DII, University of Brescia
Signals and Communication LAB

HARMONIC ANALYSIS FOR MUSIC TRANSCRIPTION AND CHARACTERIZATION

XXVII Cycle, Ph.D. In Telecommunication Engineering

April 9, 2015

Ph.D. Supervisor: Prof. **P. Migliorati**
Ph.D. Coordinator: Prof. **R. Leonardi**

Ph.D. Candidate: **Alessio Degani**

University of Brescia - Communication Technology and Multimedia
Signals and Communication LAB, **Department of Information Engineering (DII)**

Music Information Retrieval (MIR)

MUSIC INFORMATION RETRIEVAL

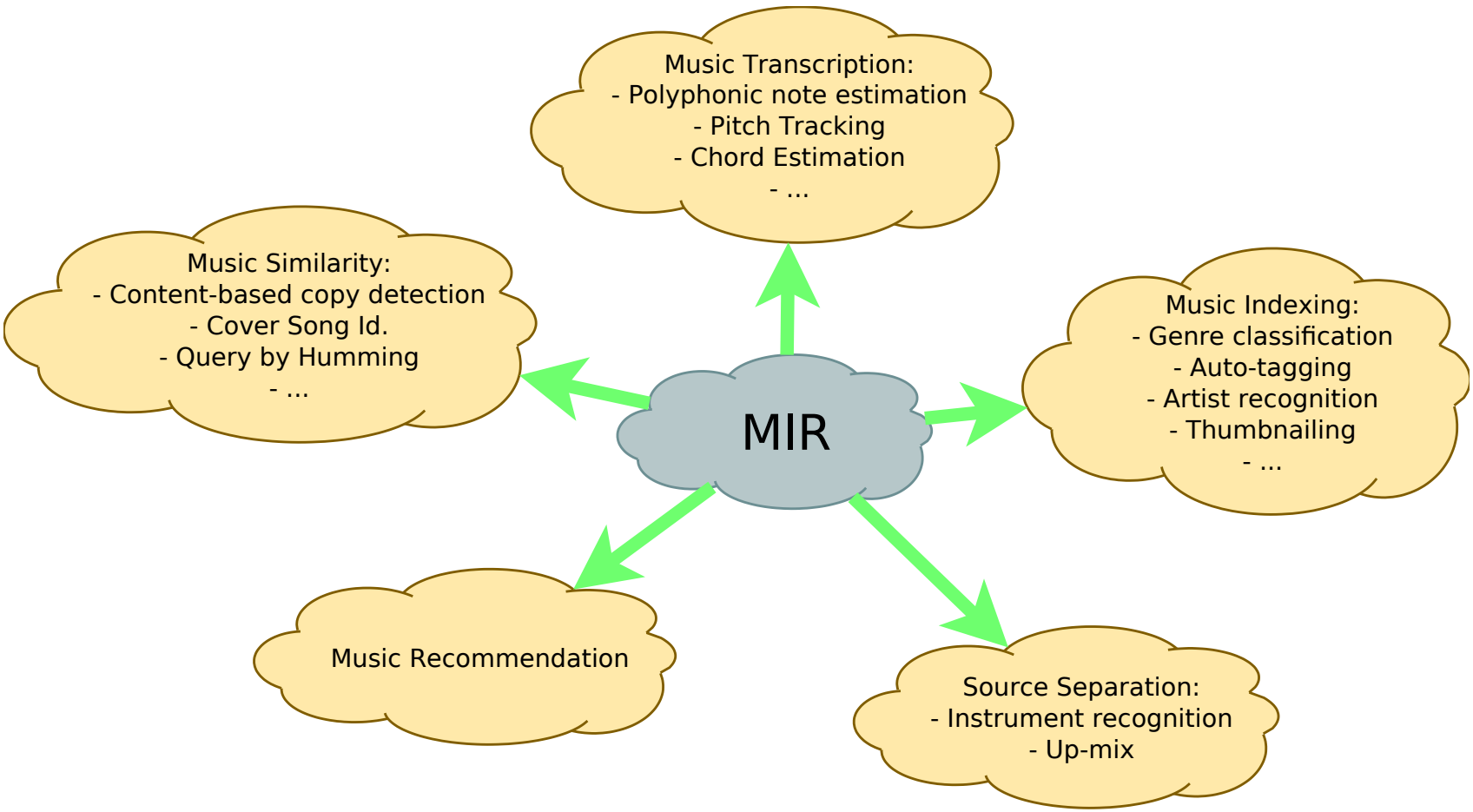
What is MIR?

↳ MIR is a interdisciplinary research field aimed to the study of information extraction methods from musical pieces, retrieval systems and all of the processes involved in those tasks in different music representation media

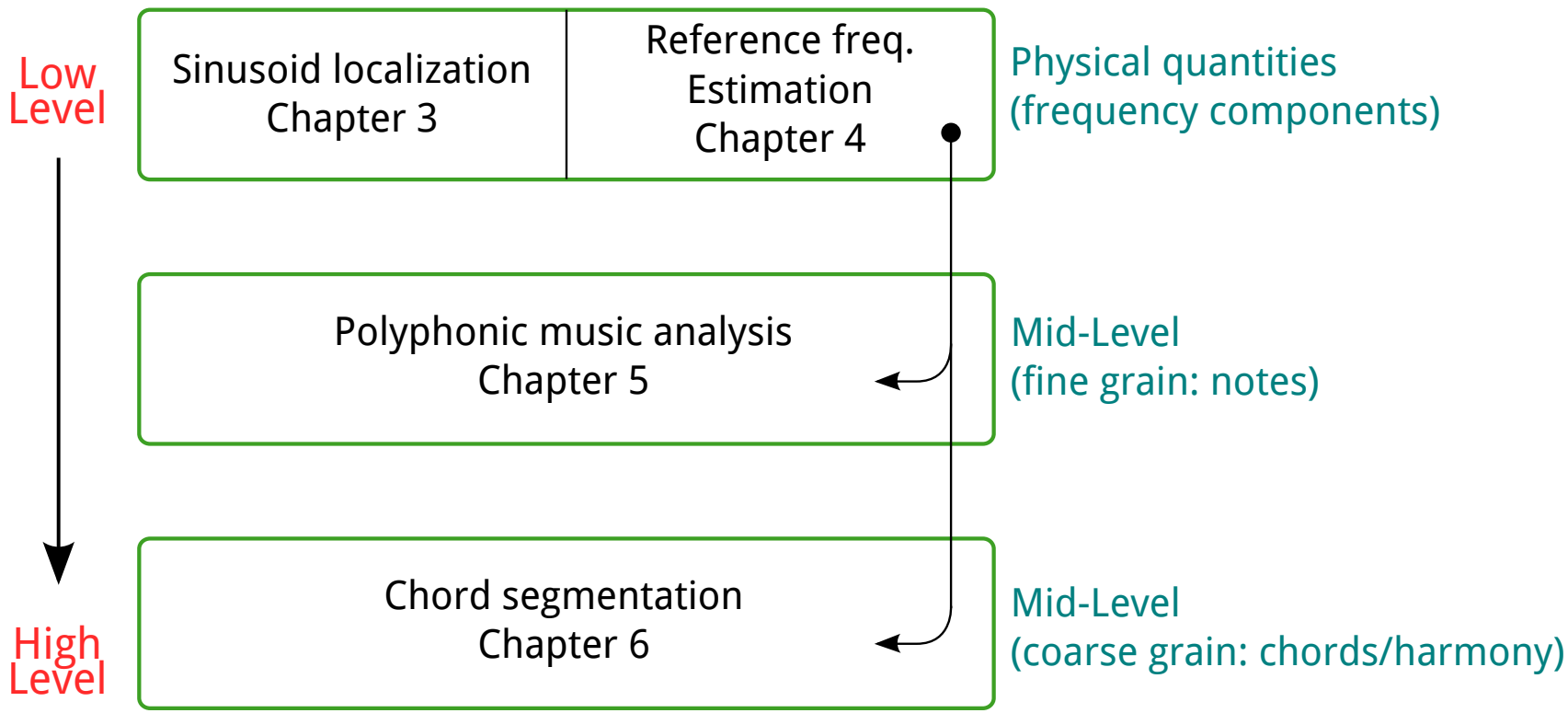
Why MIR?

↳ . . . from the necessity to manage huge collections (databases) of digital music for access, search and retrieve (indexing), preservation, research, . . .

MUSIC INFORMATION RETRIEVAL



OUTLINE


Low Level
High Level
Sinusoid localization Chapter 3
Reference freq. Estimation Chapter 4
Physical quantities (frequency components)
Polyphonic music analysis Chapter 5
Mid-Level (fine grain: notes)
Chord segmentation Chapter 6
Mid-Level (coarse grain: chords/harmony)

Phase-based Sinusoid Localization

STFT PHASE $\Longleftrightarrow$ INSTANTANEOUS FREQUENCY

$\Rightarrow$ The 1^{st}-order derivative of the STFT phase coefficients is directly correlated with the instantaneous frequency $\bar{F}$.

$$\bar{F} = N\frac{\Delta\Phi_{m_0,k_0}}{2\pi\tau}, \qquad \bar{F} \in [k_0 - \tfrac{1}{2}, k_0 + \tfrac{1}{2}]$$

$\Rightarrow$ All coefficients of the STFT are used jointly to build a measure of how likely all the frequency components are, in terms of their phase coherence evaluated in consecutive analysis window:

$\Rightarrow$ Phase Coherence Function

CONCEPT

IDEA

Combine the amplitude spectrum and the Phase Coherence Function to obtain a refined description of the spectral components of an audio signal

⇒ *A.Degani, M.Dalai, R. Leonardi and P.Migliorati*, "Time-Frequency Analysis of Musical Signals Using the Phase Coherence", **DAFx 2013**

STFT PHASE COHERENCE

The evolution of the k-th bin of the STFT as:

$$\hat{X}_{m+1,k} = X_{m,k} \cdot e^{j2\pi f_0 \tau}. \tag{1}$$

The phase evolution is coherent if:

$$\Phi_{m+1,k} - \Phi_{m,k} - 2\pi f_0 \tau = 0 \mod (2\pi). \tag{2}$$

We define the phase coherence measure as:

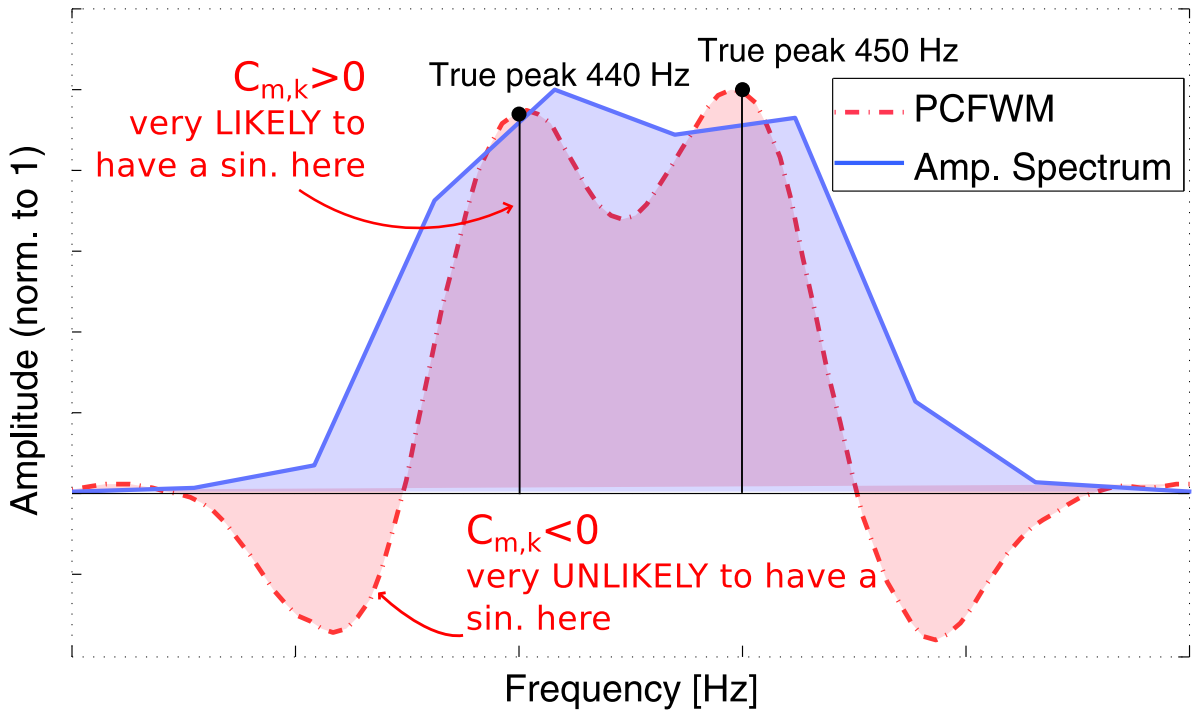
$$\begin{aligned} C_{m,k}(f) &= \cos(\Delta\Phi_{m,k} - 2\pi f \tau) \\ &= \Re\left[\frac{X^*_{m,k}}{|X_{m,k}|} \cdot \frac{X_{m+1,k}}{|X_{m+1,k}|} \cdot e^{-j2\pi f \tau}\right] \end{aligned} \tag{3}$$

