	du	
000		

Automatic Detection of Applause in the Montreux Jazz Festival Concerts

## Patrick Marmaroli<sup>1</sup> Lucas Doméjean<sup>1</sup> Hervé Lissek<sup>1</sup> Alain Dufaux<sup>2</sup> Alexandre Delidais<sup>3</sup>

<sup>1</sup>Laboratory of Electromagnetism and Acoustics (LEMA)

<sup>2</sup>MetaMedia Center (CMM)

<sup>3</sup>Vice Presidency for Innovation and Technology Transfer (VPIV)

 $2^{nd}$  Workshop on Standards and Technologies in Multimedia Archives and Records



Methodology	Results	Conclusion



- Context
- Objectives
- Applause

## 2 Methodology

- General approach
- Example

## 8 Results

- Database
- Results
- Demo



• Conclusion, difficulties and perspectives

Introduction	Methodology	Results	Conclusion

## Section 1

## Introduction

Introduction	Methodology	Results	Conclusion
000	00	000	00
Context			

Montreux Jazz Festival in numbers :

- audio and video recordings since 1966,
- 5000 hours of audio and video,
- 15 different recording formats.



Introduction	Methodology	Results	Conclusion
000	00	000	00
Objectives			

Montreux Jazz Digital Project @ MetaMedia Center :

- Archive digitization,
- Preservation and perpetuation of the archives,
- Valorization of the archives.

Introduction	Methodology	Results	Conclusion
000	00	000	00
Objectives			

## Montreux Jazz Digital Project @ MetaMedia Center :

- Archive digitization,
- Preservation and perpetuation of the archives,
- Valorization of the archives.







Introduction O • O	Methodology OO	Results 000	Conclusion OO
Objectives			
Montreux Jazz I • Archive digit • Preservation • Valorization	<b>Digital Project</b> @ MetaMo ization, and perpetuation of the a of the archives.	edia Center : archives,	
LEMA : • applause, • speech		digitalizing t	he full archives
<ul> <li>speech,</li> <li>audio cuts,</li> <li>pops and clic</li> <li>clipping.</li> </ul>	:ks,		1
		metadata inclusi audio signal prod	on through digital cessing techniques

Introduction O • O	Methodology OO	Results 000	Conclusion OO
Objectives			
Montreux Jazz <ul> <li>Archive dig</li> <li>Preservation</li> </ul>	<b>Digital Project</b> @ MetaMe itization, n and perpetuation of the a	edia Center :	
<ul> <li>Valorization</li> </ul>	of the archives.		Carlo
		digitalizing t	he full archives
LEMA :			
<ul> <li>applause,</li> </ul>			
<ul> <li>speech,</li> </ul>			
<ul> <li>audio cuts,</li> </ul>			1
<ul> <li>pops and cl</li> </ul>	icks,		
• clipping.		↓	
		metadata inclus audio signal pro	ion through digital cessing techniques

Introduction	Methodology	Results	Conclusion
000	00	000	00
Interests of app	lause sounds		



 $1^{\rm h} \, 30$ 

Introduction	Methodology OO	Results 000	Conclusion OO
Interests of ap	plause sounds		
	How many tracks	? Where are they ?	
		le par la participa de la particip	

 $1^{h}30$ 

Applause sound position : -> help-to-decision for automatic/manual track partitioning.

Introductio	n
000	

## Interests of applause sounds



Methodology	Results	Conclusion

## Section 2

## Methodology

	Methodology	Results	Conclusion
000	0	000	00
Frame by frame	analysis		



Applause or Music?

	Methodology	Results	Conclusion
000	•0	000	00
Frame by fram	e analysis		



## Binary classification problem

What is required :

- Audio features (spectral, temporal, spectro-temporal,...)
- Classifier (SVM, GMM, Decision Tree, Neural Network,...)

Introduction	Methodology	Results	Conclusion
000	•0	000	00
Frame by fram	e analysis		



### Binary classification problem

What is required :

- Audio features (spectral, temporal, spectro-temporal,...)
- Classifier (SVM, GMM, Decision Tree, Neural Network,...)

## Working philosophy :

- The least features as possible,
- The simplest classifier as possible

Introduction	Methodology	Results	Conclusion
000	•0	000	00
Frame by fram	e analysis		



### Binary classification problem

What is required :

- Audio features (spectral, temporal, spectro-temporal,...)
- Classifier (SVM, GMM, Decision Tree, Neural Network,...)

Working philosophy :

- The least features as possible,
- The simplest classifier as possible

How?: Kullback-Leibler divergence-based features optimization

Introduction	Methodology	Results	Conclusion
000	00	000	00
Example			
Spectral gravity center (SGC, in Hz)			
fmar			

$$SGC[q] = \frac{\sum_{f_{min}}^{f_{max}} f |\mathbf{Y}_q(f)|^2}{\sum_{f_{min}}^{f_{max}} |\mathbf{Y}_q(f)|^2},$$

- q : frame number,
- f : frequency (in Hz),
- $\mathbf{Y}_q(f)$  : Fourier transform of the  $q^{th}$  frame.

