Rhythm Classification Using Spectral Rhythm Patterns

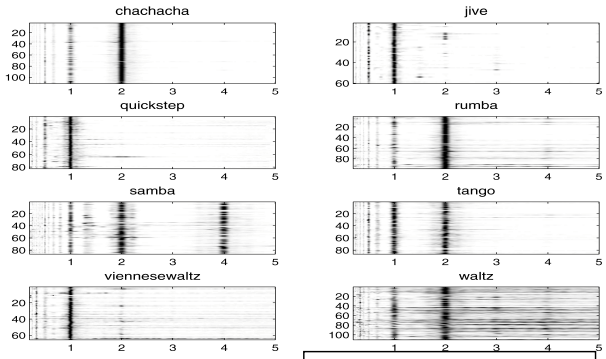
peeters@ircam.fr *IRCAM - Sound Analysis/Synthesis Team - Semantic HIFI*

Objectives

- Study the use of spectral patterns to represent the characteristics of the rhythm
- Three spectral patterns derived from the onset function
  - Discrete Fourier Transform
  - Auto-Correlation Function
  - Product of DFT and Frequency-Mapped ACF
- Evaluation for the task of rhythm classification

State of the art

- **Rhythm representations from audio signal**
  - type of information being represented,
  - how they are represented
- **Foote 2001** : beat spectrum
- **Tzanetakis 2002** : beat histogram
- **Paulus 2002** : sequence of audio features, Dynamic Time Warping
- **Gouyon 2004** : 73 features from periodicity histogram, Inter-Onset-Interval Histogram, 8 music genres of ballroom dancer database: 90,1% (ground-truth tempo), 78,9% (estimated tempo)
- **Gouyon 2004**: tempo estimation errors: 67,8 %
- **Dixon 2004**: Gouyon + temporal rhythmic patterns (energy evolution inside a bar): 96% (pattern+all features), 50% (only pattern)



Spectral rhythm patterns

Proposed method

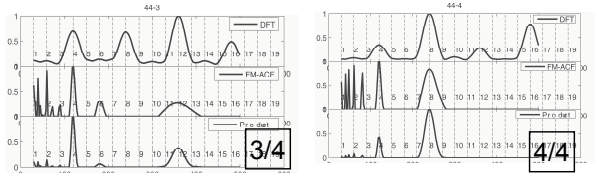
1. Tempo estimation

- **Onset energy function**: reassigned spectral energy flux (see ICMC 2005)
- **Periodicities estimation**:
  - DFT,
  - ACF,
  - product DFT and Frequency-Mapped ACF
- **Product DFT - FM-ACF ?**

  Two measures of periodicities: DFT $F(w_k,t_i)$, ACF $A(l,t_i)$ with inverse octave uncertainties -> combined both

  •1) l is mapped to the frequency domain: sr/w_k. In order to obtain the same frequencies as for the DFT: interpolation of ACF at $l=sr/w_k$

  •2) Compute the product of the DFT and ACF at each frequency w_k



- **Tempo estimation**: usage of the ground-truth tempo

2. Spectral Rhythm Patterns

- **Rhythm:** position, duration, acoustical properties
- **Here:** representation of sequence of duration
- **Sensitiveness to the sequence of duration**
  - obtained through complex DFT phase relationships
- **Independence of the tempo**
  - $Y(w_k,t_i)$ = either the DFT, ACF or DFT/FM-ACF
  - $w_{bpm}(t_i)$ = the current tempo
  - Normalized frequencies $w_k'=w_k/w_{bpm}(t_i)$ -> resampling
  - Mean of $Y(w_{k'},t_i)$ over time
  - Normalization to unit sum
- **Compactness**
  - keep only musically meaningful frequencies: 1/4, 1/3, 1/2, 2/3, 3/4, 1, 1.25, 1.5, 1.75, 2, 2.25, 2.75, 3, 3.25, 3.5, 3.75, 4
  - lower components = measure subdivision
  - upper components= beat subdivision = **Spectral Rhythm Patterns**

Evaluation: Music Genre Classification

- **Data**: Ballroom dancer database, 698 tracks, 30 s., 8 music genres (ChaChaCha, Jive, Quickstep, Rumba, Samba, Tango, VienneseWaltz, Waltz)
- **Features**:
  - DFT (18 dim.) / + tempo (19 dim.)
  - ACF (18 dim.) / + tempo (19 dim.)
  - product DFT/FM-ACF (18 dim.) / + tempo (19 dim.)
- **Classification algorithm**:
  - C4.5 decision tree algorithm,
  - Partial Decision Tree algorithm,
  - Classification via Regression algorithm
- **Results**:
  - Best classifiers: Classification via Regression
  - Best feature set: DFT
- **Comparison to the state of the art**:
  - Here: without tempo **81%,** with tempo **90,4%**
  - Gouyon 79,6%, 90,1%
  - Dixon: 50% (only pattern), 96%

	J48	PART	ClassViaReg
DFT	75,64	73,78	**80,8**
ACF	69,34	70,34	76,64
DFT/FM-ACF	65,9	65,32	75,5
DFT + tempo	90,4	88,96	**90,4**
ACF + tempo	86,67	86,67	90,25
DFT/FM-ACF + tempo	86,38	86,24	90,25
tempo	77,79	77,36	77,93

- **Confusion matrix**

classified as -->	C	J	Q	R	S	T	VW	W
ChaChaCha	**87,4%**			4,5%	0,9%	7,2%		
Jive		**86,7%**	1,7%			6,7%	5,0%	
Quickstep		1,2%	**97,6%**	1,2%				
Rumba	2,0%			**79,6%**	2,0%	3,1%		13,3%
Samba	1,2%			7,0%	**89,5%**	1,2%		1,2%
Tango	3,5%			1,2%	1,2%	**94,2%**		
Viennese Waltz		3,1%			1,5%		**95,4%**	
Waltz		0,9%		4,5%				**94,5%**

- **Best features** (CFS algorithm): 1/3, 2/3, 1, 2, 3, 3.75, 4. Recognition rate: 75.5%, 89,54%

Conclusion

- The use of simple spectral patterns allows to achieve a high recognition rate (close to the results obtained with more complex methods proposed so far)
- Future works: use estimated tempo, Evaluation on a larger set of music genres