

基本周波数情報に基づく線形予測と時間軸伸縮を利用した 非周期成分の抽出について

河原 英紀[†] 森勢 将雅^{††} 高橋 徹^{†††} 坂野 秀樹^{††††} 西村 竜一[†]
入野 俊夫[†]

[†] 和歌山大学システム工学部 〒640-8510 和歌山市栄谷 930

^{††} 関西学院大学理工学部 〒669-1337 三田市学園 2-1

^{†††} 京都大学大学院情報学研究所 〒606-8501 京都市左京区吉田本町

^{††††} 名城大学理工学部 〒468-8502 名古屋市天白区塩釜口 1-501

E-mail: †{kawahara,nisimura,irino}@sys.wakayama-u.ac.jp, ††morise@kwansei.ac.jp,

†††tall@kuis.kyoto-u.ac.jp, ††††banno@ccmfs.meijo-u.ac.jp

あらまし 高品質の音声分析変換合成のための非周期成分の抽出を検討している。本報告では、これまでに STRAIGHT で用いてきたものを中心に、従来の方法とそれらの問題点を挙げ、新しい方法を提案する。提案する方法は、基本周期程度の時間間隔を隔てた部分からの前方および後方予測可能な成分を取り除いたものを非周期成分とし、時間-帯域幅積 (TB 積) を設定するための Quadrature Mirror filter による帯域分割と、基本周波数の瞬時周波数に基づく時間軸の伸縮を併用する。これらから得られる複数の手掛かりを統合することにより、従来の方法よりも効率が良くランダムな揺らぎの影響の少ない推定が可能となった。

キーワード 音声分析, 音声合成, 声質, 予測残差, 時間軸伸縮

Aperiodicity extraction based on linear prediction and temporal axis warping using fundamental frequency information

Hideki KAWAHARA[†], Masanori MORISE^{††}, Toru TAKAHASHI^{†††}, Hideki BANNO^{††††}, Ryuichi NISIMURA[†], and Toshio IRINO[†]

[†] Faculty of Systems Engineering, Wakayama University 930 Sakaedani Wakayama, 640-8510 Japan

^{††} Department of Informatics, Kwansai Gakuin University 2-1 Gakuen, Sanda, 669-1337 Japan

^{†††} Graduate School of Informatics, Kyoto University Yoshida Sakyou-ku, Kyoto, 606-8501 Japan

^{††††} Department of Information Engineering, Meijo University Tenpaku-ku, Nagoya, 468-8502 Japan

E-mail: †{kawahara,nisimura,irino}@sys.wakayama-u.ac.jp, ††morise@kwansei.ac.jp,

†††tall@kuis.kyoto-u.ac.jp, ††††banno@ccmfs.meijo-u.ac.jp

Abstract A reliable aperiodicity extractor is crucial for high-quality speech manipulation systems. This article proposes a new extractor based on a critical review on conventional methods (mainly on our previous proposals) and fundamental issues. The proposed method uses forward and backward linear predictors with lags around fundamental period and consists of an instantaneous fundamental frequency-based temporal axis warping. The extractor also consists of Quadrature Mirror Filter for frequency band division to control TB (time-bandwidth) product for reliable estimates. Combination of multiple clues extracted using the original and the manipulated time axes yields reliable and efficient estimates of aperiodicity spectrogram.

Key words speech analysis, speech synthesis, voice quality, prediction residuals, temporal axis warping

1. はじめに

有声音の音源である声帯振動の周期は、様々な要因により変動する。また、声門や声道の狭窄部位などでは、呼気流による雑音が発生する。これらの結果、有声音であっても駆動音源は完全な周期信号にはならない。観測される音声信号では、さらに、発話に伴う声道形状の時間的変化による声道のインパルス応答の変化も加わる。これらの要因による周期性からの外れは、音声の自然性や知覚される品質に大きく影響する [1]。本資料では、音声の高品質な加工を目的として、TANDEM-STRAIGHT [2] 用に開発した非周期成分の抽出および合成音声での利用法についての検討結果を報告する。

