Extracting Personal Characteristics from Singing Voice

Yasunori Ohishi†, Hirokazu Kameoka†, Daichi Mochihashi‡, Kunio Kashino†

† NTT Communication Science Laboratories
‡ The Institute of Statistical Mathematics
Aim

- Extracting various properties from singing
- First step: Fundamental frequency (=pitch)

Pitch contours contain various dynamic components
Aim

- Extracting various **properties** from singing
- First step: Fundamental frequency (=pitch)

Properties:
- Song, Singer, Skill, Style, Genre, Emotion, …

Overshoot: deflection exceeding target note after note changes

Vibrato: quasi-periodic frequency modulation (5–8 Hz)

Portamento: pitch sliding from one note to another

Fine fluctuation
Why?

*Traditional* music identification or search:

(1) Text-based search
   - Title, Score, or Manual tags

(2) Song identification from audio
   - Watermarking
   - Feature matching or audio fingerprinting

(3) Similarity search based on audio
   - Based on features or scores extracted from signals

What we are focusing on:

Information extraction from signals
   e.g. *singer, style, emotion*, …
Application 1: Search for Emotion

- Emotions extracted from singing signals

I need another song carrying the same emotion as this one...

Search

Music Database
Application 2: Search for Skill

- Skills extracted from singing signals

I’d like to find a better singing voice than mine for this song.
Application 3: Search for Singer

- Singer characteristics extracted from singing signals

I’d love to get all the songs of this singer…

Music Database

Search
Even Applicable to Transformation…

*Voice conversion based on styles and skills*

**Let’s sing a song**

Before

Sample 1
Sample 2
Sample 3

After

Sample 1
Sample 2
Sample 3

Good !?
Skills and pitch contours (example)

Even if they are the same song, these contours are different from each other.
Skills and pitch contours (example)

1. Musical-note information
2. Various dynamic components

(1) Physical constraints of the vocal chords
   - Naturalness and singer’s individuality

(2) Singer’s musical expressive intentions (Vibrato and so on)
   - Singing styles and skills

Modeling and extracting dynamic components!
To extract dynamics from pitch contours

Process of generating singing voice

1. Musical score
2. Design how we can sing that song expressively as our intentions
3. Produce singing voices based on physical organs
4. Our voices are recorded as voice signals by microphones
Singing Pitch Contour Model

(1) **Intentions**: Pitch and duration, Expression

(2) **Physical organs**: Control system

(3) **Pitch contour**: Sum of outputs of control systems
Specifically

(1) Intentions: Stepwise signal and square signal
(2) Physical organs: Second order differential equation
(3) Pitch contour: Sum of outputs and fine fluctuation
To extract dynamic components

(1) Probabilistic modeling of generative process

(2) Parameter (dynamic components) estimation

Probabilistic modeling

Parameter estimation

Pitch contour (Observed signal)

Second order differential equation

Second order differential equation
(1) Probabilistic modeling of generative process

Hidden Markov model (HMM)

Autoregressive model

\[
\frac{\Omega^2}{s^2 + 2\zeta\Omega + \Omega^2} 
\]

Approximation

\[
u[k] = a_0 y[k] + a_1 y[k - 1] + a_2 y[k - 2]
\]
(2) Parameter estimation based on *EM algorithm*

![Diagram of parameter estimation process]

**M-step: Update of parameters**

\[ u[k] = a_0 y[k] + a_1 y[k - 1] + a_2 y[k - 2] \]

**E-step: Decomposition**

State sequence

AR parameters
Commands for Vibrato (Estimation results)

Pitch contour (Observed signal)

Vibrato: quasi-periodic frequency modulation (5–8 Hz)

Note command

Expression command
Commands for Portamento (Estimation results)

Pitch contour (Observed signal)

Portamento
(Pitch sliding from one note to another)

Note command

Expression command
Evaluation of model validity

- Expression commands
  Experiment 1: Vibrato detection
    - Determine if each frame is Vibrato frame or not
    - Calculate true positive and false negative
  Experiment 2:
  Singing skill discrimination
    - Discrimination between professional and amateur classes

<table>
<thead>
<tr>
<th>Rate [%]</th>
<th>Professional</th>
<th>Amateur</th>
</tr>
</thead>
<tbody>
<tr>
<td>True positive</td>
<td>81%</td>
<td>70%</td>
</tr>
<tr>
<td>False negative</td>
<td>5%</td>
<td>90%</td>
</tr>
</tbody>
</table>

