NOLTA 2004 – Fukuoka, Japan

Spike train clustering using a Lempel-Ziv-distance measure

Markus Christen, Thomas Ott, Ruedi Stoop
Institute of Neuroinformatics, Uni/ETH Zurich

Collaborators:

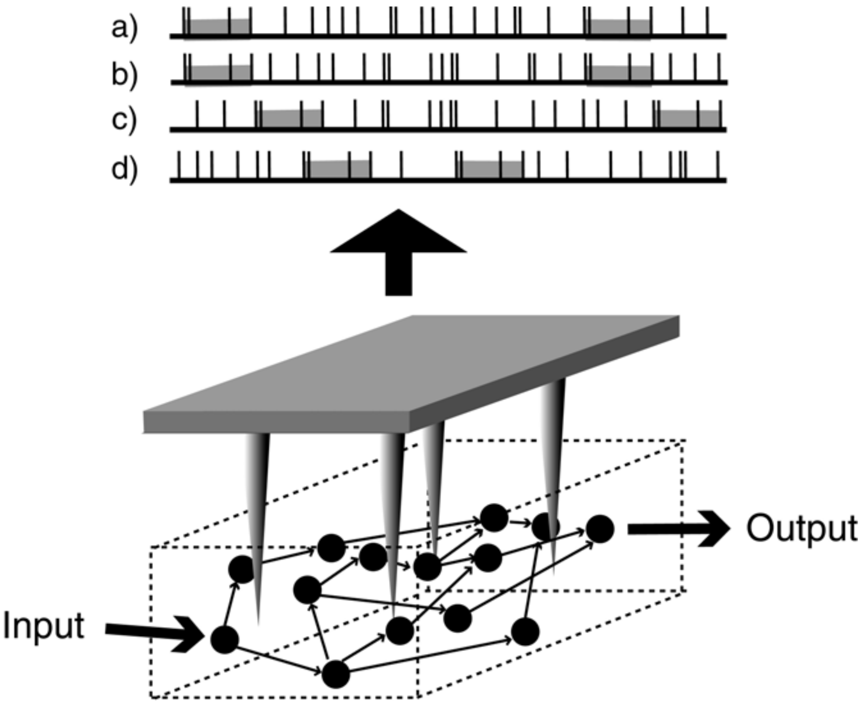
Adam Kohn, New York University
Alister Nicol, Babraham Institute Cambridge

Contents:

Multi-electrode array recordings:

Multi-electrode array recordings allow to measure many neurons simultaneously.

How to group neurons with similar temporal structure irrespective of the way the patterns are distributed?



Applications of distance measures:

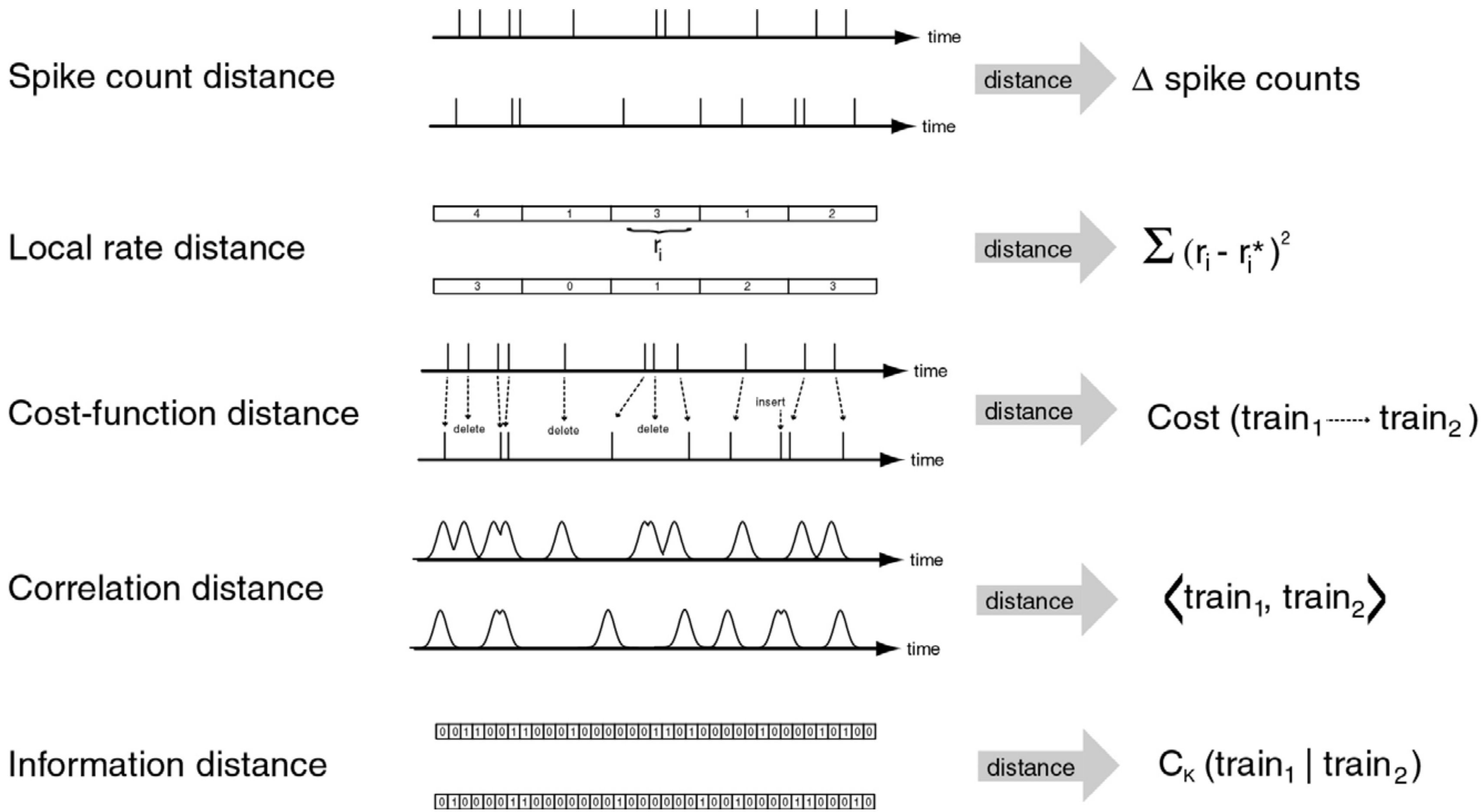
Reliability-discussion: A measure for estimating the variability of neuronal firing:
$\longrightarrow$ Distance of trains of different trials.

Spike-pattern recognition: An instrument to evaluate null-hypotheses in shuffling techniques:
$\longrightarrow$ Distance original train vs. shuffled train.

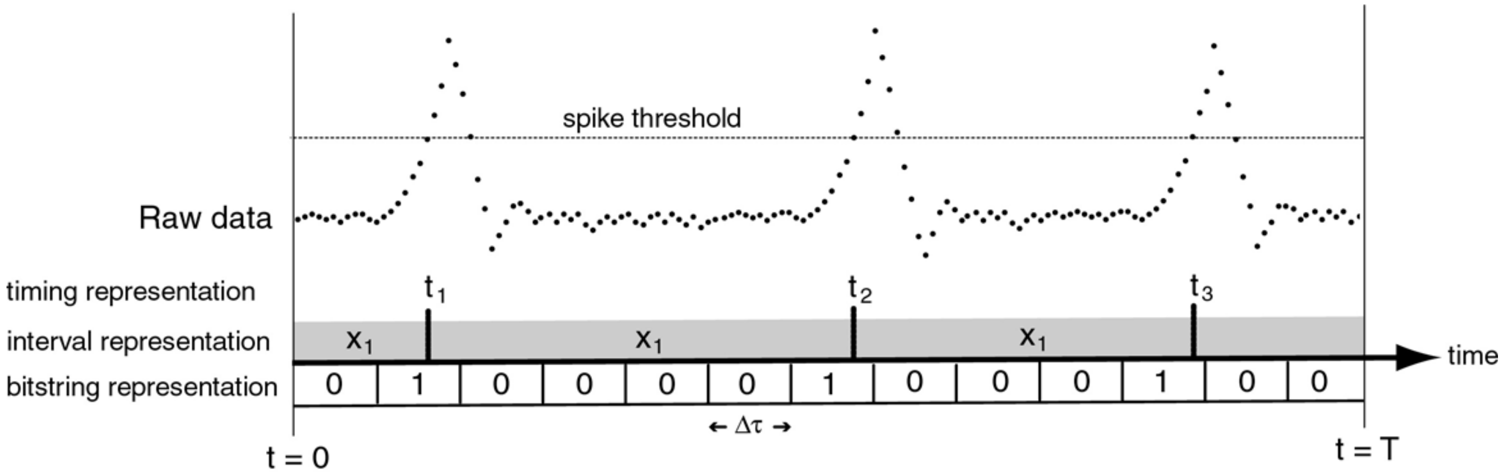
Neuronal modelling: A tool for evaluating the accuracy of neuronal models:
$\longrightarrow$ Distance of measured train vs. model train.