COMBINE PHASE COHERENCE AND AMPLITUDE SPECTRUM

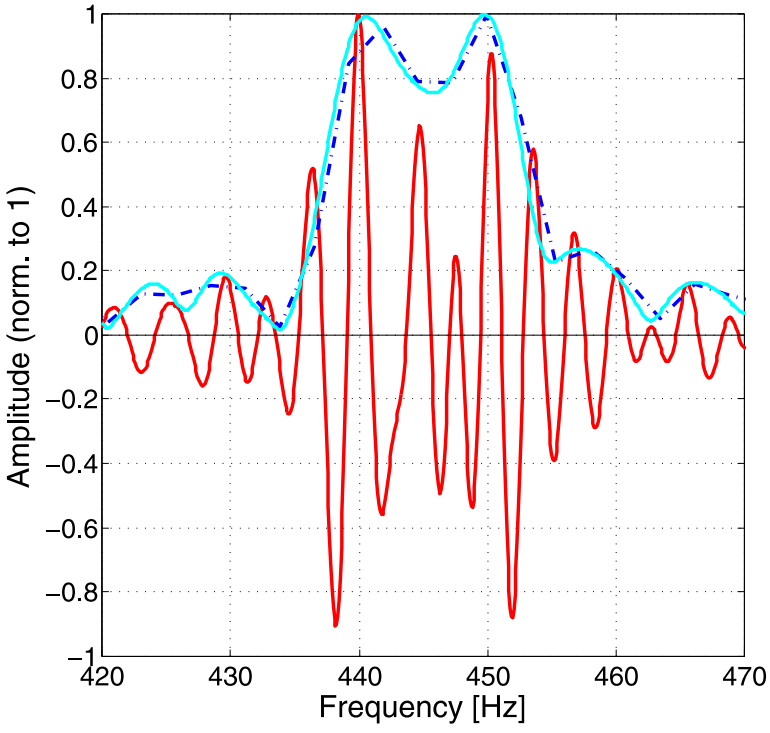
PCFWM ($\bar{\mathcal{X}}$) $\Rightarrow$

$$\bar{\mathcal{X}}_m(f) = \sum_k |X_{m,k}| \cdot W_k(f) \cdot C_{m,k}(f). \quad (4)$$

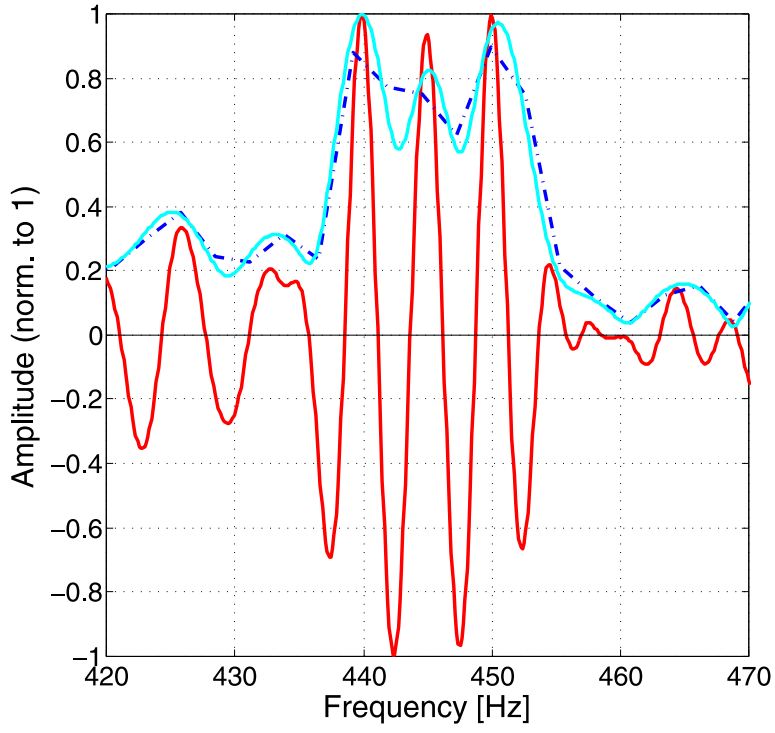
$W_k(f)$ is the DFT of the analysis window modulated at frequency f and calculated at bin k.



ACCURACY: SNR=−10DB, $N = 2048$, $F_S = 5513$ HZ



(a) $\tau = 1536$ samples

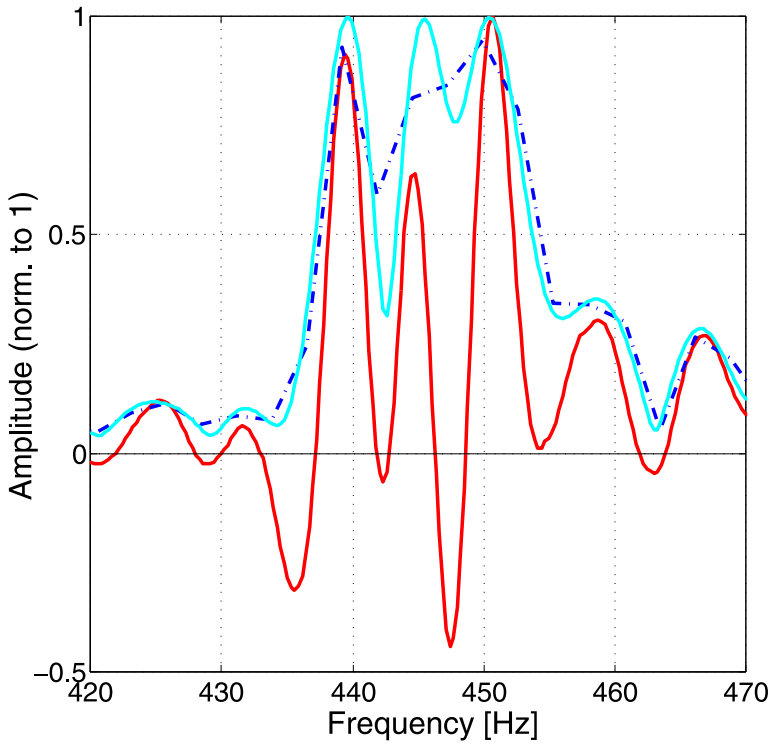


(b) $\tau = 1024$ samples

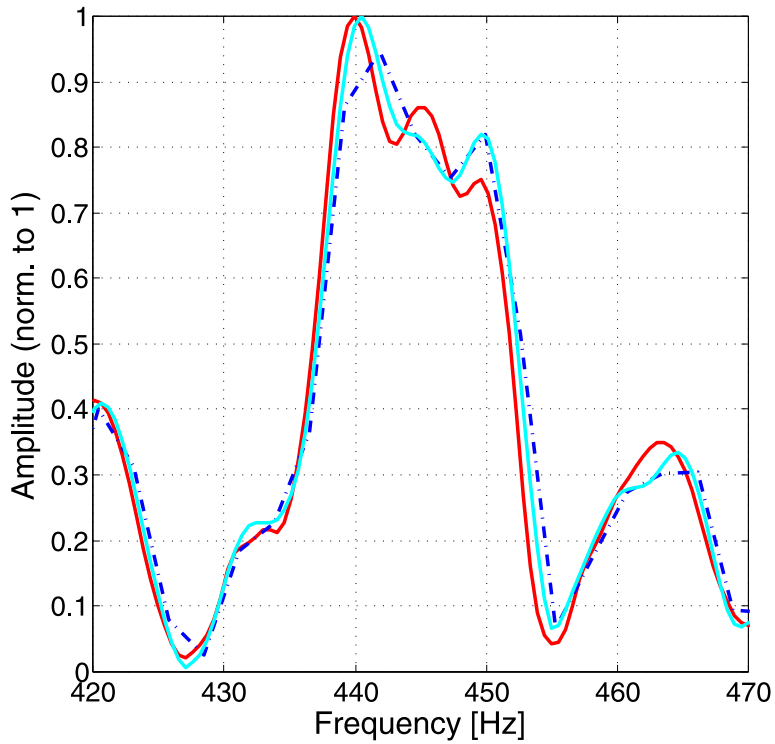
Legend:

– PCFWM, – Interpolated (zero-padding) spectrum,
-.- Amplitude spectrum

ACCURACY: SNR=−10DB, $N = 2048$, $F_S = 5513$ HZ



(c) $\tau = 512$ samples



(d) $\tau = 256$ samples

Legend:

– PCFWM, – Interpolated (zero-padding) spectrum,
-.- Amplitude spectrum

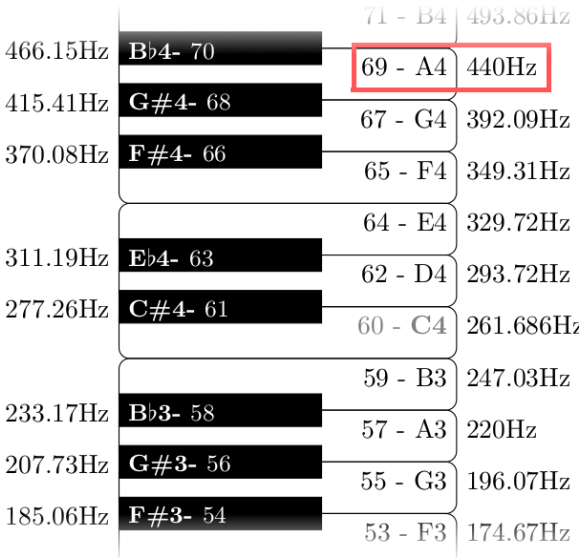
Reference Frequency Estimation

THE EQUAL TEMPERAMENT

Temperament

Set of rules that defines the correspondence between Note names and Frequencies

Equal Temperament: divides the octave in 12 equally spaced (in the logarithmic *cents* scale) intervals $\Rightarrow$ every pair of adjacent notes has an identical frequency ratio ($\sqrt[12]{2}$)



$$c = 1200 \cdot \log_2 \left(\frac{f}{f_{ref}} \right)$$

$$M = 69 + round \left[\frac{c}{100} \right]$$

WHY WE NEED TO KNOW F_{REF}? (1)

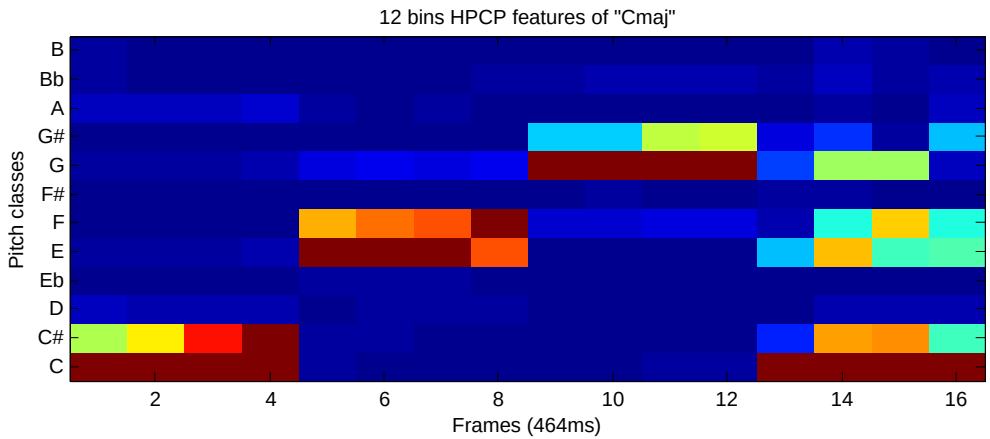
Monitor the tuning of an *A cappella* choir execution

Restore old recordings that suffer of *wow & flutter*

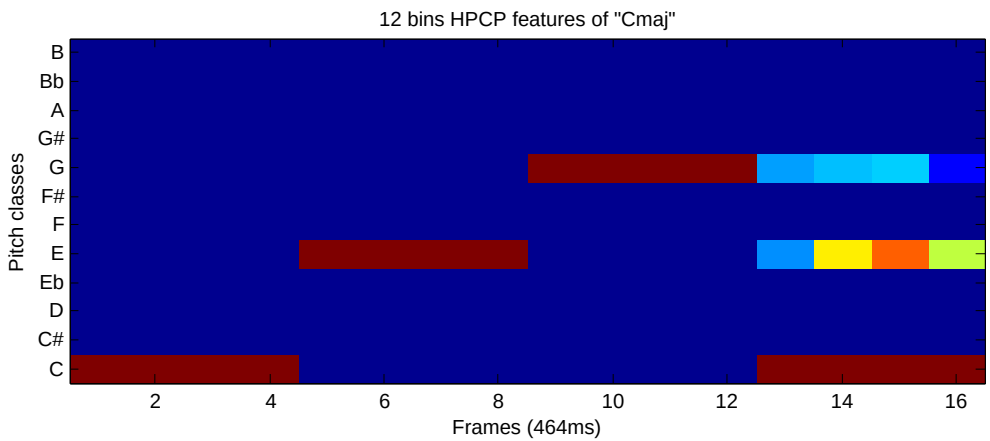
$\Rightarrow$ some applications have real-time constraints!

WHY WE NEED TO KNOW F_{REF}? (2)

Pitch Class Profile (PCP)



f_{ref} estimation



ALGORITHMS

f_{ref} has to be estimated from the data!

Estimation methods:

- ↳ Deviations Histogram[1]
- ↳ Circular Statistics[2]
- ↳ Least-Square Optimization[3]

---

[1] Emilia Gómez. "Tonal Description of Music Audio Signals". PhD thesis. Universitat Pompeu Fabra, Barcelona, Spain, 2006.

[2] Karin Dressler and Sebastian Streich. "TUNING FREQUENCY ESTIMATION USING CIRCULAR STATISTICS". . In: *Proc. of the 8th Int. Conf. on Music Information Retrieval, (ISMIR)* (2007), pp. 357–360.

[3] Volker Gnann et al. "Least-Squares Local Tuning Frequency Estimation for Choir Music". In: *Proc. of the 131st AES Convention* (2011).

