

Combining Temporal And Spectral Features in HMM-based Drum Transcription

Jouni Paulus, Anssi Klapuri

Institute of Signal Processing
Tampere University of Technology
Tampere, Finland

8th International Conference on Music Information Retrieval
25.9.2006



Drum Transcription Problem

- Input: audio
 - Anything from individual drum hits to polyphonic music
- Output: symbolic representation of the drums
 - Temporal locations of drum events
 - Content of drum events (which drums were played)
- Applications
 - Symbolic information of drum content in masses of existing audio
 - Re-using drum patterns from existing audio
 - Drum replacement in audio



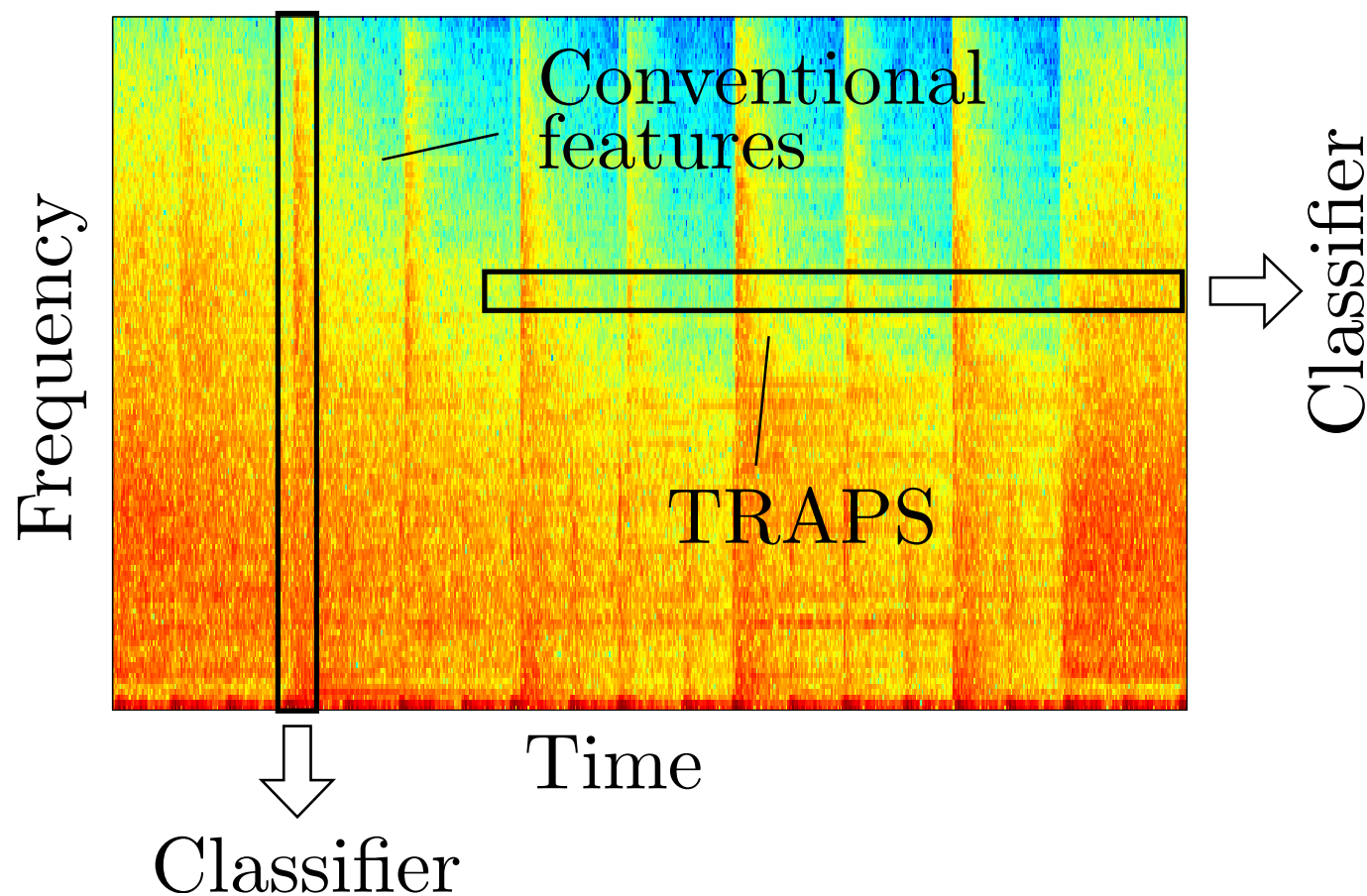
Existing Methods, some examples

- Classifiers for individual hits (Herrera et al. ICMAI 2002)
- Onset detection, classification (Gillet et al. ICASSP 2004, Tanghe et al. MIREX 2005)
