Highlights

Peripheral phoneme encoding and discrimination in aging and hearing impairment

Marjoleen Wouters, Etienne Gaudrain, Konrad Dapper, Jakob Schirmer, Deniz Başkent, Lukas Rüttiger, Marlies Knipper, Sarah Verhulst

- Speech-evoked EEG shows OHC loss-related decline of high-CF envelope encoding.
- Speech-evoked EEG shows low-CF envelope encoding stays intact with age.
- Fine-structure contrast discrimination worsens with OHC loss in quiet.
- Fine-structure contrast discrimination worsens with age in contralateral noise.
- High-frequency place-based spectral cues discrimination remains robust with age.
- Peripheral coding strength is not directly reflected at behavioral level.

Peripheral phoneme encoding and discrimination in aging and hearing impairment

Marjoleen Wouters[a,*], Etienne Gaudrain[b], Konrad Dapper[c], Jakob Schirmer[c], Deniz Başkent[d], Lukas Rüttiger[c], Marlies Knipper[c], Sarah Verhulst[a]

[a]*Hearing Technology at Wireless, Acoustics, Environment and Expert Systems, Department of Information Technology, Ghent University, Technologiepark 126, 9052, Zwijnaarde, Belgium*
[b]*Lyon Neuroscience Research Center, Université Claude Bernard Lyon 1, CNRS UMR5292, INSERM U1028, Center Hospitalier Le Vinatier-Bâtiment 462-Neurocampus, Boulevard Pinel 95, 69675 Bron CEDEX, Lyon, France*
[c]*Department of Otolaryngology, Head and Neck, University of Tübingen, Elfriede-Aulhorn-Str. 5, 72076, Tübingen, Germany*
[d]*Department of Otorhinolaryngology, University Medical Center Groningen (UMCG), University of Groningen, Hanzeplein 1, BB21, 9700, Groningen, the Netherlands*

Abstract

Speech perception difficulties in noise are common among older adults and individuals with hearing impairment, even when audiometric thresholds appear normal. We examined how aging, cochlear synaptopathy (CS), and outer hair cell (OHC) damage affect speech encoding and phoneme discrimination. Envelope-following responses (EFRs) to rectangular amplitude-modulated (RAM) tones and speech-like phoneme pairs were recorded in quiet using EEG, and behavioral discrimination was assessed in quiet, ipsilateral, and contralateral noise. Stimuli were designed to target temporal envelope (TENV) or temporal fine structure (TFS) encoding. Results showed that RAM-EFR amplitudes decreased gradually with age, consistent with emerging CS, while magnitudes of high-frequency TENV-based EFRs

*Corresponding author
Email address: marjoleen.wouters@ugent.be (Marjoleen Wouters)

in quiet were most reduced in older hearing-impaired listeners with combined CS and OHC damage. In contrast, EFRs targeting low-frequency TENV encoding in quiet remained preserved. Behaviorally, phoneme discrimination of TFS contrasts worsened with OHC loss and age in quiet and contralateral noise, respectively, while there was no significant effect of age on the discrimination of TENV contrasts. Considering that high-frequency contrasts are discriminated via place-based spectral cues, low-frequency contrasts rely on TFS, and the EFR reflects primarily TENV, this framework explains why EFRs decline for high-frequency cues without perceptual loss, while EFRs remain stable for low-frequency cues even as TFS-based discrimination deteriorates. These findings highlight the need for further investigation into how neural coding deficits relate to perceptual outcomes. Combining electrophysiological and behavioral measures might provide a sensitive framework for detecting subclinical auditory deficits to earlier diagnose age-related and hidden hearing loss.

Keywords: Cochlear synaptopathy, Age-related hearing loss, Auditory-evoked potentials, Phoneme discrimination

---

1. Introduction

Speech understanding in noisy environments is a common complaint among individuals with hearing difficulties, traditionally attributed to age-related or acquired sensorineural hearing loss (SNHL) (Furman et al., 2013; Kujawa and Liberman, 2009; Lobarinas et al., 2017). A well-established hallmark of SNHL is outer hair cell (OHC) damage, which reduces cochlear gain and elevates hearing thresholds on the pure-tone audiogram, the standard diag-

nostic tool in clinical audiology (Moore, 2007). However, mounting evidence suggests that impaired speech-in-noise perception can also occur in the absence of elevated audiometric thresholds, pointing to additional mechanisms beyond OHC damage (Plack et al., 2014; Liberman, 2017; Oxenham, 2016).

One such other aspect of SNHL is cochlear synaptopathy (CS), often called 'hidden hearing loss'. CS is characterized by degeneration of the synapses between inner hair cells (IHCs) and auditory nerve (AN) fibers, resulting in a reduced neural representation of sound without clinically relevant loss of audiometric pure-tone thresholds within the main hearing range of the human (Kujawa and Liberman, 2009; Sergeyenko et al., 2013; Liberman et al., 2015; Möhrle et al., 2016, 2019). Consequently, individuals with clinically normal audiograms may still experience substantial difficulties understanding speech in noise. CS is commonly associated with aging, noise exposure, and ototoxic drug use, and may precede OHC degeneration (Furman et al., 2013; Bramhall et al., 2017). In CS, low-spontaneous-rate (low-SR) AN fibers are typically affected first (Kujawa and Liberman, 2009). These fibers have high thresholds and are crucial for encoding sounds in noisy environments and amplitude modulation. Because they are more vulnerable to noise exposure and aging, their synapses are often lost before those of high-SR fibers, even when audiometric thresholds remain normal (Kujawa and Liberman, 2009, 2015). Given these features, CS is considered a form of subclinical hearing loss, defined as subtle auditory deficits within clinically 'normal' hearing ranges (defined as pure-tone threshold average of the frequencies 0.5, 1, 2 and 4 kHz (PTA4) $\leq$ 25 dB HL (Humes, 2018)) (Liberman et al., 2016). Individuals with subclinical hearing loss often report difficul-

ties with speech-in-noise perception, which can negatively impact cognition, psychosocial well-being, and educational outcomes, yet frequently remain undiagnosed and untreated (Bharadwaj et al., 2015).

Electrophysiological measures have proven valuable in identifying neural markers of CS. In particular, the envelope-following response (EFR) evoked by rectangularly amplitude-modulated (RAM) pure tones has been proposed as a sensitive correlate of synaptic integrity (Vasilkov et al., 2021; Keshishzadeh et al., 2021; Vasilkov and Verhulst, 2019). The RAM-EFR has been shown to be maximally insensitive to other aspects of SNHL such as OHC damage, since the sharply rising envelope of the RAM stimulus causes a large, synchronized neural response almost instantly, what minimizes the time for the cochlear amplifier to function (Keshishzadeh et al., 2020, 2021; Vasilkov et al., 2021; Van Der Biest et al., 2023). This makes the RAM-EFR a better marker for CS and nerve damage, a key feature of SNHL, as it is less influenced by the mechanical amplification provided by the OHCs. Additionally, speech-evoked auditory evoked potentials (AEPs) allow for assessment of temporal coding at different frequency ranges: low-frequency temporal fine structure (TFS) coding below the phase-locking limit (PLL, around 1.5 kHz), and high-frequency temporal envelope (TENV) coding above the PLL (Lorenzi et al., 2006; Verschooten et al., 2019). By comparing neural and behavioral responses to phoneme contrasts presented in quiet, ipsilateral noise, and contralateral noise, it is possible to probe how CS, OHC damage, and aging differentially affect speech encoding.

In the present study, we extend our previous work of Schirmer et al. (2024) and Dapper et al. (2025) by investigating the relationship between peripheral

speech encoding and phoneme discrimination across the lifespan and in the presence of hearing impairment. We focus on disentangling the effects of aging, CS, and OHC damage on (i) RAM-EFR magnitude, (ii) speech-evoked EEG responses to phoneme contrasts above and below the PLL, and (iii) behavioral phoneme discrimination quiet and in noise. Our participant groups include young normal-hearing (yNH), middle-aged normal-hearing (mNH), older normal-hearing (oNH), and older hearing-impaired (oHI) adults.

We hypothesize that: (i) CS and aging will be reflected in reduced RAM-EFR and speech-EEG responses, even in normal-hearing groups, (ii) phoneme discrimination, particularly in ipsilateral noise, will be impaired in individuals with CS, with consonant confusions being most prominent, and (iii) OHC damage will further exacerbate deficits in oHI individuals.

By investigating peripheral speech encoding and phoneme discrimination across the different participant groups that differ in age and hearing status, we expect that the yNH groups serve as a baseline control, with robust AEPs, intact speech encoding and phoneme discrimination in quiet and noise, and no CS or central gain changes (increased neural responsiveness in central auditory pathways due to SNHL in the cochlea) are expected (Möhrle et al., 2016; Sergeyenko et al., 2013; Parthasarathy and Kujawa, 2018). For the mNH group, we expect to see subtle age-related synaptopathy emerge (Möhrle et al., 2016; Sergeyenko et al., 2013; Parthasarathy and Kujawa, 2018), with possible central gain compensation (Schirmer et al., 2024). Speech coding is expected to degrade more in noise than in quiet, likely due to reduced sampling by AN fibers, because when fewer AN fibers carry detailed information, the neural representation of speech becomes less

precise, especially in challenging listening situations (Lopez-Poveda and Barrios, 2013). In the oNH group, we expect even stronger age-related effects on AEPs and speech encoding, particularly in noise. Central gain changes may partially compensate, but not restore, normal performance (Sergeyenko et al., 2013; Auerbach et al., 2014; Möhrle et al., 2016). Finally, in the oHI group, we expect combined effects of OHC damage, age-related synaptopathy, and central gain changes, with strongly degraded speech AEPs, and worse speech discrimination in quiet and noise.

By combining EEG markers of peripheral speech encoding with phoneme discrimination performance across carefully stratified age and hearing-status groups, this study aims to differentiate the contributions of aging, CS, and OHC damage to speech perception deficits. Identifying reliable neural markers for subclinical hearing loss will be essential for advancing early diagnosis and targeted interventions.

2. Materials and Methods

2.1. Ethics statement

"The study was conducted in the Department of Otolaryngology of the University of Tübingen, and approved by the ethics committee of Tübingen University (Faculty of Medicine; ethical approval number 392/2021BO2). Written, informed consent was given by all participants. All methods followed the Declaration of Helsinki by the World Medical Association (WMA) for human research ethics." (Schirmer et al., 2024)

2.2. Participant selection

In this study, we selected a subset of 49 out of the 112 participants who were recruited in the study of Schirmer et al. (2024), based on their age and audiogram. Within each group, the age range was limited in order to be able to neglect the effect of age. Audiometric hearing thresholds were assessed at 11 frequencies ranging from 0.125 to 10 kHz, as well as at 4 frequencies for the extended high-frequency range of 11.2 to 16 kHz (Schirmer et al., 2024). The pure-tone threshold average of the frequencies 0.5, 1, 2 and 4 kHz (PTA4) was calculated for each subject. The subset of subjects consists of five groups: yNH1 (young normal-hearing, 18-24 years old, n = 10, 5 females), yNH2 (young normal-hearing, 25-31 years old, n = 10, 5 females), mNH (middle-aged normal-hearing, 32-48 years old, n = 9, 7 females), oNH (older normal-hearing, 49-65 years old, n = 10, 7 females, with better PTA4), oHI (older hearing-impaired, 49-65 years old, n = 10, 8 females, with worse PTA4). Figure 1 shows the audiometric thresholds for all five groups, as well as their age distribution.

Participants in the first four groups were selected from their age group based on the lowest PTA4 values, all below 20 dB HL and hence considered normal-hearing (Humes, 2018; Von Gablenz and Holube, 2017), while participants in oHI were selected to have higher PTA4 values (the highest within their age group). The PTA4 values for each group are visualized in Figure 2a using boxplots with overlaid individual data points as solid black circles, and outliers additionally displayed as hollow circles. A Kruskal-Wallis H test, followed by Dunn's post hoc comparisons with Holm correction, implemented in Python (v.3.9.13), showed indeed a statistically significantly higher PTA4

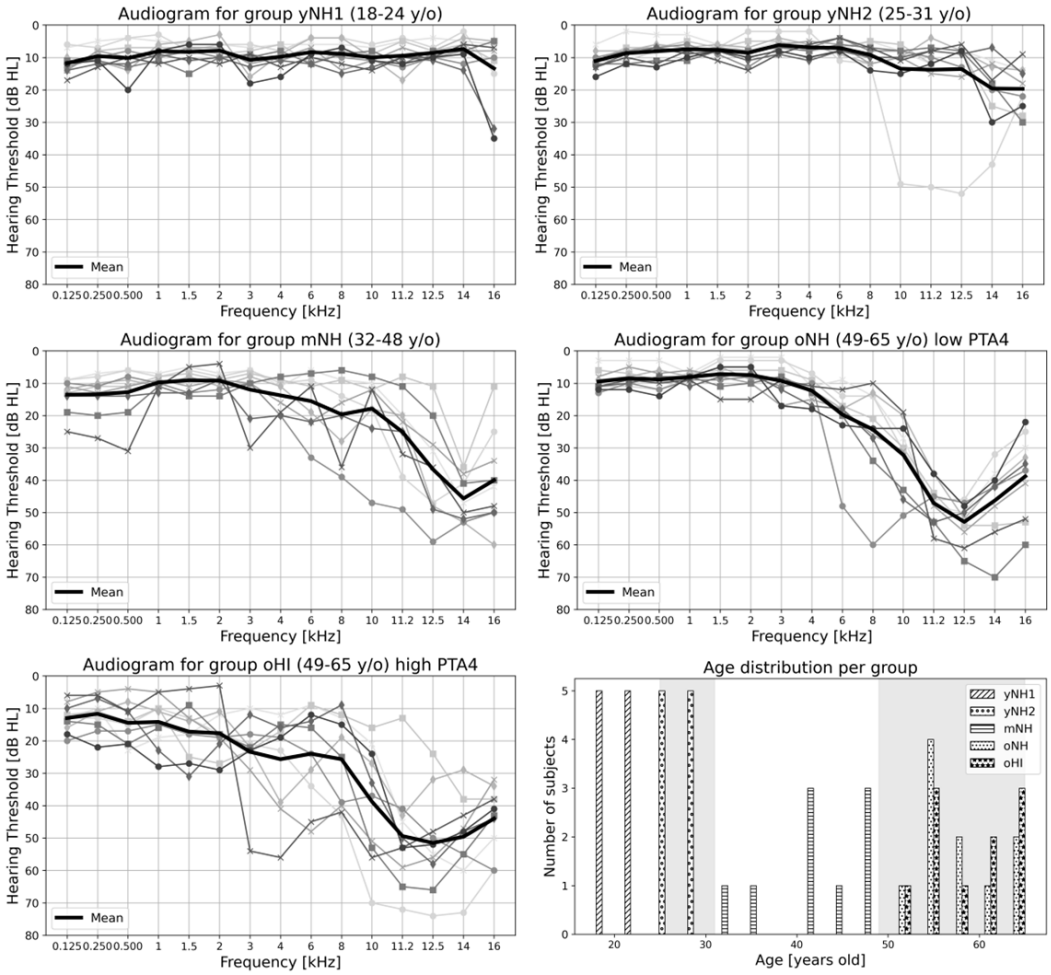
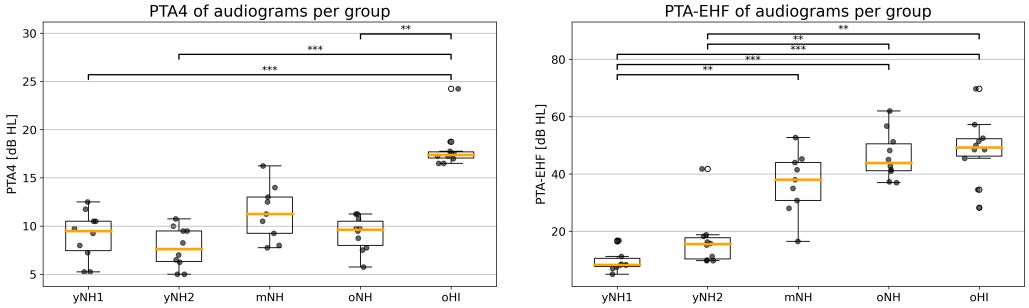


Figure 1: Audiograms and age histogram per group. In the age histogram, grey background bands were added to visually separate the predefined groups.

value in group oHI than in other groups (yNH1: $p < 0.001$, yNH2: $p < 0.001$ and oNH: $p < 0.01$). The number of participants per group is limited because of the requirement for good PTA4 values (below 20 dB HL), exclusion criteria (Schirmer et al., 2024) and the availability of complete recordings of their speech-EEG.



