

MiruSinger:

歌を「歌って/聴いて/描いて」見る歌唱力向上支援インタフェース

中野 倫靖[†]

後藤 真孝^{††}

平賀 譲[†]

MiruSinger: A Singing Skill Visualization Interface

for Singing, Listening, and Drawing Songs

TOMOYASU NAKANO[†]

MASATAKA GOTO^{††}

YUZURU HIRAGA[†]

1 はじめに

歌唱力を向上させたり、個性的で魅力溢れる歌い方を
実現させたりするためには、優れた歌唱者や自分がどの
ように歌っているのかを理解できることが望ましい。例
えば、歌唱から音符列をイメージできること、また、各
音をどのように歌っているかを認識・理解できることな
どである。音符への変換は、歌唱や楽器演奏の経験が少
なく、音楽的な訓練を受けたことがない人には困難だが、
各音の歌われ方は比較的容易に認識できると考えられる。
楽譜がイメージできなくてもカラオケで歌うことはでき
る人がこれに相当する。しかし、大まかな音の流れは認
識できても、その詳細については意識していない、もし
くは気付かないことも多い。例えば、ピブラート(音高
の周期的・規則的な変動)のある歌唱を聴いて心地良い
と感じても、「ピブラートがあった」ことはあまり意識し
ていない場合などである。このような意識していなかつ
た歌唱の特徴(歌い方)を認識・理解できることは、歌唱
力の向上や魅力的な歌い方の実現につながると思う。

そこで本研究では、ユーザが歌を歌ったり聴いたりす
るだけでなく、「見る」歌唱力向上支援インタフェース
MiruSinger を提案する。歌を「見る」従来のシステム
としては、鼻歌(ハミング)を音符列に変換する楽譜作
成ソフトウェア、カラオケの自動採点システムにおける
歌唱の点数化やピブラート検出[1]、「調子外れ」の治療
支援システムにおける音高(F_0)可視化[2]などがある。
MiruSinger は、歌唱の音高を軌跡として可視化(F_0 軌
跡)し、歌唱への理解を支援する。さらに、これまで考
えられていなかった「描く」インタラクションを導入す
ることで、音楽 CD からのボーカル抽出の精度向上と F_0
軌跡への理解をより深めることを考える。

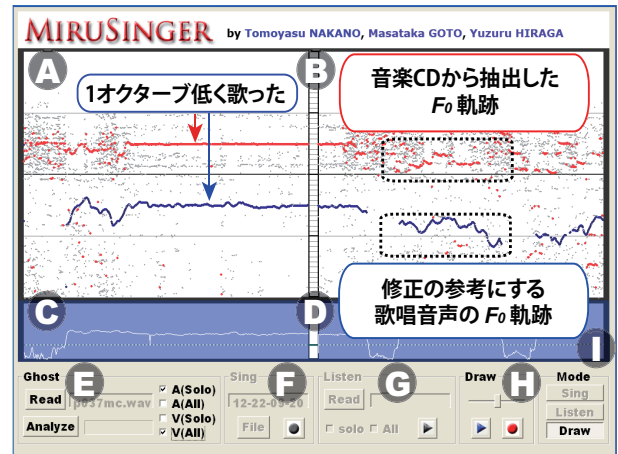


図 1: MiruSinger の表示画面例(描いて見るモード)

2 MiruSinger の機能

MiruSinger は、歌唱の F_0 軌跡を見ながら歌ったり
(歌って見るモード)、聴いたり(聴いて見るモード)、描
いたり(描いて見るモード)できる。これら各機能は、表
示画面例(図1)の右下①の“Sing”, “Listen”, “Draw” の
ボタンを押すことで、それぞれ呼び出せる。

図1上部は、以下の2つのウィンドウで構成される。

- 楽譜ウィンドウ(図1: A)

中央縦長のバー(B)が現在時刻を表し、入力された
音声や収録音声の F_0 がその右から左へと流れてゆく。
縦軸は音高で、軌跡が上にあるほど高い音となる。

- パワーウィンドウ(図1: C)

中央縦長の白いバー(D)の中に、入力された声の大
きさ(パワー)が着色して表示される。時間とともに左
へと流れてゆき、発音量の適切さをチェックできる。

さらに、音高可視化に加えて、収録済みの歌唱(過去
の自分や音楽 CD に含まれるボーカル)と、現在の自分
のどこが違うのかを比較するために、それぞれの歌唱と
その F_0 軌跡を同時に提示できる機能を実装した。本論
文では、それぞれ聴覚ゴースト・視覚ゴーストと呼ぶ。

以下、各機能と利用可能なゴーストについて述べる。

[†]筑波大学 図書館情報メディア研究科
Graduate School of Library, Information and Media
Studies, University of Tsukuba

^{††}産業技術総合研究所
National Institute of Advanced Industrial Science and
Technology (AIST)

2.1 歌って見るモード

ユーザがマイクから入力した歌唱の F_0 軌跡を、リアルタイムに描画する機能である。図 1㉔の“Sing”部から、録音ボタン を押して録音を開始する。ビブラートが検出された場合は、その区間が着色されて提示される。このように、 F_0 をリアルタイムに描画することで、歌い方がおかしかった瞬間にその部分を認識できる。

