Model-based hearing-restoration strategies for cochlear synaptopathy pathologies

Fotios Drakopoulos *, Viacheslav Vasilkov, Alejandro Osses Vecchi, Tijmen Wartenberg, Sarah Verhulst

Dept. of Information Technology, Ghent University, 9000 Ghent, Belgium

* fotios.drakopoulos@ugent.be

Abstract

It is well known that ageing and noise exposure are important causes of sensorineural hearing loss, and can result in damage of the outer hair cells or other structures of the inner ear, including synaptic damage to the auditory nerve (AN), i.e., cochlear synaptopathy (CS). Despite the suspected high prevalence of CS among people with self-reported hearing difficulties but seemingly normal hearing, conventional hearing-aid algorithms do not compensate for the functional deficits associated with CS. Here, we present and evaluate a number of auditory signal-processing strategies designed to maximally restore AN coding for listeners with CS pathologies. We evaluated our algorithms in subjects with and without suspected age-related CS to assess whether physiological and behavioural markers associated with CS can be improved. Our data show that after applying our algorithms, envelope-following responses and perceptual amplitude-modulation sensitivity were consistently enhanced in both young and older listeners. Speech intelligibility showed small improvements across participants, with the young group benefitting the most from processed speech. Our proposed hearing-restoration algorithms can be rapidly executed and can thus extend the application range of current hearing aids and hearables, while leaving sound amplification unaffected.

1 Introduction

With age, our hearing ability starts to decline: Communicating in noisy environments becomes challenging, and the hearing of faint sounds difficult. Even though there has been extensive research on the understanding and restoration of hearing [1, 2], a precise diagnostic assay of the various aspects that constitute hearing impairment is still unavailable. The resulting mixed success in the treatment of sensorineural hearing loss (SNHL) reveals an incomplete understanding of how each component of the auditory system affects human sound perception. This was recently demonstrated by the ground-breaking discovery of cochlear synaptopathy (CS) [3–5], which showed that ageing [6, 7] and noise exposure [8] can cause a selective and substantial loss of auditory nerve fibres (ANFs) in rodents as well as humans. Evidence from human and animal studies suggests that CS impacts the encoding of speech in everyday listening conditions and the comprehension of speech in noise [9–11]. The functional consequences of CS have also been referred to as "hidden hearing loss" [12], as CS cannot be detected using standard audiometric threshold measures. In animal models, the onset of CS occurs earlier in time than that of the outer-hair-cell (OHC) loss that is associated with elevated hearing thresholds [3, 4], suggesting that a large portion of the noise-exposed and ageing population may suffer from undiagnosed CS, while their audibility falls within the normal audiometric range.

Recent studies have shown that CS mainly affects the robust processing of the temporal-envelope information in sound [13,14]. Functional consequences of synaptopathy include a reduced neural representation of instantaneous supra-threshold amplitude fluctuations [10], and stochastic under-sampling of speech in noisy acoustic scenarios [15, 16]. Despite the suspected high prevalence of CS in listeners with self-reported hearing difficulties, or those with OHC damage, conventional hearing-aid algorithms do not specifically compensate for the functional deficits associated with CS. On the contrary, hearing aids usually rely on non-linear

dynamic-range compression algorithms that reduce amplitude fluctuations of the temporal envelope [17, 18], possibly resulting in a further deterioration of hearing perception in individuals with CS. Thus, a large population of those with self-reported hearing difficulties but normal audiograms, or even those with impaired audiograms, may benefit from auditory signal-processing algorithms that specifically target the functional deficits associated with CS.

Since temporal-envelope coding is degraded by CS, a prime candidate for acoustic intervention is thus to modify the speech envelope which is essential for robust speech intelligibility (SI) [19–27]. Acoustic modifications of the speech-envelope shape and periodicity can improve SI [17, 28–32], especially in connection with hearing through cochlear implants [33–38]. Here, we go beyond these conventional approaches to develop a model-based type of auditory envelope processing that operates directly on the signal waveform to counteract the functional consequences associated with CS. We focus on CS-compensating signal-processing algorithms that target the temporal envelope of the signal while avoiding effects on the temporal fine-structure information, to avoid possible decrements in SI [39].

To develop this new type of CS-compensating audio processing, we adopted a biophysically-inspired computational model of the human auditory periphery [40] that simulates how CS impacts auditory sound processing. Even though individualised models of hearing-impaired auditory processing are widely adopted in the design of hearing-aid signal processing [41–44], they are typically based on the compensation of degraded hearing thresholds or of perceived loudness, to yield optimal acoustic amplification to hearing-aid users [45, 46]. Instead, we devised a computational optimisation method to design a number of envelope-processing algorithms that maximally restored simulated CS-affected auditory-nerve (AN) responses. We evaluated our model-designed algorithms in participants with and without suspected age-related CS, to test whether our algorithms enhanced a number of biophysical and perceptual markers associated with CS. These markers included envelope-following responses (EFRs), amplitude-modulation (AM) detection sensitivity, and SI in terms of speech reception thresholds (SRTs) and word recognitions scores (WRSs). Because a direct quantification of CS relies on post-mortem imaging of the temporal bone [7, 47], we focussed on older listeners who had been shown to be affected by CS through non-invasive electroencephalographic (EEG) techniques [14, 48, 49].

2 Design of the hearing-restoration algorithms

We developed our hearing-restoration algorithms based on a biophysically-inspired auditory periphery model [40] that can simulate human auditory processing in normal-hearing (NH) and hearing-impaired (HI) subjects. The peripheral auditory model includes a transmission-line description of the cochlea and an inner-hair-cell/auditory-nerve synaptic complex model to generate AN firing rates across a number of cochlear tonotopic locations [40]. We aimed to design signal-processing strategies that can restore peripheral AN processing in CS-affected peripheries, and thereby ultimately improve speech perception and intelligibility. A general description of the stages we followed for the design of the hearing-restoration algorithms is given here and the reader is referred to Methods (Sec. 7) for a more detailed description.

2.1 Optimisation procedure

We applied an iterative optimisation approach (Fig. 1) to derive an auditory-processing algorithm that minimises the difference between simulated NH and CS-affected AN responses. The model-simulated AN responses corresponded to the summed responses of the ANF population present in a NH and a CS-affected periphery, respectively. The derivative-free optimisation that we applied is efficient at discovering local minima of continuous and smooth functions for constrained problems, but it cannot guarantee a reliable solution when the problem is more complex. To this end, we constrained the optimisation problem by using a simple stimulus, i.e., a sinusoidally-amplitude-modulated (SAM) pure tone with a carrier frequency $f_c = 4$ kHz, modulation frequency $f_m = 120$ Hz and $m = 100\%$ modulation, that can be related to many behavioural and physiological studies of temporal-envelope processing in humans [14, 50–52].

2.2 CS compensation

The optimisation procedure was applied to three peripheries that included different degrees of CS, abbreviated as 13,0,0, 10,0,0 and 7,0,0, with each number corresponding to the quantity of high (H), medium (M) and low (L) spontaneous-rate (SR) ANF innervations, respectively. When compared to the reference NH periphery (13,3,3 ANFs [40, 53]), these corresponded to a complete loss of LSR and MSR ANFs, but with either healthy HSR ANFs (13,0,0), or with losses of 23% (10,0,0) or 46% (7,0,0) of the HSR ANFs, respectively. For each CS profile, an auditory signal-processing function g was optimised to maximally enhance the respective CS-affected AN output in response to the processed stimulus (Fig. 1a).

2.3 Auditory signal processing

We chose an envelope-processing function g (Eq. 2) that preserves the stimulus envelope peaks, but modifies sound onsets and offsets to increase the resting periods between stimulation. The (remaining) ANFs are thus able to recover faster from prior stimulation to generate stronger onset responses and more synchronised

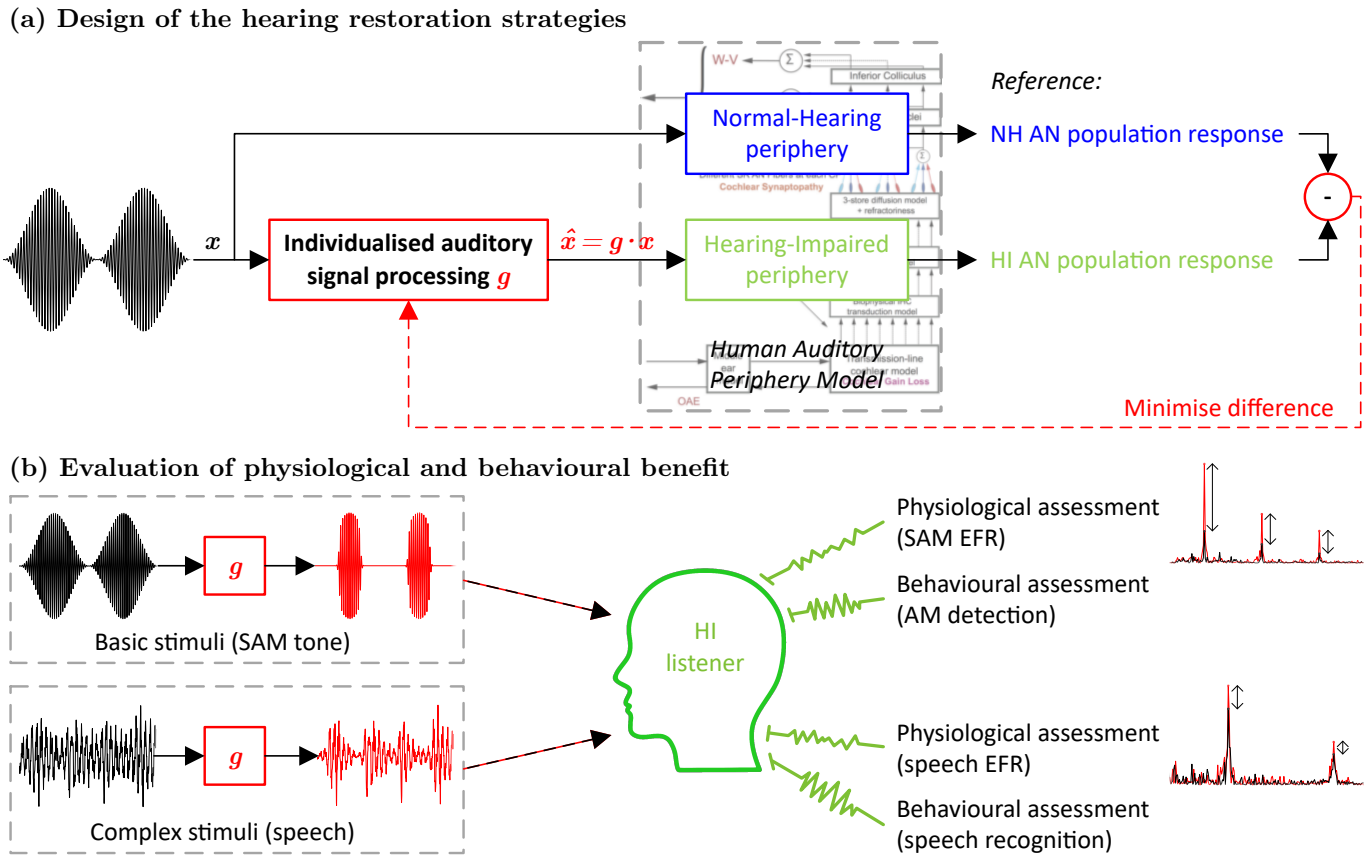


Fig 1. Framework for the design and evaluation of hearing-restoration strategies. a The schematic shows how signal-processing strategies can be designed to maximally restore a hearing-impaired (HI) AN response to the reference (NH) response. In our case, the HI periphery corresponded to a CS-affected periphery and the auditory signal processing to a non-linear envelope-processing function g. The parameters of the processing function g were optimised to generate a processed signal that minimises the difference between the two responses. The same optimisation approach can be applied to any HI periphery with different degrees of CS and/or OHC loss. **b** By applying the processing function g to different stimuli, the physiological and behavioural benefit of the processed sounds can be assessed on HI listeners. Ideally, the SNHL auditory profile of the HI listener corresponds to the simulated HI periphery that was used in the optimisation (**a**), yielding an individualised hearing restoration.

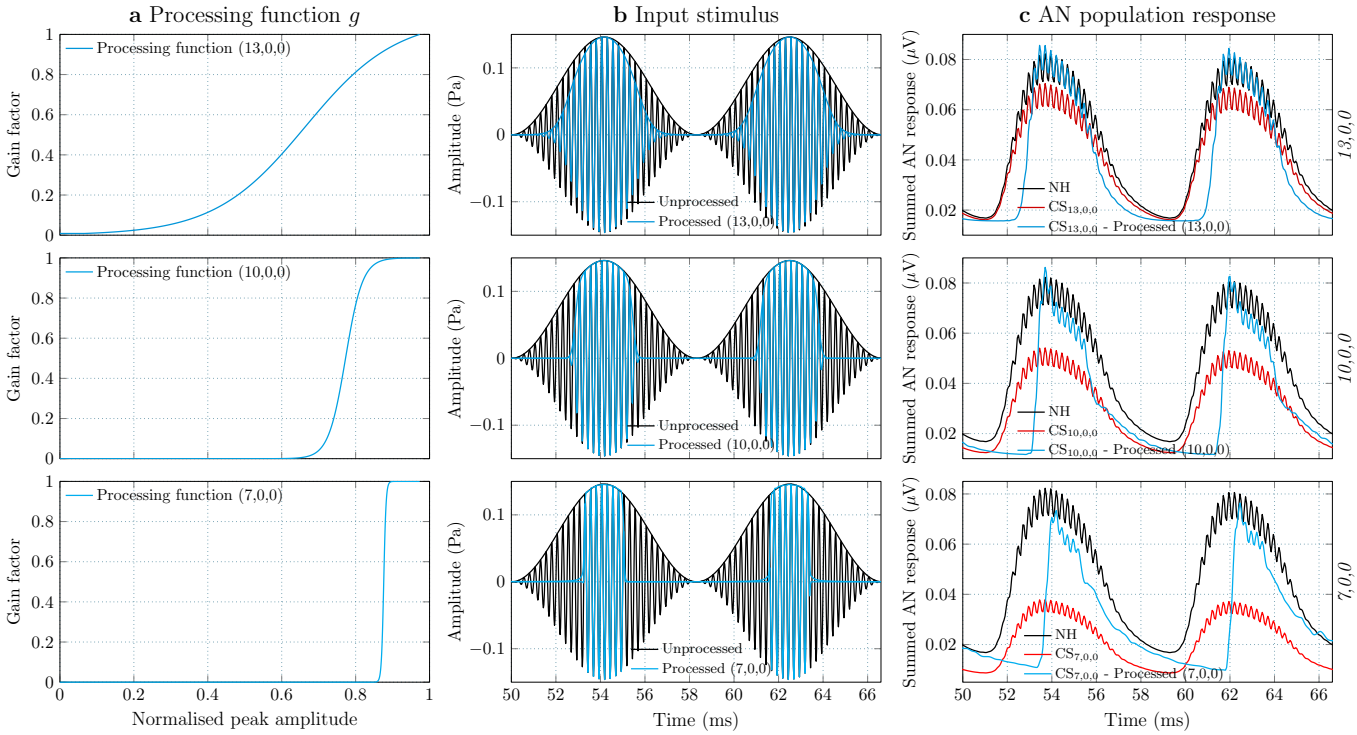


Fig 2. CS compensation of a fully-modulated SAM tone. Restoration of the AN responses of three CS-affected peripheries to a fully-modulated 70-dB-SPL SAM tone stimulus. **a** Three non-linear functions were derived from our optimisation method that maximally restore the AN responses for three degrees of CS. **b** Based on the signal envelope, the non-linear functions were applied to the SAM tones across time to derive the processed stimuli for each CS type. **c** The simulated AN responses of the NH and CS peripheries are shown in response to the unprocessed SAM stimulus (black and red) and the respective processed version (cyan), when given as input to the corresponding CS periphery.

AN activity [14, 28, 33, 34, 37, 54, 55]. During each iteration of the optimisation procedure, the parameters of the signal-processing function g were adjusted so that it could generate a processed stimulus $\hat{x} = g \cdot x$ that minimised the difference between the NH and the respective CS-affected AN response. The selected non-linear function g was easy to implement and was executed online (<5 ms latency for a 50-ms stimulus in MATLAB), allowing for future extensions in real-time applications.

Figure 2a shows the three processing functions that were derived after applying the optimisation procedure for the three selected CS profiles. The processing functions are shown as a function of the amplitude of the input stimulus x, normalised to its maximum absolute (peak) amplitude. Thus, for each CS profile, the envelope of the unprocessed SAM stimulus of Fig. 2b (modulating tone) can be used to compute the processing function g across time (Eq. *2*), which is then multiplied elementwise by the input stimulus x to derive the processed version $\hat{x} = g \cdot x$. The processed signal $\hat{x}$ is able to optimally stimulate the corresponding CS-affected periphery to generate an enhanced AN response that closely matches the NH response. Figure 2c demonstrates that simulated CS-affected AN responses can be partially or fully restored after applying our processing to fully-modulated SAM tones.

