Neural mechanisms of speech processing: time warp invariance by adaptive integration time

R. Gütig^{1,2} **and H. Sompolinsky**$^{1-3}$

[1]Racah Institute of Physics and [2]ICNC, Hebrew University of Jerusalem, Israel.
[3]Center for Brain Science, Harvard University, Cambridge, USA.

In speech processing, perceptually relevant temporal cues can require resolution of spectral transitions with millisecond precision. However, dynamic variations in speaking rate on a scale of hundreds of milliseconds introduce time warp of spoken syllables, ranging from two fold compression to two fold dilation. Although the auditory systems of adult humans, pre-articulate infants and also non-human primates have been shown to readily normalize temporal cues underlying phonetic category boundaries for speaking rate effects [1], the neural mechanisms subserving this time warp invariance have remained mysterious. In addition to time warp robustness, the neural hardware must also support processing of temporal features over different relative scales, e.g. slow formant transitions vs. fast voice on- and offset times.

To resolve these computational challenges by means of a biologically plausible spiking neuronal network, we studied a neural model of spike-based learning of auditory discrimination tasks, extending our recent tempotron model [2]. In our present model, synaptic inputs are modeled as conductances rather than current pulses, such that the time scale of the resulting voltage dynamics is governed by the total synaptic conductance. As a result, the statistics of inputs as well as the synaptic learning rule, interact with the time scale of neural information processing by changing the effective integration time constant of the cell. For instance, by balancing excitatory and inhibitory synaptic conductances a neuron can adjust its effective integration time scale to match the temporal extent of its acoustic target feature. Moreover, since in a high conductance state the effective membrane time constant of a neuron scales inversely with its total conductance, the membrane voltage exhibits a high degree of time warp invariance. Warping of input spike trains which increases the total conductance, such as time compression, or decrease it when time is stretched, are canceled by the corresponding change in the membrane integration time.

We have applied our model to speech recognition problems. Using a simple model of the auditory periphery, sound signals were converted into spike patterns by thresholding their power-spectral densities, and fed into small populations of conductance-based LIF neurons. Surprisingly, using our tempotron learning, as few as 15 neurons were sufficient to achieve perfect word recognition on the TI46 [3] digit recognition task, outperforming complex state-of-the-art hidden-markov-model word recognition systems (e.g. Spinx 4). Analysis of the spectro-temporal receptive fields of the trained neurons revealed the neural target features solving the digit recognition task. Our results demonstrate the powerful role of synaptic conductances in spike-timing based neural processing.

Acknowledgments
This work was supported by the Minerva Foundation and the Israeli Science Foundation.

References

[1] Miller JL (1981) Some effects of speaking rate on phonetic perception. Phonetica 38:159. [2] Gütig R, Sompolinsky H (2006) The tempotron: a neuron that learns spike timing-based decisions. Nat Neurosci 9: 420-428. [3] Liberman M et al. (1993) TI 46-Word. Linguistic Data Consortium, Philadelphia.