Automatic Identification for Singing Style based on Sung Melodic Characterized in Phase Plane

Tatsuya Kako†, Yasunori Ohishi††, Hirokazu Kameoka††, Kunio Kashino†† and Kazuya Takeda†

† Graduate School of Information Science ,Nagoya University

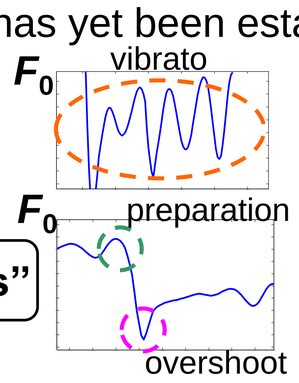
†† NTT Communication Science Laboratories, NTT Corporation

Abstract : A stochastic representation of singing styles is proposed. The dynamic property of melodic contour, i.e., fundamental frequency (F_0) sequence, is assumed to be the main cue for singing styles because it can characterize such typical ornamentations as *vibrato* . F_0 signal trajectories in the phase plane are used as the basic representation. By fitting Gaussian mixture models to the observed F_0 trajectories in the phase plane, a parametric representation is obtained by a set of GMM parameters. The effectiveness of our proposed method is confirmed through experimental evaluation where 94.1% accuracy for singer-class discrimination was obtained.

Introduction

"Singing Style" Definition has yet been established.

"Singing Style" ⟷ "Signal features"



Signal features
- Singing formant
- Vibrato
- Overshoot
- Preparation
- Fine fluctuation

◆Dynamic property of F_0 affects the perception of the individuality.

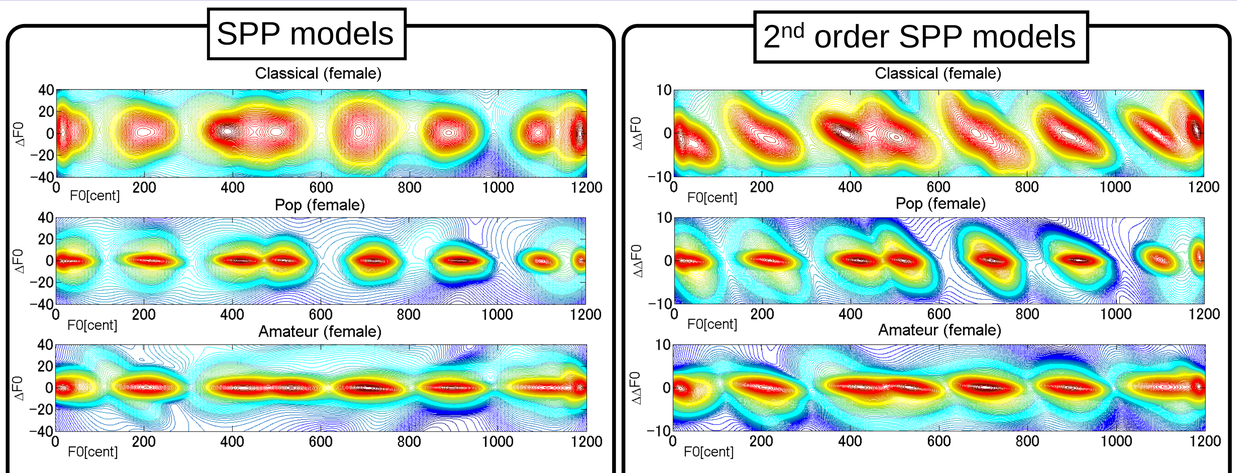
◆Singing style ⟶ Dynamics of a sung melody

Previous study: Graphical representation of the F_0 contour in phase plane [Ohishi et al., 2007]

