An Application of Empirical Mode Decomposition On Tempo Induction From Music Recordings

Aggelos Pikrakis and Sergios Theodoridis

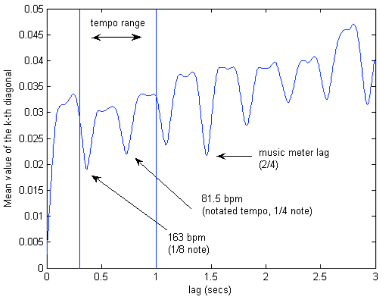
Department Of Informatics & Telecommunications, University of Athens, Greece,
e-mail: {pikrakis,stheodor}@di.uoa.gr, web: http://dsp.di.uoa.gr

Summary

This paper presents an application of Empirical Mode Decomposition (EMD) on the induction of notated tempo from music recordings. At a first stage, EMD is employed as a means to form clusters of music segments that exhibit similar rhythmic characteristics. At a second stage, EMD is used to analyze the mean "rhythmic signature" of each cluster so as to estimate the tempo of the recording. Signatures are extracted by means of Self-Similarity analysis. The proposed method has been employed on various music genres with music meters 2/4, 3/4 and 4/4. Tempo has been assumed approximately constant throughout each recording, in the range 60bpm-220bpm.

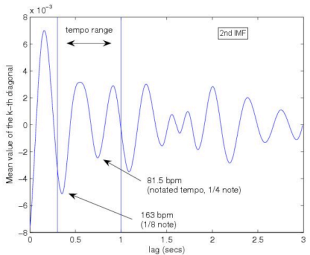
Feature Extraction

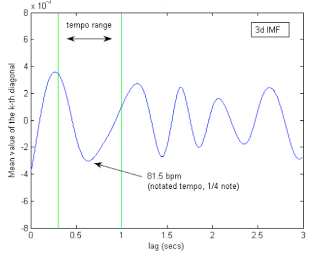
- The music recording is first split into *overlapping long-term segments*. Each long-term segment is 5 s long with 4 s overlap between successive windows.
- The *energy envelop* of each long-term window is then extracted. Suggested values for the length, w_s and hop size h_s of the short-term window are 95ms and 5ms respectively.
- The energy envelop is used to generate the *Self-Similarity Matrix (SSM)* of the segment. The Euclidean Distance function has been used as the distance metric.
- Once the SSM is generated, the mean value, B_k, k=1,…,D, of each diagonal is computed.
- If B_k is treated as a function of k, it can be observed that signal periodicities appear as local minima (valleys) of **B**.
- We define **B** to be the "rhythmic signature" of the segment.



Signature clustering using EMD

- Group "rhythmic signatures'' into clusters and compute the mean signature of each cluster.
- EMD is applied separately on each signature.
- Basic steps of EMD given a signal x(t):
  - Identify all extrema of x(t)
  - Interpolate between minima (resp. maxima), ending up with some "envelope" $e_{min}(t)$ (respectively $e_{max}(t)$).
  - Compute the average $a(t)=(e_{min}(t) +e_{max}(t))/2$
  - Extract the detail d(t)=x(t)-a(t), also known as the IMF. Iterate on the residual a(t) until a stopping criterion is satisfied, i.e., a(t) is reasonably zero everywhere
- Let $\mathbf{B}_m$ be the signature of the m-th long-term segment. Let c_m be the number of components (**IMFs**) generated by the EMD for $\mathbf{B}_m$
  - At a first step, clusters are formed by grouping $\mathbf{B}_m$'s that have generated the same c_m.
  - At a second step, we focus on each formed cluster. For each signature in such a cluster the energy of IMF's is computed and components are sorted in descending order, according to their energy values. The resulting order is then used to form sub-clusters within every cluster, i.e. segments that yield the same order of components are considered to be similar.
- The mean signature, $\boldsymbol{R_l}$, of the *l-th* cluster is computed by simply averaging signatures.





