

基本周波数パタンの大局的な制御が聴覚印象に与える効果*

青木愉子, 森大毅, 粕谷英樹 (宇都宮大・工)

1 はじめに

基本周波数パターン制御法には、アクセント句 (AP) 単位でモデル化されるパターンを、平行移動することによって生成する方法がある [1]。その平行移動の値は、ダウンステップやその他の様々な要因から決定され、決定木や数量化 I 類などを用いて機械学習される。

合成音声の自然性をさらに向上させるには、 F_0 レンジやダウンステップが支配される要因を調査する必要がある。そこで、本稿ではイントネーション句 (IP) の F_0 レンジやダウンステップがどのような要因に影響されるのかを調査した。また TTS システムの基本周波数パターン生成において、 F_0 レンジとダウンステップ量を分離したモデル化を行った。さらに主観評価実験によって、基本周波数パタンの聴覚的な効果を検証した。

2 言語的要因が IP レベルに与える影響

基本周波数パタンの模式図を Fig. 1 に示す。横軸は時間である。AP のパターン [2] を、対数周波数軸上のあるレベル (AP レベル) に平行移動することによって基本周波数パターンを構成する。AP レベルは IP レベルとダウンステップ量から求められる。

一般に F_0 レンジがリセットされた場合、IP の境界とされるが、 F_0 レンジのリセット後の IP レベルは必ずしも同じではない。

本稿では、IP レベルの違いに影響を与える要因を調査した。音声資料には成人女性 1 名が朗読したバランス文と散文からなる文セットを用意し、その中の 870 文をトレーニングセットとして使用した。残りの 97 文はテストセットに用いた。音声資料から F_0 を抽出し、各 AP の 2 モーラ目の母音中心点における F_0 値を得る。これは AP の 2 モーラ目にピークのあるものが多かったためである。

Fig. 2 は、IP レベルのヒストグラムを示している。IP レベルの平均値は 298 Hz であり、この値を 0 st としている。Fig. 2 より明らかに IP レベルに大きな幅があることがわかる。従って、IP レベルを定数としたモデル化には問題があるといえる。

Fig. 3 は、IP に含まれる AP 数と先頭 AP の総モーラ数が IP レベルに与える影響を示したものである。横軸を IP 内 AP 数、縦軸を IP レベルとし、IP の先頭 AP の総モーラ数別に平均をとっている。Fig. 3 か

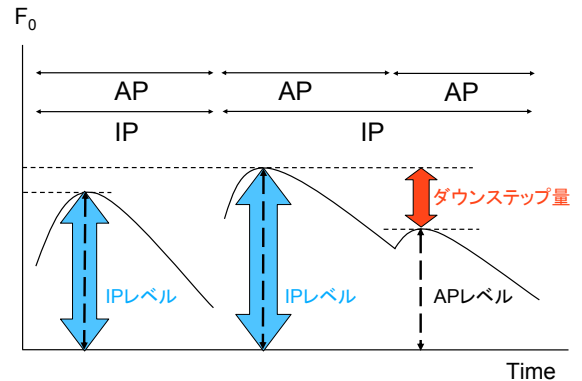


Fig. 1 基本周波数パタンの模式図

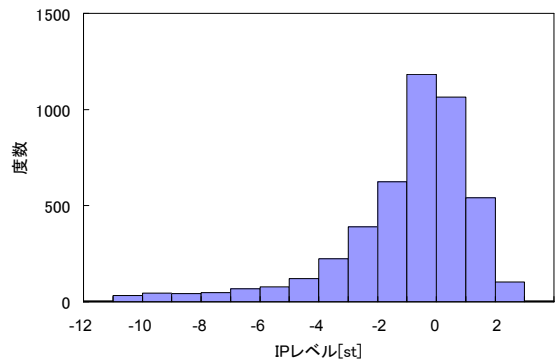


Fig. 2 IP レベルのヒストグラム

ら、IP 内 AP 数が大きくなるに従って、IP レベルが大きくなる傾向があることがわかる。また、先頭 AP の総モーラ数が小さくなるに従って、IP レベルが小さい傾向にある。これは先頭 AP が接続詞である場合に IP レベルが小さくなる傾向にあることを反映していると思われる。従って、IP 内 AP 数と先頭 AP の総モーラ数が IP レベルに影響すると考えられる。

3 基本周波数パタンのモデル化

2 で示したように、IP 内では AP のパターンを平行移動させることによって、基本周波数パターンをモデル化する。AP のパターンは、総モーラ数ごとに決定木を構築することによって生成した [2]。AP レベルは IP レベルとダウンステップ量の分離から求められる。2 で IP レベルに影響を与えることがわかった要因以外にどの要因が IP レベル及びダウンステップに影響を与えるかを調べるために、AP レベルを分離してモデル化を行った。

*The effect of global F_0 control on auditory impression. By T. Aoki, H. Mori, H. Kasuya (Utsunomiya Univ.)

