Vol.2014-MUS-103 No.55
2014/5/25

歌声にグロウルの味を加えるGUIについて

溝渕 翔平[1,a)]　西村 竜一[1,b)]　入野 俊夫[1,c)]　河原 英紀[1,d)]

概要：本研究では通常歌唱をグロウル系統の歌唱音声の印象をもつ音声に変換するシステムの検討を行っている. 先行研究では簡単な信号処理で歌唱音声にグロウルらしさを付与する方法が提案された. 本報告では提案手法で用いる特徴付与のパラメタを対話的に操作し, 歌唱音声にグロウルらしさを付与する GUI について紹介する. 提案手法は時間変調による基本周波数の高速な時間振動の付与,FIR フィルタによる処理範囲に共通した帯域強調処理, 及び近似時変フィルタによる第 3 フォルマント周辺の高速な時間変調の付与の 3 つより構 成されている. 提案手法は変換処理に分析・合成を必要としないためリアルタイム処理を可能とし, ライブで一種のエフェクターとして用いることが出来る.GUI の開発は主にデモやポスターセッションの場で本手法による処理内容と処理の影響について直感的理解を促すことを目的としている. 開発した GUI は実際にポスターセッションの場で操作し, 操作性やデザイン性についてコメントを頂きたい.

A GUI for manipulating growl-like taste in singing voice

Mizobuchi Shohei[1,a)]　Nisimura Ryuichi[1,b)]　Irino Toshio[1,c)]　Kawahara Hideki[1,d)]

Abstract: A set of GUIs is designed to add and manipulate growl-like taste in singing voice based on a set of simple signal processing procedures, proposed in our previous report. It consists of a temporal axis modulator for simulating rapid F0 variations, an equalizer to modify global spectral shape, and an approximate time varying filter for simulating rapid spectral modulation around F3 area. The proposed set of procedures is potentially applicable to realtime applications, such as live performance. This set of GUIs will be presented in the poster session for demonstrating possibilities of the proposed procedures and acquiring feedback and comments from prospective participants.

1. はじめに

歌唱では，情緒を豊かに表現するため，様々な声質が用いられている．特に強い感情を表現する場合，非周期な変化を伴う歌唱音声が用いられる．こういった特徴を持つ歌唱音声は伝統的な歌唱や民族歌唱，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．このような非周期な変化をもつ歌唱音声をグロウル系統の歌唱音声と呼び，グロウル系統の歌唱音声にみられる独特な印象を自由に操作できる技術の開発は，歌声処理の応用に大きく貢献できる可能性を有している．しかし，グロウル系統の歌唱音声の非周期的な変化には声帯振動だけではなく，声帯上部の披裂喉頭蓋ひだによる発声器官全体の相互作用によって生じられる．先行研究として，実際のグロウル系統の歌唱音声から取り出した特徴を，分析合成音声に転写してグロウル系統の歌唱を実現する方法が提案されている [1]．本研究では特にフィルタリングと変調からなる軽い処理に基づいたリアルタイム処理を可能とする手法を紹介した [2]．本手法はグロウル系統の歌唱音声に分析を加えたの分析により，特有の印象に影響を与える物理的特徴を見いだした [3]．また，グロウル歌唱の分析に関して，それらの特徴を通常歌唱音声に付与することでグロウル系統の歌唱音声のような印象をもつ歌唱音声に変換される手法を提案した．

本研究の手法を音声のエフェクターとして活用するため，最終的にユーザに音声変換技術を使用しやすい形で提供する必要がある．そこで，本手法による音声加工を対話

[1] Wakayama University, Wakayama, 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

