[招待講演] めざせ音声分析合成マスター！
－「よくわからない」から「ちょっとわかる」へのチュートリアル－

森勢 将雅†

†山梨大学大学院総合研究部 〒400-8511 山梨県甲府市武田 4-3-11

E-mail: †mmorise@yamanashi.ac.jp

あらまし Vocoder の考えに基づく音声分析合成技術は，研究用のツールとして広く利用されている．特に利用されている STRAIGHT は，音声から基本周波数(F0)，スペクトル包絡，非周期性指標を取り出し，それぞれのパラメータから音声波形を合成する機能を有する．F0 が高さであることは直感的だが，スペクトル包絡と非周期性指標に関しては，どのように変換すればどのような音色になるのかが分かりにくい．また，STRAIGHT を含む高品質音声分析合成技術については，中身をブラックボックスとする傾向があることも事実である．本講演では，音声分析合成において，それぞれのパラメータがどのように音色に影響しているかを説明し，利用者がスペクトル包絡や非周期性指標の中身を知るためのチュートリアルを行う．チュートリアルでは，筆者の開発した音声分析合成システムを利用するが，制御法に関する理論は，同一の構造を有する分析合成システム全般で利用可能である．

キーワード 音声分析合成，Vocoder，基本周波数，スペクトル包絡，非周期性指標，

Aim to be a speech analysis/synthesis master!
－I want to say that I understand a little－

Masanori MORISE†

†Faculty of Engineering, University of Yamanashi 4-3-11 Takeda, Kofu-shi, Yamahashi, 400-8511 Japan

E-mail: †mmorise@yamanashi.ac.jp

Abstract Speech analysis/synthesis systems on the basis of the ideal of Vocoder have been widely used, and several researchers can use them without enough knowledge on the principle of the systems. These systems estimate the fundamental frequency (F0), spectral envelope and aperiodicity from the speech signals and generate the signal with these three parameters. It is well-known that we can control the pitch by using the F0 information, but it is difficult to control the spectral envelope and aperiodicity with understanding in the relationship between the subjective timbre and them. The purpose of this article is to demonstrate this relationship and to give the knowledge for controlling the speech parameters to the users. This article uses a speech analysis/synthesis system named WORLD, but the knowledge in this article is generalized for being able to use other similar systems.

Keywords Speech analysis/synthesis, vocoder, fundamental frequency, spectral envelope, aperiodicity

1. はじめに

音声分析合成技術は，人間の音声知覚のメカニズムを解明する研究に有用である．とりわけ，Vocoder [1]の考えに基づく分析合成システムは，音声の特徴を段階的に変化させて知覚特性を計測する実験など様々な応用研究に利用可能である．これは，音声から人間の知覚する高さと音色に相当する音声パラメータを出力し，各パラメータから波形を合成できる特徴に基づく．

Vocoder の考えでは，音声の高さは基本周波数（F0），音色はスペクトル包絡と定義される．最近では，声の擦れの程度に相当する非周期性成分に関する 3 つ目の音声パラメータも利用される．初期の音声分析合成技術は，限られた通信能力と計算機能力で音声を効率よく伝達する観点で研究が進められており，品質が低いことが特徴として挙げられていた．1990 年以降の信号処理技術，計算機能力の発展に伴い，高品質な音声分析合成技術として STRAIGHT [2]が提案され，肉声に近い音声合成が可能となった．

STRAIGHT の発明は，音声モーフィング[3]等の新たな音声加工法の提案へと繋がり，音声分野に大きな発展をもたらした．さらに，聴覚分野における音声知覚メカニズムの解明に向けた，基盤ツールとしての側面も有する．すでに多数の事例があるが，例えば，音声のピッチや話速が印象に与える影響について調査されている[4]．聴覚特性の 1 つとして，聴覚は音を音源が有する寸法の情報を抽出可能であるという考察がなさ

れ，STRAIGHT を利用した寸法変化により検証された例もある[5]．音声知覚における Auditory adaptation の示す論文[6]など，STRAIGHT が基盤として利用された例は多数ある．品質の高さから，歌声の合成，変換技術にも利用されており，歌声のモーフィング[7]などの加工技術を支える基盤としても利用されている．

