Vol.2012-MUS-94 No.8
Vol.2012-SLP-90 No.8
2012/2/3

歌声のテクスチャに信号処理はどう迫るか

河　原　英　紀[†1]

歌唱では、声の表現の可能性が極限まで追求される。通常の会話音声や朗読のような、ほぼ周期的な駆動を前提とした方法では、シャウト等の強い表現を自由に加工することは難しい。ここでは、TANDEM-STRAIGHT の拡張に向けた最近の検討を具体例として挙げながら、この『歌声のテクスチャ』に対する信号処理からの様々なアプローチについて紹介する。

Signal processing for handling singing voice texture

Hideki Kawahara[†1]

Singers explore vocal expressions to the limit. Conventional speech processing algorithms, which were designed to handle usual conversational speech and read speech, where voiced sounds are usually quasi periodic, are not capable of flexible and versatile manipulation of such extreme expressions as shout, for example. This article discusses various signal processing approaches for handling “singing voice texture,” referring recent investigations for extending TANDEM-STRAIGHT as an example testbet.

1. はじめに

プロの歌声の表現の豊かさ[1),2)] には、畏怖を覚える。[⋆1]心を鷲掴みにされるような強烈な叫びから、不用意に触ると脆くも崩れてしまいそうな繊細な響きまで、実に多様な声の可能性が追求されている。また、プロの歌声に及ばないにしても、一人一人の歌声には、それぞれに直接聴き手に働きかけるものがある。それらの豊かな表情を分析し自由に操作するための基盤を提供することは、歌声情報処理の重要な課題の一つである。ここでは、自由に『歌声を見て触る』[3)] ことを可能にするための信号処理技術について、前回紹介することができなかった[⋆2] 課題について、未整理な状況も併せて紹介したい。

現在では、歌声を扱うことができる様々なシステムが開発されている[4)–6)]。以下では、議論がそれらのシステムにも通じる一般的なものになるように配慮しつつ、基盤を提供するための信号処理について説明する。なお、具体例に基づいて説明する際には、音声分析変換合成法 TANDEM-STRAIGHT[7)] および拡張されたモーフィング[8)] を主に用いる。

2. 歌声のテクスチャ

一様に広がる草原の写真の一部を矩形に切り取り、隣接する重ならない矩形に含まれる画像と画素毎に比較すると、一般には一致しない。しかし、両者は同じ印象を与える。同様に、一様に降り続く雨音の録音からある長さの区間を切り取り、重ならない同じ長さの区間の信号と標本値毎に比較すると、一般には一致しない。しかし、両者は同じ印象を与える。この同じ印象を与える何かをテクスチャと呼ぶことにする。

画像が先行していたテクスチャ研究の分野に、最近、音に関する興味深いアプローチ[9),10)] が提案された。この研究では、波形情報を直接用いずに、統計的なパラメタによって、様々な環境音と同じ印象を与える音の合成に成功している[11)]。ただし、この方法で扱うことができる信号は、雨音やせせらぎのような非周期的なものに限られている。歌声のようにほぼ周期的な信号は、この方法では扱うことができない。以下では、聴覚と音声生成の両方から、この制限を超えて、歌声のテクスチャに迫る方法を考えて見たい。まず、TANDEM-STRAIGHT の現状を簡単に紹介する。

3. TANDEM-STRAIGHT

有声音のようにほぼ周期的な音の印象は、信号のパワー、パワースペクトルの概形、基本周波数およびそれらの時間変化に強く影響される。TANDEM-STRAIGHT（およびその前身である STRAIGHT）の基本的なアイデアは、有声音における周期的な駆動を、背景にある滑らかな時間周波数表現を組織的に標本化するためのものと解釈するところにある[12),13)]。このアイデアに基づいて、分析位置による変動の無いパワースペクトルの計算法[14)] と、新しい観点に基づく標本化理論[15)] による補償とを組合わせることで、周期的な駆動に起因する変動が排除された滑らかな時間周波数表現が復元される。その後、$|x| << 1$

