-related hearing loss reduces the ability to detect and to discriminate speech sounds, especially in noisy environments. However, it has long been recognised that increasing age and hearing loss cannot fully account for the considerable degree of inter-individual differences observed in listening behaviour and its lifespan change[2-4].

Recent research has focused on the neurobiological mechanisms that promote successful speech comprehension by implementing 'neural filters' that segregate behaviourally relevant from irrelevant sounds. Such neural filter mechanisms act by selectively increasing the sensory gain for behaviourally relevant inputs or by inhibiting the processing of irrelevant inputs[5-7]. A growing body of evidence suggests that speech comprehension is neurally supported by an attention-guided filter mechanism that modulates sensory gain and arises from primary auditory and perisylvian brain regions: By synchronizing its neural activity with the temporal structure of the speech signal of interest, the brain 'tracks' and thereby better encodes behaviourally relevant auditory inputs to enable attentive listening[8-11].

In a large age-varying cohort (N = 155; 39–80 yrs), we have previously shown how the fidelity of this neural filtering strategy can help explain differences in listening behaviour (i) from individual to individual, and (ii) within individuals from sentence to sentence[12-14]. As participants performed a challenging dual-talker listening task, we recorded their electroencephalogram (EEG). We observed that enhanced neural filtering—defined as stronger neural tracking of attended vs. ignored speech—led to more accurate and overall faster responses. Notably, we observed both neural filtering as well as its link to behaviour to be independent of chronological age and severity of hearing loss[12] (Fig. 1A).

The observation of such brain–behaviour relationships critically advances our understanding of the neurobiological foundation of cognitive functioning by showing, for example, how neural implementations of auditory selective attention support attentive listening. They also provide fruitful ground for scientific inquiries into the translational potential of neural markers. However, the potency of neural markers to predict future behavioural outcomes is often only implicitly assumed but seldomly put to the test[15].

Using auditory cognition as a model system, we here aim to overcome this limitation by testing directly the hitherto unknown longitudinal stability of neural filtering as a neural compensatory mechanism upholding communication success. Going further, we ask to what extent an individual's attentional neural-filtering ability measured at a given moment is able to predict their future trajectory in listening

performance. Only if this is the case, and only if such an association can plausibly be assumed to be causal for future changes in communication ability, neural filtering would be a potential translational target.

We here aim to fill this gap by analysing two-year changes in the sensory, neural, and behavioural domain in a longitudinal subsample (N = 105; 39–82 yrs) of the original AUDADAPT cohort[12] (see Fig. 2A). We apply a combination of advanced cross-sectional and longitudinal modelling strategies to address the following specific research questions (see Fig. 1).

First, by focusing on each domain individually, we ask how sensory, neural and behavioural functioning evolve cross-sectionally across the middle and older adult life span (Fig. 1B). More importantly, we also ask how they change longitudinally across the studied two-year period (Fig. 1C, Q1), and whether aging individuals differ significantly in their degree of change (Q2). We expect individuals' hearing acuity and behaviour to decrease from T1 to T2. Since we previously observed inter-individual differences in neural filtering to be independent of age and hearing status, we did not expect any systematic longitudinal change in neural filtering.

Second, we test the longitudinal stability of the previously observed age- and hearing-loss–independent effect of neural filtering on both accuracy and response speed (Fig. 1A). To this end, we analyse the multivariate direct and indirect relationships of hearing acuity, neural filtering and listening behaviour within and across timepoints.

Third, leveraging the strengths of latent change score modelling[16,17], we fuse cross-sectional and longitudinal perspectives to probe the role of neural filtering as a precursor of behavioural change in two different ways: we ask whether an individual's T1 neural filtering strength can predict the observed behavioural longitudinal change (Q3), and whether two-year change in neural filtering can explain concurrent change in listening behaviour (Q4). Here, irrespective of the observed magnitude and direction of T1–T2 developments, two scenarios are conceivable: Intra-individual neural and behavioural change may be either be correlated—lending support to a compensatory role of neural filtering—or instead follow independent trajectories[18] (see Fig. 1C).

Answering these questions is vital for understanding the neurobiological mechanisms of successful communication across the lifespan. Answering them will also critically inform the development of interventions targeted at maintaining or restoring communication success and therefore concerns basic and applied researchers alike.

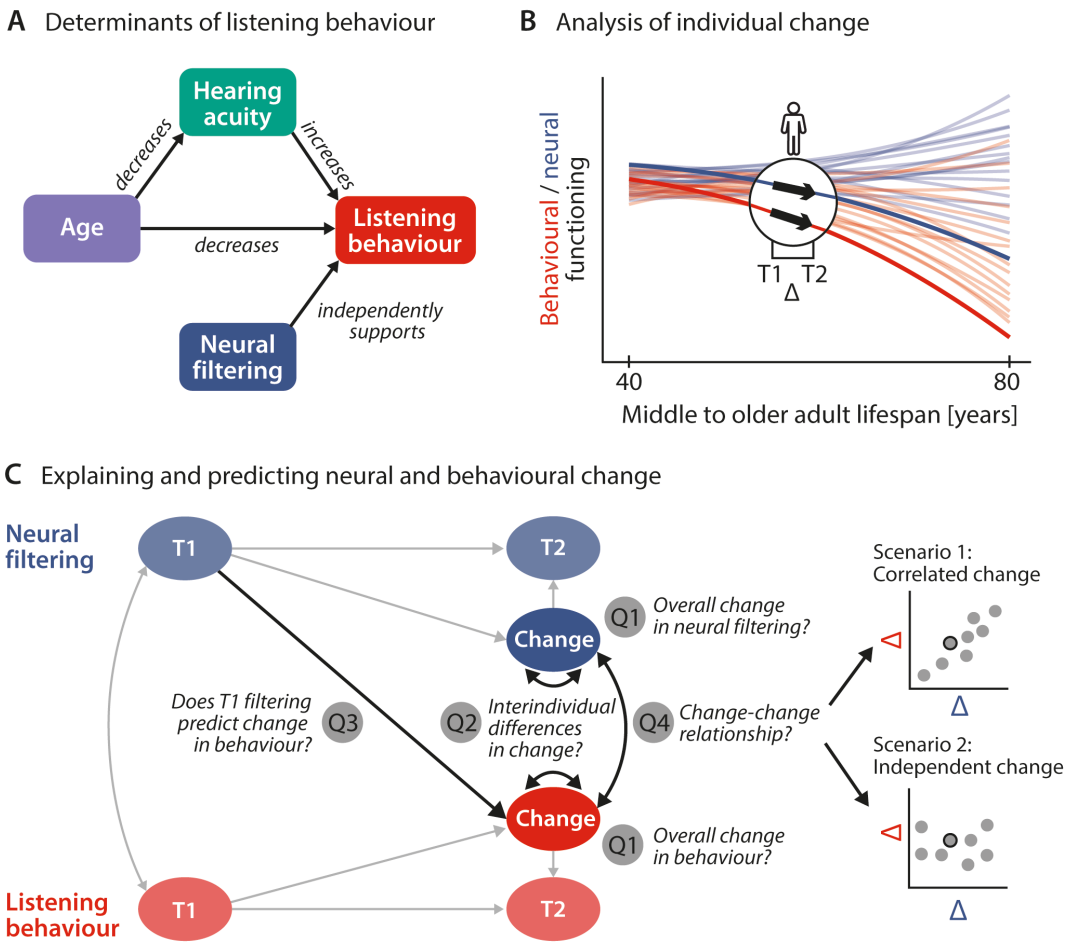

Figure 1. Schematic illustration of key assumptions and research questions. A Listening behaviour at a given timepoint is shaped by an individuals' sensory and neural functioning. Increased age decreases listening behaviour both directly, and indirectly via age-related hearing loss. Listening behaviour is supported by better neural filtering ability, independently of age and hearing acuity. **B** Conceptual depiction of individual two-year changes along the neural (blue) and behavioural (red) domain. Thin coloured lines show individual trajectories across the adult lifespan, thick lines and black arrows highlight two-year changes in a single individual. **C** Left, Schematic diagram highlighting the key research questions detailed in the introduction and how they are addressed in the current study using latent change score modelling. Right, Across individuals, co-occurring changes in the neural and behavioural domain may be correlated (top) or independent of one another (bottom).

Results

We studied an age-varying cohort of healthy middle-aged and older adults longitudinally (N = 105, 39–82 yrs, median age at T2: 63 yrs). The mean time difference between the two measurement timepoints reported here was 23.2 (sd: 4.0) months. We characterise the multivariate relationship of key measures of sensory, neural, and behavioural functioning to explain and predict individual trajectories of listening performance.

At each of two measurement timepoints, participants underwent audiological assessment followed by EEG recording during which they performed a difficult dual-talker dichotic listening task (see Fig. 2A,B)[12,19,20]. In each trial of the task, participants listened to two temporally aligned but spatially separated five-word sentences. They

then had to identify the final word in one of the two sentences from a visual array of four alternatives given a 4-sec time limit (Fig. 2C). In 50 % of the trials, a visual spatial-attention cue indicated the side of target sentence presentation, the other half of trials were preceded by an uninformative neutral cue.

We extracted individuals' mean accuracy and response speed (calculated as the inverse of reaction time) as key readouts of listening behaviour. On the basis of source-localised 1–8 Hz auditory cortical activity, we further quantified individuals' neural filtering ability as their attention-guided neural tracking of relevant vs. irrelevant speech (Fig 2D; see also ref.[12] and Methods for details). Our main analyses focus on neural filtering and listening performance averaged across all trials and thereby also across two separate spatial-attention conditions. This choice allowed us to most directly probe the trait-like nature and relationships of neural filtering. It was additionally supported by our previous observation of a general boost in behavioural performance with stronger neural filtering, irrespective of spatial attention[12].

We follow a three-step analysis strategy to address our specific research questions: First, we provide a largely descriptive overview of the observed average two-year change per studied domain. Second, we follow up on this fundamental analysis with a causal mediation analysis[21] and single-trial mixed-effects model analysis geared to assess the longitudinal stability of our recently reported effects of age, hearing acuity and neural filtering on listening task performance. Third and final, we integrate and extend the first two analysis perspectives in a joint latent change score model (LCSM)[16] to most directly probe the role of neural filtering ability as a predictor of future attentive listening ability. Addressing our key change-related research questions at the latent rather than the manifest level supersedes the manual calculation of notoriously noisy differences scores, and effectively removes the influence of each metric's reliability on the estimation of change-related relationships[16,17,22].

