Nouveaux outils de traitement du signal musical: quel utilisation pour la recherche en cognition?

JJ Aucouturier (CNRS/IRCAM)
aucouturier@gmail.com

en interaction avec: Emmanuel Bigand (LEAD, Dijon);
Petter Johansson (Lund)
Olivier Houix (IRCAM)

MIR : music information retrieval

traitement du signal sonore
+ algorithmes de décision

Napster v2.0 BETA 7

File Actions Help

Home | Chat | Library | Search | Hot List | Transfer | Discover | Help

Artist: revolution
Title:
Max Results: 100

Find it! | Clear Fields | Advanced <<

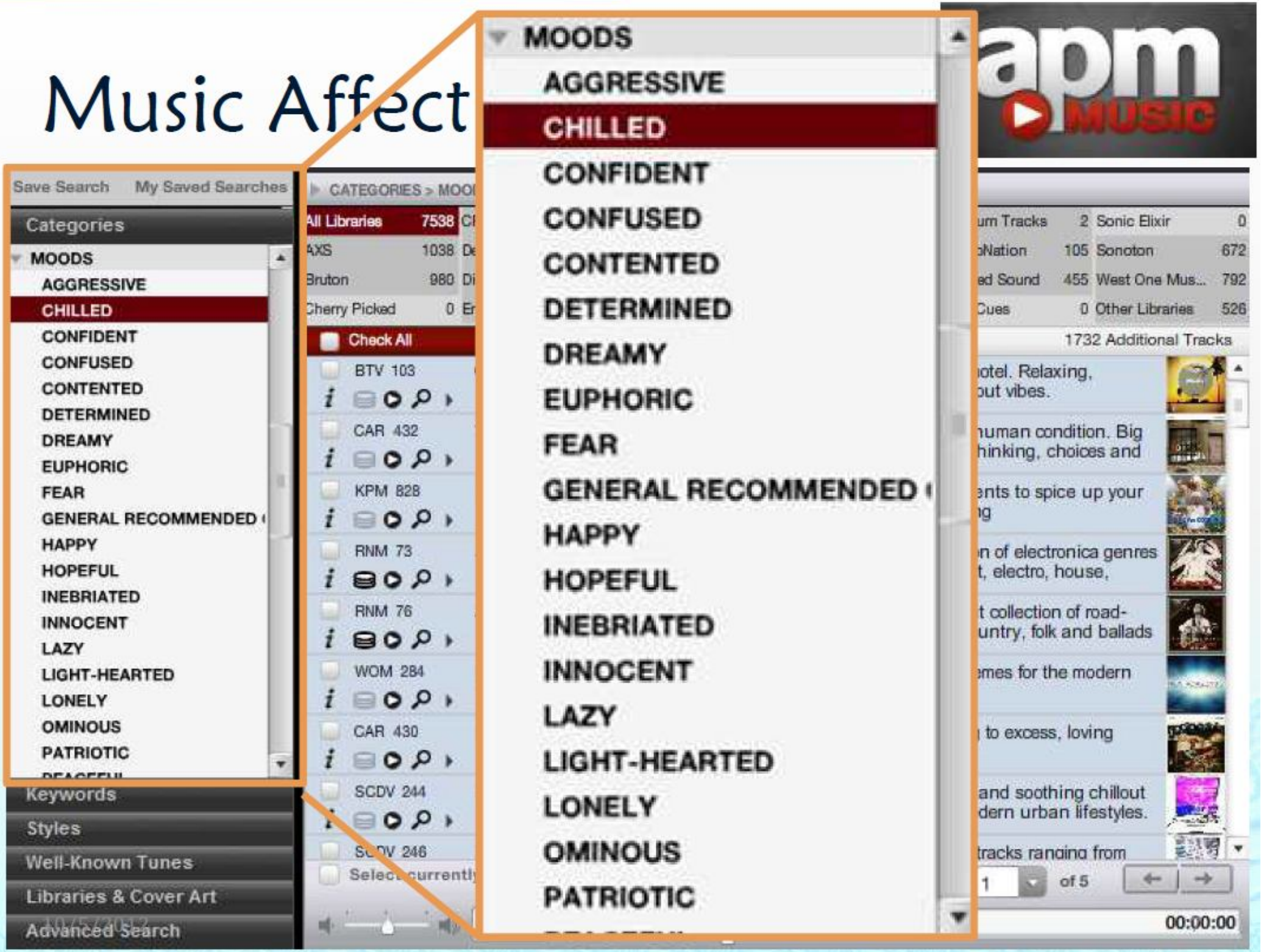
Bitrate:
Connection: AT LEAST DSL
Ping time:
☑ Ping search results

Filename	Filesize	Bitrate	Freq	Length	User	Connection	Ping
White Album\be11b - 12 - revolution 9.mp3	12,053,653	192	44100	8:16	therealh...	DSL	50
White Album\be11b - 08 - revolution 1.mp3	6,138,531	192	44100	4:15	therealh...	DSL	50
Past Masters Volume Two\be15 - 08 - revolution.mp3	4,912,821	192	44100	3:25	therealh...	DSL	50
Beatles - Revolution 1.mp3	3,068,032	96	44100	4:15	ryanm1c	T3	90
The Beatles_1 Revolution 9 master version, RS1(mono).mp3	7,489,536	128	44100	7:42	ryanm1c	T3	90
nbd\the_revolution-nbd.mp3	2,589,304	160	44100	2:12	Wisema...	DSL	100
Napster\Talkin' Bout A Revolution.mp3	3,196,928	160	44100	2:41	wakflakes	T1	110
Music\Revolution 1993.mp3	24,672,131	320	44100	10:08	yoshizumi	DSL	120
Music\Purple Rain - Prince And The Revolution - The Beautiful Ones[1].mp3	3,772,105	96	44100	5:12	mjking46	T1	131
Music\The Beatles - Revolution.mp3	3,282,624	128	44100	3:26	donnie2...	T3	131
Music\[Queensryche]Operation Mindcrime=Revolution Calling.mp3	2,473,200	128	44100	2:36	frankief...	DSL	140
Prince and the Revolution - Raspberry Beret.mp3	3,402,470	128	44100	3:33	xmurdoc	DSL	141
Sway and King Tech feat. Rza, Eminem, KRS-One, Ezibit and DJ Revolution - The Anthem.mp3	5,439,488	160	44100	4:31	JtheLover	T1	160
Coalesce 012 Revolution in Just Listening\09-They Always Come In Fall.mp3	2,118,272	128	44100	2:15	youasth...	T1	160
Coalesce 012 Revolution in Just Listening\08-Counting Murders and Drinking Beer (the $46000 escap...	2,454,384	128	44100	2:35	youasth...	T1	160
Coalesce 012 Revolution in Just Listening\07-Jesus In the Year 2000 Next On the Shit List.mp3	2,839,324	128	44100	2:59	youasth...	T1	160
Coalesce 012 Revolution in Just Listening\06-Where the Hell Is Rick Thorne These Days.mp3	1,852,523	128	44100	1:59	youasth...	T1	160
Coalesce 012 Revolution in Just Listening\05-Sometimes Selling Out Is Waking Up.mp3	3,219,249	128	44100	3:22	youasth...	T1	160
Coalesce 012 Revolution in Just Listening\04-While the Jackass Operation Spins Its Wheels.mp3	2,275,915	128	44100	2:24	youasth...	T1	160
Coalesce 012 Revolution in Just Listening\03-Burn Everything That Bears Our Name.mp3	2,300,575	128	44100	2:26	youasth...	T1	160
Coalesce 012 Revolution in Just Listening\02-cowards.com.mp3	2,330,250	128	44100	2:28	youasth...	T1	160
Coalesce 012 Revolution in Just Listening\001-What Happens On the Road Always Comes Home.mp3	2,965,548	128	44100	3:07	youasth...	T1	160
and everything else\weakerthans - ringing of revolution.mp3	3,312,326	128	44100	3:28	youasth...	T1	160
Beatles\Beatles - Revolution.mp3	4,874,385	192	44100	3:24	geenljr	T1	180
World-Wide-Message-Tribe-Revolution.mp3	3,377,663	128	44100	3:32	sh8463	DSL	190
Bad Religion\victims of the revolution.mp3	3,154,337	128	44100	3:18	robertje...	T1	240
(Dance Dance Revolution) Get Up 'n Move.mp3	1,272,740	128	44100	1:23	pmckni...	DSL	240
Music\Pantera- Revolution is my name.mp3	5,103,616	128	44100	5:17	POD4Bi...	DSL	261
Music\Pantera- Revolution is my name.mp3	5,103,616	128	44100	5:17	POD4Bi...	DSL	261
Music\Pantera- Revolution is my name.mp3	5,103,616	128	44100	5:17	POD4Bi...	DSL	261
Music\Pantera- Revolution is my name.mp3	5,103,616	128	44100	5:17	POD4Bi...	DSL	261

Returned 82 results.

Get Selected Songs | Add Selected User to Hot List

Music Affect



Thank You!
FCT
COMPETE

Features informatives
pour décider de constructions
psychologiques

i.e. chercher des représentations
« sensorielles »
ou « cognitives »

