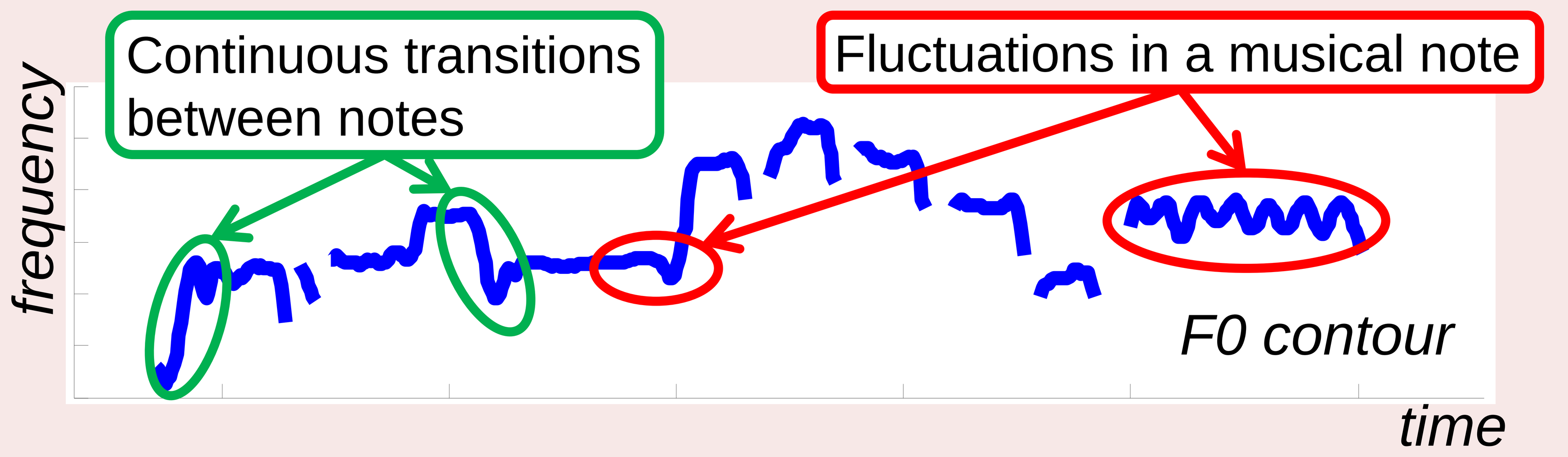
A Stochastic Representation of the Dynamics of Sung Melody

Yasunori OHISHI[1], Masataka GOTO[2], Katunobu ITOU[3], Kazuya TAKEDA[1]

[1]Nagoya University, [2] National Institute of Advanced Industrial Sci. and Tech. (AIST), [3]Hosei University

