

DECODING THE AUDITORY NERVE TO SIMULATE SENSORINEURAL PATHOLOGIES AND HELP REFINE THEIR DIAGNOSIS

Jacques Grange John Culling

School of Psychology, Cardiff University, United Kingdom
Grangeja@cardiff.ac.uk

ABSTRACT

Pathologies underlying sensorineural hearing loss (SNHL) cannot yet be differentially diagnosed. An advanced SNHL simulator takes the stimulus encoding by Meddis and colleagues' Model of the Auditory Periphery [1] at the auditory nerve (AN) level and decodes it into a signal used in psychophysical tasks to determine the signatures of specific pathologies. Speech reception thresholds (SRTs) for stimuli processed through the normal-hearing (NH) simulator, were just 1 dB higher than those obtained for unprocessed stimuli.

With both efferent reflexes disabled, SRTs grew by ~4 dB. Simulated rate-level functions illustrate how efferent reflexes enable the AN dynamic-range adaptation to context level that prevents information loss. While deactivating 70% of ANs or halving the endo-cochlear potential led to little SRT inflation, SRTs significantly rose as a result of total outer haircells knockout. 90% deafferentation was required to reflect performance found in hearing impaired listeners, a finding consistent with the expected effect of stochastic under-sampling [2]. ITD discrimination thresholds were found to be more sensitive to deafferentation than SRTs.

1. THE SIMULATOR

Our simulator makes use of a readily-available physiological model of the auditory periphery (MAP, version 1.14j) [1]. In its normal-hearing implementation, it models the outer and middle ear, the dual-resonance non-linear (DRNL) model of basilar membrane vibration [3], inner haircell transduction from stereocilia flexing to release of neurotransmitter vesicles in the cleft, AN activity across 30,000 fibers arranged over 30 best frequencies (BFs, equally spread on an ERB scale between 56 and 8000 Hz) and over 3 AN-fiber spontaneous rates (low, LSR; medium, MSR; and high, HSR fibers), and a simplified, two-layer brainstem model designed to generate both acoustic and medial olivocochlear efferent reflexes (AR and MOCR). Hence, AR and MOCR constitute feedback loops that moderate stapedial displacement and basilar membrane vibrations [via outer haircells (OHCs)], respectively. Modelled spike trains are then fed through gammatone filters centered on corresponding BFs to generate wavelets. Computed wavelet trains are then summed across BFs and SRs. Example rate-level functions (RLFs) are provided in Fig. 1, that illustrate the importance of dynamic-range

adaptation to context level [4] to the faithful coding of e.g. speech.

Summing AN response across SRs leads to a compressive response. Dynamic range adaptation to context level and resulting signal compression were found to be primarily the result of efferent reflexes. Therefore, it seemed appropriate, in order to linearize the response of a NH implementation of the simulator, that the signal be expanded by multiplying it with inverted AR and MOCR signals. A validation of the simulator was first obtained that demonstrated good simulation of normal hearing.