Psychologie

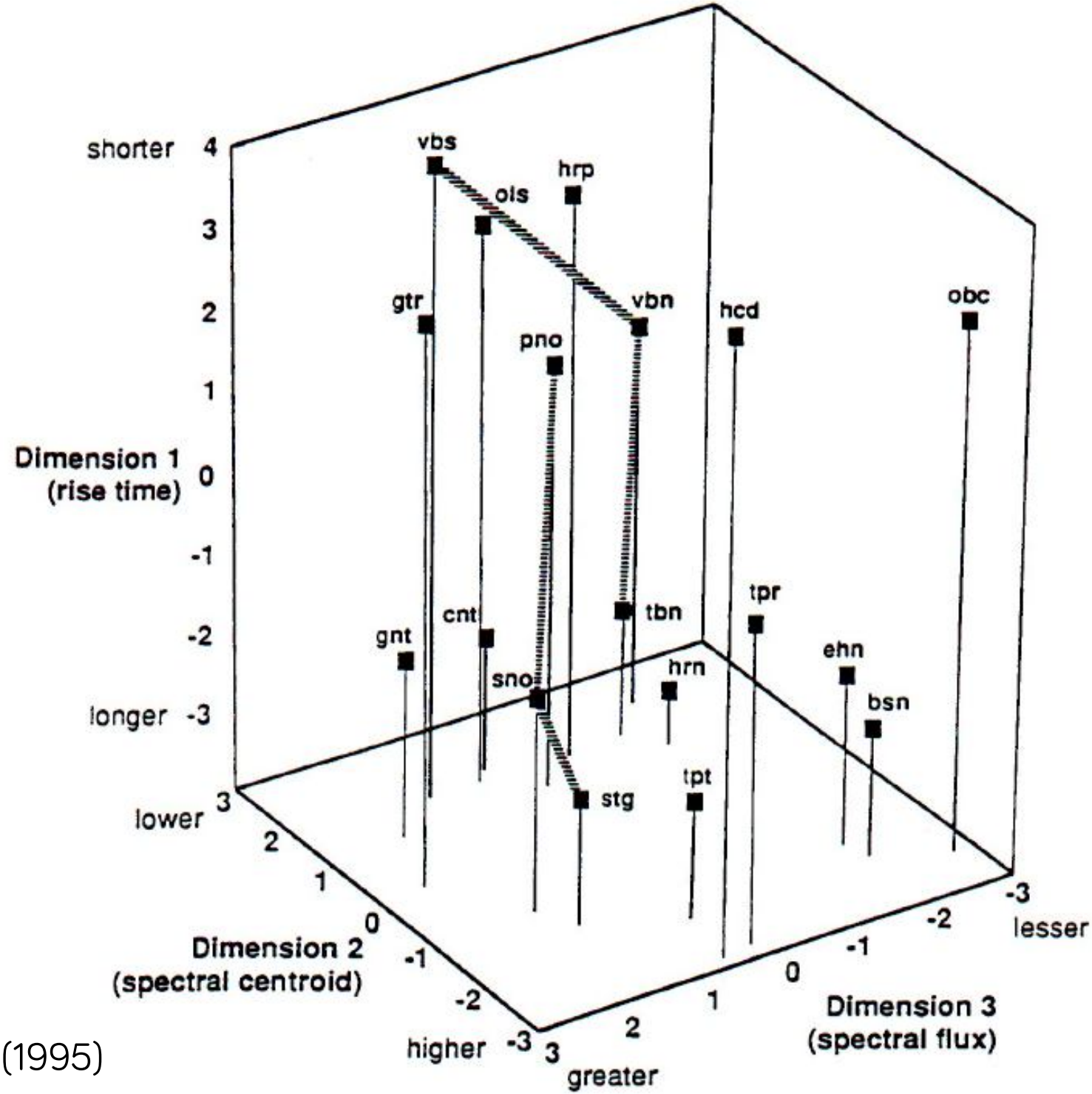
ses techniques

ex. psychoacoustique

i.e. chercher des représentations

« sensorielles »

ou « cognitives »



McAdams et al. (1995)

Psychologie

ses techniques

ex. psychoacoustique

ex. neurosciences cog

i.e. chercher des représentations

« sensorielles »

ou « cognitives »

Psychologie

ses techniques

- ex. psychoacoustique
- ex. neurosciences cog

ses inspirations

- ex. physiologie

i.e. chercher des représentations
« sensorielles »
ou « cognitives »

MIR

ses techniques

optimisation algorithmique

Psychologie

ses techniques

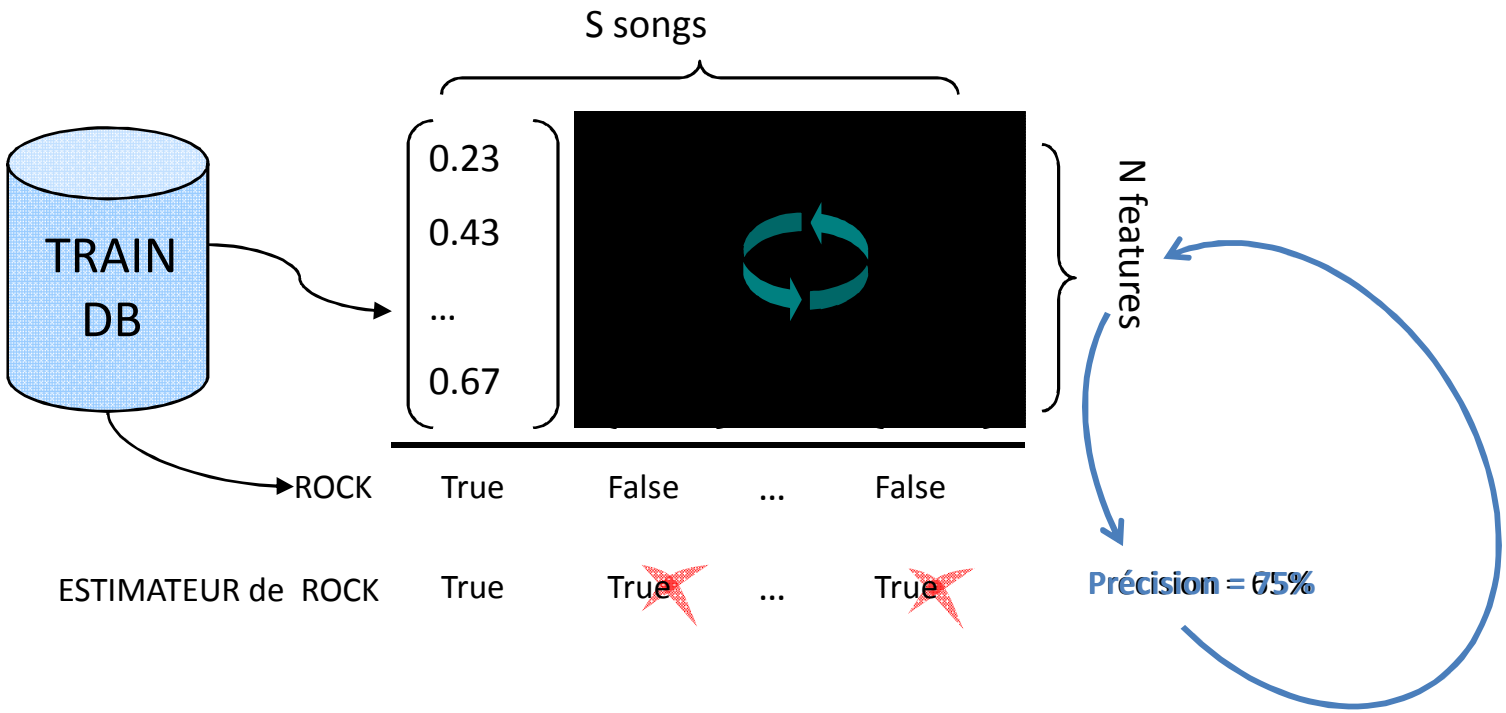
ex. psychoacoustique

ex. neurosciences cog

ses inspirations

ex. physiologie

i.e. chercher des représentations « sensorielles » ou « cognitives »


S songs
TRAIN DB
0.23
0.43
...
0.67
N features
ROCK
True
False
...
False
ESTIMATEUR de ROCK
True
True
...
True
Précision = 65%

MIR

ses techniques

optimisation algorithmique

ses inspirations

mathématiques

Psychologie

ses techniques

ex. psychoacoustique

ex. neurosciences cog

ses inspirations

ex. physiologie

i.e. chercher des représentations « sensorielles » ou « cognitives »

Spectral centroid

$$\mu = \int x \cdot p(x) \quad \delta x$$

where

- x are the observed data: $x = freq_v(x)$
- p(x) is the probability to observe x: $p(x) = \dfrac{ampl_v(x)}{\sum_x ampl_v(x)}$

Spectral entropy

Spectral spread

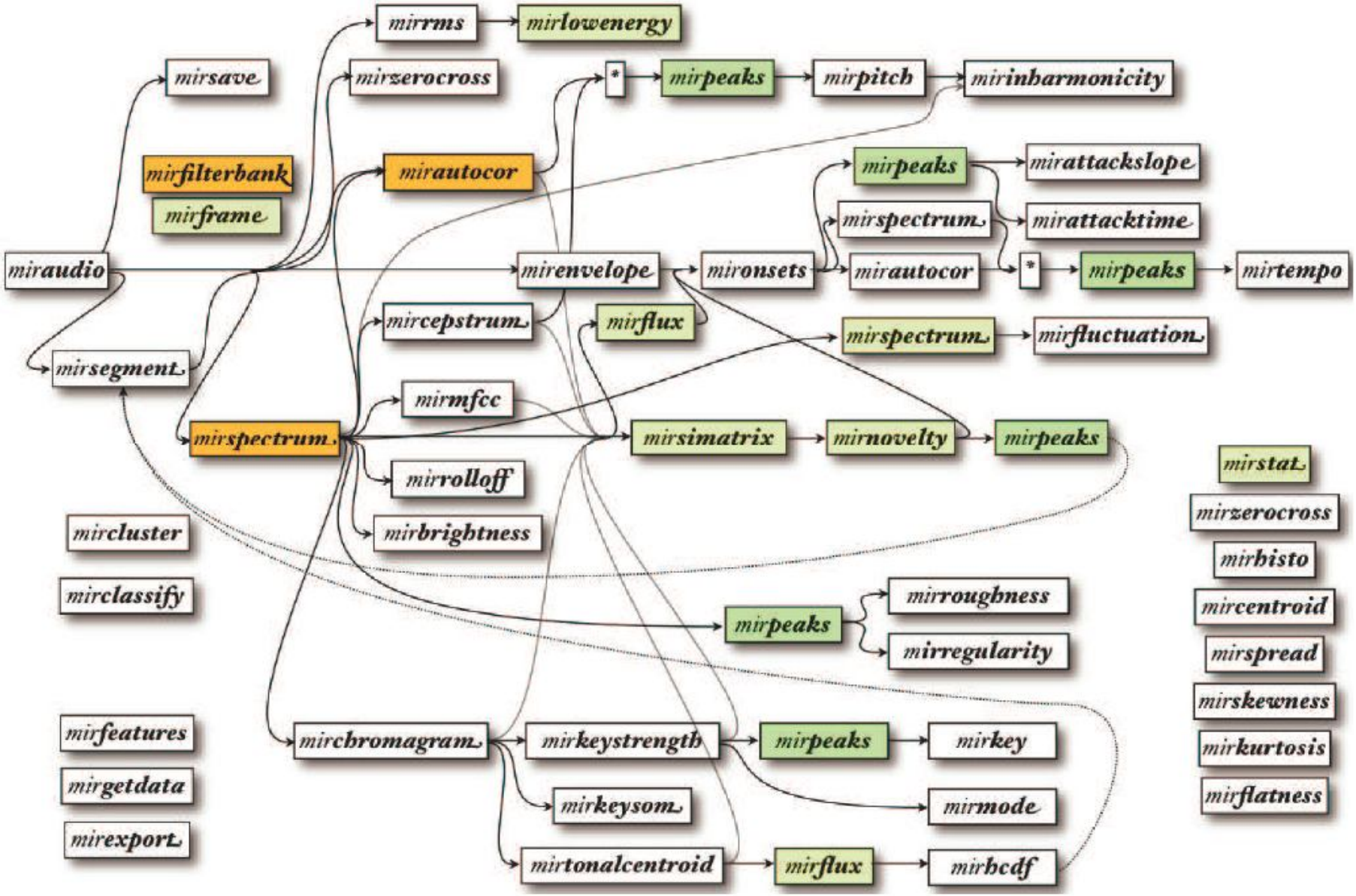
$$\sigma^2 = \int (x-\mu)^2 \cdot p(x) \quad \delta x$$

