AdaMast: A Drum Sound Recognizer based on Adaptation and Matching of Spectrogram Templates

Kazuyoshi Yoshii[1] Masataka Goto[2] Hiroshi G. Okuno[1]

[1]Kyoto University [2]National Institute of Advanced Industrial Science and Technology (AIST)

Objective: to detect onset times of the bass drums, snare drums, and hi-hat cymbals in polyphonic audio signals sampled from musical pieces of many genres.

Overview

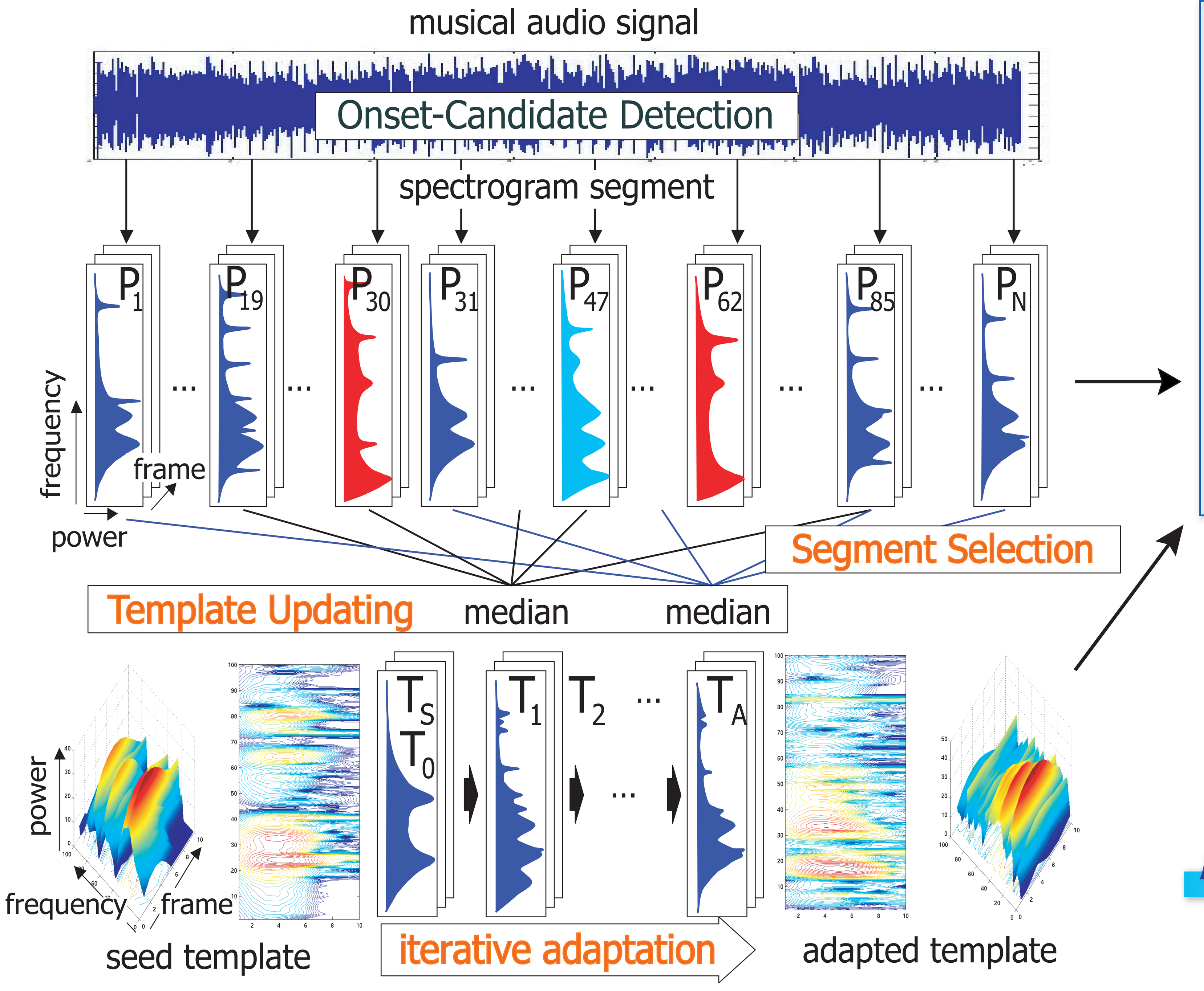
- Template-based drum sound recognition systhem: AdaMast comprises successive template-adaptation and template-matching stages, using drum-sound spectrograms as templates.
- Winner of Audio Drum Detection Contest in MIREX2005: The results of experiments using real-world musical pieces are 72.8%, 70.2%, and 57.4% in recognizing the bass drums, snare drums, and hi-hat cymbals, respectively.

