

瞬時周波数および群遅延に基づく非周期成分推定法再考

河原 英紀^{1,a)} 榎原 健一^{2,b)} 森勢 将雅^{3,c)} 坂野 秀樹^{4,d)}

概要：音声信号の駆動情報分析の枠組みとして提案された YANGsaf [1] は、(1) 基本波の分布確率マップの作成、(2) 基本波成分の追跡と初期推定、(3) 非周期成分の推定と基本周波数推定値の更新処理から構成されており、様々な処理を組み込むことができる柔軟な枠組みである。ここでは YANGsaf の第三段階の処理に、著者の一人が以前に提案した瞬時周波数に基づく方法 [2,3] を再構築し応用する。まず、FFT の各 bin の周波数からその周波数における瞬時周波数への写像から、周波数に関する導関数および、周波数と時間に関する導関数を求めて用いることで、STRAIGHT や TANDEM のような基本周波数に強く依存した処理に基づかずに、非周期成分のレベルをそれら以上に正確に求める方法を提案する。

Revisiting aperiodicity estimation based on instantaneous frequency and group delay

KAWAHARA HIDEKI^{1,a)} SAKAKIBARA KEN-ICHI^{2,b)} MORISE MASANORI^{3,c)} BANNO HIDEKI^{4,d)}

1. はじめに

声門の周期的開閉により生成される有声音には、周期的成分に加えて気流雑音などに起因するランダムに変動する成分が含まれている。また、有声音の開始や終了部分、囁れ声では、声帯の振動も非周期的になる [4,5]。分析合成音声の自然性の知覚には、これらの非周期成分が大きく寄与するため、様々な検討が進められている [2,3,6–10]。しかし、有声音では周期成分が支配的であるため、相対的にレベルの低い非周期成分を正確に推定することは簡単ではない。ここでは、支配的な周期成分の瞬時周波数の変動に注目することにより、この困難を回避することのできる方法を提案する。

2. 背景

著者の一人は、これまで音声知覚研究のための基盤と

することを狙い、分析合成型の音声分析・変換・合成系 STRAIGHT (legacy-STRAIGHT [11] および TANDEM-STRAIGHT [12]) と、その応用である音声モーフィング [13] の検討を進めてきた。legacy-STRAIGHT と TANDEM-STRAIGHT には、それぞれ全く異なった方式による非周期成分の推定法が実装されている [3,14]*¹。しかし、いずれの非周期成分推定法も、発見的な工夫の組み合わせとして実装されており、確実な数理的基盤に基づくものではなかった。

音声合成のためのコーパスに信頼できる音源情報を付与することを狙って開発した YANGsaf では、基本周波数の推定問題の根本を見直すことにより、見通しの良い枠組みが提案されている [1]*²。YANGsaf の最初の段階では、対数周波数軸上に等間隔に配置された複素数のインパルス応答を有する帯域通過フィルタを利用することにより、基本波成分の正弦波からの逸脱（基本周波数近傍の非周期成分）を、基本周波数に関する事前知識無しに、基本周波数と同時に推定する。第二段階では、これらの情報を統合して基本周波数の所在についての確率マップを求め、基本周波数の初期推定値を求める。最終段階では、FFT を用いた同様の機構によって、各調波成分近傍の非周期成分を求めるとともに、こうして求められた基本周波数の初期推定値を改

¹ 和歌山大学
Wakayama University, Wakayama, Wakayama 640–8510, Japan

² 北海道医療大学
Health Science University of Hokkaido

³ 山梨大学
University of Yamanashi

⁴ 名城大学
Meijo University

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

b) kis@hoku-iryu-u.ac.jp

c) mmorise@yamanashi.ac.jp

d) banno@meijo-u.ac.jp

*¹ TANDEM-STRAIGHT は、legacy-STRAIGHT の特許と抵触しないように、かなり無理をした方式を用いて実装されている。この制約は近日中に解消される。

*² YANGsaf を分析系として、簡単なスペクトル推定と合成系を加えたものがオープンソースの YANGvocoder として公開されている [15]。

良する。

このYANGsafの第三段階では、基本周波数の初期推定値を利用できるため、文献で提案されたものよりも簡単な方法を用いることができる。ここでは、著者の一人が以前に提案した瞬時周波数に基づく方法 [2] を見直すことにより、数理的に確実な基盤に基づく方法を実現できることを示す。

