F-1

音声に含まれる非周期成分のモデル化と評価
Modeling and evaluation of aperiodic components included in speech

大塚 貴弘 † **青山 文彦** ‡ **粕谷 英樹** ‡
Ohtsuka Takahiro Fumihiko Aoyama Kasuya Hideki

1. はじめに

気息的な声質は、女性の音声や特定の音素環境（たとえば先行の音素が無声子音）の母音に現れる音源に関する現象で、雑音的な特性を持つ。有声摩擦音や母音間にある /h/ も周期成分と非周期成分の性質を同時に示す。本論文では、この音源信号の非周期的な成分を、数式モデルで表される音源パルスの群遅延特性を高周波数領域でランダム化し模擬する。周期性が弱い（非周期的）周波数領域をフレーム毎に検出し、その領域の音源パルスの群遅延をランダム化する。われわれは、この非周期成分をテキスト音声合成システムの中で制御することでより表現力豊かな音声の合成法の確立を目指している。

2. 非周期成分のモデル化

非周期的な成分は、音声のソース・フィルタモデルにおいて音源信号 (ソース) に含まれ、その非周期的成分は基本周波数 f_0 の揺らぎ (ジッター) および緩やかな変化 (f_0 トレンド)、振幅 a_v の揺らぎ (シマー) および緩やかな変化 (a_v トレンド)、加法的な雑音 (ノイズ)、で構成されているとする。本論文では、ノイズと波形揺らぎによる非周期的成分をモデル化することを考える。

音源モデルでは、この非周期的な成分を周波数領域上で表しモデル化する。この非周期的な成分は高い周波数領域で顕著に現れ、その成分が増えるに従って段々に低い周波数領域に現れる傾向がある。そこで、この非周期的な領域の下限を与える周波数 (f_a) を設け、その非周期的成分の程度を表すパラメータとする。

3. 分析方法

非周期的周波数領域の推定は、BPF(Band-pass filter) された音声信号の修正自己相関関数 (MAF) による相関係数によって検出する方法 [1] をもとに改良を加えた。MAF による相関係数の度合いは音声信号の非周期性と関係しているが、連続音声の場合、音声の信号の時変性によって大きく影響を受ける。主に、声道の変化とピッチ周期の変化および音源振幅の変化を考慮する必要がある。

(i) 声道変化の補償: ARX 分析法 [2] を用いてフレーム毎に推定した声道パラメータをポイント毎に補間し、時変逆フィルタを構成する。声道の変化による MAF への影響をこの時変逆フィルタによって音声信号から取り除く。

(ii) ピッチ周期変化の補正: 音源パルス周期の変化の MAF への影響を取り除くために、分析フレーム内の逆フィルタ信号の音源パルスの周期が一致するように時間軸を伸縮し補正 [3] する。

(iii) 音源振幅変化の補正: 音源パルスの振幅の変化は、2 つの音源パルスのエネルギーで正規化された MAF[1] を用いる。

第 k 帯域通過信号の第 m 分析フレームにおける周期性の度合いは、次式で定義する MAF から求められる相関値とする。

†三菱電機株式会社 情報技術総合研究所
‡宇都宮大学 工学部

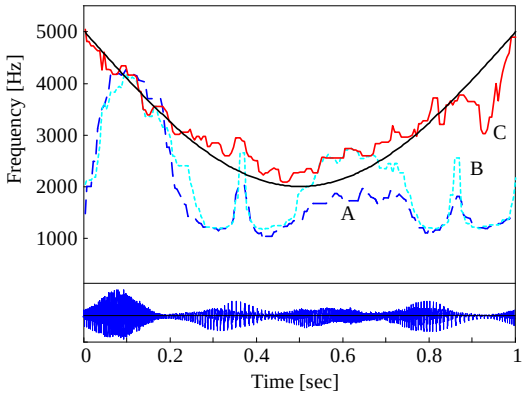

図 1: シミュレーション実験。放物線が真値、A,B,C は本文を参照。

$$R(m,k,\tau,r_\mathrm{t}) = \frac{\sum_{n=0}^{N_\mathrm{r}-1} x(m,n,k)x(m,r_\mathrm{t}n+\tau,k)}{\sqrt{\sum_{n=0}^{N_\mathrm{r}-1} x^2(m,n,k)\sum_{n=0}^{N_\mathrm{r}-1} x^2(m,r_\mathrm{t}n+\tau,k)}} \quad (1)$$