In our study: Modeling for Singing Style

Use the local dynamics of the F_0 sequence on the phase plane

Example of Stochastic Phase Plane



Classical: Large vertical deviation
Amateur: Large horizontal deviation

Classical: Strong negative correlations between F_0 and $\Delta\Delta F_0$

Stochastic Representation of the Dynamical Property of Melodic Contour

F_0 signal in the phase plane

F_0 signal

◆Dynamic property of F_0 ⟶ Main cue for singing style

◆Controlled output of the human speech production system

Dynamic characteristics ⟸ differential equation

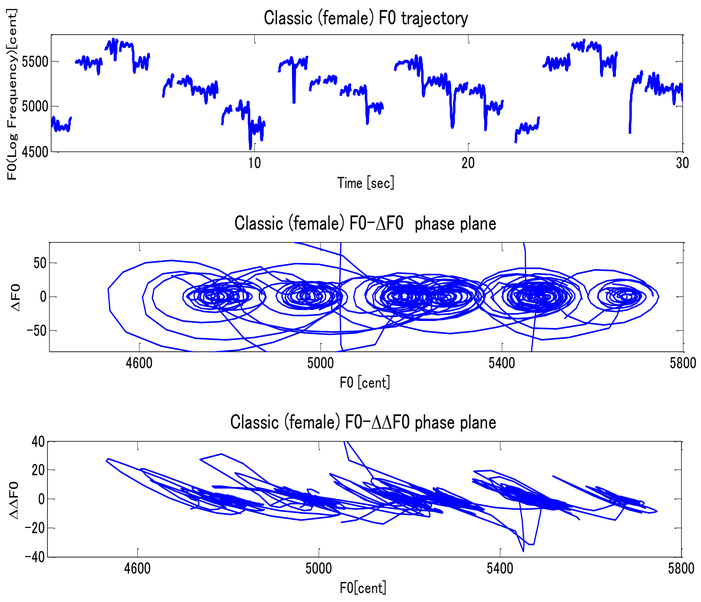
Phase Plane

◆Joint plot of a variable and its time derivative i.e., $(x, \dot{x})$

◆Time derivative using the delta- coefficient given by

$$\Delta F_0(n) = \sum_{k=-K}^{K} k \cdot F_0(n+k) \Big/ \sum_{k=-K}^{K} k^2$$

2K: window length for calculating the dynamics



Melodic contour (top) and corresponding phase plane for F_0-ΔF_0 (middle) and F_0-$\Delta\Delta F_0$ (bottm)

F_0-ΔF_0 phase plane

vibrato overshoot

- vibrato: circular trajectory centered at a target note
- overshoot: spiral pattern around the target note

F_0-$\Delta\Delta F_0$ phase plane

◆Slope of -45 degrees

$$\Delta\Delta F_0 = -A F_0$$

Assume that vibrato presents sinusoidal

$$y = \sin \omega t$$
$$y' = \omega \cos \omega t$$
$$y'' = -\omega^2 \sin \omega t$$

Stochastic representation of Phase Plane

◆Singing style ⟶ Trajectory on the phase plane

◆Model the trajectory on the phase plane with **GMM**

◆We can build a stochastic phase plane (SPP)

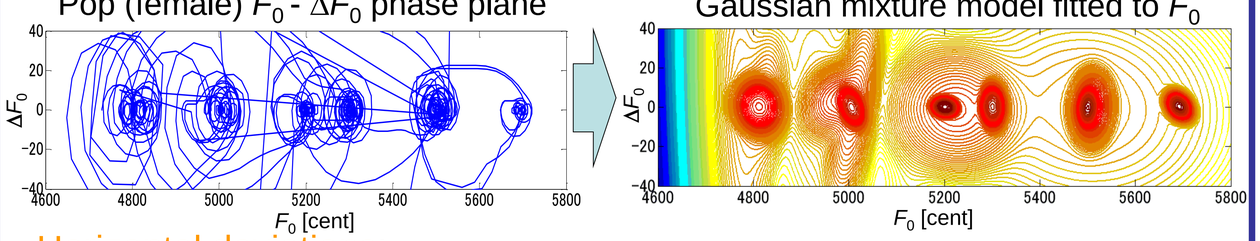
GMM

$$\sum_{m=1}^{M} \lambda_m N(\mathbf{f}_0(n); \boldsymbol{\mu}_m, \boldsymbol{\Sigma}_m) \quad \text{where} \quad \mathbf{f}_0(n) = [F_0(n), \Delta F_0(n), \Delta\Delta F_0(n)]^{\mathrm{T}}, \quad \Theta = \{\lambda_m, \boldsymbol{\mu}_m, \boldsymbol{\Sigma}_m\}_{m=1,\ldots,M}$$

A GMM trained for F_0 contours in the phase plane

Pop (female) F_0 - ΔF_0 phase plane

Gaussian mixture model fitted to F_0



Horizontal deviations: Stability of the melodic contour around the target note

Vertical deviations: Vibrato depth

Singing styles can be modeled by set of parameters Θ.