3. 瞬時周波数と群遅延

瞬時周波数と群遅延は、それぞれ位相の時間微分と周波数微分として定義された、物理的に解釈し易い量である。位相には $2\pi n$, ($n \in \mathbb{Z}$) (\mathbb{Z} は整数の集合) の多義性があるため、モーフィングなどに必要な演算が困難であるという問題がある。瞬時周波数と群遅延は、それぞれ周波数方向と時間方向の重心と解釈することができ、位相におけるような演算上の困難はない [16–18]。

3.1 特異値を回避した瞬時周波数

しかし、周期信号の瞬時周波数には、特異値が周期的に発生するという問題がある。これは、付録に示す Flanagan [19] による式 A.1 の分母にあるパワースペクトルの零点による。TANDEM-STRAIGHT のスペクトル推定 [12] の場合と同様に、基本周期の $1/2$ の間隔を隔てた二つの時刻における瞬時周波数のパワースペクトルを重みとする加重平均により、これらの特異値を解消しかつ時間変動の無い瞬時周波数を求めることができる [20]。

特異値を解消するだけであれば、付録 A.2 に示す、パワースペクトル重みを利用した平滑化 (パワースペクトル加重平滑化瞬時周波数) で十分であり、TANDEM-STRAIGHT のように基本周波数に正確に適応した処理は不要である。このように特異値を解消した瞬時周波数を用いることで、以前に提案した瞬時周波数に基づく方法 [2] を、拡張することができる。

3.2 非周期成分の推定

周期信号の各調波成分によるメインローブが重複しない時間方向に長い時間窓を用いてパワースペクトルを求め、その調波間に存在する成分から非周期成分のレベルを求めることができる [3,6]。しかし、調波間には非周期成分だけでなく、窓関数の周波数表現のサイドローブを介した他調波成分からの影響などが含まれている。また、非周期成分を観測することのできるスペクトルの谷の部分の帯域幅を広げるためには、さらに大きな時間長の窓関数を用いなければならない、時間分解能が劣化するという問題がある。帯域ごとの基本周期間隔の線形予測と F_0 に比例した時間伸縮を組み合わせた方法 [14] は、時間分解能の劣化と、 F_0 の誤差に対する脆弱性の問題がある [10]。周期信号の群遅延の性質を利用した方法 [10] でも、時間分解能の劣化の問題が残る。

3.3 瞬時周波数を利用した信号対雑音比の推定

著者らが以前に提案した方法 [2] を拡張することで、これらの問題を避けることができる。ここでは、まずアイデアの概要を説明する。帯域通過フィルタの実質的な通過域

に一個の正弦波成分 (実際には片方の複素指数関数) のみが含まれる場合には、帯域フィルタ出力の瞬時周波数は時間に依存しない定数となる。しかし、他の成分が含まれる場合には、瞬時周波数は時間的に変動する。この変動量は帯域通過フィルタの中心周波数に依存して変化する。したがって、中心周波数に関する瞬時周波数の微分 (周波数方向の傾斜) の値の大きさは、主要な成分以外の成分の相対的レベルに依存した量となる。これが基本となるアイデアである。

実用的な指標とするために、この周波数方向の微分だけではなく、それをさらに時間方向に微分したものを用意する。すると、それぞれの値の自乗の最大値と最小値は相殺する。したがって、それぞれの値の分布の中央値に基づいて校正したものの自乗平均値を指標とする (式 A.9) と、その変動はそれぞれの構成要素の変動よりも小さなものとなる。この指標 $g_A(\omega, t)$ を、以下では傾斜指標と呼ぶことにする。傾斜指標を求める出発点に、前述の特異値を回避した瞬時周波数を用いることにより、以前に提案した方法 [2] を拡張することができる。

3.3.1 群遅延に基づく解釈

群遅延を出発点として、この傾斜指標に別の解釈を与えることができる。瞬時周波数の周波数微分は、微分の順序を入れ替えると、群遅延の時間微分となる。単一の周波数成分 (複素指数関数) だけを含む信号を偶関数の窓関数で取り出したときの群遅延は、信号の包絡の重心の位置である窓の中心時刻に一致する。この主要な成分以外の成分が存在すると、唸りにより信号の包絡が時間的に変動し、重心の位置も前後に変動する。群遅延の時間変動の大きさ (群遅延の時間微分 $\Delta\tau_g$ の大きさ) は、この主要な成分の大きさに比例する。群遅延の二階時間微分 $\Delta^2\tau_g$ と、 $\Delta\tau_g$ の自乗のそれぞれの最大値と最小値は相殺する。 $\Delta^2\tau_g$ と、 $\Delta\tau_g$ の分布に基づいて、適切に校正したそれぞれの自乗平均値を指標とする (式 A.10) と、その変動はそれぞれの構成要素の変動よりも小さなものとなる。こうして、信号対雑音比の推定に用いることのできる指標を作ることができる。前の節で導入した傾斜指標は、このようにも解釈できる。