2.4 Generalising the hearing-restoration algorithms

Even though the algorithms were designed to restore ANF coding to fully-modulated SAM tones, the same processing can be applied to stimuli with different degrees of modulation, or to speech, which naturally contains envelope fluctuations of different strength. However, when applying the envelope-processing functions

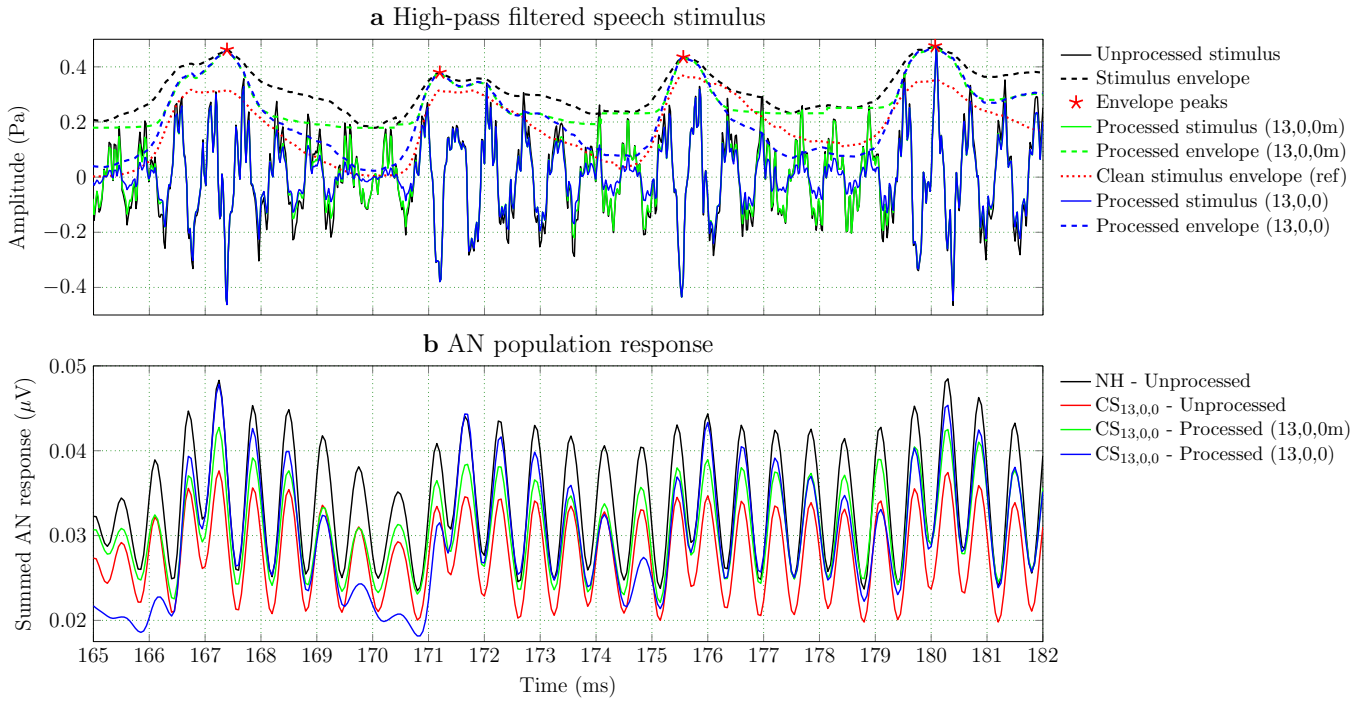


Fig 3. CS compensation of speech in noise. a The two processing types were applied for the 13,0,0 CS profile to the vowel /a/ of the word 'David', extracted from a sentence of the Flemish Matrix Test [56]. The sentences were high-pass filtered ($f_{cut} = 1.65$ kHz) and the generated speech-shaped noise was added to the word with SNR = 0 dB. **b** The unprocessed and processed stimuli were presented 20 times to the model [40], with randomly-selected segments of noise added each time. The unprocessed stimuli were given as input to the NH and CS peripheries, and the processed stimuli to the CS periphery to evaluate the achieved restoration. The simulated AN population responses were then averaged over the 20 stimulus presentations to derive the stimulus-driven responses that are shown here.

of Fig. 2a to partially-modulated stimuli, the modulation depth of the stimulus envelope will also be increased. This is not always desired, especially when the aim is to compare the effect of ANF stimulation recovery on the responses without affecting the dynamic range. Hence, we also designed a modified strategy g_m that only processes the slope of the modulation envelope while preserving the modulation depth of the original stimulus intact (Eq. 9). The resulting restoration of a partially-modulated SAM stimulus is shown in Supplementary Fig. 1, while the design of the modified strategy is explained in the Methods (Sec. 7).

The two audio processing types can be applied to any stimulus, with the option of modifying the modulation depth (and dynamic range) of the stimulus envelope or not. Both options aim to restore CS-affected AN processing, but might have qualitative differences in sound-perception outcomes. To generalise the two processing types for different audio stimuli, we first focussed on high-pass (HP) filtered speech (i.e., speech content above $f_{cut} = 1.65$ kHz) which relates more to the SAM stimuli. The intelligibility of HP-filtered speech relies on the coding of the temporal envelope of sound [57, 58] and has been linked to the EFR results of listeners in previous CS-related studies [52, 59, 60]. After establishing the envelope processing for HP-filtered speech and evaluating the simulated restoration that was achieved, the CS-compensating algorithms can be used in exactly the same way to process any stimulus. Our final goal was to optimally apply our developed strategies to natural, broadband (BB) speech stimuli and improve speech coding in noise, where CS predominantly affects speech perception [9–11, 15, 16].

An example of how our envelope-processing algorithms can be applied to any (modulated) stimulus is shown in Fig. 3a, where the two processing types were applied to a HP-filtered speech-in-noise stimulus using the 13,0,0 CS-profile parameters (Supplementary Table 3). To estimate the modulation envelope and

peaks of the noisy stimulus across time, a blind RMS-based procedure was adopted (see Methods). The envelope-processing functions g (Eq. 2) and g_m (Eq. 9) were computed between the estimated peaks and troughs of the envelope across time, and then multiplied by the noisy stimulus. The processed stimuli (and envelopes) of the two strategies are labelled as 13,0,0 and 13,0,0m, respectively. Fig. 3b shows the restoration yielded by the two algorithms on the simulated AN responses. The original processing strategy (13,0,0) was able to fully restore the CS-affected response peaks back to the NH response by extending the dynamic range, while the modified strategy (13,0,0m) achieved only a partial restoration.

Although the algorithms were originally designed for SAM tones, the processing had similar simulated outcomes when applied to HP-filtered speech in noise. In the same way, the algorithms are expected to yield similar restoration when applied to natural, BB speech. Even though our model failed to predict an improvement for natural speech after processing, we still applied our algorithms in our measurements to assess their effect. Finally, to estimate the impact of our CS-compensating algorithms on speech perception, we evaluated their processing using two objective metrics for speech quality and intelligibility, the PESQ [61] and STOI [62], respectively. Supplementary Table 1 shows the average PESQ and STOI scores before and after processing, computed for natural (BB) and HP-filtered Flemish Matrix [56] sentences in noise. To ensure the robustness of our envelope estimates in noise, we also included a reference strategy in which we applied the modified strategies using the estimated envelope of the clean-speech stimulus instead (Fig. 3a; ref). Although the 13,0,0 CS-processing algorithms mostly improved the objective evaluation metrics in BB speech, the 10,0,0 and 7,0,0 algorithms were detrimental to the PESQ and STOI scores, with the original strategies decreasing STOI in both cases. A more pronounced processing is applied to the 10,0,0 and 7,0,0 compared to the 13,0,0 CS profile (Fig. 2), which results in processed speech sentences that sound less natural and significantly differ from the unprocessed speech. For this reason, and to limit the duration of our tests, we chose to exclude most of the 10,0,0 and 7,0,0 CS-processing strategies from our SI measurements.

3 Evaluation

To evaluate the applicability of our algorithms to younger and older listeners with normal audiograms, we assessed possible improvements in physiological as well as sound-perception experiments (Fig. 1b). The evaluation included three measurement sessions (Supplementary Table 4), in which differences in human responses between processed and unprocessed sound were evaluated using measured EFRs (i.e., EEG potentials evoked by the synchronous response of neurons to the envelope of a stimulus), AM detection sensitivity, and SI. The most relevant types of processing were evaluated in each session based on the assessed auditory test, ensuring that the duration of each measurement was kept under 1.5 hours. The processing algorithms were fast to execute and were applied online to the stimuli of each trial.

NH volunteers aged between 18-25 (yNH) and 45-65 (oNH) years old participated in the study and listened to the processed and unprocessed audio for each test. Based on previous reports [14,63], we expected that the older listeners suffered from age-related CS, with correspondingly smaller EFR markers. For all subjects, audiometric thresholds were ≤ 20 dB for frequencies up to 4 kHz, except for four oNH subjects, who are indicated with a darker colour in all the following figures (Supplementary Fig. 2). These four subjects suffer from presumed age-related CS in addition to their OHC deficits.

3.1 EEG measurements

During the EEG measurement session, we assessed whether our CS algorithms yielded EFR enhancements in response to SAM tonal stimuli ($f_c = 4$ kHz, $f_m = 120$ Hz, $m = 100\%$). Four conditions were measured, corresponding to the unprocessed stimulus and the processed versions of the stimulus after applying the processing schemes for the three CS profiles (13,0,0, 10,0,0 and 7,0,0). To evaluate the applicability of our processing strategies to speech, we also assessed whether our algorithms yielded EFR enhancements to a speech stimulus. We used a word from the Flemish Matrix test [56] that was HP filtered and that had a constant fundamental frequency $f_0 \approx 220$ Hz. EFRs were recorded for the unprocessed stimulus and a

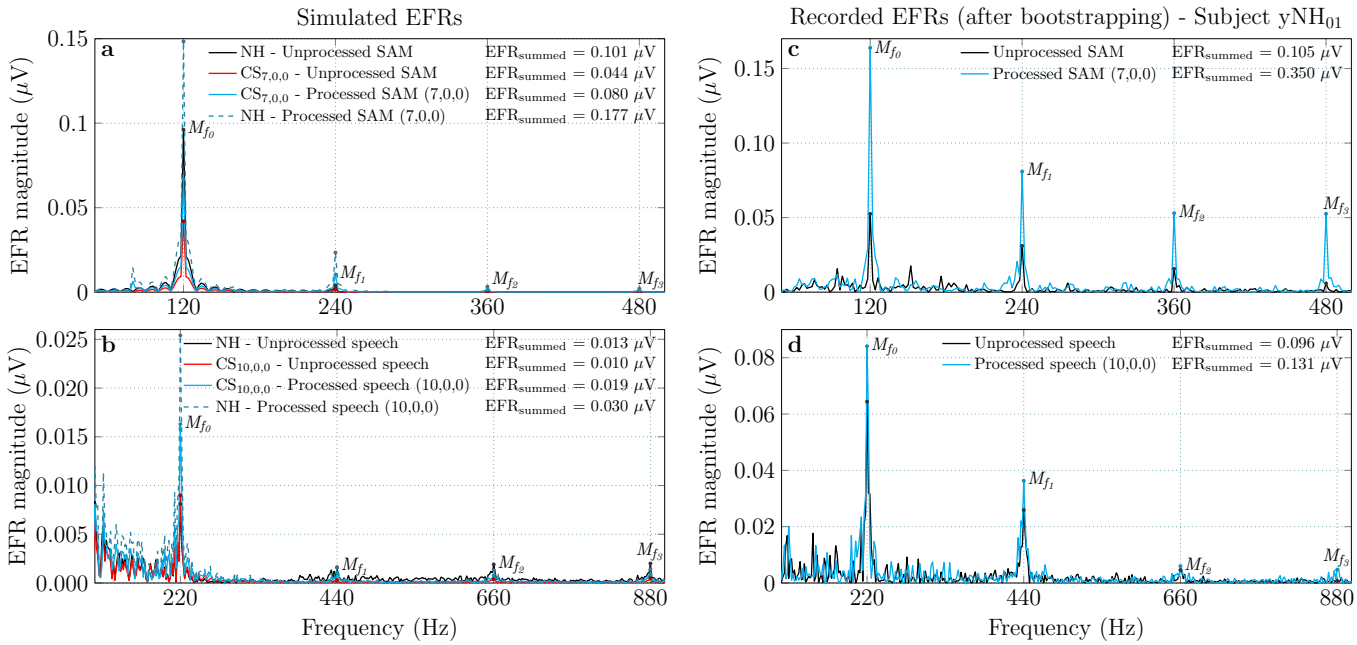


Fig 4. Comparison of simulated and recorded EFRs. a-b Simulated EFRs to the SAM tone stimulus (**a**) and the HP-filtered speech stimulus (**b**), shown for a NH periphery and a CS periphery in response to both the unprocessed and processed stimuli. **c-d** Recorded EFRs of one yNH participant (yNH_{01}), in response to the SAM stimulus (**c**) and the speech stimulus (**d**) before and after processing. In all panels, the EFR spectral peaks at the fundamental frequency of the stimulus and the next three harmonics are indicated and were used to compute the EFR_{summed} magnitudes (Eq. 10).

processed version. To limit the duration of the EEG measurement session, only one processing condition was chosen for the speech stimulus (10,0,0). More details about the EEG recording and analysis procedures can be found in the Methods (Sec. 7).

To demonstrate the EFR analysis procedure and create an expectation of what the recorded responses would look like, the EFRs of both stimuli were also simulated using our model [40]. Figures 4a,b show the frequency-domain representations of simulated EFRs to the unprocessed and processed stimuli, as presented to the NH periphery and the corresponding CS-affected periphery. Figures 4c,d also show recorded mean EFRs of one participant (yNH_{01}) for the same stimuli. For each condition of the simulated and recorded EFR spectra, the EFR_{summed} magnitude was computed using Eq. *10* to quantify the changes in the EFR magnitude after processing. The recorded EFRs corroborated the simulated enhancement trends, while the simulated responses of Fig. 4a,b suggested that the processed stimuli can also yield EFR enhancements in pristine peripheries (dashed lines). This observation follows our expectations to a certain degree, since we designed our algorithms to optimally drive the (remaining) ANFs. Consequently, our algorithms can enhance peripheral coding in any periphery.

3.2 AM detection

During the AM measurement session, we assessed whether the CS algorithms improved behavioural AM-detection sensitivity. The aim was to evaluate whether peripheral coding improvements would also benefit sound perception when adopting the same stimuli. After determining the standard AM detection thresholds (i.e., the minimal modulation depth required to distinguish a modulated tone from a pure tone) using unprocessed stimuli, the stimuli were processed to measure the AM detection thresholds for the three CS profiles. To this end, the modified processing strategies 13,0,0m, 10,0,0m and 7,0,0m were used, to preserve the modulation depth of the AM tones. For each of the three CS conditions, the processing was applied to

the modulated tone to yield a stimulus with the same modulation depth but a "sharper" envelope. Because the envelope-processed tones occupied broader FFT spectra after processing, with frequency components beyond the critical-band limits (±1 ERB), increased off-frequency listening was expected and may have introduced additional perception cues [64, 65].

3.3 Speech intelligibility assessment

During the speech intelligibility (SI) session, we assessed whether the CS algorithms improved SI for natural (BB) and HP-filtered speech. The Flemish Matrix sentence test [56] was used to derive the SRTs (i.e., the SNR required to reach 50% word recognition) of each participant for the two speech types. The SRTs were recalculated after applying one processing algorithm (13,0,0m-ref) to the noisy sentences of each trial, to roughly assess whether the processing improved the SRTs of BB and HP speech. We selected the clean-envelope 13,0,0 modified processing strategy (13,0,0m-ref) for this test, since it gave the best overall objective scores for the processed stimuli (Supplementary Table 1).

Lastly, we individually assessed the benefit to SI based on the measured SRT values of each listener. The aim was to directly compare the effect of each processing strategy on the word-recognition performance of the listeners near their SRT level. It should be noted that the Flemish Matrix test has an intrinsic SRT variability of ±0.5 dB [56], which could compromise small SI improvements after processing. Therefore, we used fixed SNRs, for which we measured the percentage of correctly recognised words before and after processing (see Methods). We chose 3 SNR conditions corresponding to the SRT, SRT-3 and SRT+3 dB levels of each participant, respectively, measured for both BB and HP speech. As noted above, we limited our experimental setup by only considering processing strategies that were not detrimental to the objective SI evaluation metrics (Supplementary Table 1). To this end, four processing conditions were evaluated here: the two 13,0,0 processing strategies (13,0,0 and 13,0,0m), the modified 10,0,0 strategy (10,0,0m), and the reference clean-envelope 13,0,0 condition (13,0,0m-ref). For each of these four processing conditions and the respective unprocessed condition, the average WRS was measured for every subject for each SNR and speech type.

4 Results

The results of our physiological and behavioural measurements (EFRs, AM detection, SRTs and SI percent scores) for the yNH and oNH listener groups are shown in Figs. 5-8. To facilitate a visual comparison between the processed conditions, the median values of the unprocessed conditions for the results of the two groups are indicated by dashed lines in each figure. Blue colours are used for the results of yNH subjects and red for oNH subjects. We adopted an independent two-sample T-test to assess statistical significance between the group results and a dependent T-test to assess statistical significance among the different conditions of each group. In all figures, statistically significant differences with p-values < 0.05 are indicated with one asterisk, p-values < 0.01 with two asterisks, p-values < 0.001 with three asterisks and p-values < 0.0001 with four asterisks. For within-group comparisons (among conditions), statistically significant differences were only computed for the comparison between the processed conditions and the unprocessed reference. Bonferroni correction was applied before establishing significance for the p-values. The correction was based on the respective number of processed conditions that were compared.

4.1 EFRs increased after CS-compensating processing

Figure 5 shows individual and median EFR magnitudes (Eq. *10*) for the two subject groups, computed from the EFR spectra of the different conditions. For both unprocessed stimuli, the oNH subjects had lower EFR magnitudes than the yNH subjects (SAM: $p = 0.0083$; speech: $p = 0.0005$), consistent with our expectations of age-related CS for the older group [14, 63]. In the case of the SAM stimulus (Fig. 5a), the 13,0,0 processing condition enhanced the recorded EFR magnitudes for the oNH group but not for the yNH group. At the

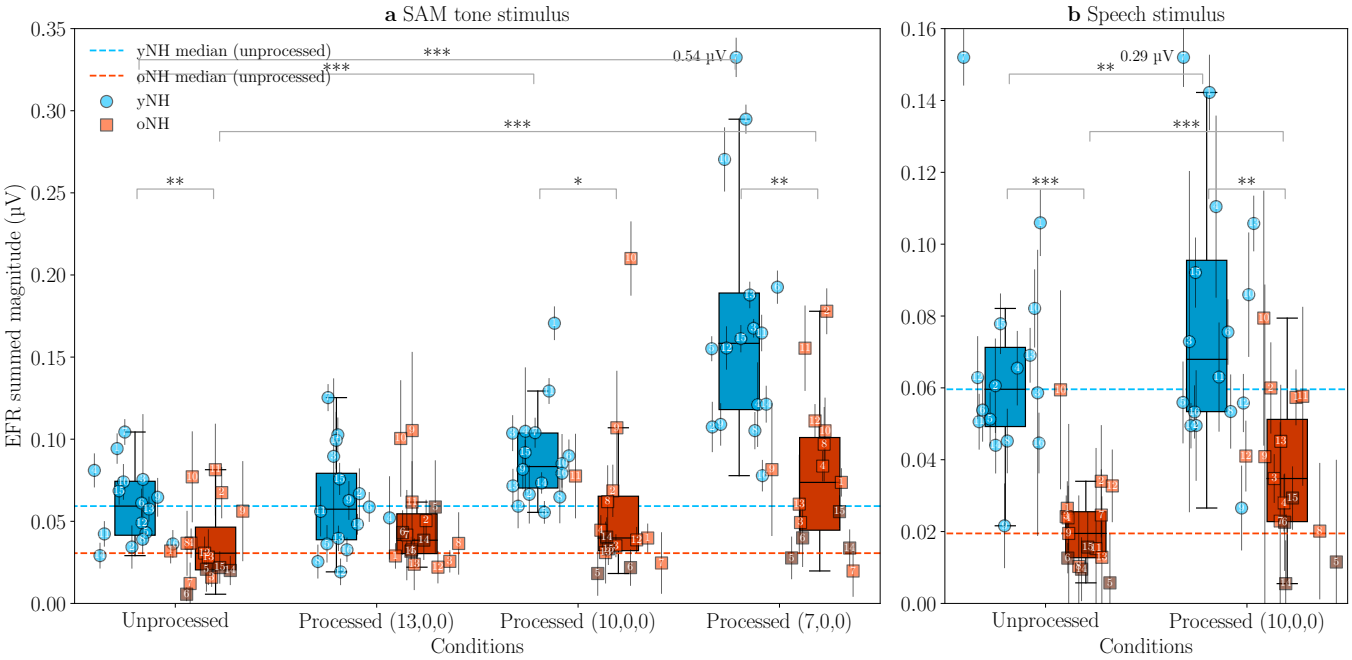


Fig 5. Effect of CS compensation on physiological EFR magnitudes. a Recorded EFR results for the SAM stimulus, before and after processing with the three CS algorithms. **b** Recorded EFR results for the speech stimulus ('David') before and after processing. In both panels, the individual points correspond to the mean and standard deviation of the summed EFR magnitudes (EFR_{summed}), computed from the 400 bootstrapped EFR spectra of each participant (Eq. 10). The box and whisker plots indicate the EFR medians and quartiles of the two groups under each condition.

same time, the spread of individual EFR magnitudes within the yNH group was increased by processing, with an interquartile range (IQR) increase from 0.033 µV (unprocessed) to 0.04 µV (13,0,0).

