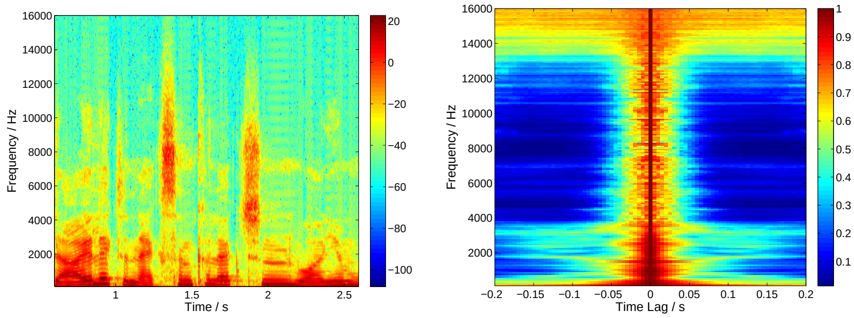
1 Introduction



(a) (b)

Figure 1: (a) the log-scaled magnitude spectrogram and (b) the frequency-wise autocorrelation function of human speech (`http://www.voxforge.org`).

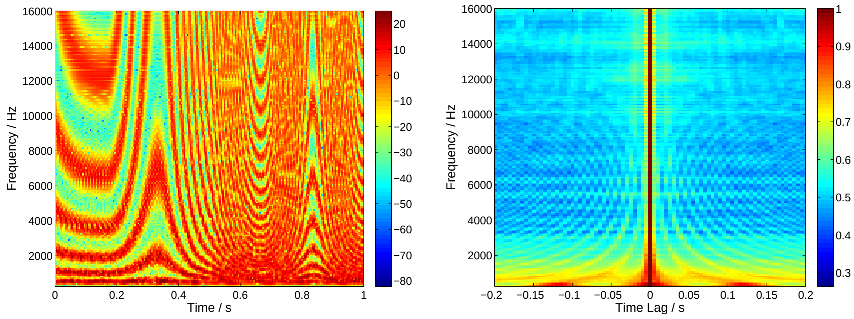
The linear auditory spectro-temporal receptive field (STRF) is a well-known approach to describe the linear relation between an acoustic stimulus and the response of a neuron evoked by that stimulus. A common STRF estimation method is reverse correlation - also known as spike-triggered average (STA) - that computes the average stimulus preceding the spikes [dK68, KDSS00]. It is assumed that patterns in the stimulus correlated with the neural response add up whereas other patterns average to zero for a large number of spikes. Hence, the STA requires stimuli with low autocorrelations, e.g. Gaussian White Noise (GWN) stimuli because otherwise patterns that are not directly correlated with the neural response will add up, too. However, GWN stimuli are not suitable to properly drive auditory neurons and alternative stimulus classes are needed for a meaningful characterization of the features encoded by cortical neurons [TSD00].

Here, we present a new stimulus class called FM banks that, similar to Dynamic Moving ripples introduced in [ES02], resembles features of conspecific vocalizations to which auditory neurons are sensitive while having low autocorrelations which makes it suitable for STRF estimation.

2 Dynamic Moving Ripples

In [ES02], a parametric stimulus class called dynamic moving ripples (DMR) has been introduced. DMRs consist of amplitude-modulated and frequency-modulated sinosoids that resemble features that are also present in vocalizations, e.g. modulated harmonic complexes.

The envelope of DMR is designed as a dynamic sinusoidal grating on an octave frequency and decibel amplitude scale. There are two parameters that define the envelope: the instantaneous ripple density, $\Omega(t)$, which defines the number of spectral peaks per octave at a given time instant and the instantaneous modulation rate, $F_m(t)$.



(a) (b)

Figure 2: (a) the log-scaled magnitude spectrogram and (b) the frequency-wise autocorrelation function of a DMR stimulus.

