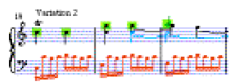
Chroma-based estimation of musical key from audio-signal analysis

Geoffroy Peeters peeters@ircam.fr

IRCAM (Sound Analysis/Synthesis Team) - CNRS (STMS)
Semantic HIFI - European IST Project



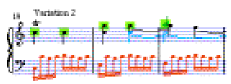
Introduction

- Goal:
  - estimate the musical key of a music audio track
    - key-note: C, C#/Db, D, D#/Eb, E, …
    - mode: Major / minor

- Applications:
  - search/ query music databases
  - automatic generation of playlists
  - automatic accompaniment

Semantic HIFI system

- Methods:
  - Derive the key from the score (symbolic representation)
    - in most cases, implies first to derive the score from the audio (multipitch) -> very costly !
  - Chroma / Pitch Class Profile (PCP) approach
    - Krumhansl & Schmukler, Temperley, … based cognitive key profiles
    - Spiral Array / Center of Effect Generator



Introduction:

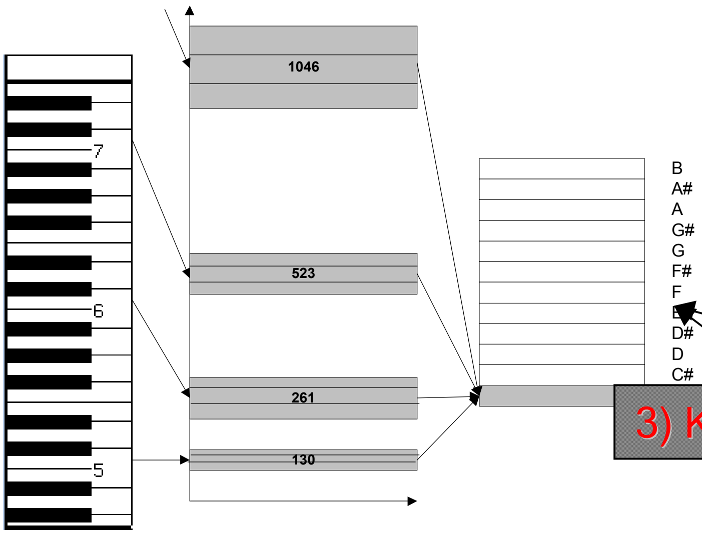
Key-estimation from chroma representation

1) Chroma / PCP representation

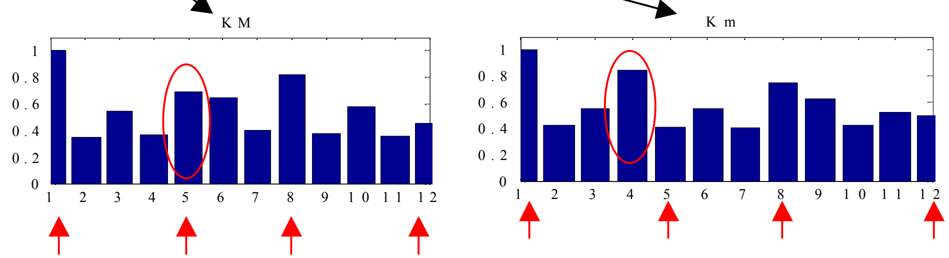
- Shepard:
  - pitch =
    tone height (octave number) +
    chroma (pitch class)
- Chroma spectrum or Pitch Class Profile (PCP) (Wakefield/Fujishima)
  - Fourier transform -> 12 semi-tones pitch classes C

2) Key-chroma profiles

- Cognitive-based approach
- Krumhansl & Schmuckler experiments
  - Tonal profiles for major and minor keys contains 12 values.
  - Values = human ratings of the degree to which each of the 12 chromatic scale tones fit a particular key.



3) Key-estimation comparison



Introduction:
current problems

Introduction

System
- Tuning
- Chroma
- HPS

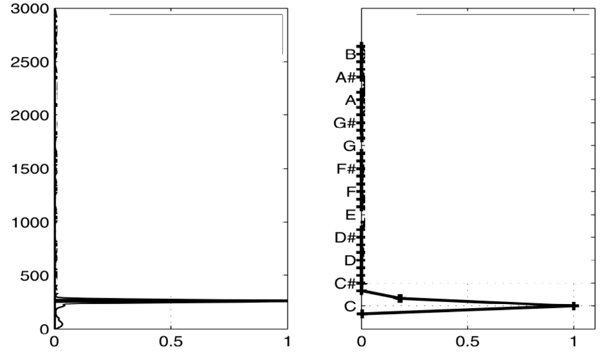
Key estimation
- Cognitive
- HMM

Evaluation

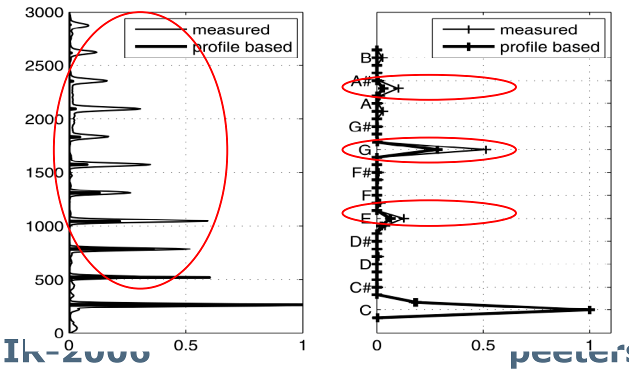
Conclusion

- Problem 1) Presence of the higher harmonics of a note
  - we do not directly observe the various pitches (but all their harmonics)
  - consequence: the direct mapping of a note spectrum to the chroma-domain will also map all the higher harmonics (fifth, third, …)

  - ideal case: the spectrum of a note played by an instrument is composed by a single partial
    - the mapping to the chroma-scale is limited to the pitch note



  - real case: : the spectrum of a note played by an instrument is composed by many partials
    - the mapping to the chroma-scale is composed of many other chromas



Introduction

System
- Tuning
- Chroma
- HPS

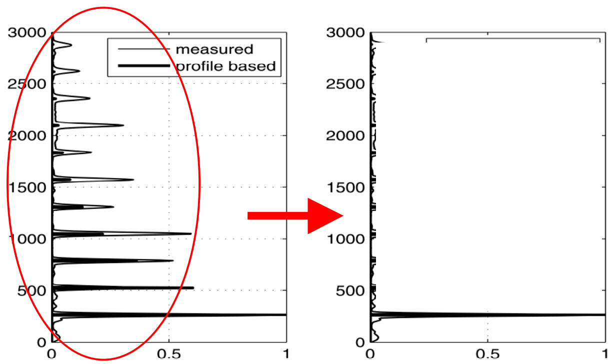
Key estimation
- Cognitive
- HMM

Evaluation

Conclusion

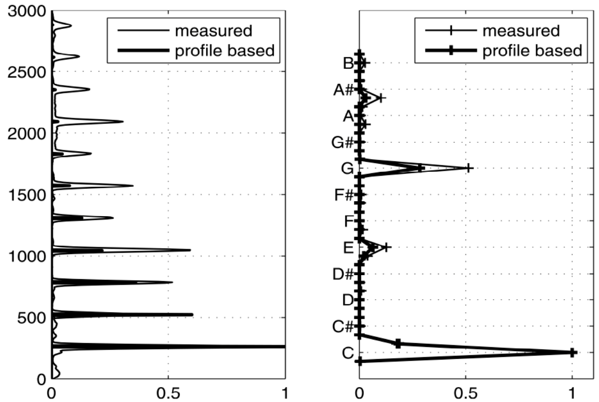
Introduction:
current problems

- Problem 1) Presence of the higher harmonics of a note
  - we do not directly observe the various pitches (but all their harmonics)
  - Solutions
  - 1) extract the various pitches, or reduce the influence of the higher harmonics



•Pauws, Chuan, Cremer, …
•Harmonic Peak Subtraction Function

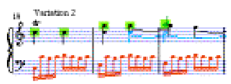
  - 2) adapt the cognitive-based profile to take into account the fact that many other chromas exist



•Gomez: extension of the Pitch-Class-Profile based on theoretical envelope contribution
Limitation of spectral envelope prediction:
Example: viola sound

•Izmirli: extension of the Pitch-Class-Profile based on measured (piano) envelope contrib.

