

アクセント型を利用した音声入力補完方式の提案

大宮 広義[†] 木田 智史[†] 荒木 雅弘[†]

京都工芸繊維大学 工芸学部 電子情報工学科[†]

概要

我々は入力補完を行うディクテーションシステムを試作し、補完候補を最初の数音節から絞り込むためにアクセント型を利用する方法の開発を行った。本システムはアクセント型を用いて補完候補を絞り込むことによって、より入力がスムーズなディクテーションシステムを目指したものである。

1. はじめに

我々はこれまで、同音異義語の区別やアクセント句境界の検出を行う目的で、日本語におけるアクセント型の認識手法の開発を行ってきた。拍を等分割した点を中心とした frames 単位に対数基本周波数とその一次、二次差分をパラメータとし、HMM を用いてアクセント句の認識実験を行った結果、前後の環境(高、低、無音)とアクセント型(0 型、1 型、N 型)を組み合わせた 28 カテゴリーでの認識精度は約 49%(オープン)であり、中心のアクセント型のみに着目した 3 カテゴリーの分類で約 66%(オープン)であった[1]。

本研究では、このアクセント型認識手法をベースに、アクセント型を利用した音声認識アプリケーションの開発を行うことを研究目的とした。対象のアプリケーションとして、入力補完を行うディクテーションシステムを試作し、補完候補を最初の数音節から絞り込むためにアクセント型を利用する方法の開発を試みた。

韻律情報を用いて音声インタフェースの機能を向上させる試みとして、後藤らは有声休止をトリガーとして音声補完を行う方式を提案している[2]。後藤らの手法は単語の補完が対象であり、ディクテーションに応用する方法が考慮されていない。そこで我々は、キーボード入力された部分文字列から曖昧検索を行い入力候補を提示するシステムと、音声認識とを結合し音声補完方式をディクテーションに適用できるように拡張した。

また音声インタフェースの機能向上に韻律情報を用いる他の方式として、高音を用いる方式[3]が提案されているが、この方式は音声入力におけるモー

ドの切り替え(通常文字入力と改行などのコマンド入力)に用いることを意図しており、自然な流れで入力補完時にのみ音高を変化させて入力するのは難しい。そこで、我々は通常入力時に自然に用いることができる無声休止(一定時間以上のポーズ)を音声補完のトリガーとする。この無声休止前の音韻とアクセント型を認識することによって、文末であれば音声補完を起動しないように工夫することが可能である。

2. 音声入力補完の方式設計

我々は、認識途中の音韻列から補完単語集合を得る方法として、キーボード入力に対して予測補完を行うシステムである POBox (Predictive Operation Based On eXample)を利用する[4]。POBox は読みやストロークを入力するたびにインクリメンタルに辞書の曖昧検索を行い、検索された候補単語から必要な単語を選択することにより単語を入力していく入力補完システムである。現在、PDA や携帯電話で広く用いられ、電子メールの入力などにその威力を発揮している。我々は将来的にはこのような携帯端末

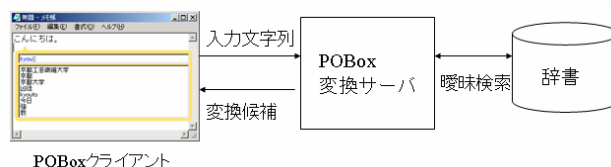


図1 POBox のシステム構成

における音声入力を補助する目的で、韻律情報を用いた音声インタフェースの高度化を目指している。

POBox での補完はクライアントーサーバ方式で行われている(図1)。クライアントはユーザからの文字入力を受け取り、エディタ等に入力及び補完された文字列を渡すものである。この補完候補の取得のためにサーバと通信し、サーバは辞書に対して曖昧検索を行い、検索結果を次回以降の候補提示順序に反映させる方式で学習を行う。

我々は、POBox クライアントの部分に音声入力が可能になるように実装した。その際に、変換候補の提示順序をアクセント型情報を用いて変更するようにし、通常のキーボード入力よりも絞り込み精度を高くすることによって、機能向上を図る。

音声入力には Julius[5]を用い、認識結果をローマ字出力させたものを整形して POBox 変換サーバへ送ることによって、変換候補集合を得ている。この変換候補集合は音韻情報のみから得られたものである。ここに韻律情報を統合し、変換候補の再ソートを行って、ヒット率を高める手法の実現を試みる。アクセント情報認識処理の流れを図 2 に示す。