Spectral kurtosis

$$m_4 = \int (x-\mu)^4 \cdot p(x) \quad \delta x$$

Spectral skewness

$$m_3 = \int (x-\mu)^3 \cdot p(x) \quad \delta x$$



https://www.jyu.fi/hum/laitokset/musiikki/en/research/coe/materials/mirtoolbox

MIR

ses techniques

optimisation algorithmique

ses inspirations

mathématiques

Psychologie

ses techniques

ex. psychoacoustique

ex. neurosciences cog

ses inspirations

ex. physiologie

i.e. chercher des représentations « sensorielles » ou « cognitives »

MIR

ses techniques

optimisation algorithmique

ses inspirations

mathématiques

Psychologie

ses techniques

ex. psychoacoustique

ex. neurosciences cog

ses inspirations

ex. physiologie

Étude de cas #1:

Un au pays des

« Tutorial on automatic mood recognition »
ISMIR 2012, Oct. 2012, Porto

Thank You!
FCT
COMPETE

Étude de cas #1:

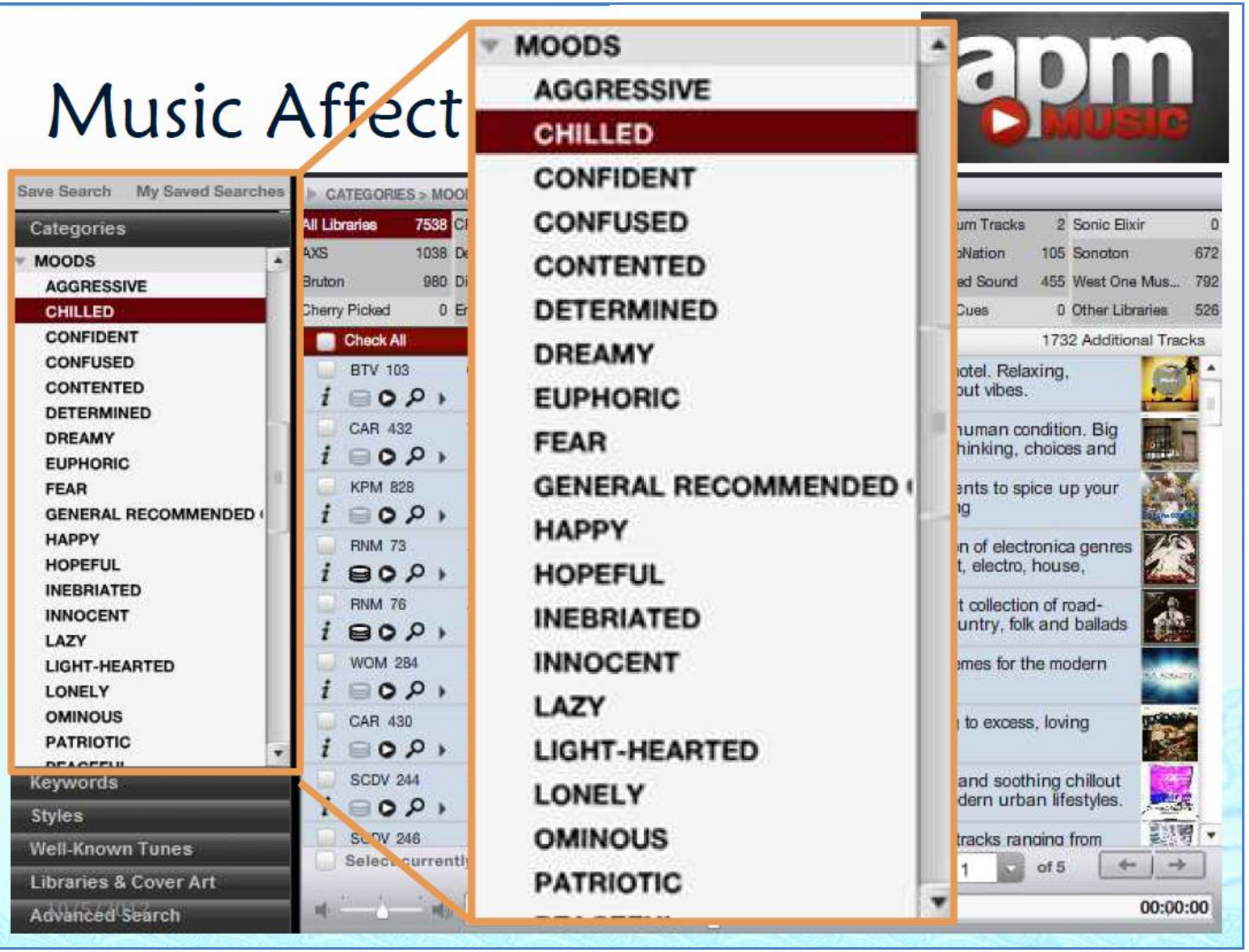
« Tutorial on automatic mood recognition »

ISMIR 2012, Oct. 2012, Porto

btw, IEEE Tr. Pattern Analysis and Machine Intelligence (IF = 4.9)

compare: Cognition (IF = 4.1)

Music Affect



Expressed or Induced

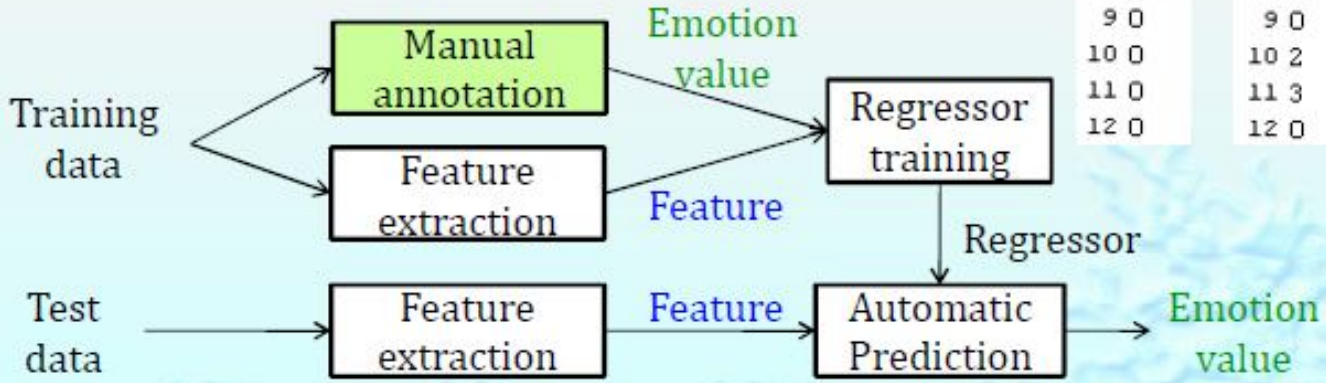
- Designated/indicated/expressed by a music piece
- Induced/evoked/felt by a listener
- Both are studied in MIR
  - Mainly differ in the ways of collecting labels
  - "indicate how you feel when listen to the music"
  - "indicate the mood conveyed by the music"

Issue 1: Difficulty of Emotion Annotation

- Rating emotion is difficult
  - User fatigue
  - Uniform ratings
  - Low quality of the ground truth
  - Difficult to create large-scale dataset

○ 1 ○ 2 ○ 3 ○ 4 ○ 5

	user A	user B
1	10	10
2	10	8
3	10	8
4	10	10
5	0	0
6	0	0
7	5	10
8	5	5
9	0	0
10	0	2
11	0	3
12	0	0



Expert Annotation: MIREX AMC

- Each expert had 250 clips
- 8 of 21 experts finished all assignments

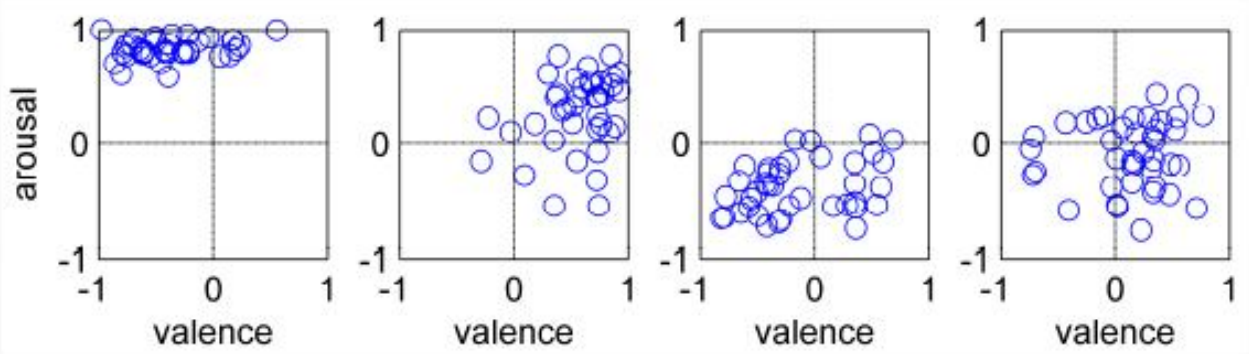
- 2468 judgments collected (3750 planned)
- Each clips had 2 or 3 judgments
- Avg. Cohen's Kappa: 0.5

Dataset built from agreements among experts

Agreements	C1	C2	C3	C4	C5	Total	Accuracy
3 of 3 judges	21	24	56	21	31	153	0.59
2 of 3 judges	41	35	18	26	14	134	0.38
2 0f 2 judges	58	61	46	73	75	313	0.54
Total	120	120	120	120	120	600	

Lessons: 1. Missed judgments -> low accuracy
2. Need **more motivated annotators**

Issue 2: Emotion Perception is Subjective



(a) Smells Like Teen Spirit (b) A Whole New World (c) The Rose (d) Tell Laura I Love Her

- A song can be perceived differently by two people, especially for emotions in the 3rd and 4th quadrants
  - Also explain why valence prediction is more challenging

Newer Challenges

- Cross-cultural applicability
  - Existent efforts focus on Western music
  - OS1 @ ISMIR 2012 (tomorrow): Yang & Hu: Cross-cultural Music Mood Classification: A Comparison on English and Chinese Songs
- Personalization
  - Ultimate solution to the subjectivity problem
- Contextualization
  - Even the same person's emotional responses change in different time, location, occasions
  - PS1 @ ISMIR 2012 (Tomorrow) Watson & Mandryk: Modeling Musical Mood From Audio Features and Listening Context on an In-Situ Data Set

