Instrogram:
A New Musical Instrument Recognition Technique Without Using Onset Detection Nor F0 Estimation

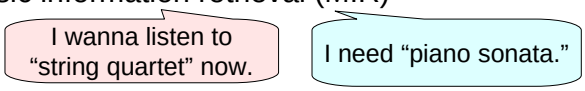
Abstract

- Task: Instrument recognition in polyphony
- Problem: Need to estimate onsets and F0s
- Solution: New framework based on *instrogram*
  - Calculate *instrument existence probabilities* for every (time, freq.)
  - Visualize them like a spectrogram
  - No need to estimate onsets nor F0s

1. Our task

Musical instrument recognition

- To recognize what instruments are played from polyphonic audio signals
- A key technology for various applications:
  - Music information retrieval (MIR)

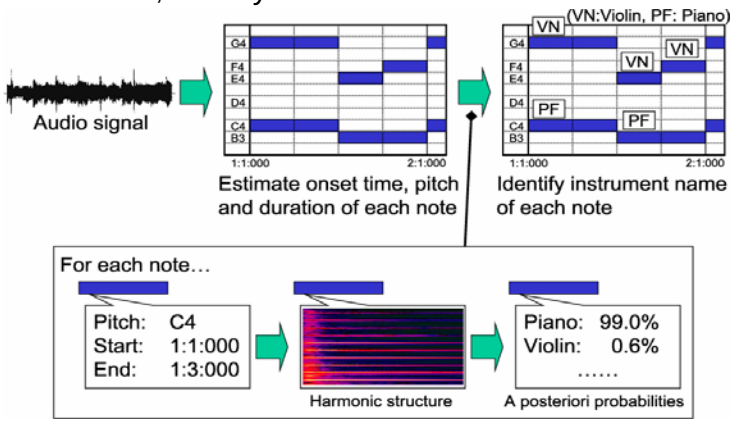


  - Multimedia content annotation (e.g. MPEG-7)
  - Automatic music transcription

2. Conventional framework

Notewise sequential framework

- First, estimate the onset time and F0 of every note
- And then, identify the instrument for each note

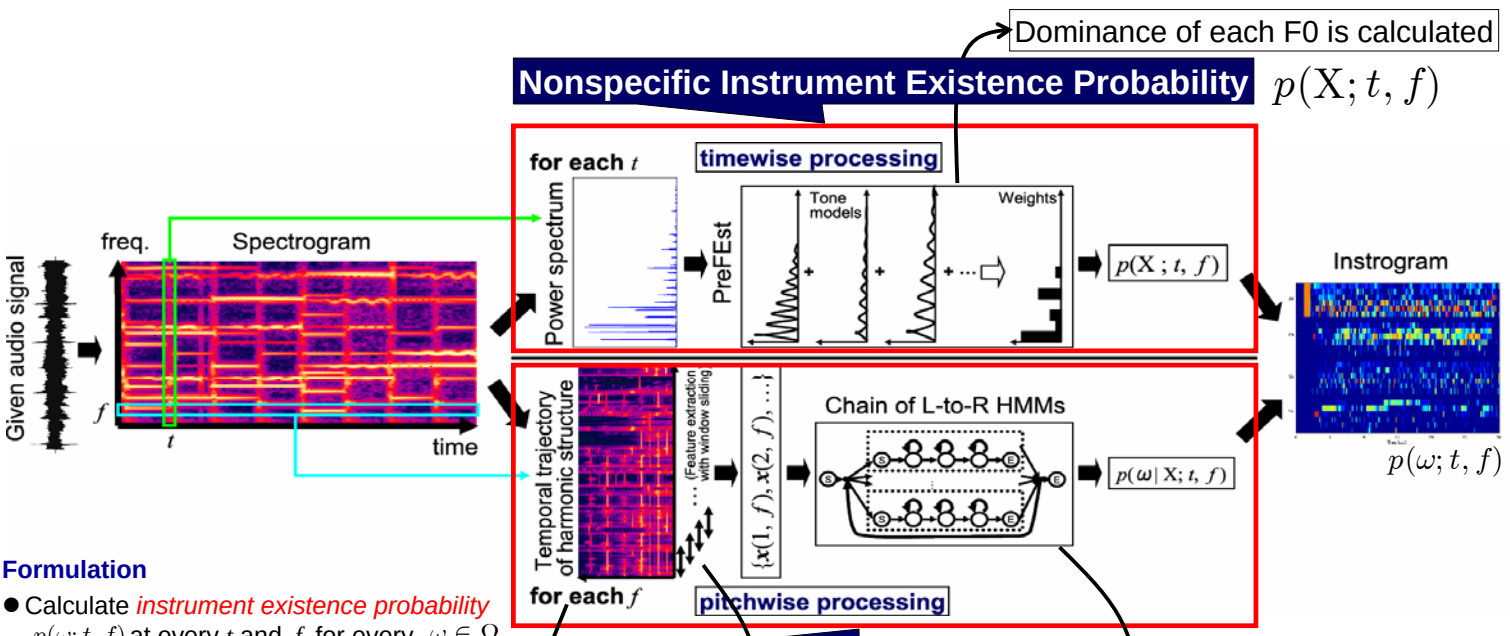


Two critical problems

- Accurate estimation of the onset time and F0 of every note is required.
  ⇨These estimation is not easy in polyphonic music
- Once the preceding estimation fails for some notes, identifying their instruments are impossible.

