

# 音声訂正：選択するだけで訂正可能な音声入力インタフェースとその評価\*

緒方 淳 後藤 真孝 (産総研)

## 1 はじめに

音声認識を日常的に使えるインタフェースにするためには、必ずどこかで生じてしまう誤認識を容易に訂正できる音声入力インタフェースが不可欠となる。本稿では、ユーザが認識誤りを選択操作だけで訂正することを可能にする「音声訂正」という音声入力インタフェース機能とその評価について述べる。

## 2 音声訂正インタフェース

「音声訂正」とは、音声認識器により引き起こされた誤認識を、ユーザとのインタラクションを介して訂正する新しい音声入力インタフェースである。以下では、音声訂正により提供される主な3つの機能について概観する(詳細な実装方法については文献[1]-[3]に委ねて省略する。)

### 2.1 音声訂正の基本機能

図1に音声訂正インタフェースの画面表示の模式図を示す。音声訂正では、従来の音声認識と異なり、最上段の通常の認識結果(単語列)に加えて、その下へ「競合候補」のリストを常に表示する。通常の認識結果は各単語の区間ごとに区切られて、その単語に対する競合候補が整列して表示される。ここで、競合候補の個数はその区間の曖昧さを反映しており、ユーザは候補が多いところに誤認識がありそうだと思って、注意深く見ることが出来る。逆に、認識結果として信頼性が高い区間は候補が少ないため、ユーザに余計な混乱を与えることが少ない。このように認識結果が提示されると、ユーザは競合候補の中から正解を「選択」する操作だけで、容易に認識誤りを訂正できる。

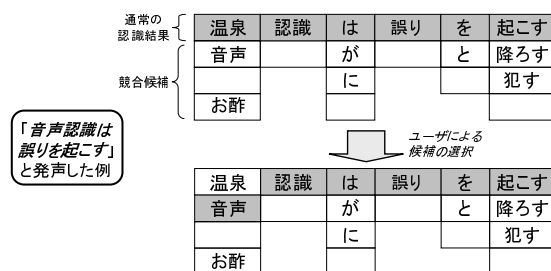


図 1: 音声訂正インタフェースの模式図

なお、図1のように、選択肢には必ず空白の候補が含まれる。これを「スキップ候補」と呼び、その候補が属する区間の認識結果をないものとする役割を持つ。これにより、最上段の認識結果に湧き出し誤りが存在しても、ユーザはスキップ候補を選択するだけで容易に削除できる。つまり単語の置き換えと削除が「選択」という一つの操作でシームレスに実行できる。また、各区間の競合候補は、上から可能性(存在確率)の高い順に並んでいるため、通常はユーザが上から下へ候補を見ていくと、早く正解にたどり着けるようになっていく。

しかし、従来、大語彙を対象とした連続音声認識では、競合候補を生成するための中間結果は非常に大規模なものとなるため、このようなシンプルな訂正イン

タフェースは実現されていなかった。それに対し音声訂正では、「confusion network[4]」を、誤り訂正インタフェースへと応用することにより、大語彙、小語彙を問わず多様な入力音声に対して上述のような効果的な候補の提示、訂正を可能にした[1]-[3]。

### 2.2 即時誤り訂正機能

使いやすいインタフェースを構築するには、ユーザの入力中に逐次現在の認識状態をフィードバックすることが重要となる。音声訂正では、ユーザに対するフィードバックを通じて、誤り訂正作業を効率的に行うことを目的とした「即時誤り訂正機能」を提供する。これは、発話中に認識の中間結果を競合候補付きでリアルタイムにフィードバックし続け、さらにユーザの選択も可能にすることで、発声の最中においても、あるいは、認識器が認識結果を確定していない段階(認識処理最中)においても、誤りを即時に訂正することを可能にする機能である。

### 2.3 発話中休止機能

前節の即時誤り訂正機能を使っていると、発話中に正しい候補を選択している間、音声認識器に一時的に続きを言うのを待って欲しくなる場面が出てくる。そこで、音声訂正では、発話中にユーザが意図した時点で、認識処理を一時停止させる「発話中休止機能」を提供する。そして次の発話が始まると、あたかも一時停止前の発話が続けていたかのように動作させる。このユーザの一時停止の意図を伝えるために、音声の中の非言語情報の1つである有声休止[5](語中の任意の母音の引き延ばし)を、発話中休止機能のトリガーとして採用した。これにより、ユーザは自然に一時停止をかけて、正しい候補を選択したり、続きの発話を考えたりできる。

## 3 評価実験

音声訂正の有効性を確認するために、文の入力をタスクとした被験者実験を行った。本実験での調査ポイントは、(1) 音声訂正により、通常の音声入力(最上単語列のみを出力する音声認識)に比べて、文入力ごとの程度効率的になるか、(2) 発話中休止機能が文入力の際に有効に働くかどうか、(3) システム全体に対してどのような印象を受けたか、の主な3点である。

今回の実験では、評価対象により幾つかの課題を行ったが、全体的な実験方法としては、提示した文を被験者に入力してもらい、一字一句間違えずに入力できた時点で1つの文を入力完了とした。また、通常の音声入力の際に認識誤りが発生した場合や、音声訂正で競合候補中に本来の正解がなくて選択できない場合には、キーボード、マウスを利用してタイプ入力にて訂正することとした。各課題の3つの文は、全被験者を通して共通だが、順番をランダムに変えたものを6通り用意して各被験者に割り当てた。