2. 背景

STRAIGHT [3], [4] は、音声中で重要な役割を担う母音等の周期性を有する信号の分析合成法として提案された。VOCODER 特有の buzzy な有声音の音色を、駆動パルスの高い周波数領域にランダムな群遅延特性を付加することで回避する [5] ことを初期の特長としていた。群遅延の操作は、オールパスフィルタであり信号のパワースペクトルには影響を与えないものの、十分に知覚できる程度に音色を変化させることができる [6]~[8]。この操作は、高い周波数領域に非周期成分を加える操作でもある。しかし、この群遅延操作の操作量を分析対象の音声から求める手段は当時提供されておらず、既定値を試行錯誤に基づいて与えていた。

周波数から瞬時周波数への写像の不動点に基づく基本周波数の抽出法の提案 [9] と同時に、音声の分析結果から非周期成分を求め、駆動信号のエネルギーを周期的パルスとスペクトル整形された広帯域雑音に配分しそれらを合成する混合駆動音源を導入した。ここでの非周期成分の抽出は、文献 [10] と同様に、調波のピークと谷の関係に基づくものであった。これにより、STRAIGHT による処理音声の品質が全般的に改善されたものの、非周期成分が主体となる部分に特有の「色づけ」が目立つこととなった。一方、群遅延の操作量の決定の手がかりとするため、パワースペクトルから求められた最小位相応答の群遅延と時間領域での時刻から平均時刻への写像の不動点に基づくイベント検出法を導出した [11]。ただし、この検出器を STRAIGHT の一部として合理的に組込むことはできなかった。非周期性の表現についてはその後も検討が続けられた [12], [13] もの、STRAIGHT に組込むには脆弱であったり、極端に計算時間がかかるなどの問題があった。

分析位置に依存しない、周期信号のパワースペクトルの新しい計算法である TANDEM [14] の発明は、STRAIGHT を根本的に再構成した [2]。^(注1)TANDEM スペクトルには、信号の周期性を反映して、周波数方向に基本周波数を周期とする余弦波状の周期的振動が重なっている。この振動は、STRAIGHT スペクトルでは、(ほぼ) 完全に取除かれている。すなわち、

(注1): 実際、TANDEM-STRAIGHT の Matlab コードには、STRAIGHT のコードは一切用いられていない。

TANDEM スペクトルを STRAIGHT スペクトルを用いて正規化することで、信号の周期性の情報だけを取り出すことができる。この性質は、基本周波数の抽出に用いることができるだけでなく、帯域毎に行うことで、非周期成分の大きさを推定することにも応用できる [15]。しかし、前述の様々な検討と同様に、TANDEM-STRAIGHT システムの一部として組込むには脆弱であった。^(注2)この脆弱性は、合成に用いるための音源の情報として用いようとする場合に必要となる、基本周波数やスペクトル包絡の時間変化による影響の補償の処理に起因する。

2.1 非周期成分抽出における問題点

非周期成分の抽出における問題点を、以下に枚挙する。次節での提案する方法の説明では、これらを必要に応じて参照する。1) 時間-帯域積 (TB 積)、2) 基本周波数の時間変化、3) 声門閉止速度等の変動、4) 声道形状の時間変化、5) 駆動点インピーダンスの周期内変動、6) 二重周期音声、などである。特に最後は、基本周波数と言う単一の量で表現することが不適切な例である。病的音声や感情音声、芸術・芸能に用いられる音声の再合成に際しては、「音源情報とは何か」という根本的な問題から問い直す必要がある。

3. 非周期成分の抽出

前節のように、非周期成分の概念自体が問題を生んでいる面があるが、ここでは音声分析合成システムへの応用を目的として、非周期成分を以下のように操作的に定義する。TANDEM-STRAIGHT における分析合成では、音声信号を駆動音源情報とスペクトル情報とに分解する。スペクトル情報からは、音源の周期性の影響が除去される。その除去のために、駆動音源情報として基本周波数が抽出される。こうして抽出されたスペクトル情報と駆動音源情報とを併せて用いても記述出来ない成分を非周期性成分とする。ただし、非周期成分は、波形情報としてではなく、簡単に計算することのできる統計量を用いて表すものとする。

3.1 線形予測残差と非線形時間軸伸縮

この方針に従い、まず、基本周波数の情報を用いることで記述出来る成分を取り除くこととする。前後の周期の信号から現在の信号を予測した場合の残差の割合を求めることとする。^(注3)