3.3.2 実装

傾斜指標の出発点となる瞬時周波数には、前の節で説明した、特異値を回避した二種類の方法を用いることができる。具体的には、時間窓としては、 \cos 級数の窓の中から、サイドローブのレベルが低くかつサイドローブの減衰速度が大きいもの (文献 [21] の表 II の 11 番目の項目) を用いることとし、TANDEM による方法と、パワースペクトル加重平滑化瞬時周波数による方法の二種類により、特異値を解消した。詳細は付録 A.2.1 に譲る。

3.4 傾斜指標と SNR : 数値例

こうして作成した傾斜指標と信号対雑音比 (SNR) の関係を、既知の SNR の信号を用いて調べる。以下では、40 Hz の周期的パルス列に白色ガウス雑音を加えて作成した試験信号を用いた。

図 1 に、SNR を 0 dB から 80 dB まで 10 dB 毎に変化させて求めたパワースペクトル加重平滑化瞬時周波数に基づく傾斜指標の値の累積分布を太線を用いて示す。図には

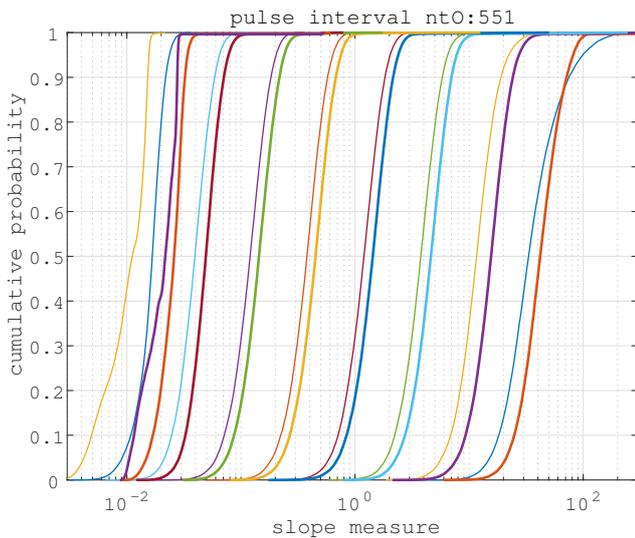


図 1 SNR 毎の傾斜指標の値の累積分布。横軸が傾斜指標の値、縦軸が累積分布の確率を表す。細線は TANDEM 型の瞬時周波数計算法によるもの、太線はパワースペクトル加重平滑化瞬時周波数によるものを示す。

Fig. 1 Cumulative distribution of the gradient measure for different SNR levels. Thin lines show results using TANDEM-based instantaneous frequency. Thick lines show results using power spectrum weighted smoothed instantaneous frequency

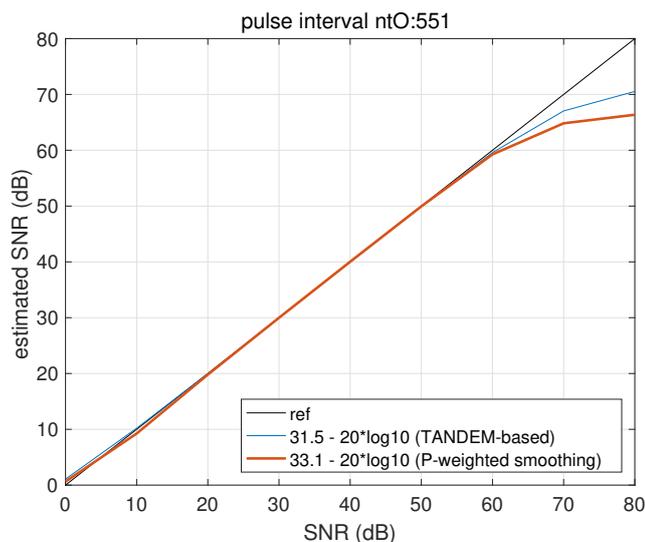


図 2 試験信号の SNR と校正した傾斜指標の中央値の関係。校正用の式を凡例に示す。

Fig. 2 Calibrated median of gradient measure and input SNR. Legend shows equations for calibration.