ここで、$x(m,n,k)$ は第 k 帯域を通過した音源信号の第 m 分析フレーム第 n 時刻、r_t は時間伸縮係数、τ は時間遅れ、N_r は分析区間の長さである。この分析長 N_r は予備実験から $N_\mathrm{r} = 2.5/f_0$ と決めた。式 (1) において、分母は、分子の $x(m,n,k)$, と $x(m,r_\mathrm{t}n+\tau,k)$ をそれぞれ規格化し、信号のシマーと振幅のトレンドの影響を相関値から取り除いている。また、$x(m,n,k)$, から $x(m,r_\mathrm{t}n+\tau,k)$ に渡るジッタと基本周期のトレンドが引き起こす波形の変形は、時間伸縮係数 r_t によって $x(m,r_\mathrm{t}n+\tau,k)$ の時間伸縮を行うことで補正する。r_t は $R(m,k,\tau,r_\mathrm{t})$ が最大値を与える r_t とする。

通過帯域信号 $x(m,n,k)$ は、推定された微分声門体積流信号 $\hat{u}(n)$ を等間隔に配置した BPF にかけ求める。微分声門体積流信号 $\hat{u}(n)$ は時変逆フィルタ法によって求める。この時変逆フィルタは、フレーム間隔 T_p ごとにあらかじめ求めておいたフォルマントとアンチフォルマントパラメータを音声標本化周期 T_s ごとに補間し構成 [3] する。

各帯域で得られた周期性の度合いを低周波数帯域から高周波数帯域まで順に閾値 θ_c と比べ下回ったときの周波数値を $f_\mathrm{a}(m)$ とする。最後に後処理として、各フレームで得られた $f_\mathrm{a}(m)$ に、非線形平滑化を行う。

従来の非周期性の分析 [4] においても、基本周期の変化を考慮し音声信号の時間伸縮を行っているが、この場合だと、フォルマントの共振特性をも変化させてしまい時間伸縮 による基本周期の変化の補償が適切に行われない場合が生じる。そこで、本論文では上述のようにフォルマントによる共振特性を逆フィルタ法によって除去し、得られた逆フィルタ信号において時間伸縮を行う。

4. 合成法

有声音源モデルとして用いる RK 音源モデルの群遅延特性を操作することで非周期的成分を付加する。RK 音源モデルによって得られた 1 周期ごとの音源波形の群遅延を文献 [5] の方法を用いてランダム化する。この方法は、任意の周波数値以上の周波数領域においてその群遅延特性を任意の大きさでランダム化できる特徴を持つ。そこで有声音源パルスの群遅延をランダム化する周波数領域の下限値を $f_{\mathrm{a}}(m)$ で制御し、時間的に変化する非周期的成分を合成音声に付加する。

5. 実験

5.1 シミュレーション

自然音声を分析して得られたパラメータを参考にして作成した合成音声によって、分析法の性能を評価する。パラメータは、フォルマントと F0 を 1 秒間、時間的に緩やかに変化させた。分析を行うときの方法として、(A) 基本周期の時間伸縮とフォルマント変化を考慮にいれない方法、(B) フォルマントの変化のみ考慮した方法、(C) 基本周期とフォルマントの両方を考慮した方法を、考える。閾値 θ_{c} はすべての方法で 0.5 に固定した。

図 1 は各方法 A, B, C によって得られた $f_{\mathrm{a}}(m)$ と、真値 (放物線) である $f_{\mathrm{a}}(m)$ を示している。基本周波数 (周期) とフォルマントの時間変化を両方考慮に入れた方法 C が他の方法に比べ明らかに真値に近く、基本周波数の補正、フォルマントの特性の除去の効果がわかる。