†1 和歌山大学
Wakayama University

⋆1 様々なメディアやライブ、演奏会での経験も、当然含まれている。

⋆2 問題が難しかったので、今でもまだ十分に整理できてはいない。もう一つの続編が必要になりそうである。

Vol.2012-MUS-94 No.8
Vol.2012-SLP-90 No.8
2012/2/3

の場合に $\log(1+x) \approx x$ が良い近似となることを利用して、処理を次のような cepstrum 上での liftering と指数変換の組合せとして実装することで、処理結果の品質向上を図って現在に至っている[16),17)]。

$$P_{ST}(\omega) = \exp\left(\mathcal{F}\left[g_1(q)g_2(q)C_T(q)\right]\right), \tag{1}$$

ここで $P_{ST}(\omega)$ は、復元された滑らかなスペクトルを表す。また、$C_T(q)$ は時間的な変動の無いパワースペクトル表現 $P_T(\omega)$ の cepstrum であり q は quefrency. を表す。記号 $\mathcal{F}[\,]$ は Fourier 変換を表す。ここで用いられている 2 つの lifter$g_1(q)$ と $g_2(q)$ は次式で定義される。

$$g_1(q) = \tilde{\alpha}_0 + 2\tilde{\alpha}_1 \cos\left(2\pi q f_0\right), \tag{2}$$

$$g_2(q) = \frac{\sin(\pi f_0 q)}{\pi f_0 q}, \tag{3}$$

ここで f_0 は、基本周波数を表す。$g_2(q)$ は、幅が f_0 の矩形の平滑化関数に対応する。$g_1(q)$ は consistent sampling に基づく補償用ディジタルフィルタ[15)] に対応するが、ここでは、理論値の代わりに $\tilde{\alpha}_0 = 1.18, \tilde{\alpha}_1 = -0.09$ とすることで、知覚される品質に影響が大きいスペクトルピーク周辺での形状の改良の手段として用いている[16)]。

3.1 駆動音源とテクスチャ

こうして求められた信号のパワー、パワースペクトルの概形、基本周波数およびそれらの時間変化は、再合成音の印象の大部分を決定する。TANDEM-STRAIGHT は、通常は VOCODER 型のシステムとして用いられ、基本周波数に対応したパルス列とスペクトル整形された雑音による混合音源により駆動される。この混合音源の周期成分と非周期成分の配分など（加えて時間的な包絡の形状）が、さらに印象を変化させる[18)]。歌声のテクスチャを議論する場合には、このような主に音源に関わる印象の操作が中心となる。

ところで、滑らかな時間周波数表現の利用法は、VOCODER 型のシステムに限られない。次のような正弦波モデル[19)] を考えた場合、成分となる正弦波の振幅の制御に $P_{ST}(\omega, t)$ を用いることができる。この式に含まれる位相項が、上で説明した以外のさらに新たな印象操作の可能性を提供する[20)]。

このモデルでは、再合成音 $x_S(t)$ は次式により与えられる。

$$x_S(t) = \sum_{k=1}^{N} a_k(t) \sin\left(\int_0^t \omega_k(\lambda)d\lambda + \varphi_k(t)\right) \tag{4}$$

ここで，N は成分の個数を表す．$\omega_k(t) = 2\pi f_k(t)$ は，k 番目の成分の瞬時角周波数[⋆1]を表しており，調波複合音の場合には，基本（瞬時）周波数 $f_0(t)$ の k 倍となる．$\varphi_k(t)$ は、k 番目の成分の位相を表す。また，$a_k(t)$ は k 番目の調波成分の瞬時振幅であり，次式により決められる．[⋆2]

$$a_k(t) = \sqrt{P_{ST}(k\omega_0(t), t)/\omega_0} \tag{5}$$

以下では、駆動音源を中心に、テクスチャの信号処理を考えていくことにする。

4. 音声の生成と知覚

声のテクスチャの加工では、生成と知覚の両面からの検討が必要となる。以下では、それぞれの側面と信号処理との関わりを概観する。

4.1 音声の生成

やや古いが、発声の生理、物理から臨床までの基礎を科学的に扱った良書として、文献 23) を挙げておく。[⋆3]詳しくは文献にゆずり、ここでは有声音のテクスチャに関連する性質について簡単に説明する。

