聴覚のフィルタバンク特性を有する最適調波抽出フィルタ

非会員 西 一 樹 (電気通信大学)

正員 安 藤 繁 (東京大学)

Optimal Harmonics Extraction Filter with Auditory Filter Bank Structure

Kazuki Nishi, Non-member (University of Electro-Communications), Shigeru Ando, Member (The University of Tokyo)

An optimal filter for extracting time-varying harmonics such as a voiced speech or vibration sound from the noise-corrupted observation is presented. The Wiener filter analysis to the quasi-periodic signal involving random amplitude/pitch fluctuations leads to the constant-BW/constant-Q mixed comb filter in which each pass-band located at harmonic frequencies has a constant bandwidth in lower frequency, constant-Q characteristic in higher frequency, and those mixture appears in the intermediate frequency. Some numerical examples show that the frequency response fits well the filter bank structure built in our auditory system and advantageous to separating a voiced speech from background noise.

キーワード：聴覚フィルタ、ウィナーフィルタ、櫛形フィルタ、調波信号、ブラウン運動、音声強調

Keywords: auditory filter, Wiener filter, comb filter, harmonic signal, Brownian motion, speech enhancement

1. ま え が き

楽音や音声、機械振動などをセンシング対象とする場合に、これらの信号波形がもつ擬似周期性（概ね周期性を保ちつつ振幅や周期がランダムに変化する性質）に有利な信号処理の仕組みを考えることは、センサの高性能化にとって不可欠な課題である。

この機能を元々備えているのが我々の聴覚である。聴覚末梢系が原理的には中心周波数の異なる多数の帯域通過フィルタで構成され、聴取音に対して周波数分析を行っていることは、生理学的にも心理学的にも明白な事実になっている⁽¹⁾。特に我々が騒音環境において周期音を聞き分けられるのは、フィルタバンクをなす蝸牛基底膜上にスペクトルの離散発火パターンが展開され、他の音のスペクトルとの明確な違いをパターンとして脳が識別するからであると考えられている⁽²⁾。この知見を踏まえ、フィルタバンクを構成し聴覚末梢系の処理を模擬したり⁽³⁾、これに基づく雑音抑圧の機能を実現しようとする試み⁽⁴⁾⁽⁵⁾がなされている。

一方、信号処理的立場から解釈すれば、特定の周期音を識別する機能とは、基本周波数（ピッチ周波数）間隔に細線スリットが並んだ櫛歯状の伝達特性をもつフィルタ（櫛形フィルタ）によって、その櫛歯間隔にちょうど合致した離散スペクトル成分を抜き出すことに基づいていると説明できる⁽⁶⁾。特に目的音が完全な周期信号で、かつその基本周波数が正確にわかっていれば、それを雑音中から分離す

るための最適フィルタは櫛形フィルタになる⁽⁷⁾。一般には、揺らぎによるスペクトル広がりを考慮し、倍音周波数毎に等帯域幅のバンドパスフィルタを配置した定 BW (Band-Width) 櫛形フィルタが用いられ、その構成法として短時間 Fourier 変換によるもの⁽⁸⁾や IIR 型のもの⁽⁹⁾などが提案されている。

これまで筆者らは、振幅が一定値、基本周波数が確率変数をとる場合の限定された条件の下で最適フィルタを求め⁽¹⁰⁾、そこで得られたフィルタの時間/周波数特性が定 Q 櫛形構造をとるという興味深い性質があることを示すとともに⁽¹¹⁾、振幅変動とピッチ変動を同時に考慮した擬似周期信号に対する Kalman フィルタを導いた⁽¹²⁾⁽¹³⁾。しかしながら、聴覚フィルタとの関係について議論するまでには至らなかった。

そこで本論文では、擬似周期信号に対する最適フィルタとして Wiener フィルタ解を新たに導くことによって、その周波数特性から聴覚のフィルタバンク構造が合理的に説明できることを示す。具体的には、複素位相が Brown 運動過程をとる擬似周期信号モデルに対してフィルタ解を求めることにより、その周波数特性が定 BW/定 Q 複合櫛形特性（低次から高次倍音になるにしたがって一定帯域幅から定 Q 特性をとるように各倍音周波数上にバンドパスフィルタが並んだ特性）をもつことを示す。特にフィルタパラメータを適切に選ぶと、臨界帯域幅に基づいて測定された聴覚フィルタバンクの周波数/帯域幅特性⁽¹⁴⁾にほぼ一致することを確かめる。さらに、このような定 BW/定 Q 櫛形

フィルタが音声認識のフロントエンドでの騒音抑圧処理において有効に働くことを検証するために、LPC ケプストラムに基づくスペクトル包絡を距離尺度としてフィルタの性能評価を行う。