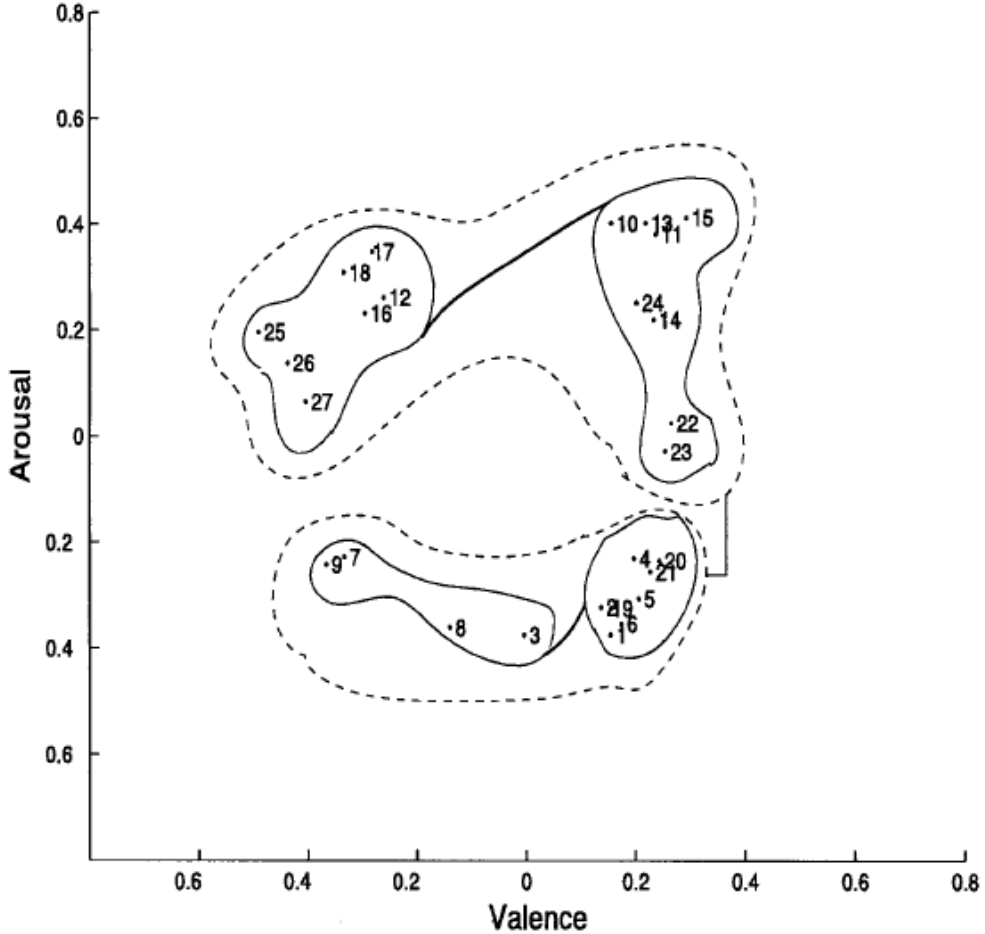
« the ultimate solution to the subjectivity problem »

?!

Étude de cas #2:

Un au pays des

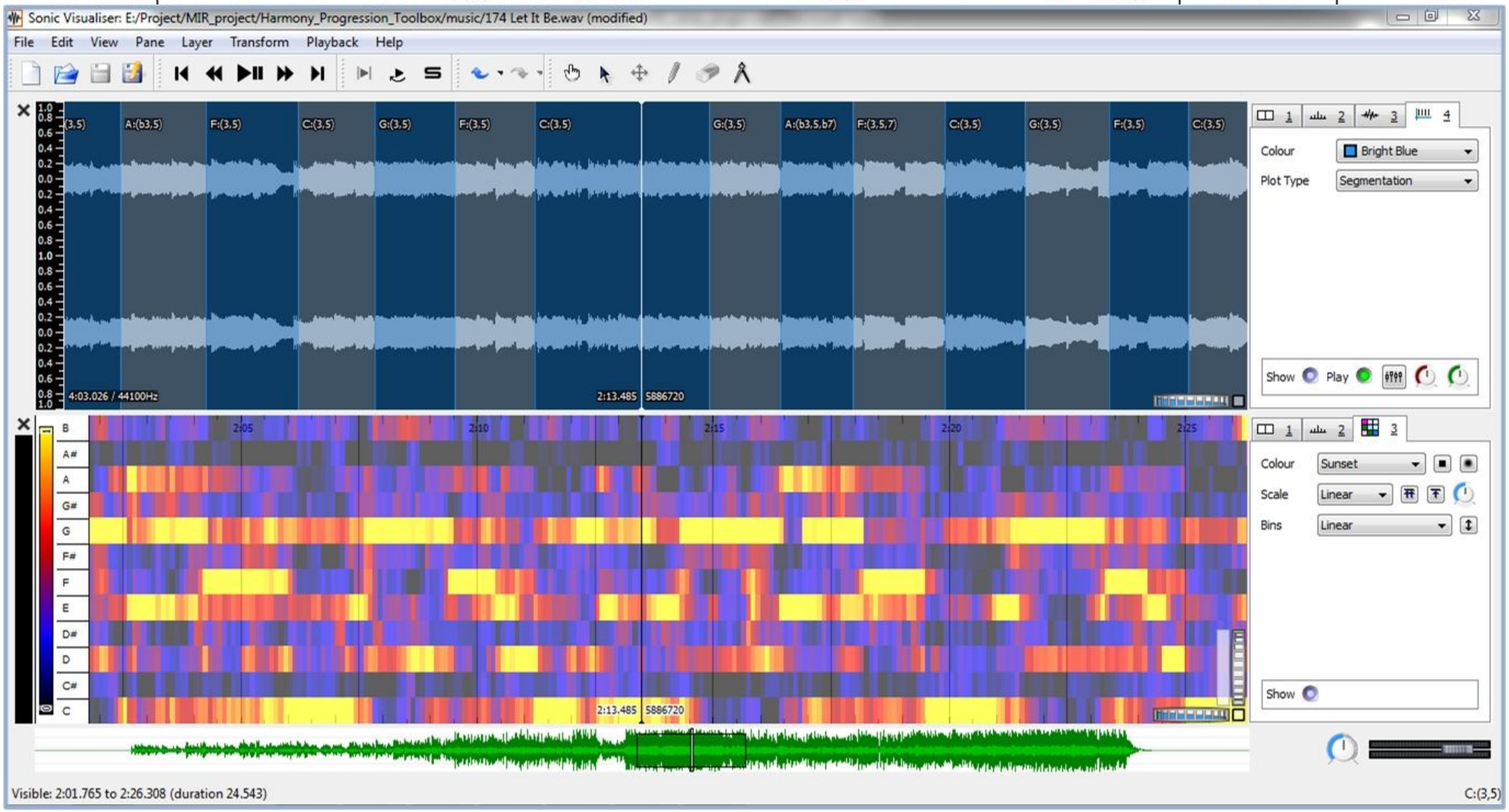
Étude de cas #2:
Bigand et al., 2005



Regression for valence	
Feature	β
tonal_chromagram_peak_PeakMagPeriodEntropy	-0.75
tonal_mode_Mean	0.13
spectral_mfcc_PeriodAmp_8 (600 Hz +/- 66)	0.12
spectral_ddmfcc_PeriodEntropy_1 (133.3)	-0.11

Regression for arousal	
Feature	β
spectral_mfcc_Std_6 (466.6)	-0.34
spectral_mfcc_Mean_3 (266.6)	0.28
tonal_keyclarity_Std	-0.28
spectral_mfcc_Std_7 (533.3)	0.24

Regression for valence	
Feature	β
tonal_chromagram_peak_PeakMagPeriodEntropy	-0.75



Regression for valence	
Feature	β
tonal_chromagram_peak_PeakMagPeriodEntropy	-0.75
tonal_mode_Mean	0.13
spectral_mfcc_PeriodAmp_8 (600 Hz +/- 66)	0.12
spectral_ddmfcc_PeriodEntropy_1 (133.3)	-0.11

Regression for arousal	
Feature	β
spectral_mfcc_Std_6 (466.6)	-0.34
spectral_mfcc_Mean_3 (266.6)	0.28
tonal_keyclarity_Std	-0.28
spectral_mfcc_Std_7 (533.3)	0.24

On « explique » tout, mais on ne comprend rien.

(avec E. Bigand, J. Intell. Inf. Systems, 2012)

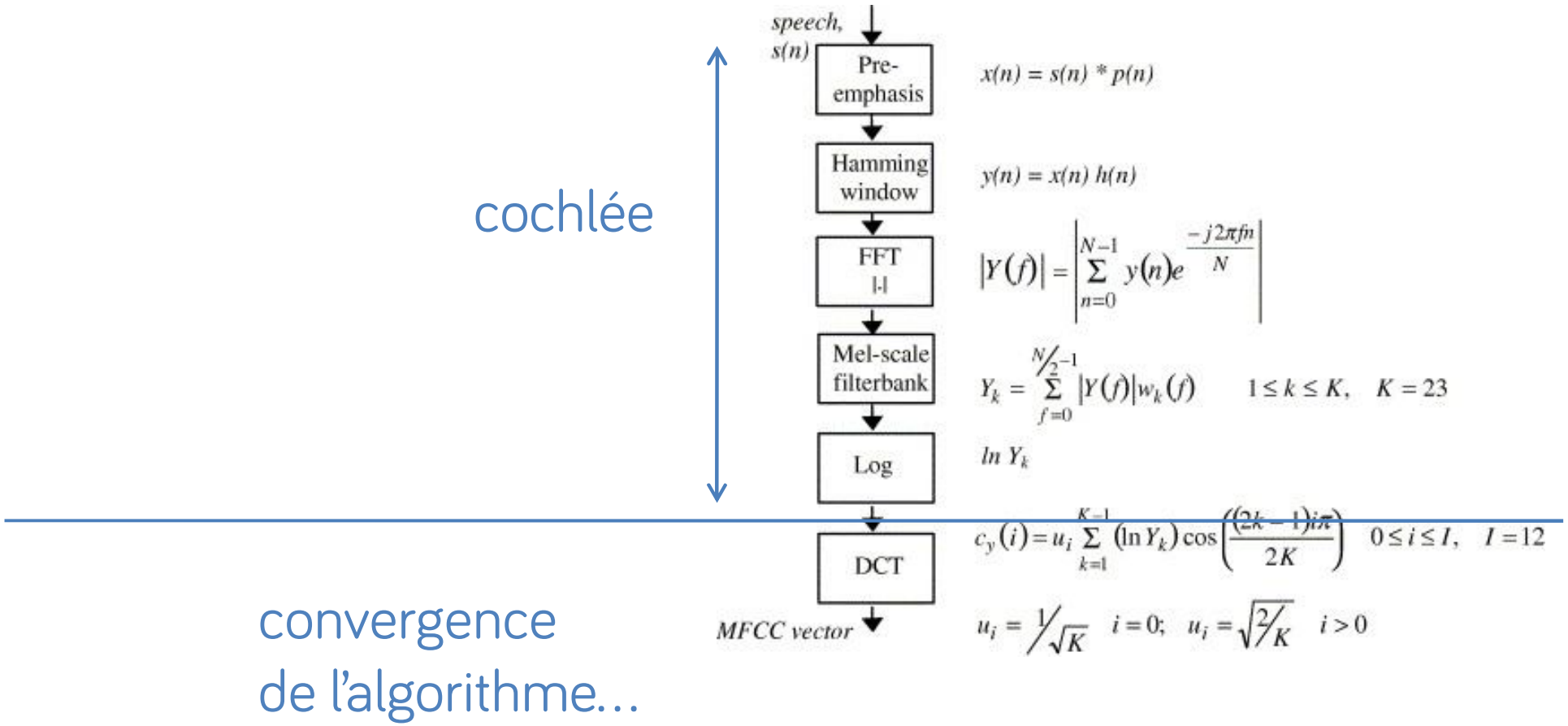
Conclusion: MIR, ça ne sert à rien.

