Automatic Detection of Applause in the Montreux Jazz Festival Concerts

Patrick Marmaroli[1] Lucas Doméjean[1] Hervé Lissek[1]
Alain Dufaux[2] Alexandre Delidais[3]

[1] Laboratory of Electromagnetism and Acoustics (LEMA)

[2] MetaMedia Center (CMM)

[3] Vice Presidency for Innovation and Technology Transfer (VPIV)

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

Section 1

Introduction

Context

Montreux Jazz Festival in numbers :

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

Objectives

Montreux Jazz Digital Project @ MetaMedia Center :

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

Objectives

Montreux Jazz Digital Project @ MetaMedia Center :

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

digitalizing the full archives

Objectives

Montreux Jazz Digital Project @ MetaMedia Center :

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

digitalizing the full archives

LEMA :

- applause,
- speech,
- audio cuts,
- pops and clicks,
- clipping.

metadata inclusion through digital audio signal processing techniques

Objectives

Montreux Jazz Digital Project @ MetaMedia Center :

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

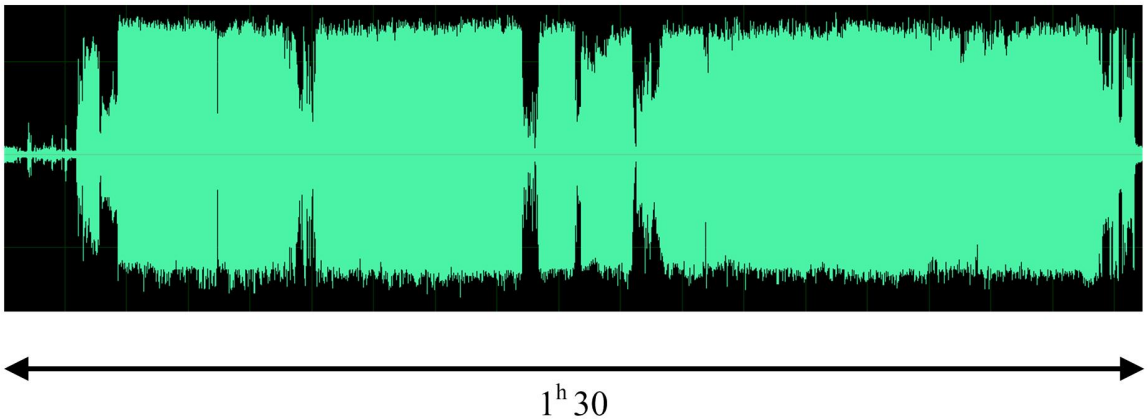
digitalizing the full archives

LEMA :

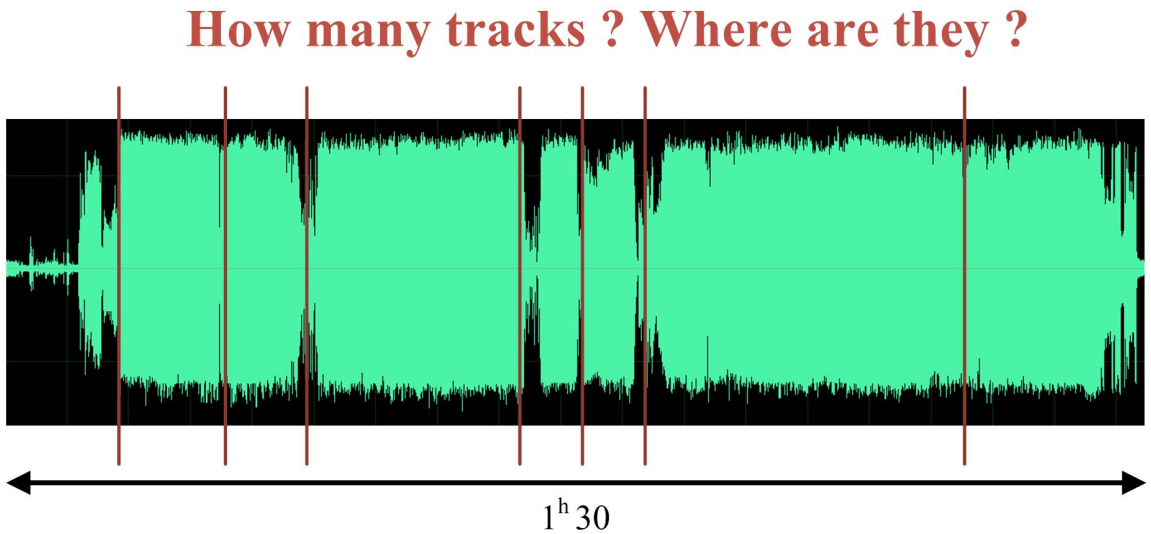
- **applause**,
- speech,
- audio cuts,
- pops and clicks,
- clipping.

