社団法人　電子情報通信学会
THE INSTITUTE OF ELECTRONICS,
INFORMATION AND COMMUNICATION ENGINEERS

信学技報
TECHNICAL REPORT OF IEICE.

有声/無声境界の動的特徴量を考慮したピッチのモデル化

全　　炳河† 　徳田　恵一† 　益子　貴史†† 　小林　隆夫†† 　北村　　正†

† 〒 466–8555 名古屋市昭和区御器所町 名古屋工業大学 知能情報システム学科
†† 〒 226–8502 横浜市緑区長津田町 4259 東京工業大学 大学院総合理工学研究科

あらまし　多空間上の確率分布に基づく HMM (multi-space probability distribution HMM: MSD-HMM) により, ピッチパターンをモデル化し, 学習した MSD-HMM からピッチパターンを出力する手法を提案し, 音声合成に適用した. しかしこれまでは, 有声/無声境界における動的特徴量を考慮せずにピッチをモデル化していたため, ピッチパターンを生成した際, 無声区間を挟んだ 2 つの有声区間の間でピッチパターンが不連続に変化する場合があった. 本論文では有声/無声境界の動的特徴量を考慮して MSD-HMM によりピッチパターンをモデル化し, より自然性の高いピッチパターンを生成する手法について述べる.

キーワード　ピッチパターン生成, 多空間確率分布, 隠れマルコフモデル, 音声合成

A Pitch Pattern Modeling Technique using Dynamic Features on the Border of Voiced and Unvoiced Segments

Heiga ZEN†, Keiichi TOKUDA†, Takashi MASUKO††, Takao KOBAYASHI††, and Tadashi KITAMURA†

† Depertment of Computer Science, Nagoya Inst. of Tech. Gokiso-cho, Showa-ku, Nagoya, 466–8555 Japan
†† Interdisciplinary Graduate School of Science and Engineering, Tokyo Inst. of Tech. 4259, Nagatsuta, Midori-ku, Yokohama, 226–8502 Japan

Abstract　We suggested that pitch pattern modeling method and pitch pattern generation method using multi-space probability distribution, and applied them to speech synthesis. However, this methods does't use dynamic features on the border of voiced/unvoiced segments, sometimes generated pitch patterns contain uncontinuous regions between two voiced segments separated by unvoiced segment. This paper describes pitch pattern modeling and generation methods using dynamic features on the border of voiced/unvoiced segments.

Key words　pitch pattern generation, multi-space probability distribution, hidden Markov model, speech synthesis

1. まえがき

HMM は音声スペクトル列の優れた統計的モデル化手法として, 音声認識の分野で広く用いられている. また我々は, HMM からのパラメータ生成手法 [1] を用いた音声規則合成システム [2] を提案し, 滑らかで自然性の高い音声スペクトル列が得られること, 話者適応技術を応用することにより容易に多様な声質で音声合成できること [3] を示した.

一方, HMM をピッチパターンの生成に用いる試みはいくつか行われている [4] [5]. ピッチパターンは, 有声区間では 1 次元の連続値, 無声区間では無声であることを表す離散シンボルとして観測されるため, 通常の離散 HMM や連続 HMM を直接用いることはできず, 何らかの工夫が必要となる. 例えば, (1) 無声区間のピッチとして分散の大きな乱数を与える [6], (2) 無声区間のピッチの値を 0 として混合分布によりモデル化する [7], (3) 無声区間のピッチの値は存在するが観測できなかったとして EM アルゴリズムを適用する [8], などの方法が用いられている.

これに対し我々は, ピッチパターンを, 有声を表す 1 次元空間の出力と無声を表す 0 次元空間からの出力が時間的に混在した系列としてとらえ, 多空間上の確率分布に基づく HMM(multi-space probability distribution HMM: MSD-HMM) [9] を用いてモデル化し, 尤度最大化基準によりピッチパターンを生成する手法 [10] を提案した.