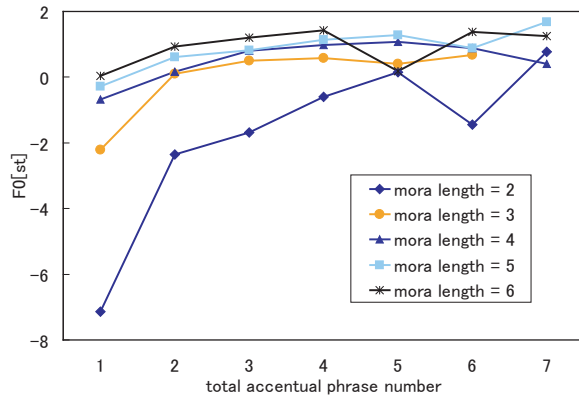


Fig. 3 IP 内 AP 数と先頭 AP の総モーラ数による影響

Table 1 IP レベルの制御要因

制御要因	
先頭 AP のアクセント型	先頭 AP のモーラ数
IP 内 AP 数	文内 IP 位置
文内 IP 数	当該 IP 関係
当該 IP の文内関係	句頭の音節の種類
先行 IP の AP 数	直前のポーズ長

IP レベルとダウンステップは、分類回帰木 (CART) を用いてそれぞれ決定木を構築した。CART プログラムは WEKA [3] を用いた。データベースには、2 で使用したトレーニングセットと同じものを使用し、無声化したものは除いた。得られた AP は 9762 個であった。

制御要因をそれぞれ Table 1, 2 に示す。制御要因は、[2, 4] で用いたものに加え、2 で IP レベルに影響を与えると推測されたものを用いた。

IP レベルを決定木によって構築し、基本周波数パターンを生成する方法を提案法と呼ぶことにする。また IP レベルを一定値として基本周波数パターンを生成する方法を従来法と呼ぶ。従来法と提案法により推定した AP レベルを、実音声との RMS 誤差と相関係数によって比較した。その結果を Table 3, 4 に示す。従来法よりも精度が向上していることがわ

Table 2 ダウンステップの制御要因

制御要因	
当該 AP のアクセント型	当該 AP のモーラ数
IP 内 AP 位置	IP 内 AP 数
文内 IP 位置	文内 IP 数
当該 IP の文内関係	句頭の音節の種類
先行 AP のアクセント型	先行 IP の AP 数
当該 IP 直前のポーズ長	IP レベル

Table 3 Close test

	RMS 誤差 [st]	相関係数 [st]
従来法	2.23	0.74
提案法	1.97	0.81

Table 4 Open test

	RMS 誤差 [st]	相関係数 [st]
従来法	2.09	0.78
提案法	1.97	0.81

かる。

また、提案法で作成した基本周波数パターンを用いて合成音声を作成し、従来法による基本周波数パターンとの主観評価実験を行った。評価用の文は、テストセットの中から 10 文を用いた。呈示回数は 2 回とした。提案法と従来法のモデルで作成した合成音声をランダムな順番で被験者 4 名に呈示し、自然性の高い音声を強制選択させた。

主観評価実験の結果、提案法がよいと判断された割合は全体の 85% (68/80 文) であった。全体的に自然性が向上しているが、特に AP の総モーラ数が小さくかつ IP を単独で構成する場合 (接続詞など) に IP レベルが小さくなり、効果が見られた。しかしながら、品詞の種類によっては自然性が低下するものもあった。従って、品詞の情報を考慮した機械学習によりさらに自然性が向上すると考えられる。

4 おわりに

基本周波数パターンの AP レベルを、IP レベルとダウンステップに分離してモデル化を行った。その結果、特に AP の総モーラ数が小さい場合や IP を単独で構成する場合に IP レベルが小さくなる傾向があり、主観評価実験の結果、自然性が向上した。今後、品詞の情報を考慮した効果についても検討したい。

参考文献

- [1] 籠嶋岳彦他, “代表パターンコードブックを用いた基本周波数制御法,” 信学論 (D-II), J85-D-II, 976-986 (2002).
- [2] 青木愉子他, “決定木を用いた F_0 パターン生成法が聴覚印象に与える効果,” 音講論 (春), pp. 181-182(2005).
- [3] I. H. Witten, E. Frank, *Data Mining: Practical machine learning tools and techniques, 2nd Edition*, (Morgan Kaufmann, San Francisco. 2005).
- [4] 渡辺健二他, “ F_0 パターンのモデル化—ダウンステップの効果—,” 音講論 (秋), pp. 389-340(2002).