- Onset detection, template adaptation recognition (Zils et al. WedelMusic 2002, Yoshii et al. ICASSP 2006)
- Onset detection, localised models (Sandvold et al. ISMIR 2004)
- Spectrogram decomposition (Virtanen ICMC 2003, FitzGerald PhD 2004, Dittmar et al. AES 2004, Paulus et al. EUSIPCO 2005)
- HMMs, no onsets (Paulus ICASSP 2006)
- Common for all: used features are from short time frames



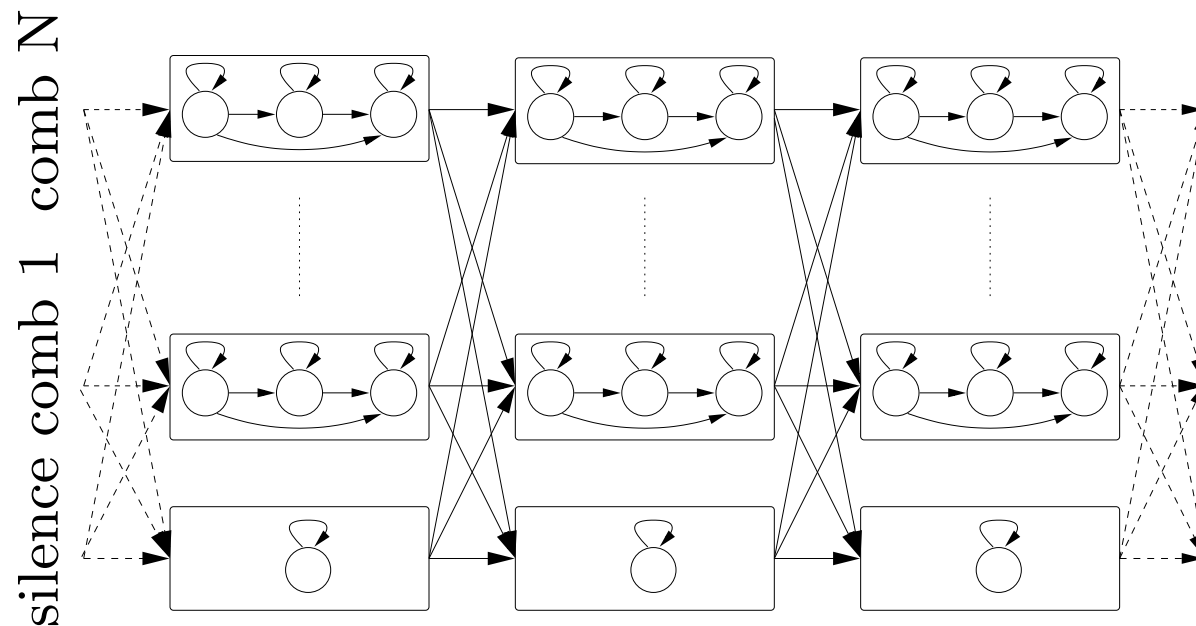
TRAPS

- TempoRAI PatternS, energy evolution on narrow subbands.
 - Human hearing bandwise.
 - Drum hits temporal events, no stationary spectrum



Base System (from ICASSP 2006)

- Model all combinations of target drums with HMMs
 - Spectral features (MFCCs etc.)
 - GMMs to model observations
- Background model when no drums are playing
- Using the models, cover the whole duration of the signal