Problems

- Individual Difference Problem:
  Acoustic features of drum sounds have the large variation and the appropriate templates are unknown in advance.
- Sound Mixture Problem:
  Acoustic features are distorted by the overlapping of other instrument sounds.

Approach

- The template-adaptation stage
  tries to obtain a semi-pure drum-sound spectrogram by estimating the common (major) structure of some sound-mixture spectrograms including the drum sound.
- The template-matching stage
  tries to detect all the onset times by using a distance measure which was designed to be robust to the spectral overlapping of other instrument sounds.



P_i: spectrogram segment extracted from the i-th onset-candidate time

T_g: intermediate template after the g-th adaptive iteration

Segment Selection

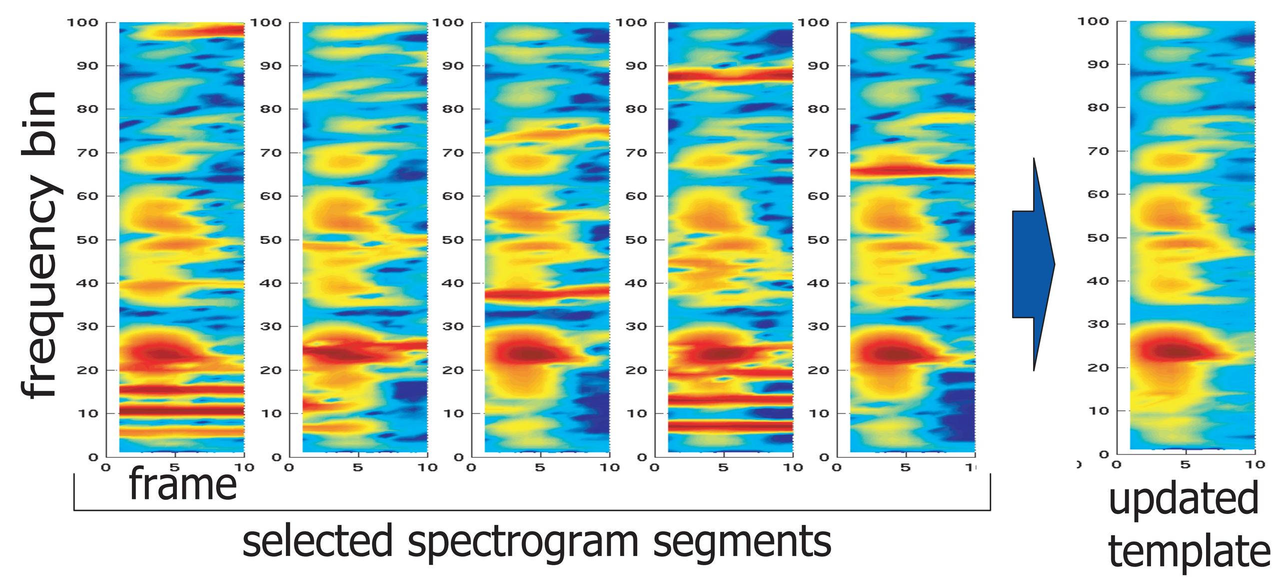
- Calculating reliability R_i that spectrogram segment P_i includes the target drum-sound spectrogram.

$$R_i = \frac{1}{\sqrt{\sum_t \sum_f \left(F_D(f)P_i(t,f) - F_D(f)T_g(t,f)\right)^2}}$$

