# 論文 / 著書情報
# Article / Book Information

| | |
|---|---|
| 論題(和文) | 話し言葉音声合成における韻律制御要因の有効性の評価 |
| Title(English) | Evaluation of the effrctiveness of prosody control factors for spontaneous speech synthesis |
| 著者(和文) | 伊藤 芳幸, 岩野 公司, 古井貞熙 |
| Authors(English) | Yoshiyuki Ito, Koji Iwano, SADAOKI FURUI |
| 出典(和文) | 日本音響学会2008年秋季講演論文集, , No. 1-4-16, pp. 271-272 |
| Citation(English) | , , No. 1-4-16, pp. 271-272 |
| 発行日 / Pub. date | 2008, 9 |

# 話し言葉音声合成における韻律制御要因の有効性の評価*

◎伊藤芳幸，岩野公司，古井 貞熙 (東工大)

## 1 はじめに

近年，多様な音声合成技術の一つとして，話し言葉音声合成が検討されている．当研究室では HMM 音声合成に基づく話し言葉音声合成システムの構築を目指しており，日本語話し言葉コーパス (CSJ) を用いて話し言葉音声のケプストラム情報 (以下ケプストラム)，音素継続時間長 (音素長)，基本周波数情報 ($F_0$) のモデル化を行っている．先行研究において各モデルのモデル化精度を調査したところ，ケプストラムは HMM によって十分な精度でモデル化されている一方，音素長，$F_0$ といった韻律情報の数量化 I 類によるモデル化には，改善の余地があることがわかった [1]．先行研究では，韻律制御に用いる制御要因の種類や数について十分な検討が行われていなかったことから，本研究では，どのような制御要因が合成音声の「話し言葉らしさ」に影響を及ぼすかを調査する．

## 2 音声合成システムとモデル構築

### 2.1 HMM 音声合成に基づく音声合成システム

本研究では，日本語話し言葉コーパス (CSJ) を用いて話し言葉音声合成のためのモデルの学習を行う．各音素の音素長とモーラ毎の $F_0$ 値は，数量化 I 類によって構築された韻律モデルを用いて推定する．その際に用いる，入力文のアクセント型，音調結合などの言語情報は，CSJ のイントネーションラベルから作成した．推定した音素長と音素 HMM を用いて，入力の音素列に対して最尤のケプストラム列を生成し [2]，それを混合励振源モデル [3] で生成された音源信号とともに MLSA フィルタ [4] に入力することで音声を合成する．

### 2.2 モデルの構築

#### 2.2.1 ケプストラムモデル

各音素のケプストラム情報はスキップのない 3 状態 4 混合 left-to-right 型の triphone HMM でモデル化する．特徴量には，STRAIGHT 分析 [5] で得られたスペクトルを変換して作成したメルケプストラムを利用する．分析には，窓幅 16ms の Blackman 窓を用い，フレーム周期は 5ms である．0〜25 次のメルケプストラムとその Δ 係数を音響パラメータとした．

#### 2.2.2 音素長モデル

音素長推定のためのモデルは，数量化 I 類によって表 1 に示す 13 の音素クラスごとに作成する．合成時にはこのモデルを用いて，音素ごとに音素長を定める [6]．利用する 21 の制御要因を表 2 に示す．表 2 では，音素長の推定対象の音素を $O$，音素 $O$ が属するアクセント句を $W$，アクセント句 $W$ が属する呼気段落（ポーズで区切られる音声区間）を $P$ とした．

#### 2.2.3 基本周波数情報モデル

$F_0$ も音素長と同様に，数量化 I 類を用いてモデル化を行う．目的変数は各モーラの母音，撥音，長音の中心時刻における（対数変換された）$F_0$ 値であり，音声合成時には推定した $F_0$ 値を直線補間することで文全体の $F_0$ パタンを生成する．用いる 21 の制御要因を表 3 に示す．表 3 では，推定対象のモーラを $M$，モーラ $M$ が属するアクセント句を $W$，アクセント句 $W$ が属する呼気段落を $P$ とする．モーラ $M$ が，アクセント句 $W$ の第 $n$ モーラであるとすると，数量化 I 類のモデルは $n = 1, 2, 3, 4, 5$ 以上，の 5 つの場合に分けて作成する．

Table 1 音素クラスの一覧

| 音素クラス | 音素 |
|---|---|
| 1. 母音 | /a/,/i/,/u/,/e/,/o/ |
| 2. 撥音 | /N/ |
| 3. 促音 | /Q/ |
| 4. 長音 | /-/ |
| 5. 有声破裂音 | /b/,/d/,/g/ |
| 6. 無声破裂音 | /p/,/t/,/k/ |
| 7. 有声摩擦音 | /z/,/j/ |
| 8. 無声破擦音 | /ch/,/ts/ |
| 9. 無声摩擦音 | /f/,/h/,/s/,/sh/ |
| 10. 鼻音 | /m/,/n/ |
| 11. 流音 | /r/ |
| 12. 半母音 | /w/,/y/ |
| 13. 拗音 | /by/,/dy/,/gy/,/py/, /ky/,/hy/,/ry/,/my/,/ny/ |