Introduction

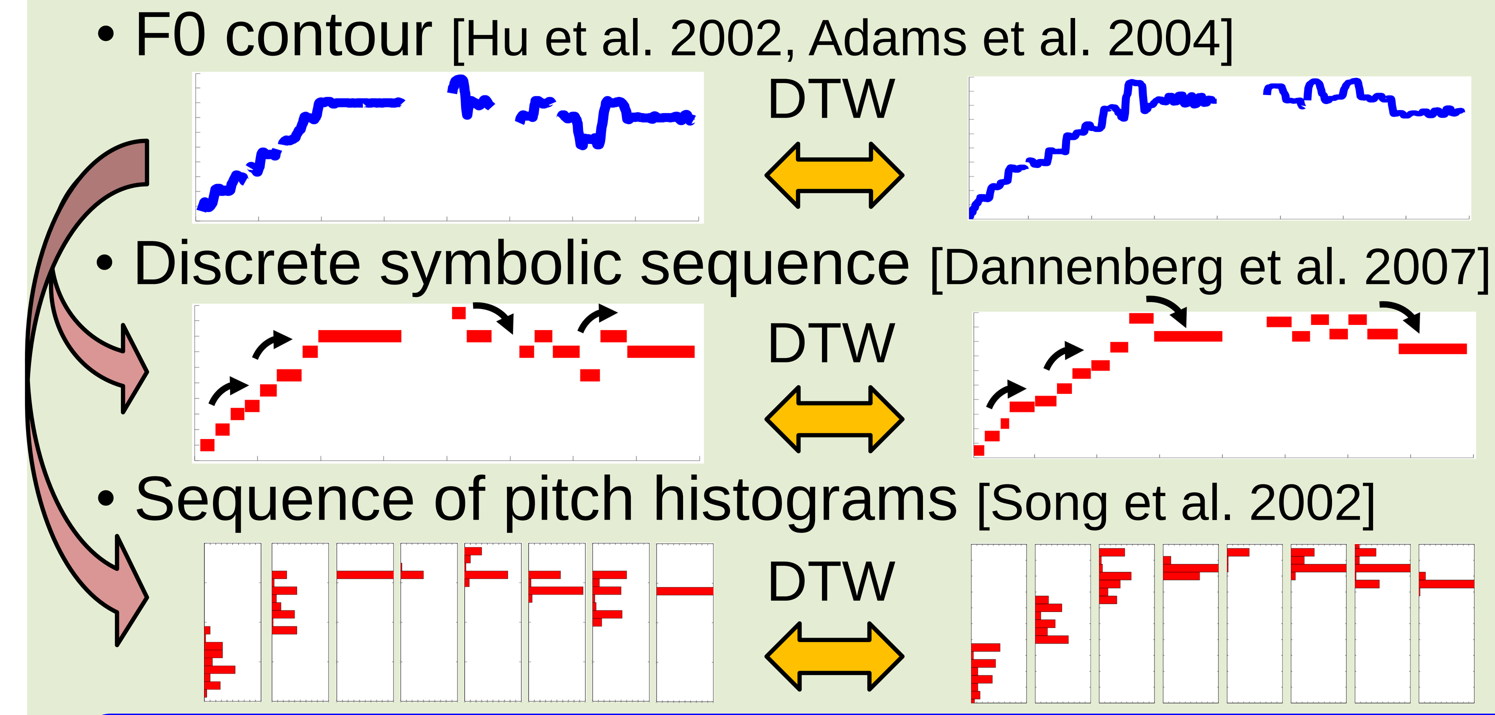
- Our goal
  - Build a model that can represent the dynamics of various singing behaviors in a sung melodic contour



- 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

Previous Work

- Melodic similarity measures for sung melodies
  - F0 contour [Hu et al. 2002, Adams et al. 2004]
  - Discrete symbolic sequence [Dannenberg et al. 2007]
  - Sequence of pitch histograms [Song et al. 2002]



These similarities are sometimes too sensitive to singing behaviors that may differ among singers

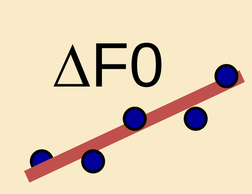
Phase Representation for Melodic Contour