比較のために、TANDEM を利用して求めた瞬時周波数に基づく傾斜指標の値の累積分布を細線を用いて示す。右端の曲線が 0 dB の SNR の結果を示し、左端の曲線が 80 dB の SNR の結果を示す。パワースペクトル加重平滑化瞬時周波数による分布は、0 dB から 60 dB の範囲でほぼ同一であり、各段階間の間隔もほぼ同一である。

図 2 に、試験信号の SNR と校正した傾斜指標の中央値の関係を示す。校正のための係数は、図 1 の分布から求めた。この結果は、前の段落での観察と整合しており、校正した傾斜指標が、信号の SNR の推定値になり得ること

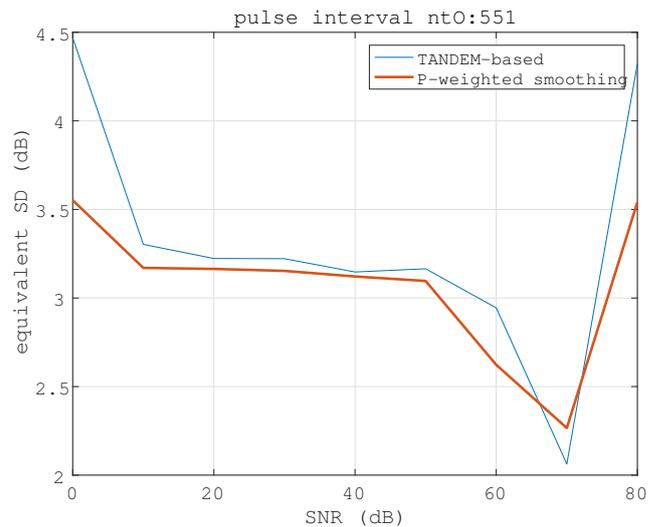


図 3 校正された傾斜指標の標準偏差。外れ値を避けて dB 軸上の正規分布を用いて近似。

Fig. 3 Standard deviation of calibrated gradient measure in dB. Central part of the distribution was approximated using Gaussian distribution for discarding outliers.

を示している。なお、SNR が 60 dB 以上の範囲では、パワースペクトル加重平滑化瞬時周波数による傾斜指標の中央値も、TANDEM を利用して求めた瞬時周波数による傾斜指標の中央値も、飽和する傾向を示している。これらは、帯域フィルタのサイドロープによるものであり、実用上は無視できる。窓関数の影響については、付録 A.2.1 に例示する。

図 3 に、それぞれの SNR の試験信号について、傾斜指標の累積分布の中央部分を近似する正規分布の標準偏差を求めた結果を示す。試験信号の SNR が 0 dB から 50 dB の範囲では、パワースペクトル加重平滑化瞬時周波数による傾斜指標の分布の広がり、ほぼ一定であり、TANDEM を利用して求めた瞬時周波数による傾斜指標よりも一貫して狭い。これらの結果は、傾斜指標を SNR の推定に用いる際には、パワースペクトル加重平滑化瞬時周波数を出発点とした方が、F0 に強く依存する TANDEM によるよりも優れていることを意味する。以下では、パワースペクトル加重平滑化瞬時周波数を出発点とした傾斜指標を SNR の推定に用いる。

3.5 非周期成分の割合

観測された各周波数におけるパワーに占める非周期成分の割合を、非周期性指標 AP とする。非周期性指標 $AP(\omega, t)$ は、 $SNR(\omega, t)$ から、次式で求められる。

$$AP(\omega, t) = \frac{1}{1 + SNR(\omega, t)} \quad (1)$$

なお、 AP と SNR の次元は、パワーの比である。

4. 実音声の分析

ここまでは、基本周波数が一定であることを前提としていた。実際の音声では、基本周波数が時間の関数となり変化するため、以下で説明するような前処理と後処理が必要となる。

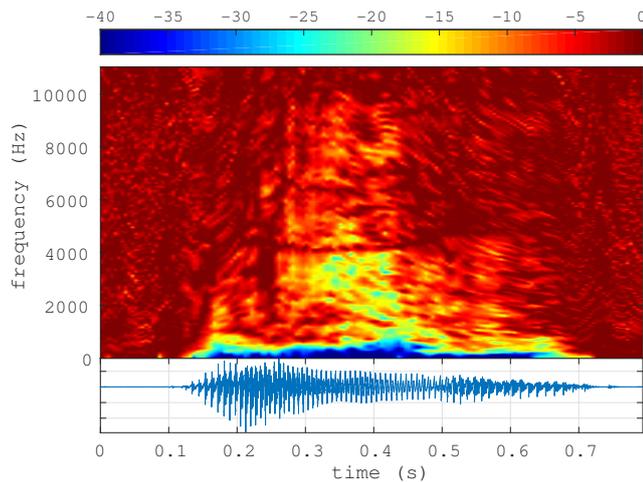


図 4 男性話者による日本語連続母音/aiueo/の非周期性指標の分析例。擬似カラーは $10 \log_{10}(AP)$ を表す。下段に音声波形を示す。

Fig. 4 Aperiodicity map of a Japanese vowel sequence /aiueo/, spoken by a male speaker. Pseudo color represents $10 \log_{10}(AP)$. The bottom plot shows sound waveform.

