論文の内容の要旨

論文題目　　ピッチ変動に着目した擬似周期信号の最適フィルタに関する研究

氏　名　　西 一樹

1　本論文の目的

本論文は、振幅変動やピッチ変動をともなった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の設計法について論じたものである。具体的には、1) 擬似周期信号とその推定問題に対する定義を与え、2) 定常信号モデルと非定常信号モデルのそれぞれに対する線形最適フィルタの設計を試みることによって、3) 最適解として得られた櫛形フィルタの時間 / 周波数特性に成り立つ関係の解明を通して、最終的には 4) ディジタルフィルタとして実現し、5) 混合音に対するピッチ推定や各擬似周期信号の個別分離アルゴリズムの開発、および 6) 音声認識のフロントエンドとしての有用性を検証するものである。

2　本論文の構成

本論文は、序論、結論および下記 6 章の，全 8 章で構成される.

1. ピッチ変動を伴う音響・音声信号のモデル
2. 定常線形最適フィルタによる音響・音声信号の分離
3. ピッチ変動に対する定 Q 櫛形フィルタの性質と効果
4. 多重擬似周期信号に対するピッチ推定と個別分離アルゴリズム
5. 非定常擬似周期信号に対する線形最適フィルタ
6. ディジタル櫛形フィルタの設計と評価

3　本論文の概要

3.1　音響・音声信号モデルと問題設定

音響・音声信号は、大まかには周期性を保ちつつ振幅やピッチが絶えず変化する擬似周期性をもっている。このような信号のモデルとして、時変パラメータをもつ波動方程式の近似解を使って

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

と定式化する。これに雑音が重畳したもの $y(t) = x(t) + v(t)$ を観測信号として、$y(t)$ から $x(t)$ を分離するための線形フィルタを求める問題を考える。本論文では、信号の時変構造を決定する振幅変動項 $\Delta c_n(s)$ およびピッチ変動項 $\omega_0(s)$ に各種の確率モデルを導入した上でこの問題を議論した。

3.2　定常擬似周期信号に対するフィルタ設計とその性質

振幅やピッチの変動には、音楽での旋律や声の抑揚などの時間的に緩やかに変化する成分と、振動体の熱的擾乱や声帯の不規則振動による微細な変動成分がある。ここでは特に、後者の微細変動がほぼ定常的であることに着目した。一つ目の信号モデルとしては、音程 (ピッチ) を一定に保ったまま発声しても、音量 (振幅) は細かく変化し、そのときの基本周波数を正確に知ることも現実には困難であることを考慮し、振幅変動が定常過程をとり基本周波数が確率変数をとるとした場合、二つ目は、実際の音声データの振舞いを比較的よく近似できる、振幅と基本周波数がともに白色 Gauss 雑音過程をとるとした場合、のそれぞれに対するフィルタ設計を行った。

その結果いずれの場合も、得られたフィルタの周波数特性には、1) 振幅変動のみを考慮した場合は定 BW(バンド幅) 型の櫛形特性、2) 基本周波数の変動のみを考慮した場合は定 Q 型の櫛形特性、3) 両者を同時に考慮した場合は低次倍音から高次倍音になるにしたがって定 BW 型から定 Q 型へ徐々に遷移する複合櫛形特性、を有することが示された。特に 3) の定 BW/ 定 Q 複合櫛形特性は、聴覚のもつ周波数分析特性と共通性があることから、本論文の議論が聴覚システムの合理性を裏付ける根拠の一つになっていると考えられる。

3.3　定 Q 櫛形フィルタに成り立つ関係

特に、ピッチ変動のみを考慮した場合の最適フィルタである定 Q 櫛形フィルタには、そのインパルス応答と周波数特性が共通の定 Q 櫛形構造をとるという興味深い性質があることを明らかにした。すなわち、インパルス応答を

$$h_Q(t) = A\delta(t) + \sum_{\substack{n=-\infty \\ n\neq 0}}^{\infty} \frac{1}{|n|} h_0(\frac{t}{n}) \qquad \text{ただし、} \quad A = \int_{t_a}^{t_b} h_0(t)dt$$

