

# [招待講演] 音声・聴覚の理解と信号処理

河原 英紀<sup>1,a)</sup>

**概要:** 「音声の多くの部分を占める有声音は、なぜ、ほぼ周期的なのか? 音声の大部分が周期的であることは、聴覚にとって良いことなのか?」という素朴な疑問から始まった STRAIGHT は、音声研究のためのツールとして広く用いられるに至っている。ここでは、STRAIGHT の基礎となっている「ピッチマークへの同期を必要としないピッチ同期分析」によるパワースペクトルと瞬時周波数の表現を簡単に振り返り、最近発見された、周期性による干渉を排除した群遅延の新しい表現を紹介する。

## 1. はじめに

音声の大きな部分を占める母音などの有声音は、ほぼ周期的な信号である。人間にとって、母音は滑らかな音として知覚されるものであり、話し手の個性や感情、ニュアンスあるいは体調などの繊細な情報を、聴き手に伝える媒体でもある。しかし、機械にとって、音声の周期性は、スペクトログラム上に周期的な干渉構造を生み出し統計的なパラメタ推定に大きなバイアスをもたらすものであり、これらの繊細な情報を推定する上での、むしろ妨害要因となっていた。STRAIGHT [1] では、有声音の周期性を、音声の生成過程の背後にある滑らかな時間周波数表現を標準化するための機構であると解釈することにより、周期的な干渉の無いスペクトル包絡を求める方法を明らかにした。TANDEM-STRAIGHT [2] では、時間変動の無いパワースペクトルの表現 [3] と、consistent sampling [4] に基づく処理の導入により、同様に周期的な干渉の無いスペクトル包絡を、より効率と見通しの良い処理により求める方法を明らかにした [2], [5]。その後、周期的な零点や無限大が問題となっていた瞬時周波数にも、TANDEM の考え方を応用した時間変動の無い表現が求められた [6]。しかし、位相情報から求められるもう一つの表現である群遅延 [7], [8], [9] については、同様な TANDEM に基づく処理 [10] を加えても、時間方向の変動の無い表現を得ることはできない状況が続いていた。

しかし、この状況は今回の音学シンポジウムに申込んだ直後に急変した。零位法による感度の高い測定を可能にする、「信号が周期的であれば 0 になる」新しい群遅延の表現が見つかったのである。ここでは、当初の予定を変更し

て、この新しい群遅延の表現を中心に紹介し、今後の応用の可能性などについて議論する。音学シンポジウムでの発表は、ムービーを中心としたものとし、発表の補足となる Matlab を用いた実装や資料へのリンクを当日までに用意することとした。

以下では、まず背景としての「周期信号の静的な表現」というアイデアから出発する。次いで、パワースペクトルと瞬時周波数の静的な表現を概観した後に、新しく求められた群遅延の表現を導入し、含まれる方式パラメタを調整するとともに実装について説明する。さらに、様々な信号に対する応答と、それらの可視化について説明する。最後に、TANDEM-STRAIGHT への導入法と、様々な分野への応用の可能性について議論する。

## 2. 周期信号の静的な情報表現

「周期信号の静的な表現」というアイデアに最初に出会ったのは、その後、長期にわたって研究で協力関係を持つことになった Patterson 博士が来日した際に行われた講演の場であった。そこでは、その後の Auditory Image Model [11] につながる pulse ribbon model [12] が紹介されていた。このモデルでは、ストロボスコープに見られるような信号によるトリガを利用した方法が、信号の静的な表現を得るための手段として用いられていた。周期信号の静的な表現は、いわゆるピッチ同期分析によっても求めることができる。しかし、実際には音声の周期は常に変動しており、ピッチマークの検出自体が困難で脆弱な処理となってしまう問題があった。

1997 年の最初の STRAIGHT についての発表では、時間方向と周波数方向の F0 適応型平滑化という大量の計算を必要とする方法により [13]、その次の論文 [1] では、相補的時間窓という技巧的な方法によって、「ピッチマークへの同期を必要としないピッチ同期分析」を実現した。しか

<sup>1</sup> 和歌山大学  
Wakayama University, Wakayama, Wakayama 640-8510,  
Japan

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

し、これらの方法が2007年のTANDEMによって完全に置き換えられたため、以下では、TANDEM-STRAIGHTとその後の拡張に基づいて、周期信号の様々な静的な表現を紹介する。

