Neural Network Models of Hearing Clarify Factors Limiting Cochlear Implant Outcomes

Annesya Banerjee (annesya_banerjee@g.harvard.edu)
Program in Speech and Hearing Bioscience and Technology, Harvard University
Cambridge, MA, 02138, USA

Department of Brain and Cognitive Sciences, MIT
McGovern Institute for Brain Research, MIT
Cambridge, MA, 02139, USA

Mark R. Saddler (msaddler@mit.edu)
Department of Brain and Cognitive Sciences, MIT
McGovern Institute for Brain Research, MIT
Center for Brains, Minds, and Machines, MIT
Cambridge, MA, 02139, USA

Josh H. McDermott (jhm@mit.edu)
Department of Brain and Cognitive Sciences, MIT
McGovern Institute for Brain Research, MIT
Center for Brains, Minds, and Machines, MIT
Cambridge, MA, 02139, USA

Program in Speech and Hearing Bioscience and Technology, Harvard University
Cambridge, MA, 02138, USA

Abstract

Cochlear implants (CIs) allow deaf individuals to hear by electrically stimulating the auditory nerve, bypassing the ear. CIs are one of the great successes of biomedical engineering, but nonetheless fail to restore normal auditory perception. Models that can predict behavioral outcomes given CI input could help diagnose the factors limiting perception and thus guide device improvements. We first built a model of normal hearing by optimizing a deep neural network to perform real-world auditory tasks using simulated auditory nerve input from an intact cochlea. We then modeled CI hearing by testing this model on simulated auditory nerve responses to CI stimulation. To simulate possible consequences of learning to hear through a CI, we re-optimized the network on CI input. When the entire network was reoptimized, the model exhibited speech intelligibility scores significantly better than typical CI users. Speech recognition on par with typical CI users was achieved only when just the late stages of the model were reoptimized. However, sound localization performance remained abnormal relative to normal hearing even when the entire network was reoptimized for CI input. The results suggest that some limitations of CIs reflect impoverished peripheral information from potentially suboptimal stimulation strategies, but that other limitations may reflect incomplete central plasticity.

Keywords: Cochlear implants; brain plasticity; neural networks; speech recognition in noise; sound localization

Introduction

Cochlear Implants (CIs) convert sound into electrical stimulation that evokes responses in auditory nerve fibers, enabling sound perception in people with sensorineural hearing loss. Despite their success in improving speech understanding, current CIs fail to restore fully normal hearing perception (Zeng, 2022). These shortcomings could arise from the devices (e.g., suboptimal strategies for processing sound and delivering electrical stimulation), from degeneration within the auditory system, or from limits on the auditory system's ability to adapt to abnormal patterns of input. Computational models that can simulate hearing behavior given CI input could potentially be used to analyze the role of different factors in shaping behavioral outcomes. Here, we trained artificial neural networks to recognize and localize sounds from simulated auditory representations (Kell et al., 2018; Saddler, Gonzalez, & McDermott, 2021; Francl & McDermott, 2022) from either a normal cochlea or a CI-stimulated auditory nerve model. We compared model speech recognition and sound localization in noise to that of CI users and analyzed the effects of different peripheral parameters and brain plasticity on model performance.

Methods

Models were built by combining a realistic simulation of the spiking auditory nerve with a neural network intended to simulate the central auditory pathway.

Normal cochlea auditory nerve model

Input sounds were passed through a filter bank modeled on the human cochlea. Filter responses were converted to firing rates using rate-level functions modeled on those of the auditory nerve. Normal hearing auditory "nervegrams" were generated by sampling spikes from these firing rates (**Figure 1A**).

CI-stimulated auditory nerve model

We used a standard CI processor (**Figure 1B**) consisting of subband envelope extraction, compression, and amplitude modulation of pulse-trains. We simulated the electrode-nerve interface by imposing spatial spread of excitation and reduced dynamic range (<15 dB) of electrically stimulated auditory nerve fibers.