Merci pour votre attention.

3 pistes pour dépasser tout ça

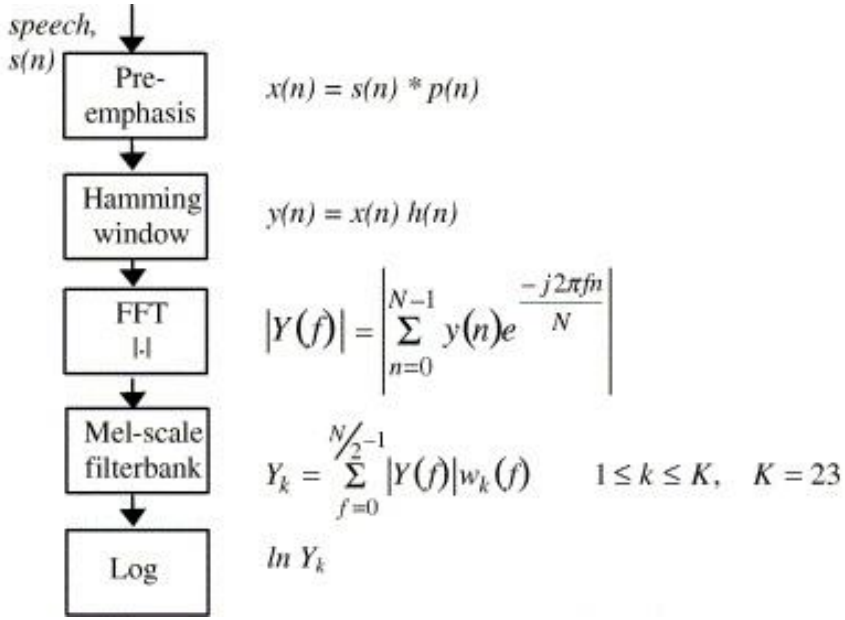
1) Features n’ont pas une modularité neurale/cognitive claire

MFCCs (Mel-Frequency Cepstrum Coefficients)

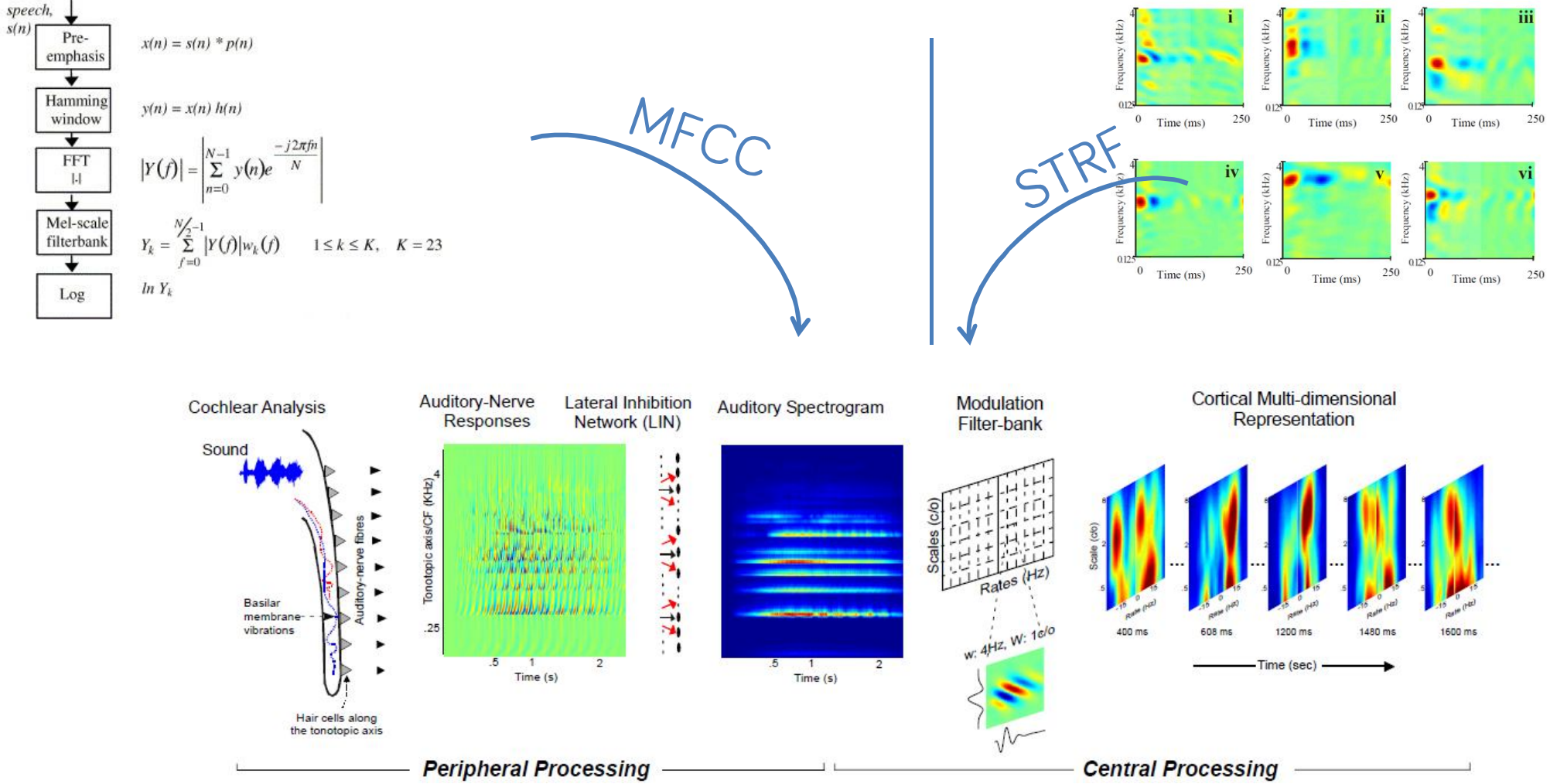


Que peut-on apporter? Un peu de modération

cochlée

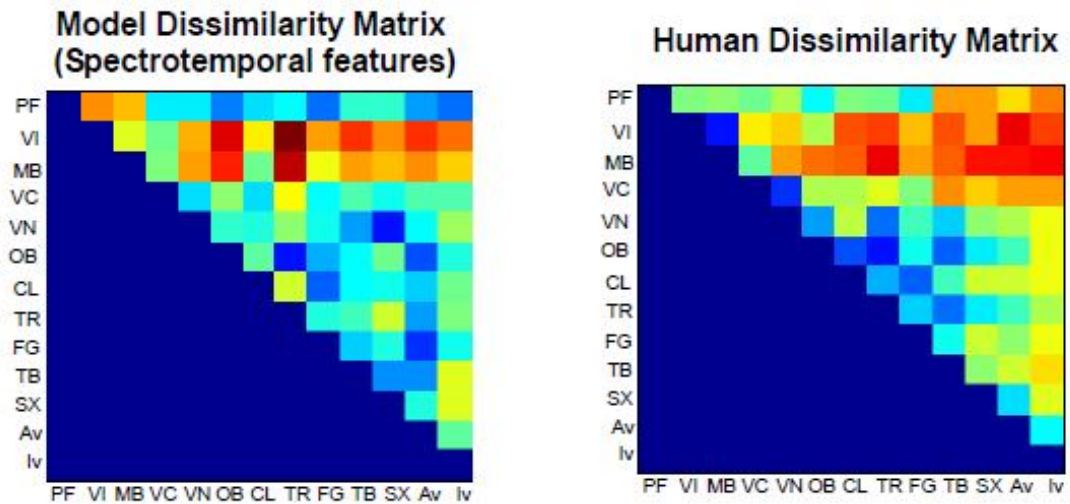


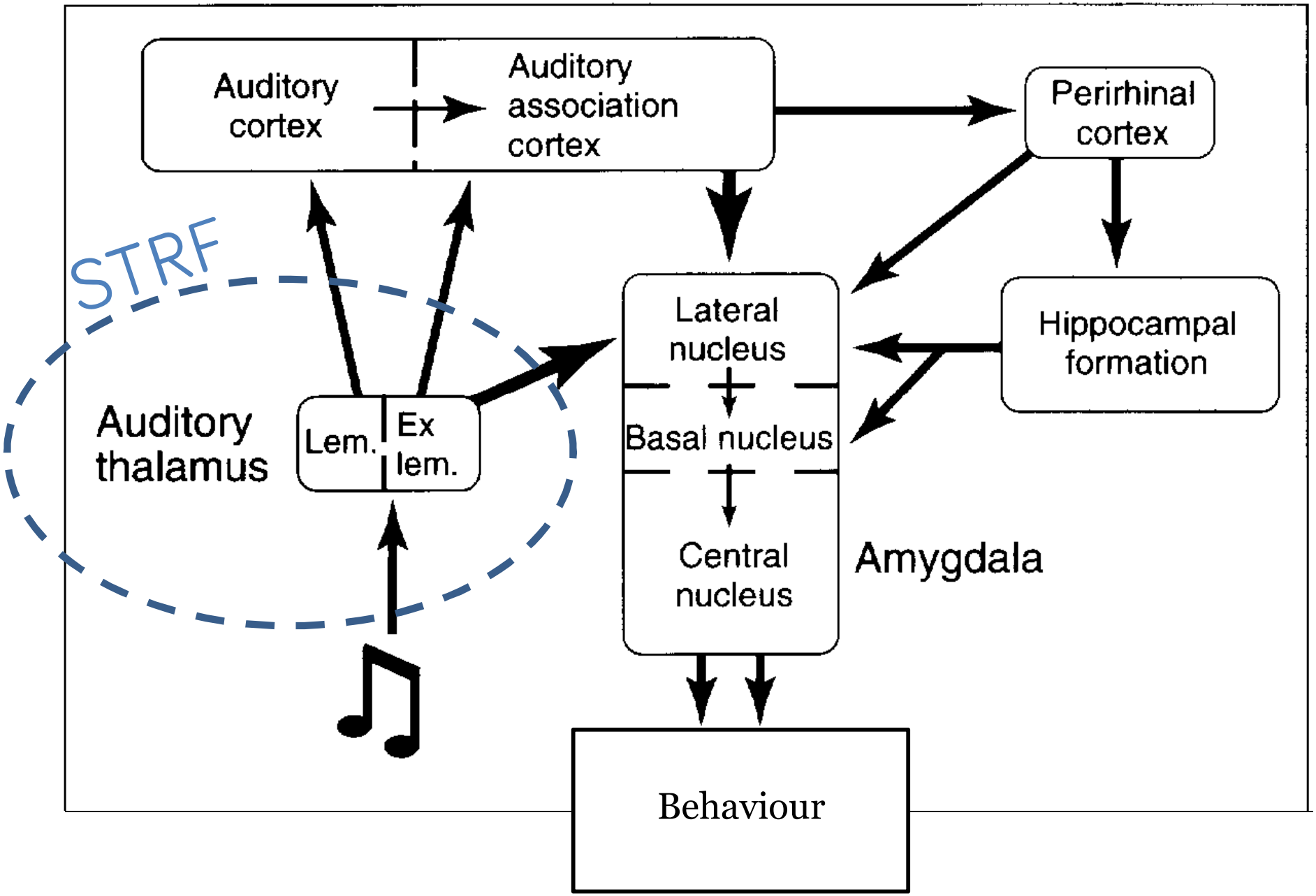
stop, on se calme!



