Dealing with nonstationarities in biomedical signals

Patrick Flandrin

CNRS & École Normale Supérieure de Lyon
Physics Department, SiSyPh Group, Lyon, France

ENS DE LYON

EMBC-10 (Buenos Aires), Aug. 31 – Sept. 4, 2010

Why being interested in nonstationarity?

Common concepts attached to **nonstationarity**:

- changes
- evolutions
- modifications
- disturbances

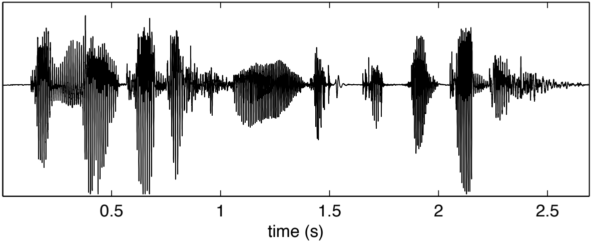
with respect to **many different** situations:

- signals (mean, variance, spectrum, . . .)
- systems (models, . . .)
- measurements (experimental conditions, baseline, . . .)

Nonstationarity is the **rule**, not the exception ("*Stationarity is a fairy tale for graduate students*" (D.J. Thomson, *Proc. IEEE ICASSP-94))*

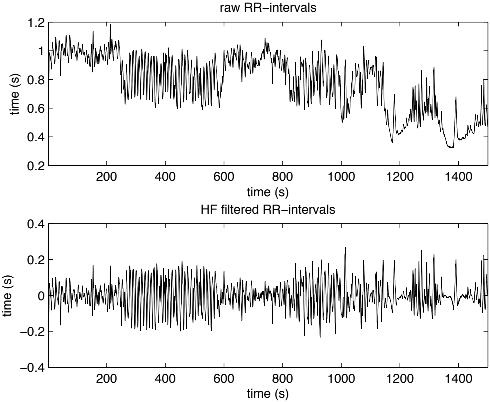
Why being interested in nonstationarity?

Speech as a typical example



Why being interested in nonstationarity?

Trend and fluctuations



(courtesy of M. Orini)

How to deal with nonstationarities?

Observation

Two *key concepts attached to stationarity:*

1. *(Fourier)* ***frequency***
2. ***time-invariance***

Principle

Wedding *time and frequency (at large)*

- ***Time-Frequency*** *(TF) and* ***Time-Scale*** *(TS) methods*
- ***time-dependent*** *models and/or algorithms (****adaptivity****)*

TF/TS: what's new?

1. **30 years** of theory and applications
   - energy distributions
   - wavelets
   - time-varying models
2. **New advances** and challenges
   - TF resolution and estimation trade-offs (reassignment/synchrosqueezing + multitapering)
   - data-driven decompositions (Empirical Mode Decomposition)
   - bivariate signals (TF coherence and EMD)
   - decision (stationarity tests + TF machines)
   - scaling beyond self-similarity (multifractals)

Today's agenda

1. **30 years** of theory and applications
   - **energy distributions**
   - wavelets
   - time-varying models
2. **New advances** and challenges
   - TF resolution and estimation trade-offs (**reassignment**/synchrosqueezing + **multitapering**)
   - data-driven decompositions (**Empirical Mode Decomposition**)
   - bivariate signals (TF coherence and EMD)
   - decision (**stationarity tests** + TF machines)
   - scaling beyond self-similarity (multifractals) → P. Abry's talk

Quadratic TF distributions

Observation

*Many quadratic TF distributions have been proposed in the literature since more than half a century (e.g., spectrogram and Wigner-Ville): **none fully extends the notion of spectrum density to the nonstationary case**.*

Principle of conditional unicity — Classes of quadratic distributions of the form $\rho_x(t,f) = \langle x, \mathbf{K}_{t,f} x \rangle$ can be constructed based on **covariance requirements** :

$$\begin{array}{ccc} x(t) & \rightarrow & \rho_x(t,f) \\ \downarrow & & \downarrow \\ (\mathbf{T}x)(t) & \rightarrow & \rho_{\mathbf{T}x}(t,f) = (\tilde{\mathbf{T}}\rho_x)(t,f) \end{array}$$

Classes of quadratic TFDs

- **Cohen's class** — Covariance w.r.t. TF **shifts** $(\mathbf{T}_{t_0,f_0}x)(t) = x(t-t_0)\exp\{i2\pi f_0 t\}$ leads to **Cohen's class** (Cohen, '66) :

$$C_x(t,f) := \iint W_x(s,\xi)\,\Pi(s-t,\xi-f)\,ds\,d\xi,$$

  with $W_x(t,f)$ the Wigner-Ville distribution (WVD) and $\Pi(t,f)$ "arbitrary" (to be specified via additional constraints).
- **Variations** — Other choices possibles, e.g., $(\mathbf{T}_{t_0,f_0}x)(t) = (f/f_0)^{1/2}x(f(t-t_0)/f_0) \rightarrow$ **affine class** (Rioul & F, '92), etc.

An example of definition

Spectrogram — If we consider the case of the **spectrogram** with window $h(t)$, one can write:

$$\begin{aligned} S_x^{(h)}(t,f) &= \left| \int x(s)\, \overline{h(s-t)}\, e^{-i2\pi fs}\, ds \right|^2 \\ &= \left| \langle x, \mathbf{T}_{t,f} h \rangle \right|^2 \\ &= \iint W_x(s,\xi)\, W_{\mathbf{T}_{t,f}h}(s,\xi)\, ds\, d\xi \\ &= \iint W_x(s,\xi)\, W_h(s-t,\xi-f)\, ds\, d\xi \end{aligned}$$

$\Rightarrow$ a spectrogram is a member of Cohen's class, with kernel

$$\Pi(t,f) = W_h(t,f)$$

Cohen's class and smoothing

- **From Wigner-Ville to spectrograms** — A generalization amounts to choose a smoothing function $\Pi(t, f)$ allowing for a **continuous** and **separable** transition between Wigner-Ville and a spectrogram (**smoothed pseudo-Wigner-Ville** distributions) :

$$\begin{array}{ccccccc} \mathit{WVD} & \ldots & \rightarrow & \mathit{SPWVD} & \ldots & \rightarrow & \mathit{spectrogram} \\ & & & & & & \\ \delta(t)\,\delta(f) & & & g(t)\,H(f) & & & W_h(t, f) \end{array}$$