4. Algorithm for calculating instrogram

The **instrument existence probability** is calculated as the product of a **nonspecific instrument existence probability** and a **conditional instrument existence probability**



Formulation

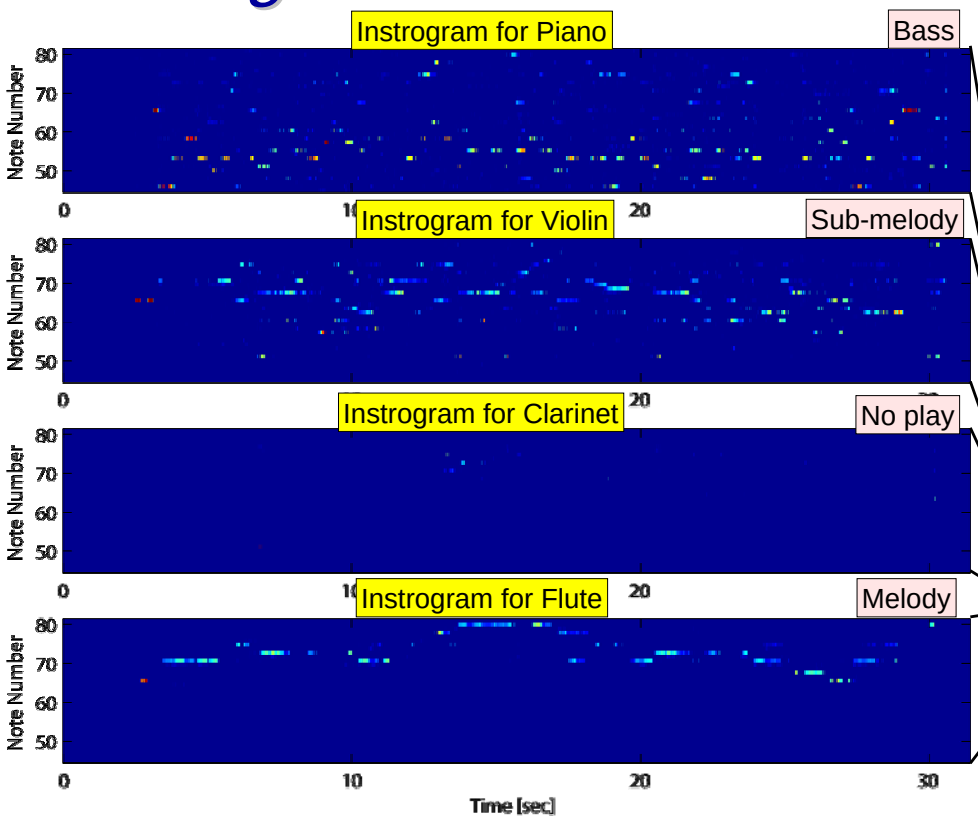
- Calculate *instrument existence probability* $p(\omega;t,f)$ at every t and f for every $\omega \in \Omega$
  t: time, f: frequency, Ω: the set of target instruments $(\Omega = \{\omega_1, \cdots, \omega_m\})$
- Assume that more-than-one instruments are not played at the same time and frequency
  $\forall \omega_i, \omega_j \in \Omega : p(\omega_i \cap \omega_j;t,f) = 0$ if $i \neq j$
  ⇨ $p(\omega;t,f) = p(\mathrm{X};t,f)\ p(\omega|\mathrm{X};t,f)$
  (X: Symbol standing for the existence of *some* instrument, i.e. $\mathrm{X} = \omega_1 \cup \cdots \cup \omega_m$)

The Int'l Conf. on Acoustics, Speech and Signal Processing (14th-19th May 2006 in Toulouse, France)

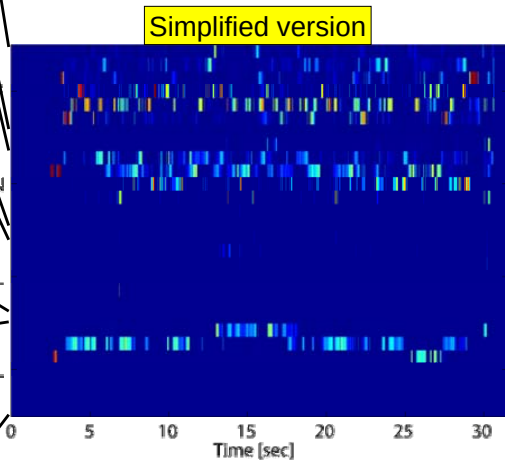
Tetsuro Kitahara, Masataka Goto**, Kazunori Komatani*, Tetsuya Ogata* and Hiroshi G. Okuno**
**Graduate School of Informatics, Kyoto University, Japan*
***National Institute of Advanced Industrial Science and Technology, Japan*
kitahara@kuis.kyoto-u.ac.jp, http://winnie.kuis.kyoto-u.ac.jp/~kitahara/

