Compression and amplification algorithms in hearing aids impair the selectivity of neural responses to speech

Alex Armstrong[&,1], Chi Chung Lam[&,1], Shievanie Sabesan[&,1], Nicholas A. Lesica[1*]
[&]These authors contributed equally
[1]Ear Institute, University College London, UK
[*]n.lesica@ucl.ac.uk

In quiet environments, hearing aids improve the perception of low-intensity sounds. However, for high-intensity sounds in background noise, the aids often fail to provide a benefit to the wearer. Here, by using large-scale single-neuron recordings from hearing-impaired gerbils — an established animal model of human hearing — we show that hearing aids restore the sensitivity of neural responses to speech, but not their selectivity. Rather than reflecting a deficit in supra-threshold auditory processing, the low selectivity is a consequence of hearing-aid compression (which decreases the spectral and temporal contrasts of incoming sound) and of amplification (which distorts neural responses, regardless of whether hearing is impaired). Processing strategies that avoid the trade-off between neural sensitivity and selectivity should improve the performance of hearing aids.

Hearing loss is one of the most widespread and disabling chronic conditions in the world today. Approximately 500 million people worldwide are affected, making hearing loss the fourth leading cause of years lived with disability[1] and imposing a substantial economic burden with estimated costs of more than $750 billion globally each year[2]. Hearing loss has also been linked to declines in mental health; in fact, a recent commission identified hearing loss as the leading modifiable risk factor for incident dementia [3]. As the societal impact of hearing loss continues to grow, the need for improved treatments is becoming increasingly urgent.

Hearing aids are the current treatment of choice for the most common forms of hearing loss that result from noise exposure and aging. But only a small fraction of people with hearing loss (15-20%) use hearing aids [4,5]. There are a number of reasons for this poor uptake, but one of the most important is lack of benefit in listening environments that are typical of real-world social settings. The primary problem associated with hearing impairment is loss of audibility, i.e. loss of the ability to detect low-intensity sounds [6,7]. As a result of cochlear damage, sensitivity thresholds are increased and low-intensity sounds can no longer be perceived. Fortunately, hearing aids are generally able to correct this problem by providing amplification. But perception often remains impaired even after audibility is restored. It is well established that hearing aids improve the perception of low-intensity sounds in quiet environments but often fail to provide benefit for high-intensity sounds in background noise[8,9].

The reasons for this residual impairment remain unclear, but one possibility is the existence of additional deficits beyond loss of audibility that impair the processing of high-intensity sounds. Many such deficits have been reported such as broadened frequency tuning [10] and impaired temporal processing [11,12]. But these deficits are typically observed when comparisons between normal and impaired hearing are made at different sound intensities to control for differences in audibility. This approach confounds the effects of hearing loss with the effects of intensity; amplification to high intensities impairs auditory processing even with normal hearing[13–15]. In fact, when listeners with mild-to-moderate hearing loss (typical of the vast majority of impairments) and normal hearing listeners are compared at the same high intensities, the performance of the two groups is often similar in both simple tasks such as tone-in-noise detection [16] and complex tasks such as speech-in-noise perception[14,17–20].

Another possibility is that the residual problems that persist after restoration of audibility are caused by the processing in the hearing aid itself. Most modern hearing aids share the same core processing algorithm known as multi-channel wide dynamic range compression (WDRC). This algorithm provides listeners with frequency-specific amplification based on measured changes in their sensitivity thresholds. It also provides compression by varying the amplification of each frequency over time based on the incoming sound intensity such that amplification decreases as the incoming sound intensity increases. This algorithm is designed to mimic the amplification and compression that normally take place within a healthy cochlea but are

compromised by hearing loss. However, it ignores many other aspects of auditory processing that are also impacted by hearing loss[21] and modifies the spectral and temporal properties of incoming sounds in ways that may actually be detrimental to perception[22,23].

Identifying the factors responsible for the failure of hearing aids to restore normal auditory perception through psychophysical studies has proven difficult. We approached the problem from the perspective of the neural code — the activity patterns in central auditory brain areas that provide the link between sound and perception. Hearing loss impairs perception because it causes distortions in the information carried by the neural code about incoming sounds. The failure of current hearing aids to restore normal perception suggests that there are critical features of the neural code that remain distorted. An ideal hearing aid would correct these distortions by transforming incoming sounds such that processing of the transformed sounds by the impaired system would result in the same neural activity patterns as the processing of the original sounds by the healthy system; current hearing aids fail to achieve this ideal.

Little is known about the specific distortions in the neural code caused by hearing loss or the degree to which current hearing aids correct them. The effects of hearing loss on the neural code for complex sounds such as speech have been well characterized at the level of the auditory nerve[24], but its impact on downstream central brain areas remains unclear as there have been few studies of single neuron responses with hearing loss and even fewer with hearing aids. Auditory processing in humans involves many brain areas from the brainstem, which performs general feature extraction and integration, to the cortex, which performs context- and language-specific processing. While large-scale studies of single neurons in these areas in humans are not yet possible, animal models can serve as a valuable surrogate, particularly for the early stages of processing which are largely conserved across mammals and appear to be the primary source of human perceptual deficits [7]. Prior work has already shown that classifiers trained to identify speech phonemes based on neural activity patterns recorded from animals perform similarly to human listeners performing an analogous task [25]. Thus, comparisons of the neural code with and without hearing loss and a hearing aid in an animal model can provide valuable insight into which distortions in the neural code underlie the failure of hearing aids to restore normal perception.

The neural code is transformed through successive stages of processing from the auditory nerve to the auditory cortex. At the level of the auditory nerve, some of the important effects of hearing loss that underlie impaired perception are not yet manifest [26], while at the level of the thalamus and cortex, neural activity is modulated by contextual and behavioural factors (e.g. attention) that complicate the study of the general effects of hearing loss on the neural representation of acoustic features. We chose to study the neural code in the inferior colliculus (IC), the midbrain hub of the central auditory pathway that serves as an obligatory relay between the early brainstem and the thalamus. The neural activity in the IC reflects the integrated effects of processing in several peripheral pathways but is still primarily determined by the acoustic features of incoming sounds.

We focused our study on mild-to-moderate sensorineural hearing loss, which reflects relatively modest cochlear damage[27]. Because peripheral processing is still highly functional with this form of hearing loss, there is potential for a hearing aid to provide substantial benefit. We found that most of the distortions in the neural code in the IC that are caused by hearing loss are, in fact, corrected by a hearing aid, but a loss of selectivity in neural responses that is specific to complex sounds remains. Our analysis suggests that the low selectivity of aided responses does not reflect a deficit in supra-threshold auditory processing, but is instead a consequence of the strategies used by current hearing aids to restore audibility. Our findings support the wide provision of simple devices to address the growing global burden of hearing loss in the short term and provide guidance for the development of improved hearing aids in the future.

Results

To study the neural code with high spatial and temporal resolution across large populations of neurons, we made recordings using custom-designed electrodes with a total of 512 channels spanning both brain hemispheres in gerbils, a commonly used animal model for studies of low-frequency hearing (Figure 1a; Figure S1). We used these large-scale recordings to study the activity patterns of more than 5,000 neurons in the IC. To induce sloping mild-to-moderate sensorineural hearing loss, we exposed young-adult gerbils to

broadband noise (118 dB sound pressure level (SPL), 3 hours). Compared to normal hearing gerbils, the resulting pure-tone threshold shifts measured one month after exposure using auditory brainstem response (ABR) recordings typically ranged from 20-30 dB at low frequencies to 40-50 dB at high frequencies (Figure 1b). Pure-tone threshold shifts with hearing loss were also evident in frequency response areas (FRAs) measured from multi-unit activity (MUA) recorded in the IC, which illustrate the degree to which populations of neurons were responsive to tones with different frequencies and intensities (Figure 1c).

For gerbils with hearing loss, we presented sounds both before and after processing with a multi-channel WDRC hearing aid. The amplification and compression parameters for the hearing aid were custom fit to each ear of each gerbil based on the measured ABR threshold shifts. The hearing aid amplified sounds in a frequency-dependent manner, with amplification for sounds at moderate intensity typically increasing from approximately 10 dB at low frequencies to approximately 20 dB at high frequencies (Figure 1b). This amplification was sufficient to restore the pure-tone IC MUA thresholds with hearing loss to normal (Figure 1c).

To begin our study of the neural code, we first presented speech to normal hearing gerbils at moderate intensity (62 dB SPL; typical of a conversation in a quiet environment). We used a set of nonsense consonant-vowel syllables, as is common in human studies that focus on acoustic cues for speech perception rather than linguistic or cognitive factors. The set of syllables consisted of all possible combinations of 12 consonants and 4 vowels, each spoken by 8 different talkers. For individual neurons, individual instances of different syllables elicited complex response patterns (Figure 2a). For a population of neurons, the response patterns can be thought of as trajectories in a high-dimensional space in which each dimension corresponds to the activity of one neuron and each point on a trajectory indicates the activity of each neuron in the population at one point in time. To visualize these patterns, we performed dimensionality reduction via principal component analysis, which identified linear combinations of all neurons that best represented the full population. Within the space defined by the first three principal components, the responses to individual instances of different syllables followed distinct trajectories that were reliable across repeated trials (Figure 2b, left).

To assess the degree to which the neural code allowed for accurate identification of consonants, we used a classifier to identify the consonant in each syllable based on the population response patterns. Despite the variability in the responses to each consonant across syllables with different vowels and talkers, the average responses to different consonants were still distinct (Figure 2b, right). We trained a support vector machine to classify the first 150 ms of single-trial responses represented as spike counts with 5 ms time bins. We formed populations of 150 neurons by sampling at random, without replacement, from neurons from all normal hearing gerbils until there were no longer enough neurons remaining to form another population. The classifier identified consonants with high accuracy (Figure 2c) and error patterns that reflected confusions within consonant classes as expected from human perceptual studies [28,29]. Accuracy was high for the sibilant fricatives (*ʃ, ʒ, s, z*), moderate for the stops (*t, k, b, d*), and low for the nasals (*n, m*) and the non-sibilant fricatives (*v, ð*).

We presented the same set of syllables to gerbils with hearing loss before and after processing with the hearing aid. The mean spike rate of individual neurons was decreased by hearing loss but restored to normal by the hearing aid (Figures 3a,b; for full details of all statistical tests including sample sizes and p-values, see Table S1). A classifier trained to detect speech in silence based on the neural response patterns of individual neurons confirmed that the hearing aid restored audibility to normal (Figure 3b, right). Consonant identification was also impacted by hearing loss but, unlike audibility, remained well below normal even with the hearing aid (Figure 3c). The hearing aid failed to restore consonant identification not only for speech in quiet, but also for speech presented in the presence of either a second independent talker or multi-talker noise. This failure was evident across a range of different classifiers, neural representations, and population sizes (Figures S2,S3) and, thus, reflects a general deficit in the neural code.

Hearing aids fail to restore the selectivity of responses to speech

To understand why the hearing aid failed to restore consonant identification to normal, we investigated how different features of the neural response patterns varied across hearing conditions. Accurate auditory

perception requires the response patterns elicited by different sounds to be distinct and reliable. For consonant identification, the response to a particular instance of a consonant must be similar to responses to other instances of that consonant but different from responses to other consonants.

In the context of any perceptual task, a neural response pattern can be separated into signal and noise, i.e. the components of the response which are helpful for the task and the components of the response which are not (Figure 4a). For consonant identification, the signal can be further divided into a *common signal,* which is common to all consonants, and a *differential signal*, which is specific to each consonant. The common signal reflects the average detectability (that is, audibility) of all consonants, while the differential signal determines how well different consonants can be discriminated.

The noise can also be further divided based on the different sources of variability in neural response patterns. The first source of variability is *nuisance noise,* which arises because consonants are followed by different vowels or spoken by different talkers (note that while this component of the response serves as noise for this task, it could also serve as signal for a different task, such as talker identification). The second source of variability is *internal noise,* which reflects the fundamental limitations on neural coding due to the stochastic nature of spiking and other intrinsic factors**.** For speech in the presence of additional sounds, there is also *external noise*, which is the variability in responses that is caused by the additional sounds themselves.

All of these signal and noise components have the potential to influence consonant identification through their impact on the neural response patterns and together they form a complete description of any response. To isolate each of these components in turn, we computed the covariance between response patterns with different forms of shuffling across consonants, vowels, and talkers. We performed this decomposition of the responses in the frequency domain by computing spectral densities in order to gain further insight into which features of speech were reflected in each component.

