Drum Transcription with Non-negative Spectrogram Factorisation

Jouni Paulus and Tuomas Virtanen

{jouni.paulus, tuomas.virtanen}@tut.fi

Institute of Signal Processing, Tampere University of Technology, Tampere, Finland

Introduction

- Aim to transcribe polyphonic drum sequences from audio to a symbolic representation.
- Overlapping sounds are problematic with many current systems.
  - Individual drum hits can be recognised reliably.
- Use source separation to segregate individual instruments to own streams.

Signal model

- Signal represented as magnitude spectrogram.
- Drum sounds' spectral content is relatively static over the whole hit.
- Sum of N sources with fixed spectra $S_n(f)$ and time-varying gains $a_{n,t}$

$$X_t(f) \approx \sum_{n=1}^{N} a_{n,t} S_n(f) \qquad (1)$$

- Spectrogram wih very coarse frequency resolution: 5 approximately log-spaced bands.

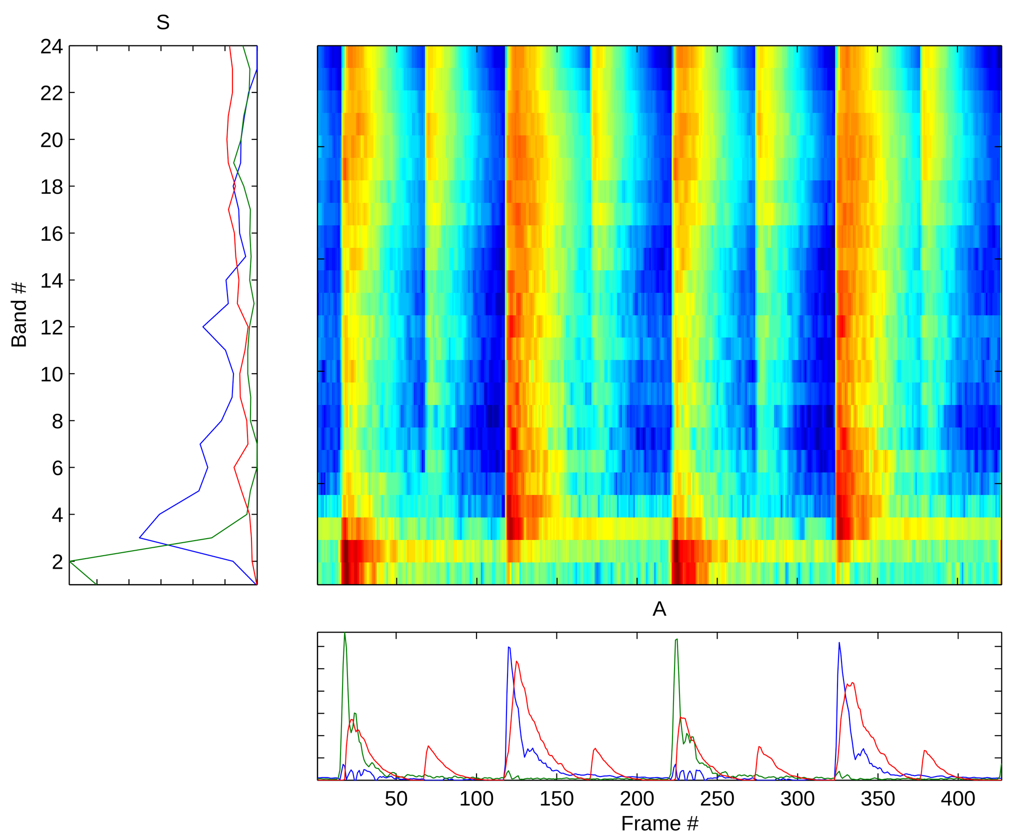

FIGURE 1: An example decomposition of a drum loop into three sources, Bark band frequency resolution.

Method

- Calculate template source spectra S from training samples.
  - Normal NMF used to separate each example hit to one source.
  - Source spectra of each drum type are averaged over examples.
- Estimate corresponding time-varying gains a to fit model and input signal together.
  - Minimise cost function (divergence) between original spectrogram X and the estimate M.
  - Iterative updates until convergence:

$$a_{n,t} \leftarrow a_{n,t} \frac{\sum_f X_t(f) S_n(f) / M_t(f)}{\sum_f S_n(f)}. \qquad (2)$$

- Detect onset times from estimated gains a.

