

# 韻律の成分に着目した日本語学習支援システム

A Japanese study support system which paid its attention to the ingredient of prosody

中村 宏      秋葉 裕司      藤井 雅弘      伊丹 誠      伊藤 紘二  
Hiroshi NAKAMURA    Yuji AKIBA    Masahiro FUJII    Makoto ITAMI    Kohji ITOH

東京理科大学 基礎工学部 電子応用工学科

Department of Applied Electronics, Tokyo University of Science

<あらまし> 本研究では、日本語表現獲得学習における音声の貢献に着目し、学習例文に対して模範音声とその韻律の視覚提示、並びに構文情報の提示を行なう学習支援システムを提案する。提案システムでは、韻律の高低特徴である基本周波数からフレーズ成分とアクセント成分を分離し、音声認識エンジン「AmiVoice」を用いて文字対応を付けて視覚提示することによって、学習者へ文章に添った形での発話韻律情報の提供が可能となった。

<キーワード> 第二言語学習, 学習支援システム, 日本語教育, 音声教育, 高低アクセント, 韻律

## 1 はじめに

近年、第二言語学習の分野においてコンピュータを用いて学習を支援する (CALL: Computer-Assisted Language Learning) 研究が盛んになってきている。CALLは時間・場所・経済的制約が少なく、言語処理技術等の導入によって双方向性が実現すれば、高い学習効果が期待できる。

我々は第二言語学習における音声の役割を、発音や聴解の学習項目としてだけでなく、文字を伴う作文・読解においても、記憶に際して刺激を与え、表現獲得を互助するものとして考えている。

そこで音声言語学習支援とともに、読解・作文学習環境 [1] における表現獲得支援としての模範音声の再生及びその特徴の視覚提示の一手法を提案する。

## 2 韻律的特徴の表示

音声の韻律、特にマクロレベルに時間変化する韻律はその言語を大きく特徴付けており、例えば英語学習においては、等時的な強勢アクセントが作るリズムの習得が重要であると我々は考え、学習支援システムを構築してきた [2]。

対して日本語の韻律は高低変化を特徴として持つ。外国人学習者に対する教育現場では、音声教育の必要性は以前から唱えられているものの、未だ定期的かつ系統だった教育が一般化しているとは言いがたい。特に単音や単語の発音と比べ、ポーズやイントネーションといった文としての韻律的特徴 (プロソディ) の指導にあまり時間をかけられないのが現状である。そこで学習支援システムにおいて、音声とともに、プロソディの時間的変化を文の文字列と対応させて視覚表示していく。

ここでイントネーションやアクセントといった音声の高低特徴は、結果として現れた基本周波数 (以下  $F_0$ ) パターンのみの提示よりも、発音時にどこを意識して抑揚をつけるのかを文字言語情報とともに与えるほうが、より学習効果が高いと考えられる。

音声生成モデルの一つである藤崎モデルでは  $F_0$  パターンを、ゆっくりと起伏する土台の成分「フレーズ成分」とアクセント型と対応して比較的急速に上下する成分「アクセント成分」との和から成るものと考えている [3]。アクセント成分とフレーズ成分との別個の把握や、学習例文の構文情報との比較参照から、学習者への提示の着目点を明確化できる。

## 3 模範音声の処理

システムは学習例文のテキスト情報とともに、模範音声と韻律情報を保有する。韻律情報は  $F_0$  パターンからアクセント及びフレーズ成分を推定した結果を用いて、モーラ単位での特徴を分かりやすく視覚化したグラフを作成する [4]。曲線を区分するために、音声とテキスト情報は予め同期をとる。

### 3.1 音声とテキストとの同期

音声認識エンジン「AmiVoice」 [5] では、発話速度や韻律に依らない高精度な認識が実現されている。また、使用者のエンロールメントや事前学習の必要がないため、不特定な学習者を想定した場合においても利便性が高い。AmiVoiceでは、言語モデルとして事前に記述されたルールグラムに則って、入力音声の認識を行なう。そのため、学習に用いられる例文に従ったルールグラムを予め生成しておくことによって、語音誤りによる認識の失敗が抑えられる。

また AmiVoice では、音声認識単位「トークン」毎にそのタイムスタンプを取得できる。ここで、トークンをモーラ毎に設定すれば、入力音声を強制的に区分し、セグメンテーションすることが可能である。図 1 に時間情報出力の例を示す。