5.2 自然音声分析

女性が発声した「遥か遠くの海の深い深いそこに、人魚の王様のお城がありました」を ARX 分析合成法 [3] によって分析合成した。図 2,3,4 は、それぞれ原音声、非周期成分なし、あり場合の合成音声「遥か」の狭帯域のサウンドスペクトログラム (SSG) である。原音声では、ハーモニクス（縞として観察される）が時間的に高周波数領域から低周波数領域まで変化している。非周期成分がない合成音声では、ナイキスト周波数までハーモニクスが観察され、原音声の特性と異なっている。非周期成分を付加した合成音声では、原音声の非周期成分の時間変化をうまく模擬していることがわかる。

6. 結果とまとめ

シミュレーション実験から、音声の動特性を考慮することで非周期成分境界周波数を精度良く安定に求めることができることと、自然音声の分析合成実験での SSG の観察から、本提案法の有効性を示した。非周期成分の聴覚的な印象の効果については、[6] で検討され、本提案法の有効性が示されている。

参考文献

[1] McCree, A. V. and Barnwell, T. P. III, “A Mixed Excitation LPC Vocoder Model for Low Bit Rate Speech Coding,” IEEE Trans. Speech and Audio Processing, Vol. 3, NO. 4 (1995).

[2] 大塚貴弘，粕谷英樹，“音源パルス列を考慮した頑健な ARX 音声分析法,” 日本音響学会誌，58-07，386-397 (2002).

[3] 大塚 貴弘, 粕谷 英樹, “音源・声道モデルに基づいた頑健な音声分析合成法とその応用,” 電子情報通信学会技術研究報告, SP 2001-21 (2001-5).

[4] Yang, C. and Kasuya, H., “Least squares estimation of laryngeal noise in speech signals,” J. Acoust. Soc. Jpn. (E) 16, 123-126 (1995).

[5] Kawahara, H., “Speech representation and transformation using adaptive interpolation of weighted spectrum: Vocoder revisited,” Proc. ICASSP 97, 1303-1306 (1997).

[6] 青山 文彦, 大塚 貴弘, 粕谷英樹 “感情を込めて朗読した童話音声に含まれる非周期的成分の分析・合成・知覚,” 音講論 (2002-9).

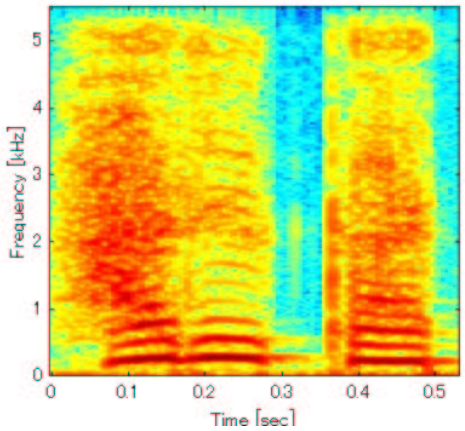

図 2: 原音声

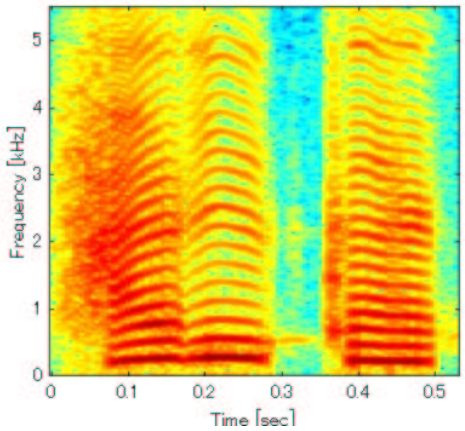

図 3: 分析合成音声（非周期成分なし）

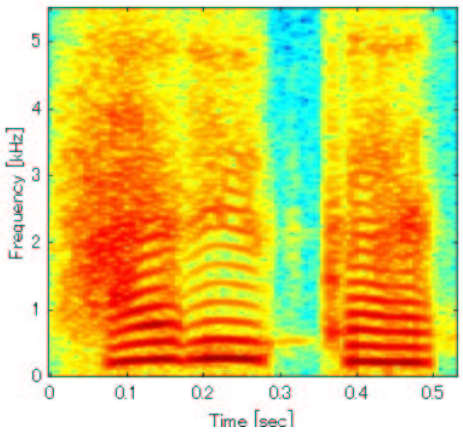

図 4: 分析合成音声（非周期成分あり）