2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

A Machine Learning-based Neural Implant Front End for Inducing Naturalistic Firing

Cynthia R. Steinhardt* and Gene Y. Fridman, *Member, IEEE*

***Abstract*—Despite being able to restore speech perception with 99% success rate, cochlear implants cannot successfully restore pitch perception or music appreciation. Studies suggest that if auditory neurons were activated with fine timing closer to that of natural responses pitch would be restored. Predicting the timing of cochlear responses requires detailed biophysical models of sound transmission, inner hair cell responses, and outer hair cell responses. Performing these calculations is computationally costly for real time cochlear implant stimulation. Instead, implants typically modulate pulse amplitude of fixed pulse rate stimulation with the band-limited envelopes of incoming sound. This method is known to produce unrealistic responses, even to simple step inputs. Here we investigate using a machine learning algorithm to optimize the prediction of the desired firing patterns of the auditory afferents in response to sinusoidal and step modulation of pure tones. We conclude that a trained network that consists of 25 GRU nodes can reproduce fine timing with 4.4 percent error on a test set of sines and steps. This trained network can also transfer learn and capture features of natural sounds that are not captured by standard CI algorithms. Additionally, for 0.5 second test inputs, the ML algorithm completed the sound to spike rate conversion in 300x less time than the phenomenological model. This calculation occurs at a real-time compatible rate of 1 ms for 1 second of spike timing prediction on an i9 microprocessor. This suggests that this is a feasible approach to pursue for real-time CI implementation.**

***Index Terms* – Cochlear implant, fine structure, pitch perception, machine learning, recurrent neural network**

I. Introduction

Cochlear implants (CIs) are arguably the most successful neural implant with nearly 40 years of innovation and over 736,900 devices implanted as of December 2019 [1], [2]. CIs significantly improve speech recognition and comprehension in children and adult users [1], [3]. However, they are considerably less successful at restoring pitch of sound. This poses major issue for CI users who speak tonal languages, such as Mandarin, as it result in difficulties with speech comprehension [4]. It also creates a lesser but significant quality of life deficiency by limiting music appreciation [5].

Until recently, the focus of improving cochlear implants has been on preventing current spread from distorting perceived sound. Hardware innovations were implemented to minimize electrode distance from the modiolar wall to more directly target spiral ganglion neurons, and algorithms were modified to avoid electrical interference by ensuring no electrodes delivered current simultaneously [6], [7]; these improvements led to significant gains in fidelity of targeting neurons for spatial encoding of sound to give the percept in sound of a certain frequency. These improvements paired with the continuous interleaved sampling (CIS) strategy, modulation of amplitude of fixed-rate pulsatile stimulation to the envelope of sound, have led to highly accurate English speech comprehension in CI users.

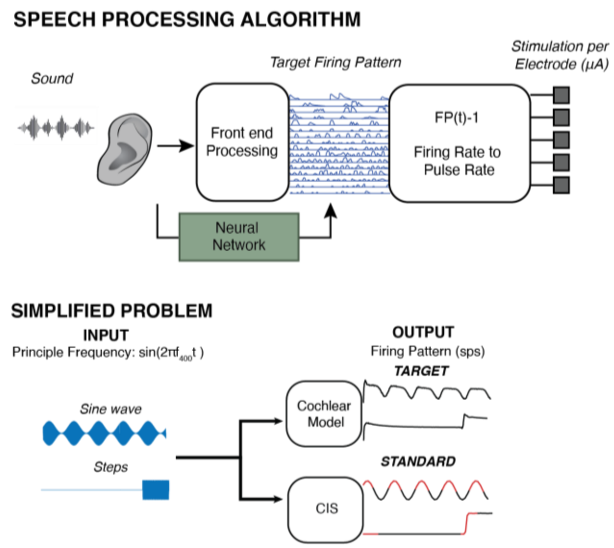

Figure 1. Study Design. We aim to create a front end sound processing algorithm for a cochlear implant to transforms natural sound into a target population firing pattern for the cochlea. This pattern could then be transformed into a stimulation pattern that induces a response with naturalistic fine timing. A neural network will be used to learn the relationship between sound and firing rate from a realistic phenomenological model of the cochlea (top). In this paper, we test algorithm performance on a simplified problem, producing single auditory fiber responses to sine wave and step stimuli, because the CIS algorithm does not replicate firing for these inputs (red) but the cochlear model does.