Introduction	Methodology	Results	Conclusion
000	00	000	00
Example			
Spectral gravity	center (SGC in Hz)		

$$SGC[q] = \frac{\sum_{f_{min}}^{f_{max}} f |\mathbf{Y}_q(f)|^2}{\sum_{f_{min}}^{f_{max}} |\mathbf{Y}_q(f)|^2},$$

standard definition

• 
$$f_{max} = f_s/2$$
 Hz.

- q : frame number,
- f : frequency (in Hz),
- $\mathbf{Y}_q(f)$  : Fourier transform of the  $q^{th}$  frame.

OOO	Methodology ○●	Results 000	OO
Example			
Spectral gravit	y center (SGC, in Hz)		
$SGC[q] = \frac{\sum_{f_{min}}}{\sum_{f_{min}}}$	$\sum_{i=1}^{n} \frac{\left \mathbf{Y}_{q}(f)\right ^{2}}{\left \mathbf{Y}_{q}(f)\right ^{2}},$	<ul> <li>q : frame number,</li> <li>f : frequency (in Hz),</li> <li>Y<sub>q</sub>(f) : Fourier transform of the q<sup>th</sup> f</li> <li>Music — Applause</li> </ul>	rame.
standard defini • $f_{min} = 0$ • $f_{max} = f_s$	ition Hz, ./2 Hz.	Normalization of the second se	

00

 $f_s/2$ 

Frequency [Hz]

0	00 00	000	00
E	xample		
	Spectral gravity center (SGC, in Hz)	1	
	$\sum_{r=1}^{f_{max}} f  \mathbf{Y}_{c}(f) ^{2}$	• a : frame number	

 $SGC[q] = \frac{\overbrace{f_{min}}^{f_{min}}}{\sum_{j=1}^{f_{max}} \left|\mathbf{Y}_q(f)\right|^2}$ 

Methodology

standard definition

• 
$$f_{max} = f_s/2$$
 Hz.

 $f_{min}$ 

## fine-tuning

• 
$$f_{min} = f_1^{opt}$$
 Hz,  
•  $f_{max} = f_2^{opt}$  Hz.

- *f* : frequency (in Hz),
- $\mathbf{Y}_q(f)$  : Fourier transform of the  $q^{th}$  frame.



Introduction 000	Methodology ○ ●	Results 000	Conclusion OO
Example			
Spectral grav	vity center (SGC, in Hz)		
$SGC[q] = \frac{f}{f}$	$\sum_{\substack{min\\f_{max}\\f_{max}}}^{max} f \mathbf{Y}_q(f) ^2,$	<ul> <li>q : frame number,</li> <li>f : frequency (in Hz),</li> <li>Y<sub>q</sub>(f) : Fourier transform of the</li></ul>	$e \ q^{th}$ frame.
standard def • $f_{min} =$	inition 0 Hz,	KL div. = 1.17	

• 
$$f_{max} = f_s/2$$
 Hz.

• 
$$f_{min} = f_1^{opt}$$
 Hz,  
•  $f_{max} = f_2^{opt}$  Hz.



Methodology	Results	Conclusion

## Section 3

## Results

	Methodology	Results	Conclusion
000	00	000	00
Database			

Style	Length	Size
Нір Нор	39 min	440 Mo
Cuban music	1 h 08 min	766 Mo
Soul / Blues	1 h 02 min	709 Mo
Reggae	1 h 20 min	908 Mo
Salsa	1 h 12 min	813 Mo
Funk	1 h 32 min	702 Mo
Jazz / Bossa Nova	1 h 02 min	702 Mo
Experimental music	1 h 34 min	1064 Mo
Jazz	1 H 53 min	1271 Mo
Blues	45 min	515 Mo

	Methodology	Results	Conclusion
000	00	000	00
Database			

Γ	Style	Length	Size
	Нір Нор	39 min	440 Mo
	Cuban music	1 h 08 min	766 Mo
	Soul / Blues	1 h 02 min	709 Mo
	Reggae	1 h 20 min	908 Mo
	Salsa	1 h 12 min	813 Mo
	Funk	1 h 32 min	702 Mo
	Jazz / Bossa Nova	1 h 02 min	702 Mo
Experimental music		1 h 34 min	1064 Mo
Jazz		1 H 53 min	1271 Mo
Blues		45 min	515 Mo