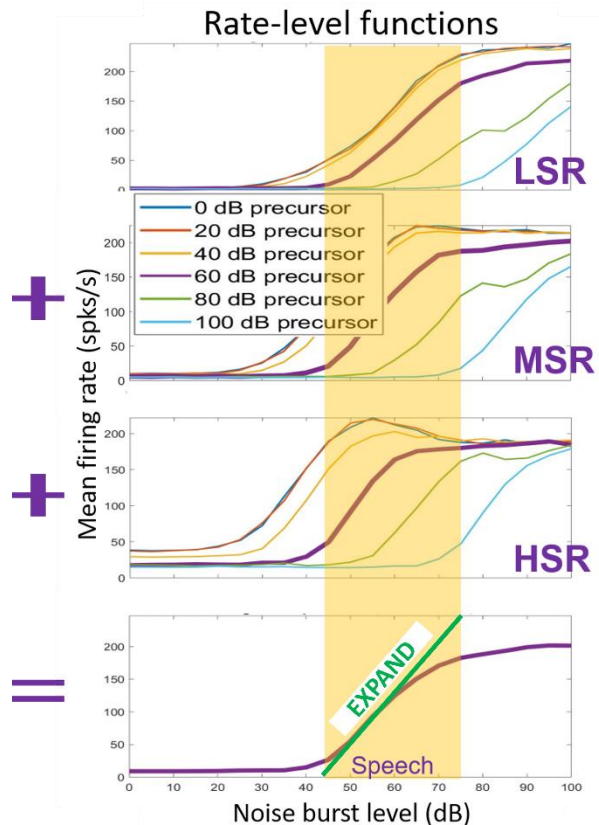


Figure 1. Example MAP-computed rate-level functions of AN fibers centered on 1000 Hz for a 50 ms noise burst, as a function of context level set by a 400ms noise precursor and for LSR, MSR and HSR fibers (three top panels). Illustration of how the simulator sums activity across fibers (bottom panel) and, for speech at 60 dB SPL, how the resulting response is compressive, thereby showing the need to expand the signal back to linearize the response of a NH implementation of the simulator.

2. SIMULATOR VALIDATION

Our first step was to validate the simulator by demonstrating high intelligibility of speech processed through a NH simulator implementation. 12 young NH adults performed a digit-triplets recognition task that enabled the adaptive measure of speech-reception thresholds (SRTs) in a steady noise colored to the target female voice. Listening conditions comprised: efferent reflexes disabled; efferent reflexes enabled without or with three types of expansion; and the original, unprocessed signal, simply band-pass filtered to the frequency bandwidth covered by the simulator. Fig. 2 illustrates resulting SRTs.

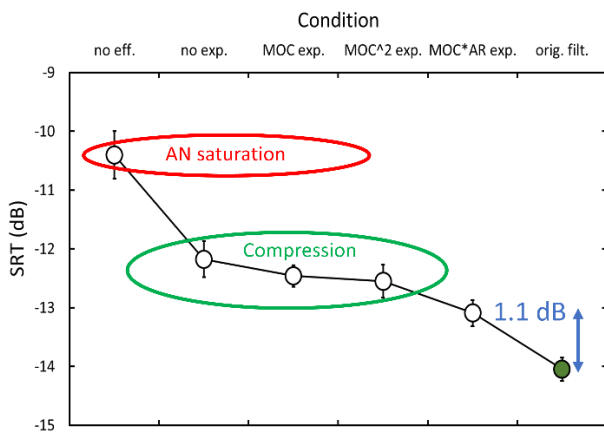


Figure 2. Digit-triplet SRTs obtained with a just-filtered original signal (orig, filt.) and simulated hearing with efferent reflexes disabled (no eff.), with them enabled but without expansion (no exp.) and with expansion based on three inverted efferent signals (MOC, MOC² and MOC*AR exp.). Standard errors of means are showed.

The absence of dynamic range adaptation caused by disabling efferent reflexes led to the largest SRT elevation (~4 dB), most likely due to the saturation of MSR and HSR fibers (ref. Fig.1). The expansion that employed both efferent signals led to the best processed SRT, only 1.1 dB above the unprocessed SRT. This constituted validation of the simulator. As removing efferent signals elevated SRTs to levels seen in mild-moderately impaired individuals, this outcome also illustrates how important efferent signals are to lossless coding of speech. Simulated NH hereafter made use of the best expansion (MOCR*AR exp.).

3. SIMULATING PATHOLOGIES

3.1 Effect of pathologies on speech intelligibility

10 young NH adults performed the digit-triplet SRT task, comparing the unprocessed condition to simulated normal hearing and three pathological conditions: 70% deafferentation (even AN-fiber knockout across BFs and SRs), all OHCs knocked out and endocochlear potential dropped from 100 to 50 mV. All three pathologies are believed to worsen with age. Fig. 3 shows resulting SRTs. While the NH-processed condition elevated SRTs more

than at validation stage, the only pathology that showed a significant SRT elevation, compared to NH-processed, was the absence of OHCs (and hence, no MOCR). This outcome suggests that speech intelligibility is robust to significant deafferentation or endocochlear potential drop but relies on the AN adaptation the MOCR provides.