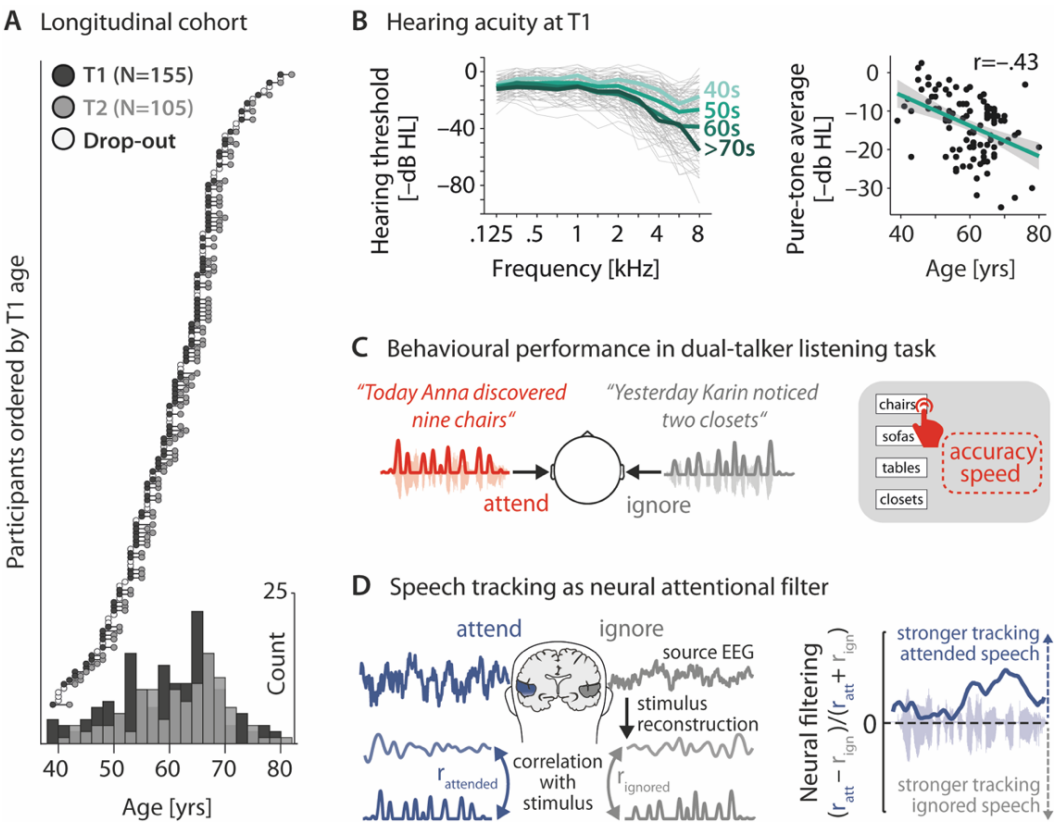

Figure 2. Key neural and behavioural metrics derived from a longitudinal cohort. A Longitudinal cohort of healthy middle-aged and older adults measured twice, two years apart. Circles represent individual participants at a given measurement time (dark grey: timepoint (T) 1, light grey: T2, white: drop-outs after T1). Bottom, Age distribution at T1 and T2 across 5-year bins. **B** Left, T1 air conduction hearing thresholds per individual (thin grey lines) and age group (thick coloured lines). Note that for didactic purposes, throughout the manuscript, thresholds are expressed as –dB HL to highlight the decrease in hearing acuity with age (left). Right, Pure-tone average hearing acuity (0.5, 1, 2, and 4kHz across both ears; higher is better) negatively correlates with age ($r = -.43$, $p = 3.73\times10^{-6}$). **C** Participants listened to two sentences presented simultaneously to the left and right ear. In 50 % of trials, a preceding visual cue indicated the to-be-attended target sentence. Listening behaviour is quantified via the accuracy and speed in identifying the final word of the target sentence. **D** Left, neural speech tracking as a proxy of an individual's neural filtering ability. Stimulus envelopes of attended and ignored sentences were reconstructed from source-localized EEG activity in auditory cortex (see Methods for details) and correlated with the actual envelopes. Right, Better neural filtering results from stronger neural tracking of attended compared to ignored speech. We analysed neural filtering derived from the entire sentence presentation period.

Listening performance remains stable despite decreased hearing acuity

In a first analysis (Fig. 3), we characterised how hearing acuity, neural filtering, and listening performance change across the middle to older adult lifespan. Additionally, we analysed longitudinal change from timepoint 1 (T1) to timepoint 2 (T2). We used the same linear mixed-effect models to test cross-sectional effects of age and longitudinal changes with time. We additionally quantified each measure's test-retest reliability as their T1–T2 Pearson's correlation.

Note that throughout the manuscript and all analyses, we reversed the sign of pure-tone average (PTA) values to express them as an index of hearing acuity rather than hearing loss (i.e., higher values indicating better acuity). Similarly, for more intuitive

interpretation, accuracy is visually presented as mean proportion correct but was logit transformed for all statistical analyses to satisfy model assumptions.

As expected, hearing acuity decreased linearly with increasing age (Fig. 3A, $\beta = -3.4$, SE (standard error) = .71, $p<.001$) and on average by 1.2 dB from T1 to T2 ($\beta = -1.18$, SE = .27, $p<.001$; $mean_{T1}$: −13.72 dB HL (sd: 7.8); $mean_{T2}$: −14.90 dB HL (8.3)). The effect of age did not change with time (age x timepoint $\beta = -.35$, SE = .28, p = .21). Assuming constant individual progression rates, this observed change corresponds to a projected average decrease in hearing acuity per decade of −6.3 (sd:15.3) dB HL (see Fig. S1). The magnitude of observed and projected hearing loss progression is well in line with recent large-sample reports[23-25]. Measurements of hearing acuity showed high test-retest reliability (r = .94, $p<.001$), underscoring the high fidelity of our audiological assessment.

In line with known deleterious effects of age, both behavioural outcomes (response speed and accuracy) declined with increasing age, and did so to a similar degree in T1 and T2 (Fig. 3C, D; Speed: $\beta = -.02$, SE = .01, p = .004; Accuracy: $\beta = -.23$, SE = .07, p = .0001; timepoint x age $p>.12$). At the same time, and contrary to our expectations, average performance levels remained stable from T1 to T2 (Speed: $\beta = .004$, SE = .01, p = .44; $mean_{T1}$: .62 s^{-1} (.08); $mean_{T2}$: .62 s^{-1} (.08); Accuracy: $\beta = .04$, SE = .05, p = .36, $mean_{T1}$: .88 % (.09); $mean_{T2}$: .88. % (.11)). Accuracy and response speed showed moderately high test-retest reliability (Speed: r = .70, $p<.001$; Accuracy: r = .72, $p<.001$).

The analysis of change in neural filtering revealed that its strength varied independently of age at both timepoints (Fig. 3B, $\beta = .0003$, SE = .0005, p = .48; timepoint x age $\beta = -.0002$, SE = .0006, p = .79), confirming our previously reported T1 results[12]. As shown in Figure 3 (bottom left panel), magnitude and direction of observed longitudinal change are highly variable across individuals and age groups, and we did not find evidence of any systematic group-level change from T1 to T2 ($\beta = .001$, SE = .001, p = .16). In addition, individual neural filtering strength correlated only weakly across time (r = .21, p = .03).

We also assessed the reliability of two established neural traits using resting-state EEG from the same recording sessions: The individual alpha frequency (IAF)[26] and the slope of 1/f neural noise[27,28]. As expected, both metrics showed high test-retest reliability (IAF: r = .83, $p<.001$; 1/f slope: r = .78, $p<.001$). These findings provide a reference level on reliability, demonstrating that the weak reliability of the neural filtering metric is not due primarily to differences in EEG signal quality across sessions.

The temporal instability of neural filtering challenges its status as a potential trait-like neural marker of attentive listening ability. At the same time, irrespective of the degree of reliability of neural filtering itself, across individuals it may still be reliably linked to the behavioural outcome (see Fig. 1). This is being addressed next.

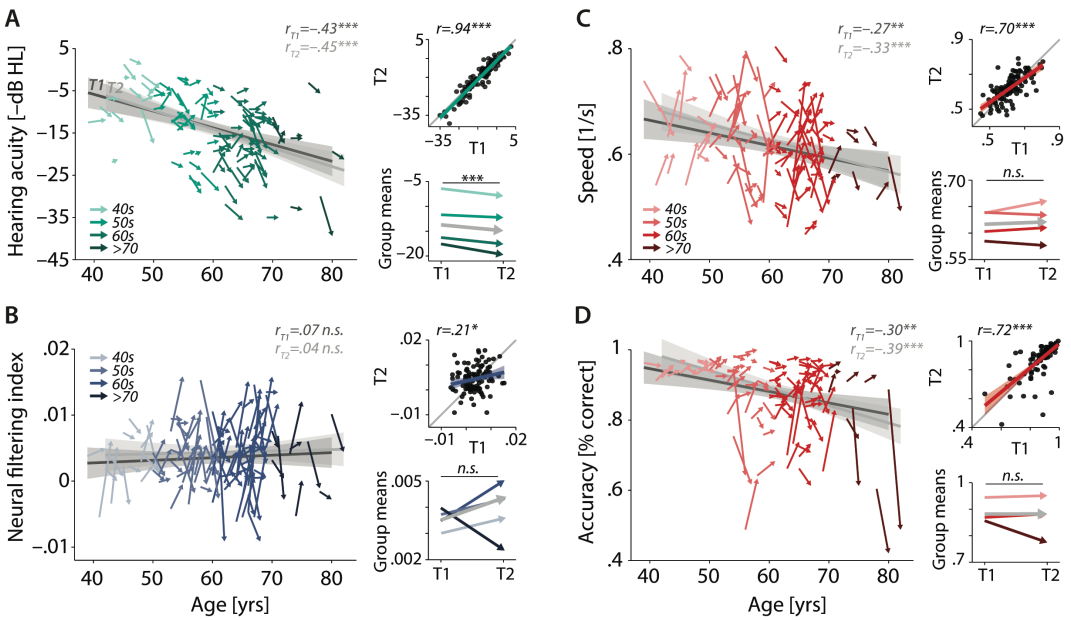

Figure 3. Characterizing cross-sectional and longitudinal change along the auditory sensory (A), neural (B), and behavioural (C, D) domain. For each domain, coloured vectors (colour-coding four age groups for illustrative purposes, only) in the respective left subpanels show an individual's change from T1 to T2 along with the cross-sectional trend plus 95% confidence interval (CI) separately for T1 (dark grey) and T2 (light grey). Top right subpanels: correlation of T1 and T2 as measure of test-retest reliability along with the 45° line (grey) and individual data points (black circles). Bottom right panels: Mean longitudinal change per age group (coloured vectors) and grand mean change (grey). Note that accuracy is expressed here as proportion correct for illustrative purposes, but was analysed logit-transformed or by applying generalized linear models.

Neural filtering reliably supports listening performance independent of age and hearing status

On the basis of the full (T1 and T2) dataset, we aimed to replicate our key T1 results and test whether the previously observed between-subjects brain-behaviour relationship would hold across time: We expected an individual's neural filtering ability to impact their listening performance (accuracy and response speed) independently of age or hearing status[12]. Given the moderately strong correlation of age and hearing acuity (r = –.43; p<.001; Fig. 2B), we employed causal mediation analysis to model the direct as well as the hearing-acuity–mediated effect of age on the behavioural outcome[21]. To formally test the stability of direct and indirect relationships across time, we used a moderated mediation analysis. In this analysis, the inclusion of interactions by timepoint tested whether the influence of age, sensory acuity, and neural filtering on behaviour varied significantly across time.