Table 2 音素長推定に用いる制御要因．括弧内はカテゴリ数．

| | |
|---|---|
| 1 | $W$ のモーラ数（9） |
| 2/3 | $P$ 内で $W$ に先行/後続するモーラ数（9） |
| 4/5/6 | 先行/当該/後続アクセント句のアクセント型（7） |
| 7 | $P$ 内で $W$ に先行する アクセント核を有する句の数（4） |
| 8/9 | $W$ 前/後の音調結合の強さ（4） |
| 10/11 | $W$ 前/後の句境界のポーズの長さ (9) |
| 12/13 | $W$ の 2 つ前/後の音調結合の強さ（5） |
| 14/15 | $W$ の 3 つ前/後の音調結合の強さ（5） |
| 16 | $O$ が属するモーラの $W$ 内のモーラ位置（9） |
| 17 | 音素 $O$ の種類 (1〜9：$O$ の音素クラスによって変化．) |
| 18/19 | 音素 $O$ の前/後の音素の種類 (18〜29：$O$ の音素種によって変化．) |
| 20/21 | 音素 $O$ の 2 つ前/後の音素の種類 (18〜29) |

Table 3 $F_0$ 推定に用いる制御要因．

| | |
|---|---|
| 1 | $W$ のモーラ数（8） |
| 2/3 | $P$ 内で $W$ に先行/後続するモーラ数（9） |
| 4/5 | 先行/後続アクセント句のアクセント型（7） |
| 6 | $P$ 内で $W$ に先行する アクセント核を有する句の数（4） |
| 7/8 | $W$ 前/後の音調結合の強さ（4） |
| 9/10 | $W$ 前/後の句境界のポーズの長さ（9） |
| 11/12 | $W$ の 2 つ前/後の音調結合の強さ（5） |
| 13/14 | $W$ の 3 つ前/後の音調結合の強さ（5） |
| 15 | トーンパタン [7] (5〜10：$n$ により異なる．) |
| 16 | 当該音素（母音，撥音，長音）の種類（8） |
| 17/18 | 16 の音素の前/後の音素の種類（13） |
| 19/20 | 16 の音素の 2 つ前/後の音素の種類（6） |
| 21 | $M$ の $W$ 内のモーラ位置（$n \geq 5$ の場合）（6） |

## 3 最適な制御要因の組合せの検討

韻律制御に有効な制御要因の種類を調べるために，21 個の制御要因から，最も重要でない制御要因 (除いたときに推定誤差の増加が最も小さい制御要因) を順に除きながら韻律モデルの学習・推定誤差の算出を行う．残っている制御要因ほど重要であると考えられるので，その結果に基づいて各制御要因を順位付けする．学習データとして，CSJ のコアに含まれる男性話者 2 名，女性話者 2 名の模擬講演音声を用い，話者ごとに韻律モデルの学習と制御要因の順位付けを行った．なお，拗音に関しては十分な学習データが存在しなかったため，誤差の算出の対象から除外した．

*Evaluation of the effectiveness of prosody control factors for spontaneous speech synthesis, by Yoshiyuki Ito, Koji Iwano and Sadaoki Furui (Tokyo Institute of Technology)

話者 4 名分の順位付けの結果から，各制御要因の平均順位を求めて順位を付け直したところ， 音素長に関する重要な制御要因の順位は，

19 18 11 21 16 20 1 17 10 5
2 6 3 4 15 13 12 7 14 9 8

$F_0$ に関しては，

15 17 3 2 1 18 4 9 10 5
6 16 11 12 13 20 19 7 21 8 14

となった．

## 4 評価実験

韻律制御に用いる制御要因数を変化させて音声を合成し，被験者による聴取実験を行うことで，制御要因数と種類の違いが話し言葉らしさに与える影響を調査する．音素長に関する調査では，上位 1, 4, 7, 21 個の制御要因を用いた場合の合成音声を作成した．この際，$F_0$ は抽出値 (正解値) を用いた．また，ポーズ長の推定は有効な手法が確立されていないため，正解値を用いることとした．$F_0$ に関する調査では，上位 1, 4, 7, 10, 21 個の制御要因を用いた場合の合成音声を作成し，その際の音素長は正解値を用いた．

話者 4 名それぞれについて，学習に用いた文のうちランダムに 5 文を選び，評価実験に用いた．同じ文を制御要因数の異なる韻律モデルで合成した音声のペアを 11 名の被験者にヘッドホンで提示し，どちらが話し言葉らしく聞こえるかを評価してもらった．音素長，$F_0$ のそれぞれの評価に対して，一人の被験者は，各制御要因数間（例えば制御要因数 1 と 7）の比較評価を 2 回行っている．合成音声の話者や文の種類，制御要因数の組み合わせやペアの提示順などはランダムで提示した．音素長の各制御要因数におけるプリファレンススコアと推定誤差の結果を図 1 に，$F_0$ の場合の結果を図 2 に示す．