The 10,0,0 and 7,0,0 conditions, however, significantly improved the EFRs in both groups (Fig. 5a), with the 10,0,0 condition showing a statistically significant benefit for the yNH group only ($p = 0.0001$) and the 7,0,0 showing statistically significant differences for both groups (yNH: $p = 0.0001$; oNH: $p < 0.0001$). Statistically significant differences were also found between the performance of the two groups (unprocessed, 10,0,0 and 7,0,0 conditions), demonstrating that there was an effect of age (or age-related synaptopathy) on the magnitude of the recorded EFRs irrespective of the processing applied. The oNH median in the 7,0,0 condition surpassed the yNH median in the unprocessed condition, showing a recovery of the EFR magnitudes in the oNH group to the level of the yNH group after processing. Overall, the SAM EFR processing benefit was more pronounced for the yNH group than for the oNH group, with a median increase of 0.099 µV and 0.043 µV for the yNH and oNH groups in the 7,0,0 processing condition, respectively. At the same time, the 7,0,0 condition significantly spread the individual EFR results (within both groups), with the IQRs increasing by 0.038 µV (116.1%) for the yNH group and by 0.03 µV (116.9%) for the oNH group. This shows that some subjects benefitted more than others from the processing, with those who had unprocessed EFR values below the group median benefitting the least.

For the speech-evoked EFR (Figure 5b), the 10,0,0 condition yielded statistically significant EFR improvements in both groups (yNH: $p = 0.0073$; oNH: $p = 0.0005$). The EFR magnitudes between the two groups also showed statistically significant differences ($p = 0.0005$ for the unprocessed condition). This suggests that pitch tracking was more affected by age-related CS, i.e., the older group relied more than the young group on envelope detection. Although the oNH EFR magnitudes were improved after processing, the (10,0,0) processed speech stimuli were not able to fully restore the oNH magnitudes to the yNH level. Once again, we report larger dispersion (spread) after processing, with the IQRs increasing by 0.02 µV (91.2%) for

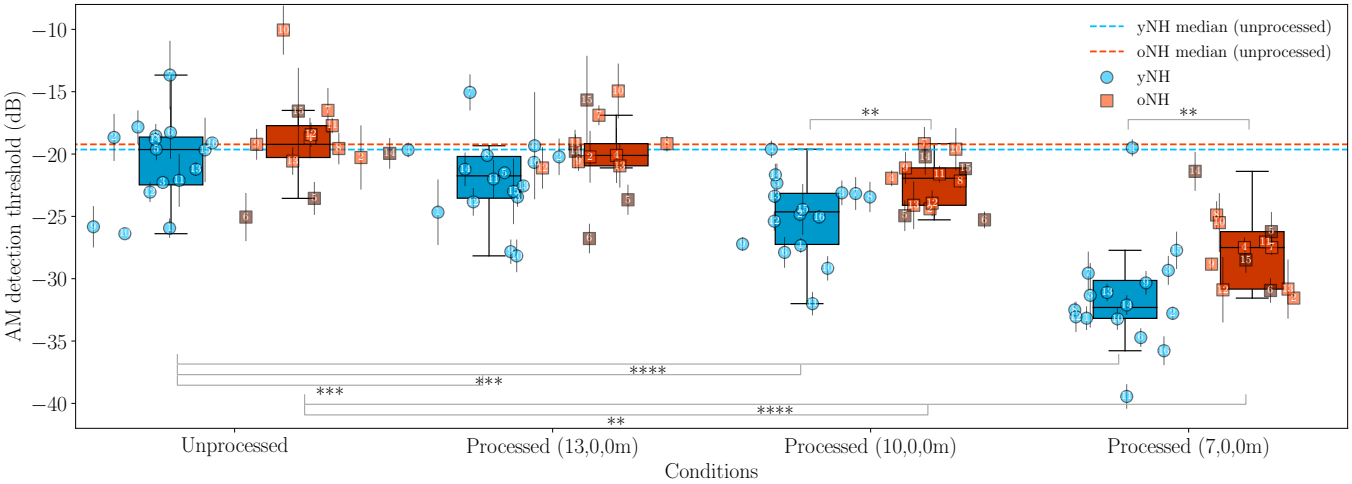


Fig 6. Effect of CS compensation on behavioural AM detection thresholds. For each condition, the mean and standard deviation of the individual AM detection thresholds are shown, computed from three trials for each participant and each condition. The box and whisker plots correspond to the AM-detection-threshold medians and quartiles of the two groups. More negative thresholds indicate better performance.

the yNH group and by 0.016 µV (122.2%) for the oNH group.

4.2 Behavioural AM detection thresholds improved after CS processing

Figure 6 shows individual and median AM detection thresholds of the yNH and oNH groups, before and after processing. All processing schemes improved the median AM thresholds and had a larger effect on the yNH group. The improvements after processing were statistically significant in all cases, except for the oNH group and the 13,0,0 condition, and the benefit increased as the processing compensated more strongly for CS. The 7,0,0 condition significantly improved the AM thresholds, with a 12.5 dB improvement for the yNH group ($p < 0.0001$) and an 8.1 dB improvement for the oNH group ($p < 0.0001$).

Statistically significant differences were also found between the performance of the two groups for the 10,0,0 and 7,0,0 processing conditions. It is noteworthy that although both age groups performed similarly for the unprocessed condition (median thresholds of -19.64 dB for the yNH group and -19.22 dB for the oNH group), the performance of the groups differed significantly after processing (median thresholds of -32.31 dB for the yNH group and -27.5 dB for the oNH group for the 7,0,0 processing condition). Once again, this shows that yNH subjects benefitted more from the CS-compensating algorithms than oNH subjects.

4.3 Mixed outcomes in SI benefit after CS compensation

Figure 7 shows individual and median SRTs for natural (BB) and HP-filtered speech in the oNH and yNH subject groups. In contrast to the observed EFR and AM detection processing benefits (Figs. 5,6), the SRTs were degraded after processing, especially for the oNH group. A small but insignificant SRT improvement was observed for the BB condition, but only among the yNH group, with a median decrease of 0.2 dB from the unprocessed BB SRT. For HP-filtered speech (which relies on temporal-envelope processing), the yNH SRTs had an almost identical median but showed a larger dispersion after processing, with an IQR increase from 1.27 dB to 3.26 dB. A statistically significant difference was also observed between the performance of the two age groups in HP-filtered speech ($p = 0.0274$), but not in BB speech. This corroborates our hypothesis suggesting degraded temporal-envelope coding in the older group, since only the high-frequency component of speech was less intelligible for these subjects. Although consistent improvements were not found for SRTs

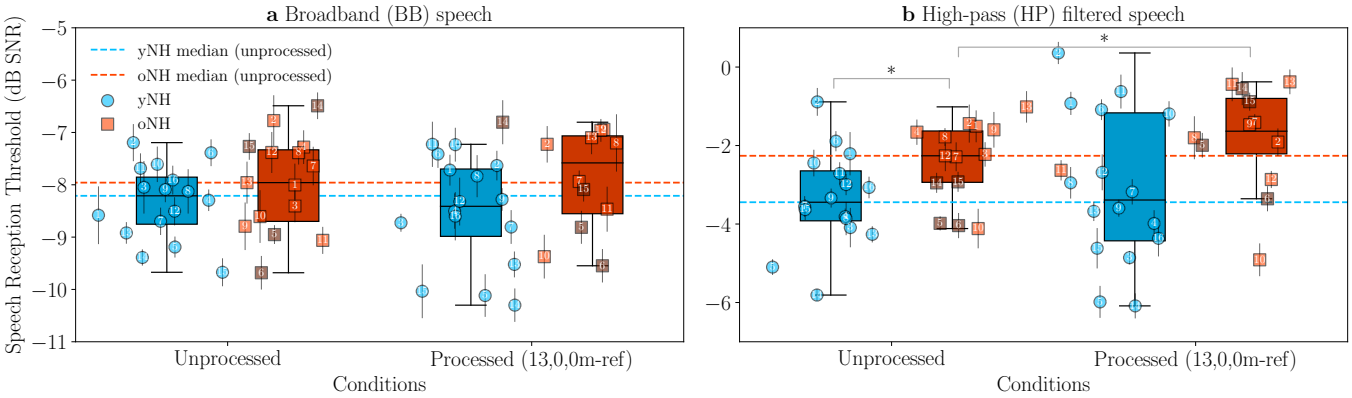


Fig 7. Effect of CS compensation on behavioural speech-reception thresholds. a Measured SRTs for BB speech before and after processing with the clean-envelope 13,0,0 modified processing strategy (13,0,0m-ref). **b** Measured SRTs for HP-filtered speech before and after processing with the 13,0,0m-ref processing strategy. For each participant and each condition, the mean and standard deviation was computed from the last 6 reversals of the respective SRT measurement. The box and whisker plots correspond to the SRT medians and quartiles of the two groups. More negative SRTs indicate better performance.

after processing, Fig. 7 shows that there were several yNH subjects who benefitted from processed speech under both speech conditions (e.g., yNH_{05}, yNH_{14}, yNH_{15}). These subjects corresponded to data points that were below the group median, i.e., subjects who had the best speech intellibiligity performance (lowest SRTs) in the unprocessed condition.

Finally, Fig. 8 shows the SI scores (WRSs) before and after processing, computed for BB and HP-filtered speech at three SNRs (SRT, SRT-3, SRT+3 dB). Since the SNR levels (at which the scores were computed) were individually adjusted to the SRTs of each participant for BB and HP-filtered speech (Fig. 7), both group medians were expected to lie at ∼50% for the BB (SRT) and HP (SRT) unprocessed conditions (Fig. 8b,e). The SNRs of ±3 dB SRTs (Fig. 8a,d and Fig. 8c,f) corresponded to two different points on the psychometric functions of the participants, and resulted in measured median scores of ∼14% and ∼32% for BB and HP-filtered speech at SRT-3 dB, respectively, and to ∼83% and ∼73% for BB and HP-filtered speech at SRT+3 dB, respectively. Assuming that the ±3 dB deviation from the individual SRT resulted in points on the linear part of the psychometric function [56], the average slope of the curve was 11.3 ± 1.5 %/dB for BB speech and 6.5 ± 0.1 %/dB for the HP-filtered speech.

For natural (BB) speech (Fig. 8a-c), the 13,0,0 and 10,0,0m processing conditions mostly decreased the individual SI scores of participants, with a less pronounced effect for the HP-filtered speech condition (Fig. 8d-f). For the 13,0,0 modified conditions (13,0,0m-ref and 13,0,0m), a similar effect as in Fig. 7 was observed, with the processing mainly improving the performance of yNH participants. The 13,0,0m processing resulted in improved median scores for yNH participants among all SNR conditions, and to small improvements for oNH participants in some cases. For the most difficult SNR condition (SRT-3 dB), the median SI increased by 5% for the yNH group ($p = 0.0155$) and by 3.3% for the oNH group ($p = 0.0824$) in BB speech (Fig. 8a), and by 8.3% for the yNH group ($p = 0.0028$) and 3.3% for the oNH group ($p = 0.5121$) in HP-filtered speech (Fig. 8d). Based on the estimated psychometric curve slopes of the BB and HP conditions, this corresponds to an SRT improvement of 0.44 dB and 0.29 dB for the yNH and oNH group, respectively (natural speech), and to an SRT improvement of 1.28 dB and 0.51 dB for the yNH and oNH group, respectively (HP-filtered speech).

In contrast to the 13,0,0m processing condition, the clean-envelope reference condition (13,0,0m-ref) did not show consistent improvements to the SI scores. Thus, processing based on a blind approximation of the noisy envelope was more robust than providing the envelope of the clean stimulus.

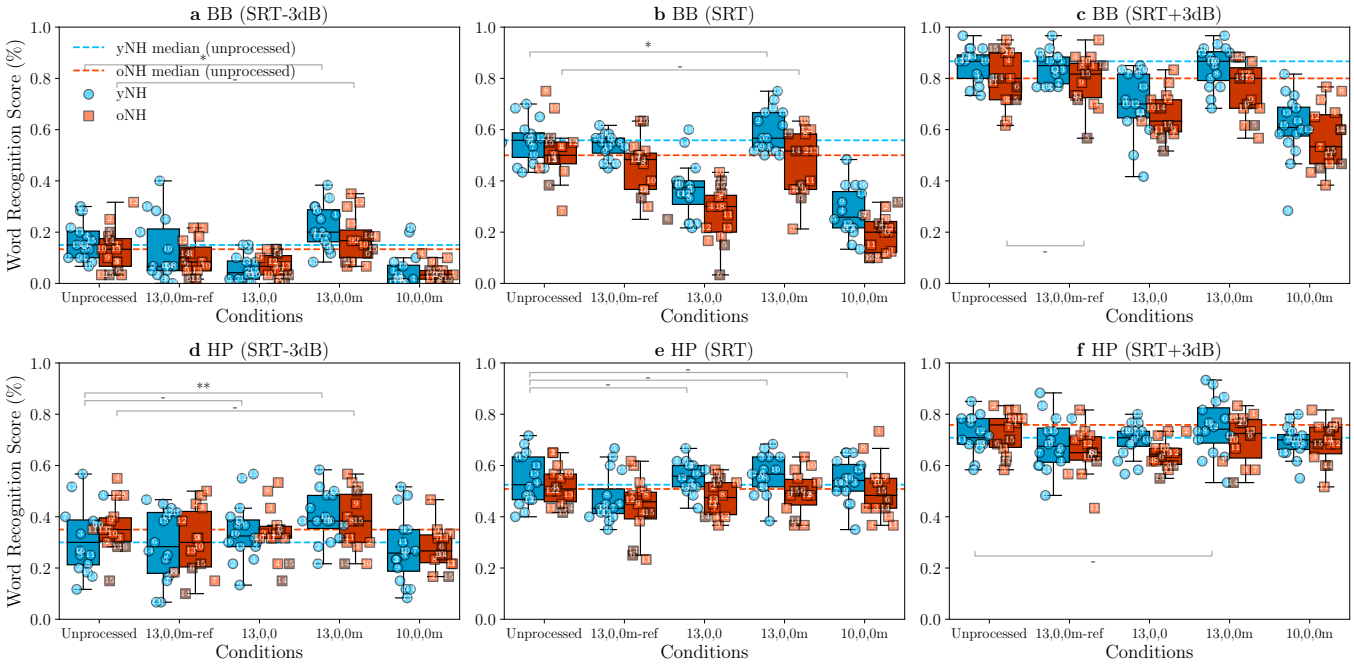


Fig 8. Effect of CS compensation on word-recognition scores near the individual SRT. For BB and HP-filtered speech at 3 SNR levels (SRT-3, SRT, SRT+3 dB), the average WRS of each participant was computed for each condition from 12 measured trials. Statistical significance between unprocessed and processed results is shown only for the cases where a higher score median was obtained after processing. The box and whisker plots correspond to the score medians and quartiles of the two groups. A higher score indicates better intelligibility.

4.4 Who benefits from CS-compensating processing?

To investigate the characteristics of those who benefitted more strongly from CS processing, we studied the relationships between the different physiological and behavioural measures using Spearman's rank-order correlation r. Correlations were computed between measured EFR magnitudes, AM detection thresholds, SRTs, WRSs, and audiometric thresholds. Supplementary Tables 5-7 show the values of Spearman's rank-order correlation r for the SAM EFR, speech EFR and AM detection results. For each metric, correlations were also computed for the improvement of each processed condition, i.e., after subtracting the respective unprocessed measures from each processed condition. Two metrics were considered for the correlations relating to audiometric thresholds: the hearing threshold at 4 kHz and the average extended-high-frequency (EHF) hearing thresholds (10-16 kHz). For the WRS results, we only considered the SI improvements of each processing condition. Finally, to better assess the results of our best-performing processing condition on SI (13,0,0m), Supplementary Table 8 summarises the correlations of the WRS improvement after processing with this condition.

Correlations between EFR magnitudes and SI EFR magnitudes showed statistically significant negative correlations ($r < 0$) to the SI score improvement, especially within the yNH group (Supplementary Tables 5,6). These negative correlations suggest that participants with low SAM EFR magnitudes benefitted more from processed speech (i.e., weaker EFR magnitudes resulted in larger SI score improvements), in line with the view that those with higher degrees of CS benefitted most from the processing.

