Two-Step Modal Identification for Increased Resolution Analysis of Percussive Sounds
(p.117)

Mathieu LAGRANGE, **Bertrand SCHERRER**

SPCL – McGill University
CIRMMT

DAFx'08 – Sept. 2nd, 2008

SPCL
CIRMMT Centre for Interdisciplinary Research in Music Media and Technology

Outline

SPCL

CIRMMT Centre for Interdisciplinary Research in Music Media and Technology

What for ?

Tool for the **modal analysis** of **percussive sounds** for:

What for ?

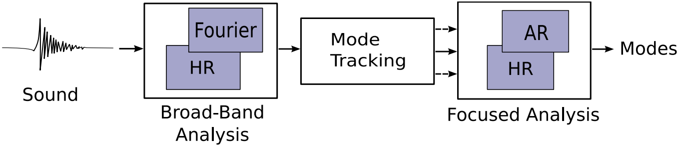
Tool for the **modal analysis** of **percussive sounds** for:

- musical instrument modeling
- synthesize impact sounds between objects for virtual environments
- ...

SPCL

CIRMMT Centre for Interdisciplinary Research in Music Media and Technology

2 steps ?

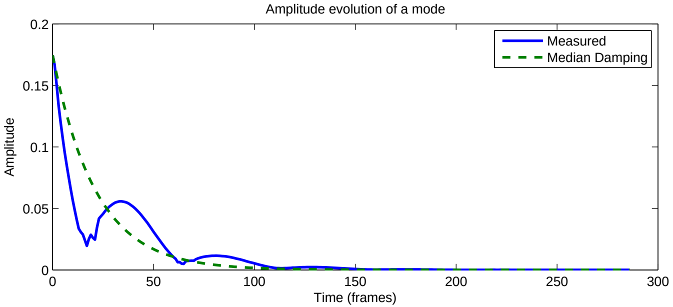


Two-step modal parameters estimation:

1. **Broad-band analysis** ⇒ global spectro-temporal structure of the sound (with limited frequency resolution)

Centre for Interdisciplinary Research in Music Media and Technology

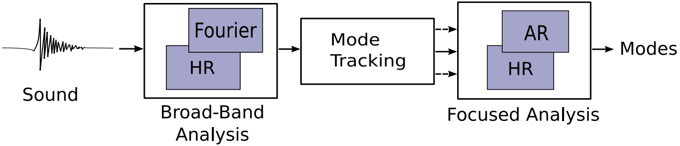
2 steps ?



SPCL

CIRMMT Centre for Interdisciplinary Research in Music Media and Technology

2 steps ?



Two-step modal parameters estimation:

1. **Broad-band analysis** ⇒ global spectro-temporal structure of the sound (with limited frequency resolution)
2. **Focused analysis** ⇒ better modeling of each component identified during the first "pass"

SPCL

CIRMMT Centre for Interdisciplinary Research in Music Media and Technology

Interest

- **Flexible method**:

Interest

- **Flexible method**:
  - **a lot of** *a priori* knowledge

Interest

- **Flexible method**:
  - **a lot of** *a priori* knowledge ⇒ physical model parameter estimation

Interest

- **Flexible method**:
  - **a lot of** *a priori* knowledge ⇒ physical model parameter estimation
  - **not much** *a priori* knowledge

SPCL

CIRMMT Centre for Interdisciplinary Research in Music Media and Technology

Interest

- **Flexible method**:
  - **a lot of** *a priori* knowledge ⇒ physical model parameter estimation
  - **not much** *a priori* knowledge ⇒ “realistic enough” sound synthesis

Sound Model

We assume that:

- the signal is made of a **sum of exponentially decaying cisoids**

Sound Model

We assume that:

- the signal is made of a **sum of exponentially decaying cisoids**
- the physical properties of the impacted object remain constant $\Rightarrow$ **mode frequencies and damping factors are constant**

SPCL

CIRMMT Centre for Interdisciplinary Research in Music Media and Technology

Sound Model

We assume that:

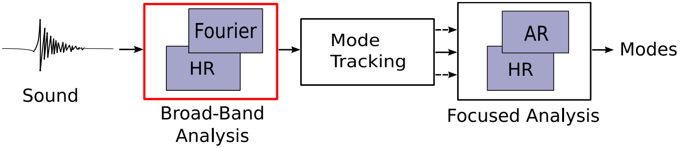
- the signal is made of a **sum of exponentially decaying cisoids**
- the physical properties of the impacted object remain constant $\Rightarrow$ **mode frequencies and damping factors are constant**

$$\hat{x}(t) = \sum_{k=1}^{K} A_k e^{\delta_k t} e^{j(2\pi f_k t + \phi_k)}$$

SPCL

CIRMMT Centre for Interdisciplinary Research in Music Media and Technology

Broad-Band Analysis

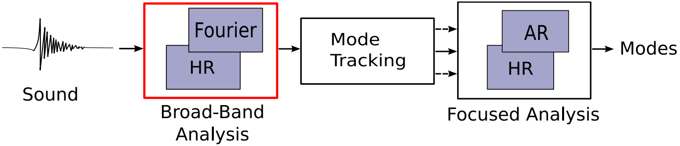


Fourier-Based Analysis:

- STFT $\Rightarrow$ one spectrum per frame
- Peaks picking using improved amplitude, frequency and phase estimation techniques [Lagrange *et al.* JAES'07]
- δ_k estimated by fitting an exponential on the amplitude profile, or Energy Decay Relief [Jot ICASSP'92] of one mode

SPCL
CIRMMT Centre for Interdisciplinary Research in Music Media and Technology

Broad-Band Analysis



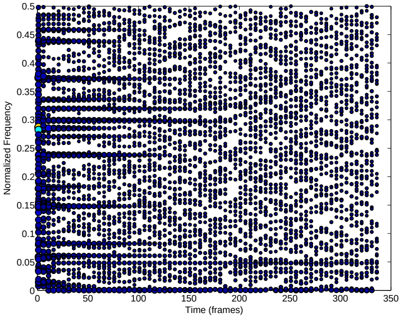
High Resolution Analysis:

- Using the adaptive implementation of the ESPRIT algorithm [Badeau *et al.* WASSPA'05]
- Frame-based analysis:
  - pre-processing
  - f_k and δ_k estimation
  - A_k and ϕ_k estimation

SPCL

CIRMMT Centre for Interdisciplinary Research in Music Media and Technology

Broad-Band Analysis

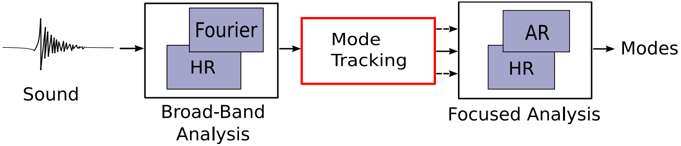


Corresponding **sound**

SPCL

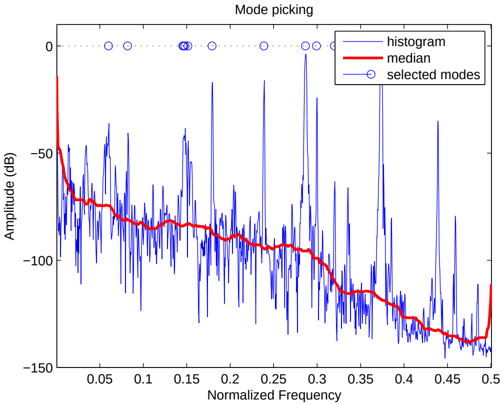
CIRMMT Centre for Interdisciplinary Research in Music Media and Technology

Mode identification and Tracking



1. Frequency marginal to identify main modes

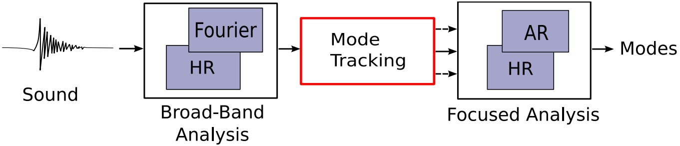
Mode identification and Tracking



SPCL

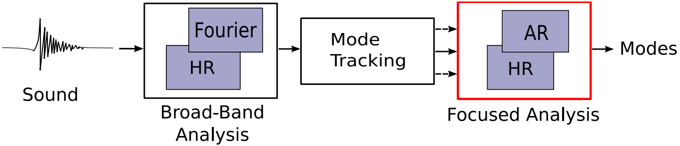
CIRMMT Centre for Interdisciplinary Research in Music Media and Technology

Mode identification and Tracking



1. Frequency marginal to identify main modes
2. Detected modes tracked over time using standard frequency proximity criterion