2. 調波抽出フィルタの基本特性

〈2・1〉 擬似周期信号モデルの一般形 理想周期信号において成り立つ Fourier 級数展開の拡張として、振幅や基本周波数の変動を考慮した時変周期信号

$$x(t) = \sum_{n=1}^N c_n e^{\int_0^t \Delta c_n(s) ds} e^{jn \int_0^t \omega_0(s) ds} \dots \dots \dots (1)$$

を考える。振幅変動がなく ($\Delta c_n(s) \rightarrow 0$)、基本周波数を一定 ($\omega_0(s) \rightarrow \bar{\omega}_0$) とした特別な場合が Fourier 級数展開になる。これに対し観測に供する信号は、背景雑音 $v(t)$ が付加されたもの

$$y(t) = x(t) + v(t) \dots \dots \dots (2)$$

とする。ただし、不規則変動を扱うため、 c_n は複素数の確率変数、 $\Delta c_n(t)$ は複素数の確率過程、 $\omega_0(t), v(t)$ は実数の確率過程をとるものとし、振幅変動 $\Delta c_n(t)$ や基本周波数 $\omega_0(t)$ の変動は十分小さく、すべてのパラメータの間は、(異なる倍音次数間、複素数の実部と虚部の間も含め) 統計的に互いに独立であるとする。ただし、波形包絡の変化など倍音間で共通した振幅変動をもつ可能性を考慮して、異なる次数間の $\Delta c_n(t)$ は相関があってもよいものとする。また、簡単のため $E[c_n] = 0$, $E[v(t)] = 0$ とする ($E[\cdot]$ は集合平均)。例えば有声音は、概ね周期性をもち振幅変動とピッチ変動が互いに無相関になることが知られているため⁽¹⁵⁾、この仮定をほぼ満たすと考えられる。

〈2・2〉 スペクトル特性 さらに音量や音程が一定とみなせる音声区間に着目すれば、振幅や基本周波数には細かな不規則変動が伴うものの⁽¹⁶⁾、その統計的性質はほぼ定常となる⁽¹⁷⁾。このことを考慮すれば、複素振幅 $c_n(t) = c_n e^{\int_0^t \Delta c_n(s) ds}$ の自己相関関数 $E[c_n(t+\tau)c_n^*(t)] \equiv R_{c_n}(\tau)$ 、およびキャリア項 $\psi_n(t) = e^{jn \int_0^t \omega_0(s) ds}$ の自己相関関数 $E[\psi_n(t+\tau)\psi_n^*(t)] \equiv R_{\psi_n}(\tau)$ を用いて、擬似周期信号 $x(t)$ の自己相関は

$$E[x(t+\tau)x^*(t)] = \sum_{n=1}^N R_{c_n}(\tau) R_{\psi_n}(\tau) \equiv R_x(\tau) \dots \dots (3)$$

と表せる。ただし、仮定より複素振幅と基本周波数間および異なる倍音次数間は互いに無相関であることを用いた。このとき、信号 $x(t)$ のスペクトル密度関数 $S_x(\omega)$ ($R_x(\tau)$ の Fourier 変換) は、(3) 式より

$$S_x(\omega) = \sum_{n=1}^N S_{x_n}(\omega), \dots \dots \dots (4)$$

$$\text{where } S_{x_n}(\omega) = \frac{1}{2\pi} S_{c_n}(\omega) * S_{\psi_n}(\omega)$$

となる。ただし、 $*$ は畳み込み演算、 $S_{c_n}(\omega), S_{\psi_n}(\omega)$ はそれぞれ $R_{c_n}(\tau), R_{\psi_n}(\tau)$ の Fourier 変換を表す。振幅変動やピッチ変動が十分小さいことを仮定しているので、 $S_{c_n}(\omega)$ は $\omega = 0$ に、 $S_{\psi_n}(\omega)$ は $\omega = n\bar{\omega}_0$ (ただし $\bar{\omega}_0 = E[\omega_0(t)]$) に分布が集中するインパルス状の関数になる。したがって (4) 式より、擬似周期信号のスペクトル密度分布 $S_x(\omega)$ は、周波数上等間隔 ($\omega = \bar{\omega}_0, \dots, N\bar{\omega}_0$) にピークが並ぶ櫛形分布となる。

〈2・3〉 調波抽出フィルタの一般解 このような信号 $x(t)$ に対する観測信号 $y(t)$ からの推定値 $\hat{x}(t)$ を、線形時不変フィルタの出力

$$\hat{x}(t) = \int_{-\infty}^{\infty} h(s)y(t-s)ds \dots \dots \dots (5)$$

として、平均二乗誤差

$$J = E[|x(t) - \hat{x}(t)|^2] \dots \dots \dots (6)$$

が最小になるように求めることを考える。インパルス応答 $h(t)$ に因果性を要求しなければ、その伝達関数 $H(\omega)$ は直ちに、

$$H(\omega) = \frac{S_x(\omega)}{S_y(\omega)} \dots \dots \dots (7)$$

与えられる⁽¹⁸⁾。ここで、観測雑音 $v(t)$ の自己相関関数 $R_v(\tau)$ に対するパワースペクトルを $S_v(\omega)$ とすれば、観測信号 $y(t)$ のパワースペクトルは

$$S_y(\omega) = S_x(\omega) + S_v(\omega) \dots \dots \dots (8)$$

と表せることに注意する。

(7) 式に (4), (8) 式を代入すれば、各倍音成分を単独で抽出するフィルタ

$$H_n(\omega) = \frac{S_{x_n}(\omega)}{S_{x_n}(\omega) + S_v(\omega)} \dots \dots \dots (9)$$