3.1.1 周期性を利用した予測

ある離散時刻 n_0 を中心とする L 個の標本からなるベクトルを $\mathbf{x}(n_0)$ と表すことにする。このベクトルを基本周期だけ離れた前後の位置付近の信号の線形結合^(注4)を用いて予測する。

$$\hat{\mathbf{x}}(n_0) = \sum \beta_k \mathbf{x}(n_0 - T_{0P} - k) + \sum \alpha_k \mathbf{x}(n_0 + T_{0S} - k) \quad (1)$$

(注2): そのため、TANDEM-STRAIGHT の試用は、著者らの周辺の狭い範囲に限らざるを得なかった。

(注3): 基本周期分だけ離れた位置からの信号を用いた予測は、Atal により、線形予測の論文 [16] に先立つ 1966 年頃の BSTJ の論文で報告されていたことを記憶している。しかし、該当する文献を確認できていないため、文献リストには挙げていない。

(注4): 基本周波数の推定値には誤差が含まれる。また、声道形状の時間変化によりインパルス応答が変化する。基本周期周辺の複数の \mathbf{x} を用いることにより、それらによる影響も含めた予測を行う。

ここで T_{0P} は、前方の基本周期、 T_{0S} は、後方の基本周期を表す。ここで $\{\dots, \beta_k, \dots\}$ と $\{\dots, \alpha_k, \dots\}$ を縦に並べて係数ベクトル \mathbf{a} を定義し、ベクトル $\{\dots, \mathbf{x}(n_0 - T_{0P} - k), \dots\}$ と $\{\dots, \mathbf{x}(n_0 + T_{0S} - k), \dots\}$ を横に並べて \mathbf{H} を定義する。この表記法を用いると、式 1 は、以下のようになる。

$$\hat{\mathbf{x}} = \mathbf{H}\mathbf{a} \quad (2)$$

これらは、注目している時刻 n_0 の関数となるが、以下では混乱のない限り、時刻の表記を省くこととする。

予測のための係数 \mathbf{a} は、評価区間の端の影響を避けるための荷重 $\mathbf{W} = \text{diag}[\dots, w_k, \dots]$ を用いた次の評価関数を最小化するように定める。

$$L_W^2 = \|\mathbf{W}\mathbf{x} - \mathbf{W}\hat{\mathbf{x}}\|^2 \quad (3)$$

この評価値を最小化する係数 $\hat{\mathbf{a}}_W$ は、次式により求められる。

$$\hat{\mathbf{a}}_W = (\mathbf{H}^T \mathbf{W}^2 \mathbf{H})^{-1} \mathbf{H}^T \mathbf{W}^2 \mathbf{x} \quad (4)$$

こうして求めた $\hat{\mathbf{a}}_W$ を用いた場合の L_W^2 から、次式により相対的残差 σ_{org}^2 を求める。

$$\sigma_{org}^2 = \frac{L_W^2}{\|\mathbf{W}\mathbf{x}\|^2} \quad (5)$$

3.1.2 時間軸の非線形伸縮

基本周期の周辺の複数の \mathbf{x} を用いることと、前方からの予測と後方からの予測に、それぞれの区間に対応する基本周期 T_{0P} と T_{0S} とを用いることで、基本周期の推定に含まれる誤差の影響と、基本周期の時間変化の影響は、軽減される。しかし、残差の評価に用いる区間長が基本周期以上である場合には、基本周期の時間変化の影響を線形予測により除くことはできない。

時間軸を、基本周波数の瞬時周波数に比例させて伸縮することにより、見かけ上の基本周波数を一定にすることができる。具体的には、新しい時間軸 $\lambda(t)$ を、以下により求める。

$$\lambda(t) = \int_0^t \frac{f_0(\tau)}{f_{ref}} d\tau \quad (6)$$

なお、目標とする見かけの基本周波数 f_{ref} は、分析対象の音声の最低の基本周波数よりも低く設定する必要がある。^(注5) このようにして新しい時間軸の上で表した信号 \mathbf{x}_{fix} を用い、前の節と同様に、基本周期の周辺の複数の \mathbf{x}_{fix} を用いた相対的残差 σ_{fix}^2 を求める。

3.2 相対的残差の統合

前の節で導入した時間軸の伸縮は、声道形状が時間的に変化せず、基本周波数が時間的に変化する場合に、見かけ上の残差の増加を招く。時間軸の伸縮により、声道伝達特性の極に対応するインパルス応答が、チャープ信号となることが、その主な原因である。しかし、声道形状が時間的に変化せず、基本周波

