Music Source Separation and its Applications to MIR

Nobutaka Ono and Emmanuel Vincent
The University of Tokyo, Japan
INRIA Rennes - Bretagne Atlantique, France

Tutorial supported by the VERSAMUS project
http://versamus.inria.fr/

Contributions from Shigeki Sagayama, Kenichi Miyamoto, Hirokazu Kameoka, Jonathan Le Roux, Emiru Tsunoo, Yushi Ueda, Hideyuki Tachibana, Geroge Tzanetakis, Halfdan Rump, Other members of IPC Lab#1

Outline

- Introduction
- Part I: Brief Introduction of State-of-the-arts
  - Singer/Instrument Identification
  - Audio Tempo Estimation
- Part II: Harmonic/Percussive Sound Separation
  - Motivation and Formulation
  - Open Binary Software
- Part III: Applications of HPSS to MIR Tasks
  - Audio Chord Estimation
  - Melody Extraction
  - Audio Genre Classification
- Conclusions

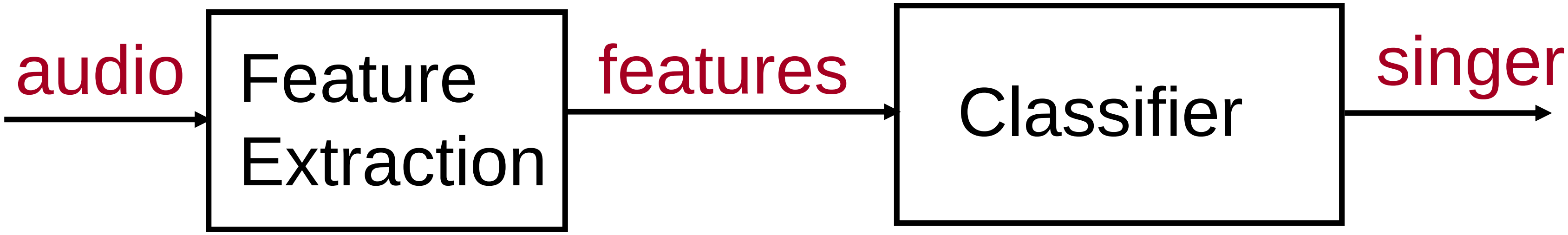
Introduction

- Focus of the second half of this tutorial is to clarify
  - What source separation has been used for MIR?
  - How does it improve performance of MIR tasks?
- Examples:
  - Multi pitch estimation
    Task itself is tightly coupled with source separation.
  - Audio genre classification
    How source separation is useful?
    Not straightforward.

Part I: Brief Introduction of State-of-the-arts

Singer Identification

- Task: Identify a singer from music audio with accompaniment
- Typical approach



Accompaniment Sound Reduction [Fujihara2005]

- Pre-dominant F0 based voice separation

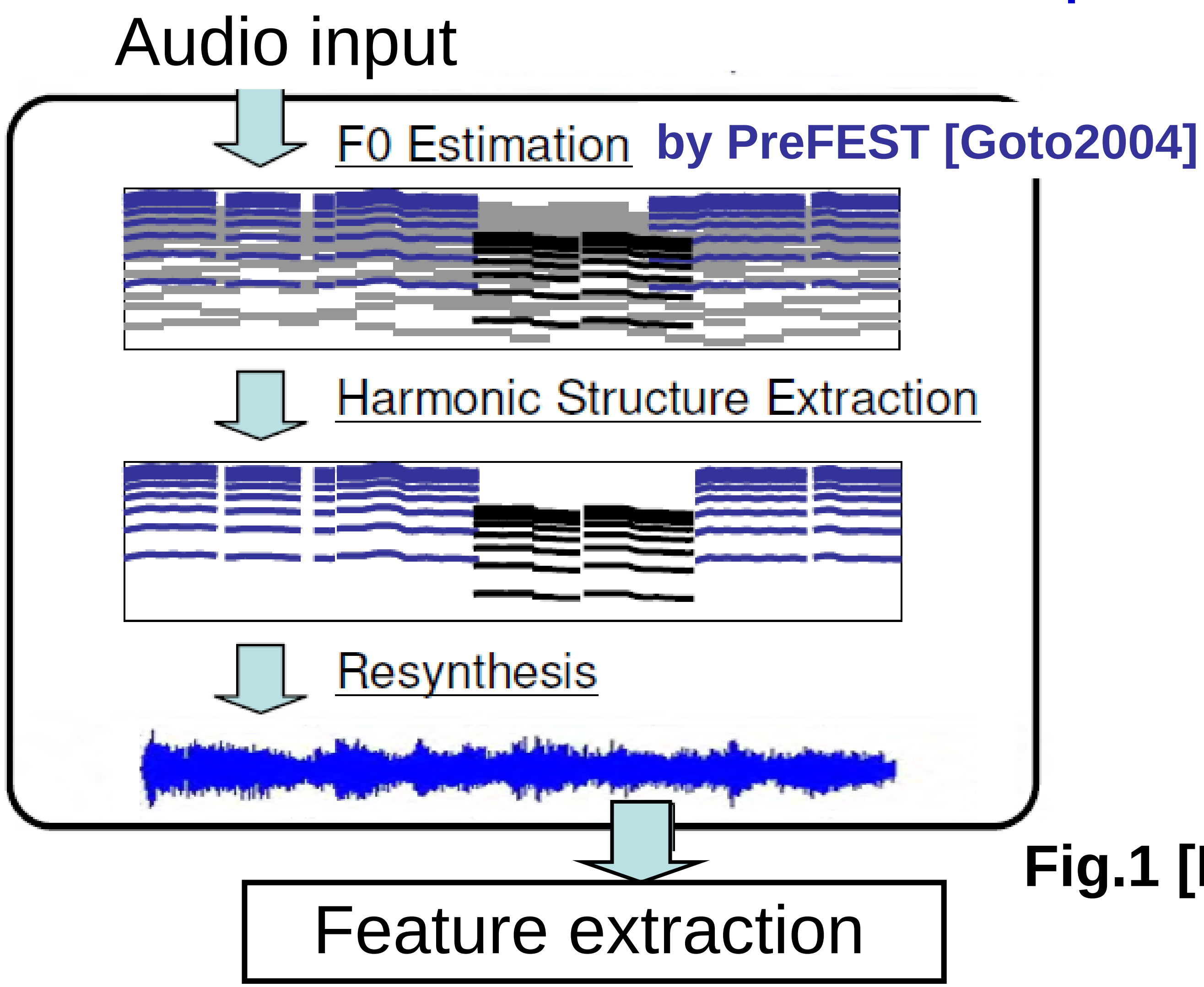


Fig.1 [Fujihara2005]

Reliable Frame Selection [Fujihara2005]

- Only reliable frame is used for classification

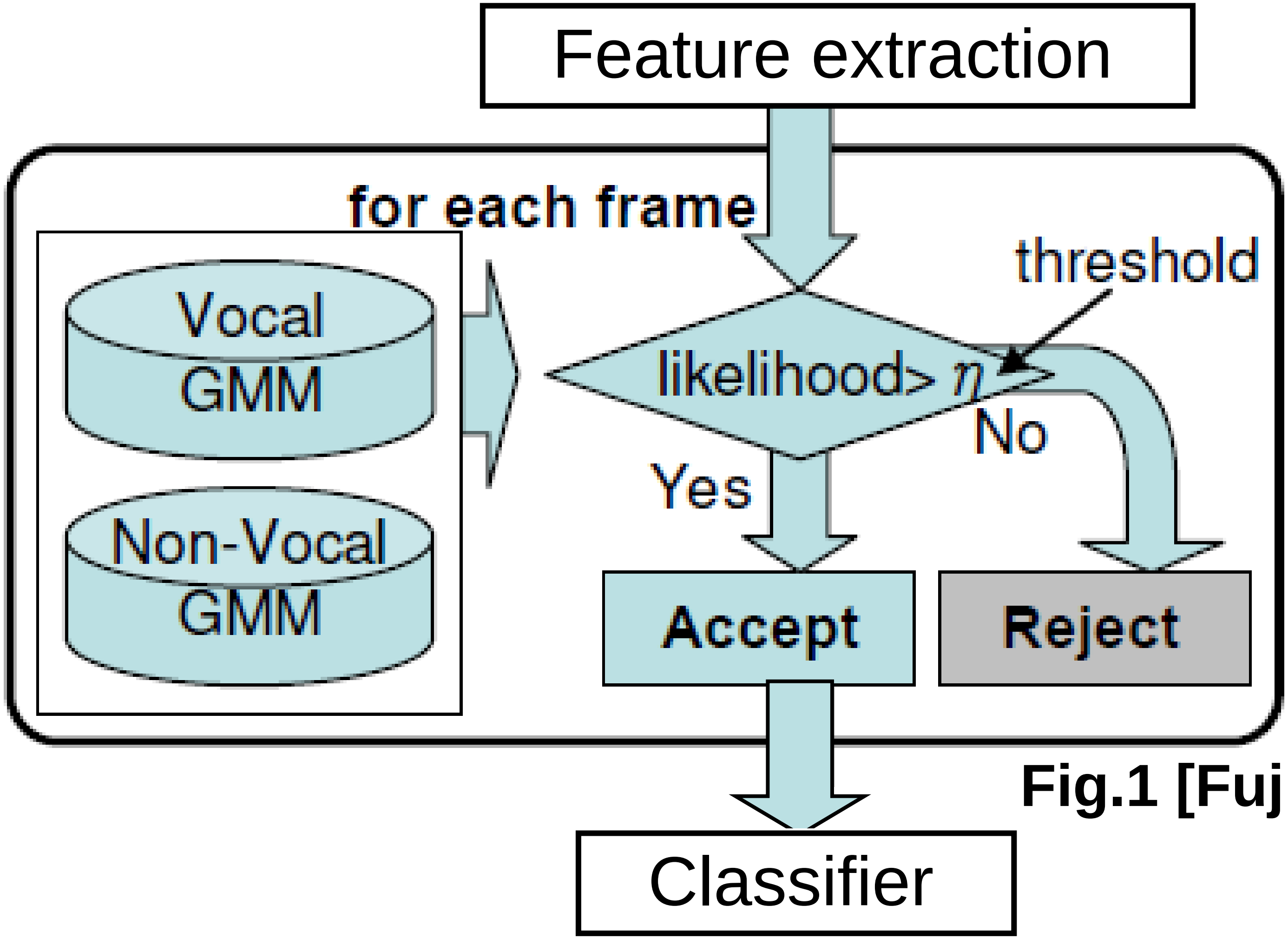
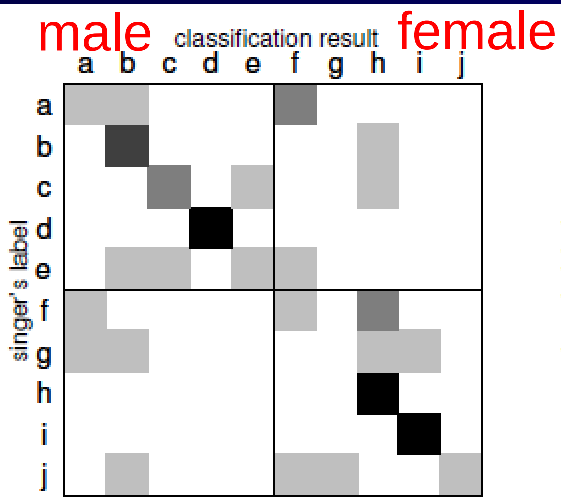
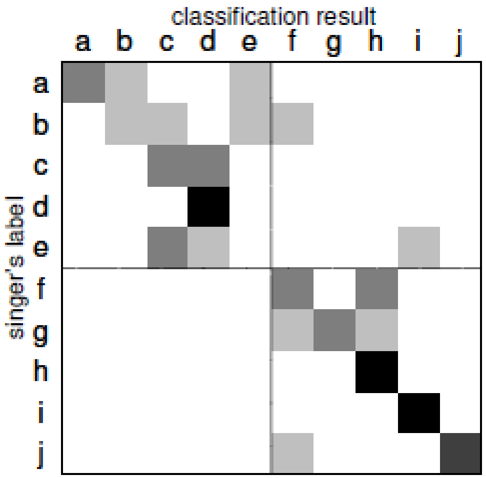


Fig.1 [Fujihara2005]

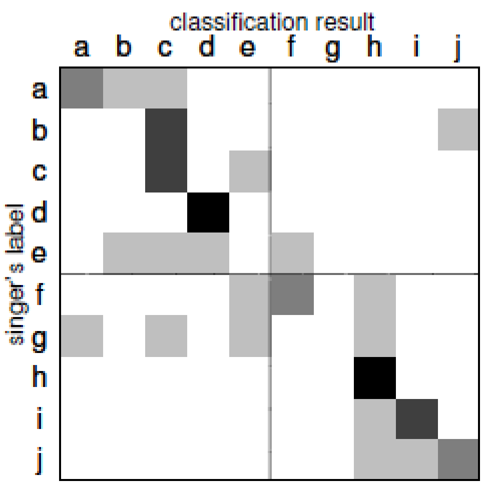
Evaluation by Confusing Matrix



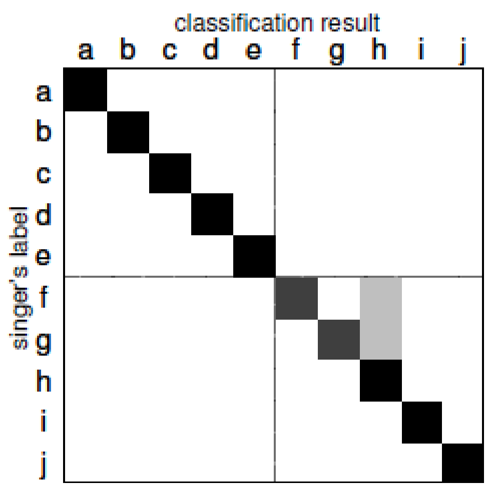
baseline



reduction only



selection only



reduction and selection

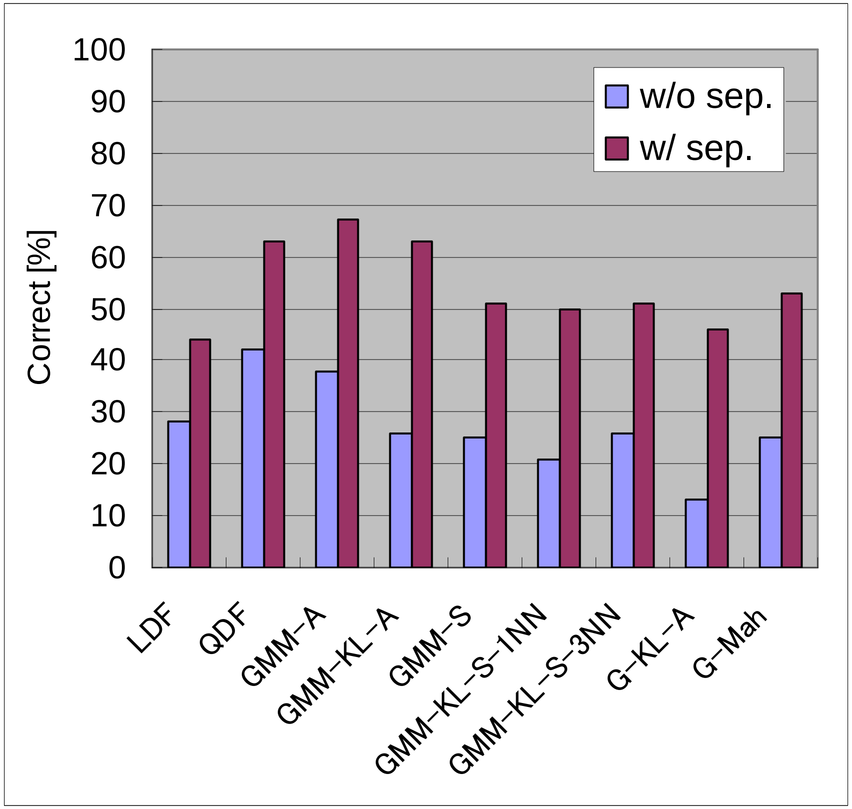
- Male/female confusion is decreased by accompaniment reduction.
- Combination of reduction and selection much improves performance.

Fig. 3 [Fujihara2005]

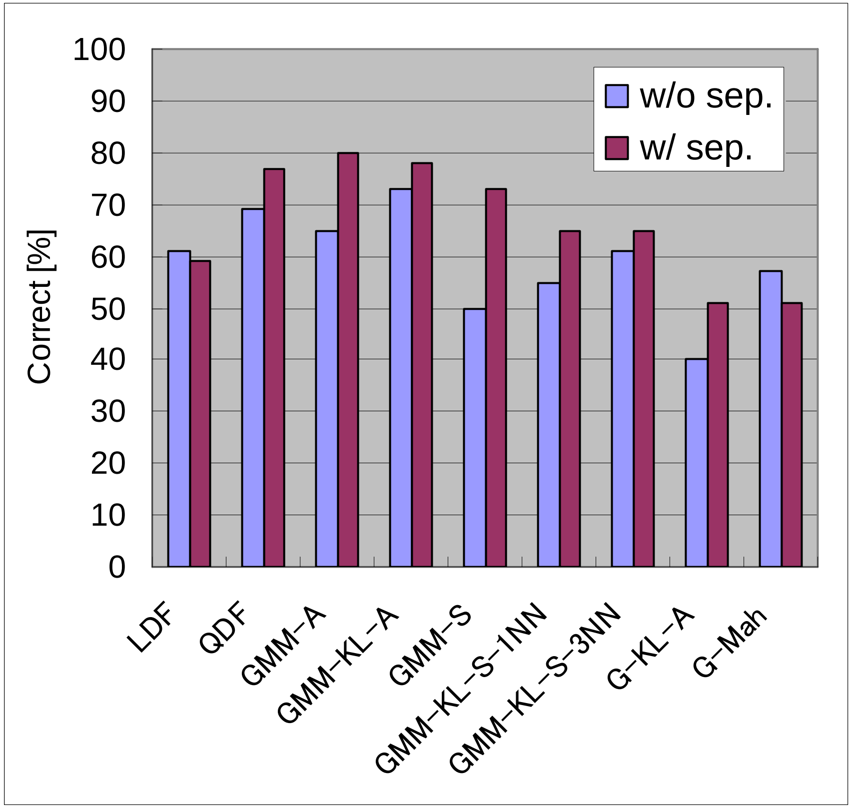
Vocal Separation Based on Melody Transcriber

- Melody-F0-based Vocal Separation [Mesaros2007]
  - Estimate melody-F0 by melody transcription system [Ryynanen2006].
  - Generate harmonic overtones at multiple of estimated F0.
  - Estimate amplitudes and phases of overtones based on cross correlation between original signal and complex exponentials.
- They evaluate the effect of separation in singer identification performance using by different classifiers.

Evaluation by Identification Rate



Singing to Accompaniment Ratio: -5dB



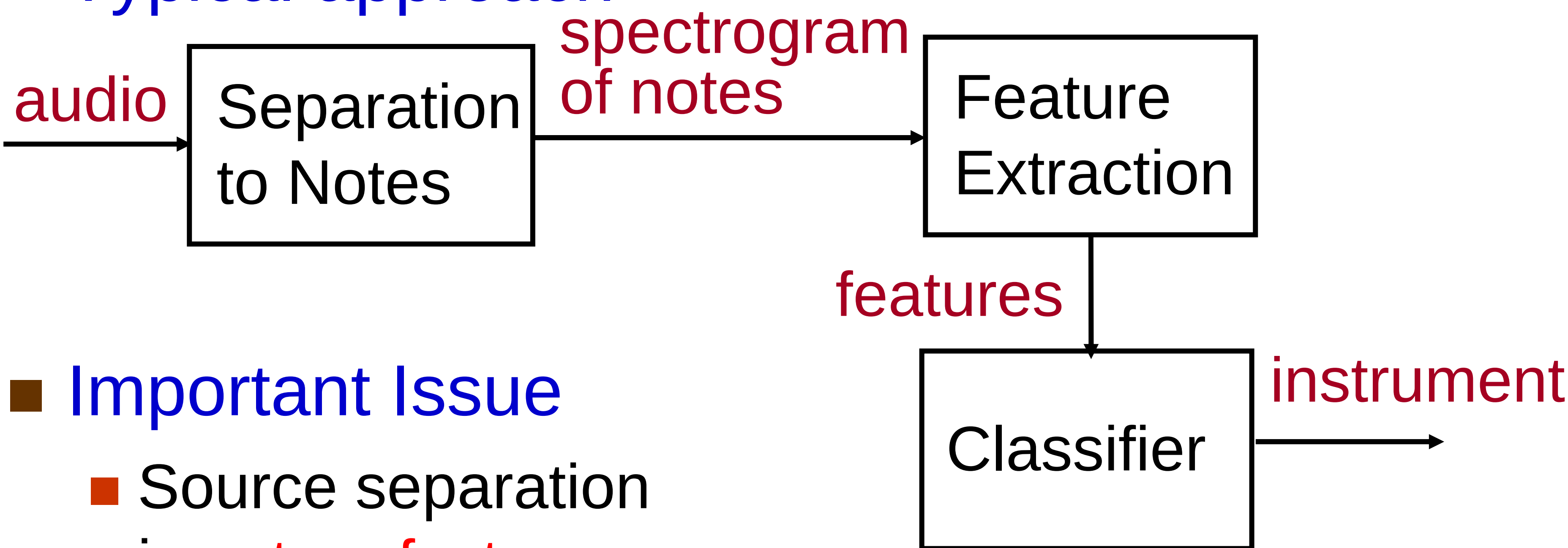
Singing to Accompaniment Ratio: 15dB

Generated by Table 1 and 2 [Mesaros2007]

Performance is much improved, especially in low singing-to-accompaniment ratio.