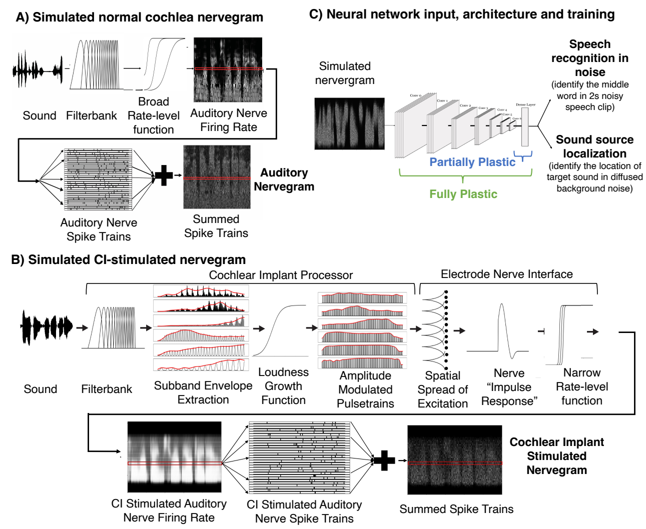

Figure 1: A) Normal hearing and B) CI-stimulated auditory nerve model stage. C) Neural network was trained to recognize words and localize sounds from auditory "nervegrams".

Model task and optimization

A feedforward convolutional neural network (Saddler, Francl, et al., 2021) was trained on either normal-hearing or CI-stimulated nervegrams. We trained the model separately on two tasks (**Figure 1C**): speech recognition in noise (reporting the word spoken at the middle of 2s audio clips) and sound localization (reporting azimuth and elevation of a target sound in diffuse background noise; (Francl & McDermott, 2022)).

Simulating the effects of neural plasticity

Speech recognition abilities improve over time following implantation and are inversely correlated with age of implantation, suggesting that neural plasticity is important for achieving the best outcomes. However, plasticity might be limited, either by sensitive periods during development and/or by some parts of the brain being less plastic than others. To understand the importance of plasticity, we explored three different training conditions. **Static CI model:** The network was optimized for normal hearing input and tested on CI input without additional optimization (potentially analogous to testing a CI user immediately after implantation). **Fully plastic CI model:** The entire neural network were reoptimized for CI input (analogous to an infinitely plastic auditory system, with task performance approaching the limit of what is possible with CI input). **Partially plastic CI model:** Only the late layers of the network were reoptimized for CI input, motivated by the hypothesis that plasticity might be more pronounced in later stages of processing, and by evidence that cortical neural responses are best predicted by relatively late stages of neural network models (Kell et al., 2018; Tuckute et al., 2023).

Results

Partially plastic model best captures CI-user speech recognition in noise

Figure 2A shows word recognition accuracy of the different models as a function of signal-to-noise ratio (SNR). The neural network trained on normal-hearing input exhibited human-level performance (black line). The static CI model (magenta line) was near chance. This result is consistent with many individuals exhibiting poor speech recognition immediately after implant activation. Performance of the fully plastic CI model (green line) was significantly better than typical CI users (dashed red line), when tested on the same task. Speech recognition performance on par with human CI users was achieved only when only the late network stages were reoptimized for CI input (blue line).

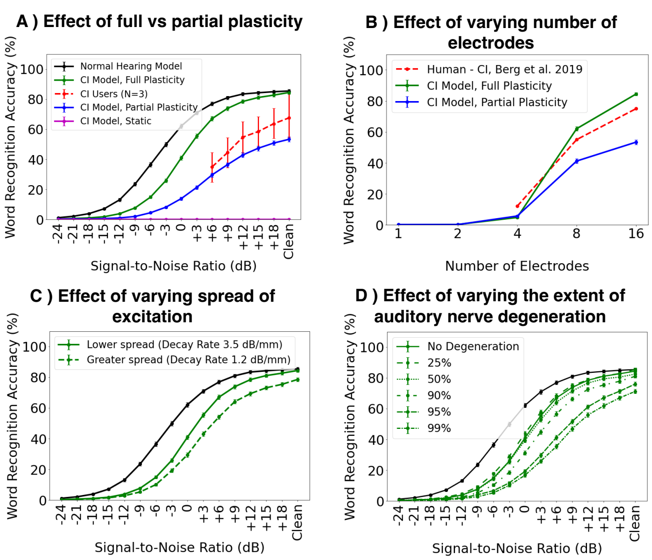

Figure 2: Speech recognition in noise results.