STRAIGHT の内容については，すでに複数の解説資料[8, 9]がある．しかしながら，その内容について把握することは容易ではなく，未だにブラックボックスとして扱われていることも多い．本講演の目的は，STRAIGHT をはじめとする高品質な音声分析合成技術について，各音声パラメータの位置付けや加工の際の問題点などを説明することである．アルゴリズムの詳細ではなく，各音声パラメータを制御することで具体的に合成時何が生じるのかなど，入門資料としての位置付けである．

2. 音声分析合成システムの構成

音声分析合成は Vocoder 以外にも，Phase vocoder [10]や Sinusoidal model [11]などが存在する．ここでは，STRAIGHT の基盤となる Vocoder (正確には Channel vocoder)方式に着目し，特に，STRAIGHT とその後継にあたる TANDEM-STRAIGHT [12,13]と WORLD [14]の 3 種をターゲットにする．上述の 3 種の方式は同様の機構を有するため，全種類を共通する呼称として STRAIGHT という用語を用いる．1999 年に提案されたものは Legacy-STRAIGHT とする．

2.1. 音声の定義と問題設定

STRAIGHT では，以下の数式により有声音$y(t)$が構成されていると仮定する．

$$y(t) = h(t) * x(t) + n(t), \tag{1}$$

$$x(t) = \sum_{k=-\infty}^{\infty} \delta(t - kT_0), \tag{2}$$

ここで，記号*は畳み込みを表し，$x(t)$は基本周期T_0の周期を有するパルス列，$h(t)$は声帯振動に相当するインパルス応答，$n(t)$は有声音中に存在する非周期的な雑音成分を表す．雑音成分が無い合成音声はブザー音的な音色（Buzzy）となるため，非周期性成分は音声の Buzzy さを低減するために重要となる．

Vocoder による音声分析では，音声波形$y(t)$から，3 つの音声パラメータを推定することを目指す．F0 は，基本周期の逆数として求める．スペクトル包絡については，声帯振動の波形$h(t)$ではなく，パワースペクトルのみが推定対象である．非周期指標についても，雑音成分$n(t)$そのものではなく，音声波形$y(t)$中の周期的な成分$h(t) * x(t)$と，非周期的な成分$n(t)$とのパワーの比として定義される．非周期性指標は帯域毎に異なるため，スペクトル状のパラメータである．

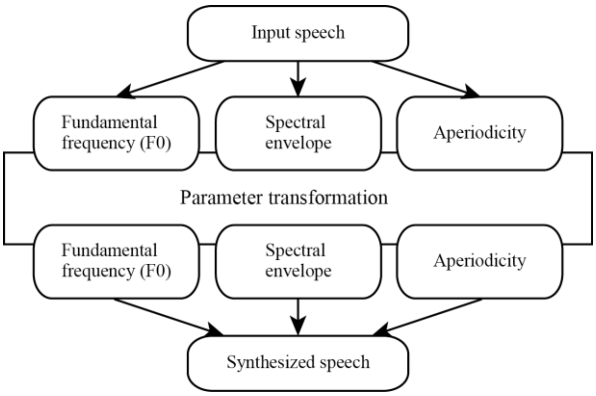

図 1：STRAIGHT の枠組み．

図 1 のように，STRAIGHT では，音声から 3 つの音声パラメータを推定するアルゴリズム，および 3 つの音声パラメータから波形を合成するアルゴリズムから構成される．以下では，Matlab のコードを含めて，各音声パラメータ推定の手順について概説する．なお，Legacy-STRAIGHT，TANDEM-STRAIGHT，WORLD のバージョンは，それぞれ STRAIGHTV40_006b，Tandem-STRAIGHTmonolithicPackage004TestRev ，v0.2.0_4 である．

2.2. F0 の推定

音声の F0 推定については，すでに膨大な研究事例が存在する広い研究領域である．STRAIGHT では，ある程度 SNR の高い音声が対象であり，混合音は非対象である．具体的に，Legacy-STRAIGHT の F0 推定は NDF [15]，TANDEM-STRAIGHT では XSX [12]，WORLD では DIO [16]というそれぞれ別の方法が採用されている．