有声音は、声帯と呼気流の相互作用により生ずる声帯の振動が、呼気流を変調することにより生み出される。文献 23) にあるように、安定した振動が持続できる条件は、広くはない。振動が停止している領域と、安定した振動が持続できる領域との間には、異なった長さの周期が交互あるいは複数個毎に生じることで振動が二重周期や三重周期になる領域や、振動がカオスになる領域が挟まることがある。さらに、これらの同期した変動に加え、強い歌唱表現の場合には、声帯の上部構造の振動による非同期な変調が加わることもある。

有声音の基本波など低次の調波成分は、声帯の開閉運動自体に大きく影響される。一方、高次の成分は、主に声門の閉止に伴う呼気流の流速の不連続によりエネルギーが供給される[26),27)]。この不連続の直後では声門が閉じているため、声道形状により定まる共鳴特性が素直に音声波形に反映される。この性質により、声門閉止時刻の精密な推定は、音声信号処理の重要な課題となっている。文献 28) では、様々な方法が紹介され比較されている。

紹介されているものの中で重要なものに、位相の周波数方向の導関数として定義される群遅延に基づく方法がある。群遅延は、それぞれの周波数におけるエネルギーの時間方向の重

⋆1 位相の時間方向の導関数として定義される瞬時周波数の概念とそれを利用した基本周波数抽出法を説明するためのムービー (DVD イメージ) が、リンク先に用意されている[21)]。右上のムービが、それである。

⋆2 この他に、破裂音などの突発現象の表現も、印象に影響を与える[22)]。なお、これらについての検討結果は、現在の TANDEM-STRAIGHT には、反映されていない。

⋆3 著者の Titze は、現在も NCSV (National Center for Voice & Speech)[24)] の所長として、vocologist（我国の言語聴覚士と一部重複）の教育と研究の指導的な立場にある。1992 年に、リンク先[25)] にあるような生成モデルに基づいた歌唱合成の素晴らしいデモが著者により行なわれている。たまたま、筆者は、このデモの会場に居合わせていた。

心と解釈することができる。音声のパワースペクトルが因果性を満たす系の応答であることを仮定し、求められた最小位相応答の群遅延を用いて観測結果を補償すると、(男性の地声の場合の多くでは) 高域の群遅延は、ほぼ直線となる[29]。これは、ある瞬間に、それらの周波数において同時にエネルギーが供給されたことを意味する[⋆1]。

もう一つの重要なものに、零周波数フィルタリングにより求めた基本波の零交差のタイミング[28] がある。零周波数フィルタリングの提案者の主張とは異なるが、この方法は、瞬時周波数を求める際に用いる、調波を単離できる帯域通過フィルタの代わりに、等価な通過帯域がその 2 倍となる低域通過フィルタを使う方法と考えることができる。帯域幅の拡大は、時間分解能の向上につながる。この高い時間分解能は、二重音声や喉頭の上部構造の振動による変調などで生ずる周期の急速な変化の検出と表現の際に役立つ。

4.2 音声の知覚

知覚される音の高さであるピッチの形成に必要となる時定数は大きい。[⋆2]言語情報を伝えるだけであれば、モーラ毎の平均値を再現できれば十分であり基本周波数の軌跡を精密に再現する必要はない。しかし、基本周波数軌跡の様々な変動は、ビブラートや、声の震え、粗さなど、ピッチ以外の属性の違いとして知覚される。テクスチャを論ずる場合には、これらの再現に必要な物理的属性を明らかにすることが必要となる。