音素長に関して，各モデル間のプリファレンススコアの差を二項分布に基づいて有意水準 5% で検定したところ，制御要因数 1 と他の制御要因数との間には有意差がみられたが，制御要因数 4, 7, 21 の間には有意差がみられなかった．この結果から，音素長のモデル化に対しては，上位 4 つ程度の制御要因（周辺音素の種類や当該/後続アクセント句間のポーズ長）で，合成音声の話し言葉らしさがほぼ最大に達していることがわかる．読み上げ調音声の音素長制御に有効な制御要因の分析結果 [6] では，周辺音素種が制御要因の上位を占め，ポーズ長は中位（8 番目程度）であったことから，読み上げ調音声の合成と比較すると，「ポーズ長」が話し言葉音声合成の音素長制御に重要であることがわかる．

$F_0$ についても同様に各スコアについて検定を行ったところ，制御要因数 1 と他の制御要因数との間には有意差がみられたが，制御要因数 4, 7, 10, 21 間には有意差はみられなかった．したがって，上位 4 つ程度の制御要因（トーンパタン，周辺音素の種類，周辺アクセント句のモーラ数）で，合成音声の話し言葉らしさがほぼ最大に達していることがわかる．なお，読み上げ調音声の $F_0$ 制御要因の分析結果 [8] と比較して，制御要因の重要性の順位には大きな違いは見られなかった．

## 5 まとめ

話し言葉合成音声における韻律（音素長・$F_0$）推定に用いる数量化 I 類の制御要因の数と種類の違いが，合成音声の話し言葉らしさに与える影響を調査した．

制御要因が異なる合成音声間の話し言葉らしさについて，被験者による対比較実験を行った結果，音素長制御には周辺音素種とポーズの長さなどが重要であることがわかった．これまでの我々の研究では，ポーズ長を話し言葉音声合成の制御要因とすることに関して十分な検討を行っていなかった．また，今回の実験ではポーズ長として正解値を利用していることから，今後はポーズ長の推定も含め，ポーズ長による音素長制御が合成音声の話し言葉らしさに与える影響について調査する必要がある．

$F_0$ 制御については，トーンパタン，周辺音素の種類，周辺アクセント句のモーラ数などが重要であることがわかった．しかし，先行研究 [1] では，これらの制御要因を用いて話し言葉音声の $F_0$ 制御を行った場合に，再合成音声と比較して十分な話し言葉らしさが得られていないことが示されている．したがって，今後は別の制御要因の導入や，数量化 I 類以外のモデル化手法の検討を行う必要がある．

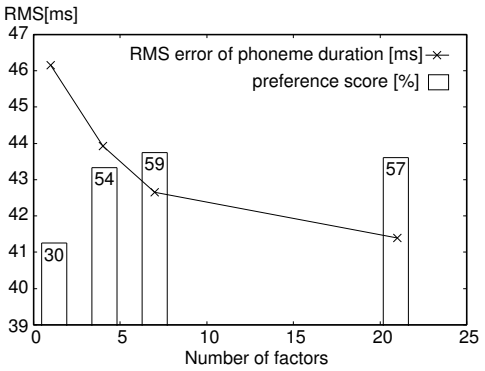


Fig. 1 音素長の推定誤差とプリファレンススコア

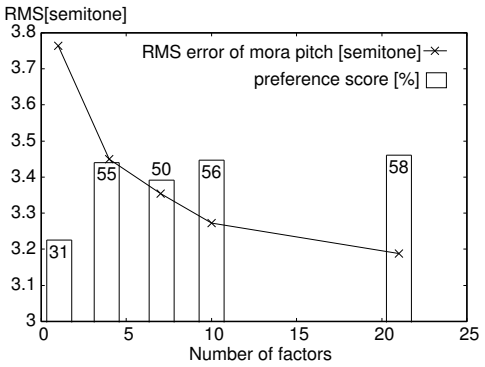


Fig. 2 $F_0$ の推定誤差とプリファレンススコア

**謝辞** STRAIGHT 分析のためのコードを御提供頂いた和歌山大学の河原英紀教授に深く感謝致します．また，実験に用いた音声合成システムの構築に多大な貢献をして下さった当研究室の神山歩相名君に感謝致します．

## 参考文献

[1] 赤川 他，音講論，1-8-5 (2007-3).
[2] 立和 他，音講論，2-3-7 (1999-3).
[3] 吉村 他，信学論，vol.101，no.325，pp.17-22 (2001).
[4] 今井 他，信学論，vol.J66-A，no.2，pp.122-129 (1983).
[5] H. Kawahara et al., Speech Communication, vol.27，pp.187-207 (1999).
[6] 岩野 他，情処研報，vol.101，no.292，pp.11-16 (2002-8).
[7] 阿部 他，音響誌，vol.49，no.10，pp.682-690 (1993-10).
[8] 山田 他，音講論，1-2-8 (2001-10).