3. Our solution

Instrogram Spectrogram-like graphical representation of *instrument existence probability*



- Recognize the instrument not for every note but for every (t, f)
- No need to estimate onsets nor F0s

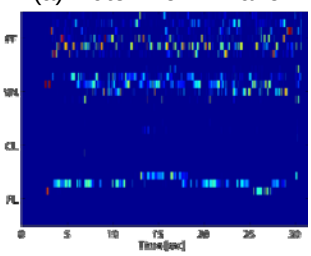


Instrogram obtained from "Auld Lang Syne" played on piano, violin, and flute

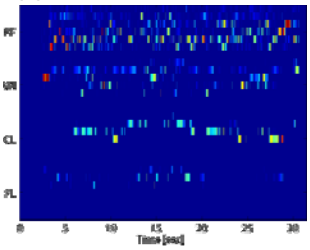
5. Experiments

I. Synthesized trio music "Auld Lang Syne"

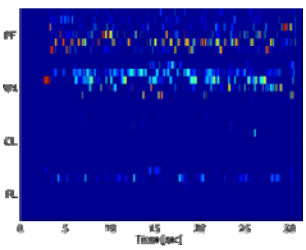
(a) Flute+Violin+Piano



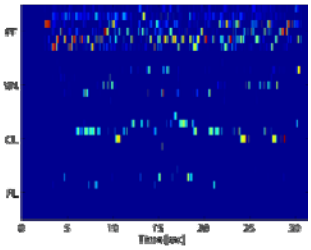
(b) Violin+Clarinet+Piano



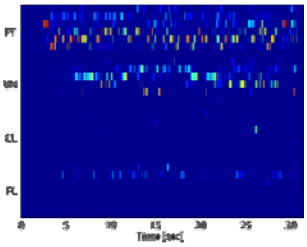
(c) Violin+Violin+Piano



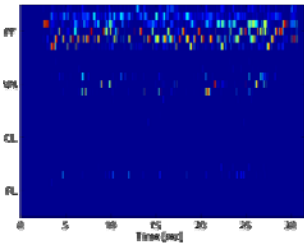
(d) Piano+Clarinet+Piano



(e) Piano+Violin+Piano

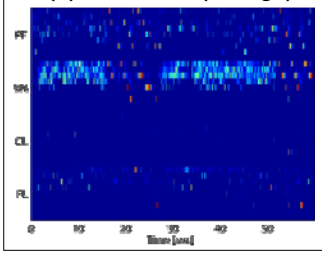


(f) Piano+Piano+Piano

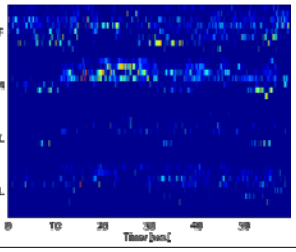


II. Real performances (taken from RWC Music DB)

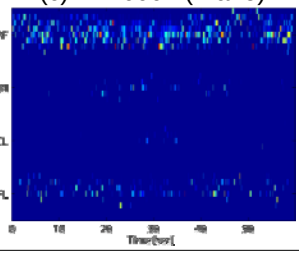
(a) RM-C014 (Strings)



(b) RM-C019 (Piano+Strings)



(c) RM-J001 (Piano)



6. Discussions

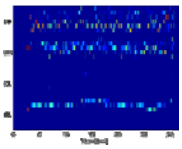
I. Relation to people's music listening

- Listening to Music ≠ Obtaining Score
  - They can understand music without mentally representing it as a score
  - They can search for piano music even if not recognize every note
  - ⇨ Instrogram enables non-score-based music understanding

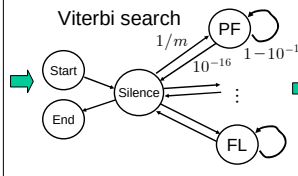
II. Potential Applications

- MPEG-7 Annotation
  (e.g. when each instrument starts & stops playing)

$$\hat{\omega}_{t,f} = \operatorname{argmax}_{\omega}(\omega; t, f)$$



PF	PF	..	..	..	..	..	..
..	VN	..	FL	FL	FL	FL	FL
..	FL	..	PF	PF	PF	..	PF
..	PF	PF	..	PF	..	..	PF
..	..	..	PF	..	..	..	..



Piano begins playing
= Trans. "Silence" to "PF"

Piano stops playing
= Trans. "PF" to "Silence"

- Instrumentation-similarity-based MIR

Please see demo on my laptop!