

楽譜情報を用いない歌唱力自動評価手法

松村晴琉

Haru Matsumura

u320062@st.pu-toyama.ac.jp

富山県立大学 工学部 情報システム工学科

14:50-18:00, Monday, December 5, 2025

N516, Toyama Prefectural University

1. 概要と研究課題

2/10

研究目的

- **楽譜情報が不要**な歌唱スキルの自動評価手法の提案.
- 評価を「うまい (good) / へた (poor)」の2クラス分類として行う.

従来研究の課題

- 歌声の特徴分析が主で, 自動評価への直接応用や, 人間による評価との定量的な比較が行われていなかった.

提案アプローチ (歌い手/メロディ非依存な特徴量)

- ピッチ間隔の正確さ (Pitch Interval Accuracy)
- ビブラート (Vibrato)

2. 評価指標の安定性検証

3/10

主観評価実験

- 目的: 人間の歌唱スキル評価が, 評価者間で安定しているかを検証.
- 方法: 22 人の被験者による歌唱サンプルの**相対評価**を実施.

評価指標: スピアマンの順位相関係数

- 評価者間の順位付けの一致度を定量化.

$$\rho = 1 - \frac{6}{N^3 - N} \sum_{i=1}^N (a_i - b_i)^2 \quad (\text{式 1})$$

- N : サンプル数, a_i, b_i : 評価者 a, b によるサンプル i の順位.

結果

- 評価者ペアの **88.9%**で、 $p < .05$ の有意な相関を確認.

3. 特徴量 1: ピッチ間隔の正確さ (I)

4/10

半音安定度 (Semitone Stability) の計算原理

- 基本周波数が、半音階の音高のどの位置に長く留まっているかを測定.
- うまい歌唱ほど、ピッチが音高の中心に集中する傾向にある.

基本周波数 (F_0) の確率密度関数 $p(x; F)$

- F_0 のセント値 x が、半音階内の位置 F からのずれを示す度合いをガウス分布でモデル化.

$$p(x; F) = \sum_{i=0}^{\infty} \frac{\omega_i}{\sqrt{2\pi}\sigma_i} \exp\left\{-\frac{(x - F - 100i)^2}{2\sigma_i^2}\right\} \quad (\text{式 2})$$

4. 特徴量 1: ピッチ間隔の正確さ (II)

5/10

半音安定度 $P_g(F, t)$ の計算

- $p(F_0(\tau); F)$ に F_0 の信頼度 $P_{F_0}(\tau)$ を乗じ, 積分時間 T で平均化.

$$P_g(F, t) = \int_{t-T}^t p(F_0(\tau); F) P_{F_0}(\tau) d\tau \quad (\text{式 3})$$

最終特徴量：集中度 (M)

- $P_g(F, t)$ の長期平均 $g(F)$ のピーク F_g 周辺での集中度を測る.
- M の値が小さいほど, 歌唱のピッチ間隔が正確.

$$M = \int_{F_g-50}^{F_g+50} (F_g - F)^2 g(F) dF \quad (\text{式 4})$$

5. 特徴量 2: ビブラートの検出

6/10

ビブラートらしさ $P_v(t)$ の計算

- F_0 の短期フーリエ変換 (STFT) から, ビブラートの周波数範囲 (約 5~8 Hz) の成分を解析.
- パワー成分 $S_v(t)$ とスペクトル変化 $\Psi_v(t)$ の積でビブラートらしさを定義.

$$P_v(t) = S_v(t)\Psi_v(t) \text{ (式 5)}$$

ビブラート区間の特定

- $P_v(t)$ が閾値を超え, かつ F_0 の変動が連続的にビブラートの特性を満たす区間を抽出.

6. 特徴量 2: ビブラートのパラメータ

7/10

抽出されるビブレート特徴量

- 検出されたビブレート区間から, その**周期 R_n ** と **振幅 E_n ** を測定し, 平均化する.

- 1 レート (**rate**): ビブラートの周波数 (Hz) .

$$\text{rate} = \frac{1}{\frac{1}{N} \cdot \sum_{n=1}^N R_n}$$

- 2 エクステント (**extent**): ビブラートの振幅 (セント単位) .

$$\text{extent} = \frac{1}{2N} \cdot \sum_{n=1}^N E_n$$

- うまい歌唱は, レートが一定で, エクステントが適度な範囲にあることが期待される.

7. 自動分類実験の設定

8/10

実験概要

- データセット: 600 の歌唱サンプル (good/poor 各 300) .
- 分類器: サポートベクターマシン (SVM).
- 特徴量: M と rate, extent を組み合わせたもの.

実験設定 (汎用性評価)

- **Closed 実験:** 歌手が学習データと評価データに重複.
- **Open 実験:** 歌手が学習データと評価データで**重複しない**. (汎用性を評価)

評価指標: 分類率 (C)

- good/poor クラスそれぞれの再現率 (R_{good}, R_{poor}) の平均.

$$C = \frac{R_{good} + R_{poor}}{2} \quad (\text{式 6})$$

8. 実験結果と考察

9/10

表 1: 歌唱スキル自動分類結果

実験設定	特徴量	分類率 (C)	特記事項
Closed	M のみ	78.8%	歌い手情報込み
	$M + \text{ビブラート}$	86.2%	
Open	M のみ	77.2%	歌い手情報なし
	$M + \text{ビブラート}$	83.5%	

- **ビブラート特徴量**を追加することで, 分類率が大幅に向上.
- **Open 実験**での 83.5%は、歌い手やメロディに依存せず, 提案特徴量が有効であることを強く示唆.

9. 結論と今後の展望

10/10

結論

- 楽譜情報が不要な歌唱スキル自動評価手法として、**ピッチ間隔の正確さとビブラート**を用いる手法を提案.
- 人間の主観評価の安定性を確認し、自動評価モデルの目標設定が妥当であることを確認.
- 汎用的な Open 実験で **83.5%**の分類率を達成し、本手法の有効性を実証した.

今後の展望

- 2 クラス分類 (good/poor) だけでなく、より****多段階****の評価レベルへの拡張.
- 歌唱時の母音や子音の発声など、更なる音響特徴量の導入.