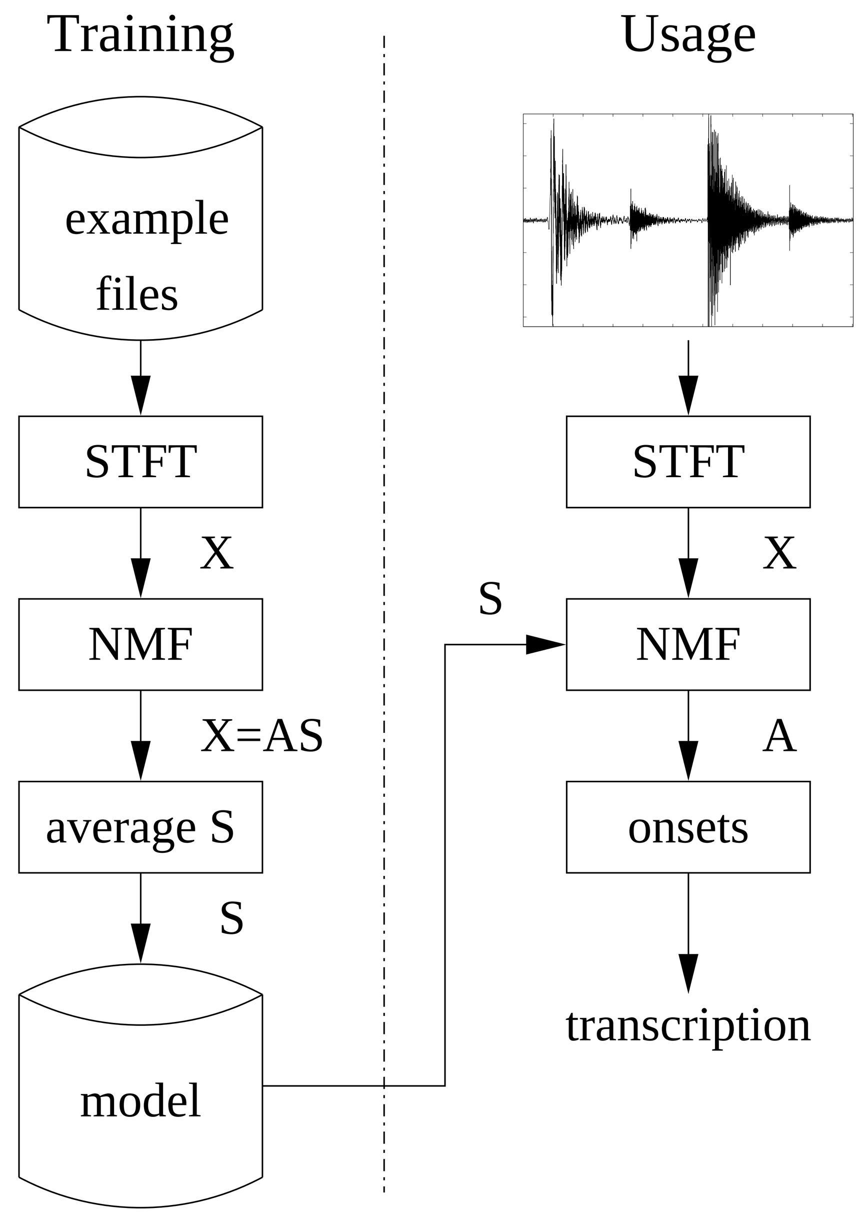

FIGURE 2: System overview.

Onset detection

- Motivated by auditory system inspired sub-band amplitude envelope method.
  - View relative difference instead of absolute.
- Spectral content fixed.
  - → Each sub-band envelope of identical form.
  - → Enough to analyse only time-varying gains a.
- Threshold value learned from training data.

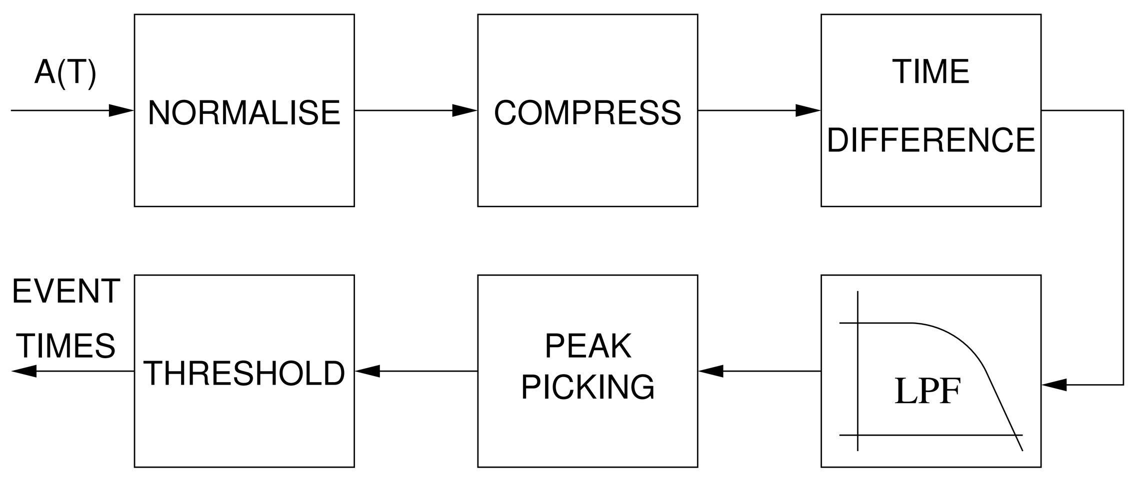

FIGURE 3: Onset detection method.

