Time-Warp-Invariant Neuronal Processing

Robert Gütig, Haim Sompolinsky

PLoS Biology, 2009

presented by Dejan Pecevski

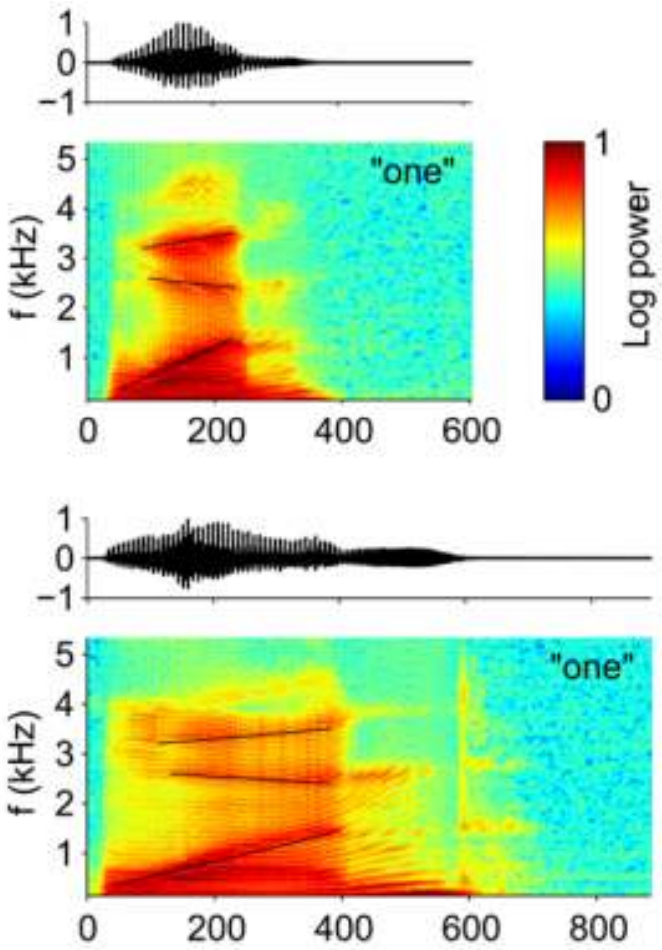
CI Seminar C, December 2009

Outline

- Introduction: auditory neuronal processing and time warping
- Conductance-based time rescaling mechanism
- The tempotron learning rule
- Simulation experiments
  - Random latency patterns discrimination
  - Spoken digit recognition
- Relation to
  - neurophysiology of the auditory system
  - psychoacoustics and speech processing
- Summary

Speech processing and time-warping

- Auditory (and speech) perception in humans relies on differences between temporal cues in the stimuli.
- E.g. recognition of consonants depends on differences in voice onsets, spectral transitions.
- Auditory stimuli and speech exhibit significant temporal-warping variability
  - Changes in speaking rate introduce 2-fold compression and 2-fold dilation



Spech processing and time-warping

- Still, the auditory neuronal processing is resilient to this variability and produces stable representations.

What are the neural mechanisms that provide this temporal variability robustness in speech (and auditory) processing?

- Proposed answer in the paper: Automatic time rescaling through synaptic shunting

The synaptic conductances adjust the effective integration time constant to match the rate of the input spike trains.

Neuron model

LIF neuron driven by exp. decaying synaptic conductances

$$g_i(t) = g_i^{max} exp(-\frac{t}{\tau_s})$$

The differential equation for the membrane potential $V(t, \beta)$

$$\frac{d}{dt}V(t,\beta) = -V(t,\beta)(g_{leak} + G_{syn}(t,\beta)) + I_{syn}(t,\beta)$$

$$G_{syn}(t,\beta) = \sum_{i=1}^{N} \sum_{t_i^j < t} g_i(t - \beta t_i^j) \quad I_{syn}(t,\beta) = \sum_{i=1}^{N} \sum_{t_i^j < t} V_i^{rev} g_i(t - \beta t_i^j)$$