(1) 前処理では、基本周波数に比例する比率で時間軸を伸縮させ、見かけの基本周波数が一定となるように変換する。併せて、本来の時間軸上で設定した分析フレームの時刻を、変換された時間軸上での時刻に変換する。(2) この変換した信号を入力とし、変換されたフレーム時刻毎に、前節で導入した方法を用いて非周期成分を求める。(3) 後処理では、変換された時間軸に対応する周波数軸上で表現されている非周期成分を、元の基本周波数の情報を用いて、本来の周波数軸上の表現に変換する。

4.1 分析例

図 4 に、分析例を示す。用いた音声は、男性の発声した日本語の母音連鎖/aiueo/である*3。標準化周波数は 22050 Hz、16 bit で量子化されている。前処理に必要な基本周波数には、YANGsaf により求めたものを用いた。なお、基本周波数の推定値には誤差が含まれているため、基本周波数の整数倍の位置で非周期性指標の値を読み取ることはできない。ここでは、パワースペクトルとスペクトル包絡との比が極大値となる位置で非周期性指標の値を読み取っている。

図 4 の上段に、非周期性指標の時間周波数表現を示す。この図で用いた、dB で表した非周期性指標の値と擬似カラーとの対応関係を、上のカラーバーに示す。最下段には、比較のために音声波形を示す。

5. 課題

図 4 を参考に、提案した方法に残された課題について説明する。非周期性指標の表示には、0.2 秒から 0.5 秒にかけて、4500 Hz 付近に非周期性が高い部分（暗く見える比較的細いほぼ水平の軌跡）が認められる。これは、梨状窩に起因する声道伝達特性の深い谷に対応する部分である。この部分では、信号のレベルが低く、背景雑音が支配的とな

*3 1996 年に筆頭著者が発声/録音し STRAIGHT と一緒に配布している音声ファイル (vaiueo2d.wav) を用いた。

ることで高い非周期性指標の値が求められる。非周期成分を比率としてではなく、信号に含まれる実際のレベルとして求めることで、既知の背景雑音による影響を補償することができる。

また、非周期性指標の表示には、調波成分に平行する構造が（例えば 0.5 秒から 0.7 秒の 2000 Hz 以上の領域）認められる。この微細な構造を残したまま、音声合成の際の駆動信号の周期成分と非周期成分にパワーを配分すると、合成された音声の品質が劣化する。音声合成の際の駆動信号への応用では、聴覚の周波数分解能を考慮した（周期成分のパワーも考慮した）平滑化などの検討が必要となる。

基本周波数の時間変化（方向・速度）と、フォルマント周波数の変化（方向・速度）とは、一般に一致しない。この食い違いにより、見かけの非周期成分は増加する。この影響を軽減するためには、前処理において声道伝達特性の極による応答を補償する必要がある。

なお、ここで提案したパワースペクトル加重平滑化瞬時周波数と同様な手法は、双対な概念である群遅延に応用することができる。著者の一人が以前に提案した駆動時刻推定法 [22] を見直し、有声音の開始/終了時やグロウル系の歌唱 [23] などに認められる非周期的な駆動 [24] の分析法を再構築することも、重要な課題である。

6. おわりに

周期信号の瞬時周波数と群遅延の性質を見直すことにより、以前に提案した非周期成分の推定法を、音声信号の客観的な計測手段として用い得るように再構築した。提案した方法は、音声コーパスの付加情報や機械学習等のための入力情報の作成に、見通しの良い信頼できる手段を提供する。しかし、合成音声の駆動情報として、この分析結果をそのまま用いることはできない。前処理では音声生成過程の影響、後処理では聴覚特性の影響などを検討する必要がある。

謝辞 本研究の一部は、科学研究費補助金基盤研究 (B) 15H03207、15H02726 と挑戦的萌芽研究 16K12464 により支援された。

参考文献

- [1] Kawahara, H., Agiomyrgiannakis, Y. and Zen, H.: Using instantaneous frequency and aperiodicity detection to estimate F0 for high-quality speech synthesis, *arXiv preprint arXiv:1605.07809*, (online), available from <http://arxiv.org/abs/1605.07809> (2016).
- [2] 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*, Budapest, Hungary, pp. 2781–2784 (online), available from <http://www.isca-speech.org/archive> (1999).
- [3] 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, *Proceedings of MAVEBA*, Firenze Italy, pp. 59–64 (2001).
- [4] Titze, I. R.: *Principles of voice production*, National Center for Voice and Speech (2000).