### 2.1 パワースペクトルの静的な表現

短時間 Fourier 変換に用いる窓関数の周波数領域での表現の実質的な定義域の幅が分析対象とする周期信号の基本周波数  $f_0 = 1/T_0$  の2倍以下である場合、パワースペクトル  $P(\omega, t)$  の時間方向の変動は、基本周期  $T_0$  を周期とする余弦関数となる。したがって、基本周期の半分の間隔を隔てた2つの時点 ( $t - \frac{T_0}{4}$  および  $t + \frac{T_0}{4}$ ) においてパワースペクトルを計算し、それらの平均  $P_T(\omega, t)$  を求めると、余弦関数は相殺して、時間変動する項は消滅する。こうして時間変動の無い静的な表現となったパワースペクトルを、TANDEM スペクトルと呼ぶこととする。TANDEM-STRAIGHT では、さらに周波数方向での周期的変動を(調波位置での値と正值性を保存しつつ)取り除くために、cepstrum 領域での F0 適応型 liftering を行っている [5]。こうして求められたパワースペクトルの次元のスペクトル包絡を、STRAIGHT スペクトル  $P_{ST}(\omega, t)$  と呼ぶこととする。

### 2.2 瞬時周波数の静的な表現

瞬時周波数は、位相スペクトログラムの時間方向の導関数として定義されている。瞬時周波数は、複素数で表されているスペクトルの対数の時間方向の導関数の虚部として、位相の unwrap に依らずに求めることができる [14]。この Flanagan による瞬時周波数の表現を、TANDEM スペクトルの場合と同様に、基本周期の半分の間隔を隔てた2つの時点 ( $t - \frac{T_0}{4}$  および  $t + \frac{T_0}{4}$ ) において求め、それぞれの時点でのパワースペクトルを用いて重み付け、加重平均を計算することにより、瞬時周波数の静的な表現を得ることができる [6]。

## 3. 周期信号の群遅延の静的な表現

周期信号の群遅延の静的な表現は、STRAIGHT での非周期成分の抽出法として検討が進められている。(非周期成分の分析は、STRAIGHT で最も問題の残る部分である。見かけ上の基本周波数を一定とする時間軸の伸縮 [15] と併用した調波構造を有するパワースペクトルの上側包絡と下側包絡から求める方法 [16] も、同様な併用を含んだピッチ間隔での線形予測に基づく方法と sigmoid によるモデル化 [17] も、妥協とアドホックな対策を組み合わせた実装となっている。)

### 3.1 unwrap によらない群遅延の計算

群遅延  $\tau_d(\omega, t)$  は、位相スペクトログラムの周波数方向の導関数に負号をつけたものとして定義されている。ここで

も、Flanagan による瞬時周波数の計算法と同様に、unwrap を必要としない以下の計算法を出発点として用いる。

$$-\tau_d(\omega, t) = \frac{d \Im[\log(X(\omega, t))]}{d\omega} = \Im \left[ \frac{1}{X(\omega, t)} \frac{dX(\omega, t)}{d\omega} \right] \\ = \frac{\Re[X(\omega, t)] \Im \left[ \frac{dX(\omega, t)}{d\omega} \right] - \Im[X(\omega, t)] \Re \left[ \frac{dX(\omega, t)}{d\omega} \right]}{|X(\omega, t)|^2}, \quad (1)$$

この  $|X(\omega, t)|^2$  は、実際にはパワースペクトル  $P(\omega, t)$  と同じものである。ここで、以下で定義される  $X(\omega, t)$  と  $X_d(\omega, t)$  を代入する。

$$X(\omega, t) = \int_{-\infty}^{\infty} w(\tau)x(\tau-t)e^{-j\omega\tau}d\tau \quad (2)$$

$$X_d(\omega, t) = \frac{dX(\omega, t)}{d\omega} = \int_{-\infty}^{\infty} w(\tau)x(\tau-t)\frac{de^{-j\omega\tau}}{d\omega}d\tau \quad (3)$$

$$= -j \int_{-\infty}^{\infty} \tau w(\tau)x(\tau-t)e^{-j\omega\tau}d\tau, \quad (4)$$

すると、効率の良い計算法である以下が得られる。

$$-\tau_d(\omega, t) = \frac{\Re[X(\omega, t)] \Im[X_d(\omega, t)] - \Im[X(\omega, t)] \Re[X_d(\omega, t)]}{|X(\omega, t)|^2}, \quad (5)$$

