

# F0 動的成分に着目した歌唱のグルーブ感の分析

新居 優<sup>1,a)</sup> 片寄 晴弘<sup>1</sup>

概要：歌唱の「うまさ」を表す指標の一つに「グルーブ感」がある。子音長などの観点から研究がなされてきたが、完全な解明には至っていない。本稿では基本周波数に見られる動的変動成分のひとつであるオーバーシュートに着目し、歌声の聴覚印象に与える影響を調査した。聴取実験の結果、オーバーシュートにより音量が大きく知覚され、その結果グルーブ感にも影響することがわかった。

## 1. はじめに

近年、「歌のうまさ」に対する関心が高まっている。カラオケ機器に採点機能が搭載されるようになり、その機能を用いて歌唱力を競う番組が放映されたりと、歌う際に評価をされる機会が増加していることがその一因である。また、歌を「作る」機会も多くなっている。ヤマハ株式会社が発売している「VOCALOID」シリーズは歌声を合成することが出来るソフトウェアであるが、その台頭により誰でも歌を「作る」ことが出来るようになった。この「VOCALOID」シリーズが様々なメディアに展開してきている今、必然的に「うまい歌」とはなにかという問題に多くの人に関心を持っていることになる。

歌のうまさを判断するための定量的な基準は存在しない。たとえば多くのカラオケ採点では、音程の正確さや、ピブラート・しゃくりの回数など、主に技巧的な面から点数が判定されるものの、このような方法により高得点と判断された歌唱が人の耳に「うまい」と聞こえるとは限らない。一般的に歌の「表情」は音響特徴からは評価しにくく、ゆえに算出基準が「歌のうまさ」の本質からは乖離しているのが現状である。歌のうまさを判定基準を人間の感性に近づけるためには、歌のうまさを構成している要因の定量的な分析が重要である。

歌のうまさを構成要因の一つにグルーブ感がある。グルーブ感とはポップス歌唱などで用いられる表現のひとつであり、主にリズムの微細な揺らぎとして説明される。より一般的な感覚ではグルーブ感とは歌唱に対して感じる生き生きとした表情のことであると言える、マディソンらの研

究では「音楽に合わせて体を動かしたくなる感覚」と説明されている [1]。

グルーブ感に関する先行研究は、リズムの観点から分析したものが多く、奥平らはドラム演奏の打点時刻と音量に着目し、打点時刻の揺らぎがグルーブ感と関係していることを示している [2]。マディソンらはシンコペーションに着目して研究を行い、シンコペーションのリズムがグルーブ感を増大させるという結果を得ている [3]。また歌声とグルーブ感の関係に着目したものとしては、的場らが歌声の母音と子音に着目した研究を行い、子音長がグルーブ感の構成要因であることを示している [4]。しかし、歌唱の他の要素に着目してグルーブ感を分析した研究は未だ行われていない。

歌声はリズム以外にも音圧・音高・音色に基づくスペクトル・言語情報など様々な要素を持っており、リズムに着目した研究はあくまで一面的な分析でしかない。著名な音楽家である坂本龍一氏も自身の番組で「リズムだけでなく、音色や音の大きさがグルーブが出てくる」[5]と語っており、現状グルーブ感の本質が説明されているとは言いがたい。

本稿では、歌唱独自の要素に着目し、グルーブ感の構成要因を探る。具体的には、F0 軌跡のオーバーシュート成分に着目し、マルコフモデルによって推定されたパラメータをコントロールして生成される歌唱データに対する聴取実験により、検討を実施する。