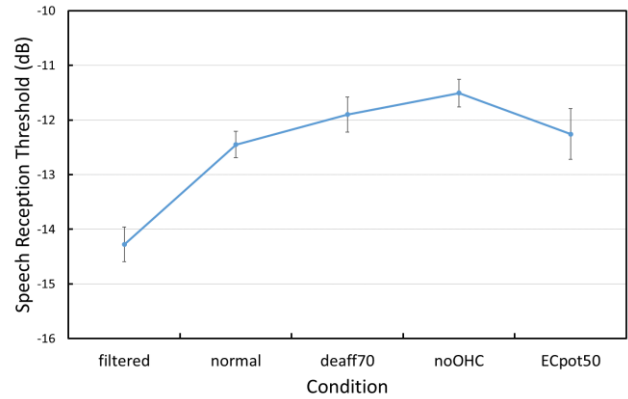


Figure 3. Digit-triplet SRTs obtained with just band-pass filtered signal (filtered), simulated NH hearing (normal), simulated 70% deafferentation (deaff70), 100% OHC knockout (noOHC) and 50% drop in endocochlear potential (ECpot50). Standard errors of means showed.

3.2 Deafferentation and speech intelligibility

Given the little impact on SRT found at 70% deafferentation, the next study set out to systematically investigate deafferentation. 10 young NH adults performed the digit-triplet SRT task at gradually increasing levels of deafferentation. A factor four in successive reduction in remaining AN fibers was applied, deafferentation ranging from 0 to 99.7%, as can be seen in Fig. 4. A highly significant effect of deafferentation was found, but the most important finding is that it took 90% of deafferentation to lead to a reliably measurable SRT elevation. This suggests that hearing loss can remain hidden well beyond previously thought levels.

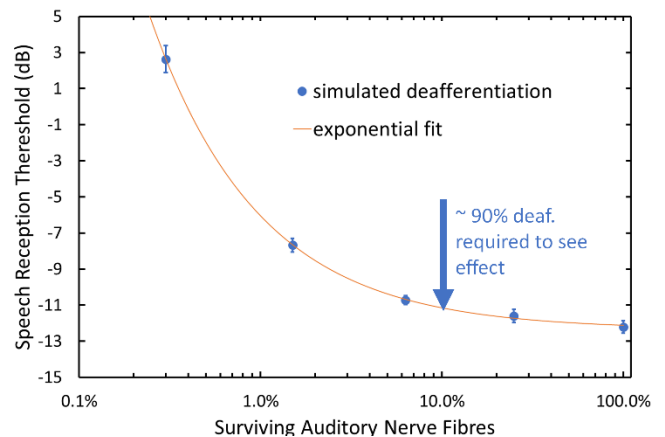


Figure 4. Digit-triplet SRTs obtained with simulated NH hearing (at 100%) and gradually diminishing percentage of surviving AN fibers. An exponential data fit shows that for P% remaining fibers, SRT (dB) can be predicted as $SRT = 0.226 * e^{-1.663 * \text{Log}(P)} - 12.34$.

3.3 Deafferentation and ITD discrimination

Our last study measured the effect of deafferentation on interaural-time-delay (ITD) discrimination. Participants judged whether a 1 KHz low-pass-filtered noise moved from left to right or vice versa over the two intervals of an adaptive, 2-AFC task. 12 young NH adults participated.

