

論文 / 著書情報
Article / Book Information

論題(和文)	HMM音声合成を利用したキャラクタの形状特徴に応じた音声合成法
Title(English)	
著者(和文)	外川 太郎, 岩野 公司, 古井 貞熙
Authors(English)	Koji Iwano, SADAOKI FURUI
出典(和文)	日本音響学会 2004年春季講演論文集, Vol. , No. 1-7-24, pp. 259-260
Citation(English)	, Vol. , No. 1-7-24, pp. 259-260
発行日 / Pub. date	2004, 9

1 はじめに

近年、多様性のある音声合成の実現が望まれている。本稿では分析合成系である HMM に基づく音声合成 [1] を用いた、キャラクタの形状特徴 (体型) に応じた音声合成法について検討する。本手法では、5 種類の体型のキャラクタに対応する HMM による音響モデル、数量化 I 類による基本周波数 (F_0) パターンモデル、音素継続時間長モデルを作成し、それらのモデルをキャラクタ間で補間することによって、中間的なキャラクタに対応する音声合成用モデルを構築する。本手法を用いて合成した様々な体型のキャラクタの音声について、主観評価実験によりキャラクタとの適合性に関する評価を行う。

2 音声合成システムの概要

本手法で用いる音声合成システムを図 1 に示す。テキスト解析部では、入力となる日本語テキストを解析して音素列とアクセント句情報を出力する。韻律生成部では、各モーラの F_0 と各音素の継続時間長 (以下、時間長) を生成する。 F_0 パターンと、時間長は数量化 I 類を用いてモデル化される。音声合成部では HMM で表現された音素モデルから最尤のケプストラム列を生成し、これに適当なピッチを与えて、MLS フィルタ [2] を用いて音声を合成する。本手法では、時間長を任意に与えたときに最尤のケプストラム列を生成する手法 [3] を用いる。

3 キャラクタの体型に基づく音声合成手法

様々な体型のキャラクタに基づく音声合成を実現するためには、与えられた体型に対応した HMM、 F_0 パターンモデル、時間長モデルという 3 つの音声合成用モデルを作成する必要がある。そこで、1) 「標準」、「太った」、「痩せた」、「背の高い」、「背の低い」キャラクタを想定して読み上げた音声データを用意して 5 種類の音声合成用モデルを作成し、2) それらのモデル間を補間することで目標とする体型のモデル



図 2. 様々な体型のキャラクタの例

を作成し、3) そのモデルを利用して音声を合成する、という音声合成法を提案する。

3.1 各体型に対応した音声合成用モデルの作成

3.1.1 音声データの収録

図 2 に示す、体型の異なる 5 種類のキャラクタを用意した。声優経験のある女性話者 1 名に、この 5 種類のキャラクタを提示し、それぞれのキャラクタの特徴がよく表れるように読み上げを依頼した。読み上げ文章には、ATR 日本語連続音声データベースの 503 文章を使用した。「標準」に対して 503 文章、他の 4 種類の体型に対しては前半の 200 文章の発声を収録した。

3.1.2 音声合成用モデルの作成

収録した音声データから 4 混合の triphone HMM を学習した。HMM の学習では、20kHz の音声信号をフレーム長 6.4ms、フレーム周期 3ms のハミング窓を用いてメルケプストラム分析し、求めた 0~24 次のメルケプストラムとそのデルタ係数を音響パラメータとした。「標準」は 503 発声、それ以外は 200 発声を用いて HMM を学習した。次に、強制切り出しにより時間長を求め、数量化 I 類で時間長モデルを作成した [4]。さらに、各モーラの F_0 をケプストラム法で抽出し、数量化 I 類で F_0 パターンモデルを作成した [5]。 F_0 パターンモデル、時間長モデルは全てのキャラクタについて、200 発声で学習した。

3.2 各体型のキャラクタ間でのモデルの補間

中間的な体型のキャラクタは、各キャラクタの変化を組み合わせで作成する。図 3 は「太った」と「背の低い」の中間の「太って低い」の例である。中間的な体型のキャラクタに対応する音声合成用モデルは、それぞれのモデルをキャラクタ間で補間することで生成する。具体的には、中間的なキャラクタの HMM は、遷移確率、混合ガウス分布の混合重み、平均値ベクトル、共分散行列の各パラメータの平均をとることで作成し、数量化 I 類によるモデルは、平均値 \bar{y} と数量 x_{fc} の平均をとることで作成した。

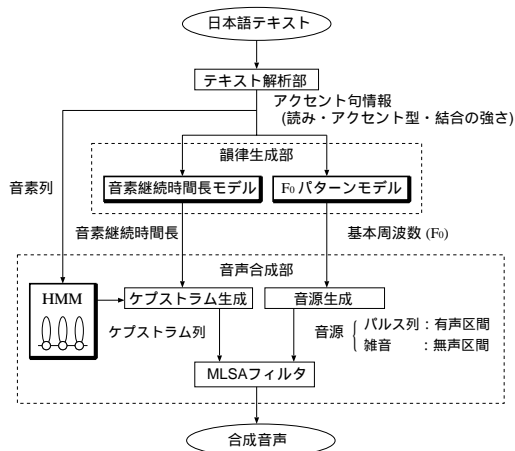


図 1. 音声合成システムの概要

* HMM-based speech synthesis corresponding to character's shapes.