調波複合音の成分の相対的位相は、知覚される音色に影響を与える[30),31)]。式 4 の正弦波モデルに基づいて、それぞれの成分の瞬時振幅に同じのものを用いた場合でも、位相項を変えることにより音声の印象が大きく変わる。全ての調波の正弦波の位相が 0 度 (正弦波) の場合と、90 度 (余弦波) は、非常に近い印象を与える。奇数番目の成分と偶数番目の成分をそれぞれ正弦波と余弦波とした場合には、大きく印象が異なる。Schroeder 位相[32] とした場合には、チャープ状の印象があり、それぞれの調波の位相を乱数により決めた場合には、摩擦音が含まれるような印象となる。ただし、この効果は、高い基本周波数の場合には消失する。[⋆3]

例えば音声信号の逆フィルタ処理により求められる駆動信号の非周期成分は、基本周期内に時間的に一様に分布している訳ではない。有声音などの周期信号に短時間のバースト状の信号を加えた場合、バースト信号と周期信号の相対位相により、マスキングの閾値が大きく変化する。この効果は、基本周波数が低い場合に顕著であり、一周期内での閾値の変化が 20 dB に達する場合がある[33]。見方を変えると、SNR が 20 dB 異なった信号が、同じものとして知覚される可能性があることになる。相対位相の場合と同様に、この効果は、高い基本周波数の場合には消失する。

5. テクスチャに迫る信号処理

歌声のテクスチャの研究には、いろいろな立場があり得る。ここでは、STRAIGHT や TANDEM-STRAIGHT の研究と同じ立場に立ち、発声機構を考慮した逆問題を解くのではなく、聴覚が同じと判断するものを同じとするような表現を追求する。

このような立場から、テクスチャの研究に必要となる信号処理を整理すると、以下のようにまとめることができる。(滑らかな時間周波数表現は、別に求められるものとする。) それらは、1) 時間分解能の高い、基本周期の検出器。2) 駆動信号の主要な周期に同期／非同期の振幅／周波数変調の分析と記述。3) 周期成分と非周期成分の分離。4) 非周期成分の基本周期内分布の分析。5) それらの変動のモデル化とパラメタ抽出。などである。現時点では、これらの全てを体系的に統合することはできていない。[⋆4] 以下では、将来の統合の基礎となるそれぞれの要素技術について、現状を紹介する。

5.1 時間分解能の高い基本周期の検出

音楽の検索や音声認識への応用であれば、人間が知覚するピッチのように、大きな時定数の平滑化あるいはダイナミクスを仮定したモデルに基づく推定により、統計的に安定した値を求めることが望ましい。しかし、テクスチャの分析と再現のためには、ピッチ以外の属性の再現が必要であり、上記の平滑化は弊害が多い。基本周期の早い変動に追従することのできる高い時間分解能が必要となる。

TANDEM によるパワースペクトルは、同一の時間変動を許容した場合、より短い持続時間の窓の利用を可能にする[7]。これを応用した基本周波数を含む周期構造の抽出法である XSX (eXcitation Structure eXtractor) も、この高い時間分解能を受け継ぐ。基本周波数の時間的変動に対する追従性能を変調周波数伝達特性として評価した場合、XSX は、従来の基本周波数推定法 (ここでは YIN[34] と SWIPE′[35] を比較対象とした) を大きく凌ぐ[36]。

前節で説明した低域通過フィルタを用いた基本波の零交差に基づいて、非常に高速な基本

⋆1 先に紹介した DVD イメージ[21] に、実際の音声での群遅延と、補償された群遅延の振舞いを示すムービーが収録されている。右下のムービーがそれである。
⋆2 定義に依存するが、50 ms から 100 ms のオーダーと考えて良い。
⋆3 http://www.wakayama-u.ac.jp/ kawahara/phaseEffect/ にデモがある。
⋆4 実際のところ、1 と 2 は、何とか目処が立ちつつある。3 は、何度も挑戦し、暫定解を出してはいる。しかし、この解が求めるものだとは思えない。4、5 は、まだアイデアの域を出ない。