## 2. F0 動的成分とグルーブ感

本研究では、歌唱の F0 軌跡おける動的な変動成分（以下、動的成分）に着目し、グルーブ感に与える影響を調査する。

### 2.1 F0 動的成分

本研究では、歌唱の F0 軌跡に着目する。F0 軌跡は、歌唱と他の楽器を大きく差別化している要素である。打楽器な

<sup>1</sup> 情報処理学会

IPSJ, Chiyoda, Tokyo 101-0062, Japan

<sup>f1</sup> 現在、関西学院大学

Presently with KWANSEI GAKUIN University

<sup>a)</sup> m.arai@kwansei.ac.jp

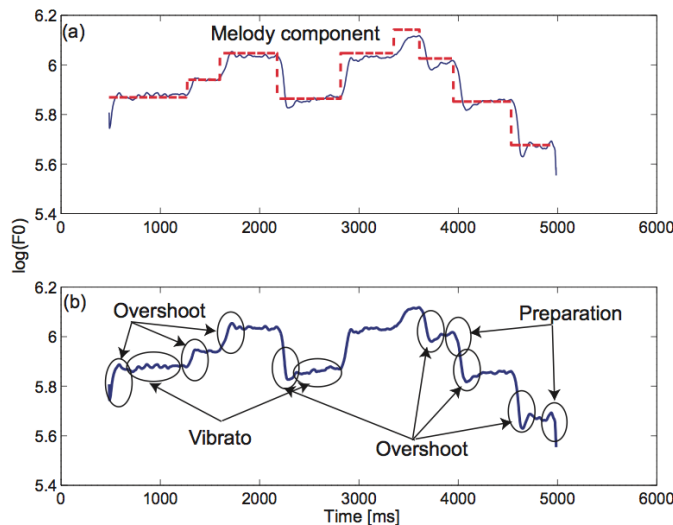


図 1 楽音成分と動的成分 [6]

どのリズム楽器と決定的に異なっている要素であることに加え、人間が発声していることにより、他の楽器の F0 軌跡とも異なった動的成分を持つことが知られている。F0 軌跡が楽音成分と動的成分に分けられる様子を図 1 に示す。

動的成分とは、歌唱の旋律成分とは別に歌唱者によって加えられる動的な変動成分のことであり、これまでたびたび研究の対象となっている。一例として、齋藤らは動的成分であるオーバーシュート/アンダーシュート・ビブラート/微細変動成分・予備的動作が、歌声の自然さに影響するとしている [7]。また中野らは、ビブラートが歌唱力の評価における重要な要素と成っていることを示した [8]。このように、動的成分は人間の歌声独特のものであり歌声知覚に影響することが分かっている。

## 2.2 グループ感の要因

グループ感を生み出す要因として考えられるもののひとつにアクセントが挙げられる。前述のマディソンらの研究 [3] では、四分音符で構成された単純な譜面とシンコペーションに変換した譜面を比較すると、シンコペーションを付与した楽譜にグループ感の増大が見られることが明らかにされた。そもそもシンコペーションとはアクセントが本来の位置とずれることであり、このことからアクセント、言い換えるならば特定の音の強調がグループ感に影響を及ぼしている可能性は大いにある。

アクセントを生み出す要因になるものとして、本研究ではオーバーシュートに着目した。これまでの研究により、動的成分のうち意図的に付与されるものであるビブラートは、歌唱における人間らしさや巧拙感など、高次の感覚に影響を与えることが知られている [7][9]。一方で、オーバーシュートは発声の構造により生じる物理的制約による動的成分である。一般に、大きな声を出そうとすればオーバーシュートの行き過ぎ量は大きくなる。このことより、オー

バーシュート行き過ぎ量の大小が音量の知覚に影響を与える可能性があると考えられる。

音量と F0 軌跡はともに歌唱の重要な要素であるが、この 2 つを結びつけて考えた研究は今までに無い。本研究では、オーバーシュートが音量の知覚に影響し、グループ感の要因になるとする仮説を構築し、これを検証するための実験を行う。