### 3.2 周波数方向の変動の除去

これを用いて、まず、周波数方向での周期的変動をパワースペクトルを重みとする加重平均により取り除く。重みを分母とすることで、式(5)の分母と相殺し、分子同士の演算となることに注意する。

$$\tau_{dF}(\omega, t) = \frac{P^{(U)}\tau_d(\omega + \frac{\omega_0}{4}, t) + P^{(D)}\tau_d(\omega - \frac{\omega_0}{4}, t)}{P^{(U)} + P^{(D)}}, \quad (6)$$

ここで  $P^{(U)}$  は  $P(\omega + \frac{\omega_0}{4}, t)$  を、 $P^{(D)}$  は  $P(\omega - \frac{\omega_0}{4}, t)$  を表す。周波数軸上の繰返し周期である  $\omega_0 = 2\pi/T_0$  は、時間軸上の繰返し周期  $T_0$  により定まる。これは、TANDEM の方法を周波数軸上で行ったことに相当する。この処理によって  $\tau_{dF}(\omega, t)$  の周波数軸上での周期的変動が除去されることの導出は文献 [10] を参照されたい。

### 3.3 時間方向の変動の除去

隣接する調波間の干渉による変動は、時間軸上での周期が  $T_0$  の正弦波となる。ここでは、パワースペクトルと同様に、 $T_0/2$  だけ時間的に隔たった分析位置で求めた上記の周波数方向の変動が取り除かれた群遅延の重み付き平均を求める。ここでは、パワースペクトル包絡を重みとして用いることとする。時間方向ならびに周波数方向の変動が取り除かれた群遅延  $\tau_{dD}(\omega, t)$  を次式のように定義する。

$$\tau_{dD}(\omega, t) = \frac{P^{B+}\tau_{dF}(\omega, t + \frac{T_0}{4}) + P^{B-}\tau_{dF}(\omega, t - \frac{T_0}{4})}{P^{B+} + P^{B-}}, \quad (7)$$

ここで  $P^{B+}$  は  $P(\omega + \frac{\omega_0}{4}, t + \frac{T_0}{4}) + P(\omega - \frac{\omega_0}{4}, t + \frac{T_0}{4})$  を表し、 $P^{B-}$  は  $P(\omega + \frac{\omega_0}{4}, t - \frac{T_0}{4}) + P(\omega - \frac{\omega_0}{4}, t - \frac{T_0}{4})$  を表す。信号が周期的であれば、 $\tau_{dD}(\omega, t) = 0$  が成立する。

ただし、残念なことに、閉じた形で項が相殺する訳ではない。ここでは、様々な窓関数を用いて、この群遅延の振舞いを調べることが必要になる。目標は、「周期信号であれば0となる」性質を持つことである。したがって、まず、0からの偏差を評価する以下の評価関数  $L$  を定義すれば良い。

$$L^2 = \frac{1}{S(\Omega, T)} \int_{\Omega} \int_T |\tau_{dD}(\omega, t)|^2 dt d\omega, \quad (8)$$

ここで  $S(\Omega, T)$  は、時間方向で  $T$ 、周波数方向で  $\Omega$  で定義される領域の大きさを表す。

### 3.4 窓関数の選択とパラメタ調整

試験用の信号として、次式に示す初期位相と振幅を制御できる調波複合音  $x_p(t)$  とガウス雑音  $x_n(t)$  の混合信号  $x_m(t)$  を用いる。

$$x_p(t) = \sum_{k=0}^{\lfloor \frac{f_s}{2f_0} \rfloor} a_k \cos(2\pi k f_0 t + \varphi_k), \quad (9)$$

ここで  $f_s$  は、標本化周波数、 $f_0$  は、信号の基本周波数である。また、 $\varphi_k$  は、 $k$  番目の調波の初期位相を示し、 $a_k$  は、 $k$  番目の調波の振幅を示す。周期的なパルス列の場合は、 $a_0 = 1/2$ 、その他の  $a_k$  は 1 である。

まず、Hann 窓、Blackman 窓、Nuttall 窓（ただし、Matlab に用意されている関数ではなく、論文 [18] の表 II の 12 番目の項目の関数）を用いた場合の、持続時間と評価値  $L$  との関係を図 1 に示す。なお、横軸は同じ持続時間を持つ矩形の窓長を用いて表している。図 2 には比較のために白色雑音を入力した場合の結果を示している。なお、雑音による評価値は、窓の形状に依存せず、持続時間のみに依存している。Nuttall 窓を用いた場合には、例えば ERW=1.3 では、提案した群遅延の大きさは、雑音の場合の 1/1000 程度であることが分る。以下では、ERW=1.3 の Nuttall 窓を使うこととする。なお、この方法は、調波成分間のレベル差が大きな場合には、変動の除去の前提が崩れる。そのため、実際の音声にも利用できるように、前処理として TANDEM-STRAIGHT の処理により求めたスペクトル包絡を用いて逆フィルタを構成し、調波レベルを等化する前処理を加えた。この等化処理に用いるスペクトル包絡の計算では、予備検討の結果に基づいて、基本周期の 3.2 倍の窓長の Blackman 窓を用い、補償用のデジタルフィルタの係数として  $q_1 = -1$  を用いることとした。