周波数推定法が開発されている[37]。ただし、この方法で用いられているフィルタ選択の指標は、折角の高い時間分解能を少し劣化させてしまい、勿体ない。選択の指標を、波形の対称性からの外れに基づくものとすることにより、この時間分解能の劣化を軽減することができる[38]。

基本波の零交差に基づく方法では、サイドローブを経由して漏れて来る他の調波成分からの影響が、推定精度の悪化に直結する。音声では調波のレベルのダイナミックレンジは、比較的近接した成分間でも 60 dB に及ぶ場合もある。そのため、60 dB よりも十分にレベルが低下したサイドローブを有し、しかも、周波数とともに、急速にサイドローブのレベルが低下する低域通過フィルタが必要となる。ただし、通過帯域の平坦性は、必要ではない。この条件を満たすために、基本波の零交差に基づく方法では、サイドローブのレベルが-90 dB 以下であり、その漸近的傾斜が-18 dB/oct である cos 級数の窓関数[39] をインパルス応答として用いることとした。

こうして求められた基本周波数の値は、その値を初期値として、複数の調波の瞬時周波数を求めて統合することにより改良される。なお、ここでの瞬時周波数の計算には、周期性に起因する変動を生じない新しい方法[40] が用いられている。この方法も TANDEM と同様の手法を用いており、同量の変動を許容する場合には、より時間分解能が高い窓を用いることができる。

5.2 駆動信号の変調構造の分析

説明が前後するが、XSX では、特定の周期を仮定して設計された複数の周期性検出器を対数周波数軸上に等間隔に配置する構造により、二重音声のように複数の周期が同時に存在する現象の検出を可能にしている。[⋆1] XSX では、候補となる周期成分と、それぞれの確からしさが併せて求められる。これまでに、XSX は、能の謡いにおける二重音声を用いた表現の記述[41] や、障害音声の分析に応用されている[36]。予備的な検討によれば、主な駆動の周期とは非同期に変化する早い変調を加えることで、シャウトなどの強い表現の特徴的なテクスチャを表現できそうである。

5.3 非周期成分の推定

調波成分の和あるいは周期的パルス列により駆動された応答により記述することのできない成分は、非周期成分としてまとめられる。この非周期成分の推定には、既に様々な方法が提案されている[18),19),42),43]。いずれの方法も、定常的な部分での問題は少ない。基本周波数あるいはスペクトル概形等、いずれかが急速に変化する場合には、様々な問題が生ずる。

例えば、基本周波数が急速に変化する場合には、特に基本周波数の低い男性の場合に、高い周波数において生ずる側帯波により非周期成分が過大に推定される[29]。この問題は、基本周波数の瞬時周波数に比例して時間軸を伸縮させることで見かけ上の基本周波数を一定とする方法で回避することができる。しかし、一方では、見かけ上のスペクトル概形が伸縮されるため、それによる副作用が生ずる。冗長ではあるが、固定された時間軸により求められた非周期成分と、伸縮された時間軸により求められた非周期成分を併用することが一つの解決策であり。

現在の TANDEM-STRAIGHT では、こうして求められた帯域毎の非周期成分のレベルをそのまま用いるのではなく、モーフィングへの応用を考慮したモデル化を行い、そのパラメタを用いている。具体的には、周期成分と非周期成分の境界周波数の値と、境界での周期成分から非周期成分への変化の傾斜をパラメタとして、sigmoid を当てはめている[44]。

なお、これまでの TANDEM-STRAIGHT の開発での経験では、二重音声や急速な基本周波数の変動、有声音の開始および終了部分での処理、また、それらの部分で生じ易い、不規則な声帯振動の処理など、実装レベルでの作り込みに、品質が依存する傾向がある。

5.4 非周期成分の基本周期内での分布

非周期成分の時間方向の包絡を整形方法と、合成音声の品質への影響が検討されている[45]。ここでは、Hilbert 変換による包絡と、エネルギーに基づく包絡が、三角形に整形した包絡よりも高い品質となるという結果が得られている。音声の知覚で説明した聴覚の特性を考慮すると、特に男性の音声について、より系統的な検討が必要になる。ただし、TANDEM-STRAIGHT では、まだ、これらの検討を進めてはいない。