(a) PTA4 values per group.

(b) PTA-EHF values per group.

Figure 2: Group comparisons for PTA4 (a) and PTA-EHF (b) ($p < 0.001$:***; $p < 0.01$:**; $p < 0.05$:*).

For all five groups, we also looked at the pure-tone threshold average for the extended high-frequencies (PTA-EHF), namely at 11.2, 12.5, 14 and 16 kHz, which are especially affected by neural degeneration (Bramhall et al., 2017; Lobarinas et al., 2013; Liberman et al., 2016). Figure 2b shows the PTA-EHF values per group. To investigate differences in PTA-EHF values across five participant groups, a Shapiro-Wilk test was first conducted to assess the normality of the data within each group. The results indicated that two groups (yNH1: W = 0.8161, p = 0.0228; yNH2: W = 0.6960, p = 0.0008) significantly deviated from normality, while the remaining groups did not ($p > 0.05$). Given this violation of the normality assumption, a Kruskal-Wallis H test was employed. The test revealed a statistically significant difference in PTA-EHF thresholds across the groups (H = 34.977, $p < 0.0001$), suggesting that at least one group may differ from the others. Pairwise comparisons were conducted using Dunn's test with Holm correction to control for multiple comparisons. The results were visualized in Figure 2b using boxplots with overlaid individual data points and annotated significance bars,

highlighting statistically significant pairwise differences. The Dunn's test showed that the PTA-EHF values of yNH1 were significantly lower than those of mNH ($p < 0.01$), oNH ($p < 0.001$) and oHI ($p < 0.001$). PTA-EHF values of yNH2 were significantly lower than those of oNH ($p < 0.01$) and oHI ($p < 0.01$). This is in correspondence with previous studies that have shown that hearing thresholds at the EHF range worsen with age, starting from around 31-40 years old, and increasing frequency (Wang et al., 2021). It should be remarked that there was no significant difference in PTA-EHF values between the mNH, oNH and oHI groups.

2.3. Phoneme stimuli

We generated four pairs of phoneme stimuli using the analysis–resynthesis framework of the WORLD vocoder (Morise et al., 2016), following the exact same approach described in Schirmer et al. (2024) and Dapper et al. (2025), to be used in a phoneme discrimination task and to investigate peripheral speech encoding of those phonemes through speech-EEG (Gaudrain et al., 2025). All stimuli were synthesized using a male voice with a fundamental frequency (F_0) of 116 Hz to match the OLSA sentences (German Oldenburger sentence test, (Wagener et al., 1999)), and a sampling frequency of 44.1 kHz. Stimuli were constructed such that some pairs differed only in spectral regions below the 1.5 kHz phase-locking limit (PLL), thereby allowing discrimination through both temporal fine structure (TFS) and temporal envelope (TENV) cues: /du/–/bu/ (like in "<u>doe</u>" versus "<u>boe</u>te" in Dutch, "do" and "fine" in English, respectively) and /o/–/u/ (like in "b<u>oo</u>t" versus "<u>oe</u>ster" in Dutch, "boat" and "oyster" in English, respectively). The other pairs differed only above the PLL, restricting discrimination to TENV cues: /di/–/bi/ (like in

"die" versus "bier" in Dutch, "that" and "beer" in English, respectively) and /i/–/y/ (like in "iets" versus "universiteit" in Dutch, "something" and "university" in English, respectively) (Weiss and Rose, 1988; Gaudrain et al., 2024, 2025). In both cases, steady-state vowel segments were standardized across syllable pairs, and stimuli were resynthesized with 30 ms onset/offset ramps and total stimulus durations of 170 ms for the vowel pairs and 230 ms for the syllable pairs (Schirmer et al., 2024; Dapper et al., 2025; Gaudrain et al., 2025).

Finally, all stimuli were spectrally equalized to achieve uniform signal-to-noise ratios across frequency when mixed with OLSA noise and were presented at 60 dB SPL. Calibration was performed using an IEC 60318–4 ear simulator (Schirmer et al., 2024; Dapper et al., 2025). Figure 3 shows the time domain waveforms of all stimulus pairs, from which we can derive that the consonants /d/ and /b/ of the syllable pairs occur until around 50-60 ms in time, but only after 100 ms there is no difference anymore between the vowel segments of the syllable pairs (Gaudrain et al., 2025). Figure 3 also shows the frequency spectra and Figure 4 shows the short-time fourier transform (STFT) of all stimuli. The STFT was calculated using a Hann window with a size of 500 samples, NFFT = 500 samples and hop length = 120 samples, with a sampling frequency of FS = 44.1 kHz. From Figures 3 and 4, we can derive that the contrast of the /o/–/u/ pair is in the lower harmonics of F_0 = 116 Hz, the envelope of /o/ is higher in amplitude than of /u/, but the latter has a higher amplitude of the peaks in between the envelope peaks. For the /i/–/y/ pair the contrast is between 2 - 4 kHz in the 2nd, 3rd and 4th formant frequency (as mentioned in Table 1 in Dapper et al.

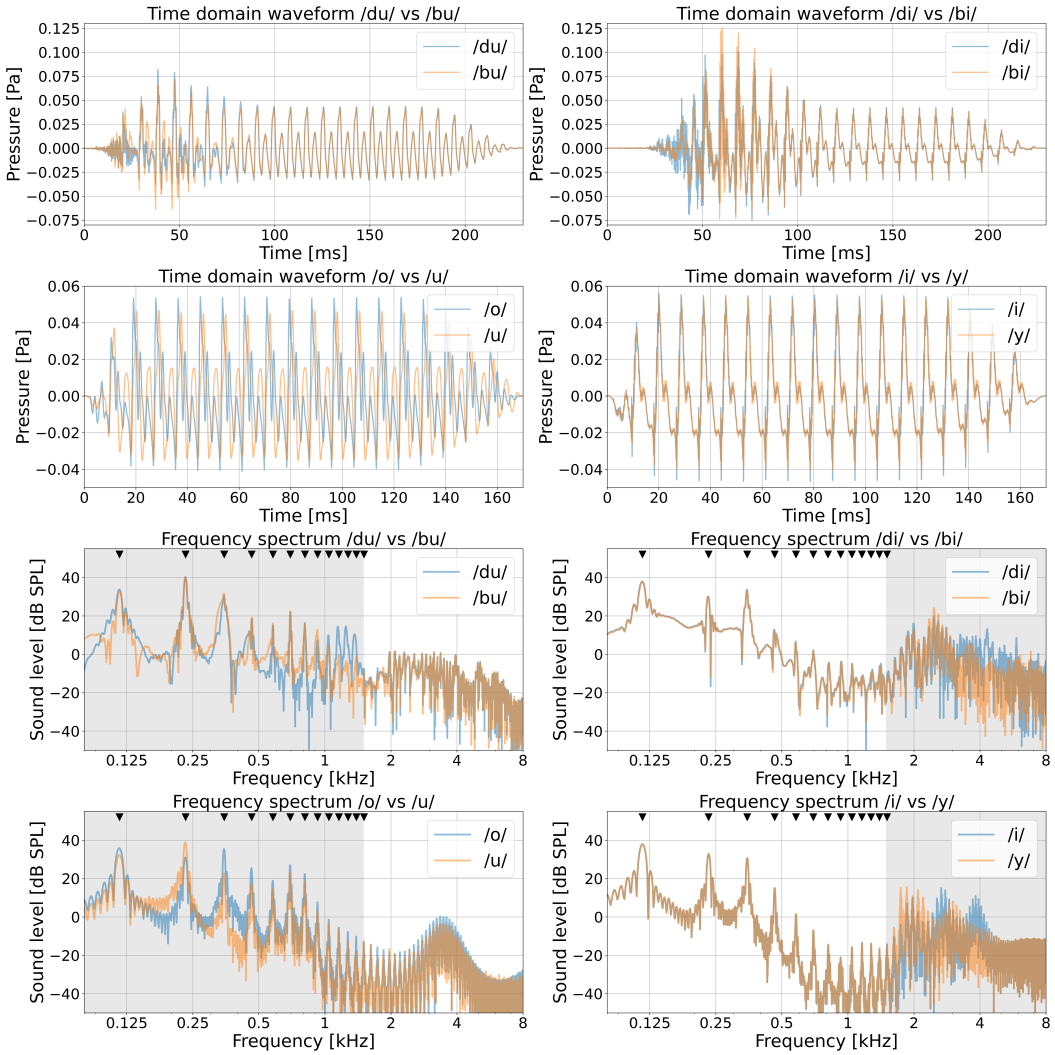


Figure 3: Time domain waveform and frequency spectra of the stimuli. The harmonics of F_0 are marked by black triangle symbols, and the grey background area marks the frequency range in which there is a contrast within the stimulus pair (below or above the phase-locking limit).

(2025)). The contrast for the /di/–/bi/ pair is in the 2nd formant around 2.4 - 2.6 kHz, and for the /du/–/bu/ pair the contrast is also in the first and

second formant around 0.8 - 1.2 kHz, as shown in Figure 1B in Gaudrain et al. (2025).

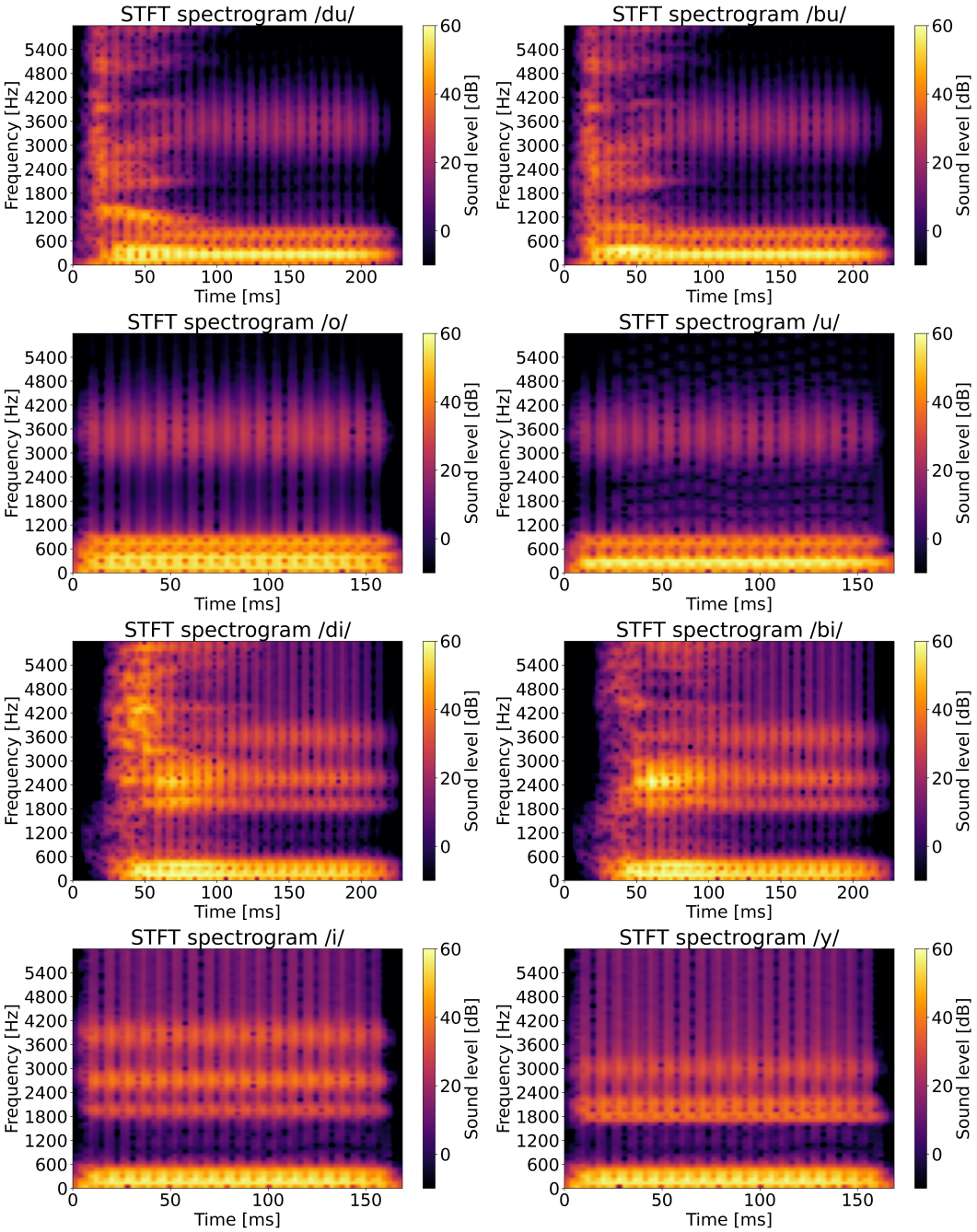


Figure 4: Short-time fourier transform of the stimuli (FS = 44.1 kHz, window shape = Hann, window size = 500 samples, NFFT = 500 samples, hop length = 120 samples).