### 3.5 性能評価

こうして選択した構成を中心に、性能評価を行った。まず、図 3 に入力信号の SNR と評価値との関係を示す。SNR と評価値の対応は安定しており、評価値を実際には観測できない SNR の推定手段として利用できることが示唆されている。

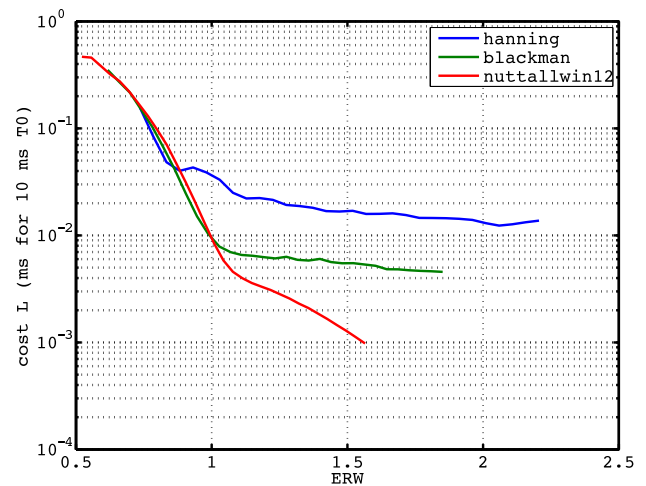


図 1 新しい関数群による窓長と評価値。入力は 100 Hz を基本周波数とする調波複合音。成分の初期位相は  $[0, 2\pi)$  に一様分布。横軸は、基本周期長の矩形窓の持続時間を単位とした持続時間。