の和として、

$$H(\omega) = \sum_{n=1}^N \frac{S_{x_n}(\omega)}{S_x(\omega) + S_v(\omega)} \simeq \sum_{n=1}^N H_n(\omega) \dots \dots (10)$$

と表せる。近似式となるのは、振幅変動やピッチ変動が十分小さいことを仮定しているため、各倍音間の干渉がほとんど無視できることによる。ここで簡単のため、観測雑音にはほぼ一様なスペクトル密度分布を仮定し、倍音周波数 $\omega = n\bar{\omega}_0$ の近傍で $S_{x_n}(\omega) \gg S_v(\omega)$ 、それ以外の周波数で $S_{x_n}(\omega) \ll S_v(\omega)$ が成り立つとすれば、(10) 式より、 $H_n(\omega)$ は倍音周波数 $\omega = n\bar{\omega}_0$ の近傍で $H_n(\omega) \simeq 1$ 、それ以外の周波数で $H_n(\omega) \simeq 0$ となるような n 次倍音を抽出する狭帯域バンドパスフィルタとなって、 $H(\omega)$ はそれらが倍音周波数間隔で並ぶ櫛形フィルタになる。

3. 定 BW/定 Q 複合櫛形フィルタ

〈3・1〉 Brown 運動過程の複素位相をもつ擬似周期信号モデル 音声にみられる振幅と基本周波数の変動の多く

は Gauss 過程をとること⁽¹⁵⁾、およびその周波数特性が概ね白色である⁽¹⁹⁾という知見を考慮し、(1) 式で

$$\begin{aligned}\Delta c_n(t) &= -\text{sign}(t)\sigma_{an}^2 + \sigma_{an}\xi_{an}(t) + j\sigma_{bn}\xi_{bn}(t), \\ \omega_0(t) &= \bar{\omega}_0 + \sigma_\omega \xi_\omega(t) \dots\dots\dots (11)\end{aligned}$$

とおいた場合を考える。ただし、 $\text{sign}(t)$ は $t \geq 0$ のとき 1, $t < 0$ のとき -1 をとる符号関数であり、 $\xi_{an}(t)$, $\xi_{bn}(t)$, $\xi_\omega(t)$ は $E[\xi_*(t)] = 0$, $E[\xi_*(t+\tau)\xi_*(t)] = \delta(\tau)$ が成り立つ互いに独立な白色 Gauss 雑音、 $\sigma_{an}, \sigma_{bn}, \sigma_\omega > 0$ はそれぞれ $\Delta c_n(t)$ の実部と虚部、 $\omega_0(t)$ に対する標準偏差である。

このとき、(1) 式は

$$x(t) = \sum_{n=1}^N c_n e^{-\sigma_{an}^2 |t| + \sigma_{an} \beta_{an}(t)} \cdot e^{j\{\sigma_{bn} \beta_{bn}(t) + n\sigma_\omega \beta_\omega(t)\}} e^{jn\bar{\omega}_0 t} \dots\dots (12)$$

となる。ただし、 $\beta_*(t)$ は

$$\beta_*(t) \equiv \int_0^t \xi_*(s) ds \dots\dots\dots (13)$$

で定義される Brown 運動過程（あるいは Wiener 過程）である⁽²⁰⁾。(12) 式中の $-\sigma_{an}^2 |t|$ は、信号が広義定常になるための補正項を意味する。実際、各倍音成分に対する自己相関が

$$\begin{aligned}E[x_n(t+\tau)x_n^*(t)] &= E[|c_n|^2]e^{-\sigma_{an}^2\{|t+\tau|+|t|\}} \\ &\cdot E[e^{\sigma_{an}\{\beta_{an}(t+\tau)+\beta_{an}(t)\}}]E[e^{j\sigma_{bn}\{\beta_{bn}(t+\tau)-\beta_{bn}(t)\}}] \\ &\cdot E[e^{jn\sigma_\omega\{\beta_\omega(t+\tau)-\beta_\omega(t)\}}]e^{jn\bar{\omega}_0\tau} \\ &= \bar{x}_n^2 e^{-\frac{1}{2}\sigma_n^2|\tau|} e^{jn\bar{\omega}_0\tau} = R_{x_n}(\tau) \dots\dots\dots (14)\end{aligned}$$