Studies indicate that the inability to correctly convey pitch is the result of unrealistic CI-evoked timing of neural responses [8]. For example, when normal hearing subjects listened to computer generated tones that deliver pulses with timing reflecting fine timing information of sound, they show improved perception of tonal language (Mandarin) [10], [11]. Thirty-years of detailed studies produced a phenomenological

*Research supported by NSF Foundation GRFP DG31746891 and NIH RN01NS092726.

C.S is with Johns Hopkins Biomedical Engineering Department, MD 21287 USA (corresponding author: 818-839-0699; e-mail: csteinh2@ jhmi.edu).

G.Y is with the Johns Hopkins Department of Otolaryngology, Biomedical Engineering, and Electrical and Computer Engineering, MD 21205 USA

model that reflects the full process of sound processing, including mechanical transduction and outer hair cell connectivity; this model produces accurate fine-timing for a single frequency of cochlear neuron[11], [12]. However, even on a powerful desktop processor, this algorithm requires considerably more time to process the sound that the duration of the sound itself. This problem worsens when considering the reduced processing power on an implanted device.

While replicating exact timing of natural spiking has not been attempted per se, high-rate pulsatile stimulation was popularized as an improvement to CIS because it leads to more desynchronized, naturalistic neural responses and in turn improved speech perception in noisy environments[13]. However, a recent study showed that reducing the number of high rate pulses by half while considering natural firing principles improves speech perception [14]; this further supports the idea that reducing the number of pulses but more accurately replicating fine timing would improve perception.

In this paper, we create a prototype of a front-end algorithm for cochlear implants that can transform any sound into the naturalistic fine timing of spikes for a fiber at real-time processing speeds using a machine learning (ML) approach (Fig. 1). Our approach is to train a recurrent neural network (RNN) to learn the sound-wave-to-spiking relationship captured in the validated Zilany 2014 version of the phenomenological model of the cochlea [12]. We will first evaluate the performance of the RNN in producing natural responses to sine waves and steps. We choose this simplified problem, because the CIS algorithm fails to capture the complexity of the natural encoding of these stimuli (Fig. 1 red). This front-end could then be included in a CI processing algorithm that (1) breaks sound into the power in spectral bands, (2) converts power to firing rate over time with a machine learning algorithm, and (3) converts induced firing rate into a pulse rate stimulation pattern, using equations relating pulse rates to induced firing rates[15] (Fig. 1). Here we focus on optimizing part (2) of this signal processing chain using ML technique.

II. Methods

A. Generation of Training and Test Waveforms

The data used to train and test the model were synthetically generated sine waves and steps in power of a 400 Hz sinusoid (Fig. 1). All inputs were generated with signal of volume V_{stim} in dB, which was converted to A_{mod} sound pressure level (SPL) with Equation (1).

$$A_{mod}\left(V_{stim}\right) = \sqrt{2}\left(20^{-6}\left(10^{\frac{V_{stim}}{20}}\right)\right) \quad (1)$$

All inputs modulated a 400 Hz wave of the form:

$$s_{base} = \sin(2\pi f_{princ} t),\ f_{princ} = 400 \text{ Hz} \quad (2)$$

Sinusoidal modulation was performed with Equation 3:

$$s_{sine} = A_{mod}(0.95(1 - d_{mod}) \sin(2\pi f_{mod} t + \phi_{mod}) + d_{mod}) s_{base} \quad (3)$$

, where depth of modulation, d_{mod}, determined the portion of modulation compared to A_{mod}, reaching up to 0.95. d_{mod}, frequency of modulation f_{mod}, phase of modulation ϕ_{mod} and A_{stim} were varied as shown in Table 1:

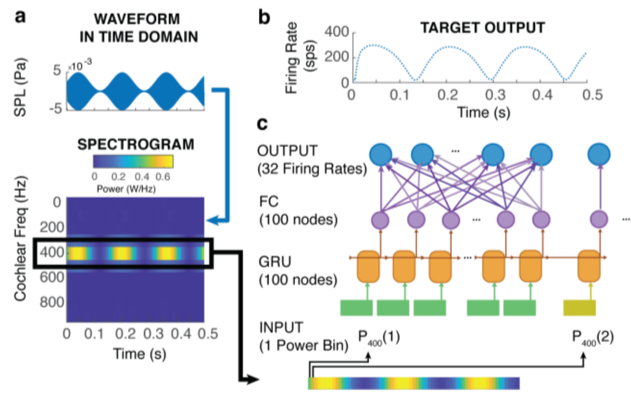

Figure 2. Network architecture for this study. (a) Waveforms are transformed in spectrograms in MATLAB and each bin of the 400 Hz power band is run through the neural network to generate induced firing rate over time. (b) Target outputs are generated by the Zilany model with smaller time bins. (c) A recurrent neural network is used to turn each time bin into thirty-two firing rate predictions over time.

	d_{mod}	$f_{mod}(Hz)$	ϕ_{mod} (radians)	V_{stim} (dB)
Min	0.5	2	0	45
Max	0.9	40	2π	95
Steps	8	10	5	10

Table 1. Parameters for sinusoidal input generation

Step modulation was performed with Equation 4, where V_{stim1} and V_{stim2}, the volume in dB of the first and second half of each step, and the shift, t_{mod}, were varied in the range of Table 2:

	V_{stim1}(dB)	V_{stim2} (dB)	t_{mod}
Min	0	45	0.2
Max	60	95	0.8
Steps	10	20	20

Table 2. Parameters for step input generation

V_{stim1} and V_{stim2} were converted to A_{mod1} and A_{mod2}, using Equation 1, and step inputs were calculated as:

$$s_{step} = (A_{mod1} + (A_{mod2} - A_{mod1})u(t - t_{mod}T))s_{base} \quad (4)$$

, where *u(t)* is the unit step, and t_{mod} is the fraction of the trial length, *T*.

Additionally, performance of the model was tested on individual spoken word recordings from the training set of the Speech Command dataset[16]. Recordings were filtered in the 400 Hz band to be equivalent to synthetic inputs.

The waveforms were converted into power by using the *spectrogram* function in MATLAB with a hamming window of length 512 (Fig. 2a). This produced 6.4 ms bins.

The Zilany 2014 model was used to generate the natural firing rate over time in responses to these stimuli. The power was used as the input for the neural network, and the firing rate over time generated by the model was used as the target for training and evaluation (Fig. 2b & 2c). The Zilany model produced data every 0.2 ms for each 6.4ms sound sample. This different bin size was addressed when designing the neural network.

For training, 100 synthetic waveforms were randomly sample from the data set, half sine waves and half steps. For testing performance on natural stimuli, 100 words from the Speech Command dataset were randomly selected.

B. Modeling Cochlear Neuron Response

A phenomenological model that we refer to as the Zilany model, of the human auditory periphery was developed over the last 30 years to replicate healthy auditory response to perception of any sound [12], [17]. This model accounts for outer hair cell and inner hair cell contributions to firing, filtering effects, and non-linearities related to synaptic and axonal activation. The model transforms sound pressure level (SPL) into spiking and firing rate over time for an auditory nerve fiber with low, medium, or high spontaneous firing (Fig. 2b). Our model consisted of 50 ganglion cells in the physiologically observed ratio of low and high spontaneous activity fibers located at the 400 Hz position along the membrane [18]. The neural responses (spikes/second over the duration of the sound stimulus) were used to create a dataset for testing and training the neural network.

C. Recurrent Neural Network (RNN) Model

Although machine learning has been used for a variety of speech processing problems, we found no evidence of it being used for optimization of calculation or for generation of neural population spiking[19]. Because this is an inherently "forwards-only" problem due to signal processing progressing from the eardrum to the ganglion cell firing pattern, we chose to use a gated recurrent unit (GRU) which incorporates the memory of past network states to generate new inputs as the core of the network design for the task. This should account for effects of history, such as past spikes affecting proceeding spikes due to refractoriness. There were 32 firing rate values for every spectrogram time bin, so a fully connected layer was used to transform the outputs of the GRU layer into 32 outputs. This also allowed additional calculations to be made to adjust firing rate predictions within several milliseconds of one another that occur within one spectral bin. During this study, we assessed model size and used GRUs with 25, 50, 100, 200, 500, and 1000 nodes. We then used a fully connected layer that reduced the GRU nodes to 32 outputs (Fig. 2c).