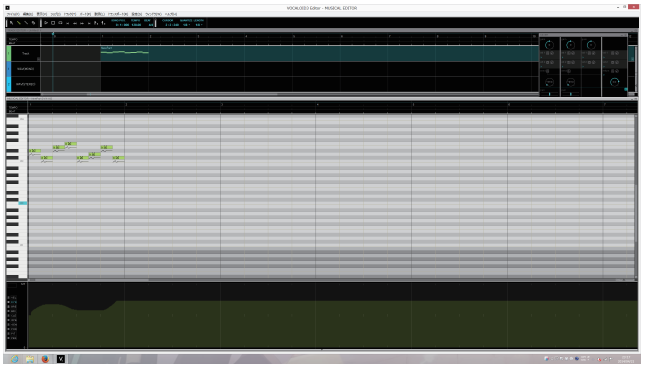

図 1 VOCALOID3 のエディター画面 [4]

的に操作出来る GUI の開発を行った．音声加工に用いられる GUI として，VOCALOID エディター (図 **??**) などが挙げられるが，本研究では，デモやポスターセッションの場で本手法による処理内容と処理の影響について直感的理解を促すことを目的とした GU の開発を目指す．本報告では提案手法による音声変換に関する処理と GUI の設計を中心に開発したシステムについて述べる．以下，2 章で本研究の GUI に必要な機能について述べ，3 章で先行研究で明らかとなったグロウル系統の歌唱音声の特徴と提案された音声変換の処理方法を解説する．4 章で 3 章の内容を統合したシステムと GUI の構成に関して述べ，最後に 5 章で全体のまとめと今後の課題を述べる．

2. 音声変換 GUI の要求仕様

音声変換用の GUI に最低求められる内容は以下の 4 つである

- 音声の入力，及び録音
- 音声変換に必要なパラメタの設定
- 音声変換
- 変換前と変換後の音声の聞き比べ

GUI 開発は Matlab の GUI 作成キットである GUIDE (GUI Development Environment) を用いた．また，設計した GUI の主な目的は，発表やデモなどの場で本研究の処理内容と処理による音声への影響に対する直感的な理解を促すことである．そのため，開発する GUI は本手法による処理内容を構造的に示し，また，パラメタ操作と処理の変化を GUI を通して視覚的に理解出来るようにするなどの工夫が要求される．

3. 特徴付与に用いられる方法

これまでの検討により，グロウル系統の歌唱音声と通常歌唱とグロウル系統の歌唱音声に特有の音響的特徴は以下の 3 種類の特徴に依存していることが示されている．

Q:基本周波数の高速な時間振動
F:基本周波数の変調に同期したスペクトルの変動
E:スペクトルにおける 2000 Hz 付近を中心とする帯域強調

提案した方法では，これらを変調とフィルタリングにより実現される．また，特徴付与処理を統合したシステム全体の構成を図 2 に示す．また，Q,F の時間的な変動の傾向に，強い歪みの無い非線形振動子を導入することでグロウル系統の歌唱音声の独特の励起振動をモデル化した．

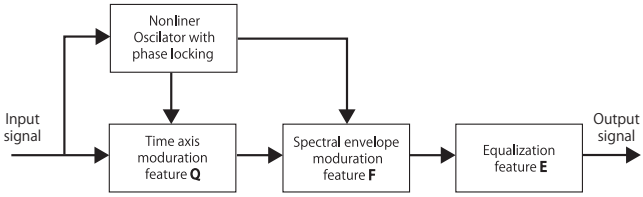

図 2 音声変換のシステム図

3.1 基本周波数の高速な時間振動:Q

通常の歌唱音声の基本周波数を時系列とみなして周波数変調のパワースペクトルを求めると，変調周波数の上昇に伴ってパワーが減少する．しかし，グロウル系統の歌唱音声では，70 Hz 付近にピークが存在していることが見いだされた．この高速な周波数変調を，時間軸を局所的に伸縮させて位相変調を加えることにことにより実現した．入力信号の瞬時周波数を $f_0^{IN}(t)$ とした際，目標とする瞬時周波数を $f_0^{MOD}(t)$ は次式で表すことが出来る．