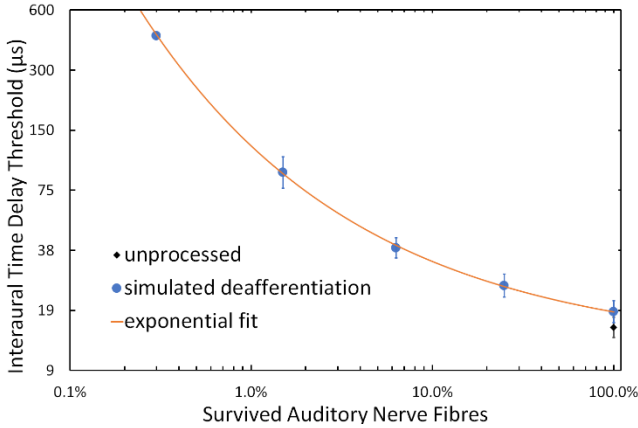


Figure 5. ITD thresholds for the unprocessed condition (black diamond) and as a function of deafferentation (blue circles). An exponential data fit shows that for P% remaining fibers, the ITD threshold can be predicted as $ITD_{Th} (\mu s) = 600 * 2^{\{0.6497 * e^{-0.83 * \text{Log}(P)} - 5.676\}}$.

ITD thresholds significantly and exponentially increased as a function of deafferentation. The very small growth of ITD threshold from 15.4 to 18.4 μs from the unprocessed to the NH-processed condition further validates the NH simulator.

4. DISCUSSION

How does the auditor periphery optimally transmit information at typical speech level? The saturation of neurons at the auditory nerve and IC levels, namely the *dynamic range problem* observed in traditional measures of rate-level functions (RLFs) in small mammals suggested that the human brain may make use of other mechanisms, such as spread of excitation or phase-locking, to code for level [5]. An engineer designing an artificial cochlea and limited by a 30 dB AN dynamic range could be tempted to employ an array of ANs with gradually increasing thresholds, such that the entire 120 dB level range that loudness perception covers could be adequately coded. However, such a system would be suboptimal in that for a specific sound level, many fibers may not fire or may be saturated. A wiser engineer would consider the stochasticity of AN fibers and keep many fibers of a restricted threshold range tuned to the same frequency to remove any risk of stochastic under-sampling [2]. Furthermore, they would address the dynamic range issue by adding feedback loops designed to prevent fiber saturation. Such feedback loops would slow down the growth of transduced level as acoustic level increases, thereby extending the dynamic range of the system to enable it to precisely code for the entire range of naturally occurring sound levels. Furthermore, such feedback loops would ensure the highest coding precision for levels most

relevant to the system's most important task. Human hearing naturally developed two feedback loops: the AR aims at protecting the inner ear from excessive vibrations that could damage the organ or Corti; the MOCR works hand in hand with the AR to ensure dynamic range matching between the range of naturally-occurring sound levels and the limited dynamic range of AN fibers. Furthermore, the MOCR fine-tunes dynamic range adaptation to ensure optimal encoding at the AN level.

This article first elucidates, via simulations based on the physiologically-inspired Model of the Auditory Periphery [1], how dynamic range adaptation may occur in Man, such that the system can remain saturation-free and no information carried by the signal's temporal modulations is lost. This is of particular importance to speech intelligibility, since temporal modulations carry most of the speech information [6]. Simulated RLFs move to higher levels as context level increases, as measured in animal studies at the AN level [4] and at the inferior colliculus level [7]. Note that MAP-predicted adaptation levels are larger than those measured in animals because efferent reflexes in these studies will have been partially suppressed by anesthetics.

With the simplest possible decoding of the auditory nerve, i.e. transforming AN spikes into gammatone wavelets or kernels [8], the response of the simulator remains compressive. Utilizing inverted AR and MOCR signals to expand the response and linearize the system proved successful in that SRTs at the simulator validation stage differed from those obtained with unprocessed stimuli by just over 1 dB. This first psychophysical study illustrates the importance of efferent reflexes to speech intelligibility. Indeed, absence of efferent reflexes would cause medium and high spontaneous rate fibers to saturate and lead to unrecoverable information loss.

Of all of three simulated pathologies in our second study, only knocking out all OHCs had a significant impact on SRTs, again due to the removal of the MOCR-based adaptation. It is worth noting that going as far as halving endocochlear potential or knocking out 70% of AN fibers (akin to significant synaptopathy) did not significantly elevate SRTs. This outcome strongly suggests that such pathologies, often associated with an ageing ear, squarely sit in the realm of hidden hearing loss.