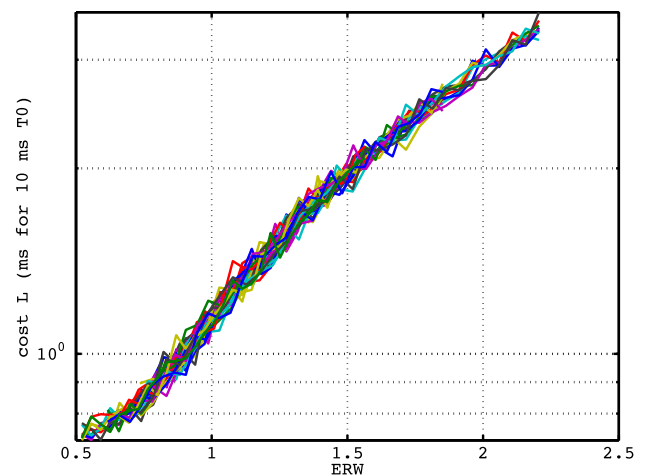


図 2 新しい関数群による窓長と評価値。入力振幅が正規分布する白色雑音。横軸は、基本周期長の矩形窓の持続時間を単位とした持続時間。

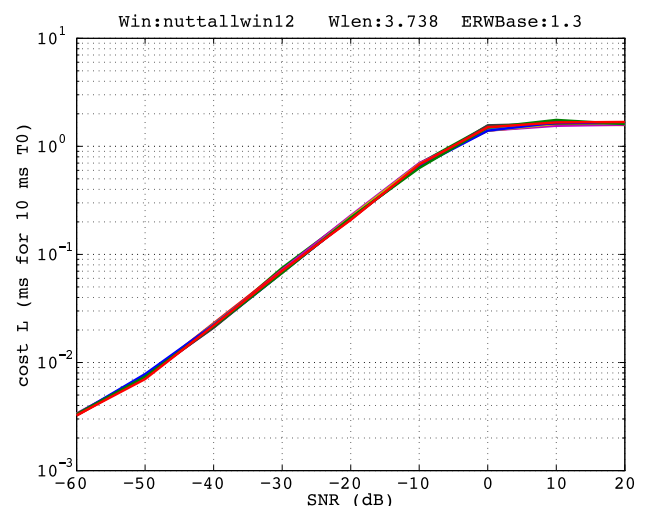


図 3 新しい関数群による SNR と評価値。入力初期位相がランダムな、100 Hz を基本周波数とする調波複合音と白色雑音との混合。

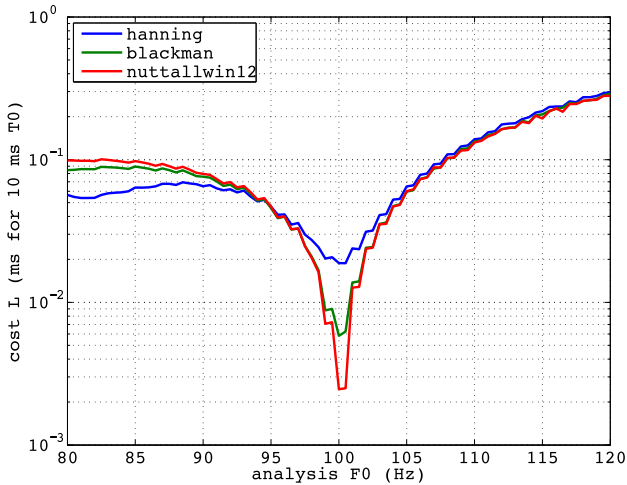


図 4 分析に用いた基本周波数が真の値と異なった場合の影響。

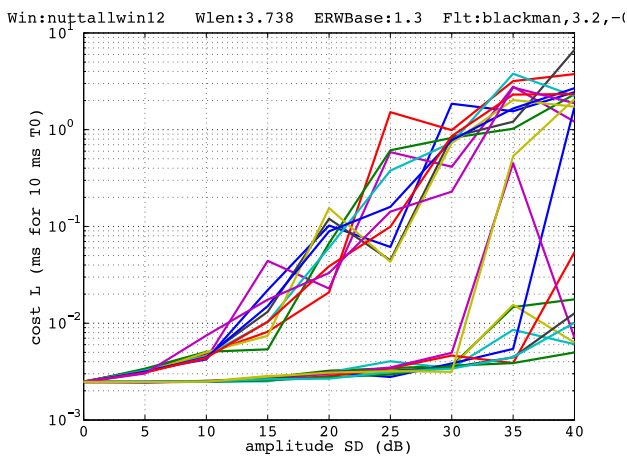


図 5 成分レベルの標準偏差と評価値。入力は初期位相がランダムで、100 Hz を基本周波数とする調波複合音の各成分のレベル (dB) が正規分布する信号を用いた。

図 4 に、分析時に想定する基本周波数の推定誤差による影響を窓毎に示す。分析時の F0 の誤差が  $\pm 5\%$  を超えた場合には、三種類の窓関数による違いは実質的には無い。また、 $\pm 2\%$  以上では、Blackman 窓関数と Nuttall 窓関数との違いは無くなる。Nuttall 窓関数を用いた場合の高い精度は、基本周波数が正確に求められている (あるいは既知) の場合に、有効になる。

図 5 に、試験信号の調波成分の振幅の対数値が正規分布する場合の分布の標準偏差と評価値の関係を示す。2 組の折れ線群のうち、上にあるものが、前処理としてのスペクトル等化処理を行っていない場合の結果、下側にあるものが、前処理を行った場合の結果である。Nuttall 窓関数の場合に、スペクトル等化を前処理として用いることで、元の信号の 30 dB までのレベルの広がりによる影響がほとんど完全に消えていることが分る。

#### 4. 静的な群遅延を用いた信号の可視化

こうして求められた群遅延を用いて、様々な信号を分析

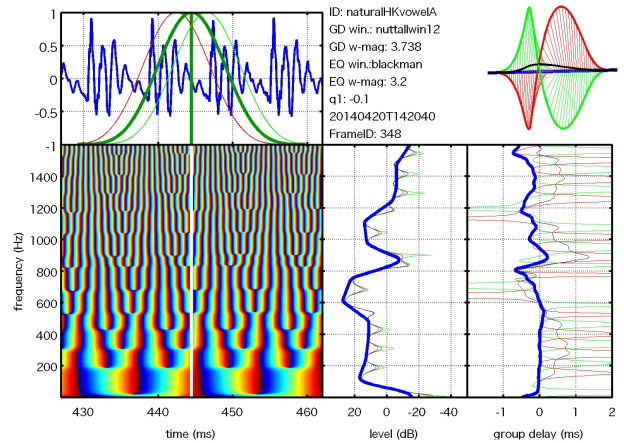


図 6 ムービーの画面レイアウト。

し、振舞いを可視化した。以下、画面に示されている要素を、左上から右上、次に下段の順に説明する。

**左上：波形と窓関数** 新しい群遅延の計算に用いた窓関数を赤および緑の線で、波形に重ねて示す。これら 2 つの窓の中央にある暗緑色の窓は、右下の位相スペクトログラム計算用に用いたものであり、実際の群遅延の計算には、赤および緑の細線で表示されているものを用いた。縦の暗緑色の線により、分析位置を表示。時間軸は左下の位相スペクトログラムと共通であるため、表示していない。

