

# 声道形状と声帯音源特性を用いたグロウル系 歌唱音声への実時間変換の提案

溝淵翔平<sup>1,a)</sup> 西村竜一<sup>1,b)</sup> 入野俊夫<sup>1,c)</sup> 河原英紀<sup>1,d)</sup>

**概要:** 本研究では通常歌唱音声をグロウル系歌唱音声の印象をもつ音声に変換するシステムについて検討している。これまでの研究よりグロウル系歌唱音声特有の物理的特徴としてスペクトル形状の高速な変動が確認された。本発表ではスペクトル形状の高速な変動を声帯音源特性と声道形状の変化としてモデル化し、グロウルの印象を付与する方法を提案する。声帯音源特性の時間変化は LF model を用いることでスペクトル傾斜の時間変化をモデル化した。声道形状の変化は入力音声について分析した声道断面積関数を操作することで実現した。提案手法による変換処理はフィルタリングで構成されているため、原理上はリアルタイム処理が可能である。

**キーワード:** グロウル, 声道音源特性, 声道形状, 声質変換, TANDEM-STRAIGHT

## Realtime singing voice conversion to growl-like singing based on vocal tract shape and glottal source characteristics

**Abstract:** Outline of a system to convert usual singing voice to growl-like performance in realtime is introduced. Relatively high-speed periodic variations (around 70Hz) in spectral shapes and fundamental frequency trajectories were found dominant features of growl-like singing in our pervious investigations. A set of simulations revealed that these spectral shape variations can be closely replicated by introducing vocal tract shape variations around spura-glottal structures and shape variations in glottal source waveform using the LF-model. Despite the fact that realtime extraction of LF parameters from input voice is not feasible, the simulation results indicated that the net effect of the variation can be represented by simple spectral slope variations. For vocal tract shape variation, several set of spectral models for approximating simulated variations can be suggested. These indicate that by using these approximated models, it is possible to design a realtime system for converting usual singing voices to growl-like voices.

**Keywords:** Growl, source characteristics, vocal tract shape, voice conversion, TANDEM-STRAIGHT

### 1. はじめに

歌唱では、情緒を豊かに表現する為に、様々な声質が用いられる。特に声帯振動に加えて、声帯上部にある披裂喉頭蓋ひだの振動との相互作用によって独特な印象を付与された歌唱音声に注目する [1]。これらの特徴をもつ歌唱音声をグロウル系歌唱音声と呼び、伝統的な歌唱や民族歌唱からロックやメタルなど幅広い音楽のジャンルで使用されて

いる。グロウル系歌唱音声にみられる独特な印象を自由に操作出来る技術の開発は、歌唱の応用に大きく貢献出来る可能性を有している。先行研究として、実際のグロウル系歌唱音声から取り出した特徴を、分析合成音声に転写することでグロウル系歌唱音声を実現する方法が提案されている [2]。また、歌声合成エンジンである VOCALOIDV4[3]では、新機能として「グロウル」が搭載されており、合成歌唱音声に「唸り声」のような印象を与えることが出来る。

我々のこれまでの研究では、特にフィルタリングと変調からなるシンプルな処理により通常歌唱音声にグロウル系歌唱音声の物理的特徴を付与する方法を提案した [4]。本手法におけるグロウル系歌唱音声特有の物理的特徴とは

<sup>1</sup> Wakayama University, Wakayama 640-8510, Japan

a) s155059@wakayama-u.ac.jp

b) nisimura@sys.wakayama-u.ac.jp

c) irino@sys.wakayama-u.ac.jp

d) kawahara@sys.wakayama-u.ac.jp

「1~4kHz 帯域の強調」、「基本周波数の高速な変動」、「スペクトル形状の高速な変動」を示す [5]。また音声の聴取実験より、これらの物理的特徴を付与された音声はグロウル系歌唱音声の印象を持つようになることがわかった [6]。本手法はリアルタイム処理を可能であるため、ライブやカラオケでの実時間処理を必要とする場で有効であると考えられる。

