Analysis-and-manipulation approach to pitch and duration of musical instrument sounds without distorting timbral characteristics

Takehiro Abe † *Katsutoshi Itoyama* † *Kazuyoshi Yoshii* ‡
Kazunori Komatani † *Tetsuya Ogata* † *Hiroshi G. Okuno* †

†Department of Intelligence Science and Technology,
Kyoto University, Japan
‡National Institute of Advanced Industrial Science
and Technology (AIST)

Demonstration: http://winnie.kuis.kyoto-u.ac.jp/~abe/DAFx-08/

Motivation

Active music listening
- Selecting from **users' requirement**
- **Changing music** to suit users' feeling

Active and exploratory listening

Conventional music listening
- Selecting from **limited playlist**
- **Only listening** after pressing "play"

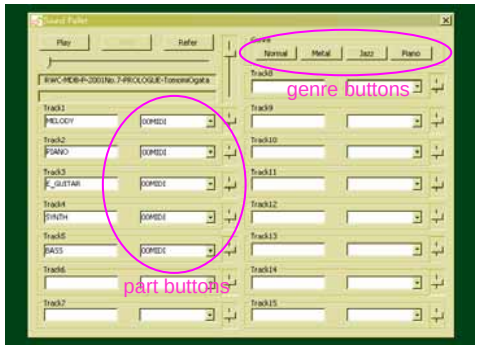
Passive and limited listening experience

Instrument equalizers have been developed

user can change	Drumix [Yoshii 07]	Itoyama's EQ. [Itoyama 08]	Our equalizer
Instruments	✕ (Drums only)	○ (All)	○
Volume	○	○	○
Timbre	○	✕	○

Replacing arbitrary part with users favorite timbre

Demonstration (Trial equalizer)



Content
- midi sound
- synthesized piano sound
- Jazz sound (synthesis)

Equalizer's sounds are synthesized from **real sounds** except midi sounds

Requirements for our equalizer

1. Sound separation from polyphonic audio to extract a musical instrument sound that users want to replace

Well studied

2. Sound manipulation from separated sounds without timbral distortion to play arbitrary phrases

The application of separated sounds is not well studied

Our research target
Difference from the sound excited by real instrument

Objective

Synthesizing **monotones excited by the same instrument** from **multiple musical instrument sounds**

Our definition of timbral features

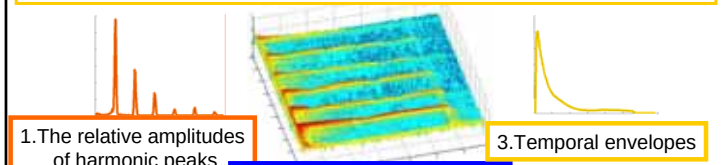
ASA's definition

The quality of a sound that distinguishes it from others of the same pitch and volume [ASA 60]

Concrete definition based on [Grey 77]

Our definition

The quality of a sound that consists of **three features** except pitch and volume



1.The relative amplitudes of harmonic peaks

2.The inharmonic component

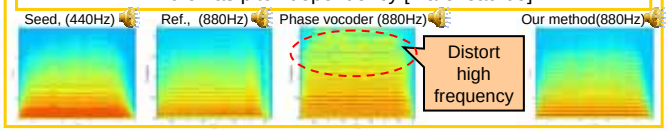
3.Temporal envelopes

We use the tonal model that can analyze these features [Itoyama 08]

Manipulation of pitch and duration

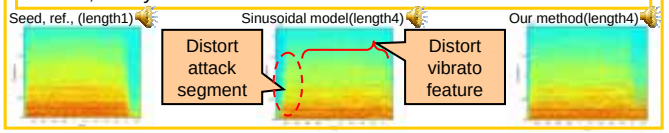
It is not proper to achieve manipulation without changing the timbral features

Timbre has pitch dependency [Marozeau 03]