音声波形を x，サンプリング周波数を fs とした場合，各システムでは以下のコマンドにより音声の F0 を推定可能である．以下，Legacy-STRAIGHT，TANDEM-STRAIGHT，WORLD の順にコードを記載する．

- [f0, ap] = exstraightsource(x, fs);
- f0 = exF0candidatesTSTRAIGHTGB(x, fs);
- f0 = Dio(x, fs);

x は波形，fs はサンプリング周波数に対応する．Legacy-STRAIGHT については，非周期性指標 ap も同時に推定される．TANDEM-STRAIGHT と WORLD の戻り値は構造体であり，メンバ変数 f0 が目的とする F0 の軌跡である．メンバ変数 temporalPositions (TANDEM-STRAIGHT)と temporal_positions (WORLD)は，f0 が推定された時刻を表す配列である．例えば，f0.f0(n)は，時刻 f0.temporal_position(n)秒の F0 を示す．Legacy-STRAIGHT については，配列の n 番目が n ミリ秒時の F0 に相当し，分析シフト量の指定はできない．TANDEM-STRAIGHT，WORLD の分析シフト量のデフォルト値は 5 ms である．

性能は，音声中の雑音量，ただし単純な SNR ではな

く音声中の非周期性成分を含む雑音量に依存する．開発者グループが非公式に行った実験では，NDF が低 SNR な音声に対しても高い精度で F0 が推定可能であることを確認している．静音環境で収録された音声では，DIO が NDF とほぼ等価な性能を達成している．分析速度については DIO が他手法より 1 桁以上高速に動作し，TANDEM-STRAIGHT，Legacy-STRAIGHT の順に遅くなる．ただし，Legacy-STRAIGHT の分析シフト量は 1 ms, TANDEM-STRAIGHT の分析シフト量は 5 ms であり，分析シフト量を揃えた場合は，TANDEM-STRAIGHT のほうが低速である．

2.3. 非周期性指標の推定

音声の非周期的な成分を扱う研究には，Mixed excitation [17]などの事例が存在する[18]．ただし，高品質音声合成のために提案された方法は少ないのが現状である．非周期性指標は，以下のコマンドにより推定される．Legacy-STRAIGHT は F0 と同時に推定されるため，ここでは省略する．

- source = aperiodicityRatioSigmoid(x, f0, 1, 2, 0);
- source = D4C(x, fs, f0);

source は構造体であり，メンバ変数には F0 を含む．Legacy-STRAIGHT と WORLD では，FFT 長に応じたスペクトル表現が結果として与えられるが，TANDEM-STRAIGHT の場合は，いくつかの帯域毎に推定を行い，帯域毎の結果に対してシグモイド関数でフィッティングを行い，そのパラメータを最終的な結果とする[19]．これは，音声は低域であるほど周期的で，高域になるほど非周期的になるという仮説に基づく．FFT 長は，後述するスペクトル包絡推定に用いる値と等しい．

音声の声帯振動は，時間的にも常に等間隔ではなく波形も毎回異なる．周期性を仮定し，周期性成分と非周期性成分のパワー比である非周期性指標を推定する場合，この声帯振動の揺らぎが結果に影響する．Legacy-STRAIGHT と TANDEM-STRAIGHT は，波形の F0 に基づいて時間伸縮を行い，F0 をフラットに変換してから推定する．Legacy-STRAIGHT では，全フレームの推定後，さらに時間方向への平滑化がなされる．

WORLD で採用しているアルゴリズム D4C [20]は，声帯振動の時間的な揺らぎに頑健なアルゴリズムを採用しているため，時間伸縮や推定後の平滑化を行うことなく非周期性指標を推定可能な特長を有する．分析合成音の品質評価については，Legacy-STRAIGHT と WORLD とがほぼ等価で，TANDEM-STRAIGHT がやや劣ることを確認している．HMM 音声合成に関して Legacy-STRAIGHT と WORLD の比較を行った実験例もあり，ほぼ同等の品質を達成している[21]．なお，有声音を無声音と誤推定した場合の品質低下は大きいが，非周期性指標推定が適切であれば，無声区間を有声区間と誤推定しても全周波数で非周期的であると推定されることから，有声音区間と判定する閾値を緩く設定することで品質が上がるという報告も存在する．