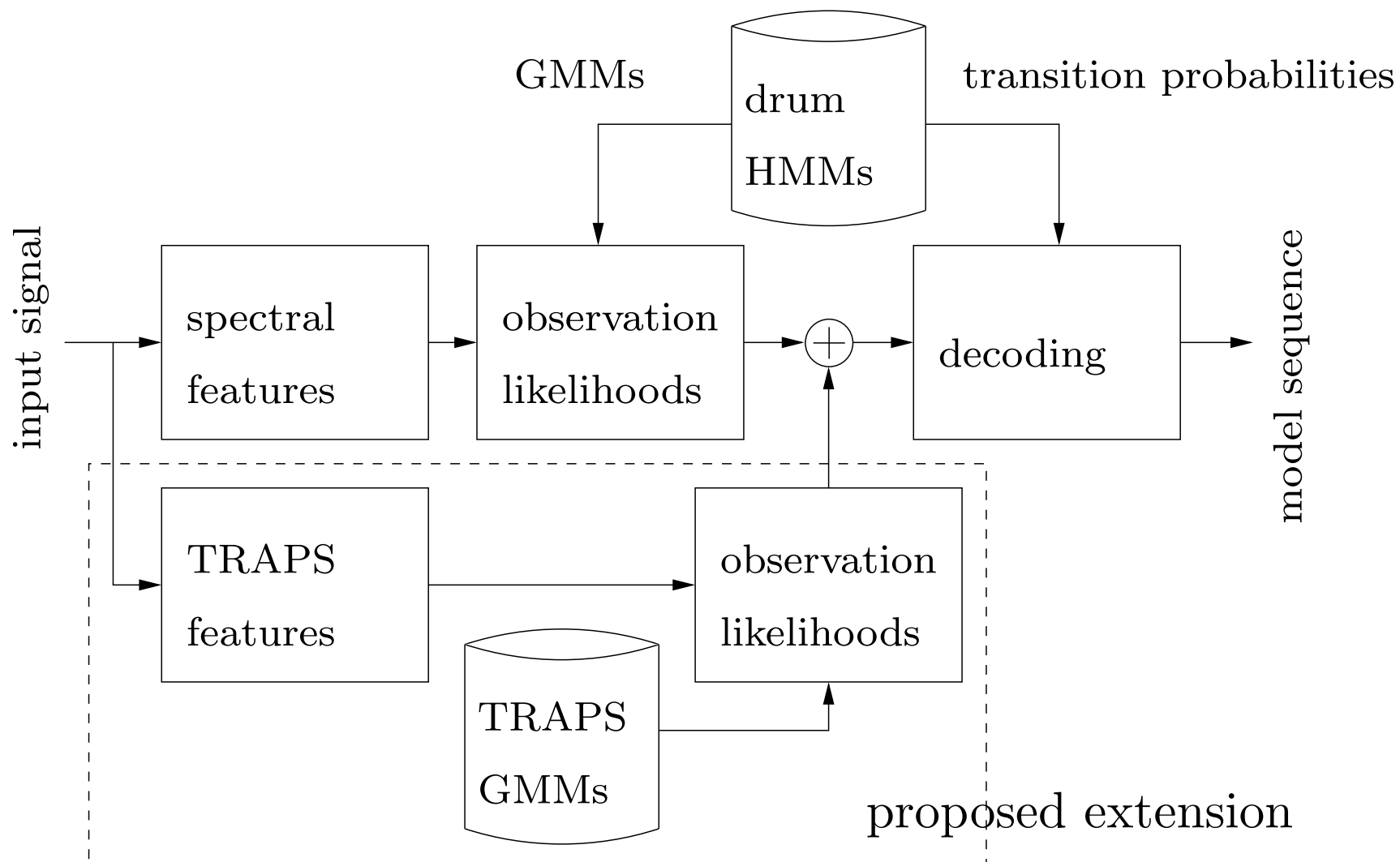


Temporal Features

- Subband envelopes
 - Bank of 1/3-octave bandpass filters,
 - Low-pass and decimate, compress, temporal differentiation
 - → Impulsive sound events visible
- Shift-invariant feature from frames of envelopes
 - Event location within frame will vary
 - Magnitude spectrum of the envelope
 - Reduce dimensionality (correlation, large amount of data)
- Combine bandwise features, train drum presence detector
GMMs for all target drums



Proposed System Block Diagram



Simulation Results

- Compare the baseline, baseline with TRAPS added, and a “detect onsets & classify” -system

F-measure (%)	simple drums	complex drums	RWC Pop
baseline HMM	93.4	84.0	66.8
HMM+TRAPS	92.9	85.2	69.7
SVM (Tanghe et al.)	85.5	76.4	65.1



Summary

- Many of earlier drum transcription systems have used only features from short frames.
- Short frames fit for stationary spectrum, drum hits are temporal events.
- Proposed incorporating long-term temporal features to HMM-based recogniser.
- The proposed addition improves results slightly.
- Demos

