A Stochastic Representation of the Dynamics of Sung Melody
Yasunori OHISHI†, Masataka GOTO‡, Katunobu ITOU‡, Kazuya TAKEDA†
†Nagoya University, ‡National Institute of Advanced Industrial Sci. and Tech. (AIST), †Hosei University

Introduction
- Our goal
  - Build a model that can represent the dynamics of various singing behaviors in a sung melodic contour
  - Define a melodic similarity measure between sung melodies
  - Useful in various applications such as query-by-humming (QBH), retrieval based on similarity in singers’ characteristics

Continuous transitions between notes
Fluctuations in a musical note
F0 contour

Phase Representation for Melodic Contour
- Graphical representation of the dynamic properties of sung melodic contours
- F0 trajectories are generated by a dynamic system and represented in a two-dimensional phase plane (F0-ΔF0)
- Represent the local direction of the F0 trajectory
  - Convert into Log - Frequency
  \[ F(t)_{\text{cen}} = 1200 \log_2 (F(t)_{\text{Hz}} / 440 \times 2^{12/4}) \]
  - Calculate F0 derivative (ΔF0)
  \[ \Delta F(t)_{\text{cen}} = \sum_{k=1}^{n} \Delta F(t+k) / \sum_{k=1}^{n} k^2 \]
- Fluctuation in a sung melody appears as a curling trajectory around a certain target point, i.e., an attractor of the system
- Vibrato (quasi-periodic modulation) Circular pattern
- Overshoot Sprial pattern
  (Exceed the F0 of a target musical note just after the note change)
- Location of each attractor corresponds to the F0 of the target musical note

Separate the dynamics of various singing behaviors from an original musical note sequence

Melodic similarity measure using the phase representation
- Stochastic phase representation (SPR)
  - Focus on the original (target) melodic information
  
  **Extract attractor points**
  1) F0 contour (short segments)
  2) Phase representation
  3) Train the F0-ΔF0 distribution
  
  **Fitting Gaussian Mixture Model**
  
  **2-Mixture GMM**
  EM algorithm
  - Gaussian weight \(w\)
  - Gaussian mean \(\mu\)
  - Gaussian variance (diagonal) \(\Sigma\)

- VQ codebook
  \[ v = \{ v_1, v_2, \ldots, v_M \} \]
  \[ p_a = \frac{1}{L} \sum_{t=1}^{L} || v(t) - v_a || \]

- Probability
  \[ p_a = \frac{1}{L} \sum_{t=1}^{L} || v(t) - v_a || \]

Alignment procedure between SPRs
- Input signal
- Reference signal

DTW

Histogram-intersection distance

Melodic similarity

Experiments
- Evaluate the potential of SPR on a small QBH
  - Song database
    - 200 short excerpts from 100 pop songs of the RWC Music Database
    - Average length of excerpts is 12 s
    - F0 contour was manually annotated
  - Query melodies
    - 75 Japanese subjects
    - Listened to each of the above 50 excerpts
    - Sang its melody with lyrics (3,750 (75 * 50) samples)
    - F0 contour was estimated for every 10 ms by using YIN
  - Preprocessing
    - Subtract the average F0 value over each contour

Results
- SPR vs. Baseline
- Window length L

<table>
<thead>
<tr>
<th>L</th>
<th>Recall rate</th>
<th>SPR (Histogram)</th>
<th>SPR (GMM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>100 ms</td>
<td>0.69</td>
<td>0.65</td>
<td>0.80</td>
</tr>
<tr>
<td>250 ms</td>
<td>0.69</td>
<td>0.65</td>
<td>0.80</td>
</tr>
<tr>
<td>500 ms</td>
<td>0.69</td>
<td>0.65</td>
<td>0.80</td>
</tr>
</tbody>
</table>

*Baseline performance: Traditional DTW using F0 contours
Novel melodic similarity measure for sung melodies
- Represent the dynamic properties of sung melodic contours using F0-ΔF0 phase plane
- Extract the original (target) melodic information in the phase plane
- QBH results: Recall Precision curve and MRR is better than the traditional DTW technique

Future work
By using phase representation
- Detect particular singing behaviors such as vibrato and overshoot automatically
- Generation of melodic contours that reflect personal singing behaviors