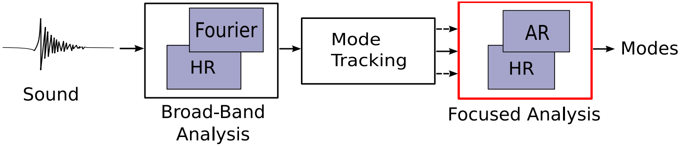
Focused Analysis



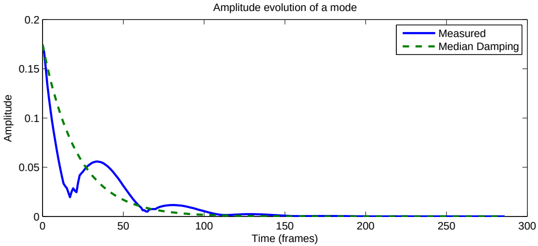
1. **Selection** of the partial track to process
   - amplitude high enough, lasts long enough
   - error between measured and estimated amplitude profiles high enough:

Centre for Interdisciplinary Research in Music Media and Technology

Focused Analysis

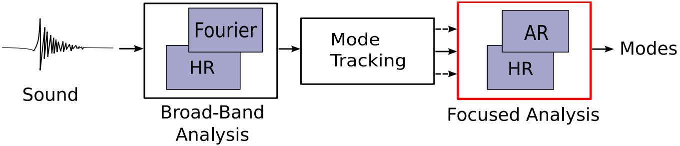


1. **Selection** of the partial track to process
   - amplitude high enough, lasts long enough
   - error between measured and estimated amplitude profiles high enough:



SPCL
CIR MMT Centre for Interdisciplinary Research in Music Media and Technology

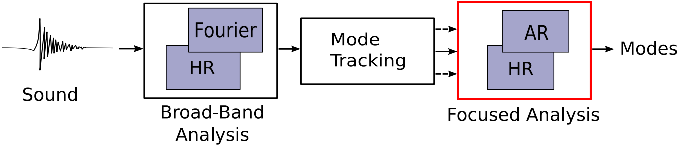
Focused Analysis



1. **Selection** of the partial track to process
2. **Pre-processing** [Laroche, JASA'93]
   - complex FIR filter around f_k
   - modulation + downsampling

SPCL
CIRMMT Centre for Interdisciplinary Research in Music Media and Technology

Focused Analysis

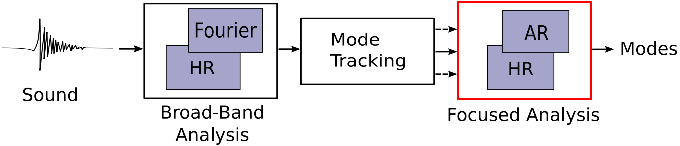


1. **Selection** of the partial track to process
2. **Pre-processing** [Laroche, JASA'93]
3. Estimation of **frequencies and damping**
   - AR analysis as done in [Karjalainen *et al.*, EUSIPCO'05]
   - HR analysis using a non adaptive ESPRIT (not frame-based)

SPCL

CIRMMT Centre for Interdisciplinary Research in Music Media and Technology

Focused Analysis

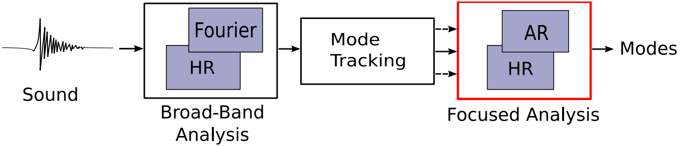


1. **Selection** of the partial track to process
2. **Pre-processing** [Laroche, JASA'93]
3. Estimation of **frequencies and damping**
4. Leat Squares **amplitude and phase** estimation

SPCL

CIRMMT Centre for Interdisciplinary Research in Music Media and Technology

Focused Analysis



1. **Selection** of the partial track to process
2. **Pre-processing** [Laroche, JASA'93]
3. Estimation of **frequencies and damping**
4. Leat Squares **amplitude and phase** estimation
5. **Sorting** components
   - discarding: $\delta_k > 0$ and $|f_{mode} - f_k| > \Delta f_{max}$
   - $\frac{A_k}{max(A_k)} >$ threshold

Centre for Interdisciplinary Research in Music Media and Technology

Synthetic Signals

- A_k: uniform pdf in [0.9, 1]
- δ_k: uniform pdf in [0.0001, 0.001]
- f_k
  - 2 freqs $\in$ [0.2499, 0.2501]
  - 20 freqs $\in$ [0.2, 0.3]