- [5] Ishi, C. T., Sakakibara, K. I., Ishiguro, H. and Hagita, N.: A Method for Automatic Detection of Vocal Fry, *IEEE Transactions on Audio, Speech, and Language Processing*, Vol. 16, No. 1, pp. 47–56 (online), DOI: 10.1109/TASL.2007.910791 (2008).
- [6] Yegnanarayana, B., D’Alessandro, C. and Darsinos, V.: An iterative algorithm for decomposition of speech signals into periodic and aperiodic components, *IEEE Transactions on Speech and Audio Processing*, Vol. 6, No. 1, pp. 1–11 (online), DOI: 10.1109/89.650304 (1998).
- [7] D’Alessandro, C., Darsinos, V. and Yegnanarayana, B.: Effectiveness of a periodic and aperiodic decomposition method for analysis of voice sources, *IEEE Transactions on Speech and Audio Processing*, Vol. 6, No. 1, pp. 12–23 (online), DOI: 10.1109/89.650305 (1998).
- [8] Deshmukh, O., Espy-Wilson, C., Salomon, A. and Singh, J.: Use of temporal information: detection of periodicity, aperiodicity, and pitch in speech, *IEEE Transactions on Speech and Audio Processing*, Vol. 13, No. 5, pp. 776–786 (online), DOI: 10.1109/TSA.2005.851910 (2005).
- [9] Drugman, T. and Dutoit, T.: The Deterministic plus Stochastic model of the residual signal and its applications, *IEEE Trans. Audio, Speech and Language Processing*, Vol. 20, No. 3, pp. 968–981 (online), DOI: 10.1109/TASL.2011.2169787 (2012).
- [10] Morise, M.: D4C, a band-aperiodicity estimator for high-quality speech synthesis, *Speech Communication*, Vol. 84, pp. 57–65 (online), DOI: 10.1016/j.specom.2016.09.001 (2016).
- [11] Kawahara, H., Masuda-Katsuse, I. and de Cheveigne, 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).
- [12] Kawahara, H., Morise, M., Takahashi, T., Nisimura, R., Irino, T. and Banno, H.: TANDEM-STRAIGHT: A temporally stable power spectral representation for periodic signals and applications to interference-free spectrum, F0 and aperiodicity estimation, *ICASSP 2008*, Las Vegas, pp. 3933–3936 (2008).
- [13] Kawahara, H., Morise, M., Banno and Skuk, V. G.: Temporally variable multi-aspect N-way morphing based on interference-free speech representations, *ASPIPA ASC 2013*, p. 0S28.02 (2013).
- [14] 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*, No. September, Tokyo, pp. 38–41 (2010).
- [15] Kawahara, H., Agiomyrgiannakis, Y. and Zen, H.: YANG vocoder, Google (online), available from (<https://github.com/google/yang-vocoder>) (accessed 2017-01-17).
- [16] Boashash, B.: Estimating and interpreting the instantaneous frequency of a signal. I. Fundamentals, *Proceedings of the IEEE*, Vol. 80, No. 4, pp. 520–538 (online), DOI: 10.1109/5.135376 (1992).
- [17] Boashash, B.: Estimating and Interpreting the Instantaneous Frequency of a Signal - Part 2: Algorithms and Applications, *Proceedings of the IEEE*, Vol. 80, No. 4, pp. 540–568 (online), DOI: 10.1109/5.135378 (1992).
- [18] Cohen, L.: *Time-frequency analysis*, Prentice Hall, Englewood Cliffs, NJ (1995).
- [19] Flanagan, J. L. and Golden, R. M.: Phase Vocoder, *Bell System Technical Journal*, Vol. 45, No. 9, pp. 1493–1509 (online), DOI: 10.1002/j.1538-7305.1966.tb01706.x (1966).
- [20] 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 (online), DOI: 10.1109/ICASSP.2011.5947584 (2011).
- [21] 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).
- [22] 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., *Icslp-2000*, Beijing, China, pp. 1–4 (online), available from (<http://www.isca-speech.org/archive/>) (2000).
- [23] 榑原健一: 世界の歌声-歌唱における supernormal な声, 日本音響学会誌, Vol. 70, No. 9, pp. 499–505 (2014).
- [24] Sakakibara, K.-I., Imagawa, H., Yokonishi, H., Kimura, M. and Tayama, N.: Physiological observations and synthesis of subharmonic voices, *Asia-Pacific Signal and Information Processing Association Annual Summit and Conference*, pp. 1079–1085 (2011).