しかし、提案手法で「スペクトル形状の高速な変動」を付与するために用いたモデルは、観察された現象を表面的に模擬するために4個のガウス関数を組み合わせたものであり、声質の表現や発声の整理機構とは無関係なものであった。そのため、見通しの良い操作が困難であるという問題をかかえていた。

そこで、先行研究では、「スペクトル形状の高速な変動」を見通しよくモデル化することを狙い、スペクトル傾斜の時間変化と声道断面積関数の時間変化に関して分析を行った [7],[8]。結果、声道断面積関数の13~15cm付近とスペクトル傾斜に周期的な変動があり、声帯音源波形と声道形状の周期的な変動が示唆された。本稿では声帯音源特性と声道形状の周期的変化を組み合わせることで通常歌唱のスペクトル形状にグロウル系歌唱音声に認められる周期的変化を付与する方法を提案する。声帯音源波形の変化には LF model を用い、声道形状の変化には声道断面積関数による特徴付与処理を行った。

2節ではこれまでの研究成果より得られたスペクトル形状の高速な変動に関する特徴について概要を述べる。3節では声道形状に着目したスペクトル形状の高速な変動の付与方法について述べる。4節では声帯音源波形の変化に着目したスペクトル形状の高速な変動の付与方法について述べる。4節では、2節と3節で述べたスペクトル形状の高速な変動の付与方法を組み合わせた実験結果について述べる。最後に5節で全体のまとめと今後の課題を述べる。

## 2. 歌唱音声の分析

### 2.1 音声資料

分析試料には歌唱の研究のために作成された J-POP の楽曲 RIDE[9] を使用した。1人の男性のプロ歌手による様々な表現の歌唱をしているデータを用いた。試料は44,100 Hz、16 bit で収録されている。歌手自身の歌唱法で演奏したものはグロウルの特徴を有していたため、これを「グロウル系歌唱音声」の試料とした。また、有名な歌手の声質を真似た歌唱データにはグロウル歌唱が含まれていないため、これを「通常歌唱音声」の試料とした。こうして選択された同じ156.8秒の長さをもつそれぞれの歌唱データを分析試料として用いた。分析時には歌唱音声が含まれている約2小節の長さの32個のフレーズに分割した。

### 2.2 歌唱音声の基本周波数とスペクトル包絡の抽出

まず、波形の対称性に基づく基本周波数抽出法 [10][11] を用い、1ms のフレーム周期で基本周波数を求めた。この方法は、波形の対象性に基づく基本周波数の初期推定と、周期性による時間変動の影響を受けない瞬時周波数計算法に基づいて、複数の長波成分の情報を利用して精度を向上させる後処理から構成されている。求められた時間分解能の高い基本周波数を用いて、TANDEM-STRAIGHT[12] によるスペクトル分析を1ms 毎に行った。この方法では、F0 に適応的に設定した時間窓と平滑化を含む処理により、周期性に起因する影響だけを選択的に除去したスペクトル形状が求められる。

### 2.3 スペクトル形状の高速な変動の分析

グロウル系歌唱音声の生成系においては、披裂喉頭蓋ひだが声帯振動の約1/3程度の周波数で振動するという特徴がある。この披裂喉頭蓋ひだが振動することにより、次の2つの変化が考えられる。

- 披裂喉頭蓋ひだによる局所的な声道形状の変化による共振特性の変化
- 披裂喉頭蓋ひだの振動による駆動インピーダンスの変化がもたらす声帯音源特性の変化

これまでの研究で、声道形状の変化と声帯音源特性まず、歌唱音声より分析した声道断面積関数の時間変化を分析することで、グロウル系歌唱音声特有の声道形状の時間変化を調査した [7]。結果は、披裂喉頭蓋ひだの影響と考えられる13~15cmあたりの周期的な声道断面積関数の変動が確認され、声道形状の変化が示唆された。また、1オクターブ幅で平滑化した歌唱音声の長時間平均対数スペクトルとフレーム毎の対数スペクトルの差分について帯域ごとのRMSを求めることで、各歌唱におけるスペクトルの変動があった [8]。結果は、グロウル系歌唱音声の長時間平均スペクトルとフレーム毎の差分は通常歌唱よりも全体域において変動量が大きく、また、LF model のシミュレーションから、高域におけるスペクトル傾斜の変化が示唆された。