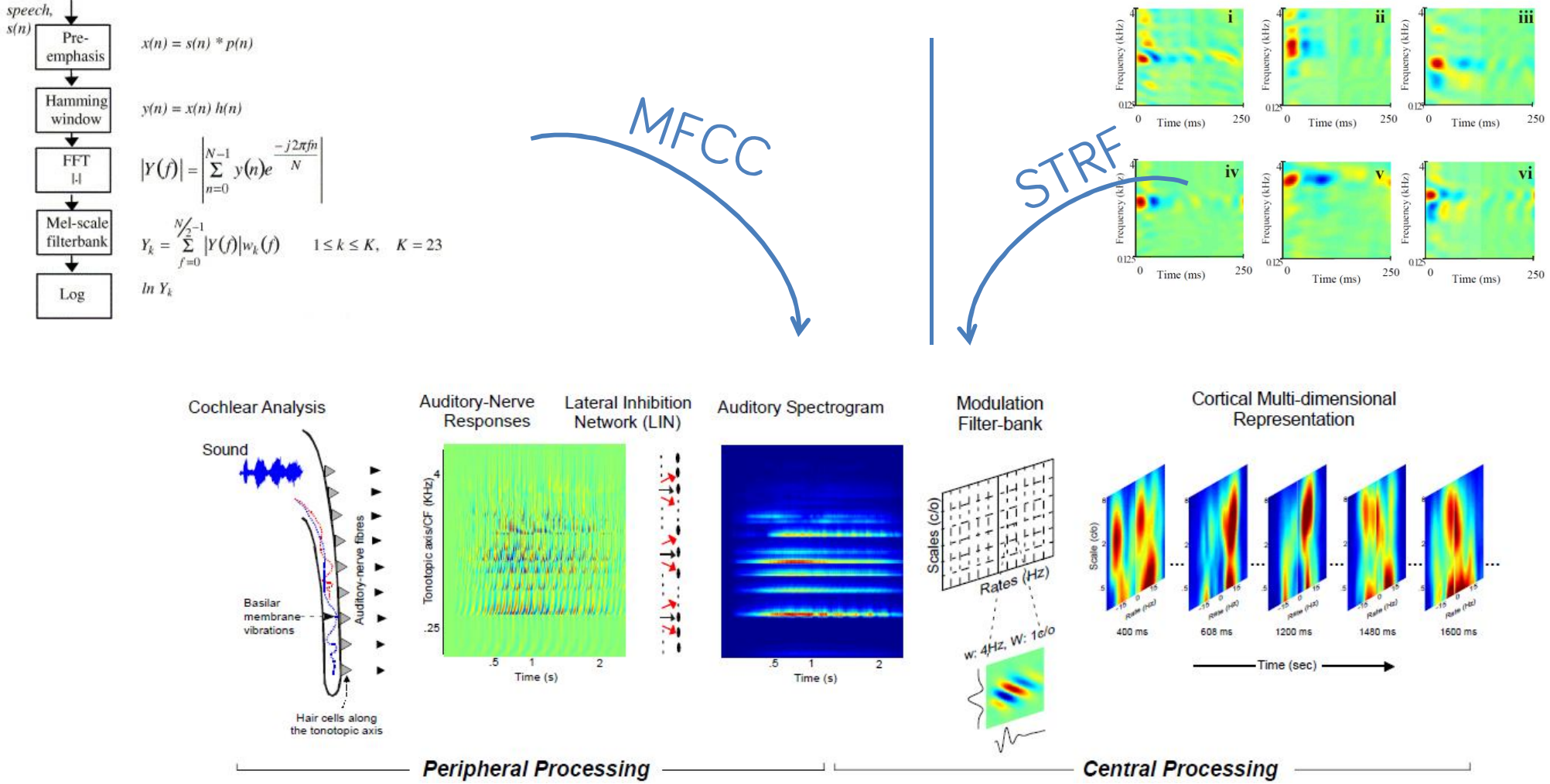
Chi, Taishih, Ru, Powen,Shamma, Shihab A. (2005) Multi-resolution spectrotemporal analysis of complex sounds JASA vol. 118(2), pp. 887-906

Patil K, Pressnitzer D, Shamma S, Elhilali M. (2012) Music in Our Ears: The Biological Bases of Musical Timbre Perception. PLoS Comput Biol 8(11)



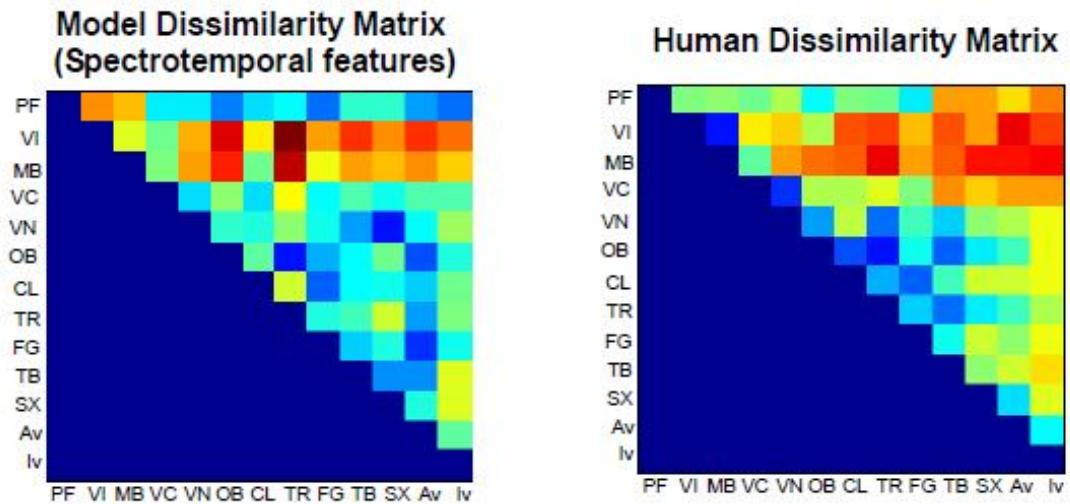

Auditory cortex
Auditory association cortex
Perirhinal cortex
STRF
Lateral nucleus
Hippocampal formation
Auditory thalamus
Lem.
Ex lem.
Basal nucleus
Central nucleus
Amygdala
Behaviour

Ledoux (1999)



Chi, Taishih, Ru, Powen,Shamma, Shihab A. (2005) Multi-resolution spectrotemporal analysis of complex sounds JASA vol. 118(2), pp. 887-906

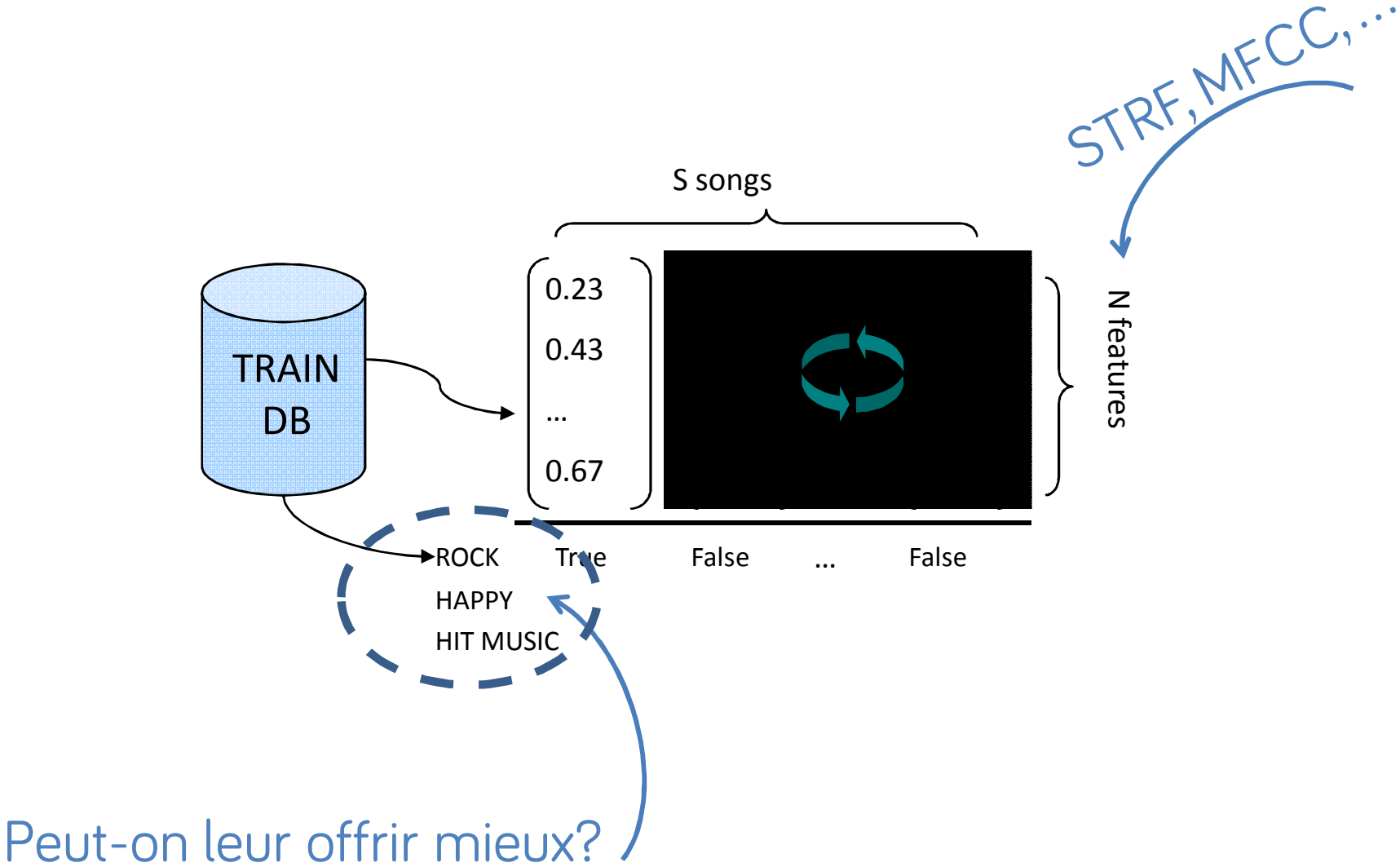
Patil K, Pressnitzer D, Shamma S, Elhilali M. (2012) Music in Our Ears: The Biological Bases of Musical Timbre Perception. PLoS Comput Biol 8(11)