We use pitch-dependency feature function for the dependency

Attack, decay and vibrato feature are similar in the same instrument



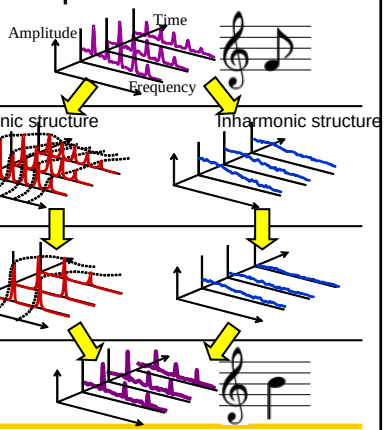
We preserve attack, decay segments and vibrato feature

Overview of our manipulation method

Step1：Analysis
Separate harmonic and inharmonic structures and extract timbral features

Step2：Manipulation
Manipulate pitch, duration, and energy of the inharmonic structure

Step3：Synthesis
Synthesize harmonic and inharmonic signals and add them



Analysis to obtain three features

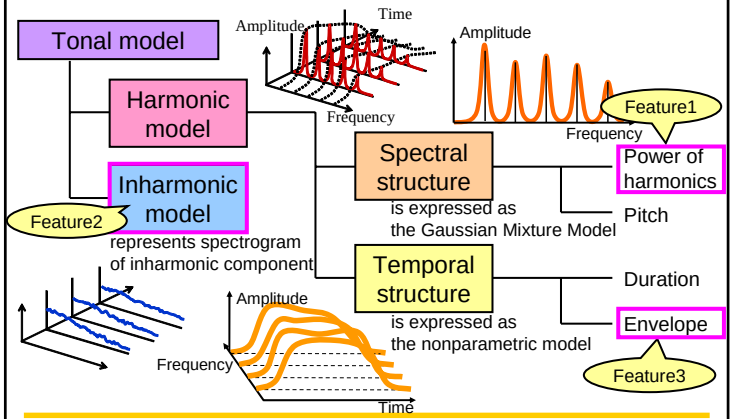
Tonal model

Harmonic model

Inharmonic model

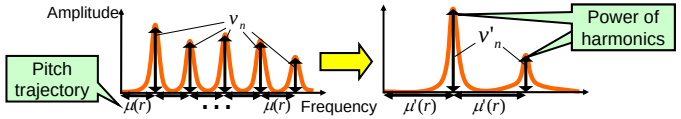
Feature2

represents spectrogram of inharmonic component

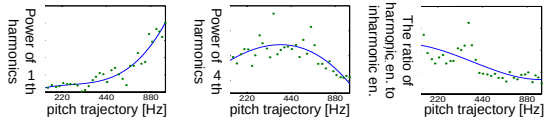


Pitch manipulation

- Manipulating the spectral envelope
  - by multiplying the pitch trajectory ($\mu(r)$) by a desired ratio
  - Obtain timbral features from pitch-dependent feature function



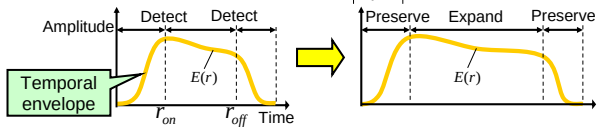
- Pitch-dependent feature function
  - approximates timbral features over pitches by polynomial function
    - power of harmonics (v_n)
    - the ratio of harmonic energy to inharmonic energy (w_H / w_I)



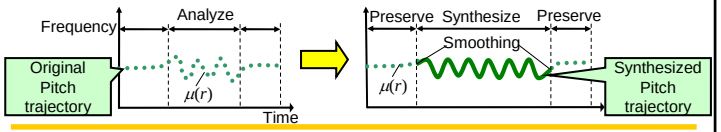
Duration manipulation

- Manipulating the temporal envelope ($E(r)$)
  - by expanding or shrinking between onset (r_{on}) and offset (r_{off})

detection equation: $\left|\frac{dE(r)}{dr}\right| < \varepsilon, E(r) > Th$

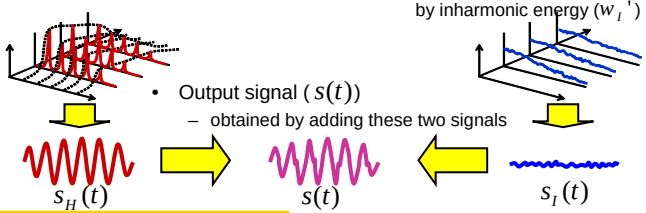


- Preserving the vibrato
  - Pitch trajectory ($\mu(r)$) is analyzed and synthesized by sinusoidal model