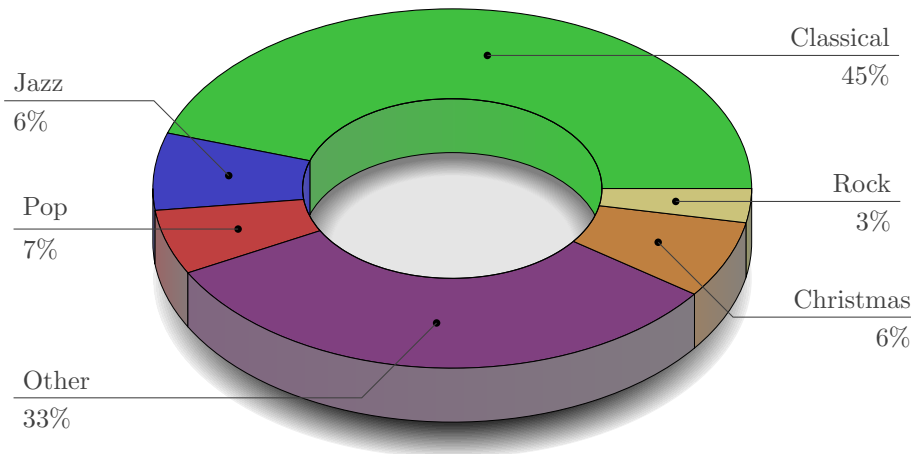
EVALUATION

- **Estimation reliability**
  - Saw sweep $s(t) = \frac{1}{H} \sum_{h=1}^{H} \frac{1}{h} \sin[2\pi h(f_0 t + \alpha t^2)]$
  - Estimation std. deviation (HYP: $f_{ref} = const$) Σ_p using $p\%$ of the song's frames
  - Datasets: *CS80* and *MS2012*
- **Local tuning estimation performances**
  - HYP: $f_{ref} \neq const$
  - Simulation of a Real-Time estimation
  - An estimation every 2 or 4 seconds
- **Computation time**
  - Using MATALB®Profiler on a Intel i5 M430 CPU, 64 bit GNU/Linux laptop with 4 GB of RAM

OUR PROPOSED MS2012 DATASET

musescore - 306 songs - Creative Commons

MIDI + MusicXML



User annotated genre tags

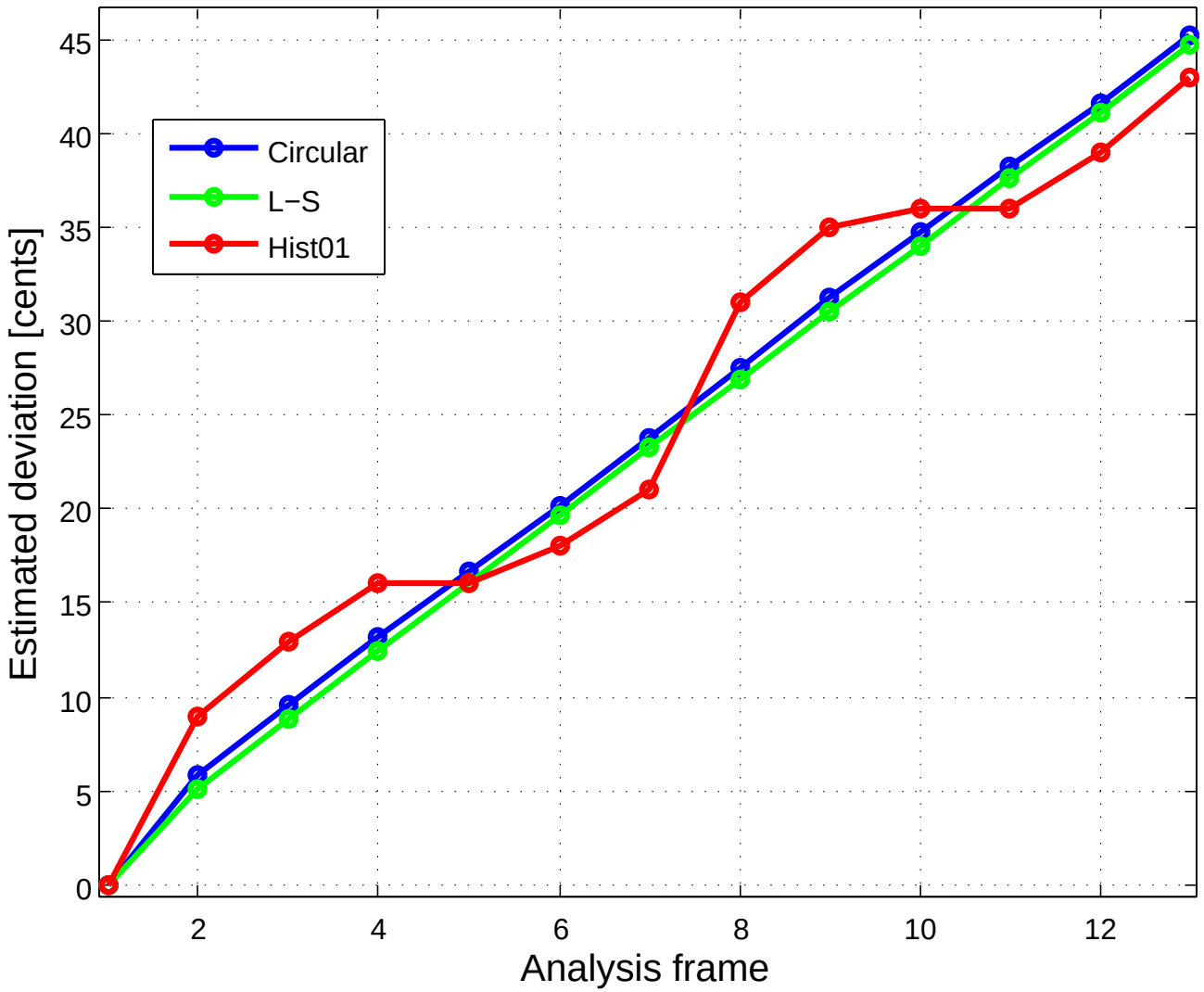
Table: Instruments

Instrument	%
Acoustic Grand Piano	26%
String Ensemble 1	19%
Choir Aahs	16%
Flute	15%
Clarinet	14%
French Horn	13%
Trumpet	12%
Trombone	11%
Violin	10%
Acoustic Guitar nylon	9%

http://www.ing.unibs.it/alessio.degani/?p=ms2012

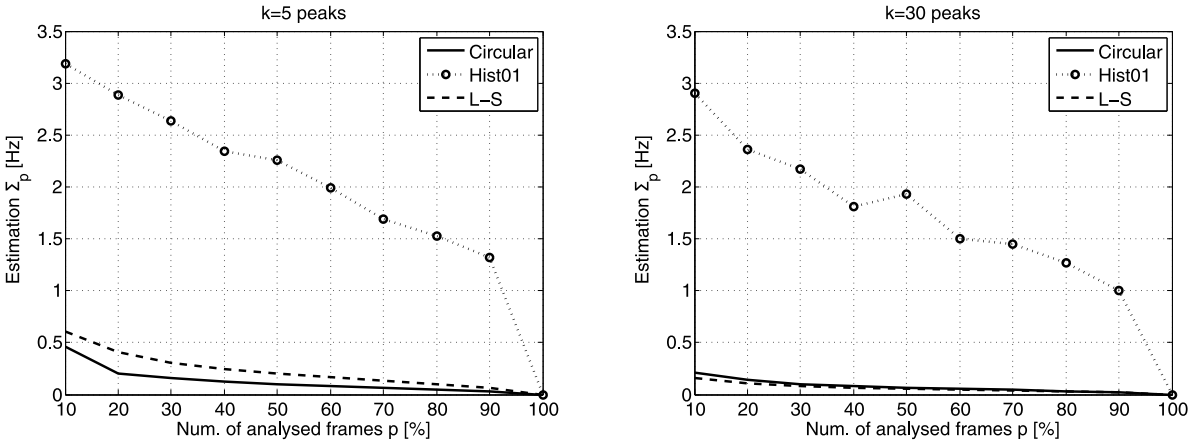
RESULTS - EST. RELIABILITY

Saw sweep results

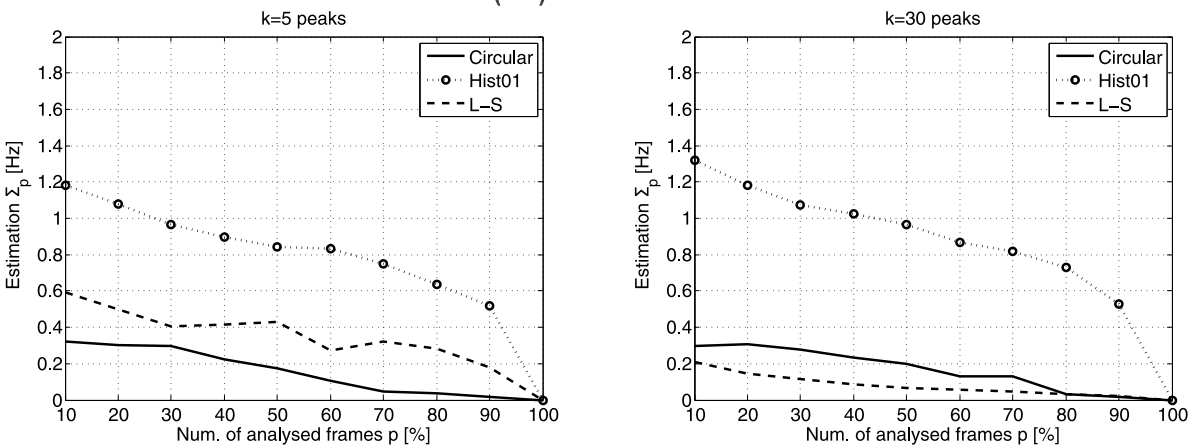


RESULTS - EST. RELIABILITY

Est. std deviation Σ_p using $k = 5$ and $k = 30$ peaks per frame



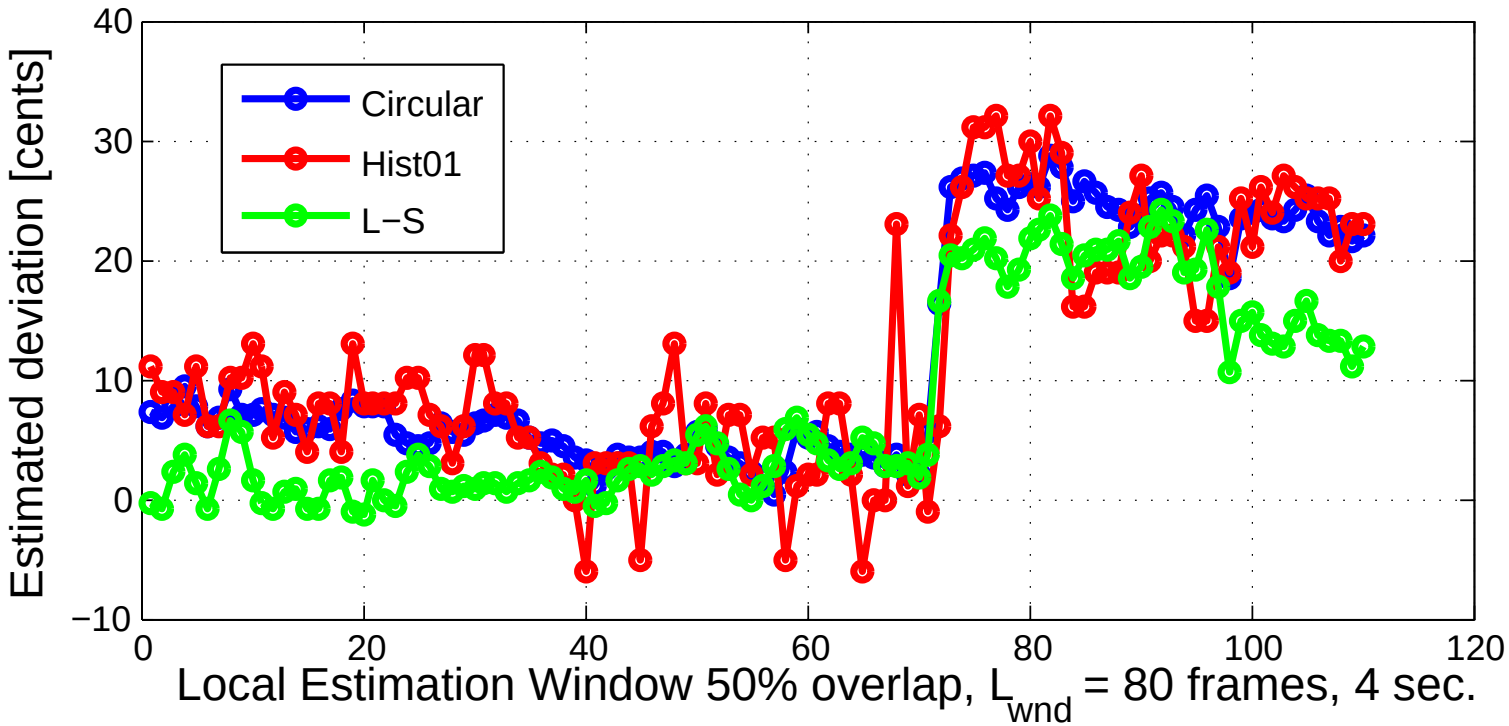
(a) CS80



(b) MS2012

RESULTS - LOCAL TUNING EST.

"Variations 16-20" in J.S. Bach, Goldberg Variations, BWV 988, played by Wanda Landowska, Paris (1933), CD version of 78 rpm recording



RESULTS - COMPUTATION TIME

Algorithm	Average Time (s)
Circular	0.19
Hist01	2.63
L-S	1.15

Table: Average execution time per song

RESULTS - CONCLUSION

The lack of a scientific ground truth makes the evaluation of the performances of tuning estimation algorithms a non trivial task $\Rightarrow$ we proposed a set of tests that allow to compare different algorithms in terms of desirable behaviours.