2.4. Auditory speech-EEG recording

In the speech-EEG recording, the speech-like phoneme stimuli (/di/, /bi/, /du/, /bu/, /o/ and /y/) were presented binaurally in quiet at 60 dB SPL in an interleaved stimuli protocol, using the methodology and equipment set-up as described by Dapper et al. (2025). Each phoneme stimulus was zero-padded to an epoch length of 250 ms, jittered in time by +/- 5ms, and was presented 3456 times, with alternating positive and negative polarity epochs (to avoid the effect of the cochlear microphonic) via insert earphones (ER1-14A) and mu-metal shielded ER2 transducers (Etymotic Research, Elk Grove Village, Illinois, USA) in a subject laying on the back in a horizontal position inside an electrically and acoustically shielded booth (Dapper et al., 2025). Recordings were performed using passive stick-on electrodes (Neuroline 720, Ambu, Steinkopfstraße 4, Bad Nauheim, Germany) positioned on both mastoids, a reference electrode at the Fpz position below the hairline, and a ground electrode in between the eyebrows, recorded at a sampling rate of FS = 50 kHz for each EEG channel (Dapper et al., 2025). Due to time constraints, the vowels /u/ and /i/ were not measured during the speech-EEG session, since their responses could be extracted from the recordings of the syllables /du/ and /di/, respectively (Dapper et al., 2025).

2.5. Pre- and post-processing of EEG data

The raw data of the envelope-following-responses (EFRs) in response to the phoneme stimuli were pre-processed in multiple steps using Python (v.3.9.13) and the MNE-Python package (v.0.23.4) (Gramfort et al., 2013, 2014). First, the Fpz channel recording was re-referenced to the average of the two mastoid electrodes. Second, the triggers of the EEG epochs were

extracted to identify to which stimulus the EFR corresponds. Subsequently, artifacts were removed from the EEG data by applying a band-pass filter between 60 and 1500 Hz, and epochs of length 250 ms (-15 to 235 ms) were extracted based on the trigger events, including 1728 positive and 1728 negative polarity epochs per phoneme stimulus. Next, the epochs were baseline-corrected. Then, for each polarity the epochs with the highest peak-to-peak amplitudes, which might correspond to artifacts, were rejected, such that 1650 epochs per polarity with the lowest peak-to-peak amplitude remained. In order to obtain the time domain waveform of the EFR to each stimulus, the total of 3300 epochs were averaged per stimulus.

Frequency domain EFR analysis

For the analysis of the EEG recordings in the frequency domain, we calculated the fast fourier transform (FFT) of each of the epochs using NFFT = 50,000 bins, to get a frequency resolution of 1 Hz for the EFR spectrum ($F_{res} = FS/NFFT$). For /u/ and /i/ we cut off the first 60 ms of the EFR to /du/ and /di/ respectively, and from the EFR in response to /o/ and /y/ we cut off the last 60 ms, to have an equal FFT length in the comparison of the pairs /o/-/u/ and /i/-/y/. We used a bootstrapping procedure of random sampling with replacement, for 500 iterations, to obtain per stimulus the EFR frequency spectrum of the signal with noise. To derive the noise-floor, we performed a bootstrapping procedure for 500 iterations on the EFR spectra of each stimulus, using the phase flip method as proposed in Zhu et al. (2013). Finally, we subtracted the noisefloor spectrum from the full EFR spectrum, and averaged it over the 500 iterations, to acquire the clean noisefloor-corrected EFR spectrum per phoneme stimulus. To illustrate this,

Figure 5 shows the EFR spectrum in response to the stimulus /bi/ for one subject of the group yNH1.

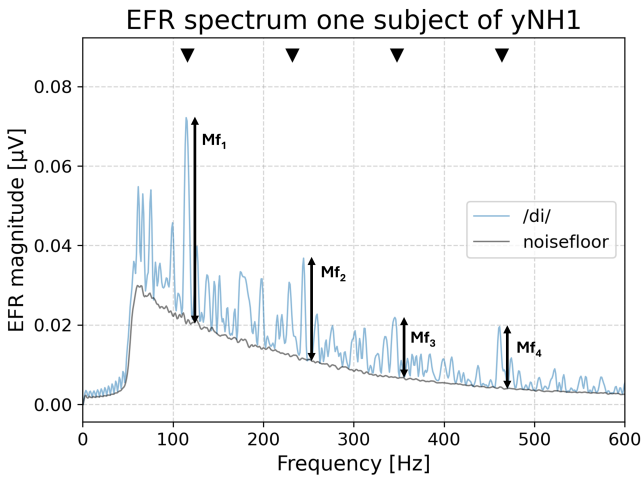


Figure 5: Envelope-following response frequency spectrum for /bi/ stimulus for one subject of the group yNH1. The harmonics of F_0 are marked by black triangle symbols.

From this noisefloor-corrected EFR spectrum, we calculate for each subject the EFR peak value as the sum of the EFR amplitude magnitudes M_{f_k} of the four harmonics of the fundamental frequency $F_0 = 116$ Hz, according to Equation 1.

$$EFR \quad harmonic \quad peak \quad magnitude = \sum_{k=1}^{4} M_{f_k}, \qquad f_k = 116 \times \mathrm{k} \qquad (1)$$

To evaluate the effects of stimulus type on EFR magnitude across different participant groups, we conducted both between-group and within-group statistical analyses in Python. For the between-group comparisons, EFR magnitudes were averaged per stimulus and assessed for normality using the

Shapiro-Wilk test. If all groups met the assumption of normality, a one-way ANOVA was performed followed by post hoc pairwise comparisons using Tukey's HSD test to identify significant differences. In cases where normality was violated, the non-parametric Kruskal-Wallis test was applied, and significant results were further explored using Dunn's post hoc test with Holm correction. For within-group comparisons, paired stimuli were analyzed using either the paired t-test or the Wilcoxon signed-rank test, depending on the normality of the difference scores. Multiple comparisons were corrected using the Benjamini-Hochberg procedure to control the false discovery rate. All statistical outcomes were visualized using boxplots with individual data points and significance markers.

For each phoneme stimulus, we additionally investigated the correlation between EFR magnitude and age. The EFR magnitude values were paired with the participant's age, and Pearson's correlation coefficient (r) was calculated to quantify the linear relationship between age and EFR magnitude. Pooled correlations were computed across all groups for each stimulus. Statistical significance was assessed using the associated p-value from the Pearson test.

Time domain EFR analysis

To assess the similarity in neural responses to different phoneme pairs, we computed normalized maximum cross-correlations between averaged EFR time domain waveforms across multiple subject groups. Higher correlation means that responses to the two stimuli are more similar, lower correlation means that responses are more distinct and hence a better discrimination. Before calculating the cross-correlation, we calculated the time shift of the

stimulus time domain waveform in response to /u/ from /du/ and /i/ from /di/, required to match the peaks in the waveforms of /o/ and /y/ respectively, and applied the same shift in the EFR time domain waveforms. For each phoneme pair, we segmented the EFR time domain waveform into consonant and vowel time windows based on a 50 kHz sampling rate. The consonant window spanned the initial 100 ms (Gaudrain et al., 2025), while vowel windows varied depending on the phoneme duration (for /du/-/bu/ and /di/-/bi/: 100–250 ms; for /o/-/u/ and /i/-/y/: 50–170 ms). The cross-correlation was calculated only within these time windows, not across the entire waveform, to isolate phoneme-specific temporal dynamics. For each subject, EFR time domain waveforms were averaged across trials, and the similarity between the specific time windows of two stimuli in each pair was quantified using the maximum value of the normalized cross-correlation function. This analysis was performed separately for each group, with subject counts adjusted accordingly, yielding correlation values for consonant and vowel segments that enabled comparative assessment across conditions.

To assess group-level differences in neural similarity of the EFR time domain waveform for each phoneme pair, we conducted separate statistical analyses for consonant and vowel segments based on cross-correlation values. For each segment type, data were grouped by phoneme pair and first tested for normality using the Shapiro-Wilk test. If all groups satisfied the assumption of normality, a one-way ANOVA was performed to evaluate overall group differences, followed by Tukey's HSD post hoc test to identify specific pairwise contrasts. In cases where normality was violated, the non-parametric Kruskal-Wallis test was applied, with post hoc comparisons conducted using

Dunn's test with Holm correction. These procedures were applied independently to both consonant and vowel data. For phoneme pairs containing only vowel segments, the same statistical workflow was applied exclusively to the vowel data. Significant pairwise differences were visualized using boxplots with annotated significance bars, where grey and black colors respectively represented consonant and vowel cross-correlation values. Scatter plots of individual subject values were overlaid to illustrate distributional variability.

2.6. Envelope-following response to a RAM pure tone stimulus

Using the same measurement setup as for the speech-EEG recordings as described in Schirmer et al. (2024), we recorded the EFR to a rectangular amplitude-modulated pure tone with a 4 kHz carrier frequency (RAM-4000) and modulation frequency of 116 Hz and modulation depth of 100 %, using 400 positive and 400 negative polarity epochs of length 500 ms (stimuli length 400 ms) at a sound level of 70 dB SPL rms and with a sampling rate of 50 kHz. The EFR to a RAM-4000 stimulus is a sensitive marker of cochlear synaptopathy, for which lower EFR magnitudes correspond to higher degrees of CS (Vasilkov et al., 2021; Keshishzadeh et al., 2021).

From the RAM-EFR recordings, the EFR peak magnitudes are calculated per subject using the Matlab (v.2022b) toolbox EEGLAB (v.2021.1). First, the EEG signals were band-pass filtered between 60 and 1000 Hz, then epoched, baseline-corrected, and the 200 epochs with the highest peak-to-peak values were rejected. The frequency spectra were calculated using Matlab's FFT function for NFFT = 50,000 bins, to obtain a frequency resolution of 1 Hz. Then, a bootstrapping procedure using the phase flip method has been executed using 400 draws to separate the noisefloor from the clean

EFR spectrum. The magnitude of the RAM-EFR of each of the 400 draws is subsequently calculated from the noisefloor-corrected spectrum as the sum of the frequency response at the first four harmonics of the fundamental frequency $F_0 = 116$ Hz, according to Equation 1, and then averaged to obtain one RAM-EFR magnitude value per subject.

To assess whether there were statistically significant differences in RAM-EFR magnitudes across subject groups, a non-parametric Kruskal-Wallis rank sum test was performed in Python. The Kruskal-Wallis test was selected due to a significant violation of the normality assumption required for parametric ANOVA, as indicated by the Shapiro-Wilk test (mNH: W = 0.8079, p = 0.0251).

2.7. Phoneme discrimination task

We employed the same phoneme discrimination paradigm as described by Schirmer et al. (2024), using a three-alternative forced-choice (3AFC) task with the four phoneme pairs as described earlier. Each pair was tested in both an easy and a difficult condition, determined based on pilot data, corresponding to a larger and smaller difference within the phoneme pair, respectively. The task was performed under three noise conditions (in quiet, ipsilateral noise, and contralateral noise), with randomized trial order and four initial training trials per condition to minimize learning effects. Stimuli were presented to the right ear and for the noise we used speech-shaped noise that was used in the OLSA test, presented at 0 dB SNR. For each stimulus pair and noise condition, phoneme discrimination scores were measured on a scale from 0 to 1. Higher scores reflect greater ability to discriminate between the two phonemes, whereas scores below 0.33 indicate performance

below chance level.

We statistically investigated the between-group differences for each stimulus pair and noise condition, and within-group differences across the noise conditions and stimulus pairs, by conducting two complementary sets of pairwise comparisons using generalized linear mixed-effects models (GLMMs) with a binomial link function in R (v.4.5.1, lme4 v.1.1.37). First, within each group, we assessed the effect of noise location (quiet, ipsilateral, contralateral) across stimulus pairs by fitting GLMMs that included noise location and stimulus pair as fixed effects, and participant nested within difficulty level as a random effect to account for repeated measures. Estimated marginal means (emmeans v.1.11.2) were computed for each noise location within each stimulus pair, and pairwise contrasts were used to determine whether specific configurations of noise significantly affected the phoneme discrimination scores. Second, to examine between-group differences, we performed pairwise comparisons of response accuracy across groups for each combination of stimulus pair and noise location. For each subset, GLMMs were fitted with group as a fixed effect and participant nested within difficulty level as a random effect. Estimated marginal means and pairwise contrasts were then used to identify significant differences in performance between groups under matched listening conditions. This dual modeling approach allowed us to disentangle the effects of noise configuration and group membership on phoneme discrimination performance.

3. Results

3.1. Envelope-following response in the frequency domain

First, we examined the RAM-EFR peak values for every subject group, shown in Figure 6, using boxplots with overlaid individual data points as solid black circles, and outliers additionally displayed as hollow circles. The Kruskal-Wallis H test revealed a marginally significant difference in EFR magnitudes across the groups (H = 10.338, p = 0.0351), but Dunn's post hoc comparisons with Holm correction did not show a significant difference between groups. This is probably because the variability within each group was substantial, reducing statistical power to detect small between-group effects. However, we noted (not significantly) smaller RAM-EFRs with increasing age.

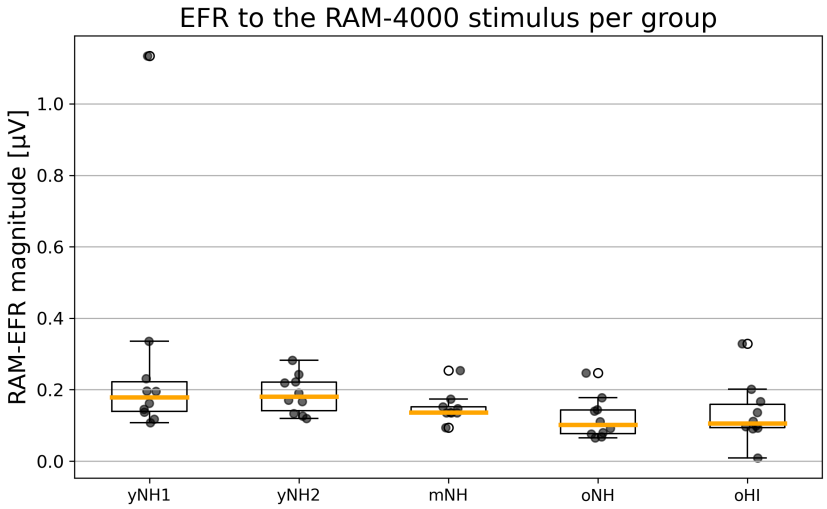


Figure 6: Envelope-following response magnitude to the RAM-4000 stimulus.