The envelope of the dynamic moving ripple is given by

$$S(t, x_k) = \frac{M}{2} \sin\left(2\,\pi\,(\Omega(t)\,x_k + \phi(t))\right) \tag{1}$$

where M is the modulation depth of the envelope in dB (30dB or 45dB) and $x_k = \log_2\left(\frac{f_k}{f_0}\right)$ describes the frequency modulation in octaves relative to f_0 (500Hz). The phase $\phi(t) = \int_0^t \mathrm{d}\tau\, F_m(\tau)$ is a function of time and controls the time-varying temporal modulation rate F_m. Both, the spectral parameter $\Omega(t)$ and the temporal parameter $F_m(t)$, are independent of each other and vary slowly ($0 \leq \Omega(t) \leq 3$Hz and $0 \leq F_m \leq 1.5$Hz) to resemble acoustic features in vocalizations.
The values are chosen from an uniform distribution such that they are unbiased in each interval. The idea is to draw the stimulus parameters from a random distribution which is a kind of white noise approach (reverse correlation) in the stimulus parameter space.

An example sound file can be found here.

3 FM Bank Stimuli

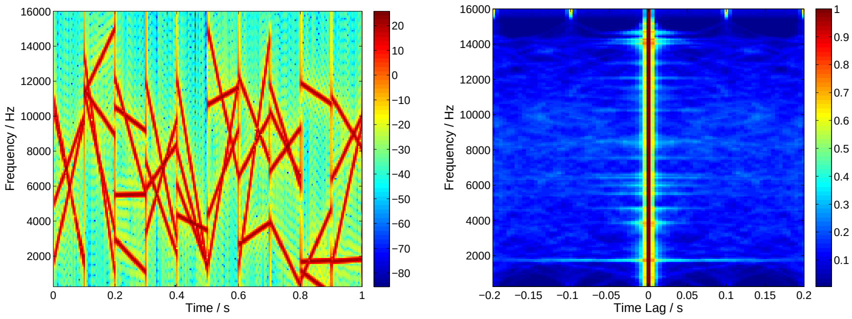
FM bank stimuli are motivated by frequency-modulated sweeps which are typical characteristics of many conspecific vocalizations, e.g. phoneme transitions in human speech.

Mathematically, a single cosine frequency sweep can be described by

$$s(t) = a \cos(2\,\pi\,\Omega(t)\,t + \phi) \tag{2}$$

where a is the amplitude, $\Omega(t)$ the instantaneous frequency at time t and ϕ the initial phase (which can also be absorbed by $\Omega(t)$). The choice of $\Omega(t)$ determines the sweep type. In this study we use linear sweeps

$$\Omega(t) = f_0 + \frac{f_1 - f_0}{t_1} t \tag{3}$$



(a) (b)

Figure 3: (a) the log-scaled magnitude spectrogram and (b) the frequency-wise autocorrelation function of a FM bank stimulus.

with starting and ending frequencies f_0 and f_1, respectively, and t_0 and t_1 are the corresponding time instants. Thus, a block of FM sweeps is given by

$$s(t) = \sum_{i=1}^{N} a_i \cos(2\,\pi\,\Omega_i(t)\,t + \phi_i). \tag{4}$$

where N is the number of sweeps that occur simultaneously. a_i is the amplitude and ϕ_i the initial phase of the i^{th} sweep.

An example sound file can be found here.

3.1 STRF estimation

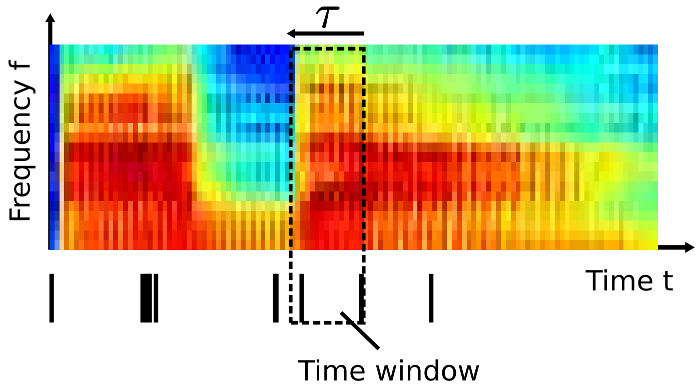


Figure 4: The principle of the STA: the parts of the stimulus preceeding spikes within a (rectangular) time window are averaged. The patterns of the stimulus that are correlated with the neural response add while other patterns average to zero for a large number of spikes. However, if there are autocorrelations in the stimulus then artefacts arise and the STA does not yield a proper estimate of the neuron's STRF.