- **Successful uses** in biomedical applications, in particular **HRV** and the quantification of LF vs. HF components during nonstationary events (see Mainardi, '09)

SPWVD quantification of HRV

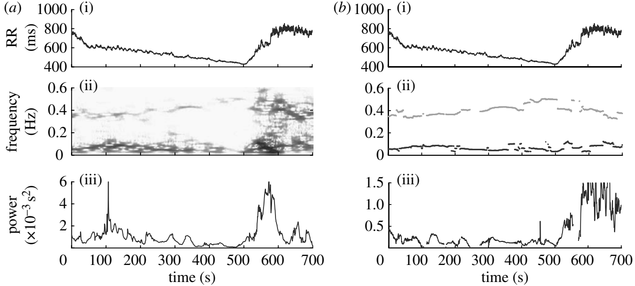


Figure 3. (*a*) (i) RR interval series obtained for a normal volunteer during stress testing, (ii) the TFR obtained by SPWV and (iii) trend of LF power. (*b*) (i) RR series, (ii) tracking of LF and HF frequencies and (iii) trend of HF power. Note that the decrease in LF power during stress (related to reduction in the overall HRV during effort) is accompanied by the persistence of the HF component even at a higher workload. The RR stress data were collected during the study by Bailón *et al.* (2006*b*) and kindly provided by the authors.

(from Mainardi, *Phil. Trans. R. Soc. A*, '09)

TF trade-offs

Back to Cohen's class — Given the smoothing relation

$$C_x(t,f) = \iint W_x(s,\xi)\,\Pi(s-t,\xi-f)\,ds\,d\xi,$$

one is faced with two types of trade-offs:

1. a **geometrical** trade-off between auto-terms localization and cross-terms interference for deterministic signals (C_x as a TFD)
2. a **statistical** trade-off between bias and variance for stochastic signals (C_x as an estimator of the WV spectrum defined as $\mathbb{E}\{W_x(t,f)\}$)

Geometrical trade-off: TF reassignment

Idea

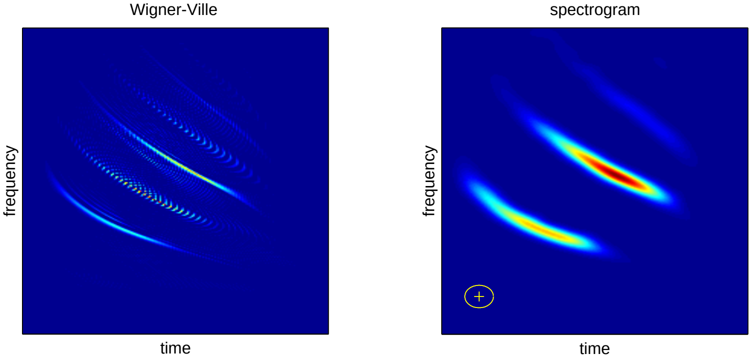
*Replace the **geometrical** center of the smoothing TF domain by the **center of mass** of the WVD over this domain, and **reassign** the value of the smoothed distribution to this local centroïd*

$$C_x(t,f) \mapsto \iint C_x(s,\xi)\, \delta\left(t - \hat{t}_x(s,\xi), f - \hat{f}_x(s,\xi)\right)\, ds\, d\xi$$

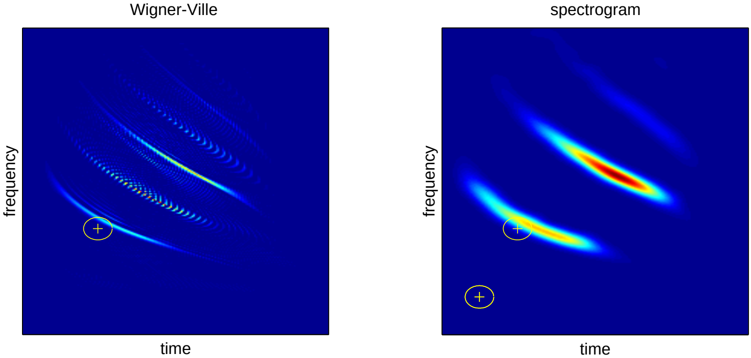
Remark

*Reassignment has been first introduced for the only spectrogram (Kodera et al., '76), but its principle has been further generalized to **any** distribution resulting from the smoothing of a localizable mother-distribution (Auger & F., '95)*

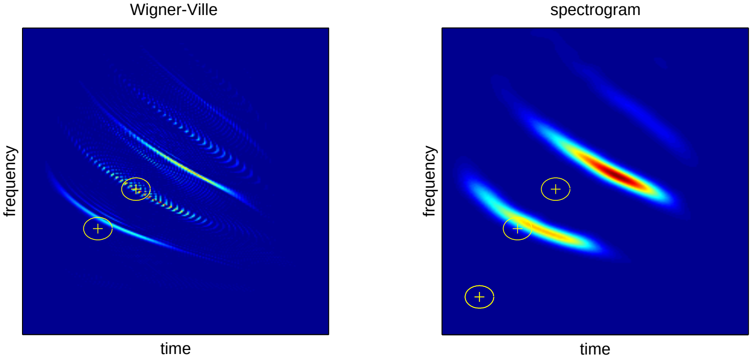
Reassignment



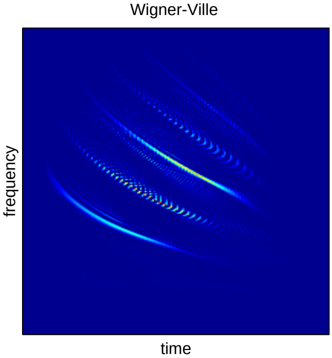
Reassignment

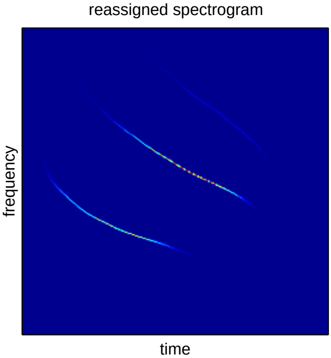


Reassignment



Reassignment





Reassignment in action

Pros

- *Efficient, **implicit** algorithms*
- *Very good properties of **localization** for chirps ($>$ spectrogram)*