(注5)：この目標とする一定の基本周波数 f_{ref} を、変換対象としている音声の有声音区間の最低の基本周波数とすることで、折り返しは生ずるものの、情報の損失を避けることができる。変換に伴う折り返しについては、予め帯域化周波数のアップコンバージョンを行っておくことにより、軽減することができる。

数が時間的に変化する場合に、元の時間軸で求められた相対的残差 σ_{org}^2 が非周期成分の正しい推定値を与える。非周期成分が白色ガウス雑音であると仮定するならば、 σ_{org}^2 と σ_{fix}^2 の値の小さな方を採用することで、正しい方の推定値を選択することができる。こうして求められた推定値の平方根を、音源として用いる非周期成分の値 σ_{AP} とする。^(注6)

3.3 非周期成分の周波数特性

合成音声の駆動音源を周期的パルスと非周期成分を表す広帯域雑音の混合音源として実装する場合、非周期成分の周波数特性を推定する必要がある。前節までの残差は、この非周期成分を白色ガウス雑音とみなしたことに相当している。しかし、実際の音声信号では、低い周波数帯域では、声門の周期的開閉による呼気流の変調が音声のパワーの大部分を占めているため、周期成分が支配的となる。一方、高い周波数帯域では、声門や声道の狭窄部位で生ずる乱流等による雑音のパワーが大部分を占めるため、非周期成分が支配的となる。したがって、全周波数帯域の非周期成分を一括して白色ガウス雑音として表すことは不適切である。

ここでは、非周期成分の周波数特性は十分に滑らかであることを仮定し、併せて、聴覚の周波数分解能がほぼ周波数に比例していることを利用し、オクターブ毎に分割した帯域毎の相対的残差を、非周期成分として用いることとする。帯域分割には、QMF (Quadrature Mirror Filter) をダウンサンプリングと組み合わせて用いる。

4. 実装

ここでは、提案する方法の実装と、パラメタの決定のために行ったシミュレーションについて補足する。

4.1 時間-帯域幅積および予測次数の設定

非周期成分はランダムな成分であるため、推定値には統計的な揺らぎが含まれている。窓が矩形の場合には、推定値の分布は χ^2 分布となる。しかし、それ以外の窓を用いた場合には分布を解析的に表現することが困難になる。さらに、予測を行うことにより、自由度が減少するため、非周期成分の推定値には偏りが生ずる。これらを、シミュレーションにより数値的に確認する。統計量として、ここでは標準偏差と中央値を用いる。また、荷重として、 \mathbf{W}^2 の対角成分が以下のような hanning 窓となるように選んだ。

$$\{\mathbf{W}^2\}_{kk} = 0.5 + 0.5 \cos\left(\frac{\pi k}{N}\right), \quad (7)$$

ただし、 k は、 $-N$ から N までの整数値を動くものとした。窓長は、 $2N + 1$ である。シミュレーションに用いた信号は正規乱数である。したがって、非周期成分の中央値の正解は 1 であり、1 との差が、推定の偏りを表す。標準偏差は、推定値の揺らぎの大きさを評価するための指標である。

(注6)：なお、残差の自由度は、予測に用いる係数の次数分だけ低下する。また、重み関数を用いたことにより、実質的な自由度は更に低下する。合成のための非周期成分の値として用いる場合には、これらによる過小推定を補償する必要がある。これらを解析的に求めることは困難であるため、実装では補償量をシミュレーションにより予め計算しておく。