$$f_0^{MOD}(t) = r_i(t) f_0^{IN}(t) \tag{1}$$

ここで $r_i(t)$ は入力信号の瞬時周波数に加える変調量を指す．この変調により得られる新しい時間軸を $t^{IN}(t)$ とすると，以下のように形式的に表される．

$$x^{OUT}(t) = x^{IN} \int_0^t \frac{1}{r_(\tau)} d\tau \tag{2}$$

この時間軸の伸縮は，区分的一次関数による補間を用いて実装している．

3.2 近似時変フィルターによる時間変調:F

グロウル系統の歌唱音声のスペクトログラムには，基本周波数の周波数変動に同期した縦縞が認められる．この縦縞は，第 3 フォルマントの形状の高速な変動によるものであり，喉頭の上部構造の振動による声道形状の変化が原因であると考えられている [3]．ここでは対数スペクトルで表現したこの特徴を，時間的に変動する 3000 Hz 付近の −10〜−15 dB 程度の鋭い谷とやや緩いピークからなるペア $R_1(f)$ と，その変動と逆の位相で変動する 2000 Hz 付近の緩いピークと谷のペア $R_2(f)$ を用いて，次式のようにモデル化することとした．

$$R_1(f) = a_{p1} \exp\left(-\frac{(f-f_{p1})^2}{\sigma_{p1}^2}\right) - a_{d1} \exp\left(-\frac{(f-f_{d1})^2}{\sigma_{d1}^2}\right) \tag{3}$$

$$R_2(f) = a_{p2} \exp\left(-\frac{(f-f_{p2})^2}{\sigma_{p2}^2}\right) - a_{d2} \exp\left(-\frac{(f-f_{d2})^2}{\sigma_{d2}^2}\right), \tag{4}$$

ここで f_{p1}, f_{p2} は，ピークの周波数，σ_{p1}, σ_{p2} はピークの広がり，a_{p1}, a_{p2} は高さを表す．同様に，f_{d1}, f_{d2} は谷の周

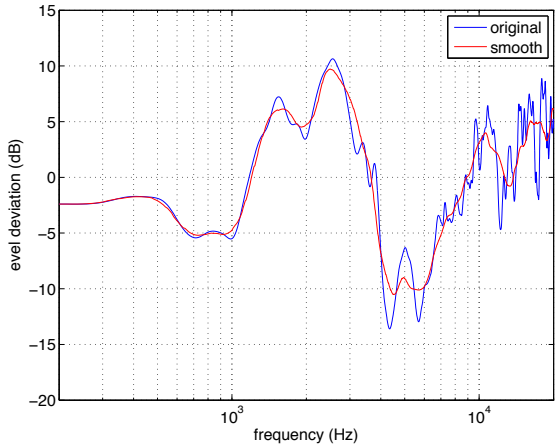

図 3 帯域強調処理に用いた FIR フィルタの周波数特性

波数，σ_{d1}, σ_{d2} は谷の広がり，a_{d1}, a_{d2} は谷の深さを表す．この 2 つの特性を，基本周波数の変調と同じ信号を値域が $[0, 1]$ になるように変換したモーフィング率 $r_m(t)$ によりモーフィングして基本周波数の変調と同じ信号を用いて補間し，$R_m(f, t)$ を作成した．

$$R_m(f,t) = r_m(t)R_1(f) + (1 - r_m(t))R_2(f) \qquad (5)$$

こうして作成した特性からフレーム毎の最小位相のインパルス応答を求め，入力信号 $x[n]$ を次式を満たすように分解した部分系列 $s^{(k)}[n]$ 毎に畳込み，OLA 処理により，時変フィルタの応答を近似的に求めた [6]．

$$x[n] = \sum_{k=-\infty}^{\infty} s^{(k)}[n] = \sum_{k=-\infty}^{\infty} x_1^{(k)}[n]w[n - kL] \qquad (6)$$

ここで，K は部分系列の長さ，L は部分系列のシフト長 (先頭位置の更新量)，$x_1^{(k)}[n]$ は，開始位置が $n = kL$ で長さが K の $x[n]$ から切出した系列を表す．具体的には，50%オーバラップの Hann 窓関数を用いて部分系列への分解を行った．