The results are shown for a typical neuron for speech in quiet in Figure 4b. We first isolated the internal noise by comparing the power spectral density (*PSD*) of responses across a single trial of every syllable with the cross spectral density (*CSD*) of responses to repeated trials of the same speech (i.e. with the order of consonants, vowels, and talkers preserved). The *PSD* provides a frequency-resolved measure of the variance in a single neural response, while the *CSD* provides a frequency-resolved measure of the covariance between two responses. For an ideal neuron, repeated trials of identical speech would elicit identical responses and the *CSD* would be equal to the *PSD*. For a real neuron, the difference between the *PSD* and the *CSD* gives a measure of the internal noise. For the example neuron, the *CSD* was less than the *PSD* at all frequencies. The difference between the *PSD* and the *CSD* increased with increasing frequency up to 80 Hz and then remained relatively constant, indicating that the internal noise was smallest (and, thus, the neural responses most reliable) at frequencies corresponding to the envelope of the speech. We next isolated the nuisance noise by comparing the *CSD* to the cross spectral density of responses to repeated trials after shuffling across vowels and talkers (denoted as $CSD_{shuff}^{V,T}$). After this shuffling, the only remaining covariance between the responses is that which is shared across different instances of the same consonants. For the example neuron, this covariance was only significant at frequencies corresponding to the speech envelope; at frequencies higher than 40 Hz, the $CSD_{shuff}^{V,T}$ dropped below chance (denoted as CSD_0). Thus, the nuisance noise, given by the difference between the *CSD* and the $CSD_{shuff}^{V,T}$, was largest at the frequencies corresponding to pitch (which is expected because pitch is reliably encoded in the response patterns but is not useful for talker-independent consonant identification).

Finally, we isolated the common signal from the differential signal by comparing the $CSD_{shuff}^{V,T}$ with the cross spectral density of the responses after shuffling across talkers, vowels, and consonants (denoted as $CSD_{shuff}^{C,V,T}$). The only covariance between the responses that remains after this shuffling is that which is shared across all syllables. For the example neuron, both the differential signal, given by the difference between the $CSD_{shuff}^{V,T}$and the $CSD_{shuff}^{C,V,T}$, and the common signal, given directly by the $CSD_{shuff}^{C,V,T}$, were significant across the full range of speech envelope frequencies.

At the population level, hearing loss impacted all components of the responses, with internal noise, nuisance noise, common signal, and differential signal all decreasing in magnitude (Figure 4c). The hearing aid increased the magnitude of both the internal noise and the nuisance noise (corresponding to the light and dark blue areas in Figure 4b, respectively), but both remained at or below normal levels. This suggests that mild-to-moderate hearing loss does not result in either fundamental limitations on neural coding or increased sensitivity to uninformative features of speech that can account for the failure of the hearing aid to restore consonant identification to normal.

The hearing aid also restored the common signal (corresponding to the dark red area in Figure 4b) to normal, but failed to increase the magnitude of the differential signal (corresponding to the light red area in Figure 4b). Thus, the key difference between normal and aided responses appears to be their selectivity, i.e. the degree to which their average responses to different consonants are distinct. This difference was most pronounced in the low-frequency component of the responses (Figure 4d, left). In fact, the same failure of the hearing aid to increase the differential signal was evident when looking only at spike counts (Figure 4d, right), suggesting that the hearing aid fails to restore even the differences in overall activity across consonants.

The selectivity of aided responses to tones is normal

One possible explanation for the low selectivity of aided responses to speech is broadened frequency tuning, which would decrease sensitivity to differences in the spectral content of different consonants and increase the degree to which features of speech at one frequency are susceptible to masking by noise at other frequencies. The width of cochlear frequency tuning can increase with cochlear damage [27] and impaired frequency selectivity is often reported in people with hearing loss [10]. However, the degree to which frequency tuning is broadened with hearing loss depends on both the severity of the hearing loss and the intensity of incoming sounds (because frequency tuning broadens with increasing intensity even with normal hearing). Forward-masking paradigms that provide psychophysical estimates that closely match neural tuning curves[30–32] suggest changes in frequency tuning may not be significant for mild-to-moderate hearing loss at moderate sound intensities[16].

To characterize frequency tuning, we examined responses to pure tones presented at different frequencies and intensities. We defined the characteristic frequency (CF) of each neuron as the frequency that elicited a significant response at the lowest intensity and the threshold as the minimum intensity required to elicit a significant response at the CF (Figure 5a). Hearing loss caused an increase in thresholds across the range of speech-relevant frequencies, but this threshold shift was corrected by the hearing aid; in fact, aided thresholds were lower than those for normal hearing for CFs at both edges of the speech-relevant range (Figure 5b).

The mean spike rate of individual neurons in response to pure tones presented at the same intensity as the speech (62 dB SPL) was decreased by hearing loss, but restored to normal by the hearing aid (Figure 5c, left). The width of frequency tuning (defined as the range of frequencies for which the mean spike rate was at least half of its maximum value) at the same relative intensity (14 dB above threshold) for each neuron was increased by hearing loss, as expected, but restored to normal by the hearing aid (Figure 5c, middle left). The width of frequency tuning at a fixed intensity of 62 dB SPL was decreased by hearing loss (Figure 5c, middle), as expected given the increased thresholds. Tuning width at this intensity was increased with the hearing aid, but remained slightly narrower than normal. This suggests that mild-to-moderate hearing loss does not result in broadened frequency tuning at moderate intensities even after amplification by the hearing aid.

To determine directly whether the selectivity of responses to pure tones was impacted by hearing loss, we again isolated the differential signal component (that is, the component of the response that varies with tone frequency). The magnitude of the differential signal was unimpacted by hearing loss and was slightly higher than normal with the hearing aid (Figure 5c, middle right), indicating that there was no loss of selectivity. To confirm the normal selectivity of aided responses to tones, we trained a classifier to identify tone frequencies based on neural response patterns. The performance of the classifier was decreased by hearing loss but returned to normal with the hearing aid (Figure 5c, right). Thus, the failure of the hearing aid to restore

consonant identification to normal does not appear to result from a general loss of frequency selectivity in neural responses.

Hearing aid compression decreases the selectivity of responses to speech

Our results thus far suggest that if the low selectivity of aided responses to speech reflects a supra-threshold auditory processing deficit with hearing loss, the deficit is only manifest for complex sounds. While this is certainly possible given the nonlinear nature of auditory processing, there is also another potential explanation: the low selectivity of responses to speech may be a result of distortions caused by the hearing aid itself [22,23]. The multi-channel WDRC algorithm in the hearing aid constantly adjusts the amplification across frequencies, with each frequency receiving more amplification when it is weakly present in the incoming sound and less amplification when it is strongly present. This results in a compression of incoming sound across frequencies and time into a reduced range. Since a pure tone is a simple sound with a single frequency and constant amplitude, this compression has relatively little impact. But for complex sounds with multiple frequencies that vary in amplitude over time, such as speech, this compression serves to decrease both spectral and temporal contrast.

The WDRC algorithm is designed to replace the normal amplification and compression that are lost because of cochlear damage. But there are two potential problems with this approach. First, whereas normal cochlear compression does decrease spectral and temporal contrast, there are also other mechanisms acting in a healthy cochlea that counteract this by increasing contrast (for example, cross-frequency suppression) that are not included in the WDRC algorithm[21]. Second, there is evidence to suggest that with mild-to-moderate hearing loss, amplification of low intensity sounds is impaired but compression of moderate and high intensity sounds remains normal[33–35]. Thus, the total compression for the aided condition with mild-to-moderate hearing loss may be higher than normal, resulting in an effective decrease in the spectral and temporal contrast of complex sounds as represented in the neural code.

To investigate the impact of the hearing aid compression on the selectivity of responses to speech, we first computed the spectrograms of each instance of each consonant before and after processing with the hearing aid and measured their contrast (Figure 6a). On average, the spectrotemporal contrast after processing with the hearing aid was 15% lower than in the original sound (Figure 6b, left). This decrease in contrast was reflected in the performance of a classifier trained to identify the consonant in each spectrogram, which also decreased after processing with the hearing aid (Figure 6b, right).

If the hearing aid compression is responsible for the low selectivity of neural responses, then it should be possible to improve selectivity (and, thus, consonant identification) by providing amplification without compression. We presented the same consonant-vowel syllables after linear amplification (with a fixed gain of 20 dB applied across all frequencies) and compared the results of classification and response decomposition to those for the original speech. Linear amplification without compression restored both classifier performance and the magnitude of the differential signal to normal (Figure 6c). Thus, the failure of the hearing aid to restore response selectivity and consonant identification for speech in quiet appears to result from hearing aid compression rather than a deficit in supra-threshold auditory processing with hearing loss. Linear amplification is able to restore the selectivity of neural responses and, consequently, consonant identification by restoring audibility without distorting the spectral and temporal features of speech.

Amplification decreases consonant identification in noise for all hearing conditions

We next investigated whether removing hearing aid compression and providing only linear amplification was also sufficient to restore consonant identification to normal for speech in the presence of additional sounds. While linear amplification was sufficient to restore consonant identification in the presence of a second independent talker, it failed in multi-talker noise (Figure 6d). This suggests that for speech in noise, there are additional reasons for the failure of the hearing aid to restore consonant identification beyond just the distortions caused by hearing aid compression.

The failure of both the hearing aid and linear amplification to restore consonant identification in noise could reflect a supra-threshold auditory processing deficit with hearing loss that is only manifest in difficult listening conditions, but this is not necessarily the case. Even with normal hearing, the intelligibility of speech in noise

decreases as overall intensity increases (an effect known as 'rollover' with a complex physiological basis[13,15,36]). When the background noise is dominated by low frequencies (as is the case for multi-talker noise), speech intelligibility decreases by approximately 5% for every 10 dB increase in overall intensity above moderate levels, even when the speech-to-noise ratio remains constant [14,37]. Thus, the differences in the perception of moderate-intensity speech-in-noise with normal hearing and that of amplified speech-in-noise with hearing loss may not reflect the effects of hearing loss per se, but rather the unintended consequences of amplifying sounds to high intensities to restore audibility.

To assess the impact of rollover on the neural code, we compared consonant identification and response decomposition with normal hearing before and after linear amplification. The amplification to high intensity did not impact consonant identification in quiet or in the presence of second talker, but decreased consonant identification in multi-talker noise (Figure 7a). This decrease in consonant identification in noise at high intensities with normal hearing appears to result from a decrease in response selectivity; the magnitude of the differential signal was significantly smaller after amplification, while the magnitudes of the common signal and total noise were unchanged (Figure 7b; note that because we did not present repeated trials of 'frozen' multi-talker noise, we cannot isolate the individual noise components but we can still measure the total magnitude of all noise components as the difference between the *PSD* and the $CSD_{shuff}^{V,T}$).

To determine whether rollover can account for the deficit in consonant identification in noise with hearing loss that remains even after linear amplification, we compared consonant identification after linear amplification for both hearing loss and normal hearing (that is, using responses to amplified speech for both conditions). When compared at the same high intensity, consonant identification with or without hearing loss was not significantly different (Figure 7c). Thus, the failure of both the hearing aid and linear amplification to restore consonant identification in noise does not appear to reflect a deficit in supra-threshold processing caused by hearing loss, but rather a deficit in high-intensity processing that is present even with normal hearing.

Taken together, our results provide a clear picture of the challenge that must be overcome to restore normal auditory perception after mild-to-moderate hearing loss. Amplification is required to restore audibility, but can also reduce the selectivity of neural responses in complex listening conditions. Thus, a hearing aid must provide amplification while also transforming incoming sounds to compensate for the loss of selectivity at high intensities. Current hearing aids provide the appropriate amplification but fail to implement the required additional transformation and, in fact, appear to further decrease selectivity through compression that decreases the spectrotemporal contrast of incoming sounds.