## 3. 声道断面積関数を用いたスペクトル形状の高速な変動の付与

### 3.1 声道断面積関数の導出

まず、それぞれの歌唱音声について TANDEM-STRAIGHT より求められたスペクトルに対して、F0 に適応した声帯音源と放射特性の補償 [13] を用いて声道特性以外の特性を抑圧した。これは、LF model[14] のシミュレーションより得られた平均的な声帯音源特性を声帯音源と放射特性を補償に用いる方法で、適応時には入力音声の基本周波数に応じてスペクトルの概形をスケールリングしてある。シミュレーションは基本周波数を50Hzに固定して時間パラメータをランダムに生成し、「地声」「フライ」「氣息

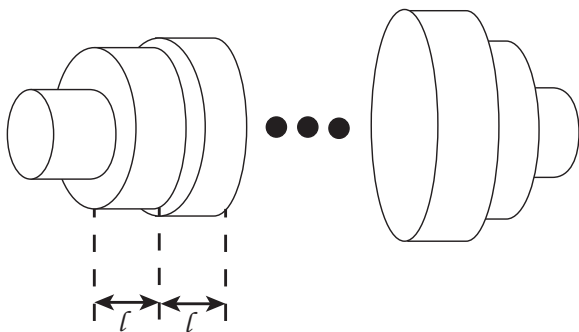


図1 Kellyの声道モデルに基づく声道断面積関数

性」の3つの声質についてそれぞれ3000回のシミュレーションを行った。時間パラメタの決定には Childers らによる平均と標準偏差 [15] を用いた。

本研究で扱う声道断面積関数は Kelly の声道モデル [16] に基づき、図1に示すような声道を断面積一定の微小音響管の縦列接続で表現したものである。図1の  $l$  は微小音響管の長さ、 $A_n$  は区間  $n$  の断面積、及び  $\kappa_n$  は区間  $n$  と区間  $n+1$  間の反射係数を示す。声道特性以外の特性を抑圧したパワースペクトルに関して 6000Hz までの帯域を取り出し、自己相関数を求め、レビンソンダービンの方法より PARCOR 係数を求めた。この時、LPC 次数は 12 とした。ここで、PARCOR 係数は反射係数と対応しており、反射係数  $\kappa_n$  と断面積  $A_n$  の間には以下の関係が成り立ち、声道断面積関数を求めることが出来る。

$$A_n = \frac{1 + \kappa_n}{1 - \kappa_n} A_{n+1} \quad (n = p, p-1, \dots, 1) \quad (1)$$

$$A_{p+1} = 1 \quad (2)$$

また声道断面積関数は体積一定の制約のもと計算している。

### 3.1.1 声道断面積関数の動的変化の付与

図1において声道断面積関数の区間の長さにあたる  $l$  は音速  $c$  m/s とサンプリング周波数  $F_s$  Hz を用いることで次式のように表現出来る。

$$l = \frac{100c}{2F_s} \quad (3)$$

披裂喉頭蓋ひだは口唇から約 13~15 cm 付近にある為、今回は時間変動を加える区間として、9 次の区間を中心にモジュレーションを付与することとした。また、式 (1) から分かるように変換後の声道断面積関数から式 (4) を用いることで変換後の反射係数を求めることが出来る。

$$\kappa_n = \frac{A_n - A_{n+1}}{A_n + A_{n+1}} \quad (4)$$

変換された反射係数から変換後の LPC スペクトルを求め、変換後の LPC スペクトルと元の音声のスペクトルの差分を声道形状の変化を与えるフィルタの周波数特性とした。