The reverse correlation method is the analytic version of the spike-triggered average stimulus (STA) which can be understood quite intuitively. It is the average of the stimulus parts that precede spikes in the time interval $[\tau_{\mathrm{max}}, 0]$. The idea is that patterns that are correlated with the neural response will add up while the other parts will average to zero for a sufficient large number of trials. Of course, it is assumend that there are no autocorrelations in the stimulus which would also add up in the cumulative averaging process.

3.1.1 Theory

The reverse correlation method can be derived from the STA using only a few assumptions (Dayan & Abbott 2005). Mathematically, the STA for a time signal is given by

$$h(\tau) = \left\langle \frac{1}{N} \sum_{i=1}^{N} s(t_i - \tau) \right\rangle \approx \frac{1}{\langle N \rangle} \left\langle \sum_{i=1}^{N} s(t_i - \tau) \right\rangle . \tag{5}$$

where N is the number of spikes in each trial and we have to average over all trials. The approximation is valid for large N because in this case the number of spikes per trial can be approximated by the average number of spikes per trial, $N \approx \langle N \rangle$.
If we assume that a spike-train can be described by a sum of delta functions the sum in Eq. (5) can be written as integral,

$$\sum_{i=1}^{n} s(t - t_i) = \int_{-\infty}^{\infty} \mathrm{d}\tau \, s(t) \, \rho(t - \tau). \tag{6}$$

Using this relation it is possible to rewrite the approximation in Eq. (5) as

$$h(\tau) = \frac{1}{\langle N \rangle} \int_0^T \mathrm{d}t \, \langle \rho(t) \rangle \, s(t-\tau) = \frac{1}{\langle N \rangle} \int_0^T \mathrm{d}t \, r(t) \, s(t-\tau). \tag{7}$$

The cross-correlation c_{rs} between the stimulus and the neural response is given by

$$c_{rs}(\tau) = \frac{1}{T} \int_0^T \mathrm{d}t \, r(t) \, s(t+\tau) \tag{8}$$

Comparing Eq. (7) and Eq. (8) yields

$$h(\tau) = \frac{1}{\langle r \rangle} c_{rs}(-\tau). \tag{9}$$

Because the argument of the cross-correlation function is $-\tau$ the STA is often called reverse correlation method and Eq. (9) is called reverse correlation function [DA05].
The stimulus has to be uncorrelated at every time instant and must span the whole stimulus space to which the neuron responds. Otherwise, the obtained receptive field does not give a proper description of the neuron's response to a novel stimulus. Usually, natural stimuli have strong autocorrelations.

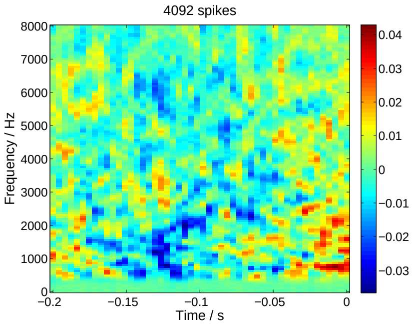
3.1.2 Algorithm

The description above explains the theoretical background of the STA. However, the actual STRF estimation is carried out using a modified version of Eq. (5) because we are interested in the time-frequency representation of the averaged stimulus parts to predict the real-valued firing rate. Hence, we use the envelope $S(t, f)$ of the discretized stimulus in the time-frequency domain and Eq. (5) becomes