- 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

$$F0[t]_{cent} = 1200\ \log_2(F0[t]_{Hz} / 440 \times 2^{\frac{3}{12}-5})$$

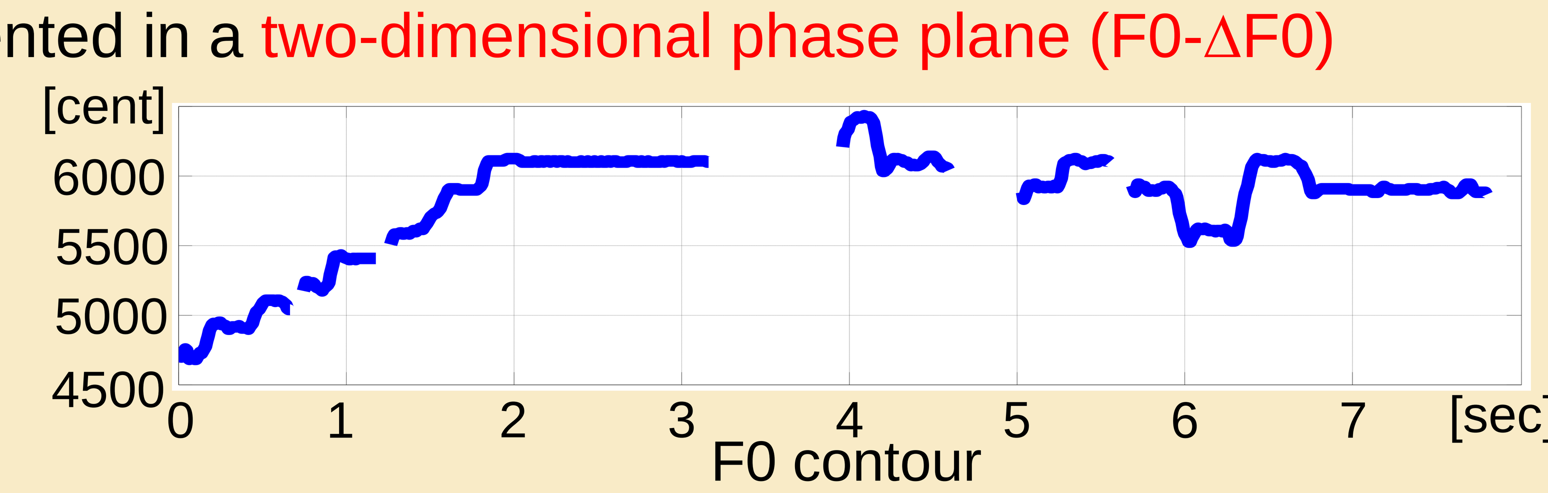
- Calculate F0 derivative (ΔF0)

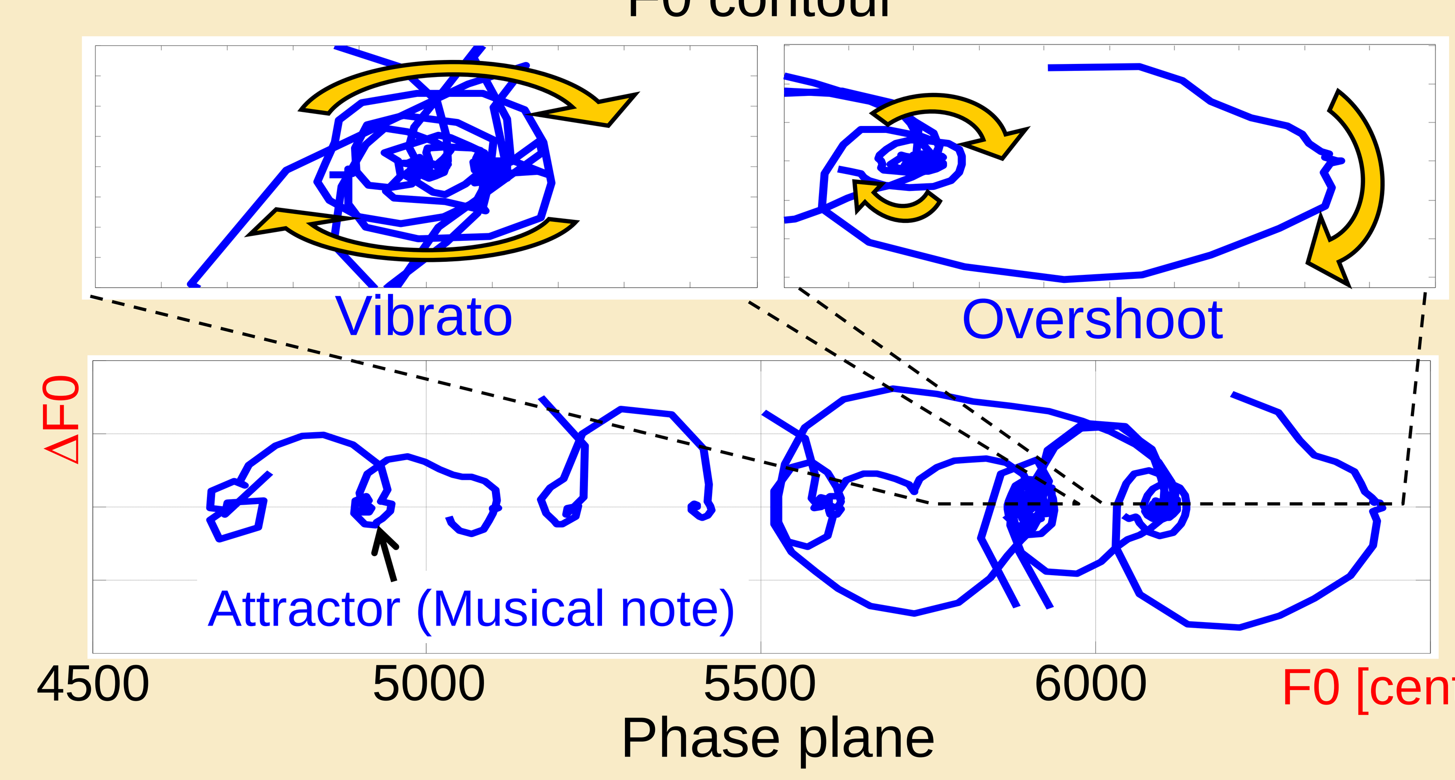
$$\Delta F0[t]_{cent} = \sum_{k=-2}^{k=2} k \cdot F0[t+k]_{cent} \Big/ \sum_{k=-2}^{k=2} k^2$$



