Empirical Signal Decomposition for Acoustic Noise Detection

L. Zão and R. Coelho

Laboratory of Acoustic Signal Processing (LASP)
Military Institute of Engineering (IME), Rio de Janeiro, Brazil
coelho@ime.eb.br
lasp.ime.eb.br

IEEE SAM 2016 Rio de Janeiro Brazil

Abstract

This work introduces an adaptive noise detection method for non-stationary acoustic noisy signals. The proposed approach is based on the empirical mode decomposition (EMD) and a vector of Hurst exponent coefficients. The EMD is a powerful tool for multiresolution analysis of nonlinear and non-stationary signals in the time domain. The proposed noise detection is evaluated considering real noisy signals with different non-stationarity degrees. The results demonstrate that the EMD-based detector enables a better separation between the clean and noisy signals when compared to two competitive methods.

Empirical Mode Decomposition

EMD method: The method decomposes a signal into a series of oscillatory intrinsic mode functions (IMF) and a residual component. The general idea is to locally analyze a signal $x(t)$ between two consecutive extrema (minina or maxima). The fast oscillations are defined as the detail components, $d_k(t)$, while the reamining slow fluctuations compose the residual or local trend, $a_k(t)$.

The first detail function, $d_1(t)$, is obtained from all the consecutive extrema of $x(t)$, such that

$$x(t) = d_1(t) + a_1(t),$$

where $a_1(t)$ denotes the first residual. In general, the separation between the fast and slow fluctuations is repeated over the residual of order $k-1$ to obtain the detail and trend of order k, i.e.,

$$a_{k-1}(t) = d_k(t) + a_k(t).$$

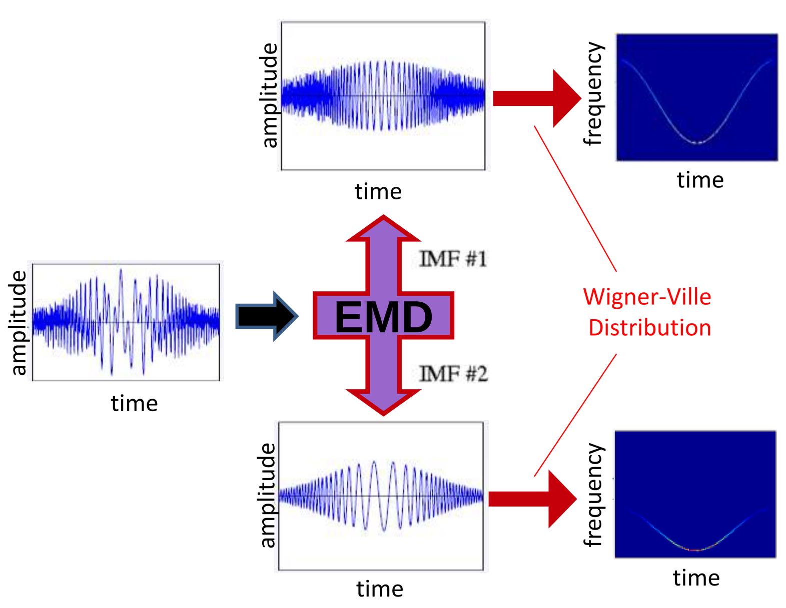

Fig. 1: Separation of two signals with different time-frequency components.

Time-Frequency Analysis

- **Fourier:** the classic theory is not suitable for non-stationary signals.
- **Wavelets:** this time-frequency method requires a set of pre-defined basis functions for the decomposition modes.

- **EMD main features:**
- ✓ **Adaptativity:** decomposition is fully data-driven.
- ✓ **Locality:** IMFs are completely based on the local properties of the input data.
- ✓ **Completeness:** If $IMF_k(t)$ denotes the k-th mode and $r(t)$ is the last residual, EMD assures that

$$x(t) = \sum_{k=1}^{K} IMF_k(t) + r(t)$$

EMD Algorithm

EMD algorithm can be described as follows:

1. Set $k = 1$ and initialize the variable $a_0(t) = x(t)$;
2. Identify all local minima and maxima of $a_{k-1}(t)$;
3. Obtain the upper ($e_{max}(t)$) and lower ($e_{min}(t)$) envelopes by cubic splines interpolation of the local maxima and minima, respectively;
4. Compute the local trend $a_k(t) = (e_{max}(t) + e_{min}(t))/2$;
5. Calculate $d_k(t) = a_{k-1}(t) - a_k(t)$ as the new detail;
6. Set $k = k + 1$ and iterate steps 2-5 on the new residual local trend $a_k(t)$.