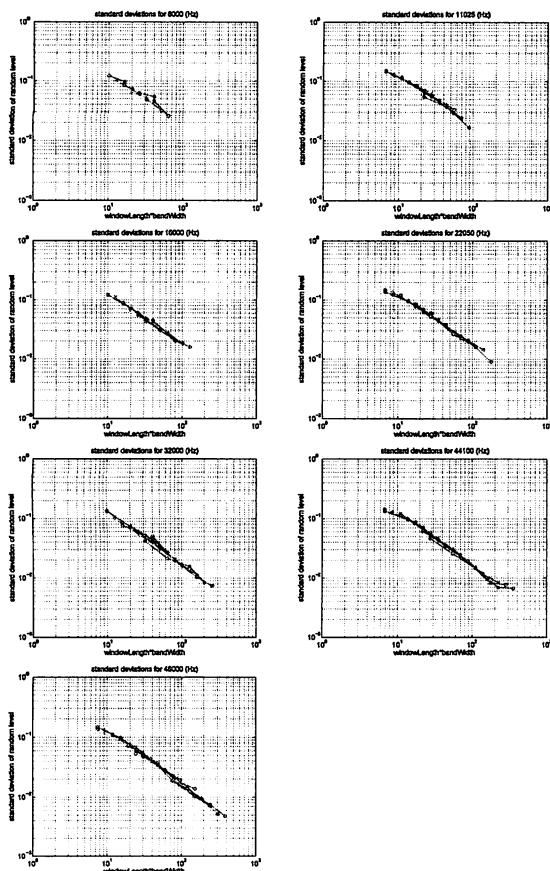


図1 窓長-帯域幅積の関数としての非周期成分推定値の標準偏差

図1と2に結果を示す。図1は、横軸を窓長-帯域幅積とし、縦軸を、推定された非周期成分の値の標準偏差とした。それぞれのプロットは、左上から右下に向けて順に、標準化周波数が8000 Hz, 11250 Hz, 16000 Hz, 22050 Hz, 32000 Hz, 44100 Hz, 48000 Hzの場合に対応する。図2は、横軸を窓長-帯域幅積とし、縦軸を、1から推定された非周期成分の値の中央値との差とした。プロットの順は、同様である。これらの図から明らかのように、推定された非周期成分の値の標準偏差および1との差は、窓長と帯域幅の積の関数となる。

予測回数も同様に、偏りに影響を与える。厳密ではないが、時間-帯域幅積 (TB積) と標準偏差の積は、窓と予測回数により決まる定数となる。また、時間-帯域幅積 (TB積) と「1と中央値との差」との積も、窓と予測回数により決まる定数となる。予測回数として10次 ($2 \times 2 + 1 + 2 \times 2 + 1$:ここでは、基本周期の前後に2標本値だけのマージンを設け、前方からの予測と後方からの予測を行った) を用いた場合のシミュレーションでは、例えば以下のような結果が得られる。

```
sideMargin =2
tbproductMedian =3.5794
tbproductSD =1.5128
itemCount =231
```

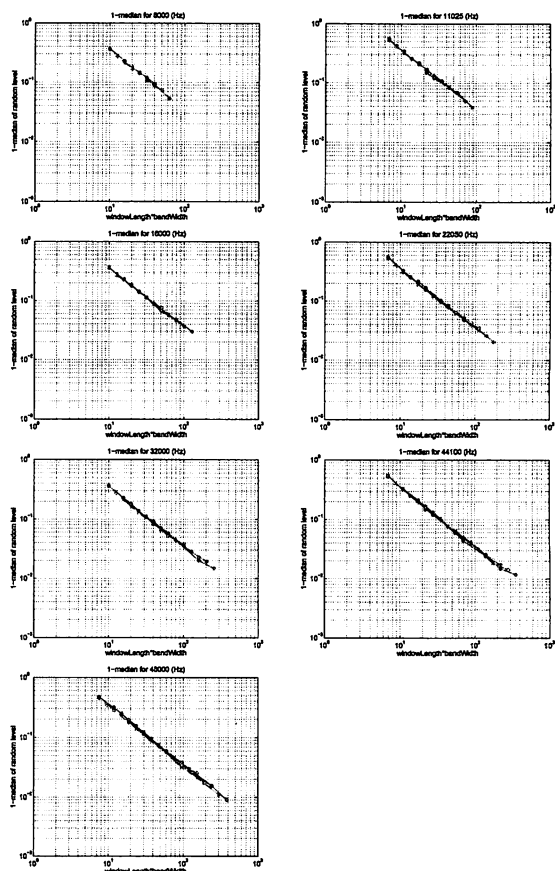


図2 窓長-帯域幅積の関数としての非周期成分推定値の中央値 (ここでは、1との差を示す)