- 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 ⇨ Spiral 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

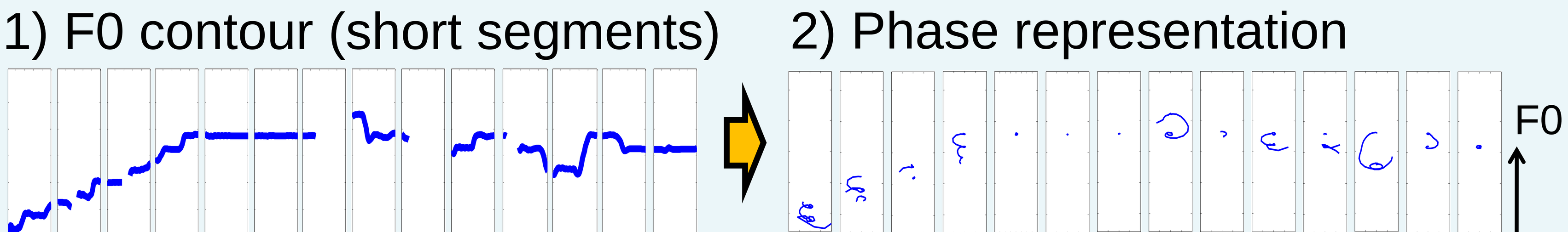




Melodic similarity measure using the phase representation

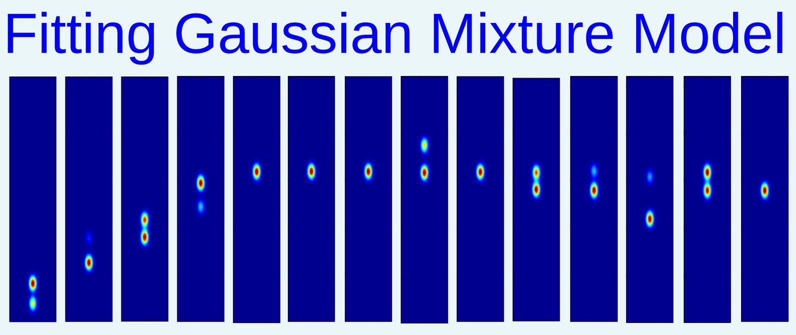
- Stochastic phase representation (SPR)
  - Focus on the original (target) melodic information
  - ⇨ Extract attractor points

