Detecting Human Frequency-Following Responses Using an Artificial Neural Network

Fuh-Cherng Jeng[1], Amanda E. Carriero[1], Sydney W. Bauer[1], and Hou-Kuang Chen[2]

[1]Communication Sciences and Disorders, Ohio University, Athens, OH 45701, USA; [2]Department of Otolaryngology-HNS, China Medical University Hospital, Taichung, Taiwan

INTRODUCTION

- The frequency-following response (FFR) is a neurophysiological measure that reflects the brain's ability to track the temporal and spectral characteristics of acoustic stimuli (Krizman & Kraus, 2019).
- Machine learning, a subset of artificial intelligence, has shown considerable promise in uncovering hidden response patterns within datasets. Various machine learning models have been explored to improve the efficiency of FFR detection (Jeng et al., 2024; Llanos et al., 2022).
- While traditional machine learning approaches have shown promising results, deep learning methods, such as neural networks, have not yet been widely explored for FFR detection.
- An artificial neural network (ANN) consists of multiple layers of interconnected neurons, mimicking the various levels of signal processing and capturing response patterns at different levels of abstraction (LeCun et al., 2015).
- The purpose of this study was threefold: first, to evaluate the feasibility of using a three-layer ANN to detect the presence or absence of an FFR elicited by the intonation of a speech stimulus; second, to determine the optimal number of inputs and hidden neurons for FFR detection; and third, to assess model performance with an increasing number of recording sweeps.

METHODS

Participants

- 60 American adults (43 females, 17 males, 24.9 ± 4.2 years old)

Stimulus

- An English vowel /i/ with a rising frequency contour (F0 ranging from 102 to 140 Hz) was utilized to elicit FFRs.
- 70 dB SPL to the right ear through an ER-3A insert earphone
- 150-ms stimulus duration + 150-ms silent interval

Recording

- 3 gold-plated surface recording electrodes
  - High forehead (+), right mastoid (-), and low forehead (gnd)
- Participants resting or fast-asleep prior to recording
- 8000 accepted sweeps from each participant

Preprocessing

- 1, 10, 100, 200, 500, 1000, 2000, 5000, and 7000 sweeps were randomly selected from a pool of the 8000 accepted sweeps.
- This resulted in a total of 9 nSweep conditions to be analyzed.
- The averaged waveform corresponding to the 150-ms stimulus presentation was treated as the experimental condition (response present), while the averaged waveform corresponding to the 150-ms silent interval was treated as the control condition (no response).
- Each averaged waveform was transformed into a narrow-band amplitude spectrogram (window size: 50 ms, step size: 0.5 ms, frequency resolution: 1 Hz), from which a vector of F0 estimates was derived.
- For each nSweep condition, a dataset containing 120 vectors of F0 estimates (60 participants x 2 conditions [experiment vs. control]) was used to train and test a three-layer ANN.

Architecture

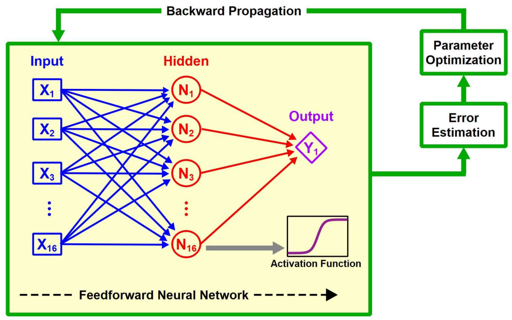

Figure 1. *Architecture of a fully connected, feedforward neural network.*

Parameter Selection and Optimization

- The number of input and hidden neurons was systematically varied between 1 and 16. The connections between the input neurons and hidden neurons were weighted; these weights were adjusted during training to optimize model performance.
- The output layer contained a single neuron that provided the network's final prediction: response present or absent.
- The output of a single neuron could be expressed by using the formula:

$$output = g(\sum_{i=1}^{n} \omega_i x_i)$$

- The learning process in this neural network was driven by backpropagation, a technique that disseminated the prediction error (i.e., loss) backward through the network to update the weights of the connections.
- During training, the BCEWithLogitsLoss function from PyTorch was used, and the AdamW optimizer was utilized with a learning rate of 0.001 (Jeng & Jeng, 2022).
- Parameter optimization was stopped when the test loss decreased to 0.01 or after 200 training iterations (Xu et al., 2007; Zhou et al., 2008).
- All vectors of F0 estimates were randomly partitioned into training (75%) and test (25%) datasets.