metadata inclusion through digital audio signal processing techniques

Interests of applause sounds

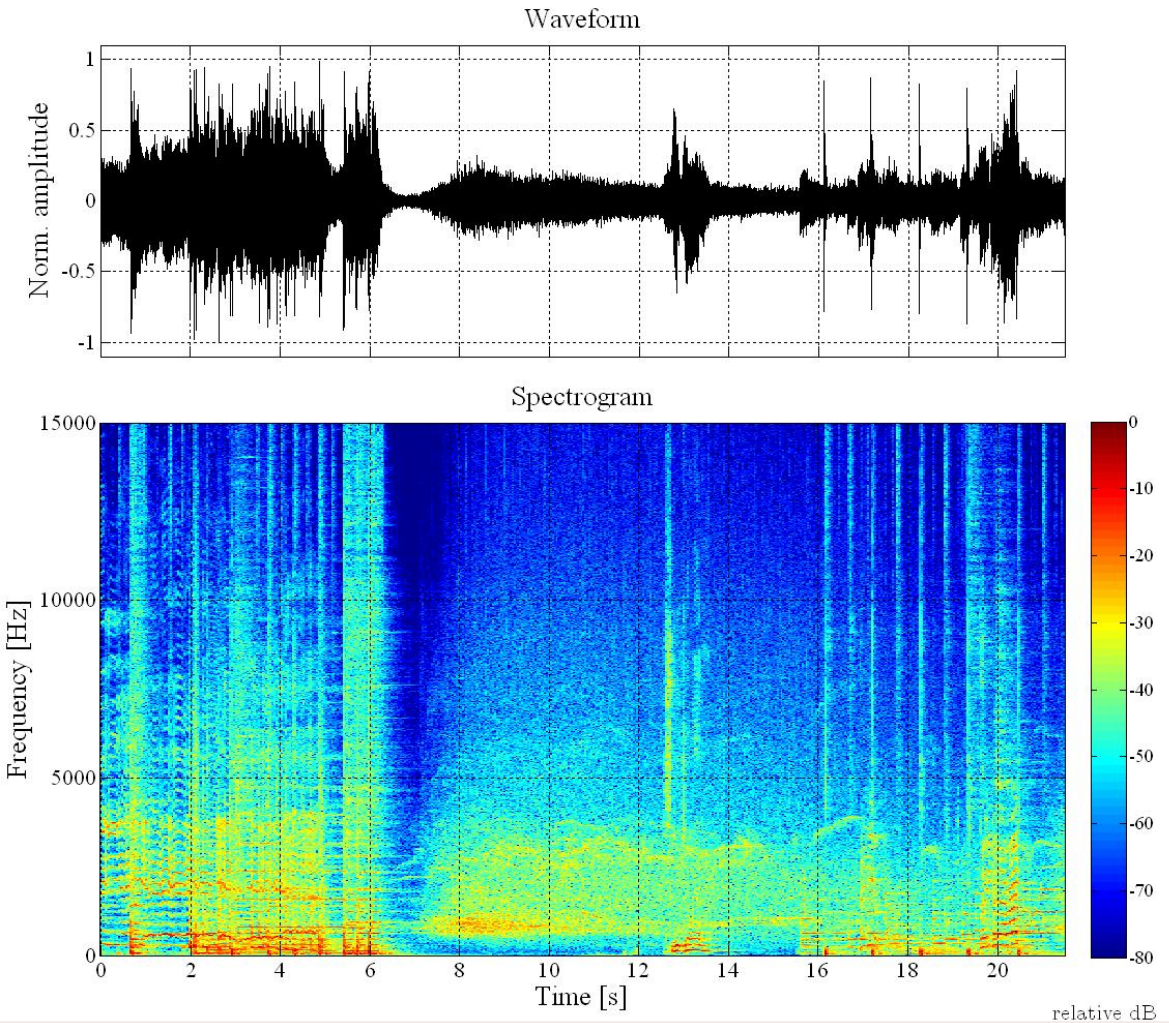


Interests of applause sounds