- ↳ All the considered algorithms perform quite well in the global f_{ref} estimation
- ↳ Remarkable difference in terms of estimation stability between the *Hist01* algorithm and the others $\Rightarrow$ *Circ* and *L-S* more suitable for real-time local frequency estimation
- ↳ For *Circ* and *L-S*, k = 5 peaks per frame are sufficient for all of estimation tasks

Polyphonic Music Analysis

POLYPHONIC PITCH SALIENCE

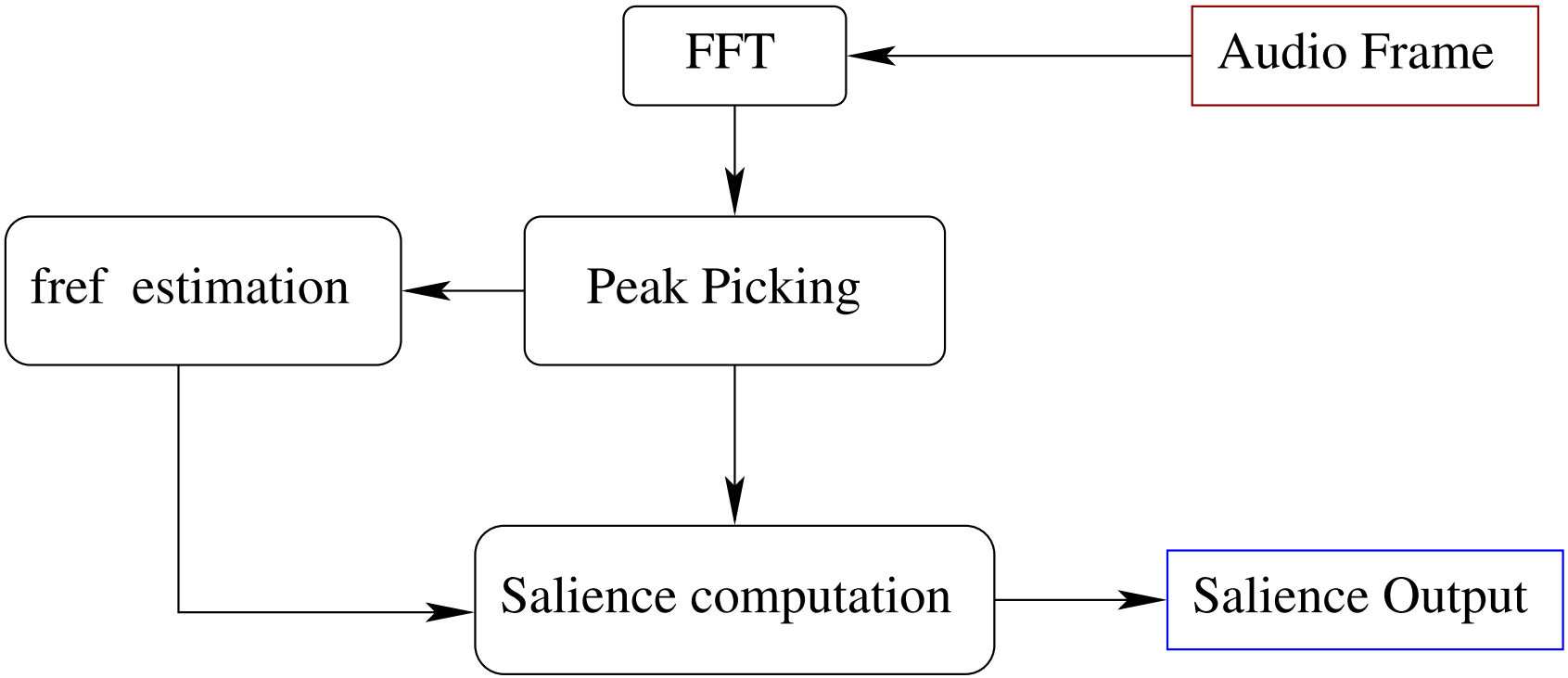
Pitch Salience

Estimates the relevance of the most salient musical pitches (here considered synonym fundamental frequencies) that are present in a certain audio excerpt

Classic Approach:

$$S[k] = \sum_{h=1}^{H} w_h |X[hk]| \tag{5}$$

PITCH SALIENCE COMPUTATION: OVERVIEW



⇒ *A.Degani, M.Dalai, R. Leonardi and P.Migliorati*, "Comparison of Tuning Frequency Estimation Methods", **Journal of Multimedia Tools and Applications, *Springer***

PARTIALS DEVIATION

$$d_h^{f_0} = 12 \log_2 \left(\frac{h\, f_0}{f_{ref}} \right) - round \left[12 \log_2 \left(\frac{h\, f_0}{f_{ref}} \right) \right] \qquad (6)$$

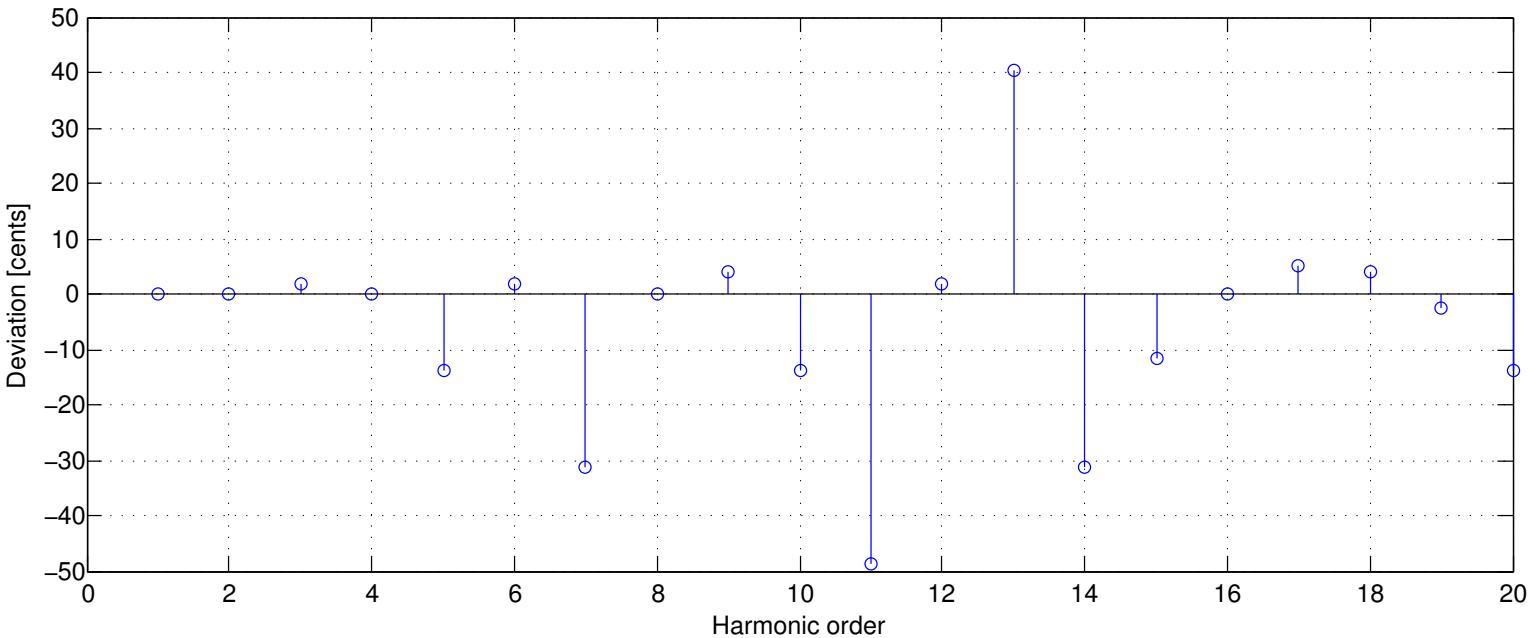
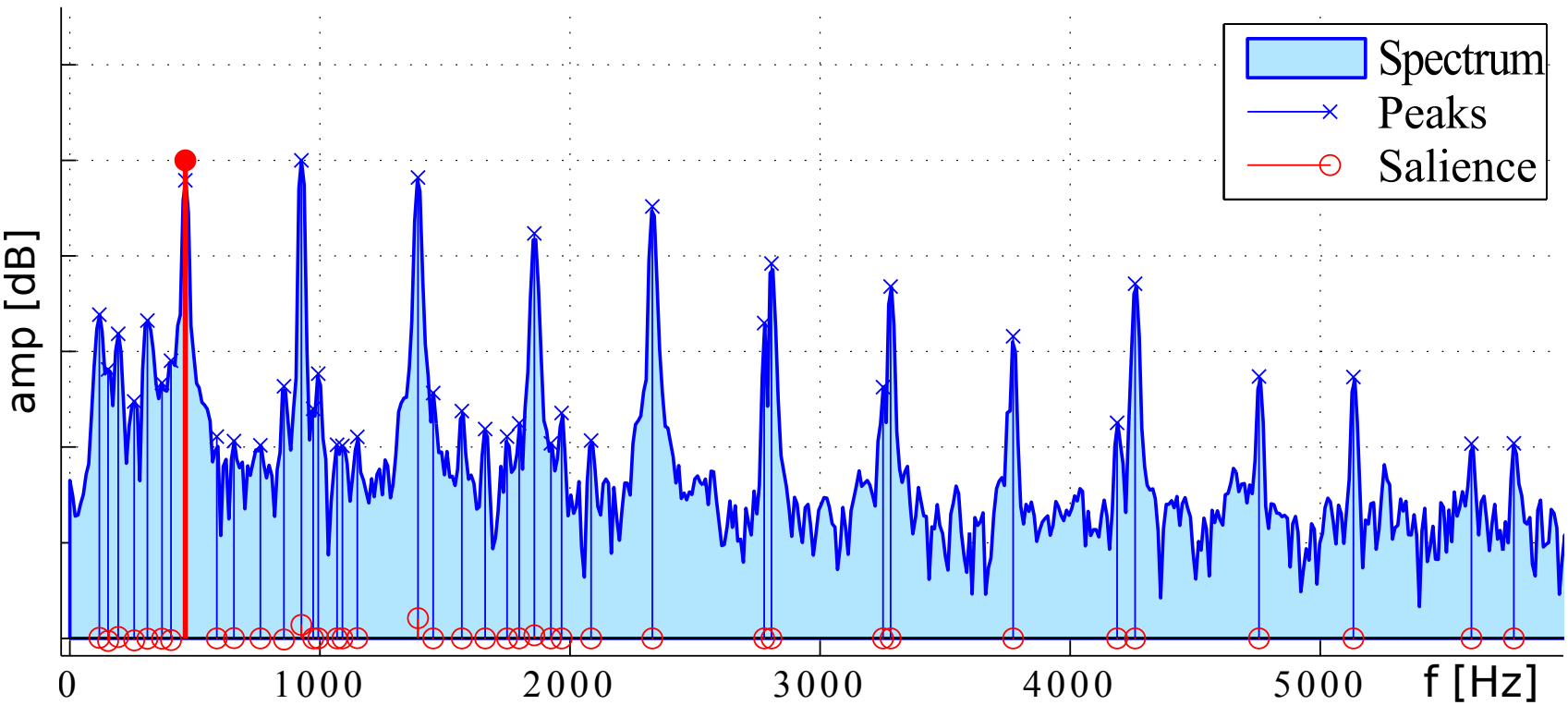


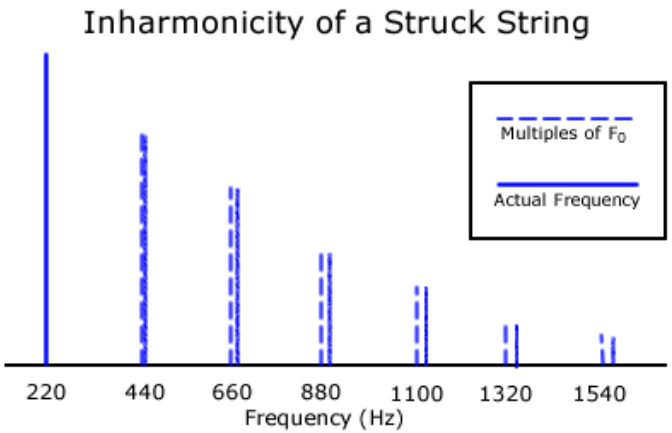
Figure: Deviation of the first 20 partials of a complex tone

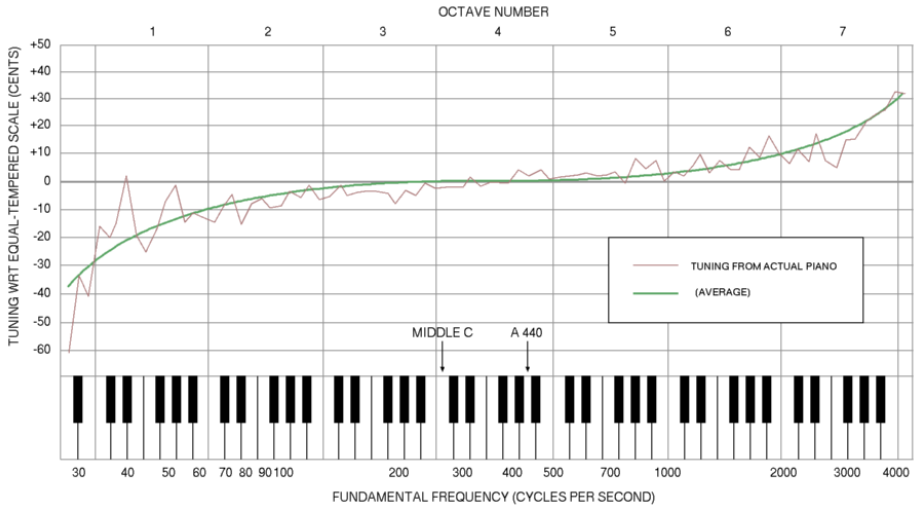
SALIENCE FUNCTION

$$\Rightarrow \quad S_p = a_p \sum_{h=1}^{H} d_h \cdot \hat{d}_h^{f_p} \tag{7}$$



INHARMONICITY...





