CROSS-RECURRENCE ANALYSIS OF SPEECH SIGNALS

Leonardo Lancia, Susanne Fuchs
leonardo.lancia@lpl-aix.fr
Centre for General Linguistics (ZAS), Berlin

1. Introduction

The temporal evolution of the speech signal's features plays an important role both in perception and production [1,2].

BUT: Measurements are often conducted at critical time-points and averaged over extended data sets.

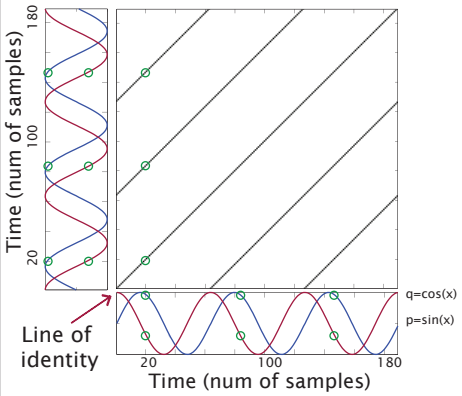
⟶ Static characterization of the observed data obscures several footprints of the underlying dynamical processes [3].

Recurrence analysis

Analysis of the repetitions occurring in time series [4,5].

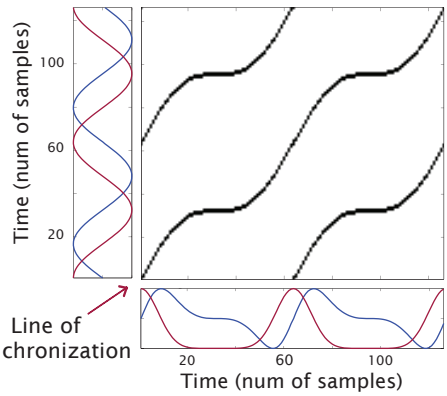
⟶ New method to measure degree of similarity of speech signals by taking into account their temporal evolution.

Recurrence plot (RP)
of a time-series shows the points in time at which a former value is repeated.

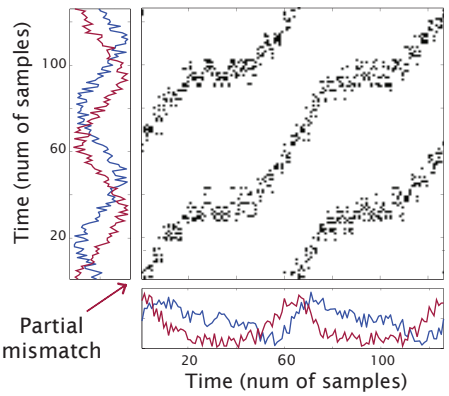


X-axis: which pair is repeated;
Y-axis: when a given pair is repeated.

Cross-recurrence plot (CRP)
of two time-series shows the points in time at which a value of one time series occurs in the other one



