Instrument Identification in Polyphonic Music: Feature Weighting with Mixed Sounds, Pitch-dependent Timbre Modeling, and Use of Musical Context

Tetsuro Kitahara, Masataka Goto**, Kazunori Komatani*, Tetsuya Ogata* and Hiroshi G. Okuno**

*(*Graduate School of Informatics, Kyoto University, Japan, **National Institute of Advanced Industrial Science and Technology, Japan)*

Introduction

Musical instrument identification

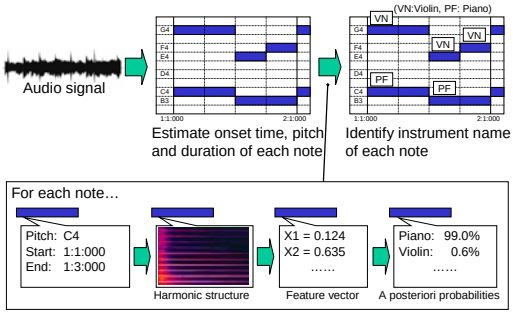
- To obtain the names of musical instruments from sounds
- A key technology for automatic music annotation
- Useful for instrument-based music information retrieval

I wanna listen to "string quartet" now.

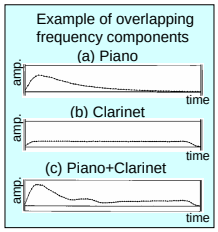
I need "piano sonata."

Issues in Handling Polyphonic music

Issue 1 Feature variations caused by sound mixtures
Issue 2 Pitch dependency of timbres
Issue 3 Musical context



Issue 1: Feature variations caused by sound mixtures



What's the problem?

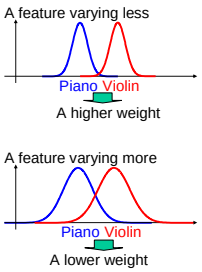
- Frequency components overlapping because of multiply instruments simultaneously playing
  ⇨*Features from such components vary*

In previous studies...

- Template matching, missing feature theory, etc.
- No attempts of feature weighting like our method

Basic Idea

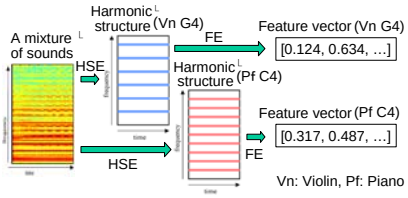
- Feature weighting based on feature variations
  - Give higher weights to features varying less
  - Give lower weights to features varying more



How we achieve the basic idea

- Use training data extracted from *mixed sounds*
- *Linear discriminant analysis* (LDA) with the data

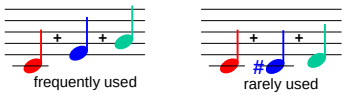
LDA is a dimensionality reduction technique that maximizes the ratio of the between-class covariance to the within-class covariance



How we collect "mixed-sound" data

- Impossible to collect all possible combinations of mixing sounds
  (instr. comb. × pitch comb. × …)
- Some combinations are frequently used, whereas other combinations are very rare
- We want to collect only frequently used ones

Creating sound mixtures based on the scores of *actual musical pieces*



Issue 2: Pitch dependency of timbres

What's the problem?

- Wide pitch ranges of musical instruments make their timbres quite different from pitch to pitch

In previous studies...

- They have not modeled how timbres changed according to the pitch

Our solution

- *F0-dependent multivariate normal distribution*
  - *F0-dependent mean function* $\mu_i(f)$
    Approximating the pitch dependency of each feature as a function (cubic polynomial) of F0
  - *F0-normalized covariance* Σ_i
    Representing the non-pitch dependency by normalizing the F0-dependent mean function

$$p(x \mid \omega_i; f) = \frac{1}{(2\pi)^{d/2}\left|\Sigma_i\right|^{1/2}} \exp\left\{-\frac{1}{2}(x-\mu_i(f))'\Sigma_i^{-1}(x-\mu_i(f))\right\}$$

Please check our ICASSP 2003 paper for more detail information.