Instrument Identification

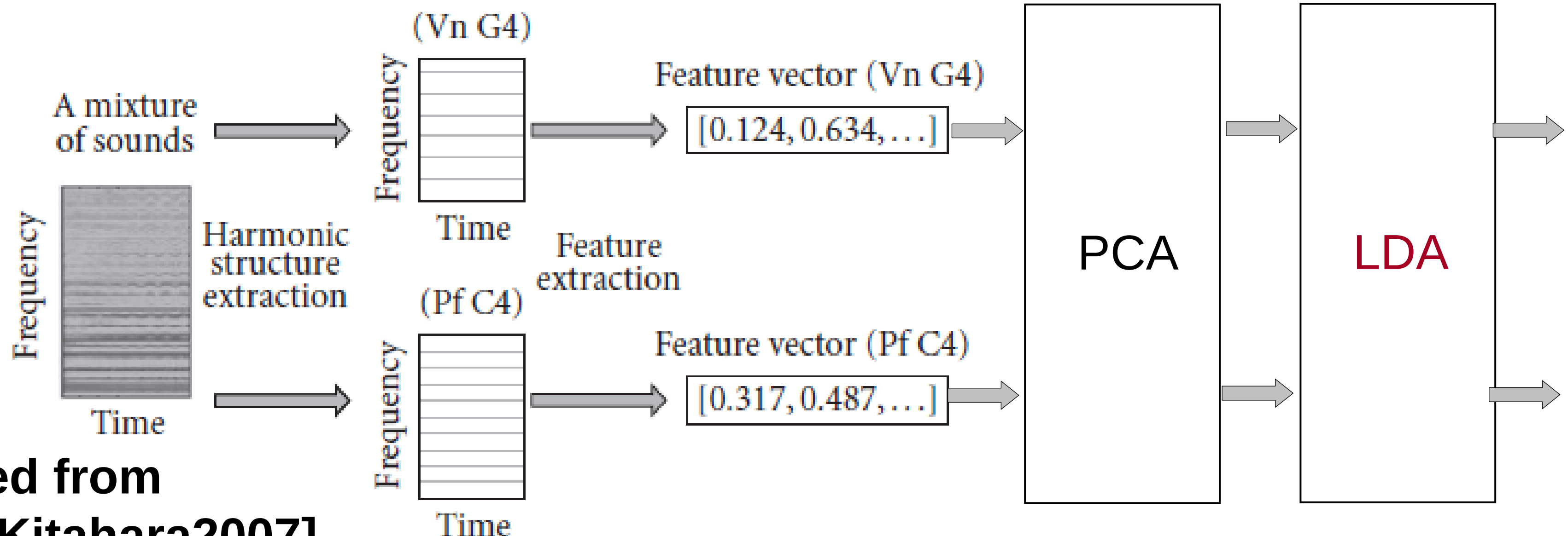
- Task: Determine instruments present in music piece
- Typical approach



- Important Issue
  - Source separation is not perfect. How to reduce errors?

Feature Weighting [Kitahara2007]

- Feature vectors of each instrument are collected from polyphonic music for training.
- Robustness of each feature is evaluated by ratio of intra-class variance to inter-class variance: Applying Linear discriminant analysis (LDA) for feature weighting.



Modified from Fig. 1 [Kitahara2007]

Effectiveness of Feature Weighting

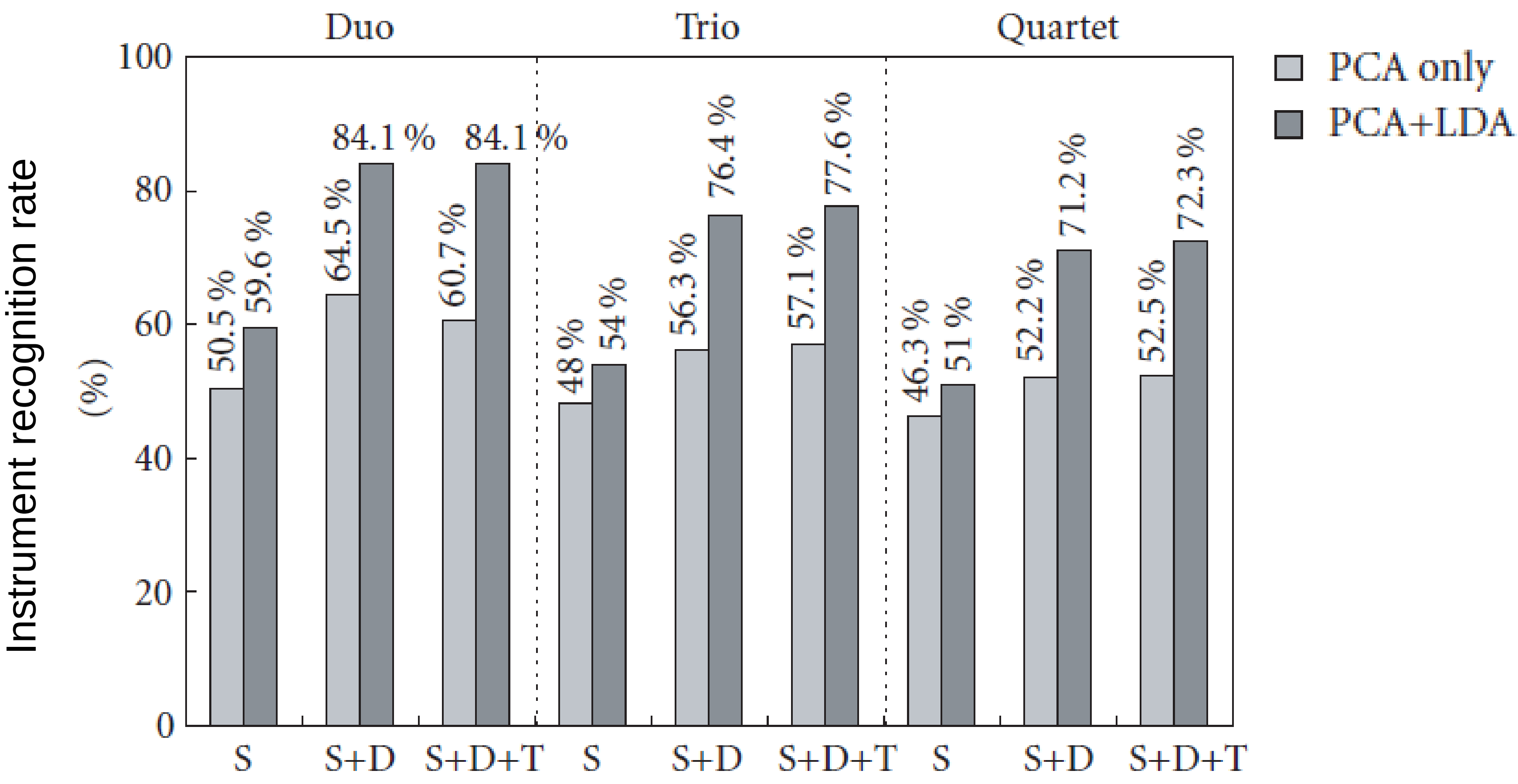
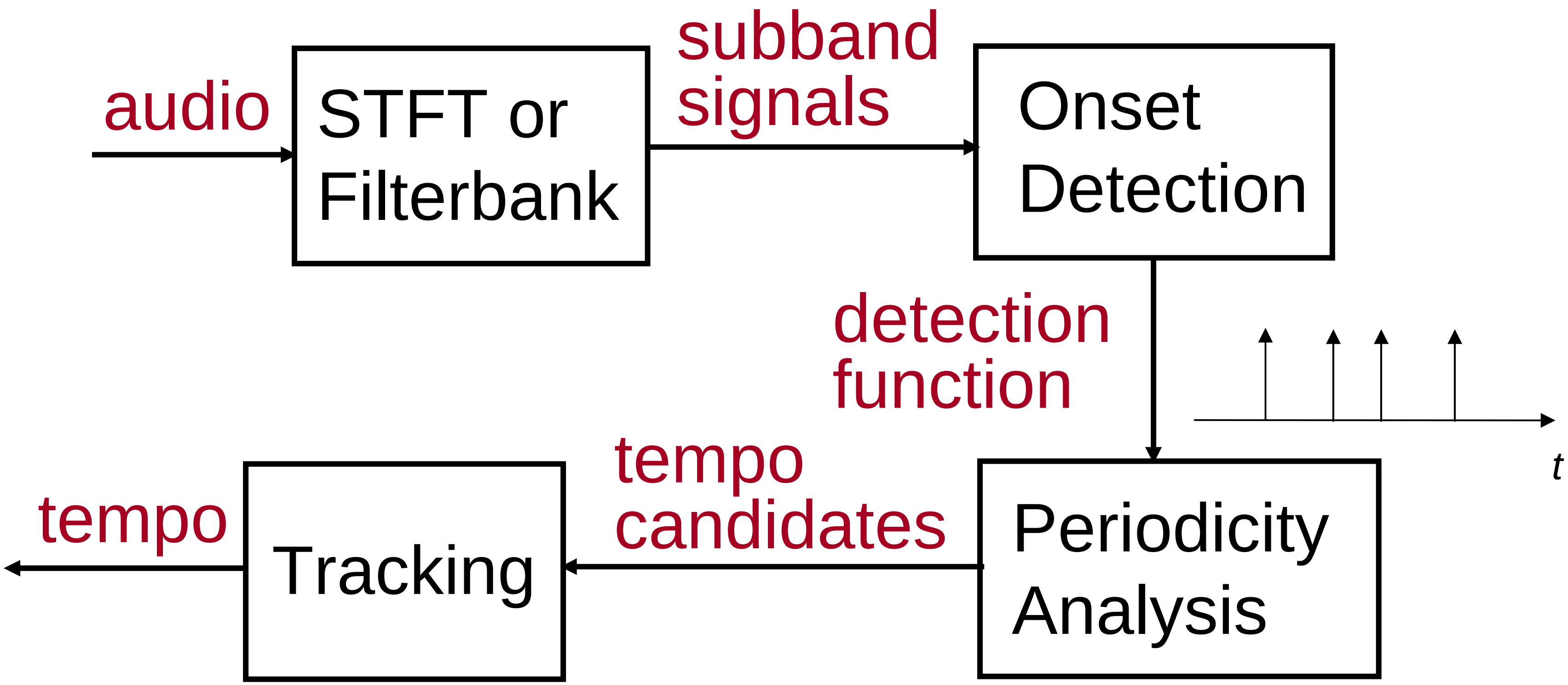


Fig. 6 [Kitahara2007]

Feature weighting by LDA improves recognition rate.

Audio Tempo Estimation

- Task: Extract tempo from musical audio
- Typical approach:



Applying Harmonic+Noise Model

- Harmonic+Noise model is applied before calculating detection function [Alonso2007]

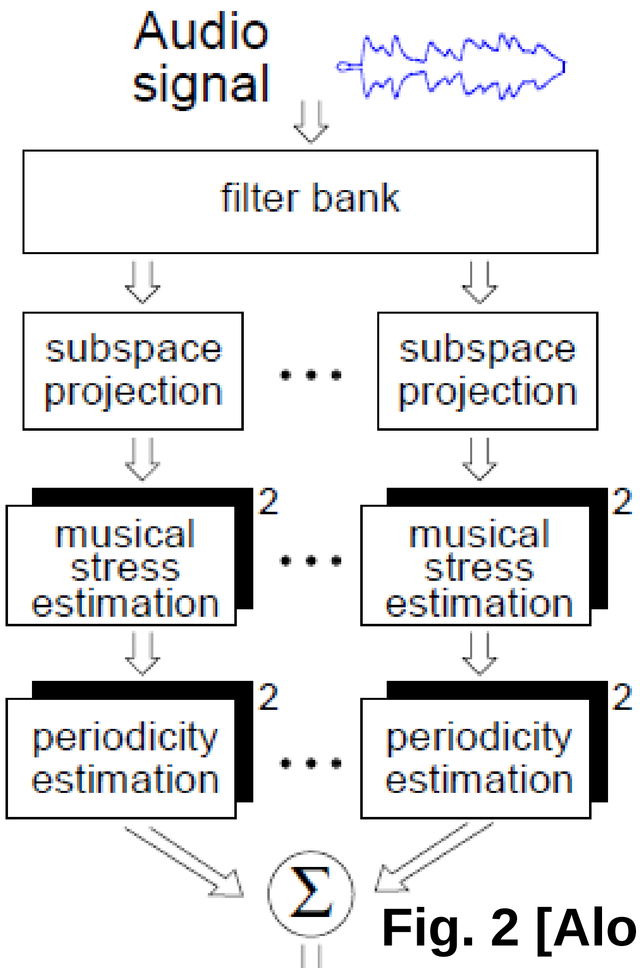
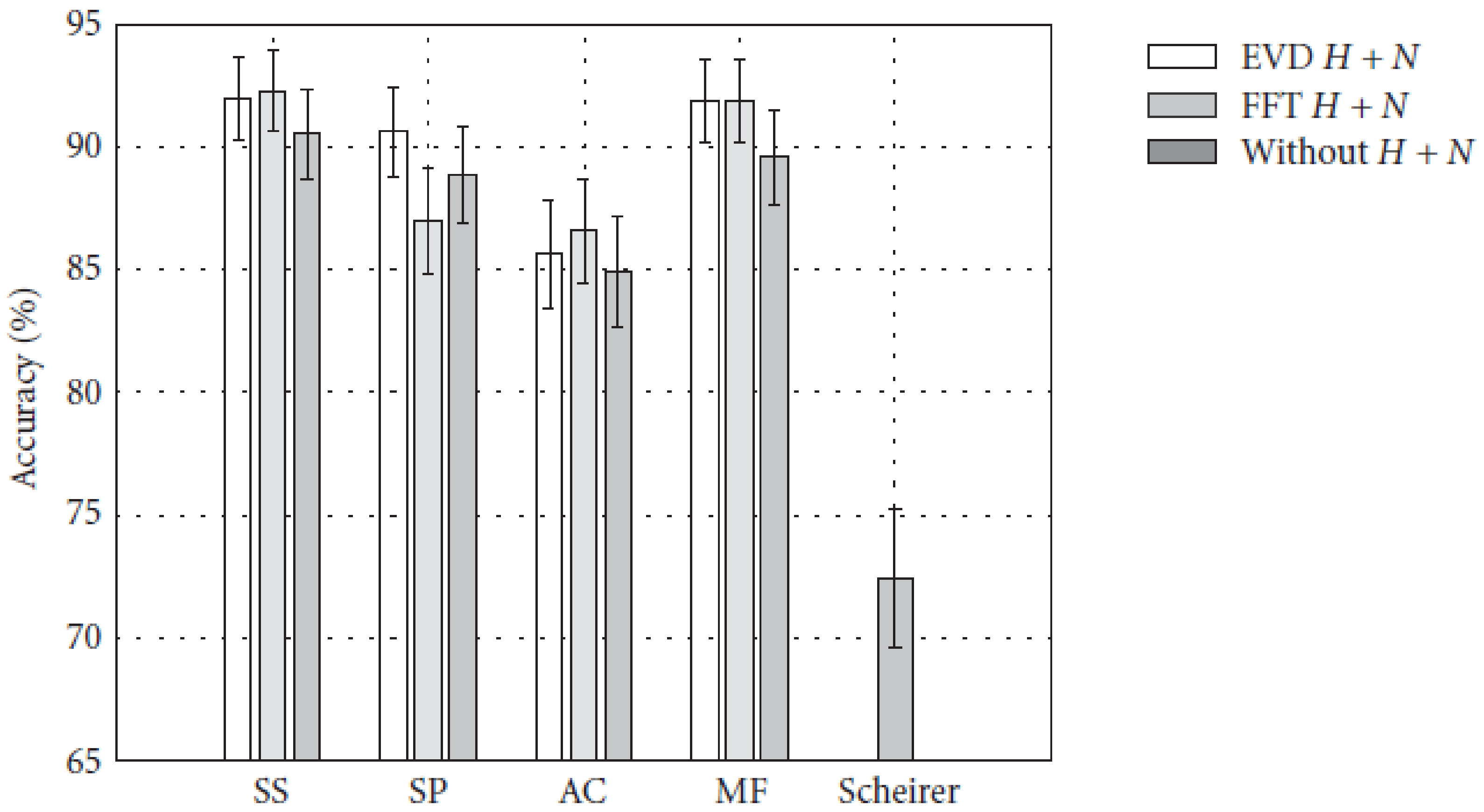


Fig. 2 [Alonso2007]

Source separation based on harmonic + noise model

Detection functions are calculated from both of harmonic component and noise component, and then, they are merged.

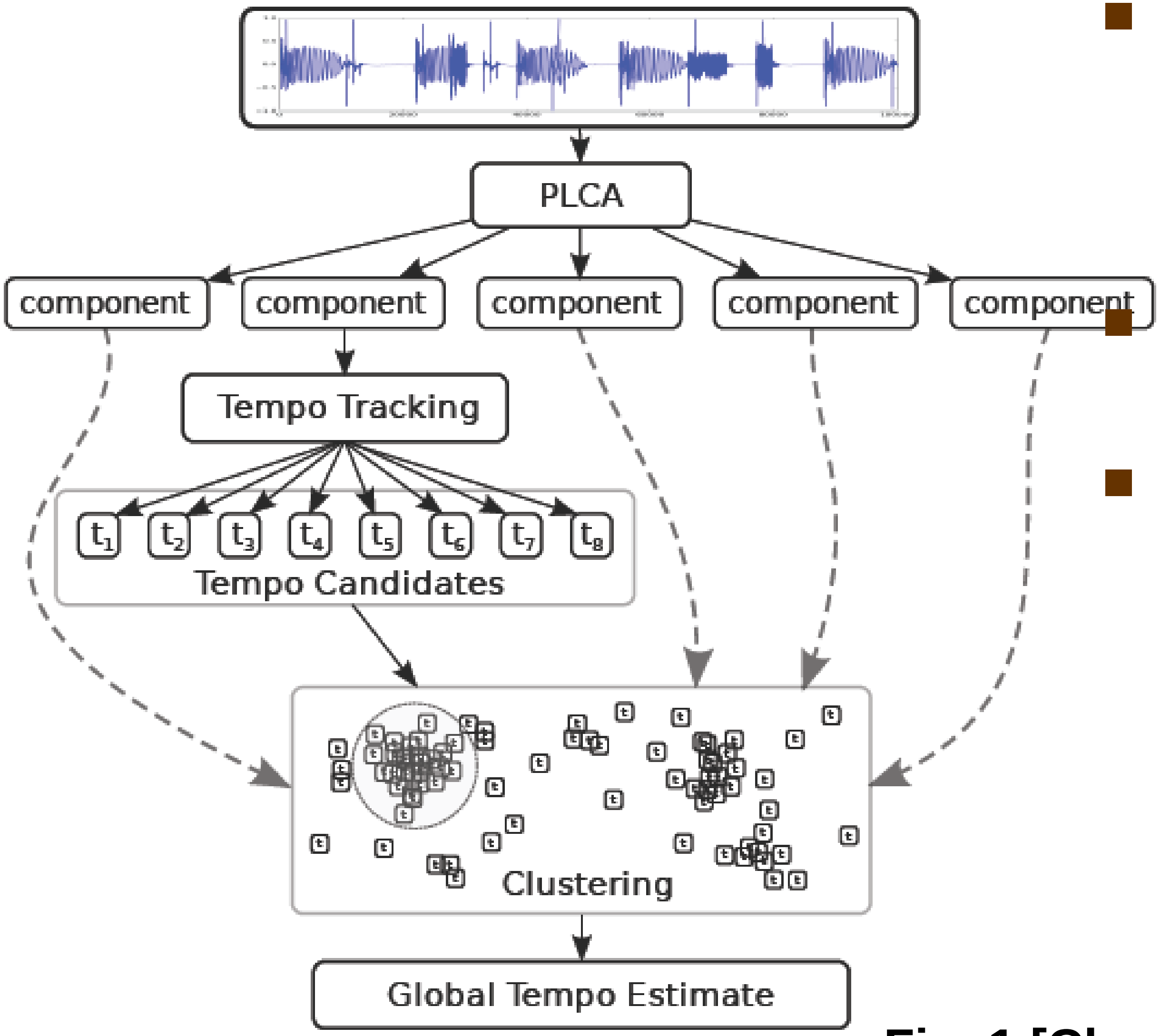
Influence of H+N Model



Algorithms of periodicity detection **Fig. 14 [Alonso2007]**

Separation based on H+N model shows better results.