となることから確かめられる。ただし、 $\bar{x}_n^2 \equiv E[|c_n|^2]$, $\sigma_n^2 \equiv \sigma_{an}^2 + \sigma_{bn}^2 + n^2\sigma_\omega^2$ とおき、導出過程で付録 1 の関係を用いた。

〈3・2〉 Wiener フィルタ解 (12) 式の信号モデルに対して、 $h(t) = 0$ ($t < 0$) を満たす因果性フィルタ（すなわち Wiener フィルタ）を設計する。

この場合も、振幅変動やピッチ変動が十分小さいことを仮定しているため、各倍音間の干渉が無視でき、(10) 式右辺と同様に各倍音成分を単独で抽出するフィルタ $H_n(\omega)$ の和として近似的に表せる。したがって、単一倍音成分のみに対する Wiener フィルタ解を求める問題に帰着する。これは、スペクトル分解⁽¹⁸⁾により

$$H_n(\omega) = \frac{1}{S_{yn+}(\omega)} \left[\frac{S_{xn}(\omega)}{S_{yn-}(\omega)} \right]_+ \dots\dots\dots (15)$$

と表せる。ただし、 $S_{yn+}(\omega)$, $S_{yn-}(\omega)$ は、 $s = j\omega$ とおいたときの s 平面の右半平面と左半平面それぞれで正則な（すなわち s 平面の右半平面と左半平面それぞれで零点および極をもたない）関数の積として、

$$S_{xn}(\omega) + S_v(\omega) = S_{yn+}(\omega) \cdot S_{yn-}(\omega) \dots\dots\dots (16)$$

図 1 最適調波抽出フィルタの極配置, (a) 定 BW 櫛形フィルタ, (b) 定 Q 櫛形フィルタ, (c) 定 BW/定 Q 複合櫛形フィルタ

Fig. 1. Pole allocation of the optimal harmonics extraction filter. (a) Constant-BW-type comb filter. (b) Constant-Q-type comb filter. (c) Constant-BW/constant-Q mixed comb filter.

と因数分解したときの各項に対応し、 $[\]_+$ は、 s 平面の左半平面と右半平面それぞれで正則な関数の和として、

$$\frac{S_{xn}(\omega)}{S_{yn}(\omega)} = \left[\frac{S_{xn}(\omega)}{S_{yn}(\omega)} \right]_- + \left[\frac{S_{xn}(\omega)}{S_{yn}(\omega)} \right]_+ \dots\dots\dots (17)$$

と表したときの後者である。

各倍音成分 $x_n(t)$ のスペクトル密度関数が、(14) 式の Fourier 変換により

$$S_{xn}(\omega) = \bar{x}_n^2 \frac{\sigma_n^2}{\frac{1}{4}\sigma_n^4 + (\omega - n\bar{\omega}_0)^2} \dots\dots\dots (18)$$

と表せることに注意し、観測雑音 $v(t)$ に白色雑音を仮定して $S_v(\omega) = \sigma_v^2$ が成り立つとすれば、(15) 式は

$$H_n(\omega) = \frac{\sqrt{\frac{\bar{x}_n^2}{\sigma_v^2} \sigma_n^2 + \frac{1}{4}\sigma_n^4 - \frac{1}{2}\sigma_n^2}}{\sqrt{\frac{\bar{x}_n^2}{\sigma_v^2} \sigma_n^2 + \frac{1}{4}\sigma_n^4} + j(\omega - n\bar{\omega}_0)} \dots\dots\dots (19)$$

と求まる。この結論は、Kalman フィルタ解析による定常解に一致する⁽¹³⁾。振幅/ピッチ変動が十分小さいため $\frac{\bar{x}_n^2}{\sigma_v^2} \sigma_n^2 \gg \frac{1}{4}\sigma_n^4$ が成り立ち、簡単のため $\bar{x}_1^2 = \dots = \bar{x}_N^2 = \bar{x}^2$, $\sigma_{a1}^2 = \dots = \sigma_{aN}^2 = \sigma_a^2$, $\sigma_{b1}^2 = \dots = \sigma_{bN}^2 = \sigma_b^2$ とすれば、(10) 式右辺より

$$H(\omega) \simeq \sum_{n=1}^N \frac{\frac{\bar{x}}{\sigma_v} \sqrt{\sigma_a^2 + \sigma_b^2 + n^2\sigma_\omega^2}}{\frac{\bar{x}}{\sigma_v} \sqrt{\sigma_a^2 + \sigma_b^2 + n^2\sigma_\omega^2} + j(\omega - n\bar{\omega}_0)} \dots\dots\dots (20)$$

を得る（ただし $\bar{x} = \sqrt{\bar{x}^2}$ ）。

$s = j\omega$ において s 平面上での伝達関数 $H(\frac{s}{j})$ の極配置を示したのが図 1 である。特に振幅変動のみを考慮した場合 ($\sigma_\omega^2 \rightarrow 0$) は、 s 平面の左半面上で虚軸と並行な直線上に等間隔に極が並ぶことにより（同図破線 (a)）、定 BW 櫛形特性（一定帯域幅のバンドパスフィルタが各倍音周波数

図2 定BW/定Q複合櫛形フィルタの周波数特性
Fig. 2. Frequency response of the constant-BW/constant-Q mixed comb filter.