しかし [10] では, 有声/無声境界の動的特徴量を計算しておらず, 有声/無声境界の動的特徴量を無声を表すシンボルとしてピッチパターンをモデル化していた. このため, ピッチパターンを生成したとき, 有声/無声境界で動的特徴量が考慮されないため, 短い無声区間を挟んだ 2 つの有声区間の間で不連続なピッチパターンが生じ, 原音声と比較してアクセントが異なって聞こえる場合があった. また, 静的特徴量が有声であるのに, 動的特徴量は無声であるといった矛盾が生じていた. そこで本論文では, 当該フレーム及び前後最近傍の有声フレームの静的特徴量より動的特徴量を求め, 有声/無声境界の動的特徴量を考慮してより精度の高いピッチパターンのモデルを構築する. また, ピッチパターン生成においても有声/無声境界の動的特徴量を考慮してパラメータ生成を行い, 不連続な区間がない自然なピッチパターンを生成する.

2. 有声/無声境界での動的特徴量

文献 [2] では図 1 のように, 当該フレームとその前後それぞれ 1 フレームずつで計算した 1 次及び 2 次の回帰係数から微分係数に対応する値を求め, 動的特徴量としてモデル化した. 時刻 t における 1 次と 2 次の動的特徴量 δp_t, $\delta^2 p_t$ は以下のようになる.

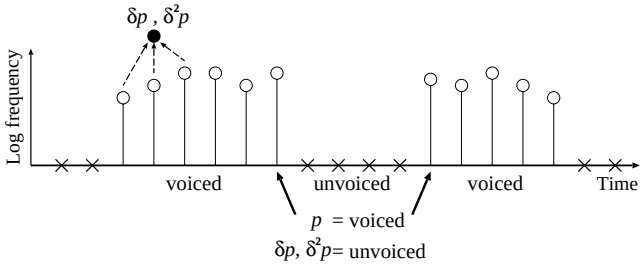

図 1　従来の動的特徴量の計算
Fig.1. conventional dynamic features

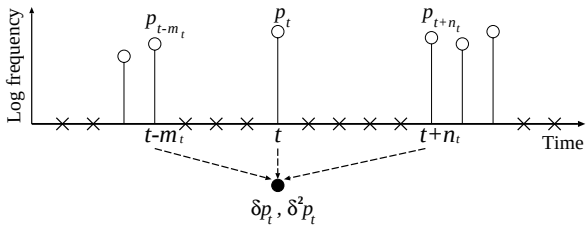

図 2　境界を考慮した動的特徴量の計算
Fig.2. dynamic features considering borders

$$\delta p_t = \frac{1}{2}p_{t+1} - \frac{1}{2}p_{t-1} \tag{1}$$

$$\delta^2 p_t = p_{t-1} - 2p_t + p_{t+1} \tag{2}$$

δp_t, $\delta^2 p_t$ はそれぞれ, 計算に利用される静的特徴量が全て有声の場合のみ計算され, 無声区間及び有声/無声境界の有声フレームについては, 動的特徴量を計算していない. そこで本論文では図 2 に示すように, 当該フレーム及び前後最近傍の有声フレームから回帰係数を計算する. 得られた回帰係数から微分係数に対応する値を求め, これを動的特徴量とする. 時刻 t の有声フレームにおいて, 前後最近傍フレームまでのフレーム数をそれぞれ m_t, n_t としたとき, 当該フレームの 1 次及び 2 次の微分係数を求める式は以下のようになる.

$$\delta p_t = \frac{m_t p_{t+n_t}}{m_t n_t + n_t^2} + \frac{(n_t - m_t)p_t}{m_t n_t} - \frac{n_t p_{t-m_t}}{m_t^2 + m_t n_t} \tag{3}$$

$$\delta^2 p_t = 2\left\{ \frac{p_{t-m_t}}{m_t^2 + m_t n_t} - \frac{p_t}{m_t n_t} + \frac{p_{t+n_t}}{m_t n_t + n_t^2} \right\} \tag{4}$$

式 (3), (4) は $m_t = n_t = 1$ のとき, それぞれ式 (1), (2) と等しい.　音声の最初の有声/無声境界では, 当該フレームより前の有声フレーム, 最後の有声/無声境界では当該フレームより後の有声フレームが存在しない. このため, 式 (3), (4) において音声の最初の

有声/無声境界では m_t, 最後の有声/無声境界では n_t を無限大として計算する. この場合の動的特徴量の計算式は以下のようになる.

