Acoustic Modelling of Drum Sounds with Hidden Markov Models for Music Transcription

Jouni Paulus
jouni.paulus@tut.fi
Institute of Signal Processing, Tampere University of Technology, Tampere, Finland

Introduction

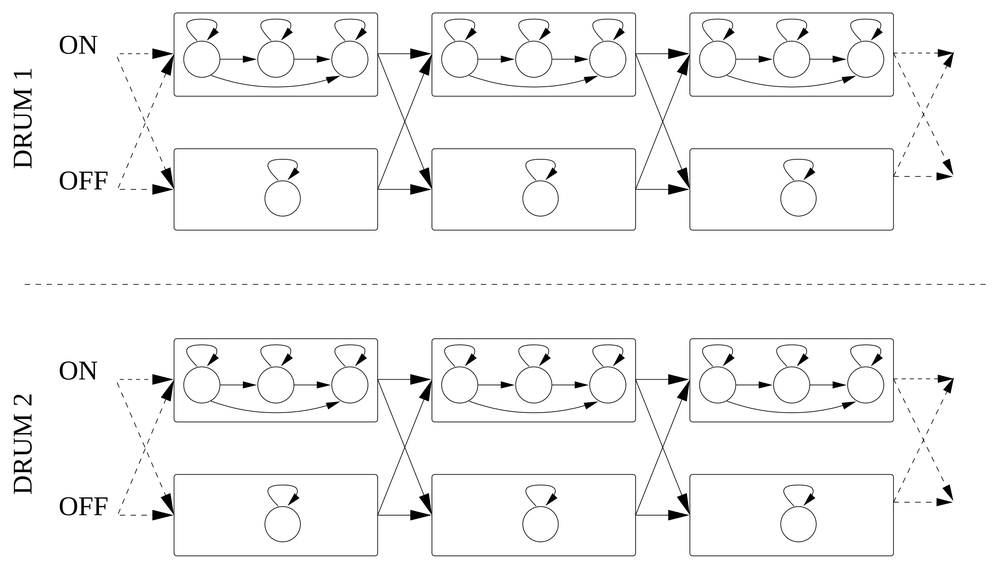
- Aim to transcription (detection and recognition) of drums from polyphonic music, e.g., from acoustic signal to MIDI file.
- Two methods for applying HMMs in acoustic modelling of drum sounds is presented: instrument-wise and combinations.
  - HMMs enable modelling evolution of features during events.

Analysis front-end

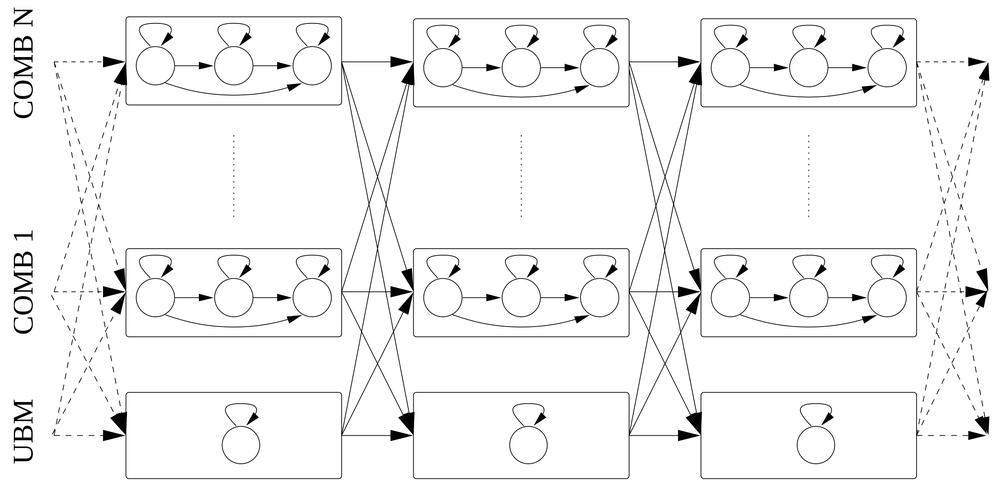
- Pre-processing with sinusoids+residual -modelling.
- Extract a set of spectral features from short, overlapping frames.
- Two linear, unsupervised feature transformations.
  - Reason: features contain redundant information $\rightarrow$ decorrelation and dimensionality reduction
  - Principal component analysis, removes second order statistical dependencies.
  - Independent component analysis, removes also higher order dependencies.

HMM architectures

- Observation distributions with Gaussian mixture models.
- Instrument-wise models
  - Each instrument is modelled independently from others, detector-like.
  - 5-state HMMs for sound events, 1-state HMM for background (UBM), UBM common for all.



- Combination modelling
  - Models for all instrument combinations.
  - All combinations need not to be modelled due rare occurrence.

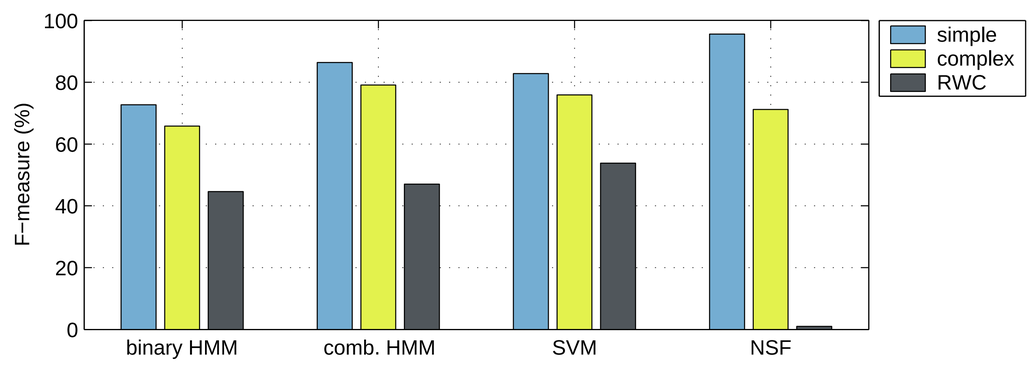


Evaluations

- Evaluated with acoustic recordings from 3 data sets with, each set with cross-validation:
  - *simple drums*, mainly target drums, simple patterns
  - *complex drums*, also non-target drums present, more complex patterns
  - *RWC Pop*, 100 polyphonic music pieces (30 s excerpts).
- Transcribe kick, snare and hi-hat.
- Performance compared with two other systems.
  - Event-based recognition: onset detection, features, classification with binary SVMs
  - Source separation with a dictionary: non-negative spectrogram factorisation (NSF), onsets from components
- Measures: precision rate P, recall rate R, and F-measure.

Results

- F-measures for the HMM and reference methods:



- Detailed results for the best performing HMM systems (combinations) for each evaluation material set:

material	metric	kick drum	snare drum	hi-hat
simple drums	P(%)	81.7	88.8	82.6
	R(%)	89.5	82.5	93.4
complex drums	P(%)	73.5	59.8	76.3
	R(%)	92.2	86.6	89.6
RWC Pop	P(%)	38.6	24.3	44.2
	R(%)	73.5	54.5	62.7

Conclusions

- Aim to transcribe drums from complex signals with two different HMM systems.
- Modelling drum combinations instead of individual drums yielded better result.
  - Probably because drums are not independent from each other.
- Main problem low precision due event insertions.
  - Probably could be alleviated with musicological modelling.