図 3. 中間的な体型のキャラクタの例

表 1. 体型と音声との適合性 (スコアの平均)

キャラクタの体型	合成音声				
	標準	太った	痩せた	高い	低い
太った	1.88	4.25	3.25	3.12	2.12
痩せた	3.12	1.12	3.12	3.00	4.38
背の高い	2.62	3.12	3.00	4.12	2.75
背の低い	2.88	1.50	2.25	2.12	4.50

表 2. 中間的な体型と音声との適合性 (スコアの平均)

キャラクタの体型	合成音声				
	標準	太った	高い	低い	太って低い
太って低い	1.62	3.38	3.00	3.38	4.38

4 主観評価実験

4.1 体型と音声の適合性に関する評価実験

学習したモデルを利用して、図 2 に示す 5 種類の体型のキャラクタに対応する音声を合成し、キャラクタの体型と合成音声との適合性に関して 5 段階 (1. 全く合っていない ~ 5. 非常に合っている) で評価を行った。「標準」を除く 4 種類の体型のキャラクタと 5 種類の合成音声との 20 の組み合わせについて、その適合性を調べた。被験者一人に対し、各組み合わせについて 1 合成音声を提示し、計 20 回の評価を行った。被験者は 8 名である。

スコアの平均値を表 1 に示す。表中の太字は、各体型について、最も適合性が高かったところを示しており、これらの数値 (スコア) は、他の合成音声のスコアに対して 5% の危険率で有意差が得られた。「太った」、「背の高い」、「背の低い」キャラクタについては、対応した合成音声に対するスコアが高いことから、それぞれのキャラクタにとって適合した合成音声が実現できているといえる。しかし、「痩せた」キャラクタについては、対応する合成音声との間に、大きな適合性が観測されなかった。

そこで、合成音声との適合性の高い 2 つのキャラクタ「太った」と「背の低い」に注目し、両者の中間的なキャラクタ「太って低い」を作成し、補間によって作成した音声合成用モデルによって合成された音声との適合性について、同様に評価を行った。スコアの平均値を表 2 に示す。この結果から、中間的なキャラクタに対しても、対応する合成音声が高最も適合性を示していることがわかる。しかし、「太った」と「背の低い」に対応する合成音声のスコアとの有意差は確認できなかった。

4.2 各モデルの主観評価への影響に関する評価実験

「太った」、「太って低い」、「背の低い」の 3 種類のキャラクタについて、ケプストラム、 F_0 パターン、時間長のどの特徴量がキャラクタの体型と合成音声と

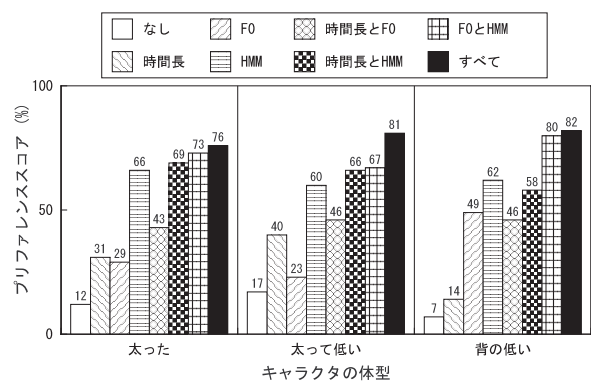


図 4. 各モデルの主観評価への影響

の適合性に影響を与えているのかを調べた。HMM、 F_0 パターン、時間長の 3 つのモデルそれぞれについて、「標準」キャラクタ用のモデルを使用した場合と、該当する体型のキャラクタ用のモデルを使用した場合を考える。この 8 通りのモデルの組み合わせで合成音声を作成し、どの組み合わせによる合成音声が意図するキャラクタとの適合性が高いかを調べた。評価は、対比較実験により行った。被験者は 3 種類のキャラクタのそれぞれについて $8C_2 = 28$ のペアの音声を聴取し、どちらの音声によりそのキャラクタに合っているかを評価した。被験者は 8 名である。

キャラクタごとの対比較実験によるプリファレンススコアを図 4 に示す。凡例中の表記は、どのモデルに意図する体型のキャラクタ用のモデルを使用したかを示している。どのキャラクタについても、全てのモデルに、意図するキャラクタ用のモデルを使用した場合に、最も適合性が高い。モデル別にみると、特に HMM による効果が大いことがわかる。 F_0 パターンモデルと時間長モデルによる効果は、HMM に比べると小さいが、使用したことによる適合性の向上が観測された。

5 まとめ

HMM 分析合成系を用いた、キャラクタの体型に応じた音声合成手法を検討した。提案法で合成した音声と体型との適合性に関する評価実験を行い、提案法の有効性を確認した。

今回は 1 名の発話者の音声データから学習した合成用モデルを扱った。今後の課題として、他の発話者へ適用したときに、同様のキャラクタと合成音声の適合性の傾向がみられるかを評価する必要がある。また、標準キャラクタのみの発声・モデルから、様々な体型に対応する音声の合成手法の検討などもあげられる。

参考文献

- [1] 益子 他, 信学論, vol.J79-D-II, no.12, pp.2183-2190 (1996-12).
- [2] 今井 他, 信学論, vol.J66-A, no.2, pp.122-129 (1983-2).
- [3] 立和 他, 音講論, 2-3-7 (1999-3).
- [4] 岩野 他, 信学技報, SP2002-73, pp.11-16 (2002-8).
- [5] 山田 他, 情処研報, vol.2001, no.100, pp.15-20 (2001-10).