Cons

*High **sensitivity to noise** ($<$ spectrogram)*

Aim

Reduce fluctuations while preserving localization

Back to spectrum estimation

- **Stationary processes** — The **power spectrum density** can be viewed as:

$$\mathbf{S}_x(f) = \lim_{T\to\infty} \mathbb{E}\left\{ \frac{1}{T} \left| \int_{-T/2}^{+T/2} x(t)\, e^{-i2\pi ft}\, dt \right|^2 \right\}$$

- **In practice** — Only one, finite duration, realization $\Rightarrow$ crude periodogram (squared FT) $=$ **non consistent estimator with large variance**

Classical way out (Welch, '67)

Principle

*Method of **averaged periodograms***

$$\hat{\mathbf{S}}_{x,K}^{(W)}(f) = \frac{1}{K} \sum_{k=1}^{K} S_x^{(h)}(t_k, f)$$

with $t_{k+1} - t_k$ of the order of the width of the window $h(t)$

Result

Bias-variance trade-off — *Given T (finite), increasing $K \Rightarrow$* ***reduces variance****, but* ***increases bias***

Multitaper solution (Thomson, '82)

Principle

Compute

$$\hat{\mathbf{S}}_{x,K}^{(T)}(f) = \frac{1}{K} \sum_{k=1}^{K} S_x^{(h_k)}(0, f)$$

with $\{h_k(t), k \in \mathbb{N}\}$ *a family of orthonormal windows extending over the whole support of the observation* $\Rightarrow$ ***reduced variance, without sacrifying bias***

- **Nonstationary extension — Multitaper spectrogram**

$$\hat{\mathbf{S}}_{x,K}^{(T)}(f) \rightarrow S_{x,K}(t, f) := \frac{1}{K} \sum_{k=1}^{K} S_x^{(h_k)}(t, f),$$

with localization controlled by most spread spectrogram

Geometrical & statistical trade-offs: multitaper reassignment (Xiao & F., '06)

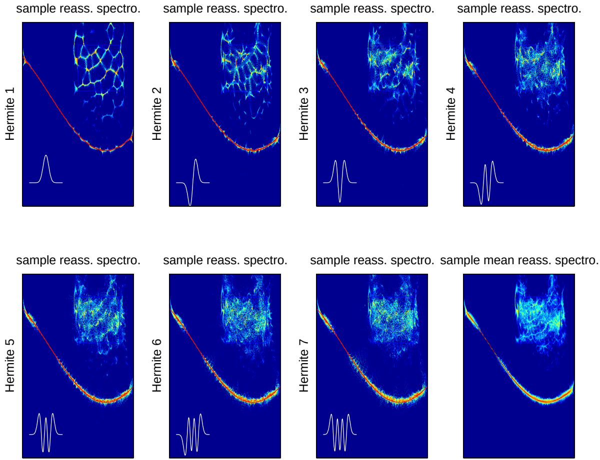
Idea

Combine the advantages of reassignment (w.r.t. localization) with those of multitapering (w.r.t. fluctuations)

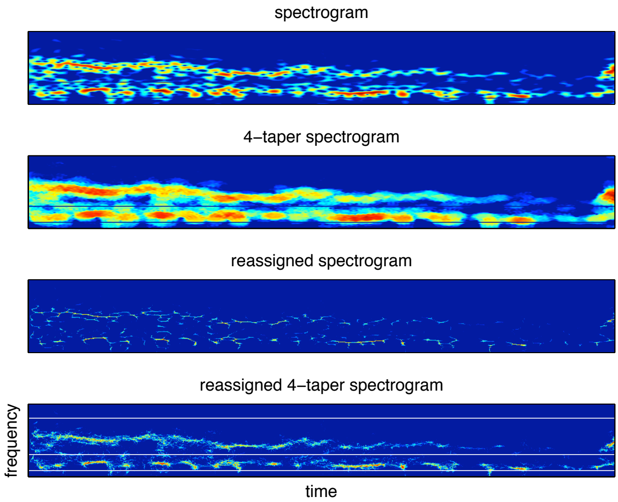
$$S_{x,K}(t,f) \rightarrow RS_{x,K}(t,f) := \frac{1}{K}\sum_{k=1}^{K} RS_x^{(h_k)}(t,f)$$

1. **coherent averaging of chirps** (localization independent of the window)
2. **incoherent averaging of noise** (different TF distributions for different windows)

Illustration



HRV example (RR intervals)



A brief summary on quadratic TFDs

Pros

- *versatile tools for **exploratory** data analysis*
- *well-established **theory**, amenable to bivariate signals: TF coherence (Orini et al., '09)*
- *allow for **decision** tasks*
  - *feature extraction: NNMF (Ghoraani & Krishnan, '09)*
  - *information measures: Rényi entropies (Baraniuk & al., '01 + Tong et al., '05)*
  - *classification: TF machines (Honeine et al., '07)*

Cons

- *no easy way back from TF to signal $\Rightarrow$ from **energy** distributions to **signal** decompositions?*

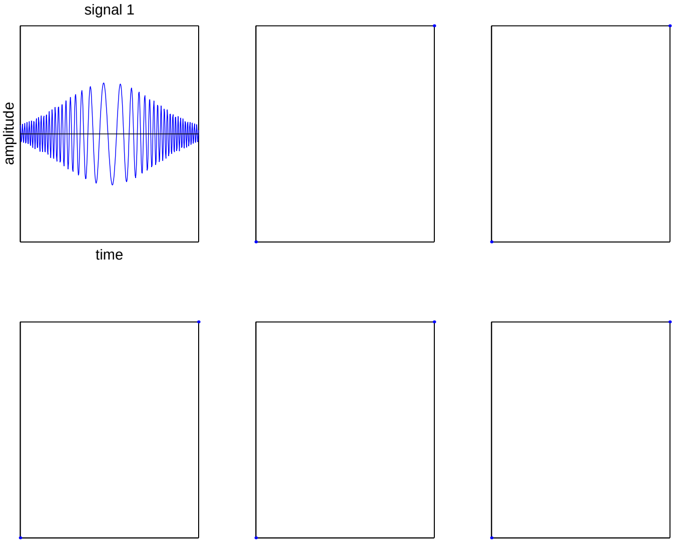
Discrete Wavelet Transform (DWT)