1) F0 contour (short segments) 2) Phase representation

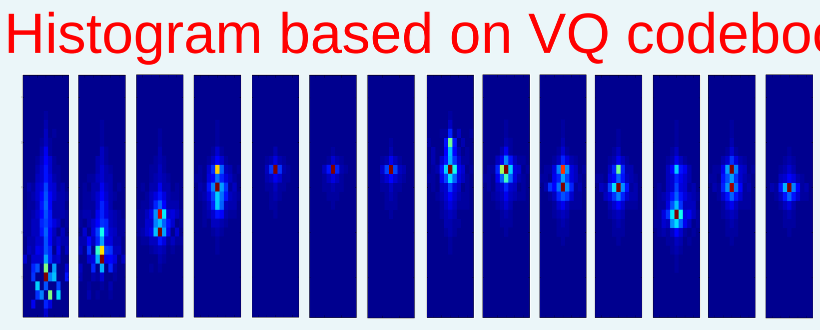


3) Train the F0-ΔF0 distribution

Fitting Gaussian Mixture Model



Histogram based on VQ codebook



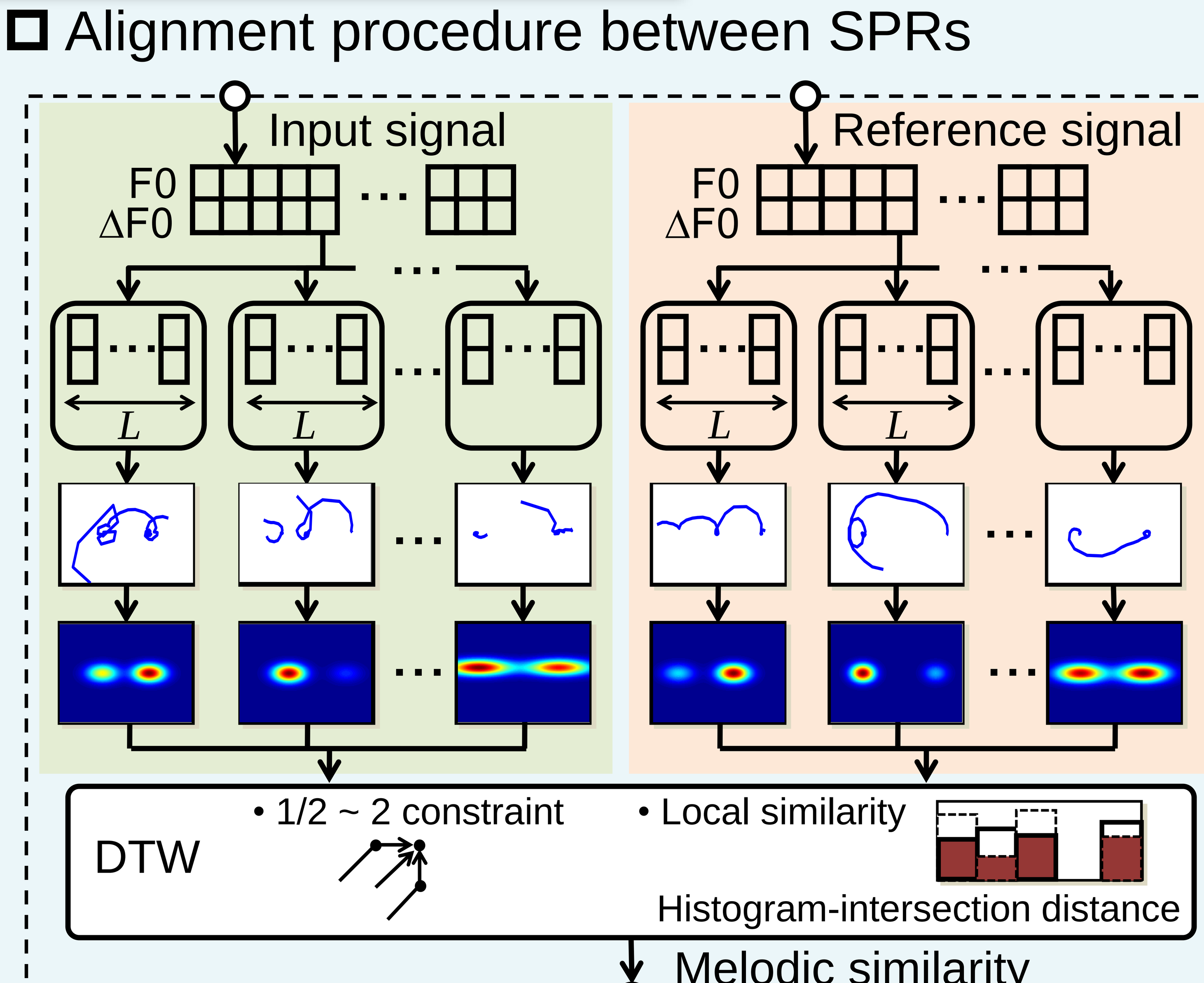
2-Mixture GMM
EM algorithm
- Gaussian weight w
- Gaussian mean μ
- Gaussian variance (diagonal) Σ