ゴーストには、聴覚・視覚ゴーストともに利用可能で、図 1 左下㉕の、“Ghost”部から読込 (“Read”)、解析 (“Analyze”) ¹、選択 (“A(solo)” / “A(All)” / “V(solo)”) できる。“A”は聴覚ゴーストで、“A(solo)”は推定された F_0 とその高調波構造から再合成した歌唱、“A(All)”は解析対象そのままの歌唱 (や音楽 CD) である。また、“V”は視覚ゴーストを表し、“V(solo)”は F_0 軌跡である。

2.2 聴いて見るモード

収録済みの歌唱を聴きながら F_0 軌跡を見る機能であり、図 1㉖の“Listen”部の読込 (“Read”) ボタンで対象を選択してから、“solo”(再合成歌唱)か“All”(そのままの歌唱 / 音楽 CD)のどちらかで再生 する。聴覚・視覚ゴーストともに利用可能であり (“A(solo)” / “A(All)” / “V(solo)”)、聴覚ゴーストのみを聴くこともできる。

歌唱力評価において、複数人が歌った歌唱を相対的に評価 (順位付け) すると、聴取者が原曲を知らず、メロディ・歌唱者の性別・歌詞の言語が一定の条件下では聴取者間の評価には高い相関があることが明らかになっている [3]。すなわち、そのようなデータセットを用意して聴いて見れば、良い歌い方が分かる可能性がある。

2.3 描いて見るモード

音楽 CD に含まれるボーカルの F_0 推定結果をマウス操作で修正できる機能である。ボーカル以外に複数音源が含まれる音楽 CD からボーカルの F_0 を自動推定することは (ボーカルが含まれているかどうかの判断を含めて) 非常に困難であるが、歌ったり聴いたりしながら見て描くことで、ボーカルの抽出精度を向上させる。ここでは、これまでゴーストであった図 1㉔の“Ghost”部が修正対象となる。“V(All)”を選択すると、楽譜ウィンドウ (A) に、 F_0 候補が全て灰色で描画され、現在選ばれている F_0 は赤で描画される (図 1)。“V(solo)”を選択した場合は、現在選ばれている F_0 のみが描画される。

図 1 右下㉗の“Draw”部にあるスライダで修正したい

¹現在の実装では、ゴーストの解析 (F_0 抽出、全 F_0 候補の高調波構造抽出、再合成) は、オフラインで処理している。

箇所を探し、修正後の F_0 軌跡をマウスの左ドラッグで描くことで、その付近で優勢な F_0 がボーカルの F_0 として再選択される。ボーカルが含まれない区間は、マウスの右ドラッグで F_0 軌跡を削除できる。録音ボタン で修正を反映したのち再生 して確認できる。ただし、 F_0 軌跡を何の基準もなしに修正することは難しいので、最も最近「歌って見るモード」でこの対象をゴーストとして歌唱した際の F_0 軌跡が同時に青で描画されるようにした。対象ボーカルの声が高くてユーザがその音高で歌えない場合でも、低く歌えば F_0 は高さが異なる同じ軌跡を描くため、参考として利用できる (図 1)。

描くという作業には、対象をより注意深く見る・聴くことが求められるため、 F_0 推定精度向上のみでなく、歌唱に対する新たな気付きや理解を促進できると考える。

3 MiruSinger の実現方法

マイクからの単独歌唱からリアルタイムに F_0 を推定する技術、複数の楽器音が混在する音源 (音楽 CD) からボーカルを抽出する技術、ビブラートを検出する技術が必要である。それぞれ、入力信号中で最も優勢な (パワーの大きい) 高調波構造を求める手法 [4]、混合音中で最も優勢な音高を推定する手法 (PreFEst)[5]、推定された F_0 軌跡の 1 次差分の短時間フーリエ変換によってビブラートらしさを得る手法 [3] を用いた。ただし、混合音からの F_0 推定では、中高域で最も優勢な高調波構造を持つ F_0 がボーカルの F_0 であると仮定した。

4 おわりに

歌を「見る」ことを「歌う」「聴く」「描く」と併用した歌唱理解・歌唱力向上を支援するインタフェース MiruSinger を提案した。描くインタラクションの導入により、普段見慣れない歌唱の F_0 軌跡への慣れ、歌唱への理解の促進、音楽 CD からのボーカル抽出の精度向上に同時に対処できた。今後は、インタフェースの有効性の評価、音高以外の歌唱情報 (声質など) の視覚化、歌唱力自動評価機能の実装などに取り組む予定である。

参考文献

- [1] 株式会社ヤマハ, 神谷, 橋: カラオケ装置, 特開 2005-107337 (2005).
- [2] 平井 他: 歌の調子外れに対する治療支援システム, 信学論, Vol. J84-D-II, No.09, pp.1933-1941 (2001).
- [3] 中野 他: 楽譜情報を用いない歌唱力自動評価手法, 情処学論, Vol.48, No.1, pp.227-236 (2007).
- [4] 後藤 他: 自然発話中の有声休止箇所リアルタイム検出システム, 信学論, Vol. J83-D-II, No.11, pp.2330-2340 (2000).
- [5] 後藤: 音楽音響信号を対象としたメロディーとベースの音高推定, 信学論, Vol. J84-D-II, No.1, pp.12-22 (2001).