RESULTS

Model Performance

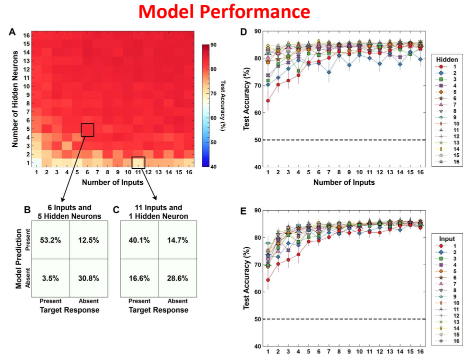

Figure 2. *Test accuracies displayed as a heat map (A), two representative examples (B & C), with increasing inputs (D), and with an increasing number of hidden neurons (E) at 7,000 sweeps.*

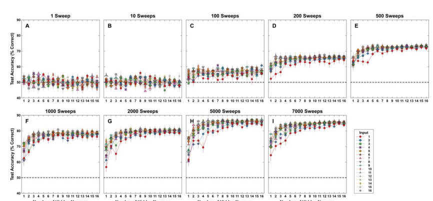

Figure 3. *Test accuracies plotted as a function of the number of hidden neurons, across different numbers of inputs and sweeps.*

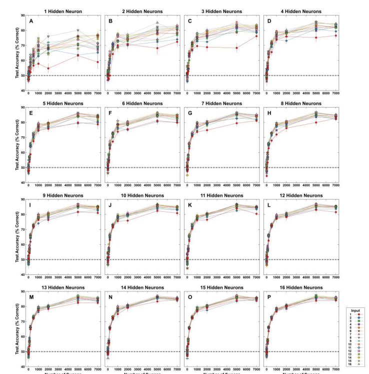

Figure 4. *Test accuracies illustrated with increasing numbers of sweeps, across different numbers of inputs and hidden neurons.*

Statistical Results

- **A three-way ANOVA** revealed a significant effect for each of the three main factors: the number of inputs (F(15, 20736) = 23.424, *p* <0.001, η_p^2 = 0.017), the number hidden neurons (F(15, 20736) = 123.432, *p* <0.001, η_p^2 = 0.082), and the number of sweeps (F(8, 20736) = 9892.434, *p* <0.001, η_p^2 = 0.792).
- **For the 7,000-sweep condition, a two-way ANOVA** demonstrated a significant effect of the number of inputs (F(15, 2304) = 9.349, *p* <0.001, η_p^2 = 0.057) and the number of hidden neurons (F(15, 2304) = 55.815, *p* < 0.001, η_p^2 = 0.267), but not for the interaction between these two factors (*p* = 1.000).

DISCUSSION

- These results provide a guideline for FFR detection in electroencephalography signals and can serve as a baseline for future studies involving similar neural network applications.
- The prediction accuracy of the ANN is significantly influenced by both the number of inputs and hidden neurons, particularly when the number of sweeps reaches 100 or more.
- For FFR detection, an optimal range of approximately 6-8 inputs and 4-6 hidden neurons is needed to maximize prediction accuracy. Beyond these ranges, adding more inputs or hidden neurons contributes minimally to improving accuracy, causing the model performance to plateau.
- The ANN achieves prediction accuracies of approximately 84% with a balanced number of inputs and hidden neurons, particularly when the signal-to-noise ratio is enhanced through a sufficient number of sweeps.

Learning Objective

- Upon completion, participants will be able to describe how deep learning models, specifically a three-layer ANN, can be used to detect the presence of FFRs, expanding their knowledge of how machine learning tools can enhance auditory signal analysis in clinical and research settings.

Clinical Takeaways

- Implementing ANN for detecting FFRs can significantly enhance the precision of auditory diagnostics, particularly in evaluating the neural encoding of speech intonation.
- This highlights the potential for integrating machine learning tools into routine audiological assessments, offering a more efficient and innovative approach for analyzing complex auditory stimuli and improving patient outcomes.

ACKNOWLEDGMENTS

This study was supported by an internal research fund (RI-4160001) and a Student Research Grant (23-X-59) at Ohio University.

REFERENCES

Jeng, F.-C., & Jeng, Y.-S. (2022). Implementation of Machine Learning on Human Frequency-Following Responses: A Tutorial. *Seminars in Hearing, 43*(3), 251–274. https://doi.org/10.1055/s-0042-1756219

Jeng, F.-C., Matzdorf, K., Hickman, K. L., Bauer, S. W., Carriero, A. E., McDonald, K., Lin, T.-H., & Wang, C.-Y. (2024). Advancing Auditory Processing by Detecting Frequency-Following Responses Through a Specialized Machine Learning Model. *Perceptual and Motor Skills, 131*(2), 417–431. https://doi.org/10.1177/00315125231225767

Krizman, J., & Kraus, N. (2019). Analyzing the FFR: A tutorial for decoding the richness of auditory function. *Hearing Research, 382*, 107779. https://doi.org/10.1016/j.heares.2019.107779

Llanos, F., Zhao, T. C., Kuhl, P. K., & Chandrasekaran, B. (2022). The emergence of idiosyncratic patterns in the frequency-following response during the first year of life. *JASA Express Letters, 2*(5), 054401. https://doi.org/10.1121/10.0010493

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature, 521*(7553), 436–444. https://doi.org/10.1038/nature14539

Xu, L., Chen, X., Zhou, N., Li, Y., Zhao, X., & Han, D. (2007). Recognition of lexical tone production of children with an artificial neural network. *Acta Oto-Laryngologica, 127*(4), 365–369. https://doi.org/10.1080/00016480601011477

Zhou, N., Zhang, W., Lee, C.-Y., & Xu, L. (2008). Lexical tone recognition with an artificial neural network. *Ear and Hearing, 29*(3), 326–335. https://doi.org/10.1097/AUD.0b013e3181662c42