However, this trend was not observed for all processing conditions. In the case of the 13,0,0m strategy, positive correlations ($r > 0$) were found between the speech EFRs and the SI score improvement in BB speech (Supplementary Table 8). Figure 9a,b shows the correlation of speech EFRs to the SI score improvement for

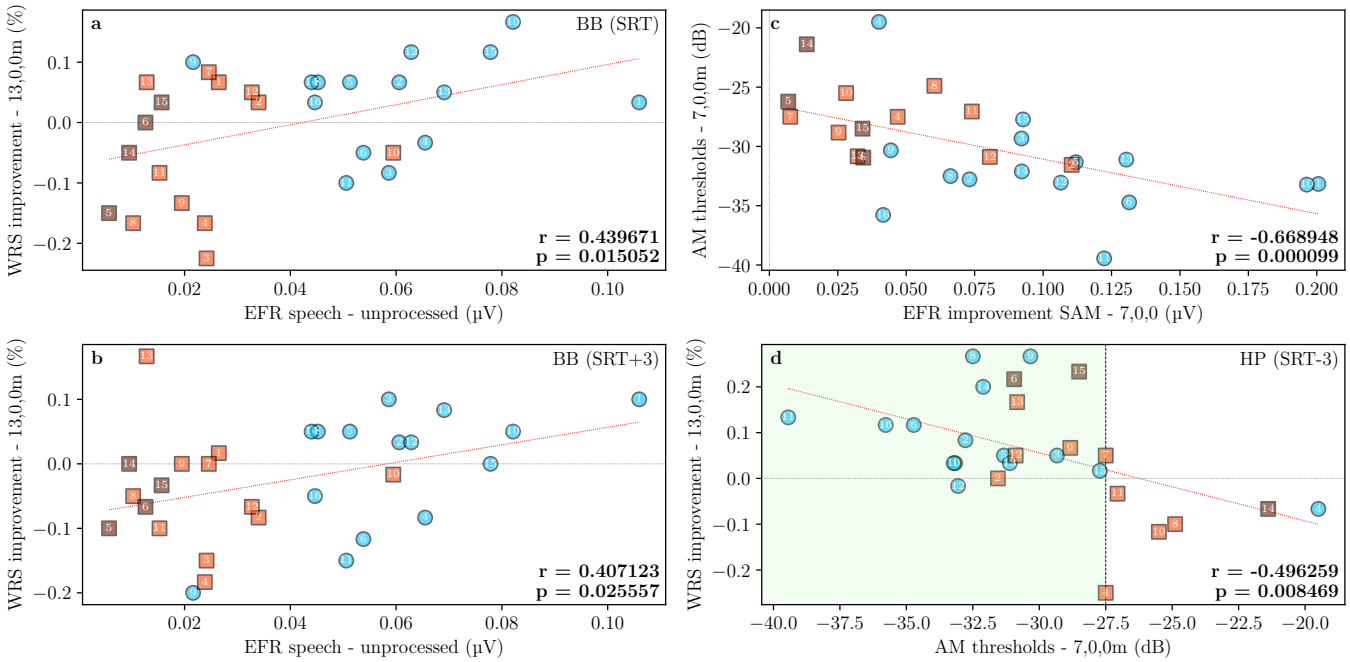


Fig 9. The most significant correlations between metrics. a,**b** Correlations between the speech EFRs (unprocessed) and the WRS improvement of the 13,0,0m strategy (13,0,0m - unprocessed) for two SNR levels of BB speech (SRT and SRT+3 dB). **c** Correlation between the SAM EFR improvement after processing with the 7,0,0 condition (7,0,0 - unprocessed) and the AM detection thresholds of the 7,0,0m condition. **d** Correlation between the AM detection thresholds of the 7,0,0m condition and the WRS improvement after processing with the 13,0,0m condition (13,0,0m - unprocessed), computed for HP-filtered speech at SNR = SRT - 3 dB. In all panels, the values of the Spearman's rank-order correlation r and the probability p are shown on the lower right corner. The red dotted lines correspond to the linear regression model fit to the data.

two SNR levels, demonstrating that subjects with strong speech EFRs (and thus good pitch tracking) were able to benefit more from the 13,0,0m strategy in natural-speech recognition in noise.

In contrast, almost no statistically significant correlations were found between measured EFRs and SRTs. The only statistically significant correlation was found between EFRs to the (unprocessed) speech stimulus and SRTs in HP-filtered (unprocessed) speech, after removing the oNH subjects with audiometric thresholds > 20 dB (Supplementary Table 6). This suggests that the EFR strength of the individual listeners does not directly relate to their SRT performance (i.e., the amount of noise they can tolerate for understanding speech), but only to the relative increment of their SI after processing (WRS improvement).

Correlations between EFR magnitudes and behavioural AM detection thresholds Strong and statistically significant correlations were found between AM detection thresholds and EFR magnitudes under most conditions (Supplementary Table 7), especially for the 7,0,0 processing. Figure 9c shows the most pronounced relationship among the measures, computed between 7,0,0m-processed AM detection thresholds and 7,0,0-processed SAM EFR improvement from the unprocessed condition (7,0,0 - unprocessed). All correlations were negative ($r < 0$), showing that AM thresholds generally decreased as the EFR magnitudes, or as the improvement of the EFR magnitudes after processing, increased (and vice versa). This indicates a strong link between physiological and behavioural metrics when the same stimuli and processing methods are adopted.

Strong correlations were also found between the 7,0,0-processed SAM EFRs and all measured AM detection thresholds after removing the oNH subjects with elevated hearing thresholds (Supplementary Tables 5,7). For

most conditions, removing these subjects (all w/o oHI) increased the statistical significance of the correlations between EFRs and AM detection thresholds, suggesting that an OHC-deficit SNHL component in addition to suspected CS may require different processing than our CS-compensation algorithms.

Correlations between AM detection thresholds and SI AM detection performance showed significant negative correlations to the SI score improvement after processing, especially for the 7,0,0m AM thresholds (Supplementary Table 7). Thus, participants with good AM detection performance (low thresholds) benefitted more from the processed speech than those with poor AM sensitivity. As an example, Fig. 9d shows the relation between HP-filtered speech at the noisiest scenario (SRT-3 dB) and the 7,0,0m AM threshold. Most participants with AM thresholds below -27.5 dB (light green part) were able to benefit from the 13,0,0m processed speech. Hence, the AM detection performance in the 7,0,0m condition could provide an indication of the expected SI score improvement after CS processing.

Correlations with audiometric thresholds When examining both age groups together, all metrics correlated well with the 4-kHz audiometric thresholds and EHFs. Statistically significant negative correlations were found between audiometric thresholds and measured EFRs (Supplementary Tables 5,6), especially for the 7,0,0 SAM EFR condition and the speech EFR (unprocessed). Thus, participants with more elevated audiometric thresholds (at 4 kHz and EHFs) had lower EFR magnitudes. Statistically significant positive correlations were also found between audiometric thresholds and measured AM detection thresholds (Table 7), especially for the 7,0,0 condition, showing that people with OHC deficits performed worse on this task.

However, when removing the oNH subjects with elevated thresholds from the statistical analyses (all w/o oHI), the statistical significance of the audiometric-threshold correlations consistently decreased (Supplementary Tables 5-8). This indicates that these four subjects could have biased the statistical analysis. At the same time, almost no correlations to the audiometric thresholds were found when examining the yNH and oNH groups separately. This suggests that the age difference between the two groups is the most important factor here and not the audiometric thresholds per se, which also implies the effect of age-related CS [6, 49, 63, 66].

5 Discussion

Our data show that physiological EFRs and perceptual AM sensitivity were enhanced in both yNH and oNH listeners when using our CS-compensating algorithms. The "mildest" CS-processing condition (13,0,0), which focussed on MSR and LSR fibre loss only, did not yield consistent improvement for all subjects, but the 10,0,0 and 7,0,0 conditions systematically improved EFR magnitudes and AM thresholds. The latter two conditions were designed to compensate for selective loss of all ANF types and applied more pronounced processing to the stimulus envelope. Furthermore, statistically significant correlations were found between AM detection thresholds and SAM EFR magnitudes (Supplementary Table 7, Fig. 9c), showing that the AM detection performance of a subject (especially in the 7,0,0m condition) can predict the strength of the recorded EFR to the same stimulus paradigm (and vice versa). The speech EFRs provided some indication of the measured SI score improvement in natural speech (Fig. 9a,b), but showed insignificant correlations to AM thresholds and SRTs (Supplementary Table 6). Since the speech-EFR strength reflects the contribution of the fundamental frequency and first formants of speech, the EFR magnitude varies naturally and could thus prove to be a poor marker [67].

Speech recognition in the Matrix sentence test showed a small, but inconsistent, improvement across participants after processing: The yNH group and those with strong speech EFR magnitudes or high AM detection sensitivity benefitted the most from processed speech (Fig. 9). The latter aspect goes in line with the findings in [18, 68], where individuals with AN deficits and poor modulation detection sensitivity performed worse in speech recognition. The evaluated processing strategies had individual effects on the SI benefit, with the 13,0,0 modified strategy (13,0,0m) yielding overall the best WRS results (median improvement

of up to 5% and 8.3% for the yNH group in the recognition of BB and HP-filtered speech, respectively). Although the 13,0,0 modified strategy improved intelligibility in most cases (Fig. 8), similar benefits were not encountered for the same processing type when applied for the 10,0,0 profile (10,0,0m). At the same time, the 13,0,0m strategy outperformed the 13,0,0 strategy, even though the simulated AN response enhancement was lower in the first case (e.g., Fig 3). Thus, perceptual markers are still necessary in the algorithm design (e.g., in terms of speech intelligibility and quality) to objectively evaluate the CS-compensating strategies alongside the simulated AN response restoration.

To this end, we chose to use the PESQ and STOI metrics for each processing strategy (Supplementary Table 1). The STOI evaluation resulted in higher scores for the modified strategies in comparison to the original strategies, which was in line with our measured SI results (when comparing the 13,0,0m to the 13,0,0 SI scores). Although our simulations suggested that the modified strategies can achieve lower CS restoration, this type of processing preserved the envelope modulation depth and resulted in smoother and more "natural-sounding" speech stimuli than the original processing strategies, which could be the reason for the higher STOI scores. On the other hand, the highest PESQ scores were achieved by the original processing strategies (BB speech), which could be justified by the possible noise suppression that these might achieve as a side effect. Supplementary Table 1 also shows that the clean-envelope 13,0,0 modified strategy (13,0,0m-ref) gave the best scores by improving both PESQ and STOI in BB speech after processing. However, this strategy showed the worst SI results for our subjects, without showing consistent improvements to the SRTs and the WRSs or significant correlations to other metrics. Thus, the evaluation of augmented speech using such objective metrics might be unsuitable, and this may relate to the fact that the PESQ and STOI metrics consider the clean-speech signal as the (ideal) reference. Processed speech would ideally require an evaluation approach that can connect the simulated AN response restoration to predicted intelligibility for the individual HI periphery. To this end, the integration of a backend in the auditory periphery model that can simulate (psychoacoustic) speech-recognition performance [69], or the use of similar human-inspired frameworks [70] might prove helpful in future studies.

Although we performed a short training at the beginning of each SI session (see Methods), stronger (or more consistent) SI improvements may be observed when allowing the subjects to get accustomed to the processed sound for a more extended time period [36]. The results of the last measurement session (WRS) were expected to reflect the performance of listeners that were already more familiar with the processed sounds (taking into account the previous SRT session), but a longer training session or a shorter interval between the sessions might still be necessary. This is better demonstrated in Supplementary Table 9, where the median WRS improvement was computed for our best-performing strategy (13,0,0m) in the two groups. Overall, the SI improvement was more significant after omitting the first four trials of the results (up to 10 % improvement in the noisiest scenario), hence there still seems to be a training effect over the first trials of the session that should be accounted for in future studies.

At the same time, the distortion of speech after processing could also be a reason for the individual differences in SI score benefit. The algorithms rely on instantaneous (fast-acting) processing of the temporal envelope, which has been argued to lead to non-linear distortion and poor sound quality in dynamic-range compression [28, 30, 31, 71–74]. Although the algorithms were optimised purely on the basis of the simulated outcomes of a biophysical model [40], additional signal-processing methods could be applied after the optimisation procedure (e.g., [34, 74, 75]) to minimise the negative effects that fast envelope processing might introduce. A possible extension of the proposed algorithms would be to process the envelope in different frequency bands (multi-channel processing [71]), to target the enhancement of both spectral and temporal contrast of the envelope modulation, and also to compensate for frequency-specific types of CS and/or OHC loss, depending on the individual SNHL type.

The design of our CS-compensating processing function focussed most strongly on stimulus envelope regions that fluctuate the most, i.e., where temporal modulations (intensity contrasts) of the signal are the strongest. As a result, this type of processing mostly improves the voiced parts of speech stimuli, resulting in more enhanced vowels and voiced consonants than unvoiced consonants. At the same time, modification of the temporal modulation characteristics of speech may attenuate short-duration parts that are low in

intensity compared to the rest of the stimulus, such as unvoiced stop consonants or rapid formant transitions between voiced vowels and consonants [34]. This could deteriorate the perception and intelligibility of consonants and introduce difficulties in word identification [76]. In future work, the processing schemes can be improved to enhance the low-intensity, short-duration consonant cues more strongly, instead of enhancing long-duration cues such as vowel formants [34], e.g., by increasing the consonant-vowel intensity ratio [33, 77]. A more thorough analysis of our SI results or a more specific measurement session (e.g., monosyllabic word tests) could provide insight regarding the individual intelligibility enhancement of vowels and consonants after processing.

Overall, the processing conditions showed larger dispersion of the individual results for each of the tests (EFRs, AM detection thresholds, SRTs and WRSs). This indicates that the processing has a strong individual component, with the variability increasing as the processing became stronger (from 13,0,0 to 7,0,0). Because the most pronounced processing condition (7,0,0) produced SAM stimuli that are similar to a rectangular-wave amplitude-modulated (RAM) stimulus [14], the evoked responses are expected to be more sensitive to CS than to OHC deficits [14, 52]. Thus, the individual degree of CS could be a possible explanation for the spread of the physiological and behavioural results as the processing became more pronounced. Listeners with higher degrees of CS (and hence smaller unprocessed EFR markers) may be those who benefitted less from CS-compensating processing, however this was not clear from the results of our statistical analysis. To this end, diagnostic assessments of the individual degree of CS in each participant [48, 49] could help to explain the individual performance differences between participants with more certainty.

6 Conclusion

Our proposed hearing-restoration strategies go beyond conventional hearing-aid algorithms that typically apply dynamic-range compression to compensate for OHC loss. Instead, our type of processing either preserved or extended the dynamic range of the temporal speech envelope, by increasing the dynamic range of the simulated CS-affected responses to compensate for selective losses of ANFs (CS). The envelope processing was designed on the basis of SAM tones and afterwards extended to HP-filtered and natural (BB) speech, targeting the enhancement of temporal modulation in the periodicity envelope. The proposed CS compensation was able to significantly improve peripheral coding in both young and older NH groups, enhancing their EFRs and AM detection performance. The strategies were not able to improve the SI of all participants, but benefitted those with strong EFRs or good behavioural performance, especially among the yNH group. For SNRs at 3 dB below the individual SRTs, our best-performing algorithm improved word recognition of natural speech by 5% in the yNH group and by 3.3% in the oNH group.

Although our proposed hearing restoration was based on the compensation of AN responses in CS-affected peripheries, the resulting algorithms were able to improve temporal-envelope processing for listeners both with and without suspected age-related CS. This could lead to restored hearing for listeners with CS pathologies or augmented hearing for NH listeners, allowing for speech-enhancement applications in acoustic scenarios where intelligibility is known to suffer (e.g., train-station announcements, smart phones, hearables). At the same time, the processing algorithms were easy to implement and fast to execute, and were used online in our measurements without adding significant delays between trials. Thus, our proposed type of sound processing may extend the application range of present-day hearing aids and improve temporal-envelope processing, while leaving sound amplification unaffected.

7 Methods

Using a biophysically-inspired model of the auditory periphery [40], we applied a numerical method that iteratively altered a sound-processing function so as to minimise the difference between simulated auditory CS responses and NH responses (Fig. 1). In each iteration, the AN population response to the modified stimulus was computed and compared to the reference response, until an error measure was sufficiently

minimised. An iterative formula was used:

$$x[n] = x[n] + \eta \cdot x_s[n], \qquad x_s[n] = \frac{e[n-1] \cdot x_s[n-1]}{e[n] - e[n-1]} - x_s[n-1], \tag{1}$$

where n is the iteration number, x the unknown parameter and x_s its respective step, η the learning rate and e the error measure to be minimised. A learning rate $\eta = 0.8$ and an initial step $x_s = 0.1$ were used, and the procedure was repeated for each unknown parameter until the step x_s was lower than 0.01.

7.1 Sound processing for fully-modulated SAM tones

The optimisation was performed for 50-ms-long SAM tonal stimuli (carrier frequency $f_c = 4$ kHz, modulation frequency $f_m = 120$ Hz and $m = 100\%$ modulation) of levels from 0 to 120 dB SPL and a step size of 1 dB SPL. Each time, the error measure e was defined as the mean-squared error (MSE) between the peak AN population responses, computed at 37-42 ms after the stimulus onset to ensure a steady-state response. The maximum values of the two responses were computed over this region and 10 samples before and after their respective maxima (± 0.5 ms) were considered for the MSE difference, accounting for possible phase differences between the two responses.

The iterative formula of Eq. *1* was used to independently derive the parameters a, b and c of a non-linear envelope-processing function g for each level:

$$g[n] = \frac{a}{1 + e^{-b \cdot E[n] + c}}, \tag{2}$$

where E is the envelope of the input auditory signal x and n is each time sample. During the optimisation, the Hilbert envelope was computed for the SAM tones, corresponding to the amplitude fluctuations of the modulating tone signal ($f_m = 120$ Hz). The parameters a, b and c were initialised at 1, 87.5, and 6.25, respectively. The optimisation procedure was applied twice to ensure that an optimal combination of the three independently-defined parameters was derived. The values of the variables a, b and c define the shape of the non-linear function: a is the variable determining the maximum gain applied, b defines the slope of the function and c the offset (lower operating point) of the function. Although the slope of the non-linear function (variable b) was kept constant for each CS profile, different values of the parameters a and c were computed for each stimulus level to achieve level-dependent processing.

After defining the optimal parameter values for each stimulus level, two exponential functions were fitted to the acquired values of the parameters a and c. These functions were selected based on the exponential shapes of the derived parameter values across level. The formulae for computing all the parameters of Eq. *2* are given below, expressed as functions of the stimulus-envelope peak. The fitted parameter values used in the formulae are given in Supplementary Table 2.

$$a(\max E) = \begin{cases} 10^{a_1 + \exp[-a_2 * log_{10}(\max E) - a_3]} & \text{if } \max E < a_4 \\ 1 & \text{if } \max E \geq a_4 \end{cases} \tag{3}$$

$$b(\max E) = \frac{b_1}{\max E} \tag{4}$$

$$c(\max E) = c_1 + \frac{c_2}{1 + \exp[-c_3 \cdot log_{10}(\max E) - c_4]} \tag{5}$$

The three formulae can be used to apply level-dependent processing to a SAM tone based on the peak (maximum value) of its envelope, but they can also be adjusted to use the respective RMS value instead. Here, we focussed on applying our processing to 70 dB-SPL stimuli and ensuring that the responses to supra-threshold stimuli can be enhanced after processing. Supplementary Table 3 shows the values of the three parameters for each chosen CS profile at 70 dB SPL. Depending on the CS profile, the respective values of the variables a, b and c can be used directly in Eq. *2* to compute the non-linear function g. The function g

can then be multiplied by the stimulus to obtain a processed signal that can compensate for the specific CS profile at 70 dB SPL. The parameter values of Supplementary Table 3 were used in the non-linear processing algorithms of each CS profile for our simulated (Sec. 2) and experimental (Sec. 3) evaluation.