- 最初の有声/無声境界

$$\delta p_t = \frac{p_{t+n_t} - p_t}{n_t} \tag{5}$$

$$\delta^2 p_t = 0 \tag{6}$$

- 最後の有声/無声境界

$$\delta p_t = \frac{p_t - p_{t-m_t}}{m_t} \tag{7}$$

$$\delta^2 p_t = 0 \tag{8}$$

3. ピッチパターンの学習

3.1 MSD-HMM によるピッチのモデル化

ピッチパターンは連続値を取る有声区間と値を持たない無声区間の時系列として表されるため, 通常の連続 HMM や離散 HMM では直接モデル化することができない. そこで, ピッチパターンを有声区間を表す k 次元空間 Ω_1 と, 無声区間に対応する 0 次元の空間 Ω_2 の二つの空間から出力される観測事象と考え, MSD-HMM によりモデル化する.

有声/無声を表す空間インデックスの集合を X , 有声区間におけるピッチの値等を含んだ特徴ベクトルを $\boldsymbol{x}$, ピッチに関する観測事象を $\boldsymbol{o} = (X, \boldsymbol{x})$ とする. $X = \{1\}$ のときには, 有声区間を表し, $\boldsymbol{x}$ はピッチの静的特徴量と動的特徴量を含んだ k 次元のベクトルである. また, $X = \{2\}$ のときには, 無声区間を表し, $\boldsymbol{x}$ は 0 次元 ($\boldsymbol{x}$ は値を持たない) となる. このとき, MSD-HMM の状態 i における観測事象 $\boldsymbol{o}$ に対する出力確率は, 次のように表される.

$$b_i(\boldsymbol{o}) = \sum_{g \in S(\boldsymbol{o})} w_{i_g} \mathcal{N}_{i_g}^{n_g}(V(\boldsymbol{o})) \tag{9}$$

但し, $V(\boldsymbol{o}) = \boldsymbol{x}$, $S(\boldsymbol{o}) = X$ であり, w_{i_g} は各空間に対する重み, $\mathcal{N}_{i_g}^{n_g}$ は各空間の分布で, $n_1 = k$, $n_2 = 0$ であり, $\mathcal{N}_{i_2}^{0}(V(\boldsymbol{o})) = 1$ とする.

各状態の出力確率分布を式 (9) で定義することにより, HMM の枠組みでピッチパターンを直接モデル化することができる.

3.2 特徴ベクトル

スペクトルとピッチをそれぞれ別々の HMM でモデル化した場合, スペクトルとピッチで音素境界を合わせるためには何らかの工夫が必要となる. また, ピッチのみを特徴ベクトルとした場合では, 有声区間, 無声区間ともに音素に関する情報が不足するため, 音素境界を適切に学習することができない. そこで, メルケプストラムと対数基本周波数を合わせて一つのベクトルとし, これを特徴ベクトルとする.

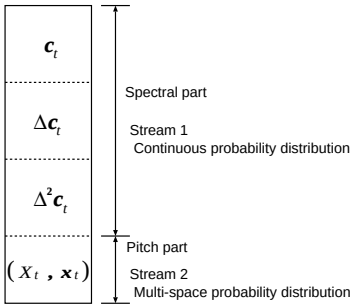

図 3　特徴ベクトル
Fig.3. feature vector

また, 式 (3),(4) を用いてデルタ及びデルタデルタメルケプストラム, デルタピッチ及びデルタデルタピッチを求め, これを動的特徴量として用いる. 従って, 時刻 t での特徴ベクトルは,

$$\boldsymbol{o}_t = [\boldsymbol{o}_t^c, \boldsymbol{o}_t^p]^\top \tag{10}$$

但し,

$$\boldsymbol{o}_t^c = \left[\boldsymbol{c}_t^\top, \Delta\boldsymbol{c}_t^\top, \Delta^2\boldsymbol{c}_t^\top\right]^\top \tag{11}$$

$$\boldsymbol{o}_t^p = (X_t, \boldsymbol{x}_t) \tag{12}$$

となる. ここで, X_t は対数基本周波数及びその 1 次, 2 次動的特徴量についての有声または無声を表す空間のインデックスを要素とする集合である. また, $\boldsymbol{x}_t$ は有声 ($X_t = \{1\}$) の場合は $\boldsymbol{x}_t = [p_t, \delta p_t, \delta^2 p_t]$ となり, 無声 ($X_t = \{2\}$) の場合は値を持たない.

