

聴覚知覚特性に基づいたピッチ検出方式

A Pitch Detection Method based on Auditory-Perceptual Characteristics

伏木田 勝信†
Katsunobu Fushikida

1. まえがき

音声のピッチ検出方式の研究は長い歴史を持ち、これまでにさまざまな方式が提案されてきた[1],[5],[6]。ここでは人間の聴覚・知覚特性に基づいたピッチ検出の方式を提案する。聴覚・知覚の特性としては、聴覚の基底膜における対数的周波数特性、基本周波数が欠落した (Missing fundamental [4]) などの周波数スペクトル歪のある場合の音声の知覚、無声・無音区間におけるピッチの補間的知覚などについて検討を行った。

対数的周波数特性を実現するためには非対称極零対フィルターを用いて任意の周波数に対してパワースペクトル成分を算出する方式を導入し、算出したパワースペクトル値を用いてピッチ検出を行った。

倍音構造の基本周波数が欠落してもピッチ知覚できるいわゆる Missing fundamental に対応するために倍音構造評価 (HSE: Harmonic Structure Evaluation) 関数を導入し、周波数軸上での欠落成分の補償を行なう方式を検討した。

一方、時間軸上の欠落補償としては無声、無音区間における補間的知覚を実現するためにピッチ遷移に制約を課しダイナミックプログラミング (D P) 法を用いて最適なピッチ系列を選択するとともに無声部における補間を行う方式を提案した。この際、フレーム間遷移に制約を設けることで連続性を確保する方式を組み込んだピッチ検出方式についての実験的検討を行った。

2. 聴覚・知覚的ピッチ検出方式

2-1. 非対称極零対フィルタースペクトル

対称な極零対フィルターによる予測残差波形の分析方式については文献 [3] で述べた。ここでは高い Q 値を持つ極 (HQP: High Q Pole) と低い Q 値を持つ零 (LQZ: Low Q Zero) (臨界フィルターでもよい) から構成される非対称な極零対フィルターによる周波数スペクトルを用いた。

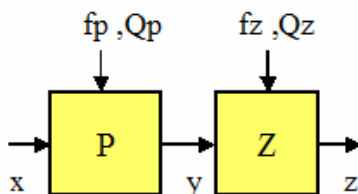


Fig. 1 Pole-Zero pair filtering (P: Pole, Z: Zero)
 $f_p = f_z, Q_p \gg Q_z$

†KF 研究所, KF Laboratory

対数スケールの周波数に対応するスペクトルを算出する方式としては等間隔でスペクトル値が得られるフーリエ変換は必ずしも適した方式ではない。そこで、任意の周波数値に対する成分を算出する方式として高い Q 値を持つ極と小さい Q 値を持つ零とを組み合わせた 極零対フィルターを用いてスペクトルを算出する方式を提案する。 x, y, z を Fig.1 の P, Z の入出力に対応する波形とすると

$$y_n = \alpha_1 x_{n-1} + \alpha_2 x_{n-2} + x_n \quad (1)$$

$$z_n = y_n - \beta_1 y_{n-1} - \beta_2 y_{n-2} \quad (2)$$

ここで、 $\{\alpha_1, \alpha_2\}, \{\beta_1, \beta_2\}$ は、それぞれ、極、零に対応する 2 次の線形予測係数を表す。

極零フィルターの出力パワー P は

$$P = \sum_{m=1}^M z_m^2 \quad (3)$$

ここで M は分析窓長を表す。

なお、1 オクターブあたりの周波数の量子化レベル数は、音楽応用との整合性などを考慮して平均律音階の整数倍 (いわゆる微分音階) とした。また、極・零に対応する 2 次の線形予測係数 (LPC) のテーブルは文献[2],[3]と同様のものを用いた。

高 Q 極・低 Q 零対フィルターの周波数特性を Fig.2 に示す。また、このフィルターにより算出された音声の周波数スペクトル例 (/mami-/の/a/の部分) を Fig. 3 に示す。

本方式によれば分析周波数を任意に設定することができ、フーリエ変換による方式に比べて音声のピッチ検出として重要な、低周波領域で分解能をあげることが容易である。

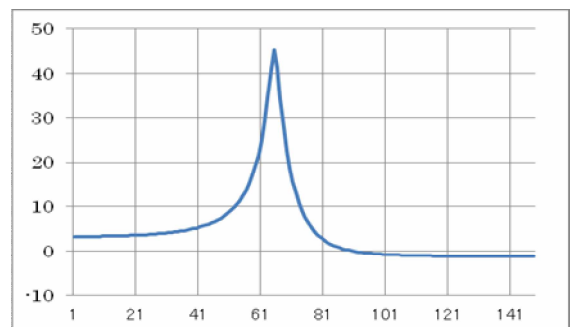


Fig. 2 High Q Pole1(Q=10) - Low Q Zero(Q=1) pair filter spectrum example.