音声変換の例として、グロウル系歌唱音声の LPC スペクトログラムと通常歌唱音声の LPC スペクトログラム、

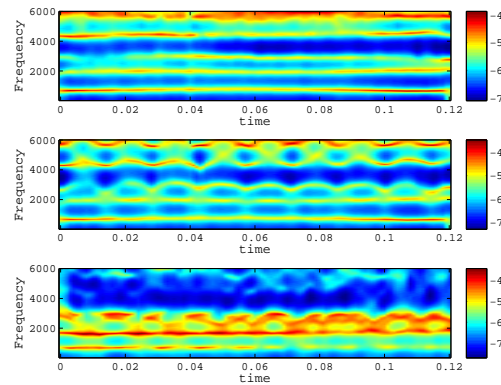


図2 声道断面積関数の操作による LPC スペクトログラムの変化の比較 (上段：通常歌唱音声、中段：特徴付与された通常歌唱音声、下段：グロウル系歌唱音声)

及び通常歌唱音声に声道形状変化の変換後の LPC スペクトログラムを図2に示す。図2はグロウル歌唱音声は特にグロウルの特徴を持つ/e/の発話区間であり、通常歌唱音声はグロウル系歌唱音声と同一の発話区間である。変換された通常歌唱音声は音声の 1/3 の周波数で振動する sin 波状に 9 区間目の声道断面積を変動させた例である。図から分かるように、通常歌唱音声のフォルマントは時間経過に対して一定であるのに対し、グロウル系歌唱音声の 3k~4k Hz における第3フォルマント周辺は時間経過に応じて波状に変化していることが分かる。変換後の音声に注目すると、第3フォルマント周辺に波状の変化が付与されていることが確認され、グロウル系歌唱音声特有のスペクトル形状の高速な変動を付与出来ることが示唆された。

## 4. LF model を用いたスペクトル形状の高速な変動の付与

声門上部の構造の振動により、声門から声道側を見た駆動点インピーダンスは、周期的に変動する。その結果、声帯振動の速度が変化し、声門閉止の鋭さが変化する。ここでは、鋭い閉止の状況を modal 発声でモデル化し、緩い閉止の状況を breathy 発声でモデル化した。この仮定のもとに 3 章で述べた声道形状の変化に合わせて、LF model による modal 時の声帯音源特性と breathy 時の声帯音源特性を発生させた。

### 4.1 LF モデルを用いた声帯音源特性

LF model [14] は、声門での体積流の波形に唇からの放射特性の影響を加えた波形を  $E(t)$  とし、次式により定義される。

$$E(t) = \begin{cases} E_0 e^{at} \sin \omega_g t & 0 \leq t \leq t_e \\ -\frac{E_c}{et_a} [e^{-\epsilon(t-t_e)} - e^{-\epsilon(t_c-t_e)}] & t_e \leq t \leq t_c < T_0 \end{cases} \quad (5)$$