Discussion

This study was designed to identify the reasons why hearing aids fail to restore normal auditory perception through analysis of the underlying neural code. Our results suggest that difficulties during aided listening with mild-to-moderate hearing loss arise primarily from the decreased selectivity of neural responses. While a hearing aid corrected many of the changes in neural response patterns that were caused by hearing loss, the average response patterns elicited by different consonants remained less distinct than with normal hearing. The low selectivity of aided responses to speech did not appear to reflect a fundamental deficit in supra-threshold auditory processing as the selectivity of responses to moderate-intensity tones was normal. In fact, for speech in quiet, the low selectivity resulted from compression in the hearing aid itself that decreased the spectrotemporal contrast of incoming sounds; linear amplification without compression restored selectivity and consonant identification to normal. For speech in multi-talker noise, however, selectivity and consonant identification remained low even after linear amplification. But linear amplification also decreased the selectivity of neural responses with normal hearing such that, when compared at the same high intensity, consonant identification in noise with normal hearing and hearing loss were similar. These results are consistent with the idea that for mild-to-moderate hearing loss, decreased speech intelligibility is primarily caused by decreased audibility [38] rather than supra-threshold processing deficits. While real-world speech perception is influenced by contextual and linguistic factors that our analysis of responses to isolated consonants cannot account for, performance in consonant identification and open-set word recognition tasks are highly correlated for both normal hearing listeners and listeners with hearing loss

[39,40]. Of course, there are many listeners whose problems go beyond audibility and selectivity for the basic acoustic features of speech: more severe or specific hearing loss may result in additional supra-threshold deficits [10]; cognitive factors may interact with hearing loss to create additional difficulties in real-world scenarios [7]; and supra-threshold deficits can exist without any significant loss of audibility for a variety of reasons [41]. But numerous perceptual studies have reported that the intelligibility of speech-in-noise at high intensities for people with mild-to-moderate hearing loss is essentially normal in both consonant identification and open-set word recognition tasks [14,17–20]. Unfortunately, because of rollover, even normal processing is impaired at high intensities. Thus, those with hearing loss must currently choose between listening naturally to low- and moderate-intensity sounds and suffering from reduced audibility, or artificially amplifying sounds to high intensities and suffering from rollover (Figure 7d).

Overcoming the current trade-off between loss of audibility and rollover is a challenge, but our results are encouraging with respect to the potential of future hearing aids to bring significant improvements. We found that current hearing aids already restore many aspects of the neural code for speech to normal, including mean spike rates, selectivity for pure tones, fundamental limitations on coding (as reflected by internal noise), and sensitivity to prosodic aspects of speech (as reflected by nuisance noise). Instead of compression, which appears to exacerbate the loss of selectivity that accompanies amplification to high intensities, the next-generation of hearing aids must incorporate additional processing to counteract the mechanisms that cause rollover. There have been a number of previous attempts to manipulate the features of speech to improve perception by, for example, enhancing spectral contrast [42–47]. But these strategies have typically been developed to counteract processing deficits that are a direct result of severe hearing loss, e.g. loss of cross-frequency suppression, that may not be present with mild-to-moderate loss. New approaches that are specifically designed to improve perception at high intensities even for normal hearing listeners may be more effective.

The mechanisms that underlie rollover are not well understood. One likely contributor is the broadening of cochlear frequency tuning with increasing sound level, which decreases the frequency selectivity of individual auditory nerve fibres and increases the spread of masking from one frequency to another [48]. But rollover is also apparent when speech is processed to contain primarily temporal cues, suggesting that there are contributions from additional factors such as increased cochlear compression at high intensities that distorts the speech envelope or reduced differential sensitivity of auditory nerve fibres at intensities that exceed their dynamic range[36]. The simplest way to avoid rollover is, of course, to decrease the intensity of incoming sounds. There are already consumer devices that seek to improve speech perception by controlling intensity through sealed in-ear headphones and active noise cancellation[49]. But for traditional open-ear hearing aids, complete control of intensity is not an option; such devices must instead employ complex sound transformations to counteract the negative effects of high intensities without necessarily changing the overall intensity itself.

The required sound transformations are likely to be highly nonlinear and identifying them through traditional engineering approaches may be difficult. But recent advances in machine learning may provide a way forward. It may be possible to train deep neural networks to learn complex sound transformations to counteract the effects of rollover in normal hearing listeners or the joint effects of rollover and hearing loss in impaired listeners. These complex transformations could also potentially address other issues that are ignored by the WDRC algorithm in current hearing aids, such as adaptive processes that modulate neural activity based on high-order sound statistics or over long timescales[50,51]. Deep neural networks may also be able to learn sound transformations that avoid the distortions in binaural cues created by current hearing aids [53,54], enabling the design of new strategies for cooperative processing between devices.

The multi-channel WDRC algorithm in current hearing aids is designed to compensate for the dysfunction of outer hair cells (OHCs) in the cochlea. The OHCs normally provide amplification and compression of incoming sounds, but with hearing loss their function is often impaired either through direct damage or through damage to supporting structures[55]. The true degree of OHC dysfunction in any individual is difficult to determine, so the WDRC algorithm provides amplification and compression in proportion to the measured loss of audibility across different frequencies, which reflects loss of amplification. But while severe hearing loss may result in a loss of both amplification and compression, several studies have found that mild-to-

moderate hearing loss appears to result in a loss of amplification only[33–35]. Thus, with mild-to-moderate loss, the use of a WDRC hearing aid can result in excess compression that distorts the acoustic features of speech[22,23]. Our results demonstrate that these distortions result in the representation of different speech elements in the neural code being less distinct from each other.

A number of studies of speech perception in people with mild-to-moderate hearing loss have found that linear amplification without compression is often comparable or superior to WDRC hearing aids[9,23,56,57]. Our analysis of the neural code provides a physiological explanation for these findings and adds support to the growing movement to increase uptake of hearing aids through the development and provision of simple, inexpensive devices that can be obtained over-the-counter[58,59]. Cost is a major barrier to hearing aid use, with a typical device in the US costing more than $2000 (ref. [60]). However, most of this cost can be attributed to associated services that are bundled with the device, e.g. testing and fitting. The hardware itself typically accounts for less than $100 (indeed, a recent study demonstrated a prototype device that provided adjustable, frequency-specific amplification costing less than $1; ref. [61]). Fortunately, neither the services nor premium features that increase cost are essential [62]. Recent clinical evaluations of over-the-counter personal sound amplification products (PSAPs) have shown that they often provide similar benefit to premium hearing aids fit by professional audiologists[63–65]. Thus, there is now compelling physiological, psychophysical, and clinical evidence to suggest that inexpensive, self-fitting devices can provide benefit for people with mild-to-moderate hearing loss that is comparable to that provided by current state-of-the-art devices.

This conclusion has important implications for strategies to combat the global burden of hearing loss. Simple devices may only be appropriate for people with mild-to-moderate loss, but this group currently includes more than 500 million people worldwide[1]. Thus, the wide adoption of simple devices could have a substantial impact, especially in low-and middle-income countries where the burden of hearing loss is largest and the uptake of hearing aids is lowest. Ideally, the next generation of state-of-the-art hearing aids will bring improvements in both benefit and affordability. But given the need for urgent action to mitigate the impact of hearing loss on wellbeing and mental health[1,3] and the potential for simple devices to provide significant benefit, promoting their use should be considered as a potential public health priority.

Methods

Experimental protocol

Experiments were performed on 35 young-adult gerbils of both sexes that were born and raised in standard laboratory conditions. Twenty of the gerbils were exposed to noise when they were 10-12 weeks old. ABR recordings and large-scale IC recordings were made from all gerbils when they were 14-18 weeks old. The study protocol was approved by the Home Office of the United Kingdom under license number 7007573. All experimental control and data analysis was carried out using custom code in Matlab R2019a.

Noise exposure

Sloping mild-to-moderate sensorineural hearing loss was induced by exposing anesthetized gerbils to high-pass filtered noise with a 3 dB/octave roll-off below 2 kHz at 118 dB SPL for 3 hours [66]. For anesthesia, an initial injection of 0.2 ml per 100 g body weight was given with fentanyl (0.05 mg per ml), medetomidine (1 mg per ml), and midazolam (5 mg per ml) in a ratio of 4:1:10. A supplemental injection of approximately 1/3 of the initial dose was given after 90 minutes. Internal temperature was monitored and maintained at 38.7° C.

Auditory brainstem responses

Animals were placed in a sound-attenuated chamber, and anesthesia and internal temperature were maintained as for noise exposure. An ear plug was inserted into one ear and a free-field speaker was placed 10 cm from the other ear. The sound level was calibrated prior to each recording using a microphone that was placed next to the open ear. Subdermal needles were used as electrodes with the active electrode placed behind the open ear, the reference placed over the nose, and the ground placed in a rear leg. Recordings were bandpass filtered between 300 and 3000 Hz. Clicks (0.1 ms) and tones (4 ms with frequencies ranging from 500 Hz to 8000 Hz in 1 octave steps with 0.5 ms cosine on and off ramps) were presented at intensities ranging from 5 dB SPL to 85 dB SPL in 5 dB steps with a 25 ms pause between

presentations. All sounds were presented 2048 times (1024 times with each polarity). Thresholds were defined as the lowest intensity at which the root mean square (RMS) of the mean response across presentations was more than twice the RMS of the mean of 2048 trials of activity recorded during silence.

Large-scale electrophysiology

Animals were placed in a sound-attenuated chamber and anesthetized for surgery with an initial injection of 1 ml per 100 g body weight of ketamine (100 mg per ml), xylazine (20 mg per ml), and saline in a ratio of 5:1:19. The same solution was infused continuously during recording at a rate of approximately 2.2 µl per min. Internal temperature was monitored and maintained at 38.7° C. A small metal rod was mounted on the skull and used to secure the head of the gerbil in a stereotaxic device. Two craniotomies were made along with incisions in the dura mater, and a 256-channel multi-electrode array (Neuronexus) was inserted into the central nucleus of the IC in each hemisphere (Figure 1a, Figure S1). The arrays were custom-designed to maximize coverage of the portion of the gerbil IC that is sensitive to the frequencies that are present in speech.

Multi-unit activity

MUA was measured from recordings on each channel of the array as follows: (1) a high pass filter was applied with a cutoff frequency of 500 Hz; (2) the absolute value was taken; (3) a low pass filter was applied with a cutoff frequency of 300 Hz. This measure of multi-unit activity does not require choosing a threshold; it simply assumes that the temporal fluctuations in the power at frequencies above 500 Hz reflect the spiking of neurons near each recording site.

Spike sorting

Single-unit spikes were isolated using Kilosort [67] with default parameters. Recordings were separated into overlapping 1-hour segments with a new segment starting every 15 minutes. Kilosort was run separately on each segment and clusters from separate segments were chained together if at least 90% of their events were identical during their period of overlap. Clusters were retained for analysis only if they were present for at least 2.5 hours of continuous recording. This persistence criterion alone was sufficient to identify clusters that also satisfied the usual single-unit criteria with clear isolation from other clusters, lack of refractory period violations, and symmetric amplitude distributions (see Figure S4).

Sounds

Sounds were delivered to speakers (Etymotic ER-2) coupled to tubes inserted into both ear canals along with microphones (Etymotic ER-10B+) for calibration. The frequency response of these speakers measured at the entrance of the ear canal was flat ($\pm$ 5 dB SPL) between 0.2 and 8 kHz. The full set of sounds presented is described below. All sounds were presented diotically except for multi-talker speech babble noise, which was processed by a head-related transfer function to simulate talkers from many different spatial locations.

(1) Tone set 1: 50 ms tones with frequencies ranging from 500 Hz to 8000 Hz in 0.5 octave steps and intensities ranging from 6 dB SPL to 83 dB SPL in 7 dB steps with 2 ms cosine on and off ramps and 175 ms pause between tones. Tones were presented 8 times each in random order.

(2) Tone set 2: 50 ms tones with frequencies ranging from 500 Hz to 8000 Hz in 0.5 octave steps at 62 dB SPL with 5 ms cosine on and off ramps and 175 ms pause between tones. Tones were presented 128 times each in random order.

(3) Consonant-vowel (CV) syllables: Speech utterances taken from the Articulation Index LSCP (LDC Catalog No.: LDC2015S12). Utterances were from 8 American English speakers (4 male, 4 female). Each speaker pronounced CV syllables made from all possible combinations of 12 consonants and 4 vowels. The consonants included the sibilant fricatives *ʃ, ʒ, s,* and *z*, the stops *t, k, b,* and *d*, the nasals *n* and *m*, and the non-sibilant fricatives *v* and *ð*. The vowels included *a, æ, i*, and *o*. Utterances were presented in random order with 175 ms pause between sounds at an intensity of 62 dB SPL (or 82 dB SPL after 20 dB linear amplification). Two identical trials of the full set of syllables were presented for each condition (e.g. 62 or 82 dB SPL, with or without second talker or multi-talker noise, with or without hearing aid). All results reported are based on analysis of only the first trial, except for those relying on computation of cross spectral densities and noise correlations for which both trials were used.