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

Interests of applause sounds

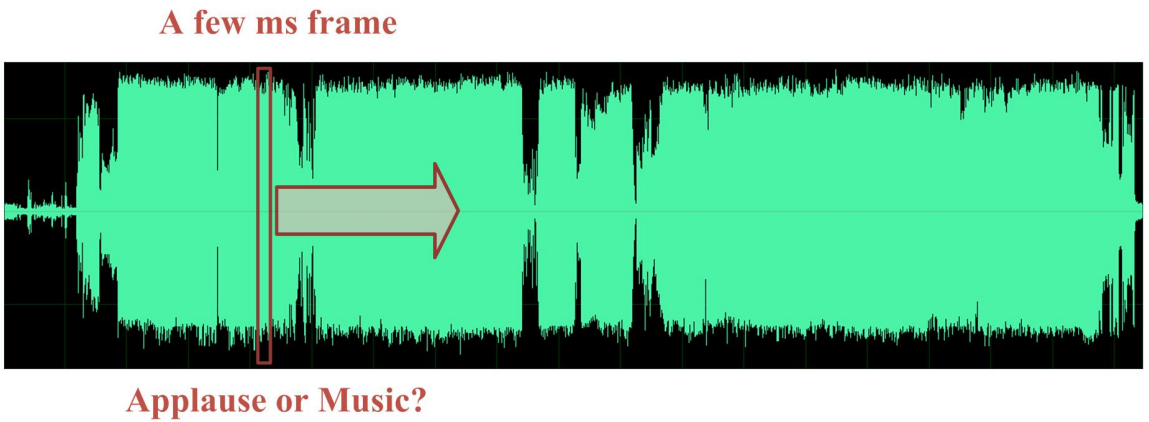


Applause sound position :
-> help-to-decision for automatic/manual track partitioning.