•Learn the adaptation:
hidden Markov modeling



System

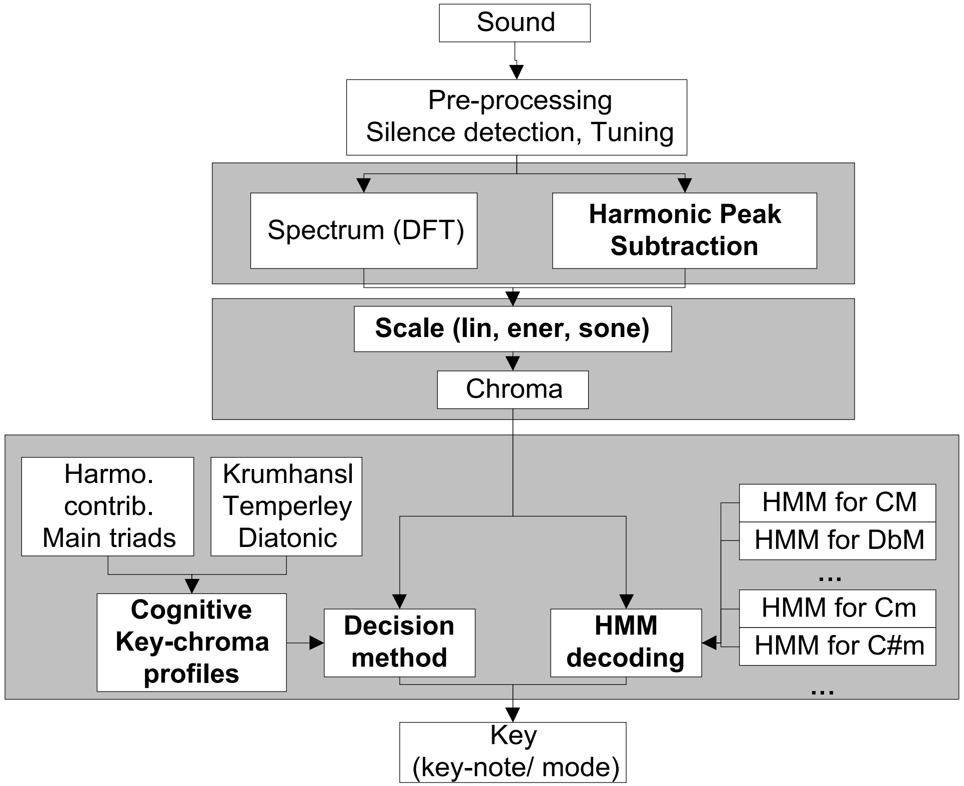
Introduction

System
- Tuning
- Chroma
- HPS

Key estimation
- Cognitive
- HMM

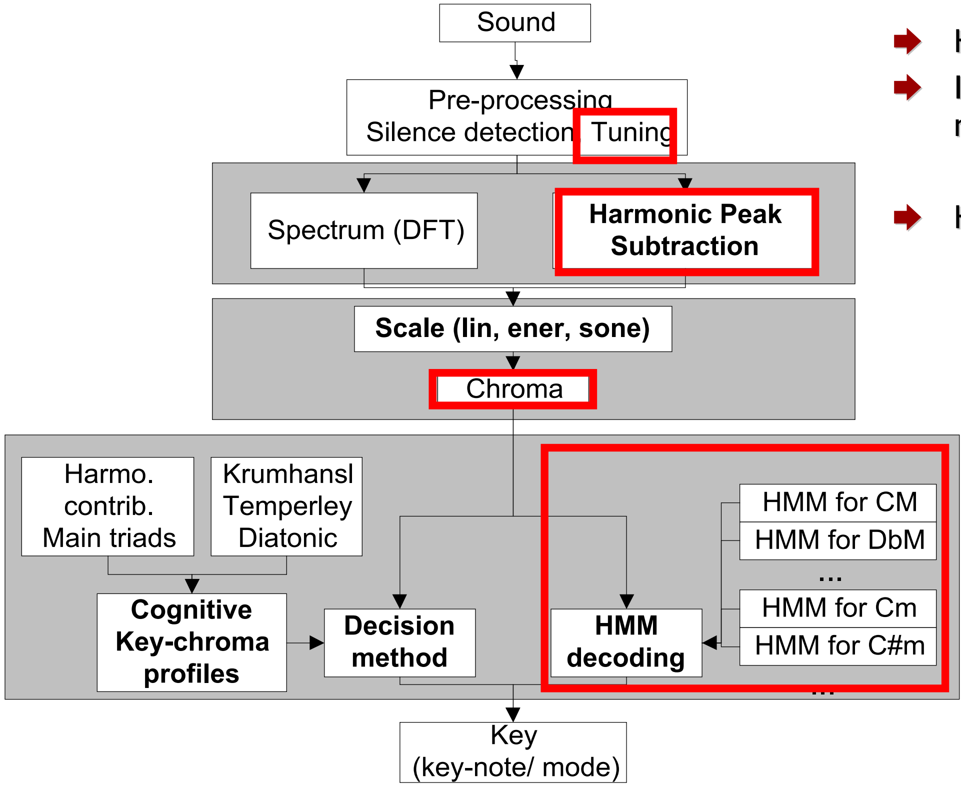
Evaluation

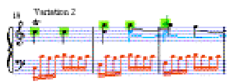
Conclusion



System
Main contributions of this research

- Main contributions of this research
  - Tuning
  - Harmonic Peak Subtraction
  - Indirect mapping to chroma (reduce noise)
  - Hidden Markov model





System

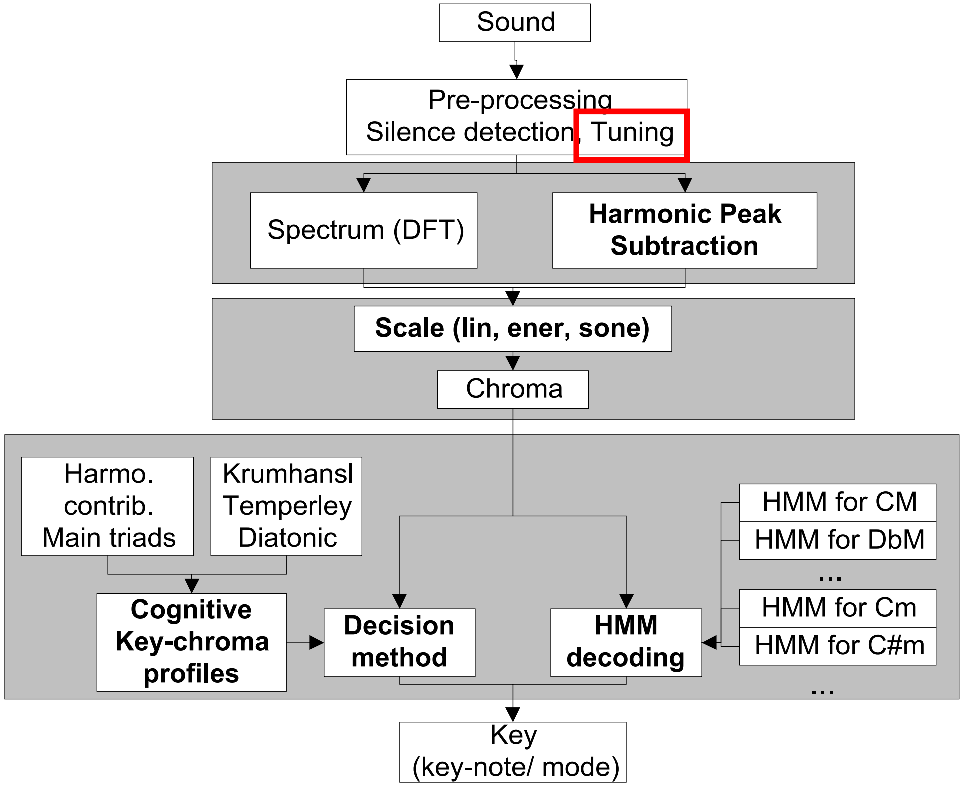
Introduction

System
- Tuning
- Chroma
- HPS

Key estimation
- Cognitive
- HMM

Evaluation

Conclusion



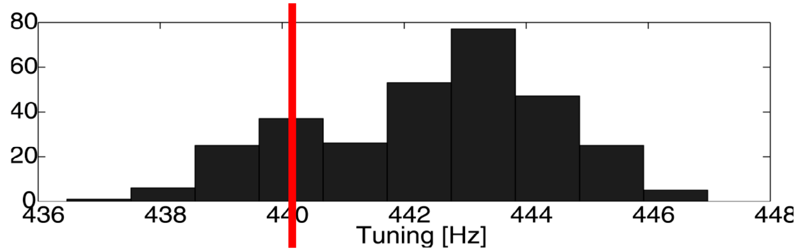
Pre-processing stages

- Estimation of the tuning of the track
  - Why ?
    - Instruments used during the recording may have used another tuning than 440Hz
    - Possible trans-coding of the audio media may have changed the tuning
  - How ?
    - Compute a modeling error given assumptions of tuning ranging between the quarter-tone below and above A4 at 440 Hz: t in [427,452] Hz
    - Take the tuning with the minimum modeling error over time

$$\epsilon(t,m) = 1 - \sum_{n} A(f_{t,n}, m) / \sum_{f} (A(f,m))$$

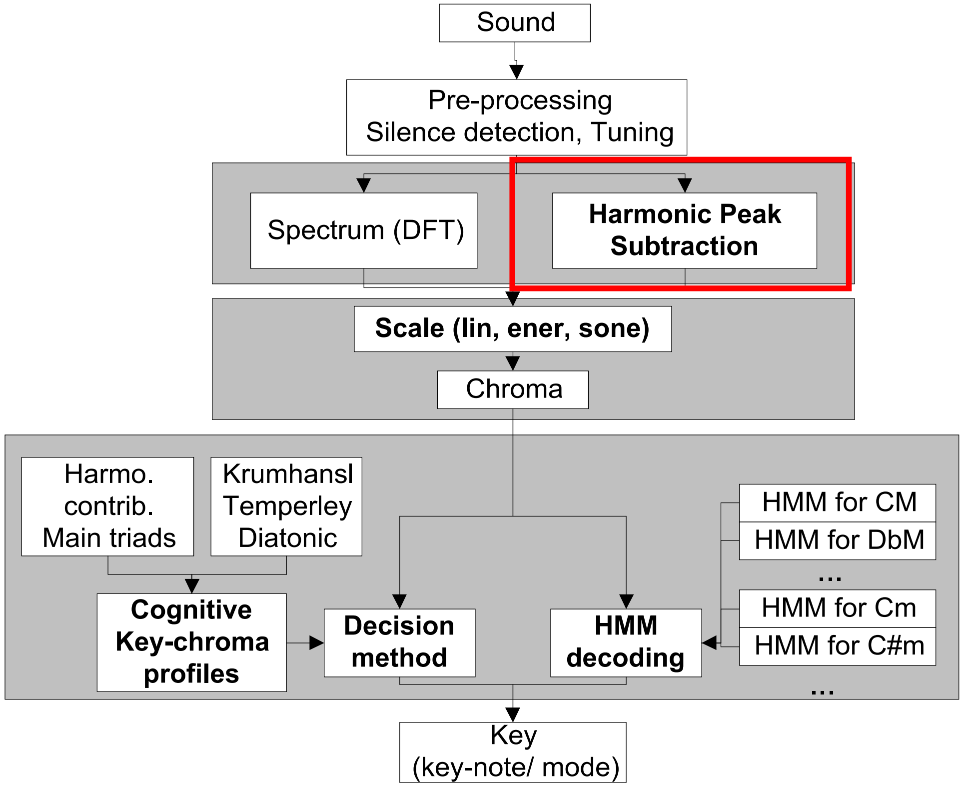
$$f_{t,n} = t \cdot 2^{(n-69)/12} \quad n \in [43, 44, \ldots, 95] \quad t \in [427, 452]$$