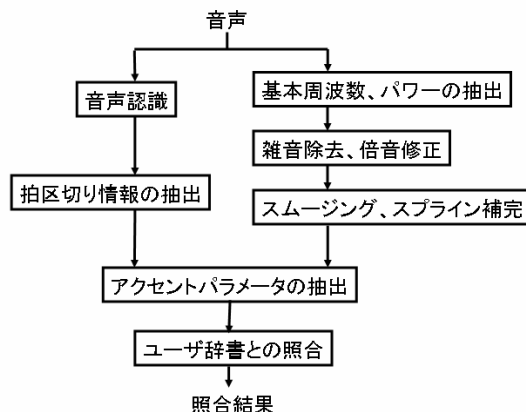


図 2 アクセント情報認識処理の流れ

入力された音声は Julius を用いて認識され、認識結果の音韻アライメントから拍区切り情報を得る。一方、韻律情報に関しては、まず自己相関法を用いて基本周波数を求め、雑音の除去、倍音・半倍音の修正を行った後、メディアンフィルタおよび最小二乗法によるスムージングを行う。さらに無声音の部分に関しては 3 次のスプライン補完を行い、広い範囲で滑らかに基本周波数の変化を観測できるようにした。

また、パワーは各フレームにおける振幅の二乗平均対数値に 10 を掛けて得た。

このようにして得られたアクセント情報パラメータを用いて、事前に入力されたユーザ履歴と照合して、補完候補の再ソートを行う。

予備実験として、最初の 2 拍が/高低/、/低高/、/高高/の 3 パターンがどれだけの特徴次元数で識別可能であるかを調査した。それぞれのアクセント型に対する複数の単語を関西方言話者 10 名に発話してもらい、韻律パラメータを調査したところ、各拍を 3 等分して得られた観測点 6 箇所に関して、基本周波数の差分とパワーの差分の組み合わせでほぼ識別可能であることがわかった。

3. 音声補完アプリケーションの実装

テスト用クライアントの実装は Java で行い、Julius は Windows 版を用いた。また、韻律情報の抽出には ATR Speech Tools を用いた。変換候補の取得には POBox のサーバを用いた。

クライアントの画面表示例を図 3 に示す。

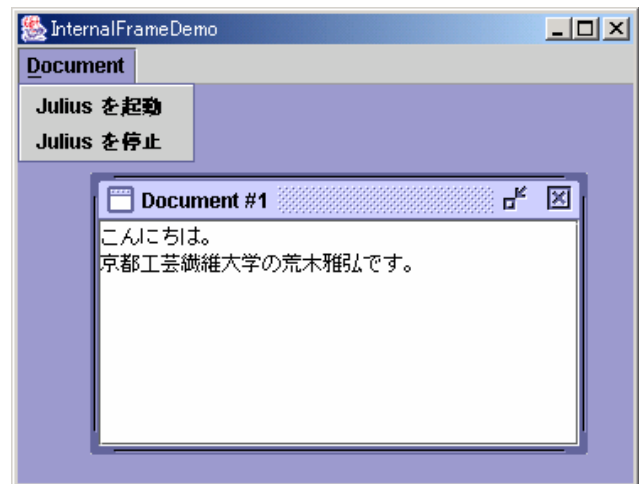


図 3 音声補完クライアントの動作例

4. 音声補完アプリケーションの実装

今後、特定話者を対象とした電子メール入力クライアントを実装し、一定期間仮想的なアカウントで辞書をパーソナライズした後に、韻律情報による変換候補の再ソートの効果を評価する。

また、現在は全ての無声休止が補完のトリガーとして取り扱っているが、無声休止前の音韻情報と韻律情報から文末であるかどうかの判定を行い、文末でなければ補完を起動する方式の実現を試みる予定である。

参考文献

- [1] 木下育子, 西本卓也, 荒木雅弘, 新美康永: 隠れマルコフモデルを用いたアクセント型の認識, 信学技報, SP2001-140, pp.37-42, 2000
- [2] 後藤真孝, 伊藤克亘, 速水悟: 音声補完: TAB on Speech, 情報処理学会研究報告, 2000-SLP-32-16, 2000.
- [3] 尾本幸宏, 後藤真孝, 伊藤克亘, 小林哲則: 音声シフト: "SHIFT" on Speech, 情報処理学会研究報告, 2002-SLP-40-3, 2002.
- [4] Toshiyuki Masui. POBox: An Efficient Text Input Method for Handheld and Ubiquitous Computers. In Proceedings of the International Symposium on Handheld and Ubiquitous Computing (HUC'99), pp. 289-300, 1999.
- [5] 河原達也他: 連続音声認識コンソーシアム 2001 年度版ソフトウェアの概要, 2002-SLP-43-3, 2002.