Inharmonicity model (one of...)

$$f_h = h f_0 \sqrt{(1 + \beta h^2)}, \qquad \beta = \frac{\pi^3 Q d^4}{64 l^2 T} \tag{8}$$

IN CASE OF INHARMONICITY? (E.G. PIANO SOUND)

Considering an inharmonic model for partials deviations

$$\forall h > 1$$

$$d_h^{f_0}(\beta) = 12\log_2\left(\frac{hf_0\sqrt{(1+\beta h^2)}}{f_{ref}}\right) + \\ - round\left[12\log_2\left(\frac{hf_0\sqrt{(1+\beta h^2)}}{f_{ref}}\right)\right] \quad (9)$$

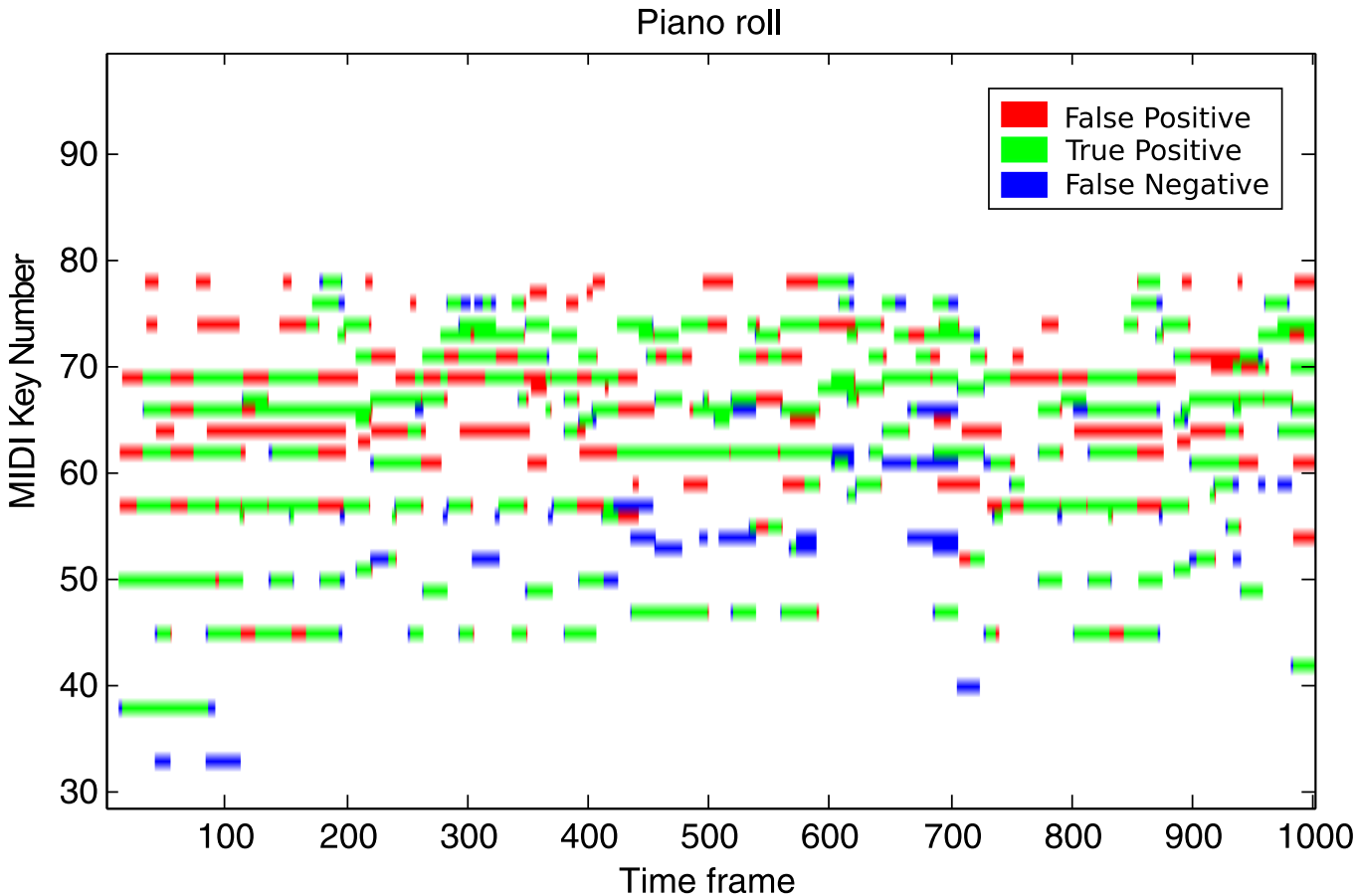
EVALUATION

Problem!

No standard evaluation procedure for **Salience Measures**
$\Rightarrow$ Evaluated as a **Multi-Pitch** estimation algorithm

Experiments are performed on the MIDI Aligned Piano Sounds test-set. **MAPS** provides CD quality piano recording (44.1 kHz, 16-bit). This test-set is available under Creative Commons license and consists of about 40GB (65 hours) of audio files recorded using both real and synthesized pianos. The aligned ground-truth is provided as MIDI

EVALUATION



$$P = \frac{TP}{TP + FP}, \; R = \frac{TP}{TP + FN}, \; F = \frac{2PR}{P + R} \qquad (10)$$

RESULTS

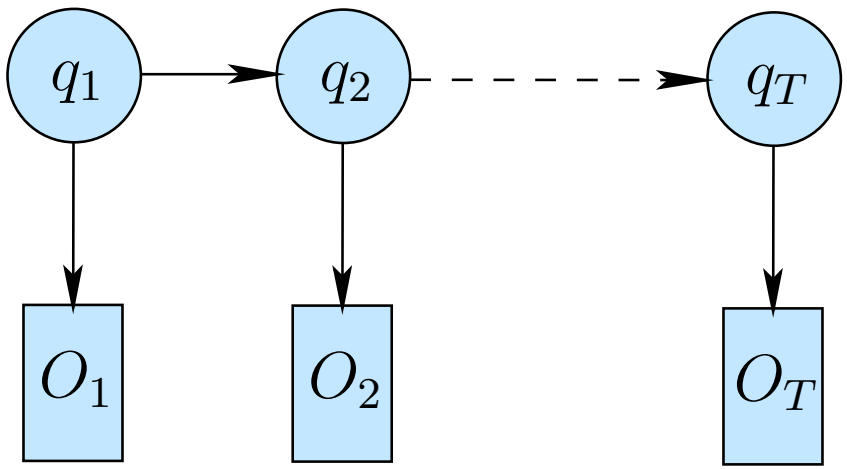
	Salience (chroma accuracy)	Salience (est. β)	Salience ($\beta = 0$)	Peaks
Precision	0.78	0.56	0.42	0.19
Recall	0.63	0.44	0.46	0.75
F-measure	0.70	0.49	0.44	0.31

- ↳ **T** is a key parameter for the *Precision*/*Recall* trade-off trade-off (same *Recall* as "peaks" but $+15\%$ *Precision*)
- ↳ Octave errors (!)
- ↳ Not a complete polyphonic pitch estimation method, but a pre-processing procedure $\Rightarrow$ maximize *Recall* and then filter out the FP

Chord Segmentation and Estimation

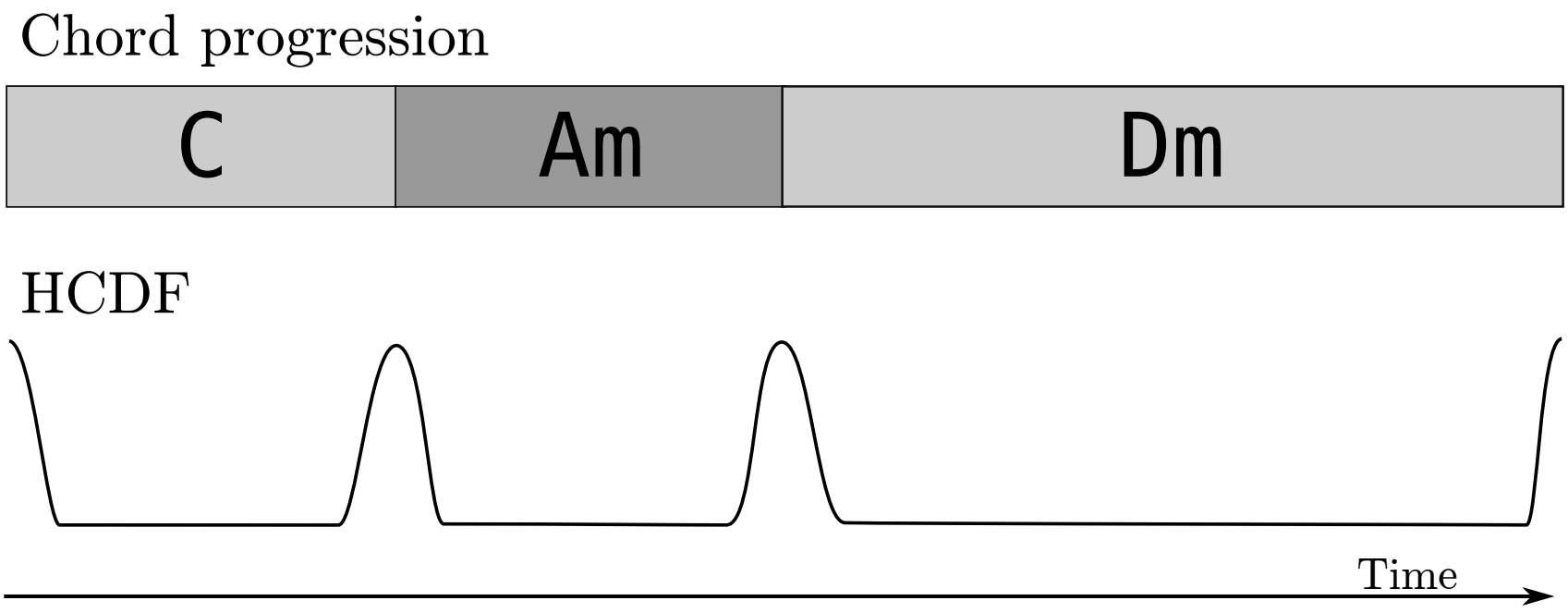
CHORD ESTIMATION

In Audio Chord Estimation (ACE), Chord Segmentation is "done" frame-wise by the estimation process (i.e. **HMM**)
$\Rightarrow$ No ad-hoc segmentation procedure



CHORD SEGMENTATION

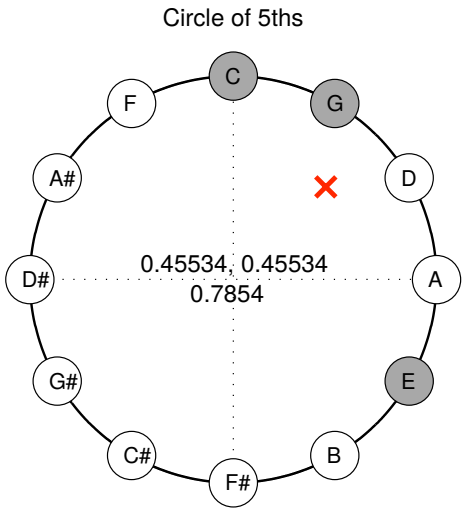
A Chord Segmentation strategy can be beneficial to the estimation process ⇒ Harmonic Change Detection Function (HCDF)

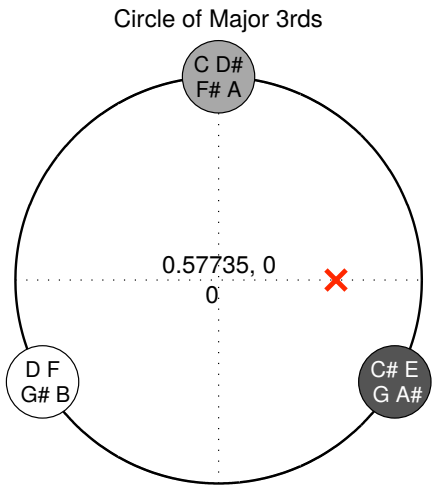


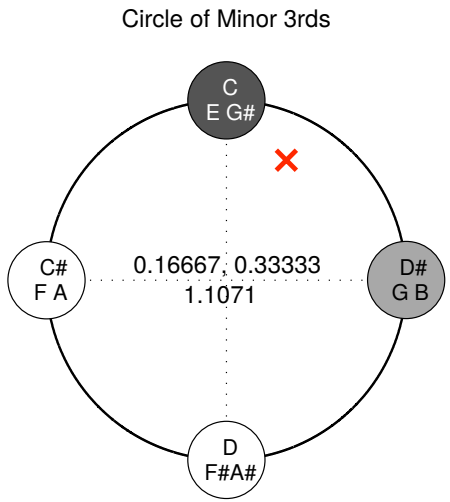
HARTE ET AL. APPROACH

Harte et al.[4] approach:

- ↳ Calculate PCP
- ↳ Map PCP to Tonal Centroid (small HC ⇒ small distance)
- ↳ Euclidean distance between adjacent TC vectors (HCDF)
- ↳ HCDF Peak finding ⇒ Chord bounds







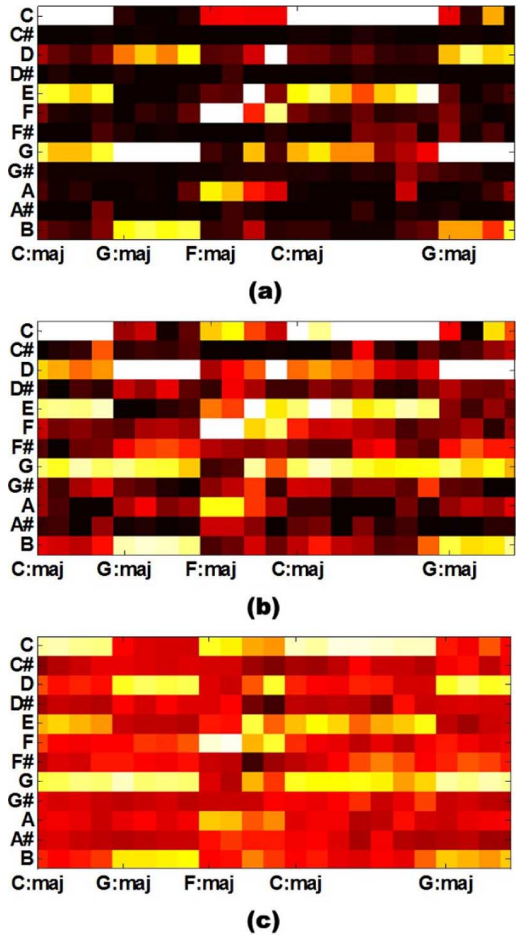
[4] Christopher Harte, Mark Sandler, and Martin Gasser. "Detecting Harmonic Change In Musical Audio". In: *Proceedings of th 1st ACM workshop on Audio and music computing multimedia* (2006), pp. 21–26.