(4) Second independent talker: Speech from 16 different talkers taken from the UCL Scribe database (https://www.phon.ucl.ac.uk/resource/scribe) provided by Prof. Mark Huckvale was concatenated to create a continuous stream of ongoing speech with one talker at a time.
(5) Omni-directional multi-talker speech babble noise: Speech from 16 different talkers from the Scribe database was summed to create speech babble. The speech from each talker was first passed through a gerbil head-related transfer function [68] using software provided by Dr. Rainer Beutelmann (Carl von Ossietzky University) to simulate its presentation from a random azimuthal angle.

Hearing aid simulation
10-channel WDRC processing was simulated using a program provided by Prof. Johsua Alexander (Purdue University) [69]. The crossover frequencies between channels were 200, 500, 1000, 1750, 2750, 4000, 5500, 7000, and 8500 Hz. The intensity thresholds below which amplification was linear for each channel were 45, 43, 40, 38, 35, 33, 28, 30, 36, and 44 dB SPL. The attack and release times (the time constants of the changes in gain following an increase or decrease in the intensity of the incoming sound, respectively) for all channels were 5 and 40 ms, respectively. The gain and compression ratio for each channel were fit individually for each ear of each gerbil using the Cam2B.v2 software provided by Prof. Brian Moore (Cambridge University) [70]. The gain before compression typically ranged from 10 dB at low frequencies to 25 dB at high frequencies. The compression ratios typically ranged from 1 to 2.5, i.e. the increase in sound intensity required to elicit a 1 dB increase in the hearing output ranged from 1 dB to 2.5 dB when compression was engaged.

Data analysis
Visualization of population response patterns
To reduce the dimensionality of population response patterns, the responses for each neuron were first converted to spike count vectors with 5 ms time bins. The responses to all syllables from all neurons across all gerbils for a given hearing condition were combined into one matrix and a principal component decomposition was performed to find a small number of linear combinations of neurons that best described the full population. To visualize responses in three dimensions, single trial or mean responses were projected into the space defined by the first three principal components.

Classification of population response patterns
Populations were formed by sampling at random, without replacement, from neurons from across all gerbils for a given hearing condition until there were no longer enough neurons remaining to form another population. (Note that each population thus contained both simultaneously and non-simultaneously recorded neurons. The simultaneity of recordings could impact classification if the responses contain noise correlations, i.e. correlations in trial-to-trial variability, which would be present only in simultaneous recordings. But we have shown previously under the same experimental conditions that the noise correlations in IC populations are negligible [71]. This was also true of the populations used in this study (Figure S5)).

Unless otherwise noted, populations of 150 neurons were used and classification was performed after converting the responses for each neuron to spike count vectors with 5 ms time bins. Only the first 150 ms of the responses to each syllable were used to minimize the influence of the vowel. The classifier was a support vector machine with a max-wins voting strategy based on all possible combinations of binary classifiers and 10-fold cross validation. To ensure the generality of the results, different classifiers, neural representations, and population sizes were also tested (see Figures S2,S3).

Computation of spectral densities
Spectral densities were computed as a measure of the frequency-specific covariance between two responses (or variance of a single response). To compute spectral densities, responses to all syllables with different consonants and vowels spoken by different talkers were concatenated in time and converted to binary spike count vectors with 1 ms time bins

$$r = [r^{c=1,v=1,t=1}\, r^{c=1,v=1,t=2} \ldots\, r^{c=C,v=V,t=T}]$$

where $r^{c,v,t} = [r^{c,v,t}[1]\, r^{c,v,t}[2] \ldots r^{c,v,t}[N]]$ is the binary spike count vector with *N* time bins for the response to one syllable composed of consonant *c* and vowel *v* spoken by talker *t*. Responses were then separated into

300 ms segments with 50% overlap and each segment was multiplied by a Hanning window. The cross spectral density between two responses was then computed as the average across segments of the discrete Fourier transform of one response with the complex conjugate of the discrete Fourier transform of the other response

$$S_{r_1,r_2}(f) = \frac{1}{M}\sum_{m=1}^{M}[F_{r_1}^{m}(f)^{*}\, F_{r_2}^{m}(f)]$$

where $S_{r_1,r_2}(f)$ is the cross spectral density between responses r_1 and r_2, M is the total number of segments, $F_{r_1}^{m}(f)^{*}$ is the complex conjugate of the discrete Fourier transform of the m^{th} segment of r_1, and $F_{r_2}^{m}(f)$ is the discrete Fourier transform of the m^{th} segment of r_2. The values for negative frequencies were discarded. The final spectral density was smoothed using a median filter with a width of 0.2 octaves and scaled such that its sum across all frequencies was equal to the total covariance between the two responses

$$\sum_{f} S_{r_1,r_2}(f) = cov(r_1, r_2).$$

Several different spectral densities were computed before and after shuffling the order of the syllables in the concatenated responses to isolate different sources of covariance as described in the Results.

PSD - the power spectral density of a single response:

r_1 = r_2= response to one trial of speech with all syllables in original order

CSD - the cross spectral density of responses to repeated identical trials:

r_1 = response to one trial of speech with all syllables in original order

r_2 = response to another trial of speech with all syllables in original order

$CSD_{shuff}^{V,T}$ - the cross spectral density of responses after shuffling of vowels and talkers, leaving the responses matched for consonants only:

r_1 = response to one trial of speech with all syllables in original order

r_2 = response to another trial of speech after shuffling of vowels and talkers

$CSD_{shuff}^{C,V,T}$ - the cross spectral density of responses after shuffling of consonants, vowels and talkers, leaving the responses matched for syllable onset only:

r_1 = response to one trial of speech with all syllables in original order

r_2 = response to another trial of speech after shuffling of consonants, vowels and talkers

CSD_0 - the cross spectral density of responses after shuffling of consonants, vowels and talkers and randomization of the phase of the Fourier transform of each response segment, leaving the responses matched for overall magnitude spectrum only:

r_1 = response to one trial of speech with all syllables in original order

r_2 = response to another trial of speech after shuffling of consonants, vowels and talkers

To isolate the differential signal component of responses to tones, the same approach was used with shuffling of frequencies.

Classification of spectrograms

To convert sound waveforms to spectrograms, they were first separated into 80 ms segments with 87.5% overlap, then multiplied by a Hamming window. The discrete Fourier transform of each segment was taken, then the magnitude was extracted and converted to a logarithmic scale. Classification was performed using a support vector machine as described above for neural responses. Only the first 150 ms of the responses to each syllable were used.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

Recordings of consonant-vowel syllables are available from the Linguistic Data Consortium (Catalog No.: LDC2015S12). Recordings of continuous speech are available from the UCL Scribe database (https://www.phon.ucl.ac.uk/resource/scribe). The database of neural recordings that were analysed in this

study is too large to be publicly shared, but is available from the corresponding author on reasonable request.

Code availability

The custom Matlab code used in this study is available at https://github.com/nicklesica/neuro.

References

1. Wilson, B. S., Tucci, D. L., Merson, M. H. & O'Donoghue, G. M. Global hearing health care: new findings and perspectives. *Lancet* **390**, 2503–2515 (2017).
2. World Health Organization. *Global costs of unaddressed hearing loss and cost-effectiveness of interventions: a WHO report*. (2017).
3. Livingston, G. *et al.* Dementia prevention, intervention, and care: 2020 report of the Lancet Commission. *Lancet* **396**, 413–446 (2020).
4. McCormack, A. & Fortnum, H. Why do people fitted with hearing aids not wear them? *Int. J. Audiol.* **52**, 360–368 (2013).
5. Orji, A. *et al.* Global and regional needs, unmet needs and access to hearing aids. *Int. J. Audiol.* **59**, 166–172 (2020).
6. Humes, L. E. Speech understanding in the elderly. *J. Am. Acad. Audiol.* **7**, 161–167 (1996).
7. Humes, L. E. & Dubno, J. R. Factors Affecting Speech Understanding in Older Adults. in *The Aging Auditory System* (eds. Gordon-Salant, S., Frisina, R. D., Popper, A. N. & Fay, R. R.) 211–257 (Springer New York, 2010). doi:10.1007/978-1-4419-0993-0_8.
8. Humes, L. E. *et al.* A comparison of the aided performance and benefit provided by a linear and a two-channel wide dynamic range compression hearing aid. *J. Speech Lang. Hear. Res.* **42**, 65–79 (1999).
9. Larson, V. D. *et al.* Efficacy of 3 commonly used hearing aid circuits: A crossover trial. NIDCD/VA Hearing Aid Clinical Trial Group. *JAMA* **284**, 1806–1813 (2000).
10. Moore, B. C. J. *Cochlear Hearing Loss: Physiological, Psychological and Technical Issues*. (John Wiley & Sons, 2007).
11. Henry, K. S. & Heinz, M. G. Effects of sensorineural hearing loss on temporal coding of narrowband and broadband signals in the auditory periphery. *Hear. Res.* **303**, 39–47 (2013).
12. Lorenzi, C., Gilbert, G., Carn, H., Garnier, S. & Moore, B. C. J. Speech perception problems of the hearing impaired reflect inability to use temporal fine structure. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 18866–18869 (2006).
13. Horvath, D. & Lesica, N. A. The Effects of Interaural Time Difference and Intensity on the Coding of Low-Frequency Sounds in the Mammalian Midbrain. *J. Neurosci.* **31**, 3821–3827 (2011).
14. Studebaker, G. A., Sherbecoe, R. L., McDaniel, D. M. & Gwaltney, C. A. Monosyllabic word recognition at higher-than-normal speech and noise levels. *J. Acoust. Soc. Am.* **105**, 2431–2444 (1999).
15. Wong, J. C., Miller, R. L., Calhoun, B. M., Sachs, M. B. & Young, E. D. Effects of high sound levels on responses to the vowel 'eh' in cat auditory nerve. *Hear. Res.* **123**, 61–77 (1998).
16. Nelson, D. A. High-level psychophysical tuning curves: forward masking in normal-hearing and hearing-impaired listeners. *J. Speech Hear. Res.* **34**, 1233–1249 (1991).
17. Ching, T. Y., Dillon, H. & Byrne, D. Speech recognition of hearing-impaired listeners: predictions from audibility and the limited role of high-frequency amplification. *J. Acoust. Soc. Am.* **103**, 1128–1140 (1998).
18. Lee, L. W. & Humes, L. E. Evaluating a speech-reception threshold model for hearing-impaired listeners. *J. Acoust. Soc. Am.* **93**, 2879–2885 (1993).
19. Oxenham, A. J. & Kreft, H. A. Speech Masking in Normal and Impaired Hearing: Interactions Between Frequency Selectivity and Inherent Temporal Fluctuations in Noise. *Adv. Exp. Med. Biol.* **894**, 125–132 (2016).
20. Summers, V. & Cord, M. T. Intelligibility of speech in noise at high presentation levels: effects of hearing loss and frequency region. *J. Acoust. Soc. Am.* **122**, 1130–1137 (2007).
21. Lesica, N. A. Why Do Hearing Aids Fail to Restore Normal Auditory Perception? *Trends Neurosci.* **41**, 174–185 (2018).
22. Souza, P. E. Effects of Compression on Speech Acoustics, Intelligibility, and Sound Quality. *Trends Amplif.* **6**, 131–165 (2002).