**中央上：情報表示** 実験条件など、テキスト情報を表示。  
**右上：変調窓** 周波数軸上での群遅延の移動のために、複素指数関数で変調した窓関数を表示。

**左下：位相スペクトログラム** 群遅延計算に用いたものと同じ窓関数を、実際に用いた 2 つの窓関数の中央において求めた位相スペクトログラム。縦の白線により、分析位置を表示。

**中央下：スペクトル断面** ここでは、スペクトル平坦化のために用いた TANDEM-STRAIGHT 処理 (簡易版: SADHANA の論文 [5] を参照) による様々なスペクトル断面 (パワースペクトル) を表示している。この例では、TANDEM 処理のために用意した 2 つの Blackman 窓を、赤と緑の細線で示し、それらから求められた TANDEM スペクトルを細い黒線で表示している。また、それらから補償用デジタルフィルタの係数として  $q_1 = -0.1$  を用いて求めた STRAIGHT スペクトルを青の太線で表示している。なお、この図の周波数軸は位相スペクトログラムと共通であり、目盛りの表示を共有することにより省略できる。

**右下：群遅延** ここでは、右上の変調窓により求められた周波数軸上で  $\pm \omega_0/4$  だけ平行移動した元の群遅延を赤と緑の細線で示し、それらのパワースペクトルによる重み付き平均を黒の細線で示している。この重み付き平均を、時間軸上で  $\pm T_0/4$  だけ離れた部分 (左上