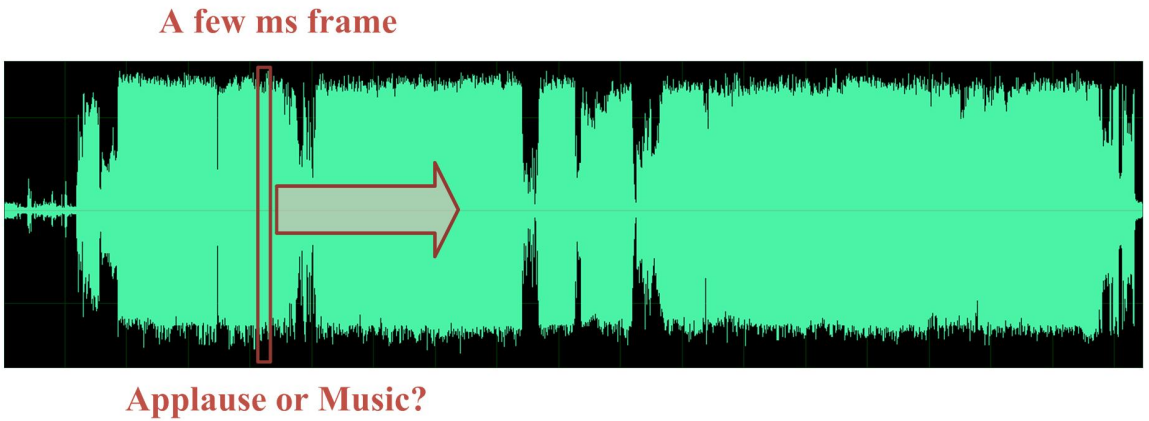
Section 2

Methodology

Frame by frame analysis



Frame by frame analysis

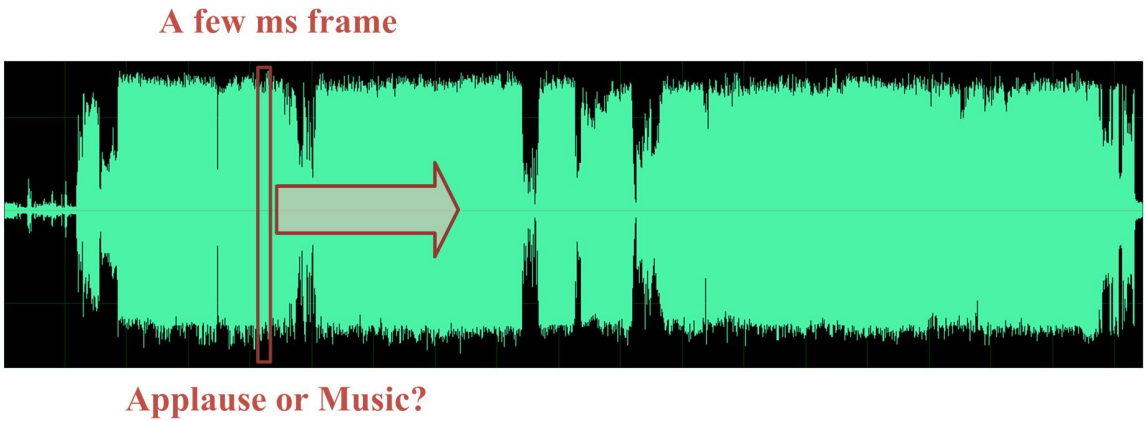


Binary classification problem

What is required :

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

Frame by frame analysis



Binary classification problem

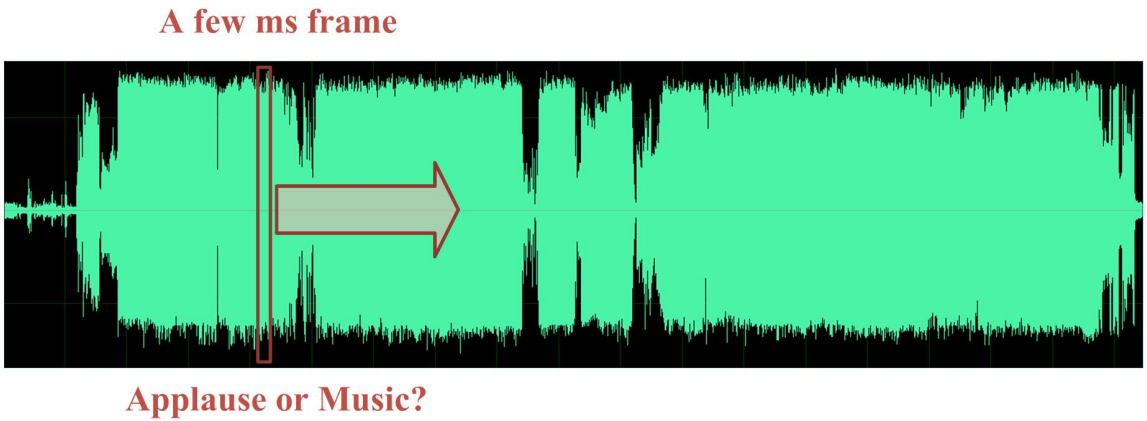
What is required :

- 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

Frame by frame 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

Example

Spectral gravity center (SGC, in Hz)

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

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

Example

Spectral gravity center (SGC, in Hz)

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

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

standard definition

- f_{min}= 0 Hz,
- f_{max}= $f_s/2$ Hz.