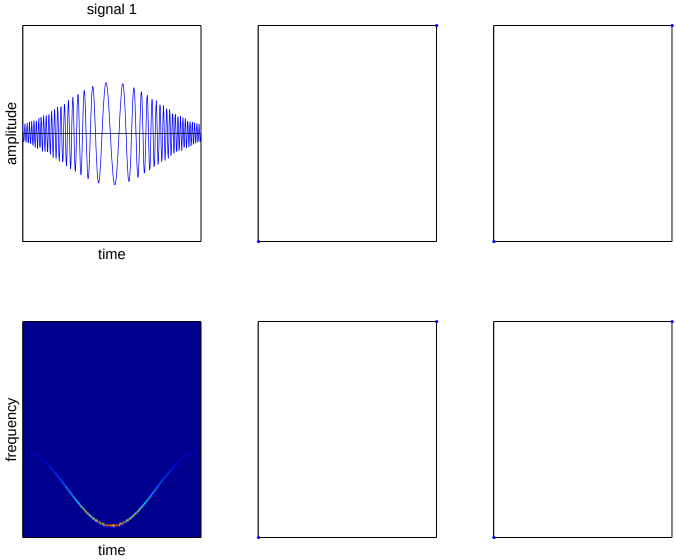
Interpretation

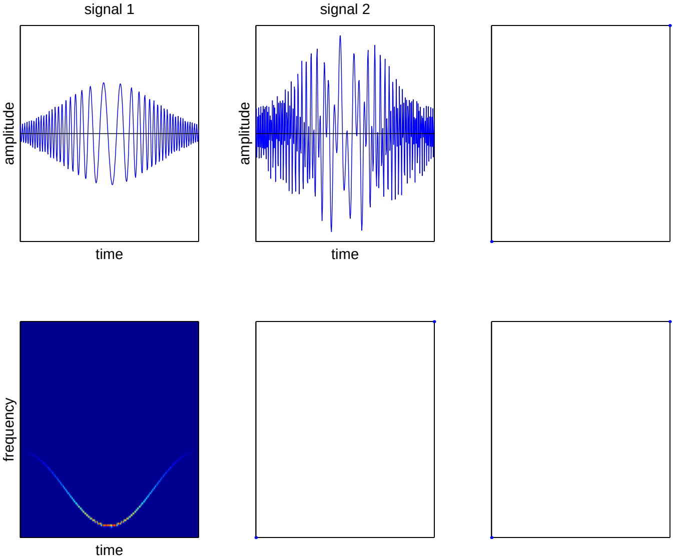
signal = approximation + detail
&
iteration

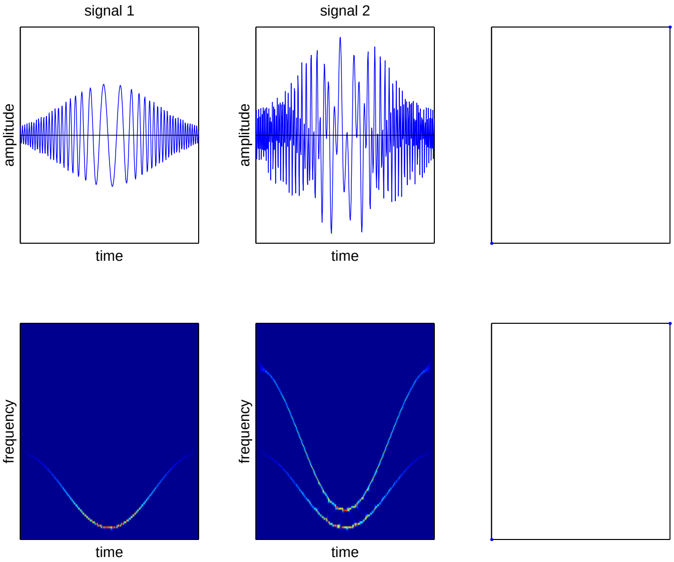
(Meyer, 85 + Mallat, ’86 + Daubechies, ’87 + . . .)

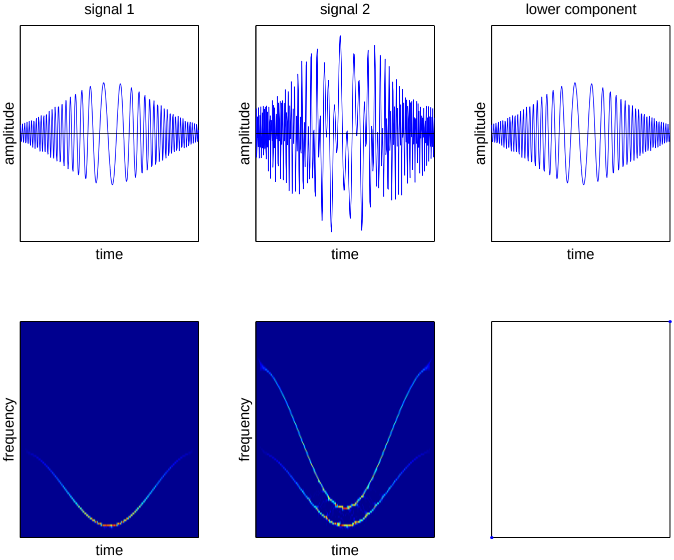
- separation “approximation vs. detail” based on **a priori** (dyadic) **filtering**
- **"global"** analysis
- other, **data-driven**, schemes?