Different time-scales, same amplitude: Bows on continuous lines

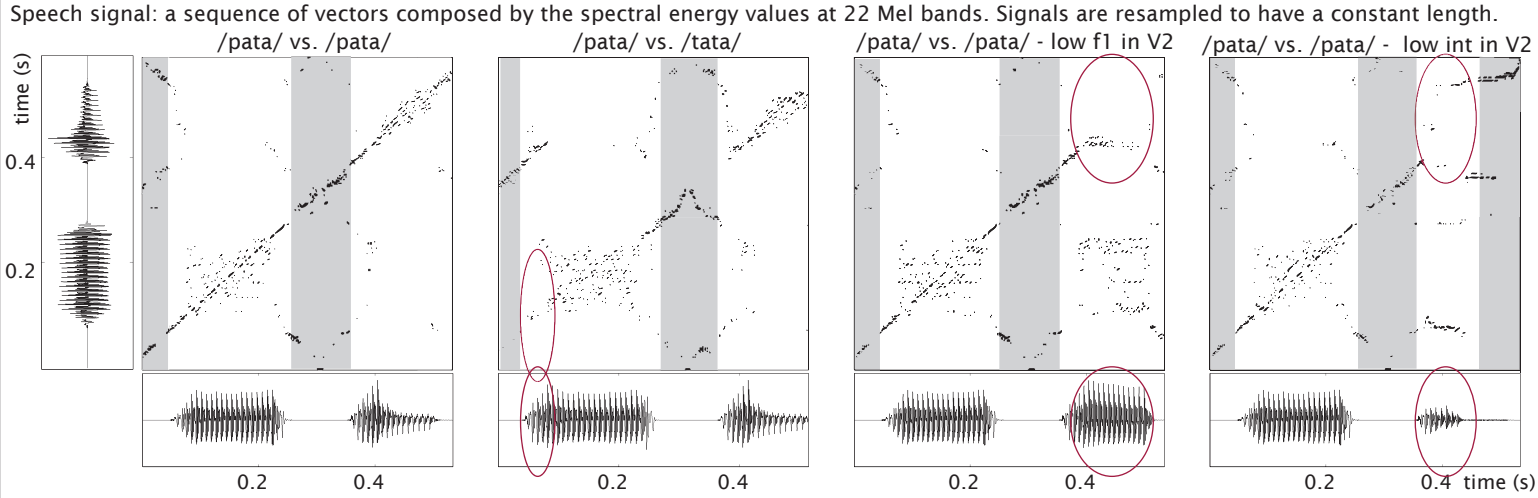


Noise: scattered black dots

2. Synthetic speech

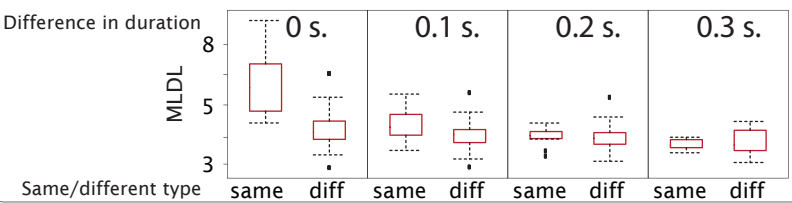
● Comparison of CVCV utterances produced trough pseudo articulatory synthesis (HLsyn [6]) .

Speech signal: a sequence of vectors composed by the spectral energy values at 22 Mel bands. Signals are resampled to have a constant length.



In presence of mismatches, recurrence points loose their diagonal structure. However, silent regions (gray bands on the RP) can be misleading.

● Is the mean length of the diagonal lines (MLDL) reliable?



256 comparisons obtained by combining 16 acoustic signals (the 4 types above and 4 different durations per type).

Linear mixed model results:

diffType:	pMCMC<0.001
diffDur:	pMCMC<0.001
diffDur*diffType:	pMCMC<0.001

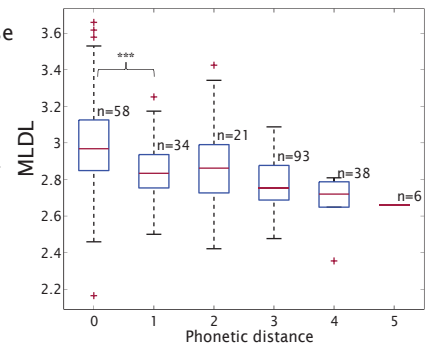
3. Real speech : Is MLDL consistent with the manual labeling of speech signals?

Task: Repeat the target utterance during 20 secs (repeated speech experiment [7,8].) while keeping pace with a metronome, whose speed first increasses then decreases.

Target utterances: /tapa/ and /pata/.

Methods:

- Produced utterances hand segmented and labeled with respect to the observed deviation from the target utterance.
- Labels = segment elision, creaky vowel, breathy vowel, devoiced vowel, raised vowel (F1<350Hz), plosive to affricate, plosive to fricative, plosive to approximant, plosive or fricative voicing.
- Labeling score (for each pair of consecutive utterances) = counting the number of deviant features which are present only in one utterance of the pair.
- Silent portions of speech signals removed trough a VAD algorithm [9].



n: Number of observation per level
*** pMCMC < 0.001; * pMCMC < 0.05

Statistics : Linear mixed model

- Fixed: Phonetic distance (coded with successive difference contrast [10]); duration of the two utterances;
- Random: combination of speaker and target utterance.

4. Summary discussion

The mean length of the diagonal lines in a CRP constitutes a reliable index of acoustic similarity among speech signals:

- Sensitive to changes in the evolution over time of speech signals.
- Effective on the basis of small data sets.
- Capable to deal with non-linearity and non-stationarity.

Further work:

- Improvement of the method trough wavelet scalogram representation of speech signals.
- Application to articulatory data and comparison with acoustic data.

Possible applications:

Imitation and inter-speaker coordination; gestural synchronization; speech errors.