HMM の学習の時には, この特徴ベクトルを図 3 で表されるように 2 つのストリームに分け, メルケプストラムに関する部分は通常の単一対角共分散ガウス分布で, また対数基本周波数に関する部分は, 有声の空間の数を 1 として MSD-HMM でモデル化する.

このように, スペクトルとピッチを一つの特徴ベクトルとすることにより, 音声の音韻情報及び韻律情報を統一的にモデル化することができ, さらに無声が続く区間においても, スペクトルパラメータの出力確率に基づいて, 適切に状態遷移が定まることが期待できる.

3.3 決定木に基づくコンテキストクラスタリング

ピッチに影響を与える変動要因 (ここではコンテキストと呼ぶ) には, アクセント型, 構文情報, 当該・先行・後続音素など, 様々な組み合わせが考えられる. モデル構築の際に, ピッチに影響を与えると考えられるコンテキストを多数用意すれば, より精度の高いモデルが得られると期待できる.

考慮するコンテキストの種類が増加するとコンテキストの組み合わせが指数的に増加するため, モデルあたりの学習データが著しく減少し, モデルパラメータの推定精度が低下する. また, 可能な全てのコンテキストの組み合わせを網羅する学習データを用意することは現実的には不可能であるため, 生成時に学習データ中に存在しないコンテキストの組み合わせが必要となった場合に, 対応するモデルを用意できずパラメータを生成することができなくなる.

この問題を解決するために, ここでは, 決定木を用いたHMMの状態のコンテキストクラスタリングを導入する. 決定木は2分木であり, それぞれの節 (node) 毎にコンテキストを二つに分割する質問が用意されている. 全てのコンテキストは根 (root node) からそれぞれの節の質問に従って木を下って行き, 葉 (leaf) のうちのどれかに達するため, いったん決定木を構築すれば, 学習データに出現しないコンテキストの組み合わせにも対応するモデル (クラスタ) が一意に決定される. このコンテキストクラスタリング手法は, 文献 [10] においてMSD-HMMに対し拡張されている. また, 決定木構築における節分割の停止条件としてはMDL基準 [2] を用いた.

3.4 MSD-HMMからのピッチパターン生成

まず, ピッチパターンを生成するターゲットとなる文章をラベル列に変換する. このラベル列に従って, 学習されたHMMを接続し, 一つの文HMMを作る. 次に, 状態継続長分布から各HMMの状態継続長を決定する. 各フレームの有声/無声は, ここでは簡単に, HMMの各状態におけるピッチのストリームに対する空間の重み (式 (9) 中の w_g) のうち, $w_1 > w_2$ となる状態を有声区間, それ以外を無声区間とする.

図4に示すように, 有声区間の各フレームに対応するHMMの状態の有声空間のガウス分布を連結し, ガウス分布系列 λ を作る. このとき, 連結前の各有声フレームの前後最近傍の有声フレームまでのフレーム数を m_t, n_t とする. 生成するパラメータ列を $\boldsymbol{O}$, 連結された有声区間のフレーム数が T とすると, $P(\boldsymbol{O} \mid \lambda)$ の対数は,

$$\begin{aligned}\log\ P(\boldsymbol{O} \mid \lambda) &= \log \prod_{t=1}^{T} b_{q_t}(\boldsymbol{o}_t) \\ &= -\frac{1}{2}(\boldsymbol{O}-\boldsymbol{M})^\top \boldsymbol{U}^{-1}(\boldsymbol{O}-\boldsymbol{M}) + \frac{1}{2}\log|\boldsymbol{U}| \\ &\quad + \text{Const} \end{aligned} \tag{13}$$