Next, we analyzed the auditory EEG recordings in response to the speech-like phoneme stimulus pairs, differing in spectral regions below (/du/-/bu/,

/o/-/u/ (from /du/), targeting both TENV and TFS contrast encoding) or above (/di/-/bi/, /i/ (from /di/) –/y/, targeting only TENV contrast encoding) the phase-locking limit of 1.5 kHz. The resulting EFR frequency spectra in response to the vowel stimulus pairs are shown per group in Appendix Figure A1, and to the syllable pairs in Appendix Figure A2. In Figure 7, statistically significant within-group differences of the EFR peak magnitude (as in Equation 1) per phoneme stimulus are shown for each group, and in Figure 8 the significant between-group differences in the EFR magnitudes are shown.

Concerning the within-group differences, for the stimulus pair /i/-/y/ the EFR peak magnitude is significantly larger in response to /y/ than to /i/ for each group, but this is only significant for the groups yNH2 ($p < 0.05$) and oHI ($p < 0.01$). In yNH1, yNH2 and mNH, the harmonics up to the fourth harmonic are clearly present in the EFR spectrum of all phoneme pairs (in the Appendix Figures A1 and A2), while the fourth harmonic disappears in oNH and the third and fourth harmonic disappear in oHI.

Concerning the between-group differences, the EFR magnitude to /i/ in yNH1 ($p < 0.01$) and yNH2 ($p < 0.05$) is significantly larger than in oHI. For the stimulus /y/, the EFR response in yNH1 is significantly larger than in oHI ($p < 0.05$). The EFR magnitude to the stimulus /di/ is significantly larger in yNH1 than in oHI ($p < 0.05$). For the stimulus /bi/, the EFR magnitude also decreased with age, but this decrease between groups was not significant.

For the stimulus pairs /o/–/u/ and /du/–/bu/, there are no significant between- or within-group differences. These results show that the stimu-

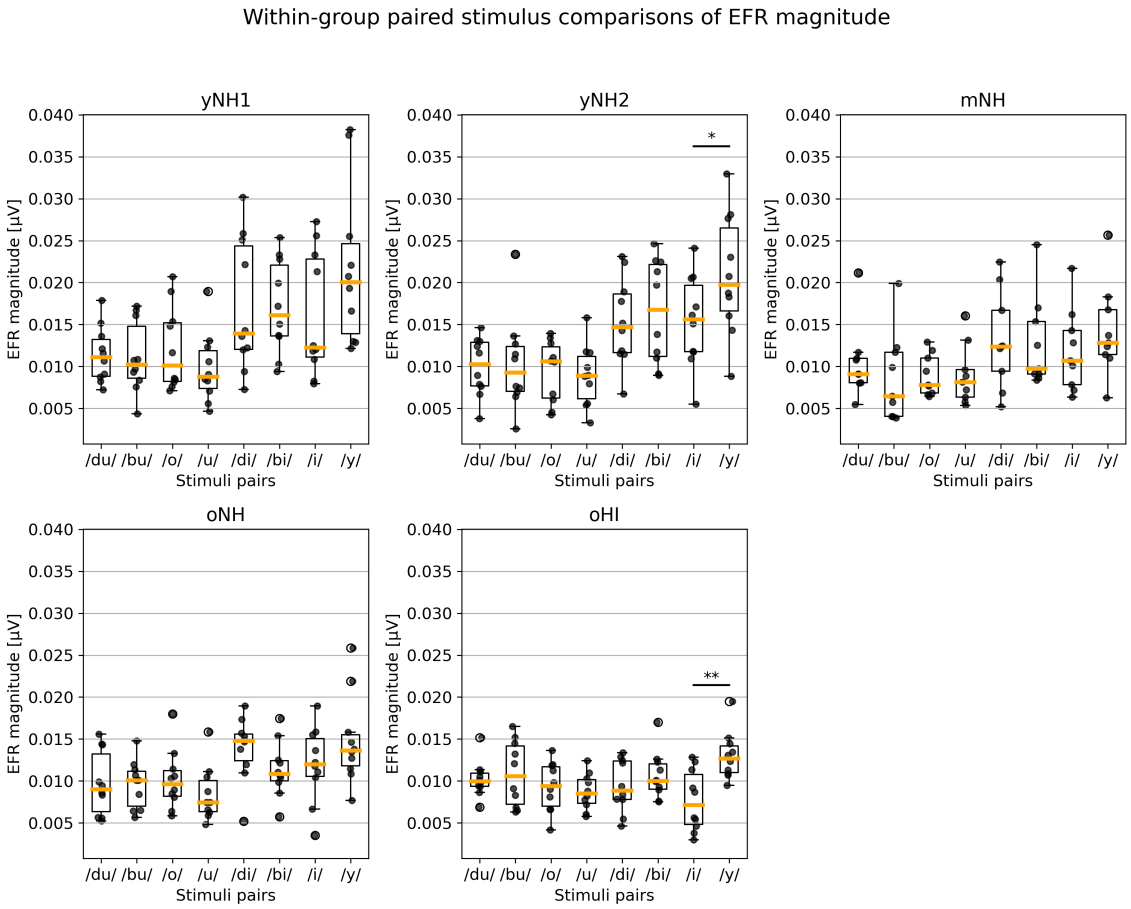


Figure 7: EFR peak magnitude within-group differences per phoneme ($p < 0.001 : ***$; $p < 0.01 : **$; $p < 0.05 : *$).

lus pairs for which the contrast is below the PLL, with most energy of the stimuli in the lower frequencies below the PLL, no significant change in the magnitude of the EFR occurs with age and the associated increase in hearing disability. The EFR predominantly reflects an envelope-following measure and does not measure TFS fidelity directly, therefore the EFR to /o/-/u/ and /du/-/bu/ remains relatively unchanged with age, since the

low-characteristic frequency (CF) auditory nerve fibers responsible for the envelope tracking are not as much affected with age than high-CF fibers. The discriminability of these phoneme contrasts as EFR magnitude at the level of the brainstem also does not change with age. Despite the differences present below the PLL in the stimulus pairs as input, the resulting EFR magnitude to these stimuli is not significantly different, but we observe that the harmonics of the fundamental frequency become less present with age, as seen in the EFR frequency spectra in Appendix Figures A1 and A2. This decrease in EFR magnitude of the harmonics represents less neural synchrony to the periodic components of the stimulus envelope at the fundamental frequency and its harmonics, which might be a consequence of a possible loss of low-spontaneous-rate auditory nerve fibers, which are crucial for encoding amplitude modulation, and reduced phase-locking (Kujawa and Liberman, 2015).

In contrast, for the stimulus pairs for which the contrast is above the PLL, and which contain most energy in the higher frequencies, we did observe between-group differences in the magnitude of the EFR for /i/, /y/ and /di/. This reflects that TENV processing by the high-CF auditory nerve fibers declines with age, as was also concluded in Dapper et al. (2025), but especially with hearing-impairment involving higher PTA4 values, since the EFR magnitudes of the young listeners only significantly differed with those of the oHI group, and not with those of the oNH group, according to Figure 8. The discriminability of the /i/–/y/ contrast as EFR magnitude at the level of the brainstem remains present across all age groups, be it only significant in groups yNH2 and oHI. For these stimulus pairs, the harmonics of the

fundamental frequency also decrease in magnitude in the frequency spectra with age, according to Appendix Figures A1 and A2.

We further investigated the effect of age by computing the pooled correlation between the EFR magnitude and age per stimulus across all groups, as shown in Appendix Figure A4. Here, we observed no significant correlation between EFR magnitude and age for the stimuli /du/, /bu/, /o/ and /u/, for which the EFR is generated by envelope-following of low-CF auditory nerve fibers, as expected from Figure 8. But for the stimuli /di/, /bi/, /i/ and /y/, which mostly contain high-CF energy, we did observe a significant negative correlation between EFR magnitude and age (/di/: $r = -0.35$, $p = 0.014$; /bi/: $r = -0.48$, $p < 0.001$; /i/: $r = -0.43$, $p = 0.002$; /y/: $r = -0.42$, $p = 0.003$), showing again that TENV processing by high-CF auditory nerve fibers significantly declines with increasing age.

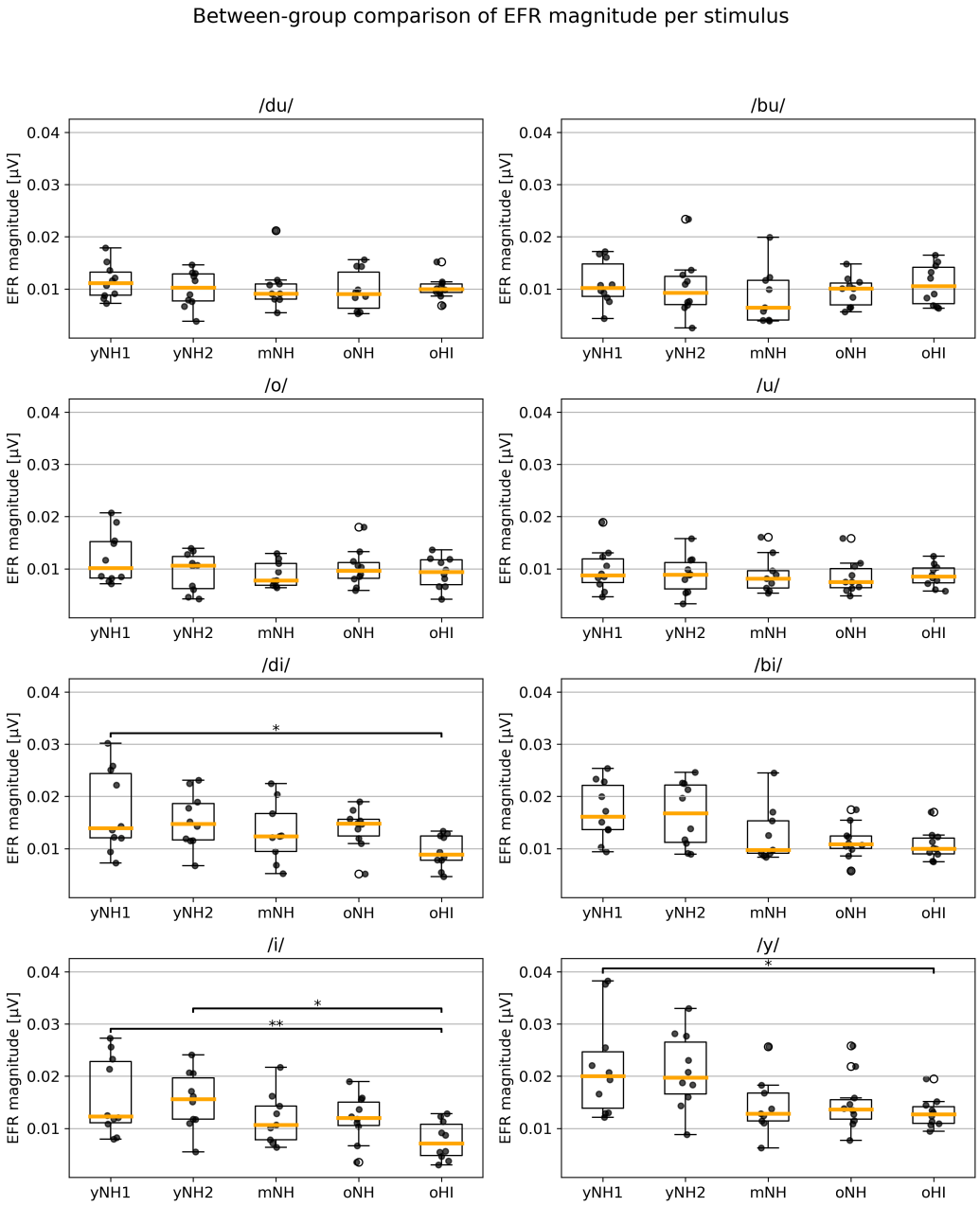


Figure 8: EFR peak magnitude between-group differences per phoneme ($p < 0.001 : ***$; $p < 0.01 : **$; $p < 0.05 : *$).

3.2. Envelope-following response in the time domain

To further investigate the differences in EFRs, we also analyzed the EFR in response to the phoneme stimuli in the time domain. Figure 9 shows the average per group of the time domain waveforms of the EFR for the syllable pairs, and in Appendix Figure A3 the time domain EFRs are shown for the vowel pairs.

The peaks observed in the EFR time domain waveform reflect neural phase-locking to the temporal envelope of the stimulus. The neural transmission delay is typically between 7 to 9 ms, due to the travel time from the ear to the midbrain (Chandrasekaran and Kraus, 2010). Synchronous firing at stimulus onset mainly depends on high-spontaneous rate, low-threshold auditory nerve fibers, which determine response latencies and perception thresholds (Rhode and Smith, 1986; Meddis, 2006; Heil et al., 2008), while the low-SR, high-threshold auditory nerve fibers contribute little to overall synchrony (Huet et al., 2019). Early peaks, occurring within the first 50 to 60 ms, primarily correspond to the consonant portion of the syllable (/d/ or /b/), which is characterized by transient acoustic bursts and onset cues. In contrast, the later, more periodic oscillations align with the vowel segment (/u/ or /i/), which has a sustained harmonic structure that drives stronger and more regular neural synchronization. Thus, the EFR waveform captures both the rapid onset features of consonants and the steady periodicity of vowels, providing insight into segment-specific auditory encoding.

