3R-9

歌声 F0 生成過程とメロディ分離手法に基づく楽譜逸脱成分推定

池宮 由楽 † 阪上 大地 ‡ 糸山 克寿 ‡ 奥乃 博 ‡

† 京都大学 工学部情報学科 ‡ 京都大学 大学院情報学研究科 知能情報学専攻

1. はじめに

人間が歌を歌うとき，その基本周波数 (F0) 軌跡は楽譜通りではなく，ビブラート・オーバーシュート・ポルタメント・微細変動といった成分を含んでいる．このような**楽譜逸脱成分**[1] は歌唱者の個人性と深く関わっている．つまり，楽譜逸脱成分を CD 楽曲から抽出することができれば，様々な歌唱者の個人性をボーカロイドや MIDI 楽曲に転写したり，歌手の歌い方に基づく楽曲検索を行うことができるようになる．本稿では，伴奏付きの歌声から F0 の楽譜逸脱成分を抽出する方法を述べる．

大石らは歌声 F0 生成過程を導入し，楽譜逸脱成分を推定した [1]．この手法の問題点は，伴奏を含む歌声に適用すると F0 を正確に抽出できず楽譜逸脱成分を推定できないことである．したがって分析対象が伴奏のない歌唱に限られてしまう．本研究の目的は，伴奏付きの歌唱であっても，伴奏音に影響されず歌声の F0 が高精度に抽出できる技法の開発である．伴奏音中の F0 抽出は，藤原ら [2] が歌詞と歌声との時間的対応付けで行っているが，F0 抽出があくまで Sinusoid 波の再合成のためであり，楽譜逸脱成分の抽出に十分な精度は要求されない．

本手法ではまずメロディ分離手法により歌声を伴奏から分離した後，歌声 F0 生成過程に基づき入力音高列と F0 軌跡のおおまかなアライメントを計算する．次に，音高列のアライメント結果から各時刻での F0 存在範囲を絞り込み，再度歌声 F0 推定を行う．最後に，歌声 F0 生成過程に基づいて楽譜逸脱成分の推定を行う．

2. 問題設定

本稿で扱う問題は以下のように要約できる．

入力 伴奏付き歌声，歌声に対応する楽譜の音高列．

出力 歌声 F0 軌跡 $\mathbf{y}$ の楽譜成分を表す階段状の信号 $\mathbf{u}$，オーバーシュートやポルタメントといった音符が別の音符に移る時の楽譜逸脱成分を含む信号 $\mathbf{y}_t$，ビブラートや微細変動といった音符の音高が安定するときの楽譜逸脱成分 $\mathbf{y}_s$．ここで，$\mathbf{y} = \mathbf{y}_t + \mathbf{y}_s$ である．

前提 歌唱者は 1 人で，歌声の有声区間は既知．

オーバーシュートは音高の変化直後に目標音高を超過する瞬時的変動，ポルタメントは音符間の滑らかな音高変化，ビブラートは同一音高中の周期的な揺れ，微細変動は F0 全体で観測される不規則で細かい変動を意味する．

3. 楽曲からの歌声楽譜逸脱成分推定

提案法のアルゴリズムは以下の通りである．

1) REPET-SIM により歌声（メロディ）と伴奏を分離．
2) Subharmonic Summation (SHS) で F0 を抽出し，ノイズ除去して F0 概形を計算．
3) 歌声 F0 生成過程を用いて，入力音高列と 2) で推定された F0 概形の時間的なアライメントを計算．
4) アライメントに基づく時間周波数的な制約条件のもとで，SHS で再度 F0 を抽出．
5) 歌声 F0 生成過程を用いて，入力音高列と 4) で推定された F0 から楽譜逸脱成分を推定．