Detail function: it will be considered as an IMF when its mean is close to zero ($< 10^{-6}$), and all its maxima and minima are positive and negative, respectively.

Sifting: while a detail function $d_k(t)$ is not considered as an IMF, steps 2-5 are repeated with $d_k(t)$ in place of $a_{k-1}(t)$.

Stopping criteria: the algorithm stops when the last residual, $a_k(t)$, has less than three extrema.

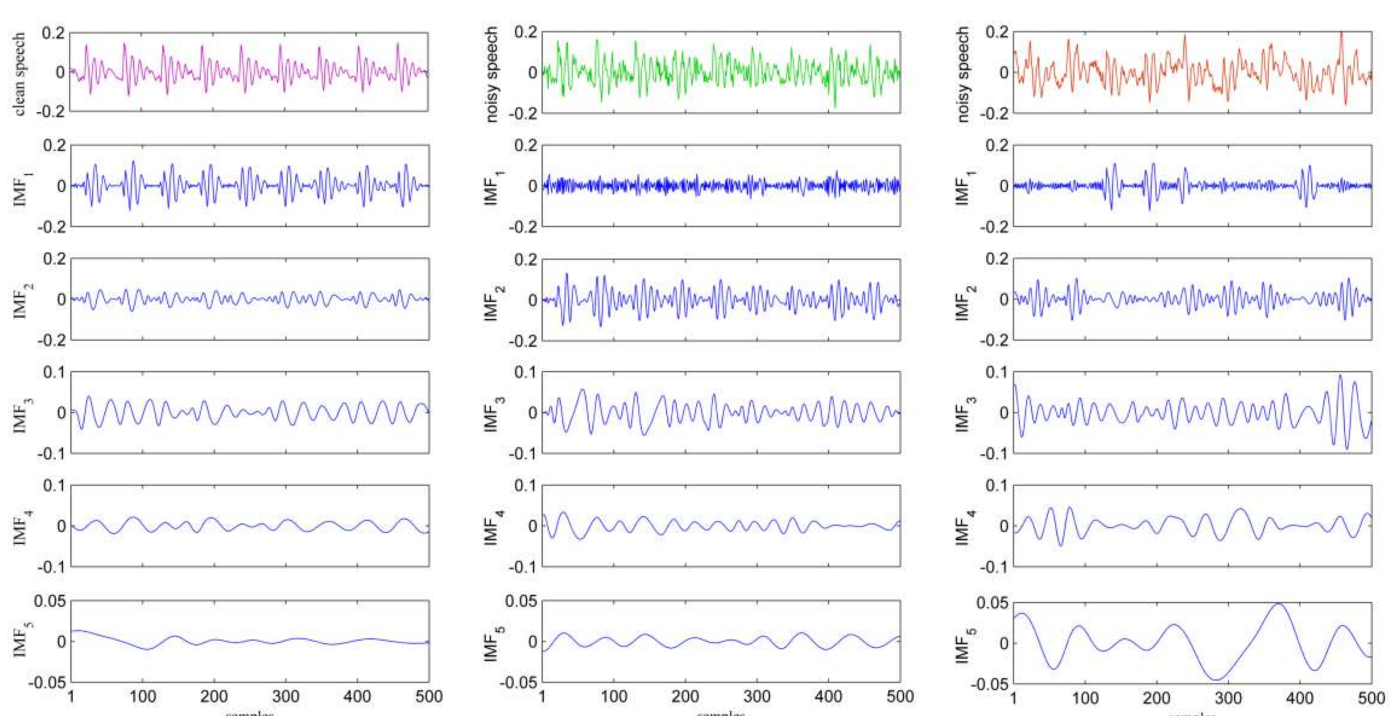

Fig. 2: The first five IMFs obtained from (a) a clean speech signal, and the same signal corrupted with (b) white noise and (c) car traffic noise.