Applying PLCA



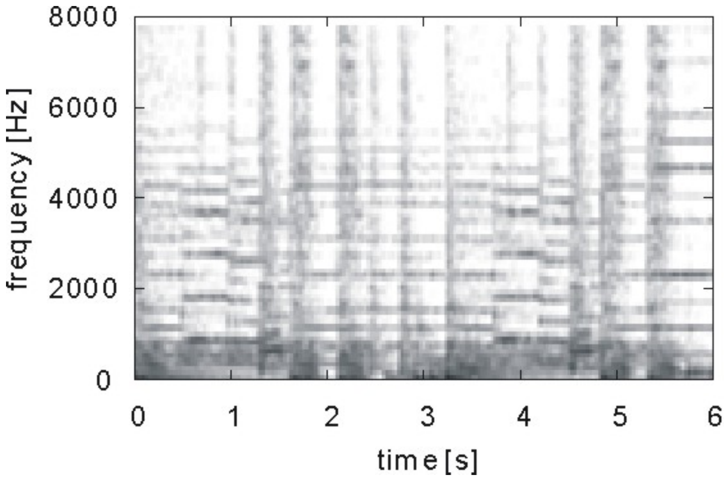
- PLCA (Probabilistic Latent Component Analysis), NMF-like method is applied.
- It increases much candidates of tempo.
- They report its effectiveness.

Fig. 1 [Chordia2009]

Part II: Harmonic/Percussive Sound Separation

Motivation and Goal of HPSS

- Motivation: Music consists of two different components



example of a popular music (RWC-MDB-P034)

harmonic component

percussive component

- Goal: Separation of a monaural audio signal into harmonic and percussive components
- Target: MIR-related tasks
  - multi-pitch analysis, chord recognition… H-related
  - beat tracking, rhythm recognition… P-related

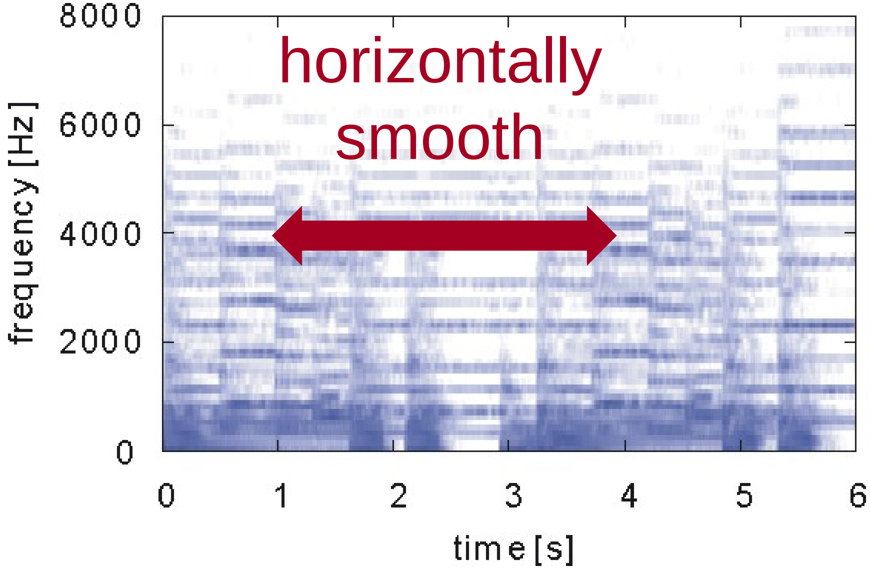
Related Works to H/P Separation

- Source separation into multiple components followed by classification
  - ICA and classification [Uhle2003]
  - NMF and classification [Helen2005]
- Steady + Transient model
  - Adaptive phase vocoder
  - Subspace projection
  - Matching persuit

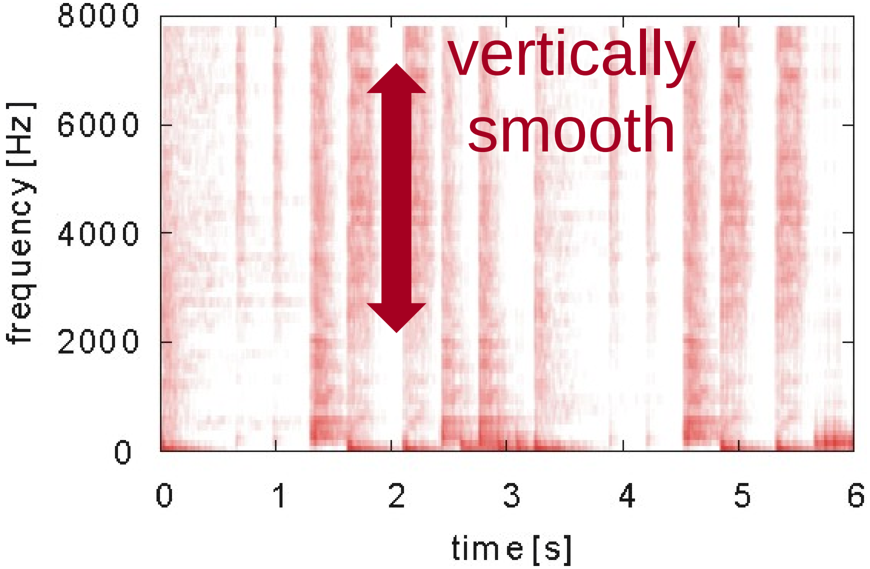
  …etc

  Good review is provided in [Daudet2005]
- Baysian NMF [Dikmen2009]

Point: Anisotropy of Spectrogram



harmonic component

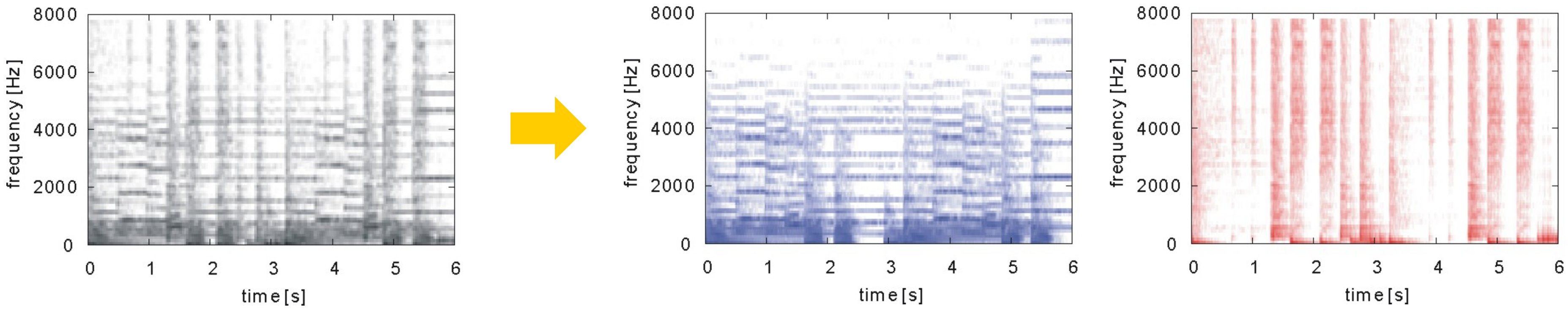


percussive component

H/P Separation Problem

- Problem:

  Find $H_{t,\omega}$ and $P_{t,\omega}$ from $W_{t,\omega}$ on power spectrogram



$W_{t,\omega}$ $H_{t,\omega}$ $P_{t,\omega}$

- Requirements:

  1) $H_{t,\omega}$: horizontally smooth

  2) $P_{t,\omega}$: vertically smooth

  3) $H_{t,\omega}$ and $P_{t,\omega}$: non-negative

  4) $H_{t,\omega}$ + $P_{t,\omega}$: should be close to $W_{t,\omega}$

Formulation of H/P Separation (1/2)

- Formulation as an Optimization Problem:
  - Objective function to minimize

$$J(\boldsymbol{H}, \boldsymbol{P}) = \underbrace{D(\boldsymbol{W}, \boldsymbol{H} + \boldsymbol{P})}_{\text{Closeness cost}} + \underbrace{C_H(\boldsymbol{H}) + C_P(\boldsymbol{P})}_{\text{Smoothness cost}}$$

  - Under constraints:
    - $H_{t,\omega} \geqq 0$
    - $P_{t,\omega} \geqq 0$

In MAP estimation context, they are corresponding likelihood term and prior term, respectively.

Formulation of H/P Separation (2/2)

- Closeness cost function: I-divergence

$$D(\boldsymbol{W}, \boldsymbol{H}+\boldsymbol{P}) = \sum_{\omega,\tau} \left\{ W_{\omega,\tau} \log \frac{W_{\omega,\tau}}{H_{\omega,\tau}+P_{\omega,\tau}} - W_{\omega,\tau} + H_{\omega,\tau} + P_{\omega,\tau} \right\}$$

- Smoothness cost function: Square of difference

$$C_H(\boldsymbol{H}) = \sum_{\omega,\tau} \underline{\frac{1}{2\sigma_H^2}} (H_{\omega,\tau-1}^{\gamma} - H_{\omega,\tau}^{\gamma})^2$$

$$C_P(\boldsymbol{H}) = \sum_{\omega,\tau} \underline{\frac{1}{2\sigma_P^2}} (P_{\omega-1,\tau}^{\gamma} - P_{\omega,\tau}^{\gamma})^2$$

$\gamma = 0.5$
for scale invariance

Weights to control two smoothness

- A variance modeling based separation using
  - Poisson observation distribution
  - Gaussian continuity priors

[Miyamoto2008, Ono2008, etc]

Update Rules

- Update alternatively two kinds of variables:
  - H and P:

$$H_{\omega,\tau} \leftarrow \left(\frac{b_{H\,\omega,\tau} + \sqrt{b_{H\,\omega,\tau}^{2} + 4a_{H\,\omega,\tau} c_{H\,\omega,\tau}}}{2a_{H\,\omega,\tau}} \right)^2$$

$$P_{\omega,\tau} \leftarrow \left(\frac{b_{P\,\omega,\tau} + \sqrt{b_{P\,\omega,\tau}^{2} + 4a_{P\,\omega,\tau} c_{P\,\omega,\tau}}}{2a_{P\,\omega,\tau}} \right)^2$$

  - Auxiliary variables:

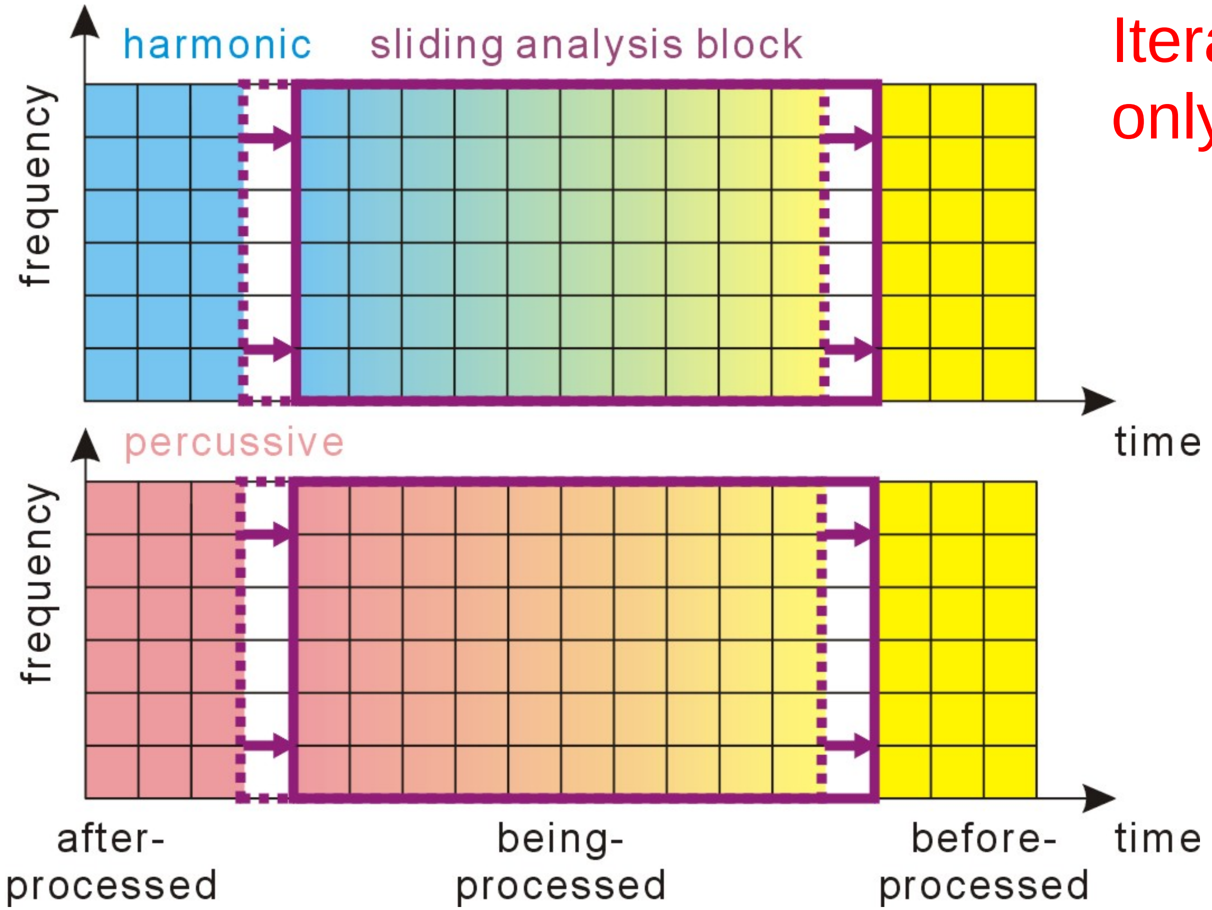
$$\begin{aligned}
m_{P\,\omega,\tau} &= \frac{P_{\omega,\tau}}{H_{\omega,\tau} + P_{\omega,\tau}} & \quad m_{H\,\omega,\tau} &= \frac{H_{\omega,\tau}}{H_{\omega,\tau} + P_{\omega,\tau}} \\
a_{P\,\omega,\tau} &= \frac{2}{\sigma_P^2} + 2 & a_{H\,\omega,\tau} &= \frac{2}{\sigma_H^2} + 2 \\
b_{P\,\omega,\tau} &= \frac{(\sqrt{P_{\omega-1,\tau}} + \sqrt{P_{\omega+1,\tau}})}{\sigma_P^2} & b_{H\,\omega,\tau} &= \frac{(\sqrt{H_{\omega,\tau-1}} + \sqrt{H_{\omega,\tau+1}})}{\sigma_H^2} \\
c_{P\,\omega,\tau} &= 2m_{P\,\omega,\tau} W_{\omega,\tau} & c_{H\,\omega,\tau} &= 2m_{H\,\omega,\tau} W_{\omega,\tau}
\end{aligned}$$

Separation Examples

Music piece	original	H	P
RWC-MDB-P-7 “PROLOGUE ”			
RWC-MDB-P-12 “KAGE-ROU ”			
RWC-MDB-P-18 “True Heart”			
RWC-MDB-P-25 “tell me”			
RWC-MDB-J-16 “Jive ”			

Real-Time Implementation

- Sliding Block Analysis

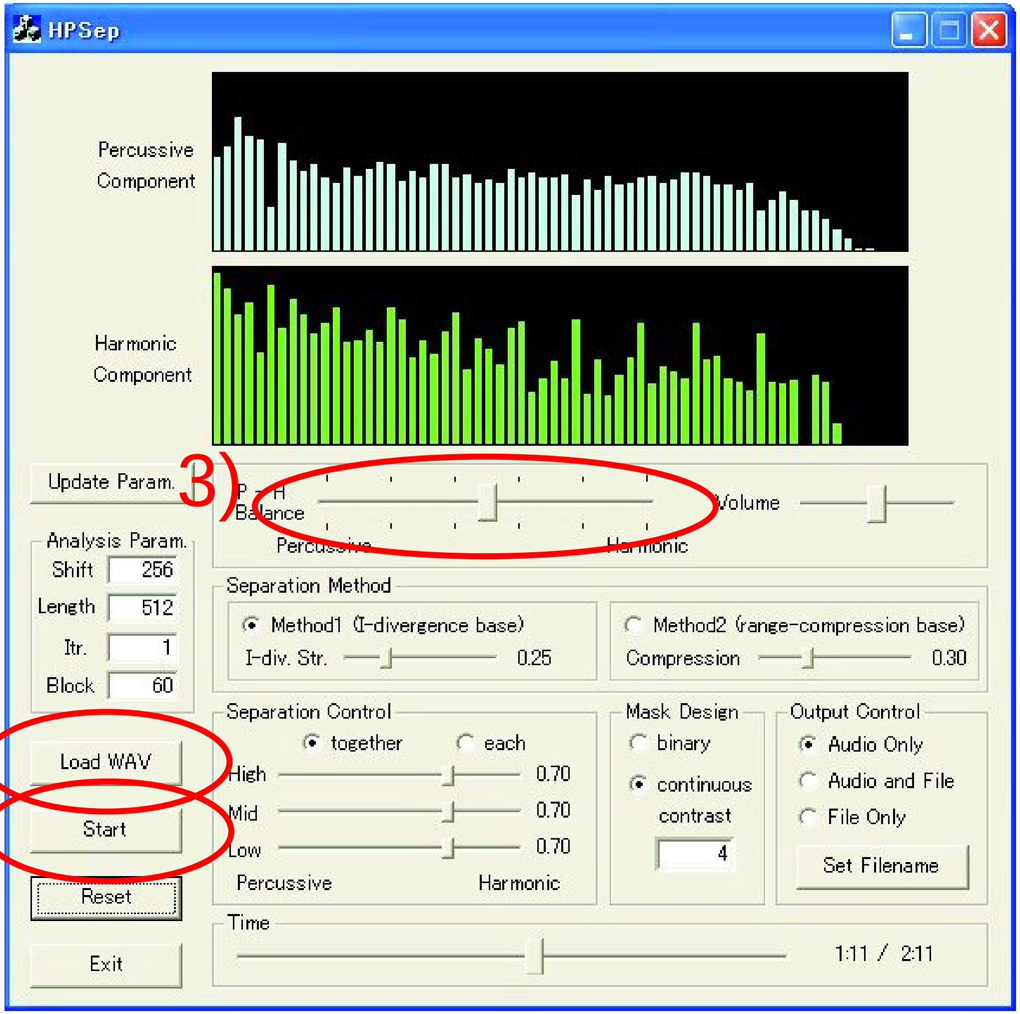


Iterations are applied only within sliding block

Open Software: Real-time H/P equalizer

Available at http://www.hil.t.u-tokyo.ac.jp/software/HPSS/

- Control H/P balance of audio signal in real time
- Simple instructions:
  1) Click “Load WAV” button and choose a WAV-formatted audio file.
  2) Click “Start” button, and then, audio starts.
  3) Slide H/P balance bar as you like and listen how the sound changes.