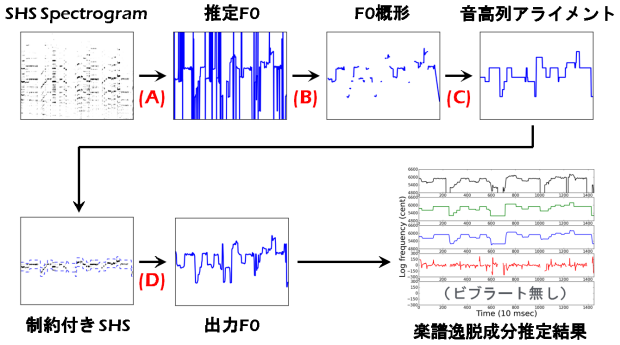


図 1: F0 推定プロセスと楽譜逸脱成分推定

3.1 メロディ（歌声）分離

歌声と伴奏の分離はメロディ抽出アルゴリズム (REPET-SIM) [3] を使用する．REPET-SIM は楽曲中に繰り返し現れる音を伴奏音とみなして除去し，楽曲からメロディを分離する手法である．伴奏が繰り返し構造を持つ限り，REPET-SIM で歌声が抽出可能である．

3.2 歌声 F0 生成過程

F0 抽出の高精度化は大石らの歌声 F0 生成過程を使用するので，本節で先に説明を行う．$\mathbf{y}_t$ は階段状の信号 $\mathbf{u}$ に 2 次伝達関数 $\mathscr{H}(s) = \Omega^2/(s^2 + 2\zeta\Omega s + \Omega^2)$ のインパルス応答 $h(t)$ を畳み込むことで表現する．ここで s は時間微分演算子，ζ は減衰率，Ω は固有周波数である．$\mathbf{y}_s$ はガウス性白色雑音であると仮定する ($\mathbf{y}_s \sim \mathcal{N}(\mathbf{0}, \sigma_s^2 \mathbf{I}_K)$)．$\sigma_s^2$ は正規分布の分散，$\mathbf{I}_K$ は $K \times K$ の単位行列である．

$\mathbf{u}$ は Left-to-Right 型 HMM でモデル化する．各状態は旋律中の個々の音符に対応しており，F0 系列とそれに対応する楽譜中の音高列の時間的なアライメントはビタビ探索で効率的に解くことができる．推定すべきパラメータは $\theta_u := \{\{s_k\}_{k=1}^K, d, \sigma_u^2\}$ となる．ここで s_k は HMM の状態系列，σ_u^2 は状態出力分布の分散，d は歌唱音高の楽譜からのずれを表す．

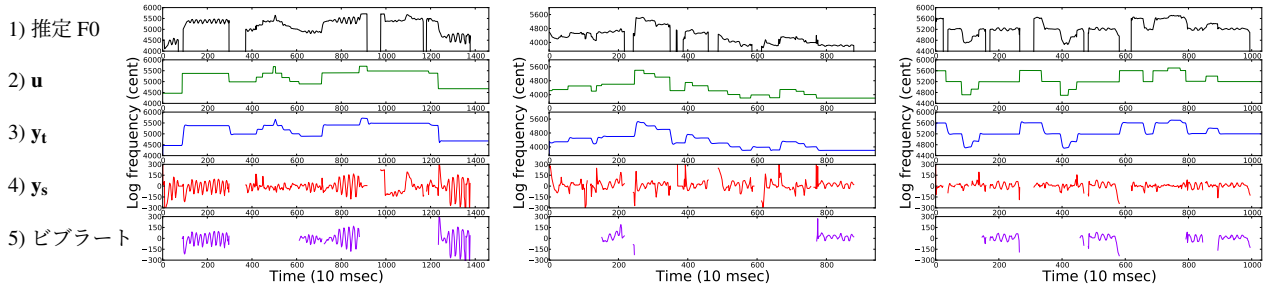
楽譜逸脱成分推定は，パラメータ $\Theta := \{\theta_u, \zeta, \Omega, \sigma_s^2\}$ を推定する問題となる．歌声 F0 は有声区間のみで観測されるため無声区間に対しては実測 F0 が存在しない確率モデルとなる．これは不完全データ問題に対する EM アルゴリズムにより解くことができる．