β is a global scaling factor for the incoming spike times,

t_i^j are the input spike times,

V_i^{rev} is the reversal potential,

g_{leak} is the leak conductance,

τ_s is the synaptic time constant,

Time rescaling

$$\frac{d}{dt}V(t,\beta) = -V(t,\beta)(g_{leak} + G_{syn}(t,\beta)) + I_{syn}(t,\beta)$$

$$I_{syn}(t,\beta) = \sum_{i=1}^{N} \sum_{t_i^j < t} V_i^{rev} g_i(t - \beta t_i^j)$$

$$G_{syn}(t,\beta) = \sum_{i=1}^{N} \sum_{t_i^j < t} g_i(t - \beta t_i^j)$$

- If one assumes that τ_s is small, then

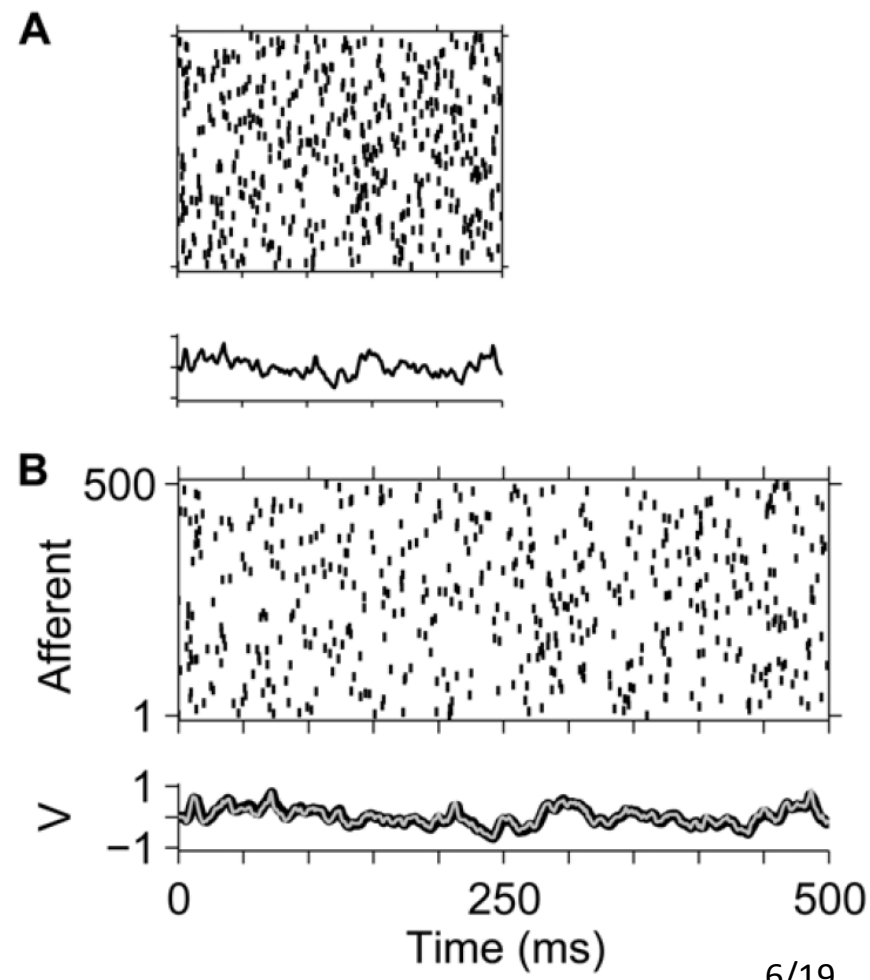
$$I_{syn}(\beta t,\beta) \approx \beta^{-1} I_{syn}(t,1)$$

$$G_{syn}(\beta t,\beta) \approx \beta^{-1} G_{syn}(t,1)$$

- And if $G_{syn}(t,\beta) \gg g_{leak}$

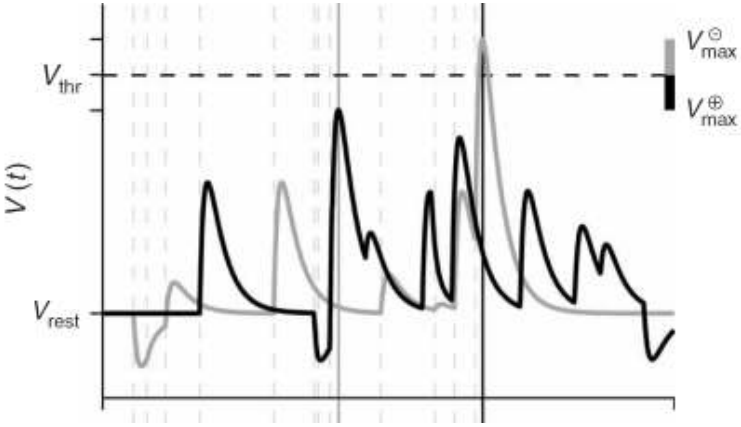
$$\tau_{eff} = \frac{1}{g_{leak} + G_{syn}(t)} \approx \frac{1}{G_{syn}(t)}$$