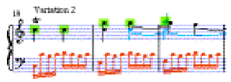
    - Histogram of the estimated tunings of the 300 test database



  - Resample (using a polyphase filter implementation) the signal in order to bring its tuning back to 440 Hz

System


Sound
Pre-processing
Silence detection, Tuning
Spectrum (DFT)
Harmonic Peak Subtraction
Scale (lin, ener, sone)
Chroma
Harmo. contrib. Main triads
Krumhansl Temperley Diatonic
Cognitive Key-chroma profiles
Decision method
HMM decoding
HMM for CM
HMM for DbM
...
HMM for Cm
HMM for C#m
...
Key
(key-note/ mode)

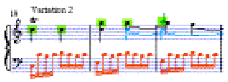


Spectral observation:
Harmonic Peak Subtraction Function

- Harmonic Peak Subtraction Function

- Why ?
  - The spectral representation will be mapped to the chroma domain
  - It must represent only information about the pitches of the various notes and not all their harmonics
    - the presence of the harmonics will distort the chroma representation
    - harmonics 3,6 will strenghten the presence of the fifth note
    - harmonic 5 of the third note

- How ?
  - Use a representation which reduce the influence of the higher harmonics
    - Extension of a mono-pitch representation algorithm [Peeters ICASSP 2006]



Spectral observation:
Harmonic Peak Subtraction Function

Introduction

System
- Tuning
- Chroma
- HPS

Key estimation
- Cognitive
- HMM

Evaluation

Conclusion

- Mono-pitch signal [Peeters, ICASSP2006]

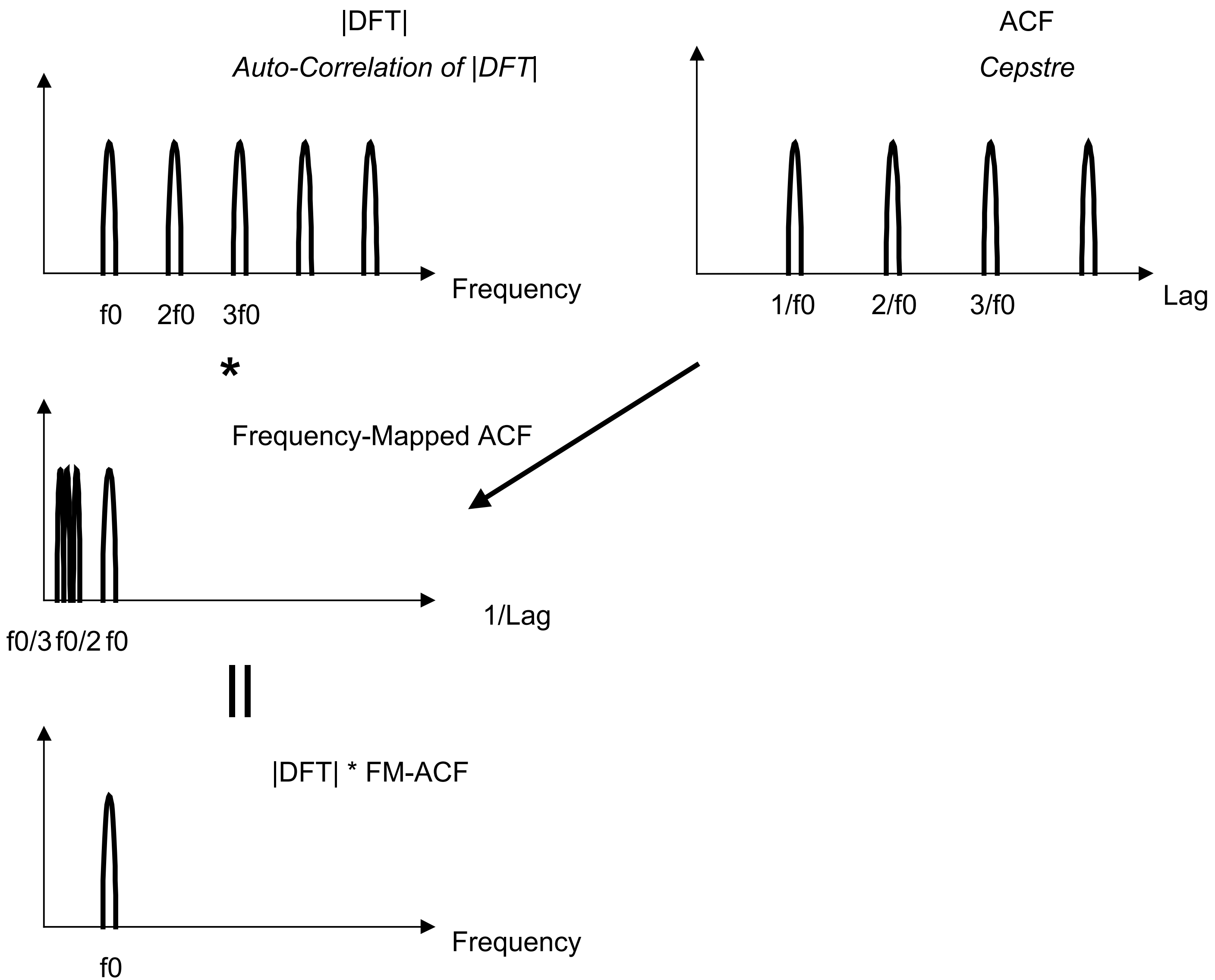
  - a representation which allows reducing the influence of the higher harmonics of a note

  - a combination of
    - a frequency representation (DFT or ACFofDFT) and
    - a temporal representation (ACF or CEP) mapped to the frequency domain (the lags are expressed as frequencies=1/lags)

  - Inverse octave errors of the DFT (in frequency) and the ACF (in lags)
    - combined both representation (octave errors cancel each others)

  - Results obtained with a large database (5371 sounds, 27 musical instrument sounds, 27.5Hz - 7900 Hz)
    - 97% - 97.3 % (Yin 94.9% - 95.5%) see ICASSP2006 for details

Spectral observation:
Harmonic Peak Subtraction Function

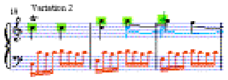


Spectral observation:
Harmonic Peak Subtraction Function

- Underlying process:
  - ACF can be understood as the projection of $S(w_k)^2$ on a set of cosine functions

$$g_\tau(f_k) = \cos(2\pi f_k \tau)$$

$$g_\tau(f_k) = g_\tau^+(f_k) - g_\tau^-(f_k)$$



Spectral observation:
Harmonic Peak Subtraction Function

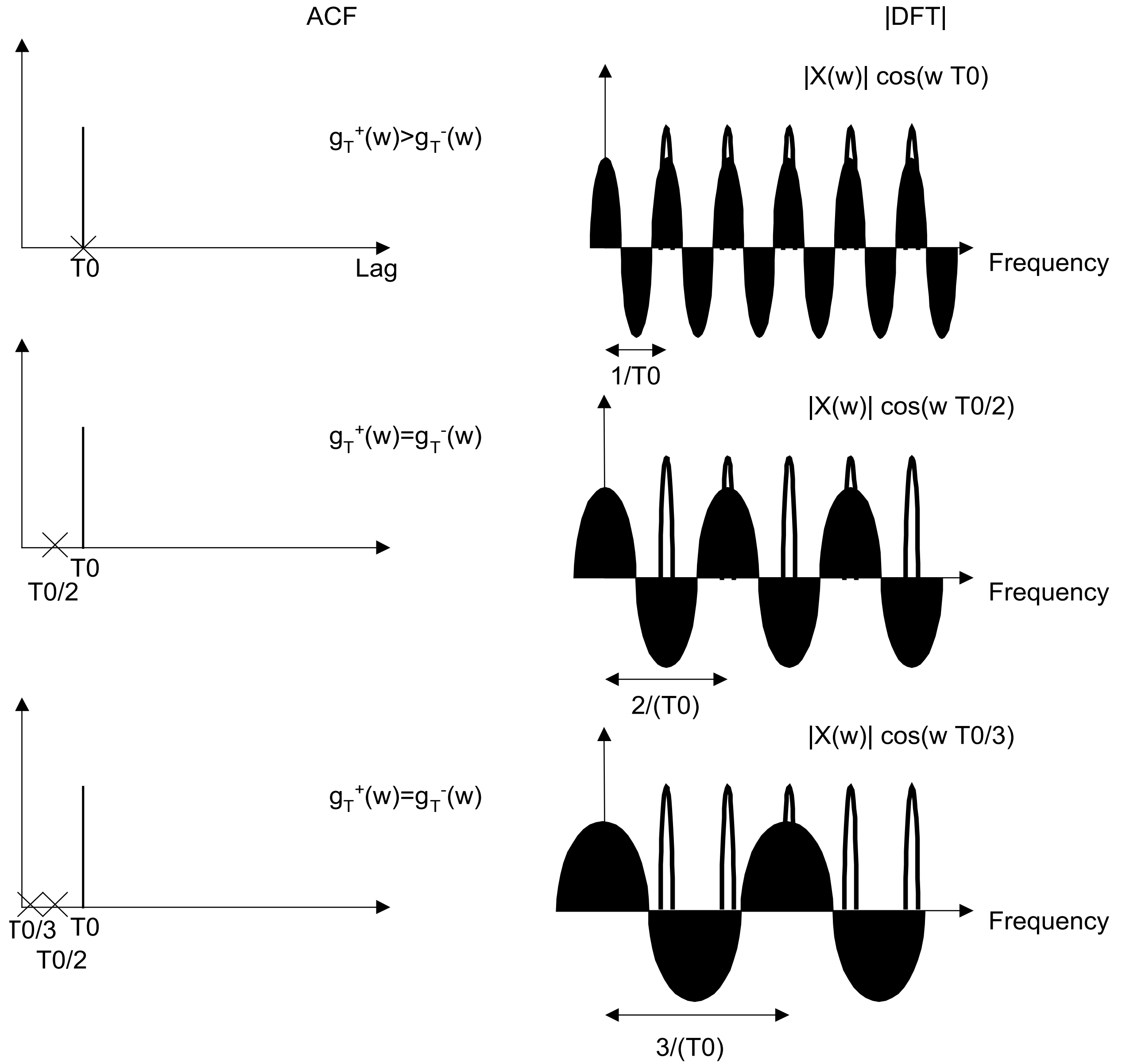
Introduction

System
- Tuning
- Chroma
- HPS

Key estimation
- Cognitive
- HMM

Evaluation

Conclusion



Spectral observation:
Harmonic Peak Subtraction Function

- Underlying process:
  - ACF can be understood as the projection of $S(w_k)^2$ on a set of cosine functions

