Speech recognition using biologically-inspired neural networks

Thomas Bohnstingl, Ayush Garg, Stanisław Woźniak,
George Saon, Evangelos Eleftheriou and Angeliki Pantazi

Paper ID: 4711

icassp 2022
Singapore, China, Virtual

2022 IEEE International Conference on Acoustics, Speech, and Signal Processing

Biologically-inspired automatic speech recognition systems (ASR), based on spiking neural networks (SNNs) are so far lagging in terms of accuracy and focus primarily on small scale applications. In this work, we substantially enhance their capabilities, by taking inspiration from the diverse neural and synaptic dynamics found in the brain.
In particular, we introduce

- A novel neural connectivity concept emulating the **axo-somatic synapses**
- A novel neural connectivity concept emulating the **axo-axonic synapses**
- A biologically-inspired RNN-T architecture with **significantly reduced computational cost** and **increased throughput**

Enhancing models for deep learning with insights from biology

Biology provides an abundance of mechanisms, which could enhance the dynamics of neurons. In this work, we resort to the diverse types of synapses and neurons present in the brain and enhance the leaky integrate-and-fire (LIF) neuron model with a threshold adaptation mechanism based on axo-somatic synapses as well as with an output modulating mechanism based on the axo-axonic synapses. We build upon the spiking neural units (SNUs) [1], which allows to leverage advanced training capabilities from the machine learning (ML) domain [2,3].

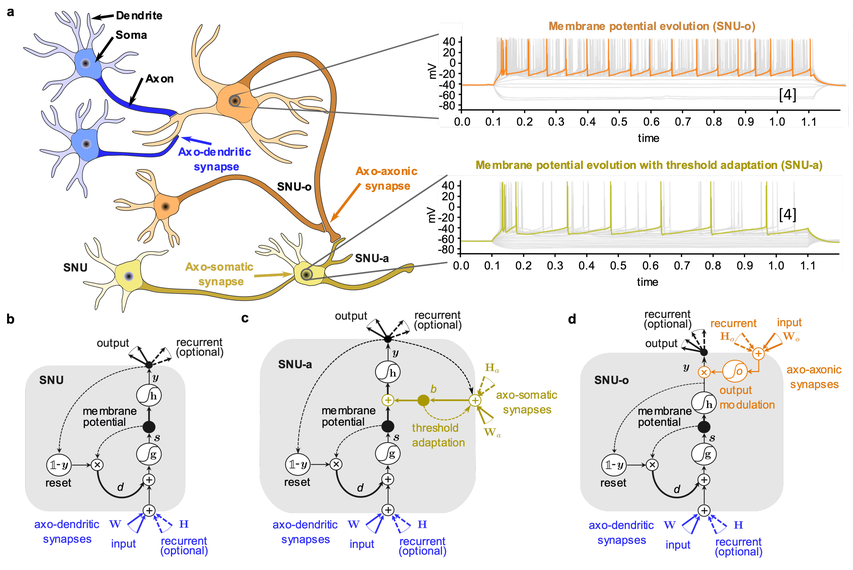


Illustration of different biological synapse and neuron types and their realization in simulations.

Biologically-inspired RNN-T achieves competitive performance with state-of-the-art

We integrate our proposed novel units in the encoding as well as in the prediction network of the recurrent neural network transducer (RNN-T), resulting in a network architecture for speech recognition solely based on biologically-inspired neurons.