2.4. スペクトル包絡の推定

スペクトル包絡推定には，線形予測(LPC: Linear predictive coding) [22]やケプストラム[23]などの代表的な方法や，改良法が提案されている．従来の音声分析では，窓関数で波形を切り出し，スペクトル包絡を推定するが，推定結果は，毎回の声帯振動が不変にも関わらず，波形を切り出す時刻に依存して変化する．

STRAIGHT は，この分析時刻に依存する成分をそれぞれ独自に定式化し，除去するようデザインされている．類似研究として，中野らの取り組み[24]が存在するが，1 sample ごとの波形切り出しが必要など計算コスト面での課題が残されている．各システムでのスペクトル包絡は，それぞれ以下のコマンドで推定する．

- spec = exstraightspec(x, f0, fs);
- spec = exSpectrumTSTRAIGHTGB(x, fs, source);
- spec = CheapTrick(x, fs, source);

Legacy-STRAIGHT 以外は構造体で結果が与えられる．また，Legacy-STRAIGHT は振幅スペクトルだが，それ以外はパワースペクトルである．FFT 長は，F0 の下限とサンプリング周波数から自動的に決定される．

スペクトル包絡推定精度については，TANDEM-STRAIGHT が Legacy-STRAIGHT を上回るという結果が得られている[25]．合成音声の品質については，3 種ともに有意差が無く，サーストンの一対比較法では，高い順に WORLD で採用されている CheapTrick [26, 27]，TANDEM-STRAIGHT，Legacy-STRAIGHT であることが示唆されている．ただし，分析合成する音声との相性があるため，大局的にはどの方法にも大きな差が存在しないという報告も寄せられている．

2.5. 3 つの音声パラメータからの波形合成

波形合成部では，音声波形から得られた 3 つの音声パラメータを入力とし，以下のコマンドで波形を出力する．なお，TANDEM-STRAIGHT の戻り値は構造体であり，メンバ変数 synthesisOut が波形である．

- y = exstraightsynth(f0, spec, ap, fs);
- y = exTandemSTRAIGHTsynthNx(source, spec);
- y = Synthesis(source, spec);