23. Kates, J. M. Understanding compression: modeling the effects of dynamic-range compression in hearing aids. *Int. J. Audiol.* **49**, 395–409 (2010).
24. Young, E. D. Neural Coding of Sound with Cochlear Damage. in *Noise-Induced Hearing Loss* (eds. Le Prell, C. G., Henderson, D., Fay, R. R. & Popper, A. N.) vol. 40 87–135 (Springer New York, 2012).
25. Mesgarani, N., David, S. V., Fritz, J. B. & Shamma, S. A. Phoneme representation and classification in primary auditory cortex. *J. Acoust. Soc. Am.* **123**, 899–909 (2008).
26. Heinz, M. G., Issa, J. B. & Young, E. D. Auditory-nerve rate responses are inconsistent with common hypotheses for the neural correlates of loudness recruitment. *J. Assoc. Res. Otolaryngol. JARO* **6**, 91–105 (2005).
27. Liberman, M. C., Dodds, L. W. & Learson, D. A. Structure-Function Correlation in Noise-Damaged Ears: A Light and Electron-Microscopic Study. in *Basic and Applied Aspects of Noise-Induced Hearing Loss* (eds. Salvi, R. J., Henderson, D., Hamernik, R. P. & Colletti, V.) 163–177 (Springer US, 1986). doi:10.1007/978-1-4684-5176-4_12.
28. Miller, G. A. & Nicely, P. E. An Analysis of Perceptual Confusions Among Some English Consonants. *J. Acoust. Soc. Am.* **27**, 338–352 (1955).
29. Phatak, S. A. & Allen, J. B. Consonant and vowel confusions in speech-weighted noise. *J. Acoust. Soc. Am.* **121**, 2312–2326 (2007).
30. Moore, B. C. & Glasberg, B. R. Auditory filter shapes derived in simultaneous and forward masking. *J. Acoust. Soc. Am.* **70**, 1003–1014 (1981).
31. Shera, C. A., Guinan, J. J. & Oxenham, A. J. Revised estimates of human cochlear tuning from otoacoustic and behavioral measurements. *Proc. Natl. Acad. Sci.* **99**, 3318–3323 (2002).
32. Sumner, C. J. *et al.* Mammalian behavior and physiology converge to confirm sharper cochlear tuning in humans. *Proc. Natl. Acad. Sci.* **115**, 11322–11326 (2018).
33. Dubno, J. R., Horwitz, A. R. & Ahlstrom, J. B. Estimates of Basilar-Membrane Nonlinearity Effects on Masking of Tones and Speech. *Ear Hear.* **28**, 2–17 (2007).
34. Lopez-Poveda, E. A., Plack, C. J., Meddis, R. & Blanco, J. L. Cochlear compression in listeners with moderate sensorineural hearing loss. *Hear. Res.* **205**, 172–183 (2005).
35. Plack, C. J., Drga, V. & Lopez-Poveda, E. A. Inferred basilar-membrane response functions for listeners with mild to moderate sensorineural hearing loss. *J. Acoust. Soc. Am.* **115**, 1684–1695 (2004).
36. Dubno, J. R., Ahlstrom, J. B., Wang, X. & Horwitz, A. R. Level-Dependent Changes in Perception of Speech Envelope Cues. *J. Assoc. Res. Otolaryngol.* **13**, 835–852 (2012).
37. Hornsby, B. W. Y., Trine, T. D. & Ohde, R. N. The effects of high presentation levels on consonant feature transmission. *J. Acoust. Soc. Am.* **118**, 1719–1729 (2005).
38. Zurek, P. M. & Delhorne, L. A. Consonant reception in noise by listeners with mild and moderate sensorineural hearing impairment. *J. Acoust. Soc. Am.* **82**, 1548–1559 (1987).
39. Woods, D. L., Yund, E. W. & Herron, T. J. Measuring consonant identification in nonsense syllables, words, and sentences. *J. Rehabil. Res. Dev.* **47**, 243–260 (2010).
40. Woods, D. L. *et al.* Aided and unaided speech perception by older hearing impaired listeners. *PloS One* **10**, e0114922 (2015).
41. Parthasarathy, A., Hancock, K. E., Bennett, K., DeGruttola, V. & Polley, D. B. Bottom-up and top-down neural signatures of disordered multi-talker speech perception in adults with normal hearing. *eLife* **9**, e51419 (2020).
42. Baer, T., Moore, B. C. & Gatehouse, S. Spectral contrast enhancement of speech in noise for listeners with sensorineural hearing impairment: effects on intelligibility, quality, and response times. *J. Rehabil. Res. Dev.* **30**, 49–72 (1993).
43. May, T., Kowalewski, B. & Dau, T. Signal-to-Noise-Ratio-Aware Dynamic Range Compression in Hearing Aids. *Trends Hear.* **22**, (2018).
44. Miller, R. L., Calhoun, B. M. & Young, E. D. Contrast enhancement improves the representation of /epsilon/-like vowels in the hearing-impaired auditory nerve. *J. Acoust. Soc. Am.* **106**, 2693–2708 (1999).
45. Moore, B. C. Enhancement of spectral contrast and spectral changes as approaches to improving the intelligibility of speech in background sounds. *J. Acoust. Soc. Am.* **139**, 2043–2043 (2016).
46. Rallapalli, V. H. & Alexander, J. M. Effects of noise and reverberation on speech recognition with variants of a multichannel adaptive dynamic range compression scheme. *Int. J. Audiol.* **58**, 661–669 (2019).
47. Rasetshwane, D. M., Gorga, M. P. & Neely, S. T. Signal-processing strategy for restoration of cross-channel suppression in hearing-impaired listeners. *IEEE Trans. Biomed. Eng.* **61**, 64–75 (2014).

48. Dubno, J. R., Horwitz, A. R. & Ahlstrom, J. B. Word recognition in noise at higher-than-normal levels: decreases in scores and increases in masking. *J. Acoust. Soc. Am.* **118**, 914–922 (2005).
49. Bose. Hearphones. https://www.bose.com/en_us/support/products/bose_wellness_support/hearphones.html (2020).
50. King, A. J. & Walker, K. M. Listening in complex acoustic scenes. *Curr. Opin. Physiol.* **18**, 63–72 (2020).
51. McWalter, R. & McDermott, J. H. Adaptive and Selective Time Averaging of Auditory Scenes. *Curr. Biol.* **28**, 1405-1418.e10 (2018).
52. Cai, S., Ma, W.-L. D. & Young, E. D. Encoding intensity in ventral cochlear nucleus following acoustic trauma: implications for loudness recruitment. *J. Assoc. Res. Otolaryngol. JARO* **10**, 5–22 (2009).
53. Akeroyd, M. A. An Overview of the Major Phenomena of the Localization of Sound Sources by Normal-Hearing, Hearing-Impaired, and Aided Listeners. *Trends Hear.* **18**, (2014).
54. Brown, A. D., Rodriguez, F. A., Portnuff, C. D. F., Goupell, M. J. & Tollin, D. J. Time-Varying Distortions of Binaural Information by Bilateral Hearing Aids. *Trends Hear.* **20**, (2016).
55. Gates, G. A. & Mills, J. H. Presbycusis. *Lancet* **366**, 1111–1120 (2005).
56. Davies-Venn, E., Souza, P., Brennan, M. & Stecker, G. C. Effects of audibility and multichannel wide dynamic range compression on consonant recognition for listeners with severe hearing loss. *Ear Hear.* **30**, 494–504 (2009).
57. Shanks, J. E., Wilson, R. H., Larson, V. & Williams, D. Speech recognition performance of patients with sensorineural hearing loss under unaided and aided conditions using linear and compression hearing aids. *Ear Hear.* **23**, 280–290 (2002).
58. Committee on Accessible and Affordable Hearing Health Care for Adults, Board on Health Sciences Policy, Health and Medicine Division & National Academies of Sciences, Engineering, and Medicine. *Hearing Health Care for Adults: Priorities for Improving Access and Affordability*. (National Academies Press, 2016).
59. Warren, E. & Grassley, C. Over-the-Counter Hearing Aids: The Path Forward. *JAMA Intern. Med.* **177**, 609–610 (2017).
60. President's Council of Advisors on Science and Technology, (PCAST). Aging America & Hearing Loss: Imperative of Improved Hearing Technologies. (2016).
61. Sinha, S., Irani, U. D., Manchaiah, V. & Bhamla, M. S. LoCHAid: An ultra-low-cost hearing aid for age-related hearing loss. *PLOS ONE* **15**, e0238922 (2020).
62. Cox, R. M., Johnson, J. A. & Xu, J. Impact of Hearing Aid Technology on Outcomes in Daily Life I: The Patients' Perspective. *Ear Hear.* **37**, e224-237 (2016).
63. Brody, L., Wu, Y.-H. & Stangl, E. A Comparison of Personal Sound Amplification Products and Hearing Aids in Ecologically Relevant Test Environments. *Am. J. Audiol.* **27**, 581–593 (2018).
64. Cho, Y. S. *et al.* Clinical Performance Evaluation of a Personal Sound Amplification Product vs a Basic Hearing Aid and a Premium Hearing Aid. *JAMA Otolaryngol Head Neck Surg* **145**, 516–522 (2019).
65. Humes, L. E. *et al.* The Effects of Service-Delivery Model and Purchase Price on Hearing-Aid Outcomes in Older Adults: A Randomized Double-Blind Placebo-Controlled Clinical Trial. *Am. J. Audiol.* **26**, 53–79 (2017).
66. Suberman, T. A. *et al.* A Gerbil Model of Sloping Sensorineural Hearing Loss. *Otol. Neurotol.* **32**, 544–552 (2011).
67. Pachitariu, M., Steinmetz, N., Kadir, S., Carandini, M. & Harris, K. Kilosort: realtime spike-sorting for extracellular electrophysiology with hundreds of channels. *bioRxiv* 061481 (2016) doi:10.1101/061481.
68. Maki, K. & Furukawa, S. Acoustical cues for sound localization by the Mongolian gerbil, Meriones unguiculatus. *J. Acoust. Soc. Am.* **118**, 872–886 (2005).
69. Alexander, J. M. & Masterson, K. Effects of WDRC release time and number of channels on output SNR and speech recognition. *Ear Hear.* **36**, e35-49 (2015).
70. Moore, B. C. J., Glasberg, B. R. & Stone, M. A. Development of a new method for deriving initial fittings for hearing aids with multi-channel compression: CAMEQ2-HF. *Int. J. Audiol.* **49**, 216–227 (2010).
71. Garcia-Lazaro, J. A., Belliveau, L. A. C. & Lesica, N. A. Independent population coding of speech with sub-millisecond precision. *J. Neurosci.* **33**, 19362–19372 (2013).

Acknowledgements
This work was supported by a Wellcome Trust Senior Research Fellowship [200942/Z/16/Z]. The authors thank J. Linden, S. Rosen, D. Fitzpatrick, B. Moore, J. Alexander, M. Huckvale, K. Harris, G. Huang, T. Keck and R. Beutelmann for their advice.

Author contributions
N.A.L. and C.C.L. conceived and designed the experiments. N.A.L., C.C.L., A.A., and S.S. performed the experiments. N.A.L. analyzed the data and wrote the paper.

Competing Interests
N.A.L. is a co-founder of Perceptual Technologies Ltd. A.A., C.C.L. and S.S. declare no competing interests.

Additional information
Supplementary information is available for this paper at https://doi.org/10.1038/s41551-01X-XXXX-X.
Correspondence and requests for materials should be addressed to N.A.L.
Peer review information *Nature Biomedical Engineering* thanks Hubert Lim, David McAlpine and the other, anonymous, reviewers for their contribution to the peer review of this work.

Figure captions

Fig. 1 | Large-scale recordings of neural activity from the inferior colliculus with normal hearing and mild-to-moderate hearing loss. a, Schematic diagram showing the geometry of custom-designed electrode arrays for large-scale recordings in relation to the inferior colliculus in gerbils. **b**, Threshold shifts with hearing loss and corresponding hearing aid amplification. Top: Hearing loss as a function of frequency in noise-exposed gerbils (mean ± standard error, n = 20). The values shown are the ABR threshold shift relative to the mean of all gerbils (n = 15) with normal hearing. Bottom: Hearing aid amplification as a function of frequency for speech at 62 dB SPL with gain and compression parameters fit to the average hearing loss after noise exposure. The values shown are the average across 5 minutes of continuous speech. **c**, MUA recorded in the inferior colliculus during the presentation of tones. Left, The MUA FRAs for 16 channels from a normal hearing gerbil. Each subplot shows the average activity recorded from a single channel during the presentation of tones with different frequencies and intensities. The colormap for each plot is normalized to the minimum and maximum activity level across all frequencies and intensities. Middle: MUA FRAs for 16 channels from a gerbil with hearing loss. Right, The average MUA FRAs across all channels from all gerbils for each hearing condition. The lines indicate the lowest intensity for each frequency at which the mean MUA was more than 3 standard deviations above the mean MUA during silence. The line for normal hearing is shown in blue on all three subplots.

Fig. 2 | Single-trial responses to speech can be classified with high accuracy. a, Single-unit responses to speech. Each column shows the sound waveform for one instance of a syllable and the corresponding raster plots for repeated presentations of that syllable for two example neurons from a gerbil with normal hearing. **b**, Left, Low-dimensional visualization of population single-trial responses to speech. Each line shows the responses from all neurons from all gerbils with normal hearing after principal component decomposition and projection into the space defined by the first three principal components. Responses to two repeated presentations for each of three syllables (indicated by the three colours) are shown. The time points corresponding to syllable onset are indicated by t = 0 s. Right, Low-dimensional visualization of mean population response to each consonant. Each line shows responses as in b after averaging across all presentations of syllables with the same consonant. Mean responses to each of 12 consonants are shown, with colours corresponding to consonant categories: sibilant fricatives (orange), stops (pink), and nasals and non-sibilant fricatives (blue). **c**, Performance and confusion patterns for a support-vector-machine classifier

trained to identify consonants based on population single-trial responses to speech at 62 dB SPL. Left, Each row shows the frequency with which responses to one consonant were identified as that consonant (diagonal entries) or other consonants (off-diagonal entries) by the classifier. The values on the diagonal entries are the F1 score computed as 2 x (precision x recall) / (precision + recall), where precision = true positives / (true positives + false positives) and recall = true positives / (true positives + false negatives). The values shown are the average across all populations. Right, Consonants were assigned angles along a unit circle (indicated by black letters). For each single-trial response for a given actual consonant, a vector was formed with magnitude 1 and angle corresponding to the consonant that the response was identified as by the classifier. The positions of the coloured letters indicate the sum of these vectors across all responses for each consonant.