## 3. 歌唱の分析・再合成

F0 動的成分によるグループ感への影響を調べるため、動的成分を含む F0 軌跡のモデルを構築する。その上で F0 動的成分を編集し、再合成した歌唱を作成する。

### 3.1 歌唱データとソフトウェア

本研究では、プロ歌唱者による同一楽曲のグループ感有無歌い分け歌唱を対象として分析を行う。この歌唱データは、的場らによる歌唱のグループ感に関する先行研究 [4] で収録されたもので、グループ感のありなし歌唱を歌い分け可能な歌唱者に 2 つの楽曲 (楽曲 A: LaLaLaLoveSong (久保田利伸), 楽曲 B: LoveRain 恋の雨 (久保田利伸)) を歌い分けさせたものである。

また、本研究は F0 の動的成分という観点からグループ感にアプローチしているため、歌唱の F0 を高い精度で分析した上で独立に F0 を操作した歌唱を再合成できることが求められる。よって歌唱の分析・再合成は STRAIGHT[10] を用いて行う。

### 3.2 分析手法

F0 動的成分が歌唱の知覚に影響するという仮説の検証のためには、F0 動的成分を個別に編集する必要がある。そこで本研究では、楽音ごとの動的成分を分析できるよう、歌唱の F0 軌跡をモデル化した。歌唱の F0 軌跡は、階段状に変化する楽譜上の楽音成分と、それに付随する動的成分とで構成されるため、二次系の伝達関数と HMM を導入する (図 2)。なお、以上のモデル構築に際しては大石らの提唱した手法 [11] を踏襲しつつ、一部の値やアルゴリズムに筆者による調整を加えている。これは F0 軌跡へのフィッティングが一部うまくいかなかったため、対象としている歌唱ジャンルの違いが原因であると考えられる。

#### 3.2.1 HMM による楽音単位の分割

まず、読み込んだ F0 系列  $o_{Hz}$  を対数スケールの周波数  $o_{cent}$  に変換する。変換式は以下の通りである。

$$o_{cent} = 1200 \log_2 \frac{o_{Hz}}{440 \times 2^{\frac{3}{12} - 5}} \quad (1)$$

次に、HMM を用いて楽音の推移を推定する。HMM は各状態がひとつの音高に対応するエルゴディック HMM で、3000[cent] から 7000[cent] まで 100[cent] 刻みで 42 個の状態を持つ。各状態は、平均が音高値、分散が 100 の正規分布

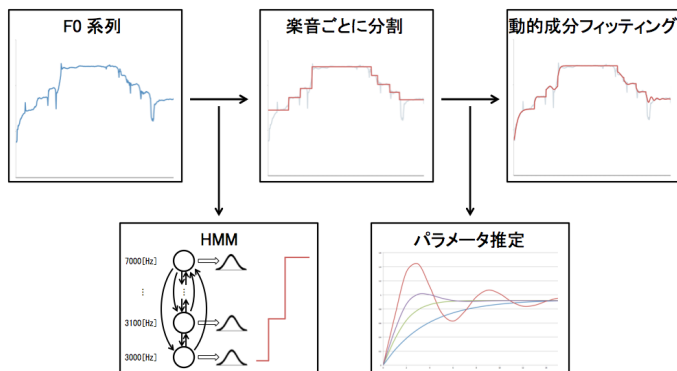


図 2 F0 軌跡分析処理の流れ

に従う出力を持つ。状態遷移確率は自己遷移確率が 0.9, その他の状態への遷移確率は一律で 0.1/41 とした。そして楽譜情報をもとに Viterbi 探索を行い状態経路を推定することで, F0 系列を楽音単位に分割する。なお, 出力分布の分散や遷移確率などの値は実験的に決定した。

### 3.2.2 最小二乗法による動的成分のフィッティング

続いて, 各楽音ごとに伝達関数のパラメータを推定し, 動的成分へのフィッティングを行う。伝達関数は以下の式で表される。

$$H(s) = \frac{\Omega^2}{s^2 + 2\zeta\Omega s + \Omega^2} \quad (2)$$

実際のフィッティングに際して, 前処理として各楽音の F0 軌跡を以下の方法で正規化する。

- 該当の楽音が最初の音であれば, 該当楽音の F0 系列からその先頭の値を減算する。
- 最初の音でなければ, 該当楽音の F0 系列の先頭の値及び最初の音から該当の音の一つ前の楽音までの音高値を減算する。