図3 フィルタバンクの中心周波数と占有帯域幅の関係
Fig. 3. Relationship between center frequency and bandwidth of the filter banks.

上に並ぶ特性)を有し、逆にピッチ変動のみを考慮した場合 ($\sigma_a^2, \sigma_b^2 \rightarrow 0$) は、原点を通る斜め直線上に等間隔に極が並ぶことにより (同図破線 (b)), 定Q櫛形特性 (倍音次数に比例して帯域幅が広がるバンドパスフィルタが各倍音周波数上に並ぶ特性)を有することになる。さらに両者をともに考慮した場合は、同図破線 (c) のような極配置となり、低次倍音では相対的に定BW特性が、高次倍音では逆に定Q特性が支配的になって、低次から高次になるにつれて定BW特性から定Q特性へと徐々に遷移する定BW/定Q複合櫛形特性を有することになる (図2参照)。

4. 聴覚の周波数特性との整合性

聴覚末梢系を構成するフィルタバンクの中心周波数と占有帯域幅の関係は、Zwicker⁽¹⁴⁾によって「臨界帯域幅」と呼ばれる基準を用いて精密に測定されている。そこで、定BW/定Q複合櫛形フィルタの周波数特性がこの関係によく適合することを以下で確かめる。

(19)式のパンドパスフィルタに対して適当な閾値以上の利得をとる占有帯域幅を求め、

$$\Delta\omega_n = k \sqrt{\frac{\bar{x}^2}{\sigma_v^2} \sigma_n^2 + \frac{1}{4} \sigma_n^4} \dots \dots \dots (21)$$

この値をもってフィルタ入力に作用する実効帯域幅と定義

図4 各フィルタ係数に対する出力信号のSN比, “X”は入力信号に与えた振幅/ピッチ変動の標準偏差を示す
Fig. 4. Signal-to-Noise ratios of output signals computed with various filter parameters, where “X” indicates normal deviations of amplitude/pitch fluctuations included in the input signal.

する。ただし、特定の倍音構造を分析対象とする訳ではないので、(19)式で $\bar{x}_n^2 = \bar{x}^2$, $\sigma_n^2 = \sigma_a^2 + \sigma_b^2 + n^2 \sigma_\omega^2$ とおいた。 k は実効帯域幅の基準を決めるための利得閾値であり、Zwickerが測定した周波数/臨界帯域幅特性⁽¹⁴⁾の最小周波数側で帯域幅が一致するように、フィルタの最大利得から-27dB減衰する値を基準として選んだ。また、任意の基本周波数に対して櫛形フィルタが連続的に選択できる構成をとるために、十分密に並んだフィルタバンクが予め用意されていると考える。このことを等価的に実現するために、基本周波数を $\omega_0/2\pi = 100[\text{Hz}]$ と固定するかわりに、上式で倍音次数 n が実数値をとるものとする。

図3は、各パラメータを $\sigma_a^2 + \sigma_b^2 = 0.1[\text{s}^{-1}]$, $\sigma_\omega = 0.002[\text{s}^{-1}]$, $\bar{x}^2/\sigma_v^2 = 6[\text{dB}]$ と設定し、(21)式の $n\omega_0$ を横軸、 $\Delta\omega_n$ を縦軸にとってプロットした結果である。ほぼ全ての周波数にわたってZwickerの周波数/臨界帯域幅特性⁽¹⁴⁾とよく一致する。特に、勾配が徐々にきつくなる高周波数側の曲線部分においても聴覚特性とよく一致するという事実は、定Qの比例関係よりさらに高次の帯域広がりを生じさせるBrown運動によるピッチ変動モデル ((12)式)が、聴覚にとって標準的な分析対象になっている可能性を示唆している。

5. 有声音抽出への応用

〈5・1〉 デジタルフィルタの導出 上記の定BW/定Q複合櫛形フィルタを計算機上で実現するには、デジタル形式に変換しておく必要がある。これは、インパルス不変法⁽²¹⁾に基づいて(20)式をs-z変換することにより直ちに、

$$H_D(z) = \sum_{n=1}^N \frac{1 - \gamma_n}{1 - \gamma_n z_0^n z^{-1}}, \dots \dots \dots (22)$$

図5 LPC ケプストラムによるフィルタ性能比較
Fig. 5. Comparison of the filter performance based on the LPC cepstrum.

$$\text{where } \gamma_n = 1 - \frac{\bar{x}}{\sigma_v} \sqrt{\sigma_a^2 + \sigma_b^2 + n^2 \sigma_\omega^2} \cdot T_s, \quad z_0 = e^{j\bar{\omega}_0 T_s}$$