3.3 2000 Hz 付近の帯域の強調:E

図 3 の青線にグロウル系統の歌唱音声と通常歌唱のパワースペクトルのレベル差を示す．これからグロウル系統の歌唱では，2000 Hz 付近を中心として，ほぼ 1000 Hz から 4000 Hz の範囲が 10 dB 程度強調されていることが分かる．ここでは，図 3 の赤線が示す 1/3 オクターブ毎の長時間スペクトルの差を平滑化した特性を用いて FIR フィルタを設計し，この特徴を付加することとした．

3.4 非線形自励振動子

本研究ではグロウル歌唱特有の喉頭上部の励起振動を正弦波状の変化とそれに非線形振動子を加えたものの 2 つを用いて実装する．喉頭上部の励起振動は特に特徴 Q と特徴 E におけるそれぞれの時間振動やスペクトルの時間変調に大きな影響を与えていると考えられる．具体的には次式のように，正弦波状の振動子 x に強い歪みを無くした非線形振動子を組み合わせたものを用いる．

$$\frac{d^2x}{dt^2} + \epsilon\left(\left(\frac{dx}{dt}\right)^2 + x^2 - 1\right)\frac{dx}{dt} + x = 0 \qquad (7)$$

ϵ は非線形振動子における非線形性の減衰の強さを示す．ここで $y = \frac{dx}{dt}$ とすると 7 は以下のような形式で表すことが出来る．

$$y = \frac{dx}{dt} \qquad (8)$$

$$\frac{dy}{dt} = \epsilon\left(1 - y^2 - x^2\right)y - x \qquad (9)$$

また，式 7 に振動の ON/OFF をコントロールする $s(t)$ を導入すると以下のように変形出来る．

$$y = \frac{dx}{dt} \qquad (10)$$

$$\frac{dy}{dt} = s(t)\epsilon\left(1 - y^2 - x^2\right)y - x - (1 - s(t))\epsilon y \qquad (11)$$

ここで，式 11 は $s(t) = 1$ のとき，式 9 となり，正弦波状の非線形振動を始める．$s(t) = 0$ のとき，式 11 は非線形振動を止める．次に，振動子の時間進行に応じた振動周波数 $f_m(t)$ を制御する方法を考える．$f_m(t)$ の角周波数を $\omega_m(t) = 2\pi f_m(t)$ とすると，式 11 の時間的な振動周波数は $\omega_m(t)$ を用いて次のように直接制御することが出来る．

$$\frac{dx}{dt} = \omega_m(t)y \qquad (12)$$

$$\frac{dy}{dt} = \omega(t)\left(s(t)\epsilon\left(1 - y^2 - x^2\right)y - x - (1 - s(t))\epsilon y\right) + F(t) \qquad (13)$$

ここで $F(t)$ は外部の力を表す．実際の処理では入力信号の基本波を用いている．

4. 音声変換システムの GUI 設計

2 章でグロウル系統の歌唱音声の特徴を通常歌唱に付与する方法を紹介した．本研究におけるユーザが操作できるパラメタは数が多く，1 つ 1 つのパラメタの持つ意味も分かりにくい．それぞれの処理に対して入力を必要とするパラメタをまとめたものを図 4 に示す．

そこで，開発した GUI はパラメタ操作のしやすさを重視して設計している [7]．GUI は一覧性 (システム全体の機能の把握しやすさ) を重視したデザインとシステムの機能をツリー型の階層構造としたデザインの 2 つを設計した．本章ではそれぞれ設計した GUI の機能とねらいについて紹介する．