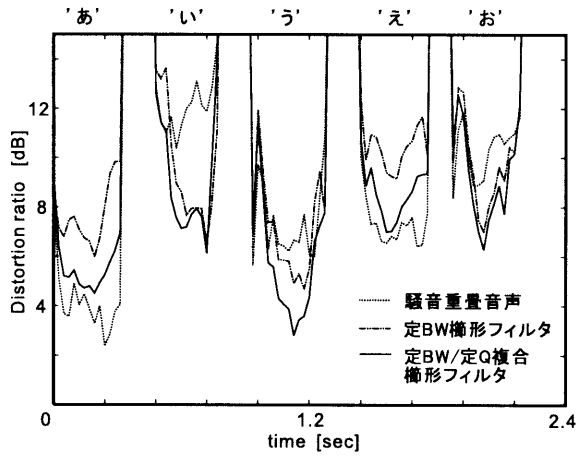
と表したとき ($h_0(t)$ は $t_a < t < t_b$ でコンパクトな台をもつ区分的に滑らかな任意の関数)、そのフーリエ変換つまり伝達関数は

$$H_Q(\omega) = B\delta(\omega) + \sum_{\substack{n=-\infty \\ n\neq 0}}^{\infty} \frac{2\pi}{|\omega|} h_0(n\frac{2\pi}{\omega}) \qquad \text{where} \quad B = 2\pi \int_{t_a}^{t_b} \frac{h_0(t)}{t} dt$$

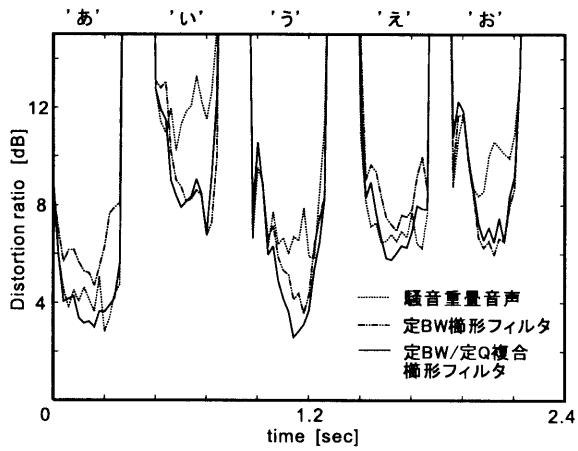
で与えられ、再び定 Q 櫛形関数になるという関係を示した。

3.4　混合音に対するピッチ推定と個別分離アルゴリズム

櫛形フィルタを構成するには、ピッチパラメータを事前に推定しておかなければならない。しかも一般の聴取音には、複数の音源からの信号が重畳していると考えられる。この場合を想定し、混合音中の複数のピッチ候補から目的ピッチを選択的に追跡しつつ、その軌跡上の調波成分を分離抽出するアルゴリズムを開発した。



(a) LPC ケプストラムによる比較



(b) LPC メルケプストラムによる比較

図 1: スペクトル包絡間の距離尺度を用いたフィルタ性能比較

具体的には、複数のピッチ候補を尤度関数の多峰性により表現し、その中でピッチ変化のダイナミクスに従うものだけが最終的な目的ピッチとして選別される仕組みを Non-Parametric Kalman フィルタにより実現した。目的信号抽出のための定 Q 櫛形フィルタ演算は、 Wavelet 空間においてピッチ軌跡から決まる各倍音周波数上の複素振幅を読みとり合成することによって等価的に実現できる。本アルゴリズムによって混合音中のピッチ軌跡が選択的に補間追跡でき、目的ストリームが分離再生できることを、シミュレーションおよび実音声実験により検証した。

3.5 非定常擬似周期信号に対するフィルタ設計

音声や楽音などの実信号では、振幅やピッチの変動の統計的性質 (期待値や分散) が時間的に変化していくことが普通である。このような 3.2節では扱えなかった非定常信号モデルに対するフィルタ設計を行った。