7.2 Sound processing for partially-modulated stimuli

To apply the processing functions to SAM tones of <100% modulation without affecting the modulation depth, we modified the range of the non-linear function to have a different lower operating point. First, the envelope E of the modulated signal was scaled to remove the envelope offset (modulation depth):

$$E_{scaled}[n] = \frac{(E[n] - \min E) \cdot \max E}{\max (E - \min E)} \tag{6}$$

Then, the processing function of Eq. 2 was applied to the scaled envelope E_{scaled} instead:

$$\hat{E}_{scaled}[n] = g_{scaled}[n] \cdot E_{scaled}[n], \qquad g_{scaled}[n] = \frac{a}{1 + e^{-b \cdot E_{scaled}[n] + c}} \tag{7}$$

The operation of Eq. 6 was reversed to rescale the processed envelope $\hat{E}_{scaled}[n]$ back to the range of the envelope E of the original signal:

$$E_m[n] = \frac{\hat{E}_{scaled}[n] \cdot \max (E - \min E)}{\max E} + \min E \tag{8}$$

The resulting envelope E_m corresponds to the desired processed envelope, which differs from the original envelope E only in the regions between the envelope peaks and troughs. By diving the derived envelope E_m by the original envelope E we get the desired processing function:

$$g_m[n] = \frac{E_m[n]}{E[n]} \tag{9}$$

The modified processing function g_m can thus be applied to a modulated signal and preserve the original modulation depth of its envelope (Supplementary Fig. 1). By using the modified function g_m, the processing is applied solely to the slope of the modulation envelope. However, as shown in Supplementary Fig. 1, the processed stimuli are unable to fully restore the CS-affected AN responses to the NH responses. As the modulation depth decreases, the processing of the envelope gradually becomes less effective, until the envelope becomes flat ($m = 0$) and our algorithms apply no processing to the signals. It should be noted that in the case of fully-modulated stimuli, both the original and the modified functions apply the same processing.

7.3 Speech processing

The evaluation of the algorithms for speech was performed using the material of the Flemish Matrix test [56]. The Flemish Matrix includes 260 sentences recorded by a female speaker, each comprised of a 5-word combination from a closed set of 50 Flemish words. To generate the HP speech material, the (BB) Flemish Matrix sentences were HP filtered (zero-phase digital filtering) using an FIR filter with an order of 1024 and a cutoff frequency $f_{cut} = 1650$ Hz (as in [59, 60]). Then, speech-shaped noise (SSN) material was generated from the long-term average speech spectrum of all sentences (as in [56]), separately for the BB and HP-filtered material. In each case, the BB SSN generated was used for the evaluation of BB speech in noise and the HP SSN for the evaluation of HP-filtered speech in noise.

To apply our envelope processing to speech, we performed an automatic RMS-based envelope estimation from the speech waveform using a running RMS window of 25 samples (∼0.57 ms for a sampling frequency $f_s = 44{,}100$ Hz). The estimated envelope was then smoothed using a moving-window function of the same size (25 samples), to remove the high-frequency fluctuations of the envelope and capture the speech pitch

periodicity [20]. To find the operating points of the non-linear functions, we used the peaks of the envelope that had an amplitude larger than 1.15 times the amplitude of their neighbouring samples. The corresponding troughs between each pair of peaks were then computed from the local minima of the envelope between peaks, and the envelope was normalised to the local maxima of the signal. Finally, based on the estimated peak and trough positions, the attenuation function g or g_m (Eqs. 2,9) was computed between each pair of troughs using the normalised envelope, and was multiplied by the original stimulus to generate the envelope-processed speech.

The aforementioned process was applied in the same way for quiet or noisy stimuli and was visualised in Fig. 3a for a speech-in-noise stimulus. In the case of stimuli in noise, the two processing functions were directly applied to the estimated envelope of the noisy stimulus without having access to the clean stimulus or envelope (blind processing). However, we also included a reference condition that used the estimated envelope of the clean stimulus (ref). In this case, the envelope-processing function was computed across time from the peaks and troughs of the clean envelope and then applied to the noisy signal to derive the processed version for the desired CS profile (abbreviated as 13,0,0m-ref, 10,0,0m-ref and 7,0,0m-ref for the modified strategy).

7.4 Evaluation

For the evaluation of our algorithms, participants were recruited into two groups: 16 young normal-hearing (yNH: 22.5 ± 2.3 years, 9 female) and 15 older normal-hearing (oNH: 51.3 ± 4.4 years, 12 female). Volunteers with a history of hearing pathology or ear surgery were excluded as per a recruitment questionnaire. Audiograms were measured in a double-walled, sound-attenuating booth, using an Interacoustics Equinox audiometer. The hearing thresholds of the participants were assessed at 12 standard audiometric frequencies between 0.125 and 16 kHz. The stimuli were presented monaurally to the best ear, determined on the basis of their audiogram and tympanogram. Audiometric thresholds were below 20 dB hearing-loss (HL) for frequencies up to 4 kHz in the yNH group (Supplementary Fig. 2). In the oNH group, audiometric thresholds were ≤ 20 dB HL for frequencies up to 4 kHz for all subjects except for four subjects whose results were tagged by a darker colour.

The experimental data were collected in a sound-proof and electrically-shielded booth and the sounds were presented monaurally via Etymotic ER-2 insert earphones connected to a Fireface UCX external sound card (RME) and a TDT-HB7 headphone driver. The EEG recordings were conducted while subjects were sitting in a comfortable chair with a head rest and were watching muted movies. The processing algorithms were implemented in MATLAB and the participants' psychoacoustic responses were collected using the AFC toolbox for MATLAB [78]. Participants were informed about the experimental procedure according to the ethical guidelines at Ghent University and Ghent University Hospital (UZ-Gent), and were paid for their participation. Written, informed consent was obtained from all participants.

An overview of the three measurement sessions is shown in Supplementary Table 4, with each session explained in detail in the following sections. Although all participants completed the three sessions, incomplete or corrupt data were found for some participants among the oNH group: subjects 1 and 3 had incomplete AM detection data, subjects 1, 3 and 4 did not perform the SRT processed measurement, and subject 5 was missing WRS data for the HP-filtered speech conditions. The EFR, AM detection, SRT and WRS results in Figures 5-8 are shown without these missing data.

7.5 EEG measurements

For the first part of the EEG session, a SAM tone with carrier frequency $f_c = 4$ kHz, modulation frequency $f_m = 120$ Hz and $m = 100\%$ modulation was generated in MATLAB and presented using PLAYREC with a sampling rate of 48 kHz. The processing algorithms were implemented in MATLAB and were applied online to the stimulus before presenting them to the listener. The stimuli were 500-ms long and repeated 1000 times with alternating polarities. A uniformly-distributed, random, silence jitter was applied between consecutive epochs (100 ± 10 ms) of the 1000 stimulus presentations.

The calibration of the SAM stimuli was carried out based on their peak amplitudes (RMS of the carrier tone), rather than the RMS of each modulated stimulus separately. The reason for avoiding re-calibrating to the RMS was that, after processing the modulated stimulus using our algorithms (Fig. 2), we could keep the same peak amplitude among the different conditions (before and after processing) and thus make a fair comparison between their envelopes and responses. To this end, the $f_c = 4$ kHz pure-tone carrier that was used to generate the unprocessed SAM tone was calibrated to 70 dB SPL using a B&K sound-level meter type 2606. Then, the calibrated pure tone was modulated to generate the SAM stimulus, which had an RMS of 71.76 dB SPL, and RMS values of 70.27, 70.05 and 68.91 dB SPL after processing with the 13,0,0, 10,0,0 and 7,0,0 conditions, respectively.

For the second part of the session, EEG responses were recorded to the HP-filtered word 'David', before and after processing. It has been shown that high-frequency ANFs do not phase lock to the fine-structure content of the sound, but rather follow the envelope changes of the voiced portion of speech [79]. Thus, EFRs to HP-filtered speech stimuli stem from the contribution of mid-frequency fibres and their periodic bursts that follow the fundamental frequency of the stimulus. To this end, the word 'David' was selected because of its length and spectral properties, with the voiced parts (/a/ and /i/) containing mostly energy at the fundamental frequency and its harmonics. We isolated the word from a Flemish Matrix sentence [56] and used Praat [80] to monotonise its fundamental frequency f_0 to ∼220 Hz. Then, the stimulus was HP filtered in the same way as before ($f_{cut} = 1.65$ kHz) to remove all content related to the modulation (fundamental) frequency. The sentence was extracted from the calibrated material of the Flemish Matrix, which was based on a 70-dB-SPL calibration of the noise material (SSN) of the test. For the speech stimulus that we used, this resulted in an RMS of 67.93 dB SPL for the unprocessed version and in an RMS of 65.72 dB SPL for the 10,0,0 processed version. The stimuli were 500-ms long ($f_s = 44{,}100$ Hz) and were repeated 1200 times with alternating polarities. A uniformly-distributed, random, silence jitter was applied between consecutive epochs (200 ± 20 ms) of the 1200 stimulus presentations.

Scalp-recorded potentials were obtained with a 64-Channel Biosemi EEG recording system and a custom-built trigger-box using a sampling frequency of 16,384 Hz. The electrodes were placed according to the 10-20 standard, using highly conductive gel (Signa gel). The Common-Mode-Sense (CMS) and Driven-Right-Leg (DRL) electrodes were placed on top of the head. Two external electrodes were connected to the earlobes as references. During the EFR analysis, all channels were re-referenced to the average of the two earlobe electrodes, with all EFR results representing the Cz channel recordings.

7.6 EFR analysis

The EFR analysis was performed according to the procedures of [63]. EEG responses were first filtered with an 800th order Blackman window-based FIR filter. Zero-phase digital filtering was applied between 60 and 600 Hz for the SAM responses, and between 60 and 2500 Hz for the speech responses (to preserve the higher harmonics of the f_0 of the speech stimulus). Signals were broken into epochs of 500 ms relative to the trigger onset, and the first 100 ms of each epoch were omitted in the analysis of the SAM responses. Baseline correction was applied by subtracting the mean of each epoch, before epochs were averaged across trials. Local epoch outliers were then determined using a moving median window of 200 samples (epochs). The peak-to-peak amplitudes of epochs that exceeded the median threshold, corresponding to three times the median absolute deviation, were rejected. An average of 66.4 ± 36.4 epochs were rejected for the recorded SAM conditions of all participants (6.64 ± 3.64 %), and an average of 92.4 ± 53 epochs for the recorded speech conditions (7.7 ± 4.42 %).

The bootstrapping approach proposed in [81] was employed to estimate the noise-floor (NF) component and remove it from the averaged trials. The bootstrapping procedure was repeated 400 times, resulting in 400 NF-corrected EFR spectra. Each time, n_e epochs were drawn randomly with replacement from the epochs after rejection, where n_e is the number of remaining epochs after removing the outliers of each measured condition. The FFT spectra were computed with $n_{FFT} = 8192$ points and the NF component was estimated by repeating the resampling procedure 800 times (with equal numbers of polarities in the draw). A more

detailed explanation of the bootstrapping procedure can be found in [49].

After applying the bootstrapping approach, the summed magnitude of the EFR was computed from each of the 400 NF-corrected EFR spectra using:

$$EFR_{summed} = \sum_{i=0}^{N=3} M_{f_i}, \tag{10}$$

where M_{f_i} are the EFR spectral-peak (to NF) values at the modulation frequency f_0 of the stimulus and the next $N = 3$ harmonics of f_0. For each bootstrapped EFR spectrum, only the peaks that exceeded one standard deviation of the average NF were considered. Then, the mean and standard deviation across the 400 summed EFR magnitudes was computed to define the mean and variability of the EFR_{summed} metric (Fig. 5).

7.7 AM detection

The AM detection test was implemented as a 3-AFC experiment in the AFC toolbox of MATLAB [78]. Participants were instructed to select the signal that sounded different to the other two, each time choosing between three randomly-presented signals containing two pure tones and a modulated tone of variable modulation depth. An 1-up, 2-down adaptive procedure was used with step sizes of [10, 5, 3, 1] dB and a starting modulation depth of -6 dB. The task was repeated three times for each condition, and the average value and standard deviation of the three trials was calculated (Fig. 6).

In each trial, two 70 dB-SPL, 4-kHz pure tones (reference signals) and a modulated tone (target signal) were generated. The same peak-to-peak calibration as that applied to the EFR stimuli was used, to ensure that the same carrier signal was used among the three signals within each trial. Thus, the modulated tone was calibrated each time to have the same carrier level as the reference pure tones, rather than an RMS calibration that is commonly used in AM detection experiments. As noted above, in the case of a fully-modulated tone calibrated in this way, the difference between the pure-tone and modulated-tone stimuli is less than 2 dB SPL. Even though level cues might be introduced in this worst-case scenario, and during the first trials of the AM detection task, the RMS differences of the modulated stimuli close to the AM detection thresholds are much lower. For the upper limit of our measured AM thresholds (m = 3% / -30.46 dB modulation; Fig. 6), the RMS levels were 70, 69.92, 69.89 and 69.85 dB SPL for the unprocessed, and 13,0,0m, 10,0,0m and 7,0,0m processed stimuli, respectively. Thus, level cues are not expected to have any effect close to the detection thresholds, since the largest difference in this case was ∼0.15 dB, which is well below the perceivable loudness difference at all frequencies (∼0.5 dB [82]).

7.8 SI assessment

A custom implementation of the Flemish Matrix Test [56] on the AFC toolbox [78] was used for assessing SI performance. The first test included a standard SRT measurement in BB and HP-filtered speech (before and after processing), while the second test included a word-recognition measurement among the different processing conditions at fixed SNRs of BB and HP-filtered speech. Before presenting them to the listener, the processing algorithms were applied online to the noisy mixture of each sentence. The processing between each step did not significantly increase the total duration of the test. The BB and HP speech material were calibrated based on a separate 70-dB-SPL calibration of the generated BB and HP SSN, respectively, using a B&K sound-level meter type 2606.

At the beginning of each SI test, each subject was presented with a training double list (20 sentences) comprised of all different speech conditions in a random order, so that the subjects could get familiar with the test and the processed speech. For the SRT measurement of each subject, two double lists for BB speech and two double lists for HP speech were randomly presented. For each double list, the SRT measurement was performed using an adaptive procedure to estimate the SNR level that corresponded to ∼50% word recognition for each subject (SRT). The initial SNR was 0 dB and the noise level was adapted after each

trial to reach the desired SNR. For each measured condition, the last 6 reversals were considered to derive the SRT (and devation) of each participant (Fig. 7), using only the results of the second list for each of the two speech types to account for the training effect of the test [56]. The test was also repeated after applying one processing algorithm (13,0,0m-ref) to the noisy sentences of each trial, with one double list used for each speech type in this case.

During the last SI measurement session, the average WRS was derived for each of the 5 processing conditions (unprocessed, 13,0,0m-ref, 13,0,0, 13,0,0m, 10,0,0m) at 3 SNRs of BB and 3 SNRs of HP-filtered speech. For each participant, the SNRs were individually defined based on the estimated SRT in BB and HP speech (SRT and SRT$\pm$3 dB). Twelve sentences were randomly presented for each processing condition, resulting in three double lists ($5 * 12 = 60$ sentences) for each SNR condition that were randomly selected for each run. Then, the average SI score was computed from the 12 measured trials of each condition, separately for each SNR and speech type (Fig. 8).

7.9 Statistical analysis

Supplementary Tables 5-8 show the values of Spearman's rank-order correlation r for the SAM EFR, speech EFR, AM detection and 13,0,0m WRS improvement results. Correlations were computed between the yNH and oNH groups separately, between all subjects, and finally between all subjects without taking the four oNH subjects that had HL $>$ 20 dB (all w/o oHI) into account. In each table, results are shown only for the combinations for which a statistically significant correlation was found. The correlations with probability values $p < 0.05$, $p < 0.01$, $p < 0.001$ and $p < 0.0001$ are indicated with one, two, three and four asterisks, respectively. The number in the parentheses indicate the degrees of freedom (N-2), where N is the number of points used to compute each correlation. Statistically significant correlations are indicated in bold font. For the EFR conditions, the results of yNH subject 7 were not taken into account in the correlation estimations, since the EFR magnitude was significantly higher than the rest of the subjects for three of the five conditions.

8 Data availability

The evaluation results of our participants will be made available upon request. A public repository containing the whole dataset will be made available upon acceptance of the paper. Figures 5, 6, 7, 8, 9 and Supplementary Fig. 2 in this paper can be reproduced using this repository. The source data underlying the remaining figures will also be made available in the same repository.

9 Code availability

The MATLAB implementations of the presented hearing-restoration algorithms, as well as the scripts that were used to record and analyse the EFRs, can be made available upon request or upon acceptance of the paper. The source code of the auditory periphery model v1.2 is available via `https://doi.org/10.5281/zenodo.3717800` or `https://github.com/HearingTechnology/Verhulstetal2018Model`.

10 Acknowledgments

This work was supported by the European Research Council (ERC) under the Horizon 2020 Research and Innovation Programme (grant agreement No 678120 RobSpear). We thank Heleen Van Der Biest for assisting in the participant recruitment and data collection. English language services were provided by stels-ol.de.

11 Author contributions

F.D.: Conceptualisation, Methodology, Software, Validation, Formal analysis, Investigation, Data Curation, Data Collection, Writing: Original Draft, Visualisation; V.V.: Conceptualisation, Methodology, Experimental design; A.O.V.: Software, Experimental design; T.W.: Conceptualisation, Experimental design; S.V.: Conceptualisation, Resources, Supervision, Project administration, Funding acquisition, Writing: Review, Editing.

References

1. Cunningham, L. L. & Tucci, D. L. Hearing loss in adults. *New England Journal of Medicine* **377**, 2465–2473 (2017).

2. Rubel, E. W., Furrer, S. A. & Stone, J. S. A brief history of hair cell regeneration research and speculations on the future. *Hearing research* **297**, 42–51 (2013).

3. Kujawa, S. G. & Liberman, M. C. Adding insult to injury: cochlear nerve degeneration after "temporary" noise-induced hearing loss. *Journal of Neuroscience* **29**, 14077–14085 (2009).

4. Sergeyenko, Y., Lall, K., Liberman, M. C. & Kujawa, S. G. Age-related cochlear synaptopathy: an early-onset contributor to auditory functional decline. *Journal of Neuroscience* **33**, 13686–13694 (2013).

5. Furman, A. C., Kujawa, S. G. & Liberman, M. C. Noise-induced cochlear neuropathy is selective for fibers with low spontaneous rates. *Journal of neurophysiology* **110**, 577–586 (2013).

6. Parthasarathy, A. & Kujawa, S. G. Synaptopathy in the aging cochlea: Characterizing early-neural deficits in auditory temporal envelope processing. *Journal of Neuroscience* **38**, 7108–7119 (2018).

7. Wu, P. *et al.* Primary neural degeneration in the human cochlea: evidence for hidden hearing loss in the aging ear. *Neuroscience* **407**, 8–20 (2019).

8. Shaheen, L. A., Valero, M. D. & Liberman, M. C. Towards a diagnosis of cochlear neuropathy with envelope following responses. *Journal of the Association for Research in Otolaryngology* **16**, 727–745 (2015).