合成処理は，(1) F0 軌跡から声帯振動が生じた時刻を計算，および(2) 各声帯振動が生じた時刻における有声音，無声音の合成の 2 ステップで構成される．

初めに，F0 軌跡から声帯振動の生じる時刻を推定する方法を述べる．F0 軌跡$f_0(t)$から，以下の式によりパラメータ$\theta_c(t)$を計算する．

$$\theta_c(t) = \int_0^t f_0(\tau)\, d\tau. \qquad (3)$$

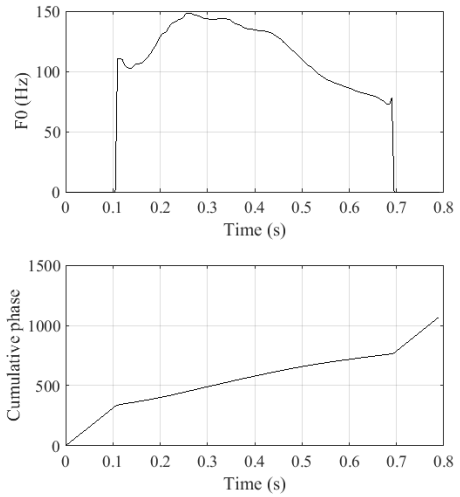

図 2：F0 軌跡（上段）と式(3)により得られた結果（下段）．無音区間の F0 は 500 Hz として計算している．

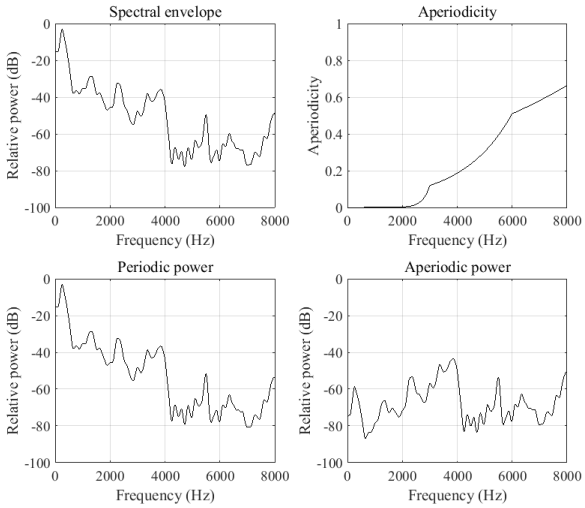

図 3：あるフレームのスペクトル包絡（上段左），非周期性指標（上段右），周期性スペクトル（下段左），非周期性スペクトル（下段右）．

図 2 に，ある音声を分析した結果の F0 軌跡$f_0(t)$と，パラメータ$\theta_c(t)$の例を示す．無声音に F0 は存在しないが，破裂音の合成で瞬時に音が発生することに対応するため，高い F0 の値（WORLD では 500 Hz）に置き換えて計算する．時刻 0 の値を初期値とし，縦軸が2π変動するのに要した時間間隔が基本周期となる．このアルゴリズムにより時刻 0 から声帯振動の生じる時刻を計算する．このアルゴリズムは，F0 制御においては，大局的な制御が重要であることを示唆する．微細変動は，毎回の声帯振動時刻を微細に変動させることになるため，品質を損なう原因となり得る．声帯振動の生じる時刻が得られた後は，各時刻について有声音と無声音の合成，および得られた結果を Overlap-add の考え方に基づいて加算する．

声帯振動の位相を推定していないため，位相はパワースペクトルから計算される最小位相とする．STRAIGHT では，スペクトル包絡をそのまま用いて有声音を合成するのではなく，スペクトル包絡$S_e(\omega)$と非周期性指標$ap(\omega)$から，周期性スペクトルを求めて利用する．スペクトル包絡と非周期性指標，周期性・非周期性スペクトルは以下の関係式となる．

$$S_e(\omega) = S_e(\omega)ap(\omega) + S_e(\omega)(1 - ap(\omega)), \qquad (4)$$

右辺の第一項が非周期性スペクトルであり，第二項が周期性スペクトルである．非周期性指標は 0 から 1 の範囲の値（Legacy-STRAIGHT は対数となっているため負の値）であり，非周期性指標が 0 であることは，スペクトル包絡が全て周期的であることを示す．

図 3 は，特定のフレームについて計算された，スペクトル包絡，非周期性指標，周期性・非周期性スペクトルを示す．非周期性指標は滑らかであるが，スペクトル包絡との乗算で周期性・非周期性スペクトルを計算するため，フォルマントのピークへの影響は小さい．

非周期性指標が品質に与える影響は，他の 2 つに比べると小さいため，非周期性指標をどのように与えるかについては，現在までに一定の結論には至っていない．ただし，D4C では 3 kHz 毎の中心周波数について計算し，0 Hz の値を-60 dB，ナイキスト周波数の値を 0 dB として与えて補間することにより，全離散周波数について値を求める Legacy-STRAIGHT と等価かやや上回る品質を達成している．これは，周期性・非周期性スペクトル形状の複雑さがスペクトル包絡側で決定するため，非周期性指標については，概形のみ推定できれば充分である可能性を示唆する．

3. 音声加工の実例と応用例

ここでは，比較的容易な変換法やそれを用いた研究事例を紹介する．

3.1. 話速制御

話速変化は，比較的容易に実装できる変換技術の 1 つである．Legacy-STRAIGHT ではフレームシフトが 1 ms に固定されているため，例えば話速を N 倍にする場合は各パラメータを時間方向に N 倍へ伸縮する必要がある．一方，WORLD では，source のメンバ変数 temporal_positions に時間の情報が格納されているため，話速を N 倍にしたい場合，temporal_positions を N 倍すれば良い．TANDEM-STRAIGHT も temporalPositions を N 倍することで発話速度を伸縮できる．spec メンバ変数にも temporal_positions は存在するが，こちらは合成時に利用されない．ただし，伸縮が線形の場合は子音部や調音区間も線形に伸縮されるため，変化率が大