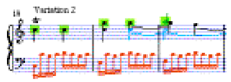
$$g_\tau(f_k) = \cos(2\pi f_k \tau)$$

$$g_\tau(f_k) = g_\tau^+(f_k) - g_\tau^-(f_k)$$

- positive values: when projection on g+ larger than on g-
  - sub-harmonics of f0: tau=k/f0
- non-positive values: when projection on g- larger than or equal to on g-
  - harmonics of f0: tau=1/(kf0)

  - $S(w_k)^2$ positive values at the harmonics of f0

- Combined function $h(f_k) = S(f_k) \cdot r(1/f_k)$



ircam Centre Pompidou

Spectral observation:
Harmonic Peak Subtraction Function

Introduction

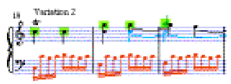
System
- Tuning
- Chroma
- HPS

Key estimation
- Cognitive
- HMM

Evaluation

Conclusion

- Multi-pitch signal
  - Problem:
    - we cannot apply directly the combined function to multi-pitch signals because the relationship between r(tau) and the periodicity of the various pitches are intricated
  - Solution:
    - use the same underlying principle
  - Principle ?
    - test the hypothesis that f_k is a pitch
      - value given by the projection on g+
    - against the hypothesis that f_k is a higher harmonic of a lower harmonic
      - value given by the projection on g-
    - avoid the detection of low-harmonics (multiplication by S(fk))



Spectral observation:
Harmonic Peak Subtraction Function

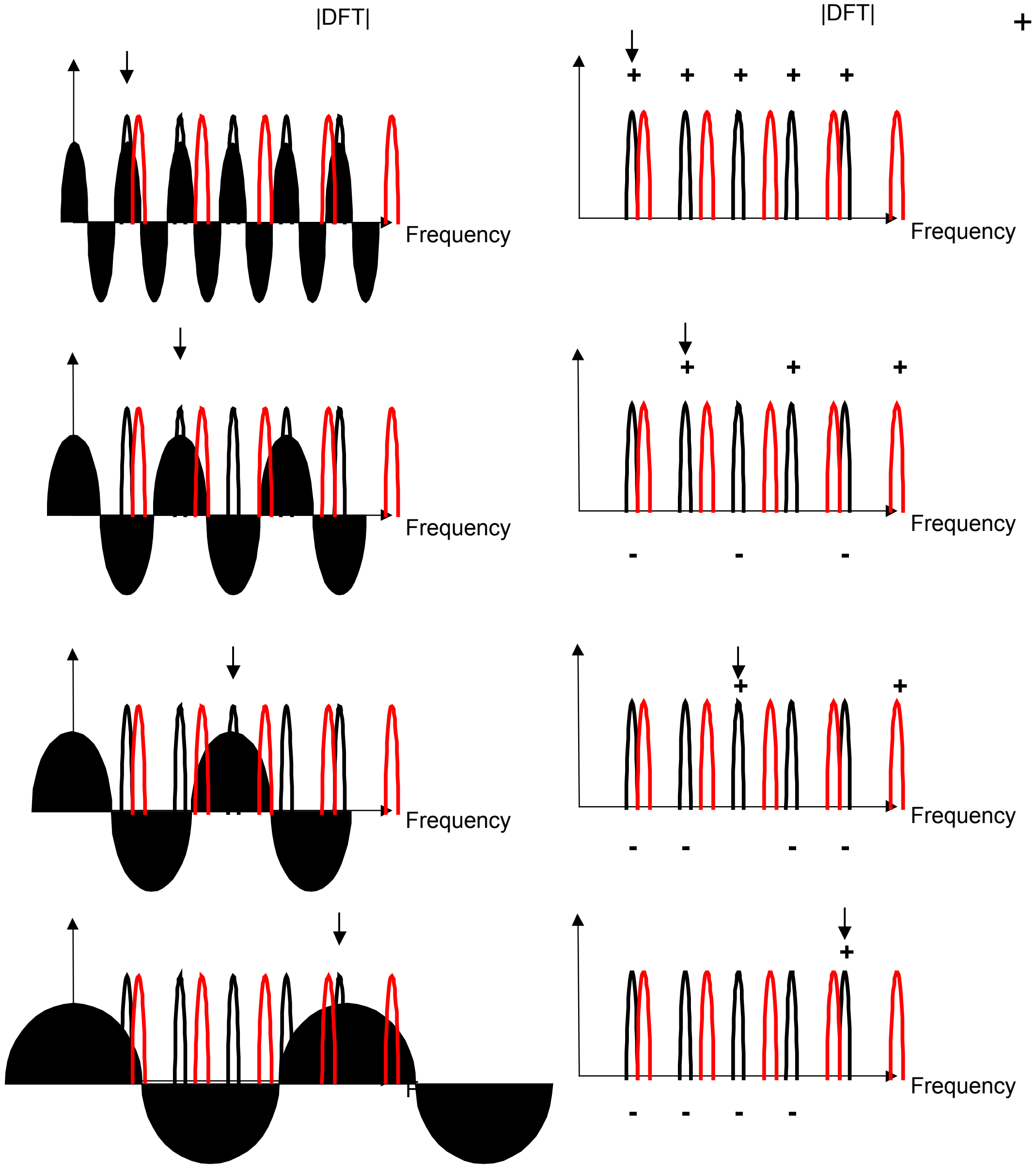
Introduction

System
- Tuning
- Chroma
- HPS

Key estimation
- Cognitive
- HMM

Evaluation

Conclusion



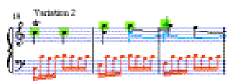
+ hypothesis that f_k is a pitch

- hypothesis is an even harmonic of a lower pitch

- hypothesis is the third harmonic of a lower pitch

+

- hypothesis is the fifth harmonic of a lower pitch



Spectral observation:
Harmonic Peak Subtraction Function

Introduction

System
- Tuning
- Chroma
- HPS

Key estimation
- Cognitive
- HMM

Evaluation

Conclusion

- Multi-pitch signal
  - HPS formula

$$\hat{r}(f_k) = \sum_{h=1}^{H} A(hf_k) - \max\{\alpha(f_k), \beta(f_k), \gamma(f_k)\}$$

$$\alpha(f_k) = \sum_{h=0}^{H-1} A\left(\left(h + \frac{1}{2}\right) f_k\right)$$

$$\beta(f_k) = \min_{h \in \{\frac{1}{3}, \frac{2}{3}, \frac{4}{3}, \frac{5}{3}\}} A(hf_k)$$

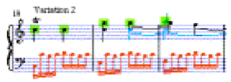
$$\gamma(f_k) = \min_{h \in \{\frac{1}{5}, \frac{2}{5}, \frac{3}{5}, \frac{4}{5}\}} A(hf_k)$$

Equivalent to the projection on g+ (sum over the harmonics of f_k)

Penalizes frequencies which are <u>even</u> (2,4,6,…) harmonics of a lower pitch

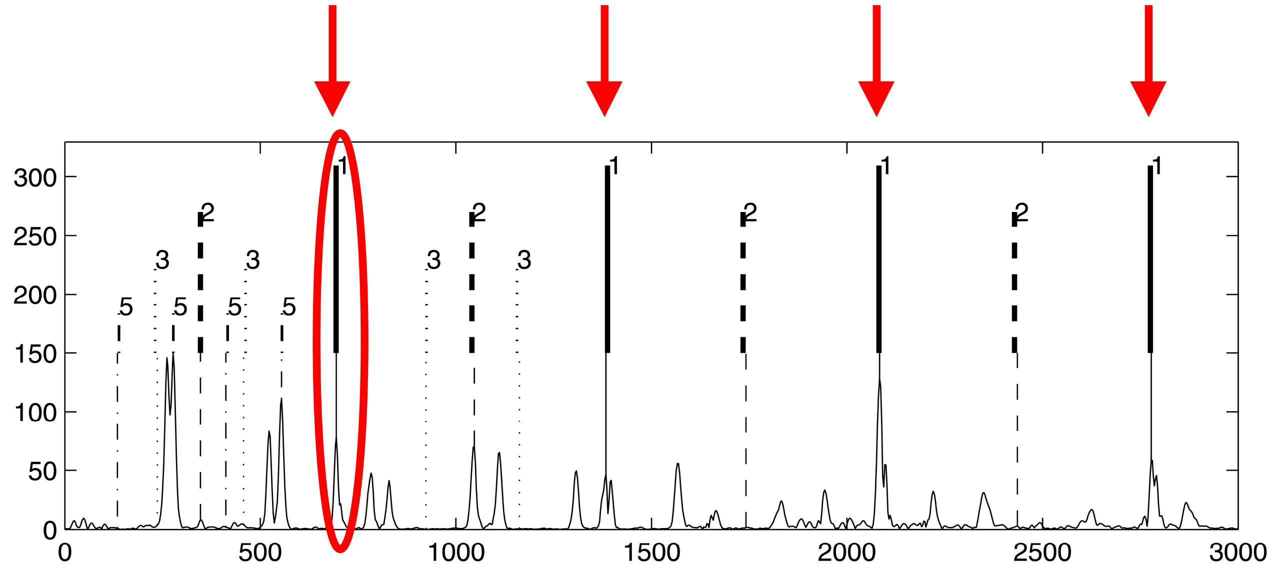
Penalizes frequencies which have <u>third</u> harmonic relationship with a lower pitch (assumption of envelope continuity)

Penalizes frequencies which have <u>fifth</u> harmonic relationship with a lower pitch (assumption of envelope continuity)



Spectral observation:
Harmonic Peak Subtraction Function

- Examples



- C4 (261.6Hz),
  C#4 (277.2Hz),
  F5 (698.5Hz)
  viola sounds.