The model was created using the Python Pytorch package. To train this model, the mean squared error (MSE) was used for backpropagation, using the "MSELoss" criteria.

D. Performance Metrics

To assess performance on the test and training data, the rms between the target firing rate in spikes per second (sps) of the Zilany model and the output of the RNN was used as a measure of error. During testing, the model was assessed on 100 waveforms (49 sinewaves). Transfer learning as also evaluated on 100 speech command recordings in the 400 Hz frequency. The rms between 10 predictions of the same response to sound with the Zilany model was used as a measure of the variance in natural responses to sound. The rms was transformed into percent error by dividing by the rms of the firing rate over time.

Statistical testing between models and performance was computed with a paired t-test for comparing model size performance and a two-sided t-test when comparing performance on sinusoidal versus step modulation.

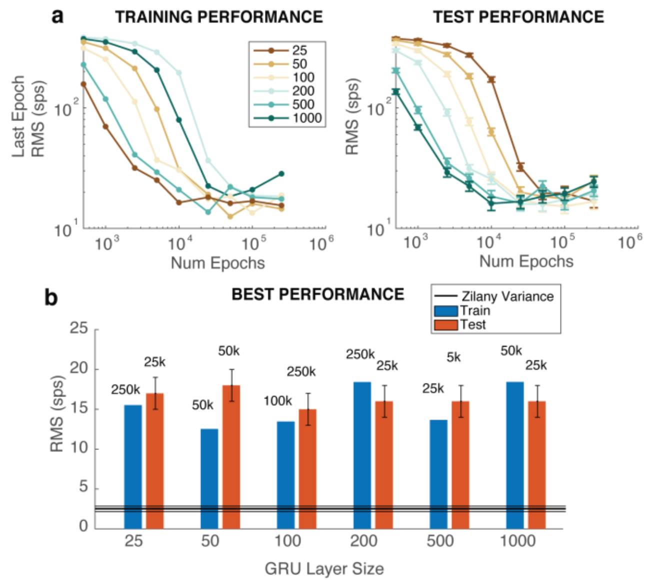

Figure 3. Loss/Performance with Epochs. (a) Left. The rms by the last epoch of training on networks with 25 to 1000 GRU nodes. Right. Test performance of each of the trained models on 100 novel sine and step modulated waveforms. (b) The best performance of each size network on the training (blue) and test (red) data with the number of training epochs at which it best performed written above. Error bars are SEM.

This study is attempting to understand whether a machine learning based front-end could be implemented in real-time in a cochlear implant. Thus, in addition to determining the minimum number of nodes necessary to predict responses to sinusoidal and step modulation, the computation time for the model was also assessed compared to the computation time of running the Zilany 2014. Run-time was evaluated on one CPU from a 2.4 GHz 8-Core Intel Core i9 Processor on a 2019 15-inch MacBook Pro when the trained RNN and Zilany model perform a prediction in response to the same 0.5 second sound 10 times. We used the ratio of speed as a metric in the results.

Additionally, we created RNNs of different sizes to determine the minimum number of nodes necessary to reach an acceptable loss value. We also assessed whether better trained networks involved more calculations and therefore led to significantly slower run times by comparing performance between RNNs trained with 500 and 250,000 epochs.

III. Results

Studies indicate restoration of pitch perception requires fine timing of cochlear neuron firing. Standard cochlear implant algorithms, such as the CIS algorithm do not attempt to replicate this fine timing because it is computationally intensive. In this study, we attempt to perform the same computation as in the phenomenological model of cochlear response from Zilany 2014 in real-time by training a neural network to learn the computations performed in Zilany 2014. We assess performance of our RNN first on predicting responses to sinusoidal and step modulation of a 400 Hz sound, a simplified task with observable transformations compared to natural sound. We then determine whether