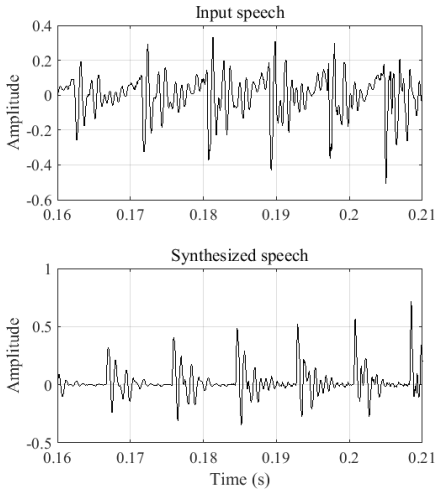

図 4：音声波形（上段）と合成波形（下段）．波形のエンベロープが異なる．

きい場合は自然性が低下する．自然性を保ったまま話速を変換するためには，調音速度や子音・母音区間を加味した非線形な伸縮が必要になる．

3.2. 寸法の制御

各フレームに対するスペクトル包絡を線形伸縮することは，寸法（声道長）の制御に対応する．線形伸縮を行う場合は，Matlab の interp1 関数を利用することで容易に実装可能である．また，スペクトル包絡の周波数伸縮は，対数パワーに対して行うことが望ましいといえる．STRAIGHT による合成音声は，フォルマントピークの鋭さが鈍ることで品質が劣化するため，対数パワーによる処理でこの影響を低減できる．

3.3. 声道断面積関数の制御

声道断面積関数(VTAF: vocal tract area function)を用いることで，声門から口唇までの声道形状を近似することが可能となる．近年では音声波形から VTAF を推定する技術が提案されている[28]. VTAF は全極スペクトルでの近似になるため，STRAIGHT により得られたスペクトル包絡から VTAF を推定すると，スペクトル包絡を VTAF 由来のものと残差に分離することになる．筆者らの検討では，音声の「はきはき」「もごもご」感には口の開き方の時間的な変化量が重要であることを示唆している[29]．VTAF 制御による声質変換[28]は，口の開き方に対応するため直感的である．

4. 音声分析合成システムの展望と限界

ここでは，現状の STRAIGHT で残された課題と音声分析合成技術の限界について述べる．

4.1. 分析合成方式に残された課題

音声分析合成には，音声波形の位相をどの様に扱うべきかという共通の課題が存在する．中野らの取り組み[24]はあるものの，概ね各声帯振動に相当するインパルス応答の位相には最小位相を利用している．一方，図 4 からも明らかに，実音声の波形と合成された波形のエンベロープは異なる．

聴覚は位相の違いを知覚することが可能である[30]．また，聴覚野の神経細胞応答を計測する研究では，波形のエンベロープにより反応を変化させる神経細胞応答の存在が示唆されている[31]．波形のエンベロープは，波形のエネルギーが時間的にどの程度散らばっているかに相当する．波形の時間的な散らばりはパワースペクトルと群遅延から求められるため，同一のスペクトル包絡に対するエンベロープは，群遅延操作で制御することが可能である[32]．知覚的に重要なことがエンベロープのみである場合，波形のエンベロープを制御しやすいような群遅延，あるいは位相のモデリングを行うことが新たな課題と言える．

4.2. 分析合成方式の限界

Vocoder の構造に基づく音声分析合成システムは，音声波形が有する位相情報の扱いに限界がある．また，声帯振動が周期的であるという前提で理論が構築されているが，実際の音声は声帯振動の生じる時間間隔がばらついており，声帯振動波形も毎回同一とはならない．短時間で分析を行うため，周期性の仮定は分析結果に大きな影響を与えないが，能[33]のような特殊発声では，この仮定が成立しない．また，グロウル・シャウトのような演奏表現においても，同様に現状の音声分析合成技術で解析することは不可能である．これらの音声を解析するためには，音声の周期性を仮定しない理論を構築し直すことが必要になる．