4.1 一覧性を重視した GUI

図 5 にグロウル音声への変換システムのためのプロトタイプ GUI を示す．図に示されているようにこの GUI は音

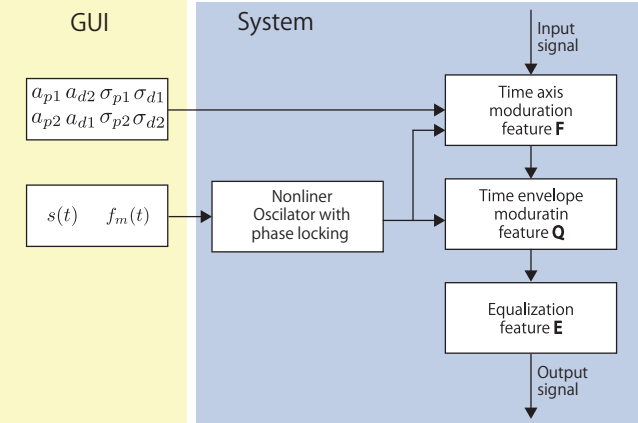

図 4 GUI で入力されるパラメタとシステムの関係

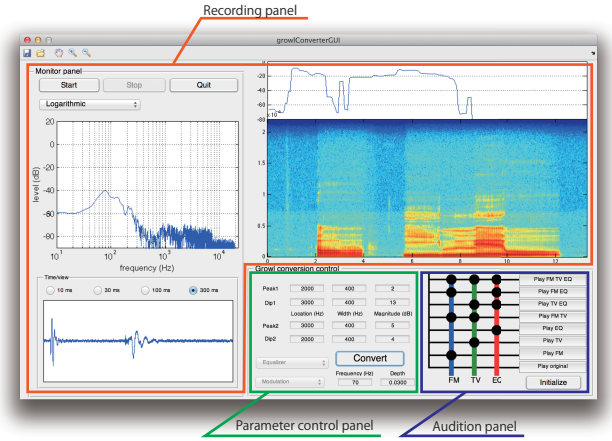

図 5 プロトタイプ GUI のスクリーンショット

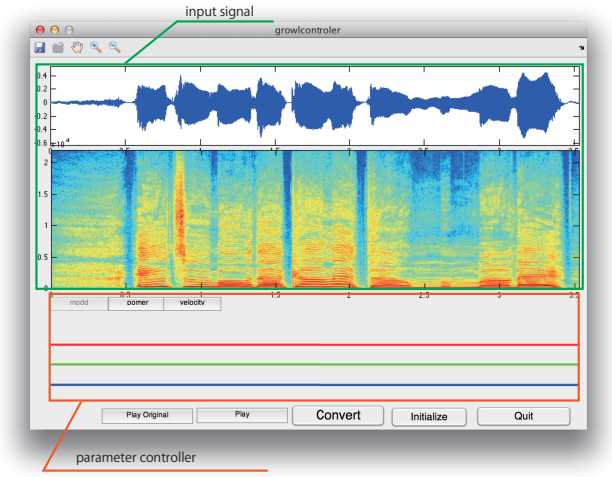

図 6 mainGUI のスクリーンショット

声録音部，パラメタ操作部，変換音声の試聴部の 3 つから成り立っている．また，音声の入力は WAVE ファイルでも可能となっており，変換音声は WAVE ファイルで書き出すことができる．

音声録音部の機能について紹介する．音声録音部では「Start」で録音を開始し，「Stop」で録音を停止する．録音中はそれぞれのグラフで録音中の音声の特性を調べることが出来る．音声録音部の左上のグラフでは FFT した入力信号をリアルタイムでプロットしている．また，左下は時間振幅を観察でき，横軸のスケールを 10ms, 30ms, 100ms, 300ms と変更することが出来る．右上はフレーム毎の振幅の最大値をデシベルで表現し，リアルタイムでプロットしている．右下は入力信号のスペクトログラムである．これは，信号の入力が終了すると，入力信号の分析が行われ，スペクトログラムが表示される．パラメタ操作部には近似時変フィルタの周波数特性と時励振動子のパラメタ設定が設けられている．変換音声の試聴部では 3 つの特徴付与処理の全組み合わせの音声を聴くことが出来る．それぞれ FM，TV，EQ が時間変調処理 (Q)，近似時変フィルタ処理 (F)，帯域強調処理 (E) にあたる．また，FM，TV，EQ を青，緑，赤のラインで示し，それぞれの試聴ボタンの横に，処理が適応されている特徴のライン上に黒丸をプロットしている．これはどの試聴ボタンが特徴付与処理の組み合わせに対応しているかを視覚的に理解しやすくするためである．