Fig. 3 | Hearing aids restore speech audibility but not consonant identification. a, Single-unit responses to speech. Each column shows the sound waveform for one instance of a syllable and the corresponding raster plots for repeated presentations of that syllable for two example neurons from a gerbil with hearing loss, without and with a hearing aid. **b**, Left, Spike rate of single-unit responses to speech at 62 dB SPL. Results are shown for neurons from normal hearing gerbils (NH) and gerbils with hearing loss without (HL) and with (HA) a hearing aid (mean ± 95% confidence intervals derived from bootstrap resampling across neurons; *** indicates $p < 0.001$, ** indicates $p < 0.01$, * indicates $p < 0.05$, ns indicates not significant; for sample sizes and details of statistical tests for all figures, see Table S1). Right, Performance of a support-vector-machine classifier trained to detect speech at 62 dB SPL in silence based on individual single-unit responses (the first 150 ms of single-trial responses represented as spike counts with 5 ms time bins), presented as in the panel on the left. **c**, Performance of a support-vector-machine classifier trained to identify consonants based on population single-trial responses to speech at 62 dB SPL. Results are shown for three conditions: speech in quiet, speech in the presence of ongoing speech from a second talker at equal intensity, and speech in the presence of multi-talker babble noise at equal intensity (values for each population are shown along with mean ± 95% confidence intervals derived from bootstrap resampling across populations).

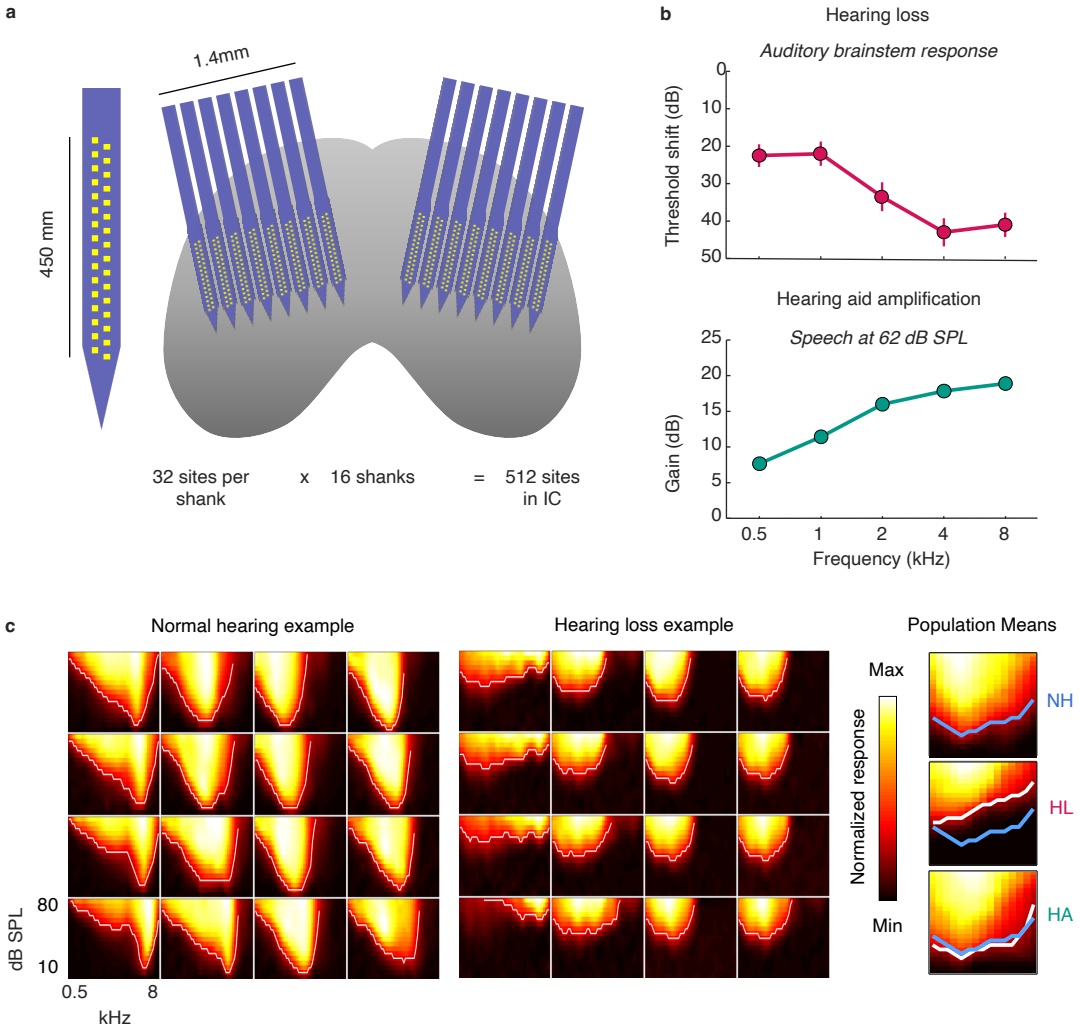
Fig. 4 | Hearing aids fail to restore the selectivity of neural responses to speech. a, The different signal and noise components of neural responses in the context of a consonant identification task. **b**, Spectral decomposition of responses for an example neuron. Each line shows a power spectral density or cross spectral density computed from responses before and after different forms of shuffling, and each filled area indicates the fraction of the total response variance corresponding to each response component. The superscripts *C*, *V*, and *T* denote consonants, vowels, and talkers, respectively. CSD_0 was computed after shuffling and phase randomization in the spectral domain. **c**, Magnitude of different response components for single-unit responses to speech at 62 dB SPL. Results are shown for neurons from normal hearing gerbils (NH) and gerbils with hearing loss without (HL) and with (HA) a hearing aid (mean ± 95% confidence intervals derived from bootstrap resampling across neurons). **d**, Left, Magnitude of the differential signal component as a function of frequency for single-unit responses to speech at 62 dB SPL (mean ± 95% confidence intervals derived from bootstrap resampling across neurons indicated by shaded regions). Right, Magnitude of the differential signal component for single-unit spike counts (the total number of spikes in the response to each syllable) for speech at 62 dB SPL, presented as in **c**.

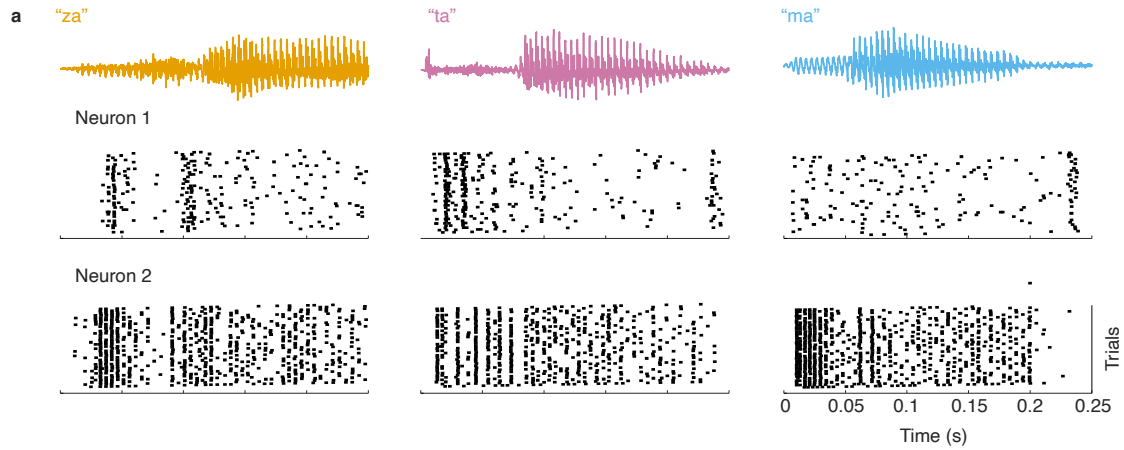
Fig. 5 | Hearing aids restore the selectivity of neural responses to tones. a, Single-unit responses to tones. The FRA for an example single-unit from a gerbil with hearing loss showing the mean spike rate during the presentation of tones with different frequencies and intensities without (left) and with (right) a hearing aid. The center frequency (CF) and threshold (white dot) are indicated. The lines indicate the lowest intensity for each frequency at which the response was significantly greater than responses recorded during silence (probability of observed spike count $p < 0.01$ assuming Poisson-distributed counts; no correction was made for multiple comparisons). The colormap for each plot is normalized to the minimum and maximum spike rate across all frequencies and intensities. **b**, Threshold shift as a function of frequency for single-unit responses to tones. Results are shown for neurons (n = 2664) from gerbils with hearing loss without (HL) and with (HA) a hearing aid. The values shown are the threshold shift relative to the mean of all neurons from all gerbils with normal hearing (mean ± 95% confidence intervals derived from bootstrap resampling across neurons). **c**, Left, Spike rate of single-unit responses to tones at 62 dB SPL. Results are shown for

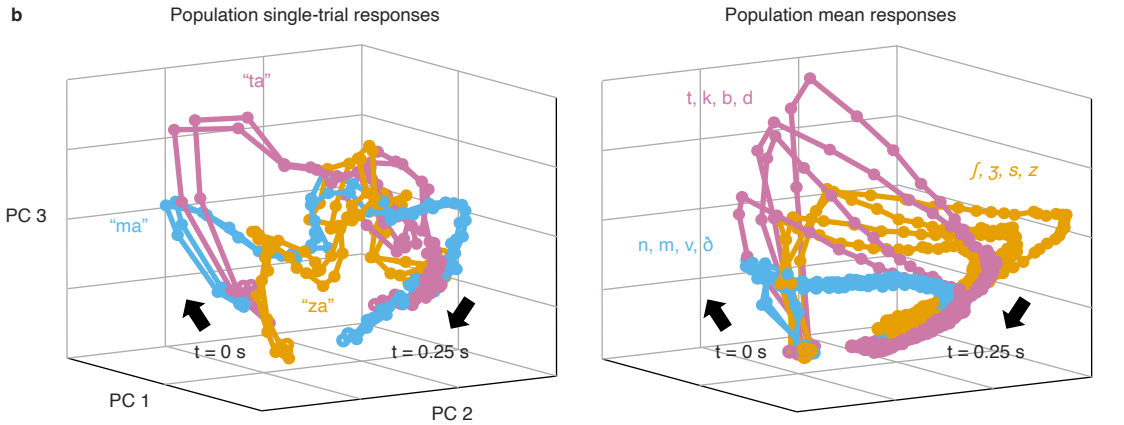
neurons from normal hearing gerbils (NH) and gerbils with hearing loss without (HL) and with (HA) a hearing aid (mean ± 95% confidence intervals derived from bootstrap resampling across neurons). Middle left, Tuning width of single-unit responses to tones at 14 dB above threshold, presented as in the leftmost panel. The values shown are the range of frequencies for which the mean spike rate during the presentation of a tone was at least half of its maximum value across all frequencies. Middle, Tuning width of single-unit responses to tones at 62 dB SPL, presented as in the leftmost panel. Middle right, Magnitude of the differential signal component for single-unit responses to tones at 62 dB SPL, presented as in leftmost panel. Right, Performance of a support-vector-machine classifier trained to identify tone frequency based on population single-trial responses (represented as spike counts with 5 ms time bins) to tones at 62 dB SPL (mean ± 95% confidence intervals derived from bootstrap resampling across populations). Populations of 10 neurons were formed by sampling at random, without replacement, from neurons from all gerbils until there were no longer enough neurons remaining to form another population. A population size of 10 was used to allow for accurate classifier performance for all conditions while avoiding the 100% ceiling for any condition.