	Methodology	Results	Conclusion
000	00	000	00
Database			

Style	Length	Size
Нір Нор	39 min	440 Mo
Cuban music	1 h 08 min	766 Mo
Soul / Blues	1 h 02 min	709 Mo
Reggae	1 h 20 min	908 Mo
Salsa	1 h 12 min	813 Mo
Funk	1 h 32 min	702 Mo
Jazz / Bossa Nova	1 h 02 min	702 Mo
Experimental music	1 h 34 min	1064 Mo
Jazz	1 H 53 min	1271 Mo
Blues	45 min	515 Mo

## Test database

	Methodology	Results	Conclusion
000	00	000	00
Database			

Style	Length	Size
Нір Нор	39 min	440 Mo
Cuban music	1 h 08 min	766 Mo
Soul / Blues	1 h 02 min	709 Mo
Reggae	1 h 20 min	908 Mo
Salsa	1 h 12 min	813 Mo
Funk	1 h 32 min	702 Mo
Jazz / Bossa Nova	1 h 02 min	702 Mo
Experimental music	1 h 34 min	1064 Mo
Jazz	1 H 53 min	1271 Mo
Blues	45 min	515 Mo

## Test database

#### **Training database**

Frame size (in samples / in ms)	2048 / 42
Frames «applause»	12 757
Frames «music»	399 500

	Methodology	Results	Conclusion
000	00	000	00
Database			

Style	Length	Size
Нір Нор	39 min	440 Mo
Cuban music	1 h 08 min	766 Mo
Soul / Blues	1 h 02 min	709 Mo
Reggae	1 h 20 min	908 Mo
Salsa	1 h 12 min	813 Mo
Funk	1 h 32 min	702 Mo
Jazz / Bossa Nova	1 h 02 min	702 Mo
Experimental music	1 h 34 min	1064 Mo
Jazz	1 H 53 min	1271 Mo
Blues	45 min	515 Mo

## Test database

#### **Training database**

Frame size (in samples / in ms)	2048 / 42
Frames «applause»	12 757
Frames «music»	399 500

#### Test database

Frame size (in samples / in ms)	2048 / 42
Frames «applause»	29 288
Frames «music»	425 516

	Methodology		Results	Conclusion
000	00		000	00
Classification	results			
	Spectral Gravity Center	True Detection: 94 False Alarm: 3%	1%	









	Methodology	Results	Conclusion
000	00	000	00
Video demo			













START STOP

Methodology	Results	Conclusion

## Section 4

## Conclusion

Introduction	Methodology	Results	Conclusion
000	00	000	•0
Conclusion			

- light and robust applause sound detection system,
- KL-based optimization features,
- implemented on a real-time Matlab environment.

Introduction	Methodology	Results	Conclusion
000	00	000	0
Conclusion			

- light and robust applause sound detection system,
- KL-based optimization features,
- implemented on a real-time Matlab environment.

Remaining difficulties

	Methodology	Results	Conclusion
000	00	000	•0
Conclusion			

- light and robust applause sound detection system,
- KL-based optimization features,
- implemented on a real-time Matlab environment.

## Remaining difficulties

applause within song (break,...),

	Methodology	Results	Conclusion
000	00	000	•0
Conclusion			

- light and robust applause sound detection system,
- KL-based optimization features,
- implemented on a real-time Matlab environment.

## Remaining difficulties

applause within song (break,...), brutal changes of songs (no applause),

	Methodology	Results	Conclusion
000	00	000	•0
Conclusion			

- light and robust applause sound detection system,
- KL-based optimization features,
- implemented on a real-time Matlab environment.

## Remaining difficulties

applause within song (break,...), brutal changes of songs (no applause), speech + applause and other events.

	Methodology	Results	Conclusion
000	00	000	•0
Conclusion			

- light and robust applause sound detection system,
- KL-based optimization features,
- implemented on a real-time Matlab environment.

## Remaining difficulties

### Perspectives

applause within song (break,...), brutal changes of songs (no applause), speech + applause and other events.

Introduction	Methodology	Results	Conclusion
000	00	000	0
Conclusion			

- light and robust applause sound detection system,
- KL-based optimization features,
- implemented on a real-time Matlab environment.

## Remaining difficulties

### Perspectives

applause within song (break,...), brutal changes of songs (no applause), speech + applause and other events. song difference analyser,

	Methodology	Results	Conclusion
000	00	000	•0
Conclusion			

- light and robust applause sound detection system,
- KL-based optimization features,
- implemented on a real-time Matlab environment.

## Remaining difficulties

### Perspectives

applause within song (break,...), brutal changes of songs (no applause), speech + applause and other events. song difference analyser,

	Methodology	Results	Conclusion
000	00	000	•0
Conclusion			

- light and robust applause sound detection system,
- KL-based optimization features,
- implemented on a real-time Matlab environment.

## Remaining difficulties

### Perspectives

applause within song (break,...), brutal changes of songs (no applause), speech + applause and other events. song difference analyser,

investigate more complex classifiers.

Methodology	Results	Conclusion
		00

# Thanks for your attention !