Our expectations on the direct relationships were indeed borne out by the data: higher age was associated with poorer hearing ability (β = –.43, SE = .09, p<.001) and listening performance (Speed: β = –.33, SE = .06, p<.001; Accuracy: β = –.26, SE = .06, p<.001). Better hearing ability, on the other hand, boosted accuracy but not response speed (Accuracy: β = .30, SE = .1, p = .003; Speed: β = –.06, SE = .1, p = 0.56). These direct effects remained stable from T1 to T2 (all interactions by timepoint p>.56; all log Bayes Factors ($\log BF_{01}$)>2.5).

Age also impacted accuracy indirectly: The total effect of age was partially mediated via its detrimental effect on hearing acuity (average causal mediation effect (ACME), $\beta = -.12$, SE = .04, $p<.001$). We did not find evidence for an analogous indirect effect on speed (ACME: $\beta = .008$, SE = .03, $p = .77$). Again, the hearing-acuity mediated effect of age on accuracy did not change from T1 to T2 as evidenced by moderated mediation analysis (interaction by timepoint $p = .73$).

Speaking to the robustness of our previous results, we observed the beneficial effect of stronger neural filtering fidelity on both measures of listening performance (Accuracy: $\beta = .21$, SE = , $p = .02$; Speed: $\beta = .33$, SE = .09, $p<.001$). Note that the magnitude of this direct brain–behaviour effect is comparable to that of the direct effect of age. Alternative models that included indirect, neural filtering-mediated paths from either age or hearing acuity to behaviour did not reveal any significant mediation effects.

Most importantly, the longitudinal stability of the observed direct brain-behaviour link was further supported by the absence of any significant changes with time (interactions by timepoint; all $p>.28$, all $\log BF_{01}>2.1$).

In our previous T1 analysis (N = 155)[12], we had found evidence for the here analysed brain-behaviour link at two different levels of observation: (i) at the trait level—individuals with overall stronger neural filtering also performed better overall—and (ii) at the state-level—stronger neural filtering in a given trial raised the chances of responding correctly. Aiming at replication of the state-level (i.e., within-participant) relationship, we ran a single-trial linear mixed-effects model analysis on our longitudinal N = 105 sample. This analysis utilised single-trial data of both T1 and T2.

Lending credibility to our previous results, stronger single-trial neural filtering was associated with higher listening success at both T1 and T2 (logistic mixed-effects model; within-participant effect of neural filtering on accuracy: odds ratio (OR) = 1.08, SE = .02, $p<.001$; interaction neural filtering x timepoint: OR = .99, SE = .03, $p = .82$; see also Fig. 4C).

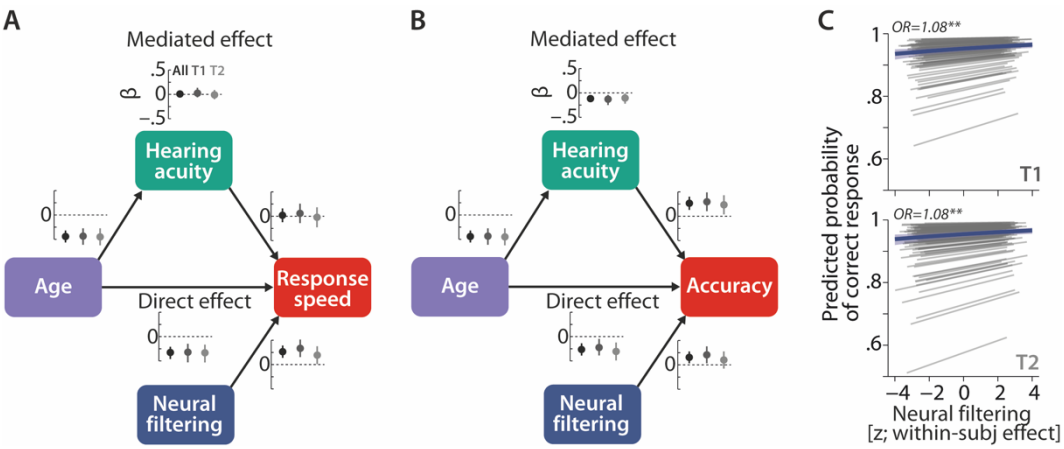

Figure 4. Longitudinal stability of sensory and neural determinants of listening behaviour. Causal mediation analysis of age, hearing acuity, and neural filtering on response speed (**A**) and logit-transformed accuracy (**B**). Graphs next to each path indicate standardised coefficients plus 95% CIs separately for the full dataset (black), T1 (dark grey), and T2 (light grey). We did not find any significant modulation of the observed effects with time (see text for results). **C** Neural filtering strength was found to be predictive of accuracy at the single-trial level at both T1 (top) and T2 (bottom). Grey lines show individual effects, blue thick line shows the group-level fixed effect along with 95 % confidence interval. OR = odds ratio, ** p<.01.

Accuracy is longitudinally stable but speed and neural filtering increase at T2

Having established the longitudinal stability of the beneficial impact of intact neural filtering on listening performance, we turned to our final, most comprehensive analysis.

In a latent change score model (LCSM), in its bivariate form sometimes termed a parallel process model[16,17], we connected the neural and behavioural domain. This allowed us to most directly probe the potential role of neural filtering as a precursor of behavioural changes. Specifically, we asked: (i) Is an individual's baseline (T1) level of neural filtering ability predictive of their two-year change in behaviour; and (ii) are individual differences in longitudinal dynamics in the behavioural domain associated with those in the neural domain?

As a technical note, it is worth reiterating that in the present data, the highly variable, weakly reliable surface measure of neural filtering was nonetheless robustly connected to the behavioural outcome (see above; see Fig. 4). It is in such scenarios that the LCSM framework comes with particular methodological benefits: By expressing individuals' T1 and T2 levels, as well as their T1–T2 change as latent variables instead of manifest indicators, these types of models circumvent the calculation of notoriously unreliable noisy difference scores. They also avoid potential regression to the mean due to random errors[17,22]. Instead, the measurement errors of both, latent variables and their associated indicators, are explicitly modelled and thus effectively removed from the estimates of individual differences and relationships of interest[16].

Accordingly, for our metrics, we estimated T1 and T2 latent variables of behavioural and neural filtering from two manifest indicators each. These indicators were the average of each metric across the first and second half of the experiment, respectively (see Methods for details). For all measurement models, the standardized factor loadings were significant (all ps<.05; all standardized λ>.55). The assumption of

strict factorial invariance across time could be maintained for all models (all $\Delta\chi^2_{(df = 1)} < 3.4$, all $p > .07$).

We then constructed univariate latent change score models (LCSM) to test for significant mean change in each metric from T1 to T2 while adjusting for their respective baseline (T1) level. The univariate models had acceptable (Speed: $\chi^2_{(df = 5)} = 9.7$, $p = .085$, CFI = .988, RMSEA = .094] to excellent fit [Accuracy: $\chi^2_{(df = 5)} = 5.1$, $p = .40$, CFI = .999, RMSEA = .016; Neural filtering: $\chi^2_{(df = 5)} = .6$, $p = .99$, CFI = 1, RMSEA = 0) according to established indices[29].

On average, listening task accuracy remained stable ($b_0 = .053$, SE = .045, $\Delta\chi^2_{(df = 1)} = 1.33$, $p = .25$). Response speed, on the other hand, showed a significant mean increase over time ($b_0 = .13$, SE = .03, $\Delta\chi^2_{(df = 1)} = 12.79$, $p<.001$; Fig. 5A). Similarly, the model of neural filtering showed a (marginally) significant mean increase ($b_0 = .24$, SE = .11, $\Delta\chi^2_{(df = 1)} = 2.98$, $p = .08$; Fig. 5A).

As a proof of principle, we extracted latent factor scores for T1/T2 neural filtering and response speed from their respective univariate latent change score model. By correlating T1 and T2 factor scores per domain, we show how the explicit modelling of measurement errors helps to improve the test-retest reliability compared to that observed at level of manifest variables (see insets in Fig. 5A; neural filtering: $r = .65$, $p<.001$, response speed: $r = .75$, $p<.001$).