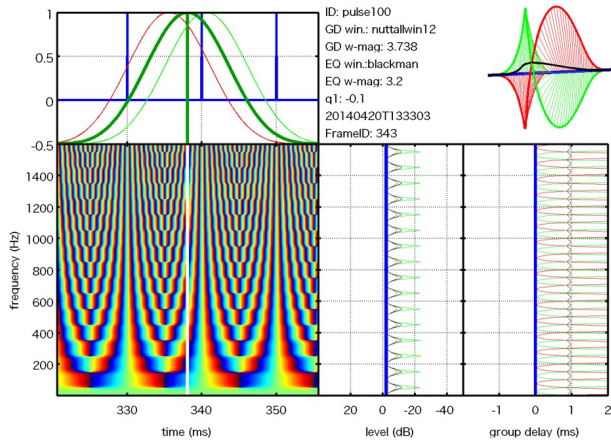


図 7 基本周波数が 100 Hz のパルス列の分析例。

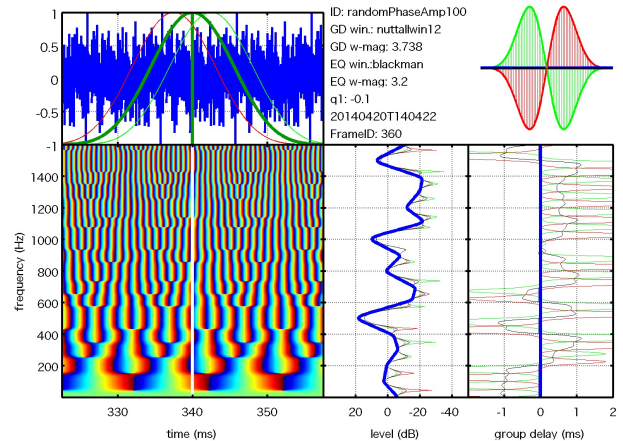


図 9 新しいスクリプトを用いた分析例。信号は、基本周波数が 100 Hz で各成分の初期位相および振幅がランダムな複合正弦波。

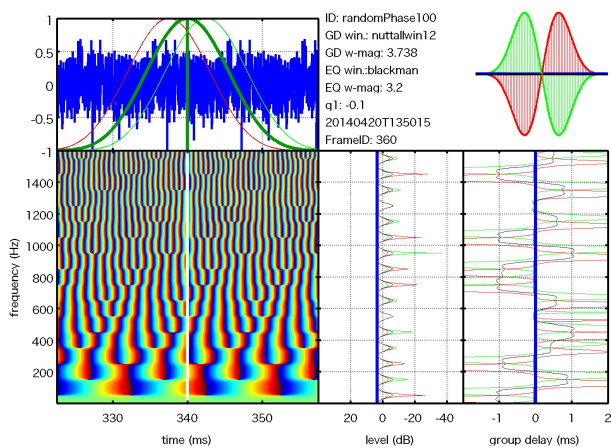


図 8 新しいスクリプトを用いた分析例。信号は、基本周波数が 100 Hz で各成分の初期位相がランダムな複合正弦波。

の図の赤線と緑線の窓)で求め、それらをパワースペクトルにより重み付けて求めた平均を青の太線で表示している。これが、提案した新しい群遅延である。なお、この図でも同様に、周波数軸は位相スペクトログラムと共通であり、目盛りの表示を共有することにより省略できる。

#### 4.1 分析例

以下、幾つか特徴的な可視化の例を示す。これらは、会場でデモするとともに、ダウンロードできるようにしておく。

##### 4.1.1 基本周波数 100 Hz のパルス列の分析例

まず、基本周波数が 100 Hz の周期的パルス列を分析し、ムービーを作成した。図 7 にパルスが分析位置と一致しない場合の例を示す。

##### 4.1.2 基本周波数が 100 Hz のランダム位相複合正弦波の分析例

図 8 に、調波成分の振幅が同一で初期位相が  $[0, 2\pi)$  の一様分布により与えられる複合正弦波の分析例を示す。こ

こでは、APSIPA2012 の方法で求められた群遅延 (黒い細線) は、周波数毎に大きく異なる。しかし、提案した表現では、0 のまま変動しない。

##### 4.1.3 基本周波数が 100 Hz のランダム位相ならびにランダム振幅の複合正弦波の分析例

図 9 に、調波成分の振幅が標準偏差 10 dB の正規分布に従い、初期位相が  $[0, 2\pi)$  の一様分布に従う複合正弦波の分析例を示す。この複合正弦波は、中央に示した STRAIGHT スペクトルを用いて作成した逆フィルタにより等化された後、群遅延が求められた。ここでも、APSIPA2012 の方法で求められた群遅延 (黒い細線) は、周波数毎に大きく異なる。しかし、提案した表現では、0 のまま変動しない。これは、隣接する 500 Hz の成分と 600 Hz の成分のように 40 dB 程度の大きなレベル差がある部分においても同様である。

##### 4.1.4 実音声の分析例

図 10 に、男性の発声した日本語母音/a/の分析例を示す。ここでは、隣接する周期の声門閉止時刻に分析位置を設定した。試験用の周期信号の分析例とは大きく異なり、求められた群遅延は 0 から逸脱している。これらは、実際の声帯振動における周期毎の閉止状況の変動を反映したものと解釈することができる。(この例では、基本波での群遅延が一個置きに前後しており、サブハーモニックを表している可能性がある。この変動は、1000 Hz 付近での群遅延の変動とも同期しているようであり、この推測を補強する。) 今後、他の観測手段と組み合わせることにより、非侵襲的な声の診断手段としての有用性を検討する価値があると思われる。また、TANDEM-STRAIGHT における非周期成分の推定にも、この新しい群遅延の表現は有用であると考えられる。