そして, 伝達関数のインパルス応答を畳み込んだ信号と, 実際の F0 系列との残差の二乗和が最も小さくなるように, パラメータ  $\zeta$  と  $\Omega$  を推定する。

これにより, 各楽音ごとに F0 動的成分を分析・編集し, 再合成することが可能となる。

## 4. 聴取実験

本研究では, オーバーシュートがグループ感に与える影響についての仮説を検証するため, 二つの聴取実験を行う。まず, オーバーシュートが音量の知覚に与える影響を調査するため, フレーズ中一箇所の動的成分を編集して聴取実験を行う。次に, オーバーシュートがグループ感に与える影響を調査するため, フレーズ全体の動的成分を編集して聴取実験を行う。

### 4.1 オーバーシュートと音量の知覚に関する実験

オーバーシュートが大きくなると音量も大きく知覚されるとの仮説を証明するため, フレーズ中のオーバーシュ

表 1 オーバーシュートと音量の知覚に関する実験の結果

	仮説と一致する回答	仮説と一致しない回答	計
観測度数	55	53	108
期待度数	36	72	108

トを一箇所増幅し, その音の音量について聴取実験を行う。

#### 4.1.1 歌唱の合成

プロの歌唱者による「グループあり」歌唱と「グループなし」歌唱のデータセットを用いて, 「グループなし」歌唱の 6 箇所のフレーズに対して F0 動的成分の編集を行う。仮説に基づき, オーバーシュートの行き過ぎ量が「グループあり」歌唱と等しくなるようにして再合成した歌唱を作成する。編集するフレーズは, グループあり歌唱とグループなし歌唱の間で特にオーバーシュート部分に差が見られたものを選定した。

#### 4.1.2 聴取実験

もとの歌声 (A) と F0 動的成分を増幅した歌声 (B) の組 (A,B),(B,A),(A,A) をランダムに提示し, 計 18 組に関して「1 番目が大きい」「どちらも同じ」「2 番目が大きい」で回答してもらったのち, 「A,B で変化していると感じた点」を自由記述してもらう。また, 被験者は正常な聴力を有した大学生 6 名である。

#### 4.1.3 実験結果

得られた 108 の回答の内訳は, 増幅した歌声のほうが大きいと答えた回答, すなわち仮説に一致する回答が 55, それ以外の回答が 53 であった。選択肢が 3 つであるため, 期待度数は仮説と一致する回答が 36, 一致しない回答が 72 であると考え, この結果に関して  $\chi^2$  検定を行ったところ 5%水準で有意であった (表 1)。つまり, オーバーシュートが音量の知覚に影響を与えていると言える。

また, (A,B) の組を聴いて「1 番目が大きい」とした回答及び (B,A) の組を聴いて「2 番目が大きい」とした回答は合わせて全体の 10%であった。

### 4.2 オーバーシュートとグループ感に関する実験

オーバーシュートがグループ感に与える影響を調査するため, フレーズに含まれるオーバーシュートを全て増幅し, 聴取実験を行う。

#### 4.2.1 歌唱の合成

さきの実験と同じく, 「グループなし」歌唱の 6 箇所のフレーズに対して F0 動的成分の編集を行う。フレーズ内の全てのオーバーシュート行き過ぎ量を「グループあり」歌唱と揃えて再合成した歌唱を作成した。

#### 4.2.2 聴取実験

再合成した歌声をもとに聴取実験を行う。もとの歌声 (A) と F0 動的成分を増幅した歌声 (B) をランダムに提示し, 12 個の歌唱についてグループ感を 0 点から 9 点で評価してもらう。そしてその後, グループ感を判定した際の基