ここで、sideMarginは、前後に設ける予測のマージン、tbproductMedianは、時間-帯域幅積 (TB積) と「1と中央値との差」との積を表し、tbproductSDは、時間-帯域幅積 (TB積) と標準偏差との積を表す。itemCountは、シミュレーションで得られた事例の数を表す。このsideMarginを2, 3, 4とした場合の積を表1に示す。窓長を T_w 、帯域幅を B_w とし、時間-帯域幅積 (TB積) と「1と中央値との差」との積を Q_{med} とした場合、「1と非周期成分の推定値中央値 (A_P) との差」とは、次の関係がある。

$$Q_{med} = T_w B_w (1 - \bar{A}_P) \quad (8)$$

従って、 \bar{A}_P が定まると、帯域幅 B_w は、次のように求められる。

$$B_w = \frac{Q_{med}}{T_w (1 - \bar{A}_P)} \quad (9)$$

例えば、1と中央値との差として0.2までを許容する場合、窓長を30 msとすると、帯域幅は、これらの値を許容値0.2で割り、さらに窓長で割ることにより、表2に示す値以上である必要があることが分かる。この帯域幅は、QMFによる分割の最後の分割の境界の周波数を与える。

表 1 時間-帯域幅積 (TB 積) と、それと中央値および標準偏差との積の平均値

	2, (10)	3, (14)	4, (18)
TB 積と「1 と中央値との差」の積	3.5794	5.1728	6.4941
TB 積と標準偏差の積	1.5128	1.7188	8.0622

表 2 窓長を 30 ms とし、中央値の偏差を 0.2 以下とする場合の境界周波数 (Hz)

	2, (10)	3, (14)	4, (18)
境界周波数	597	862	1082

表 3 窓長を基本周期の 2 倍とし、中央値の偏差を 0.2 以下とする場合の境界周波数 (Hz)

基本周波数	2, (10)	3, (14)	4, (18)
50	447	647	812
100	895	1293	1624
200	1790	2586	3247
400	3579	5173	6494

ここで、QMF と組み合わせる窓長を、基本周期の 2 倍とすることとすると、一定値に設定する基本周波数が 50, 100, 200, 400 Hz のそれぞれの場合の境界周波数は、表 3 に示す値以上である必要があることが分かる。8000 Hz が対象とする信号の標準化周波数の下限であるとする、帯域分割が有効に使えるためには、一定値に設定する基本周波数を 200 Hz 以下とすべきであることが分かる。

なお、時間軸の非線形伸縮を行った場合には、伸縮率に応じて、見かけの境界周波数が変化することになる。合成時の駆動音源に非周期性指標を用いる場合には、この伸縮率に応じた補償を行う必要がある。

4.2 安全策

周期信号であれば、基本周期の整数倍だけ時刻を移動させた信号は、元の信号と一致する。その場合、ランク落ちのために予測係数を求めるための逆行列を求めることができないという問題が生ずる。実装では、特異値分解による最大特異値と最小特異値の比として定義される行列の条件数を用いて、悪条件になる場合に別の処理を行っている。

5. 数値例

提案した非周期成分の推定法を TANDEM-STRAIGHT に組み込んで、実音声の分析合成を行った。予備的な試聴では、これまでの STRAIGHT で耳障りだった非周期成分が主となる部分での「色づき」が軽減された印象がある。幾つかの分析例を示す。

図 3 に、男性の発声した連続母音「あいうえお」の分析例を示す。上は、対応する STRAIGHT スペクトログラムである。音声試料の標準化周波数は 22050 Hz である。基本周波数は 50 Hz から 150 Hz であり、見かけ上の基本周波数を一定にする場合の目標値として、約 52 Hz が選択された。QMF による帯域分割では、元の時間軸の信号も、非線形時間軸伸縮を行った時間軸の信号も、ともに 5 つの帯域に分割された。表示され

