

研究速報

平方根及び4乗根パワースペクトルの自己相関に基づくピッチ抽出

吉尾 重治[†] 趙 奇方^{††}(正員)
 島村 徹也^{†††a)}(正員) 鈴木 誠史^{†††}(正員)

Pitch Detection Based on Autocorrelation of Root and Fourth-Root Power Spectra

Shigeharu YOSHIO[†], Nonmember, Qifang ZHAO^{††},
 Tetsuya SHIMAMURA^{†††a)}, and
 Jouji SUZUKI^{†††}, Regular Members

[†] (有)ヒュー, 東京都

Hue Co., Ltd., Tokyo, 164-0013 Japan

^{††} (株)NTTデータ情報科学研究所, 東京都

Laboratory for Information Technology, NTT DATA Co.,
 Ltd., Tokyo, 104-0033 Japan

^{†††} 埼玉大学工学部情報システム工学科, 浦和市

Faculty of Engineering, Saitama University, Urawa-shi,
 338-8570 Japan

^{††††} 日本工業大学工学部情報工学科, 埼玉県

Faculty of Engineering, Nippon Institute of Technology,
 Saitama-ken, 345-8501 Japan

a) E-mail: shima@sie.ics.saitama-u.ac.jp

あらまし 本論文では, 音声信号のピッチ周期/基本周波数の新しい推定法を提案する. スペクトルを波形とみなし, その周期を求めることにより基本周波数を算出する立場をとる. ケプストラム法及びその拡張法である ACLOS 法がパワースペクトルの対数値を基本とするのに対し, 提案法はパワースペクトルの平方根及び4乗根を基本とする. 実験結果は, 提案する方法が優れた耐雑音性を有することを示す.

キーワード 基本周波数, ケプストラム, 自己相関関数, 平方根パワースペクトル, 4乗根パワースペクトル

1. まえがき

ピッチ抽出は多くの音声処理システムにおける基本処理であり, その重要性は古くから認識されている. しかし, 現実問題としてのピッチ抽出処理は, 音声信号が本来有する非定常性を原因として, 困難を来す場合が多い. 数多くのピッチ抽出アルゴリズム (PDA) が今までに提案されているが [4], [5], 付加雑音にロバストに高精度な抽出結果を与える確約されるべき方法はいまだ存在しないと考えられる.

ケプストラム法 [1] と自己相関関数法 [2] は, 現在広く用いられている PDA である. ケプストラム法は, ホルマントの影響を受けにくいという長所を有する反面, 雑音に弱く, 信号対雑音比 (S/N) の劣化とともに, 抽出誤りが急激に増大するという短所がある. 一

方, 自己相関関数法は比較的雑音に対してロバストであるが, ホルマントの影響を受けるといった性質を持ち併せる. 文献 [5] においては, 無雑音の場合にはケプストラム法が最も優れた結果を与え, 雑音が付加されると自己相関関数法が優位であるとの報告がなされている.

一方, 最近, ケプストラム法と自己相関関数法の長所を組み合わせた PDA, 対数スペクトル自己相関関数法 (AutoCorrelation function of LOf Spectrum: ACLOS) が提案された [6]. ACLOS は, 音声信号の対数スペクトルを波形とみなし, その自己相関関数を算出することによりピッチ周期を求める手法であり, 雑音にロバストな実行結果を与える. また, 長い基本周期を有する男声に有利であるケプストラム法及び自己相関関数法と異なり, ACLOS は基本周波数をスペクトル上で求めるために, 高い基本周波数を有する女声に有利である. 本論文では, ACLOS のスペクトルの自己相関関数を計算する原理を保持しつつ, そのピッチ抽出精度を更に改善する PDA を提案する.

対数スペクトルは, 無雑音の状況下においては音声信号がもつ調波構造を明確に表現する. この性質により, ケプストラム法は高精度なピッチ抽出結果を与えられと考えられる. しかし, 周知のように, 対数スペクトルは付加雑音に敏感である. したがって, その対数スペクトルを, より雑音にロバストなスペクトル表現に置き換えることにより, ピッチ抽出精度は向上されと考えられる. そこで本論文では, パワースペクトルの平方根及び4乗根, すなわち平方根パワースペクトル及び4乗根パワースペクトルを取り上げることにする. そして, ケプストラム法でのフーリエ変換または逆フーリエ変換によるケフレンシーへの変換操作の代わりに, ACLOS での自己相関関数処理を用い, 雑音にロバストなピッチ抽出を図る.