$$V(\beta t,\beta) \approx V(t,1)$$



The Tempotron Learning Rule

- Supervised learning, two classes of patterns, target $\oplus$ and null $\ominus$
- The neuron learns to output
  - only one spike for $\oplus$ patterns
  - no spikes for $\ominus$ patterns



- From (Gütig, 2006), with current based neuron

$$V(t) = \sum_i \omega_i \sum_{t_i^j} K(t - t_i^j) + V_{rest}$$

$$\Delta\omega_i = \lambda \sum_{t_i^j < t_{max}} K(t_{max} - t_i^j) = \lambda \frac{\partial V(t_{max})}{\partial w_i}$$

where

$\lambda > 0$ if the neuron is silent on a $\oplus$ pattern

$\lambda < 0$ if the neuron spikes on a $\ominus$ pattern

$\lambda = 0$ if the neuron spikes on a $\oplus$ pattern or is silent on a $\ominus$ pattern

t_{max} is the time where *V(t)* reaches maximum

Conductance-based Tempotron

- The peak synaptic conductance changes after each trial by the amount

$$\Delta g_i^{max} = \lambda \frac{dV(t_{max})}{dg_i^{max}}$$

where t_{max} is the time where the membrane potential is maximum and λ changes as in the current-based tempotron.

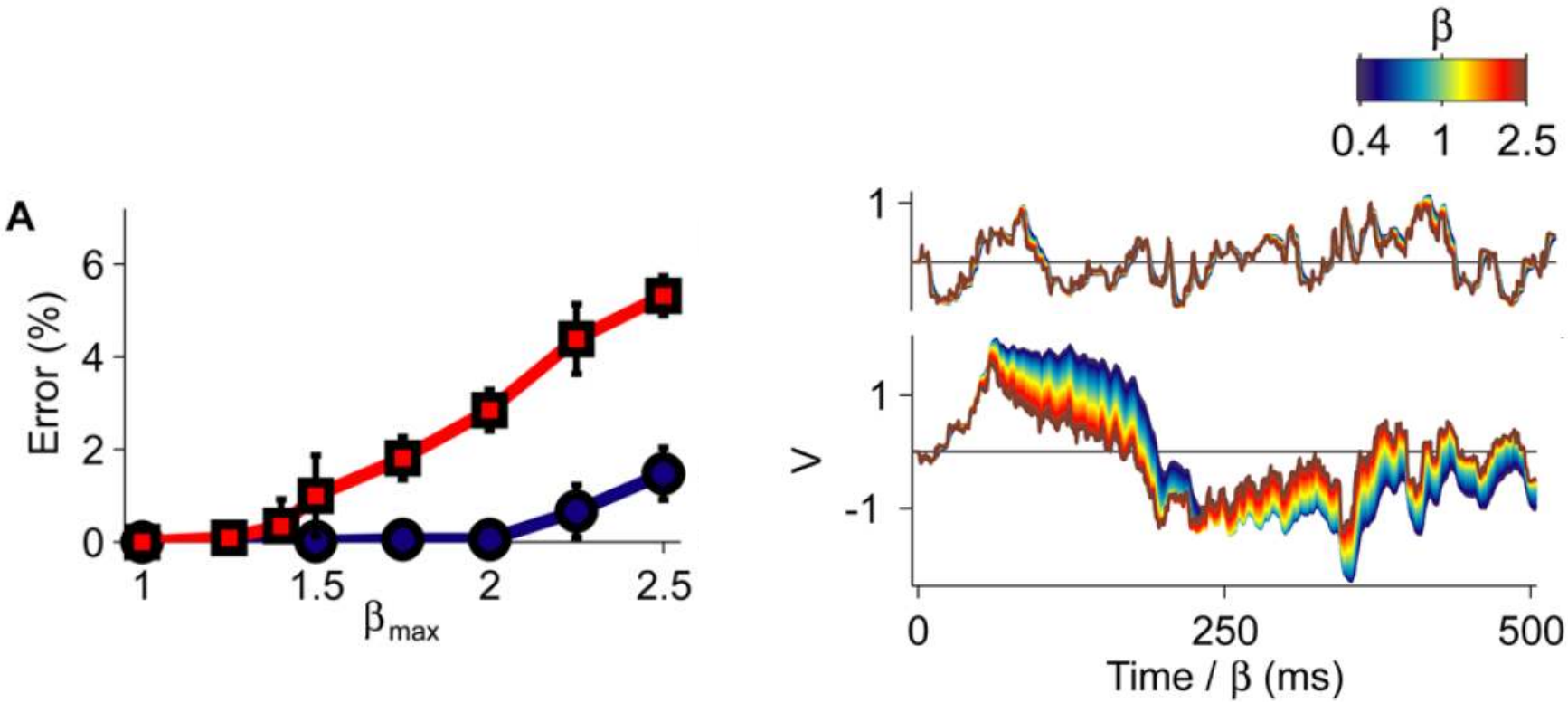
Classifying Random Latency Patterns

- 1250 spike pattern templates randomly assigned to target and null classes
- 500 afferents, one spike per afferent
- Spike times drawn from a uniform distribution between 0 and 500 ms
- Upon each presentation, the templates underwent temporal warping with a random scaling factor β between $1/\beta_{max}$ and β_{max} .
- Multiple experiments were performed where β_{max} took different values between 1 and 2.5.

Results

- The conductance based neuron (blue line) is robust to time-warping, with learning error remaining zero for up to $\beta_{max} = 2$.

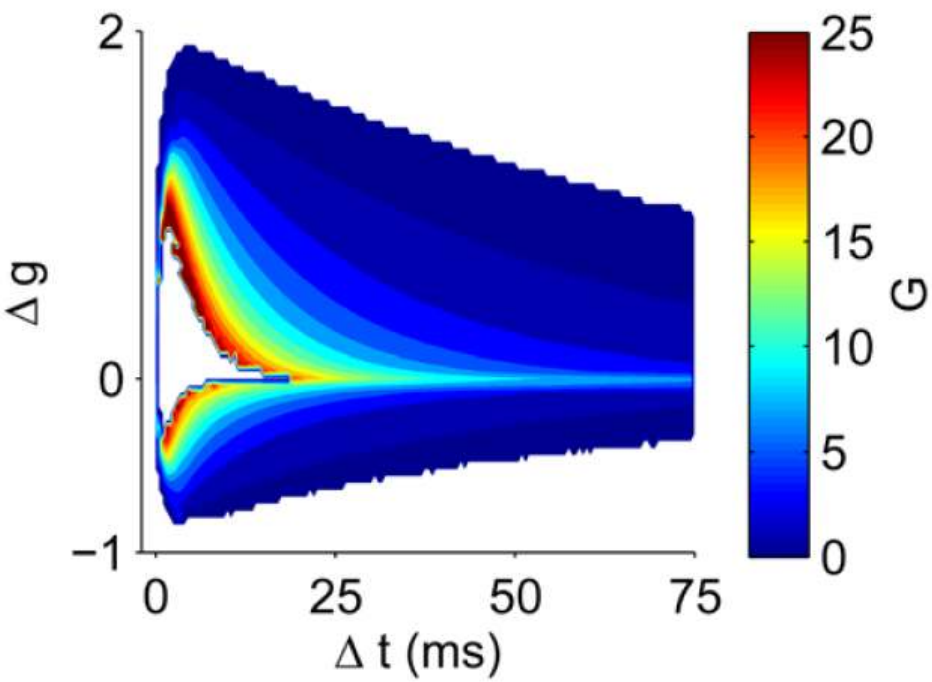
- The current-based neuron (red line) is much more sensitive to time-warping.



Adaptive learning kernel

- The plasticity window changes its effective width according to the current total synaptic conductance in the neuron.

- Important for learning, since the rule should credit appropriately the synapses that contributed to the maximum of the membrane potential.