Spectral observation:
Harmonic Peak Subtraction Function

Introduction

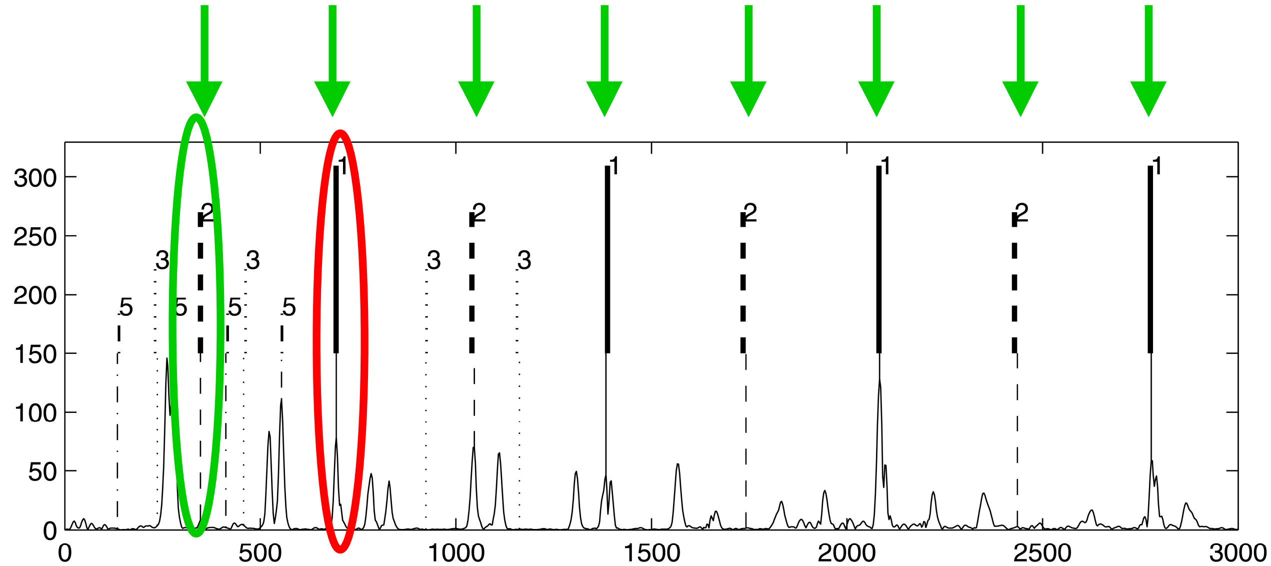
System
- Tuning
- Chroma
- HPS

Key estimation
- Cognitive
- HMM

Evaluation

Conclusion

- Examples



Spectral observation:
Harmonic Peak Subtraction Function

Introduction

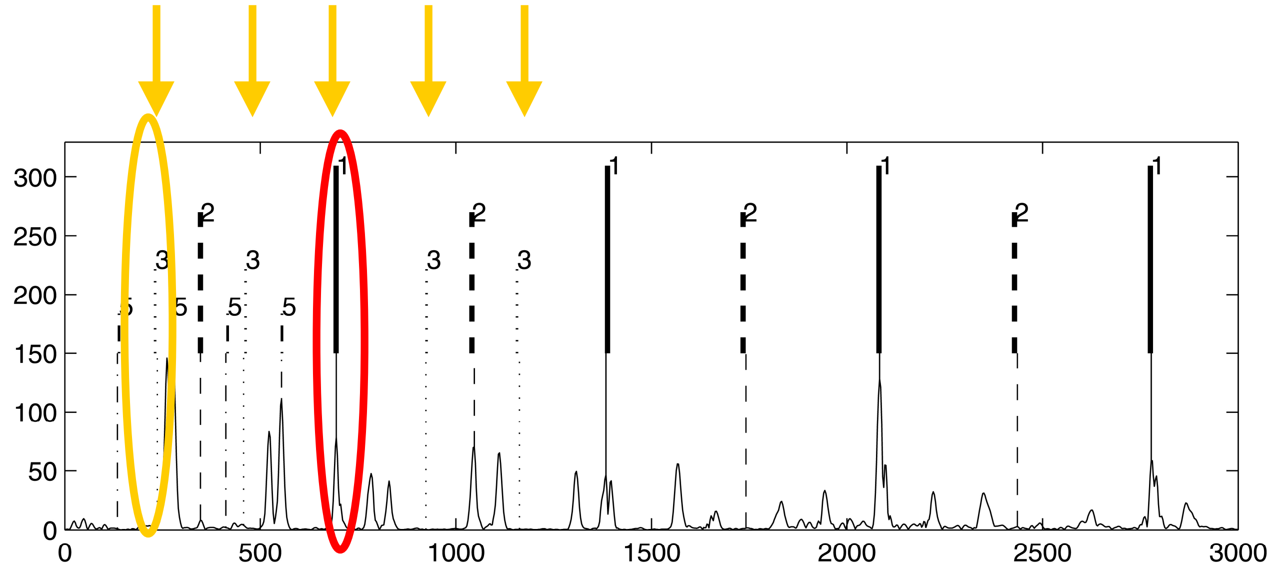
System
- Tuning
- Chroma
- HPS

Key estimation
- Cognitive
- HMM

Evaluation

Conclusion

- Examples



Introduction

System
- Tuning
- Chroma
- HPS

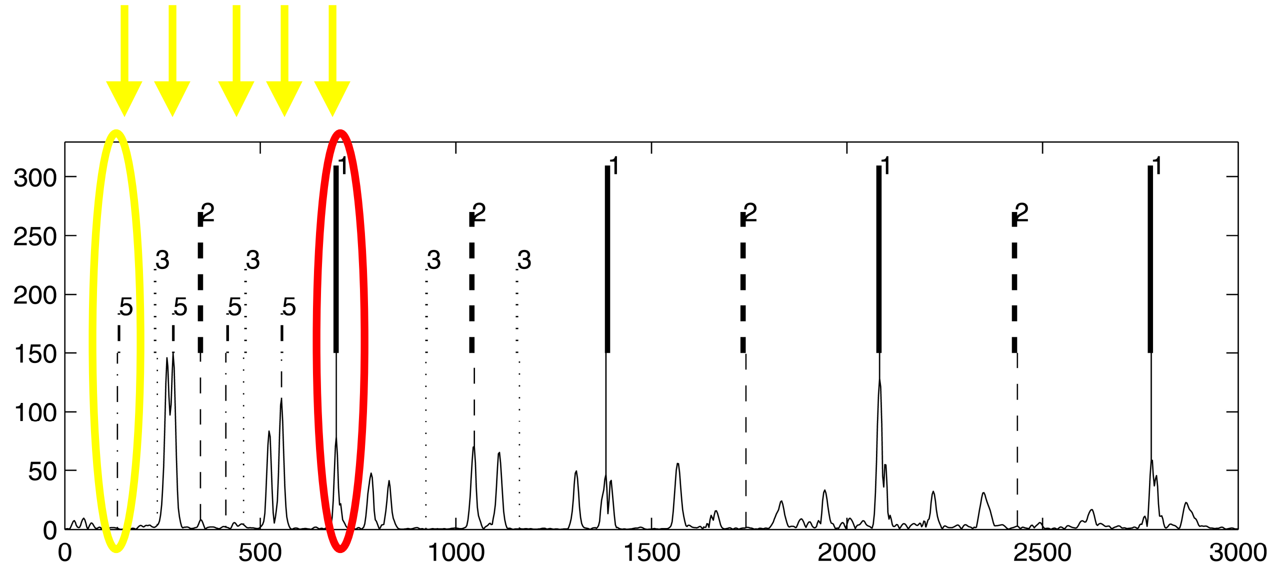
Key estimation
- Cognitive