Focusing on deafferentation in our third study suggested that as much as 90% deafferentation might be required in order to measure appreciable SRT elevation. This shows how redundancy in stochastic sampling makes the afferent system extremely robust against the effects of ageing or noise exposure. Conversely, it suggests that human noise-induced, measurable hearing loss may be caused by much higher levels of deafferentation than previously thought or measured in animal studies, if not primarily by deafferentation due to OHC loss. Lopez-Poveda and Barrios [2] simulated the effect of deafferentation predicted by a simple model of stochastic under-sampling. There, the authors showed SRTs were elevated by 1.7 and 10.15 dB, respectively, by simulating the signal and going from unprocessed to 300 and from 300 to 10 stochastic samplers per BF over 10 BFs. Here, going from 1000 to 100 and from 100 to 3.33 AN fibers across 30 BFs, our simulated SRTs are predicted (via exponential-fit

interpolation) to rise by 2.1 dB and 12.7 dB, respectively. This demonstrates a very close match to Lopez-Poveda and Barrios' results.

The simulated effect of deafferentation on ITD discrimination was investigated in our last psychophysical study so we could compare sensitivity to deafferentation between monaural speech intelligibility and a binaural measure. This measure was found to be much more sensitive to deafferentation than SRTs, since 90% deafferentation nearly doubled the simulated ITD threshold (going from 18 to 33 μ s, compared to 15 μ s in the unprocessed condition). Note that contralateral MOCR was not modelled: the ears were modelled as two independent monaural systems, because simulating the contralateral MOCR would have doubled up on the contralateral MOCR of our participants. In any case, the signal level was the same in both ears for all our tasks, and hence, any contralateral efferent reflex would have been symmetrical in our studies. This, however, highlights a limitation in the application of our current pathological simulations to binaural hearing, especially when stimulus levels differ between the ears.

In conclusion, the model predictions and studies presented herein (1) form a framework for advanced simulations of sensorineural pathologies, (2) open the door to establishing the psychophysical signatures of specific pathologies, such that they may be differentially diagnosed, (3) confirm the importance of efferent reflexes in the efficient coding of sound at the AN level, (4) show how simulated OHC knockout leads to a degradation in speech intelligibility comparable to that seen in mild-to-moderately hearing impaired individuals and (5) reveal that very high levels of deafferentation (or synaptopathy) would be required in order for a so-called hidden hearing loss to become measurable.

5. REFERENCES

- [1] R. Meddis *et al.*, "A computer model of the auditory periphery and its application to the study of hearing," in *Advances in Experimental Medicine and Biology*, vol. 787, pp. 11–20, 2013.
- [2] E. A. Lopez-Poveda and P. Barrios, "Perception of stochastically undersampled sound waveforms: A model of auditory deafferentation," *Front. Neurosci.*, vol. 7, no. 7 JUL, pp. 1–13, 2013.
- [3] E. A. Lopez-Poveda and R. Meddis, "A human nonlinear cochlear filterbank.," *J. Acoust. Soc. Am.*, vol. 110, no. 6, pp. 3107–3118, 2001.
- [4] B. Wen, G. I. Wang, I. Dean, and B. Delgutte, "Dynamic range adaptation to sound level statistics in the auditory nerve," *J. Neurosci.*, vol. 29, no. 44, pp. 13797–13808, 2009.
- [5] C. J. Plack, Chap.6, "Loudness and Intensity Coding" in *The Sense of Hearing*, 2018.
- [6] R. Drullman, J. Festen, and R. Plomp, "Effect of temporal envelope smearing on speech reception," *J. Acoust. Soc. Am.*, vol. 95, no. 2, pp. 1053–1064, 1994.
- [7] I. Dean, N. S. Harper, and D. McAlpine, "Neural population coding of sound level adapts to stimulus statistics," *Nat. Neurosci.*, vol. 8, no. 12, pp. 1684–1689, 2005.
- [8] M. S. Lewicki, "Efficient coding of natural sounds," *Nat. Neurosci.*, vol. 5, no. 4, pp. 356–363, 2002.