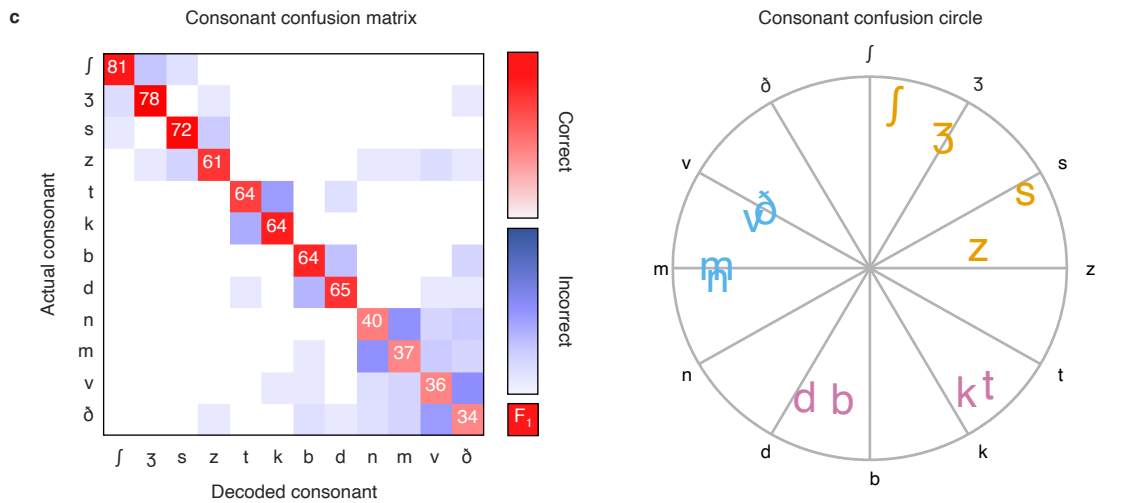
Fig. 6 | Hearing aid compression decreases the selectivity of neural responses to speech. a, Spectrograms showing the log power across frequencies at each time point in one instance of the syllable "za" before and after processing with a hearing aid. **b**, Left, percent change in RMS contrast of all syllables (n = 384 instances with 16 consonants followed by each of 4 vowels spoken by each of 8 talkers) after processing with a hearing aid. Only the first 150 ms of each syllable were used. Right, Performance of a support-vector-machine classifier trained to identify consonants based on spectrograms either before (Original) or after (HA) processing with a hearing aid (mean ± standard error across 10 different held-out samples). **c**, Left, Performance of a support-vector-machine classifier trained to identify consonants based on population single-trial responses to speech at 62 dB SPL. Results are shown for normal hearing gerbils (NH) and gerbils with hearing loss without a hearing aid (HL), with a hearing aid (HA), and with linear amplification (HL+20dB) (values for each population are shown along with mean ± 95% confidence intervals derived from bootstrap resampling across populations). Right**,** Magnitude of the differential signal component for single-unit responses to speech at 62 dB SPL (mean ± 95% confidence intervals derived from bootstrap resampling across neurons). **d**, Performance of a support-vector-machine classifier trained to identify consonants based on population single-trial responses to speech at 62 dB SPL. Results are shown for speech in the presence of ongoing speech from a second talker at equal intensity and speech in the presence of multi-talker babble noise at equal intensity, presented as in **c**.

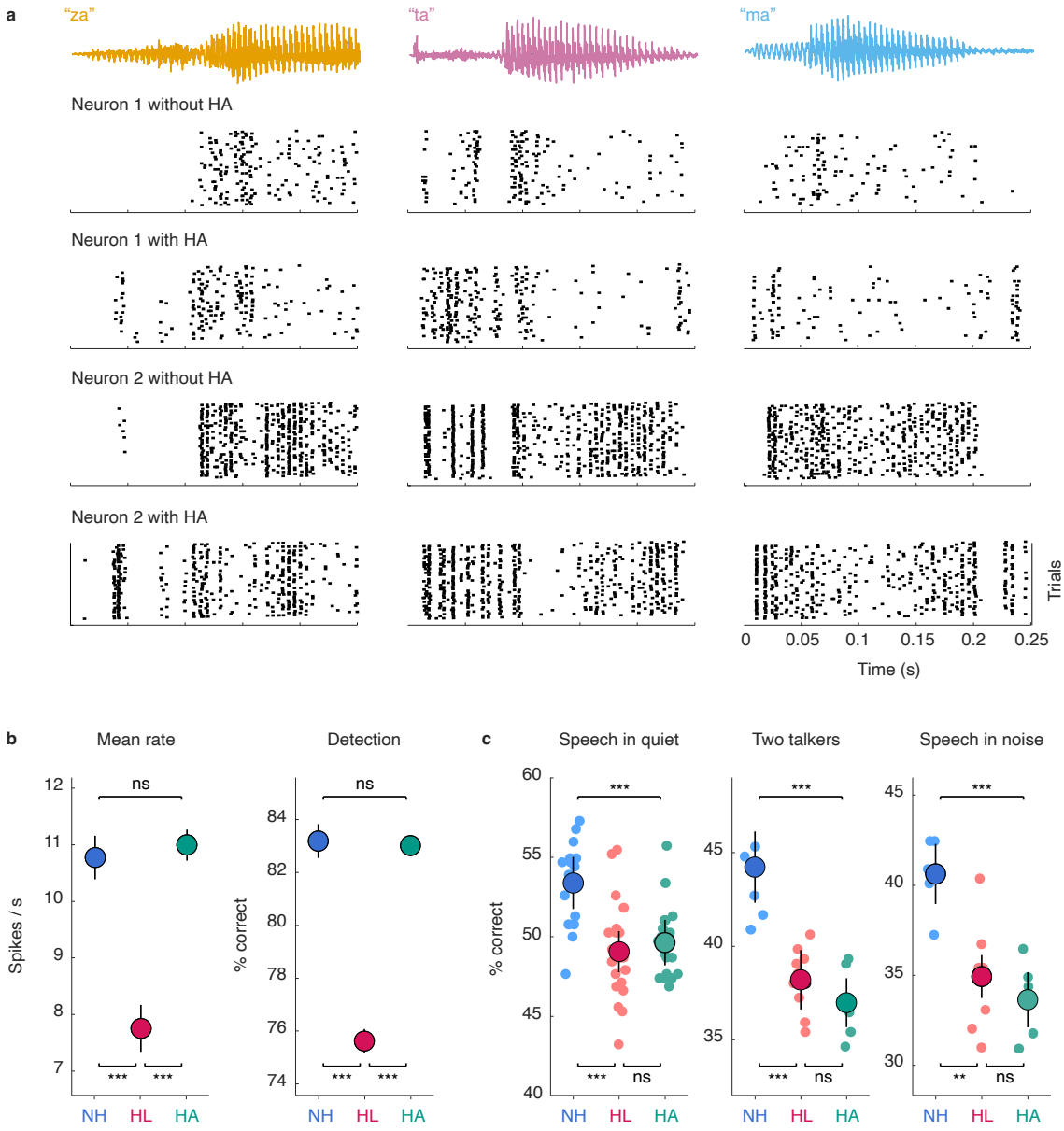
Fig. 7 | Amplification decreases consonant identification even with normal hearing. a, Performance of a support-vector-machine classifier trained to identify consonants based on population single-trial responses to speech at 62 dB SPL. Results are shown for normal hearing gerbils without (NH) and with (NH+20dB) linear amplification (values for each population are shown along with mean ± 95% confidence intervals derived from bootstrap resampling across populations) for three conditions: speech in quiet, speech in the presence of ongoing speech from a second talker at equal intensity, and speech in the presence of multi-talker babble noise at equal intensity. **b**, Magnitude of different response components for single-unit responses to speech at 62 dB SPL in multi-talker babble noise (mean ± 95% confidence intervals derived from bootstrap resampling across neurons). **c**, Performance of a support-vector-machine classifier trained to identify consonants based on population single-trial responses to speech in noise at 62 dB SPL with linear amplification. Results are shown for normal hearing gerbils (NH+20dB) and gerbils with hearing loss (HL+20dB), presented as in **a**. **d**, Schematic showing the effects of intensity on speech intelligibility with and without hearing loss and amplification. The range of intensities of typical speech is shown in gray. Left, The loss of intelligibility with hearing loss that results from the loss of audibility without amplification. Right, The loss of intelligibility with hearing loss that results from rollover with amplification.

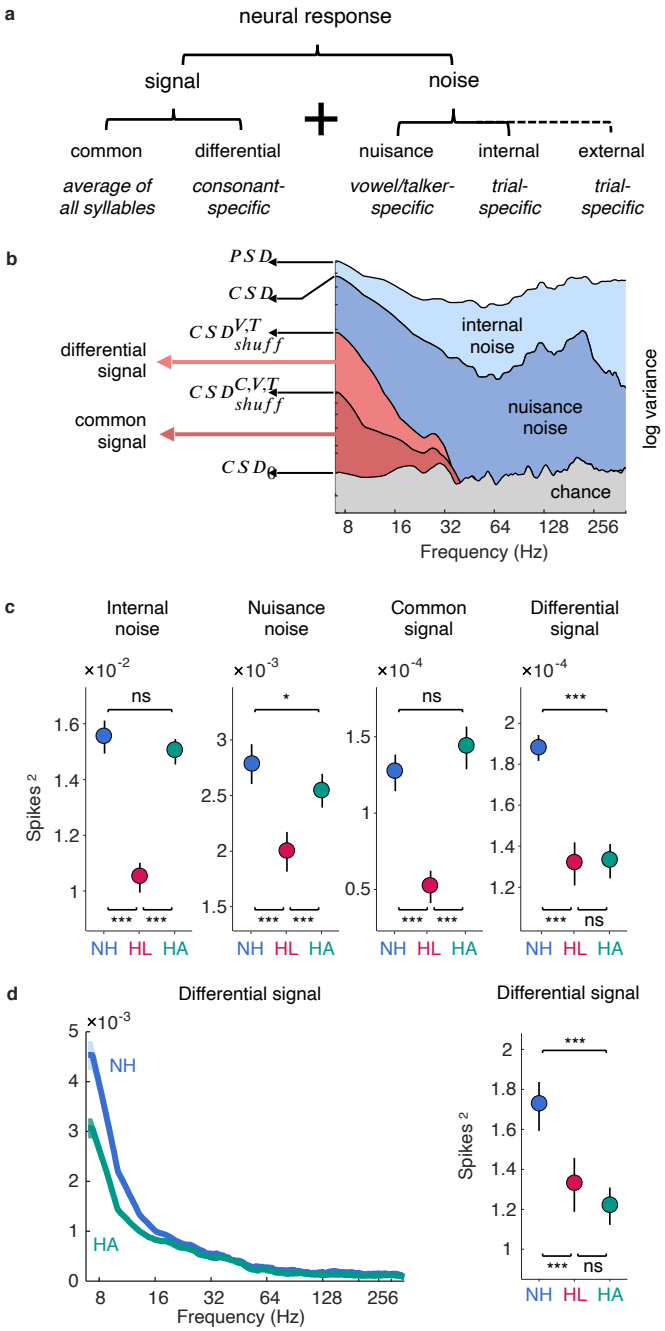

a
1.4mm
450 mm
32 sites per shank
x
16 shanks
=
512 sites in IC
b
Hearing loss
Auditory brainstem response
Threshold shift (dB)
0
10
20
30
40
50
Hearing aid amplification
Speech at 62 dB SPL
Gain (dB)
25
20
15
10
5
0
0.5
1
2
4
8
Frequency (kHz)
c
Normal hearing example
Hearing loss example
Population Means
Max
Normalized response
Min
NH
HL
HA
dB SPL
80
10
0.5
8
kHz

