T2R2 東京科学大学 リサーチリポジトリ
Science Tokyo Research Repository

論文 / 著書情報
Article / Book Information

論題(和文)	ＨＭＭ音声合成における残差駆動による自然性向上の検討
Title(English)	
著者(和文)	小池宗幸, 岩野 公司, 古井 貞熙
Authors(English)	Koji Iwano, SADAOKI FURUI
出典(和文)	日本音響学会 2003年春季講演論文集, Vol. , No. 1-6-10, pp. 241-242
Citation(English)	, Vol. , No. 1-6-10, pp. 241-242
発行日 / Pub. date	2003, 3

Powered by T2R2 (Science Tokyo Research Repository)

HMM 音声合成における残差駆動による自然性向上の検討 *

◎小池 宗幸　岩野 公司　古井 貞熙 (東工大)

1　はじめに

本稿では，対数振幅近似 (LMA) フィルタ [1] を用いた HMM 音声合成 [2][3] において，合成音声の自然性向上を目的とし，残差波形を利用した音源生成法を提案する．

これまで，当研究室では多様な形態での音声合成の利用を考え，声質や韻律などの柔軟な制御が可能な HMM によるテキスト音声合成システムを構築してきた [4][5][6]．ボコーダ型音声合成であるこの手法は，自然音声に比べ音質が劣化するという問題があり，自然性向上が望まれている．音質劣化の原因としては，周期的インパルスと白色雑音から生成された音源の問題が取り上げられており，その改善策として音源に残差波形を利用する方式が検討されている [7][8]．残差波形は分析で得られた声道フィルタの逆特性を持つフィルタを構成し，これに自然音声を通して得られる波形である．従って，残差波形を音源としてフィルタを駆動すれば，自然音声を完全に復元できる．

そこで，本稿では HMM 音声合成において，音源に残差波形を利用することを考える．自然音声から合成逆フィルタにより残差波形を抽出，データベース化し，合成時に適当な残差波形素片を選択，接続し，自然に近い音源を生成することで，合成音声の自然性向上をはかる．

以下では，残差波形データベース構築法，音源の生成法，さらに本稿で提案した音源を使用した合成音声の評価について報告する．

2　残差波形データベース構築

まず，自然音声から残差波形を抽出し残差波形データベース (以下，残差波形 DB と呼ぶ) を構築する．

LMA 逆フィルタの伝達関数は次式で定義される．

$$H(z) = \exp(\sum_{m=0}^{M}(-c[m])z^{-m}) \qquad (1)$$

ここで，フィルタ係数 $c[m]$ を自然音声から求めた 0〜24 次のケプストラム系列 (分析周期 12.8ms，分析窓長 25.6ms) とし自然音声から残差波形を抽出する．

次に，残差波形内の励振状態に目視によりマーキングを行う．前後の励振状態から基本周期ずれた時刻付近で最も振幅が大きい励振状態をマーキングした．また，各励振状態にラベル情報を付与する．ラベル情報は，励振状態が検出された区間の音素，その区間より分析された 0〜24 次のケプストラム系列，基本周波数を用いた．

3　音源生成

合成時に，作成した残差波形 DB を利用し音源を生成する．入力テキストの解析によって得られた音韻・韻律情報をもとに波形選択，波形接続の順で音源を生成する．ただし，無声音については従来と同様に白色雑音を用いた．

3.1　波形選択部

波形選択部では，残差波形 DB から最適な励振状態を選択する．

まず，合成時に得られるケプストラム・基本周波数をもとに，音素単位で平均ケプストラム，平均対数基本周波数を求める．次に，音素ごとに，残差波形 DB から最適となる励振状態を 1 つ選択する．具体的には，合成対象音素と一致し，かつ，ケプストラム距離と対数基本周波数誤差の重み付け和が最小となるラベルを検索し，最適な励振状態を決定する．