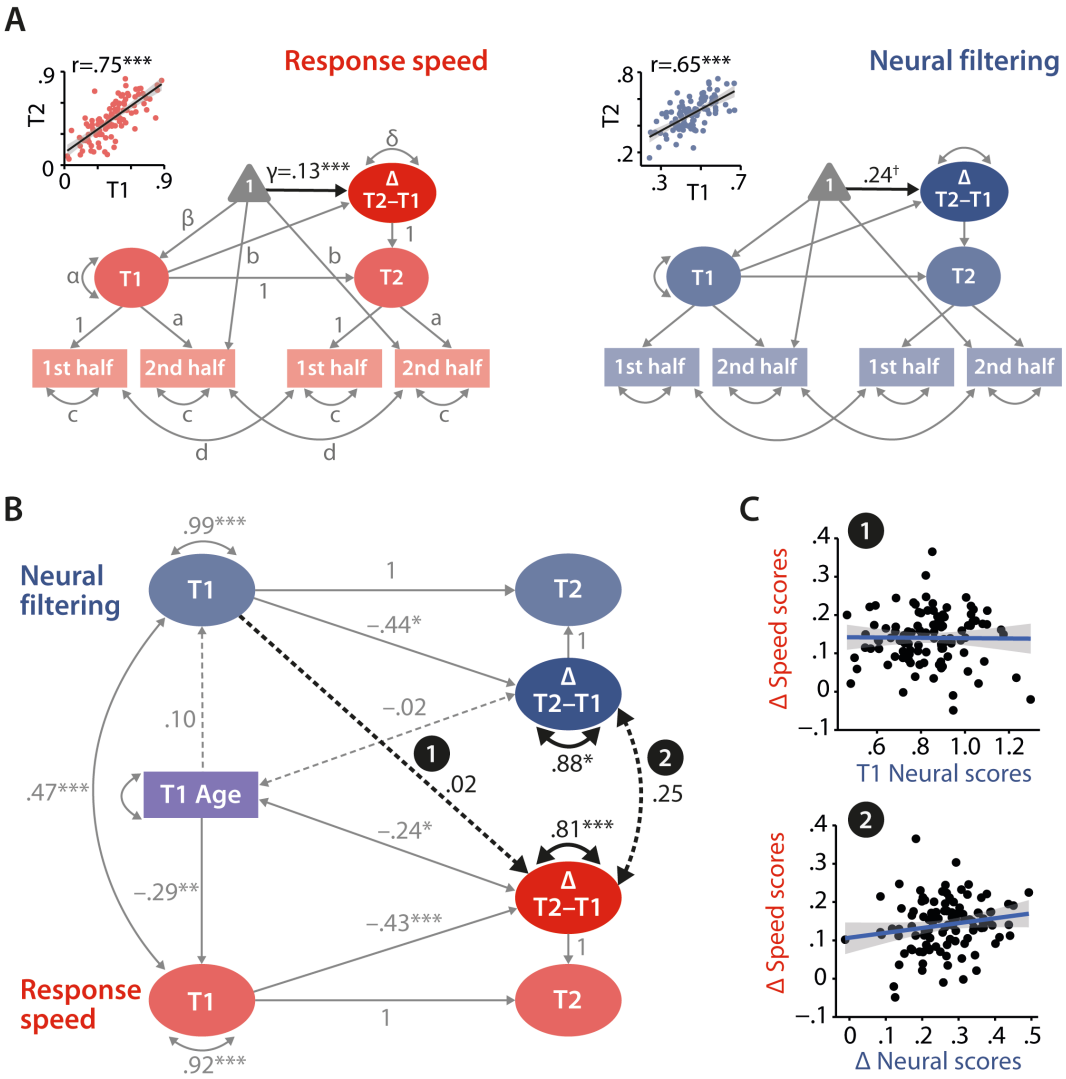

Figure 5. Modelling of univariate and bivariate change. A Univariate latent change score models for response speed (left) and neural filtering (right). All paths denoted with Latin letters refer to freely estimated but constrained to be equal parameters of the respective measurement models. Greek letters refer to freely estimated parameters of the structural model. Highlighted in black is the estimated mean longitudinal change from T1 to T2. Scatterplots in the top left corner illustrate how capturing T1 and T2 neural and behavioural functioning as latent factors improves their respective test-retest reliability. **B** Latent change score model (LCSM) relating two-year changes in neural filtering strength to changes in response speed. Black arrows indicate paths or covariances of interest. Solid black arrows reflect freely estimated and statistically significant effects, dashed black arrows reflect non-significant effects. All estimates are standardised. Grey arrows show paths that were freely estimated or fixed as part of the structural model but that did not relate to the main research questions. For visual clarity, manifest indicators of the measurement model and all symbols relating to the estimated mean structure are omitted but are identical to those shown in panel A. ***p<.001, **p<.01, *p<.05, †p = .08. **C** Scatterplots of model-predicted factor scores that refer to the highlighted paths in panel **B**. Top panel shows that baseline-level neural filtering did not predict two-year change in behavioural functioning, bottom panel shows the absence of a significant change-change correlation.

Baseline neural functioning does not predict future change in listening behaviour

Based on the univariate results, we then connected the two metrics that showed a significant mean change, namely speed and neural filtering, in a bivariate model of change (Fig. 5B; see also Fig. S2 for full model details).

In line with our hypotheses, we modelled the longitudinal impact of T1 neural functioning on the change in speed, and tested for a change-change correlation. Since the analyses conducted up to this point have either directly shown or have suggested that longitudinal change per domain may be affected by age, we included individuals' age as a time-invariant covariate in the final model. We modelled the influence of age on neural and behavioural functioning at T1 but also on individual change per domain. By accounting for linear effects of age on longitudinal change, we also minimize its potential impact on the estimation of change-change relationship of interest. Note that we refrained from fitting separate models per age group due to both limited and different number of data points per age group. The model fit the data well ($\chi^2_{(df = 27)}$ = 25.65, p = .54, CFI = 1, RMSEA = 0, 95 % CI[0 .07]).

Having ensured factorial invariance and goodness of fit, we can confidently interpret the estimates of individual differences and bivariate relationships that speak to our specific research questions. Crucial to our change-related inquiries, we observed reliable variance (i.e., individual differences) in the longitudinal change in both speed (standardized estimate ϕ = 0.88, SE = .07, $\Delta\chi^2_{(df = 5)}$ = 94.07, p<.01) and neural filtering (ϕ = .81, SE = .14, $\Delta\chi^2_{(df = 4)}$ = 25.64, p<.001).

Individuals' baseline levels of both speed and neural filtering strength were predictive of their respective longitudinal change: Individuals with relatively strong neural filtering or fast responses at T1 showed a smaller increase from T1 to T2, possibly indicating ceiling effects (Speed: β = –.43, SE = .12, $\Delta\chi^2_{(df = 1)}$ = 8.75, p = .003; Neural filtering: β = –.44, SE = .16, $\Delta\chi^2_{(df = 1)}$ = 4.17, p = .04).

We also observed that participants' age at T1 covaried with the individual degree of change in speed but not with that in neural filtering: The older a participant at T1, the smaller their longitudinal increase in speed (Speed: ϕ = –.24, SE = .11, $\Delta\chi^2_{(df = 1)}$ = 4.38, p = .037; Neural filtering: ϕ = –.02, SE = .14, $\Delta\chi^2_{(df = 1)}$ = .02, p = .89). Importantly, however, an individual's latent T1 level of neural filtering strength was not predictive of the ensuing latent T1–T2 change in response speed (β = .02, SE = .16, $\Delta\chi^2_{(df = 1)}$ = .02, p = .90). We did not have any *a priori* hypotheses on the influence of T1 speed on the individual T1–T2 change in neural filtering. Still, in a control analysis that freely estimated the respective path, we found that an individual's latent T1 level of response speed was not predictive of the ensuing latent T1–T2 change in neural filtering (β = –.11, SE = .21, $\Delta\chi^2_{(df = 1)}$ = .31, p = .58).

Neural filtering ability and listening behaviour follow independent developmental trajectories in later adulthood

Finally, we turn to the last piece in our investigation where we address the question of whether individual differences in the neural and behavioural longitudinal change are

connected. In other words: Are the contemporaneous changes along the two studied domains correlated or do they occur largely independently of one another?

Change score modelling revealed that longitudinal change in the neural and the behavioural domain occurred largely independent of one another despite their systematic relationship within each separate measurement timepoint (ϕ = .25, SE = .15, $\Delta\chi^2_{(df = 1)}$ = 2.74, p = .1). In other words, those individuals who showed the largest change in neural filtering were not necessarily the ones who also changed the most in terms of their behavioural functioning (see Fig. 5C bottom panel and Fig. S3).

Control analyses: The weak correlation of behavioural and neural change is robust against different quantifications of neural filtering

Taken together, our main analyses revealed that inter-individual differences in behavioural change could only be predicted by baseline age and baseline behavioural but not neural functioning, and did not correlate with contemporaneous neural changes.

However, one could ask in how far core methodological decisions taken in the current study, namely our focus on (i) the *differential* neural tracking of relevant vs. irrelevant speech as proxy of neural filtering, and (ii) on its trait-level characterization that averaged across different spatial-attention conditions may have impacted these results. Specifically, if the neural filtering index (compared to the neural tracking of attended speech alone) is found to be less stable generally, would this also impact the chances of observing a systematic change-change relationship? Relatedly, did the analysis of neural filtering across all trials underestimate the effects of interest?

To evaluate the impact of these consideration on our main findings, we conducted two additional control analyses: First, we repeated the main analyses using the neural filtering index (and response speed) averaged across selective-attention trials, only. Second, we repeated the main analyses focused on the neural tracking of attended speech only, again averaged across selective-attention trials.

As shown in Figure 6, taken together, the control analyses provide compelling empirical support for the robustness of our main results: Linking response speed and neural filtering under selective attention strengthened their relationship at T1 (ϕ = .54, SE = .15, $\Delta\chi^2_{(df = 1)}$ = 2.74, p = .1; see. Fig 6B) but did not yield any significant effects for the influence of T1 neural filtering on behavioural change (β = .13, SE = .21, $\Delta\chi^2_{(df = 1)}$ = .43, p = .51), or for the relationship of neural and behavioural change (ϕ = .26, SE = .14, $\Delta\chi^2_{(df = 1)}$ = 3.1, p = .08; please note the close correspondence to path estimates reported in Fig. 5).

The second control analysis revealed a substantially higher manifest-level test-retest reliability of neural tracking of attended speech (r = .65, p<.001; Fig. 6C) compared to that of the neural tracking index. However, when linked to longitudinal changes in response speed, this analysis provided even less evidence for systematic change-related relationships: Baseline-levels of attended-speech tracking did not predict future change in response speed (β = .18, SE = .11, $\Delta\chi^2_{(df = 1)}$ = 2.73, p = .10), and changes in neural and

behavioural functioning occurred independently of one another (ϕ = –.03, SE = .12, $\Delta\chi^2_{(df = 1)}$ = .06, p = .81).

In sum, the two control analyses provide additional empirical support for the results revealed by our main analysis.

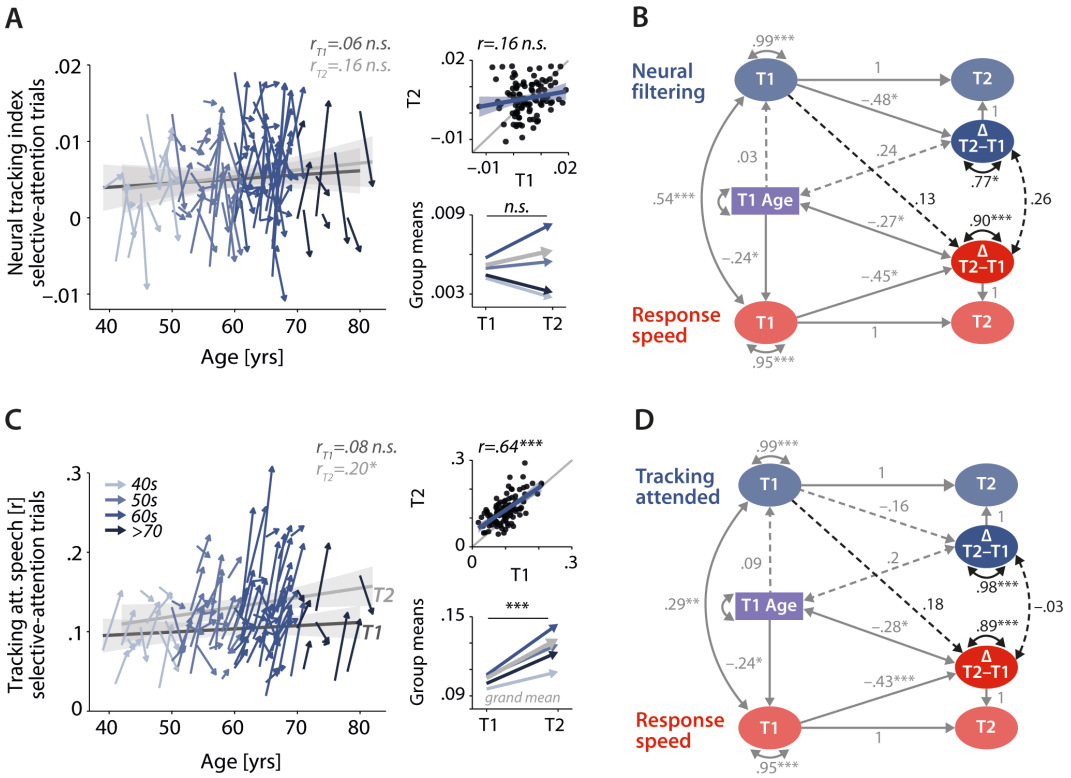

Figure 6. Control analyses corroborate the independence of neural and behavioural trajectories under selective attention. Cross-sectional and longitudinal change in neural filtering **(A)** and neural tracking of attended speech **(C)** averaged across selective-attention trials, only. Coloured vectors (colour-coding four age groups for illustrative purposes, only) in the left subpanels show individual T1–T2 change along with the cross-sectional trend plus 95% confidence interval (CI) separately for T1 (dark grey) and T2 (light grey). Top right, correlation of T1 and T2 as measure of test-retest reliability along with the 45° line (grey) and individual data points (black circles). Bottom right, mean longitudinal change per age group and grand mean change (grey). **B, D** Latent change score model (LCSM) relating two-year changes in neural filtering **(B)** /neural tracking **(D)** strength to changes in response speed. Black arrows show the paths or covariances of interest that were freely estimates, grey arrows show paths that were freely estimated or fixed as part of the structural model but did not relate to the main research questions. Solid arrows indicate statistically significant effects, dashed arrows reflect non-significant paths. All estimates are standardised. ***p<.001, **p<.01, *p<.05.