と書くことができる. ここで,

$$\boldsymbol{O} = \left[\boldsymbol{o}_1^\top, \boldsymbol{o}_2^\top, \ldots, \boldsymbol{o}_T^\top\right]^\top \tag{14}$$

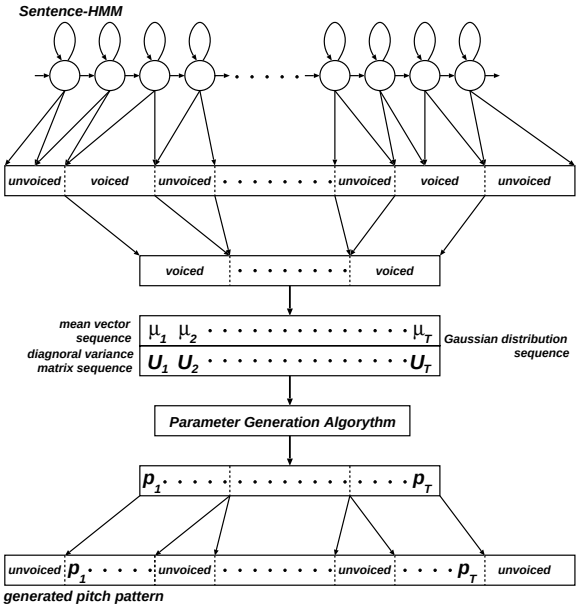

図4　ピッチパターン生成
Fig:4 pitch pattern generation

$$\boldsymbol{M} = \left[\boldsymbol{\mu}_{q_1}^\top, \boldsymbol{\mu}_{q_2}^\top, \ldots, \boldsymbol{\mu}_{q_T}^\top\right]^\top \tag{15}$$

$$\boldsymbol{U} = \text{diag}\left[\boldsymbol{U}_{q_1}, \boldsymbol{U}_{q_2}, \cdots, \boldsymbol{U}_{q_T}\right] \tag{16}$$

であり, $\boldsymbol{\mu}_{q_t}$ と $\boldsymbol{U}_{q_t}$ はそれぞれ, 時刻 t におけるガウス分布の平均ベクトルと共分散行列である. 式 (17), (18) の制約を考えないとき, $P(\boldsymbol{O} \mid \lambda)$ は $\boldsymbol{O} = \boldsymbol{M}$ のときに最大化されることは明らかである. これは, 出力ベクトル系列が平均ベクトルの系列によって与えられることを意味する.

この問題は, 音声認識で広く用いられている動的特徴 [11] を考慮することにより解決される. つまり, 出力ベクトル $\boldsymbol{o}_t$ は, 静的特徴量 p_t と, 動的特徴量 δp_t および $\delta^2 p_t$ で構成され, $\boldsymbol{o}_t = [p_t, \delta p_t, \delta p_t]^\top$ で表されるとする. 但し, δp_t 及び $\delta^2 p_t$ の値は, 静的特徴量 p_t から以下のように計算される.

$$\delta p_t = \sum_{\tau=\{-m_t,0,n_t\}} w_t^{(1)}(\tau) p_{t+\tau} \tag{17}$$

$$\delta^2 p_t = \sum_{\tau=\{-m_t,0,n_t\}} w_t^{(2)}(\tau) p_{t+\tau} \tag{18}$$

但し, $w_t^{(1)}(\tau)$, $w_t^{(2)}(\tau)$ はそれぞれ, 時刻 t における1次及び2次の動的特徴量を計算するための重み係数であり,

$$\begin{aligned}&\left[w_t^{(1)}(-m_t), w_t^{(1)}(0), w_t^{(1)}(n_t)\right] \\ &\quad = \left[-\frac{n_t}{m_t^2 + m_t n_t}, \frac{n_t - m_t}{m_t n_t}, \frac{m_t}{m_t n_t + n_t^2}\right] \end{aligned} \tag{19}$$

$$\left[w_t^{(2)}(-m_t), w_t^{(2)}(0), w_t^{(2)}(n_t)\right]$$

$$= 2\left[\frac{1}{m_t^2 + m_t n_t}, -\frac{1}{m_t n_t}, \frac{1}{m_t n_t + n_t^2}\right] \quad (20)$$

とする.