Tempo Induction

- $\boldsymbol{R_l}$ is decomposed with EMD. The two IMF's of $\boldsymbol{R_l}$ that possess the higher energy values are then chosen.
- The following procedure is applied on each one of the two IMF's:
  - All valleys of the IMF are detected, including valleys to the right of the tempo region.
  - Each valley in the tempo region, is then examined against all valleys with larger lags. Let $\mathbf{k}_m$ be the lag of a valley in the tempo region and $\mathbf{k}_i$ the lag of the valley against which it is examined. If the roundoff error of $\mathbf{k}_i/ \mathbf{k}_m$ is less than **0.1**, then $\mathbf{k}_i$ is considered to be a multiple of $\mathbf{k}_m$, i.e., $\mathbf{k}_m$ is treated as a fundamental periodicity and $\mathbf{k}_i$ as its multiple. This is repeated for all possible pairs, yielding a set $\mathbf{L}_{\mathbf{k}_m}$ of multiples for $\mathbf{k}_m$. The following sum is then computed:
  - $P_{k_m} = c_{k_m} + \sum_{\forall k_i \in L_{k_m}} c_{k_i}$, where $\mathbf{c}_{\mathbf{k}_i}$ is the value of the IMF for lag $\mathbf{k}_i$.
  - The above step is repeated for all valleys that fall within the allowable tempo limits. The valley with the highest sum is selected as the winner and the corresponding lag as the periodicity of the tempo estimate.
- Tempo estimates from all clusters are placed in a histogram and the tempo corresponding to the highest peak is selected as the tempo of the music recording. It is often the case, that the histogram exhibits lobes around peaks because EMD tends to slightly displace periodicities. This is why an averaging of the lobes (histogram smoothing) is needed prior to selecting the highest peak.

Results

- Tempo is estimated from clusters with at least 3 signatures. At an average, 25% of the signatures in each audio recording is grouped into such clusters. Large clusters are likely to contain approximately 20 signatures.
- The average accuracy of the extracted tempo value lies within 3.5% of the respective notated tempo value.
- Notated tempo was successfully inducted from 79% of the recordings.
- Cases of failure: **(a)** twice or half the notated value, **(b)** three times the notated value (for fast contemporary music of meter 3/4, **(c)** in Greek Folk music of meter 2/4, the dotted quarter-note is returned as the dominant periodicity.

	Music Meter		
Broad Music Genre	$\frac{2}{4}$	$\frac{3}{4}$	$\frac{4}{4}$
Contemporary Pop/Rock	20%	5%	40%
Traditional Greek Folk music	10%	15%	10%

Distribution of music tracks among genres and music meters.

	Music Meter		
Broad Music Genre	$\frac{2}{4}$	$\frac{3}{4}$	$\frac{4}{4}$
Contemp. Pop/Rock (2xbpm)	3%	0%	2%
Contemp. Pop/Rock ($\frac{1}{2}$bpm)	1.5%	0%	3%
Contemp. Pop/Rock (3xbpm)	0	1.5%	0%
Greek Folk Dances (2xbpm)	1%	0%	3%
Greek Folk Dances ($\frac{1}{2}$bpm)	2%	0%	2%
Greek Folk Dances (3xbpm)	0%	0%	0%
Greek Folk Dances (1.5xbpm)	2%	0%	0%

Distribution of tempo induction failures.

References:
[1] A. Pikrakis, I. Antonopoulos and S. Theodoridis, "Music Meter and Tempo Tracking from raw polyphonic audio", Proceedings of ISMIR, Barcelona, Spain,2004.
[2] N.E. Huang et al., "The empirical mode decomposition and Hilbert spectrum for nonlinear and nonstationary time series analysis", Proceedings of R. Soc. London, vol. 454, pp. 903-995, 1998.
[3] P. Flandrin, G. Rilling and P. Concalves, "Empirical mode decomposition as a filter bank", IEEE Signal Processing Letters, vol 11(2), pp. 112-114, Feb. 2004.