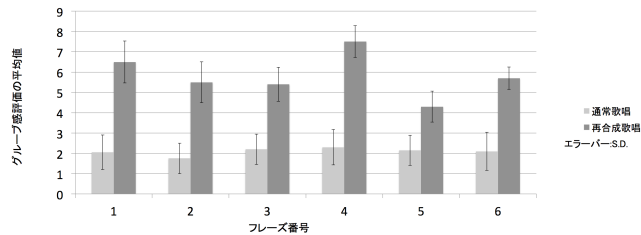


図 3 グループ感の印象評価結果

準について記述してもらおう。また、実験には Web インターフェイスを用い、静かな環境とヘッドホンまたは大きいスピーカーにて実験を行ってもらおう。被験者は正常な聴力を有した大学生 10 名である。

#### 4.2.3 結果

聴取実験の結果を図 3 に示す。横軸がフレーズ番号、縦軸がグループ評価の平均値である。この結果に関して t 検定を行ったところ、フレーズ 1,2,3,4,6 が 5%水準で有意であった。つまり、オーバーシュートがグループ感に影響していると言える。

### 5. 考察

#### 5.1 オーバーシュートと音量の知覚に関する実験の考察

表 1 に示した結果の通り、オーバーシュートの行き過ぎ量と音量の知覚の間に相関が見られた。特に、オーバーシュートを増幅したほうが小さく聞こえるとした回答が非常に少なかったことは、オーバーシュートと音量の知覚に正の相関があることを示している。また、自由記述では、「音量が大きいというより、アクセントやスピード感が付与されているように感じる」という意見が得られた。このことは、実際の音量は変わらないにもかかわらずオーバーシュートにより音量が大きく知覚されるとした仮説を支持している。

以上のことより、オーバーシュートは音量の知覚に影響し、行き過ぎ量が大きくなれば音量も大きく知覚されると言える。

#### 5.2 オーバーシュートとグループ感に関する実験の考察

図 3 に示した結果の通り、全てのフレーズにおいて再合成した歌唱のグループ感の評価が上昇し、そのうちほとんどで有意な差が見られた。これは、オーバーシュートを増幅することによりグループ感も増加するということを示している。

また、グループありなしの判断基準として自由に記述してもらったコメントとしては、「歌詞の節ごとの歌声の大きさ」など音量に関するものが見られた。これらのように、一部の音における音量評価が基準であるとしたコメントは、オーバーシュートによる音量知覚の増大がグループ感の要因となっていることを示唆している。

以上のことより、オーバーシュートを増幅するとグループ感も増大して知覚されると言える。

表 2 各フレーズのオーバーシュート内訳と平均点の差

フレーズ番号	上向き	下向き	平均点の差
1	2	0	4.45
2	1	1	3.75
3	1	1	3.2
4	2	1	5.2
5	1	1	2.15
6	1	1	3.6

#### 5.3 フレーズごとの差に関する考察

全てのフレーズでグループ感の評価は向上したものの、フレーズごとにその上昇量は異なっており、有意な差が見られなかったフレーズも存在した。この原因として考えられるのが、含まれるオーバーシュートの内訳である(表 2)。本研究では音が上昇する際のオーバーシュートと下がる際のオーバーシュートを同一として扱い全てに変更を加えたが、フレーズ内に含まれる上向きオーバーシュートと下向きオーバーシュートの数には差があり、これがグループ感知覚の差になっているものと考えられる。

最も差が大きかったフレーズ 4 は、上向きと下向きを合わせた数が 3 であり最も多い。また、その他のフレーズは合計数が 2 であるが、そのうち最も差が大きいフレーズ 1 は上向きが 2 で下向きが 0 である。このことより、含まれるオーバーシュートの数がグループ感の知覚に大きく影響し、更に上向きオーバーシュートより下向きオーバーシュートの方が影響が大きいものと考えられる。