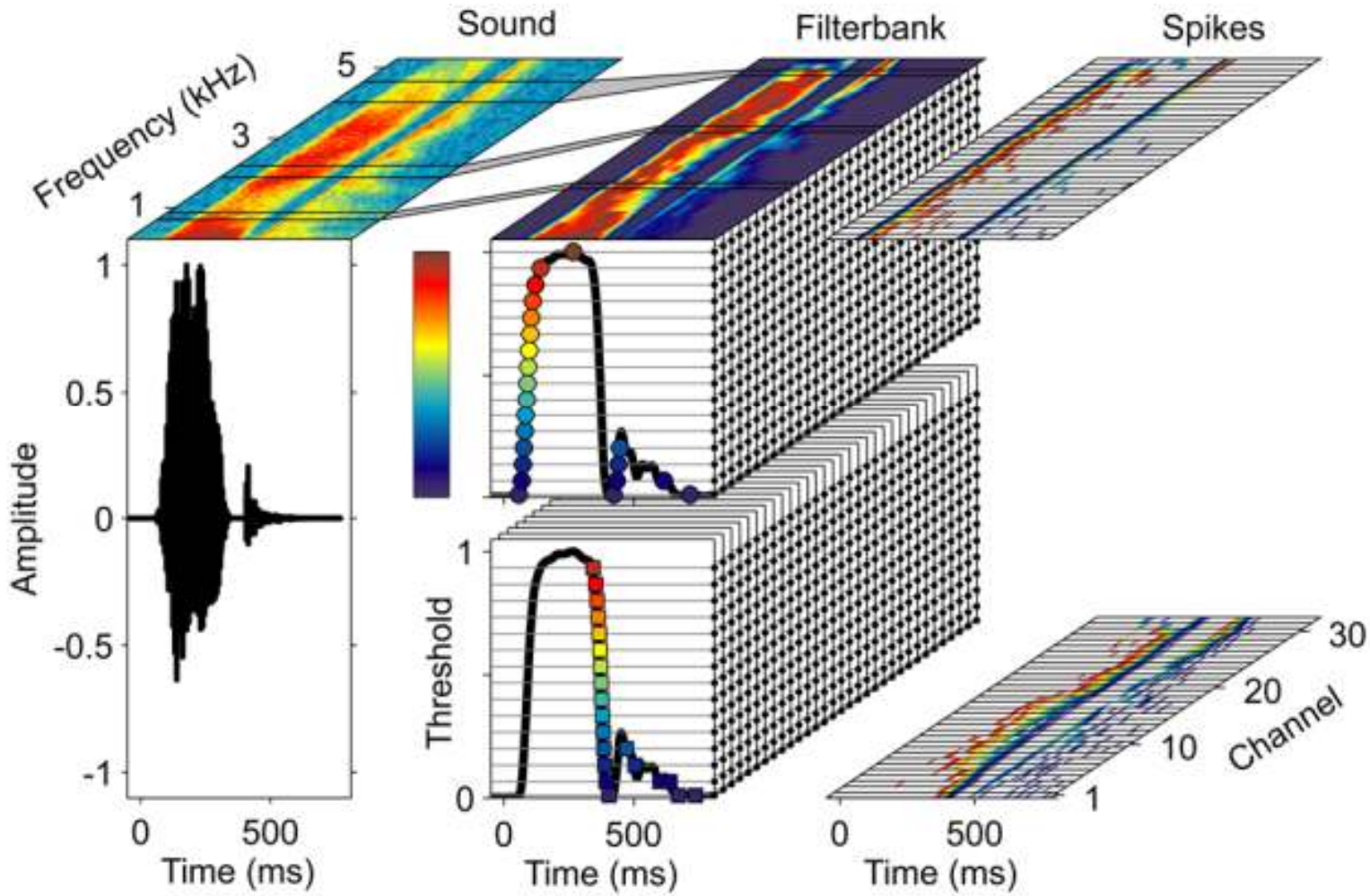


$$\Delta g_i^{max} = \lambda \frac{dV(t_{max})}{dg_i^{max}}$$

Spoken Digit Recognition Task

- The dataset was a subset of the TI46 speech recognition database (only spoken digits considered)
- Neuronal architecture of two processing stages
- The first layer of neurons converts the acoustic signal into spike patterns.
- The second layer consists of 20 conductance-based tempotrons
  - Each is trained to fire a spike in response to utterances of one digit and speaker gender, and
  - remain quiescent for other utterances