Example

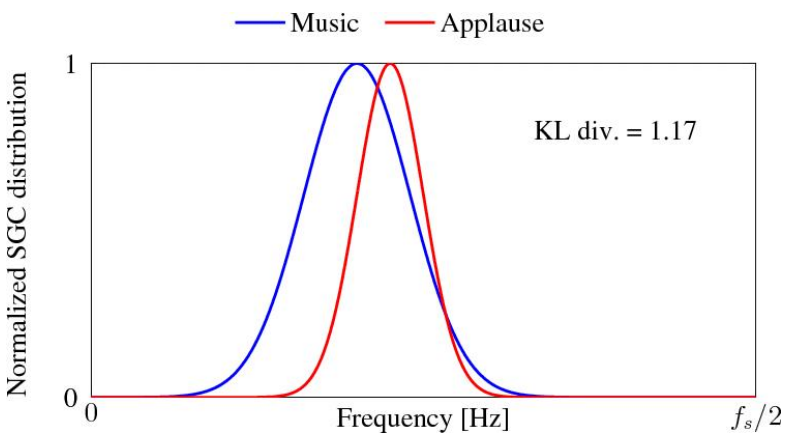
Spectral gravity center (SGC, in Hz)

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

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

standard definition

- f_{min}= 0 Hz,
- f_{max}= $f_s/2$ Hz.

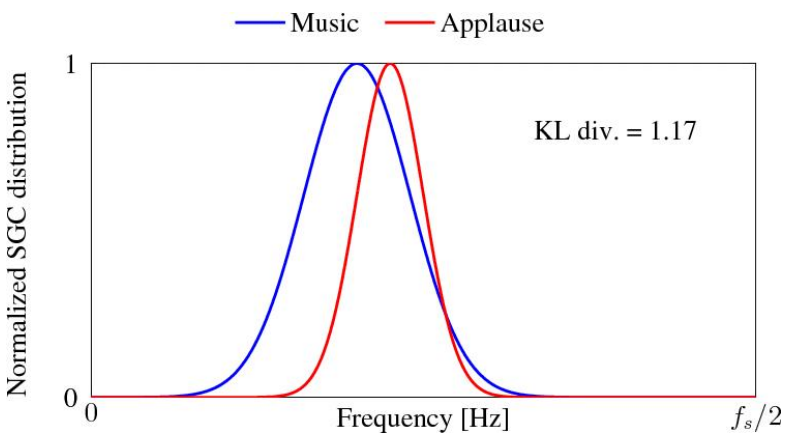


Example

Spectral gravity center (SGC, in Hz)

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

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



standard definition

- f_{min}= 0 Hz,
- f_{max}= $f_s/2$ Hz.

fine-tuning

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

Example

Spectral gravity center (SGC, in Hz)

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

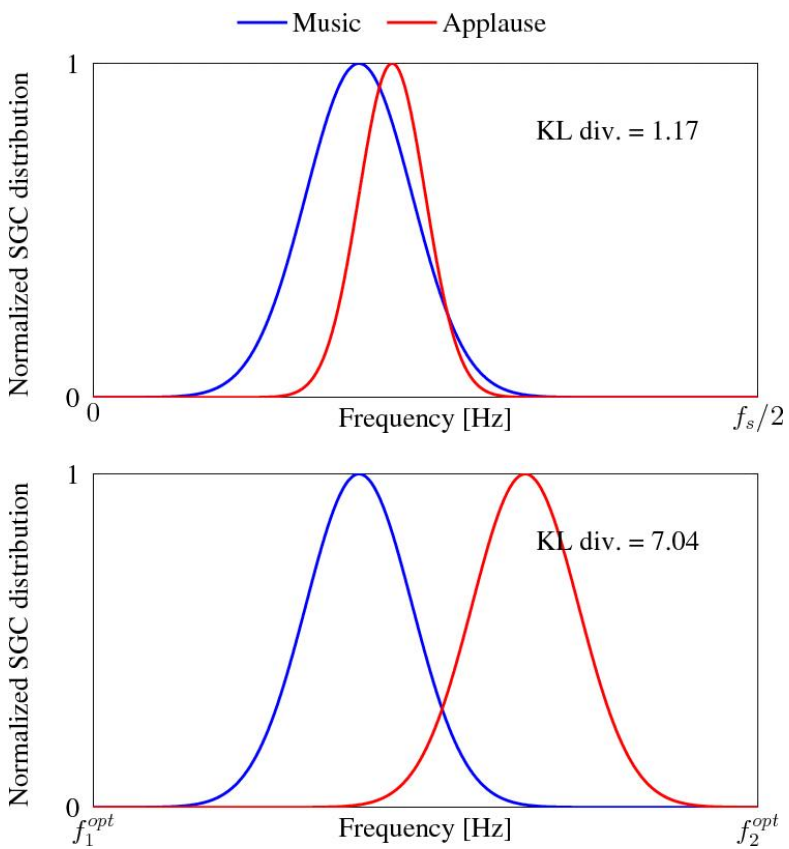
- q : frame number,
- f : frequency (in Hz),
- $\mathbf{Y}_q(f)$: Fourier transform of the q^{th} frame.