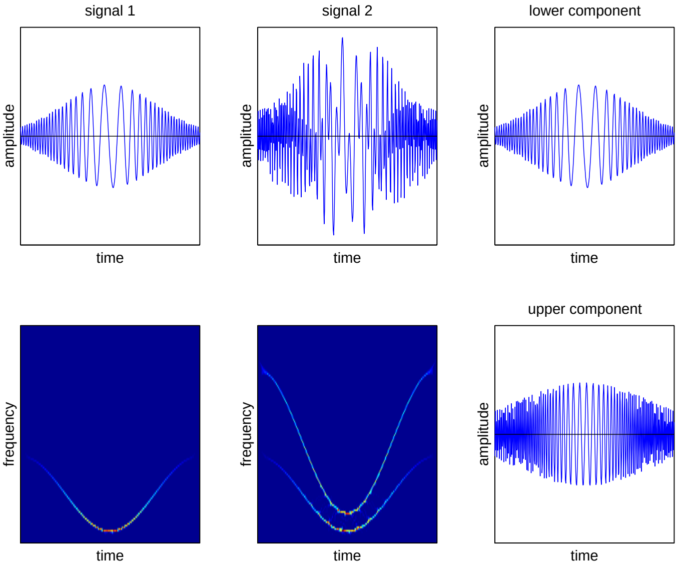

signal 1
amplitude
time

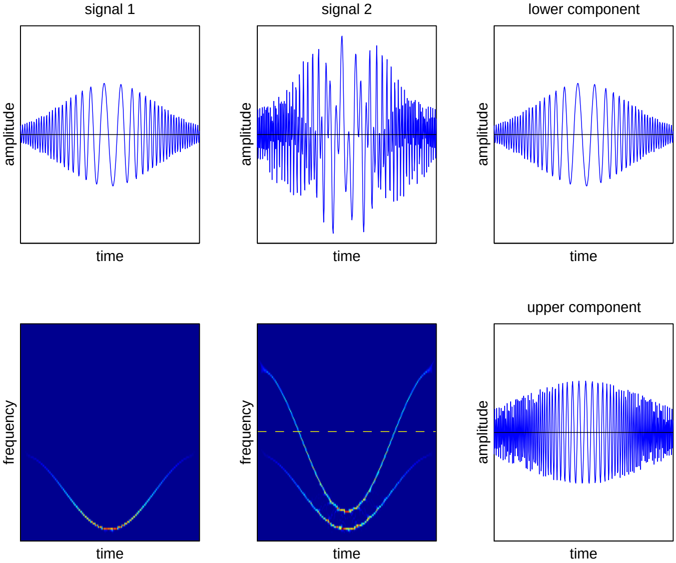

signal 1
amplitude
time
frequency
time

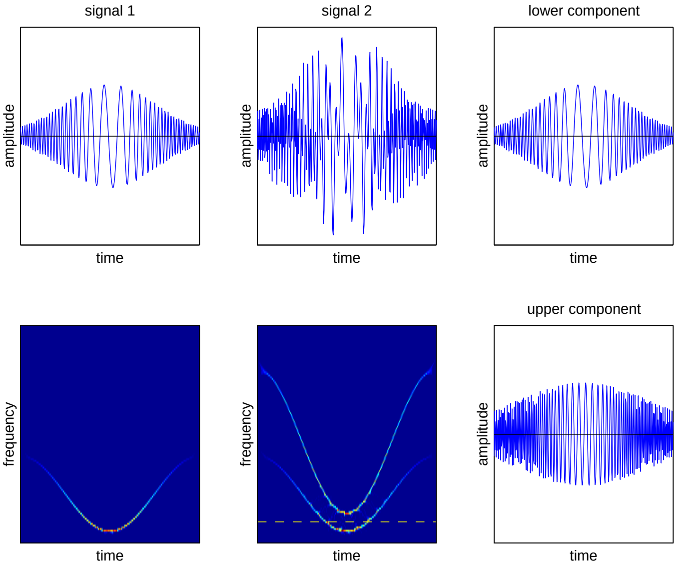

signal 1
signal 2
amplitude
amplitude
time
time
frequency
time

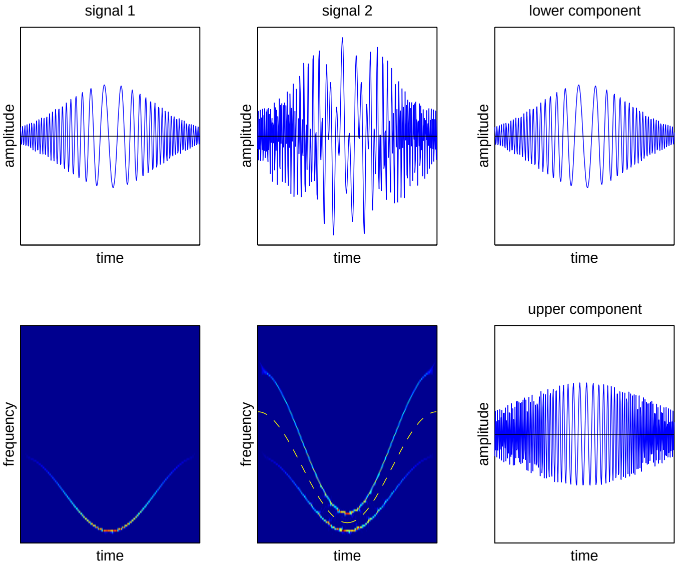

signal 1
signal 2
amplitude
amplitude
time
time
frequency
frequency
time
time


signal 1
signal 2
lower component
amplitude
time
frequency


signal 1
amplitude
time
signal 2
amplitude
time
lower component
amplitude
time
frequency
time
frequency
time
upper component
amplitude
time


signal 1
amplitude
time
signal 2
amplitude
time
lower component
amplitude
time
frequency
time
frequency
time
upper component
amplitude
time


signal 1
amplitude
time
signal 2
amplitude
time
lower component
amplitude
time
frequency
time
frequency
time
upper component
amplitude
time


signal 1
amplitude
time
signal 2
amplitude
time
lower component
amplitude
time
frequency
time
frequency
time
upper component
amplitude
time

Empirical Mode Decomposition (EMD) as an alternative

Idea

signal = fast oscillation + slow oscillation

&

iteration

(Huang *et al.*, *Proc. Roy. Soc. A., '98)*

- separation "fast vs. slow" **data driven**
- **"local"** analysis based on extrema
- still open question: which **theoretical framework**?