Spoken Digit Recognition: First Processing Stage

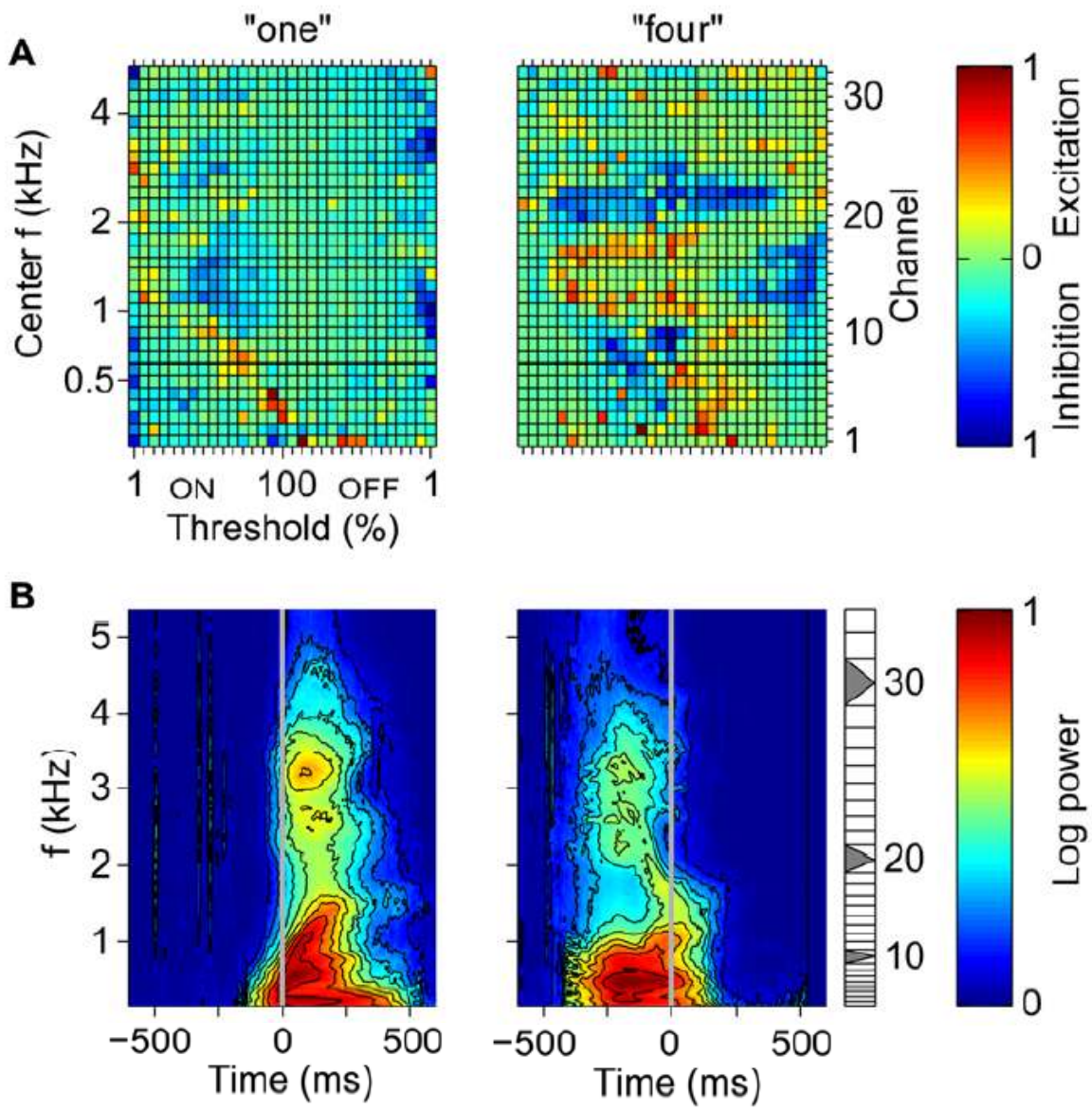


Classification Performance

- The full digit classifier by using all the detector neurons achieved test error of 0.0017.
- Matches the error rates of state-of-the-art HMM-based Sphinx-4 (0.0017) and HTK (0.0012)
- 70% of the individual selector neurons achieved 0 test error.