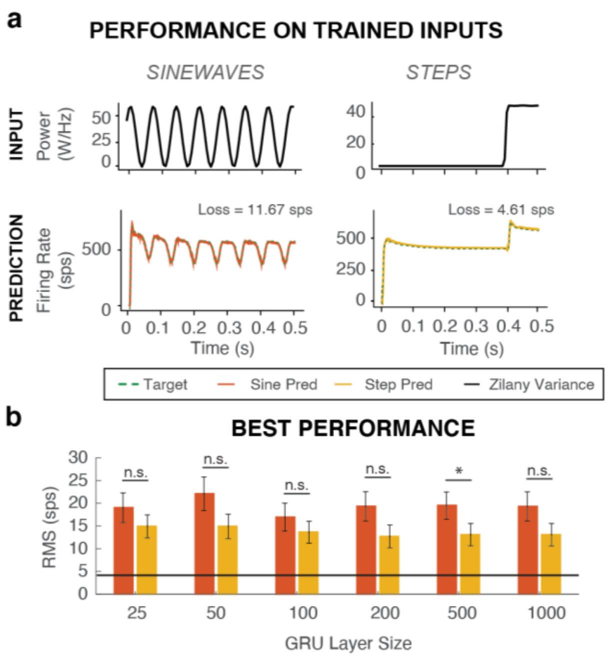

Figure 4. Relative Performance on Sine and Step Waveforms. (a) Input power signal (top) and prediction (red/yellow) and target firing response generate by Zilany model (green dash). (b) Test performance across models on predicting responses to sinusoidal (red) versus step (yellow) inputs with networks trained for number of epochs of best overall performance. Statistics are two-sample t-test. *, $p < 0.1$.

learned transformations apply to natural sounds in the 400 Hz frequency and produce neural responses to natural inputs that are not captured by the existing CIS algorithm.

A. Performance on Synthetic Sounds

We first examined the training time and network size required to create an RNN that can perform this task. The RNN contains a GRU layer and fully connected layer. We attempted to train networks with as few as 25 GRU nodes and as many as 1000 GRU nodes for up to 250,000 epochs. The smaller models trained and reached the lowest training errors after fewer epochs (Fig. 3a left). However, all models converged to approximately the same performance by 250,000 training epochs. Larger models reached lower test error more quickly. However, by 250,000 training epochs, all model size performances were approximately the same (Fig. 3a right). We speculate that there are fewer weights to adjust so these models converge more quickly to an optimization minimum. However, ultimately, even a 25-node GRU layer learned this transformation after a reasonable number of training epochs. Although the best performance occurred after different numbers of training epochs, depending on model size, all models had test performance comparable to training performance of approximately a rms of 10 sps (Fig. 3b).

We compared this to the minimum achievable error, the rms between multiple simulations of the natural response to a sound with the Zilany model, which reaches a minimum of 3.0 ± 0.2 (SEM) sps. Compared to the rms of the signal, the models on average have a test error of 4.20 ± 0.03 %.

The models were trained to infer responses to both sinusoidal (red) and step (yellow) modulation (Fig. 4a). The RNN was able to generate both types of responses with high fidelity to the outputs generated by the Zilany model (green dash) (Fig. 4a). The model appeared to predict step input responses more accurately. However, difference in performance were not statistically significant except for the 500 node GRU model (Fig. 4b). We expect performance to converge with a larger number of training epochs and more training data.

B. Performance on Natural Sounds

We then evaluated the relative difficulty of learning responses to natural sounds and consistency of cochlear neuron transformations by using the same models (Fig. 4) without retraining to predict responses to recorded speech in the same 400 Hz auditory fiber bundle. Audio recordings from the Speech Command dataset of male and female subjects saying individual words were inputted into the RNN. The amount of transfer learning was again measured with the rms between the prediction of the RNN and the output of the Zilany model for these natural inputs (Fig. 5a).

Without retraining, the model is capable of transfer learning and capturing complex structure in the response not captured using the CIS algorithm (Fig. 5a). The error primarily comes from offsets in predicted firing rate not inability to capture complexity. This leads to rms increases of up to 80 sps across models, and the minimum percent error across models averaged 46.1 ± 0.76 % (Fig. 5a-b). Model size shows some significant effect. The 50-node RNN significantly outperforms all models except for the 500-node RNN (Fig. 5b). However, we speculated that the large