- HMM

Evaluation

Conclusion

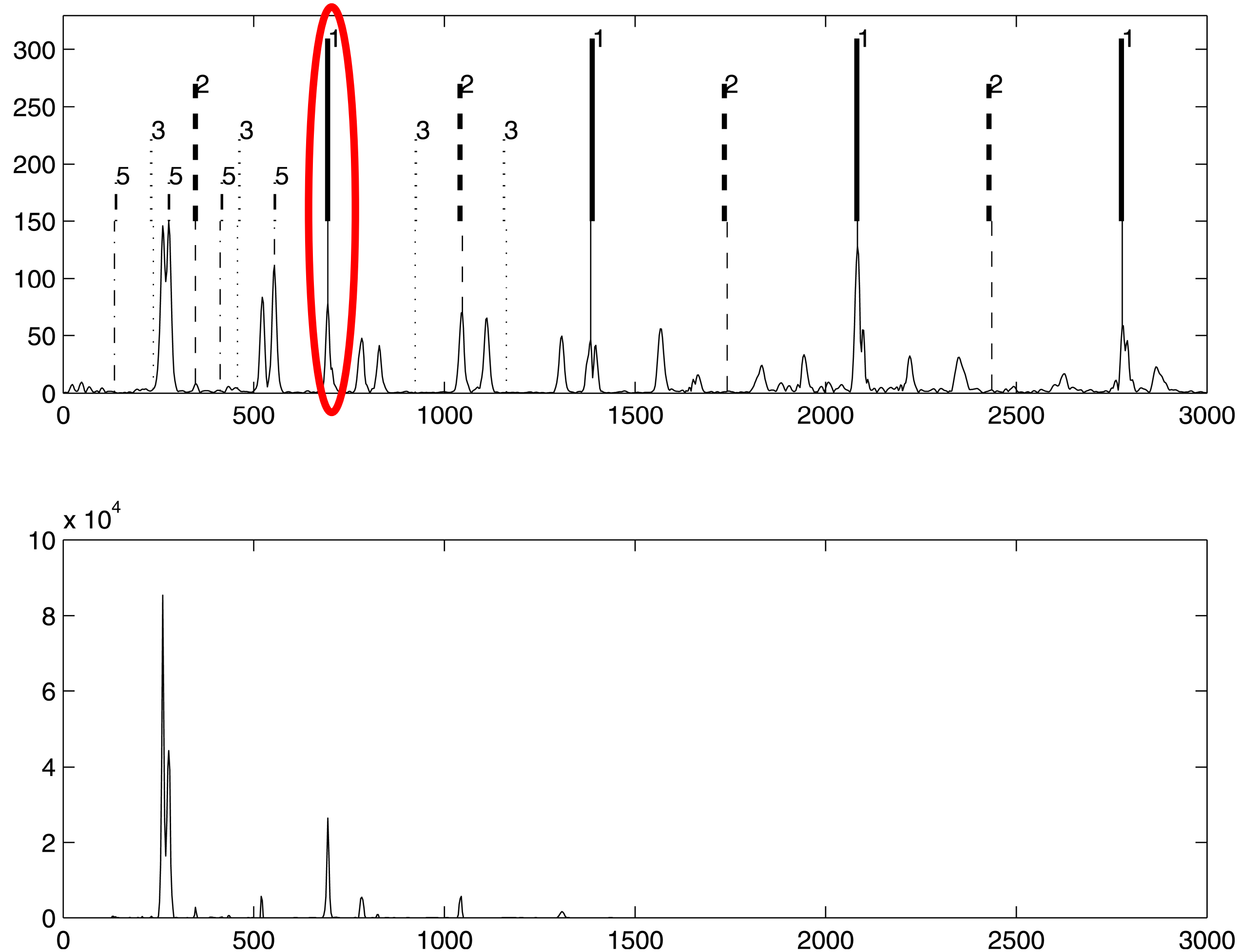
Spectral observation:
Harmonic Peak Subtraction Function

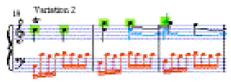
➡ Examples



Spectral observation:
Harmonic Peak Subtraction Function

- Examples





Introduction

System
- Tuning
- Chroma
- HPS

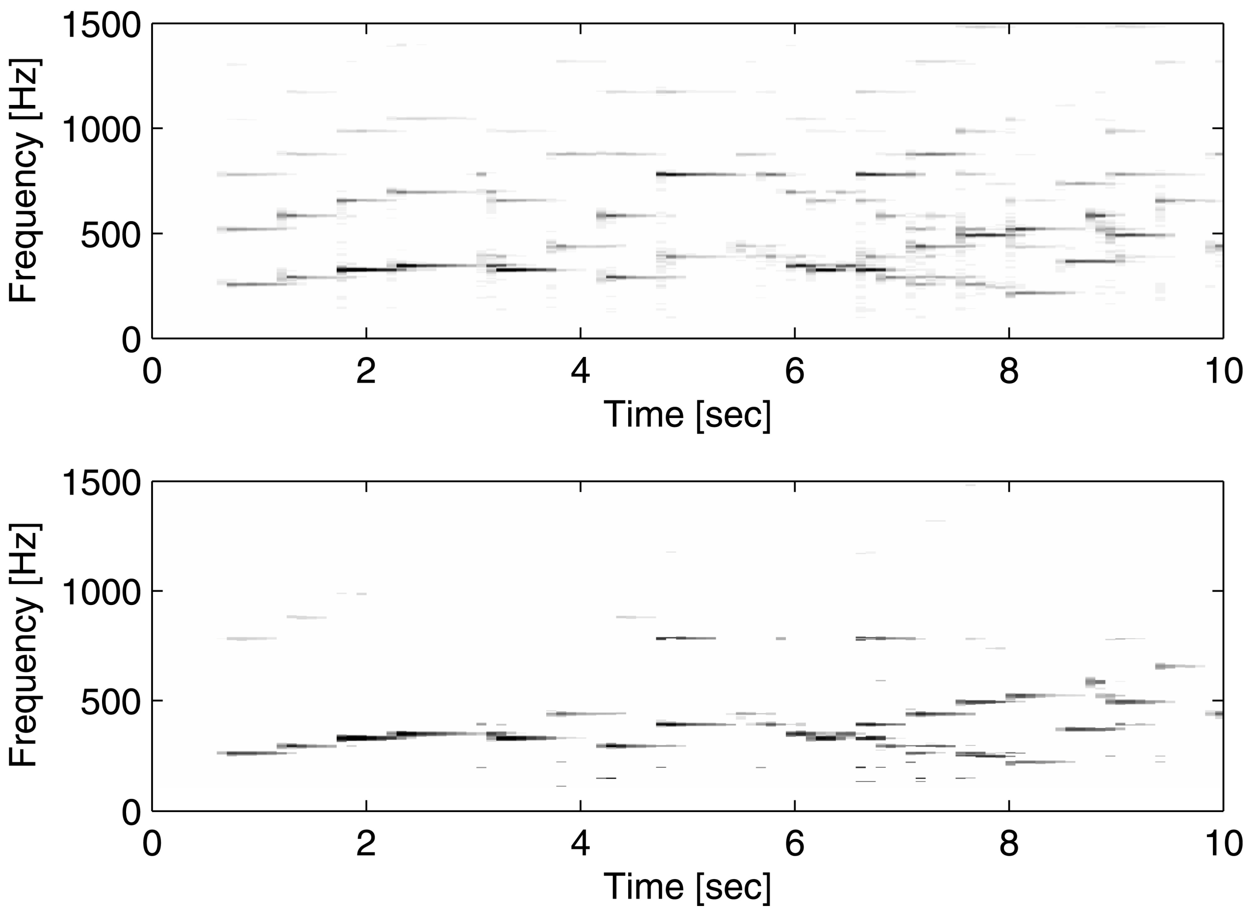
Key estimation
- Cognitive
- HMM

Evaluation

Conclusion

Spectral observation:
Harmonic Peak Subtraction Function

- Examples:
  first 10s of J.S. Bach, Well-Tempered Clavier, 02 Fugue in CM.