と表せる。ただし、 T_s は離散化の際のサンプリング間隔である。

〈5・2〉フィルタ設計の最適性 有声音「あ」の1周期分を切り出したものに対して Fourier 係数 c_n を求め、(12) 式によって振幅/ピッチ変動 ($\sigma_a = 0.0005$, $\sigma_b = 0$, $\sigma_\omega = 0.0005$) を与え擬似周期信号を合成する。これに白色雑音を付加したものを観測信号として上記のフィルタに入力し、フィルタ係数 $\sigma_a^2, \sigma_b^2, \sigma_\omega^2$ をいろいろ変えて出力信号を求め、各 SN 比を計算した結果が図4である。フィルタ出力が最大の SN 比をとる点と、入力信号に与えた振幅/ピッチ変動の標準偏差を示す点がほぼ一致することから、ほぼ理論通りの最適フィルタが実現できているといえる。

〈5・3〉LPC ケプストラムによるフィルタ性能評価 男性アナウンサー1名による有声音「あいうえお」に白色雑音を付加したものに対して、従来型の定 BW 楕形フィルタおよび新提案の定 BW/定 Q 複合楕形フィルタの両者を使って音声抽出を実験を行った。

(22) 式のピッチパラメータ $\bar{\omega}_0$ には、事前に論文⁽²²⁾の手法により推定したものを用いる。また各楕形フィルタの帯域幅を決めるパラメータ $\sigma_a^2, \sigma_b^2, \sigma_\omega^2$ は、雑音重畳音から直接推定することは困難なので、原音声から直接求めた振幅変動とピッチ変動の分散値により決定した。サンプリング周波数 $1/T_s$ を 16kHz、フィルタの倍音次数 N を 30 とした。このときのフィルタ出力波形に対する評価基準には、音声認識のための特徴量として利用されている LPC ケプストラム⁽²³⁾を用いる。原音声、雑音重畳音およびフィルタ出力信号の各データに対して LPC ケプストラム係数を求め、原音声のそれとのユークリッド距離を計算した結果が図5である。

振幅変動だけでなく基本周波数推定時の誤差やピッチ変

動も考慮した定 BW/定 Q 複合楕形フィルタの方が、振動変動のみを考慮した定 BW 型のものに比べ全体的に良好な結果を示している。ただし、「あ」や「え」の部分では何もしない方がむしろ良いという結果が現れているのは、定 BW/定 Q 複合楕形フィルタであっても吸収しきれないピッチ誤差や変動があるためと考えられる。それでもなお、定 BW 楕形フィルタよりは良い結果を示していることから、常に安定した性能を発揮する定 BW/定 Q 複合楕形フィルタを用いることによって、騒音環境下での認識率を全体的に向上できる可能性が期待される。

以上の実験は、高次倍音を多く含むプロのアナウンサーによる音声を用いているため、定 Q 楕形フィルタの効果が比較的現れ易い実験例であると言える。実際、普通の人の音声に対して行った同様な実験では、定 BW 型と定 BW/定 Q 複合型との間には上述したほどの明確な違いは表れなかった。しかしながら、音声だけでなく豊かな音色をもつ楽音などへの適用も考えた場合、全周波数領域にわたって頑健性を有する定 BW/定 Q 複合楕形フィルタの方が最終的に優位であると考えられる。

6. む す び

振幅や基本周波数の変動を伴う擬似周期信号を雑音環境下で抽出するための最適フィルタとして、一般的な信号モデルの場合と複素位相が Brown 運動過程をとる場合の両者に対して Wiener フィルタを導いた。その周波数特性はいずれも楕形特性を有し、特に後者のモデルにおいては、低次倍音から高次倍音になるにつれて定 BW 型から定 Q 型に徐々に遷移する複合楕形特性を有することを示すとともに、このような定 BW/定 Q 複合特性が聴覚フィルタバンクの周波数/帯域幅特性にほぼ一致することを確認した。さらに、合成した擬似周期信号を用いたシミュレーションにより理論通りの最適抽出が可能なこと、および LPC ケプストラムを評価基準として用いた有声音に対するフィルタ性能比較実験では、ピッチ推定誤差や変動に対する定 BW/定 Q 複合楕形フィルタの頑健性を検証した。

今後は、フィルタ係数を決める際に不可欠な振幅/ピッチ変動の標準偏差の推定法、あるいはフィルタ係数の自動調整アルゴリズム、さらには非定常信号成分に対するフィルタ処理との統合による調波信号センシングシステムの確立を目指したい。

(平成 13 年 8 月 28 日受付, 同 13 年 12 月 25 日再受付)