9. Ruggles, D., Bharadwaj, H. & Shinn-Cunningham, B. G. Normal hearing is not enough to guarantee robust encoding of suprathreshold features important in everyday communication. *Proceedings of the National Academy of Sciences* **108**, 15516–15521 (2011).

10. Bharadwaj, H. M., Verhulst, S., Shaheen, L., Liberman, M. C. & Shinn-Cunningham, B. G. Cochlear neuropathy and the coding of supra-threshold sound. *Frontiers in systems neuroscience* **8**, 26 (2014).

11. Mepani, A. M. *et al.* Envelope following responses predict speech-in-noise performance in normal-hearing listeners. *Journal of Neurophysiology* **125**, 1213–1222 (2021).

12. Plack, C. J., Barker, D. & Prendergast, G. Perceptual consequences of "hidden" hearing loss. *Trends in hearing* **18**, 2331216514550621 (2014).

13. Parthasarathy, A., Bartlett, E. L. & Kujawa, S. G. Age-related changes in neural coding of envelope cues: peripheral declines and central compensation. *Neuroscience* **407**, 21–31 (2019).

14. Vasilkov, V., Garrett, M., Mauermann, M. & Verhulst, S. Enhancing the sensitivity of the envelope-following response for cochlear synaptopathy screening in humans: the role of stimulus envelope. *Hearing Research* **400**, 108132 (2021).

15. Lopez-Poveda, E. A. & Barrios, P. Perception of stochastically undersampled sound waveforms: a model of auditory deafferentation. *Frontiers in neuroscience* **7**, 124 (2013).

16. Resnik, J. & Polley, D. B. Cochlear neural degeneration disrupts hearing in background noise by increasing auditory cortex internal noise. *Neuron* **109**, 984–996 (2021).

17. Drullman, R., Festen, J. M. & Plomp, R. Effect of reducing slow temporal modulations on speech reception. *The Journal of the Acoustical Society of America* **95**, 2670–2680 (1994).

18. Rance, G., McKay, C. & Grayden, D. Perceptual characterization of children with auditory neuropathy. *Ear and hearing* **25**, 34–46 (2004).

19. Plomp, R. The role of modulation in hearing. In *Hearing—Physiological bases and psychophysics*, 270–276 (Springer, 1983).

20. Rosen, S. Temporal information in speech: acoustic, auditory and linguistic aspects. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* **336**, 367–373 (1992).

21. Shannon, R. V., Zeng, F.-G., Kamath, V., Wygonski, J. & Ekelid, M. Speech recognition with primarily temporal cues. *Science* **270**, 303–304 (1995).

22. Smith, Z. M., Delgutte, B. & Oxenham, A. J. Chimaeric sounds reveal dichotomies in auditory perception. *Nature* **416**, 87–90 (2002).

23. Samuel, A. G. & Kraljic, T. Perceptual learning for speech. *Attention, Perception, & Psychophysics* **71**, 1207–1218 (2009).

24. Füllgrabe, C., Stone, M. A. & Moore, B. C. Contribution of very low amplitude-modulation rates to intelligibility in a competing-speech task. *The Journal of the Acoustical Society of America* **125**, 1277–1280 (2009).

25. Kubanek, J., Brunner, P., Gunduz, A., Poeppel, D. & Schalk, G. The tracking of speech envelope in the human cortex. *PloS one* **8**, e53398 (2013).

26. Decruy, L., Vanthornhout, J. & Francart, T. Hearing impairment is associated with enhanced neural tracking of the speech envelope. *Hearing Research* **393**, 107961 (2020).

27. Souza, P. E., Ellis, G., Marks, K., Wright, R. & Gallun, F. Does the speech cue profile affect response to amplitude envelope distortion? *Journal of Speech, Language, and Hearing Research* **64**, 2053–2069 (2021).

28. Drullman, R. Temporal envelope and fine structure cues for speech intelligibility. *The Journal of the Acoustical Society of America* **97**, 585–592 (1995).

29. Verschuure, J. *et al.* Compression and its effect on the speech signal. *Ear and Hearing* **17**, 162–175 (1996).

30. Hohmann, V. & Kollmeier, B. The effect of multichannel dynamic compression on speech intelligibility. *The Journal of the Acoustical Society of America* **97**, 1191–1195 (1995).

31. van Buuren, R. A., Festen, J. M. & Houtgast, T. Compression and expansion of the temporal envelope: Evaluation of speech intelligibility and sound quality. *The Journal of the Acoustical Society of America* **105**, 2903–2913 (1999).

32. Narne, V. K. & Vanaja, C. Perception of speech with envelope enhancement in individuals with auditory neuropathy and simulated loss of temporal modulation processing. *International journal of audiology* **48**, 700–707 (2009).

33. Geurts, L. & Wouters, J. Enhancing the speech envelope of continuous interleaved sampling processors for cochlear implants. *The Journal of the Acoustical Society of America* **105**, 2476–2484 (1999).

34. Vandali, A. E. Emphasis of short-duration acoustic speech cues for cochlear implant users. *The Journal of the Acoustical Society of America* **109**, 2049–2061 (2001).

35. Milczynski, M., Wouters, J. & Van Wieringen, A. Improved fundamental frequency coding in cochlear implant signal processing. *The Journal of the Acoustical Society of America* **125**, 2260–2271 (2009).

36. Francart, T., Osses, A. & Wouters, J. Speech perception with f0mod, a cochlear implant pitch coding strategy. *International journal of audiology* **54**, 424–432 (2015).

37. Azadpour, M. & Smith, R. L. Enhancing speech envelope by integrating hair-cell adaptation into cochlear implant processing. *Hearing research* **342**, 48–57 (2016).

38. Steinmetzger, K. & Rosen, S. The role of envelope periodicity in the perception of masked speech with simulated and real cochlear implants. *The Journal of the Acoustical Society of America* **144**, 885–896 (2018).

39. Lorenzi, C., Gilbert, G., Carn, H., Garnier, S. & Moore, B. C. Speech perception problems of the hearing impaired reflect inability to use temporal fine structure. *Proceedings of the National Academy of Sciences* **103**, 18866–18869 (2006).

40. Verhulst, S., Altoe, A. & Vasilkov, V. Computational modeling of the human auditory periphery: Auditory-nerve responses, evoked potentials and hearing loss. *Hearing research* **360**, 55–75 (2018).

41. Scollie, S. *et al.* The desired sensation level multistage input/output algorithm. *Trends in amplification* **9**, 159–197 (2005).

42. Keidser, G., Dillon, H., Flax, M., Ching, T. & Brewer, S. The nal-nl2 prescription procedure. *Audiology research* **1**, 88–90 (2011).

43. Kates, J. M. & Arehart, K. H. The hearing-aid speech perception index (haspi). *Speech Communication* **65**, 75–93 (2014).

44. Moore, B. C., Glasberg, B. R. & Stone, M. A. Development of a new method for deriving initial fittings for hearing aids with multi-channel compression: Cameq2-hf. *International Journal of Audiology* **49**, 216–227 (2010).

45. Smeds, K. *et al.* Proprietary hearing aid gain prescriptions: Changes over time. *Hearing Review* **22**, 16–22 (2015).

46. Marriage, J. E., Vickers, D. A., Baer, T., Glasberg, B. R. & Moore, B. C. Comparison of different hearing aid prescriptions for children. *Ear and hearing* **39**, 20–31 (2018).

47. Lin, H. W., Furman, A. C., Kujawa, S. G. & Liberman, M. C. Primary neural degeneration in the guinea pig cochlea after reversible noise-induced threshold shift. *Journal of the Association for Research in Otolaryngology* **12**, 605–616 (2011).

48. Buran, B. N., McMillan, G. P., Keshishzadeh, S., Verhulst, S. & Bramhall, N. Predicting synapse counts in living humans by combining computational models with auditory physiology (2020).

49. Keshishzadeh, S., Garrett, M. & Verhulst, S. Towards personalized auditory models: Predicting individual sensorineural hearing-loss profiles from recorded human auditory physiology. *Trends in Hearing* **25**, 2331216520988406 (2021).

50. Bharadwaj, H. M., Masud, S., Mehraei, G., Verhulst, S. & Shinn-Cunningham, B. G. Individual differences reveal correlates of hidden hearing deficits. *Journal of Neuroscience* **35**, 2161–2172 (2015).

51. Keshishzadeh, S., Vasilkov, V. & Verhulst, S. Tonotopic sensitivity to supra-threshold hearing deficits of the envelope following response evoked by broadband stimuli. In *23rd International Congress on Acoustics (ICA 2019)*, 6513–6518 (Deutsche Gesellschaft für Akustik, 2019).

52. Garrett, M. *et al.* Speech-in-noise intelligibility difficulties with age: The role of cochlear synaptopathy. *bioRxiv* (2020).

53. Liberman, M. C. Auditory-nerve response from cats raised in a low-noise chamber. *The Journal of the Acoustical Society of America* **63**, 442–455 (1978).

54. Van Canneyt, J., Hofmann, M., Wouters, J. & Francart, T. The effect of stimulus envelope shape on the auditory steady-state response. *Hearing research* **380**, 22–34 (2019).

55. Dreyer, A. & Delgutte, B. Phase locking of auditory-nerve fibers to the envelopes of high-frequency sounds: implications for sound localization. *Journal of neurophysiology* **96**, 2327–2341 (2006).

56. Luts, H., Jansen, S., Dreschler, W. & Wouters, J. Development and normative data for the flemish/dutch matrix test (2014).

57. Ardoint, M. & Lorenzi, C. Effects of lowpass and highpass filtering on the intelligibility of speech based on temporal fine structure or envelope cues. *Hearing research* **260**, 89–95 (2010).

58. Ardoint, M., Agus, T., Sheft, S. & Lorenzi, C. Importance of temporal-envelope speech cues in different spectral regions. *The Journal of the Acoustical Society of America* **130**, EL115–EL121 (2011).

59. Verhulst, S. & Warzybok, A. Contributions of low-and high-frequency sensorineural hearing deficits to speech intelligibility in noise. *bioRxiv* 358127 (2018).

60. Maele, T. V. *et al.* The variability in potential biomarkers for cochlear synaptopathy after recreational noise exposure. *bioRxiv* (2021).

61. Rix, A. W., Beerends, J. G., Hollier, M. P. & Hekstra, A. P. Perceptual evaluation of speech quality (pesq)-a new method for speech quality assessment of telephone networks and codecs. In *2001 IEEE international conference on acoustics, speech, and signal processing. Proceedings (Cat. No. 01CH37221)*, vol. 2, 749–752 (IEEE, 2001).

62. Taal, C. H., Hendriks, R. C., Heusdens, R. & Jensen, J. An algorithm for intelligibility prediction of time–frequency weighted noisy speech. *IEEE Transactions on Audio, Speech, and Language Processing* **19**, 2125–2136 (2011).

63. Keshishzadeh, S., Garrett, M., Vasilkov, V. & Verhulst, S. The derived-band envelope following response and its sensitivity to sensorineural hearing deficits. *Hearing Research* **392**, 107979 (2020).

64. Moore, B. C. & Glasberg, B. R. Temporal modulation transfer functions obtained using sinusoidal carriers with normally hearing and hearing-impaired listeners. *The Journal of the Acoustical Society of America* **110**, 1067–1073 (2001).

65. Kohlrausch, A., Fassel, R. & Dau, T. The influence of carrier level and frequency on modulation and beat-detection thresholds for sinusoidal carriers. *The Journal of the Acoustical Society of America* **108**, 723–734 (2000).

66. Johannesen, P. T., Buzo, B. C. & Lopez-Poveda, E. A. Evidence for age-related cochlear synaptopathy in humans unconnected to speech-in-noise intelligibility deficits. *Hearing research* **374**, 35–48 (2019).

67. Easwar, V., Banyard, A., Aiken, S. J. & Purcell, D. W. Phase-locked responses to the vowel envelope vary in scalp-recorded amplitude due to across-frequency response interactions. *European Journal of Neuroscience* **48**, 3126–3145 (2018).

68. Zeng, F.-G., Oba, S., Garde, S., Sininger, Y. & Starr, A. Temporal and speech processing deficits in auditory neuropathy. *Neuroreport* **10**, 3429–3435 (1999).

69. Osses Vecchi, A., Ernst, F. & Verhulst, S. Hearing-impaired sound perception: What can we learn from a biophysical model of the human auditory periphery? In *23rd International Congress on Acoustics (ICA 2019)*, 678–685 (Deutsche Gesellschaft für Akustik, 2019).

70. Schädler, M. R., Warzybok, A., Ewert, S. D. & Kollmeier, B. A simulation framework for auditory discrimination experiments: Revealing the importance of across-frequency processing in speech perception. *The journal of the acoustical society of America* **139**, 2708–2722 (2016).

71. Plomp, R. The negative effect of amplitude compression in multichannel hearing aids in the light of the modulation-transfer function. *The Journal of the Acoustical Society of America* **83**, 2322–2327 (1988).

72. Stone, M. A. & Moore, B. C. Side effects of fast-acting dynamic range compression that affect intelligibility in a competing speech task. *The Journal of the Acoustical Society of America* **116**, 2311–2323 (2004).

73. Herzke, T. & Hohmann, V. Effects of instantaneous multiband dynamic compression on speech intelligibility. *EURASIP Journal on Advances in Signal Processing* **2005**, 1–10 (2005).

74. Stone, M. A. & Moore, B. C. Effects of spectro-temporal modulation changes produced by multichannel compression on intelligibility in a competing-speech task. *The Journal of the Acoustical Society of America* **123**, 1063–1076 (2008).

75. Goldstein, J., Oz, M., Gilchrist, P. & Valente, M. Signal processing strategies and clinical outcomes for gain and waveform compression in hearing aids. In *The Thrity-Seventh Asilomar Conference on Signals, Systems & Computers, 2003*, vol. 1, 391–398 (IEEE, 2003).

76. Stilp, C. E. & Kluender, K. R. Cochlea-scaled entropy, not consonants, vowels, or time, best predicts speech intelligibility. *Proceedings of the National Academy of Sciences* **107**, 12387–12392 (2010).

77. Kennedy, E., Levitt, H., Neuman, A. C. & Weiss, M. Consonant–vowel intensity ratios for maximizing consonant recognition by hearing-impaired listeners. *The Journal of the Acoustical Society of America* **103**, 1098–1114 (1998).

78. Ewert, S. D. Afc—a modular framework for running psychoacoustic experiments and computational perception models. In *Proceedings of the international conference on acoustics AIA-DAGA*, 1326–1329 (2013).

79. Henry, K. S., Kale, S. & Heinz, M. G. Distorted tonotopic coding of temporal envelope and fine structure with noise-induced hearing loss. *Journal of Neuroscience* **36**, 2227–2237 (2016).

80. Boersma, P. Praat, a system for doing phonetics by computer. *Glot. Int.* **5**, 341–345 (2001).

81. Zhu, L., Bharadwaj, H., Xia, J. & Shinn-Cunningham, B. 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 (2013).

82. Miller, G. A. Sensitivity to changes in the intensity of white noise and its relation to masking and loudness. *The Journal of the Acoustical Society of America* **19**, 609–619 (1947).

Supplementary Table 1. Objective speech evaluation of the different processing strategies. The average PESQ [62] and STOI [61] scores were estimated for the Flemish Matrix sentences [56] of the BB and HP conditions, after adding SSN at SNRs of -8 and -3 dB, respectively (roughly corresponding to the average measured SRTs). In each case, the PESQ and STOI scores were computed using the clean sentence as the reference, with values ranging from 0 (very bad) to 5 (very good) and 0 (very bad) to 1 (very good), respectively.

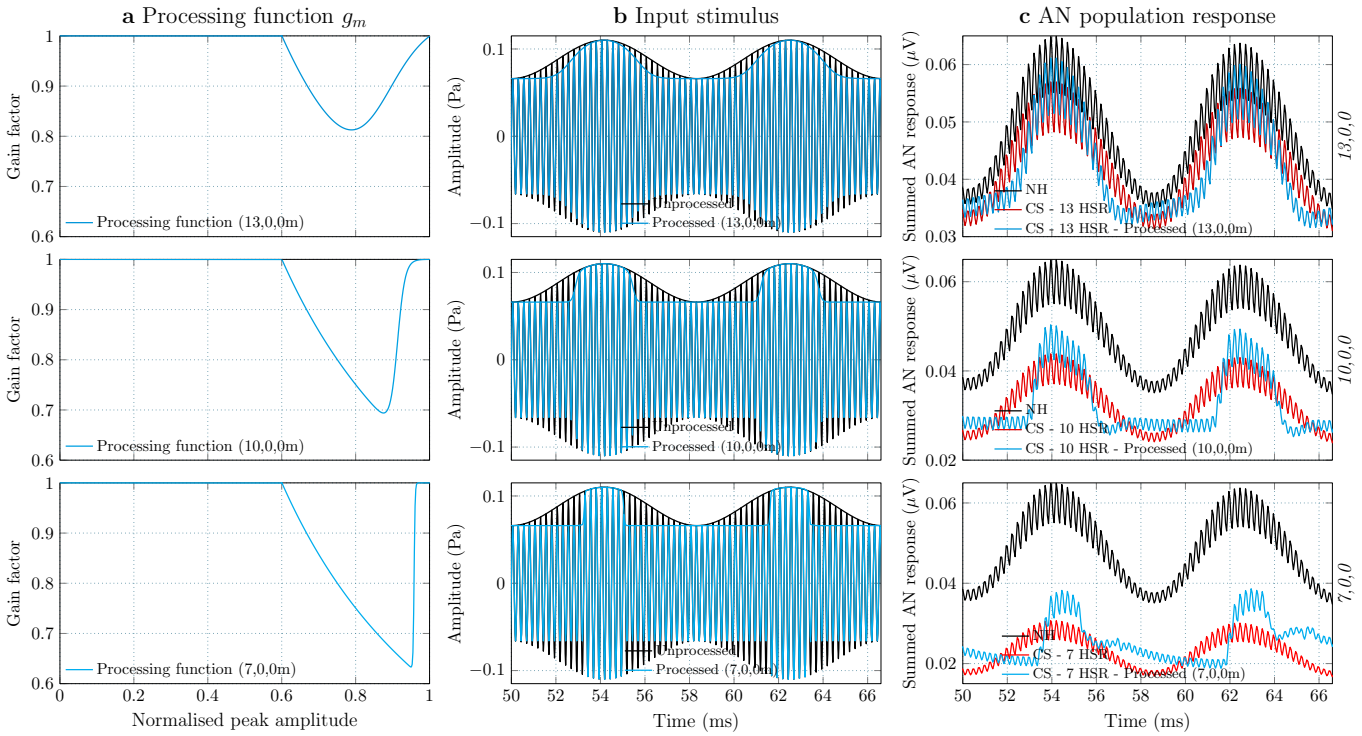
	BB		HP	
	PESQ	STOI	PESQ	STOI
Unprocessed	0.6062	0.5198	1.4806	0.7727
$13, 0, 0$	**0.6193**	0.5085	1.3846	0.7154
$13, 0, 0$m	**0.6145**	0.5148	1.4415	0.7454
$13, 0, 0$m-ref	**0.6140**	**0.5260**	1.4421	0.7237
$10, 0, 0$	0.5954	0.4637	1.2175	0.6691
$10, 0, 0$m	0.5731	0.4822	1.3798	0.7298
$10, 0, 0$m-ref	0.5972	0.5050	1.3509	0.7054
$7, 0, 0$	0.6061	0.4469	1.0706	0.6557
$7, 0, 0$m	0.5105	0.4627	1.3382	0.7294
$7, 0, 0$m-ref	0.5670	0.4754	1.3062	0.7007

Supplementary Table 2. Fitted parameters of Eqs. *3-5* for three CS profiles. The parameters of Eqs. 3-5 were estimated for each CS profile after fitting the functions to the values obtained from the optimisation procedure. The fitted values can be used to compute the parameters a, b and c for each CS profile.