Discussion

Successfully comprehending speech in noisy environments is a challenging task, particularly for aging listeners whose hearing ability gradually declines[30]. A much-researched neural support mechanism here is attention-guided neural 'tracking' of behaviourally relevant speech signals, as one neural strategy to maintain listening success[7-10,14,31-33].

However, to date, it is unknown if the fidelity with which an individual implements this filtering strategy does represent a stable neural trait-like marker of individual attentive listening ability. Of direct relevance to any future translational efforts building on neural speech tracking, it is also unknown whether differences in neural filtering strength observed between aging listeners are predictive at all of how their attentive listening ability will develop in the future. We here have addressed these questions leveraging a new representative prospective cohort sample of healthy middle-aged to older listeners.

Over two years from T1 to T2, individuals' hearing ability worsened as expected[23-25]. Their listening performance, however, stayed stable. In addition, an individual's baseline (T1) neural filtering strength proved to be a strikingly poor indicator of their future (T2) level of neural filtering. On the other hand, bolstering previous results, neural filtering reliably supported listening behaviour within the same session—both at T1 and T2—at two levels of granularity: Individuals with generally stronger neural tracking of target vs. distractor speech performed the listening task on average more accurately and faster. They were also more likely to respond correctly in a given trial with relatively stronger neural filtering in that trial.

Crucially, however, momentary states of neural functioning were not predictive of future behavioural change, and the dynamics of longitudinal change at the neural and behavioural level appear to follow largely independent trajectories. Notably, these key findings hold for different definitions of neural-attentional filtering, and under different spatial-attention conditions (see "Control analyses" in Results).

Neural filtering fidelity as a trait-like neural marker of individual attentive listening ability?

In recent years, the enhanced representation of behaviourally relevant sounds via their prioritised neural tracking has been reported in numerous listening studies investigating different acoustic environments, participant populations, and stages of auditory processing[12,13,34-39].

This neural signature is commonly interpreted as a neural instantiation of selective auditory attention in the service of successful speech comprehension. Note however that its link to behaviour is not always explicitly established[12-14,40,41]. Given how robustly the current data show the enhanced neural tracking of attended vs. ignored speech at the group level, the weak reliability of individual neural filtering strength may come as a surprise. At the same time, stronger neural filtering was reliably linked to better behavioural performance within both T1 and T2.

How can these two findings be reconciled? Based on the current and previous results, what may be concluded about the role of neural filtering as a potential neural marker of individual attentive listening ability?

Previous studies on attention-guided neural speech tracking have not provided any direct evidence on the temporal stability of neural filtering nor on its relationship with behaviour. Studies on related neural signatures such as speech-aligned auditory

brainstem responses, or the entrainment of auditory cortical activity to rhythmic (non-speech) stimulation reported moderate to high reliability[42-45]. However, these studies have (i) investigated the temporal stability across sessions spaced only days or weeks apart, (ii) focused on younger normal-hearing populations, or (iii) quantified the neural encoding of speech or non-speech stimulation that involved less or no attentional control. In contrast, we here explicitly focused on a definition of neural filtering that incorporated the neural tracking of attended and ignored speech to highlight the dependence of communication success on the attention-guided differential neural encoding of relevant vs. irrelevant input. These differences render a direct comparison to our approach difficult, but there is reason to consider a model-derived, latent representation of neural filtering as employed here the more generalisable metric[46]. Leaving the latent-variable framework aside, it is worth emphasizing that the neural tracking of attended speech alone proved to be substantially more stable over time, with a re-test reliability in the $r_{tt} = .6$ range.

Our current results broadly align with a view of attention-guided neural speech tracking as a form of 'neural entrainment in the broad sense' that reflects a listener's neural attentional state[10,47,48].

Under this interpretation, the stable link of neural filtering to listening behaviour at different levels of granularity is noteworthy: An individual's ability to exert top-down selective attention to prioritize the neural encoding of behaviourally relevant information is far from stable but it fluctuates at different time scales[49-52]. This entails that at a longer time scale, here captured by two distinct measurement timepoints, aging individuals will differ with respect to their overall level of neural filtering and associated listening behaviour. Their level of neural and behavioural functioning will differ from other individuals' levels at the same timepoint but also from their own level at a different timepoint. Moreover, a listener's behavioural outcome at either timepoint is not only shaped by their broad neural attentional state with which they enter a communication situation. It is also critically influenced by short-term fluctuations in neural filtering strength around their current overall level of neural functioning (see Fig. 4C)[12]. As our control analyses have revealed, these fluctuations are more pronounced if neural filtering is defined by the tracking of both attended and ignored speech rather than by attended-speech tracking alone.

What does this mean for the potential translational value of neural tracking? Their highly dynamic nature gives neural tracking-based metrics value as online neural indicators of a listener's momentary attentional focus. As such, they could serve as critical neural read-outs in novel brain-computer interfaces such as neurally-steered hearing aids[53-55]. At the same time, however, their non-stationarity over time limits their potential as translational targets for diagnosis and therapeutic intervention[45,56,57].

575 ***Individual trajectories of listening behaviour cannot be explained by changes within***
576 ***a single domain***

577 As a second central query of the current study, we went beyond the establishment of
578 robust brain–behaviour relationships and directly probed the potency of neural filtering
579 to predict behavioural change over time[15]. We asked whether individual trajectories of
580 listening behaviour could be predicted by past levels of neural filtering, or by co-
581 occurring changes in neural filtering.

582 Past studies have observed enhanced cortical speech tracking in aging
583 compared to young adults. This suggests a compensatory role of increased speech-
584 brain coupling to counteract the deleterious effect of age or of hearing loss on speech
585 comprehension[41,58-62]. As a corollary of this relationship, typically observed cross-
586 sectionally, one might expect an individual's neural filtering strength to be connected
587 not only to present but also to future trajectories of listening behaviour.

588 Such relationships, if shown to be causal, would provide the strongest evidence
589 for the role of neural speech tracking as a neural compensatory mechanism supporting
590 communication success[30]. However, when analysed in our longitudinal sample of aging
591 listeners, irrespective of the precise definition of neural filtering, we did not find
592 evidence for a predictive role of neural filtering despite supporting brain–behaviour
593 links observed within each time point. What are potential reasons for this absent
594 connection (Fig. 5)?

595 One obvious explanation, both in statistical and substantive terms, may lie in
596 the low retest reliability of our neural filtering metric as discussed above. Analytically,
597 however, we were able to mitigate this problem by adopting a modelling approach
598 which effectively removes the influence of measurement error[22] (shown in Fig. 5A). Still,
599 individual change in response speed could only be predicted by an individual's baseline
600 speed and age but not by baseline neural filtering nor by its longitudinal change.
601 Moreover, our control analysis of attended-speech tracking provided additional
602 empirical support for the absence of such a predictive link despite the neural metric's
603 moderate manifest-level reliability. These findings call for a more substantive
604 explanation that goes beyond methodological details.

605 While most desirable from a translational perspective and a core quest in the
606 cognitive neuroscience of aging, predicting change in cognitive functioning, here
607 listening behaviour, from baseline or longitudinal change in brain function or structure
608 is a non-trivial endeavour. Connecting individual trajectories of neural or cognitive
609 functioning goes beyond the establishment of domain-specific age trends[63,64]. It also
610 goes beyond the mere extrapolation of (age-independent) brain–behaviour
611 relationships observed at a given moment[65,66]. Indeed, empirical evidence—and to
612 some degree also theoretical grounds—for robust brain–behaviour baseline–change or
613 change–change associations are limited[18].

614 Most empirical studies reporting such significant cross-domain change-change
615 correlations have in fact connected behavioural change to alterations in brain structure
616 rather than brain function[67-72]. Focusing on structural change may be advantageous: not
617 only can structural feature be quantified more directly and reliably, they also follow

systematic age-dependent trajectories thereby providing clearer causal pathways for ensuing behavioural change[73,74]. Still, less than half of the studies testing such cross-domain associations have indeed observed them[18]. From the perspective of theoretical models of neurobiological and cognitive aging[75-77], the absence of correlated trajectories of neural and cognitive functioning may indeed be the more expected result. These models highlight the multifaceted nature of healthy cognitive aging in which environmental variables, neurobiology, and cognition are dynamically interrelated[78]. Neural compensatory mechanisms, such as the neural filtering correlate targeted here, are thought to offset structural decline but are themselves influenced by a number of factors. This leads to increased inter-individual variability that may circumvent the emergence of group-level correlated change relationships.

Additionally, it is important to bear in mind that the behavioural outcome of interest, that is, the speed and accuracy with which an aging individual solves a difficult listening situation, involves the orchestration of different perceptual and cognitive processes[79,80]. We here focused on one candidate neurobiological implementation of an auditory attentional filter to help explain inter-individual differences in listening behaviour and its lifespan trajectories. Yet, aging individuals may rely on different alternative neural or cognitive strategies[81]. A complete understanding of inter-individual differences in listening behaviour in aging adults will therefore depend on a number of different factors among which the attention-modulated tracking of relevant speech constitutes one, potentially necessary, but not sufficient neural correlate[7,10,12,82,83].

Not least, there are a number of methodological choices that might constrain the conclusions afforded by of our current study. First, the current study was limited to two distinct timepoints spaced only two years apart. This limits the ability to model linear as well as non-linear dynamics of change. Second, it also does not allow the separation of distinct patterns of change co-occurring at the same time: one continuous, constant change with age along with a separate process in which relative change is proportional to the level observed at prior timepoints[84]. Third, we here focus on a single dichotic listening task from which both neural and behavioural functioning is derived. We therefore cannot assess whether the observed pattern would generalize to other listening tasks. Lastly, denser sampling across a longer time interval would have also increased statistical power to detect correlated change[85]. It would have also allowed to more directly test hypotheses on causal pathways by which change in the neural domain should precede change in the behavioural domain.