文 献

- (1) 境 久雄, 聴覚と音響心理, コロナ社 (1992)
- (2) A.S.Bregman, *Auditory Scene Analysis - the perceptual organization of sound*, MIT Press (1990)
- (3) R.Patterson, "The deterioration of hearing with age: Frequency selectivity, the critical ratio, the audiogram, and speech threshold," *J. Acoust. Soc. Am.*, vol.72, no.6, pp.1788-1803 (1982)
- (4) T.Irino, "Noise suppression using a time-varying, anal-

- ysis/synthesis Gammachirp filterbank," *Proc. IEEE Int. Conf. on Acoust. Speech and Signal Process.*, pp.97-100 (1999)
- (5) 鶴木 祐史, 赤木 正人, "聴覚の情景解析に基づいた雑音下の調波複合音の一抽出法", 電子情報通信学会論文誌, vol.J82-A, no.10, pp.1497-1507 (1999)
- (6) A. de Cheveigne, "Separation of concurrent harmonic sounds: fundamental frequency estimation and a time-domain cancellation model of auditory processing," *J. Acoust. Soc. Am.*, vol.93, no.6, pp.3271-3290 (1993)
- (7) 小畑 秀文, "最尤推定法に基づく音声波形の復元," 音響学会誌, vol.50, no.10, pp.809-815 (1994)
- (8) 長淵 裕実, "コームフィルタリング雑音抑圧法の基本性能と PCM 音声の品質改善への応用," 電子情報通信学会論文誌, vol.J65-A, no.6, pp.572-578 (1982)
- (9) A.Nehorai and B.Porat, "Adaptive comb filtering for harmonic signal enhancement," *IEEE Trans. Acoust., Speech and Signal Process.*, vol.ASSP-34, no.5, pp.1124-1138 (1986)
- (10) K.Nishi and S.Ando, "An Optimal Comb Filter for Time-Varying Harmonics Extraction," *IEICE Trans. Fundamentals*, vol.E81-A, no.8, pp.1622-1627 (1998)
- (11) 西, 安藤, "定 Q 楕円フィルタとその時間/周波数特性 - ピッチ変動誤差に頑健な調波信号推定のためのフィルタ特性 -, " 電子情報通信学会論文誌, vol.J83-A, no.2, pp.152-160 (2000)
- (12) 西, 安藤, "ピッチ変動と振幅変動に頑健性を有する調波信号抽出フィルタ", 電気学会論文誌, vol.119-E, no.4, pp.215-220 (1999)
- (13) K.Nishi, "Kalman filter analysis for quasi-periodic signals," *Proc. IEEE Int. Conf. on Acoust. Speech and Signal Process.*, pp.3937-3940 (2001)
- (14) E.Zwicker and E.Terhardt, "Analytical expressions for critical-band rate and critical bandwidth as a function of frequency," *J. Acoust. Soc. Am.*, vol.68, no.5, pp.1523-1525 (1980)
- (15) 小室 修, 粕谷 英樹, "基本周期のゆらぎの性質とそのモデル化に関する検討," 音響学会誌, vol.47, no.12, pp.928-934 (1991)
- (16) 北風 裕教, 赤木 正人, "基本周波数の微細変動成分に対する知覚," 電子情報通信学会技術研究報告, SP98-168, pp.9-16 (2000)
- (17) 青木 直史, 伊福部 達, "持続発声母音における振幅ゆらぎ及びピッチゆらぎの周波数特性とその音響心理的效果," 電子情報通信学会論文誌, vol.J82-A, no.5, pp.649-657 (1999)
- (18) A. Papoulis, *Signal Analysis*, McGraw-Hill (1977)
- (19) 坂本 勤, 粕谷 英樹, "合成音の自然性に寄与する基本周期の変動特性," 音響学会秋季研究発表会講演論文集, pp.169-170 (1985)
- (20) 小倉 久直, (続) 確率過程論, コロナ社 (1987)
- (21) 三谷 政昭, デジタルフィルタデザイン, 昭晃堂 (1997)
- (22) 西, 安藤, 安藤, "聴覚情景分析のための多重ピッチ追跡と調波分離アルゴリズム", 計測自動制御学会論文集, vol.34, no.6, pp.483-490 (1998)
- (23) 今井 聖, 音声信号処理, 森北出版 (1996)
- (24) B. エクセンダー著, 谷口 説男訳, 確率微分方程式, シュプリンガー・フェアラーク東京 (1999)

付 録

1. Brown 運動過程に成り立つ関係

(13) 式の Brown 運動過程に対する差分

$$\Delta\beta(t_1, t_2) \equiv \beta(t_2) - \beta(t_1) = \int_{t_1}^{t_2} \xi(s) ds \cdots (\text{付 1})$$