2. 提案法の原理

音声に雑音加わった場合, 対数スペクトルはその谷の部分から周期性が損なわれてしまう. これは, 対数が大きい振幅値を急激に圧縮してしまう性質を有するためである. 結果として, 小さい振幅値を有する雑音の影響は相対的に増大されてしまう. そこで, 対数の代わりに平方根または4乗根を用いることを本論文では考える. 図1にパワースペクトル, 対数スペクトル, 平方根スペクトル, 4乗根スペクトルのそれぞれの波形の違いを示す. この図を見るとわかるとおり, パワースペクトルの平方根や4乗根をとると, 高周

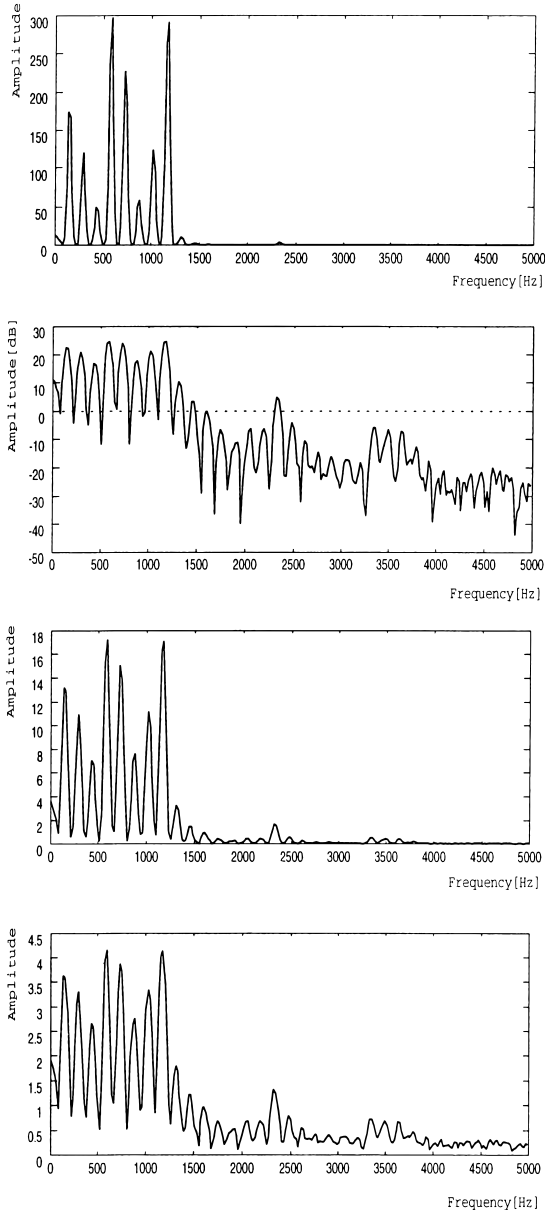


図1 スペクトルの波形の違い：男声 /a/，上：パワースペクトル，中上：対数パワースペクトル，中下：平方根パワースペクトル，下：4乗根パワースペクトル
 Fig. 1 Difference of spectral shapes: male voice /a/, upper: power spectrum, central upper: log power spectrum, central lower: root power spectrum, lower: fourth-root power spectrum.

波数域の振幅が強調され，調波構造がはっきりする。この性質を利用して，本論文では，耐雑音性の向上を図る目的でパワースペクトルの平方根及び4乗根を

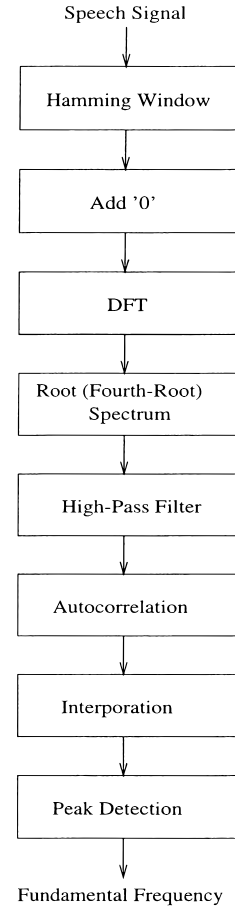


図2 提案法の流れ図
 Fig. 2 Flowchart of the proposed method.