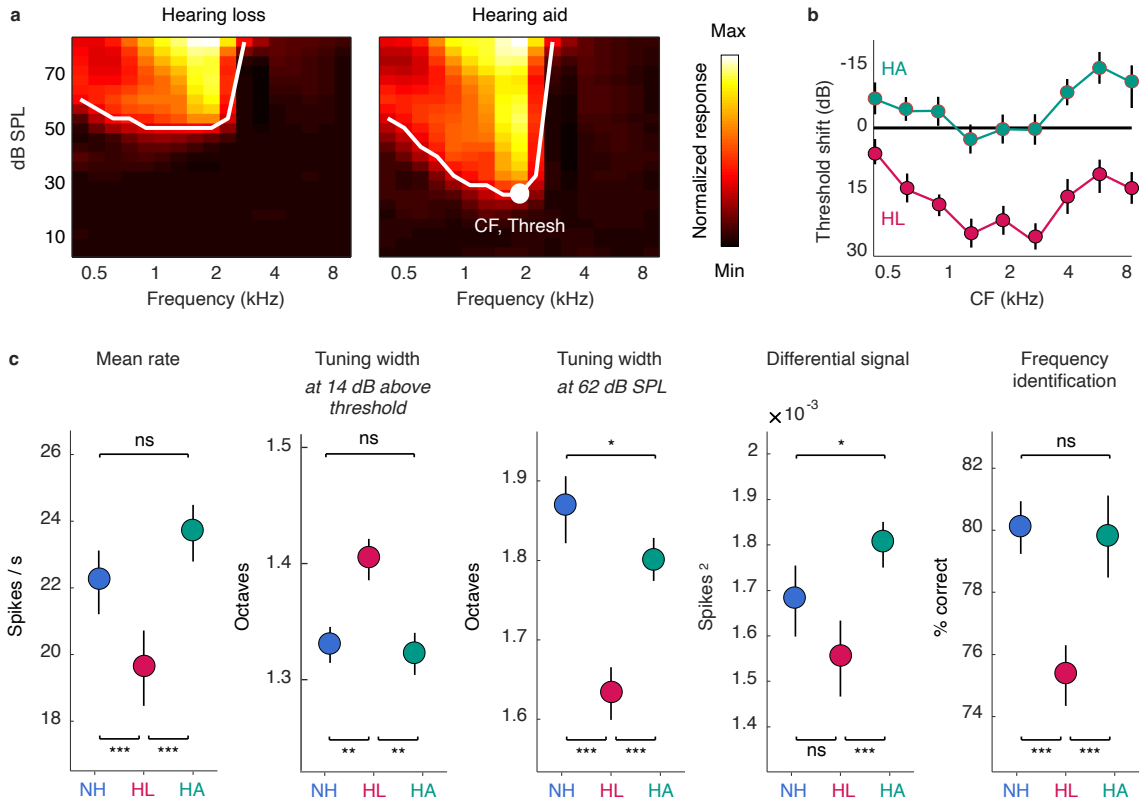

a
“za”
“ta”
“ma”
Neuron 1
Neuron 2
Trials
0
0.05
0.1
0.15
0.2
0.25
Time (s)

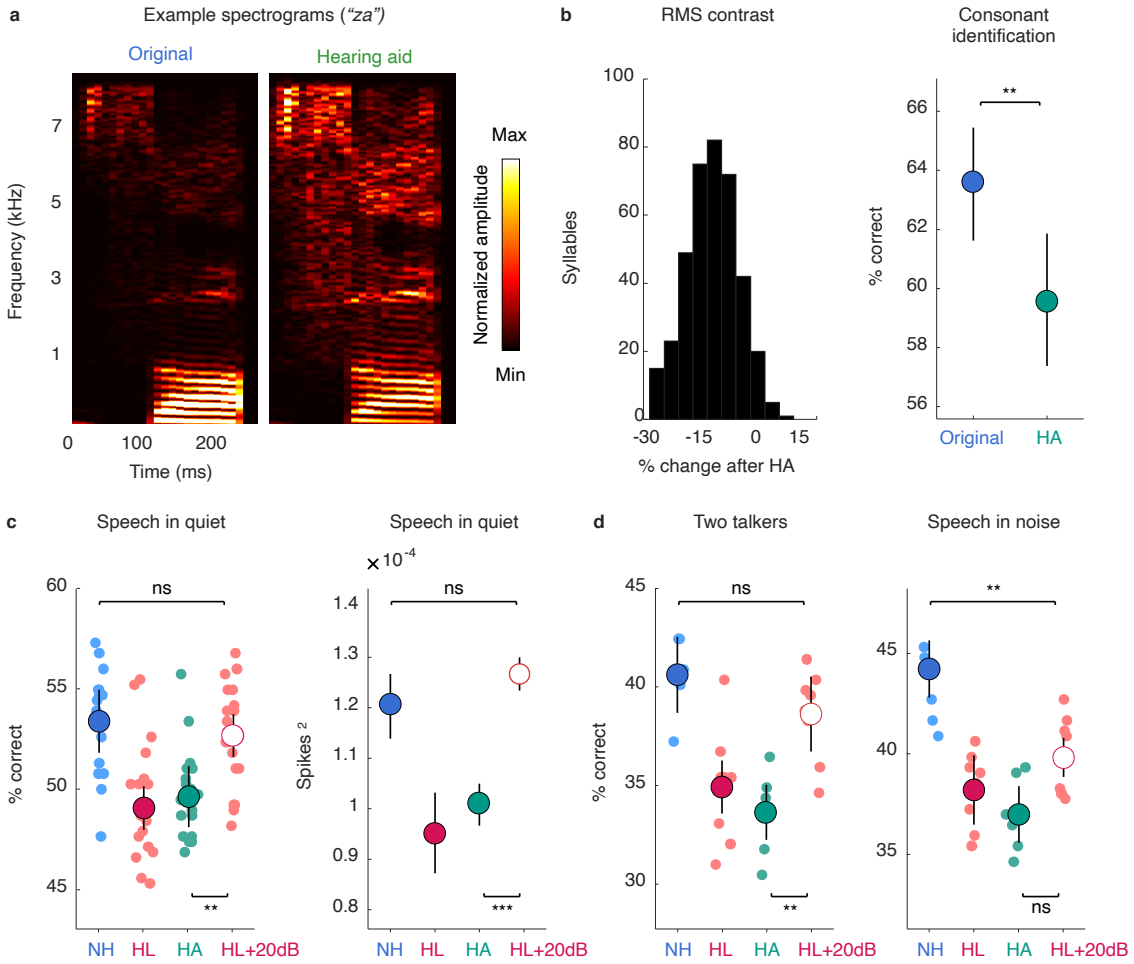

b
Population single-trial responses
Population mean responses
PC 3
PC 1
PC 2
“ta”
“ma”
“za”
t = 0 s
t = 0.25 s
t, k, b, d
ʃ, ʒ, s, z
n, m, v, ð

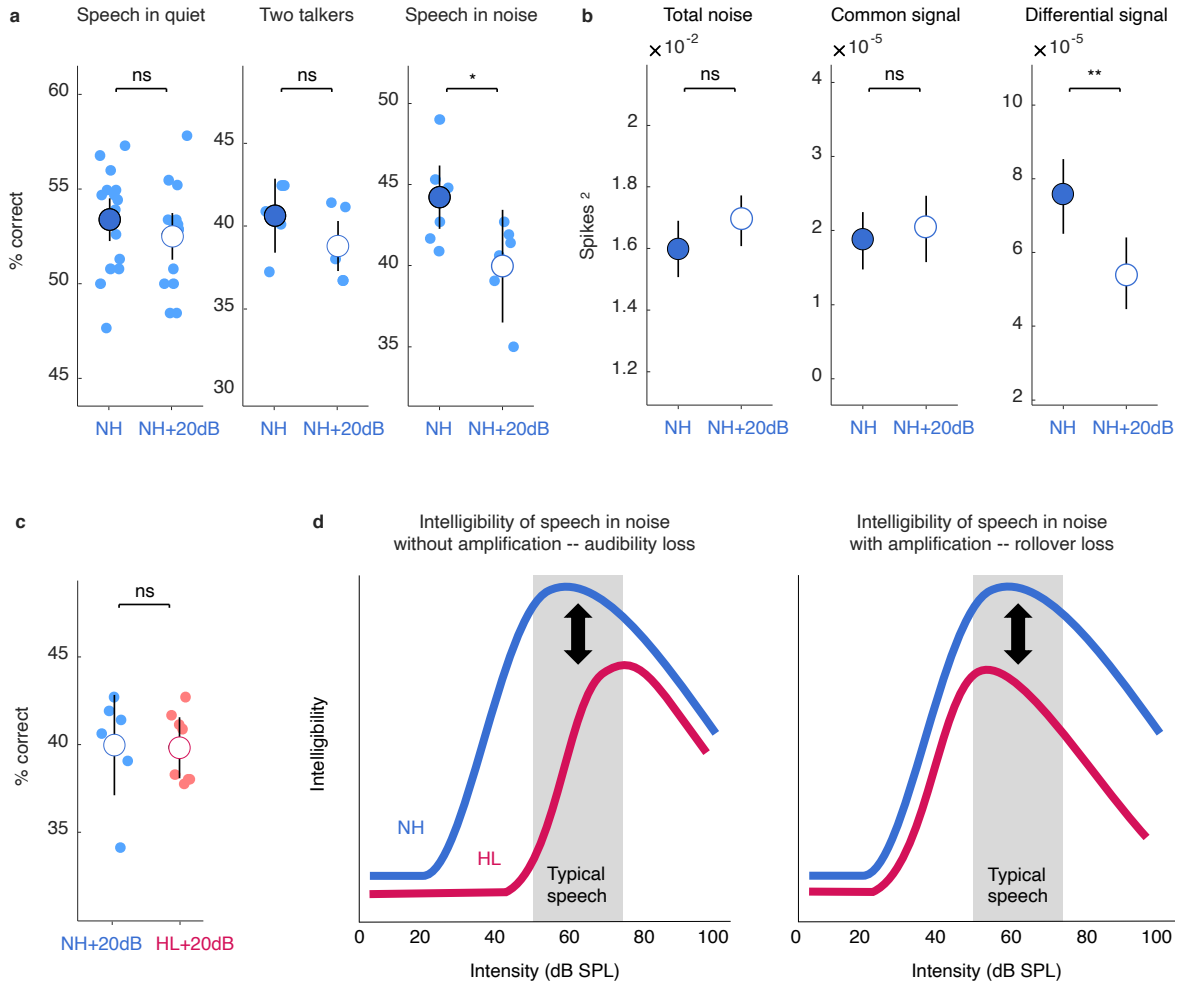

c
Consonant confusion matrix
Actual consonant
Decoded consonant
Correct
Incorrect
F1
ʃ ʒ s z t k b d n m v ð
81 78 72 61 64 64 64 65 40 37 36 34
Consonant confusion circle


a
"za"
"ta"
"ma"
Neuron 1 without HA
Neuron 1 with HA
Neuron 2 without HA
Neuron 2 with HA
Trials
0
0.05
0.1
0.15
0.2
0.25
Time (s)
b
Mean rate
Detection
Spikes / s
% correct
ns
***
NH
HL
HA
c
Speech in quiet
Two talkers
Speech in noise
% correct
**


a
neural response
signal
noise
common
differential
nuisance
internal
external
average of all syllables
consonant-specific
vowel/talker-specific
trial-specific
trial-specific
b
PSD
CSD
CSD^{V,T}_{shuff}
CSD^{C,V,T}_{shuff}
CSD_0
differential signal
common signal
internal noise
nuisance noise
chance
log variance
8 16 32 64 128 256
Frequency (Hz)
c
Internal noise
Nuisance noise
Common signal
Differential signal
×10^-2
×10^-3
×10^-4
×10^-4
ns
*
ns
***
Spikes^2
NH HL HA
d
Differential signal
×10^-3
NH
HA
Frequency (Hz)
Differential signal
***
ns
Spikes^2
NH HL HA


a
Hearing loss
Hearing aid
dB SPL
70
50
30
10
0.5
1
2
4
8
Frequency (kHz)
CF, Thresh
Max
Normalized response
Min
b
HA
HL
-15
0
15
30
Threshold shift (dB)
CF (kHz)
c
Mean rate
Spikes / s
Tuning width
at 14 dB above threshold
Octaves
Tuning width
at 62 dB SPL
Octaves
Differential signal
×10-3
Spikes 2
Frequency identification
% correct
ns
*
**
***
NH
HL
HA


a
Example spectrograms ("za")
Original
Hearing aid
Frequency (kHz)
7
5
3
1
0
100
200
Time (ms)
Max
Normalized amplitude
Min
b
RMS contrast
Syllables
100
80
60
40
20
0
-30
-15
0
15
% change after HA
Consonant identification
% correct
66
64
62
60
58
56
**
Original
HA
c
Speech in quiet
% correct
60
55
50
45
ns
**
NH
HL
HA
HL+20dB
Speech in quiet
Spikes ²
×10⁻⁴
1.4
1.3
1.2
1.1
1
0.9
0.8
ns
***
NH
HL
HA
HL+20dB
d
Two talkers
% correct
45
40
35
30
ns
**
NH
HL
HA
HL+20dB
Speech in noise
45
40
35
**
ns
NH
HL
HA
HL+20dB


a
Speech in quiet
Two talkers
Speech in noise
% correct
ns
ns
*
NH
NH+20dB
b
Total noise
Common signal
Differential signal
Spikes ²
×10^-2
×10^-5
×10^-5
ns
ns
**
c
HL+20dB
d
Intelligibility of speech in noise
without amplification -- audibility loss
with amplification -- rollover loss
Intelligibility
NH
HL
Typical
speech
Intensity (dB SPL)