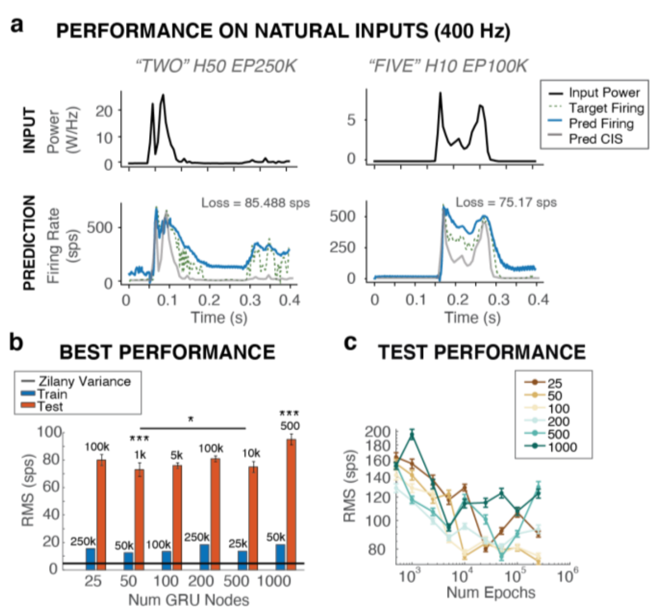

Figure 5. Performance on Natural Sounds. (a) Example natural inputs with word, GRU size and number of training epochs written above. Target response (green dash) and inferred response (blue) of RNN. (b) Best training performance on synthetic sounds (blue) of each model and test performance (red) on natural sounds for each network size. Best model training epochs written above. *, $p < 0.1$; ***, $p < 0.01$ with paired t-test. If there is no bar (as for the 50 node GRU), stars indicate t-test compared to all models. (c) Test performance across models for predicting responses to natural inputs.

models had not converged and learned the rules as accurately as the smaller models did with fewer weights and biases to train. Plotting the minimum loss achieve for each network size when the network was trained for up to 250,000 epochs supported this idea, as the rms still showed higher loss values and high variance than when the model was trained on sine and step inputs (Fig. 3a). Longer training epochs are therefore required to determine the ideal network size, but this implies a network larger than 50-100 nodes is not necessary to learn cochlear responses to natural stimuli.

Observing differences between the inferred response by the RNN (blue) and target response (green dash), we find the model captured non-linear transformations of the sound into firing rate (Fig. 5a left .05-1.5 s & .3-.5 s, right .15-.3 s). The model appeared to accumulate the most error for portions of response that were not scaled accurately. However, it captured complexities in shape that would not be captured with a CIS model (grey), which linearly maps the sound amplitude envelope to pulse amplitude. These results suggest that essential transformations were learned from sine and step inputs alone. Additionally, because the models have not yet converged, with more training epochs, the RNN will likely capture both shape transformation and scaling accurately, as it was able to learn offsets in the step response (Fig. 4a right). How significantly the present differences in scaling influence pitch perception is yet to be determined.

C. *Real-time Applicability*

We evaluated the potential of these RNNs to be used in a real-time implementation on the same 0.5 second sound. The Zilany model required 1.47 ± 0.01 seconds to predict the neural response of a single fiber. The 25-node network required 4.73 ± 0.02 milliseconds. We plotted this improvement as a ratio of time to perform the task with the Zilany model over the time to perform the task with the RNN (Fig. 6). The RNN was 335.4 ± 4.54 times faster with a 25-node network trained with 50,000 epochs. We chose to evaluate performance with a minimum 50,000 epochs, because the performance of the RNNs converged by 50,000 epochs across models on the synthetic data (Fig. 3a). So, models of these size produced reasonable predictions of responses.

The number of training epochs did not significantly influence run-time for most models (Fig. 6 grey v. blue). For models with a GRU layer with less than 200 nodes, run-time was approximately the same. As the model approached 100 nodes, the relative gain in computation speed was significantly reduced (Fig. 6). As performance was consistently low when the RNN has less than 200 nodes (Fig. 5b), we do not anticipate requiring a network that is less than 200 times faster than the Zilany model. At these speeds, the model can perform a computation in approximately 1/100 of the length of the stimulus. If we assume this processing speed scales with sound size, because the GRU steps are an iterative process, we anticipate these computation speeds to be within the range of real-time.

The computation speed was evaluated on a 2019 MacBook Pro with an Intel Core i9 with 2.4GHz Processor (I909980HK). These processors are clocked at 478 GFLOPS.