Correct rate [%]

Professional 70%
Amateur 90%
Evaluation of model validity

- **Expression commands**

Experiment 1: Vibrato detection
- Determine if each frame is Vibrato frame or not
- Calculate true positive and false negative

Experiment 2:
**Singing skill discrimination**
- Discrimination between professional and amateur classes

<table>
<thead>
<tr>
<th>Rate [%]</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>True positive</td>
<td>81%</td>
</tr>
<tr>
<td>False negative</td>
<td>5%</td>
</tr>
</tbody>
</table>

We tested **behavior of parameter estimation algorithm** and verified **effectiveness of estimated expression commands** through automatic discrimination experiments.
Concluding remarks

- Extracting various properties from singing
- First step: Fundamental frequency (=pitch)
  - Singing pitch contour model
  - Statistical framework for parameter estimation
- Quantitative evaluation of model validity using the large data sets
- Learning properties (singer, skill, style...) using the expression commands
- Application to singer, style, and emotion identification and search
Thank you!
Estimation results

Pitch contour (Observed signal)

Note command

Expression command
Synthesized pitch contour

Estimated note command

Estimated expression command

Fine fluctuation

Second order differential equation

Second order differential equation

Synthesized pitch contour
Estimation results

Pitch contour (Observed signal)

Note command

Expression command

Synthesized contour

Similar!
Command generation using HMM

Musical-note “G”

Musical-note “B”

Musical-note “A”

Start

End
Command generation using HMM

Musical-note "G"

\begin{align*}
  S_{1,1} \
  S_{1,2} \
  S_{1,3}
\end{align*}

Start

Note command

Expression command

Time

\begin{align*}
  B \
  A \quad \uparrow \
  G
\end{align*}

End

\begin{align*}
  S_{2,1} \
  S_{2,2} \
  S_{2,3}
\end{align*}

"B"

\begin{align*}
  S_{3,1} \
  S_{3,2} \
  S_{3,3}
\end{align*}

"A"
Command generation using HMM

Musical-note “G”

Note command

Expression command

Time

Time
Command generation using HMM

Musical-note “G”

Start $S_{1,1}$ $S_{1,2}$ $S_{1,3}$

“B”

$S_{2,1}$ $S_{2,2}$ $S_{2,3}$

“A”

$S_{3,1}$ $S_{3,2}$ $S_{3,3}$

Note command

Expression command

B A G

Time

Time
Command generation using HMM

Musical-note “G”

“B”

“A”

Note command

Expression command

Time

Time
Command generation using HMM

Musical-note “G”

Start $S_{1,1}$ $S_{1,2}$ $S_{1,3}$

“B”

$S_{2,1}$ $S_{2,2}$ $S_{2,3}$

“A”

$S_{3,1}$ $S_{3,2}$ $S_{3,3}$

Note command

Expression command

Time

B

A

G

Start

End
Command generation using HMM

Musical-note “G”

Note command

Expression command

Pitch and duration

Expression

Time

Time
Command generation using HMM

Musical-note “G”

Note command

Expression command

Start → \( S_{1,1} \) → \( S_{1,2} \) → \( S_{1,3} \) → \( S_{1,1} \) → End

Start → \( S_{2,1} \) → \( S_{2,2} \) → \( S_{2,3} \) → \( S_{2,1} \) → End

Start → \( S_{3,1} \) → \( S_{3,2} \) → \( S_{3,3} \) → \( S_{3,1} \) → End

Note command:
- B
- A
- G

Expression command:
- Time
Command generation using HMM

Musical-note “G”

Start

Note command

Expression command

“B”

“A”

End

Time

Time
Command generation using HMM

Musical-note “G”

Note command

Expression command

Time

Start → End
Command generation using HMM

Musical-note “G”

“B”

“A”

Note command

Expression command
Command generation using HMM

Musical-note “G”

“B”

“A”

Note command

Expression command

Time

Time

Start

End

S_{1,1}

S_{1,2}

S_{1,3}

S_{2,1}

S_{2,2}

S_{2,3}

S_{3,1}

S_{3,2}

S_{3,3}
Command generation using HMM

Musical-note “G”

“B”

“A”

Note command

Expression command

Pitch and duration

Expression
Demonstration

Voice conversion based on singing styles

Let’s sing a song

Conversion!

$F_0$ contour

Commands

Professional

Good !?