付 録

A.1 FFT に基づく瞬時周波数の導出

複素数を値とする信号 $x(t) = r(t)e^{j\theta(t)}$ について、Flanagan による瞬時周波数 $\omega_i(t)$ は以下により求められる [19]。

$$\begin{aligned} \omega_i(t) &= \frac{d\theta(t)}{dt} = \Im \left[\frac{d \log(x(t))}{dt} \right] = \Im \left[\frac{1}{x(t)} \frac{dx(t)}{dt} \right] \\ &= \frac{\Re[x(t)]\Im \left[\frac{dx(t)}{dt} \right] - \Im[x(t)]\Re \left[\frac{dx(t)}{dt} \right]}{|x(t)|^2}. \quad (\text{A.1}) \end{aligned}$$

窓関数 $w(t)$ を用いた短時間 Fourier 変換の成分 $X(\omega, t)$ を上記の複素数信号とみなして、式 A.1 を用いることでそれぞれの周波数における瞬時周波数 $\omega_i(\omega, t)$ を求めることができる。

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

ここで窓関数は偶関数を用いる。成分 $X(\omega, t)$ の時間に関する導関数 $X_d(\omega, t)$ を具体的に求めると、以下が得られる。

$$\begin{aligned} X_d(\omega, t) &= \frac{dX(\omega, t)}{dt} \\ &= j\omega \int_{-\infty}^{\infty} w(t - \tau) \exp(-j\omega\tau) x(\tau) d\tau \\ &\quad - \int_{-\infty}^{\infty} w_d(\tau - t) \exp(-j\omega\tau) x(\tau) d\tau, \quad (\text{A.3}) \end{aligned}$$

ここで $w_d(t)$ は、窓関数の導関数である。

$$w_d(t) = \frac{dw(t)}{dt}. \quad (\text{A.4})$$

これらの結果をまとめると、以下が得られる。

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

A.2 パワースペクトル加重平滑化瞬時周波数

瞬時周波数の特異値は、式 A.5 の分母が 0 となることによる。分子はパワースペクトルであり、常に非負値をとる。パワースペクトルの値が 0 となる可能性のある区間幅よりも広い幅の定義域で正值となる平滑化関数 $w_S(\omega)$ とパワースペクトルを畳み込めば、分母が常に正值となることを保証できる。パワースペクトルを重みとして瞬時周波数の加重平均を求めると、式 A.5 の分母が相殺される。したがって、式 A.5 の分子を $w_S(\omega)$ により平滑化したものを分子とし、平滑化したパワースペクトルを分母とすることで、発散しない瞬時周波数 $\omega_S(\omega, t)$ を求めることができる [16, 17]。

$$\omega_S(\omega, t) = \frac{\int_{-\infty}^{\infty} w_S(\nu - \omega) P(\nu, t) \omega_i(\nu, t) d\nu}{\int_{-\infty}^{\infty} w_S(\nu - \omega) P(\nu, t) d\nu}, \quad (\text{A.6})$$

なお、ここでは $P(\omega, t) = |X(\omega, t)|^2$ と定義した。

A.2.1 窓関数の選択

図 A-1 に、パワースペクトル加重平滑化瞬時周波数計算における窓関数と平滑化区間の影響を示す。図 A-1 では Hann 窓と、Nuttall 窓（文献 [21] の表 II の 11 番目の項目）を比較している。入力信号は、40 Hz の周期的パルス列に正規乱数を加えて作成した*4。窓長は、周波数領域での最初の零点が基本周波数となるように設定した。分析に用いる窓とパルス列の相対位相は、 $[0, 2\pi)$ の区間で一様分布するようにランダム化した。

隣接する調波との中間の周波数において生ずる特異値のため、 $0.5f_0$ と $1.5f_0$ で瞬時周波数は、大きく変動する。高 SNR の場合には、中心周波数 f_0 での瞬時周波数は、一定値 f_0 となる。Hann 窓の場合には、サイドローブのレベルが高いため、高 SNR の場合でも、FO 周辺での写像の傾斜は大きく変動する。中段の右側に示すように、サイドローブのレベルが十分に低い Nuttall 窓とパワースペクトル加重平滑化の組み合わせにより、TANDEM のような時間分解能の劣化を招かずに、分析位置による変動を抑圧することができる。なお、平滑化には引数が $\pm\pi$ の区間内で定義される raised cos 関数 $0.5 + 0.5 \cos(2\pi f/b_w)$ を用いた。このように Nuttall 窓と平滑化を用いることで、雑音成分による影響のみを傾斜の変動として観測することができる。