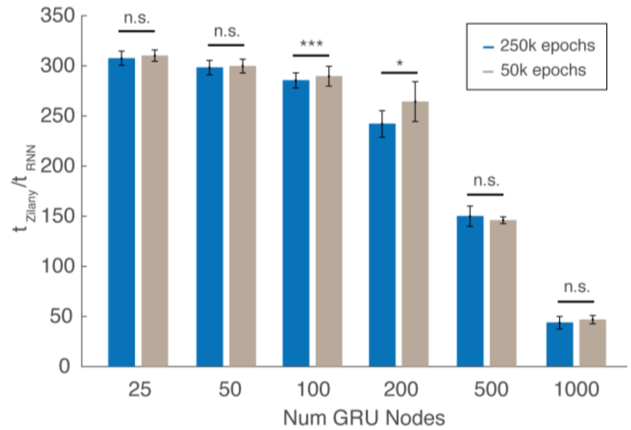

Figure 6. Run-time Evaluation of RNN. The ratio of run-time for calculating the response to a 0.5 second synthetic sound was measures across 10 runs with the Zilany model and the trained RNN one 1 CPU from a 2.4 GHz 8-Core Intel Core i9 Processor on a 2019 15-inch MacBook Pro.

If we implement the RNN using fundamental blocks rather than the ones provided by Python libraries, we can calculate the number of operations for each sound sample. This calculation yields the following number of operations for each sound sample: (One GRU node calculation = 48 operation) + (32 Linear operations, one for each output node: 32(2N), where N is the number of GRU nodes). For a 25 node RNN, we expect only 48+25*2*32 = 1648 operations. If these operations are to be completed in 6.4ms, the processor must be able to execute 1648/0.0064 = 250,000 operations per second. Assuming the typical average 4 cycles/operation, the clock speed of this processor must be 250K*4=1MHz. If we assume there are 20 channels that must execute at the same time, one for each electrode, we will need a 20MHz microprocessor. While many highly powerful microcontrollers exist that function at 40 or 80 MHz, this is a comfortable execution speed for even a common modern microcontroller, such as MSP430 which executes at 24 MHz[20]. At 142 µA/MHz, a typical cochlear implant battery with 126mAh would have a 126mAh/(20 MHz*142 µA/MHz) = 44 hour battery life [21].

With these results, we feel this is a promising approach for creating a real-time front end for a cochlear implant that can generate realistic target responses. To use this novel front end to the benefit of patients, algorithms also require accurate transformation of a predicted firing pattern to a stimulation pattern that can evoke this firing patter in actual neurons. These algorithms also need to be able to incorporate complexities of how stimulation parameters, such as pulse amplitude and rate affect induced firing rate; however, recent studies have begun to explore these exact issues [15]. Ultimately, efficacy of these approaches will require clinical evaluation of speech and pitch perception using the novel CI processing algorithm.

IV. Conclusions

A real-time front end for a CI was created using an RNN. The RNN could use less than a 100-node GRU layer and a fully connected layer and perform the transformation of sinusoidal and step neural response prediction with less than 5 percent error. Additionally, the relationships between sound and predicted firing pattern on this simplified task transfers to natural sound and captures a number of non-linearities in the transformation of sound into firing rate encoding by the cochlea. These RNNs can run over 300 times faster than the only existing phenomenological model that perform this task, can accurately produce natural cochlear responses to sound, and run at real-time speeds on a typical cochlear implant microprocessors with reasonable battery life. This is the first step towards creating a novel neural implant algorithm that generates neural responses with the fine timing of natural population responses in the body and we hope the first step to restoring pitch perception in CI users.

REFERENCES

[1] Y. R. Semenov, R. Martinez-Monedero, and J. K. Niparko, “Cochlear Implants. Clinical and Societal Outcomes,” *Otolaryngologic Clinics of North America*, vol. 45, no. 5. Elsevier, pp. 959–981, 01-Oct-2012.

[2] “What Are Cochlear Implants for Hearing? | NIDCD.” [Online]. Available: https://www.nidcd.nih.gov/health/cochlear-implants. [Accessed: 23-Jul-2021].

[3] B. Wilson, M. D.-J. R. R. Dev, and undefined 2008, “Cochlear implants: current designs and future possibilities,” *books.google.com*.