3.3 歌声 F0 抽出

雑音を含むスペクトログラムから歌声 F0 軌跡を取り出すため，Subharmonic Summation (SHS) [4] を用いる．入力音声を定 Q 変換して得られる対数周波数スペクトログラムから F0 を抽出する．

歌声 F0 生成過程を用いて SHS に音高列による制約を加える手順を 図 1 に示す．SHS Spectrogram とは足し合わせ後のスペクトログラムを表す．(A) REPET-SIM を適用した音響信号に対して SHS を行うとノイズを含む歌声 F0 が推定．(B) 推定 F0 に対して以下の処理を行い，歌声 F0 概形を得る．

1. メジアンフィルタを適用．



(a) 美空ひばり 「川の流れのように」　(b) ブルーハーツ 「リンダリンダ」　(c) スピッツ 「クリスピー」

図 2: 3 種類の楽曲に対する歌声楽譜逸脱成分推定の例．一番上は 1) 推定 F0，中央 3 つは 1) に対する楽譜逸脱成分推定結果でそれぞれ 2) **u**，3) $\mathbf{y}_t$，4) $\mathbf{y}_s$，一番下は 5) 抽出したビブラートを表示している．

表 1: F0 推定精度 (%)

楽曲	P-002	P-007	P-011	P-012	P-014	P-016	P-018
精度	85.65	92.41	82.24	85.20	90.00	85.63	91.41

表 2: 2 次系パラメータの誤差の累積度数 (%)

楽曲	∼5(%)	∼10	∼15	∼20
P002	32.00	48.00	57.67	69.33
P007	22.11	41.06	52.38	64.49
P011	23.85	34.23	43.85	50.77
P012	23.85	35.77	45.39	54.62
P014	34.17	49.73	63.34	71.28
P016	37.81	48.75	56.56	64.06
P018	29.13	48.91	61.74	68.91

Ω

楽曲	∼5(%)	∼10	∼15	∼20	∼25	∼30
P002	20.33	33.00	41.67	49.33	54.67	60.00
P007	12.11	23.69	33.43	39.48	45.26	49.74
P011	15.77	26.54	33.85	40.00	45.00	53.08
P012	11.92	22.69	30.38	35.76	40.77	46.15
P014	22.50	34.44	42.22	49.44	53.89	58.06
P016	22.19	31.25	41.56	47.50	53.75	58.13
P018	16.52	30.65	40.87	48.26	51.96	58.26

ζ

2. 入力音高列の最低音より 200 cent (二半音) 以上低い値と最高音から 200 cent 以上高い値を除去．
3. 1 秒に 1800 cent 以上の変化率で音高差が 200 cent 以上の部分を除去．
4. F0 軌跡中の素片のうち，長さが 100 ms 以下で前後の素片の音高が 350 cent 以上離れているもの，長さが 150 ms 以下で前後の素片の音高が 500 cent 以上離れているものを除去．

(C) 歌声 F0 生成過程を用い，F0 概形と音高列との時間的なアライメント（音高列アライメント）を推定する．
(D) SHS Spectrogram の各時間フレームで音高列アライメントから ±400 cent の範囲で argmax を計算することで歌声 F0 推定の精度を向上．手順中の定数は実験的に決定した．

4. 評価実験

4.1 実験条件

入力として楽曲から無声区間を手作業で取り除いたものを使用した．また，歌声 F0 生成過程のモデルの性質から，入力音高数が増えるほどアライメント推定が難しくなるので，入力データを 10 秒程度とした．