ここでのポイントは、複素位相項にランダム性をもつ信号モデルを伊藤確率微分方程式により記述しておくことにある。これは、各倍音成分において

$$
\begin{aligned}
dx_n(t) \quad = \quad & \left[\frac{\dot{\bar{x}}_n^2(t)}{2\bar{x}_n^2(t)} - \frac{1}{2}\{\sigma_{an}^2(t) + \sigma_{bn}^2(t) + n^2\sigma_{\omega}^2(t)\} + jn\bar{\omega}_0(t)\right] x_n(t)dt \\
& + \sigma_{an}(t)x_n(t)d\beta_{an}(t) + j\sigma_{bn}(t)x_n(t)d\beta_{bn}(t) + jn\sigma_{\omega}(t)x_n(t)d\beta_{\omega}(t)
\end{aligned}
$$

と記述できる。$\sigma_*(t)$ が振幅やピッチ周波数の標準偏差を表し、これを時変パラメータとしたことが3.2節との本質的な違いになっている。$d\beta_*(t)$ はWiener積分の微小要素である。特に右辺第2項以降は、ともに確率過程をとる信号と雑音の積で駆動される非線形な状態依存性雑音を表し伊藤積分の意味をもつ。この伊藤積分項に注意し、上式を状態空間モデルとしてKalmanフィルタを導いた結果、フィルタ方程式

$$\frac{d\hat{\boldsymbol{x}}(t)}{dt} = \left[\Omega(t) - \frac{1}{\sigma_v^2(t)}P(t)\mathbf{1}\mathbf{1}^T\right]\hat{\boldsymbol{x}}(t) + \frac{1}{\sigma_v^2(t)}P(t)\mathbf{1}y(t)$$

およびRiccati方程式

$$\frac{dP(t)}{dt} = \Omega(t)P(t) + P(t)\Omega^*(t) - \frac{1}{\sigma_v^2(t)}P(t)\mathbf{1}\mathbf{1}^T P(t) + \Sigma^2(t)$$

が得られることを詳細な解析の下で示した。ただし、$\Sigma^2(t) = \mathrm{diag}\Big\{\{\sigma_{a1}^2(t) + \sigma_{b1}^2(t) + \sigma_\omega^2(t)\}\overline{x}_1^2(t), \cdots,$ $\{\sigma_{aN}^2(t) + \sigma_{bN}^2(t) + N^2\sigma_\omega^2(t)\}\overline{x}_N^2(t)\Big\}$ である。

3.6 ディジタルフィルタの実現と評価

さらに上式を離散化近似すると、ディジタルフィルタの形式として

$$H(z) = \sum_{n=1}^{N}\frac{1-\gamma_n(kT_s)}{1-\gamma_n(kT_s)z_0^n(kT_s)z^{-1}},\ \text{ただし、}\ \gamma_n(kT_s) = 1 - \frac{\overline{x}_n(kT_s)}{\sigma_v(kT_s)}\sigma_n(kT_s)T_s,\ z_0(kT_s) = e^{j\overline{\omega}_0(kT_s)T_s}$$

のような時変なフィルタ係数をもつ伝達関数が導けることを示した。3.2節の定BW/定Q複合櫛形フィルタが単一極のディジタルフィルタとしての並列接続で実現でき、各時刻において振幅やピッチの変動量に応じてフィルタの中心周波数や帯域幅が決定できる。

振幅やピッチの変動に最適なフィルタとして得られた上式の結果が、騒音環境下での音声認識の前処理としても有用であることを確かめた。図1は、LPCケプストラムとLPCメルケプストラムにより音声「あいうえお」のスペクトル包絡を各フィルタ出力に対して求め、原音声のそれとの差を計算したものである。振幅変動とピッチ変動の両者を考慮した複合型の方が振幅変動のみを考慮した定BW型のものに比べて、いずれも全体的に良好な結果を示している。このことから、比較的安定した性能をもつ定BW/定Q複合櫛形フィルタを音声認識の前処理として用いることによって、騒音環境下での認識率を向上できる可能性が期待される。