3 pistes pour dépasser tout ça

1) Features n'ont pas une modularité neurale/cognitive claire

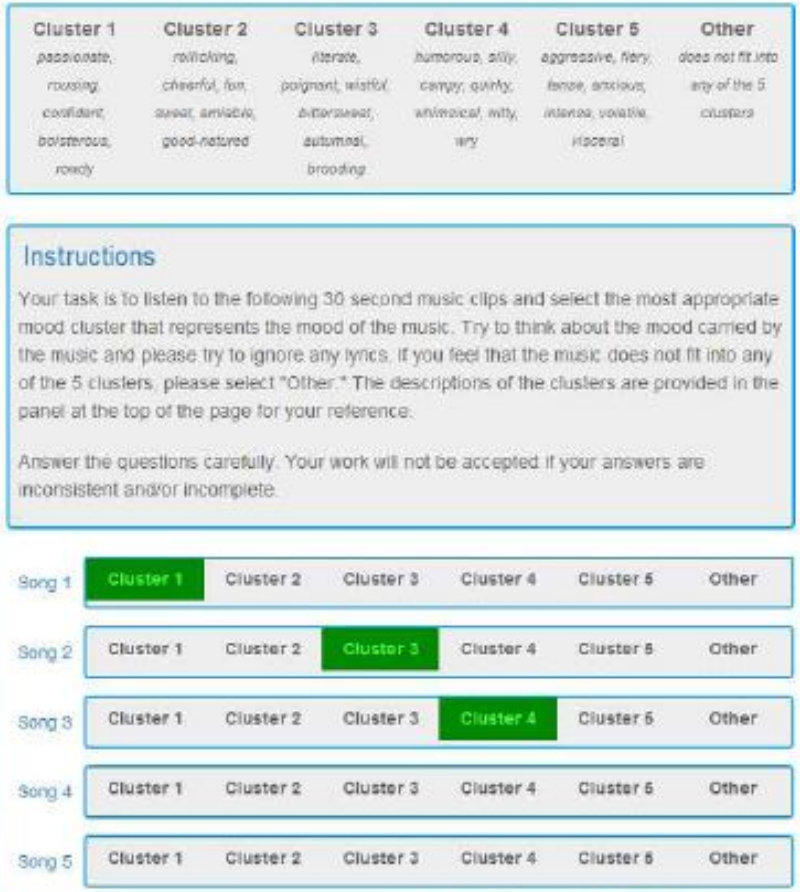
2) Pas validées au bon niveau psychologique

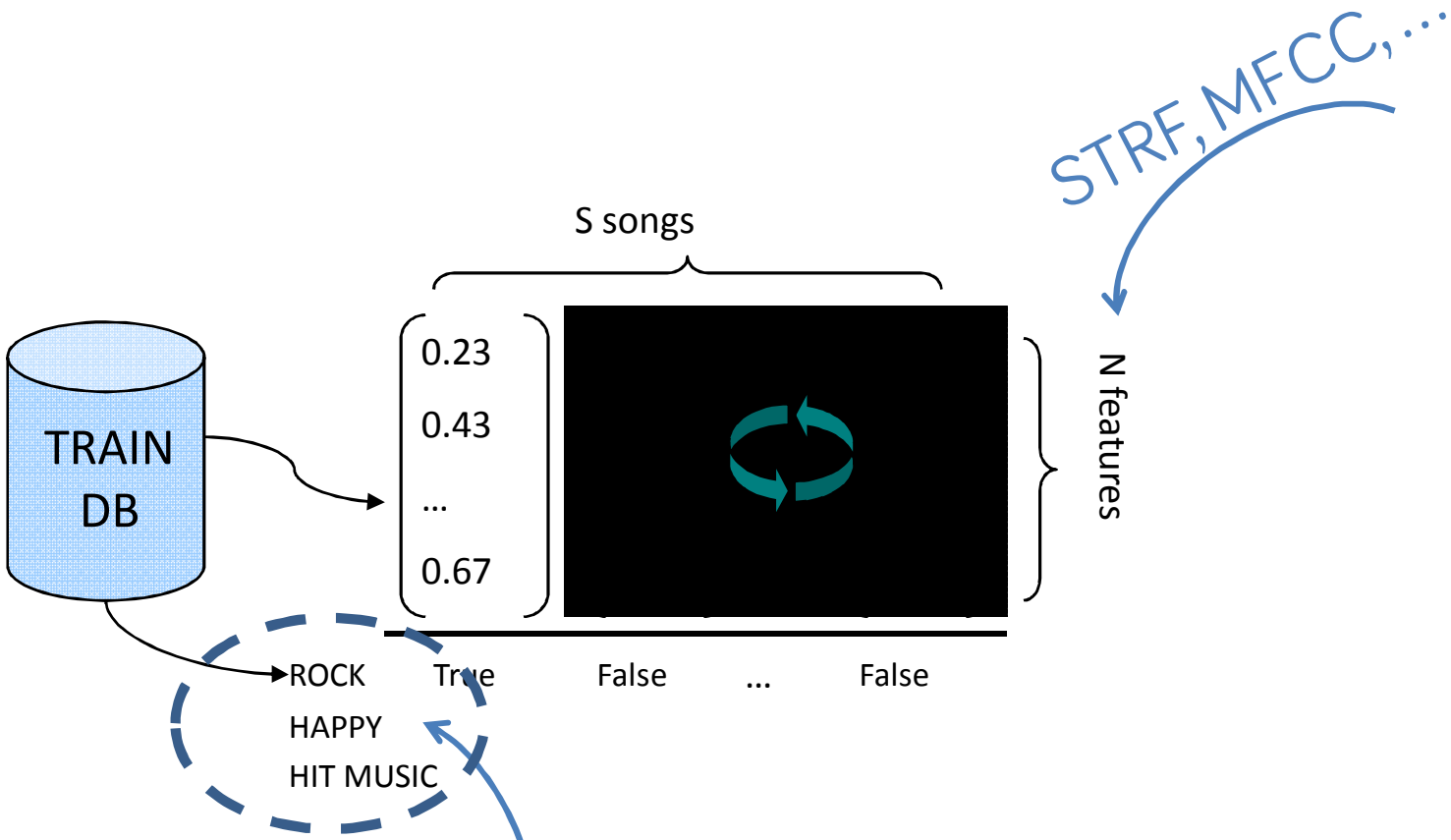

STRF, MFCC, ...
S songs
TRAIN DB
0.23
0.43
...
0.67
N features
ROCK
HAPPY
HIT MUSIC
True
False
...
False
Peut-on leur offrir mieux?

Annotation: Amazon Mechanic Turk

- Human Intelligence Task (HIT)
- Each HIT had 27 clips
- 2 duplicates for consistency check
- Each clips had 2 judges
- Paid 0.55 USD for 1 HIT

- Qualification test before proceeding to task
- 186 HITs collected
- 100 HITs accepted
- Avg. Cohen's kappa: 0.48



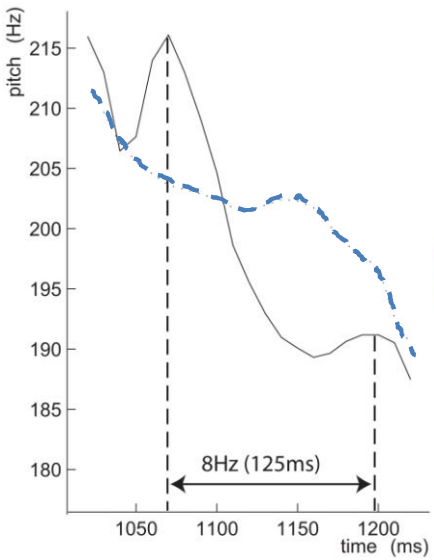


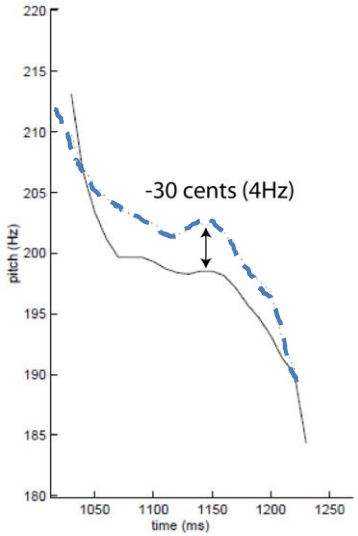
Peut-on leur offrir mieux?

- des paradigmes pré-attentifs (MMN)
- des réponses sous-corticales (CS/UCS)
- des paradigmes inconscients, etc...

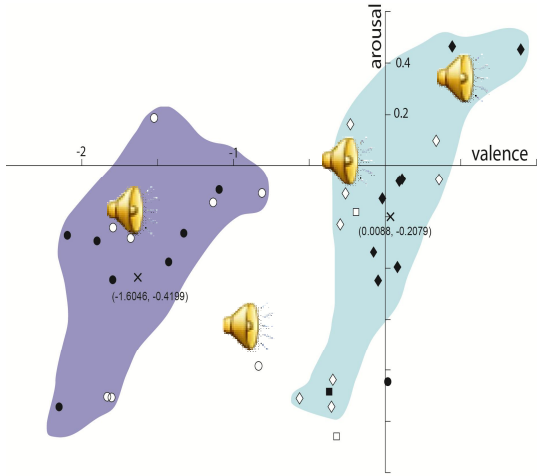
Exemple: voix émotionnelle

Features MIR d'émotionalité (vocale): pitch, vibrato, HFC



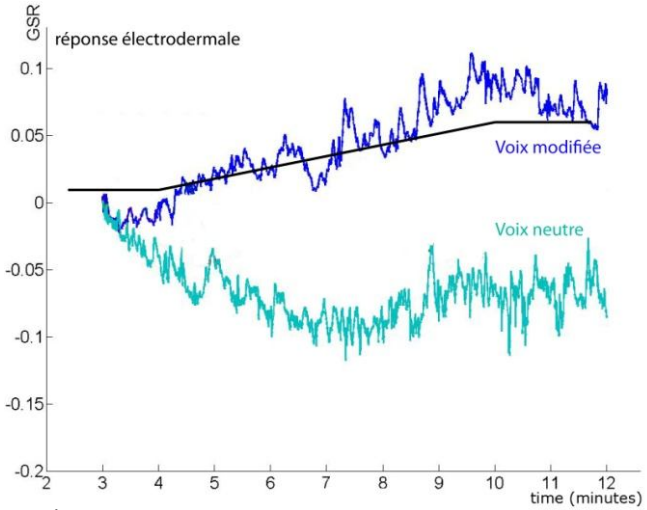


test d'écoute



Baisse les seuils jusqu'à ce que non détecté, mais…

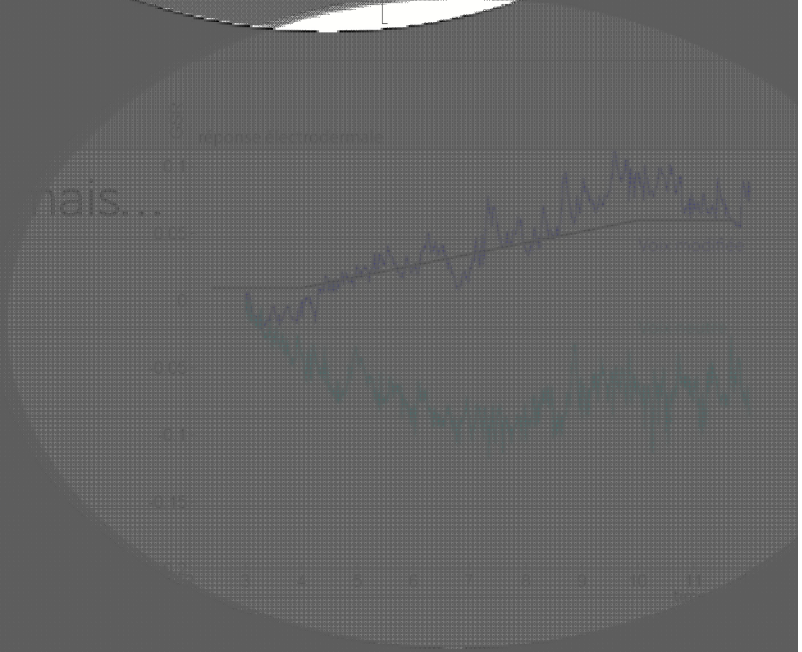
feedback auditif modifié



(avec P. Johansson, L. Mercadie et al., en prép. 2012)

features: MFCC

pitch, vibrato, HFC



mais...