Multi-train analysis: A prerequisite for spike train clustering:
$\longrightarrow$ Distance matrix.

Standard distance measures:



LZ parsing procedures:



Bitstring: $X_n = (x_1, ., x_i, x_{i+1}, ... x_j .. x_n)$ $x_i \in \{0,1\}$

Phrase $X_n(i,j) = (x_i, ... x_j)$ P_{Xn} : set of phrases

Vocabulary $V_{Xn} = \{\text{all substrings of } X_n\}$

Let $X_n(1,i) = (x_1, ... x_i)$ be parsed. New phrase $X_n(i+1,j)$?

LZ-76 parsing: first j such that $X_n(i+1,j) \notin V_{X(1,i)}$

LZ-78 parsing: first j such that $X_n(i+1,j) \notin P_{X(1,i)}$

LZ complexity:

Information theory: Distinct parsings of bitstrings X_n, which are the result of stationary, ergodic processes with entropy rate *H* have the property of asymptotic optimality:

$$K(X_n) = \limsup_{n\to\infty} c(X_n) * \log c(X_n) / n \to H$$

where $c(X_n)$ is the size of P_{Xn}.

$K(X_n)$ is the Lempel-Zif-complexity of the string X_n.

$K(X_n)$ can be used to calculate the entropy rate of a spike train (problem: stationarity!).

LZ distance:

Basic idea: Let X, Y be two bitstrings of equal length. The distance should tell us, to what extend the parsing of Y helps to parse X. We define:

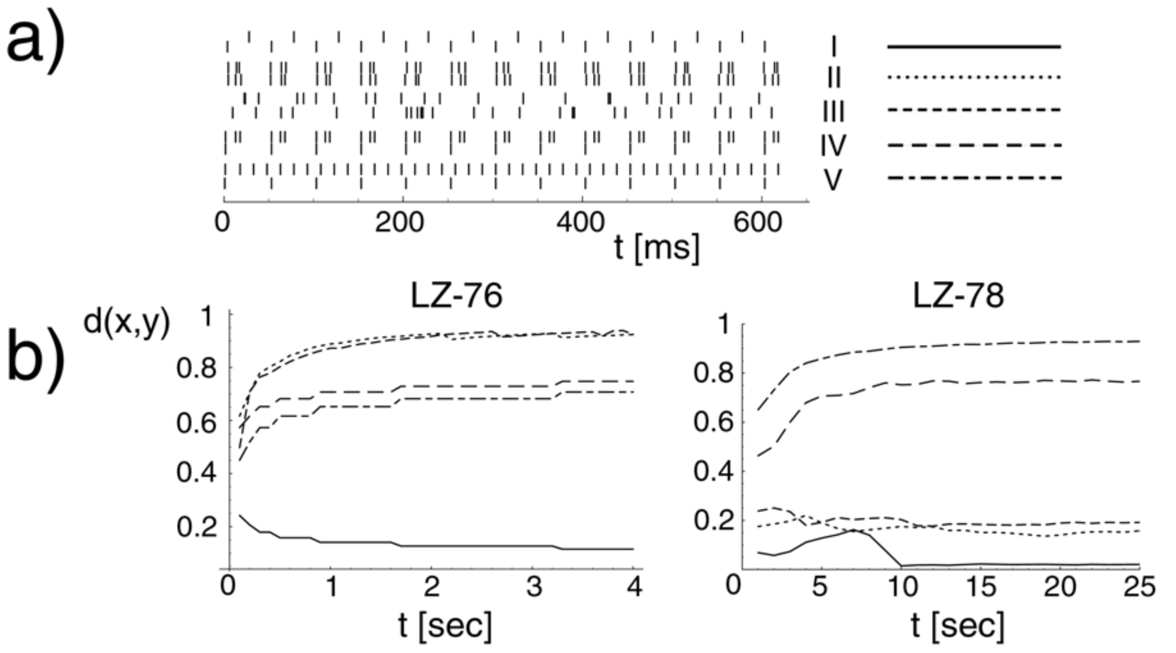
$$d(X,Y) = 1 - \min\left\{\frac{K(X) - K(X|Y)}{K(X)}, \frac{K(Y) - K(Y|X)}{K(Y)}\right\}$$