CS profile	a_1	a_2	a_3	a_4	b_1	c_1	c_2	c_3	c_4
$13, 0, 0$	0.0931	2.5804	10.2873	0.1160	7.3543	-95.8223	100.6698	283.8136	312.0264
$10, 0, 0$	-0.1644	0.5473	1.5698	1.6390	39.3868	-42.7324	73.6083	13.8984	17.5337
$7, 0, 0$	-0.4947	0.4224	0.6477	0.2063	291.4200	132.0206	128.2104	20.4308	24.6311

Supplementary Table 3. Parameters for 70 dB-SPL processing. The values of the parameters b and c of Eq. 2 were computed from Eqs. 4,5 for a 70 dB-SPL SAM tone (f_c = 4 kHz, f_m = 120 Hz). The stimulus envelope had a maximum amplitude of 0.14577. Although the computed value for a was close to 1, a constant value of 1 was used across profiles to avoid applying any gain and keep exactly the same peak.

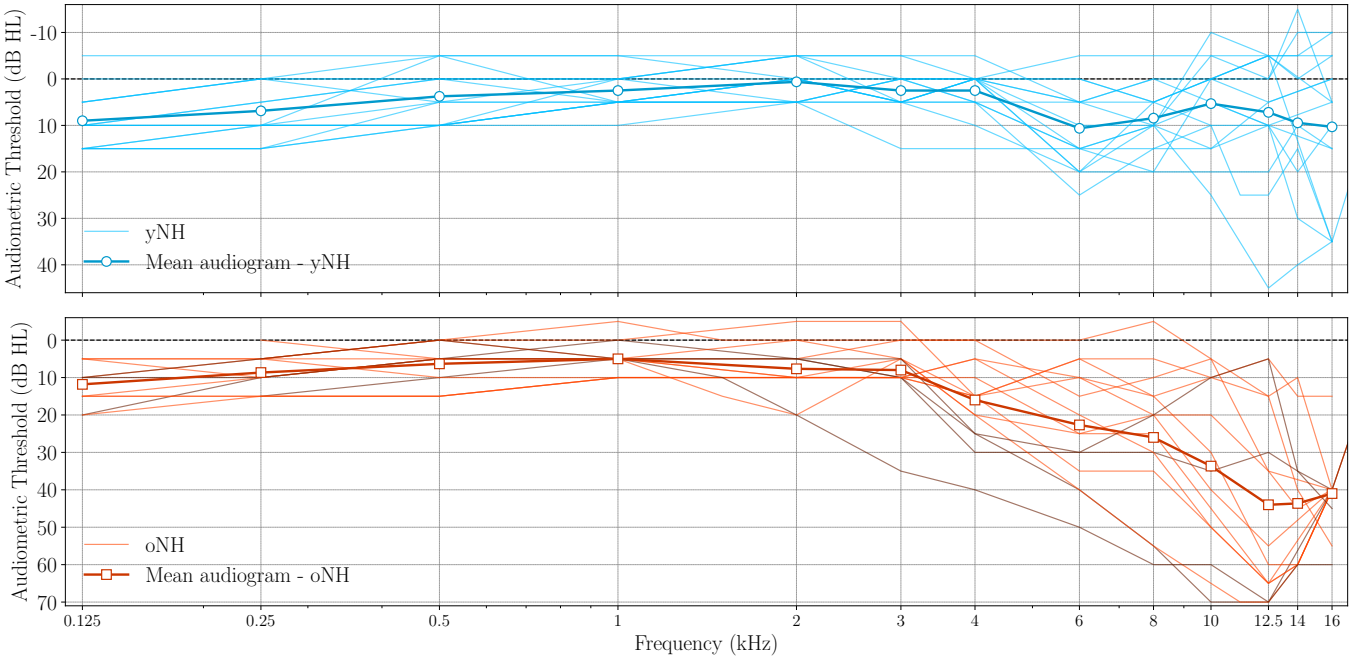
CS profile	a	b	c
$13, 0, 0$	1	50.4500	4.8476
$10, 0, 0$	1	270.1903	30.6769
$7, 0, 0$	1	1999.1160	260.1633



Supplementary Figure 1. CS compensation of a SAM tone without affecting the modulation depth. The modified processing strategies were applied to a SAM tone stimulus to preserve its modulation depth. The results for a 70-dB-SPL SAM tone with -12 dB modulation depth (m = 25 %) are shown. **a** The modified non-linear functions only processed the slope of the envelope, leaving the envelope troughs (and peaks) intact. **b** For each CS type, the non-linear functions were applied to the SAM tones to derive the respective processed stimuli. **c** The simulated AN responses of the NH and CS peripheries are shown in response to the unprocessed SAM stimulus and the respective processed version, given as input to the corresponding CS periphery.

Supplementary Table 4. Overview of the measurement sessions. Our measurement protocol was divided into three sessions, each lasting approximately 1.5 hours. For each session, the conditions evaluated are summarised in brackets.

1st session (EEG)	SAM tone stimulus (unprocessed, 13,0,0, 10,0,0, 7,0,0) "David" speech stimulus (unprocessed, 10,0,0)
2nd session (AM-SRT)	AM detection (unprocessed, 13,0,0m, 10,0,0m, 7,0,0m) SRT measurement for BB and HP speech (unprocessed, 13,0,0m-ref)
3rd session (SI scores)	SI of BB and HP speech (unprocessed, 13,0,0m-ref, 13,0,0, 13,0,0m, 10,0,0m) at 3 fixed SNRs (SRT, SRT-3 dB, SRT+3 dB)



Supplementary Figure 2. Audiograms of participants. The individual audiograms of the participants are shown in thin lines for the two age groups. The thick lines represent the grand-averaged audiogram for each age group. The darker thin lines for the oNH group indicate participants with HL $>$ 20 dB at frequencies up to 4 kHz.

Supplementary Table 5. Statistical analysis of SAM EFRs. Spearman's r values between the EFR results of the four SAM conditions and the remaining metrics. Statistically significant correlations are indicated in bold, with the asterisks corresponding to the probability degree p.

EFR SAM (unprocessed)	yNH	oNH	all	all w/o oHI
AM thresholds (10,0,0m - unprocessed)	$\boldsymbol{r_{(13)}}$ **= 0.66****	$\boldsymbol{r_{(11)}}$ **= -0.59***	$r_{(26)}$ = -0.04	$r_{(22)}$ = 0.29
AM thresholds (7,0,0m - unprocessed)	$\boldsymbol{r_{(13)}}$ **= 0.78*****	$r_{(11)}$ = -0.36	$r_{(26)}$ = 0.10	$\boldsymbol{r_{(22)}}$ **= 0.45***
WRS (13,0,0 - unprocessed) - HP (SRT+3)	$\boldsymbol{r_{(13)}}$ **= -0.57***	$r_{(12)}$ = -0.39	$r_{(27)}$ = -0.22	$r_{(24)}$ = -0.29
WRS (10,0,0m - unprocessed) - HP (SRT+3)	$\boldsymbol{r_{(13)}}$ **= -0.75****	$r_{(12)}$ = -0.50	$\boldsymbol{r_{(27)}}$ **= -0.47****	$\boldsymbol{r_{(24)}}$ **= -0.46***
WRS (13,0,0 - unprocessed) - BB (SRT-3)	$r_{(13)}$ = -0.35	$r_{(13)}$ = -0.18	$\boldsymbol{r_{(28)}}$ **= -0.37***	$r_{(24)}$ = -0.28
WRS (13,0,0 - unprocessed) - HP (SRT-3)	$\boldsymbol{r_{(13)}}$ **= -0.85******	$r_{(12)}$ = -0.25	$r_{(27)}$ = -0.34	$r_{(24)}$ = -0.38
WRS (13,0,0m - unprocessed) - HP (SRT-3)	$\boldsymbol{r_{(13)}}$ **= -0.55***	$r_{(12)}$ = -0.46	$r_{(27)}$ = -0.33	$r_{(24)}$ = -0.34
Audiometric thresholds (4 kHz)	$r_{(13)}$ = 0.00	$\boldsymbol{r_{(13)}}$ **= -0.73****	$\boldsymbol{r_{(28)}}$ **= -0.55****	$r_{(24)}$ = -0.32
EFR SAM (13,0,0)	**yNH**	**oNH**	**all**	**all w/o oHI**
WRS (13,0,0 - unprocessed) - HP (SRT+3)	$\boldsymbol{r_{(13)}}$ **= -0.55***	$r_{(12)}$ = -0.11	$r_{(27)}$ = -0.20	$r_{(24)}$ = -0.23
WRS (10,0,0m - unprocessed) - HP (SRT+3)	$\boldsymbol{r_{(13)}}$ **= -0.69****	$r_{(12)}$ = -0.17	$r_{(27)}$ = -0.33	$r_{(24)}$ = -0.30
WRS (13,0,0 - unprocessed) - HP (SRT-3)	$\boldsymbol{r_{(13)}}$ **= -0.70****	$r_{(12)}$ = -0.18	$r_{(27)}$ = -0.33	$r_{(24)}$ = -0.36
EFR SAM (10,0,0)	**yNH**	**oNH**	**all**	**all w/o oHI**
WRS (10,0,0m - unprocessed) - BB (SRT)	$r_{(13)}$ = -0.26	$\boldsymbol{r_{(13)}}$ **= -0.58***	$r_{(28)}$ = -0.24	$r_{(24)}$ = -0.24
WRS (13,0,0 - unprocessed) - BB (SRT+3)	$r_{(13)}$ = -0.25	$r_{(13)}$ = -0.08	$r_{(28)}$ = -0.17	$\boldsymbol{r_{(24)}}$ **= -0.40***
WRS (13,0,0 - unprocessed) - HP (SRT+3)	$\boldsymbol{r_{(13)}}$ **= -0.67****	$r_{(12)}$ = -0.36	$r_{(27)}$ = -0.11	$r_{(24)}$ = -0.18
WRS (10,0,0m - unprocessed) - HP (SRT+3)	$\boldsymbol{r_{(13)}}$ **= -0.71****	$r_{(12)}$ = -0.39	$r_{(27)}$ = -0.28	$r_{(24)}$ = -0.20
WRS (13,0,0 - unprocessed) - HP (SRT-3)	$\boldsymbol{r_{(13)}}$ **= -0.61***	$r_{(12)}$ = -0.31	$r_{(27)}$ = -0.20	$r_{(24)}$ = -0.19
WRS (13,0,0m - unprocessed) - HP (SRT-3)	$r_{(13)}$ = -0.13	$\boldsymbol{r_{(12)}}$ **= -0.61***	$r_{(27)}$ = -0.15	$r_{(24)}$ = -0.04
Audiometric thresholds (4 kHz)	$r_{(13)}$ = -0.08	$\boldsymbol{r_{(13)}}$ **= -0.69****	$\boldsymbol{r_{(28)}}$ **= -0.61*****	$\boldsymbol{r_{(24)}}$ **= -0.43***
Audiometric thresholds (EHF)	$r_{(13)}$ = 0.13	$r_{(13)}$ = -0.03	$\boldsymbol{r_{(28)}}$ **= -0.40***	$r_{(24)}$ = -0.28
EFR SAM (7,0,0)	**yNH**	**oNH**	**all**	**all w/o oHI**
AM thresholds (unprocessed)	$r_{(13)}$ = -0.51	$r_{(11)}$ = 0.21	$r_{(26)}$ = -0.29	$\boldsymbol{r_{(22)}}$ **= -0.48***
AM thresholds (13,0,0m)	$r_{(13)}$ = -0.45	$r_{(11)}$ = 0.12	$r_{(26)}$ = -0.36	$\boldsymbol{r_{(22)}}$ **= -0.52****
AM thresholds (10,0,0m)	$r_{(13)}$ = -0.39	$r_{(11)}$ = -0.13	$\boldsymbol{r_{(26)}}$ **= -0.45***	$\boldsymbol{r_{(22)}}$ **= -0.52****
AM thresholds (7,0,0m)	$r_{(13)}$ = -0.16	$r_{(11)}$ = -0.23	$\boldsymbol{r_{(26)}}$ **= -0.51****	$\boldsymbol{r_{(22)}}$ **= -0.43***
AM thresholds (10,0,0m - unprocessed)	$r_{(13)}$ = 0.25	$\boldsymbol{r_{(11)}}$ **= -0.64***	$r_{(26)}$ = -0.15	$r_{(22)}$ = 0.10
WRS (10,0,0m - unprocessed) - HP (SRT+3)	$r_{(13)}$ = -0.44	$\boldsymbol{r_{(12)}}$ **= -0.57***	$r_{(27)}$ = -0.32	$r_{(24)}$ = -0.30
WRS (13,0,0 - unprocessed) - BB (SRT-3)	$r_{(13)}$ = -0.17	$r_{(13)}$ = -0.35	$\boldsymbol{r_{(28)}}$ **= -0.47****	$\boldsymbol{r_{(24)}}$ **= -0.43***
WRS (10,0,0m - unprocessed) - BB (SRT-3)	$r_{(13)}$ = -0.26	$r_{(13)}$ = -0.14	$r_{(28)}$ = -0.35	$\boldsymbol{r_{(24)}}$ **= -0.40***
WRS (10,0,0m - unprocessed) - HP (SRT-3)	$r_{(13)}$ = -0.22	$\boldsymbol{r_{(12)}}$ **= -0.58***	$r_{(27)}$ = -0.26	$r_{(24)}$ = -0.16
Audiometric thresholds (4 kHz)	$r_{(13)}$ = -0.08	$\boldsymbol{r_{(13)}}$ **= -0.71****	$\boldsymbol{r_{(28)}}$ **= -0.64*****	$\boldsymbol{r_{(24)}}$ **= -0.46***
Audiometric thresholds (EHF)	$r_{(13)}$ = -0.02	$r_{(13)}$ = -0.45	$\boldsymbol{r_{(28)}}$ **= -0.60*****	$\boldsymbol{r_{(24)}}$ **= -0.49***

Supplementary Table 6. Statistical analysis of speech EFRs. Spearman's r values between the EFR results of the two speech conditions and the remaining metrics. Statistically significant correlations are indicated in bold, with the asterisks corresponding to the probability degree p.

EFR speech (unprocessed)	yNH	oNH	all	all w/o oHI
AM thresholds (unprocessed)	$r_{(13)}$ = -0.31	$\boldsymbol{r_{(11)}}$ **= 0.62***	$r_{(26)}$ = -0.09	$r_{(22)}$ = -0.28
AM thresholds (13,0,0m)	$r_{(13)}$ = -0.32	$r_{(11)}$ = 0.34	$r_{(26)}$ = -0.27	$\boldsymbol{r_{(22)}}$ **= -0.43***
AM thresholds (7,0,0m)	$r_{(13)}$ = 0.11	$r_{(11)}$ = -0.39	$\boldsymbol{r_{(26)}}$ **= -0.46***	$r_{(22)}$ = -0.36
AM thresholds (10,0,0m - unprocessed)	$r_{(13)}$ = 0.30	$\boldsymbol{r_{(11)}}$ **= -0.75****	$r_{(26)}$ = -0.21	$r_{(22)}$ = 0.05
AM thresholds (7,0,0m - unprocessed)	$r_{(13)}$ = 0.44	$\boldsymbol{r_{(11)}}$ **= -0.86*****	$r_{(26)}$ = -0.23	$r_{(22)}$ = 0.03
SRT HP (unprocessed)	$r_{(13)}$ = 0.30	$r_{(13)}$ = 0.24	$r_{(28)}$ = -0.20	$\boldsymbol{r_{(24)}}$ **= -0.40***
SRT HP (13,0,0m-ref - unprocessed)	$r_{(13)}$ = 0.32	$\boldsymbol{r_{(10)}}$ **= -0.64***	$r_{(25)}$ = -0.25	$r_{(21)}$ = 0.03
WRS (13,0,0m - unprocessed) - BB (SRT)	$r_{(13)}$ = 0.19	$r_{(13)}$ = 0.31	$\boldsymbol{r_{(28)}}$ **= 0.44***	$r_{(24)}$ = 0.37
WRS (13,0,0m-ref - unprocessed) - HP (SRT)	$r_{(13)}$ = -0.10	$\boldsymbol{r_{(12)}}$ **= 0.59***	$r_{(27)}$ = 0.21	$r_{(24)}$ = 0.12
WRS (13,0,0m - unprocessed) - BB (SRT+3)	$r_{(13)}$ = 0.42	$r_{(13)}$ = 0.03	$\boldsymbol{r_{(28)}}$ **= 0.41***	$\boldsymbol{r_{(24)}}$ **= 0.40***
WRS (13,0,0 - unprocessed) - BB (SRT-3)	$r_{(13)}$ = -0.22	$r_{(13)}$ = -0.36	$\boldsymbol{r_{(28)}}$ **= -0.47****	$\boldsymbol{r_{(24)}}$ **= -0.43***
WRS (13,0,0 - unprocessed) - HP (SRT-3)	$r_{(13)}$ = -0.60*	$r_{(12)}$ = 0.13	$r_{(27)}$ = 0.02	$r_{(24)}$ = 0.02
WRS (13,0,0m - unprocessed) - HP (SRT-3)	$\boldsymbol{r_{(13)}}$ **= -0.84*****	$r_{(12)}$ = -0.08	$r_{(27)}$ = -0.13	$r_{(24)}$ = -0.11
Audiometric thresholds (4 kHz)	$r_{(13)}$ = -0.10	$\boldsymbol{r_{(13)}}$ **= -0.67****	$\boldsymbol{r_{(28)}}$ **= -0.69******	$\boldsymbol{r_{(24)}}$ **= -0.54****
Audiometric thresholds (EHF)	$r_{(13)}$ = 0.05	$\boldsymbol{r_{(13)}}$ **= -0.58***	$\boldsymbol{r_{(28)}}$ **= -0.68******	$\boldsymbol{r_{(24)}}$ **= -0.60****
EFR speech (10,0,0)	**yNH**	**oNH**	**all**	**all w/o oHI**
AM thresholds (10,0,0m - unprocessed)	$r_{(13)}$ = 0.30	$\boldsymbol{r_{(11)}}$ **= -0.81*****	$r_{(26)}$ = -0.25	$r_{(22)}$ = -0.00
AM thresholds (7,0,0m - unprocessed)	$r_{(13)}$ = 0.39	$\boldsymbol{r_{(11)}}$ **= -0.79****	$r_{(26)}$ = -0.21	$r_{(22)}$ = 0.06
WRS (13,0,0 - unprocessed) - BB (SRT-3)	$r_{(13)}$ = -0.43	$r_{(13)}$ = -0.33	$\boldsymbol{r_{(28)}}$ **= -0.46***	$\boldsymbol{r_{(24)}}$ **= -0.42***
WRS (13,0,0 - unprocessed) - HP (SRT-3)	$\boldsymbol{r_{(13)}}$ **= -0.59***	$r_{(12)}$ = 0.18	$r_{(27)}$ = -0.03	$r_{(24)}$ = -0.06
WRS (13,0,0m - unprocessed) - HP (SRT-3)	$\boldsymbol{r_{(13)}}$ **= -0.70****	$r_{(12)}$ = 0.03	$r_{(27)}$ = -0.16	$r_{(24)}$ = -0.16
Audiometric thresholds (4 kHz)	$r_{(13)}$ = 0.05	$\boldsymbol{r_{(13)}}$ **= -0.70****	$\boldsymbol{r_{(28)}}$ **= -0.63*****	$\boldsymbol{r_{(24)}}$ **= -0.44***
Audiometric thresholds (EHF)	$r_{(13)}$ = 0.38	$\boldsymbol{r_{(13)}}$ **= -0.60***	$\boldsymbol{r_{(28)}}$ **= -0.48****	$r_{(24)}$ = -0.32

Supplementary Table 7. Results of the statistical analysis of AM detection thresholds. Spearman's r values between the AM detection thresholds of the four conditions and the remaining results. Statistically significant correlations are indicated in bold, with the asterisks corresponding to the probability degree p.