この時、 $t_p$  は声門体積流が最大となる時点、 $t_e$  は声門の閉止時点、 $t_a$  は声門閉止の時定数、 $t_c$  は声門が完全に閉止す

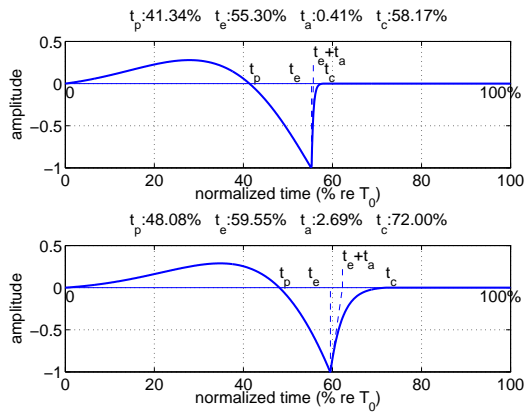


図3 平均的な modal 時の LF model による微分声門体積流

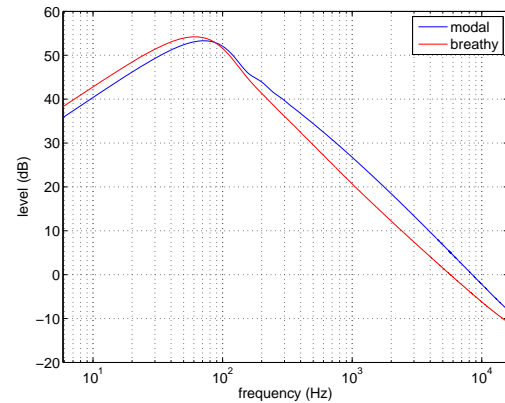


図4 平均的な modal 時の LF model による微分声門体積流

る時点を指す。また、 $T_0$  は声帯振動の一周期を指し、負の最大値を  $E_e$  としている。ここで、放射特性が周波数 0 に零点を有することから、以下の拘束条件があることがわかる。

$$\int_0^{T_0} E(t)dt = 0 \quad (6)$$

式 5 に含まれる  $\omega_g$  や  $\epsilon$  は、波形のタイミングを表す  $t_p, t_e, t_a, t_c$  から一意に定まる。また、 $E_e$  は、 $t_e$  における連続性から求められる。ここでは、シミュレーションにおいて実現不可能な音源波形を生成しない為に、各パラメータ  $t_p, t_e, t_a, t_c$  について以下の拘束条件を設けた。

- 体積流は正でなければならない。したがって、 $2t_p > t_e$
- 声門閉止の開始時点は、最大体積流の時点よりも後でなければならない。 $t_p < t_e$
- 声門閉止途上での体積流の減少速度は加速しない。 $t_e + t_a < t_c$
- 声門閉止の時定数は正でなければならない。 $t_a > 0$
- 声門が完全に閉じた後に、声門は開き始める。 $t_c < T_0$

図3に LF model を利用した平均的な modal 時と breathy 時の声帯音源波形を示す。また、基本周波数は 50Hz としている。breathy は modal に比べ、 $t_p$  と  $t_a$  の値が大きいことが分かる。これは breathy の場合、声門開口と声門閉止の力が弱く、時間がかかる為である。この影響は声帯音源特性にも現れる。図4に modal と breathy の平均的な声帯音源特性を示す。実際の基本周波数を考慮して移動させると、600Hz から 3000Hz の間では、breathy は、modal よりもスペクトルの傾斜が急であり、高域のレベルが低くなっていることが分かる。また、基本波と第二次調波のレベルを比較すると、breathy の方が基本波のレベルが高くなっていることが分かる。

#### 4.2 声帯音源特性の時間変化の付与

前説で述べたように声帯音源特性の時間変化の付与は 3 章で述べた声道断面積関数の操作に依存する。ここでは仮に、声道形状の変化が少ない部分は modal の声帯音源波

形に近づき、声道形状の変化が大きいは breathy の声帯音源特性に対応させるように設定した。よって、本研究では声道形状を操作するモジュレーション信号  $m(t)$  が  $0 \leq m(t)$  のとき modal に近づき、 $m(t)$  が  $0m(t)$  のとき breathy に近づくよう LF model のパラメータが変化するように設計する。例えば、 $t_p(t)$  の場合であれば

$$t_p(t) = m(t)t_p^m + (1 - m(t))t_p^b \quad (7)$$

とする。ここで、 $t_p^m$  は modal 時の  $t_p$  の平均値であり、 $t_p^b$  は breathy 時の  $t_p$  の平均値を示す。ここで、それぞれの時間パラメータの平均値は Childers らによる平均 [15] を用いた。 $t_e, t_a, t_c$  についても同じようにパラメータを求めた。こうして設定した時間パラメータを用いて得られたフレーム毎の声帯音源特性と平均的な modal 時の声帯音源特性との差分を特徴付与フィルタとして用いた。