To calculate K(X|Y), the size of the difference set $P_X \setminus P_Y$ (= c(X|Y)) is used. d(X,Y) fulfills the distance axioms.

If Y provides no information about X, then $P_X \setminus P_Y = P_X$ and the distance is 1.

If Y provides complete information about X, then $P_X \setminus P_Y = \emptyset$ and the distance is 0.

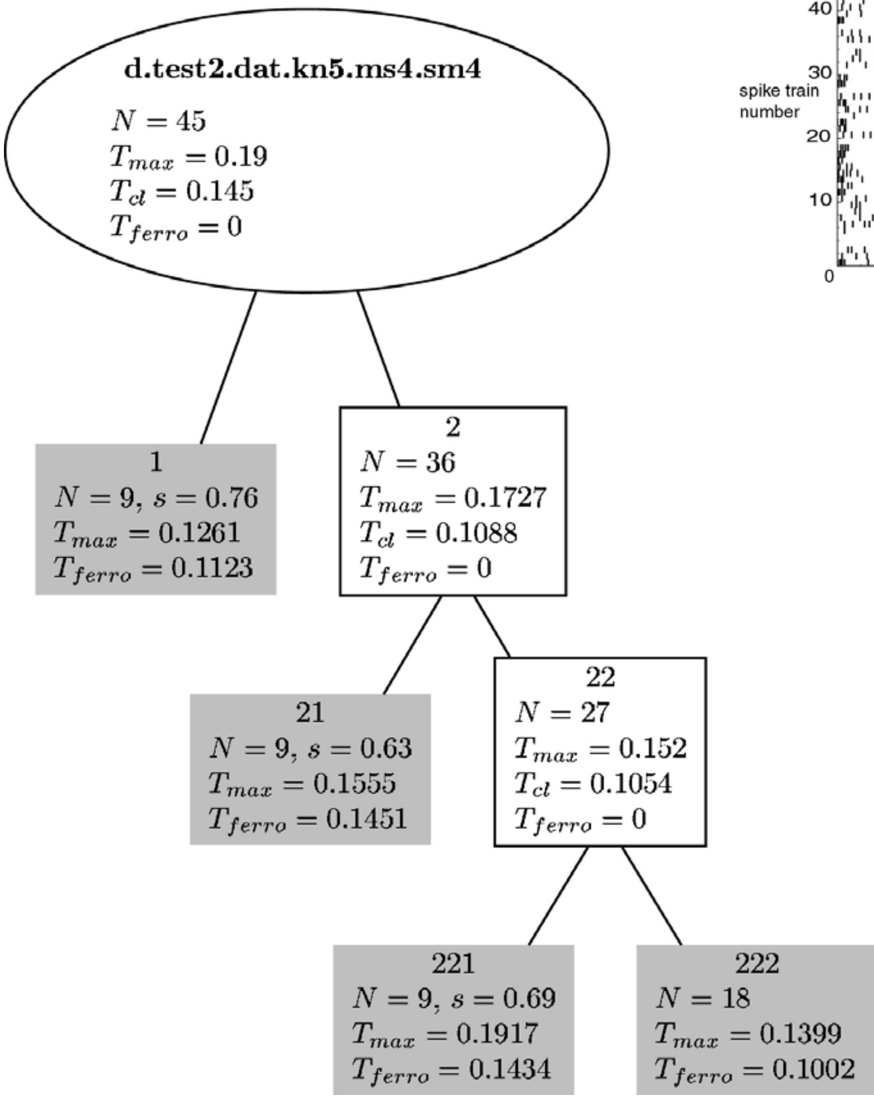
Choosing the parsing:

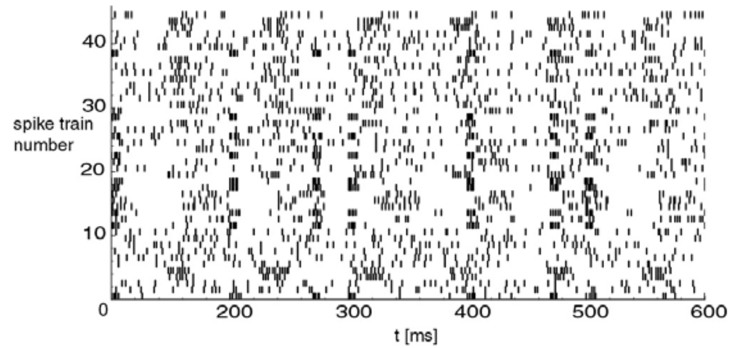


LZ-76 parsing converges faster.

LZ-78 parsing is more noise-robust and computationally much cheaper $\Longrightarrow$ we use LZ-78 parsing.

Clustering:





5 Testtrains (all with similar firing rate):

I: Homogenous Poisson with abs. & rel. refractory period.
II: Inhomogenous Poisson driven like V1 simple cell.
III: Noisy burst pattern train.
IV: Complex cell V1.
V: Simple cell V1, sinusoidally driven.

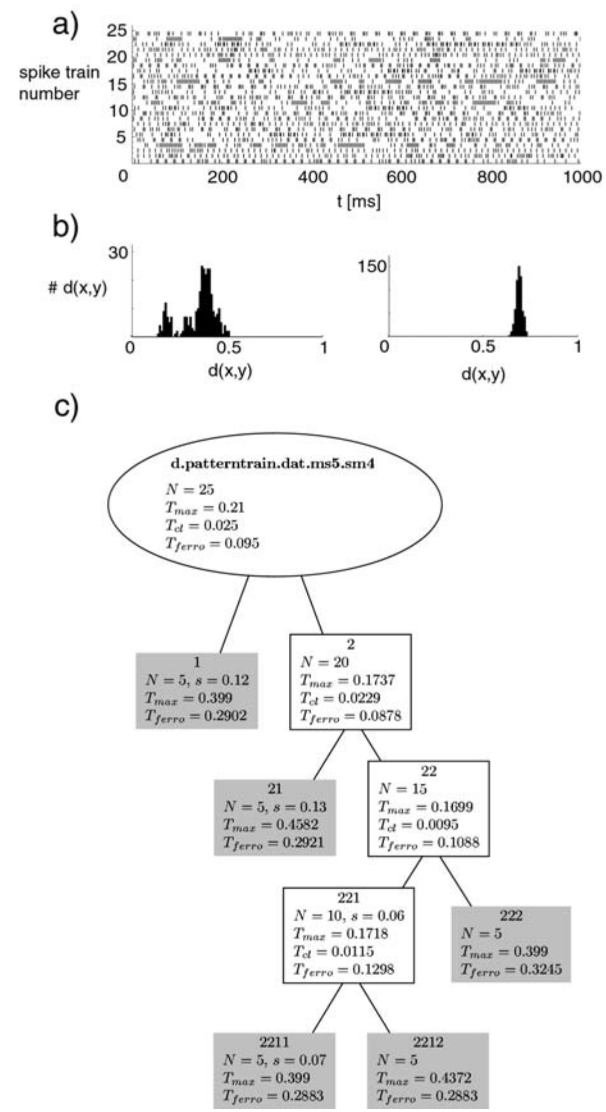
All tested distances lead to comparable results.