この GUI は 1 つのウィンドウで全ての機能表示しているため，ユーザは操作の全体量を把握しやすい．また，それぞれの特徴付与処理の組み合わせによる出力音声の違いを試聴することができる．次に，それぞれの機能別にウィンドウを分け，画面の簡潔性を重視した GUI 設計を紹介する．

4.2 機能を構造化した GUI

改変した GUI は主に時間的なパラメタの変化を操作できる「mainGUI」と音声の録音ができる「recordingGUI」，近似時変フィルタのパラメタ設定を行う「subGUI」の 3 つの画面から成り立つ．これは，1 つの画面で操作するパラメタの数を削減することによってインタフェースの情報量を減らし，ユーザの操作性を向上させるためである．また，全ての操作画面において，画面の下部にパラメタ操作，上部に GUI のフィードバックを集中させることで，操作画面に一貫性を持たせた．「recordingGUI」(図 7) に関する機能は前節で紹介したプロトタイプ GUI の音声収録部の内容と同じであるため，紹介を省く．

4.3 時間変化を伴うパラメタに対する対話的操作:mainGUI

「mainGUI」を図 6 に示す．「mainGUI」におけるパラメタ操作は，一番下のパラメタを実際にペンツールを用いて，パラメタの変化を描くことである．操作出来るパラメタは「自励振動子の周波数」，「自励振動子の ON/OFF」，「Q,F の時間変調の深さ」の 3 つである．また，パラメタと時間軸が同期した入力音声のスペクトログラムと時間波形をユーザに提供することで，実際の音声の情報を確認しながらパラメタ操作を行うことができる．

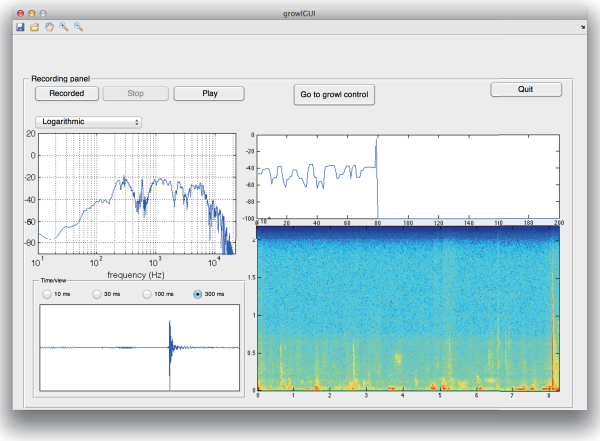

図 7 recordingGUI のスクリーンショット

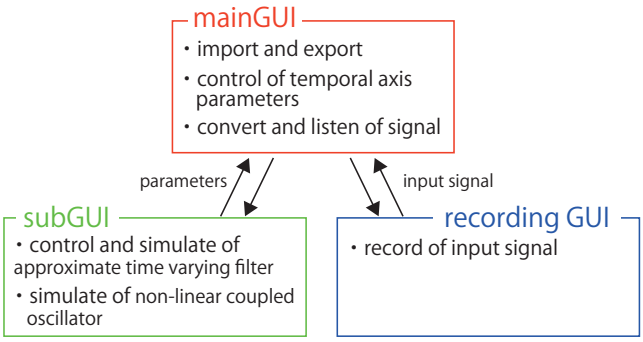

図 8 ツリー型のＧＵＩのシステム図

「mainGUI」では音声変換や音声ファイルの操作やその他の操作画面への移行を行うことができ，本システムは「mainGUI」を最上位にもつツリー型の階層構造となっている．図 8 に GUI のシステム図を示す．これは，全てのシステムの機能へのリンクを「mainGUI」に集約することで，システムの一覧性を持たせるためである．