### 5. 特徴付与処理の組み合わせに関する検討

3 節と 4 節でグロウル系歌唱音声の動的変化を付与する方法を提案した。スペクトル形状の高速な変動は他のグロウル系歌唱音声の特徴である基本周波数の振動とほぼ同周期であることが確認されている。これまでの検討で、「基本周波数の時間振動」の特徴は入力音声に位相変調することで、基本周波数に時間振動を与えていた。しかし、基本周波数の高低の変化とスペクトルのパワーの変動の関係は示されていない。そこで、本研究ではグロウル系歌唱音声の基本周波数軌跡をモジュレーション信号とし、本手法により特徴付与された通常歌唱音声と同区間におけるグロウル系歌唱音声のフレーム毎のスペクトル距離を評価した。

#### 5.1 モジュレーション信号の作成

まず、基本周波数軌跡から直流成分を差し引いたものを実信号  $s(t)$  とする。次に実信号  $s(t)$  をヒルベルト変換し、90 度位相が回転した信号  $\mathcal{H}[s(t)]$  を得る。よって実信号  $s(t)$  に比べ、位相が  $\phi$  ずれたモジュレーション信号  $m^{(\phi)}(t)$  は以下の式より得る。



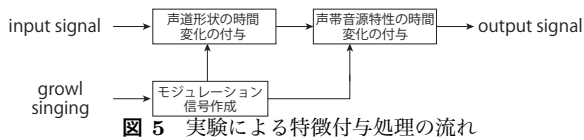


図5 実験による特徴付与処理の流れ

$$m^{(\phi)}(t) = \cos(\phi)s(t) + \sin(\phi)\mathcal{H}[s(t)] \quad (8)$$

本研究では、位相を 0.1rad ずつ増加させ、 $\phi=6.2$  までのモジュレーション信号を発生させた。また、モジュレーション信号は最大値が 1 になるように正規化した。

## 5.2 評価実験

本実験の流れを図5に示す。まず、予めグロウル系歌唱音声进行分析し、得られた基本周波数軌跡の情報を元に前節の方法でモジュレーション信号を作成する。次に TANDEM-STRAIGHT を用いて通常歌唱音声进行分析し、SRAIGHT スペクトルを抽出する。この時、スペクトル形状の高速な変動以外の両歌唱のスペクトルの差を軽減する為に、両歌唱の長時間平均スペクトルの差分を通常歌唱音声のスペクトルに付与した。次に通常歌唱音声のフレーム毎のスペクトルから声道断面積関数を分析し、モジュレーション信号  $m^{(\phi)}(n)\{n = 1, 2, \dots, N\}$  に応じて、声道断面積関数の 9 区間目を操作し、声道形状の変化に対応した補償フィルタを作成する。

補償フィルタを用いて変換された通常歌唱音声とグロウル系歌唱音声のスペクトル距離を評価した。スペクトル距離はフレーム毎のグロウル系歌唱音声と変換された通常歌唱音声の対数スペクトルの差分の RMS の平均を用いる。位相が  $\phi$  ずれたモジュレーション信号を用いて変換された通常歌唱音声とグロウル系歌唱音声のスペクトル距離を考える。n フレーム目のグロウル系歌唱音声のスペクトル  $f^{(\phi,n)}(\lambda)$  と変換された通常歌唱音声のスペクトル  $g^{(\phi,n)}(\lambda)$  の対数スペクトルの差  $V^{(\phi,n)}(\lambda)$  は

$$V^{(\phi,n)}(\lambda) = 10 \log_{10} f^{(\phi,n)}(\lambda) - 10 \log_{10} g^{(\phi,n)}(\lambda) \quad (9)$$

で表すことが出来る。位相が  $\phi$  ずれたモジュレーション信号の時の各帯域ごとの RMS は

$$R^{(\phi)}(\lambda) = \left( \frac{1}{N} \sum_{n=1}^N V^{(\phi,n)}(\lambda)^2 \right)^{(1/2)} \quad (10)$$