The conclusion stands, though, that individual trajectories in listening behaviour cannot be explained by longitudinal change along a single dimension. Instead, a better understanding of the influences shaping individual listening behaviour across the adult lifespan will critically rely on uncovering the relative contribution and age-dependent dynamics of sensory, neural, and cognitive factors.

Conclusion

The results presented here support the role of attention-guided neural filtering as a readout of an individual's neural attentional state. At the same time, the state-like nature of neural tracking-based metrics limits their translational potential as predictors of longitudinal change in listening behaviour over middle to older adulthood. Our data caution against explaining audiology-typical listening performance solely from changes in aspects of neural functioning, as listening behaviour and neural filtering ability follow largely independent developmental trajectories. Our results critically inform translational efforts aimed at the preservation and restoring of speech comprehension abilities in aging individuals.

Materials and Methods

Data collection

The analysed data are part of a large-scale longitudinal study on the neural and cognitive mechanisms supporting adaptive listening behaviour in a prospective cohort healthy middle-aged and older adults ("The listening challenge: How ageing brains adapt (AUDADAPT)"; https://cordis.europa.eu/project/rcn/197855_en.html). This project encompasses the collection of different demographic, audiological, behavioural, and neurophysiological measures across initially two time points spaced approximately two years apart. The analyses carried out on the data aim at relating adaptive listening behaviour to changes in different neural dynamics. Given the longitudinal nature of the current study, all procedures concerning data collection, as well as EEG recording and analysis are identical to those detailed in our recently published analysis of T1 data using the same experimental paradigm (see Fig. S6)[12].

Participants and procedure

We here report on a total N = 105 right-handed German native speakers (median age at timepoint 2 (T2) 63 years; age range 39–82 years; 61 females) who underwent audiological, behavioural, and electrophysiological (EEG) assessment at two separate timepoints. On average, the measurement timepoints were spaced 23.2 (± SD 4.0) months apart.

At timepoint 1 (T1), we had screened a total of N = 184 participants. Included participants had normal or corrected-to-normal vision, and did not report any neurological, psychiatric, or other disorders. They were also screened for mild cognitive impairment using the German version of the 6-Item Cognitive Impairment Test (6CIT[86] and the MoCA[87]). Only participants with normal hearing or age-adequate mild-to-moderate hearing loss were included (see Fig. 2B for individual audiograms at T1). Handedness was assessed using a translated version of the Edinburgh Handedness Inventory[88]. As a result of the initial screening procedure, 17 participants were excluded prior to EEG recording due to a medical history or non-age-related hearing loss. Three participants dropped out of the study prior to EEG recording and an additional 9 participants were excluded from analyses after EEG recording due to incidental findings after structural MR acquisition (N = 3), or due to EEG data quality issues (N = 9). Again, all detailed criteria are to be found in ref.[12].

At T2, N = 115 participants returned for follow-up measurements. All individuals passed the repeat screening procedures identical to those at T1. Ten participants had to be excluded from the analyses reported here: Three participants had dropped out prior to EEG recording, three participants were excluded due to EEG data quality issues, and four participants because their EEG data had been excluded at T1. This resulted in a final longitudinal sample of N = 105 individuals.

Dropout at T2 could not be predicted from participants' T1 age, hearing loss, behavioural performance (accuracy, speed) or neural filtering strength (all p>.13). This

indicates that compared to the full T1 cohort reported on in previous studies[12,19] our reduced longitudinal sample was not biased in terms of sensory, cognitive, and neural functioning.

At each measurement timepoint, participants underwent detailed pure-tone and speech audiometric measurements, along with an extensive battery of cognitive tests and personality profiling (see ref.[81] for details). On a separate day, we recorded participants' electroencephalogram (EEG) during rest (5 min each of eyes-open and eyes-closed measurements) followed by six blocks of the same dichotic listening task (see Fig. 2C and ref.[12] for details).

Participants gave written informed consent and received financial compensation (10€ 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

At each timepoint, participants performed a previously established dichotic listening task[20]. We provide full details on trial structure, stimulus construction, recording and presentation in our previously published study on the first (N = 155) wave of data collection (but see also Fig. S6)[12].

In short, in each of 240 trials, participants listened to two competing, dichotically presented five-word sentences spoken by the same female speaker. They were probed on the sentence-final noun in one of the two sentences. Participants were instructed to respond within a given 4 s time window beginning with the onset of a probe screen showing four alternatives. They were not explicitly instructed to respond as quickly as possible. The probe screen showed four alternative words presented either on the left or right side of the screen, indicating the probed ear. 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, but irrelevant to the current study, 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. All participants listened to the same 240 sentence pairs at each of the two measurement timepoints. The order of sentence pair presentation was randomised for each participant and at each timepoint.

To account for differences in hearing acuity within our group of participants, all stimuli were presented 50 dB above the individual sensation level.

EEG recording and analysis

The approach for EEG recording, pre-processing and subsequent analysis is identical to the procedures carried out for T1 data collection and analysis[12].

In short, 64-channel EEG data were recorded, cleaned for artifacts using a custom ICA-based pipeline, down-sampled to 125 Hz, filtered between 1–8Hz, and cut

into single-trials epochs covering the presentation of auditory stimuli. Following source-localisation via beamforming, we focused on auditory cortical activity to train and test decoding models of attended and ignored speech using cross-validated regularised regression. Models were trained on selective-attention trials, only, but then also tested on divided-attention trials. As results, we obtained single-trial reconstruction accuracy (Pearson's r) estimates as metrics of the degree of attended and ignored neural speech tracking, respectively.

We then calculated a neural filtering index across the entire sentence presentation period. The index quantifies the difference in neural tracking of the to-be-attended and of the to-be-ignored sentence (Neural filtering index = $(r_{attended} - r_{ignored})/(r_{attended} + r_{ignored})$, and thus indexes the strength of neural filtering at the single-trial level. Positive values indicate successful neural filtering in line with the behavioural goal.

The EEG analyses were carried out in Matlab 2016b using the Fieldtrip toolbox (v. 2017-04-28), the Human Connectome Project Workbench software (v1.5), FreeSurfer (v.6.0), and the multivariate temporal response function (mTRF) toolbox (v1.5)[89].

Behavioural and audiological data analysis

We evaluated participants' behavioural 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.

We defined participants' hearing acuity as their pure-tone average (PTA) composed of (air-conduction) hearing thresholds at the frequencies of .5, 1, 2, and 4 kHz. Individual PTA values were then averaged across the left and right ear.

Statistical analysis

For statistical analyses focused on between-participant ('trait-level') effects, behavioural performance metrics and neural filtering index values were averaged across all trials and experimental conditions to arrive at one trait-level estimate per participants. This approach was also motivated by previous results based on T1 data: Here, we had observed that stronger neural speech tracking led to overall faster and to more trials with accurate responses irrespective of the specific cue-cue condition[12]. Accuracy was logit-transformed for statistical analysis but expressed as proportion correct for illustrative purposes. Similarly, for more intuitive interpretation, we reversed the sign of PTA values for higher values to correspond to better hearing ability.

All analyses were performed in R (v.4.2.2)[90] using the packages lme4 (v.1.1-31)[91], mediation (v4.5.0)[92], lavaan (v0.6-12)[93], and OpenMx (v2.21.8)[94].

Linear mixed-effect modelling

As the first step of our three-part analysis approach, we applied general linear mixed-effect models to tested for cross-sectional and longitudinal changes in trait-level sensory acuity, neural filtering, and listening performance. These models included age, timepoint and their interaction as fixed effect regressors and allowed random participant-specific intercepts.

In the second step of the analysis, we also aimed at replicating the previously observed single-trial state-level relationship of neural filtering and accuracy. To this end, we applied a single generalised linear mixed-effects model (binomial distribution, logit link function) on single-trial data of both T1 and T2. This model represents an adapted version of the brain–behaviour model reported in Tune et al. (2021)[12]. In short, we included all experimental manipulation predictors, as well as age, hearing acuity and neural filtering metrics. We omitted previously shown to be non-significant higher-order interactions and additionally included interactions by timepoint to directly test for any longitudinal change in the effect of neural filtering on behaviour.

To tease apart state-level (i.e., within-participant) and trait-level (i.e., between-participants) effects, we included two separate neural regressors: For the between-participant effect regressor, we averaged single-trial neural filtering values per individual across all trials. By contrast, the within-participant effect of interest was modelled by the trial-by-trial deviation from the subject-level mean[95]. We included participant- and item-specific random intercepts as well as random slopes for the effect of the spatial-attention cue and the probed ear.

We used deviation coding for categorical predictors and z-scored all continuous regressors. P-values for individual model terms in general linear mixed-effect models are based on the Satterthwaite approximation for degrees of freedom, and on z-values and asymptotic Wald tests for the generalised linear mixed-effect model of accuracy[96].

Causal mediation analysis

We performed causal mediation analysis to model the direct as well as the hearing-acuity–mediated effect of age on accuracy and response speed[21]. Critically, these models also included a direct (i.e. independent of age and hearing loss) path of neural filtering on behaviour. To formally test the stability of direct and indirect relationships across time, we used a moderated mediation analysis[97]. In this analysis, the inclusion of interactions by timepoint tested whether the influence of age, sensory acuity, and neural functioning on behaviour varied across time. We z-scored all dependent and independent variables, and estimated the magnitude of direct and mediated effects along with percentile-based confidence intervals on the basis of 1,000 replications.

Bayes Factor calculation

To facilitate interpretation of non-significant effects, we calculated the Bayes factor (BF) based on the comparison of Bayesian information criterion (BIC)[98]. We report log Bayes Factors, with a log BF_{01} of 0 representing equal evidence for and against the null

hypothesis; log BF_{01} with a positive sign indicating relatively more evidence for the null hypothesis than the alternative hypothesis, and vice versa. Magnitudes > 1 are taken as moderate, > 2.3 as strong evidence for either of the alternative or null hypotheses, respectively[99].

Latent change score modelling

In the third and final step of our analysis approach, we used structural equation modelling (SEM) to investigate the role of neural filtering as a predictor of behavioural change. We used latent change score modelling (LCSM)[16,17] to test (i) whether an individual's T1 neural filtering strength can be predictive two-year changes in behaviour, and (ii) whether two-year changes in neural filtering and listening behaviour are systematically related. All models were specified and fitted with the R-based package lavaan[93] using maximum likelihood estimation.

To bring all manifest variables onto the same scale while preserving mean differences over time, we first stacked them across timepoint and then rescaled them using the proportion of maximum scale ('POMS') method[100,101]. We assessed model fit using established indices including the χ^2 test, the root mean square error of approximation (RMSEA), and the comparative fit index (CFI)[29,102]. Likelihood-ratio tests helped us decide whether (i) constraining individual parameters to be equal significantly decreased model fit, and (ii) individual parameter estimates were statistically significant. We report standardized parameter estimates.