4.4 近似時変フィルタにおける周波数特性の対話的操作:subGUI

式 3 と式 4 から分かるように，近似時変フィルターに用いる周波数特性の設定パラメタは非常に多い．そこで，フィルタの周波数特性を常に表示し，ユーザの操作に応じて，フィルタの周波数特性が変化する「subGUI」を設計した．これによって，ユーザの操作に素早くフィードバックを返すことで操作に対する変化をすぐに確認することができる．図 9 に実際に設計した GUI を示す．上部は GUI の下部に設置したパラメタをもとに式 3 と式 4 を用いて生成された近似時変フィルターの周波数特性である．下部のパラメタはそれぞれ自由に変更でき，変更に応じて上記の周波数特性も再プロットされる．今後は，上記の周波数特性の図を直接操作して周波数特性を変更出来るよう，よりインタラクティブな操作に改良する．

また，近似時変フィルタによる処理は周波数特性の時間

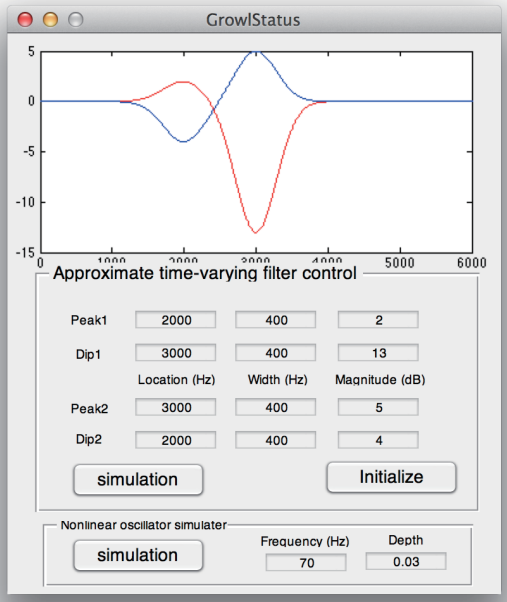

図 9 subGUI のスクリーンショット

変化を考慮してシミュレーションしなければならない．そこで，近似時変フィルタの時間的な変動を視覚的に捉えるために図 9 の下部にある「simulation」から図 10 のような時間的な特性変化を確認することができる．これは設定したパラメタから得られる近似時変フィルタの周波数特性の時間変化をスペクトログラムで表したものである．図から分かるように 0.05 秒間に 4 周期程度変化していることから近似時変フィルタは正常に機能していることがわかる．また，非線形振動子に対する直感的な理解を促すため，「Nonliner oscillator simulater」より非線形振動子のパラメタ操作による挙動の変化を観察出来るシミュレーターを実装した．図 9 で設定されている値で「simulation」から図 11 のような結果が出力される．赤線は非線形振動子の ON/OFF を制御する $s(t)$，青線は非線形振動子の応答を示す．図 11 から分かるように ON と OFF の間で滑らかに変化していることが分かる．

5. おわりに

今回は通常歌唱からグロウル歌唱への音声変換を対話的に操作出来る GUI を紹介した．開発した GUI はデモやポスターセッションの場で本手法による処理内容と処理の影響について直感的理解を促すことを目的としている．GUI は大きく分けて「音声録音」と「パラメタ操作」と「変換音声の試聴」の 3 つの機能を実装した．また，GUI は画面の一覧性を重視したものと画面を分け，それぞれの機能を構造化したものの 2 種類を作成した．画面の一覧性を重視した GUI では，すべての機能を 1 つの画面で実現し，GUI