Centre for Interdisciplinary Research in Music Media and Technology

Synthetic Signals

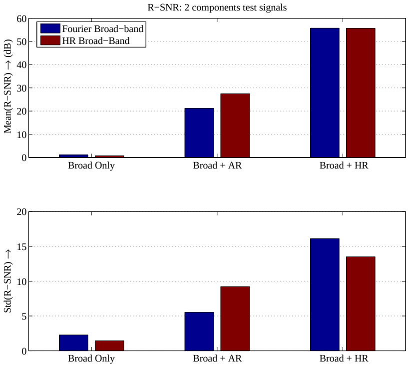
- A_k: uniform pdf in [0.9, 1]
- δ_k: uniform pdf in [0.0001, 0.001]
- f_k
  - 2 freqs $\in$ [0.2499, 0.2501]
  - 20 freqs $\in$ [0.2, 0.3]

$\Rightarrow$ 1000 sounds

SPCL

CIRMMT Centre for Interdisciplinary Research in Music Media and Technology

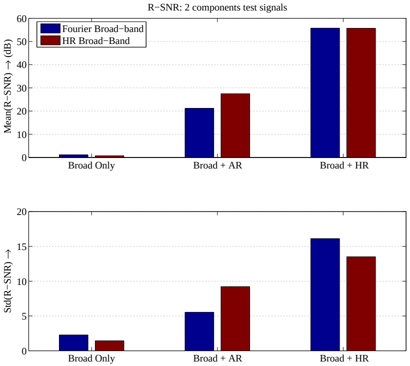
Results: 2 components:



SPCL

CIRMMT Centre for Interdisciplinary Research in Music Media and Technology

Results: 2 components:

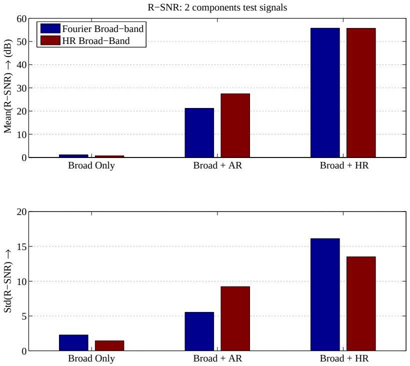


- AR $<$ HR: over-estimation of the number of components using our empirical metric

SPCL

Centre for Interdisciplinary Research in Music Media and Technology

Results: 2 components:

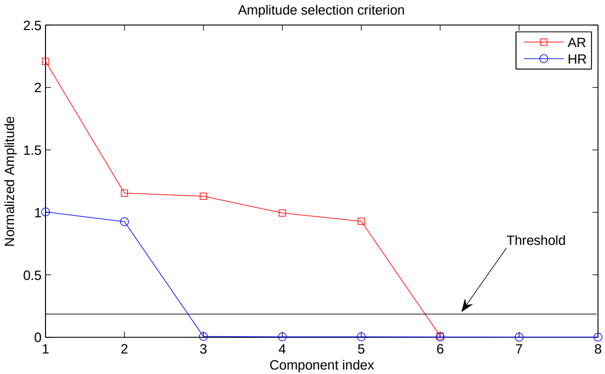


- AR $<$ HR: over-estimation of the number of components using our empirical metric
- HR broad + HR $\approx$ Fourier broad + HR

SPCL

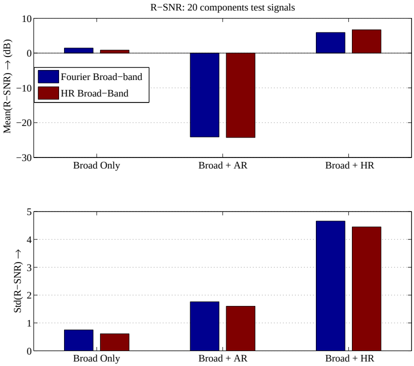
CIRMMT Centre for Interdisciplinary Research in Music Media and Technology

Results: 2 components:



Centre for Interdisciplinary Research in Music Media and Technology

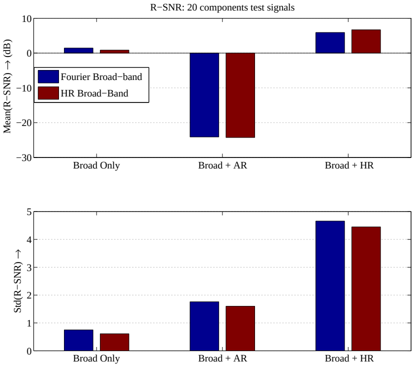
Results: 20 components:



SPCL

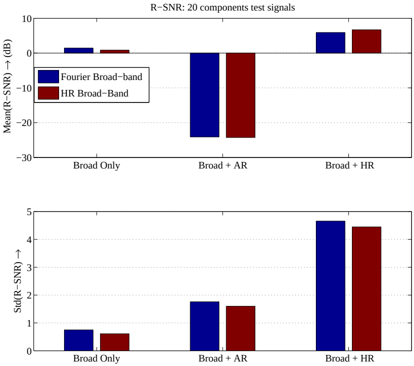
CIRMMT Centre for Interdisciplinary Research in Music Media and Technology