To quantitatively investigate whether the differences present within the stimulus pairs can also be observed in their EFR time domain response, we performed cross-correlation on the separate consonant and vowel segments

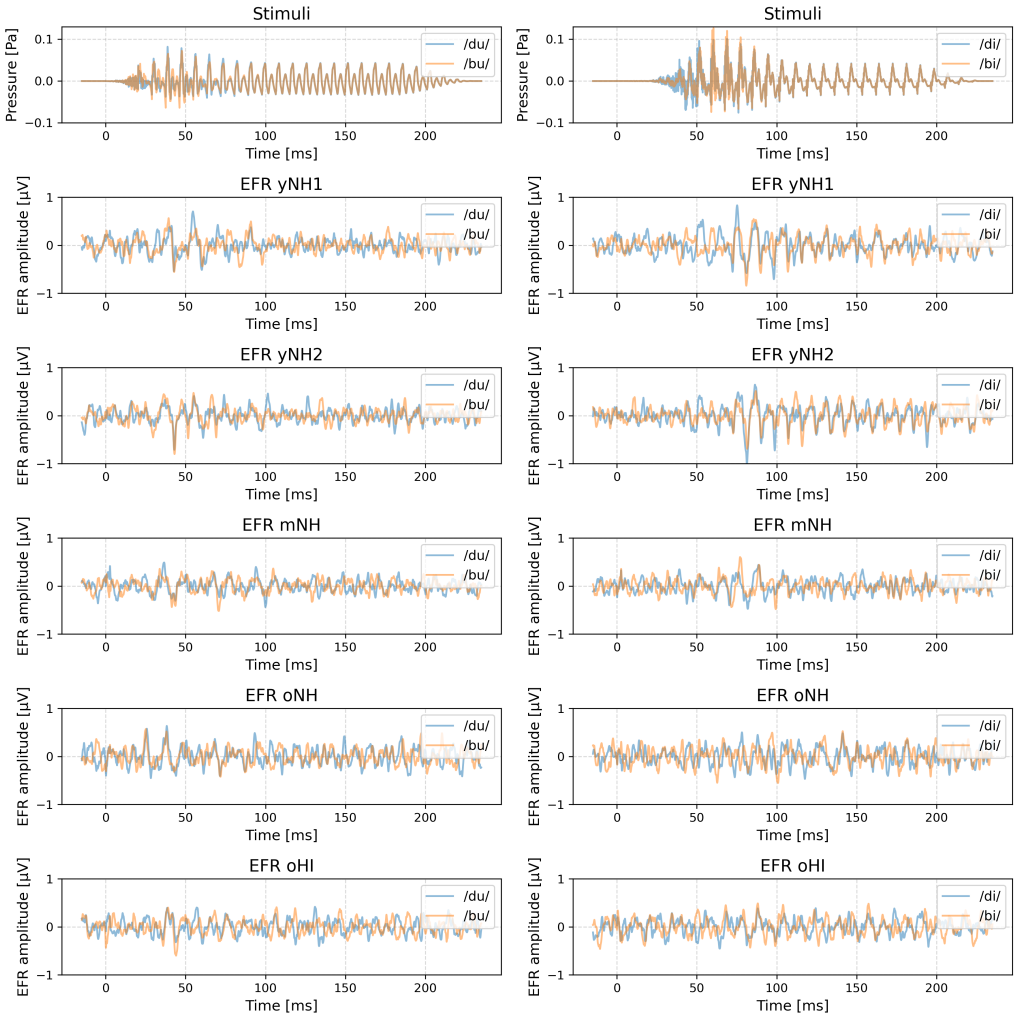


Figure 9: Time domain waveform of the envelope-following response for syllable contrast.

of the EFR time domain waveforms. Figure 10 shows the cross-correlation values for each phoneme pair per subject group. For the phoneme pair /du/-/bu/, the cross-correlation values for consonant segments showed a significant group effect. Although most groups passed the Shapiro-Wilk normality test, one group did not (oHI: $p = 0.0348$), prompting the use of the Kruskal-Wallis

test, which revealed a significant difference across groups (p = 0.0147). Post hoc Dunn's test indicated marginally significant differences between oHI and yNH2 (p = 0.0440) and between oHI and mNH (p = 0.0440), suggesting reduced neural synchrony in the oHI group. For the vowel segments of the same pair, normality was violated in oNH (p = 0.0048), and the Kruskal-Wallis test did not reach significance (p = 0.0874), indicating no strong group-level differences.

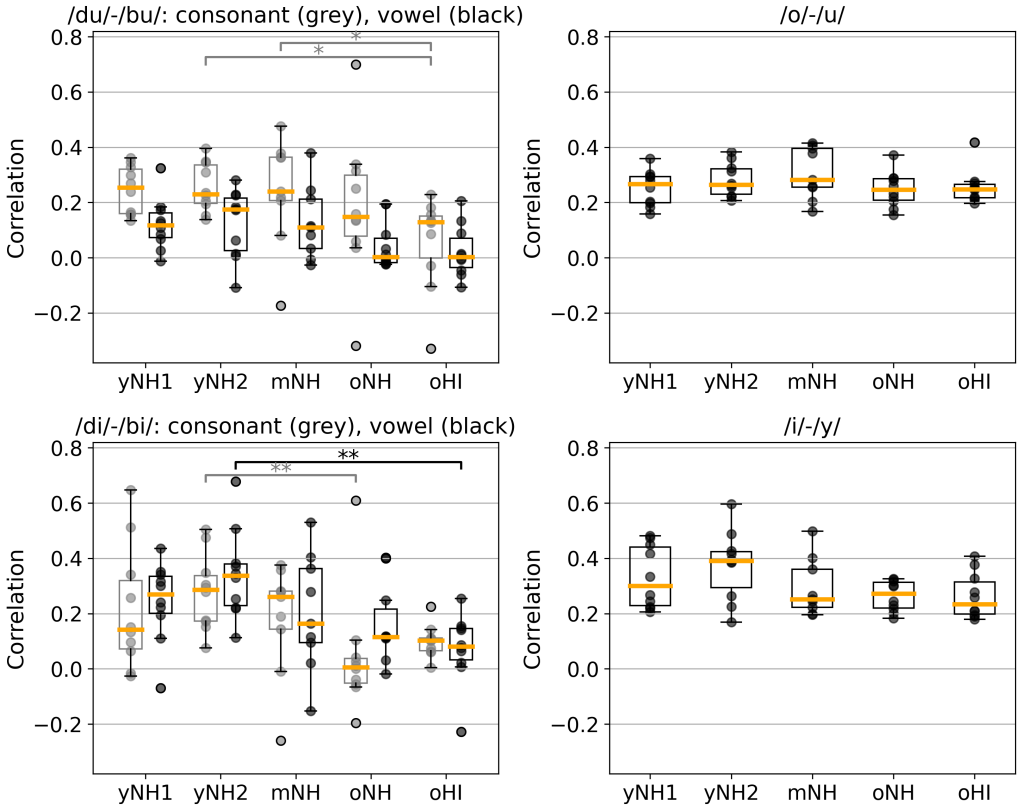


Figure 10: Cross-correlation within phoneme pairs ($p < 0.001 : ***$; $p < 0.01 : **$; $p < 0.05 : *$). Correlation values for consonant sections are shown in grey, values for vowel sections are shown in black.

In the pair /di/-/bi/, the consonant segment analysis showed a violation of normality in oNH ($p = 0.0018$), and the Kruskal-Wallis test was significant ($p = 0.0049$). Dunn's test revealed a significant difference between oNH and yNH2 ($p = 0.0036$), indicating altered consonant processing in the oNH group. For the vowel segment, normality was violated in oNH ($p = 0.0497$), and the Kruskal-Wallis test was again significant ($p = 0.0152$). Dunn's test showed a significant difference between oHI and yNH2 ($p = 0.0096$), suggesting reduced vowel similarity in the hearing-impaired group.

For the vowel-only pair /o/-/u/, all groups passed the normality test, allowing for a one-way ANOVA, which did not reveal significant differences ($p = 0.2045$), suggesting comparable vowel segment similarity across groups. Finally, for the vowel-only pair /i/-/y/, all groups passed the normality test, and the ANOVA did not reveal significant differences ($p = 0.5679$). None of the age groups showed correlation in EFR time domain waveform for the vowel segments within the stimulus pairs.

In summary, consonant segments showed that yNH listeners exhibited higher correlations than older and hearing-impaired groups, both for stimulus pairs with contrast below (/du/-/bu/) the PLL and above (/di/-/bi/). This suggests that younger listeners produce more stable and consistent neural responses across similar consonant contexts, whereas older and hearing-impaired listeners show greater variability, leading to lower correlation values. Importantly, this does not necessarily reflect better discrimination in younger listeners. It may indicate stronger encoding fidelity and less neural noise. Vowel-only pairs (/o/-/u/, /i/-/y/) showed no correlation, and also no significant group differences, confirming that age-related changes primar-

ily affect the encoding of rapid consonant cues rather than sustained vowel periodicity.

3.3. Phoneme discrimination

The phoneme discrimination scores for each subject group is shown in Figure 11 per noise condition and stimulus pair. For the stimulus pair /du/-/bu/, we can observe that the phoneme discrimination score in ipsilateral noise is significantly larger than in quiet for the groups yNH1 ($p < 0.05$), yNH2 ($p < 0.001$) and mNH ($p < 0.001$). For the oNH and oHI groups, discrimination in ipsilateral noise is also better than in quiet, but not significantly. For the groups yNH2 ($p < 0.001$) and mNH ($p < 0.001$), the score is also larger in ipsilateral noise than in contralateral noise. There were no significant between-group differences for this stimulus pair, in each noise condition, all groups performed similar for this stimulus pair. It is worth noting that in quiet and contralateral noise, discrimination was close to chance level for all groups. The results show that this stimulus contrast was the most difficult one to discriminate.

For the stimulus pair /o/-/u/, we observe that for group yNH1, the phoneme discrimination score in ipsilateral noise is significantly smaller than in quiet ($p < 0.001$) and contralateral noise ($p < 0.001$). For group yNH2, the score in ipsilateral noise is also significantly smaller than in contralateral noise ($p < 0.05$). Concerning the between-group differences, we observe in quiet that the score for yNH1 is significantly larger than in yNH2 ($p < 0.001$) and oHI ($p < 0.05$). This worsening of TFS encoding with increasing age in quiet agrees with findings in Schirmer et al. (2024). In contralateral noise, the score for yNH1 is significantly larger than in yNH2 ($p < 0.01$), oNH

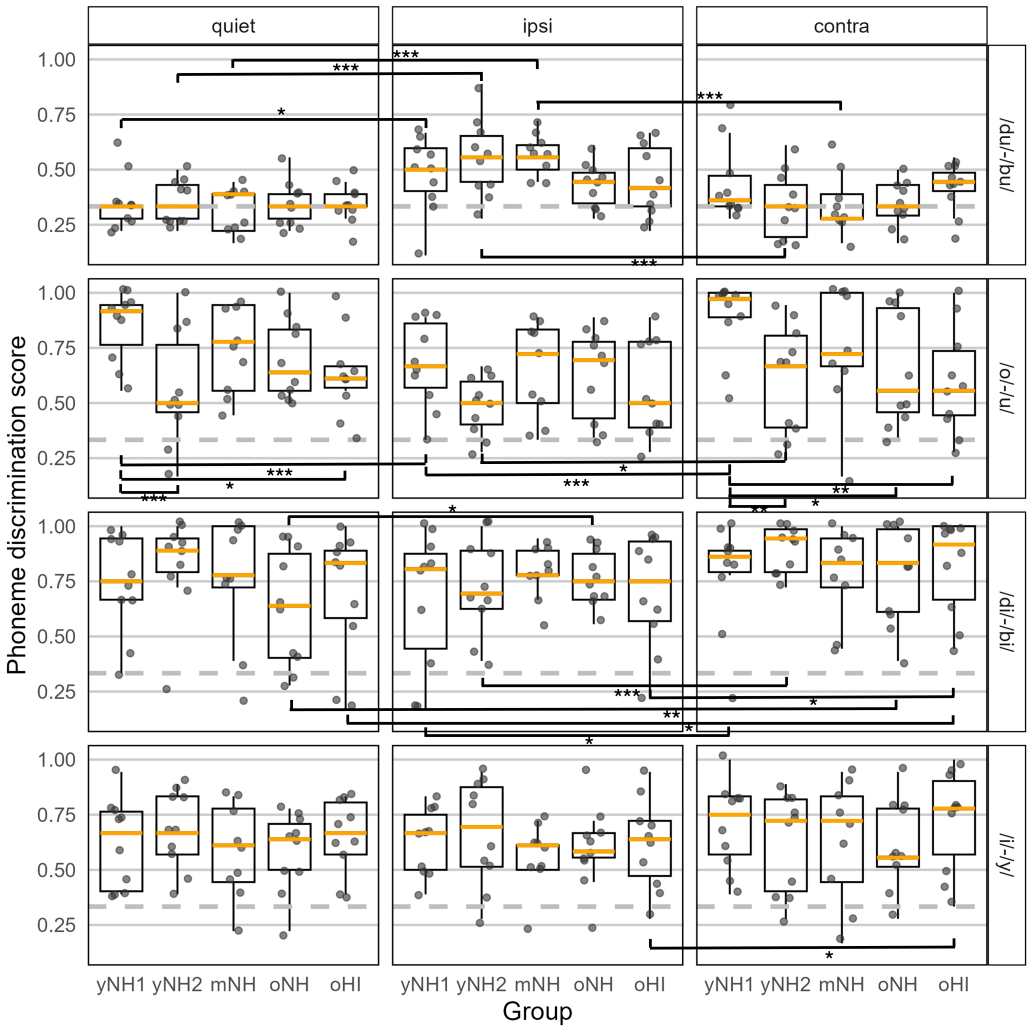


Figure 11: Phoneme discrimination scores per group, stimulus type and noise condition ($p < 0.001 : ***$; $p < 0.01 : **$; $p < 0.05 : *$).

($p < 0.05$) and oHI ($p < 0.01$). In ipsilateral noise, the phoneme discrimination performance of all groups was similar.

For the stimulus pair /di/-/bi/, all groups perform similarly per noise condition, but there are significant within-group differences. For the yNH1

group, the discrimination score in contralateral noise is significantly larger than in ipsilateral noise ($p < 0.05$), which is also the case for the yNH2 group ($p < 0.01$) and oHI group ($p < 0.05$). For the oNH group, the score in quiet is significantly smaller than in ipsilateral noise ($p < 0.05$). For the oNH ($p < 0.01$) and oHI ($p < 0.05$) groups, the score in contralateral noise is significantly larger than in quiet. We did not observe this improvement in discrimination in contralateral noise in comparison to quiet in young listeners, which might be due to the 'ceiling' effect in quiet and contralateral noise conditions for this stimulus pair.

Finally, for the stimulus pair /i/-/y/, we could only observe that the phoneme discrimination score in contralateral noise for the oHI group was significantly larger than in ipsilateral noise ($p < 0.05$), in correspondence with findings in Schirmer et al. (2024). Besides this difference, the noise condition had no influence on the score for any of the groups.

4. Discussion

The present study examined how age-related cochlear synaptopathy and outer hair cell damage affect peripheral speech encoding and phoneme discrimination. Using electrophysiological markers (RAM-EFR and speech-evoked EFRs) and behavioral discrimination tasks in several noise conditions, we characterized five participant groups differing in age and hearing status (yNH1, yNH2, mNH, oNH, oHI), in which all groups are considered NH based on the audiogram, but with higher PTA4 values in the oHI group. Our results demonstrate distinct effects of aging, CS, and OHC dysfunction on temporal envelope coding by low-CF versus high-CF auditory nerve

fibers, and on temporal fine structure coding, extending the recent findings of Schirmer et al. (2024) and Dapper et al. (2025) on cortical and midbrain correlates of degraded auditory encoding.