5.5 テクスチャに関わる変動のモデル化とパラメタ抽出

テクスチャのモーフィングでは、時間分解能の高い基本周期の検出結果を直接用いるのではなく、補間性の良い表現とすることが必要になる。ビブラートのモデル化[?),46] と同様の検討が、今後の課題となる。

6. おわりに

歌唱音声のテクスチャの表現と操作は、歌声情報処理の新しい重要な課題である。この困難な課題への挑戦に、様々なアイデアで参加する方々が多数表れることを期待したい。特に、簡単に触れただけになっているテクスチャに関わる変動のモデル化は、統計的な手法の得意な、若手の方々の力をぜひ発揮して頂きたい。

⋆1 http://www.wakayama-u.ac.jp/%7ekawahara/HowTANDEMSTRAIGHTworks/に、説明用のムービーがある。

謝辞 TANDEM-STRAIGHT は、様々な支援と利用する方々からの刺激を受けて、現在も変化し続けている。直近では本研究の一部は、挑戦的萌芽研究 22650042 の支援を受けている。

参 考 文 献

1) 中山一郎：日本語を歌・唄・謡う，Audio CD 株式会社アド・ポポロ (2002).
2) 後藤真孝，橋口博樹，西村拓一，岡 隆一：RWC 研究用音楽データベース: 研究目的で利用可能な著作権処理済み楽曲・楽器音データベース，情報処理学会論文誌， Vol.45, No.3, pp.728–738 (2004).
3) 河原英紀，森勢将雅：歌声を見て触る: TANDEM-STRAIGHT と時変モーフィングが提供する基盤，情報処理学会研究報告. [音楽情報科学]， Vol.2010-MUS-86, No.6, pp.1–6 (2010).
4) 剣持秀紀：歌声合成技術の動向 :「初音ミク」を支える技術，日本音響学会誌，Vol.67, No.1, pp.46–50 (2011).
5) 大浦圭一郎，間瀬絢美，山田知彦，徳田 恵，後藤真孝：Sinsy:「あの人に歌ってほしい」をかなえる HMM 歌声合成システム，情報処理学会研究報告. [音楽情報科学], Vol.2010-MUS-86, No.1, pp.1–8 (2010).
6) Villavicencio, V., Röbel, A. and Rodet, X.: Applying improved spectral modeling for high quality voice conversion, *ICASSP2009*, pp.4285–4288 (2009).
7) Kawahara, H., Morise, M., Takahashi, T., Nisimura, R., Irino, T. and Banno, H.: A temporally stable power spectral representation for periodic signals and applications to interference-free spectrum, F0 and aperiodicity estimation, *ICASSP'2008*, pp.3933–3936 (2008).
8) Kawahara, H., Nisimura, R., Irino, T., Morise, M., Takahashi, T. and Banno, B.: Temporally variable multi-aspect auditory morphing enabling extrapolation without objective and perceptual breakdown, *ICASSP2009*, pp.3905–3908 (2009).
9) McDermott, J.H., Oxenham, A.J. and Simoncelli, E.P.: Sound texture synthesis via filter statistics, *Proc. IEEE WASPAA*, pp.297–300 (2009).
10) Ellis, D., Xiaohong, Z. and McDermott, J.: Classifying soundtracks with audio texture features, *ICASSP2011*, pp.5880 –5883 (online), DOI:10.1109/ICASSP.2011.5947699 (2011).
11) : Texture Synthesis Examples - Page 1 (McDermott and Simoncelli), New York University (online), available from ⟨http://www.cns.nyu.edu/~jhm/texture_examples/⟩ (accessed 2012-01-09).