AM thresholds (unprocessed)	yNH	oNH	all	all w/o oHI
EFR SAM (7,0,0)	$r_{(13)}$ = -0.51	$r_{(11)}$ = 0.21	$r_{(26)}$ = -0.29	**$r_{(22)}$ = -0.48***
EFR speech (unprocessed)	$r_{(13)}$ = -0.31	**$r_{(11)}$ = 0.62***	$r_{(26)}$ = -0.09	$r_{(22)}$ = -0.28
EFR SAM (7,0,0 - unprocessed)	$r_{(13)}$ = -0.47	$r_{(11)}$ = 0.01	$r_{(26)}$ = -0.32	**$r_{(22)}$ = -0.52****
WRS (13,0,0m-ref - unprocessed) - HP (SRT+3)	$r_{(14)}$ = -0.06	**$r_{(10)}$ = -0.70***	$r_{(26)}$ = -0.37	$r_{(23)}$ = -0.36
Audiometric thresholds (4 kHz)	$r_{(14)}$ = 0.32	$r_{(11)}$ = -0.31	$r_{(27)}$ = 0.19	**$r_{(23)}$ = 0.40***
AM thresholds (13,0,0m)	**yNH**	**oNH**	**all**	**all w/o oHI**
EFR SAM (7,0,0)	$r_{(13)}$ = -0.45	$r_{(11)}$ = 0.12	$r_{(26)}$ = -0.36	**$r_{(22)}$ = -0.52****
EFR speech (unprocessed)	$r_{(13)}$ = -0.32	$r_{(11)}$ = 0.34	$r_{(26)}$ = -0.27	**$r_{(22)}$ = -0.43***
EFR SAM (7,0,0 - unprocessed)	$r_{(13)}$ = -0.38	$r_{(11)}$ = -0.06	**$r_{(26)}$ = -0.40***	**$r_{(22)}$ = -0.55****
WRS (13,0,0m - unprocessed) - BB (SRT)	$r_{(14)}$ = -0.48	$r_{(11)}$ = -0.03	$r_{(27)}$ = -0.35	**$r_{(23)}$ = -0.47***
WRS (13,0,0m-ref - unprocessed) - HP (SRT+3)	$r_{(14)}$ = -0.03	$r_{(10)}$ = -0.53	**$r_{(26)}$ = -0.38***	**$r_{(23)}$ = -0.42***
Audiometric thresholds (4 kHz)	$r_{(14)}$ = 0.25	$r_{(11)}$ = -0.23	$r_{(27)}$ = 0.27	**$r_{(23)}$ = 0.41***
Audiometric thresholds (EHF)	$r_{(14)}$ = 0.15	$r_{(11)}$ = 0.05	**$r_{(27)}$ = 0.37***	**$r_{(23)}$ = 0.48***
AM thresholds (10,0,0m)	**yNH**	**oNH**	**all**	**all w/o oHI**
EFR SAM (7,0,0)	$r_{(13)}$ = -0.39	$r_{(11)}$ = -0.13	**$r_{(26)}$ = -0.45***	**$r_{(22)}$ = -0.52****
EFR SAM (7,0,0 - unprocessed)	$r_{(13)}$ = -0.50	$r_{(11)}$ = -0.39	**$r_{(26)}$ = -0.56****	**$r_{(22)}$ = -0.63*****
SRT BB (unprocessed)	**$r_{(14)}$ = -0.58***	$r_{(11)}$ = 0.23	$r_{(27)}$ = -0.01	$r_{(23)}$ = -0.29
SRT BB (13,0,0m-ref)	**$r_{(14)}$ = -0.80*****	$r_{(10)}$ = 0.25	$r_{(26)}$ = -0.09	$r_{(22)}$ = -0.29
SRT HP (unprocessed)	**$r_{(14)}$ = -0.81*****	$r_{(11)}$ = -0.15	$r_{(27)}$ = -0.14	$r_{(23)}$ = -0.23
SRT HP (13,0,0m-ref)	**$r_{(14)}$ = -0.75*****	$r_{(10)}$ = 0.25	$r_{(26)}$ = -0.17	$r_{(22)}$ = -0.36
SRT BB (13,0,0m-ref - unprocessed)	**$r_{(14)}$ = -0.70****	$r_{(10)}$ = -0.23	$r_{(26)}$ = -0.33	$r_{(22)}$ = -0.25
SRT HP (13,0,0m-ref - unprocessed)	**$r_{(14)}$ = -0.57***	$r_{(10)}$ = 0.10	$r_{(26)}$ = -0.12	$r_{(22)}$ = -0.25
WRS (13,0,0 - unprocessed) - HP (SRT)	**$r_{(14)}$ = 0.62***	$r_{(10)}$ = -0.37	$r_{(26)}$ = 0.10	$r_{(23)}$ = 0.10
WRS (13,0,0m-ref - unprocessed) - BB (SRT-3)	**$r_{(14)}$ = -0.55***	$r_{(11)}$ = 0.41	$r_{(27)}$ = -0.11	$r_{(23)}$ = -0.13
WRS (10,0,0m - unprocessed) - BB (SRT-3)	$r_{(14)}$ = -0.04	**$r_{(11)}$ = 0.61***	$r_{(27)}$ = 0.26	$r_{(23)}$ = 0.27
WRS (10,0,0m - unprocessed) - HP (SRT-3)	**$r_{(14)}$ = -0.54***	$r_{(10)}$ = 0.10	$r_{(26)}$ = -0.33	**$r_{(23)}$ = -0.41***
Audiometric thresholds (EHF)	$r_{(14)}$ = -0.00	$r_{(11)}$ = 0.31	**$r_{(27)}$ = 0.48****	**$r_{(23)}$ = 0.50***
AM thresholds (7,0,0m)	**yNH**	**oNH**	**all**	**all w/o oHI**
EFR SAM (7,0,0)	$r_{(13)}$ = -0.16	$r_{(11)}$ = -0.23	**$r_{(26)}$ = -0.51****	**$r_{(22)}$ = -0.43***
EFR speech (unprocessed)	$r_{(13)}$ = 0.11	$r_{(11)}$ = -0.39	**$r_{(26)}$ = -0.46***	$r_{(22)}$ = -0.36
EFR SAM (7,0,0 - unprocessed)	$r_{(13)}$ = -0.42	$r_{(11)}$ = -0.44	**$r_{(26)}$ = -0.67******	**$r_{(22)}$ = -0.62****
SRT BB (unprocessed)	**$r_{(14)}$ = -0.55***	$r_{(11)}$ = 0.05	$r_{(27)}$ = -0.05	$r_{(23)}$ = -0.19
SRT BB (13,0,0m-ref)	**$r_{(14)}$ = -0.64****	$r_{(10)}$ = 0.04	$r_{(26)}$ = -0.07	$r_{(22)}$ = -0.16
SRT HP (unprocessed)	**$r_{(14)}$ = -0.61***	$r_{(11)}$ = -0.43	$r_{(27)}$ = -0.10	$r_{(23)}$ = -0.07
SRT HP (13,0,0m-ref)	**$r_{(14)}$ = -0.59***	$r_{(10)}$ = 0.17	$r_{(26)}$ = -0.08	$r_{(22)}$ = -0.25
WRS (13,0,0 - unprocessed) - BB (SRT)	**$r_{(14)}$ = -0.51***	$r_{(11)}$ = -0.07	**$r_{(27)}$ = -0.46***	**$r_{(23)}$ = -0.55****
WRS (13,0,0m - unprocessed) - BB (SRT)	$r_{(14)}$ = 0.11	$r_{(11)}$ = -0.54	**$r_{(27)}$ = -0.41***	$r_{(23)}$ = -0.31
WRS (13,0,0 - unprocessed) - HP (SRT)	**$r_{(14)}$ = 0.51***	$r_{(10)}$ = -0.38	$r_{(26)}$ = 0.04	$r_{(23)}$ = 0.06
WRS (13,0,0m-ref - unprocessed) - HP (SRT+3)	$r_{(14)}$ = -0.24	$r_{(10)}$ = -0.41	$r_{(26)}$ = -0.46*	**$r_{(23)}$ = -0.53****
WRS (13,0,0 - unprocessed) - HP (SRT+3)	$r_{(14)}$ = -0.02	$r_{(10)}$ = -0.25	**$r_{(26)}$ = -0.47***	**$r_{(23)}$ = -0.45***
WRS (13,0,0m-ref - unprocessed) - BB (SRT-3)	**$r_{(14)}$ = -0.75*****	$r_{(11)}$ = 0.39	$r_{(27)}$ = -0.21	$r_{(23)}$ = -0.20
WRS (13,0,0m-ref - unprocessed) - HP (SRT-3)	**$r_{(14)}$ = -0.59***	$r_{(10)}$ = -0.25	**$r_{(26)}$ = -0.57****	**$r_{(23)}$ = -0.59****
WRS (13,0,0 - unprocessed) - HP (SRT-3)	$r_{(14)}$ = -0.35	$r_{(10)}$ = -0.34	**$r_{(26)}$ = -0.51****	**$r_{(23)}$ = -0.51****
WRS (13,0,0m - unprocessed) - HP (SRT-3)	$r_{(14)}$ = -0.33	**$r_{(10)}$ = -0.64***	**$r_{(26)}$ = -0.50****	**$r_{(23)}$ = -0.54****
WRS (10,0,0m - unprocessed) - HP (SRT-3)	$r_{(14)}$ = -0.49	$r_{(10)}$ = -0.26	$r_{(26)}$ = -0.37	**$r_{(23)}$ = -0.46***
Audiometric thresholds (4 kHz)	$r_{(14)}$ = 0.40	$r_{(11)}$ = 0.15	**$r_{(27)}$ = 0.59*****	**$r_{(23)}$ = 0.56****
Audiometric thresholds (EHF)	$r_{(14)}$ = 0.18	$r_{(11)}$ = 0.46	**$r_{(27)}$ = 0.68******	**$r_{(23)}$ = 0.62****

Supplementary Table 8. Statistical analysis of the WRS improvement in the six SNR conditions using the 13,0,0 modified (13,0,0m) algorithm. Spearman's r values between the WRS improvement of the 13,0,0m processing algorithm and the remaining metrics.

WRS (13,0,0m - unprocessed) - BB (SRT)	yNH	oNH	all	all w/o oHI
EFR speech (unprocessed)	$r_{(13)}$ = 0.19	$r_{(13)}$ = 0.31	$\boldsymbol{r_{(28)}}$ **= 0.44***	$r_{(24)}$ = 0.37
EFR SAM (10,0,0 - unprocessed)	$\boldsymbol{r_{(13)}}$ **= -0.63***	$r_{(13)}$ = -0.23	$r_{(28)}$ = -0.15	$r_{(24)}$ = -0.24
EFR speech (10,0,0 - unprocessed)	$\boldsymbol{r_{(13)}}$ **= -0.67****	$r_{(13)}$ = 0.18	$r_{(28)}$ = -0.29	$\boldsymbol{r_{(24)}}$ **= -0.40***
AM thresholds (13,0,0m)	$r_{(14)}$ = -0.48	$r_{(11)}$ = -0.03	$r_{(27)}$ = -0.35	$\boldsymbol{r_{(23)}}$ **= -0.47***
AM thresholds (7,0,0m)	$r_{(14)}$ = 0.11	$r_{(11)}$ = -0.54	$\boldsymbol{r_{(27)}}$ **= -0.41***	$r_{(23)}$ = -0.31
AM thresholds (13,0,0m - unprocessed)	$\boldsymbol{r_{(14)}}$ **= -0.50***	$r_{(11)}$ = -0.05	$r_{(27)}$ = -0.33	$r_{(23)}$ = -0.29
Audiometric thresholds (EHF)	$r_{(14)}$ = -0.19	$r_{(13)}$ = -0.30	$\boldsymbol{r_{(29)}}$ **= -0.48****	$\boldsymbol{r_{(25)}}$ **= -0.40***
WRS (13,0,0m - unprocessed) - BB (SRT+3)	yNH	oNH	all	all w/o oHI
EFR speech (unprocessed)	$r_{(13)}$ = 0.42	$r_{(13)}$ = 0.03	$\boldsymbol{r_{(28)}}$ **= 0.41***	$\boldsymbol{r_{(24)}}$ **= 0.40***
Audiometric thresholds (EHF)	$r_{(14)}$ = -0.38	$r_{(13)}$ = 0.07	$\boldsymbol{r_{(29)}}$ **= -0.39***	$\boldsymbol{r_{(25)}}$ **= -0.40***
WRS (13,0,0m - unprocessed) - BB (SRT-3)	yNH	oNH	all	all w/o oHI
SRT BB (13,0,0m-ref - unprocessed)	$r_{(14)}$ = 0.37	$r_{(10)}$ = 0.31	$r_{(26)}$ = 0.33	$\boldsymbol{r_{(22)}}$ **= 0.43***
WRS (13,0,0m - unprocessed) - HP (SRT)	yNH	oNH	all	all w/o oHI
AM thresholds (13,0,0m - unprocessed)	$r_{(14)}$ = -0.37	$r_{(10)}$ = -0.15	$r_{(26)}$ = -0.36	$\boldsymbol{r_{(23)}}$ **= -0.44***
SRT HP (unprocessed)	$r_{(14)}$ = -0.26	$r_{(12)}$ = -0.35	$\boldsymbol{r_{(28)}}$ **= -0.38***	$\boldsymbol{r_{(25)}}$ **= -0.39***
Audiometric thresholds (EHF)	$r_{(14)}$ = -0.33	$r_{(12)}$ = -0.46	$\boldsymbol{r_{(28)}}$ **= -0.42***	$\boldsymbol{r_{(25)}}$ **= -0.46***
WRS (13,0,0m - unprocessed) - HP (SRT-3)	yNH	oNH	all	all w/o oHI
EFR SAM (unprocessed)	$\boldsymbol{r_{(13)}}$ **= -0.55***	$r_{(12)}$ = -0.46	$r_{(27)}$ = -0.33	$r_{(24)}$ = -0.34
EFR SAM (10,0,0)	$r_{(13)}$ = -0.13	$\boldsymbol{r_{(12)}}$ **= -0.61***	$r_{(27)}$ = -0.15	$r_{(24)}$ = -0.04
EFR speech (unprocessed)	$\boldsymbol{r_{(13)}}$ **= -0.84*****	$r_{(12)}$ = -0.08	$r_{(27)}$ = -0.13	$r_{(24)}$ = -0.11
EFR speech (10,0,0)	$\boldsymbol{r_{(13)}}$ **= -0.70****	$r_{(12)}$ = 0.03	$r_{(27)}$ = -0.16	$r_{(24)}$ = -0.16
AM thresholds (7,0,0m)	$r_{(14)}$ = -0.33	$\boldsymbol{r_{(10)}}$ **= -0.64***	$\boldsymbol{r_{(26)}}$ **= -0.50****	$\boldsymbol{r_{(23)}}$ **= -0.54****
WRS (13,0,0m - unprocessed) - HP (SRT+3)	yNH	oNH	all	all w/o oHI
SRT BB (unprocessed)	$\boldsymbol{r_{(14)}}$ **= -0.57***	$r_{(12)}$ = -0.24	$\boldsymbol{r_{(28)}}$ **= -0.37***	$\boldsymbol{r_{(25)}}$ **= -0.42***
Audiometric thresholds (4 kHz)	$\boldsymbol{r_{(14)}}$ **= 0.55***	$r_{(12)}$ = 0.11	$r_{(28)}$ = 0.02	$r_{(25)}$ = 0.02
Audiometric thresholds (EHF)	$r_{(14)}$ = -0.08	$r_{(12)}$ = -0.22	$\boldsymbol{r_{(28)}}$ **= -0.38***	$r_{(25)}$ = -0.36

Supplementary Table 9. Word-recognition improvement for the 13,0,0m strategy (cf. Fig. 8). The median SI improvement achieved with the 13,0,0m strategy is shown for each SNR condition, after subtracting the median WRS in the unprocessed condition. The median scores were defined from the average results of all subjects in each group, which were computed either using all 12 trials of each subject (trials 1-12) or after omitting the first 4 trials of the results (trials 5-12). For each group in each condition, the best SI improvement is indicated by bold font.

	BB			HP		
	SRT -3 dB	SRT	SRT +3 dB	SRT -3 dB	SRT	SRT +3 dB
yNH (trials 1-12)	5 %	**0.83 %**	0 %	8.33 %	4.17 %	**3.33 %**
oNH (trials 1-12)	3.33 %	**1.67 %**	0 %	3.33 %	-1.67 %	-3.33 %
yNH (trials 5-12)	**6.25 %**	0 %	**1.25 %**	**10 %**	**7.5 %**	0 %
oNH (trials 5-12)	**7.5 %**	-2.5 %	**2.5 %**	**3.75 %**	**2.5 %**	**-1.25 %**