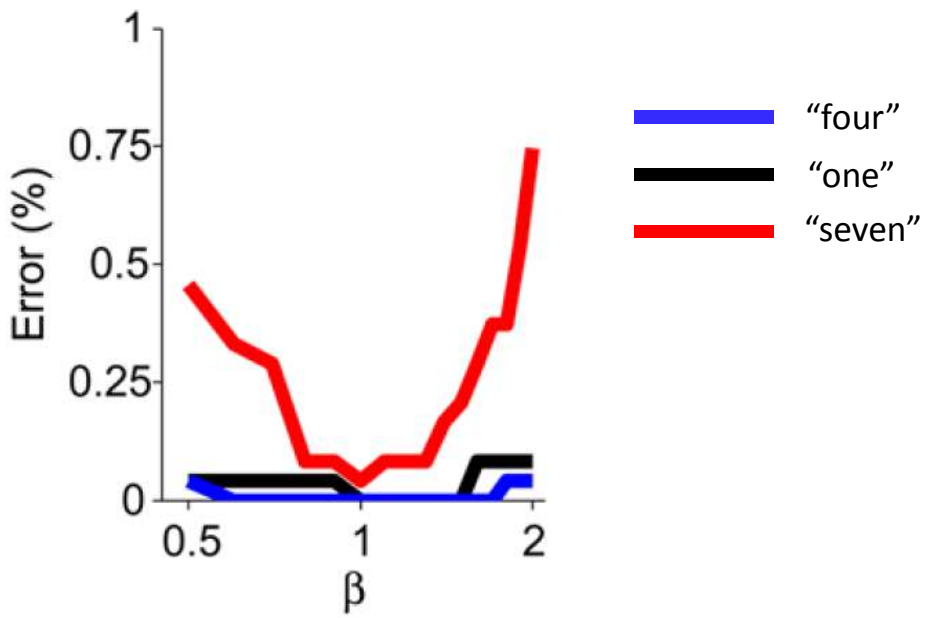
Digit	Male	Female
0	0.0	0.0
1	0.0	0.0
2	0.0008	0.0017
3	0.0	0.0
4	0.0	0.0
5	0.0029	0.0062
6	0.0	0.0
7	0.0004	0.0008
8	0.0	0.0
9	0.0	0.0

Learned Spectrotemporal Features



Generalization properties

- Achieved test error rate 0.0949 on TIDIGITS database (trained on TI46)
- HTK achieved error rate 0.2156 on the same test.
- The error remains low for time-warped versions of the trained input spike patterns.



Relations to Neurophysiology

Mapping to brain structures:

- Neurons in **Inferior Colliculus** of the auditory midbrain have frequency tuned onset and offset firings with different thresholds
  - similar to the first layer in the proposed architecture

- The second layer of time-warp invariant feature detectors are localized in the **primary auditory cortex**

- Studies have found functional lateralization in auditory processing
  - left auditory cortices process features at time scales of 10 ms
  - right auditory cortices process longer temporal features

- Corollary from the model:
  - Right auditory cortex neurons operate with low synaptic conductances
  - Left auditory cortex neurons are faster time-warp invariant feature detectors: high-conductance regimes

Implications for Psychoacoustics and Speech Processing

Studies have suggested that perceptual normalization of temporal speech cues is

- involuntary
- controlled by physical (rather than perceived) speaking rate
- not specific to speech sounds
- operational in prearticulate infants

$\Rightarrow$ Consistent with the proposed conductance-based time-rescaling mechanism

- Hypothesis from the model:
  Impairment in speech comprehension in elderly listeners is a result of
  - downregulation of inhibitory neurotransmitter systems in aging mammalian auditory pathways that affects the neurons ability to generate high-conductances.

$\Rightarrow$ Consistent with studies that link decline in central auditory processing and discrimination of fine temporal cues with diminished speech comprehension

Summary

- A conductance-based **time rescaling mechanism** has been proposed based on the **change of the effective time constant** of neurons induced by synaptic conductances **.**

- This time rescaling mechanism was accompanied by a **conductance-based tempotron learning rule** that learns time-warp invariant spike pattern recognition.

- The time-invariant processing capabilities of the model were demonstrated on **spoken word recognition task** where they achieved performance comparable to the state-of-the-art HMM-based speech processing algorithms.

- The model is **consistent** with studies in neurophysiology of the auditory system, psychoacoustics and speech processing and it **generates testable hypotheses.**