Learning reconstruction and prediction of natural stimuli by a population of spiking neurons – Supplementary methods –

Michael Gutmann and Aapo Hyvärinen

Main article in Proc. ESANN2009

Preprocessing and settings for the encoding, decoding, and the learning We used speech samples from the Festvox KED timit database. [1] The speech samples are downsampled from 16kHz to 8kHz using MATLAB's decimate function. We extracted then randomly speech segments of length $T = 800$ time units (100ms), where we excluded segments of silence. Additionally, the segments were multiplied with half a period of a shifted raised cosine window (period: 160 time units) so that the segments start from zero. The segments were then smoothed with a Gaussian kernel (variance: 4 time units) and normalized to maximal value 1.

For learning, we used 7000 speech segments $x(t)$. The step sizes were $\mu_w = 0.01$ and $\mu_h = 0.001$. Delay T_d was 64 time units, T_p valued 32 time units. The number of neurons was $M = 15$. The encoding filters w_m had length $T_w = 64$ time units. For initialization of the learning rules, the initial values of the discretized w_m and h_m were drawn from a Gaussian distribution with standard deviation 0.01 and mean 0. These initial values of the encoding filters w_m are too small to trigger spikes. To have spikes even for $w_m = 0$, we set $u_m^n(t)$ to be a ramp voltage with a randomly chosen slope: For each neuron m the slope was drawn from an uniform distribution on $[\theta/240, \theta/480]$. The threshold θ was set to 4, $\eta_0 = -8$, and τ was 10 time units.

[1] Free download at http://festvox.org/dbs/dbs_kdt.html