3.2　波形接続部

波形接続部では，以下のような操作を用いて音源を生成する．

1) 波形選択部で決定された励振状態を中心とする，基本周期の約 2 倍の領域（以下，励振波形と呼ぶ）を残差 DB から切り出す．

2) 得られた励振波形を，基本周波数から求めた励振状態の位置に割り当てるが，その際，隣接する音素との音源波形の不連続性が生じないようにする必要がある．そのため，当該音素の励振波形について，前後の音素で選択された励振波形との補間処理を行う．音素内での位置によって，前後の音素の励振波形との補間係数を調整し，なめらかに音源波形の形状が変化するようにする．

3) 補間により得られた素片波形を PSOLA 方式 [9] で接続する．

4　評価実験

4.1　実験条件

実験には ATR 日本語音声データベース 503 文を用いた．HMM の学習には 493 文，残差波形 DB には学習に用いた 493 文中の 36 文を用いた．

評価文として，HMM の学習，残差波形 DB の構築両方に用いた 5 文 (closed-closed)，HMM 学習に用いたが残差波形 DB 構築には用いていない 5 文 (closed-open)，HMM 学習にも残差波形 DB 構築にも用いていない 5 文 (open-open) を用いた．

また，ケプストラム系列として，自然音声を分析し求めたケプストラム系列 (C_{ana})，HMM 合成により生成したケプストラム系列 (C_{HMM})，音源として，周期的インパルス及び白色雑音による音源 (V_{pul})，提案法による音源 (V_{pro}) を用意し，それらの組合せによる合成音声 4 種を評価音声として用いた．

以下の実験では，被験者にテキストを提示した上で，各音声を被験者ごとにランダムに提示し，明瞭性，自然性，総合評価の 3 項目について 5 段階 (1. 非常に悪い 2. 悪い 3. 普通 4. 良い 5. 非常に良い) で評価を行った．合成音声はヘッドホンを用いて提示した．被験者数は 14 名である．

∗ Improving naturalness using residual exicitation for HMM-based speech synthesis

By Muneyuki Koike, Koji Iwano and Sadaoki Furui (Tokyo Institute of Technology)

各音素の継続時間長は HMM により強制切り出しした結果をそのまま利用した．基本周波数は ATR 日本語音声データベース付随のデータを用いた．

4.2 従来音源との比較

男性話者 1 名 (MHT) が発声した音声データを用いて，HMM の学習及び残差波形 DB を構築し，評価実験を行った．図 1 に実験結果 (評価項目:自然性) を示す．ただし，横軸は各評価音声を表し，縦軸は MOS を表す．図中の系列 average は各評価音声についての全文章に対する MOS である．

全ての条件で，V_{pro} を用いた合成音声のほうが V_{pul} を用いた合成音声より高いスコアとなっていることが分かる．この結果について性能の優劣が見出せるかどうかを有意水準 5 %の 2 標本検定を行ったところ提案手法の優位差が認められた．また，$C_{ana}+V_{pro}$ 及び $C_{HMM}+V_{pro}$ の結果を見ると，音源について closed の場合 (closed-closed) と open の場合 (closed-open) に有意差が見られなかった．用いた残差波形 DB は少量のデータから構築されたものにも関わらず，今回の実験では提案手法の様々な文章の合成に対する有効性が確認された．明瞭性，総合評価についても同様の結果が得られている．

4.3 残差波形 DB の汎用性

特定話者 (話者 MHT) の音声で構築した残差波形 DB の汎用性を調査するために，他話者 (男性話者 MYI，女性話者 FKN) の音声から作成されたケプストラム系列を用いて評価実験を行った．評価文は open-open を用いた．図 5 に実験結果 (評価項目:自然性) を示す．比較のために話者 MHT の音声から作成されたケプストラム系列を用いた場合の結果も同図に示す．

話者 MYI，FKN どちらの話者においても，V_{pro} を用いた合成音声のほうが V_{pul} を用いた合成音声より高いスコアとなっていることが分かる．また，有意水準 5 %の 2 標本検定において 2 手法間の有意差が認められた．明瞭性，総合評価についても同様の結果が得られた．よって，他話者の残差波形 DB を用いて本手法を適用しても，周期的インパルス及び白色雑音による音源を使用するより有効であると判断できる．また，$C_{HMM}+V_{pro}$ の結果を見ると，ケプストラム系列と残差音源に同一話者データを使用した場合と異なる話者を使用した場合で有意な差が見られない．自然性に関する今回の評価では，HMM 音声合成において残差駆動を行った場合に，特定話者による残差波形 DB の汎用性が認められた．しかし，話者 MYI における明瞭性と総合評価に関しては，有意差が認められており，今後，残差波形 DB の量については，さらに検討を進めていく必要がある．

