Towards Classifying Human Phonemes without encodings via spatiotemporal liquid state machines

Extended Abstract

Alex Frid[1], Hananel Hazan[23] and Larry Manevitz[3]

Abstract— The abstract is extended below

Index Terms— : Liquid State Machine, speech synthesis, unit selection, Parkinson's Disease, classification, Machine Learning, spatiotemporal

1 Extended Abstract

Recently, we showed that liquid state machines [1] can be adapted to give robust pattern recognition of temporal patterns [2]. Moreover we showed how certain natural modifications in the neurons enables the liquid state machine to compute on complex continuous real valued patterns without the need for discretization and digital encoding [3]. In attempting to apply this technique to signal processing on phoneme recognition from a continuous voice signal, we found this to be intractable; because of problems in both making the system accurate while maintaining good generalizability properties. It seems that the combination of the separability of the liquid with the digital encoding conflict with the generalizability.

Current machine learning and other techniques are fairly accurate in classifying phonemes after substantial preprocessing. However this is not a very natural methodology and can not be related to human decoding of speech.

In this work we investigate whether recent spatiotemporal methodologies can successfully decode the raw speech signal into phonemes without such special preprocessing as depicted in Figure 1. Our methodology uses a recent robust version of Liquid State Machines [4].

We have successfully shown that the method (i) can decode and categorize synthetic signals which are statistically similar to natural voiced phonemes and (ii) can successfully decode and categorize artificially induced noisy versions of those signals and thus has good generalization capabilities. (iii) Can successfully classify between three representative phonetic groups of vowels (the front, mid and back groups) to 77% level. The phonemes were extracted from natural speech recorded in TIMIT Acoustic-Phonetic Continuous Speech Corpus [5].

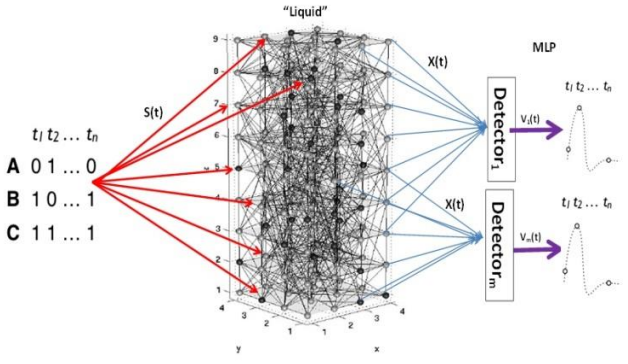

Figure 1: *Diagram of the Liquid / Echo State Machine with the input afferents and read out detectors. In our work, the elements of the liquid have been examined with different properties as has the connectivity and strength between the members of the liquid.*

References

[1] W. Maass, T. Natschläger, and H. Markram, "Real-time Computing Without Stable States: A New Framework for Neural Computation Based on Perturbations," Neural Computation, vol. 14, no. 11, pp. 2531–2560, 2002.

[2] L. Manevitz and H. Hazan, "Stability and Topology in Reservoir Computing," in Advances in Soft Computing, vol. 6438, G. Sidorov, A. Hernández Aguirre, and C. Reyes García, Eds. Springer Berlin / Heidelberg, 2010, pp. 245–256.

[3] A. Frid, H. Hazan, and L. Manevitz, "Temporal pattern recognition via temporal networks of temporal neurons," in 2012 IEEE 27th Convention of Electrical Electronics Engineers in Israel (IEEEI), 2012, pp. 1–4.

[4] H. Hazan and L. M. Manevitz, "Topological constraints and robustness in liquid state machines," Expert Systems with Applications, vol. 39, no. 2, pp. 1597–1606, Feb. 2012.

[5] J. S. Garofolo, L. F. Lamel, W. M. Fisher, J. G. Fiscus, and D. S. Pallett, "DARPA TIMIT acoustic-phonetic continous speech corpus CD-ROM.

1. *Edmond J. Safra Brain Research Center for the Study of Learning Disabilities, University of Haifa, Israel Email: alex.frid@gmail.com*
2. *Lorrey L. Lokey Network Biology Laboratories, Technion - Israel Institute of Technology, Haifa, Israel. Email: hananel@hazan.org.il*
3. *Department of Computer Science, University of Haifa, Israel. Email: dan.hilu@hotmail.com*

NIST speech disc 1-1.1," NASA STI/ Recon Technical Report N, vol. 93, p. 27403, Feb. 1993.