Spiking Neural Network for Speech Intonation Modelling

Loïc Jeanningros, Philip N. Garner

Idiap Research Institute, Martigny, Switzerland

***Abstract*—**

I. Introduction

Intonation is a prosodic feature of speech that carries non-linguistic information such as emphasis and emotion. As a distorded pitch can change the meaning of a sentence, or can reveal the speaker's emotional state, a good model of intonation is crucial for speech-to-speech translation systems that intend to transfer paralinguistics between languages.

In previous work with colleagues [4], we investiguated a physiologically plausible intonation ($F0$) model based on the Command-Response model of Fujisaki [2]. We presented a Generalized CR intonation contour using a matching pursuit algorithm [8].

The task to perform then consists of learning temporal sequences of spikes, i.e. spike trains. Hence, we propose to resolve that learning task directly with a Spiking Neural Network (SNN) model.

Most important problem of supervised learning for SNN is binary character of of spiking neurons [3]. No obvious differentiable error function can be backpropagated with gradient techniques. SpikeProp algorithm [1] employed difference between desirable and actual time of output spike as error but does not enable learning patterns composed of more than one spike. A supervised spike-based learning algorithm [11] that optimizes the likelihood of postsynaptic firing by gradient ascent allows to learn few spikes, but it is hard to estimate a potential to learn complex spike trains. A gradient descent technique [5] allows to learn spike trains but it breaks the all-or-none nature of synaptic currents.

Nevertheless, reservoir computing is a new trend of understanding training that has been started with Liquid State Machine (LSM) [7] for the spike-based version. The basic idea behind reservoir computing is that, as long as an Recurrent Neural Network (RNN) possesses certain generic properties, supervised adaptation of all interconnection weights is not necessary, and only training a memoryless supervised readout from it is enough to obtain excellent performance in many tasks [6].

The LSM is a task-independent high-dimensional network based neocortex microcircuits observation. The first component of a LSM is a filter, the liquid itself, outputs of that filter called the liquid state. The liquid state has only one attractor: the resting state. Perturbed states of the liquid represent present and past inputs. It has proven to have universal computational power and universal analog fading memory.

The second component of an LSM is a memoryless readout map that is trained to exctract information from the liquid state. Two similar learning rules are used in this purpose, ReSuMe [13], [12] and SPAN [9], [10].

Reservoir computing has never been used for speech synthesis. Moreover, learning rules used here are suited for learning complex spike trains, but it is not obvious wether such algorithm are capable of generalization, able to predict unseen samples dynamic.

SNN has proved itself adequate for a number of computation and engineering problems, It is considered as a suitable tool to perform temporal pattern recognition and real-time computation [9].

Due to the ability of the method to operate online and due to its fast convergence, the method is suitable for real-life applications. SNN trained with ReSuMe become efficient neurocontrollers for movement generation and control [13].

APPENDIX

ACKNOWLEDGMENT

References

[1] Sander M Bohte, Joost N Kok, and Han La Poutre. Error-backpropagation in temporally encoded networks of spiking neurons. *Neurocomputing*, 48(1-4):17–37, 2002.

[2] Hiroya Fujisaki, Sumio Ohno, and Changfu Wang. A command-response model for f0 contour generation in multilingual speech synthesis. In *The Third ESCA/COCOSDA Workshop (ETRW) on Speech Synthesis*, 1998.

[3] Andrey V Gavrilov and Konstantin O Panchenko. Methods of learning for spiking neural networks. a survey. In *2016 13th International Scientific-Technical Conference on Actual Problems of Electronics Instrument Engineering (APEIE)*, volume 2, pages 455–460. IEEE, 2016.

[4] Pierre-Edouard Honnet, Branislav Gerazov, and Philip N Garner. Atom decomposition-based intonation modelling. In *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 4744–4748. IEEE, 2015.

[5] Dongsung Huh and Terrence J Sejnowski. Gradient descent for spiking neural networks. In *Advances in Neural Information Processing Systems*, pages 1440–1450, 2018.

[6] Mantas Lukoševičius, Herbert Jaeger, and Benjamin Schrauwen. Reservoir computing trends. *KI-Künstliche Intelligenz*, 26(4):365–371, 2012.

[7] Wolfgang Maass, Thomas Natschläger, and Henry Markram. Real-time computing without stable states: A new framework for neural computation based on perturbations. *Neural computation*, 14(11):2531–2560, 2002.

[8] Stéphane G Mallat and Zhifeng Zhang. Matching pursuits with time-frequency dictionaries. *IEEE Transactions on signal processing*, 41(12):3397–3415, 1993.

[9] Ammar Mohemmed, Stefan Schliebs, Satoshi Matsuda, and Nikola Kasabov. Span: Spike pattern association neuron for learning spatio-temporal spike patterns. *International journal of neural systems*, 22(04):1250012, 2012.

[10] Ammar Mohemmed, Stefan Schliebs, Satoshi Matsuda, and Nikola Kasabov. Training spiking neural networks to associate spatio-temporal input–output spike patterns. *Neurocomputing*, 107:3–10, 2013.

[11] Jean-Pascal Pfister, Taro Toyoizumi, David Barber, and Wulfram Gerstner. Optimal spike-timing-dependent plasticity for precise action potential firing in supervised learning. *Neural computation*, 18(6):1318–1348, 2006.

[12] Filip Ponulak. Supervised learning in spiking neural networks with resume method. *Phd, Poznan University of Technology*, 46:47, 2006.

[13] Filip Ponulak and Andrzej Kasiński. Supervised learning in spiking neural networks with resume: sequence learning, classification, and spike shifting. *Neural computation*, 22(2):467–510, 2010.