[4] C. G. Wei, K. Cao, and F. G. Zeng, “Mandarin tone recognition in cochlear-implant subjects,” *Hear. Res.*, vol. 197, no. 1–2, pp. 87–95, Nov. 2004.

[5] P. E. Riley, D. S. Ruhl, M. Camacho, and A. M. Tolisano, “Music Appreciation after Cochlear Implantation in Adult Patients: A Systematic Review,” *Otolaryngology - Head and Neck Surgery (United States)*, vol. 158, no. 6. SAGE Publications Inc., pp. 1002–1010, Jun-2018.

[6] J.-J. Sit, A. M. Simonson, A. J. Oxenham, M. A. Faltys, and R. Sarpeshkar, “A Low-Power Asynchronous Interleaved Sampling Algorithm for Cochlear Implants That Encodes Envelope and Phase Information,” *IEEE Trans. Biomed. Eng.*, vol. 54, no. 1, 2007.

[7] B. S. Wilson and M. F. Dorman, “Cochlear implants: a remarkable past and a brilliant future.,” *Hear. Res.*, vol. 242, no. 1–2, pp. 3–21, Aug. 2008.

[8] A. H. Mehta and A. J. Oxenham, “Vocoder Simulations Explain Complex Pitch Perception Limitations Experienced by Cochlear Implant Users,” *JARO - J. Assoc. Res. Otolaryngol.*, vol. 18, no. 6, pp. 789–802, Dec. 2017.

[9] F. Chen and Y. T. Zhang, “A novel temporal fine structure-based speech synthesis model for cochlear implant,” *Signal Processing*, vol. 88, no. 11, pp. 2693–2699, Nov. 2008.

[10] N. Lan, K. B. Nie, S. K. Gao, and F. G. Zeng, “A Novel Speech-Processing Strategy Incorporating Tonal Information for Cochlear Implants,” *IEEE Trans. Biomed. Eng.*, vol. 51, no. 5, 2004.

[11] M. S. A. Zilany, I. C. Bruce, P. C. Nelson, and L. H. Carney, “A phenomenological model of the synapse between the inner hair cell and auditory nerve: Long-term adaptation with power-law dynamics,” *J. Acoust. Soc. Am.*, vol. 126, no. 5, pp. 2390–2412, Nov. 2009.

[12] M. S. A. Zilany, I. C. Bruce, and L. H. Carney, “Updated parameters and expanded simulation options for a model of the auditory periphery,” *Cit. J. Acoust. Soc. Am.*, vol. 135, p. 283, 2014.

[13] C. L. Runge, F. Du, and Y. Hu, “Improved Speech Perception in Cochlear Implant Users With Interleaved High-Rate Pulse Trains,” *Otol. Neurotol.*, vol. 39, no. 5, pp. e319–e324, Jun. 2018.

[14] W. Lamping, T. Goehring, J. Marozeau, and R. P. Carlyon, “The effect of a coding strategy that removes temporally masked pulses on speech perception by cochlear implant users,” *Hear. Res.*, vol. 391, p. 107969, Jun. 2020.

[15] C. R. Steinhardt and G. Y. Fridman, “Predicting Response of Spontaneously Firing Afferents to Prosthetic Pulsatile Stimulation,” in *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, 2020, vol. 2020-July, pp. 2929–2933.

[16] P. Warden, “Speech Commands: A Dataset for Limited-Vocabulary Speech Recognition,” *arXiv*. arXiv, 09-Apr-2018.

[17] L. H. Carney, “A model for the responses of low-frequency auditory-nerve fibers in cat,” *J. Acoust. Soc. Am.*, vol. 93, no. 1, pp. 401–417, Jan. 1993.

[18] M. Wirtzfeld, “Predicting speech intelligibility and quality from model auditory nerve fiber mean-rate and spike-timing activity,” 2017.

[19] M. J. Bianco *et al.*, “Machine learning in acoustics: Theory and applications,” *arXiv*. arXiv, 10-May-2019.

[20] “Product Folder Order Now MSP430FR235x, MSP430FR215x Mixed-Signal Microcontrollers 1 Device Overview,” 2018.

[21] “What’s the State of Rechargeable Batteries for Hearing Aids?” [Online]. Available: https://www.hearingreview.com/hearing-products/hearing-aids/whats-state-rechargeable-batteries-hearing-aids. [Accessed: 28-Jul-2021].