Dyadic filterbank structure:
When applied over fractional Gaussian noise (fGn) processes, EMD behaves like a dyadic filterbank with overlapping band-pass filters.

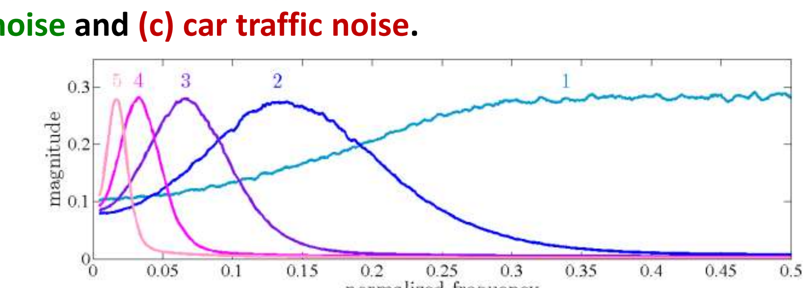

Fig. 3: Average magnitudes of IMFs obtained from a fGn white noise sample sequence.

Non-Stationarity

The index of non-stationarity (INS) is a time-frequency approach to objectively examine the non-stationarity of a target signal. For each window length T_h a threshold γ is defined for the stationarity test, such that:

$$\text{INS} \begin{cases} \leq \gamma: \text{ signal is stationary;} \\ > \gamma: \text{ signal is non-stationary.} \end{cases}$$

INS of Acoustic Signals and Noises: Speech signal and chainsaw noise are highly non-stationary (**HNS**; $\text{INS}_{max} > 80$); jackhammer noise is non-stationary (**NS**; $40 < \text{INS}_{max} < 50$); babble signal and car traffic noise are moderately non-stationary (**MNS**; $\text{INS}_{max} < 20$).