Issue 3: Musical context

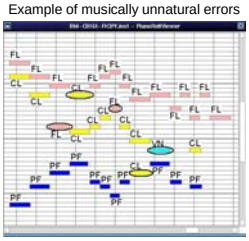
What's the problem?

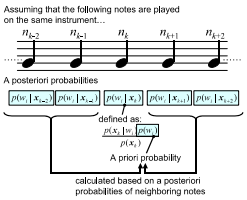
- Individually identifying the instrument of each note sometimtes causes musically unnatural errors (e.g. only one clarinet note in a melody on a flute)

In previous studies...

- Bayesian Network [Kashino '99]
- No attempts of simpler frameworks

Example of musically unnatural errors



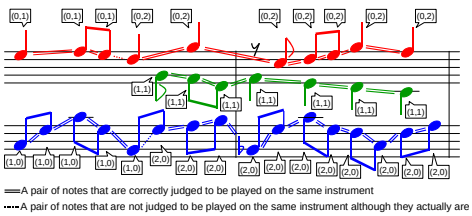


Basic Idea

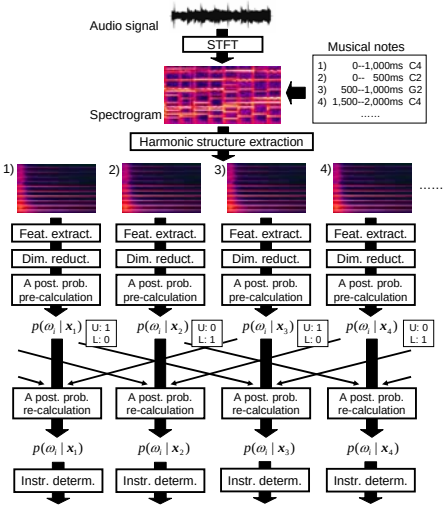
- Apply the *a posteriori* probabilities of *temporally neighboring notes* to the *a priori* probability

How we find "temporally neighboring notes"

- Their instruments should not be different from the target note
- Use musical heuristics that pitch crossing in simultaneous melodies rarely occurs



System Overview and Acoustic Features



Spectral Features	
[1]	Spectral centroid
[2]	Relative power of fundamental component
[3]--[10]	Relative cumulative power from fundamental to i-th components
[11]	Relative power in odd and even components
[12]--[20]	Number of components whose duration is p% longer than the longest duration
Temporal Features	
[21]	Gradient of straight line approximating power envelope
[22]--[30]	Average differential of power envelope durating t-sec interval from onset time
[31]--[39]	Ratio of power at t-sec after onset time
Modulation Features	
[40], [41]	Amplitude and Frequency of AM
[42], [43]	Amplitude and Frequency of FM

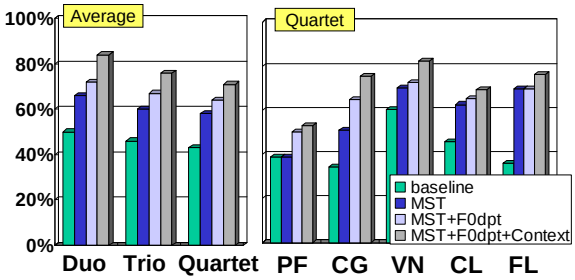
Experiments

Data for experiments

- Test set: Duo, trio and quartet music (3 pcs. each)
- Training data: Solo and duo music (2 pieces other than the test for each piece)
- Generated by mixing audio data in RWC-MDB-I-2001 according to SMFs on a computer.
- Piano, classical guitar, violin, clarinet and flute

Experimental results

- The recognition rate was improved
- Even if the combinations of instruments were limited, the recognition rates were improved.
- Without LDA, the recognition rates were improved only by a few percents.
  ⇨LDA with mixed sounds were effective.



Conclusions

- We presented three methods that work in combination to automatically generate the description of musical instrument names for music information retrieval.
- Experimental results with our methods showed the improvement of musical instrument identification.
- Future work: to integrate our methods with a note estimation method manually performed.

The 6th Int'l Conf. on Music Information Retieval (11th-15th Sept. 2005 in London, UK)