VQ codebook $\mathbf{V} = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_{512}\}$

Probability $\mathbf{P} = \{p_1, p_2, \ldots, p_{512}\}$

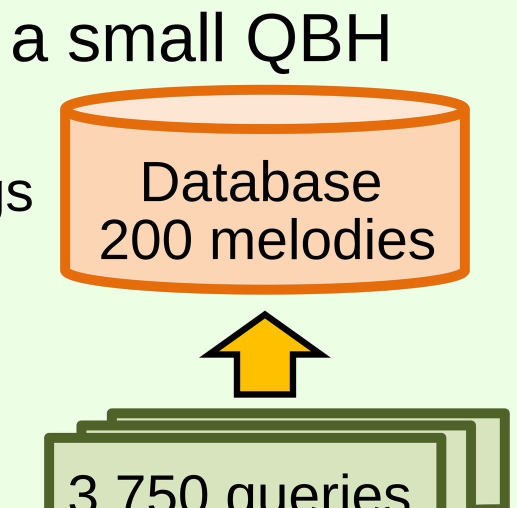
$$w_{lm} = \frac{1/\|\mathbf{x}_l - \mathbf{v}_m\|_2}{\sum_{n=1}^{M} 1/\|\mathbf{x}_l - \mathbf{v}_n\|_2} \qquad p_m = \frac{\sum_{l=1}^{L} w_{lm}}{L}$$