DIFFERENT FEATURES/DISTANCES

PCP used by Harte et al. are “simple”

Our approach:

- ↳ Use advanced PCP (NNLS(a), LBC(b) , CRP(c), HPCP)
- ↳ Use different distance measure (Correlation, Cosine)
- ↳ Study the effect of TC mapping
- ↳ Evaluate how good is the segmentation and how the ACE accuracy is influenced



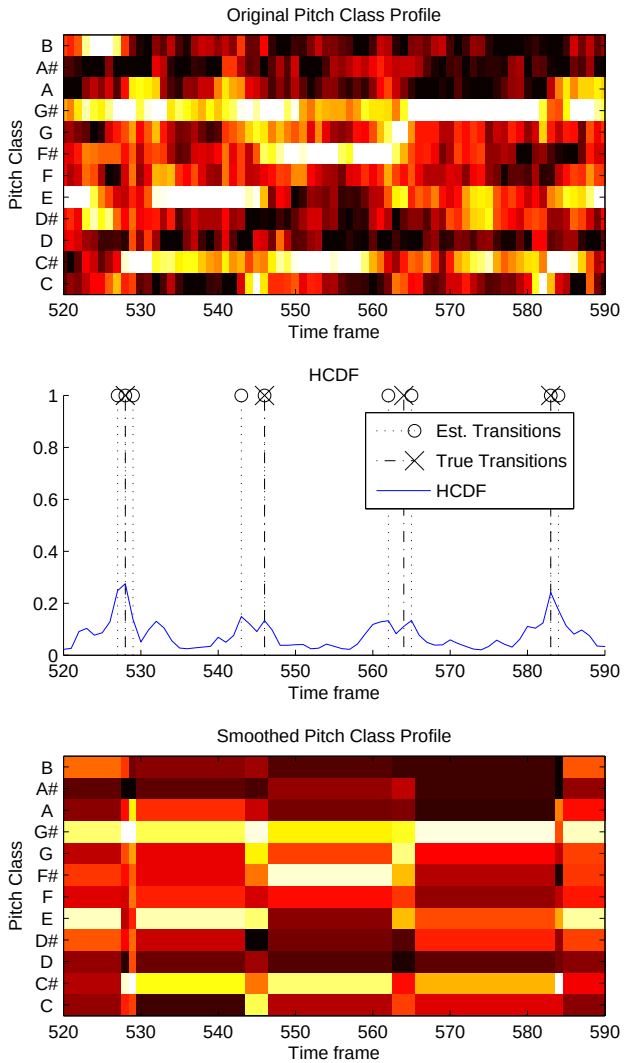
EVALUATION

Chord bounds detection accuracy

- ↳ Precision, Recall, F-Measure
- ↳ Small dataset (16 Beatles songs)
- ↳ Comparison (Harte's and Hainsworth's)

ACE accuracy

- ↳ With and without Chord bound-based smoothing
- ↳ Beat-based smoothing
- ↳ 176 Beatles songs



ACCURACY

Type	Distance	Precision	Recall	F-Meas.
CRP	Euclidean	0.78	0.8	0.79
CRP	**Cosine**	**0.67**	**0.96**	**0.79**
LBC	Euclidean	0.67	0.77	0.72
LBC	**Cosine**	**0.62**	**0.96**	**0.75**
Hainsworth et al.	-	0.31	0.88	0.46
Harte et al.	Euclidean	0.53	0.84	0.65

Method details	Mirex2010	Relative improvement
CRP, no smooth	0.659	-
CRP, smooth	0.670	1.1%
LBC, no smooth	0.752	-
LBC, smooth	**0.765**	**1.3%**
LBC, beatsynch	0.754	0.2%

CONCLUSION

- ↳ Advanced PCP are better for Chord Bound Detection
- ↳ TC + Euclidean ⇒ Good *Recall* but poor *Precision* and *F-Measure*
- ↳ Cosine distance (best *F-Measure*) ⇒ Good *Recall* without sacrifice *Precision* to much
- ↳ For ACE, Recall value must be optimized for bounds detection
- ↳ **Chord Change Based Smoothing better than Beat Based Smoothing**

⇒ *A.Degani, M.Dalai, R. Leonardi and P.Migliorati*, "Harmonic Change Detection for Musical Chords Segmentation", **ICME 2015** (to appear)

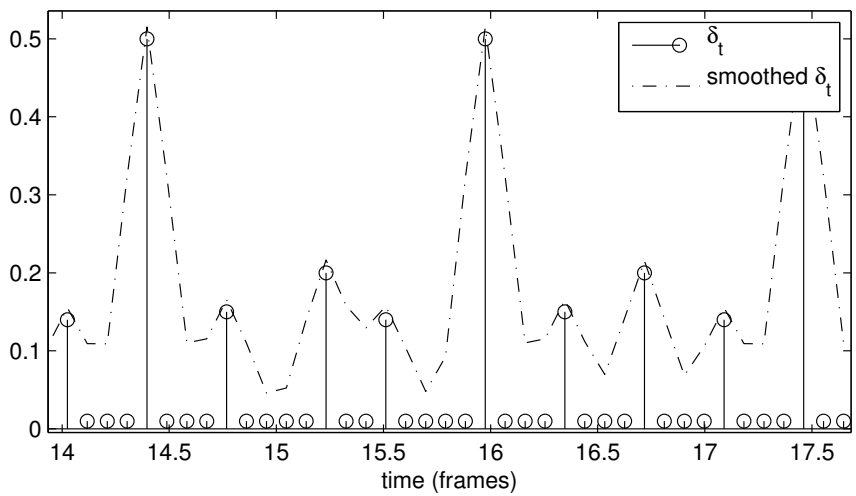
Current Research

WHY MUSICAL METER?

Musical meter structure of a song have been proven to be beneficial for Audio Chord Estimation

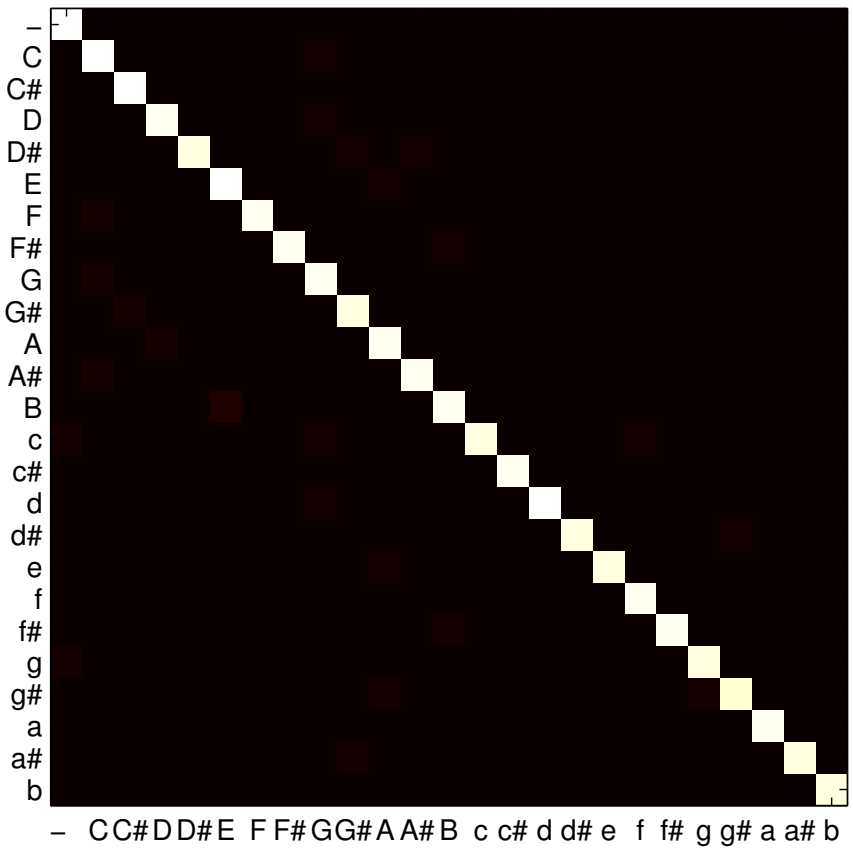
IDEA: Chord transitions are more likely on the first beat of a measure

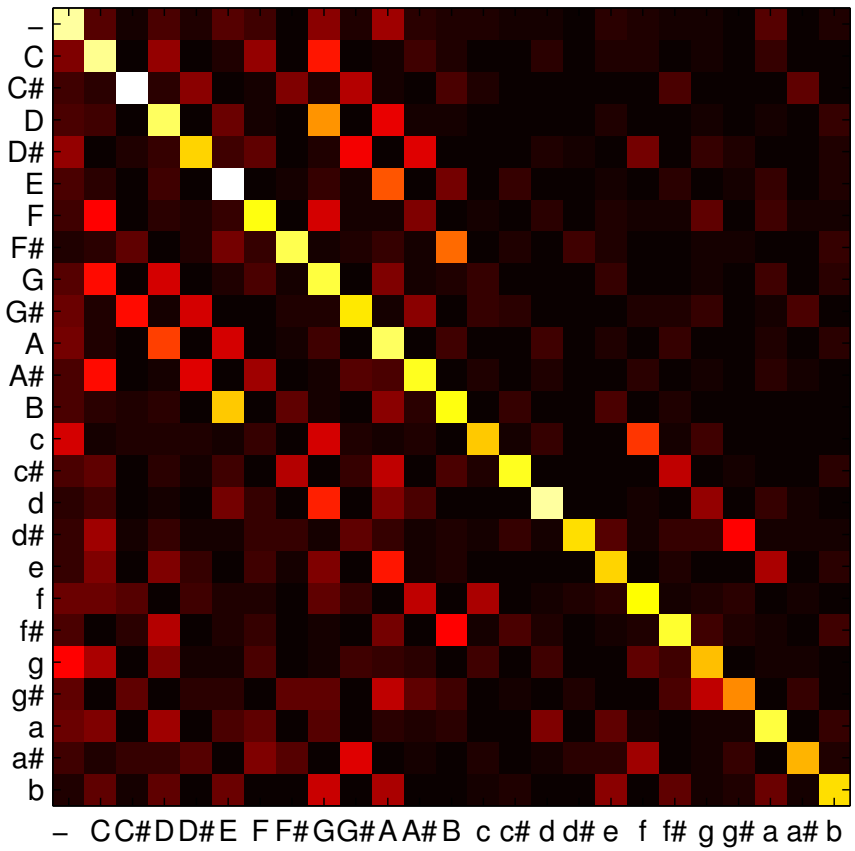
- ↳ Estimate the meter and beats position
- ↳ Calculate **P(chord change|meter position)** at each time instant
- ↳ Modulate the Transition Probability Matrix A of the HMM



TIME-VARYING TRANSITION PROB. MATRIX

Modulated A, elements on the diagonal are the probabilities of *no chord change*





TWO-STAGE DECODING

- Usually, **Viterbi algorithm** is used in the decoding stage of HMM for Chord Estimation
- Viterbi $\Rightarrow$ *Maximum likelihood* path (sequence)
- **Before** Viterbi $\Rightarrow$ forward-backward estimation
- Forward-backward $\Rightarrow$ *Maximum-a-posteriori* at each time instant

Score	HMM	HMM+B	HMM+2S	HMM+B+2S
Mirex2010	0.751	0.763	0.763	0.773
Improve.	-	1.2%	1.2%	2.2%

Conclusion and Future Perspectives

CONCLUSION

- ↳ Use the phase coefficients to refine the amplitude spectrum representation
- ↳ Proposed a new testing procedure for f_{ref} estimation algorithms
- ↳ Proposed a new Salience Measure for polyphonic music analysis
- ↳ Proposed a modification to current HCDF calculation strategies and tested the benefits for ACE algorithms

FUTURE PERSPECTIVES

- The proposed Salience Measure is promising, but an ad-hoc post processing technique must be developed in order to obtain a “full” multi-pitch algorithm
- Ongoing research in extending the features of ACE algorithms (MIR “hot” topic)
- MIR research offers a wide spectrum of perspectives and finds a lot of application in the “digital era” of music distribution media (Spotify, . . .) and “digital musicians” (AnteScoFo, . . .)

Publications

PUBLICATIONS

Conference Papers

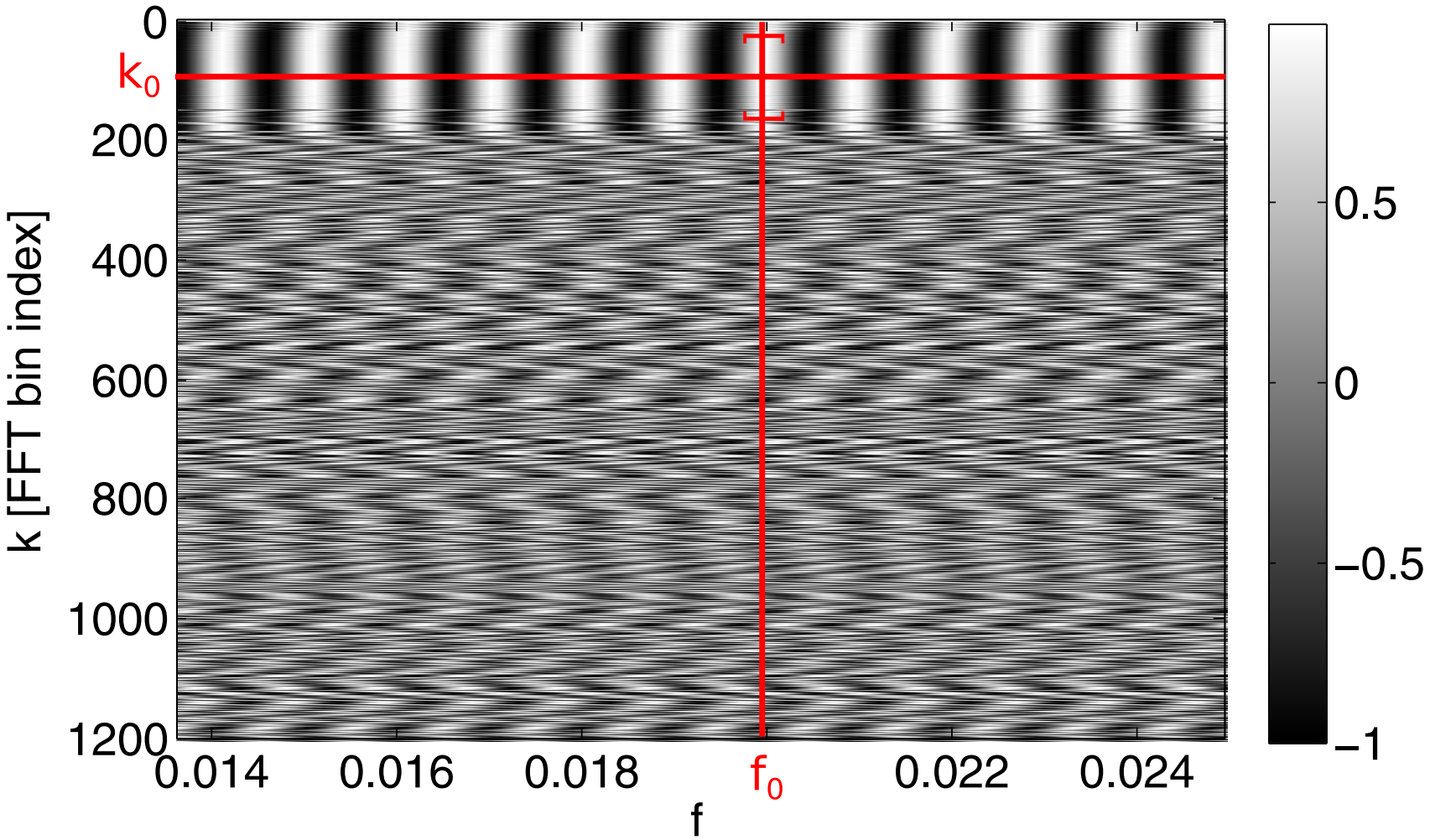
- *A.Degani, M.Dalai, R. Leonardi and P.Migliorati*, "Harmonic Change Detection for Musical Chords Segmentation", **ICME 2015** (to appear)
- *A.Degani, M.Dalai, P.Migliorati and G. Peeters*, "A Pitch Salience Function Derived from Harmonic Frequency Deviations for Polyphonic Music Analysis", **DAFx 2014**
- *A.Degani, M.Dalai, R. Leonardi and P.Migliorati*, "Time-Frequency Analysis of Musical Signals Using the Phase Coherence", **DAFx 2013**
- *A.Degani, M.Dalai, R. Leonardi and P.Migliorati*, "A Heuristic for Distance Fusion in Cover Song Identification", **WIAMIS 2013**
- *A.Degani, M.Dalai, R. Leonardi and P.Migliorati*, "Real-Time Performance Comparison of Tuning Frequency Estimation Algorithms", **ISPA 2013**