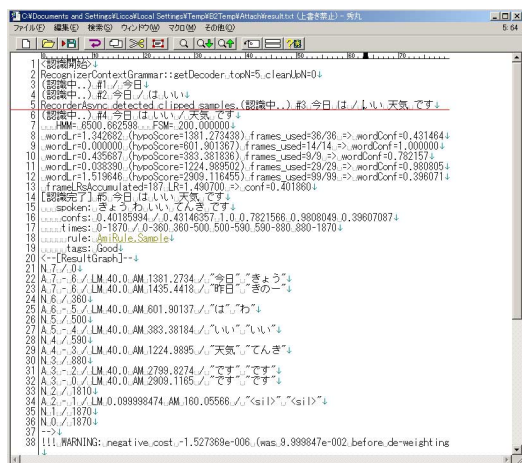


図 1: AmiVoice でのセグメンテーション結果

### 3.2 運動指令の推定

$F_0$  パターン中のフレーズ成分は、該当する韻律句・韻律節の分節の特徴にやや先行して始まり、一般的には上昇しながら最大値に達した後、緩やかに下降してピッチ最低値に漸近する。

まず、音声は 10kHz, 16bit で A / D 変換し可変時間窓長の自己相関関数法によって、10ms 間隔で基本周波数を抽出する。このままでは雑音や、ダブルピッチ・ハーフピッチといった gross error, microprosody が含まれているため、音声認識によるモーラ区分検出情報と組合せて修正し、さらに無声区間を補間して全体の平滑化を行なう。

こうして得られた平滑化  $F_0$  パターンを微分して、アクセント指令を得る。

アクセント指令から生成されるアクセント成分を先の平滑化  $F_0$  パターンから除去し、残差からフレーズ指令を推定し、さらに推定されたフレーズ成分を残差パターンから差し引いて積分を行ない、新たにフレーズ指令を得る [3]。この処理を反復して行なって得られたフレーズ指令とアクセント指令で逐次近似しながら、もとの音声の運動指令を推定していく。

アクセント指令は韻律上に、ほとんどずれずに現れるため、テキストに併せて学習者に提示することも可能だが、フレーズ指令は分節的特徴に先行するため、実位置を表示すると学習に悪影響を及ぼす可

能性がある。学習者へのフィードバックには、平滑化  $F_0$  パターン及びフレーズ成分・アクセント成分を分かりやすく表示したものを作成する。

### 3.3 構文情報との照合

ポーズ、即ちフレーズ成分の終端から次のフレーズ成分の開始位置までのブランクは、構文上での係り受け関係が希薄な語の間に生ずると考えられる。そこで、基本周波数の解析結果に、言語情報である構文木の係り受けとの関係を明示し提示する。

模範音声のポーズ位置を、音声認識による文字列との対応を用いて、係り受けと照合し表示することによって、学習者自身が係り受けとの関係を自ら汎化することができると思われる。

### 4 今後の課題

- 提案システムでの、フレーズ成分及び構文木表示による学習の有効性を検討。
- 学習に用いる例文の選定とデータベースの構築。
- 学習者音声の解析のための、各処理の自動化
- 作文・読解支援環境に組込んで、表現獲得のための音声の貢献を評価。

### 謝辞

広島大学教育学部の松崎寛先生と横浜国立大学人間科学部の河野俊之先生には音声教育の現場の視点から色々アドバイスを頂いた。ここに謝意を表する。

### 参考文献

- [1] 掛川淳一, 石川賢太郎, 海野俊介, 藤井雅弘, 伊丹誠, 伊藤紘二: “比較による表現獲得を支援する日本語読解・作文学習環境”, 言語処理学会第 10 回年次大会併設ワークショップ, e-Learning における自然言語処理 (2004).
- [2] 秋葉裕司, 中村宏, 小松崎聖, 掛川淳一, 藤井雅弘, 伊丹誠, 伊藤紘二: “リズム表示と音声の読み聞かせを組み合わせた英語表現獲得支援システム”, 教育システム情報学会第 29 回全国大会講演論文集, pp.101-102(2004).
- [3] H. Fujisaki and S. Narusawa: “Automatic extraction of model parameters from fundamental frequency contours of speech,” Proc. 2001 2nd Plenary Meeting and Symposium on Prosody and Speech Processing, pp. 133-138 (2002.1).
- [4] 松崎寛, 河野俊之: “発音指導法が変わる - プロソディーグラフを用いた音声教育”, 『月刊日本語 10 月号』 pp.26-33, アルク (1997).
- [5] 音声認識エンジン AmiVoice:  
株式会社 アドバンスド・メディア,  
<http://www.advanced-media.co.jp/amivoice/index.html>