(Vertical: Power (dB), Horizontal: Log scale frequency)

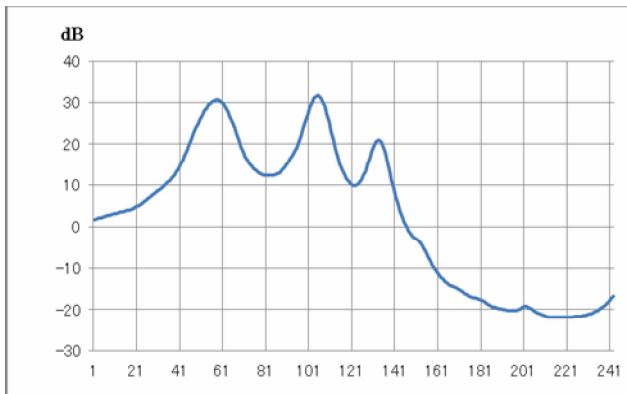


Fig. 3 Frequency spectrum analysis example for /m/ using HQP(Q=10)-LQZ(Q=1) pair filter.
(Vertical: Power, Horizontal: Log frequency 50Hz~1.6kHz.)

2-2. 倍音構造評価関数

失われた基本周波数成分を補うために整数倍の周波数成分における非対称極零フィルターの出力の和を倍音構造評価 (HSE) 関数 H として定義して用いる。(4) 式に H の定義式を示す。

$$H = \sum_{k=1}^K h_k \quad (4)$$

ここで h_k : Output power of Pole-Zero pair filter of kf_0
(f_0 : fundamental pitch frequency candidate)

この評価関数 H を用いれば基本ピッチ周波数成分が欠落している場合でも、高次のピッチ周波数成分により倍音構造の評価を補償して行うことができる。また、 H の値は無声部において小さくなる性質があり、有声・無声識別にも有効に用いることができる。

2-3. ピッチ系列の最適選択方式

知覚的には、隣接した時間フレームにおけるパラメータ値は比較的近く、連続性の高いものが同一系列として選択、知覚されやすいと考えられる。

自己相関値を用い動的計画法 (ダイナミックプログラミング: Dynamic Programming: DP) によって最適なピッチ周期系列を選択する方法は、以前に提案され、量子化ノイズなどに影響されにくくロバストなピッチ検出方式であることが示されている[6]。

ここでは、倍音構造評価関数 H を評価値として用い、連続性の制約のもとで各フレームにおける H の総和が最大になるようにピッチ系列を選択する方式を提案する。この際、 H が閾値以下の場合、無声・無音区間として、一定値に置き換える。これにより無声・無音部のピッチは前後の有声部のピッチを補間した値として得られる。

Fig. 4 に DP によるピッチ系列最適選択方式の評価値・DP パステブルの概念図を示す。 $G(n, i)$ の算出には前フレームの該当部分 (黄色で示す) が用いられる。

DP 法によれば、最適なピッチデータは次の式転化式(5)を解くことにより、最終段 ($n=N$) の最大評価値 ((6) 式) に付随して求められる。

$$G(n, i) = \max_{l_{min} < l < l_{max}} \{G(n-1, i+l)w(l) + H(n, i)\} \quad (5)$$

ここで、 G は評価関数、 l_{min} と l_{max} はピッチ遷移の制約を表す。 $w(l)$ は連続性に対する制約 (重み関数) を表す。

$$G_{max} = \max_{l_{min} < l < l_{max}} \{G(N, I+l)\} \quad (6)$$

ここで I は総分析フレーム数を表す。

Pitch level

I						
..						
.						
I+2			$G(n-1, i+2)$			
i+1			$G(n-1, i+1)$			
i			$G(n-1, i)$	$G(n, i), H(n, i)$		
i-1			$G(n-1, i-1)$			
i-2			$G(n-1, i-2)$			
..						
.						
1						
	1	...	n-1	n	...	N
	frame number					

Fig. 4 Schematic diagram of criterion value and DP-path table (where $l_{min} = -2$, $l_{max} = 2$)