となる。結果、指標となるスペクトル距離  $D^{(\phi)}$  は  $R^{(\phi)}(\lambda)$  の  $\phi$  毎の平均とした。また、声道形状の変化の付与に関するスペクトル距離の評価は 6000Hz までの帯域を評価対象とした。

## 5.3 実験結果

以下は、特にグロウルの特徴を有していた歌詞「戯れ言も辛い」の/re/に注目した結果を示す。図6にモジュレーション信号の位相差  $\phi$  とモジュレーション信号と帯域毎

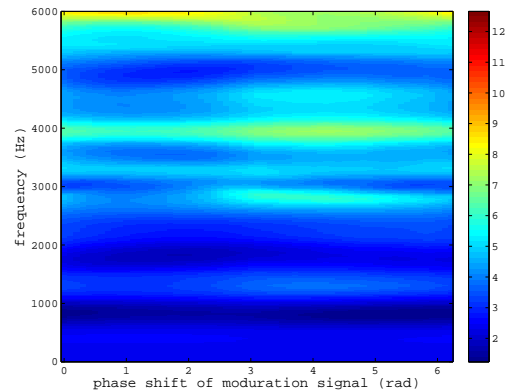


図6 モジュレーション信号の位相差  $\phi$  とモジュレーション信号と帯域毎の両歌唱の差の RMS である  $R^{(\phi)}(\lambda)$  の関係

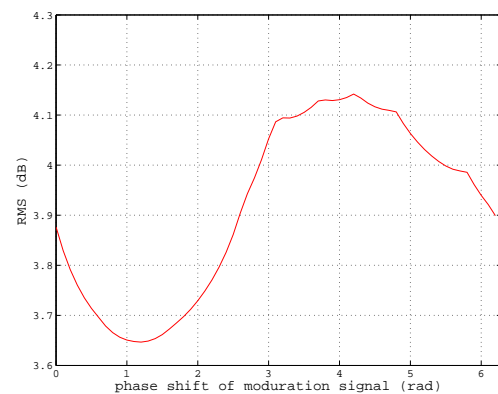


図7 モジュレーション信号の位相差  $\phi$  と  $D^{(\phi)}$  の関係

の両歌唱の差の RMS である  $R^{(\phi)}(\lambda)$  の関係を示す。横軸はモジュレーション信号の位相差  $\phi$ 、横軸は周波数、色の变化は RMS の大きさを示す。特にグロウル系歌唱音声においてフォルマント付近の変動が大きかった 3k~4kHz 帯域に注目する。約  $0.43\pi$  rad 位相をずらすことで 3~4kHz 帯域のスペクトル距離が小さくなっていることがわかる。また、図7にモジュレーション信号の位相差  $\phi$  と  $D^{(\phi)}$  の関係を示す。図7から分かるように、基本周波数軌跡をモジュレーション信号として利用した場合、約  $0.43\pi$  rad 位相をずらすことで最もグロウル系歌唱音声と特徴付与した通常歌唱音声のスペクトル距離が小さくなった。

## 6. おわりに

本資料では、声道形状の周期的変化と声帯音源波形の周期的変化を組み合わせるにより、通常歌唱のスペクトル形状に、グロウル系歌唱に認められる周期的変化を付与できることを客観的に示した。これらは、記録された歌唱音声のポストプロダクションに応用することのできる技術である。実時間でのグロウル系歌唱への変換の場合には、ここで得られた知見を、実時間処理に適した形で実装する必要がある。LF モデルによるスペクトル形状の変化は、スペクトル傾斜の変化により近似することができ、簡単な