12) Kawahara, H., Masuda-Katsuse, I. and de Cheveigné, A.: Restructuring speech representations using a pitch-adaptive time-frequency smoothing and an instantaneous-frequency-based F0 extraction, *Speech Communication*, Vol.27, No.3-4, pp.187–207 (1999).
13) 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, No.8, pp.442–449 (2007).
14) 森勢将雅， 高橋徹，河原英紀，入野俊夫：窓関数による分析時刻の影響を受けにくい周期信号のパワースペクトル推定法，電子情報通信学会論文誌 D， Vol.J 90-D, No.12, pp.3265–3267 (2007).
15) Unser, M.: Sampling–50 Years After Shannon, *Proceedings of the IEEE*, Vol.88, No.4, pp.569–587 (2000).
16) 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，電子情報通信学会 論文誌 A， Vol.J94-A, No.8, pp.557–567 (2011).
17) Kawahara, H. and Morise, M.: Technical foundations of TANDEM-STRAIGHT, a speech analysis, modification and synthesis framework, *SADHANA - Academy Proceedings in Engineering Sciences*, Vol.36, No.5, pp.713–722 (2011).
18) 河原英紀：聴覚における情報表現に基づく音声信号の分解 : 周期性からの逸脱をどう扱うか，電子情報通信学会技術研究報告（音声），Vol.110, No.297, pp.23–28 (2010).
19) Laroche, J., Stylianou, Y. and Moulines, E.: HNS: Speech modification based on a harmonic+noise model, *Acoustics, Speech, and Signal Processing, 1993. ICASSP-93., 1993 IEEE International Conference on*, Vol. 2, pp. 550 –553 vol.2 (online), DOI:10.1109/ICASSP.1993.319365 (1993).
20) 河原英紀，ロイ・パターソン，森勢将雅，坂野秀樹，津崎 実，高橋 徹，西村竜一，入野俊夫：成分位相の制御により声の肌触りを変える，インタラクション 2011 論文集，No. 2CR3-7， pp. CD–ROM & web（オンライン），入手先⟨http://www.interaction-ipsj.org/archives/paper2011/⟩ (2011).
21) : STRAIGHT introduction (DVD イメージのアーカイブ)，Add data for field: Organization（オンライン），入手先⟨http://www.wakayama-u.ac.jp/RAIGHT_2001_OCT_DEMO.zip⟩（参照 2012-01-09）.
22) 河原英紀，森勢将雅，高橋 徹，坂野秀樹，西村竜一，入野俊夫：尖度に基づく音響的イベントの検出と音声分析変換合成システムへの応用について，日本音響学会 2010 年度春季研究発表会，No.1-7-16, pp.315–316 (2010).
23) Titze, I.R.: *Principles of voice production*, Prentice Hall (1994). （邦訳：新美他（訳）：音声生成の科学–発声とその障害、医歯薬出版、2003）.
24) : NCVS: Giving Voice to America, NCVS (online), available from ⟨http://www.ncvs.org/index.html⟩ (accessed 2012-01-09).
25) : Ingo Titze and Pavarobotti singing Nessun Dorma, YouTube (online), available from ⟨http://www.youtube.com/watch?v=UQw03TXZsHA⟩