System

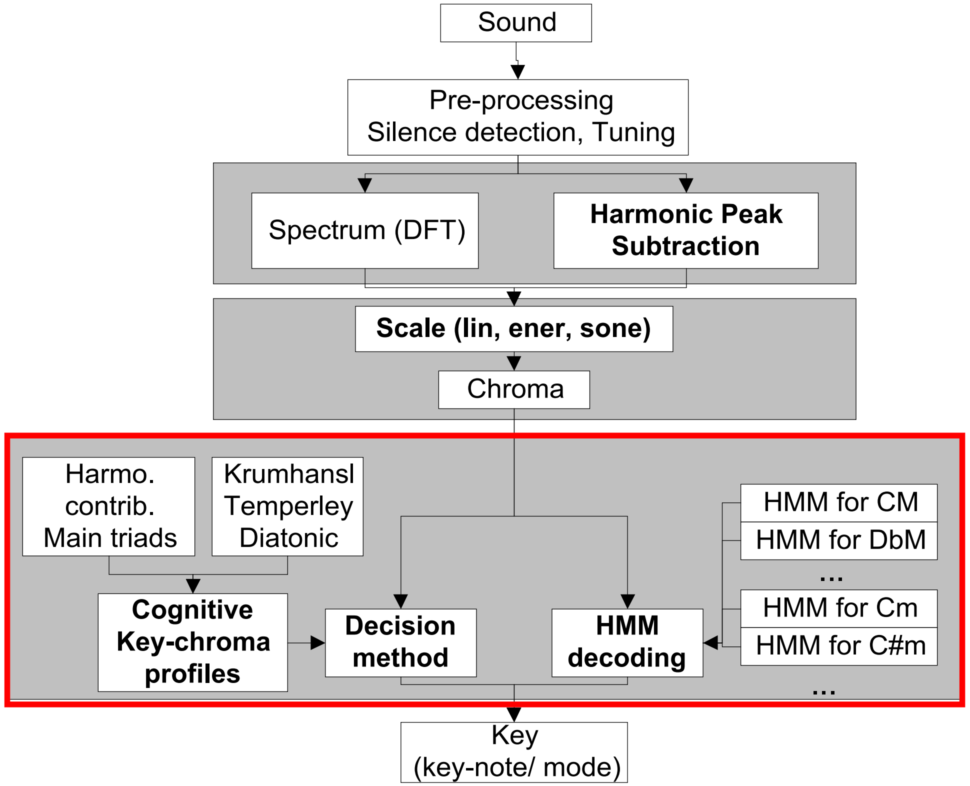
Introduction

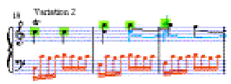
System
- Tuning
- Chroma
- HPS

Key estimation
- Cognitive
- HMM

Evaluation

Conclusion





System

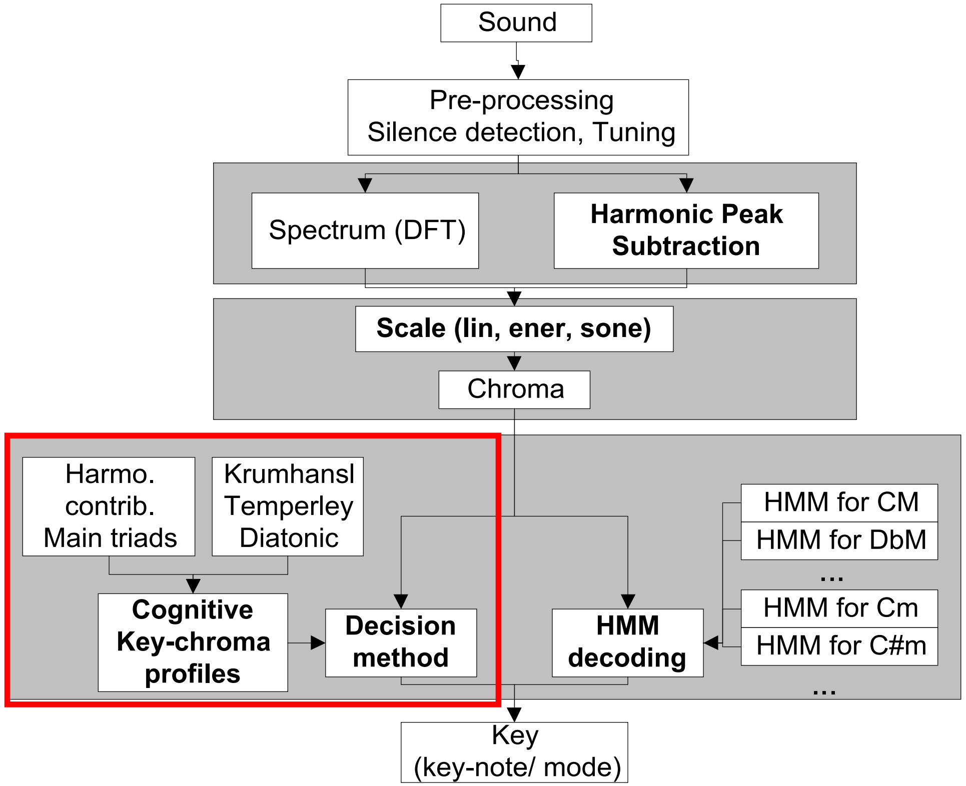
Introduction

System
- Tuning
- Chroma
- HPS

Key estimation
- Cognitive
- HMM

Evaluation

Conclusion



Key-estimation
1) Cognitive-based approach

Introduction

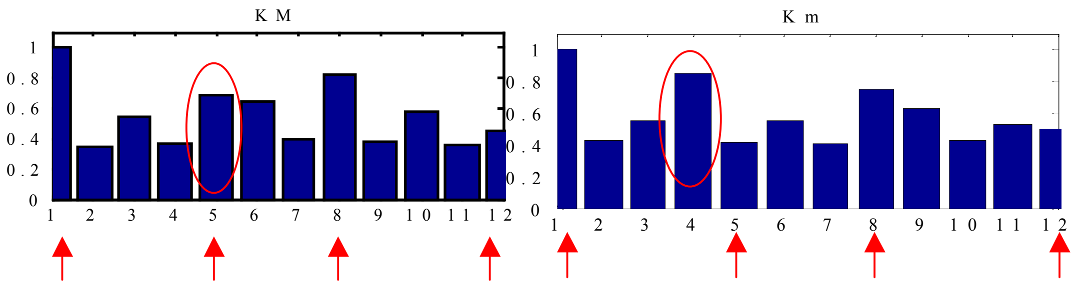
System
- Tuning
- Chroma
- HPS

Key estimation
- Cognitive
- HMM

Evaluation

Conclusion

- Cognitive-based approach
  - 1) key-chroma profiles creation
    - Krumhansl & Schmuckler experiments
      - Tonal profiles for major and minor keys contains 12 values.
      - Values = human ratings of the degree to which each of the 12 chromatic scale tones fit a particular key.



    - Extend Krumhansl & Schmukler (or Temperley, Diatonic) [Gomez]
      - to the polyphonic (several notes) case
      - to the audio (harmonics of each note) case

Key-estimation
1) Cognitive-based approach

Introduction

System
- Tuning
- Chroma
- HPS

Key estimation
- Cognitive
- HMM

Evaluation

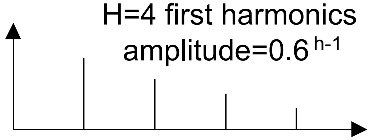
Conclusion

C Major -> 3 main triads:

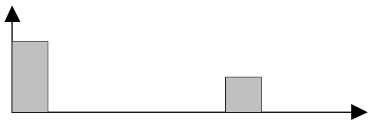
C Major
key-chroma profile

C-E-G

C note theoretical spectrum
H=4 first harmonics
amplitude=0.6 $^{h-1}$



C note theoretical chroma



* Krumhansl CM 1st coef

E note theoretical spectrum → E note theoretical chroma

G note theoretical spectrum → G note theoretical chroma

sum = triad-chroma vector

G-B-D

G note theoretical chroma

B note theoretical chroma

D note theoretical chroma

* Krumhansl CM 8th coef

sum = triad-chroma vector

F-A-C

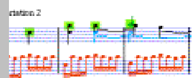
F note theoretical chroma

A note theoretical chroma

C note theoretical chroma

* Krumhansl CM 6th coef

sum = triad-chroma vector



Key-estimation
1) Cognitive-based approach

Introduction

System
- Tuning
- Chroma
- HPS

Key estimation
- Cognitive
- HMM

Evaluation

Conclusion

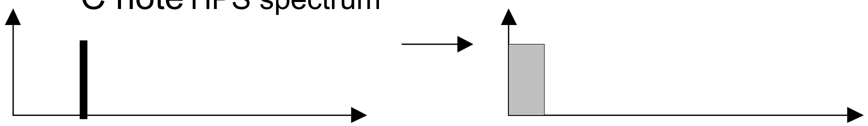
C Major -> 3 main triads:

C Major key-chroma profile

Krumhansl CM 1st coef *

C-E-G

C note HPS spectrum → C note theoretical chroma



E note HPS spectrum → E note theoretical chroma

G note HPS spectrum → G note theoretical chroma

sum = triad-chroma vector

Krumhansl CM 8th coef *

G-B-D

G note theoretical chroma

B note theoretical chroma

D note theoretical chroma

sum = triad-chroma vector

Krumhansl CM 6th coef *

F-A-C

F note theoretical chroma

A note theoretical chroma

C note theoretical chroma

sum = triad-chroma vector

Introduction

System
- Tuning
- Chroma
- HPS

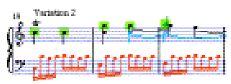
Key estimation
- Cognitive
- HMM

Evaluation

Conclusion

Key-estimation
1) Cognitive-based approach

- Cognitive-based approach
  - 2) key decision method
    - CorrelMeanChroma [Gomez]
      - Key-chroma profile which has the highest correlation with an averaged over-time chroma-vector
    - MeanInstCorrel
      - Maximum of the average correlation between key-chroma profile and instantaneous chroma-vectors
    - ScoreCorrelCumul [Izmirli]
      - At each time, estimate the key-chroma profile that has the highest correlation with a cumulated-over-time chroma vector
      - Assign to it a score (reliability coefficient): distance between the 1st and 2nd correlation
      - Take the key-chroma profile with the highest score over the first 20s



System

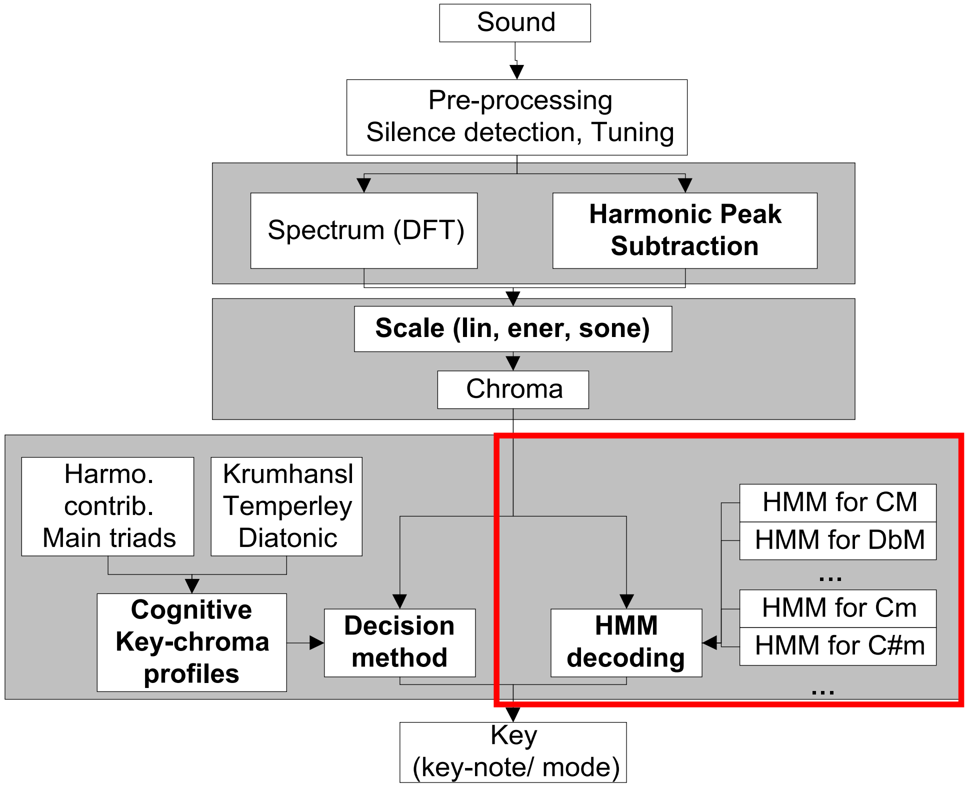
Introduction

System
- Tuning
- Chroma
- HPS

Key estimation
- Cognitive
- HMM

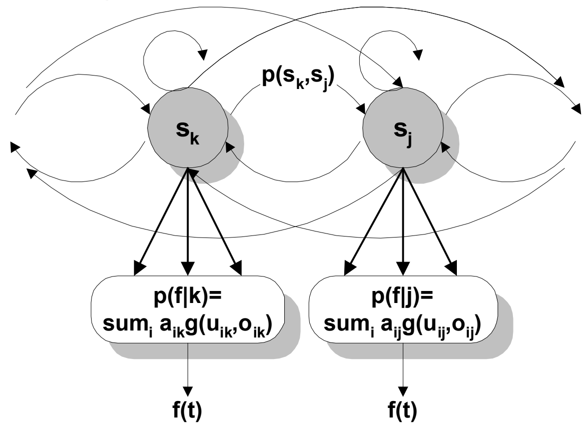
Evaluation

Conclusion



Key-estimation
2) HMM-based approach

- HMM based approach
  - Goal:
    - avoid to make assumptions about the presence of the harmonics of pitch notes, about specific polyphony
    - avoid the choice of a specific pitch distribution profile (Krumhansl, Temperley or Diatonic)
    - allows to take into account possible modulation of key over time
  - Method:
    - learn everything from a music database
    - train a set of hidden Markov models corresponding to the 24 possible keys (12 key-notes * 2 modes)