- Alignment procedure between SPRs



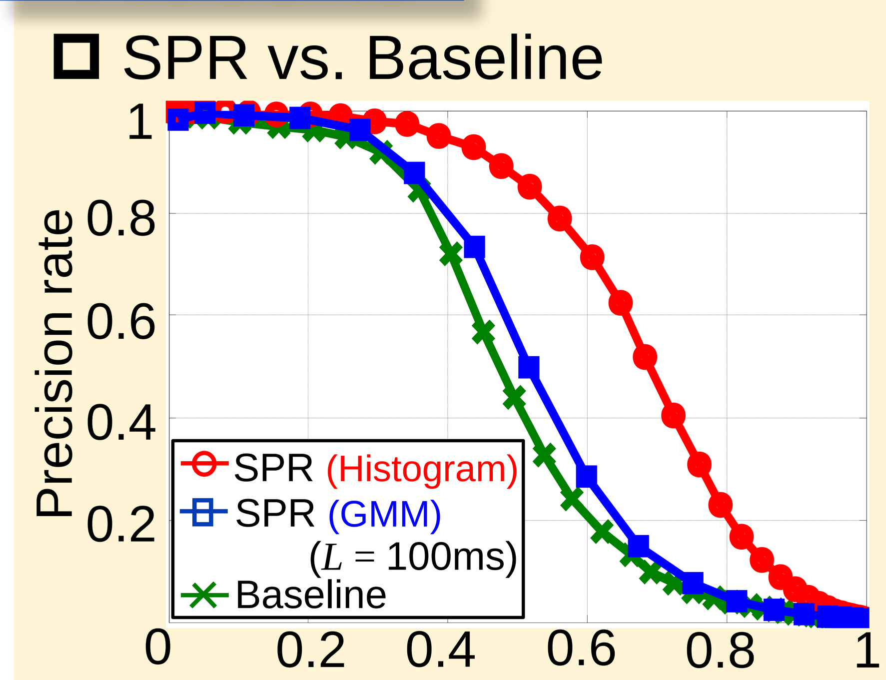
Experiments

- 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

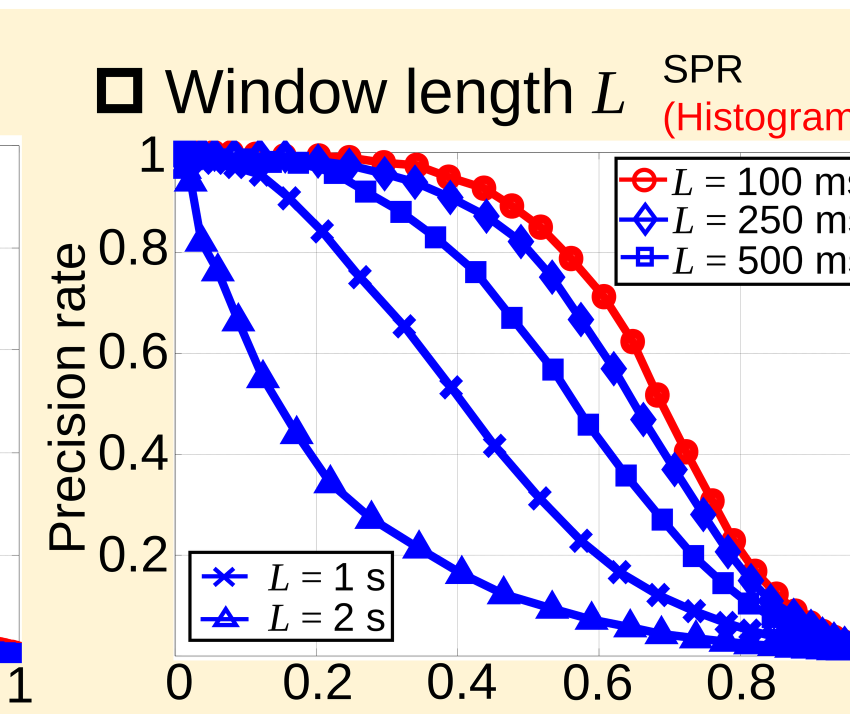


Results

- SPR vs. Baseline



- Window length L



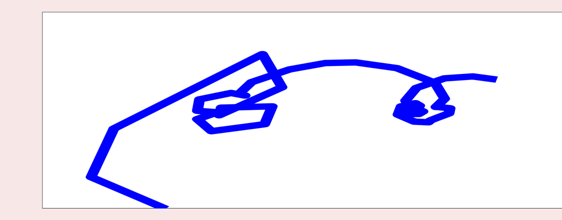
	Baseline	SPR (GMM)	SPR (Histogram)
Mean Reciprocal Rank	0.69	0.65	0.80

* Baseline performance : Traditional DTW using F0 contours

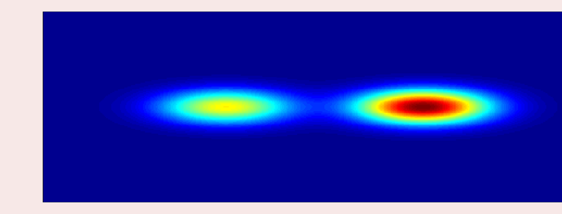
Conclusion

- **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



Phase representation



Stochastic
phase representation