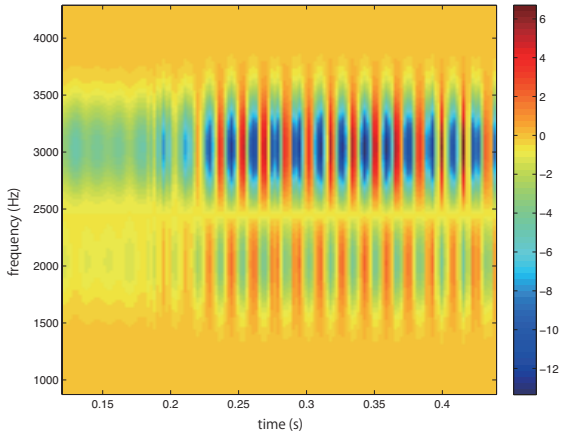

図 10 入力パラメタよりシミュレーションされた近似時変フィルタの時間特性

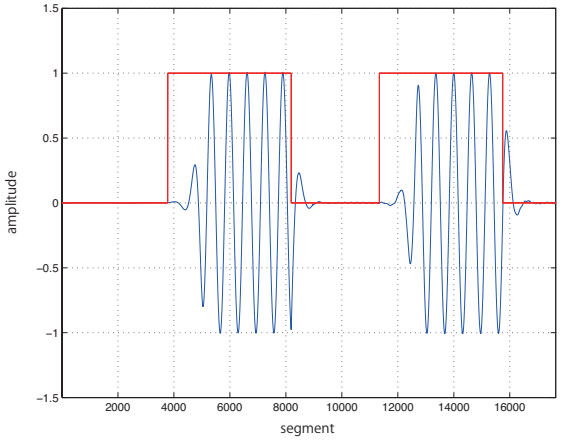

図 11 入力パラメタよりシミュレーションされた非線形自励振動子の概形

の機能が全てひとつの画面で完結するようにした．簡潔性を重視した GUI は時間的なパラメタ変化の操作とシステムの中心となる役割を果たす「mainGUI」と音声の録音を行う「recordingGUI」と近似時変フィルタの周波数特性を操作する「subGUI」の 3 つの画面から成り立っている．また，簡潔性を重視した GUI は「mainGUI」を最上位とツリー型の構成となっており，「mainGUI」に実装されていない機能は全て「mainGUI」から派生出来るようになっている．「subGUI」では，ユーザの操作に対して素早くフィードバックを返すことで，パラメタ操作による処理の変化を直感的に把握することができるように設計した．

提案手法であるグロウル歌唱の特徴付与は設定するパラメタが非常に多く，適切な設定を行うことが難しいため，操作パラメタを減らす必要がある．また，提案手法による変換音声は湿性嗄声のような印象があるなどの批判があり，品質をさらに向上させる必要がある．GUI の設計においてもまだ操作性やデザイン性について課題が多く，インタフェースの改良や市場にある DAW(Digital Audio Workstation) のデザインに関する調査など検討の必要がある．

謝辞 本研究の一部は，科学研究費基盤 (B)24300073 の支援を受けた．

参考文献

[1] B. Jordi, B. Merljin, 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, 音楽情報科学研究会, pp.1-6 (MUS), 東京 (2013).
[2] 河原英紀，溝渕翔平，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”, 音楽情報科学研究会, pp.1-6 (MUS), 東京 (2013).
[3] 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 2013*, pp.256-263,(2013).
[4] http://www.vocaloid.com/.
[5] Kawahara, Hideki and Morise, Masanori and Nisimura, Ryuichi and Irino, Toshio.: “ IDeviation measure of waveform symmetry and its application to high-speed and temporally-fine F0 extraction for vocal sound texture manipulation”, *Proc. Interspeech 2012*, O2d.05,(2012).
[6] Oppenheim, A. V. and Shafer, R. W. .: “Discrete-Time Signal Processing”, Prentice-Hall, Englewood Cliffs, NJ(1987).
[7] 山岡俊樹:人間工学講義, 武蔵野美術大学出版局 , pp.271-289(2002).