Musical Instrument Recognizer "Instrogram and Its Application to Music Retrieval based on **Instrumentation Similarity**

Tetsuro Kitahara* Masataka Goto** Kazunori Komatani* Tetsuya Ogata* Hiroshi G. Okuno* *Grad. Sch'l of Informatics, Kyoto University, Japan **Nat'l Inst. of Adv. Ind. Sci. and Tech. (AIST), Japan

Similarity-based MIR (MIR: Music Information Retrieval)

- To retrieve musical pieces similar to the piece(s) specified by the user
- No special skill necessary to use it
- Issue: What kind of cues should be used for measuring similarity?



If the same piece is played on different instruments, you should have different impressions!

Demo

- Retrieve musical pieces that have similar instrumentation to that specified by the user
- Visualize how likely each instrument is played in parallel to playback



How to Achieve This?

i. Conventional instrument recognition

Onset & pitch estimation needed in advance

It is still a challenging problem for polyphonic music



ii. Low-level features (e.g. MFCCs)

➡ Suffer from other elements, e.g. harmony

iii. New recognition scheme

Instrogram

- New instrument recognition scheme
- Represents how likely each instrument is played as probabilities
- Instrument existence probability (IEP)
 p(ω_i; t, f) ... the probability that the sound of instrument ω_i with pitch of f exists at time t
 Calculate this probability for every time & pitch
- No need to estimate onset times or pitches

Instrogram (1/2)

Visualization of IEPs on time-frequency plane Example: Auld Lang Syne (Flute-Violin-Piano)



Instrogram (2/2)

Visualization of IEPs on time-frequency plane Example: Auld Lang Syne (Flute-Violin-Piano)



Formulation

IEP $p(\omega_i; t, f)$ can be decomposed as follows: $p(\omega_i; t, f) = p(X; t, f) \quad p(\omega_i | X; t, f)$ $\therefore \omega_i \cap X = \omega_i$

- $\mathbf{X} = \omega_1 \cup \cdots \cup \omega_m$ (Existence of *some* instrument)
- p(X; t, f) Nonspecific IEP (NIEP)
 Prob. that the sound of *some* instrument exists
 p(ω_i|X; t, f) Conditional IEP (CIEP)
 Cond. prob. that if the sound of some instrument exists, the instrument is ω_i

Overview of Algorithm





Nonspecific Instr. Exist. Prob.

 Calculate how likely a sound exists at each pitch from the observed power spectrum every frame
 ⇒ Use PreFEst [Goto 2004]



Consider an observed spectrum to be a weighted mixture of tone models with different pitches and estimate the weights using EM algorithm

Conditional Instr. Exist. Prob.

Clarinet

CIPF.inst - PianoRollViewer

• For each pitch...

The temporal trajectory of the harmonic structure is modeled as follows:



Similarity between Instrograms

- 1. Get the vector consisting of IEPs every frame
- 2. Define the distance of vectors as cosine distance
- 3. Apply *dynamic time warping* (DTW)



Flute's IEP

Dynamic time warping Musical piece $\{q_t\}$

Vector q_t Use cosine distance Vector p_t Piano's IEP



Musical piece $\{p_t\}$

Experiments: Making Instrograms

- Real performances of classical/jazz music from solo to quartet (taken from RWC-MDB)
- Training data: Trio music created by mixing isolated monophonic tones
- Targets: Piano, Violin, Clarinet, Flute

Classical	(i)	No. 12, 14, 21, 28	Strings
	(ii)	No. 19, 40	Piano+Strings
	(iii)	No. 43	Piano+Flute
Jazz	(iv)	No. 1, 2, 3	Piano solo

Results (1/2)



- High IEPs (light color) for violin
- Piano's IEPs for the right were higher than those for the left

Results (2/2)



- Left: Initially only piano
 strings begin to play
- Right: Only piano from beginning to end

Retrieval Results & Discussions

- Results of instrumentation-similarity-based MIR
 - All of 3-best-similarity pieces from strings music were strings music
 - All of 3-best-similarity pieces from non-strings music were non-strings music
 - Better than MFCCs (typical low-level features)
- Instrogram is a middle-level descriptor
 - Music descriptors should be
 - musically meaningful (intuitive)
 - Each describing a single musical aspect

Conclusions

- Instrogram: Spectrogram-like representation of instrument existence probability (IEP)
 - Nonspecific IEP × Conditional IEP
 - No need to estimate onsets and F0s
 - Non-symbolic representation of instrumentation
- MIR based on instrogram similarity
- Future work: Tests on various music (e.g. ones containing drums and vocals), etc.