本方式によれば、無声・無音区間におけるピッチの補間を DP アルゴリズムに組み込んだ形で同時に行うことができる。

3. ピッチ検出実験と結果

(実験条件) 音声サンプルとしては男女各 3 名が発声した日本語の連続有声音節連鎖 (/mV/ など)、文章サンプル約 30 秒を用いた。サンプリング周波数は 22.05kHz、フレーム周期は 200 sample points (約 9.1msec)、分析ウィンドー長は 1024 sample points を用いた。 f_0 の変化範囲は 50~400Hz、ピッチの量子化レベル数は 48 levels/octave、倍音構造評価 (HSE) 関数 (H) の倍音数は $K=1 \sim 4$ ((4) 式) を用いた。連続性に対する重みとしては、予備実験の結果をふまえて $w(0) = 1$, $w(\pm 1) = 0.999$, $w(\pm 2) = 0.995$ 近傍の値を用いた。

3-1. 倍音構造評価関数の評価実験

Missing fundamental に対する効果を確かめるため High pass filter (Cut-off 400Hz, 12dB/octave) によりピッチの基本

周波数成分を抑圧した音声についてピッチ検出を行った。まず、あらかじめ、これらの音声を原音と対比較で聴しピッチの知覚的差異はないことを確認した。Fig.5 に HQP-LQZ 対スペクトルの例を示す。

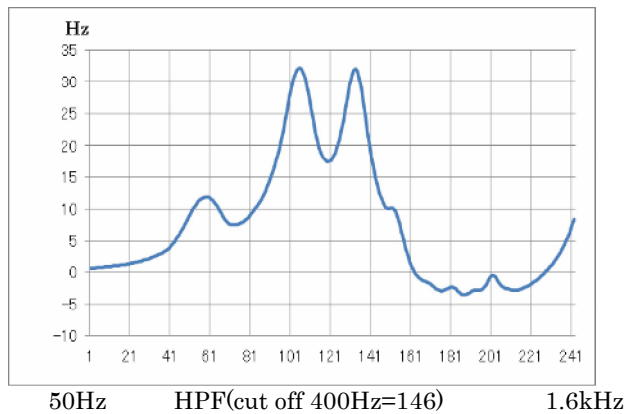


Fig.5 HQP-LQZ frequency spectrum example / (m) a/.
(Vertical: Power (dB), Horizontal: Log frequency.)

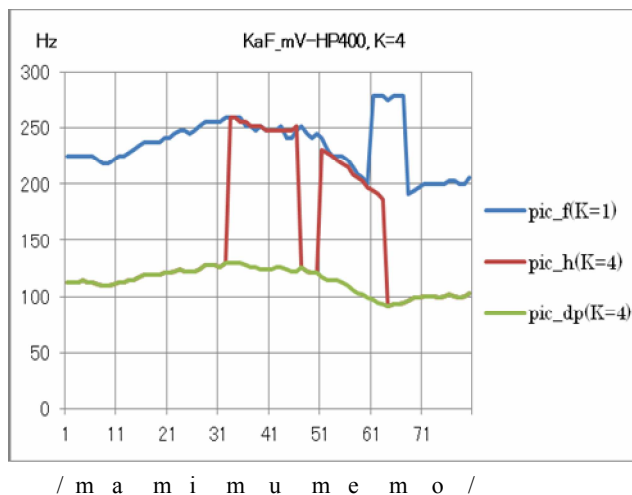


Fig.6 Comparison of 3 pitch detection methods.
(Vertical: Pitch frequency, Horizontal: Frame number)
(pic_f: by K=1, pic_h: by HSV(K=4), pic_dp: DP method)

Fig. 6 に/mamimu'memo/ のピッチ検出実験の結果例を示す。Fig. 6によれば、基本波のみの方式 ($K = 1$, pc_f) に比べて倍音構造評価値 (HSE) 方式 ($K = 4$, pc_h) では倍ピッチへのエラーが減少しており、DP による方式 (pc_dp , $K=4$) ではグロスエラー (Gross error : 1/2 or 2 倍音などへのエラー) はなく、知覚特性と一致した検出結果が得られた。他の音声サンプルについてもほぼ同様の結果が得られており、本方式は、周波数歪みによる知覚的ピッチ検出の誤りを低減するため有効であると考えられる。

3-2. 通常音声サンプルのピッチ検出実験

スペクトル歪のない通常の音声サンプルに対して、①基本波のみの方式 ($K = 1$, pc_f)、②倍音構造評価値 (HSV) 方式 ($K=2,3,4$, pc_h)、③DP の併用による方式 (pc_dp) の3方式についてピッチ検出実験を行った。