Synthesis from harmonics and inharmonics

- Harmonic signal ($s_H(t)$)
  - using sinusoidal model
- Inharmonic signal ($s_I(t)$)
  - from inharmonic model weighted by inharmonic energy (w_I ')



- Output signal ($s(t)$)
  - obtained by adding these two signals

※" ' " parameter is a manipulated parameter.

Equations for harmonic signal

Harmonic signal: $s_H(t) = \sum_n A_n(t)\exp[j\phi_n(t)]$

Instance amplitude: $A_n(t) = w_H v_n' E_n(t)$

Instance phase: $\phi_n(t) = \phi_n(0) + \int_0^t n\mu'(\tau)d\tau$

Harmonic energy: w_H

Power of harmonics: v_n'

Temporal envelope: $E_n(t)$

Pitch trajectory: $\mu'(\tau)$

Evaluation in pitch manipulation

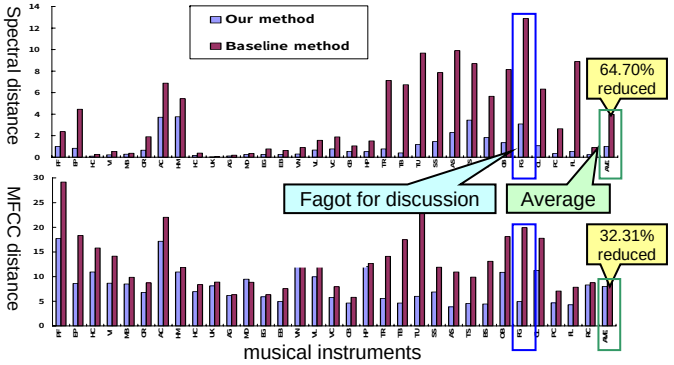
- Baseline method = Sophisticated sinusoidal model
  - Our method **without pitch-dependent feature function**
- Criteria
  - Spectral distance: evaluation of **harmonic component difference**
  - Mel-Frequency Cepstrum Coefficient (MFCC) distance:
    - quantitative auditory measurement
    - evaluation of **harmonic and inharmonic components differences**

$$D = \sum_{f,t} (C_{real}(f,r) - C_{syn}(f,r))^2 / T \qquad C_i$$

Real sound: $C_{real}(f,r)$ — Synthesis sound: $C_{syn}(f,r)$ — Frames: T — C_i Spectrum or MFCC

- Conditions
  - 32 instruments from RWC-MDB (forte, normal articulation)
    - 3 individuals for each instrument
  - 10-fold cross validation (10%:90% = [evaluation data]:[learning data])