standard definition

- f_{min}= 0 Hz,
- f_{max}= $f_s/2$ Hz.

fine-tuning

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



Section 3

Results

Database

Style	Length	Size
Hip Hop	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

Database

Training database

Style	Length	Size
Hip Hop	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

Database

Training database

Style	Length	Size
Hip Hop	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

Database

Training database

Style	Length	Size
Hip Hop	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**

Database

Training database

Style	Length	Size
Hip Hop	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**

Classification results

Spectral Gravity Center

True Detection: 94%
False Alarm: 3%

Classification results

Spectral Gravity Center

True Detection: 94%
False Alarm: 3%

How increase it ?

Addition of a second feature

Classification results

Spectral Gravity Center

True Detection: 94%
False Alarm: 3%

How increase it ?

Addition of a second feature

Spectral Gravity Center
+
Spectral Roll-Off

True Detection: 98%
False Alarm: 4%

Classification results

Spectral Gravity Center

True Detection: 94%
False Alarm: 3%

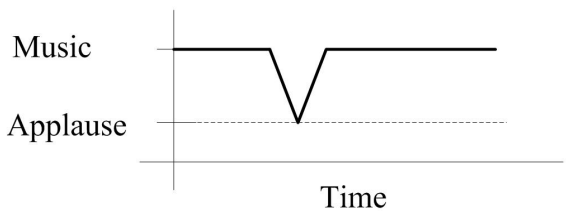
How increase it ?

Addition of a second feature

Spectral Gravity Center + Spectral Roll-Off

True Detection: 98%
False Alarm: 4%

How decrease it ?



Suppress artefacts

Classification results

Spectral Gravity Center

True Detection: 94%
False Alarm: 3%

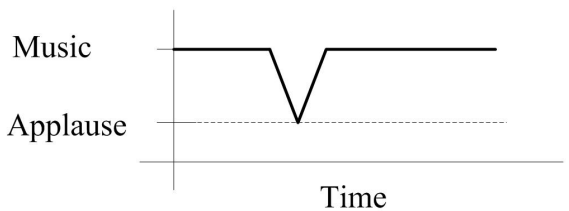
How increase it ?

Addition of a second feature

Spectral Gravity Center + Spectral Roll-Off

True Detection: 98%
False Alarm: 4%

How decrease it ?