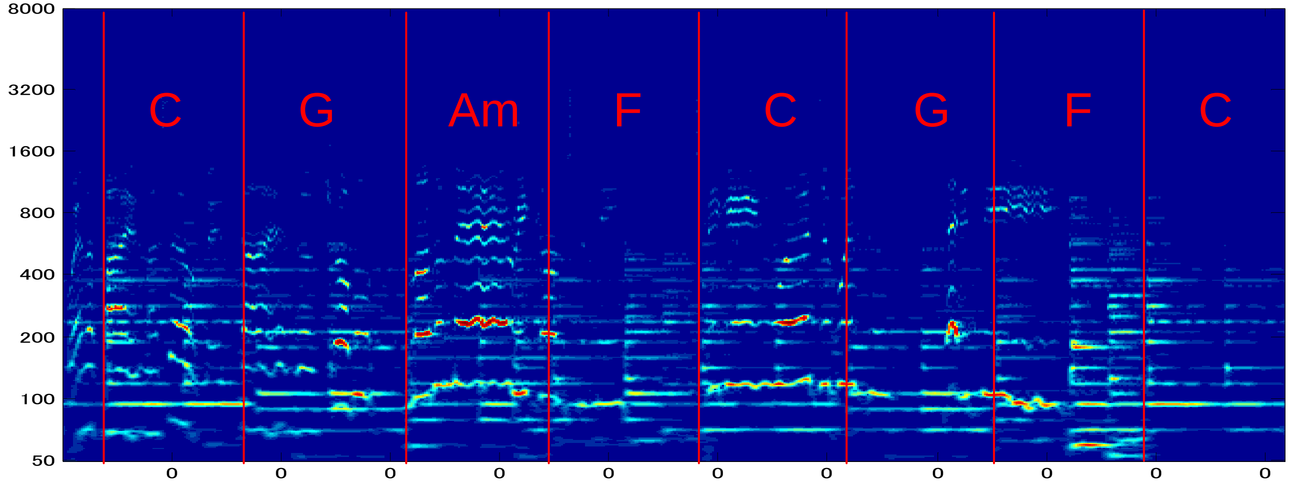


Part III: Applications of HPSS to MIR Tasks

III-1: Audio Chord Detection

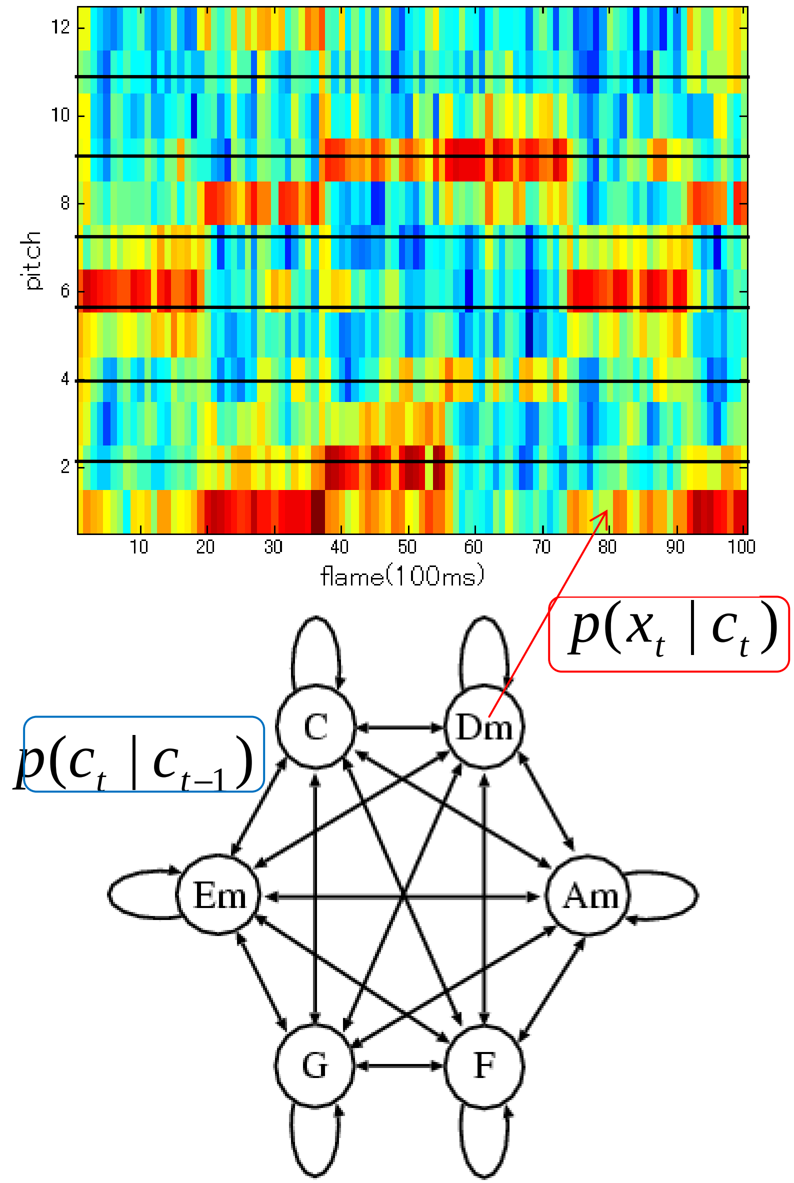
Audio Chord Detection

- Task: Estimate chord sequence and its segmentation from music audio



Typical Approach: Chroma Feature + HMM

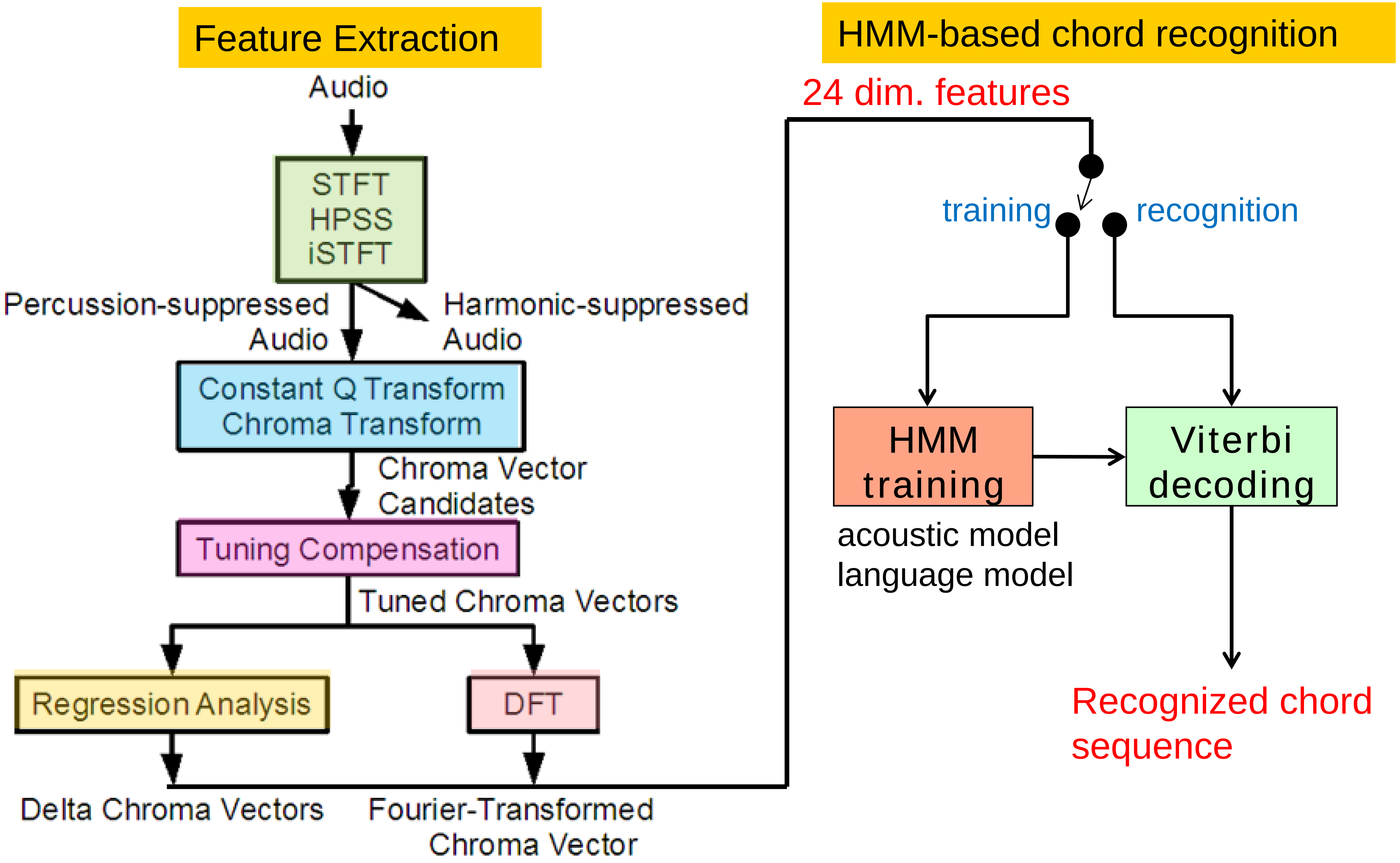
- Feature: chroma [Fujishima1999]
  - Chroma observation probability $p(x_t \mid c_t)$
- Transition: chord progression
  - Bigram probability $p(c_t \mid c_{t-1})$
- Maximum a Posteriori Chord Estimation [Sheh2003]
  - Viterbi algorithm for …



$$\operatorname*{argmax}_{c}\, p(x_0|c_0)\boxed{p(c_0)}\prod_{t=1}^{T}\boxed{p(x_t|c_t)}\boxed{p(c_t|c_{t-1})}$$

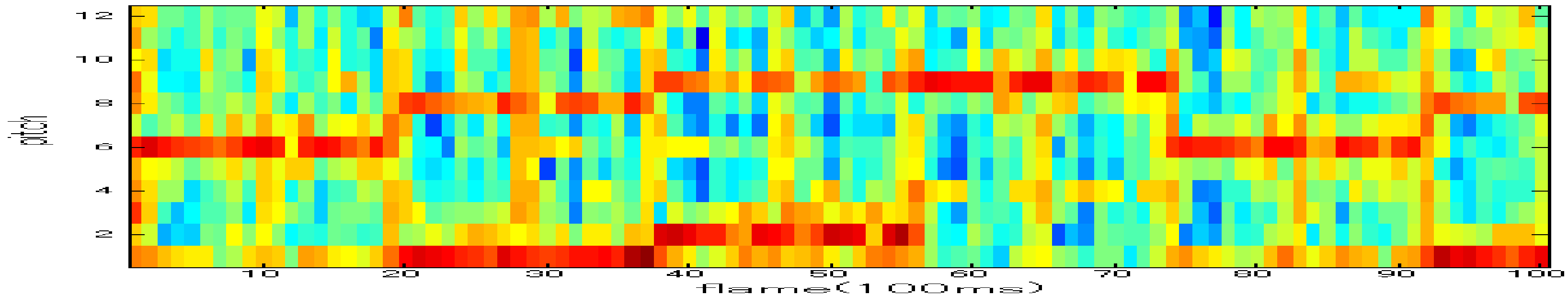
Initial prob. emission transition

Feature-refined System [Ueda2009]

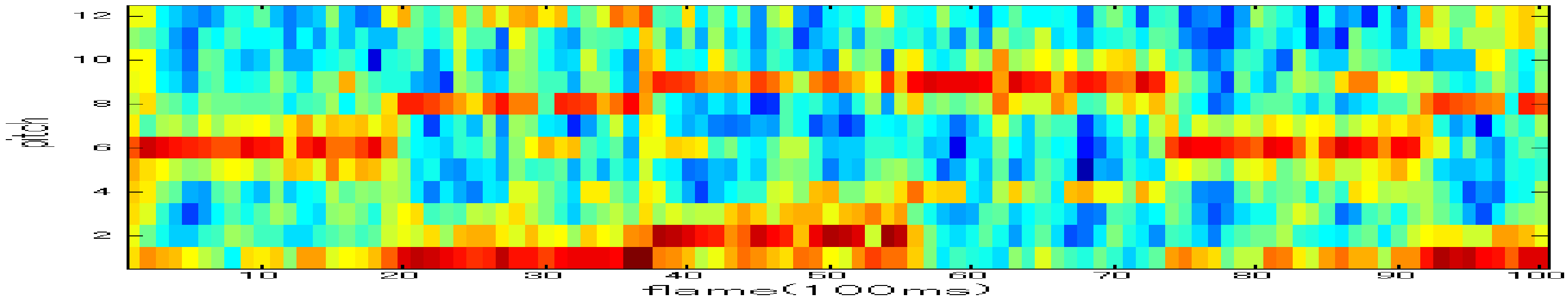


Suppressing Percussive Sounds

- Percussive sounds are harmful in chord detection

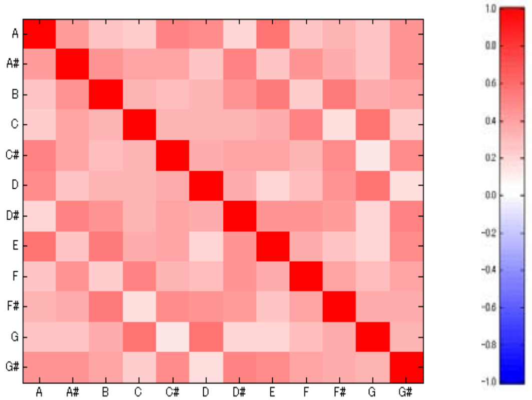


Emphasize harmonic components by HPSS

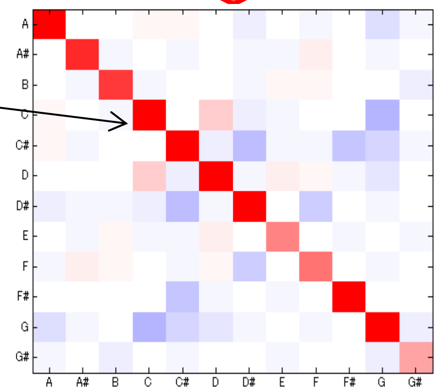


Fourier-transformed Chroma

- Covariance matrix of chroma
  - Highly correlated components: diagonal-only approximation infeasible
    - Caused by harmonic overtones or some pitches performed at the same time
    - Results in large number of parameters
- Covariance matrix is near circulant
  - Assuming …
    - Harmonic overtones of all pitches have the same structure
    - The amount of occurrence of the same intervals is the same
  - Circulant matrix diagonalized by DFT
- Diagonal approximation of FT-Chroma covariance
  - Reduces the number of model parameters (statistically robust)



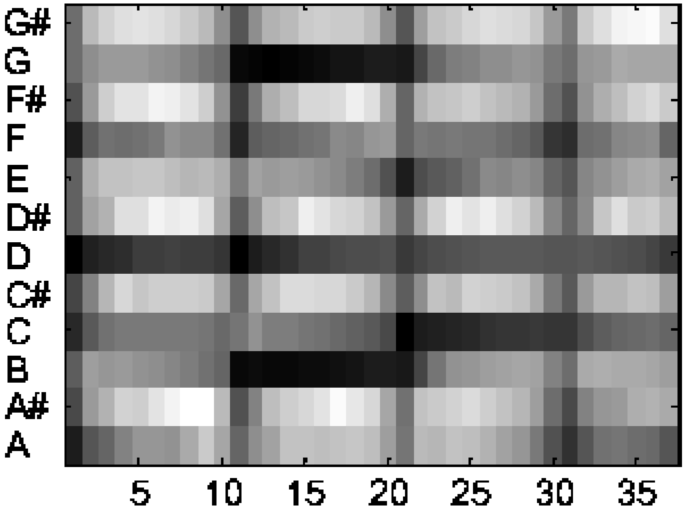
Chroma covariance

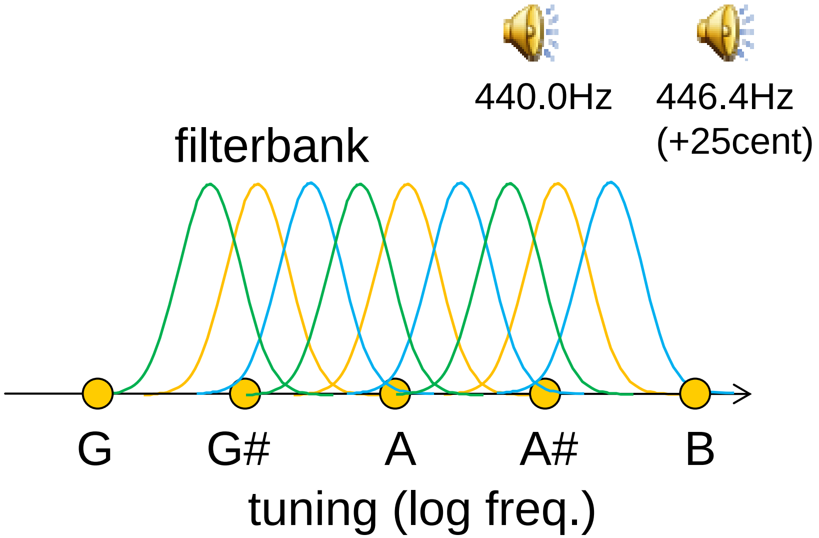


FT-Chroma covariance

Tuning Compensation

- Tuning difference among songs
  - Neglecting this may blur chroma features
- Choose best tuning from multiple candidates
  - Find maximum chroma energy (sum of all bins of chroma)
  - Assume: tuning does not change within a song



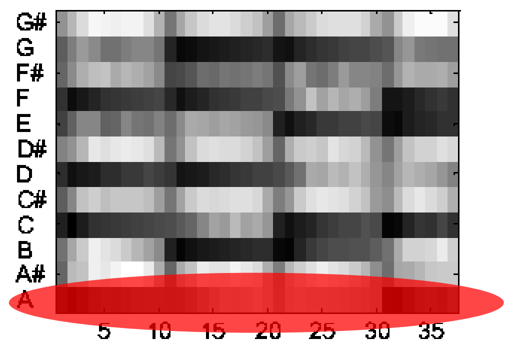


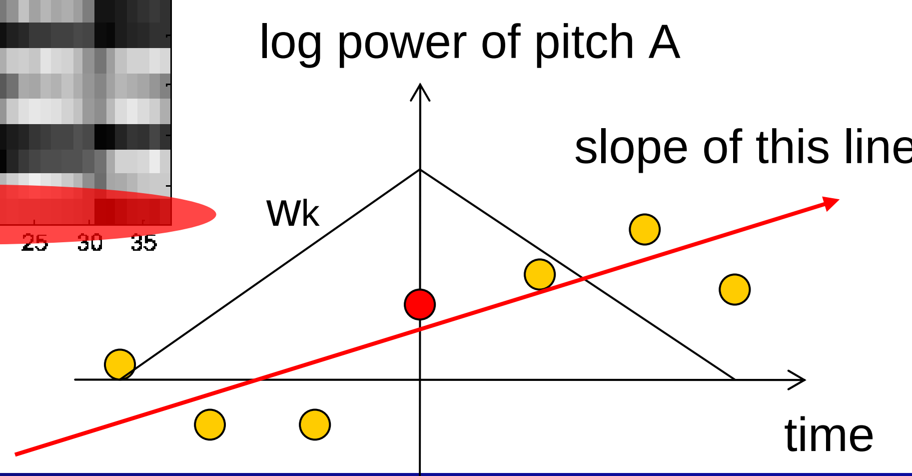
Delta Chroma Features

- Improve chord boundary accuracy
  - by features representing chord boundaries
- Chord tones largely changes at chord boundary
  - Delta chroma：derivative of chroma features
  - Cf. Delta cepstrum (MFCC)：Effective features of speech recognition
- Calculated by regression analysis of δ sample points [Sagayama&Itakura1979]
  - Robust to noise

$$\Delta C(i,n) = \frac{\sum_{k=-\delta}^{\delta} k w_k C(i,t+k)}{\sum_{t=-\delta}^{\delta} k^2 w_k}$$