Distance measures allow to assess the accurracy of models.

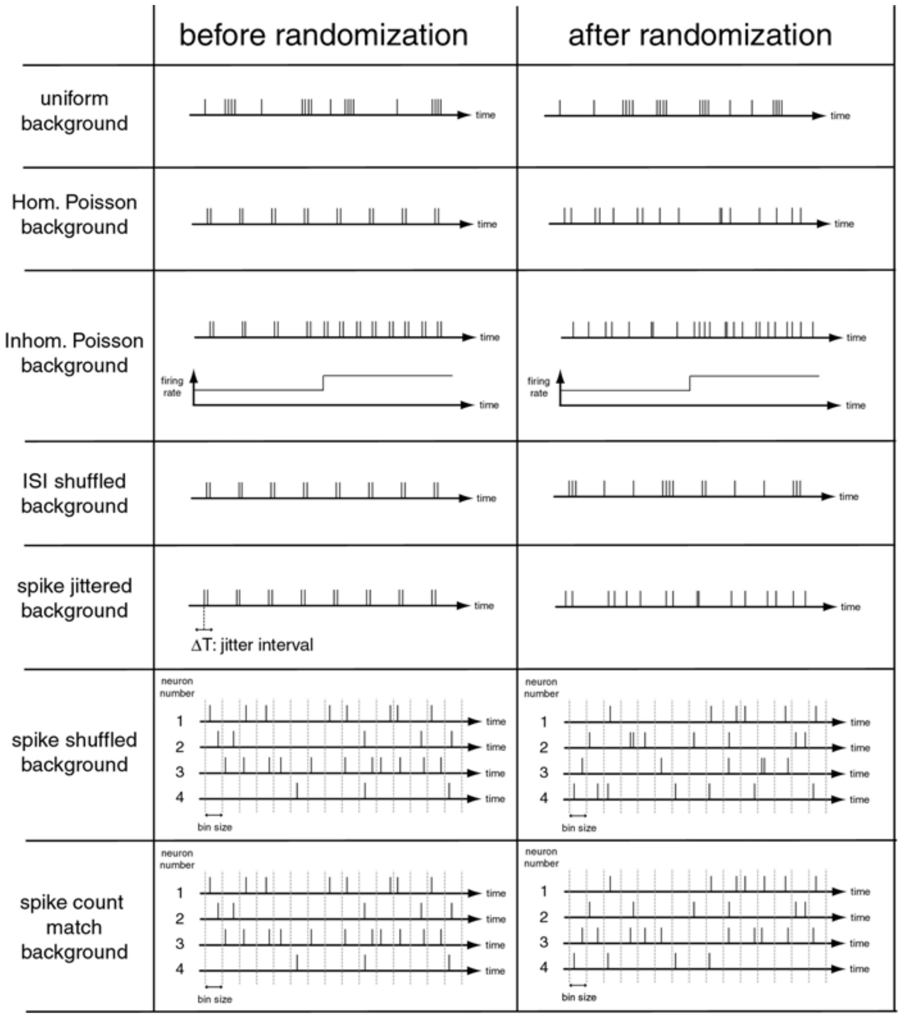
Comparison with other distances:

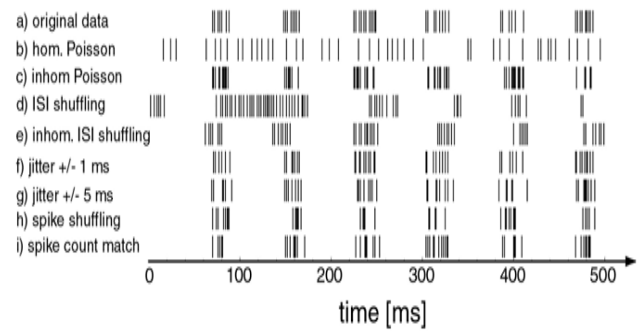
Fife different spike patterns embedded in a random (Poisson) background such that the firing rates of the spike trains are similar.

