

高ピッチ音声のフォルマント周波数の抽出に関する検討*

○ 粕谷英樹, 太田香苗, 森大毅 (宇都宮大・工) 出口利定, 大山玄 (東京学芸大・教育)

1 はじめに

音源パラメータである基本周波数（以下ピッチ周波数と略す）と声道パラメータであるフォルマント周波数は音声の基本的な物理量である。特に幅広い年齢層にわたる母音のピッチ周波数やフォルマント周波数の変化の傾向は、日本語音声の基本的な資料として教科書などいろいろな場面で利用されている。

フォルマント周波数の測定には、古くはサウンドスペクトログラムのセクションパターンから視察によって求める方法[1], AbS (Analysis-by-Synthesis) 法[2], 線形予測分析法[3]などがあるが、最近ではデジタル化された音声については、測定の容易さから線形予測分析法が用いられることが多い。しかし、線形予測分析法には後述するようにいろいろな問題があることが知られている。

われわれはシステム同定で広く使われている ARX (Autoregressive with exogenous input) モデルによって音声生成過程をモデル化し、AbS 法によって音源・声道パラメータを推定する方法を提案した[4-6]。われわれの最新の方法では、分析窓内の声道パラメータの定常性と音源波形の時変性を仮定することによって、ピッチ周波数の高い音声でも線形予測分析法より精度が良くしかも安定にフォルマントが推定できることを示した[6]。

本報告では、フォルマント推定における ARX 分析法の利点と推定精度を上げるために検討したことについて述べる。

2 線形予測分析法と ARX 分析法

線形予測分析法と ARX 音声分析法の背景にある音声生成モデルを Fig. 1 に示す。また両者の特徴をまとめて Table 1 に示す。これらの図表から明らかなように、線形予測分析の音源は**単一インパルス**（有声音）を仮定するのに対し、ARX 分析では**微分声門体積流パルス列**（有声音源）を仮定する。線形予測分析（自己相関法）では、このような単純な音源モデルのために、高ピッチ音声では優勢な高調波成分にフォルマントが引き寄せられ

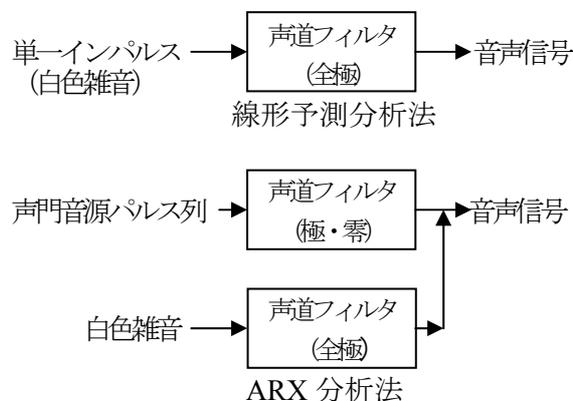


Fig.1 線形予測分析法と ARX 分析法の背景にある音声生成モデル

Table 1 Comparison between LPC and ARX model

事 項	線形予測分析	ARX 分析
有声音源	単一インパルス	微分声門音源パルス列
無声音源	白色雑音	白色雑音
声道フィルタ	全極型 (有声・無声)	極・零型 (有声) 全極 (無声)
誤差評価	相関関数 パワースペクトル	位相を含む音声波形
解法	連立方程式	AbS (非線形最適化)

(単一パルスモデルのため)、更に微分声門音源波のピークが存在する 700 Hz 以下にある第 1 フォルマント周波数の推定誤差が大きくなる（インパルスモデルのため）。一方 ARX 分析では、より現実に近い音源モデルを採用しているためこのようなことは起こりにくい。

Fig. 2 は、直線的に変化する第 1 フォルマント周波数(250→1,000 Hz), 第 2 フォルマント周波数(3,000→1,500 Hz)と、一定の第 3 フォルマント周波数(3,200 Hz), 第 4 フォルマント周波数(4,500Hz)から合成した擬似音声信号(A)に対する線形予測分析(B)及び ARX 分析(C)によって推定したフォルマント周波数軌跡を示す[6,7]。ピッチ周波数は 447 Hz である。線形予測分析によると(B), 第 3 フォルマントは第 7 高調波に、第 4 フォルマントは第

*Considerations on measurement of formant frequencies of high-pitched speech, by KASUYA, Hideki, OHTA, Kanae, MORI, Hiroki (Utsunomiya University), DEGUCHI, Toshisada and OHYAMA, Gen (Tokyo Gakugei University).

10 高調波に引き寄せられている。また、第2フォルマント周波数は階段状に高調波をたどりながら下降している。逆に上昇する第1フォルマント周波数も似た傾向を示すとともに、第1フォルマント周波数が基本周波数(447 Hz)以下になることはない。それに対して ARX 分析では、比較的良好に擬似音声のフォルマント周波数を追跡している。

3 ARX 分析による第1フォルマント周波数の高精度推定法

ARX 分析における音源・声道パラメータの推定は、逐次的な声門閉鎖時刻(glottal closure instant, GCI)の最適化によって行われる[6,7]。音源パラメータの中では声門開放率(Open Quotient, OQ)が特に第1フォルマント周波数の推定精度に影響を与える。Fig.3 はいろいろな OQ(図では r_0)に対する音源スペクトル形状である。/a/の第1フォルマントは OQ の影響を受けにくい。子供のような高ピッチ音声进行分析する場合は[8]、いろいろな OQ の初期値(50, 60, 70, 80%)によって/a/を分析し、文献[6,7]で述べられている方法で自動的に推定した OQ の値の平均値を OQ の初期値として与えて、すべての母音进行分析する方法が効果的である。文献[8]のような比較的单純な発話では、OQ は母音にはほとんどよらない。