In a first step, we specified separate unstructured measurement models of accuracy, speed, and neural filtering to establish factorial invariance across time using a series of tests[17,103]. Latent variables representing each metric at T1 and T2 were constructed from the observed individual means averaged across the first and second half of the experiment, respectively. Covariances between T1 and T2 latent variables were freely estimated, residual covariances were set to be equal for the first and second half of the experiment. The factor loading and mean of the first manifest variable were set to 1 and 0, respectively, to ensure model identification[101]. Given our choice of POMS-transformation of raw values to preserve mean differences over time, the mean of the second manifest variable had to be freely estimated to avoid model misfit. We sequentially tested for metric (i.e., identical factor loadings), strong (i.e., identical means), and strict (i.e., identical residuals) time invariance using likelihood ratio tests.

Next, for those metrics surviving factorial invariance testing, we constructed separate (i.e., univariate) change score models to test for group-level mean change following the tutorial by Kievit et al. (2018). We included a regression path from the T1 latent variable to the latent change variable to test how baseline functioning impacted the degree of longitudinal change. A likelihood ratio test of nested models (with mean change being freely estimated vs. constrained to be 0) determined the significance of group-level mean change.

In the last and final step, we constructed the bivariate LCSM that connected changes in response speed and neural filtering, which both significantly increased over time (see Fig. S2 for full model details). We modelled the covariance between the baseline neural filtering and speed, as well as the covariance between neural and behavioural change. In addition, we included a regression path from T1 neural filtering to the T1-T2 change in response speed to directly test the predictive potency of baseline neural functioning to explain behavioural change.

Following the logic of a generalised variance test[104], we tested for significant inter-individual differences in neural and behavioural change. Note that such reliable variance in change is a necessary prerequisite for testing change-change correlations. In short, per domain (neural, behaviour), a likelihood ratio test compared the full model with a restricted model in which all paths pointing to or connected with the latent change variable are fixed to 0.

Finally, we asked whether neural and behavioural change covaried significantly. We ran a likelihood ratio test on the full model compared to a restricted model in which the covariance parameter was fixed to 0. For visualisation of results, we re-fit the final model using OpenMx[94] to predict latent factor scores based on maximum likelihood.

Data and code 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. Lindenberger, U. Human cognitive aging: Corriger la fortune? *Science* **346,** 572–578 (2014).
2. Akeroyd, M. A. Are individual differences in speech reception related to individual differences in cognitive ability? A survey of twenty experimental studies with normal and hearing-impaired adults. in **47,** S53–71 (2008).
3. Houtgast, T. & Festen, J. M. On the auditory and cognitive functions that may explain an individual's elevation of the speech reception threshold in noise. *Int J Audiol* **47,** 287–295 (2008).
4. Humes, L. E. *et al.* Central presbycusis: A review and evaluation of the evidence. *J Am Acad Audiol* **23,** 635–666 (2012).
5. Cherry, E. C. Some experiments on the recognition of speech, with one and with two ears. *J. Acoust. Soc. Am.* **25,** 975–979 (1953).
6. Broadbent, D. E. *Perception and communication*. (Pergamon Press, 1958). doi:10.1186/s12889-020-10047-2
7. Obleser, J. & Erb, J. in *The Cognitive Neurosciences* (eds. Poeppel, D., Mangun, G. S. & Gazzaniga, M. S.) (MIT Press, 2020). doi:10.7551/mitpress/11442.001.0001
8. Ding, N. & Simon, J. Z. Emergence of neural encoding of auditory objects while listening to competing speakers. *Proc. Natl. Acad. Sci. U.S.A.* **109,** 11854–11859 (2012).
9. Zion Golumbic, E. M. Z. *et al.* Mechanisms Underlying Selective Neuronal Tracking of Attended Speech at a '"Cocktail Party"'. *Neuron* **77,** 980–991 (2013).
10. Obleser, J. & Kayser, C. Neural Entrainment and Attentional Selection in the Listening Brain. *Trends Cogn. Sci.* **23,** 913–926 (2019).
11. Lakatos, P., Karmos, G., Mehta, A. D., Ulbert, I. & Schroeder, C. E. Entrainment of neuronal oscillations as a mechanism of attentional selection. *Science* **320,** 110–113 (2008).
12. Tune, S., Alavash, M., Fiedler, L. & Obleser, J. Neural attentional-filter mechanisms of listening success in middle-aged and older individuals. *Nat Commun* 1–14 (2021). doi:10.1038/s41467-021-24771-9
13. O'Sullivan, J. A. *et al.* Attentional Selection in a Cocktail Party Environment Can Be Decoded from Single-Trial EEG. *Cereb. Cortex* **25,** 1697–1706 (2014).
14. Mesgarani, N. & Chang, E. F. Selective cortical representation of attended speaker in multi-talker speech perception. *Nature* **485,** 233–236 (2012).
15. Woo, C.-W., Chang, L. J., Lindquist, M. A. & Wager, T. D. Building better biomarkers: brain models in translational neuroimaging. *Nat. Neurosci.* **20,** 365–377 (2017).
16. McArdle, J. J. Latent Variable Modeling of Differences and Changes with Longitudinal Data. *Annu. Rev. Psychol.* **60,** 577–605 (2009).
17. Kievit, R. A. *et al.* Developmental cognitive neuroscience using latent change score models: A tutorial and applications. *Dev Cogn Neurosci* **33,** 99–117 (2018).
18. Oschwald, J. *et al.* Brain structure and cognitive ability in healthy aging: A review on longitudinal correlated change. *Reviews in the Neurosciences* (2019). doi:10.1515/revneuro-2018-0096
19. Alavash, M., Tune, S. & Obleser, J. Dynamic large-scale connectivity of intrinsic cortical oscillations supports adaptive listening in challenging conditions. *PLoS Biol* **19,** e3001410 (2021).
20. Alavash, M., Tune, S. & Obleser, J. Modular reconfiguration of an auditory control brain network supports adaptive listening behavior. *Proc. Natl. Acad. Sci. U.S.A.* **116,** 660–669 (2019).
21. Imai, K., Keele, L. & Tingley, D. A general approach to causal mediation analysis. *Psychol. Methods* **15,** 309–334 (2010).
22. McArdle, J. J. & Nesselroade, J. R. Using multivariate data to structure developmental change. *Life-span developmental psychology: Methodological contributions* 223–267 (1994).
23. Linssen, A. M., Van Boxtel, M. P. J., Joore, M. A. & Anteunis, L. J. C. Predictors of hearing acuity: Cross-sectional and longitudinal analysis. *J. Gerontol. A Biol. Sci. Med. Sci.* **69,** 759–765 (2014).

24. Rigters, S. C., Van Der Schroeff, M. P., Papageorgiou, G., Baatenburg De Jong, R. J. & Goedegebure, A. Progression of hearing loss in the aging population: Repeated auditory measurements in the Rotterdam study. *Audiol Neurootol* **23,** 290–297 (2019).
25. Wiley, T. L., Chappell, R., Carmichael, L., Nondahl, D. M. & Cruickshanks, K. J. Changes in hearing thresholds over 10 years in older adults. *J Am Acad Audiol* **19,** 281–92– quiz 371 (2008).
26. Corcoran, A. W., Alday, P. M., Schlesewsky, M. & Bornkessel-Schlesewsky, I. Toward a reliable, automated method of individual alpha frequency (IAF) quantification. *Psychophysiology* **57,** e13064–21 (2018).
27. Donoghue, T. *et al.* Parameterizing neural power spectra into periodic and aperiodic components. *Nat. Neurosci.* 1–24 (2020). doi:10.1038/s41593-020-00744-x
28. Voytek, B. *et al.* Age-related changes in 1/f neural electrophysiological noise. *J. Neurosci.* **35,** 13257–13265 (2015).
29. Raykov, T. & Marcoulides, G. A. On multilevel model reliability estimation from the perspective of structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal* **13,** 130–141 (2006).
30. Peelle, J. E. & Wingfield, A. The Neural Consequences of Age-Related Hearing Loss. *Trends Neurosci.* **39,** 486–497 (2016).
31. Power, A. J., Foxe, J. J., Forde, E.-J., Reilly, R. B. & Lalor, E. C. At what time is the cocktail party? A late locus of selective attention to natural speech. *Eur. J. Neurosci.* **35,** 1497–1503 (2012).
32. O'Sullivan, J. *et al.* Hierarchical Encoding of Attended Auditory Objects in Multi-talker Speech Perception. *Neuron* **104,** 1195–1204 (2019).
33. Gross, J. *et al.* Speech Rhythms and Multiplexed Oscillatory Sensory Coding in the Human Brain. *PLoS Biol* **11,** e1001752–14 (2013).
34. Fiedler, L., Wöstmann, M., Herbst, S. K. & Obleser, J. Late cortical tracking of ignored speech facilitates neural selectivity in acoustically challenging conditions. *NeuroImage* **186,** 33–42 (2019).
35. Kaufman, M. & Golumbic, E. Z. Listening to two speakers: Capacity and tradeoffs in neural speech tracking during Selective and Distributed Attention. *NeuroImage* **270,** 119984 (2023).
36. Kraus, F., Tune, S., Ruhe, A., Obleser, J. & Wöstmann, M. Unilateral Acoustic Degradation Delays Attentional Separation of Competing Speech. *Trends in Hearing* **25,** 23312165211013242 (2021).
37. Brodbeck, C., Jiao, A., Hong, L. E. & Simon, J. Z. Neural speech restoration at the cocktail party: Auditory cortex recovers masked speech of both attended and ignored speakers. *PLoS Biol* **18,** e3000883–22 (2020).
38. Forte, A. E., Etard, O. & Reichenbach, T. The human auditory brainstem response to running speech reveals a subcortical mechanism for selective attention. *eLife* **6,** (2017).
39. Fuglsang, S. A., Märcher-Rørsted, J., Dau, T. & Hjortkjær, J. Effects of Sensorineural Hearing Loss on Cortical Synchronization to Competing Speech during Selective Attention. *J. Neurosci.* **40,** 2562–2572 (2020).
40. Orf, M., Wöstmann, M., Hannemann, R. & Obleser, J. Target enhancement but not distractor suppression in auditory neural tracking during continuous speech. *ISCIENCE* **26,** 106849 (2023).
41. Schmitt, R., Meyer, M. & Giroud, N. Better speech-in-noise comprehension is associated with enhanced neural speech tracking in older adults with hearing impairment. *CORTEX* **151,** 133–146 (2022).
42. Song, J. H., Nicol, T. & Kraus, N. Test-retest reliability of the speech-evoked auditory brainstem response. *Clin Neurophysiol* **122,** 346–355 (2011).
43. Easwar, V., Scollie, S., Aiken, S. & Purcell, D. Test-Retest Variability in the Characteristics of Envelope Following Responses Evoked by Speech Stimuli. *Ear and Hearing* **41,** 150–164 (2020).
44. Cabral-Calderin, Y. & Henry, M. J. Reliability of Neural Entrainment in the Human Auditory System. *J. Neurosci.* **42,** 894–908 (2022).