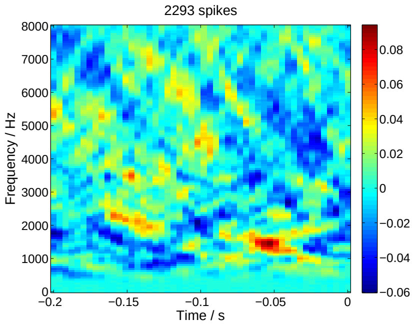
$$H(\tau, f) = \left\langle \frac{1}{N} \sum_{i=1}^{N} S(t_i - \tau, f) \right\rangle \approx \frac{1}{\langle N \rangle} \left\langle \sum_{i=1}^{N} S(t_i - \tau, f) \right\rangle. \tag{10}$$

It is easy to see that the energy density distribution of the STA depends on the energy density distribution of the stimulus parts used for the calculation. To compensate for this, we substract the mean stimulus energy in every frequency band from the spectrogram parts used for STRF estimation.

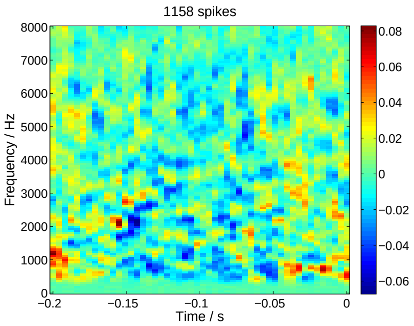
3.2 Examples



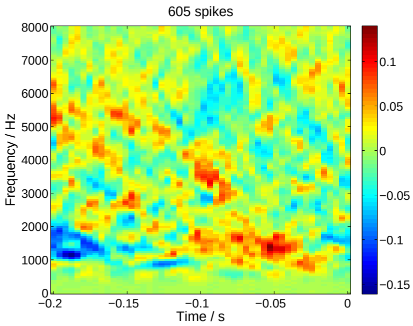
(a) DMR, Unit 1



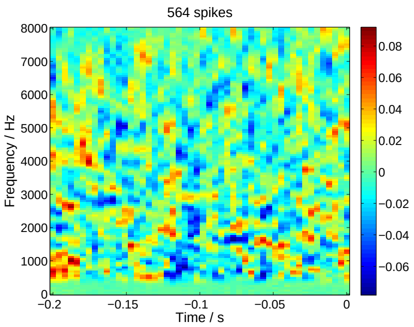
(b) FM banks, Unit 1



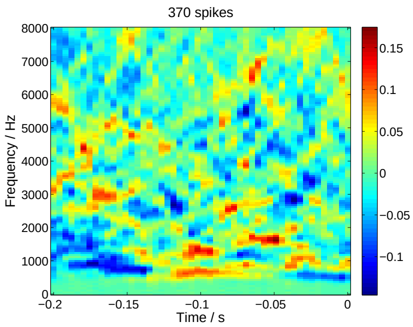
(c) DMR, Unit 2



(d) FM banks, Unit 2

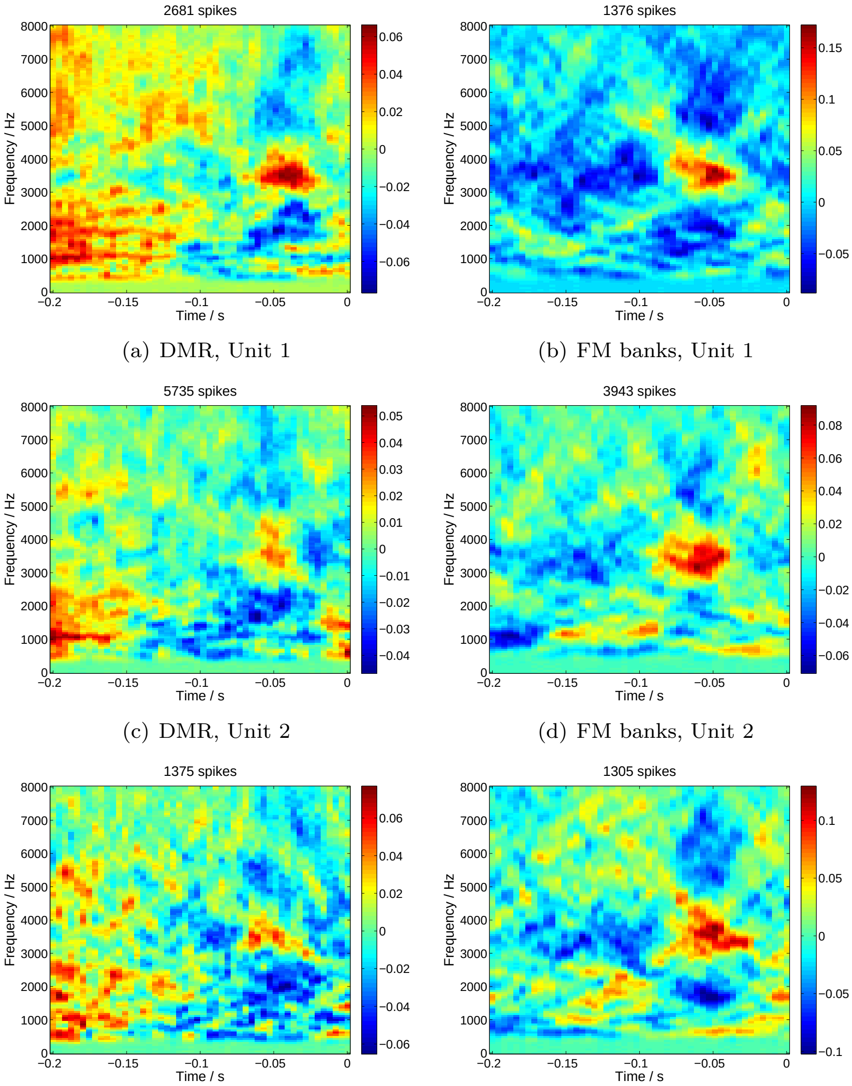


(e) DMR, Unit 3



(f) FM banks, Unit 3

Figure 5: STRFs estimated using DMR and FM-bank stimuli. The left column shows DMR-based STRF and the right column shows FM bank-based STRFs where each row corresponds to one unit measured at the same electrode. The number above the pictures indicates the number of spikes used for STRF estimation.



(a) DMR, Unit 1 (b) FM banks, Unit 1

(c) DMR, Unit 2 (d) FM banks, Unit 2