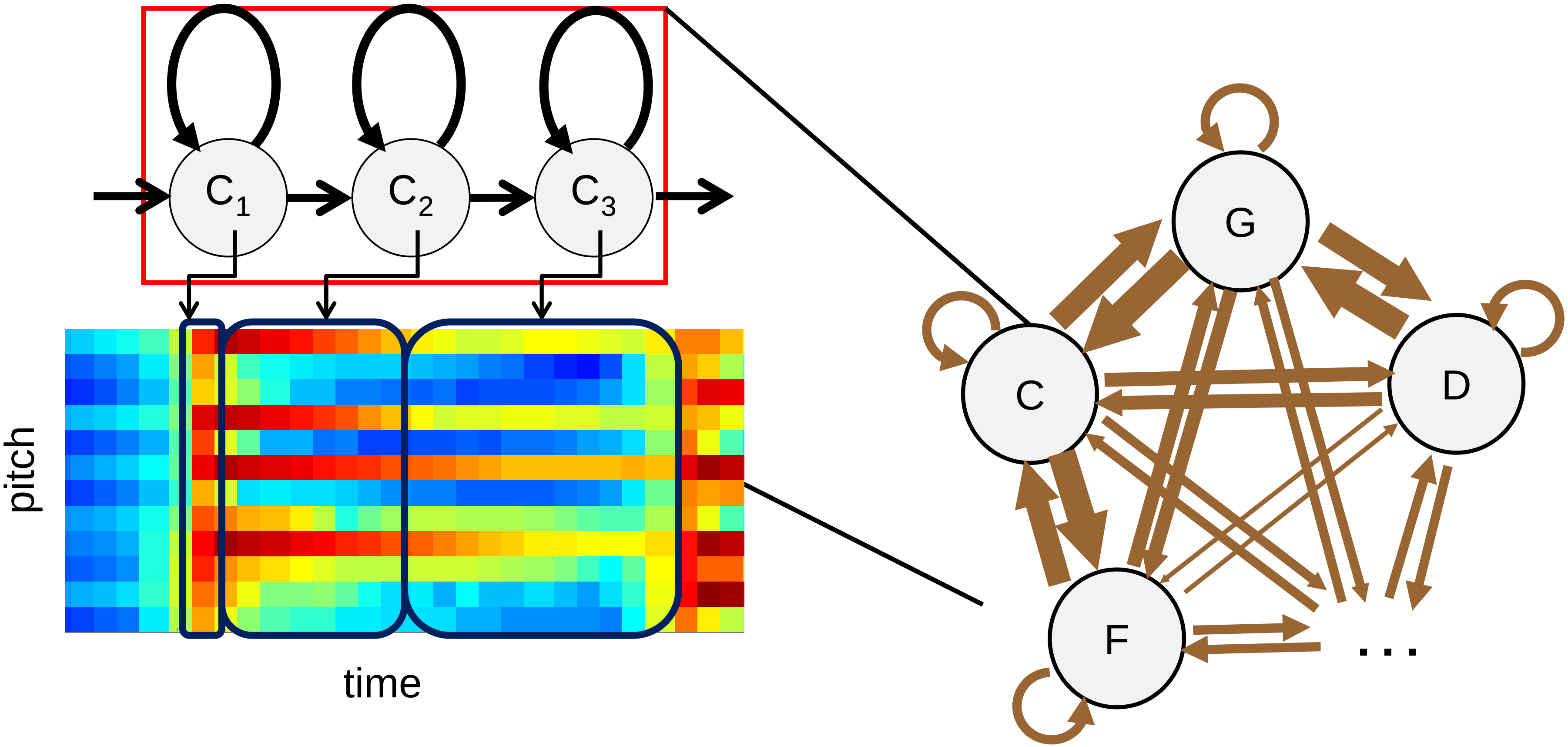
$$i = 1,\cdots,12$$





Multiple States per Chord

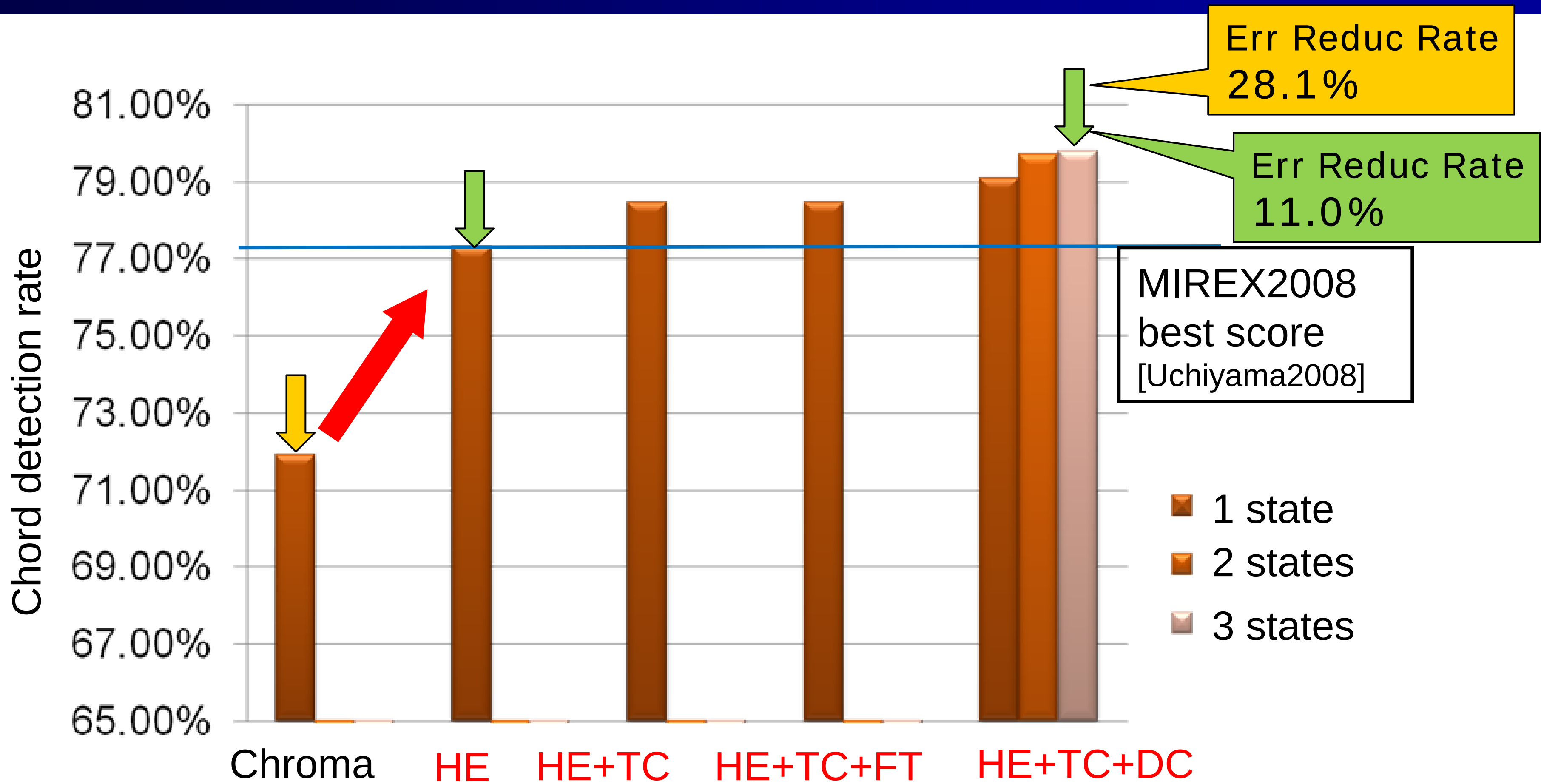
- Chroma changes from “onset” to “release”
  - capture the change by having multiple states per chord
  - tradeoff between data size and the number of states



Experimental Evaluation

- Test Data
  - 180 songs (12 albums) of The Beatles (chord reference annotation provided by C. Harte)
  - 11.025 kHz sampling, 16bit, 1ch, WAV file
  - Frequency range: 55.0Hz-1661.2Hz (5 octaves)
- Labels
  - 12 × major/minor =24 chords + N (no chord)
- Evaluation
  - Album filtered 3-fold cross validation
    - 8 albums for training, 4 albums for testing
  - Frame Recognition Rate
    = (#correct frames) / (#total frames)
  - Sampled every 100ms

Chord Detection Results



HE：harmonic sound emphasized
TC: tuning compensation
FT: FT chroma (diagonal covariance)
DC: Delta chroma

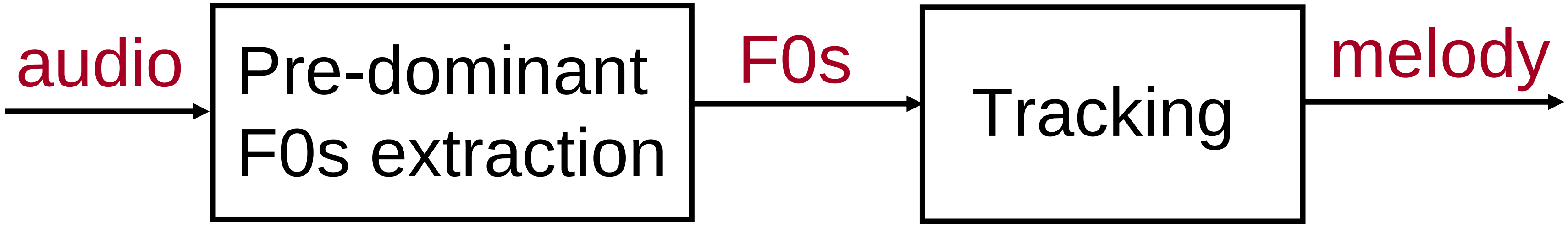
HPSS improves chord detection performance

Part III: Applications of HPSS to MIR Tasks

III-2: Melody Extraction

Melody Extraction

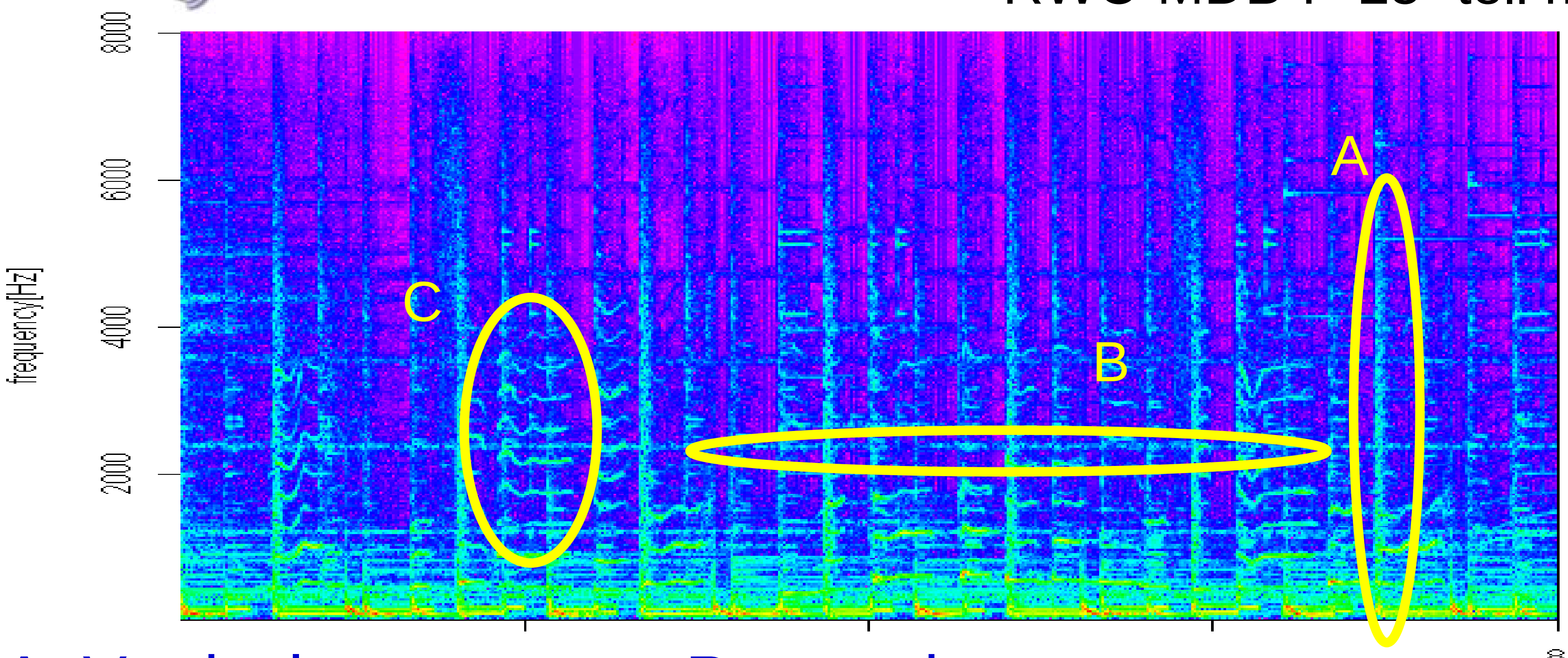
- Task: Identify a melody pitch contour from polyphonic musical audio
- Typical approach:



  - Singing voice enhancement will be useful pre-processing.

Singing Voice in Spectrogram

RWC-MDB-P-25 “tell me”



A. Vertical component: Percussion

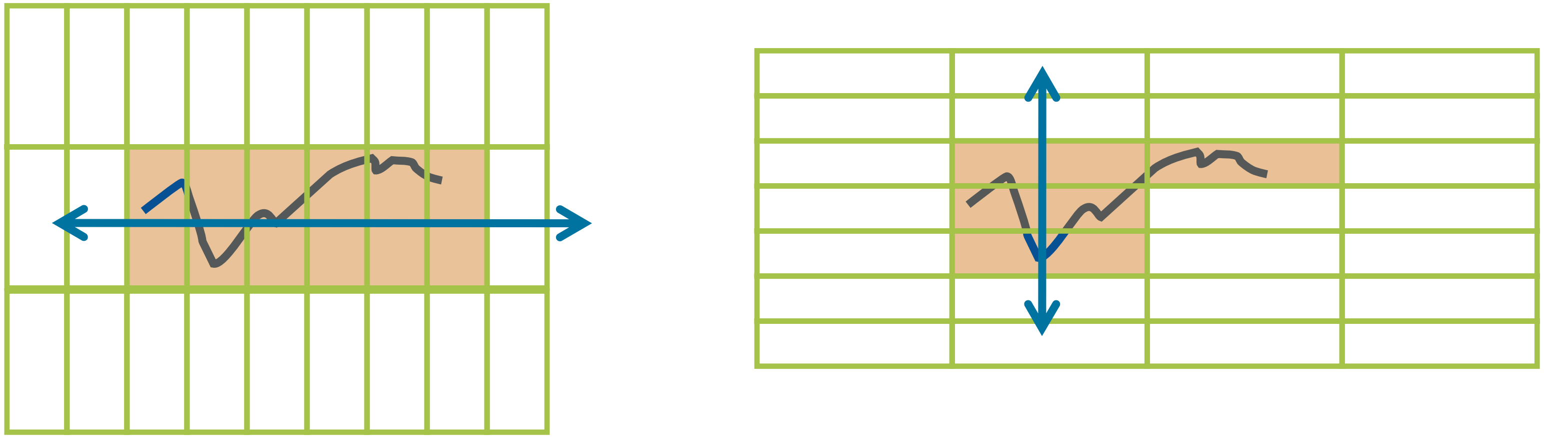
B. Horizontal component: Harmonic instrument (piano, guitar, etc..)

C. Fluctuated component: Singing voice

Is Voice Harmonic or Percussive?

Depends on spectrogram resolution (frame-length)

- On short-frame STFT domain, voice appears as "H" (time direction clustered).
- On long-frame STFT domain, voice appears as "P" (frequency direction clustered).



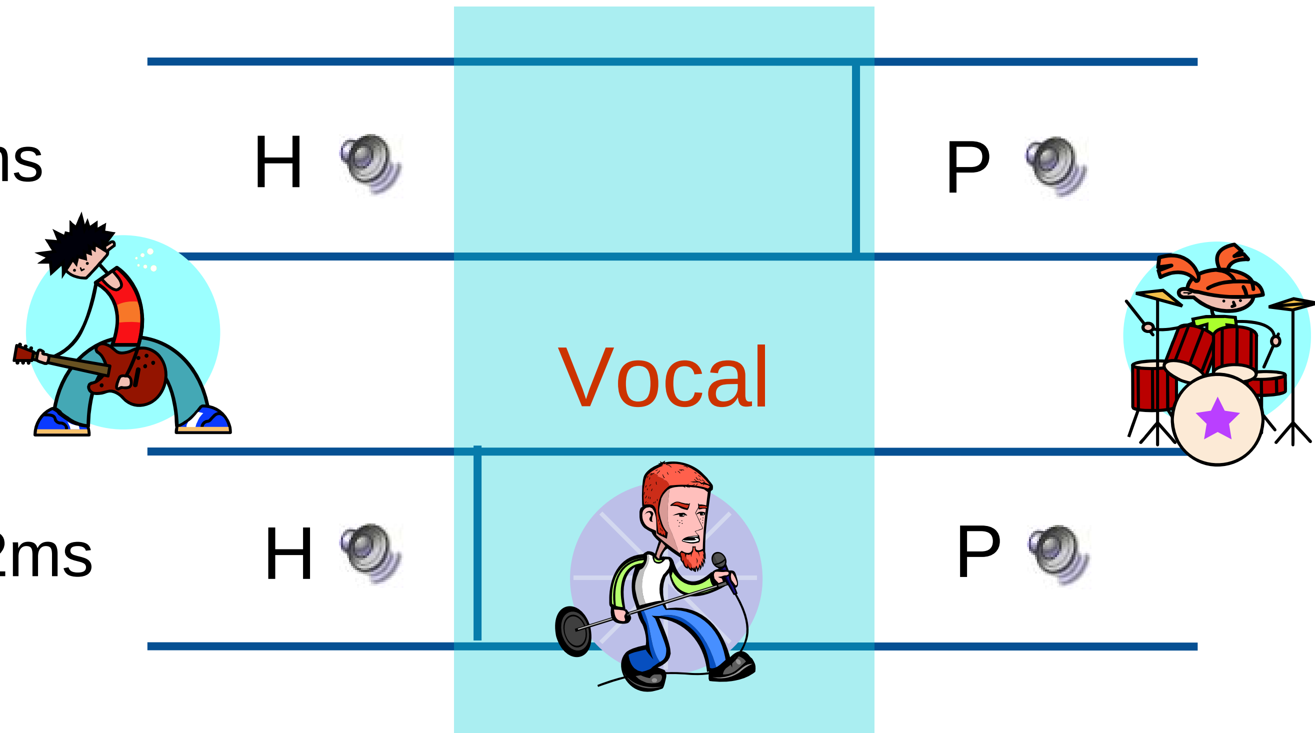
"Harmonic" "Percussive"

HPSS Results with Different Frame Length

Example

Frame length: 16ms

Frame length: 512ms



Two-stage HPSS [Tachibana2010]

Original

HPSS with short frame

Sinusoidal Sound

Percussive Sound

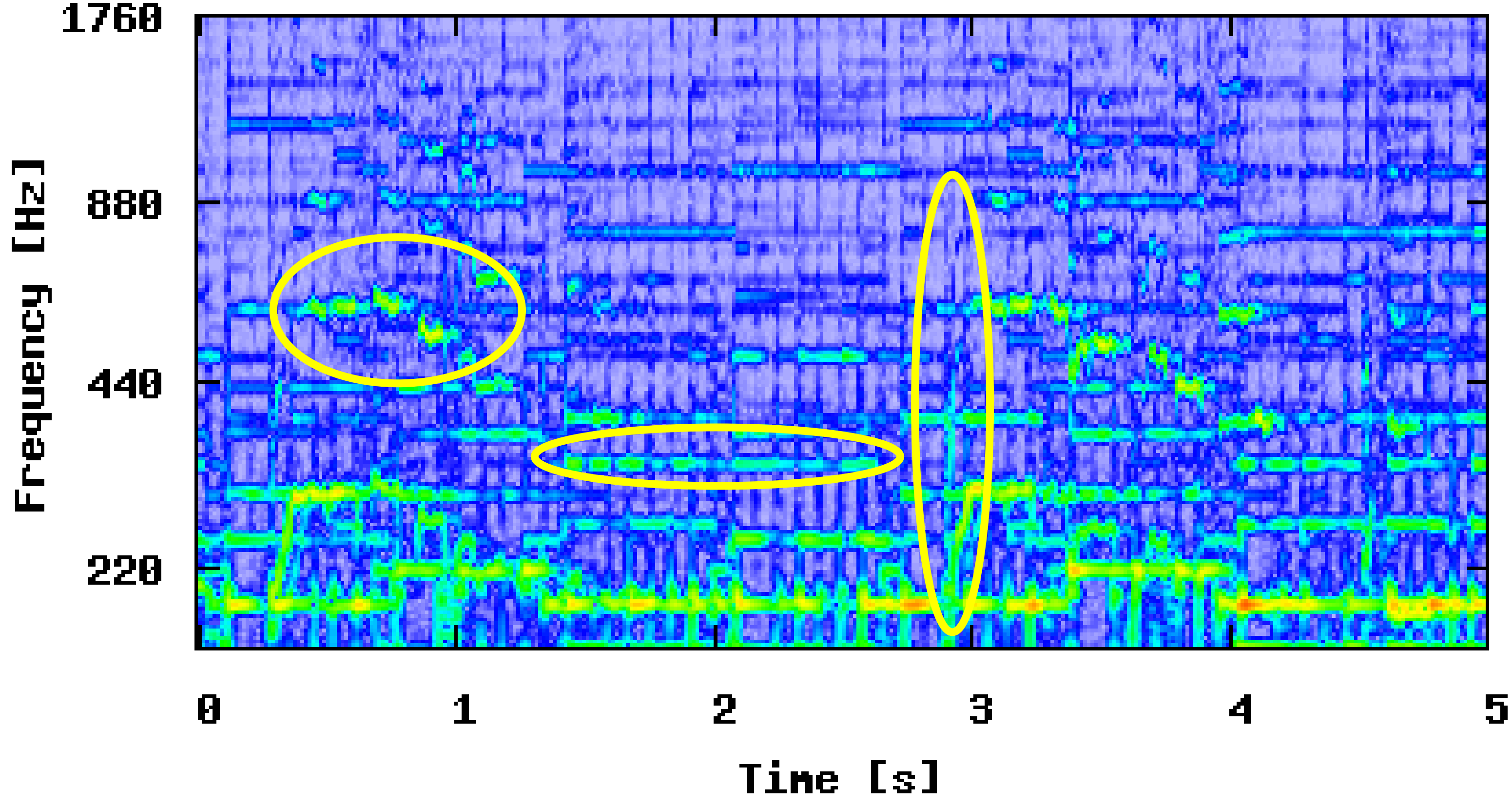
HPSS with long frame

Stationary-sinsoidal Sound

Fluctuating-sinusoidal Sound (≒singing voice)