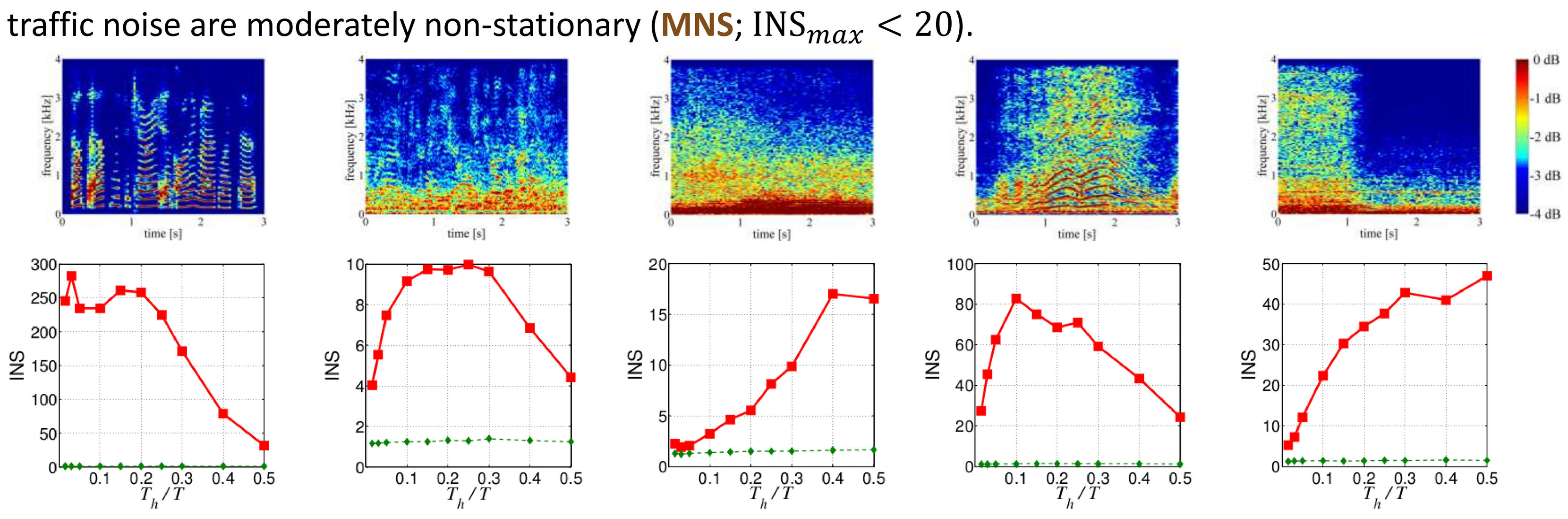

Fig. 4: Spectrogram and index of non-stationarity of acoustic signals: (a) speech, (b) babble, and noises: (c) car traffic, (d) chainsaw, and (e) jackhammer. Dashed lines indicate the values of γ.

Acoustic Noise Detection

❑ Detection of the Noise Components:

The Hurst exponent ($0 \leq H \leq 1$) is a time-frequency coefficient and it is related to the power spectral characteristics of a signal $x(t)$. The power spectral density $S_x(f) \propto f^{1-2H}$ when $f \to 0$.

$H < 1/2$: $S_x(f)$ is mostly concentrated at the high frequencies;
$H = 1/2$: $S_x(f)$ is approximately constant over the entire spectrum (e.g., white noise);
$H > 1/2$: low frequencies are prominent, in particular when $H \to 1$ ($1/f$ or pink noise).

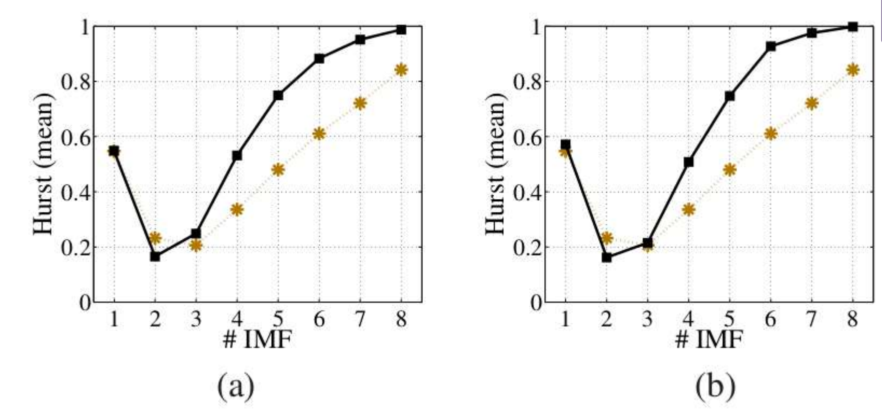

Fig. 5: Values of H estimated from IMFs obtained from the speech signal corrupted with acoustic noises with signal-to-noise ratio of 5 dB: (a) car traffic (MNS) and (b) jackhammer (NS). Dashed lines refer to the clean speech.

Signal Reconstruction: The noise components are assumed to be mostly concentrated at the IMFs with $H \cong 1$ (low-frequency). After the decomposition of the noisy signal, each IMF is divided into short-time frames. For each frame, the Hurst exponent defines an index L such that the target signal is reconstructed using only the first $L-1$ modes: $\hat{x}(t) = \sum_{k=1}^{L-1} IMF_k(t)$.

Experiments and Results

❑ Acoustic Noise Detection: Competitive Criteria

1. Variance: identifies the first IMF with index L ($L \geq 4$) where the variance is greater than the adjacent modes, i.e.,

$$\text{Var}[IMF_L(t)] > \text{Var}[IMF_{L-1}(t)] \text{ and } \text{Var}[IMF_L(t)] > \text{Var}[IMF_{L+1}(t)].$$

2. StdMean: is defined as the ratio of the mean and the standard deviation of an IMF. This criterion identifies L as the first index for which StdMean is greater than the root mean square of the standardized mean of the first four modes.