(e) DMR, Unit 3 (f) FM banks, Unit3

Figure 6: STRFs estimated using DMR and FM-bank stimuli. The left column shows DMR-based STRF and the right column shows FM bank-based STRFs where each row corresponds to one unit measured at the same electrode. The number above the pictures indicates the number of spikes used for STRF estimation.

4 Downloads

- Neural datasets
- Matlab code

References

[DA05] Peter Dayan and Larry F. Abbott. *Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems*. The MIT Press, 2005.

[dK68] Egbert deBoer and Paul Kuyper. Triggered correlation. *IEEE Transactions on Biomedical Engineering*, BM15(3):169–179, 1968.

[ES02] Monty A. Escabi and Christoph E. Schreiner. Nonlinear spectrotemporal sound analysis by neurons in the auditory midbrain. *J Neurosci*, 22(10):4114–4131, May 2002.

[KDSS00] David J. Klein, Didier A. Depireux, Jonathan Z. Simon, and Shihab A. Shamma. Robust spectrotemporal reverse correlation for the auditory system: Optimizing stimulus design. *Journal of Computational Neuroscience*, 9(1):85–111, 2000.

[TSD00] Frederic E. Theunissen, Kamal Sen, and Allison J. Doupe. Spectral-temporal receptive fields of nonlinear auditory neurons obtained using natural sounds. *J Neurosci*, 20(6):2315–2331, Mar 2000.