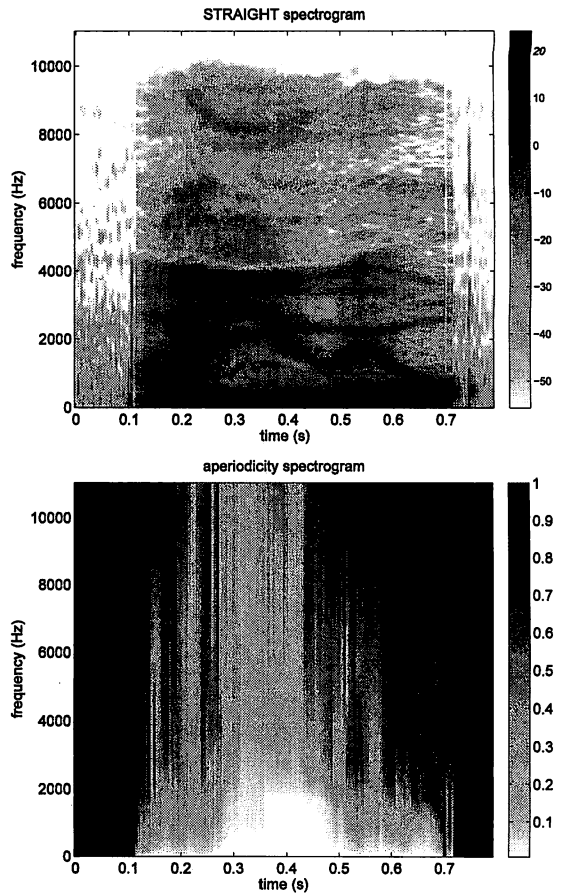


図 3 STRAIGHT spectrogram (upper plot) and corresponding aperiodicity component (lower plot). (Japanese vowel sequence /aiueo/ spoken by a male speaker)

た非周期性指標は、二つの時間軸のそれぞれから求められた残差を帯域毎に周波数方向に線形補間し、それらの結果から値が小さな方を各周波数毎に選択したものである。合成の際には、この操作が、音源パルスの発生する時刻毎に行われる。

図 4 に、女性の発声した“Open the crate but don't break the glass.”の冒頭の部分 (“Open the”) の分析例を示す。最後の部分の有声/無声判定に誤りがあるが、他の部分では母音部と鼻子音の遷移部分など、音素の切換えに影響されずに非周期成分の推定が行われていることが分かる。0.8 秒付近の高域での非周期成分の増加は、直前の有声摩擦子音の雑音成分の影響を示すものでしょう。

6. まとめと課題

基本周波数情報に基づく線形予測と時間軸伸縮を利用した非周期成分の抽出法を提案した。提案した方法は、第 2.1 節で挙げた 1) から 4) までの問題に対する一つの実践的な回答である。5) については、線形モデルでは記述できないが、実用上は音源波形の一部として扱うことで、提案した方法には

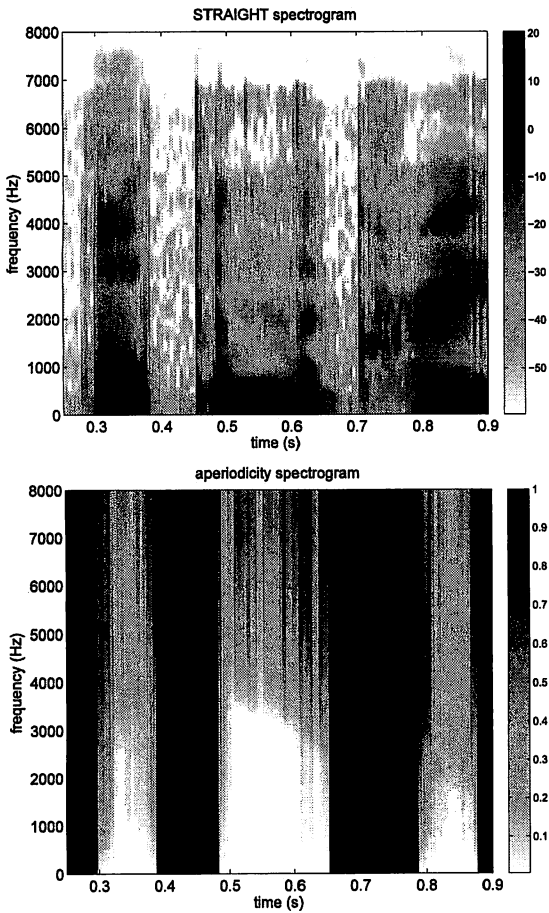


図 4 STRAIGHT spectrogram (upper plot) and corresponding aperiodicity component (lower plot). (English utterance "Open the ..." spoken by a female speaker)

影響を与えない。6)には、TANDEM-STRAIGHTに基づく基本周波数抽出法が基本周期以外の周期構造を抽出できることを利用することで、対応出来る可能性がある。これにより、TANDEM-STRAIGHTとして一貫して動作するシステムを構成することができたことになる。今後は、システムパラメタの調整による各部の校正と、主観評価実験に基づく調整が当面の課題となる。