式 (17), (18) の条件は行列形式により,

$$\boldsymbol{O} = \boldsymbol{W}\boldsymbol{P} \quad (21)$$

と線形変換の形で書くことができる. 但し,

$$\boldsymbol{P} = [p_1, p_2, \ldots, p_T]^\top \quad (22)$$

とする. $\boldsymbol{P}$, $\boldsymbol{O}$ は, それぞれ, T 次元, $3T$ 次元である. $\boldsymbol{W}$ は, $3T \times T$ の行列であり, 1 部の要素に係数 $1, w_t^{(1)}(\tau), w_t^{(2)}(\tau)$ を持ち, 他の多くの要素は 0 となる. 式 (21) の条件の下で, $P(\boldsymbol{O} \mid \lambda)$ を最大にする $\boldsymbol{P}$ は,

$$\frac{\partial \log P(\boldsymbol{W}\boldsymbol{P} \mid \lambda)}{\partial \boldsymbol{P}} = \boldsymbol{0}, \quad (23)$$

とおくことによって得られる線形方程式

$$\boldsymbol{W}^\top \boldsymbol{U}^{-1}\boldsymbol{W}\boldsymbol{P} = \boldsymbol{W}^\top \boldsymbol{U}^{-1}\boldsymbol{M}^\top. \quad (24)$$

により定められる. $\boldsymbol{W}^\top \boldsymbol{U}^{-1}\boldsymbol{W}$ が $T \times T$ の行列であることから, 式 (24) を解くためには $O(T^3)$ の演算量を必要とする. しかし, $\boldsymbol{W}^\top \boldsymbol{U}^{-1}\boldsymbol{W}$ の特別な性質を利用して, コレスキー分解を用いて $O(T)$ の演算量で解くことができ, 有声区間におけるピッチパターン $\boldsymbol{P}$ を得ることができる.

得られたピッチパターンは無声区間を含んでいないため, 無声区間を除いた部分に無声シンボルを入れて, 全体のピッチパターンが生成される.

4. 実　　験

4.1 実 験 条 件

HMM の学習データとして, ATR 日本語音声データベース B セットの話者 MHT による音韻バランス文 503 文のうち 450 文を用いた. 音声データのサンプリング周波数は 16kHz, 分析周期は 5ms で, 長さ 25ms のブラックマン窓を使用した. スペクトル分析には, 24 次のメルケプストラム分析を行った. メルケプストラムおよびそのデルタ, デルタデルタとピッチ及びそのデルタ, デルタデルタを合わせた全 78 次元のベクトルを特徴ベクトルとして HMM をトレーニングした. 使用した HMM は, 対角共分散単一ガウス分布を持つ 5 状態 left-to-right モデルであり, 状態継続長モデルは 5 次元ガウス分布である. スペクトル, ピッチモデルはそれぞれ, 音素内での状態位置毎にクラスタリングを行い, 計 5 つの決定木を作成し, 状態継続長モデルは全体で一つの決定木を作成した. クラスタリングの結果, スペクトル, ピッチモデル, 状態継続長モデルの総分布数はそれぞれ, 928, 1700, 1019 となった. 学習データには含まれない 53 文章を入力とし, 音声を合成した.