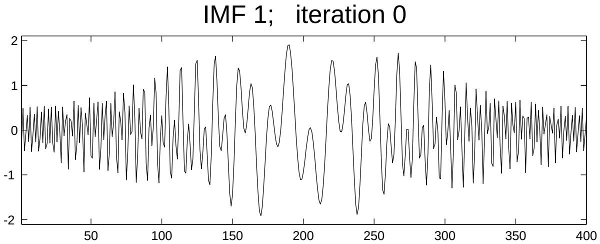
EMD algorithm

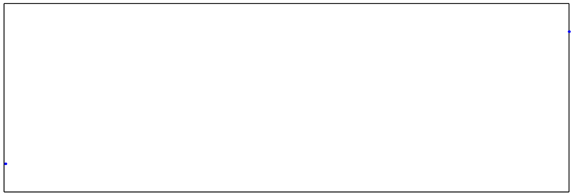
1. identify local maxima and local minima
2. deduce an upper envelope and a lower envelope by interpolation (cubic splines)
    1. subtract the mean envelope from the signal
    2. iterate until "mean envelope $= 0$" (*sifting*)
3. subtract the obtained mode from the signal
4. iterate on the residual

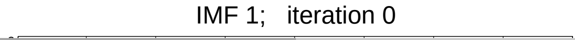
$$\begin{aligned} x(t) &= c_1(t) + r_1(t) \\ &= c_1(t) + c_2(t) + r_2(t) \\ &= \ldots\ldots\ldots\ldots\ldots\ldots \quad = \sum_{k=1}^{K} c_k(t) + r_K(t), \end{aligned}$$

with the $c_k(t)$'s referred to as **Intrinsic Mode Functions** (IMFs)

EMD algorithm in action







EMD features

- *Locality* — The method operates at the scale of **one oscillation**
- *Adaptivity* — The decomposition is fully **data-driven**
- *Multiresolution* — The iterative process explores **sequentially** the “natural" constitutive scales of a signal
- *Oscillations of any type* — **No assumption** on the (e.g., harmonic) nature of oscillations $\Rightarrow$ 1 nonlinear oscillation $=$ 1 mode
- *Instantaneous frequency* — By construction, IMFs are **zero-mean** time-varying waveforms $\Rightarrow$ **Hilbert transform** analysis (so-called “Hilbert-Huang Transform”)

EMD vs. wavelets

1. *similarity*: both achieve a decomposition into "**fluctuations**" and "**trend**"

$$
\begin{aligned}
x(t) &= \sum_k c_k(t) + r_K(t) && \text{(EMD)} \\
&= \sum_k d_k(t) + a_K(t) && \text{(DWT)} \\
\text{with } d_k(t) &= \sum_n \langle x, \psi_{kn} \rangle \, \psi_{kn}(t) \\
\text{and } a_K(t) &= \sum_n \langle x, \varphi_{Kn} \rangle \, \varphi_{Kn}(t)
\end{aligned}
$$

2. *difference*: scales are **pre-determined** for DWT ($\{\varphi, \psi\}_{kn}(t) = 2^{-k/2}\{\varphi, \psi\}(2^{-k}t - n)$) and **adaptive** (data-driven) for EMD

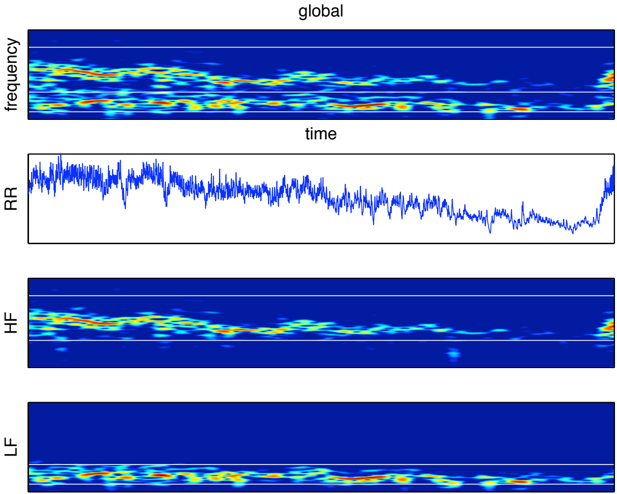
HRV analysis

- **Sympatho-vagal balance** quantified by comparing the low-frequency (LF) and high-frequency (HF) components of RR intervals
- Frequency bands usually **fixed** *a priori* (LF = 0.04-0.15 Hz and HF = 0.15-0.4 Hz) $\Rightarrow$ use of pre-determined, time-invariant filters
- EMD data-driven $\Rightarrow$ automatic, adaptive selection of **time-varying frequency bands** $\Rightarrow$ possibility of dealing with postural changes in tilt tests

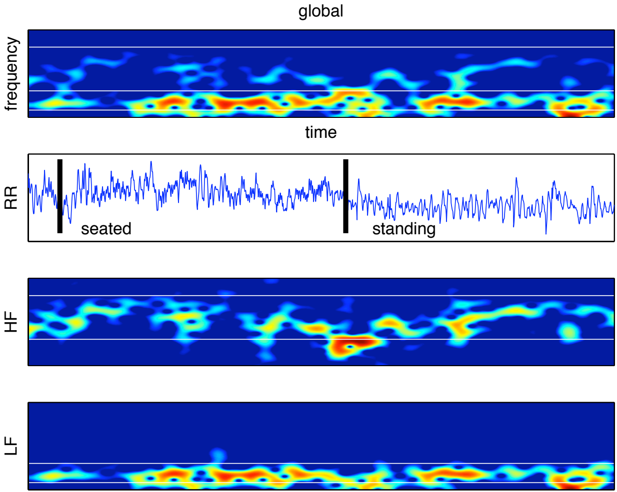
Idea

HF vs. LF $\rightarrow$ *fast vs. slow oscillations + locality*

HRV analysis. Example 1



HRV analysis. Example 2 (Souza Neto *et al.*, '02)



Extensions and uses

- Some extensions
  - **Ensemble EMD** — Increased robustness by adding controlled noise to data (Wu & Huang, ’09)
  - **Bivariate and multivariate EMD** — Analysis of bivariate or complex-valued signals (oscillations → **rotations**) (Rilling *et al.*, ’07) + Rehman & Mandic, ’10
  - **Synchrosqueezing** — Variant of (wavelet-based) reassignment performing an EMD-like decomposition (Daubechies *et al.*, ’09)
- Some uses
  - **Pre-processing** — Baseline removal, signal disentanglement, selection of significant IMFs
  - **Post-processing** — Hilbert transform of IMFs, grouping of significant IMFs, (local) trend removal, denoising from partial coarse-to-fine reconstruction, scaling analysis