4.1. Age-related EHF degradation and RAM-EFR sensitivity to CS

The progressive elevation in PTA-EHF thresholds from the mNH group onward, significantly higher than in young listeners, confirms early high-frequency degradation with age, consistent with histopathological and physiological evidence that extended-high-frequency hearing is particularly vulnerable to synaptic loss and neural degeneration (Liberman et al., 2016; Bramhall et al., 2017). As also observed in Schirmer et al. (2024), the observation that RAM-EFR magnitudes, although not significantly different between groups, showed a decreasing trend with age, aligns with previous reports that RAM-EFR amplitude is a sensitive indicator of synaptopathic loss (Keshishzadeh et al., 2020, 2021; Van Der Biest et al., 2023; Vasilkov et al., 2021). The absence of strong group-level significance likely reflects inter-individual variability and partial central compensation (Auerbach et al., 2014; Johannesen and Lopez-Poveda, 2021). Together, these findings support the view that measurable CS begins around midlife, before measurable audiometric loss, and progresses alongside EHF threshold elevation.

4.2. EFR in the frequency domain: high-CF versus low-CF TENV coding

EFR magnitudes to speech phonemes, as shown in Figure 8, revealed differential vulnerability of low-CF and high-CF auditory nerve fiber coding. Stimuli containing most energy below the phase-locking limit (/du/-/bu/,

/o/-/u/), relying on low-CF fiber envelope-tracking for the EFR, showed preserved EFR magnitudes in quiet across age, whereas stimuli containing most energy above the PLL (/di/(-/bi/), /i/-/y/), relying exclusively on TENV encoding by high-CF fibers, exhibited significant group differences, particularly between yNH and oHI listeners. This decline with age in EFR magnitude in response to phonemes relying on high-CF TENV coding was also observed in Dapper et al. (2025). This pattern indicates that high-frequency TENV processing is more susceptible to age and hearing impairment than low-frequency TENV coding, confirming both physiological and modeling results showing early decline of envelope-phase locking with increased age or severity of SNHL (Henry and Heinz, 2012; Parthasarathy and Kujawa, 2018).

The finding that only the oHI group, characterized by both CS and OHC damage, showed robust reductions in EFR magnitude above the PLL suggests an additive effect of OHC dysfunction on high-CF envelope encoding. Thus, while RAM-EFR magnitudes remained relatively insensitive to OHC loss, speech-evoked EFRs targeting high-CF TENV coding appear to capture combined degradations of neural synchrony and cochlear mechanics. This interpretation is consistent with recent work demonstrating that while RAM-EFR paradigms are optimized to isolate CS and remain relatively insensitive to OHC loss, EFRs elicited by more complex or speech-like stimuli reflect broader cochlear mechanics and neural synchrony, making them vulnerable to both pathologies (Van Der Biest et al., 2023; Keshishzadeh et al., 2020; Vasilkov et al., 2021).

4.3. EFR in the time domain: neural synchrony

Analysis of the EFR time-domain waveforms showed that cross-correlations between responses to phoneme pairs were highest in yNH listeners and significantly reduced in oNH and oHI groups in consonant segments, as shown in Figure 10. These findings point to increased neural noise and less precise phase locking in older participants, consistent with reduced auditory-nerve synchrony and temporal jitter described in both animal and human aging studies (Furman et al., 2013; Sergeyenko et al., 2013). The fact that vowel-only segments showed smaller age effects, reinforces the interpretation that consonant encoding, which depends heavily on accurate envelope onset timing, is particularly vulnerable to CS (Bharadwaj et al., 2015).

4.4. Phoneme discrimination and the role of noise configuration

Behavioral phoneme discrimination revealed specific interactions between stimulus type, age, and noise configuration, as shown in Figure 11. For the /du/-/bu/ pair, which engages both TFS and TENV contrast coding, yNH and mNH listeners showed significantly better discrimination in ipsilateral noise than in quiet or contralateral noise. This significant effect was absent in the older groups (oNH and oHI), suggesting that intact stimulus onset (since the consonant discrimination is in the onset) synchrony in younger listeners may facilitate envelope-based segregation when target and masker are co-located (Parthasarathy and Kujawa, 2018; Plack et al., 2014). However, the absence of a significant effect in the older groups might be due to their performance being at chance level, which limits the ability to detect group differences. Higher participant numbers may bring the discrimination in

ipsilateral noise for the /du/-/bu/ pair to a significance between the young and old groups.

For the /o/-/u/ pair, which also targets both TFS and TENV contrast coding but without consonant segment, the pattern was reversed: discrimination in ipsilateral noise was worse than in quiet and contralateral noise for yNH listeners, whereas older groups showed no such significant effect by noise condition. This may reflect that younger listeners rely strongly on low-frequency TFS cues, which are more easily masked when noise is presented to the same ear, whereas older listeners—already affected by CS—have reduced access to those fine-structure cues and thus show less change across conditions. Significant between-group differences were present for the discrimination of the /o/-/u/ stimulus pair. In quiet, yNH groups performed significantly better than oHI listeners, but not than oNH, indicating that the combined presence of CS and OHC loss in oHI leads to worse TFS coding. In contralateral noise, yNH groups performed significantly better than all older listeners, indicating that age in general affects discrimination in contralateral noise for TFS cues.

For the /di/-/bi/ stimulus pair targeting primarily TENV processing, older groups exhibited poorer consonant discrimination in quiet and ipsilateral noise compared to contralateral noise. For the /i/-/y/ vowel discrimination targeting TENV coding, we observed worsening of discrimination score in ipsilateral noise compared to contralateral noise for oHI participants. This agrees with the expectation that speech-in-noise understanding depends critically on intact TENV encoding and that degradation of high-CF envelope coding due to CS disproportionately affects performance in conditions where

peripheral channels are challenged by energetic masking (Henry and Heinz, 2012; Oxenham, 2016). This better performance in contralateral noise might be a consequence of better concentration to the ear in which the stimulus pairs were presented. We did not observe this improvement in discrimination in contralateral noise in comparison to quiet in young listeners, which might be due to the 'ceiling' effect in quiet and contralateral noise conditions for this stimulus pair.

4.5. Relation of phoneme discrimination and EFR magnitude

The observed effects of age and hearing impairment on the EFR magnitude and performance of the phoneme discrimination task remain puzzling. On one hand, stimuli that rely on temporal envelope coding (/di/, /i/-/y/) show significantly reduced EFR representation in older hearing-impaired listeners (Figure 8). On the other hand, this neural decline does not translate into significantly poorer behavioral discrimination, as performance is similar across age groups (Figure 11). These findings raise the possibility that listeners compensate for degraded envelope encoding through alternative acoustic cues or central mechanisms (Auerbach et al., 2014; Pichora-Fuller et al., 2016; Peelle, 2018; Goossens et al., 2018; Parthasarathy et al., 2019). Furthermore, EFRs targeting TFS contrast coding (/o/-/u/) remain stable across age and hearing status (Figure 8), yet psychophysical performance changes significantly in quiet and contralateral noise (Figure 11).

The /i/-/y/ stimulus contrast concerns high-frequency formant frequencies (2-4 kHz) which excite different tonotopic regions along the basilar membrane. Behavioral /i/-/y/ discrimination therefore uses place-based spectral cues along the basilar membrane, which are more robust against aging than

TFS-based cues. Based on the age-related high-CF fibers decline we observed in Figure 2b, we expect that the envelope tracking ability in response to /i/-/y/ declines, which could corroborate the observed decline of the EFR measured by EEG. The /o/-/u/ stimulus contrast concerns the lower harmonics of F_0 (116 Hz). Discrimination of this low-frequency contrast in sound perception relies more on low-frequency TFS coding (precise carrier-phase timing below the phase locking limit), which we expect to degrade with age (Ponsot et al., 2025; Moore et al., 2012) and may be related to degraded behavioral discrimination. However, the EFR predominantly reflects an envelope-following measure and does not measure TFS fidelity directly, therefore the /o/-/u/ EFR remains relatively unchanged with age, since the low-CF fibers responsible for the envelope tracking are not as much affected with age than high-CF fibers, despite TFS-based perceptual decline. This disconnect between TFS and TENV processing in the EFR might explain why for the /o/-/u/ contrast, the EFR responses remain unaffected by age while the discrimination perception does show an effect of age and SNHL on TFS discrimination. Together, the place-based spectral coding for /i/-/y/ and TFS reliance for /o/-/u/ for discrimination, plus the TENV-dominant nature of the EFR, might offer an explanation as to why /i/-/y/ shows reduced EFR without behavioral loss, whereas /o/-/u/ shows stable EFR with behavioral decline.

The dissociation between neural and behavioral measures observed here closely parallels the contradiction reported by Ponsot et al. (2025), who demonstrated robust age- and hearing-loss–related reductions in neural phase locking to both TENV and TFS components, yet found no relationship be-

tween these electrophysiological markers and psychophysical performance. Furthermore, Goossens et al. (2018) demonstrated that stronger neural envelope encoding, measured at both subcortical and cortical levels, can paradoxically predict poorer speech perception in noise in both normal-hearing and hearing-impaired adults, highlighting that enhanced or degraded neural synchrony does not necessarily reflect functional benefit.

The mismatch between electrophysiological and psychophysical outcomes suggests complex interactions between peripheral encoding and perceptual strategies (Parthasarathy and Kujawa, 2018; Plack et al., 2014). These unresolved discrepancies highlight the need for further investigation into how neural coding deficits relate to perceptual outcomes, and whether redundancy of cues (Plack et al., 2014) or top-down processes mitigate the impact of degraded TENV or TFS encoding (Pichora-Fuller et al., 2016).

4.6. Limitations and future directions

While the present study provides valuable insights into how aging, CS and OHC damage jointly shape peripheral speech encoding, several limitations should be acknowledged. The group sizes were modest, which may have limited the statistical power to detect subtle between-group differences in electrophysiological measures such as the RAM-EFR. Interindividual variability in EFR magnitude, likely influenced by anatomical and recording factors, may also have masked smaller effects of CS.

It is also worth noting that the behavioral phoneme discrimination performance was at chance for most participants in some conditions (/du/-/bu/), or had a 'ceiling' effect (/di/-/bi/) which might have masked a more complex picture. Further studies would likely yield richer results by using more stimuli

that are more discriminable in silence. Furthermore, while RAM-EFRs and speech-evoked EFRs offer complementary insights into subcortical temporal processing, they cannot fully isolate peripheral from central contributions. This might have to be addressed in the future. In addition, future work should further investigate how neural coding deficits relate to perceptual outcomes.

5. Conclusion

This study investigated how aging and hearing impairment affect peripheral speech encoding and phoneme discrimination, with a particular focus on cochlear synaptopathy and outer hair cell damage, across five listener groups differing in age and hearing impairment. Although RAM-EFR magnitudes did not differ significantly between groups, they showed a gradual decline with age, suggesting progressive synaptic loss beginning in midlife, before measurable audiometric deterioration. This trend was accompanied by a worsening of extended high-frequency audiometric thresholds, consistent with previous evidence that EHF regions are particularly sensitive to synaptopathy and neural degeneration (Bramhall et al., 2017; Lobarinas et al., 2013; Liberman et al., 2016).

Speech-evoked EFRs provided further insight into the specific coding mechanisms affected. Phoneme stimuli containing most energy below the PLL, targeting low-CF TENV coding, showed largely preserved neural encoding across age groups. In contrast, stimuli containing most energy above the PLL, targeting high-CF TENV coding, exhibited reduced EFR magnitudes with age, especially in the older hearing-impaired group. These results

suggest that high-CF envelope encoding is more vulnerable to the combined effects of CS and OHC damage, whereas low-CF envelope coding remains comparatively robust. This pattern aligns with findings from Dapper et al. (2025) and Schirmer et al. (2024), who observed that age-related reductions in cochlear synchrony and midbrain onset coding primarily impact temporal envelope contrast representations.

Analysis of the EFR time-domain waveforms supported this interpretation: cross-correlations between responses to phoneme pairs were highest in young listeners and declined with age, particularly during consonant segments. This reduction in encoding fidelity indicates increased neural noise and degraded phase locking, consistent with known physiological markers of synaptopathy (Furman et al., 2013; Sergeyenko et al., 2013).

Behaviorally, phoneme discrimination did not fully mirror the electrophysiological findings, revealing a complex relationship between neural coding and perception (Parthasarathy and Kujawa, 2018; Plack et al., 2014; Ponsot et al., 2025; Goossens et al., 2018). While EFRs to phonemes targeting high-CF fiber envelope coding declined significantly with age, EFRs to phonemes targeting low-CF fiber envelope coding remained relatively stable. However, phoneme discrimination involving TENV-only contrasts did not show significant effects of age and hearing impairment in any of the noise conditions, while discrimination of TFS contrasts showed behavioral decline with age. This pattern can be understood by noting that high-frequency contrasts are discriminated primarily through place-based spectral cues, which remain relatively robust with age, whereas low-frequency contrasts rely on TFS coding, a mechanism known to degrade with aging. At the same time,

the EFR predominantly reflects TENV coding, and because age-related neural decline is most pronounced in high-CF regions, EFRs to high-frequency contrasts show the largest age-related reductions even though their perceptual discrimination remains relatively preserved.

Taken together, these results demonstrate that age-related synaptopathy emerges before measurable audiometric loss, progressively degrading high-CF TENV encoding at the brainstem level while leaving low-CF TENV encoding relatively intact. Behavioral discrimination of contrasts above the PLL remained stable, likely because the high-frequency contrast in the formants uses place-based spectral cues, while discrimination of contrasts above the PLL, targeting TFS coding, worsened with age. These findings underscore the complex interplay between peripheral neural synchrony, central compensation, and perceptual strategies (Parthasarathy and Kujawa, 2018; Plack et al., 2014; Ponsot et al., 2025; Goossens et al., 2018). Integrating electrophysiological measures such as RAM- and speech-evoked EFRs with behavioral phoneme discrimination thus provides a powerful framework to identify and characterize subclinical hearing deficits. Future work should further investigate the link between neural coding deficits and perceptual outcomes. These complementary approaches might offer sensitive tools for early diagnosis, monitoring neural decline, and tailoring interventions to preserve communication ability in aging and hearing-impaired populations.

Author declarations

Author contributions

Conceptualization: S.V., E.G., D.B., L.R. and M.K.; Data curation: J.S. and K.D.; Methodology: M.W., K.D., J.S., E.G., D.B. and S.V.; Formal Analysis: M.W., K.D. and J.S.; Writing - original draft: M.W.; Writing - review and editing: M.W., L.R., M.K., E.G., D.B. and S.V.; Supervision: S.V. All authors have read and agreed to the published version of the manuscript.