  - Problem:
    - number of instances strongly differs in the database among the 24 keys
  - Solution:
    - train only two models (one Major and one minor mode model)
    - map them to the 12 key-notes

Key-estimation
2) HMM-based approach

Introduction

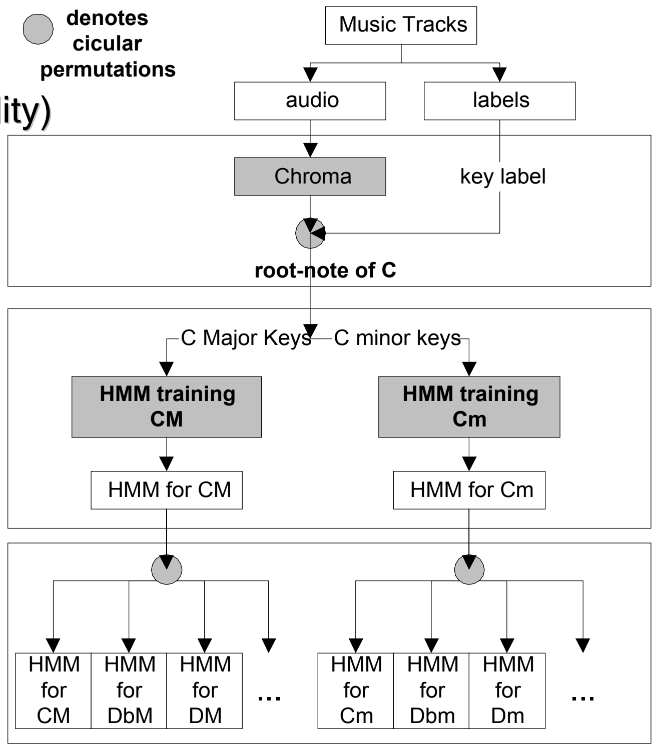
System
- Tuning
- Chroma
- HPS

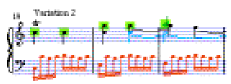
Key estimation
- Cognitive
- HMM

Evaluation

Conclusion

- HMM training
  - 1) map the chroma-vectors of all the tracks to a key-note of C (using circular permutation of chroma-vectors)
  - 2) train one HMM for C Major and one for C minor
  - 3) construct the other Major (minor) keys by mapping the two models to the various key-note (using circular permutation of the mean vector and covariance matrices of the state observation probability)
  - Baum-Welsh algorithm
- Key decision method
  - evaluate the log-likelihood of the chroma-vectors sequence given each of the 24 HMMs
  - forward algorithm
- HMM configuration
  - number of states: S=3,6,12
  - number of mixture per state: M=1,3
  - covariance matrix: full/diagonal





Evaluation

test set

Introduction

System
- Tuning
- Chroma
- HPS

Key estimation
- Cognitive
- HMM

Evaluation

Conclusion

- Test set:
  - 302 European baroque, classical, romantic music extracts
  - Composers: Bach (48), Corelli (12), Handel (16), Telleman (17), Vivaldi(6), Beethoven (33), Haydn (23), Mozart (33), Brahms (32), Chopin (29), Dvorak (18), Schubert (23), Schuman (7)
  - Instruments: solo keyboard (piano, harpichord), chamber and orchestra music
  - no opera or choir music, only first movement (label of the piece)
  - Source: Naxos web radio service
  - Remark: tuning of part of the baroque pieces were based on A4=415Hz

	Keyboard	Chamber	Orchestra	
Baroque	61	37	6	104
Classical	42	N/A	47	89
Romantic	46	10	53	109
	149	47	106	

Evaluation
method

Introduction

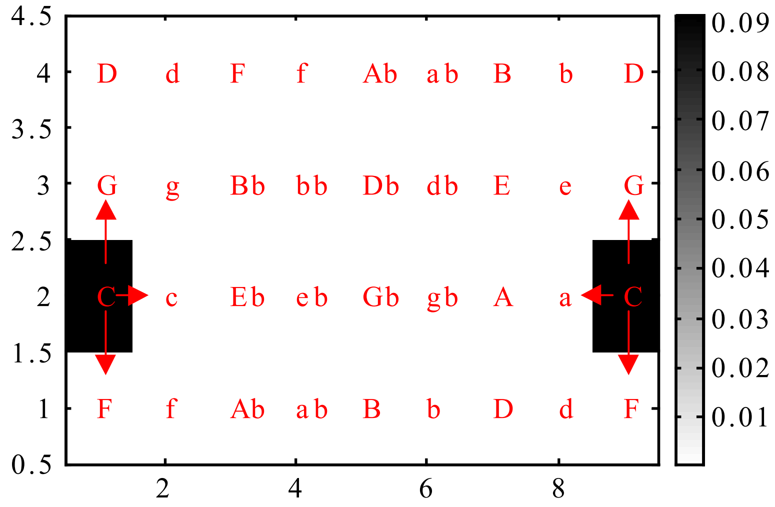
System
- Tuning
- Chroma
- HPS

Key estimation
- Cognitive
- HMM

Evaluation

Conclusion

- Evaluation method
  - correct key
  - correct key-note
  - correct mode



- MIREX 2005 key estimation contest score
  - 1 for correct key estimation,
  - 0.5 for perfect fifth relationship between estimated and ground-truth key,
  - 0.3 if detection of relative major/minor key,
  - 0.2 if detection of parallel major/minor key.

- HMM evaluation: ten-fold cross-validation

Evaluation
results

Introduction

System
- Tuning
- Chroma
- HPS

Key estimation
- Cognitive
- HMM

Evaluation

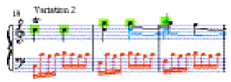
Conclusion

- Results for HPS

		MeanInstCorrel				ScoreCorrelCumul			
		MIREX	Correct key	Correct key-note	Correct mode	MIREX	Correct key	Correct key-note	Correct mode
DFT ampl	H=1	86,1	79,8	82,8	91,4	86,7	81,8	84,4	91,1
	H=4	**88,4**	83,4	86,4	92,1	**87,9**	**83,8**	86,1	91,7
DFT ener	H=1	84,9	78,5	80,8	90,7	80,5	73,2	75,2	86,4
	H=4	85,1	78,8	80,8	91,1	79,7	71,9	74,2	86,1
DFT sone	H=1	84,6	76,5	79,8	90,7	86,6	81,8	85,4	90,7
	H=4	**88**	82,5	84,8	**92,7**	**87,6**	**83,8**	**87,1**	**92,1**
HPS ampl	H=1	86,4	80,5	83,1	91,4	84,3	79,5	82,8	88,1
	H=4	86	80,1	82,8	91,4	81,9	75,5	80,1	86,8
HPS ener	H=1	84,6	77,8	80,8	90,4	82,7	76,5	79,8	87,4
	H=4	81,6	73,2	76,8	88,4	76,2	67,5	71,9	82,5
HPS sone	H=1	85,9	80,5	83,1	90,4	**89,1**	**84,8**	**87,7**	**93**
	H=4	87,8	83,1	85,4	92,1	88,2	84,1	87,1	92,1
HMM DFT sone						85,5	81	87,4	88

- Best results (KE and MIREX) using the HPS in sone scale + ScoreCorrelCumul (MI=89.1%, KE=84.8%)

- Changing only one process (scale, H, decision method) can drastically change the results
- Choice of a specific decision method: no clear trend (depending on MI or KE)
- Choice of the scale: energy scale decreases both MI and KE, amplitude and sone close results for DFT; best results with sone for the HPS
- Choice of H: H=4 for the DFT, H=1 for the HPS
- Choice of the observation: the HPS but in sone scale (amplitude compression, problem of estimating the amplitude for the HPS)



Evaluation

results by music genre

Introduction

System
- Tuning
- Chroma
- HPS

Key estimation
- Cognitive
- HMM

Evaluation

Conclusion

- Results by Music Genre

	Keyboard	Chamber	Orchestral	
Baroque	89,8	94,6	100	92,1
Classical	96,2	N/A	93	94,5
Romantic	85,4	92	76,8	81,8
	90,3	94	85,3	

- Results by music genre / instrumentation type
  - the results strongly depends on the considered music genre
  - lowest recognition rate for the romantic period (81.8%)
  - Brahms, Schuman contains mainly a neighboring tonality in the first 20s

Conclusion and future works

Conclusion

- best results obtained with the HPS function (89.1%)
- estimation of key based on HMM (85.5%) very promising (no knowledge was introduced !)
  but remains lower than the one obtained with cognitive-based approach (87.9%)
- results strongly depend on the music period considered (romantic music)
  -> limitation of such a straightforward approach

Future works

- Improving the amplitude associated to the peaks of the HPS
- Testing/training the HMMs on whole track duration
- Testing HMM with HPS
- Testing the performances of a multi-pitch detection algorithm mapped to the chroma domain in order to know the limits of the chroma based approach