特殊発声を分析合成システムに入力して分析することは可能であり，周期性の逸脱も小さければ高品質な音声合成が可能である．Legacy-STRAIGHT，TANDEM-STRAIGHT，WORLD それぞれに相性の良い音声があることは，このような周期性の逸脱が原因であることが考えられる．

5. おわりに

本稿では，高品質音声分析合成システムとして STRAIGHT，TANDEM-STRAIGHT，WORLD の 3 つを対象とし，音声分析により得られるパラメータが合成時にどのように利用されるのかを説明した．特に，非周期性指標の扱いについて解説し，各パラメータが合成結果にどのような影響を与えるかについて述べた．応用研究に向けて，何を変換するとどのような結果が得られるのか「ちょっとわかる」ようになれば幸いである．

6. 謝辞

本研究は，科研費 15H02726，26540087，および東北大学電気通信研究所 共同プロジェクト（H25/A08）の支援を受けて実施された．

文　献

[1] H. Dudley,“Remaking speech,”J. Acoust. Soc. Am., vol. 11, pp. 169-177, 1939.

[2] H. Kawahara, I. Masuda-Katsuse, and A. De Cheveigné, “Restructuring speech representations using a pitch-adaptive time-frequency smoothing and an instantaneous-frequency-based F0 extraction,” Speech Communication, vol. 27, pp. 187-207, 1999.

[3] H. Kawahara and H. Matsui,“Auditory morphing based on an elastic perceptual distance metric in an interference-free time-frequency representation,” Proc. ICASSP2003, pp. 256-259, 2003.

[4] 内田照久，“音声の発話速度の制御がピッチ感及び話者の性格印象に与える印象，”音響学会誌，vol. 56, pp. 396-405, 2000.

[5] D. R. Smith, R. D. Patterson, R. Turner, H. Kawahara, and T. Irino,“The processing and perception of size information in speech sounds,”J. Acoust. Soc. Am., vol. 117, pp. 305-318, 2005.

[6] S. R. Schweinberger, C. Casper, N. Hauthal, J. M. Kaufmann, H. Kawahara, N. Kloth, D.M.C. Robertson, A. P. Simpson and R. Zäske,“Auditory Adaptation in Voice Perception,”Current Biology, vol. 18, pp. 684-688, 2008.

[7] M. Morise, M. Onishi, H. Kawahara, and H. Katayose, “v.morish'09: A morphing-based singing design interface for vocal melodies,” Lecture Notes in Computer Science, LNCS 5709 (in Proc of ICEC 2009), pp. 185-190, 2009.

[8] 河原英紀，“Vocoder のもう一つの可能性を探る－音声分析変換合成システム STRAIGHT の背景と展開－，”日本音響学会誌，vol. 63 pp. 442-449, 2007.

[9] H. Kawahara,“STRAIGHT, Exploration of the other aspect of VOCODER: Perceptually isomorphic decomposition of speech sounds,”Acoustic Science and Technology, vol. 27, pp. 349-353, 2006.

[10] J. L. Flanagan and R. M. Golden,“Phase vocoder,” Bell System Technical Journal, vol. 45, pp. 1493-1509, 1966.

[11] R. McAulay and T. F. Quatieri,“Speech analysis/synthesis based on a sinusoidal representation,”IEEE Trans. Acoust., Speech, Sig. Process. vol. 34, pp. 744-754, 1986.

[12] H. Kawahara, M. Morise, T. Takahashi, R. Nisimura, T. Irino and H. Banno,“TANDEM-STRAIGHT: A temporally stable power spectral representation for periodic signals and applications to interference-free spectrum, f0, and aperiodicity estimation,”Proc. ICASSP 2008, pp. 3933-3936, 2008.

[13] H. Kawahara and M. Morise,“Technical foundations of TANDEM-STRAIGHT, a speech analysis, modification and synthesis framework,”SADHANA - Academy Proceedings in Engineering Sciences, vol. 36, pp. 713-728, 2011.