5 まとめ

HMM 音声合成において，合成音声の自然性向上を目的とし，残差波形を用いた音源生成法を提案し，その有効性を評価実験を行い確認した．

今後の課題としては，残差波形 DB の量による音質の変化の検証や，ケプストラムの次元数の変化に対する本手法の有効性についての検証，子音に関する有効な駆動音源制御法の導入などが考えられる．

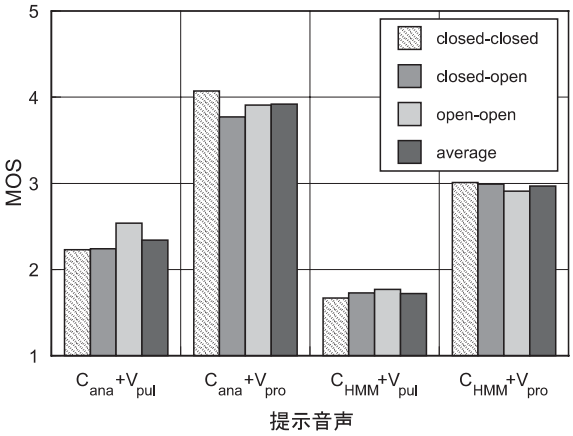

図 1. 従来音源との比較 (評価項目:自然性)

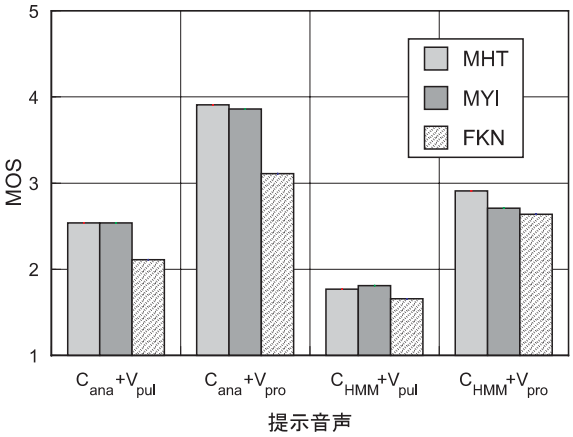

図 2. 残差波形 DB の汎用性 (評価項目:自然性)

参 考 文 献

[1] 今井，“対数振幅近似 (LMA) フィルタ,” 信学論, Vol.J63-A, No.12, pp.886-893 (1980).

[2] 益子 他，“動的特徴量を用いた HMM に基づく音声合成,” 信学論，vol.J79-D-II，no.12，pp.2184-2190 (1996-12).

[3] 徳田 他，“動的特徴量を用いた HMM からの音声パラメータ生成アルゴリズム,” 音響誌，vol.53，no.3，pp.192-200 (1997-3).

[4] 立和 他，“HMM による規則音声合成の検討,” 音講論, 2-3-7, pp.239-240 (1999-3).

[5] 山田 他，“数量化 I 類を用いたピッチパターン生成における制御要因の検討,” 音講論, 1-2-8, pp.221-222 (2001-10).

[6] 外川 他，“数量化 I 類を用いた発話速度制御法,” 音講論, 3-10-9, pp.345-346 (2002-9).

[7] 佐藤，“CVC と音源要素に基づく (SYMPLE) 音声合成,” 音声研資, S83-69, pp.541-546 (1984-1).

[8] 匂坂，“規則による音声合成のための駆動音源作成法,” 音講論, 2-3-19, pp.191-192 (1985-9).

[9] E.Moulines, et al., “Pitch-synchronous wave-form processing techniques for text-to-speech synthesis using diphones,” Speech Communication, vol.9, nos.5-6, pp.453-467 (1990-12).