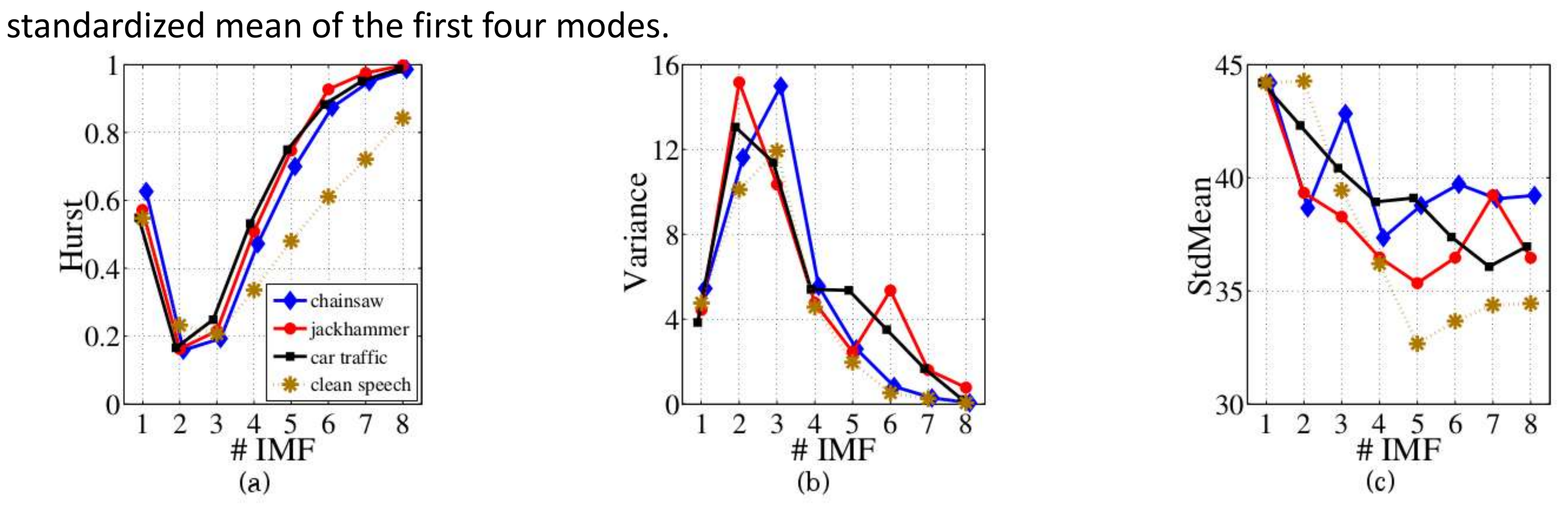

Fig. 6: Average values of (a) Hurst, (b) Variance, and (c) StdMean adopted for acoustic noise detection.

❑ Reconstructed Signal Quality (SNR and SegSNR)

Target Signal Reconstruction: the Hanning window is used to avoid discontinuities after the concatenation of the reconstructed frames. The reconstructed signals are evaluated in terms of signal-to-noise ratio (SNR) and segmental SNR (SegSNR).

Tab. 1: SNR (dB) of the reconstructed signals.

Noise	SNR	speech signal			babble signal		
		Hurst	Variance	StdMean	Hurst	Variance	StdMean
car traffic (MNS)	-5	**6.7**	-4.2	-4.0	**2.5**	-4.9	-4.0
	0	**9.9**	2.2	1.4	**3.7**	0.6	1.1
	5	**12.9**	7.5	6.7	5.3	4.4	4.6
jackhammer (NS)	-5	**3.8**	-0.4	-1.5	**1.6**	-2.3	-3.1
	0	**7.6**	5.6	3.2	**4.3**	2.4	1.5
	5	**11.7**	9.7	7.8	**6.7**	5.1	5.4
chainsaw (HNS)	-5	**-3.4**	-4.7	-4.7	-4.3	-4.9	-4.8
	0	0.8	0.4	0.3	0.0	-0.4	-0.3
	5	5.3	5.2	5.2	**4.9**	3.7	3.8

Tab. 2: SegSNR gain (dB) of the reconstructed signals.