Conflict of interest

The authors have no conflicts of interest to declare. All co-authors have seen and agreed with the manuscript's contents, and there is no financial interest to report.

Data availability

The raw data supporting the conclusions of this article will be made available by the authors on request.

Acknowledgments

This work was supported by the Deutsche Forschungsgemeinschaft DFG KN 316/13–1, DFG RU 713/6–1, KL 1093/12–1; ERA-NET NEURON JTC 2020: BMBF 01EW2102 CoSySpeech and FWO G0H6420N; IZKF Promotionskolleg of the Faculty of Medicine, University Hospital of Tübingen. VICI Grant (Grant No. 918–17–603), Netherlands Organization for Scientific Research (NWO) and the Netherlands Organization for Health Research and

Development (ZonMw). The funding sources are not involved in study design; data collection, analysis and interpretation, or writing and submitting the manuscript.

Appendix

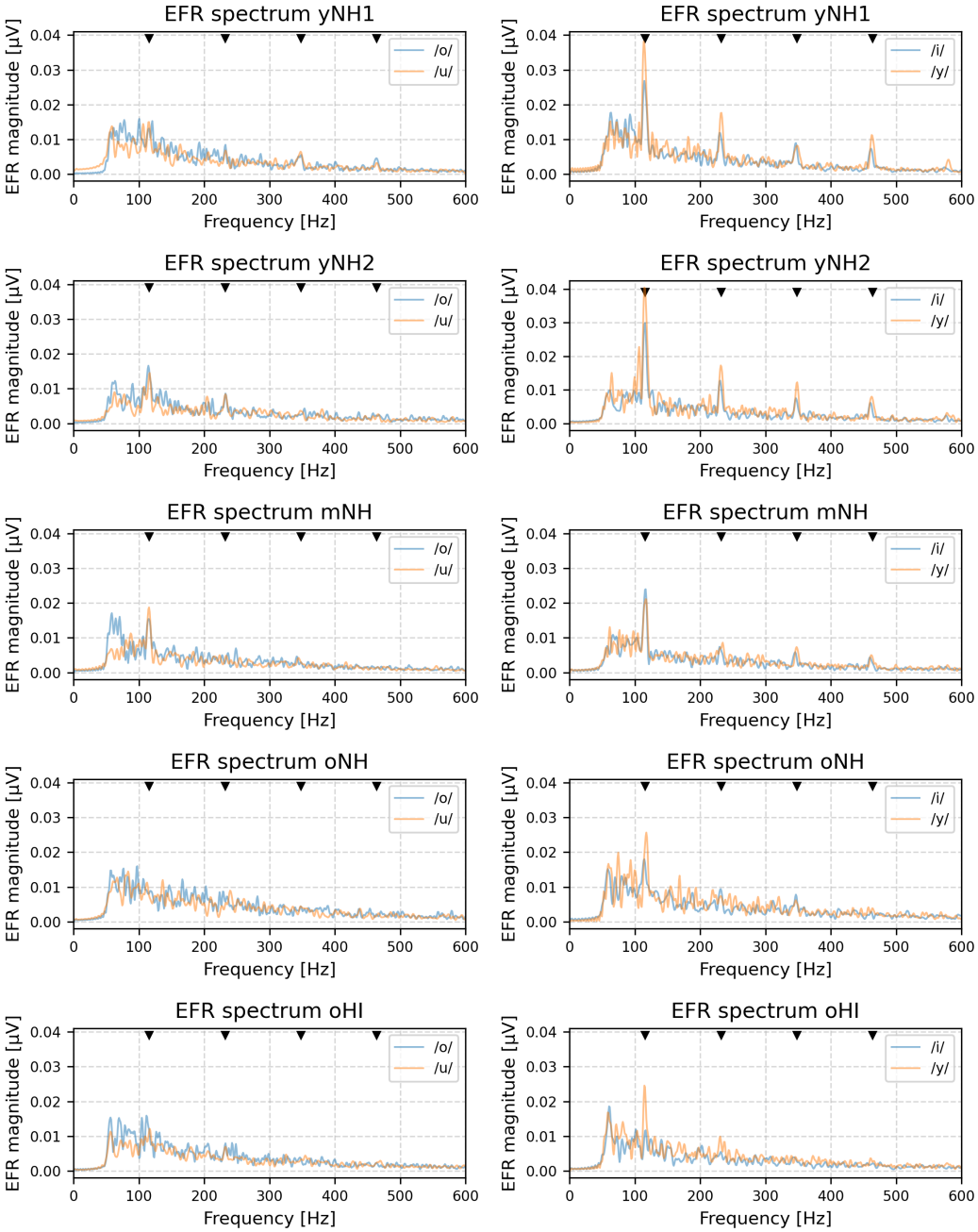


Figure A1: Envelope-following response frequency spectrum for vowel contrast. The harmonics of F_0 are marked by black triangle symbols.

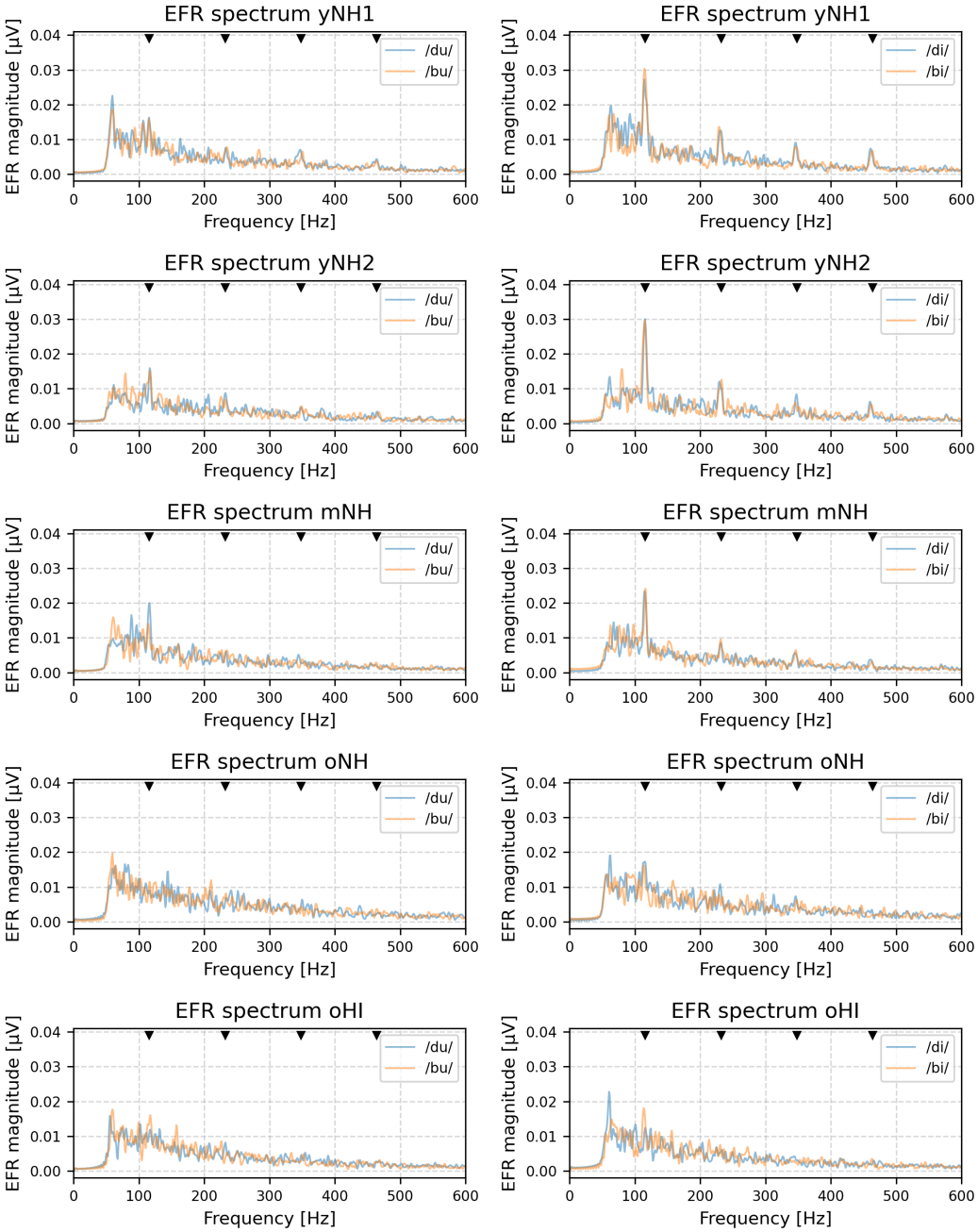


Figure A2: Envelope-following response frequency spectrum for syllable contrast. The harmonics of F_0 are marked by black triangle symbols.

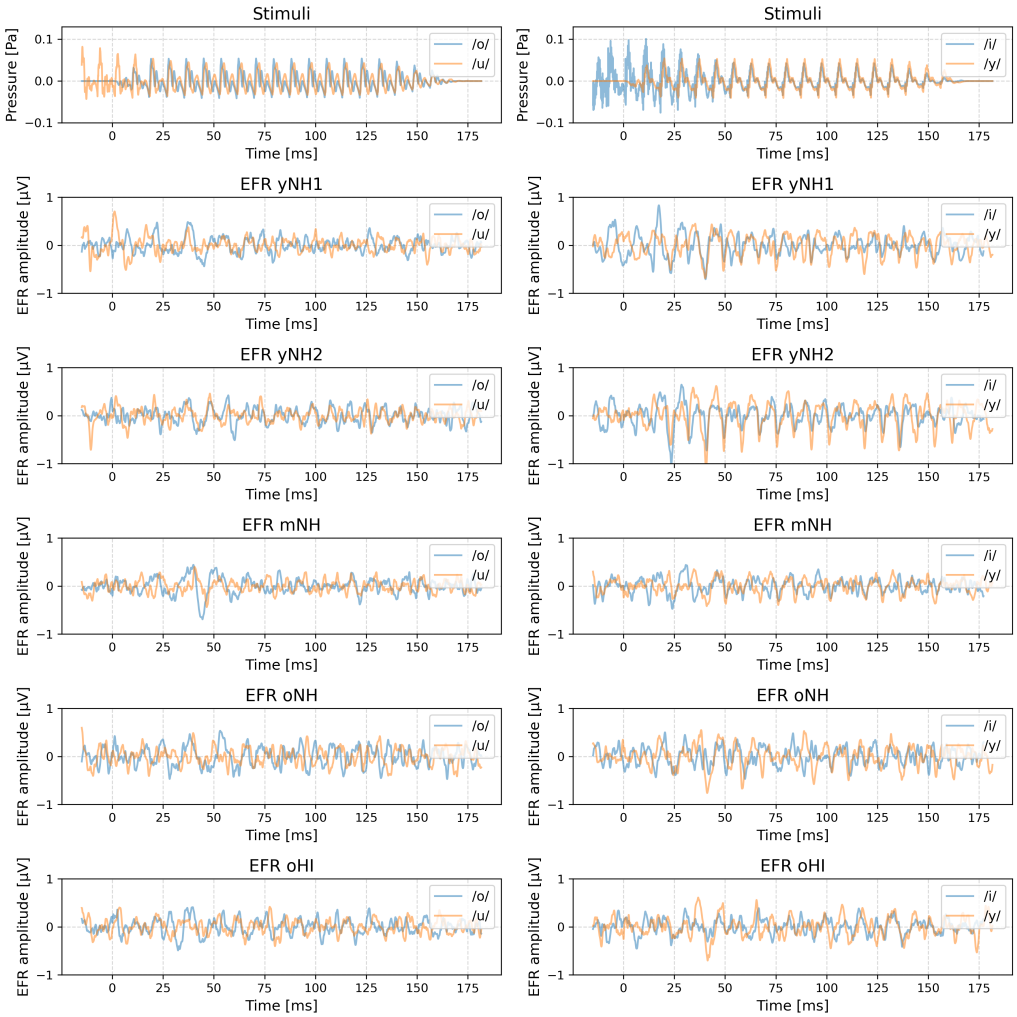


Figure A3: Time domain waveform of the envelope-following response for vowel contrast.

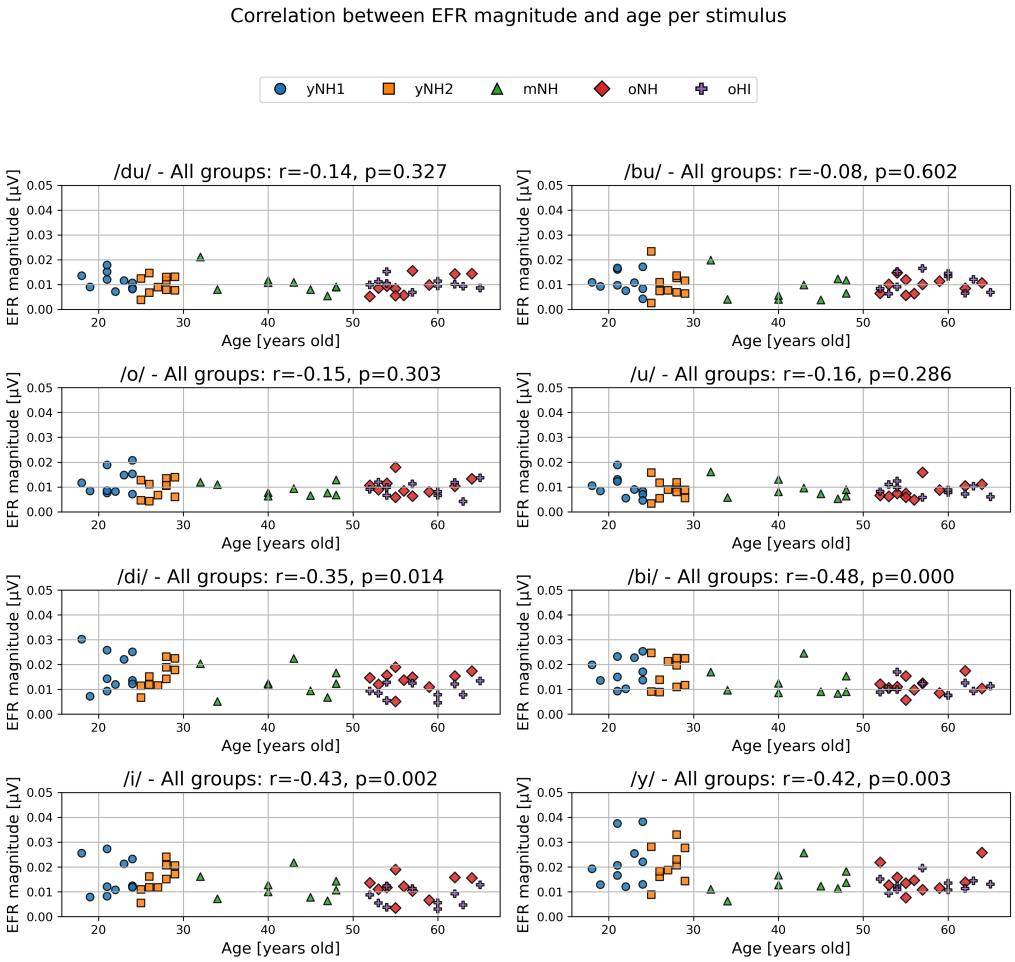


Figure A4: Correlation between EFR magnitude and age per stimulus. Pooled Pearson's correlation coefficients (r) were calculated across all groups, with associated p-values.