ACLOSに組み入れる方法を提案する。

図2に提案するPDAの処理の流れを示す。まず，標本化された音声信号を窓関数で切り出す。分析精度を高めるために0を付加して離散フーリエ変換(DFT)を行う。この後，ACLOSでは対数スペクトルを求めるが，ここでは平方根または4乗根スペクトルを求める。しかし，これらのスペクトルは対数スペクトルと同様ホルマントの影響を受けるため，スペクトルを波形とみなして高域通過フィルタ処理を行い，声帯音源の周期信号成分を強調する。そして，自己相関関数を計算し，内挿処理により精度を高めつつ，ピーク位置を検出することにより基本周波数を求める。

提案法における平方根または4乗根スペクトルへの高域通過フィルタ処理は，ACLOSにおける対数スペクトルへの高域通過フィルタ処理(これがリフタ処理

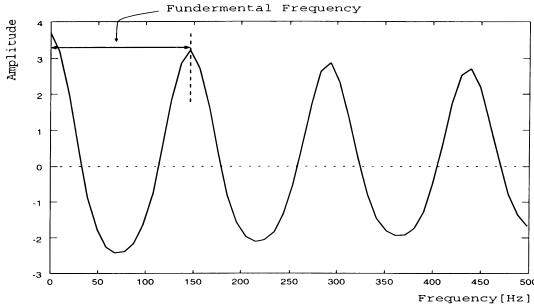


図3 高域通過フィルタリング後の4乗根パワースペクトルの自己相関波形: 男声 /a/

Fig. 3 Autocorrelation of fourth-root power spectrum after high-pass filtering: male voice /a/.

に対応する)と厳密には異なる性質を有する。しかし、スペクトル包絡の形成に支配的である低周波数成分を除去することにより、基本的にはスペクトル波形が有する周期成分を強調することができる。現に、図3はその正当性を示している。図3は4乗根パワースペクトルの高域通過フィルタ処理後の自己相関波形である。ここには、本来有すべき基本周波数に対応する明瞭な周期信号が観測される。

提案法における基本周波数は、図3に示されるように、スペクトルの自己相関波形の極大値から求められる。基本周波数のこの抽出方法は、ACLOSと同様、内挿に適している。また、自己相関関数は、ランダム成分をラグ0付近に集中させる性質があるため、スペクトル上の周期性は、その自己相関関数を求めることにより強調される。これが、結局、付加雑音に対するロバスト性を与えることになる。

3. 実験結果

本章では、提案法の有効性を検証するために、自己相関関数法、ケプストラム法、ACLOS法との抽出精度の比較を行う。使用した音声資料は、NTTアドバンステクノロジ(株)の「20か国語音声データベース」に収録されている男女各1名が発声した約10秒の短文である。発話内容は、

“彼は以前から、科学技術の進歩と人間の勇気が、はるかな宇宙への旅を可能にしたのだと考えていました。”

である。これらは、3.4kHzに帯域制限し、10kHzの標準化周波数で標準化して用いられる。

雑音下での実効性能を調べるために、上記の音声データに S/N がそれぞれ 10dB, 5dB, 0dB とな

るように白色雑音を加えた。これらの雑音データも、音声データ同様、3.4kHzに帯域制限し、10kHzの標準化周波数で標準化して用いられている。このとき、 S/N は

$$SNR = 10 \log_{10}(P_s/P_n) \quad (1)$$

のように計算される。ここで、 P_s 及び P_n は、それぞれ音声信号と雑音の平均電力を表す。

基準とする基本周波数は、音声の時間波形の視察から求めた基本周期の逆数を用いた。ただし、音声データの有声音区間は、分析フレーム内の最大振幅が信号全体の最大振幅の -30dB 以上のときとして、それ以外は無声音区間として扱った。