4 おわりに

高ピッチ音声のフォルマント推定において、線形予測分析法より ARX 分析法が優れている理由について述べ、ARX 分析法で第1フォルマントを精度良く推定するための声門開放率の扱いについて述べた。実際の高ピッチ音声に適用した例は[8]に示されている。

謝辞

有益なコメントを頂いた大塚貴弘氏(三菱電機(株))に厚くお礼申し上げます。

参考文献

- [1] 粕谷, 他, 音響誌, 24 巻, 355-364, 1968.
- [2] Bell, C., et al., JASA, 33, 1725-1736, 1961.
- [3] Markel, J. & Gray, A., Jr., *Linear Prediction of Speech*, Springer-Ver., 1976.
- [4] Ding, W. & Kasuya, H., IEICE Trans. Inf & Syst., 78, 738-743, 1995.
- [5] Zhu, W. & Kasuya, H., JASJ(E), 19, 223-230, 1998.
- [6] 大塚, 粕谷, 音響誌, 58 巻, 386-397, 2002.
- [7] 大塚, 宇都宮大学博士論文, 2002-3.
- [8] 粕谷, 他, 音講論(春), 2-4-12, 2006.

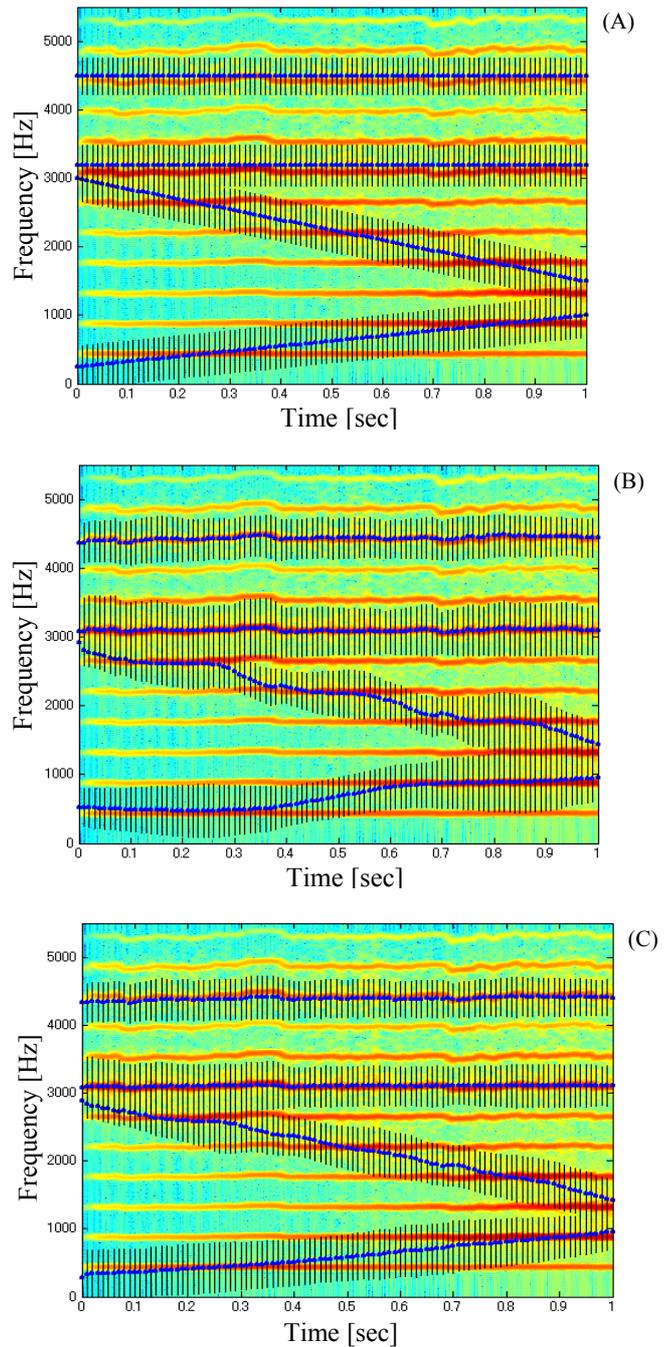


Fig. 2 推定したフォルマント周波数軌跡
(a) 合成音声(試験音)のスペクトル図
(b) 線形予測分析法による分析結果
(c) ARX 分析法による分析結果

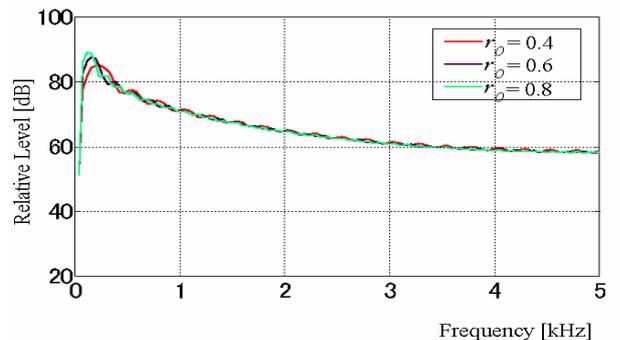


Fig. 3 OQ に対する音源波のスペクトル形状