Results: 20 components:



- HR focused analysis improves results for both broad-band analyses

Results: 20 components:



- HR focused analysis improves results for both broad-band analyses
- **Fourier Broadband + HR focused analysis** = good compromise: computation load / good results

SPCL

CIRMMT Centre for Interdisciplinary Research in Music Media and Technology

“Real world” cases

Type of sound:

- Metallic plate struck by ceramic hammer

Performance Assessment

- **Context:** Excitation estimation using standard deconvolution method [Laroche *et al.* TSAP'94]

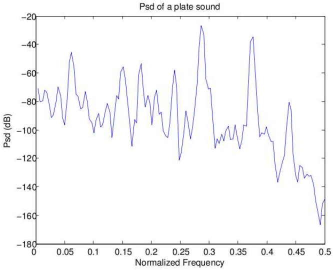
Performance Assessment

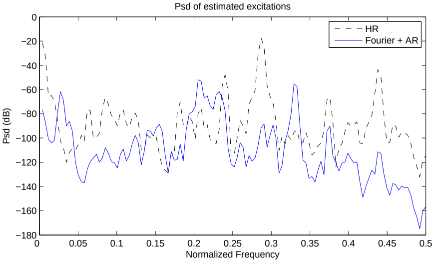
- **Context:** Excitation estimation using standard deconvolution method [Laroche *et al.* TSAP'94]
- We compare HR broadband (state of the art in terms of resolution) and Fourier broadband + AR focused analysis

Performance Assessment

- **Context:** Excitation estimation using standard deconvolution method [Laroche *et al.* TSAP’94]
- We compare HR broadband (state of the art in terms of resolution) and Fourier broadband + AR focused analysis
- We compare HR broadband and Fourier broadband + HR focused analysis

HR vs. Fourier-AR:





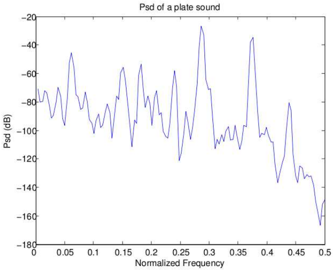
Original

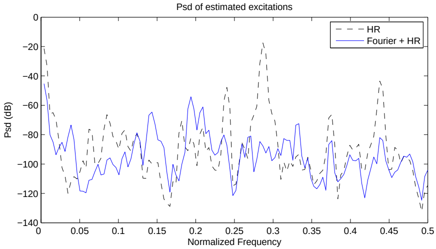
HR excitation
Fourier + AR excitation

SPCL

CIRMMT Centre for Interdisciplinary Research in Music Media and Technology

HR vs. Fourier-HR:





HR excitation
Fourier + HR excitation

Original

Centre for Interdisciplinary Research in Music Media and Technology

Results

- Main modes of the sounds are generally absent from excitation signals
- Presence of strong modes in excitation $=$ artifact of deconvolution method

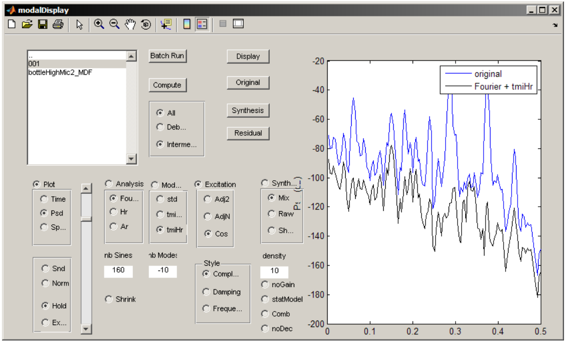
Conclusion

A two-step analysis scheme for the estimation of modal parameters from recorded sounds.

- alleviating parametrization and manual post-processing
- interest demonstrated in the case of synthetic signals
- interest demonstrated in the context of excitation estimation for source filter modeling

SPCL

CIRMMT Centre for Interdisciplinary Research in Music Media and Technology



Toolbox available upon request: mathieu.lagrange@mcgill.ca

Centre for Interdisciplinary Research in Music Media and Technology

Thank you for your attention.

SPCL

CIRMMT Centre for Interdisciplinary Research in Music Media and Technology

Questions, comments ?

SPCL

CIRMMT Centre for Interdisciplinary Research in Music Media and Technology