3 pistes pour dépasser tout ça

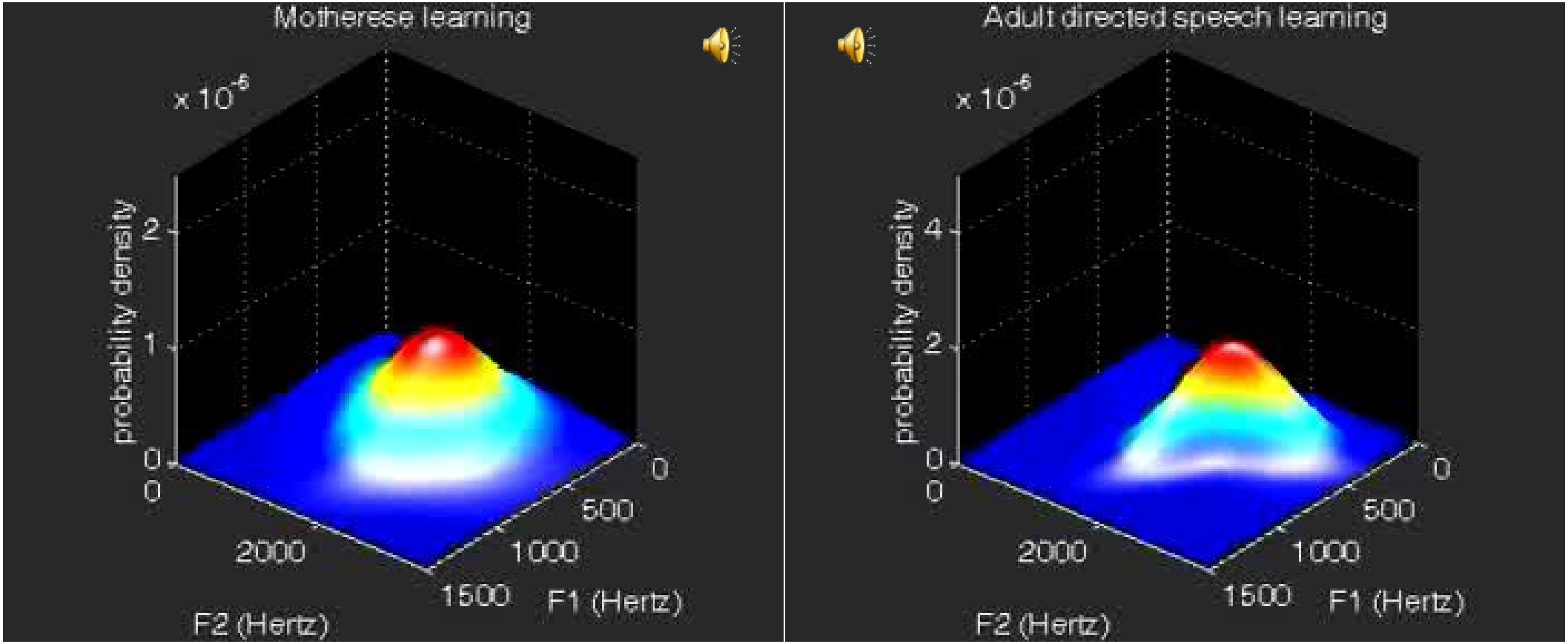
1) Features n'ont pas une modularité neurale/cognitive claire

2) Pas validées au bon niveau psychologique

3) MIR comme modèle de traitement de l'information

Exemple: motherese (de Boer & Kuhl, 2003)

Performances de l'algo supérieures sur infant-directed > adult-directed speech

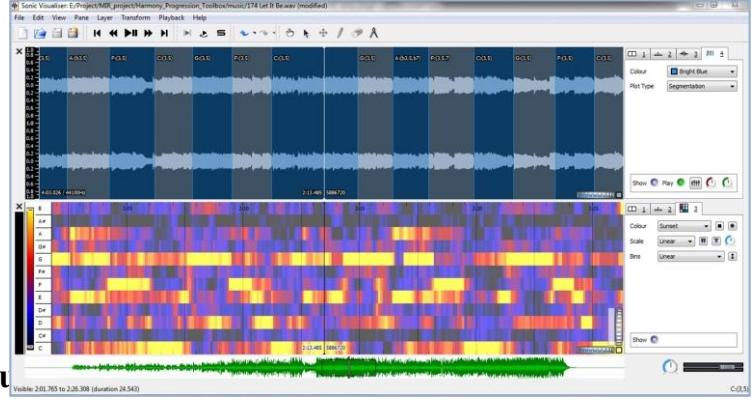


L'information acoustique existe dans le motherese pour rendre l'apprentissage des phonèmes plus facile

	play	fight	alone	stranger	walk	ball
play	6%	13%	4%	25%	40%	13%
fight	1%	74%	0%	14%	6%	6%
alone	4%	7%	15%	43%	16%	16%
stranger	1%	13%	4%	63%	6%	13%
walk	11%	11%	5%	30%	30%	12%
ball	2%	7%	5%	38%	12%	36%

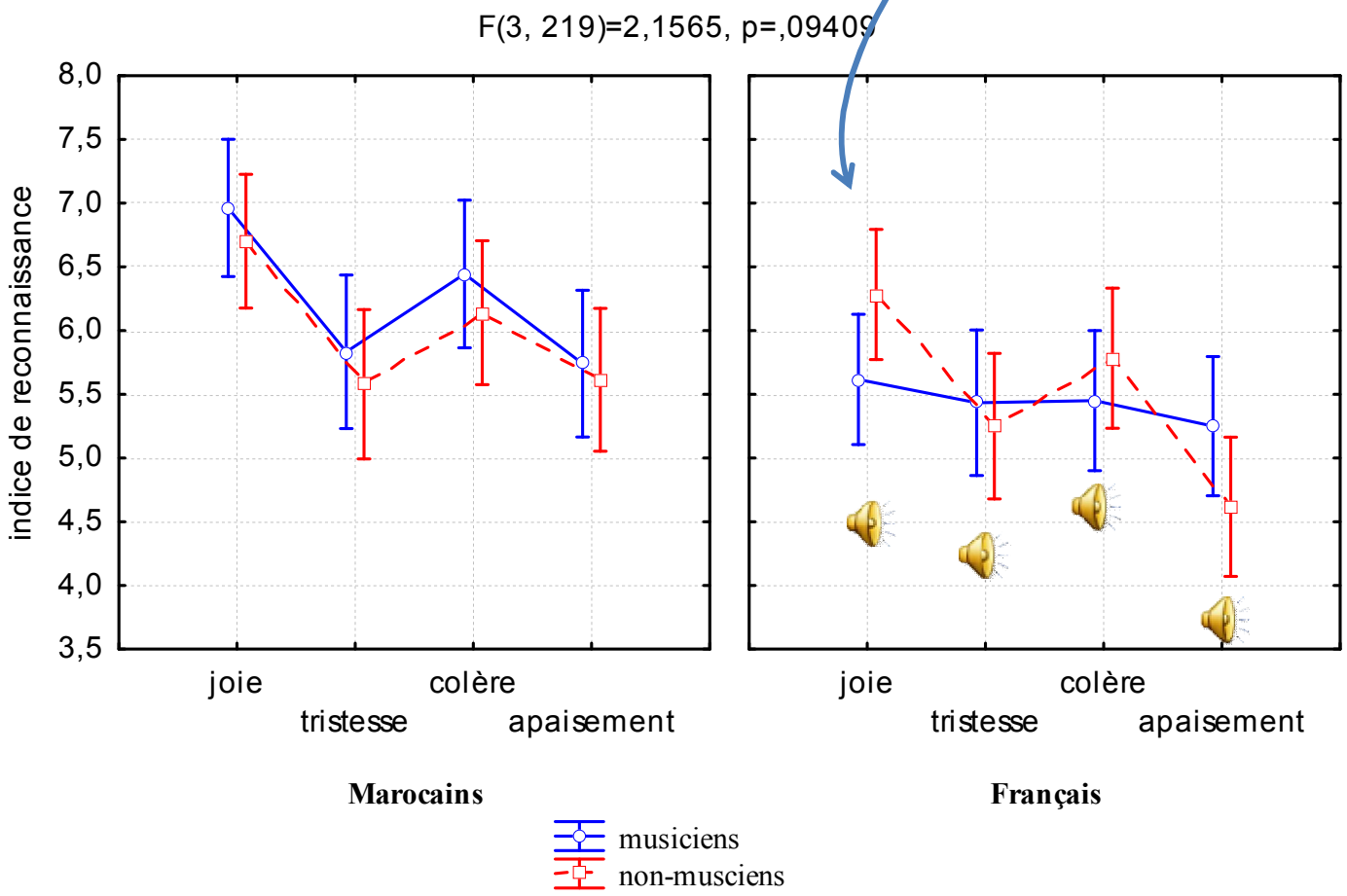
(Molnar, Kaplan et al., 2008)

Reconnaissance d'émotion en musique traditionnelle arabe



Chromagram: traitement de l'information « occidentalisé »

Figure 3: Indice de reconnaissance en fonction des émotions, de la cu d'appartenance et de l'expertise musicale



(avec F. Matallah, E. Bigand, en prép. 2012)

3 pistes pour dépasser tout ça

1) Features n'ont pas une modularité neurale/cognitive claire

collaborer
pour s'en assurer
(cf. STRF)

2) Pas validées au bon niveau psychologique

Leur fournir les bons
Paradigmes expérimentaux

3) MIR comme modèle de traitement de l'information

Mettre ça
dans la boite à outils

Quelques ressources:

- proceedings ISMIR (2000-2012)

http://www.ismir.net/proceedings

- MIRToolbox (gratuit)

https://www.jyu.fi/hum/laitokset/musiikki/en/research/coe/materials/mirtoolbox

- ISMIR review

http://teachingmir.wikispaces.com

- mel & ann

http://jjtok.io/papers/ISMIR-2012.pdf