(accessed 2012-01-09).
26) Childers, D.G. and Ahn, C.: Modeling the glottal volume-velocity waveform for three voice types, *J. Acoust. Soc. Am.*, Vol.97, No.1, pp.505–519 (1995).
27) Kawahara, H., Atake, Y. and Zolfaghari, P.: Accurate vocal event detection method based on a fixed-point analysis of mapping from time to weighted average group delay, *Proc. ICSLP'2000*, Beijin China, pp.664–667 (2000).
28) Murty, K. and Yegnanarayana, B.: Epoch Extraction From Speech Signals, *Audio, Speech, and Language Processing, IEEE Transactions on*, Vol.16, No.8, pp.1602 –1613 (online), DOI:10.1109/TASL.2008.2004526 (2008).
29) Kawahara, H., Katayose, H., de Cheveigné, A. and Patterson, R.D.: Fixed Point Analysis of Frequency to Instantaneous Frequency Mapping for Accurate Estimation of F0 and Periodicity, *Proc. EUROSPEECH'99*, Vol.6, ESCA, pp.2781–2784 (1999).
30) Plomp, R. and Steeneken, H. J.M.: Effect of Phase on the Timbre of Complex Tones, *J. Acoust. Soc. Am.*, Vol.46, No.2B, pp.409–421 (1969).
31) Patterson, R.D.: A pulse ribbon model of monaural phase perception, *J. Acoust. Soc. Am.*, Vol.82, No.5, pp.1560–1586 (1987).
32) Schroeder, M.R.: Synthesis of low-peak-factor signals and binary sequences with low autocorrelation, *IEEE Trans. Information Theory*, Vol.16, No.1, pp.85–89 (1970).
33) Skoglund, J. and Kleijn, W.: On time-frequency masking in voiced speech, *Speech and Audio Processing, IEEE Transactions on*, Vol.8, No.4, pp.361 –369 (online), DOI:10.1109/89.848218 (2000).
34) de Chevengné, A. and Kawahara, H.: YIN, a fundamental frequency estimator for speech and music, *J. Acoust. Soc. Am.*, Vol.111, No.4, pp.1917–1930 (2002).
35) Camacho, A. and Harris, J.G.: A sawtooth waveform inspired pitch estimator for speech and music, *J. Acoust. Soc. Am.*, Vol.124, No.3, pp.1638–1652 (2008).
36) 和田芳佳，森勢将雅，河原英紀：複数の周期成分を持つ音声のための周期構造抽出法と障害音声分析への応用について，電子情報通信学会技術研究報告（応用音響），Vol.111, No.175, pp.81–86 (2011).
37) 森勢将雅，河原英紀，西浦信敬：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, No.2, pp.109–117 (2010).
38) 河原英紀，森勢将雅，西村竜一，入野俊夫：基本波の FM と AM 成分に基づく高速な基本周波数推定法について，日本音響学会聴覚研究会資料 H-2011-121， Vol.41, No.9, pp.679–684 (2011).
39) Nuttall, A.H.: Some windows with very good sidelobe behavior, *IEEE Trans. Audio Speech and Signal Processing*, Vol.29, No.1, pp.84–91 (1981).
40) Kawahara, H., Irino, T. and Morise, M.: An interference-free representation of instantaneous frequency of periodic signals and its application to F0 extraction, *ICASSP 2011*, pp.5420–5423 (2011).
41) Fujimura, O., Honda, K., Kawahara, H., Konparu, Y., Morise, M. and Williams, J.: Noh Voice Quality, *J. Logopedics Phoniatrics Vocology*, Vol.34, No.4, pp.157–170 (2009).
42) Yegnanarayana, B., d'Alessandro, C. and Darsinos, V.: An iterative algorithm for decomposition of speech signals into periodic and aperiodic components, *Speech and Audio Processing, IEEE Transactions on*, Vol.6, No.1, pp.1 –11 (online), DOI:10.1109/89.650304 (1998).
43) Hermus, K., Van hamme, H. and Irhimeh, S.: Estimation of the Voicing Cut-Off Frequency Contour Based on a Cumulative Harmonicity Score, *Signal Processing Letters, IEEE*, Vol. 14, No. 11, pp. 820 –823 (online), DOI:10.1109/LSP.2007.898854 (2007).
44) Kawahara, H., Morise, M., Takahashi, T., Banno, H., Nisimura, R. and Irino, T.: Simplification and extension of non-periodic excitation source representations for high-quality speech manipulation systems, *Proc. Interspeech2010*, ISCA, pp.38–41 (2010).
45) Pantazis, Y. and Stylianou, Y.: Improving the modeling of the noise part in the harmonic plus noise model of speech, *ICASSP'2008*, pp.4609–4612 (2008).
46) 右田尚人，森勢将雅，西浦敬信：歌唱データベースを用いたヴィブラートの個人性の制御に有効な特徴量の検討，情報処理学会論文誌， Vol.52, No.5, pp.1910–1922 (2011).