Evaluations

- System evaluated with acoustic recordings.
  - Four data sets with differring instruments and acoustic environments.
  - 4-fold cross-validations, performance metrics calculated over all folds.
- Performance compared with two other systems with same material.
  - Prior Subspace Analysis (PSA)
    - FitzGerald et al., "Prior subspace analysis for drum transcription", AES 2003.
    - Same signal model as in proposed method.
    - Solve time-varying gains with matrix inverse and ICA, onset detection with thresholding.
  - Event-based recognition
    - Gillet & Richard, "Automatic transcription of drum loops", ICASSP 2004.
    - More traditional pattern recognition approach.
    - Detect onsets from acoustic signal, segment, extract features, classify with N binary SVMs.

Results

- Calculated hit rate measure $R_h = 1 - \frac{insertions+deletions}{reference}$

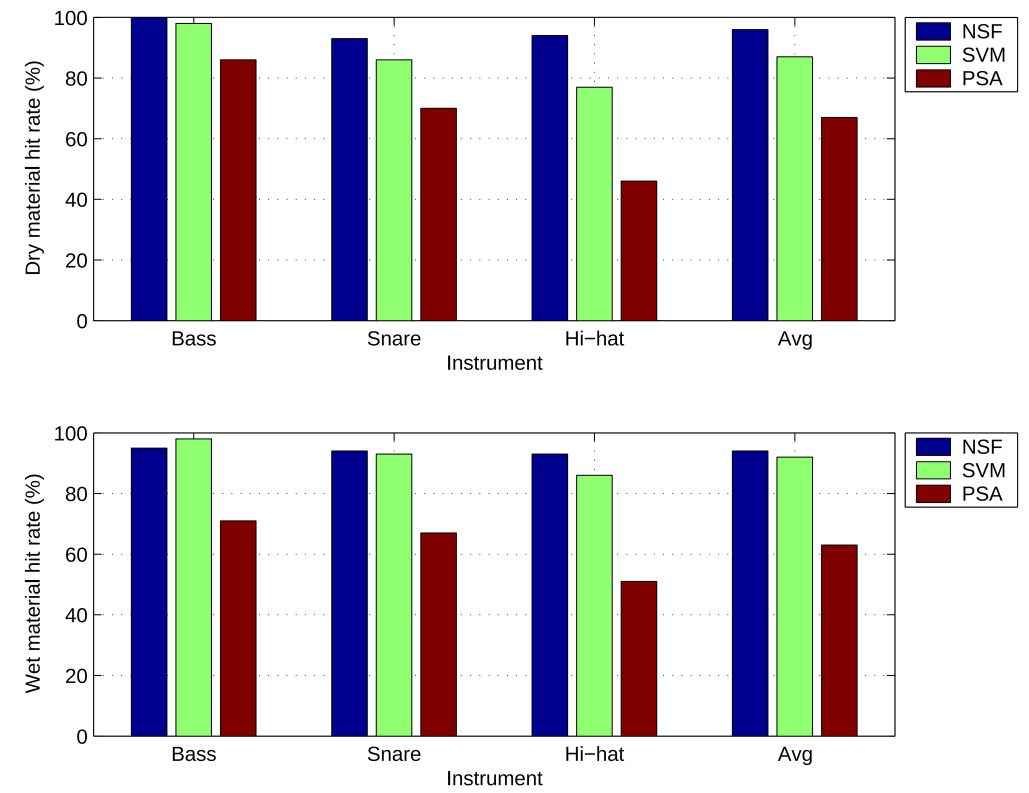

FIGURE 4: Evaluation results.

Conclusions

- Aim to transcribe signals containing bass drum, snare drum and hi-hats.
- Use fixed template spectra and estimate time-varying gains with algorithm stemming from non-negative matrix factorisation.
- Detect onsets from gains with auditory motivated method.
- When target signal and used model match, the result is good.
  - Signals containing also other instruments will result into a failure.

Demonstrational signals are available at http://www.cs.tut.fi/~paulus/research.html