ISMIR Lecture Notes

# Signal Processing Methods for Sound Recognition

Dr. E. Benetos

Conferences – ISMIR 2005, ICA 2007, MPEG 2009, CMMR 2012, WAC 2017, Audio Mostly 2017, SysMus 2017

Software – Sonic Visualiser (+Vamp plugins), SoundBite, Tony

**Machine Listening for Sound Scenes**

* Sound scene recognition
* Sound event detection
* Source separation
* Noise monitoring/reduction

**Sound Event Detection ( SED )**

Main goal: identify labels and start/end times for sound events

**Matrix Factorization**

Spectrogram ( frequency-time representation ) – treat as a matrix and factorize into two matrices

NMF ( Nonnegative Matrix Factorization)

VFxT = WFxK x HKxT

Nonnegative matrices – W (dictionary of templates containing signatures of events) , H ( activation matrix)

PLCA – Probabilistic Latent Component Analysis

* Decompose a bivariate distribution over frequency and time
* Probabilities of frequencies over particular component ( W )
* Probabilities of component over time ( H )

**Recognizing multiple overlapping sounds**

Mel and ERB spectrograms – what are the difference ?

Mel- frequency cepstrum : representation of the short term power spectrum of a sound.

* P(t) : spectrogram energy ( known quantity )
* First term : spectral template for event class s , exemplar c, and sound state q ( fixed, pre-extracted )
* Second term : event activation over time
* Third term : exemplar contribution for each event class, over time
* Fourth term : sound state contribution for each event class, over time. Controlled by an event HMM

Application source code : <https://code.soundsoftware.ac.uk./projects/sound-event-detection-plca>

Tracking sequences over time using Hidden Markov Models ( discrete ) or a State Space Model ( SSM )

**Linear Dynamical System** is a state space model where all the conditional probability distributions are linear-Gaussian.

* Given some output you can infer information about some hidden variables/features.
* The hidden variables are linked by a transition matrix

Approach : Use LDS to track multiple concurrent sound events

Motivation : Assume the posteriogram is the observation in an LDS, with the latent states corresponding to the ‘clean’ posteriogram. Essentially clean up the model which was unsatisfying and noisy.

**Future perspectives**

* Use of context to better inform models
* Sound event taxonomy
* Language/ library modeling

# Automatic Music Transcription by Dr. Emmanouil Benetos – Part 1

Dr. E. Benetos