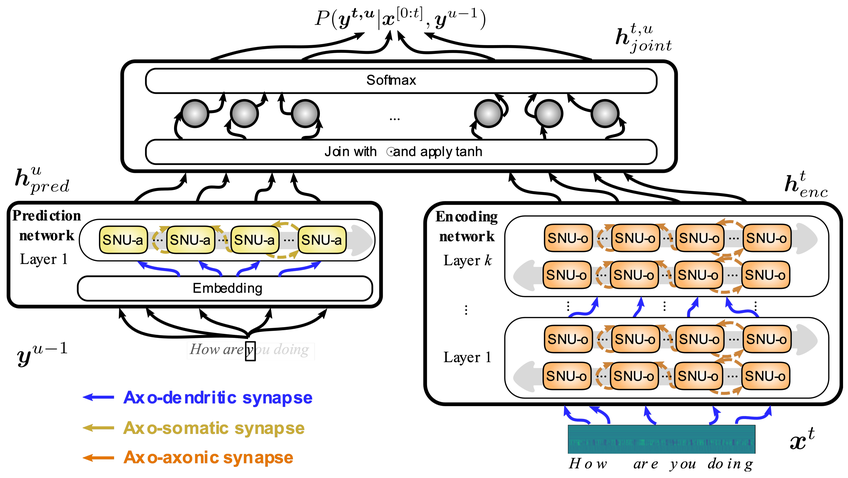


Illustration of the RNN-T architecture using biologically-inspired units

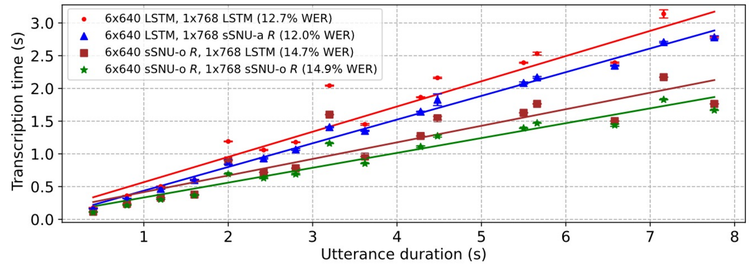
We tested our model on the Switchboard speech corpus and were able to show that end-to-end speech recognition with biologically-inspired units is possible and that they achieve competitive WER performance compared to the LSTM-based RNN-T.

Enc. RNN	Pred. RNN	WER (%)	# Param.	(%)	# Multipl.	(%)	t_{inf}(s)	(%)
LSTM	LSTM	12.7	2.39M	100	2.39M	100	2.78	100
	sSNU	15.1	8.45k	< 1	9.2k	< 1	2.71	97
	sSNU *R*	12.4	0.60M	25	0.60M	25	2.76	99
	sSNU-a	12.1	8.45k	< 1	11.52k	< 1	2.73	98
	sSNU-a *R*	12.0	0.60M	25	0.60M	25	2.78	100
	sSNU-o	12.6	16.90k	< 1	17.66k	<1	2.75	98
	sSNU-o *R*	12.4	1.20M	50	1.20M	50	2.76	99
LSTM	LSTM	12.7	54.20M	100	54.20M	100	2.78	100
sSNU-a *Ra*		25.2	18.47M	34	18.50M	34	1.23	44
sSNU-o		23.2	17.27M	32	17.28M	32	1.22	44
sSNU-o *R*		14.7	27.10M	50	27.11M	50	1.76	63
LSTM	LSTM	12.7	56.59M	100	56.58M	100	2.78	100
sSNU-o *R*	sSNU-a *R*	16.0	27.70M	49	27.71M	49	1.64	59
sSNU-o *R*	sSNU-o *R*	14.9	28.30M	50	28.30M	50	1.66	60

- Diverse types of neurons and synapses provide powerful enhancements for the neuronal dynamics
- An RNN-T network with biologically-inspired units is competitive in a large-scale speech recognition task
- Using biologically-inspired units can significantly reduce the computational cost as well as the inference time

Energy-efficiency and transcription time is significantly improved

The inference time and the computational cost are critical metrics for speech recognition. With our proposed units, the inference time as well as the computational cost can be reduced by up to 50% and 40%, respectively.



Comparison of the transcription time of various architectures.

References

[1] Stanisław Woźniak et al. (2020). "Deep learning incorporating biologically inspired neural dynamics and in-memory computing" In: Nature Machine Intelligence, pp. 325-336

[2] George Saon et al. (2021) "Advancing RNN Transducer Technology for Speech Recognition". In: IEEE International Conference on Acoustics, Speech and Signal Processing, pp. 5654-5658

[3] Diederik P. Kingma et al. (2015). "Adam: A Method for Stochastic optimization" In: 3rd International Conference on Learning Representations

[4] A. I. for Brain Science (2010). "Allen Human Brain Atlas"