Experimental Evaluation

Effectiveness of using SPP ⇒ Discriminate different singing style

Experimental set up

◆Singing signals: 6 singers [Classical, Pop, Amateur]

◆Songs: "*Twinkle- Twinkle, Little Star*", "*Ode to Joy*" and 5 etudes with Japanese Lyrics and hummed

◆With/without musical accompaniment

Total of 132 song signals

Condition for F_0 estimation [Goto et al,1999]
- Signal sampling frequency: 16 kHz
- F_0 estimation window length: 64 ms
- window function: Hanning window
- window shift: 10 ms
- F_0 contour smoothing: 50 ms MA filter
- Δ coefficient calculation: K = 2

F_0 Normalization

F_0 frequency [Hz] convert [cent]

$$1200 \times \log_2 \frac{F_0}{440 \times 2^{3/12-5}}$$

F_0 value is limited to (0, 100) in [cent]

$$\mathrm{mod}(F_0 + 50, 100)$$

Discrimination Experiment

◆Discrimination of 3 singer classes based on **MAP**

$$\hat{s} = \arg\max_s \left[p(s | \{F_0, \Delta F_0, \Delta\Delta F_0\}) \right]$$
$$= \arg\max_s \left[\frac{1}{N} \sum_{n=1}^{N} \log p(\mathbf{f}_0(n) | \Theta_s) + \log p(s) \right]$$

Professional Classical / Professional Pop / Amateur

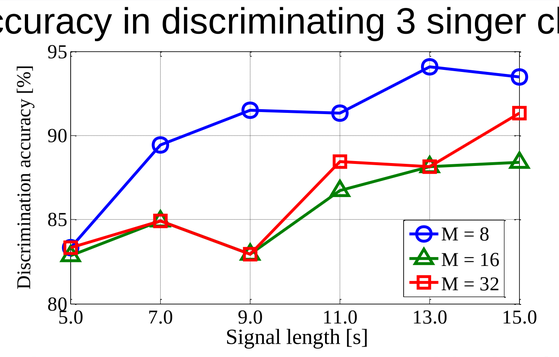
s: Singer class ID
Θ_s: Model parameters of s^{th}
N: Length of the signal

◆Training : "*Twinkle-Twinkle, Little Star*" and 5 etudes

◆Testing : "*Ode to Joy*"

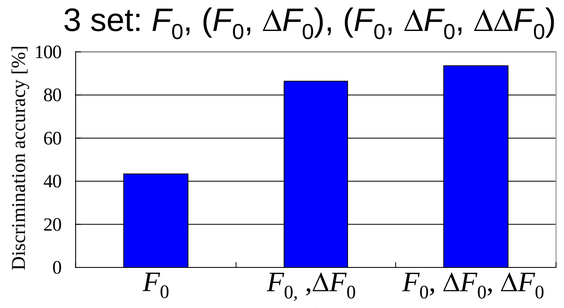
Result

Accuracy in discriminating 3 singer classes



Comparing accuracy

3 set: F_0, (F_0, ΔF_0), (F_0, ΔF_0, $\Delta\Delta F_0$)

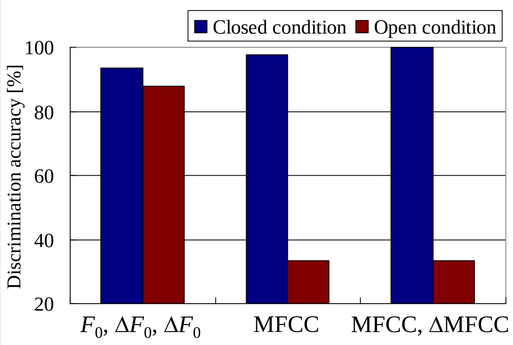


◆The accuracy increases with the length of the test signal.

◆The best performance of 94 % (13 [s], 8-mixture GMM)

◆Using (F_0, ΔF_0, $\Delta\Delta F_0$) is the best accuracy.

} SPP effectively characterizes singing style



Comparing SPP with MFCC under two conditions

◆Closed: Training ⟶ 6 singers (male and female)
Testing ⟶ Same singers

◆Open: Training ⟶ Female singer data
Testing ⟶ Male singer data

Op⟶ MFCC-GMM 33.3% SPP-GMM **87.9**%

SPP can classify new singer's data

Summary and future works

◆Singing style is represented as a phase plane trajectory.

◆Model the F_0 trajectory on the phase plane with a GMM.

◆More than 90% accuracy can be achieved in discriminating the 3 classes.

◆Increasing the number of singers and singer classes is critical future work.