Fig. 7 に有声音系列の音声サンプルのピッチ検出結果例を示す。方式②では方式①に比べてグロスエラーは低減している。また、③のDP方式においてはグロスエラーは見られない。

ピッチ検出誤り (グロスエラー) 率は基本周波数のみで行った場合①に対して、倍音構造評価値を用いた場合②で減少しており、DP法を併用した場合③では、今回用いた音声サンプルにおいては、グロスエラーは見られなかった。Fig.8 に無声音を含む音声サンプルのピッチ検出結果例を示す。無声部・無音部においてもピッチは前後の有声音部に対して連続的にめらかに補間されていることが分かる。他の音声サンプルにおいてもほぼ同様の結果が得られている。

これらの実験結果は、倍音構造評価値とDP方式を併用した方式によればグロスエラーがなくロバストな知覚的ピッチ検出を行うことができることを示している。[7]

4. あとがき

人間のピッチ周波数の知覚特性を反映した方式として高いQ値を持つ極と低いQ値を持つ零とを組み合わせた極零対フィルターを用いる分析方式、倍音構造評価値とDPアルゴリズムによる連続性制約、補間を組み込んだピッチ検出方式を提案した。ピッチ検出実験を行った結果、提案方式にグロスエラーは見られず、ロバストな知覚的ピッチ検出方式であることがわかった。

参考文献

- [1] L. R. Rabiner, M. J. Cheng, A. E. Rosenberg and C. A. McGonegal, "A Comparative Performance Study of Several Pitch Detection Algorithms," IEEE Trans. on Acoustics, Speech, and Signal Processing, vol. 24, no. 5, pp. 399-418, 1976.
- [2] 伏木田「自己相関領域で逆フィルタリングを用いたホルモントの多段推定方式」, 日本音響学会, 音声研究会資料, S81-41, pp. 323-329, 1981-10.
- [3] 伏木田「音声の全極型予測残差の極零対フィルターによる分析」, 情報処理学会第72回全国大会, 2D-5, 2010-3,
- [4] Chao He and Laurel J. Trainor, "Finding the Pitch of the Missing Fundamental in Infants", The Journal of Neuroscience, June 17, 2009, 29(24):7718-8822.
- [5] M.Slaney et al., "A Perceptual Pitch Detector," ICASSP-90, S6b.3, 1990-3.
- [6] 伏木田「ピッチ周期系列最適選択の一方式」 日本音響学会講演論文集 (昭和 52 年 10 月), 1-1-18, pp. 35-36, 1977.
- [7] <http://www.5b.biglobe.ne.jp/~hfytkf/Laboratory/Laboratory.htm>