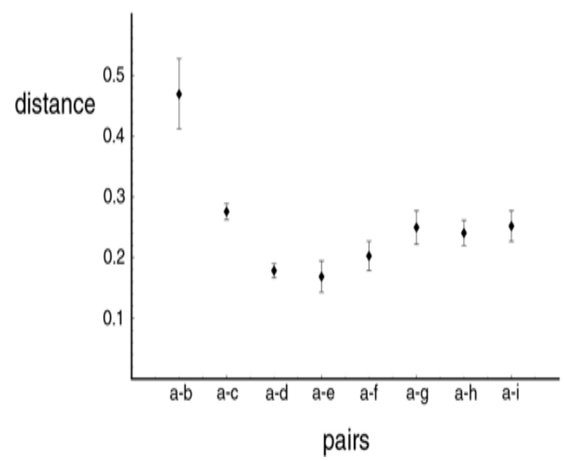
LZ-distance allows classification, correlation-based distances do not (histogram: tescaling would not help).



Example 1:


before randomization
after randomization
uniform background
Hom. Poisson background
Inhom. Poisson background
ISI shuffled background
spike jittered background
spike shuffled background
spike count match background
time
firing rate
ΔT: jitter interval
neuron number
1
2
3
4
bin size

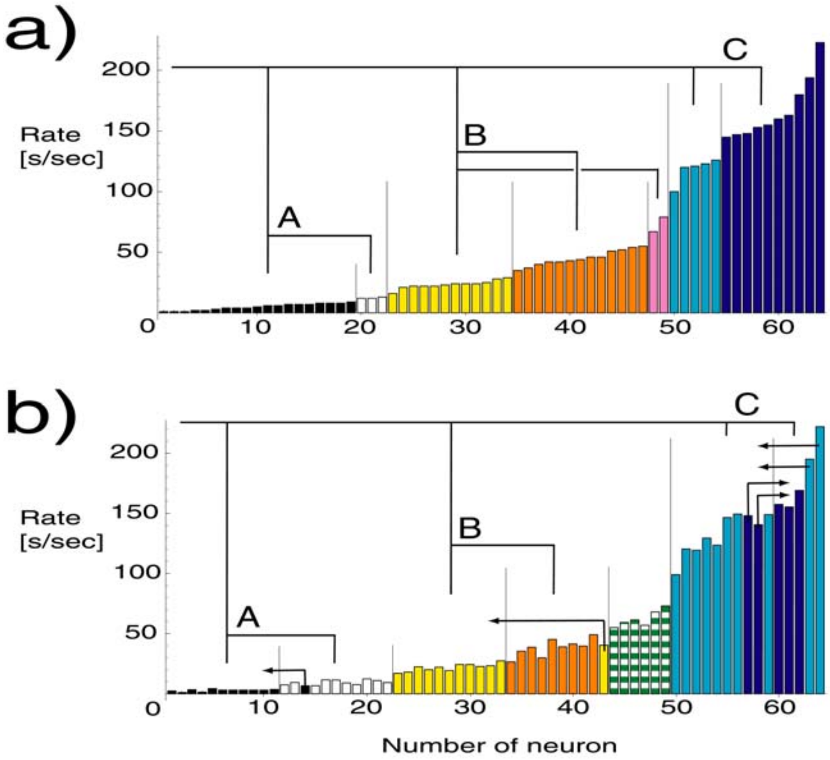

a) original data
b) hom. Poisson
c) inhom Poisson
d) ISI shuffling
e) inhom. ISI shuffling
f) jitter +/- 1 ms
g) jitter +/- 5 ms
h) spike shuffling
i) spike count match
0
100
200
300
400
500
time [ms]


distance
0.5
0.4
0.3
0.2
0.1
a-b
a-c
a-d
a-e
a-f
a-g
a-h
a-i
pairs

Example 2:

64 neurons in the olfactory bulb of the rat, before (a) and during (b) stimulus presentation:

Identification of „interesting neurons“.



Conclusions:

LZ distance has the following properties:

- It needs a minimum number of parameters (bin-size for bitstring-generation).
- It is easy to implement and computationally cheap.
- It fullfills the mathematical requirements for a distance measure.
- It performs at least as good as other distance measures in standard spike train clustering problems.
- It allows to cluster spike trains with siminar, but randomly distributed patterns.