なお、この図 10 では、TANDEM-STRAIGHT におけるスペクトル包絡推定の問題が、典型的に示されている。

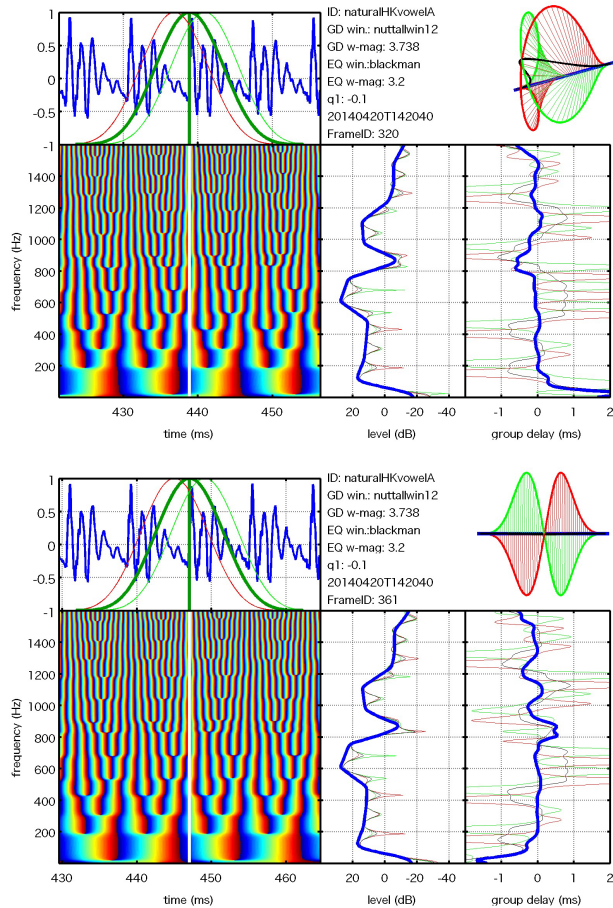


図 10 新しいスクリプトを用いた分析例。男性の発声した日本語母音/a/。ここでは、隣接する周期での声門閉止時刻における分析例を示す。

## 5. まとめ

音声の大きな部分を占める周期的な信号の群遅延から、周期性の影響を取り除き静的な表現を得ることに成功した。これにより、TANDEM-STRAIGHTを構成する処理全てを、静的な表現により再構成することが可能となった。より広い枠組みで考えると、音声を表現するパラメタ推定におけるバイアスの主要な要因であった周期性の影響が、取り除かれたことになる。バイアスからの自由を得たこれらの表現と、急速に進化している機械学習との出会が何を生み出すかが楽しみである。

**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。本稿をまとめる上で、TANDEMの発明者である森勢博士との議論が非常に有益であった。ここに感謝する。

## 参考文献

[1] 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).

[2] 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, *Proc. ICASSP 2008*, pp. 3933–3936 (2008).

[3] 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. J90-D, No. 12, pp. 3265–3267 (2007).

[4] Unser, M.: Sampling—50 Years After Shannon, *Proceedings of the IEEE*, Vol. 88, No. 4, pp. 569–587 (2000).

[5] Kawahara, H. and Morise, M.: Technical foundations of TANDEM-STRAIGHT, a speech analysis, modification and synthesis framework, *Sadhana*, Vol. 36, No. 5, pp. 713–727 (2011).

[6] Kawahara, H., Irino, T. and Morise, M.: An interference-free representation of instantaneous frequency of periodic signals and its application to F0 extraction, *Proc. ICASSP 2011*, pp. 5420–5423 (online), DOI: 10.1109/ICASSP.2011.5947584 (2011).

[7] Boashash, B.: Estimating and interpreting the instantaneous frequency of a signal. I. Fundamentals, *Proceedings of the IEEE*, Vol. 80, No. 4, pp. 520–538 (1992).

[8] Boashash, B.: Estimating and interpreting the instantaneous frequency of a signal. II. Algorithms and applications, *Proceedings of the IEEE*, Vol. 80, No. 4, pp. 540–568 (1992).

[9] Cohen, L.: *Time-frequency analysis*, Prentice Hall, Englewood Cliffs, NJ (1995).

[10] Kawahara, H., Morise, M., Nisimura, R. and Irino, T.: An interference-free representation of group delay for periodic signals, *Proc. APSIPA ASC 2012*, pp. 1–4 (2012).

[11] Bleack, S., Ives, T. and Patterson, R. D.: Aim-mat: the auditory image model in MATLAB, *Acta Acustica United with Acustica*, Vol. 90, No. 4, pp. 781–787 (2004).

[12] Patterson, R. D.: A pulse ribbon model of monaural phase perception, *J. Acoust. Soc. Am.*, Vol. 82, No. 5, pp. 1560–1586 (1987).

[13] Kawahara, H.: Speech representation and transformation using adaptive interpolation of weighted spectrum: vocoder revisited, *ICASSP-97*, Vol. 2, IEEE, pp. 1303–1306 (1997).

[14] Flanagan, J. L. and Golden, R. M.: Phase Vocoder, *Bell System Technical Journal*, pp. 1493–1509 (1966).

[15] Abe, T., Kobayashi, T. and Imai, S.: The IF Spectrogram: A New Spectral Representation, *Proc. ASVA-97*, Tokyo, pp. 423–430 (1997).

[16] Kawahara, H., Estill, J. and Fujimura, O.: Aperiodicity extraction and control using mixed mode excitation and group delay manipulation for a high quality speech analysis, modification and synthesis system STRAIGHT, *Proc. MAVEBA*, pp. 13–15 (2001).

[17] 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. Interspeech 2010*, pp. 38–41 (2010).

[18] 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).