F_D: filter function

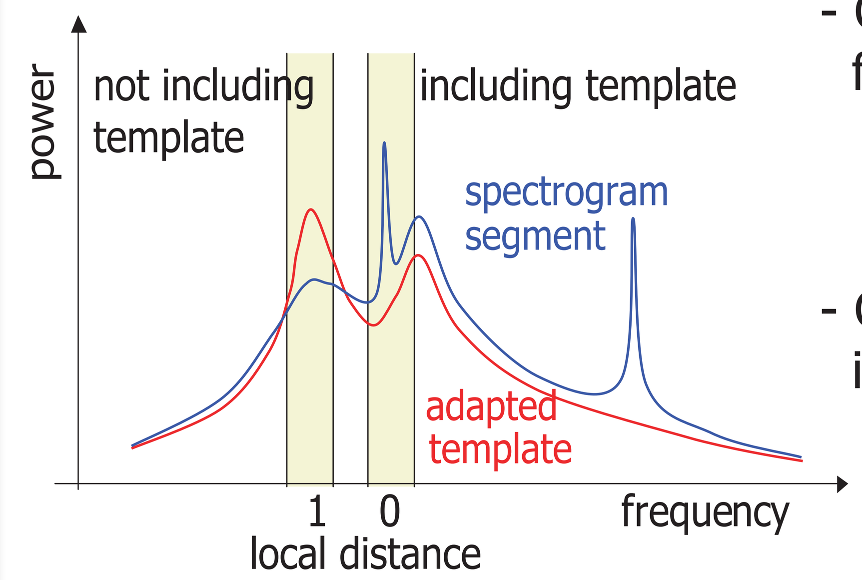
- Selecting spectrogram segments with high reliabilities.

Template Updating



- Calculating the median power at each frame and each frequency bin among the selected spectrogram segments.

Distance Calculation



- Calculating local distance γ at each frame and each frequency bin.

$$\gamma_i(t,f) = \begin{cases} 0 & \text{if } P_i(t,f) - T_A(t,f) \geq \Psi \\ 1 & \text{otherwise} \end{cases}$$

- Calculating total distance Γ by integrating γ in time-frequency domain.

$$\Gamma_i(t,f) = \sum_t \sum_f \underbrace{F_D(f)\, T_A(t,f)}_{\text{weight function}} \gamma_i(t,f)$$

Alternative Decision: yes/no judgment

Does each spectrogram segment include the adapted template ?

- This method can appropriately deal with the simultaneous sounds of multiple drum instruments.

Evaluation

Practical Testset: full CD-quality music

- 30-[s] fragments in C. Dittmar (CD) and K. Tanghe (KT) collections and entire songs in M. Goto (MG) collection (from RWC-MDB-P-2001) .
- Live and sequenced music of many genres.

Test results (F-measures)

Database	Total	BD	SD	HH
Overall	0.670	0.728	0.702	0.574
CD	0.690	0.714	0.811	0.533
KT	0.617	0.686	0.652	0.481
MG	0.716	0.776	0.710	0.661

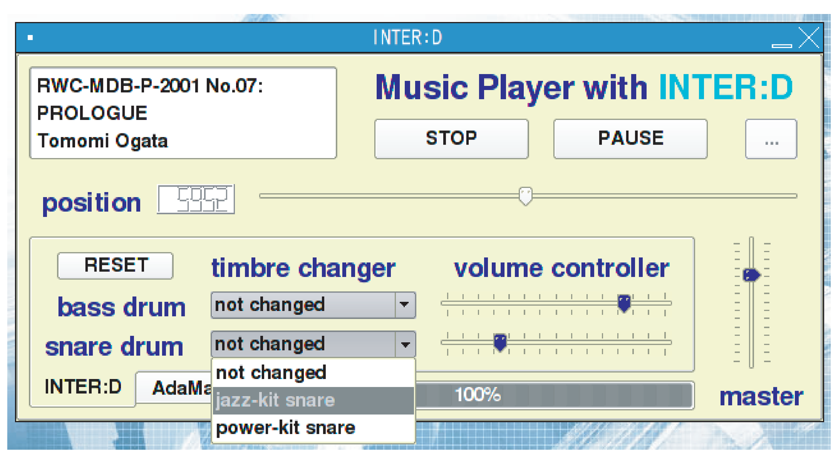
- Promising results.
- Better performance for longer musical pieces (MG).
- It is difficult to absorb the large variations of hi-hat cymbal spectrograms.

Future Work

- Dealing with musical pieces which do not include all the drum types.
- A criterion for evaluating the correctness of the adaptation.

Application

INTER:D — A Drum Sound Equalizer for Controlling Volume and Timbre of Drums



Volume Controller: Listeners can cut or boost the volume of each drum by moving its corresponding slider.

Timbre Changer: Listeners can easily switch timbres by selecting a favorite timbre from the dropdown list.