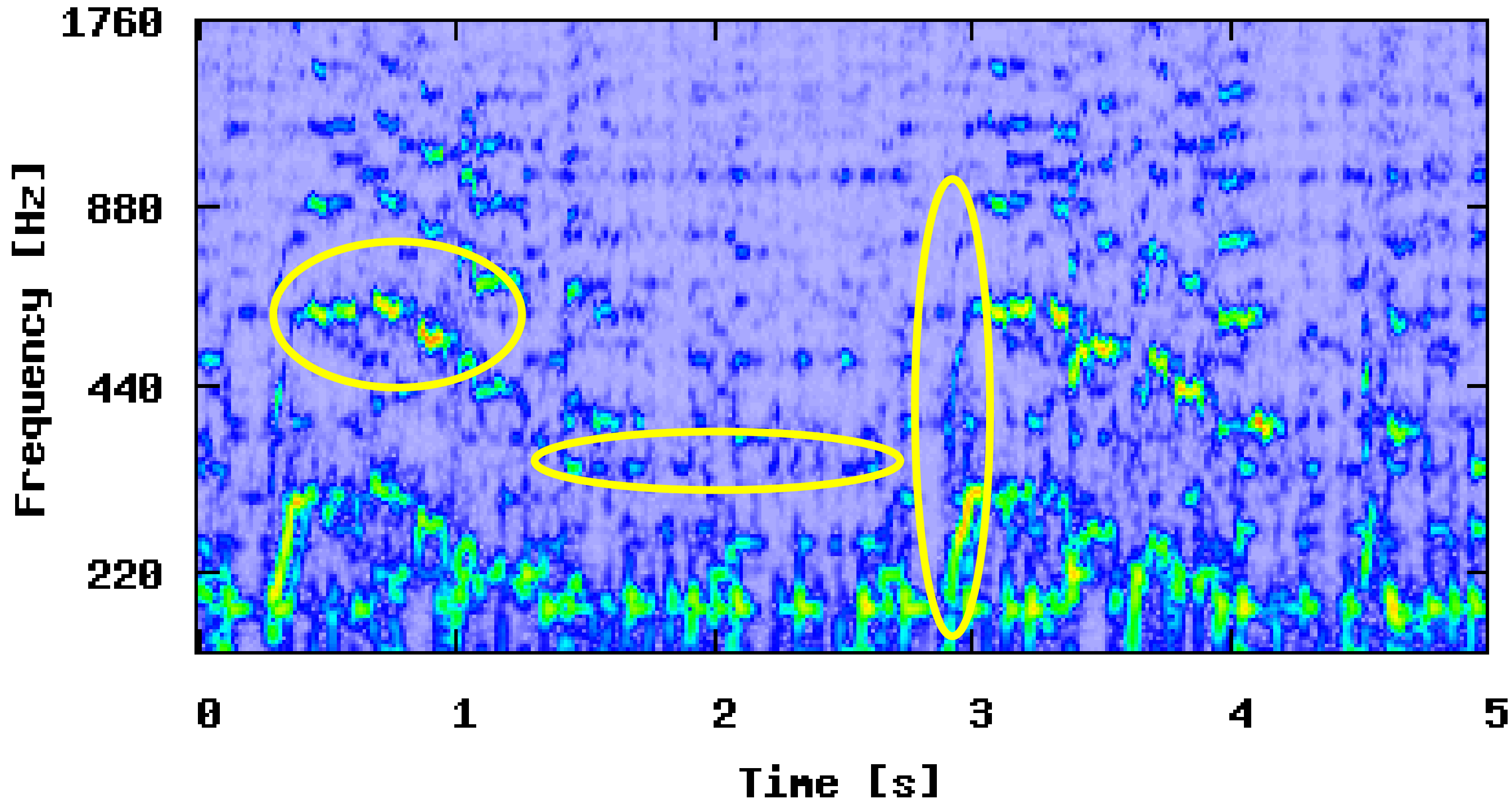
Spectrogram Example

Original signal (from LabROSA dataset)



Spectrogram Example

Voice-enhanced signal (by two-stage HPSS)

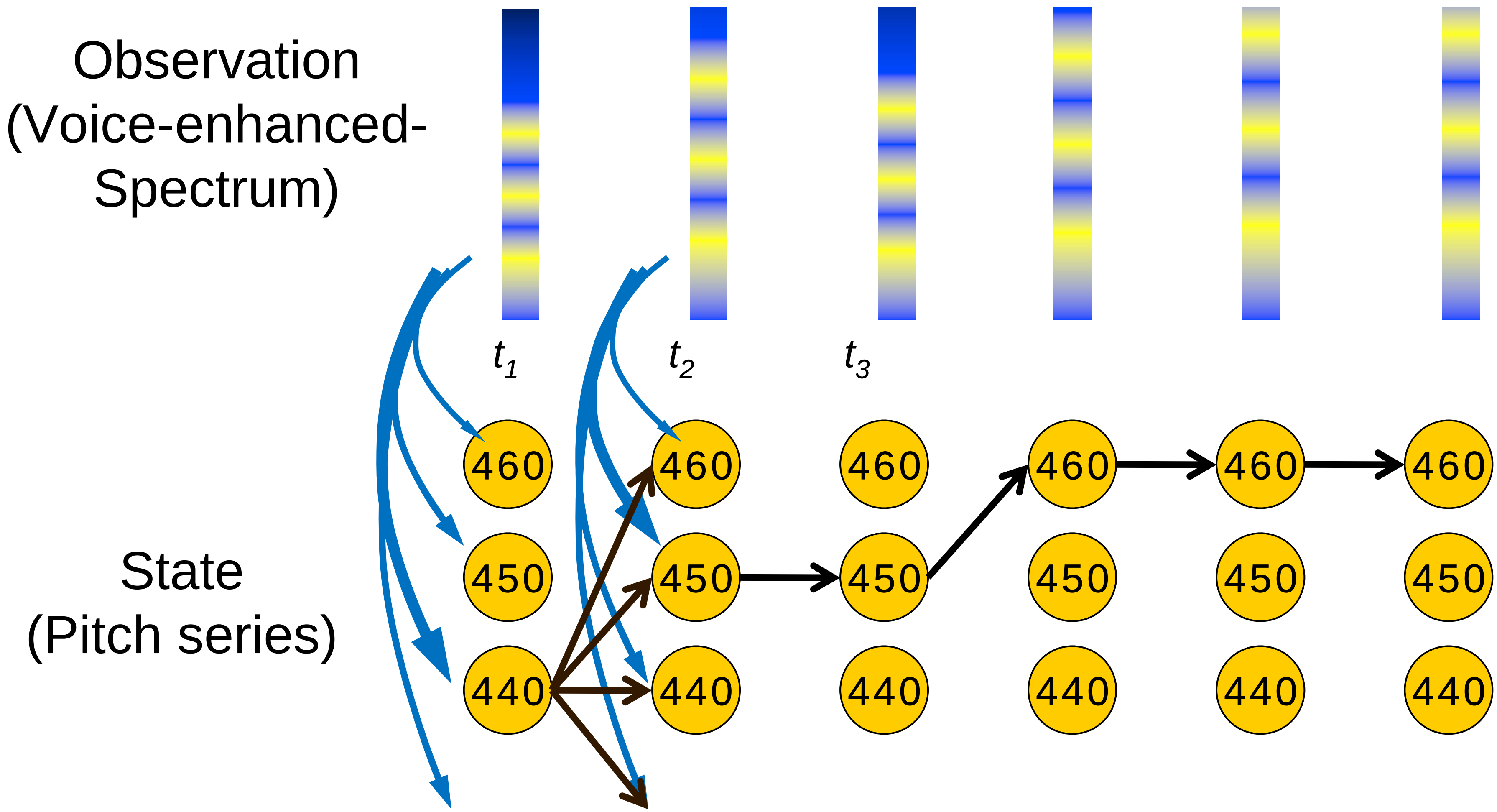


Separation Examples

title	original	Extracted Vocal	Vocal Cancelled*	Genre
“tell me”				F, R&B
“Weekend”				F, Euro beat
“Dance Together”				M, Jazz
“1999”				M, Metal rock
“Seven little crows”				F, Nursery rhyme
“La donna è mobile” from Verdi’s opera “Rigoletto”				M, Classical

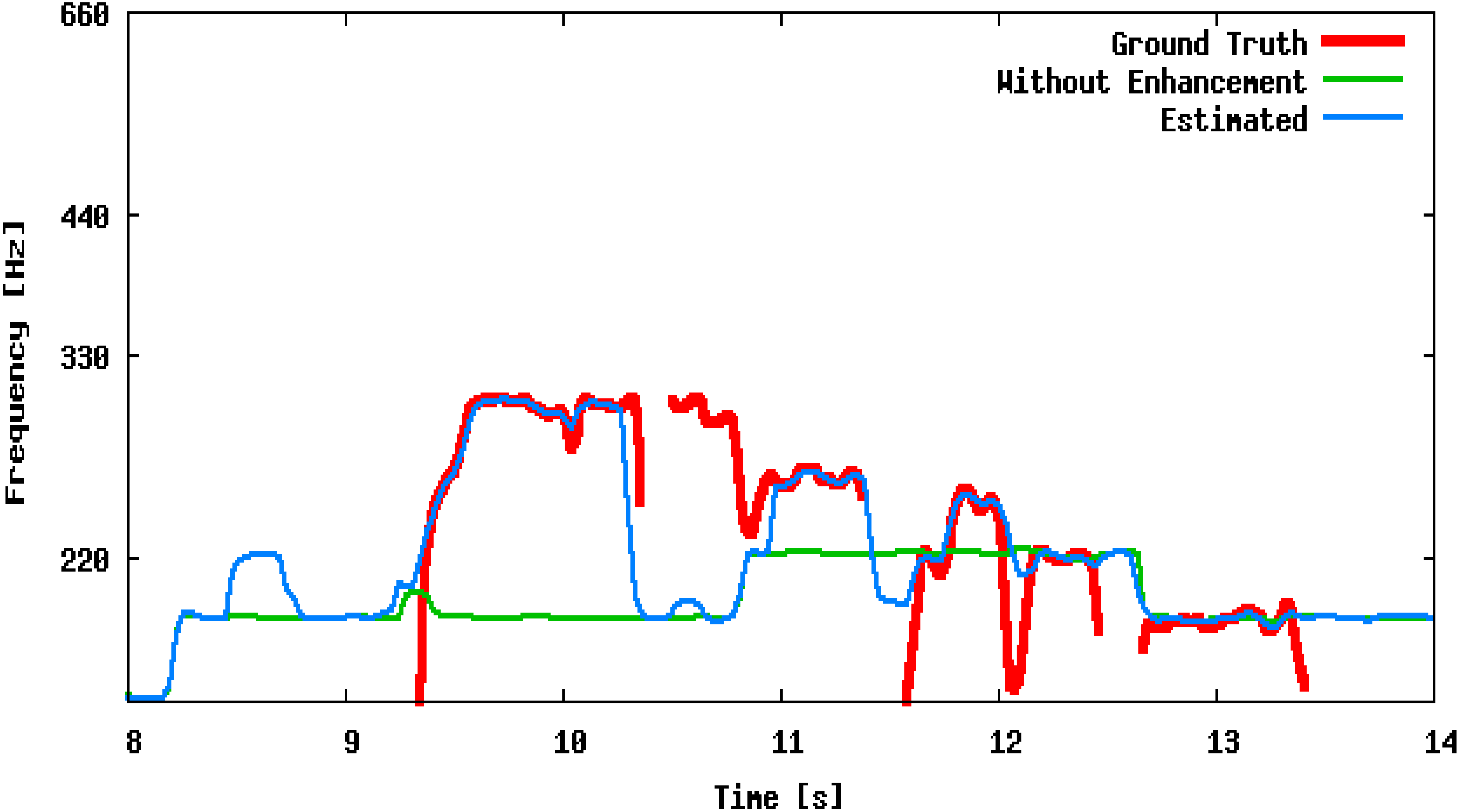
Melody Tracking by DP [Tachibana2010]

- Estimating hidden states by dynamic programming



Example of Melody Tracking

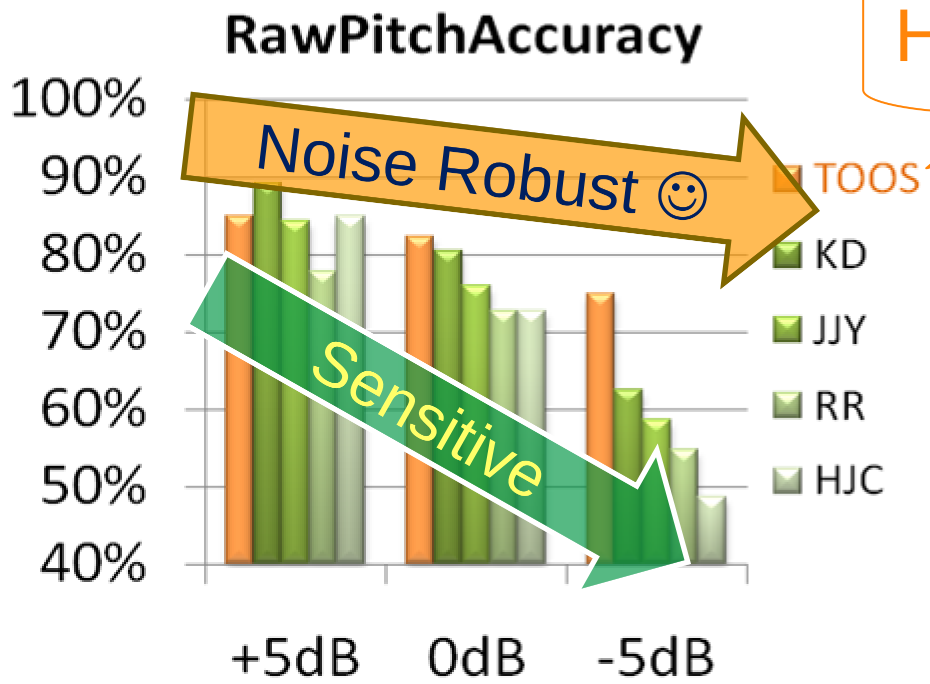
- train06.wav, distributed by LabROSA database



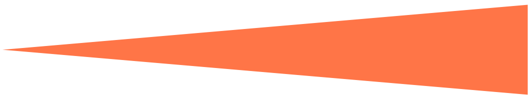
Results in MIREX 2009

- Data: 379 songs, mixed in +5 dB, 0dB, and -5 dB.

HPSS-based method



	+5dB	0dB	-5dB
original			
processed			



Accompaniments

Robustness to large singer-to-accompaniment ratio is greatly improved.

Part III: Applications of HPSS to MIR Tasks

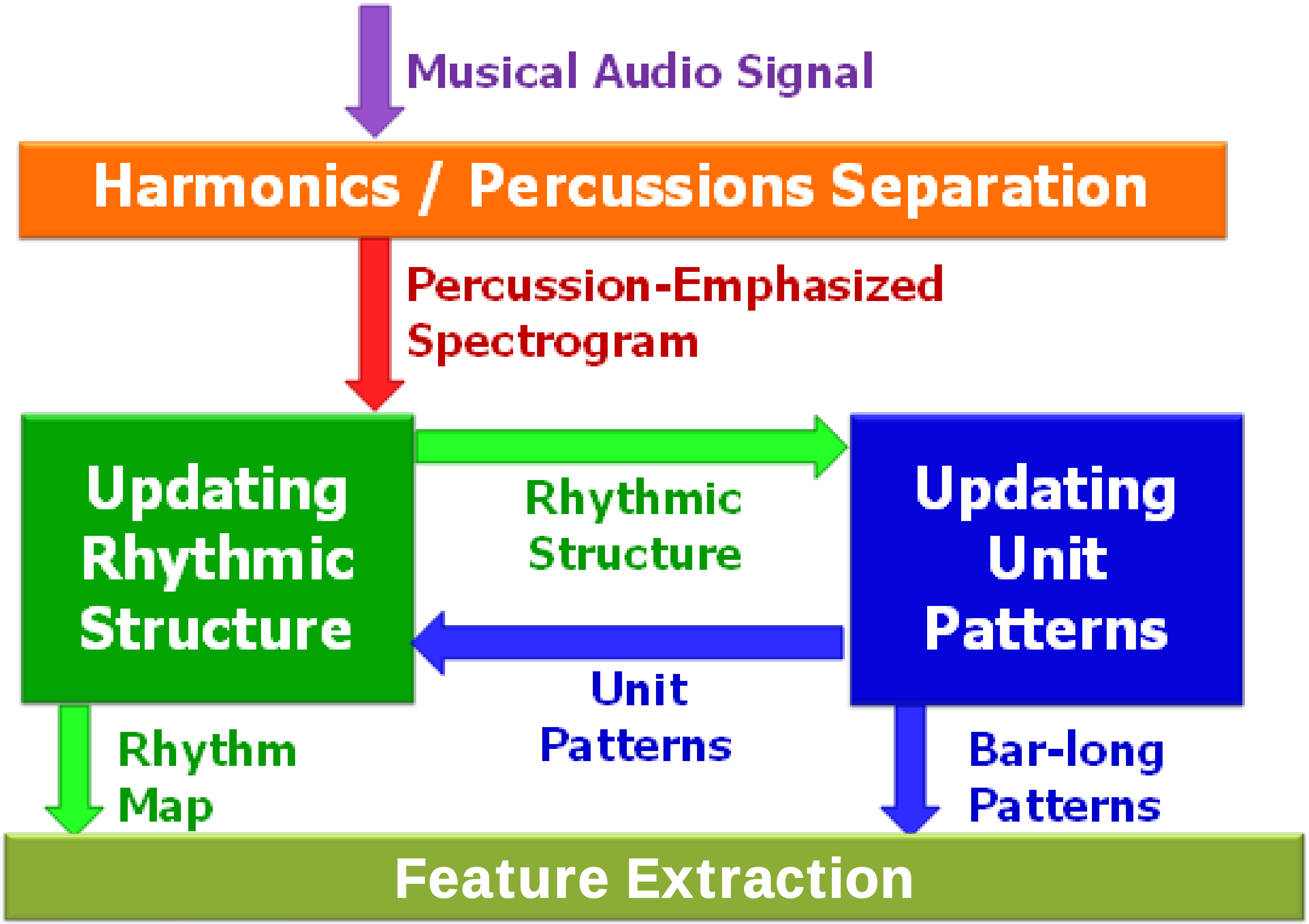
III-3: Audio Genre Classification

Audio Genre Classification

- Task: estimate genre from music audio
  - Blues, classical, jazz, rock, ...
- Typical approach

- Example of features [Tzanetakis2001]
  - Timbral information (MFCC, etc.)
  - Melodic information
  - Statistics about periodicities: Beat histogram

New Features I: Percussive Patterns



[Tsunoo2009]

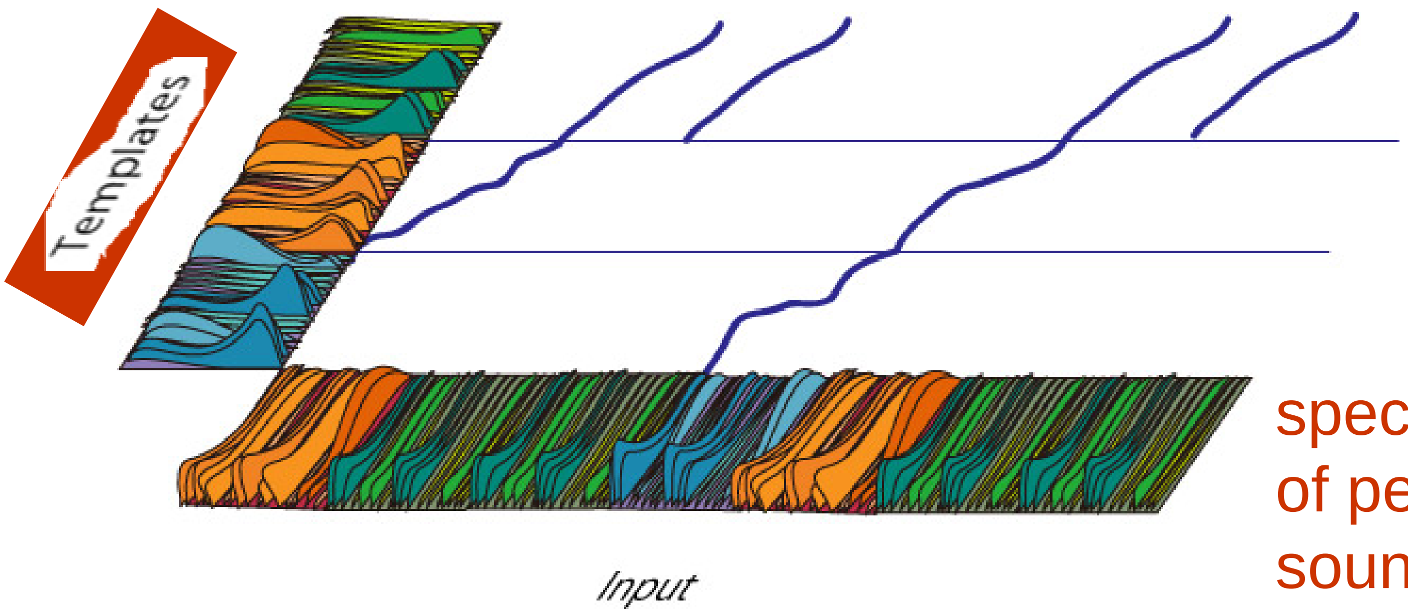
Motivation for Bar-long Percussive Patterns

- Bar-long percussive patterns (temporal information) are frequently characteristic of a particular genre
- Difficulties
  1) Mixture of harmonic and percussive components
  2) Unknown bar-lines
  3) Tempo fluctuation
  4) Unknown multiple patterns