Some biomedical EMD applications

- **HRV and baroreflex** (Souza Neto *et al.*, ’02-’04-’07; Magagnin *et al.*, ’08-’09; Ihlen, ’09; Yeh *et al.*, ’10) + **ECG: baseline removal** (Lemay & Vesin, ’06; Pan *et al.*, ’07) + **QRS detection and ventricular fibrillation** (Hadj Slimane & Naït-Ali, ’09) + **Sleep apnea** (Corthout *et al.*, ’08) + **Cardiorespiratory synchronization** (Wu & Huang, ’09)
- **EEG: seizure detection and ocular artifacts** (McKeown *et al.*, ’05; Pachori, ’08; Raghavendra & Dutt, ’07) + **Event-Related Potential classification** (Liang *et al.*, ’05; Williams *et al.*, ’09)
- **Postural stability analysis** (Amoud *et al.*, ’08)
- **Laser-Doppler flowmetry** (Roulier *et al.*, ’05)
- **Esophageal data analysis** (Liang *et al.*, ’05)

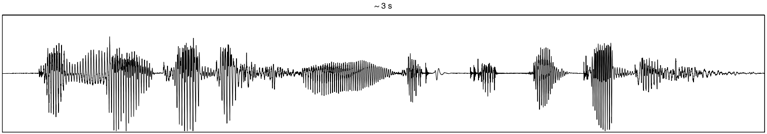
Nonstationarity?

1. Nonstationarity is a **non-property** $\Rightarrow$ no unique form, but infinite number of possibilities
2. **Nonstationarities** to be contrasted with **stationarity**

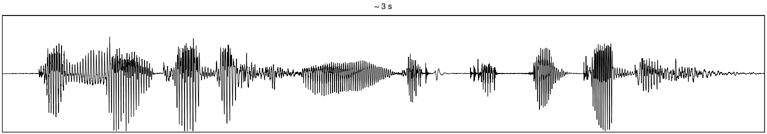
Observation

*Although well-defined in theory, stationarity **doesn't exist** in practice $\Rightarrow$ dealing with nonstationarity/ies cannot be disentangled from revisiting stationarity in some **operational sense***

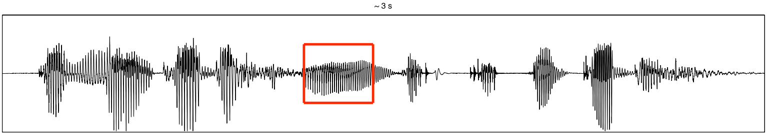
Back to the speech example



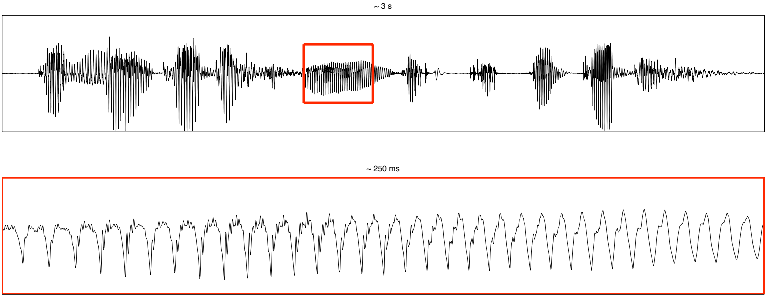
Nonstationary



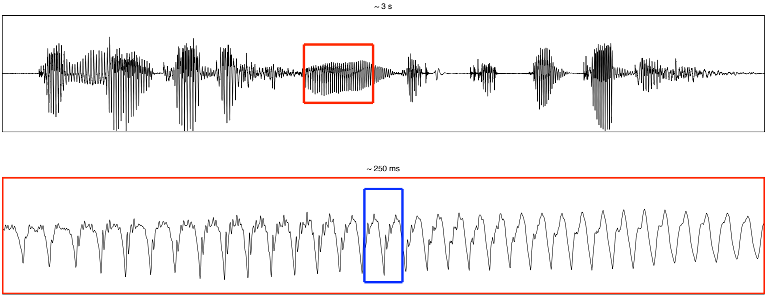
?


~ 3 s

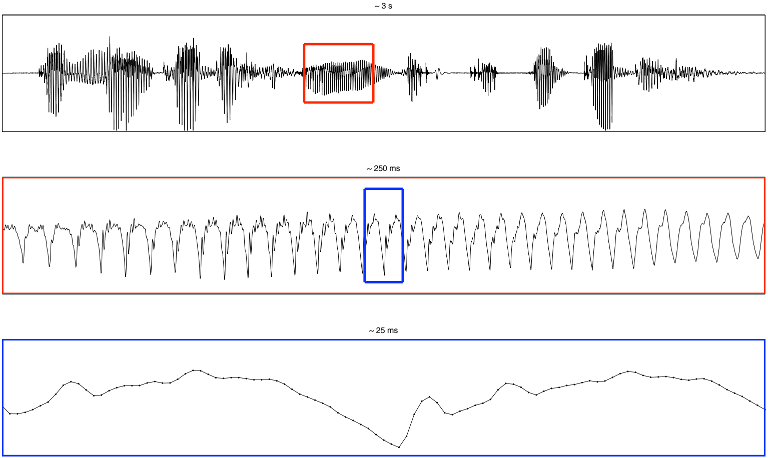
Stationary



?


~ 3 s
~ 250 ms

Nonstationary again!