45. Panela, R. A., Copelli, F. & Herrmann, B. Reliability and generalizability of neural speech tracking in younger and older adults. *Nature Communications* 2023.07.26.550679 (2023). doi:10.1101/2023.07.26.550679
46. Hertzog, C. & Schaie, K. W. Stability and change in adult intelligence: 2. Simultaneous analysis of longitudinal means and covariance structures. *Psychol Aging* **3,** 122–130 (1988).
47. Lakatos, P. *et al.* Global dynamics of selective attention and its lapses in primary auditory cortex. *Nat. Neurosci.* **19,** 1707–1717 (2016).
48. Lakatos, P., Gross, J. & Thut, G. A New Unifying Account of the Roles of Neuronal Entrainment. *Curr. Biol.* **29,** R890–R905 (2019).
49. Schroeder, C. E. & Lakatos, P. Low-frequency neuronal oscillations as instruments of sensory selection. *Trends Neurosci.* **32,** 9–18 (2009).
50. Schroeder, C. E., Wilson, D. A., Radman, T., Scharfman, H. & Lakatos, P. Dynamics of Active Sensing and perceptual selection. *Current Opinion in Neurobiology* **20,** 172–176 (2010).
51. Helfrich, R. F. *et al.* Neural Mechanisms of Sustained Attention Are Rhythmic. *Neuron* **99,** 854–865.e5 (2018).
52. Buschman, T. J. & Kastner, S. From Behavior to Neural Dynamics: An Integrated Theory of Attention. *Neuron* **88,** 127–144 (2015).
53. Van Eyndhoven, S., Francart, T. & Bertrand, A. EEG-informed attended speaker extraction from recorded speech mixtures with application in neuro-steered hearing prostheses. *IEEE Transactions on Biomedical Engineering* **64,** 1045–1056 (2016).
54. Ceolini, E. *et al.* Brain-informed speech separation (BISS) for enhancement of target speaker in multitalker speech perception. *NeuroImage* **223,** (2020).
55. O'Sullivan, J. *et al.* Neural decoding of attentional selection in multi-speaker environments without access to clean sources. *J Neural Eng* **14,** 056001 (2017).
56. Gillis, M., Vanthornhout, J., Simon, J. Z., Francart, T. & Brodbeck, C. Neural Markers of Speech Comprehension: Measuring EEG Tracking of Linguistic Speech Representations, Controlling the Speech Acoustics. *J. Neurosci.* **41,** 10316–10329 (2021).
57. Di Liberto, G. M., Hjortkjær, J. & Mesgarani, N. Editorial: Neural Tracking: Closing the Gap Between Neurophysiology and Translational Medicine. *Front. Neurosci.* **16,** 872600 (2022).
58. Gillis, M., Decruy, L., Vanthornhout, J. & Francart, T. Hearing loss is associated with delayed neural responses to continuous speech. *Eur. J. Neurosci.* **55,** 1671–1690 (2022).
59. Presacco, A., Simon, J. Z. & Anderson, S. Speech-in-noise representation in the aging midbrain and cortex: Effects of hearing loss. *PLoS ONE* **14,** e0213899–26 (2019).
60. Presacco, A., Simon, J. Z. & Anderson, S. Evidence of degraded representation of speech in noise, in the aging midbrain and cortex. *J. Neurophysiol.* **116,** 2346–2355 (2016).
61. Decruy, L., Vanthornhout, J. & Francart, T. Evidence for enhanced neural tracking of the speech envelope underlying age-related speech-in-noise difficulties. *J. Neurophysiol.* **122,** 601–615 (2019).
62. Decruy, L., Vanthornhout, J. & Francart, T. Hearing impairment is associated with enhanced neural tracking of the speech envelope. *Hearing Research* **393,** 107961 (2020).
63. Lindenberger, U., Oertzen, von, T., Ghisletta, P. & Hertzog, C. Cross-Sectional Age Variance Extraction: What's Change Got To Do With It? *Psychol Aging* **26,** 34–47 (2011).
64. Tucker-Drob, E. M. *et al.* A strong dependency between changes in fluid and crystallized abilities in human cognitive aging. *Sci Adv* **8,** eabj2422 (2022).
65. Boker, S. M. & Martin, M. A Conversation between Theory, Methods, and Data. *Multivariate Behav Res* **0,** 1–14 (2018).
66. Raz, N. & Lindenberger, U. Only time will tell: Cross-sectional studies offer no solution to the age-brain-cognition triangle: Comment on salthouse (2011). *Psychological Bulletin* **137,** 790–795 (2011).
67. Raz, N. Regional brain changes in aging healthy adults: general trends, individual differences and modifiers. *Cereb. Cortex* **15,** 1676–1689 (2005).

68. McArdle, J. J. *et al.* Structural modeling of dynamic changes in memory and brain structure using longitudinal data from the normative aging study. *J Gerontol B Psychol Sci Soc Sci* **59,** P294–304 (2004).
69. Bender, A. R., Prindle, J. J., Brandmaier, A. M. & Raz, N. White matter and memory in healthy adults: Coupled changes over two years. *NeuroImage* **131,** 193–204 (2016).
70. Ritchie, S. J. *et al.* Coupled changes in brain white matter microstructure and fluid intelligence in later life. *J. Neurosci.* **35,** 8672–8682 (2015).
71. Lövdén, M. *et al.* Changes in perceptual speed and white matter microstructure in the corticospinal tract are associated in very old age. *NeuroImage* **102,** 520–530 (2014).
72. Persson, N. *et al.* Regional brain shrinkage and change in cognitive performance over two years: The bidirectional influences of the brain and cognitive reserve factors. *NeuroImage* **126,** 15–26 (2016).
73. Bennett, I. J. & Madden, D. J. Disconnected aging: cerebral white matter integrity and age-related differences in cognition. *Neuroscience* **276,** 187–205 (2014).
74. Grady, C. The cognitive neuroscience of ageing. *Nat. Rev. Neurosci.* **13,** 491–505 (2012).
75. Reuter-Lorenz, P. A. & Park, D. C. How Does it STAC Up? Revisiting the Scaffolding Theory of Aging and Cognition. *Neuropsychol Rev* **24,** 355–370 (2014).
76. Cabeza, R. *et al.* Maintenance, reserve and compensation: the cognitive neuroscience of healthy ageing. *Nat. Rev. Neurosci.* **19,** 1–10 (2018).
77. Stern, Y., Barnes, C. A., Grady, C., Jones, R. N. & Raz, N. Brain reserve, cognitive reserve, compensation, and maintenance: operationalization, validity, and mechanisms of cognitive resilience. *Neurobiology of Aging* **83,** 124–129 (2019).
78. Freund, J. *et al.* Emergence of individuality in genetically identical mice. *Science* **340,** 756–759 (2013).
79. 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* **300,** 18–32 (2013).
80. Peelle, J. E. Language and aging. The Oxford Handbook of Neurolinguistics (de Zubicaray and Schiller, eds), pp. 295–316. Oxford University Press.
81. 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.* **48,** 2537–2550 (2018).
82. Gillis, M., Van Canneyt, J., Francart, T. & Vanthornhout, J. Neural tracking as a diagnostic tool to assess the auditory pathway. *Hearing Research* 108607 (2022).
83. Strauß, A., Wöstmann, M. & Obleser, J. Cortical alpha oscillations as a tool for auditory selective inhibition. *Front. Hum. Neurosci.* **8,** 1–7 (2014).
84. Jacobucci, R., Serang, S. & Grimm, K. J. A Short Note on Interpretation in the Dual Change Score Model. *Structural Equation Modeling: A Multidisciplinary Journal* **0,** 1–7 (2019).
85. Rast, P. & Hofer, S. M. Longitudinal design considerations to optimize power to detect variances and covariances among rates of change: simulation results based on actual longitudinal studies. *Psychol. Methods* **19,** 133 (2014).
86. Jefferies, K. & Gale, T. M. in *Cognitive Screening Instruments* 209–218 (Springer London, 2012). doi:10.1007/978-1-4471-2452-8_11
87. Nasreddine, Z. S. *et al.* The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. *J Am Geriatr Soc* **53,** 695–699 (2005).
88. Oldfield, R. C. The assessment and analysis of handedness: The Edinburgh inventory. **9,** 97–113 (1971).
89. Crosse, M. J., Di Liberto, G. M., Bednar, A. & Lalor, E. C. The Multivariate Temporal Response Function (mTRF) Toolbox: A MATLAB Toolbox for Relating Neural Signals to Continuous Stimuli. *Front. Hum. Neurosci.* **10,** 604 (2016).
90. R Core Team. R: A language and environment for statistical computing. (2019). doi:https://www.R-project.org/
91. Bates, D., Mächler, M., Bolker, B. & Walker, S. Fitting Linear Mixed-Effects Models Using lme4. *J. Stat. Soft.* **67,** 1–48 (2015).
92. Tingley, D., Yamamoto, T., Keele, L. & Imai, K. Mediation: R package for causal mediation analysis. *J. Stat. Soft.* **59,** 1–38 (2014).

93. Rosseel, Y. lavaan: An R package for structural equation modeling. *J. Stat. Soft.* **48,** 1–36 (2012).
94. Neale, M. C. *et al.* OpenMx 2.0: Extended Structural Equation and Statistical Modeling. *Psychometrika* **81,** 535–549 (2016).
95. Bell, A., Fairbrother, M. & Jones, K. Fixed and random effects models: making an informed choice. *Qual Quant* **53,** 1051–1074 (2018).
96. Luke, S. G. Evaluating significance in linear mixed-effects models in R. *Behav Res* **49,** 1494–1502 (2017).
97. Preacher, K. J., Rucker, D. D. & Hayes, A. F. Addressing Moderated Mediation Hypotheses: Theory, Methods, and Prescriptions. *Multivariate Behav Res* **42,** 185–227 (2007).
98. Wagenmakers, E.-J. A practical solution to the pervasive problems of p values. *Psychonomic Bulletin & Review* **14,** 779–804 (2007).
99. Lee, M. D. & Wagenmakers, E.-J. *Bayesian Cognitive Modeling*. (Cambridge University Press, 2014).
100. Moeller, J. A word on standardization in longitudinal studies: don't. *Front. Psychology* **6,** 1389 (2015).
101. Little, T. D. *Longitudinal structural equation modeling*. (Guilford press, 2013).
102. Kline, R. B. *Principles and practice of structural equation modeling*. (2023).
103. Polk, S. E. *et al.* Change in Latent Gray-Matter Structural Integrity Is Associated With Change in Cardiovascular Fitness in Older Adults Who Engage in At-Home Aerobic Exercise. *Front. Hum. Neurosci.* **16,** 852737 (2022).
104. Brandmaier, A. M., Oertzen, von, T., Ghisletta, P., Lindenberger, U. & Hertzog, C. Precision, reliability, and effect size of slope variance in latent growth curve models: Implications for statistical power analysis. *Front. Psychology* **9,** (2018).

Acknowledgments

Research was funded by the European Research Council (grant no. ERC-CoG-2014-646696 "Audadapt" awarded to J.O.). All data were acquired with the help of Franziska Scharata. The comments of three anonymous reviewers helped improve the manuscript.