

音声のための基本周波数推定法の検討

Study on Fundamental Frequency Estimation Method for Speech Signal

桐生 恵佑¹⁾

指導教員 三輪 賢一郎

1) サレジオ工業高等専門学校 機械電子工学科 情報コミュニケーション研究室

キーワード：基本周波数, 音声, 非同期検波

1. 緒言

基本周波数 (F0) とは音声や楽器といった様々な音源での信号処理で利用される重要な特徴量である。実環境で F0 推定を正確に行うことができれば、対象音に対する雑音除去や音源分離などを効果的に実現できることが期待される。

現在までに数多くの基本周波数推定法が提案されてきており、静音環境や雑音環境における F0 推定法はほぼ確立されているが、残響環境や雑音残響環境に適応できるものはほとんど存在しないとあってよい。

雑音・残響環境に一定の耐性を持つとされる手法に FreeDAM がある[1]。FreeDAM はミッシングファンダメンタル現象を計算機上で模擬したもので、隣り合う 3 本の調波の組を AM 信号に見立てて、それを AM ラジオのように復調することで F0 を特定するという手法である。しかし、FreeDAM は復調方式に同期検波を用いているため、計算量が多いという欠点があった。そこで本研究室では、復調方式を非同期検波に変更することで計算量を大幅に低減した FreeDAM Lite を提案している[2]。FreeDAM Lite は、対象信号を理想的な調波複合音とした場合には雑音や残響に対して一定の頑健性を持つことが確認されているが、音声に対応できるかどうかは明らかではない。そこで本稿では、FreeDAM Lite が音声信号に対して基本周波数推定が可能であるか、シミュレーションにより検証する。

2. 研究方法

MATLAB 上で FreeDAM Lite を実装し、音声信号を入力したときの F0 推定結果をシミュレーションにより検証した。音声信号は、東北大学電気通信研究所で制作されたデータベース[3]中の短母音 /a//i//u//e//o/ について、それぞれ男声と女声の計 10 種類を用いることとし、各々 0.1 秒間を抽出した上で分析した。図 1 に検証に用いた音声信号の F0 軌跡の一例を示しており、赤枠は抽出した 0.1 秒間を示す。なお、同データベースには F0 の情報は無いことから、従来法である TEMPO2[4]で別途計測したものを真値とした。評価指標は、真値との誤差率[%]を用いた。今回実行した処理のフローを図 2 に示す。

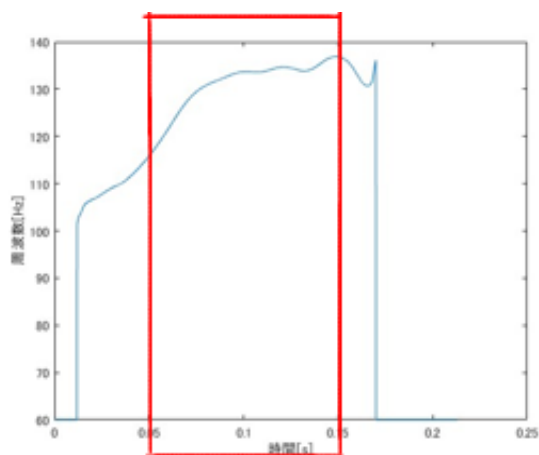


図1 音声信号/o/ (男声) の F0 軌跡

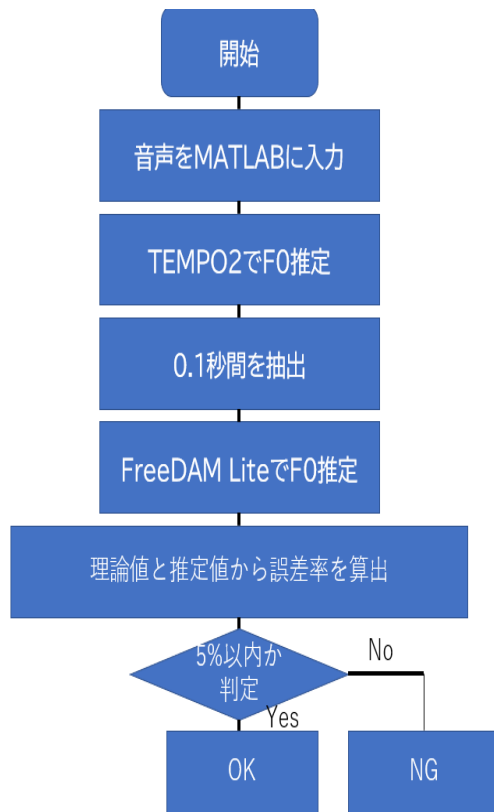


図2 本研究のフローチャート

3. 結果

実験結果の誤差率を表1に示す。/a/以外は、男声の方が誤差が小さい結果となった。また、女声の/i/以外はいずれも5 [%]以内に収まっており、音声に対して比較的正確なF0推定が行えたことが確認できた。

表1 音声信号のF0推定結果（誤差率 [%]）

	誤差率[%]				
	/a/	/i/	/u/	/e/	/o/
男声	1.9	0.2	1.4	0.9	0.4
女声	0.3	7.0	4.8	3.2	3.7

4. 結言

本稿では、FreeDAM Liteが音声信号に対して基本周波数推定が可能かどうかをシミュレーションにより検証した。結果から、比較的正確なF0推定が行えたことを確認できた。

5. 今後の予定

今後は、連母音等を対象に、基本周波数推定が可能かどうかを検証してく予定である。

文献

- [1] 三輪賢一郎, 鷗木祐史, "振幅変調音のピッチ知覚に基づいた調波複合音の基本周波数推定法," 電子情報通信学会論文誌A, Vol. J98-A, No.12, pp. 668--679, 2015年12月
- [2] 山川拓真, 三輪賢一郎, "雑音あるいは残響に頑健な基本周波数推定法の研究," 音学シンポジウム2023, 信学技報, vol. 123, no. 88, SP2023-14, pp. 68--72, 2023年6月
- [3] 天野成昭, 近藤公久, 坂本修一, 鈴木陽一, "親密度別別単語理解度試験用音声データセット(FW03)," NII 音声資源コンソーシアム, 2006
- [4] H. Kawahara, I. Masuda-Katsuse, and A. Cheveigné, "Restructuring speech representations using a pitch-adaptive time-frequency smoothing and an instantaneous-frequency-based F0 extraction: Possible role of a repetitive structure in sounds," Speech Commun., vol. 27, pp. 187-207, 1999.