References

Auerbach, B.D., Rodrigues, P.V., Salvi, R.J., 2014. Central gain control in tinnitus and hyperacusis. Frontiers in Neurology 5, 206. doi:10.3389/fneur.2014.00206.

Bharadwaj, H.M., Masud, S., Mehraei, G., Verhulst, S., Shinn-Cunningham, B.G., 2015. Individual differences reveal correlates of hidden hearing deficits. Journal of Neuroscience 35, 2161–2172. doi:10.1523/JNEUROSCI.3915-14.2015.

Bramhall, N.F., Konrad-Martin, D., McMillan, G.P., Griest, S.E., 2017. Auditory brainstem response altered in humans with noise exposure despite normal outer hair cell function. Ear and Hearing 38, e1–e12. doi:10.1097/AUD.0000000000000370.

Chandrasekaran, B., Kraus, N., 2010. The scalp-recorded brainstem response to speech: Neural origins and plasticity. Psychophysiology 47, 236–246. doi:10.1111/j.1469-8986.2009.00928.x.

Dapper, K., Wolpert, S.M., Schirmer, J., Fink, S., Gaudrain, E., Başkent, D., Singer, W., Verhulst, S., Braun, C., Dalhoff, E., Rüttiger, L., Munk, M.H.J., Knipper, M., 2025. Age-dependent deficits in speech recognition in quiet and noise are reflected in mgb activity and cochlear onset coding. NeuroImage 305, 120958. doi:10.1016/j.neuroimage.2024.120958.

Furman, A.C., Kujawa, S.G., Liberman, M.C., 2013. Noise-induced cochlear neuropathy is selective for fibers with low spontaneous rates. Journal of Neurophysiology 110, 577–586. doi:10.1152/jn.00164.2013.

Gaudrain, E., Başkent, D., Verhulst, S., 2024. Cosyspeech: Material for vowel and consonant contrasts targeting specific peripheral coding mechanisms. Zenodo .

Gaudrain, E., Verhulst, S., Başkent, D., 2025. Speech stimulus design to study the neural coding of speech and the impact of cochlear synaptopathy, in: Interspeech 2025, pp. 1283–1287. doi:10.21437/Interspeech.2025-2199.

Goossens, T., Vercammen, C., Wouters, J., van Wieringen, A., 2018. Neural envelope encoding predicts speech perception performance for normal-hearing and hearing-impaired adults. Hearing Research 370, 189–200. doi:10.1016/j.heares.2018.07.012.

Gramfort, A., Luessi, M., Larson, E., Engemann, D.A., Strohmeier, D., Brodbeck, C., Goj, R., Jas, M., Brooks, T., Parkkonen, L., Hämäläinen, M., 2013. MEG and EEG data analysis with MNE-Python. Frontiers in Neuroscience doi:10.3389/fnins.2013.00267.

Gramfort, A., Luessi, M., Larson, E., Engemann, D.A., Strohmeier, D., Brodbeck, C., Parkkonen, L., Hämäläinen, M.S., 2014. MNE software for processing MEG and EEG data. NeuroImage 86, 446–460. doi:10.1016/j.neuroimage.2013.10.027.

Heil, P., Neubauer, H., Brown, M., Irvine, D.R., 2008. Towards a unifying basis of auditory thresholds: Distributions of the first-spike latencies of auditory-nerve fibers. Hearing Research 238, 25–38. doi:10.1016/j.heares.2007.09.014.

Henry, K.S., Heinz, M.G., 2012. Diminished temporal coding with sensorineural hearing loss emerges in background noise. Nature Neuroscience 15, 1362–1364. doi:10.1038/nn.3216.

Huet, A., Batrel, C., Wang, J., Desmadryl, G., Nouvian, R., Puel, J.L., Bourien, J., 2019. Sound coding in the auditory nerve: From single fiber activity to cochlear mass potentials in gerbils. Neuroscience 407, 83–92. doi:10.1016/j.neuroscience.2018.10.010.

Humes, L., 2018. The world health organization's hearing-impairment grading system: an evaluation for unaided communication in age-related hearing loss. International Journal of Audiology 58, 1–9. doi:10.1080/14992027.2018.1518598.

Johannesen, P.T., Lopez-Poveda, E.A., 2021. Age-related central gain compensation for reduced auditory nerve output for people with normal audiograms, with and without tinnitus. iScience 24. doi:10.1016/j.isci.2021.102658.

Keshishzadeh, S., Garrett, M., Vasilkov, V., Verhulst, S., 2020. The derived-band envelope following response and its sensitivity to sensorineural hearing deficits. Hearing Research 392. doi:10.1016/j.heares.2020.107979.

Keshishzadeh, S., Garrett, M., Verhulst, S., 2021. Towards personalized auditory models: Predicting individual sensorineural hearing-loss profiles from recorded human auditory physiology. Trends in Hearing 25. doi:10.1177/2331216520988406.

Kujawa, S.G., Liberman, M.C., 2009. Adding insult to injury: cochlear nerve degeneration after "temporary" noise-induced hearing loss. Journal of Neuroscience 29, 14077–14085. doi:10.1523/JNEUROSCI.2845-09.2009.

Kujawa, S.G., Liberman, M.C., 2015. Synaptopathy in the noise-exposed and aging cochlea: Primary neural degeneration in acquired sensorineural hearing loss. Hearing Research 330, 191–199. doi:10.1016/j.heares.2015.02.009.

Liberman, L.D., Suzuki, J., Liberman, M.C., 2015. Dynamics of cochlear synaptopathy after acoustic overexposure. Journal of the Association for Research in Otolaryngology 16, 205–219. doi:10.1007/s10162-015-0510-3.

Liberman, M.C., 2017. Noise-induced and age-related hearing loss: New perspectives and potential therapies. F1000Research 6, 927. doi:10.12688/f1000research.11310.1.

Liberman, M.C., Epstein, M.J., Cleveland, S.S., Wang, H., Maison, S.F., 2016. Toward a differential diagnosis of hidden hearing loss in humans. PLOS ONE 11, e0162726. doi:10.1371/journal.pone.0162726.

Lobarinas, E., Salvi, R., Ding, D., 2013. Insensitivity of the audiogram to carboplatin induced inner hair cell loss in chinchillas. Hearing Research 302, 113–120. doi:10.1016/j.heares.2013.03.012.

Lobarinas, E., Spankovich, C., Prell, C.G.L., 2017. Evidence of "hidden hearing loss" following noise exposures that produce robust TTS and ABR

wave-I amplitude reductions. Hearing Research 349, 155–163. doi:10.1016/j.heares.2016.12.009.

Lopez-Poveda, E.A., Barrios, P., 2013. Perception of stochastically undersampled sound waveforms: A model of auditory deafferentation. Frontiers in Neuroscience 7, 124. doi:10.3389/fnins.2013.00124.

Lorenzi, C., Gilbert, G., Carn, H., Garnier, S., Moore, B.C.J., 2006. Speech perception problems of the hearing impaired reflect inability to use temporal fine structure. Proceedings of the National Academy of Sciences 103, 18866–18869. doi:10.1073/pnas.0607364103.

Meddis, R., 2006. Auditory-nerve first-spike latency and auditory absolute threshold: A computer model. Journal of the Acoustical Society of America 119, 406–417. doi:10.1121/1.2139628.

Möhrle, D., Hofmeier, B., Amend, M., Wolpert, S., Ni, K., Bing, D., Klose, U., Pichler, B., Knipper, M., Rüttiger, L., 2019. Enhanced central neural gain compensates acoustic trauma-induced cochlear impairment, but unlikely correlates with tinnitus and hyperacusis. Neuroscience 407, 146–169. doi:10.1016/j.neuroscience.2018.12.038.

Möhrle, D., Ni, K., Varakina, K., Bing, D., Lee, S.C., Zimmermann, U., Knipper, M., Rüttiger, L., 2016. Loss of auditory sensitivity from inner hair cell synaptopathy can be centrally compensated in the young but not old brain. Neurobiology of Aging 44, 173–184. doi:10.1016/j.neurobiolaging.2016.05.001.

Moore, B.C.J., 2007. Cochlear Hearing Loss: Physiological, Psychological and Technical Issues. 2 ed., Wiley.

Moore, B.C.J., Vickers, D.A., Mehta, A., 2012. The effects of age on temporal fine structure sensitivity in monaural and binaural conditions. International Journal of Audiology 51, 715–721. doi:10.3109/14992027.2012.690079.

Morise, M., Yokomori, F., Ozawa, K., 2016. WORLD: A Vocoder-Based High-Quality Speech Synthesis System for Real-Time Applications. IEICE Transactions on Information and Systems E99.D, 1877–1884. doi:10.1587/transinf.2015EDP7457.

Oxenham, A.J., 2016. Predicting the perceptual consequences of hidden hearing loss. Trends in Hearing 20, 2331216516686768. doi:10.1177/2331216516686768.

Parthasarathy, A., Bartlett, E.L., Kujawa, S.G., 2019. Age-related changes in neural coding of envelope cues: Peripheral declines and central compensation. Neuroscience 407, 21–31. doi:10.1016/j.neuroscience.2018.12.007.

Parthasarathy, A., Kujawa, S.G., 2018. Synaptopathy in the aging cochlea: Characterizing early-neural deficits in auditory temporal envelope processing. Journal of Neuroscience doi:10.1523/JNEUROSCI.3240-17.2018.

Peelle, J.E., 2018. Listening effort: How the cognitive consequences of acoustic challenge are reflected in brain and behavior. Ear and Hearing 39, 204–214. doi:10.1097/AUD.0000000000000494.

Pichora-Fuller, M.K., Kramer, S.E., Eckert, M.A., Edwards, B., Hornsby, B.W.Y., Humes, L.E., Lemke, U., Lunner, T., Matthen, M., Mackersie, C.L., Naylor, G., Phillips, N.A., Richter, M., Rudner, M., Sommers, M.S., Tremblay, K.L., Wingfield, A., 2016. Hearing impairment and cognitive energy: The framework for understanding effortful listening (fuel). Ear and Hearing 37 Suppl 1, 5S–27S. doi:10.1097/AUD.0000000000000312.

Plack, C.J., Barker, D., Prendergast, G., 2014. Perceptual consequences of "hidden" hearing loss. Trends in Hearing 18, 2331216514550621. doi:10.1177/2331216514550621.

Ponsot, E., Devolder, P., Dhooge, I., Verhulst, S., 2025. Age-related decline in neural phase-locking to envelope and temporal fine structure revealed by frequency following responses: A potential signature of cochlear synaptopathy impairing speech intelligibility. Journal of the Association for Research in Otolaryngology 26, 249–270. doi:10.1007/s10162-025-00985-2.

Rhode, W.S., Smith, P.H., 1986. Encoding timing and intensity in the ventral cochlear nucleus of the cat. Journal of Neurophysiology 56, 261–286. doi:10.1152/jn.1986.56.2.261.

Schirmer, J., Wolpert, S., Dapper, K., Rühle, M., Wertz, J., Wouters, M., Eldh, T., Bader, K., Singer, W., Gaudrain, E., Başkent, D., Verhulst, S., Braun, C., Rüttiger, L., Munk, M.H.J., Dalhoff, E., Knipper, M., 2024. Neural adaptation at stimulus onset and speed of neural processing as critical contributors to speech comprehension independent of hearing threshold or age. Journal of Clinical Medicine 13. doi:10.3390/jcm13092725.

Sergeyenko, Y., Lall, K., Liberman, M.C., Kujawa, S.G., 2013. Age-related cochlear synaptopathy: An early-onset contributor to auditory functional decline. Journal of Neuroscience 33, 13686–13694. doi:10.1523/JNEUROSCI.1783-13.2013.

Van Der Biest, H., Keshishzadeh, S., Keppler, H., Dhooge, I., Verhulst, S., 2023. Envelope following responses for hearing diagnosis: Robustness and methodological considerations. The Journal of the Acoustical Society of America 153, 191–208. doi:10.1121/10.0016807.

Vasilkov, V., Garrett, M., Mauermann, M., Verhulst, S., 2021. Enhancing the sensitivity of the envelope-following response for cochlear synaptopathy screening in humans: The role of stimulus envelope. Hearing Research 400, 108132. doi:10.1016/j.heares.2020.108132.

Vasilkov, V., Verhulst, S., 2019. Towards a differential diagnosis of cochlear synaptopathy and outer-hair-cell deficits in mixed sensorineural hearing loss pathologies. medRxiv preprint doi:10.1101/19008680.

Verschooten, E., Shamma, S., Oxenham, A.J., Moore, B.C., Joris, P.X., Heinz, M.G., Plack, C.J., 2019. The upper frequency limit for the use of phase locking to code temporal fine structure in humans: A compilation of viewpoints. Hearing Research 377, 109–121. doi:10.1016/j.heares.2019.03.011.

Von Gablenz, P., Holube, I., 2017. Hearing loss and speech recognition in the elderly. Laryngo- Rhino- Otologie 96, 759 – 764. doi:10.1055/s-0043-119388.

Wagener, K., Brand, T., Kollmeier, B., 1999. Development and evaluation of a German sentence test Part III: Evaluation of the Oldenburg sentence test. Zeitschrift Fur Audiologie 38, 86–95.

Wang, M., Ai, Y., Han, Y., Fan, Z., Shi, P., Wang, H., 2021. Extended high-frequency audiometry in healthy adults with different age groups. Journal of Otolaryngology - Head & Neck Surgery 50, 52. doi:10.1186/s40463-021-00534-w.

Weiss, T., Rose, C., 1988. A comparison of synchronization filters in different auditory receptor organs. Hearing Research 33, 175–179. doi:10.1016/0378-5955(88)90030-5.

Zhu, L., Bharadwaj, H., Xia, J., Shinn-Cunningham, B., 2013. A comparison of spectral magnitude and phase-locking value analyses of the frequency-following response to complex tones. The Journal of the Acoustical Society of America 134, 384–395. doi:10.1121/1.4807498.