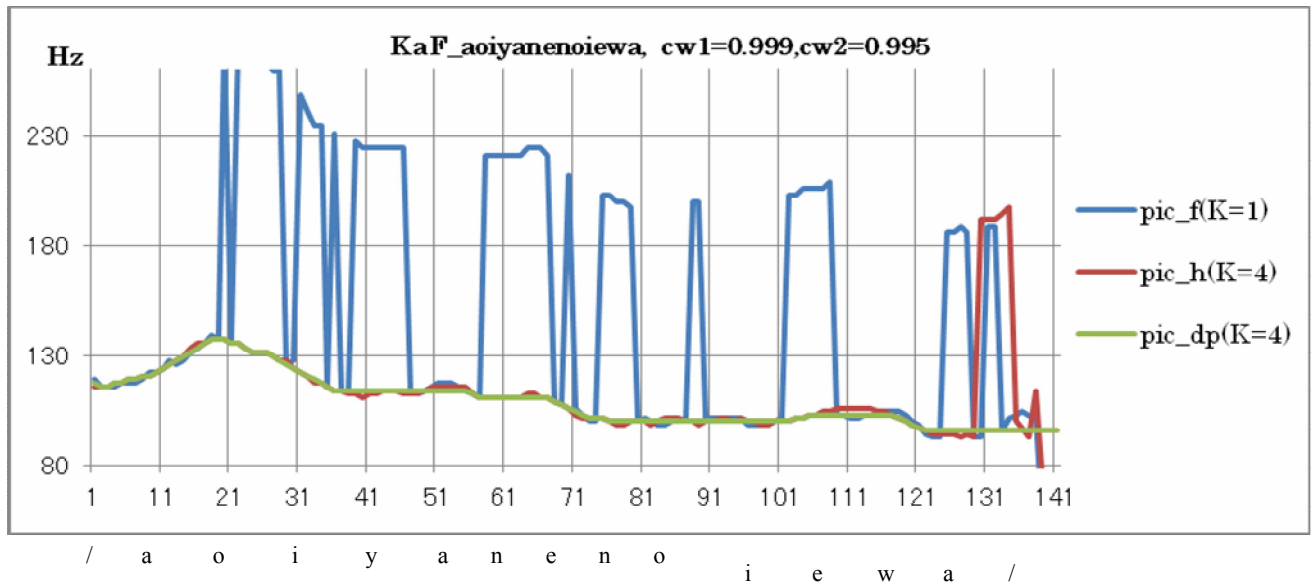


Fig. 7 Pitch detection result (pic_f: by fundamental frequency,
pic_h: harmonic structure, pic_dp: DP method)

Horizontal axis shows frame number.
Speech sample : /aoiyanenoiwa/

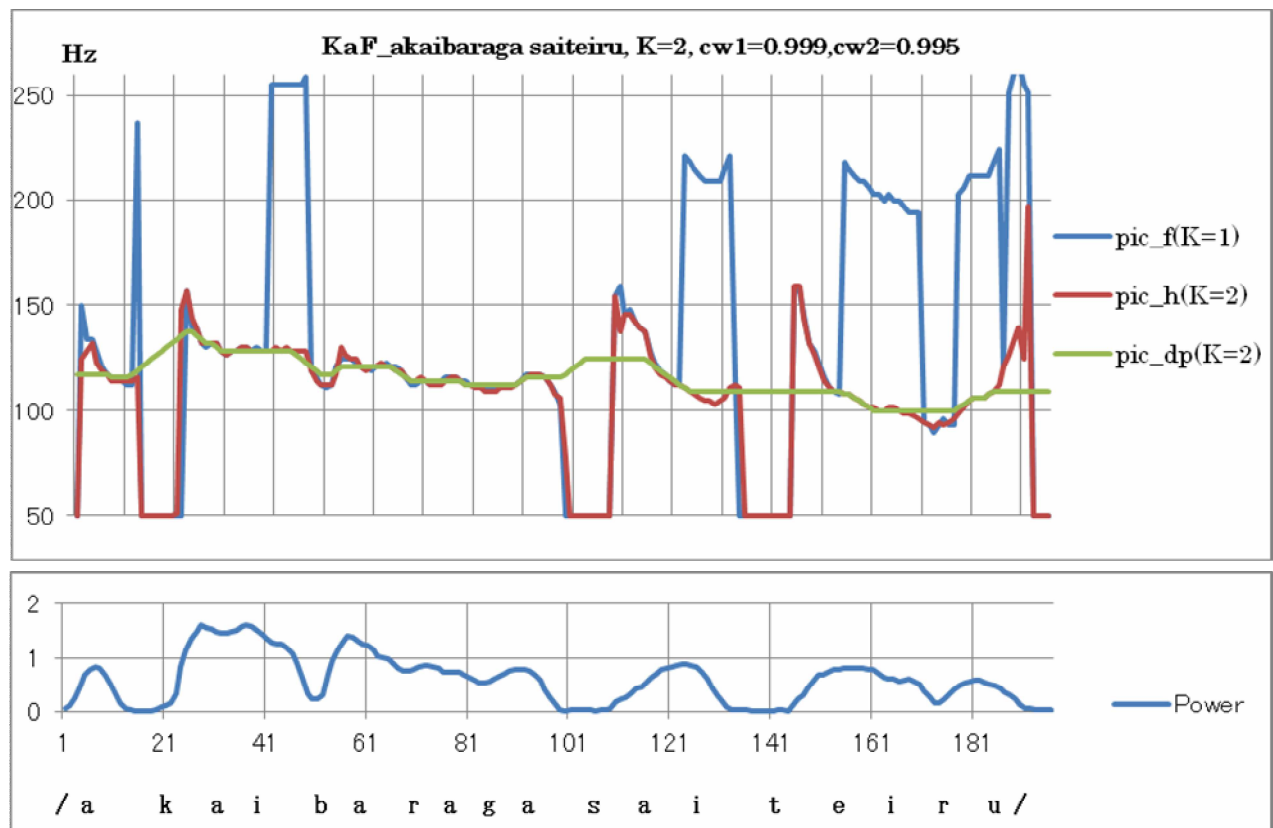


Fig. 8 Pitch detection result (containing unvoiced segments)
Lower figure shows averaged power pattern.

Horizontal axis shows frame number.
Sentence speech sample: /akaibaraga saiteiru/