しかし、もう一つの大きな問題が残されている。一基本周期内におけるバースト信号に対するマスキング閾値の大きな変動[4]である。この変動の大きさは、20dBを超えることもある。この効果の大きさは、非周期成分の推定精度の議論を無意味なものとする。これまで STRAIGHT に合理的に組み込むことが出来なかった音響イベント検出法[11]は、この問題に対する解を与える可能性がある。これらについては、稿を改めて報告したい。

謝 辞

本研究は、科学研究費 基盤 (A)19200017 と科学技術振興

機構による戦略的創造研究推進事業のデジタルメディア領域 CrestMuse プロジェクトのの支援を受けて行われた。非周期成分について議論して頂いている、藤村靖オハイオ州立大学名誉教授に感謝する。

文 献

- [1] O. Fujimura: "An approximation to voice aperiodicity", *IEEE Trans. Aud. Eng.*, **16**, pp. 68-72 (1968).
- [2] 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, f_0 , and aperiodicity estimation", *Proc. ICASSP 2008*, pp. 3933-3936 (2008).
- [3] H. Kawahara, I. Masuda-Katsuse and A. de Cheveigné: "Restructuring speech representations using a pitch-adaptive time-frequency smoothing and an instantaneous-frequency-based F_0 extraction", *Speech Communication*, **27**, 3-4, pp. 187-207 (1999).
- [4] 河原: "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—", *音響学会誌*, **63**, 8, pp. 442-449 (2007).
- [5] H. Kawahara: "Speech representation and transformation using adaptive interpolation of weighted spectrum: vocoder revisited", *Proc. ICASSP 1997*, Vol. 2, pp. 1303-1306 (1997).
- [6] 河原, 津崎, R. D. Patterson: "オールパスフィルタの位相操作による時間構造制御とその知覚への影響について", *聴覚研究会資料*, H-96-79, pp. 1-8 (1996).
- [7] 坂野, 陸, 中村, 鹿野, 河原: "時間領域平滑化群遅延による位相制御を用いた声質制御方式", *信学論*, **J83-DII**, 11, pp. 2276-2282 (2000).
- [8] S. Uppenkamp, S. Fobel and R. D. Patterson: "The effects of temporal asymmetry on the detection and perception of short chirps", *Hearing Research*, **158**, 1-2, pp. 71-83 (2001).
- [9] H. Kawahara, H. Katayose, A. de Cheveigné and R. D. Patterson: "Fixed point analysis of frequency to instantaneous frequency mapping for accurate estimation of F_0 and periodicity", *Proc. Eurospeech'99*, Vol. 6, pp. 2781-2784 (1999).
- [10] B. Yegnanarayana, C. d'Alessandro and V. Darsinos: "An iterative algorithm for decomposition of speech signals into periodic and aperiodic components", *IEEE Trans. on Speech and Audio Processing*, **6**, 1, pp. 1-11 (1998).
- [11] H. Kawahara, Y. Atake and P. Zolfaghari: "Accurate vocal event detection method based on a fixed-point analysis of mapping from time to weighted average group delay", *Proc. ICSLP'2000*, Beijing China, pp. 664-667 (2000).
- [12] H. Kawahara, J. Estill and O. Fujimura: "Aperiodicity extraction and control using mixed mode excitation and group delay manipulation for a high quality speech analysis, modification and synthesis system STRAIGHT", *Proc. 2nd MAVEBA*, Firenze, Italy (2001). [CD ROM].
- [13] H. Kawahara, M. Morise, T. Takahashi, T. Irino, H. Banno and O. Fujimura: "Group delay for acoustic event representation and its application for speech aperiodicity analysis", *Proc. EUSIPCO 2007*, Poznan Poland, pp. 2219-2223 (2007).
- [14] 森勢, 高橋, 河原, 入野: "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", *信学論*, **J90-D**, 12, pp. 3265-3267 (2007).
- [15] 英紀, 将雅, 徹, 竜一, 秀樹, 俊夫: "TANDEM-STRAIGHT に基づく基本周波数の抽出と評価について (音声合成・韻律, 第9回音声言語シンポジウム)", *電子情報通信学会技術研究報告*. SP, 音声, **107**, 406, pp. 259-264 (2007).
- [16] B. S. Atal and S. L. Hanauer: "Speech analysis and synthesis by linear prediction of speech wave", *J. Acoust. Soc. Am.*, **50**, 2 pt.2, pp. 637-655 (1971).