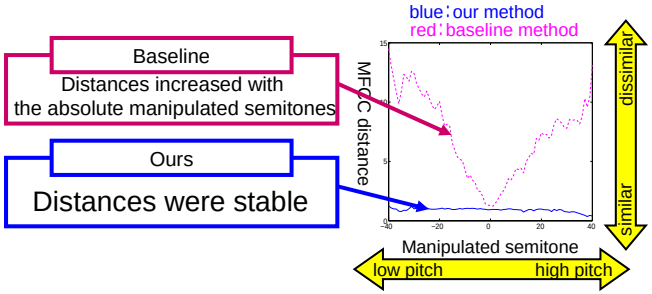
Quality in pitch manipulation



There was good improvement for the whole musical instruments

Discussion on good improvement

- The result of the fagot

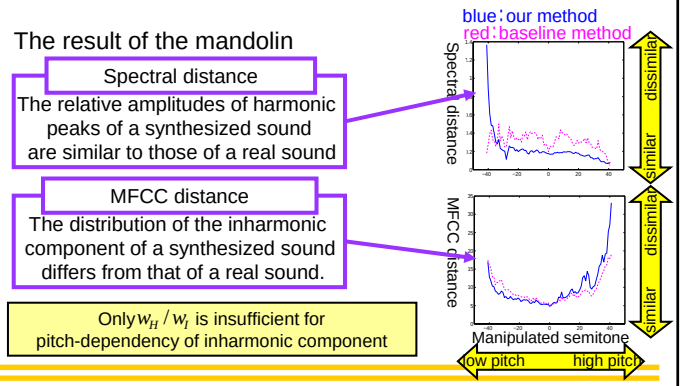


The result demonstrated the validity of our method, which considering pitch dependency of timbre

Discussion on poor improvement

There was poor improvement for instrument sounds that have a lot of inharmonic component in attack segment…

- The result of the mandolin



Only w_H / w_l is insufficient for pitch-dependency of inharmonic component

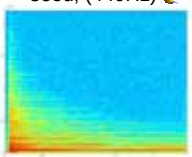
Conclusion

- Objective
  - Manipulating pitch and duration of a musical instrument sound **using multiple instrument sounds without distorting timbral characteristics**
- Approach
  - We **defined and analyzed timbral features**.
  - In pitch manipulation, we use **pitch-dependency of timbre** as a pitch-dependent feature function
  - In duration manipulation, we **preserve attack, decay and the vibrato**
- Future work
  - Incorporating other dependencies (e.g., volume)
  - Evaluating our method for duration manipulation
  - Applying our method to musical instrument parts separated from the polyphonic audio signals of commercial CD recordings

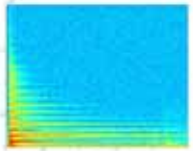
Pitch manipulation demo. for piano

Real sounds

seed, (440Hz)

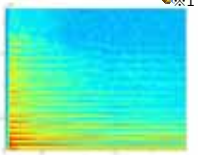


ref., (880Hz)

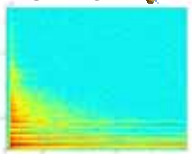


Synthesized sounds (880Hz)

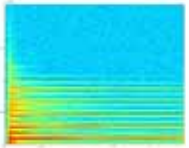
Phase vocoder ※1



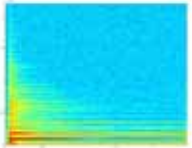
STRAIGHT



Sinusoidal model



Our method ※2



Digital Audio Effects DAFx-2008

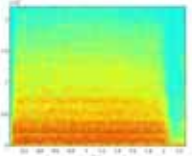
※1 from MARSYAS

※2 do not use sound Ref. as learning data

Duration manipulation demo. for violin

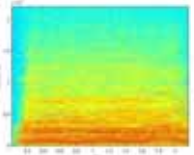
Real sound

seed, ref., (length1)

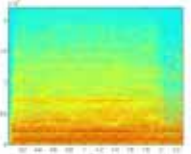


Synthesized sounds(length4)

Sinusoidal model



Our method

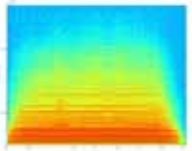


Digital Audio Effects DAFx-2008

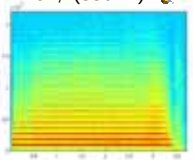
Pitch manipulation demo. for trumpet

Real sounds

seed, (440Hz)

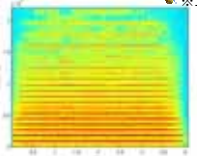


ref., (880Hz)

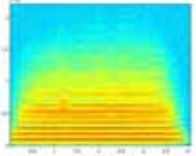


Synthesized sounds (880Hz)

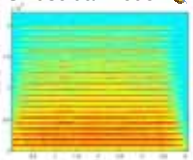
Phase vocoder ※1



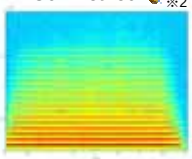
STRAIGHT



Sinusoidal model



Our method ※2



Digital Audio Effects DAFx-2008

※1 from MARSYAS

※2 do not use sound Ref. as learning data