A A A A B A A A C C C C

Rhythmic Structure Analysis by One-pass DP algorithm

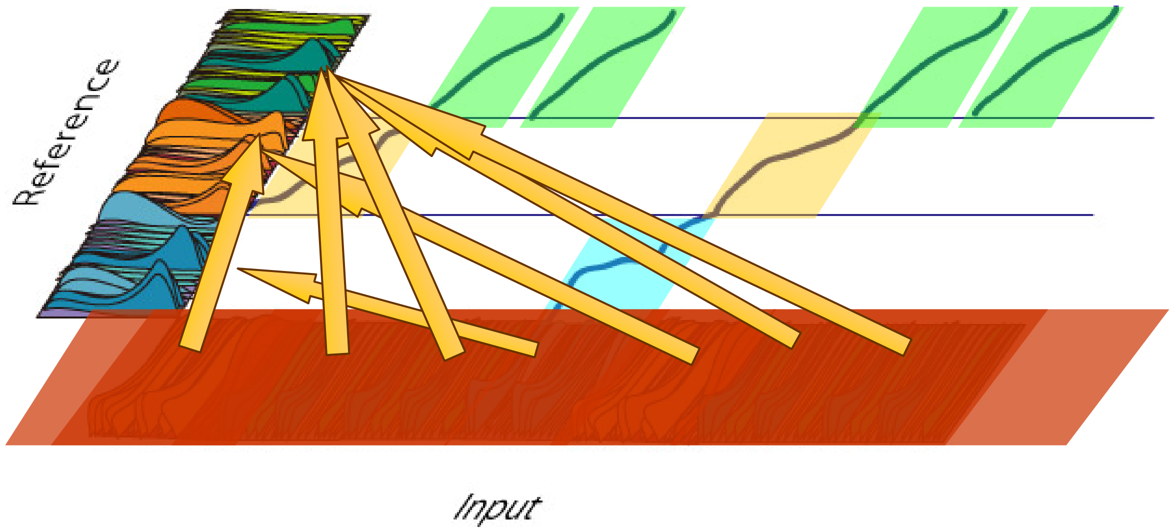
- Assume that correct bar-line unit patterns are given.
- Problem: tempo fluctuation and unknown segmentation
  - Analogous to continuous speech recognition problem
  - One-pass dynamic programming algorithm can be used to segment



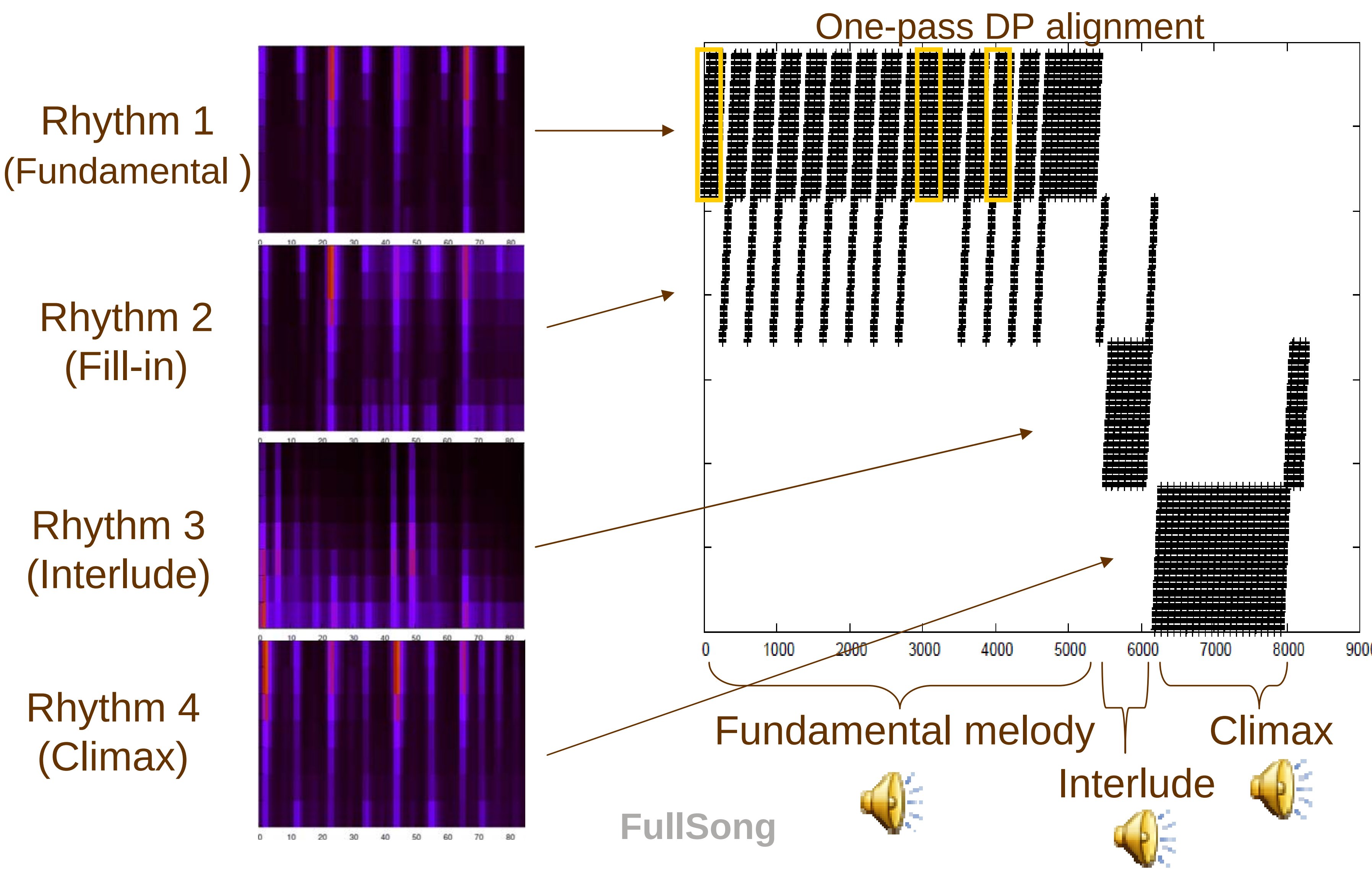
spectrogram of percussive sound

Dynamic Pattern Clustering [Tsunoo2009]

- Actually, unit patterns also should be estimated.
  - Chicken-and-egg problem
  - Analogous to unsupervised learning problem
- Iterative algorithm based on k-means clustering
  - Segment spectrogram using one-pass DP algorithm
  - Update unit patterns by averaging segments
- Convergence is guaranteed mathematically

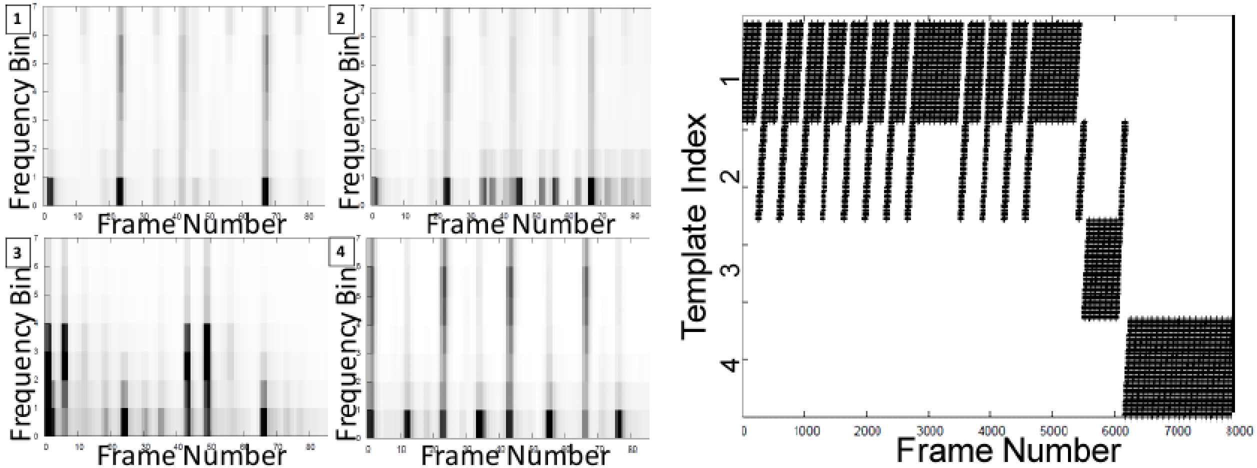


Example of "Rhythm Map"

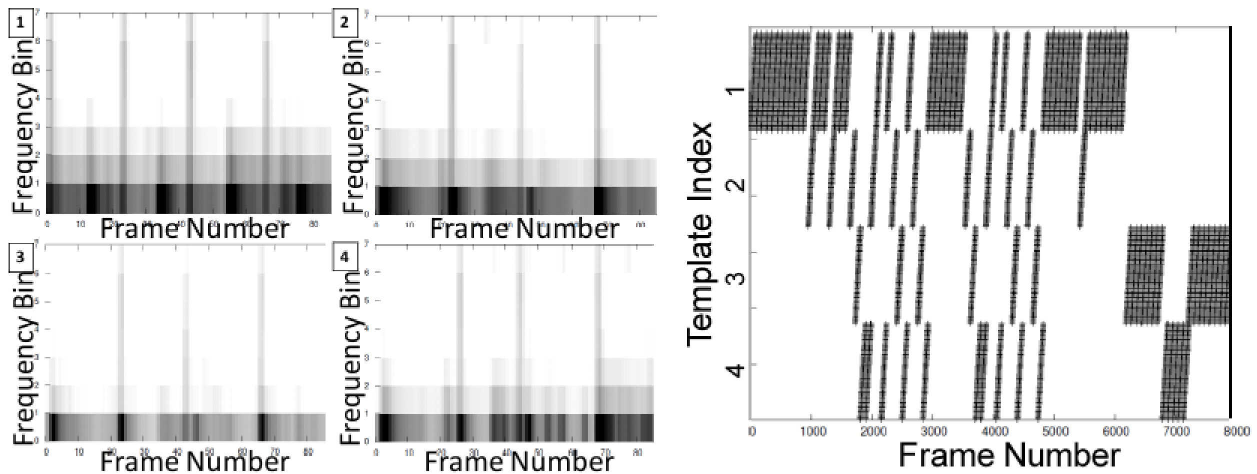


Necessity of HPSS in Rhythm Map

With HPSS



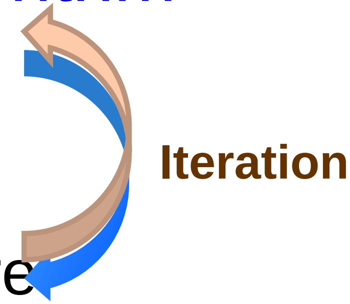
Without HPSS

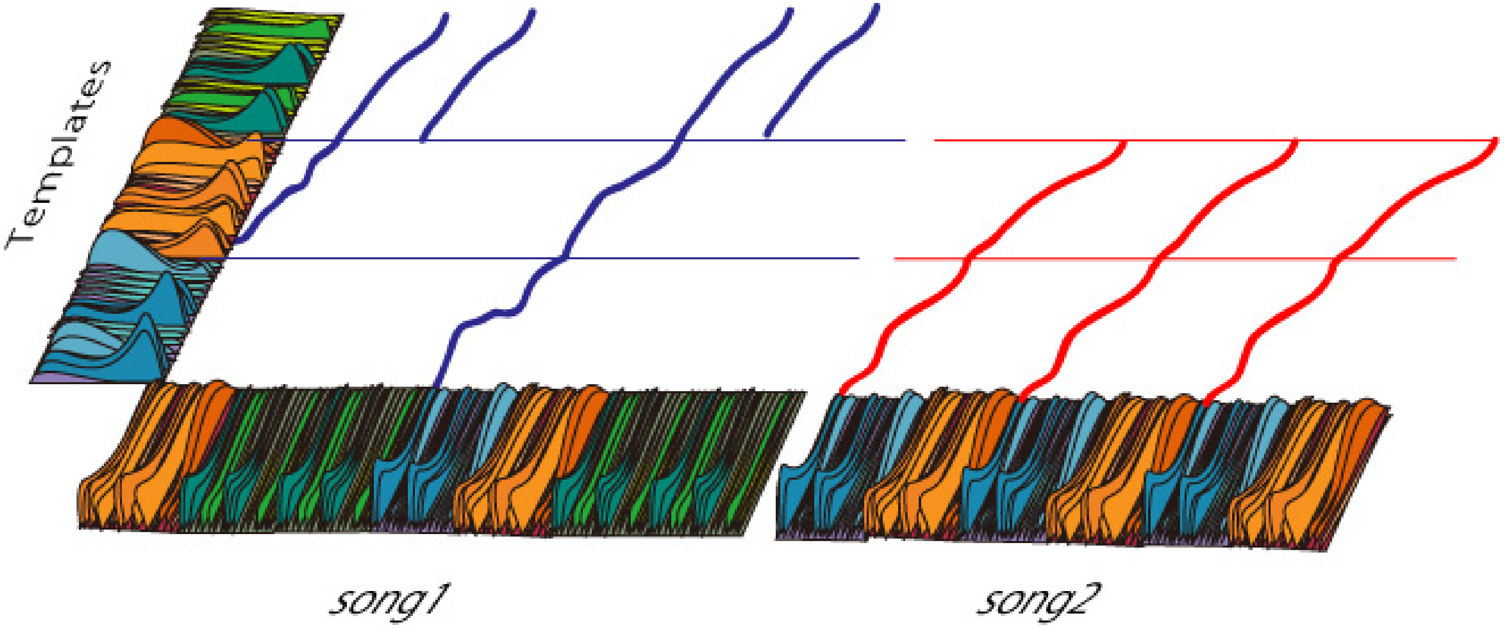


Rhythm patterns and structures are not extracted without HPSS!

Extracting Common Patterns to a Particular Genre

- Apply to a collection of music pieces
- Alignment calculation by one-pass DP algorithm
  - Use same set of templates
- Updating templates by *k*-means clustering
  - Use whole music collection of a particular genre

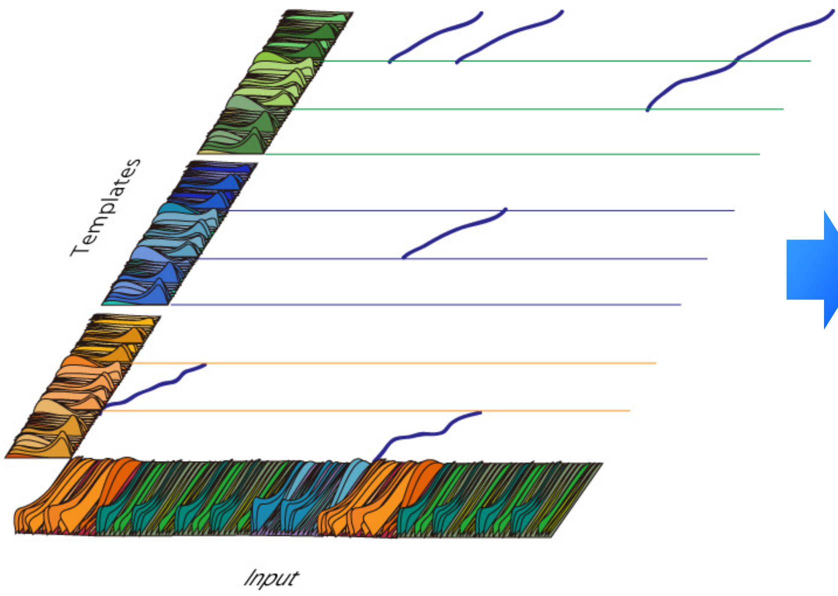




60

Features and Classifiers

- Feature Vectors:
  Genre-pattern Occurrence Histogram (normalized)
- Classifier: Support Vector Machine (SVM)



4	**4/7**
1	**1/7**
2	**2/7**
Histogram	**Normalize**

61

Experimental Evaluation

- Dataset

(standard)

  - GTZAN dataset
  - 22050Hz sampling, 1ch
  - 30 seconds clips
  - 10 genres
    - {blues, classical, country, disco, hiphop, jazz, metal, pop, reggae, rock}
  - 100 songs per genre: total 1000 songs

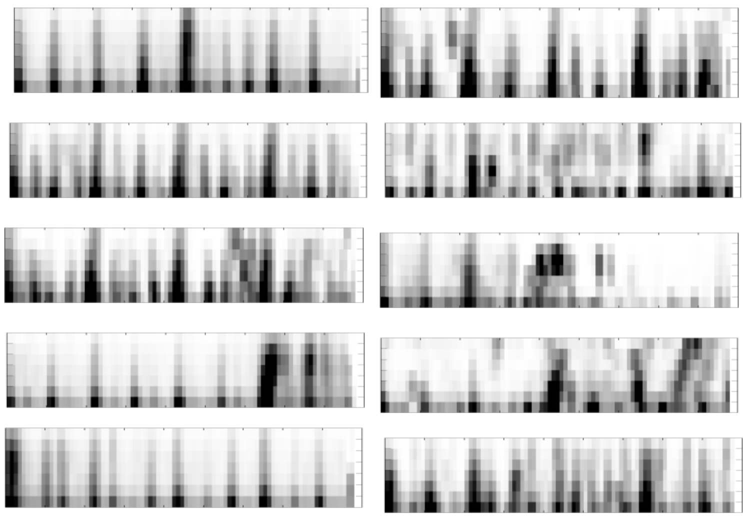
(rhythm-intensive)

  - Ballroom dataset
  - 22050Hz sampling, 1ch
  - 30 seconds clips
  - 8 styles
    - {chacha, foxtrot, quickstep, rumba, samba, tango, viennesewaltz, waltz}
  - 100 songs per style: total 800 songs

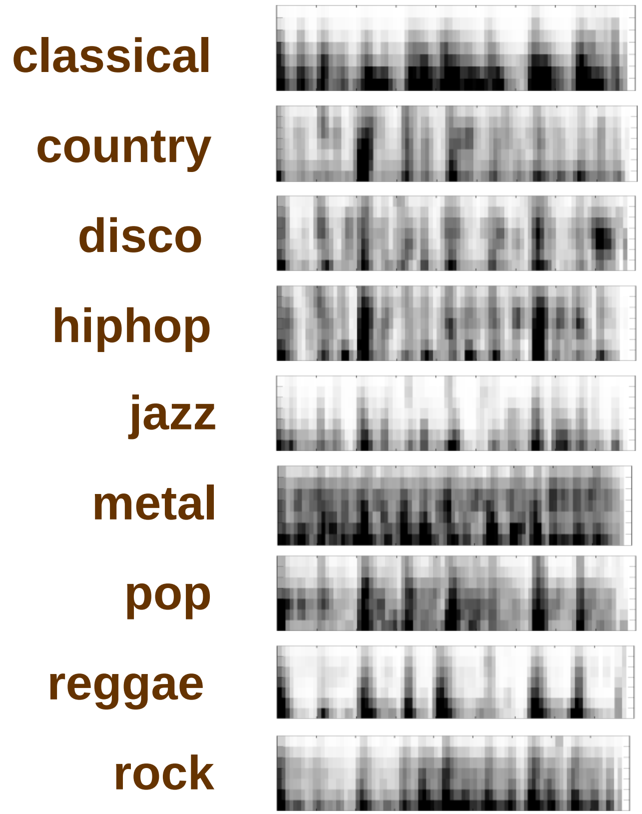
- Evaluation
  - 10-fold cross-validation
  - Classifier: linear SVM (toolkit “Weka” used)

Extracted Percussive Patterns

- Pattern set
  - Divided the datasets into 2 parts and obtained 2 sets of 10 templates for each genre
- Example of learned templates



10 templates learned from “blues”



Genre Classification Accuracy

- Percussive pattern feature only

Features [number of dim.]	GTZAN dataset	Ballroom dataset
Baseline (Random)	10.0%	12.5%
Rhythmic (from template set #1) [10/8]	43.6%	54.0%
Rhythmic (from template set #2) [10/8]	42.3%	55.125%