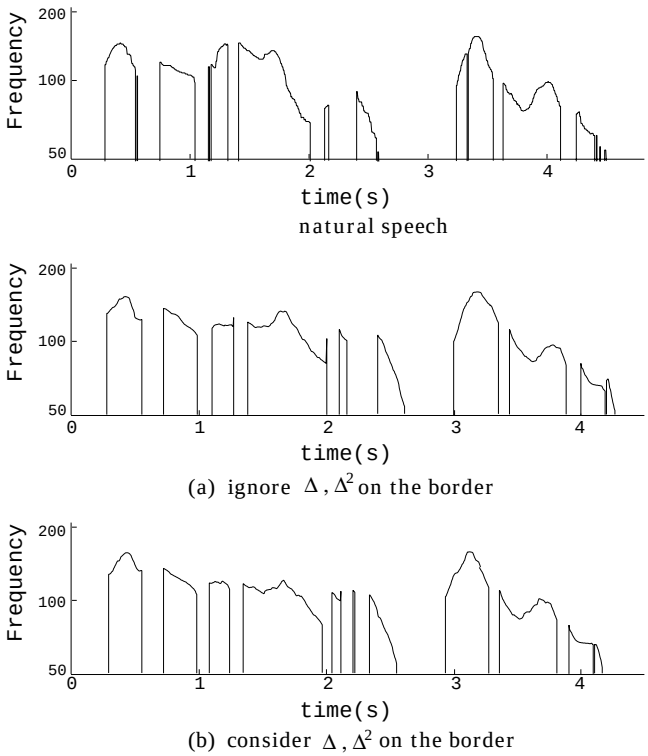

図 5　生成されたピッチパターン
「部屋いっぱいにタバコの濃霧がたちこめ, ゆるやかに動いている」

Fig:5. Generated pitch pattern for a sentence
“ heya ippaini tabakono noumuga tachikome yuruyakani ugoite iru”

4.2 有声/無声境界における動的特徴量の効果

図 5, 6 はそれぞれ, 生成された合成音声のピッチパターンである. 文章は学習データに存在しないものを用いている. それぞれ, (a) 有声/無声境界における動的特徴量を考慮してモデル化及びピッチパターン生成を行ったもの, (b) 考慮せずにモデル化及びピッチパターン生成を行ったものの 2 通りの出力を行った. (a) では, 有声区間の終端及び短い無声区間を挟んだ有声区間において不連続な区間が幾つか生じている. これに対して (b) では, このような有声/無声境界における不連続な区間の多くが解消されている. しかし (b) では, 有声/無声判定の誤りが生じているフレームの最近傍有声フレームにおいて, (a) では生じていない不連続なピッチパターンが生じている. これは有声/無声判定の誤りが生じているフレームに対して, 動的特徴量を考慮して滑らかなピッチパターンを生成しようとするため生じていると思われる.

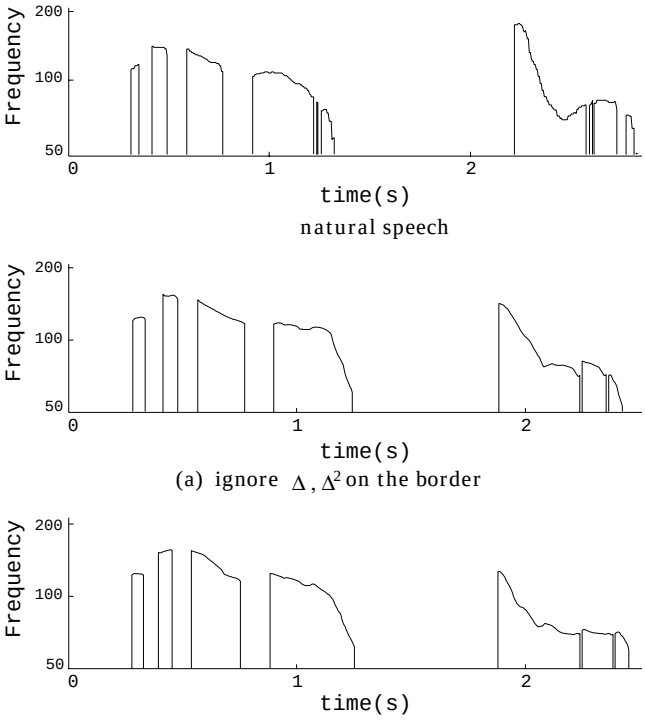

図 6　生成されたピッチパターン
「畑は干上がり土は割れる」
Fig:6. Generated pitch pattern for a sentence
“ hatakewa hiagari tuchiwa wareru ”