以上のことより、オーバーシュートの数、および上向きオーバーシュートと下向きオーバーシュートの違いがグループ感に多寡に影響していると考えられる。

#### 5.4 総合的考察

本研究では、歌唱のグループ感の構成要因を解明するべく音楽情報処理の観点から分析を行った。その際、歌唱における独自の構成要因を見いだすため、すでにくつか研究がなされているリズムという側面ではなく F0 軌跡に着目して分析を行い、聴取実験にて有意な結果を得た。これにより、オーバーシュートはグループ感の構成要因であり、歌唱におけるグループ感の知覚に影響を与えるということが分かった。

### 6. おわりに

本研究では歌声の F0 における動的変動成分に関して、(1) 音量知覚への影響、(2) グループ感への影響、について調査した。

音量知覚への影響 オーバーシュートが音量の知覚に与える影響を調査した。STRAIGHT によりオーバーシュートを大きくした歌唱を合成し聴取実験を行ったところ、有意な差が得られた。すなわち、オーバーシュートが大きくなるにつれて音量も大きく知覚されるという結

果が得られた。

グループ感への影響 オーバーシュートによる音量の増加がグループ感に与える影響を調査した。STRAIGHTによりオーバーシュートを大きくした歌唱を合成し聴取実験を行ったところ、こちらも有意な差が得られた。オーバーシュートが大きくなり音量が大きくなると、グループ感も増大して感じられるということが確認された。

このふたつの実験を通じて、F0 動的成分のひとつであるオーバーシュートと、グループ感の知覚の間には相関があると示された。

今後の課題として、さらに多方面からグループ感の構成要因を検討することが必要である。F0 動的成分がグループ感の構成要因の一つであるという結果が出たが、グループ感全貌の解明には至っていないため、歌唱の他の要素にも着目したさらなる分析が必要である。

謝辞 本研究の考察にあたり、成山隆一様(ヤマハ株式会社)の助言を得ました。

#### 参考文献

- [1] Guy, Madison. et al.: Experiencing groove induced by music: consistency and phenomenology, *Music Percept*, 2006.
- [2] 奥平啓太, 平田圭二, 片寄晴弘, ポップス系ドラム演奏の打点時刻及び音量とグループ感の関連について, 情報処理学会研究報告, 2005-MUS-59-5, 2005.
- [3] Guy, Madison. et al.: Syncopation creates the sensation of groove in synthesized music examples, *frontiers in PSYCHOLOGY*, 2014.
- [4] 的場達矢, 歌唱のグループ感の構成要因の分析, 情報処理学会研究報告, 2014-MUS-102-12, 2014.
- [5] NHK BS2「スコラ 坂本龍一 音楽の学校」(2010年6月19日放映)
- [6] 齋藤毅, 歌声における動的変動成分の抽出と制御モデル, 北陸先端科学技術大学院大学 博士論文, 2006.
- [7] 齋藤毅, 鶴木祐史, 赤木 正人, 歌声における動的変動成分の抽出と制御モデル, 音響学会春季講演論文集, 2002.
- [8] 中野倫靖, 後藤真孝, 平賀讓, 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, 2007.
- [9] 右田尚人, 森勢将雅, 西浦敬信, ヴィブラート歌唱における基本周波数制御に有効な特徴量の検討, 音響学会秋季講演論文集, 2010.
- [10] Kawahara, H.: STRAIGHT, Exploration of the other aspect of VOCODER: Perceptually isomorphic decomposition of speech sounds, *Acoustic Science and Technology*, Vol.27, pp.349-353, 2006.
- [11] 大石康智, 亀岡弘和, 持橋大地, 永野秀尚, 柏野邦夫, Vocal Dynamics Controller: 歌声の F0 動特性をノート単位で編集し, 合成できるインタフェース, 情報処理学会研究報告, 2010-MUS-86-9, 2010.