得られた結果の評価には、Rabinerら[3]の gross pitch error と fine pitch error を用いた。すなわち、基準となる基本周波数の逆数である基本周期を \hat{T}_i (i はフレーム周期の番号)、各抽出法により求めた基本周期を T_i とすると、抽出誤差 e_i は

$$e_i = T_i - \hat{T}_i \quad (2)$$

で与えられる。このとき、 $|e_i| > 1$ ms であれば Gross Pitch Error (GPE)、 $|e_i| \leq 1$ ms であれば Fine Pitch Error (FPE) とする。そして、GPE の場合はその割合をパーセントで求め、FPE の場合は平均誤差と標準偏差を算出する。

本実験における諸元は次のとおりである。窓関数は 51.2ms のハミング窓、分析フレーム周期は 10ms、DFT のポイント数は 1024、内挿はピーク位置の前後3点からなるラグランジェ補間である。従来法としては、自己相関関数法、ケプストラム法、ACLOS法を用いた。いずれの方法も、提案法と分析精度が対等になるように、3点のラグランジェ補間を行った。また、DFTは1024ポイントで計算された。しかし、ACLOS法においては、文献[6]とは多少異なり、クリッピングと帯域制限の処理を施していない。これは、提案法における平方根及び4乗根スペクトルと、ACLOSにおける対数スペクトルとの比較を明確にするためである。

表1から表3に本実験で得られた結果を示す。

まず、表1より、ACLOS法がケプストラム法より耐雑音性に優れていることがわかる。本論文で提案する平方根及び4乗根スペクトルを用いる方法は、更に改善を与えている。また、従来、ランダム性の雑音に強いとされる自己相関関数法に対しては、男声の場合には提案法は劣るが、女声の場合には明らかに

表 1 GPE の割合 (%) (上段 : 男声 , 下段 : 女声)
Table 1 Rate of GPE (%) (Upper: male speech, Lower: female speech)

SN 比	自己相関関数法	ケブストラム法	ACLOS 法	提案法 (平方根)	提案法 (4 乗根)
∞	4.7	4.2	4.4	4.9	4.7
	7.0	4.9	0.2	0.0	0.0
10 dB	5.2	11.0	9.1	7.2	7.0
	11.0	14.4	9.4	0.4	0.9
5 dB	5.9	26.7	17.8	13.3	11.7
	15.3	22.0	9.4	1.3	2.3
0 dB	7.9	41.8	31.5	25.0	24.6
	26.5	35.6	18.6	5.6	7.4

表 2 FPE の平均値 (ms) (上段 : 男声 , 下段 : 女声)
Table 2 Averages of FPE(ms), (Upper: male speech, Lower: female speech)

SN 比	自己相関関数法	ケブストラム法	ACLOS 法	提案法 (平方根)	提案法 (4 乗根)
∞	-0.001	0.020	-0.093	-0.061	-0.068
	0.044	-0.003	0.019	0.024	0.021
10 dB	-0.012	-0.025	-0.078	-0.061	0.078
	-0.046	-0.007	0.040	0.034	0.030
5 dB	-0.015	-0.030	-0.089	-0.071	-0.081
	-0.043	-0.019	0.066	0.041	0.043
0 dB	-0.013	-0.040	0.086	-0.079	-0.089
	0.048	0.009	0.085	0.049	0.049

表 3 FPE の標準偏差 (ms) (上段 : 男声 , 下段 : 女声)
Table 3 Standard deviations of FPE(ms), (Upper: male speech, Lower: female speech)

SN 比	自己相関関数法	ケブストラム法	ACLOS 法	提案法 (平方根)	提案法 (4 乗根)
∞	0.196	0.120	0.197	0.225	0.209
	0.073	0.005	0.085	0.095	0.090
10 dB	0.201	0.173	0.254	0.248	0.233
	0.075	0.141	0.168	0.133	0.124
5 dB	0.213	0.205	0.302	0.278	0.267
	0.075	0.181	0.212	0.160	0.164
0 dB	0.227	0.240	0.391	0.313	0.305
	0.087	0.262	0.271	0.185	0.189