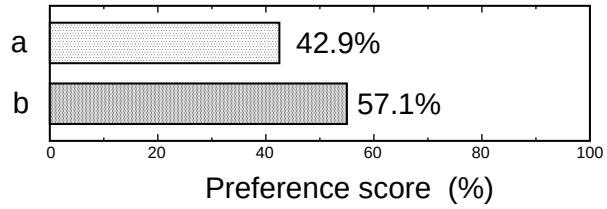

図 7　境界における動的特徴量の効果
Fig.7. effect of dynamic features on the border

4.3　主観評価実験

有声/無声境界の動的特徴量を考慮してモデル化及びピッチパターン生成を行う有効性を確認するため, 主観評価実験を行った. 受聴試験に用いた文章は, 53 文章の中から被験者ごとにランダムに 30 文章を選んだ. 受聴試験のサンプルとしては, (a) モデル化及びピッチパターン生成の際に有声/無声境界の動的特徴量を考慮せずに合成した音声, (b) 考慮して合成した音声の 2 種類のサンプルを用いた. 被験者 8 名に対比較試験によりスペクトル, ピッチ, 状態継続長などを総合的に評価させた.

プレファレンススコアを図 7 に示す. 結果から, モデル化及びピッチパターン生成を行う際に有声/無声境界の動的特徴量を考慮した場合に, 合成音声の品質が向上していることがわかる.

5. むすび

本論文では, MSD-HMM を用いたピッチパターンのモデル化及びピッチパターン生成において, 有声/無声境界での動的特徴量を考慮する手法を提案した. 提案した手法では短い無声区間を挟んだ有声区間の境界において不連続な区間が生ぜず, 自然で滑らかなピッチパターンが生成されることを確認した.

今回の実験では, 有声/無声を簡単にピッチのストリームの空間重みから定めた. このため, 有声/無声の誤りがいくつかのフレームで生じており, 不連続な区間が生ずる原因となっている. そこで, 尤度最大化基準に基づく有声/無声の決定などについて検討を行う必要があると思われる.

文　献

[1] K. Tokuda, T. Yoshimura, T. Masuko, T. Kobayashi, T. Kitamura, “Speech parameter generation algorithms for HMM-based speech synthesis,” Proc. of ICASSP, vol.III, pp.1315–1318, June 2000.

[2] 吉村 貴克, 徳田 恵一, 益子 貴史, 小林 隆夫, 北村 正, “HMM に基づく音声合成におけるスペクトル・ピッチ・継続長の同時モデル化,” 信学論 (D-II), vol.J83-D-II, no.12, pp.2099–2107, Nov. 2000.

[3] 田村正統, 益子貴史, 徳田恵一, 小林隆夫, “HMM 音声合成における MLLR を用いたピッチ・スペクトルの話者適応” 信学技報, SP2001-11, 2001.

[4] A. Ljolje and F. Fallside, “Synthesis of natural sounding pitch contous in isolated utterances using hidden Markov models,” Proc. of IEEE Trans. Acoust., Speech & Signal Process., ASSP-34, pp.1074–1080, 1986.

[5] T. Fukada, Y. Komori, T. Aso and Y. Ohora, “A study of pitch pattern generation using HMM-based statistical information,” Proc. of ICSLP, pp.723–726, 1994.

[6] G.J. Freij and F. Fallside, “Lexical stress recognition using hidden Markov models,” Proc. of ICASSP88, pp.135–138, April 1988.

[7] U. Jensen, R.K. Moore, P. Dalsgaard, and B. Lindberg, “Modeling intonation contours at the phrase level using continuous density hidden Markov models,” Computer Speech and Language, vol.8, no.3, pp.247–260, July 1994.

[8] K. Ross, and M. Ostendorf, “A dynamical system model for generating F_0 for synthesis,” Proc. ESCA/IEEE Workshop on Speech Synthesis, pp.131–134, 1994.

[9] 徳田 恵一, 益子 貴史, 宮崎 昇, 小林 隆夫, “多空間上の確率分布に基づいた HMM,” 電子情報通信学会論文誌, J83-D-II, No. 7, pp.1579–1589, 2000.

[10] 益子 貴史, 徳田 恵一, 宮崎 昇, 小林 隆夫, “多空間確率分布 HMM によるピッチパタン生成,” 電子情報通信学会論文誌, J83-D-II, No. 7, pp.1600–1609, 2000.

[11] S. Furui, “Speaker independent isolated word recognition using dynamic features of speech spectrum,” IEEE Trans. Acoust., Speech, Signal Processing, vol.34, pp.52–59, 1986.