フィルタで実現することができる。声道形状の変化によるスペクトルの変化は、LPCに基づく処理が軽いものであることから、そのまま実時間システムに組み込むことも可能である。また、通常歌唱とグロウル系歌唱の違いは、見通しの良い簡単なフィルタを用いて近似することも可能である。これらの近似の際に必要な基本周波数の周期的変化との位相差の設定は、基本周波数に引き込まれる自励振動子の二つの状態の値を利用することができる。これらを組み合わせて実際に動作するシステムとして実装することが、今後の課題である。

**謝辞** 本研究の一部は、科研費基盤 (B)24300073 と挑戦的萌芽 26540087 の支援による

## 参考文献

- [1] K. Sakakibara, et al.: Growl voice in ethnic and pop styles, *Proc. Int. Symp. on Musical Acoustics* (2004).
- [2] B. Jordi, B. Merlijn, 才野, 久湊: スペクトルモーフィングによるグロウル系統の歌唱音声合成, 情報処理学会研究報告, MUS-100-24, pp.1-6 (2013).
- [3] <http://www.vocaloid.com/vocaloid4/>
- [4] 河原英紀, 溝淵翔平, 森勢将雅, 榊原健一, 西村竜一, 入野俊夫: 非線形振動子による変調と近似時変フィルタに基づくグロウル系統の歌唱への実時間変換の定式化について, 情報処理学会研究報告, MUS-102-14, pp.1-6 (2013).
- [5] H. Kawahara and M. Morise and K. Sakakibara : Interference-free observation of temporal and spectral features in “shout” singing voices and their perceptual roles, *Proc. SMAC-SMC*, pp.256-263 (2013).
- [6] 溝淵翔平, 西村竜一, 入野俊夫, 河原英紀: グロウル系統の歌唱音声にみられるスペクトルの時間変動に注目した分析と再現の検討, 日本音響学会 2014 年春季研究発表会講演論文集, pp.499-500 (2014).
- [7] 溝淵翔平, 西村竜一, 入野俊夫, 河原英紀: グロウル系歌唱から求められる声道断面積関数の特徴について, 電子情報通信技術研究報告, SP2014-80, pp.29-30 (2014).
- [8] 溝淵翔平, 西村竜一, 入野俊夫, 河原英紀: 声道形状と声帯音源特性を利用したグロウル系歌唱音声への変換について, 日本音響学会 2015 年春季研究発表会講演論文集 (2015). (発表予定)
- [9] 河原ら,.:モーフィングに基づく歌唱デザインインタフェースの提案と初期検討, 情報処理学会論文誌, vol. 48, no. 12, pp.3637-3648 (2007).
- [10] H. Kawahara, et al.: Higher order waveform symmetry measure and its application to periodicity detectors for speech and singing with fine temporal resolution, *Proc. ICASSP*, pp.6797-6801 (2013).
- [11] H. Kawahara, et al.: Temporally fine F0 extractor applied for frequency modulation power spectral analysis of singing voices, *Proc. MAVEBA*, pp.125-128 (2013).
- [12] H. Kawahara, et al.: Tandem-STRAIGHT: A temporally stable power spectral representation for periodic signals and applications to interference-free spectrum, F0, and aperiodicity estimation, *Proc. ICASSP*, pp.3933-3936 (2008).
- [13] 伊佐衣代, 西村竜一, 入野俊夫, 河原英紀: 声道断面積関数推定における声帯音源特性の補償について, 日本音響学会 2015 年春季研究発表会講演論文集, (2015). (発表予定)
- [14] G. Fant.: The LF-model revisited. Transformations and frequency domain analysis., *Speech Trans. Lab. Q. Rep., Royal Inst. of Tech.*, vol. 2-3, pp. 121-156 (1995).
- [15] D G Childers and C K Lee.: Vocal quality factors: analysis, synthesis, and perception, *The Journal of the Acoustical Society of America*, vol. 90, pp. 2394-2410 (1991).
- [16] J. L. Kelly and C. C. Lochbaudsondm: Speech synthesis, *Proc. 4th International Congress on Acoustics*, G42, pp. 1-4 (1962).