4.2 定量評価

実験には RWC Music Database [5] のポピュラー 7 曲を歌声音高数が 20 となるように分割した計 117 データを使用した．正解 F0 には AIST アノテーション [6] を使用し，2 次系パラメータは，正解 F0 からの [1] による推定結果を正解データとした．本手法による F0 推定精度と，推定された 2 次系パラメータ Ω，ζ の誤差累積度数をそれぞれ表 1，2 に示す．F0 推定 では正解 F0 から 50 cent の範囲を正解とし，Ω，ζ の誤差の度数は正解データ値に対する割合を表す．Ω，ζ の全正解データの標準偏差はそれぞれ平均値の約 25 %，36 % であった．実験の結果，F0 推定は約 80∼90 % の精度を実現し，Ω，ζ の誤差がそれぞれ 20 %，30 %を下回るものは 50∼70 %，50∼60 % 程度となった．全体の標準偏差の八割程度を下回るデータが半数を超えており，本手法による個人の歌唱特性の分析・転写の可能性を示唆している．現在，推定精度の上限は音高列アライメントの精度によって決まっているため，雑音に頑健なアライメント手法を開発し，楽譜逸脱成分の高精度な推定を目指したい．

4.3 市販楽曲に対する楽譜逸脱成分推定

本手法の定性的評価を，(a) 美空ひばり「川の流れのように」，(b) ブルーハーツ「リンダリンダ」，(c) スピッツ「クリスピー」を対象として行った．それぞれサビ部・B メロ部・サビ部の 10 秒程度と，歌声に対応する楽譜の音高列を使用した．図 2 に各曲の F0 推定結果，それらに対する楽譜逸脱成分推定結果，ビブラート区間の推定結果 [7] を示す．3 曲とも楽譜逸脱成分を含めて安定に F0 が推定できており，本手法により市販の楽曲から高精度に F0 が推定でき，楽譜逸脱成分も抽出できることが分かった．

5. おわりに

本稿では，楽曲中から歌声楽譜逸脱成分を推定する手法について述べた．本手法では，メロディ分離手法を用いて歌声を分離し，SHS と歌声 F0 生成過程モデルを利用して伴奏音に頑健な推定を実現した．評価実験では本手法によって十分な歌声 F0 推定精度を実現し，市販の楽曲から楽譜逸脱成分を抽出できることを確認した．本稿では有声区間を手動で取り除いているため，短い無声区間を多く含む楽曲では手間が多くなる．したがって，今後は自動有声区間検出にも取り組みたい．なお，本研究の一部は科研費 24220006, 24700168 の支援を受けた．

参考文献

[1] 大石康智，亀岡弘和，持橋大地，永野秀尚，柏野邦夫 : “歌声 F0 系列からの楽譜逸脱成分の抽出 -動特性モデルに基づく楽譜との時間的対応付け-”, 日本音響学会 2011 年秋季研究発表会, 1-8-19, pp.279–282, Sep. 2011.

[2] H. Fujihara, M. Goto, J. Ogata, K. Komatani, T. Ogata and H. G. Okuno: “Automatic synchronization between lyrics and music CD recordings based on Viterbi alignment of segregated vocal signals”, *ISM 2006*, pp.257–264.

[3] Zafar Rafii and Bryan Pardo: “Music/Voice Separation using the Similarity Matrix,”, *ISMIR 2012*, pp.583–588.

[4] Dik J. Hermes: “Measurement of pitch by subharmonic summation”, *J. Acoust. Soc. Am.* 83, 257–264, 1988.

[5] M. Goto, H. Hashiguchi, T. Nishimura and R. Oka: “RWC Music Database: Popular, Classical, and Jazz Music Databases”, *ISMIR 2002*, pp. 287–288.

[6] Masataka Goto: “AIST Annotation for the RWC Music Database”, *ISMIR 2006*, pp.359-360.

[7] 中野倫靖，後藤真孝，平賀譲: “楽譜情報を用いない歌唱力自動評価手法”, 情報処理学会論文誌, Vol.48, No.1, pp.227–236, 2007.