を考える。 $\xi(s)$ が Gauss 過程をとるから上式も Gauss 過程となり, その平均および分散は

$$\begin{aligned} E[\Delta\beta(t_1, t_2)] &= \int_{t_1}^{t_2} E[\xi(s)] ds = 0 \\ E[(\Delta\beta(t_1, t_2))^2] &= \int_{t_1}^{t_2} \int_{t_1}^{t_2} E[\xi(s)\xi(u)] ds du \\ &= \int_{t_1}^{t_2} \int_{t_1}^{t_2} \delta(s-u) ds du \\ &= |t_1 - t_2| \cdots \cdots (\text{付 2}) \end{aligned}$$

となる。特に, $t_1 < t_2 < t_3$ あるいは $t_1 > t_2 > t_3$ のとき,

$$E[\Delta\beta(t_1, t_2)\Delta\beta(t_2, t_3)] = 0 \cdots \cdots (\text{付 3})$$

が成り立つから, $\Delta\beta(t_1, t_2)$ と $\Delta\beta(t_2, t_3)$ は統計的に互いに独立になる。

特に, Wiener 過程を指数部にもつ形式

$$\phi_\sigma(t) \equiv e^{\sigma\beta(t)} \cdots \cdots (\text{付 4})$$

は, 幾何 Brown 運動過程と呼ばれる⁽²⁴⁾。 $\beta(t)$ が平均 0, 分散 $|t|$ の Gauss 分布をとることから, 上式の期待値は

$$\begin{aligned} E[\phi_\sigma(t)] &= \frac{1}{\sqrt{2\pi|t|}} \int_{-\infty}^{\infty} e^{\sigma B} e^{-\frac{B^2}{2|t|}} dB \\ &= \frac{1}{\sqrt{2\pi|t|}} \int_{-\infty}^{\infty} e^{-\frac{1}{2|t|}\{B-\sigma|t|\}^2} dB \cdot e^{\frac{1}{2}\sigma^2|t|} \\ &= e^{\frac{1}{2}\sigma^2|t|} \cdots \cdots (\text{付 5}) \end{aligned}$$

と求められる。ただし, $\beta(t)$ の実現値を B とした。以上の関係を用いると, $\phi_\sigma(t)$ に対する自己相関は,

1) $t_1 > t_2 > 0$ または $t_1 < t_2 < 0$ のとき,

$$\begin{aligned} E[\phi_\sigma(t_1)\phi_\sigma^*(t_2)] &= E[e^{\sigma\{\beta(t_1)-\beta(t_2)\}}]E[e^{2\sigma\beta(t_2)}] \\ &= e^{\frac{1}{2}\sigma^2|t_1-t_2|}e^{2\sigma^2|t_2|} \cdots \cdots (\text{付 6}) \end{aligned}$$

2) $t_1 > 0 > t_2$ または $t_1 < 0 < t_2$ のとき,

$$\begin{aligned} E[\phi_\sigma(t_1)\phi_\sigma^*(t_2)] &= E[e^{\sigma\beta(t_1)}]E[e^{\sigma\beta(t_2)}] \\ &= e^{\frac{1}{2}\sigma^2|t_1|}e^{\frac{1}{2}\sigma^2|t_2|} \cdots \cdots (\text{付 7}) \end{aligned}$$

と表せる。

また, 位相変調型の幾何 Brown 運動過程

$$\phi_{j\sigma}(t) \equiv e^{j\sigma\beta(t)} \cdots \cdots (\text{付 8})$$

に対する自己相関は, Fourier 変換対

$$\frac{1}{\sqrt{2\pi|\tau|}} e^{-\frac{1}{2}\frac{\tau^2}{|\tau|}} \xleftrightarrow{\mathcal{F.T.}} e^{-\frac{1}{2}\omega^2|\tau|} \cdots \cdots (\text{付 9})$$

の関係より,

$$E[\phi_{j\sigma}(t_1)\phi_{j\sigma}^*(t_2)] = e^{-\frac{1}{2}\sigma^2|t_1-t_2|} \cdots \cdots (\text{付 10})$$

と表せる。

西 一 樹 (非会員) 1984 年電気通信大学電波通信学科卒業、86 年同大学大学院修士課程終了。同年 NTT 入社、88 年より電気通信大学電子情報学科 (現情報通信工学科) 助手となり現在に至る。音響信号処理、画像処理の研究に従事。計測自動制御学会、電子情報通信学会、IEEE などの会員 (工学博士)。

安 藤

繁 (正員) 1974 年東京大学工学部計数工学科卒業、
79 年同博士課程修了。87 年より東京大学工学部
助教授、96 年より同教授、現在に至る。画像処
理、信号処理、光および音響計測、計測基礎論、
電子回路などの研究と教育に従事。センサとセン
シングの知能化、人間の基礎的な感覚・知覚機能
の工学的実現を目指している。昭和 53 年、昭和

62 年、平成 5 年 (友田賞) 計測自動制御学会論
文賞受賞。著書に画像処理論 (昭晃堂)、電子回路 (培風館) 等。日
本音響学会、日本光学会、情報処理学会、計測自動制御学会、IEEE、
OSA などの会員 (工学博士)。