PUBLICATIONS

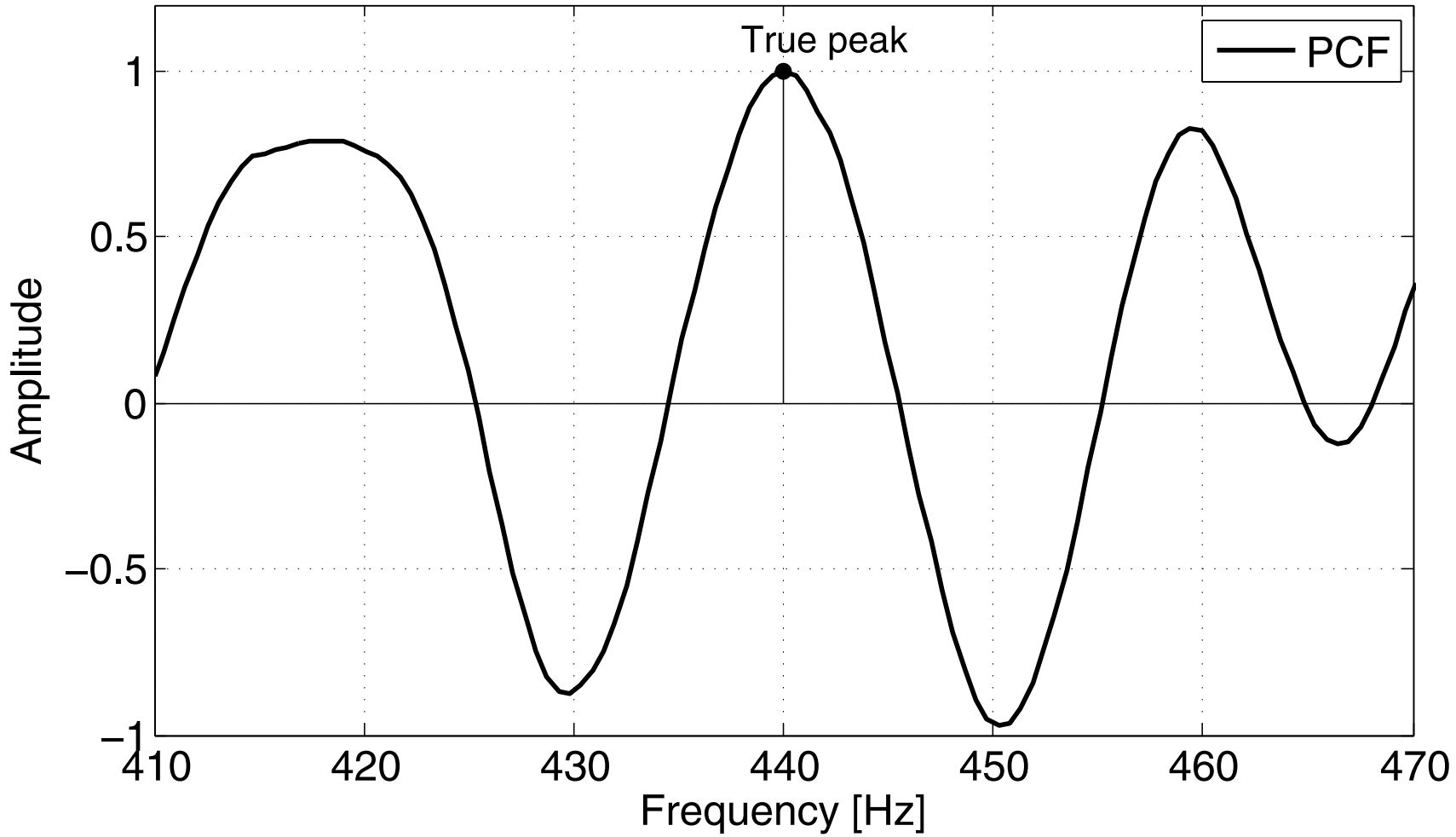
Journal Papers

- *A.Degani, M.Dalai, R. Leonardi and P.Migliorati*, "Comparison of Tuning Frequency Estimation Methods", **Journal of Multimedia Tools and Applications, *Springer***
- *A.Degani, M.Dalai, R. Leonardi and P.Migliorati*, "Chord Estimation with Meter Modelling and 2-Stage Decoding", **(To be submitted...)**

PHASE COHERENCE FUNCTION



PHASE COHERENCE FUNCTION



(1) DEVIATIONS HISTOGRAM

Build an histogram $\mathcal{H}$ of the deviations d_i of each spectral peak from the equal-tempered scale.

$$d_i = \frac{c_i - round(c_i)}{100} \in [-0.5, 0.5[\tag{11}$$

The overall estimated deviation $\hat{d}$ is the deviation associated to the histogram bin with the maximum value

$$\hat{d} = \arg\max(\mathcal{H}) \tag{12}$$

The reference frequency is computed as:

$$f_{ref} = 440 \cdot 2^{\frac{\hat{d}}{12}} \tag{13}$$

(2) CIRCULAR STATISTICS

Each peak's frequency converted in cents c_i is mapped onto a unit circle 100 cents periodic and represented as a unit vector u_i

$$u_i = 1 \cdot e^{j\phi_i} \tag{14}$$

$$\phi_i = \frac{2\pi}{100} \cdot c_i \tag{15}$$

The overall deviation $\hat{d}$ is the angle of the mean vector $\hat{u}$ of all circular quantities

$$\hat{d} = \frac{1}{2\pi} \angle(\hat{u}) \tag{16}$$

The reference frequency is computed as:

$$f_{ref} = 440 \cdot 2^{\frac{\hat{d}}{12}} \tag{17}$$

(3) LEAST-SQUARE OPTIMIZATION

Deviation s_i in (integer) semitones index are calculated at each peak frequency f_i

$$s_i = round\left[12 \cdot \log_2\left(\frac{f_i}{f_{ref}^{k-1}}\right)\right], \tag{18}$$

then the new reference frequency f_{ref}^k is estimated in a Least-Square sense minimizing the squared position error

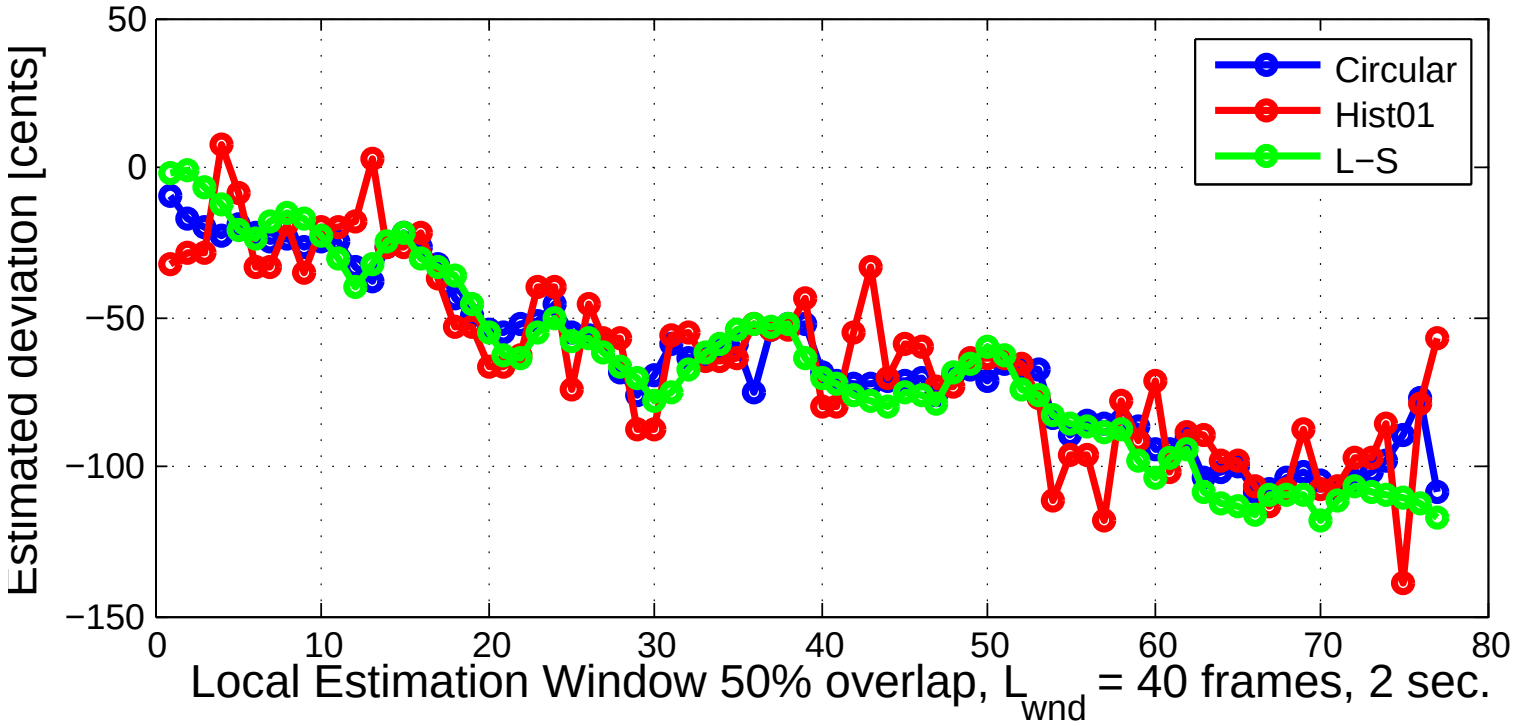
$$E = \sum_i \left(f_i - f_{ref}^k 2^{\frac{s_i}{12}}\right)^2. \tag{19}$$

by setting $\frac{\delta E}{\delta f_{ref}^k} = 0$, the estimation of f_{ref}^k is

$$f_{ref}^k = \frac{\sum_i (f_i \cdot 2^{\frac{s_i}{12}})}{\sum_i 2^{\frac{s_i}{6}}} \tag{20}$$

RESULTS - LOCAL TUNING EST.

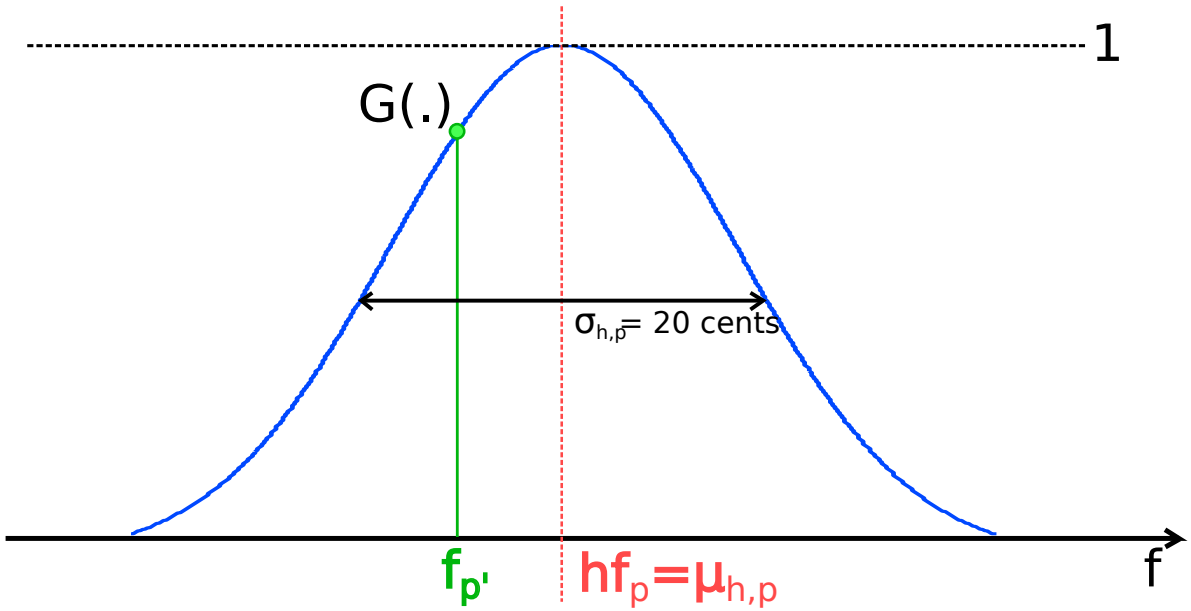
Choir with a pitch drift during the execution



