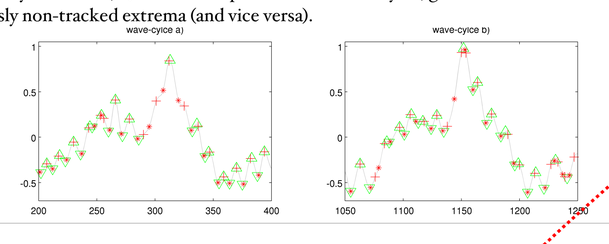
Practical Empirical Mode Decomposition for Audio Synthesis

Niklas Klügel, Technische Universität München, kluegel@in.tum.de

Modification I

Problem: Riding waves may not produce local extrema (e.g. saddle points only) *if immersed* in local trend.
Alternative: Use inverse of the second derivative, $-\frac{d^2}{dx^2}f(x)$; for pure sinusoids local extrema stay the same.
Example: A signal and its local extrema (green markers) and those of the alternative solution (red markers): a) wave cycle of a signal with local extrema of riding component partially immersed; b) later development of the wave-cycle; green markers indicate previously non-tracked extrema (and vice versa).



Modification II

Problem: Usually, $\mathcal{S}$ is a fixed amount of iterations. Large overshoots of the spline approximation may *degenerate* the fitted interpolant.
Solution: Abort current iteration and recover last one, if area under the spline increased. Since the filtering removes the local trend, the area should always decrease in subsequent iterations.

Modification III

Problem: *Modified* (e.g. resampled) instantaneous attributes for resynthesis may result in severe phase distortion / amplitude modulation due to intra-cycle modulations in the analyzed signal.
Solution: Smooth the instantaneous attributes in a *time-variant* way (∼lossy compression scheme):

- use whole proposed method for each instantaneous attribute of the mono-components → creates *sets* of mono-components (of instantaneous attributes)
- for each set calculate the overall energy, discard all elements except those contributing most to the overall energy (e.g. keep 0.9 percentile)
- resynthesize modified sets as new instantaneous attributes

-Motivation-

The synthesis by analysis approach for sounds with non-stationary & non-linear changing attributes still poses problems that are to a large degree rooted in the properties of the underlying Time Frequency Representation (TFR) methods.

Problematic mathematical properties of common TFRs:

- basis chosen *a priori* may color or deprive the TFR of physical meaning
- Heisenberg-Gabor limit forces a trade-off for time/frequency resolution
- presence of interference terms

Goal: High resolution Time and Frequency representation

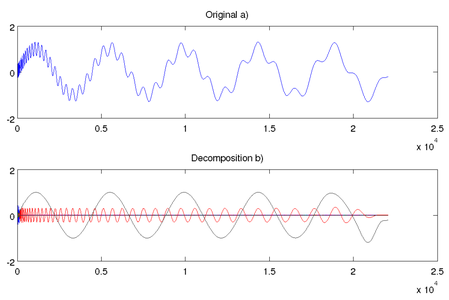
Focus: derive instantaneous attributes of the individual components of signal $s(t)$

For the analysis one is interested to decompose (the multi-component signal) $s(t)$ into individual mono-components. A mono-component $h_k(t)$ is a sinusoid whose amplitude and phase varies with time, so $s(t)$ can be expressed with a residue $r(t)$ as:

$$s(t) = \sum_{k=1}^{n} h_k(t) + r(t) = \sum_{k=1}^{n} A_k(t)\cos\phi_k(t) + r(t) \quad (1)$$

If $h_k(t)$ exhibits a well-behaved Hilbert-Transform (HT) the analytic signal $H_k(t)$ and its instantaneous attributes can be constructed:

$$|H_k(t)| = \sqrt{h_k^2(t) + \tilde{h_k}^2(t)} = A_k(t),\ \phi_k(t) = \arctan(\frac{\tilde{h_k}(t)}{h_k(t)}) \quad (2)$$



A chirp sinusoid riding on top of a frequency invariant sinusoid a), its decomposition b) with the proposed method resulting in two mono-components (red & black) and a residue (blue)

⇒ The Empirical Mode Decomposition (EMD) and Hilbert Vibration Decomposition (HVD) allow for such a demodulation of wide-band signals.

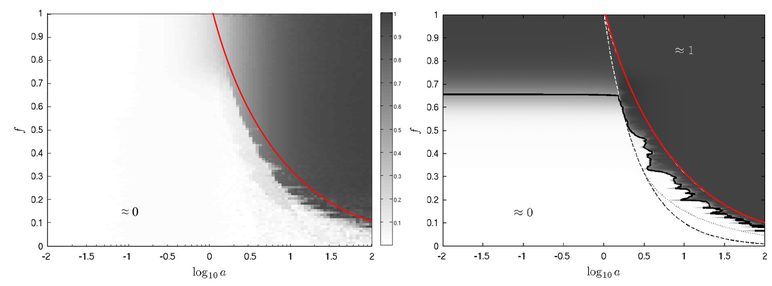
	EMD	HVD
Idea	determine base functions (= mono-components) *a posteriori* during decomposition	
Philosophy	innate undulations belong to diff. relative freq. scales, discern these waves	instantaneous attributes of the component with highest energy change slowly cp. to their sum of underlying components
Method	recursively reconstruct riding waves on local trend by (spline) approximation	recursively low-pass filter instantaneous frequency of Hilbert-Transformed signal
Frequency Resolution	f_{res} determined by approximation, e.g. cubic spline: $1/(5\Delta t)$ Hz, *theoretical limit* of discrimination for two comps.: $\frac{A_1}{A_2}(\frac{f_2}{f_1})^2 = 1$, for noise roughly ≥ 1 octave	bound to low-pass filter steepness; discrimination for two comps. typically around ≥ 0.5 octave for filter with cutoff freq. $f_{min} \geq 0.02Fs$

	Advantages	Disadvantages
EMD	creates up to $k \leq \log_2 N$ mono-components insensitive to noise & bursts	spectra of close components can become mixed
HVD	can decompose close spectra of components	sensitive to noise creates only a few mono-components $k \leq 7$ introduces errors to instantaneous attributes due to applied FIR filtering

Combination of EMD & HVD

Main idea: use the EMD as main method but embed HVD to generate masking signals that emphasize components with high energy during decomposition

Result: Decomposition performance *increased* to near theoretical limit of EMD



Decomposition performance of bi-component signals; the method proposed a) and original EMD b); input $x(t)$ consists of a fixed parameter HF signal $x_h(t)$ and a variable parameter (axes of the plots) LF signal $x_l(t)$ relative to $x_h(t)$. A lower z-value means the first component extracted is more similar to $x_h(t)$. The red lines show the theoretical limit ($a_l f_l^2 = 1$) for the decomposition.

-Synthesis-

- mono-components can be seen as time-variant partials → no tracking of spectral peaks needed
- resampling & scaling $A_k(t)$ & $\phi_k(t)$ (unwrapped) allows to alter the sound for resynthesis (time stretching/compressing and pitch shifting)
- to change the scaling ratio of phase functions relative to the fundamental (e.g. mono-component with *largest* $A_k(t)$ at time instant t) phase scaling functions $\Delta_k(t)$ can be defined and applied for:
  - pitch shifting with spectrum dilation
  - frequency shifting
  - treating the partials as fixed formants
  - a mix of the above
- currently a dedicated noise model is missing

Applied Informatics / Cooperative Systems AICOS

Contact, Examples & Code

Technische Universität München