[14] http://ml.cs.yamanashi.ac.jp/world/ 最新版は Web で公開しており，最新版のシステム全体をまとめた資料はまだ存在しない．

[15] H. Kawahara, A. de Cheveigne, H. Banno, T. Takahashi and T. Irino,“Nearly Defect-free F0 Trajectory Extraction for Expressive Speech Modifications based on STRAIGHT,” Proc. Interspeech2005, pp. 537-540, 2005.

[16] 森勢将雅，河原英紀，西浦敬信，“基本波検出に基づく高 SNR の音声を対象とした高速な F0 推定法，”電子情報通信学会 論文誌 D，vol. J93-D, pp. 109-117, 2010.

[17] A. V. McCree and T. P. Barnwell III,“A mixed excitation LPC vocoder model for low bit rate speech coding,”IEEE Trans. on Speech Audio Process., vol. 3, pp. 242-250, 1995.

[18] D. W. Griffin and J. S. Lim,“Multiband excitation vocoder,”IEEE Trans. on Acoust. Speech, and Signal Process., vol. 36, pp. 1223-1235, 1988.

[19] H. Kawahara and M. Morise,“Simplified aperiodicity representation for high-quality speech manipulation systems,”Proc. ICSP2012, pp. 579-584, 2012.

[20] 森勢将雅，“帯域毎の非周期性指標推定法とその誤差評価，”信学技報，vol. 115, pp. 13-18, 2015.

[21] 高道慎之介，戸田智基，森勢将雅，中村哲，“HMM 音声合成における音声分析合成器 STRAIGHT と WORLD の比較，”音講論(秋), pp. 271-272, 2015.

[22] B. S. Atal, S. L. Hanauer,“Speech analysis and synthesis by linear prediction of the speech wave,”J. Acoust. Soc. Am., vol. 50, pp. 637-655, 1971.

[23] A. V. Oppenheim,“Speech analysis-synthesis system based on homomorphic filtering,”J. Acoust. Soc. Am., vol. 45, pp. 458-465, 1969.

[24] T. Nakano and M. Goto,“A spectral envelope estimation method based on f0-adaptive multi-frame integration analysis,”Proc. SAPA-SCALE2012, pp. 11-16, 2012.

[25] 赤桐隼人，森勢将雅，入野俊夫，河原英紀，“スペクトルピークを強調した F0 適応型スペクトル包絡抽出法の最適化と評価，”信学論 A, vol. J94-A, pp. 557-567, 2011.

[26] M. Morise,“CheapTrick, a spectral envelope estimator for high-quality speech synthesis,”Speech Communication, vol. 67, pp. 1-7, 2015.

[27] M. Morise,“Error evaluation of an F0-adaptive spectral envelope estimator in robustness against the additive noise and F0 error,”IEICE transactions on information and systems, vol. E98-D, pp. 1405-1408, 2015.

[28] A. Arakawa, Y. Uchimura, H. Banno, F. Itakura, and H. Kawahara,“High quality voice manipulation method based on the vocal tract area function obtained from sub-band LSP of straight spectrum,” Proc. ICASSP2010, pp. 4834-4837, 2010.

[29] M. Morise, S. Tsuzuki, H. Banno, and K. Ozawa, “Muffled and brisk speech evaluation with criterion based on temporal differentiation of vocal tract area function,”IEICE transactions on information and systems, vol. E97-D, pp. 3230-3233, 2014.

[30] R. Promp and H. J. M. Steeneken,“Effect of phase on the timbre of complex tones,”J. Acoust. Soc. Am., vol. 46, pp. 409-421, 1969.

[31] 森勢将雅，大久保快走，地本宗平，佐藤悠，小澤賢司，“ソース・フィルタ型音声合成における有声音の位相が聴覚野の神経細胞応答に与える影響について　～覚醒ネコ第一次聴覚野の神経細胞応答に基づく検討～，”信学技報，vol. 114, pp. 41-46, 2014.

[32] L. コーエン，“時間-周波数解析，”朝倉書店, 1998.

[33] O. Fujimura, K. Honda, H. Kawahara, Y. Konparu, M. Morise and J.C. Williams,“Noh voice quality,” Logopedics Phoniatrics Vocology, vol. 34, pp. 157-170, 2009.