SALIENCE OF SPECTRAL PEAK P

$$S_p = a_p < \mathbf{d}, \hat{\mathbf{d}}^p >= a_p \sum_{h=1}^{H} d_h \cdot \hat{d}_h^{f_p} \tag{21}$$

$$\hookrightarrow \ \hat{d}_h^{f_p} = \sum_{p'=1}^{P} G(f_{p'}; \mu_{h,p}, \sigma_{h,p}) \cdot \bar{d}_{p'}$$



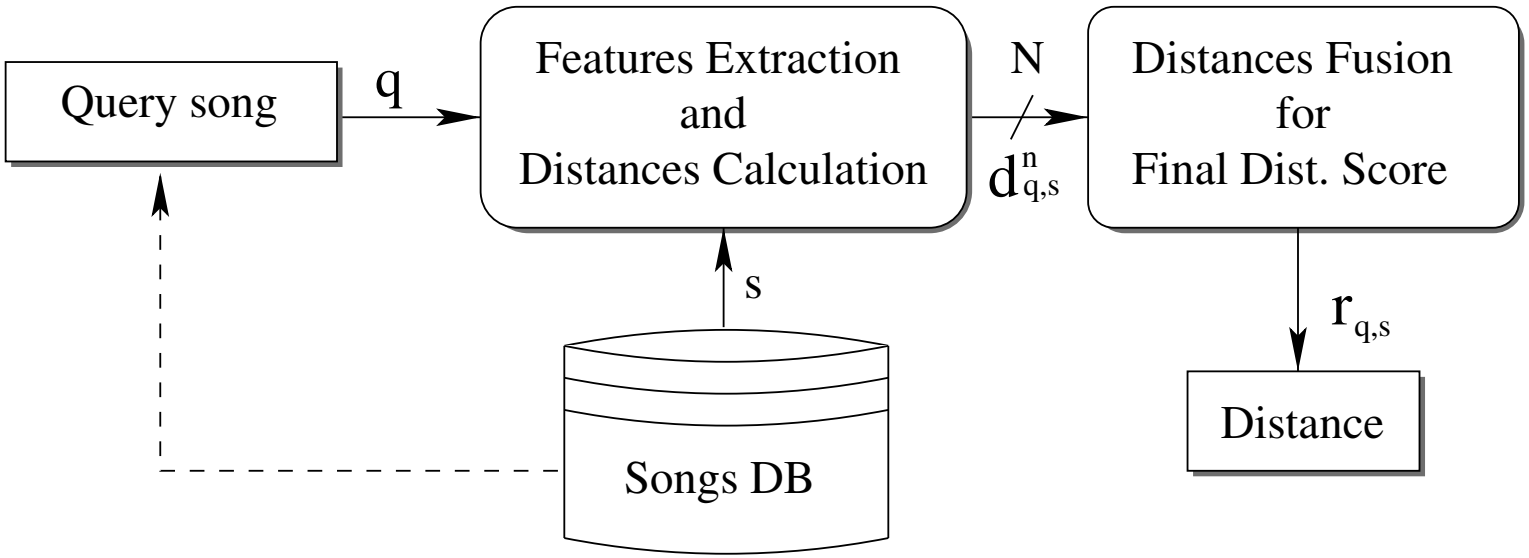
COVER SONG: DEFINITION

OBJ : find an “alternative” version of a given query song

Differences between original and alternative?

Change	Don’t Change
-	Title
Tempo	-
Instruments	-
Chord/Key ?	Chord/Key
Main melody ?	Main melody
Lyrics ?	Lyrics

COVER SONG IDENTIFICATION

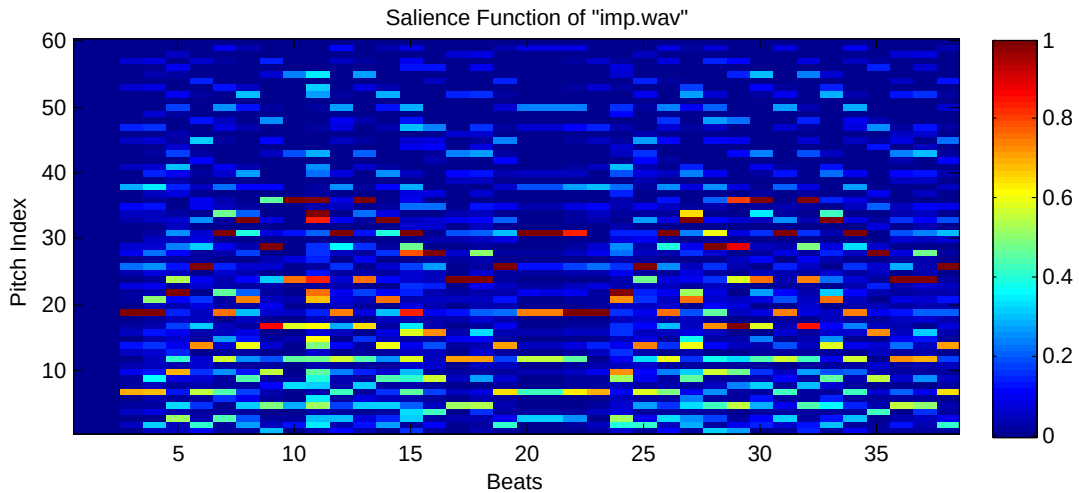


Audio feature extraction | Distances computation

DESCRIPTORS

Instruments (spectral envelope) invariant harmonic/melodic descriptors

- ↳ HPCP[5] (H)
- ↳ Salience[6] (S)

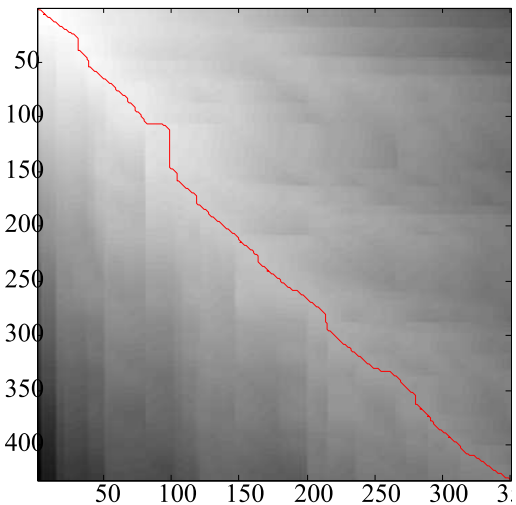


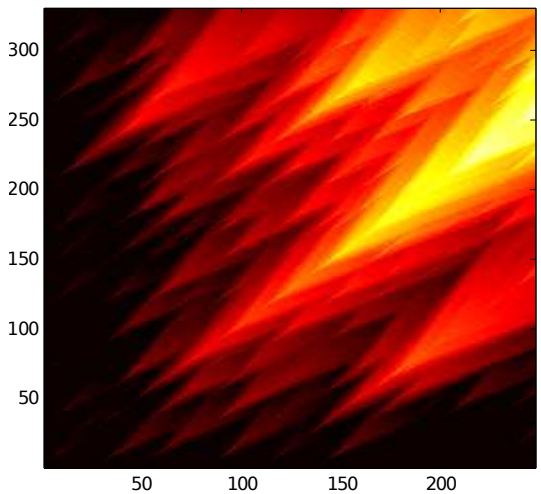
[5] Emilia Gómez. "Tonal Description of Music Audio Signals". PhD thesis. Universitat Pompeu Fabra, Barcelona, Spain, 2006.

[6] Justin Salamon, Emilia Gómez, and Jordi Bonada. "Sinusoid Extraction and Salience Function Design for Predominant Melody Estimation". In: *Proc. of the 14th Int. Conference on Digital Audio Effects, (DAFx)* (2011).

Tempo/Key Insensitive distance measures

- ↳ Dynamic Time Warp[7] (D)
- ↳ Q_{max}[8] (Q)





[7] Meinard Müller. *Information Retrieval for Music and Motion*. Springer, 2007.

[8] Joan Serrà, Xavier Serra, and Ralph G. Andrzejak. "Cross recurrence quantification for cover song identification". In: *New Journal of Physics* 11 (2011).

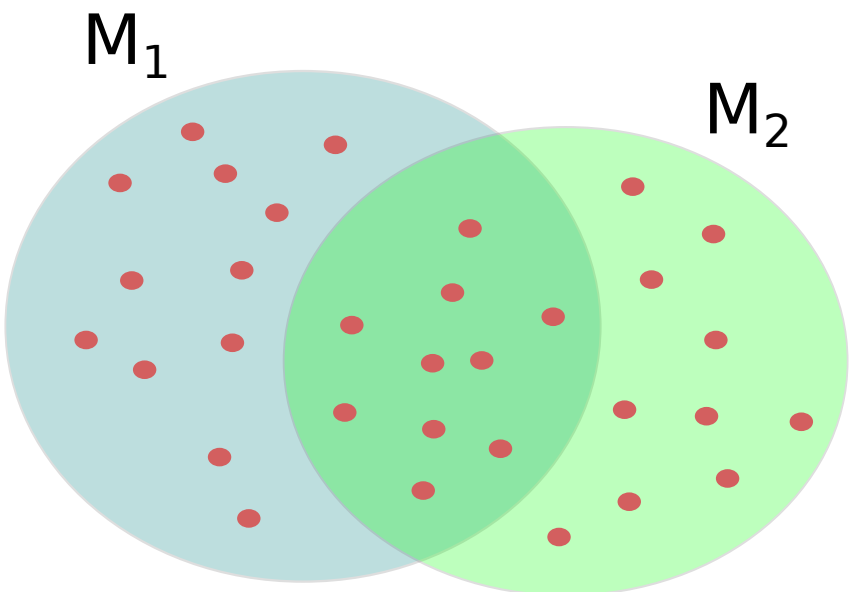
DISTANCE FUSION

OBJ: take the advantages of N different distance measures

For $\mathbf{N = 2}$ and

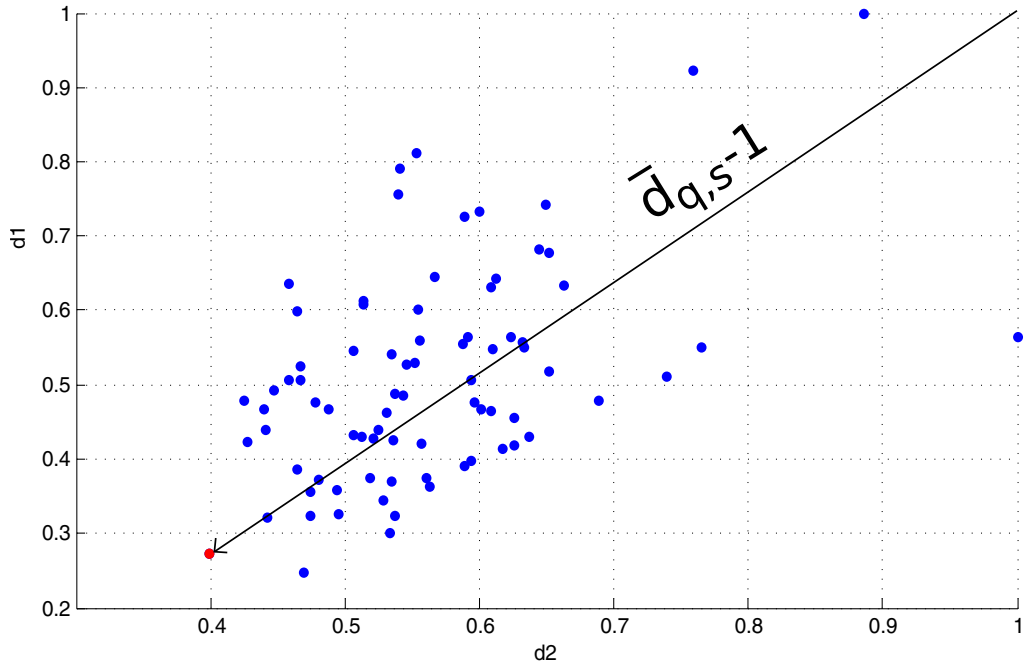
- $M_1 = \{$correct covers detected by method $n = 1\}$
- $M_2 = \{$correct covers detected by method $n = 2\}$
- $T =$ percent of virtual total unique correct identified cover

$$\Rightarrow T = \frac{|M_1 \cup M_2|}{N_S}, \quad N_S\text{: Total number of songs}$$



DISTANCE SPACE (2D)

$$r_{q,s} = ||\mathbf{1}|| - ||\bar{\mathbf{d}}_{q,s} - \mathbf{1}||\,, \tag{22}$$



⇒*A.Degani, M.Dalai, R. Leonardi and P.Migliorati*, "A Heuristic for Distance Fusion in Cover Song Identification", **WIAMIS 2013**

RESULTS

MR1st: mean rank of 1st correct id. cover
MAP: Mean Average Precision

Combination	N	Precision	MR1st	MAP	T
(S,D)	1	0.47	12	0.55	-
(S,Q)	1	0.52	15	0.60	-
(H,D)	1	0.35	17	0.43	-
(H,Q)	**1**	**0.59**	**9**	**0.65**	-
(S,D)+(S,Q)	2	0.61	12	0.66	0.61
(S,D)+(H,D)	2	0.46	12	0.53	0.55
(S,D)+(H,Q)	**2**	**0.65**	**7**	**0.71**	**0.66**
(S,Q)+(H,D)	2	0.56	13	0.63	0.60
(S,Q)+(H,Q)	2	0.64	10	0.69	0.69
(H,D)+(H,Q)	2	0.60	9	0.66	0.60
(S,D)+(S,Q)+(H,D)	3	0.60	11	0.66	0.66
(S,D)+(S,Q)+(H,Q)	**3**	**0.65**	**9**	**0.71**	**0.74**
(S,Q)+(H,D)+(H,Q)	**3**	**0.66**	**9**	**0.70**	**0.70**
(S,D)+(H,D)+(H,Q)	3	0.60	7	0.67	0.68
(ALL)	**4**	**0.66**	**8**	**0.72**	**0.75**