Noise	SNR	speech signal			babble signal		
		Hurst	Variance	StdMean	Hurst	Variance	StdMean
car traffic (MNS)	-5	**5.2**	1.2	0.9	**4.8**	0.3	0.9
	0	**4.1**	1.8	1.1	**2.2**	0.5	0.7
	5	**2.4**	1.4	1.0	0.0	-0.5	-0.4
jackhammer (NS)	-5	**4.6**	3.6	2.4	**4.7**	3.1	2.3
	0	**3.4**	3.3	1.8	**3.0**	2.1	1.4
	5	**2.2**	1.9	1.1	**0.9**	0.1	0.2
chainsaw (HNS)	-5	**0.7**	0.3	0.2	0.2	0.0	0.0
	0	0.3	0.2	0.2	0.0	-0.6	-0.3
	5	0.1	0.1	0.1	**0.2**	-1.1	-0.8

EMD: Trends and Challenges

The EMD enables the analysis of numerous natural and artificial:

biomedical engineering   image processing   seismic
speech processing   pattern recognition   financial market

❑ Challenges:

- **Mode Mixing Problem.**
  - ✓ Ensemble EMD (EEMD);
  - ✓ Complete EEMD with Adaptive Noise (CEEMDAN);
  - ✓ Multivariate EMD (MEMD).
- **Envelope Computation.**
  - ✓ Optimization-based Mode Decomposition (OMD);
  - ✓ Sequential Variational Modal Decomposition (Seq-VMD).

Conclusion

This work presented a time-domain noise detection scheme for signals corrupted by non-stationary acoustic noise. The proposal is derived from a two steps procedure composed by the empirical mode decomposition and a Hurst exponent vector. A SNR gain of 1.6 dB is obtained for the highly non-stationary chainsaw noise source. Moreover, it can be very promising for speech enhancement solutions.

References: Highlights

❑ N. Huang, et al, "The Empirical Mode Decomposition and Hilbert Spectrum for Nonlinear and Non-Stationary Time Series Analysis", Proceedings of the Royal Society, vol. 454, no. 1971, pp. 903-995, 1998.

❑ P. Flandrin, G. Rilling and P. Gonçalvès, "Empirical Mode Decomposition as a Filter Bank", IEEE Signal Processing Letters, vol. 11, no. 2, pp. 112-114, 2004.

❑ P. Borgnat, P. Flandrin, P. Honeine, C. Richard and J. Xiao, "Testing Stationarity With Surrogates: A Time-Frequency Approach," IEEE Transactions on Signal Processing, vol.58, no.7, pp. 3459-3470, July 2010.

❑ P. Flandrin, P. Gonçalvès and G. Rilling, "Detrending and Denoising with Empirical Mode Decomposition", Proc. of the Eur. Sig. Proc. Conf. (EUSIPCO 2004), pp. 1581-1584, 2004.

❑ N. Chatlani and J. Soraghan, "EMD-based Filtering (EMDF) of Low-Frequency Noise for Speech Enhancement", IEEE Transactions on Audio, Speech and Language Processing, vol. 20, no. 4, pp. 1158-1166, May 2012.

❑ L. Zão, R. Coelho and P. Flandrin, "Speech Enhancement with EMD and Hurst-Based Mode Selection," IEEE/ACM Transactions on Audio, Speech and Language Processing, vol. 22, no. 5, pp. 899-911, May 2014.

❑ R. Coelho and L. Zão, "Empirical mode decomposition theory applied to speech enhancement", in Signals and Images: Advances and Results in Speech, Estimation, Compression, Recognition, Filtering, and Processing, eds. R. Coelho, V. Nascimento, R. Queiroz, J. Romano and C. Cavalcante, CRC Press, 2015.

❑ R. Sant Ana, R. Coelho and A. Alcaim, "Text-Independent Speaker Recognition Based on the Hurst Parameter and the Multi-Dimensional Fractional Brownian Motion Model", IEEE Transactions on Audio, Speech and Language Processing, vol. 14, no. 3, pp. 931-940, May 2006.