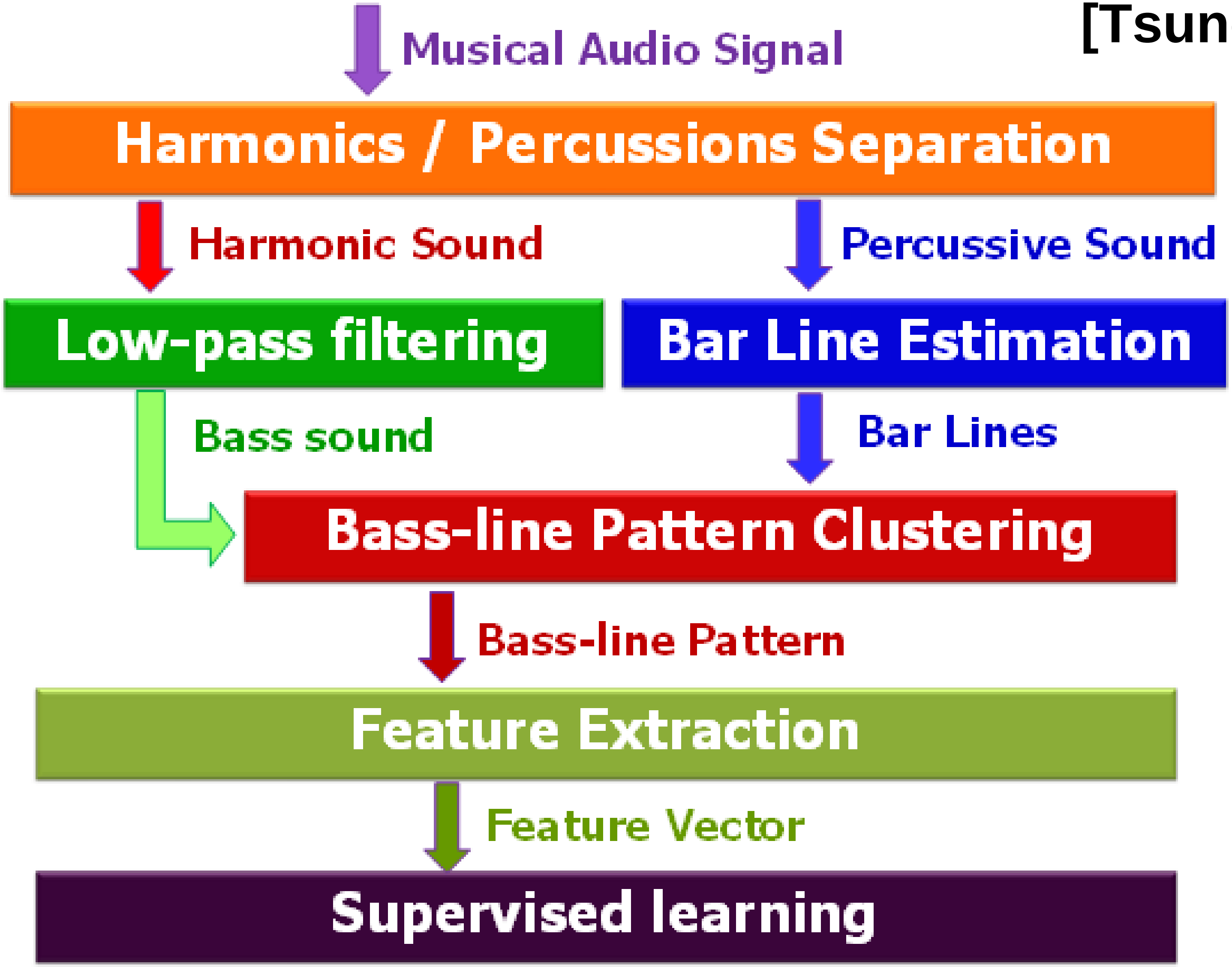
- Merged with timbral features
  - Statistic features such as MFCC, etc. (68 dim.) [Tzanetakis 2008]
  - Performed well on audio classification tasks in MIREX 2008

Features [number of dim.]	GTZAN dataset	Ballroom dataset
Existing (Timbre) [68]	72.4%	57.625%
Merged (from template set #1) [78/76]	76.1%	70.125%
Merged (from template set #2) [78/76]	76.2%	69.125%

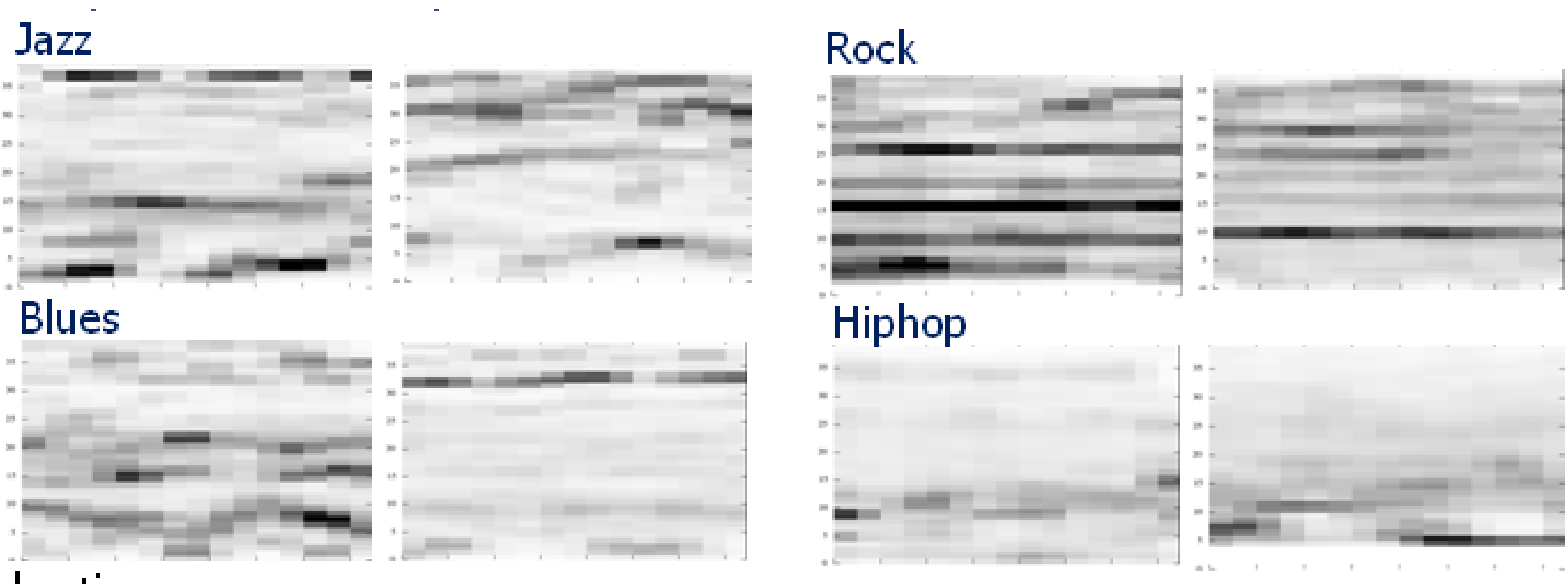
Classification accuracy is improved
by combining percussive pattern features.

New Features II: Bass-line Patterns

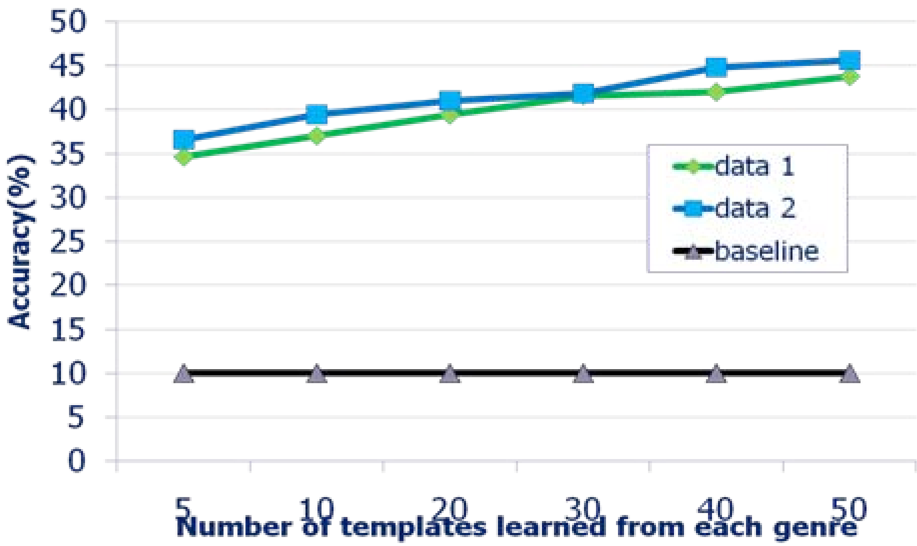
[Tsunoo2009]



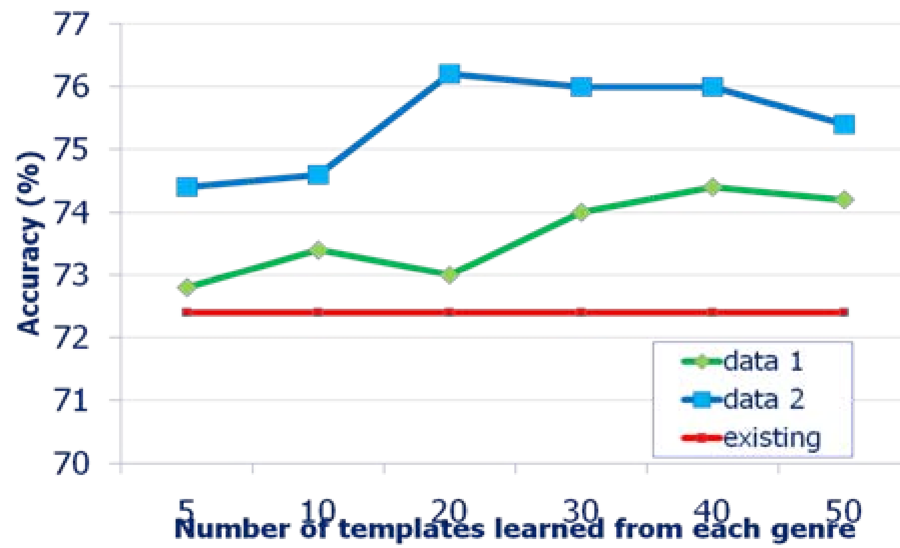
Examples of Extracted Bass-line Patterns



Genre Classification Accuracy



Classification accuracy with only bass-line features

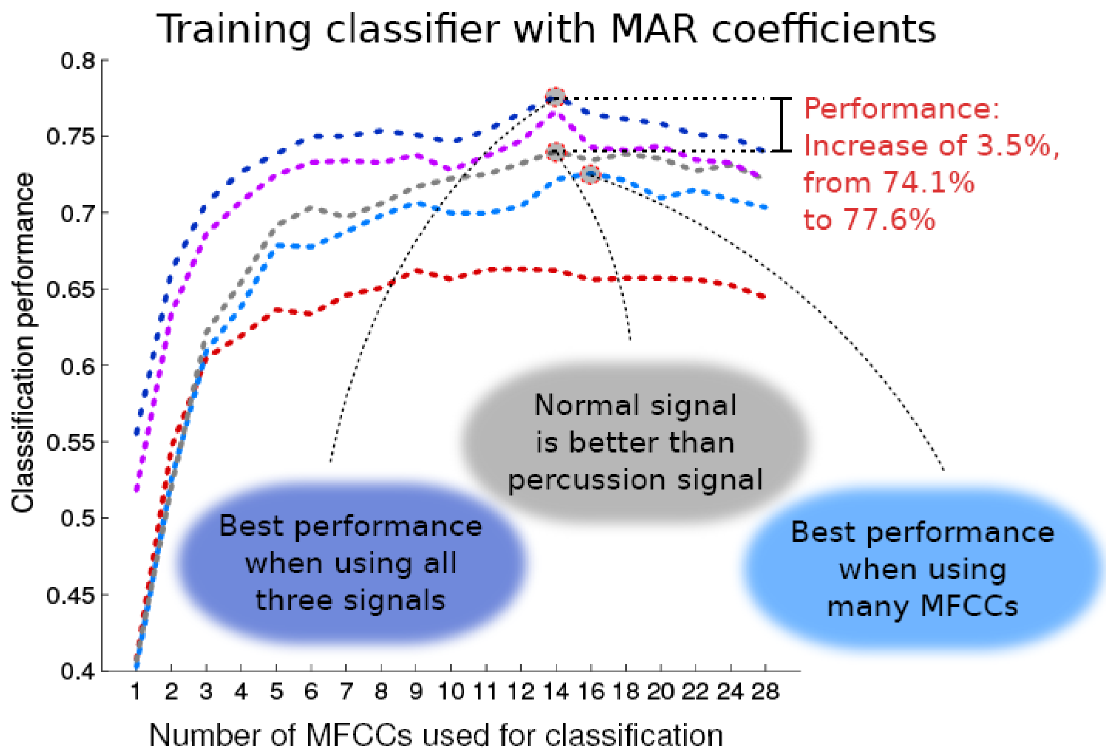


Classification accuracy merged with timbre features

Features	GTZAN dataset	Ballroom dataset
Baseline (random classifier)	10.0%	10.0%
Only bass-line (400 dim.)	42.0%	44.8%
Existing (timbre, 68 dim.)	72.4%	72.4%
Merged (468 dim.)	74.4%	76.0%

Another Application of HPSS [Rump2010]

- Autoregressive MFCC Model applied to Genre Classification
- HPSS increases the number of channels mono -> three (original, harmonic, percussive) and improves performance



Conclusions

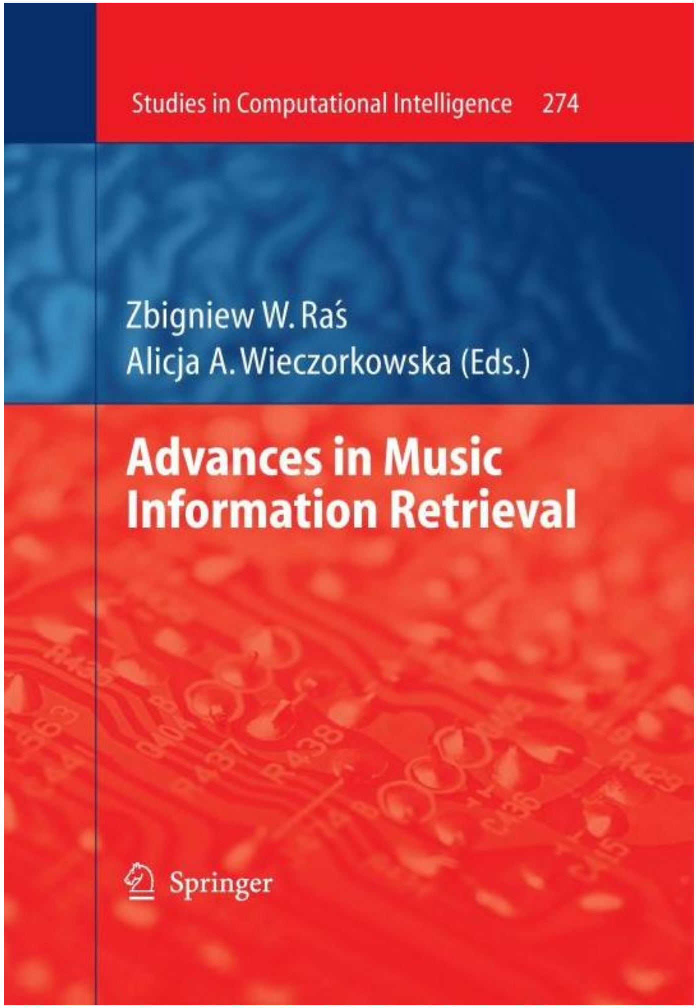
- Source separation techniques used to MIR
  - F0-based harmonic separation
  - Non-negative matrix factorization or PLCA
  - Sinusoid + Noise model
  - Harmonic/percussive sound separation
- Source separation is useful
  - To enhance specific components
  - To increase the number of channels and the dimension of feature vectors
  - To generate new features

Future Works

- Application of source separation to other MIR tasks
  - Cover song identification, audio music similarity,...
- Improvement of separation performance itself by exploiting musicological knowledge
- Using spatial (especially stereo) information
  - Current works are limited to monaural separation
- Feature weighting technique for overcoming errors due to imperfect source separation

Reference Book Chapter

- *Advances in Music Information Retrieval, ser. Studies in Computational Intelligence, Z. W. Ras and A. Wieczorkowska, Eds. Springer*, 274
  - N. Ono, K. Miyamoto, H. Kameoka, J. Le Roux, Y. Uchiyama, E. Tsunoo, T. Nishimoto and S. Sagayama, "Harmonic and Percussive Sound Separation and its Application to MIR-related Tasks," pp.213-236



Available Separation Softwares

- Harmonic Percussive Sound Separation (HPSS)
  - http://www.hil.t.u-tokyo.ac.jp/software/HPSS/
- ICA Central: Early software restricted to mixtures of two sources
  - http://www.tsi.enst.fr/icacentral/algos.html
- SiSEC Reference Software: Linear modeling-based software for panned or recorded mixtures
  - http://sisec2008.wiki.irisa.fr/tiki-index.php?page=Under-determined+speech+and+music+mixtures
- QUAERO Source Separation Toolkit: Modular variance-modeling based software implementing a range of structures: GMM, NMF, source-filter model, harmonicity, diffuse mixing, etc
  - To be released Fall 2010: watch the music-ir list for an announcement!

Advertisement: LVA/ICA 2010

- LVA/ICA 2010 will be held in St. Malo, France on September 27-30, 2010.
- More than 20 papers on music and audio source separation will be presented.

References

- Singer/Instrument Identification

  - H. Fujihara, T. Kitahara, M. Goto, K. Komatani, T. Ogata and H. Okuno, ”Singer Identification Based on Accompaniment Sound Reduction and Reliable Frame Selection, “ Proc. ISMIR, 2005.
  - M. Goto, “A real-time music-scene description system: predominant-F0 estimation,” Speech Communication, vol. 43, no. 4, pp. 311–329, 2004.
  - A. Mesaros, T. Virtanen and A. Klapuri, “Singer identification in polyphonic music using vocal separation and pattern recognition methods,” *Proc. ISMIR*, pp. 375-378, 2007.
  - M. Ryynanen and A. Klapuri, ”Transcription of the Singing Melody in Polyphonic Music”, *Proc. ISMIR, 2006.*
  - T. Kitahara, M. Goto, K. Komatani, T. Ogata and H. G. Okuno, “Instrument identification in polyphonic music: feature weighting to minimize influence of sound overlaps,” EURASIP Journal on Applied Signal Processing, vol. 2007, 2007, article ID 51979.

References

- Audio Tempo Estimation

  - M. Alonso, G. Richard and B. David, “Accurate tempo estimation based on harmonic + noise decomposition,” EURASIP Journal on Advances in Signal Processing, Volume 2007 (2007), Article ID 82795
  - P. Chordia and A. Rae, "Using Source Separation to Improve Tempo Detection," Proc. ISMIR, pp. 183-188, 2009.

- Related Works to H/P Separation

  - C. Uhle, C. Dittmar, and T. Sporer, “Extraction of drum tracks from polyphonic music using independent subspace analysis," Proc. ICA, pp. 843-847, 2003.
  - M. Helen and T. Virtanen, "Separation of drums from polyphonic music using non-negative matrix factorization and support vector machine," Proc. EUSIPCO, Sep. 2005.
  - L. Daudet, "A Review on Techniques for the Extraction of Transients in Musical Signals," Proc. CMMR, pp. 219-232, 2005.
  - O. Dikmen, A. T. Cemgil, “Unsupervised Single-channel Source Separation Using Basian NMF,” *Proc. WASPAA*, pp. 93-96, 2009.

References

- Harmonic/Percussive Sound Separation

  - K. Miyamoto, H. Kameoka, N. Ono and S. Sagayama, “Separation of Harmonic and Non-Harmonic Sounds Based on Anisotropy in Spectrogram, *Proc. ASJ*, pp.903-904, 2008. (in Japanese)
  - N. Ono, K. Miyamoto, J. Le Roux, H. Kameoka and S. Sagayama, “Separation of a Monaural Audio Signal into Harmonic/Percussive Components by Complementary Diffusion on Spectrogram,” *Proc. EUSIPCO*, 2008.
  - N. Ono, K. Miyamoto, J. Le Roux, H. Kameoka and S. Sagayama, “A Real-time Equalizer of Harmonic and Percussive Components in Music Signals,” *Proc. of ISMIR*, pp.139-144, 2008.
  - N. Ono, K. Miyamoto, H. Kameoka, J. Le Roux, Y. Uchiyama, E. Tsunoo, T. Nishimoto and S. Sagayama, “Harmonic and Percussive Sound Separation and its Application to MIR-related Tasks,” *Advances in Music Information Retrieval, ser. Studies in Computational Intelligence, Z. W. Ras and A. Wieczorkowska, Eds. Springer*, 274, pp.213-236, Feb., 2010.

References

- Applications of HPSS to MIR Tasks

  - Y. Ueda, Y. Uchiyama, T. Nishimoto, N. Ono and S. Sagayama, “HMM-Based Approach for Automatic Chord Detection Using Refined Acoustic Features,” *Proc. ICASSP*, pp.5518-5521, 2010.
  - J. Reed, Y. Ueda, S. M. Siniscalchi, Y. Uchiyama, S. Sagayama, C. -H. Lee, “Minimum Classification Error Training to Improve Isolated Chord Recognition,” *Proc. ISMIR*, pp.609-614, 2009.
  - H. Tachibana, T. Ono, N. Ono and S. Sagayama, “Melody Line Estimation in Homophonic Music Audio Signals Based on Temporal-Variability of Melodic Source,” *Proc. ICASSP*, pp.425-428, 2010.
  - H. Rump, S. Miyabe, E. Tsunoo, N. Ono and S. Sagayama, “On the Feature Extraction of Timbral Dynamics,” *Proc. ISMIR*, 2010.

References

- Applications of HPSS in MIR Tasks

  - E. Tsunoo, N. Ono and S. Sagayama, “ Rhythm Map: Extraction of Unit Rhythmic Patterns and Analysis of Rhythmic Structure from Music Acoustic Signals,” *Proc. ICASSP*, pp.185-188, 2009.
  - E. Tsunoo, G. Tzanetakis, N. Ono and S. Sagayama, “Audio Genre Classification Using Percussive Pattern Clustering Combined with Timbral Features,” *Proc. ICME*, pp.382-385, 2009.
  - E. Tsunoo, N. Ono and S. Sagayama, “Musical Bass-Line Pattern Clustering and Its Application to Audio Genre Classification,” *Proc. ISMIR*, pp.219-224, 2009.
  - E. Tsunoo, T. Akase, N. Ono and S. Sagayama, “Music Mood Classification by Rhythm and Bass-line Unit Pattern Analysis,” *Proc. ICASSP*, pp.265-268, 2010.