勝っている。男声及び女声の結果の平均値を求めるとすると、提案法は自己相関関数法よりも好結果を与えていることがわかる。したがって、提案法は雑音にロバストな方法であるといえよう。

提案法が男声より女声の場合に良好な結果を与える理由は、次のように考えられよう。提案法を用いれば、周波数軸上での自己相関関数ピークから基本周波数が抽出されるために、男声と女声で周波数軸上に同等の誤差が生じたとしても、その逆数の基本周期で評価される本実験においては、高い基本周波数を有する女声の方が本質的に精度が高くなると考えられる。これに加えて、ランダム性の雑音が付加された場合には、その自己相関関数は周波数軸で 0 周波数に対応するラグ 0 付近で支配的となるために、0 周波数からより離れた位置に基本周波数を有する女声の方が、男声より雑音の影響を受けなくてすむと思われる。

次に表 2 であるが、ACLOS 法自体が、自己相関関数法、ケブストラム法に比べ、より大きな FPE を与えている。提案法により、改善はされるものの、自己相関関数法とケブストラム法を上回るには至っていない。表 3 においても、女声の場合において、提案法がケブストラム法よりも好結果を与えているが、これは上記の提案法の女声に対する優位性の一結果と思われる。全体としては表 2 と同様の結果を与えている。したがって、基本周波数の抽出の精度に関しては、提案法には改善の余地が残されているといえよう。

二つの提案法の比較に関しては、表 1 より、男声においては 4 乗根スペクトルの方が、女声においては平方根スペクトルの方が雑音にロバストであるといえよう。これは、男声とは異なり女声のスペクトルにおいては、基本周波数の調波構造に対応するスペクトルピークの間にはしばしば小さなピークが形成されること

に起因していると考えられる。スペクトルを圧縮しすぎると、その小さなスペクトルピークは、雑音に乱されるときに基本周波数抽出に悪影響を及ぼす可能性がある。と、容易に推測される。しかし、表 2 の FPE の平均値に関しては平方根スペクトルが高精度である。一方で、表 3 の標準偏差では明らかに 4 乗根スペクトルが優位である。したがって、二つの提案法の優劣を論じるには、更なる実験的検討が必要と思われる。

4. むすび

本論文では、先に提案した ACLOS 法をもとに、パワースペクトルに対して対数ではなく平方根または 4 乗根を施す PDA を提案した。そして、自己相関関数法、ケプストラム法、ACLOS 法と実効性能を比較することにより、耐雑音特性において優位性を示すことができた。また、特に女声の場合に、提案法は好結果を与えることがわかった。今後は、より多くの音声資料をもとに、更に抽出特性を分析する予定である。また、白色雑音以外の雑音や妨害波についても検討する予定である。

文 献

- [1] A.M. Noll, "Cepstrum pitch determination," J. Acoust. Soc. Am., vol.41, pp.293-309, 1967.
- [2] L.R. Rabiner, "On the use of autocorrelation analysis for pitch detection," IEEE Trans. Acoust., Speech & Signal Process., vol.ASSP-25, no.1, pp.24-33, Feb. 1977.
- [3] L.R. Rabiner, M.J. Cheng, A.E. Rosenberg, and A. McGonegal, "A comparative study of several pitch detection algorithms," IEEE Trans. Acoust., Speech & Signal Process., vol.ASSP-24, no.5, pp.399-413, Oct. 1976.
- [4] W.J. Hess, Pitch Determination of Speech Signals — Algorithms and Devices, Springer-Verlag, Berlin, 1983.
- [5] W.J. Hess, "Pitch and voicing determination," in Advances in Speech Signal Processing, ed. S. Furui and M.M. Sondhi, pp.3-48, Marcel Dekker, New York, 1992.
- [6] 國枝伸行, 島村徹也, 鈴木誠史, "対数スペクトルの自己相関関数を利用したピッチ抽出法," 信学論 (A), vol.J80-A, no.3, pp.435-443, March 1997.

(平成 12 年 8 月 8 日受付, 10 月 20 日再受付)