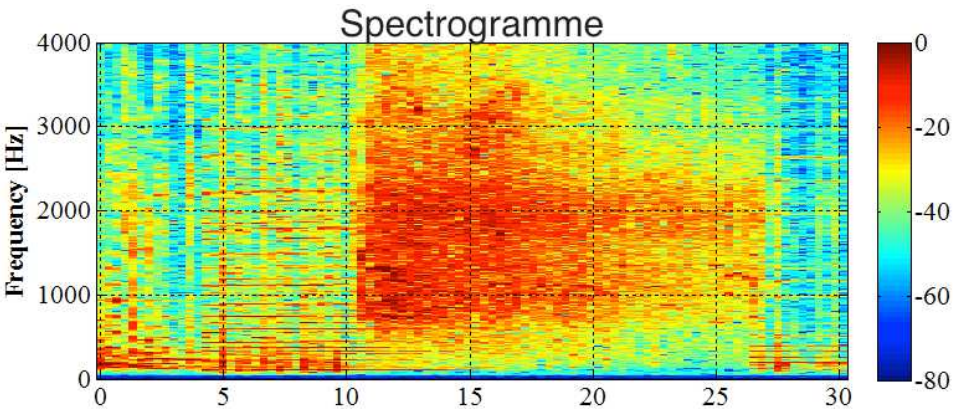


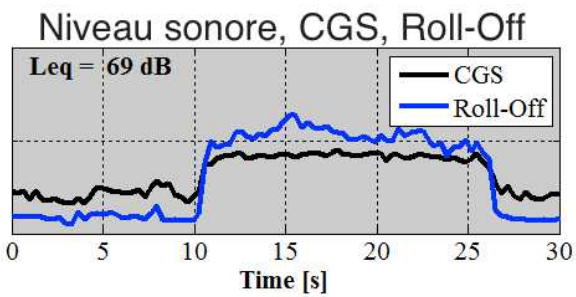
Suppress artefacts

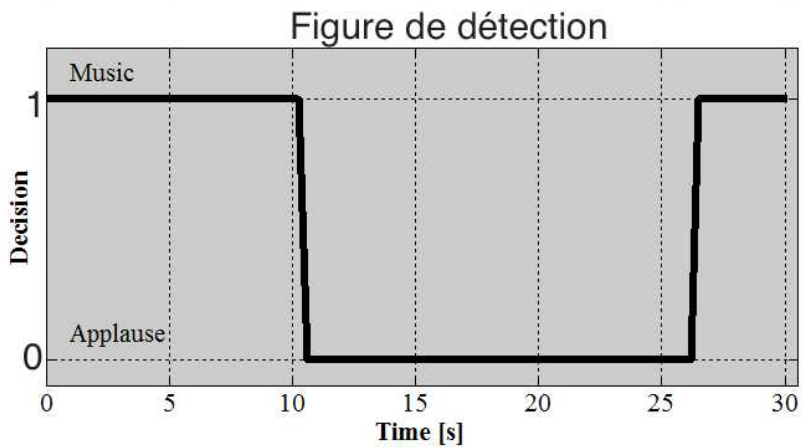
Spectral Gravity Center + Spectral Roll-Off + Temporal consistency

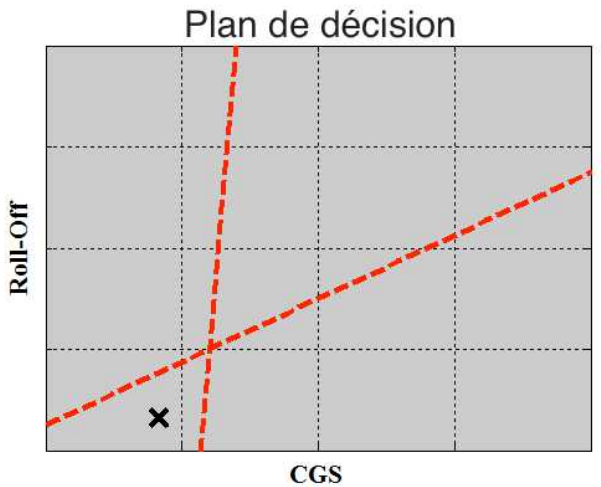
True Detection: **99%**
False Alarm: **2%**

Video demo









START STOP

Section 4

Conclusion

Conclusion

Done :

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

Conclusion

Done :

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

Remaining difficulties

Conclusion

Done :

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

Remaining difficulties

applause within song (break,...),

Conclusion

Done :

- 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),

Conclusion

Done :

- 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.

Conclusion

Done :

- 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.

Perspectives

Conclusion

Done :

- 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.

Perspectives

song difference analyser,

Conclusion

Done :

- 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.

Perspectives

song difference analyser,
" "

Conclusion

Done :

- 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.

Perspectives

song difference analyser,
" "
investigate more complex classifiers.

Thanks for your attention !