A.3 傾斜指標

このパワースペクトル平滑化瞬時周波数 $\omega_S(\omega, t)$ を出発点として、傾斜指標を以下のように定義する。まず、周波数傾斜 $g_F(\omega, t)$ と時間周波数傾斜 $g_{TF}(\omega, t)$ を以下のように定義する。

$$g_F(\omega, t) = \frac{d\omega_S(\omega, t)}{d\omega} = -\frac{d\tau_g(\omega, t)}{dt} \quad (\text{A.7})$$

$$g_{TF}(\omega, t) = \frac{d}{dt} \left[\frac{d\omega_S(\omega, t)}{d\omega} \right] = -\frac{d^2\tau_g(\omega, t)}{dt^2}, \quad (\text{A.8})$$

*4 実験に用いた標準化周波数 44100 Hz で周期が整数サンプル数となるように設定したため、実際には 39.9819 Hz。

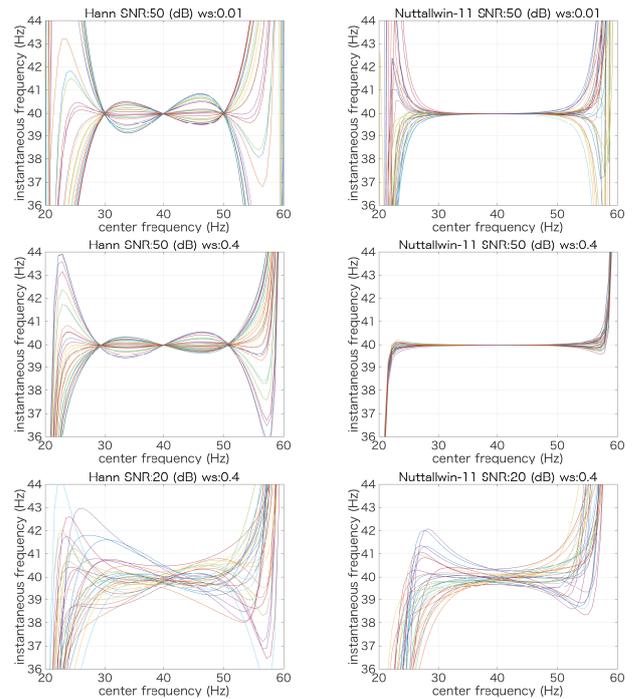


図 A-1 パワースペクトル加重平滑化瞬時周波数計算における窓関数と平滑化区間の影響。左側が Hann 窓、右側が Nuttall 窓によるもの。最上段の平滑化区間長は、基本周波数の 0.01 倍、中段と下段は 0.4 倍。SNR は、最上段と中段が 50 dB、下段が 20 dB。

Fig. A-1 Effects of time windowing function and smoothing width on power spectrum weighted average instantaneous frequency. Left plots show the mapping from center frequency to instantaneous frequency using Hann window. Right plots show those using Nuttall window. The top plots uses smoothing length $0.01f_0$. The middle and bottom plot uses $0.4f_0$. SNRs are 50 dB for top and middle plots and 20 dB for bottom plots.

ここで、 $\tau_g(\omega, t)$ は、対応する位相の周波数微分に負号をつけたものとして定義される群遅延を表す。

これらを用いて、傾斜指標 $g_A(\omega, t)$ を以下により定義する。

$$g_A(\omega, t) = \sqrt{|g_F(\omega, t)|^2 + c_f |g_{TF}(\omega, t)|^2}, \quad (\text{A.9})$$

ここで、 c_f は、校正のための係数であり、 $|g_F(\omega, t)|^2$ の分布の中央値と $|g_{TF}(\omega, t)|^2$ の分布の中央値が一致するようにシミュレーションに基づいて決定する。群遅延としての解釈に基づけば、以下のように表すこともできる。

$$g_A(\omega, t) = \sqrt{|\Delta\tau_g(\omega, t)|^2 + c_f |\Delta^2\tau_g(\omega, t)|^2}, \quad (\text{A.10})$$

A.3.1 実装

このパワースペクトル平滑化瞬時周波数 $\omega_S(\omega, t)$ は、時間方向と周波数方向の双方で既に平滑化されているため、離散化された値の差分で微分を近似することができる。実装では、この性質を用いている。