Human-like dependence on the number of active electrodes. **Figure 2B** depicts the word recognition accuracy as a function of the number of active CI electrodes. The full and partially plastic model both performed worse when the number of electrodes was reduced from 16 to 4, as found in human CI users ((Berg et al., 2019); note that the human data was obtained with a different speech task, such that absolute performance is not comparable).

Performance degrades with increasing spread of excitation. To investigate whether more spatially focused CI stimulation might improve speech recognition, we manipulated the extent of spatial spread of excitation in our CI model (determined in part by the nature of the electrical stimulation). We simulated broad (green dashed line) and narrow (green solid line) spatial spread, mimicking monopolar and bipolar CI stimulation, respectively (**Figure 2C**). Model speech performance became closer to that of normal hearing as the spread of excitation was reduced.

Performance degrades with auditory nerve degeneration. We evaluated the effect of auditory nerve degeneration on word recognition performance by reducing the number of simulated nerve fibers (by 25%, 50%, 90%, 95%, and 99%). Our results (**Figure 2D**) showed modest drop as the extent of nerve degeneration was increased, with the effect being prominent only for >95% degeneration (Cheng & Svirsky, 2021), suggesting that nerve degeneration on its own is unlikely to count for suboptimal performance of human CI users.

CI model fails to achieve near-normal sound localization performance

Our CI hearing model exhibited similar deficits in localization behavior as CI users. We analyzed localization error as a function of the stimulus SNR (**Figure 3**). The normal hearing model (black line) showed improved performance as the SNR increased, and the localization error was very low for clean stimuli. The fully plastic CI model (green line) exhibited a similar overall trend, but the localization error was at least 4 to 5 times higher than the normal hearing condition. Performance was worse for the partially plastic CI model (blue line) and close to chance for the static condition (magenta line). The localization deficits with the fully plastic CI model indicate that limitations in CI users' localization abilities is at least partly because of limits on the information conveyed by the CI processing strategy.

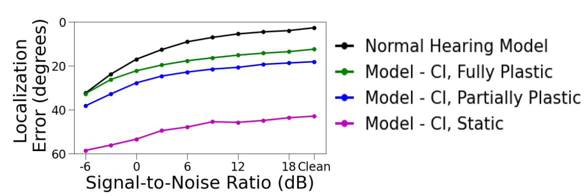

Figure 3: Sound localization results for normal hearing model and CI models with different extent of plasticity.

Conclusion

This work provides initial validation of machine learning-based models of CI-mediated perception. Our results clarify the roles of impoverished peripheral information and incomplete central plasticity in limiting CI users' performance of realistic auditory tasks.

References

Berg, K. A., Noble, J. H., Dawant, B. M., Dwyer, R. T., Labadie, R. F., & Gifford, R. H. (2019). Speech recognition as a function of the number of channels in perimodiolar electrode recipients. *The Journal of the Acoustical Society of America*, *145*(3), 1556–1564.

Cheng, Y.-S., & Svirsky, M. A. (2021). Meta-analysis—correlation between spiral ganglion cell counts and speech perception with a cochlear implant. *Audiology Research*, *11*(2), 220–226.

Francl, A., & McDermott, J. H. (2022). Deep neural network models of sound localization reveal how perception is adapted to real-world environments. *Nature human behaviour*, *6*(1), 111–133.

Kell, A. J., Yamins, D. L., Shook, E. N., Norman-Haignere, S. V., & McDermott, J. H. (2018). A task-optimized neural network replicates human auditory behavior, predicts brain responses, and reveals a cortical processing hierarchy. *Neuron*, *98*(3), 630–644.

Saddler, M. R., Francl, A., Feather, J., Qian, K., Zhang, Y., & McDermott, J. H. (2021). Speech Denoising with Auditory Models. In *Proc. interspeech 2021* (pp. 2681–2685). doi: 10.21437/Interspeech.2021-1973

Saddler, M. R., Gonzalez, R., & McDermott, J. H. (2021). Deep neural network models reveal interplay of peripheral coding and stimulus statistics in pitch perception. *Nature communications*, *12*(1), 7278.

Tuckute, G., Feather, J., Boebinger, D., & McDermott, J. H. (2023). Many but not all deep neural network audio models capture brain responses and exhibit correspondence between model stages and brain regions. *Plos Biology*, *21*(12), e3002366.

Zeng, F.-G. (2022). Celebrating the one millionth cochlear implant. *JASA express letters*, *2 7*, 077201.