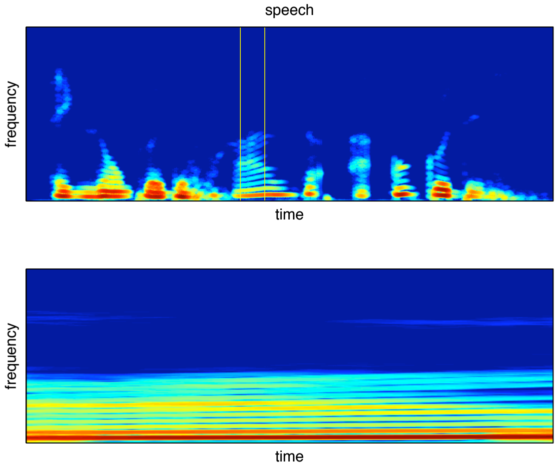
Relative stationarity

1. Form of **invariance**, with respect to
   - time and/or space evolutions
   - more general transformations (e.g., self-similarity)
2. **Relative** concept, depending on "scale"
3. Could be **stochastic** (e.g., mean or variance) as well as **deterministic** (e.g., AM/FM)

Idea

Use Time-Frequency (TF) or Time-Scale (TS) as a unified framework

Relative stationarity in the TF plane

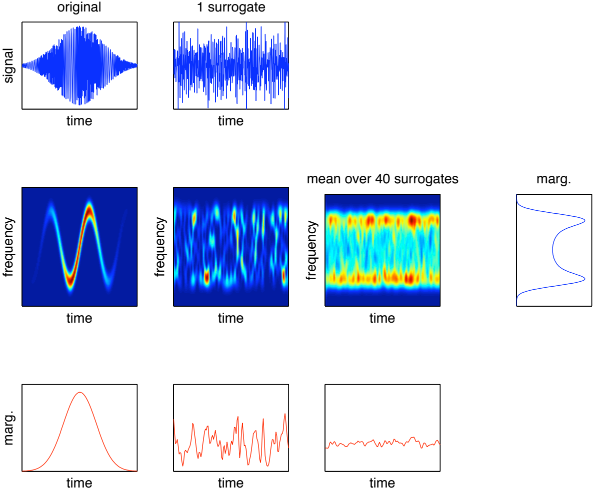


Testing for stationarity: local = global?

- test calls for a stationary reference: how?
  - nonstationarity encoded in **time evolution** or, equivalently, in **spectrum phase**
  - stationarization via spectrum **phase randomization**
- new use of **surrogate data** technique (Theiler *et al.*, '92)
- basic algorithm:

$\hat{x} = \text{FFT}(x)$ **%** x **= original data**
draw WGN $\epsilon(t)$ and compute $\hat{\epsilon} = \text{FFT}(\epsilon)$
$\hat{x} \leftarrow |\hat{x}| \exp\{i \arg \hat{\epsilon}\}$
$y = \text{IFFT}(\hat{x})$ **%** y **= surrogate data**

Stationarization via surrogates 1.



The proposed approach

Principle

1. *compute, from the data, a set of* ***stationary surrogates***
2. *attach to both data and surrogates a series of* ***features*** *aimed at comparing* ***local vs. global*** *behaviors*
3. *construct a test based on the* ***empirical statistical characterization*** *of such features for surrogates (null hypothesis of stationarity)*

(Xiao *et al.*, EUSIPCO'06]; Xiao *et al.*, IEEE-SSP'07; Borgnat *et al.*, IEEE-TSP'10)

Two variations

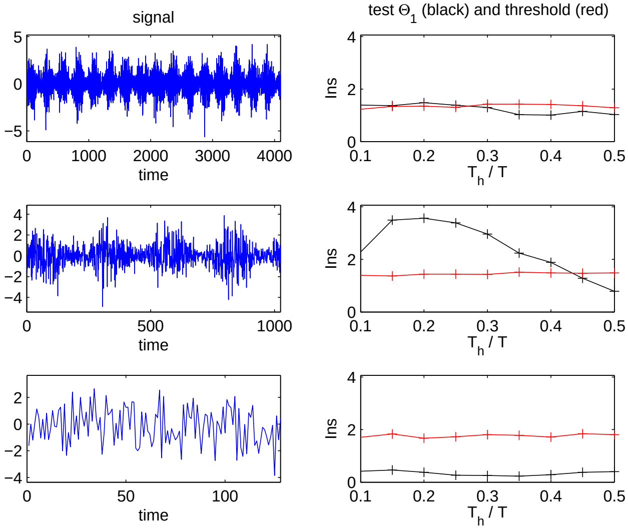
Approach 1

*Compute a **distance** between global and local spectra, and construct a one-sided test based on the **surrogates distribution***

Approach 2

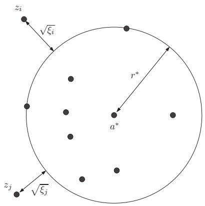
*Consider surrogates as a stationary **learning set**, and construct an outlier detection test by using the machinery of **one-class SVM***

Variation 1: “local-global” distance

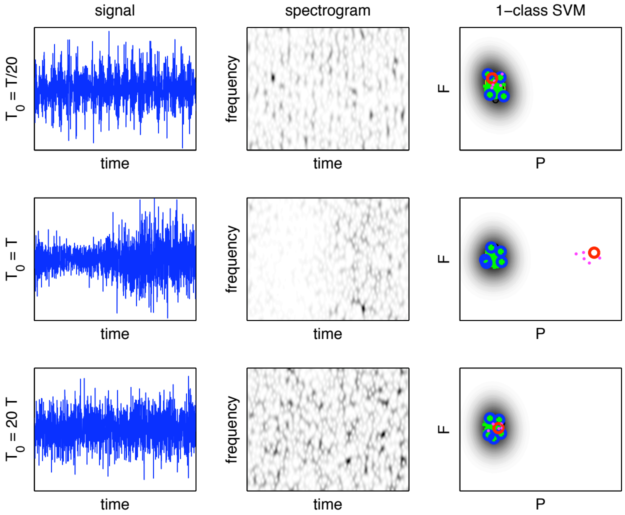


Variation 2: one-class SVM

- **rationale**: determine the minimum volume hypersphere that encloses (most of) the training points, up to a small fraction of data excluded from the domain.
- **optimization**: trade-off between minimizing the radius r^* of the enclosing hypersphere and controlling the sum of the slack variables ξ_i^* associated with each outlier.



Variation 2: one-class SVM



Concluding remarks

1. **Comprehensive** approaches for nonstationary signals
   - **TF**/TS, **decompositions**, models
   - from exploratory data **analysis** to **processing** and **decisions**
2. From theory to applications, **and back**
   - well-established standard methodologies, equipped with algorithms, freewares, ... $\Rightarrow$ improved **toolkit** for biomedical signals
   - specific biomedial problems $\Rightarrow$ sources of inspiration for new **dedicated** tools!