音声を用いた実験では、音声認識単体の基本性能差による影響を排除するために、認識デコーダ、音響モデル、言語モデルに関しては、通常の音声入力、音声訂正の各実験ともに同一のものを用いた。本実験には、音声入力ソフトウェアを普段利用していない、20代の25名の被験者(男性14名、女性11名)が参加した。

\*Speech Repair: New Speech Input Interface Capable of Repairing Recognition Errors Just by Using Selection Operation and Its Evaluation.  
by Jun OGATA, Masataka GOTO (AIST)

表 1: 認識率の比較

	認識率 (%)
通常音声入力	86.12
音声訂正 (訂正前)	85.03
音声訂正 (訂正後)	97.70

### 3.1 実験方法

まず、音声訂正と通常の音声入力の比較を目的とした課題として、新聞記事から選定した3つの文を、各入力手法を使用して入力した。被験者は、課題を行う直前に、通常の音声入力、音声訂正(基本機能と即時誤り訂正機能)についての説明を受け、それぞれの入力手段について、上記の3文とは別の2文を用いて練習を行った。また、各入力手法の順番は、被験者ごとに交互に変わるように設定した。本課題の終了後には、音声訂正の基本機能、即時誤り訂正機能それぞれについてのアンケート(6項目7段階評価)を実施した。

次に、発話中休止機能の評価を目的とした課題を行った。本研究で提案した発話中休止機能は、2.3節で述べたように、ユーザが発声中に音声認識処理を意図的に中断させることが可能となるものである。このような機能の実際の利用状況として、我々は、自発的に発声された音声に対する入力インタフェースを想定しており、そのため、休止状態にするためのトリガーとしては言い淀みを採用している。したがって、発話中休止機能の評価を目的とした本課題においては、被験者が自発的に発声する状況になるべく近くなるように、頭の中にある文を発声して入力している場面を想定し、以下のような手順で実験を行った。

まず、発話中休止機能についての説明を行い、その後被験者は実際に本機能を練習した。次に、本課題についての説明を行った後、発話中休止機能を使うかどうかは被験者の自由という条件で、最初の課題とは別の3文の入力を行った。ただし、被験者には、各々の文を入力する前に、提示した入力対象文を記憶して(頭の中に入れて)もらった。このとき、記憶に要する時間は特に制限せず、被験者の自由とした。文を入力する際の制約として、発声している最中には、入力対象文は確認できないこととした。ただし、発話中休止機能を使って休止状態にし、発声がなされていないときには、入力対象文を確認することは可能とした。本課題の終了後には、発話中休止機能について、また音声訂正全般についてのアンケートを実施した。

### 3.2 実験結果

表1に、通常音声入力、音声訂正の各認識率を示す。表より、これまでに報告した音声データベース(JNAS, CSJ)での結果[1]-[3]と同様に、実際の被験者実験においても、音声訂正はほとんどの誤りを訂正可能とし、高い訂正能力を示した。また、音声訂正では、通常の音声入力に比べて約31%の入力にかかる所要時間を削減できていた(通常音声入力:平均14.41秒, 音声訂正:平均9.94秒)。なお、同じ3文をキーボードで入力したときの平均入力時間は23.65秒であった(ただし、あくまでも訂正の際のタイプ入力の練習を兼ねて本課題の直前に実施したものであり、入力時間に関する厳密な比較対象にはならない)。以上の結果より、音声訂正は優れた訂正能力を持ち、より効率的に文入力が可能であることがわかる。発話中休止機能の評価のための課題では、本機能が使用されたのは、全被験者の全発声のうちで約61%(46/75)であった。また、本課題の3文を入力するにあたり、発話中休止機能を全く使用しなかったのは全25名のうち4名であり、今回実験に参加した多くの被験者が本機能を利用していたことがわかった。

図2に音声訂正使用後のアンケート結果(各項目とも-3~+3の7段階尺度)を示す。(A)の音声訂正の基

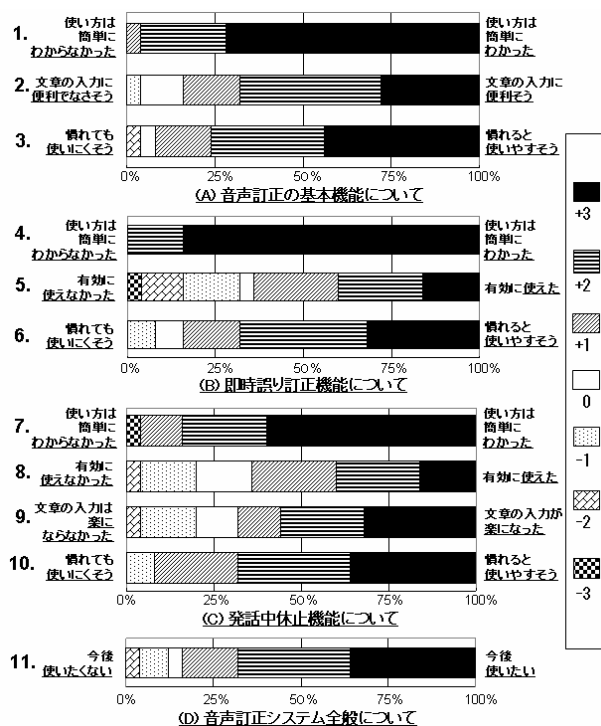


図 2: アンケートの集計結果

本機能については、項目1, 2, 3.ともに評価値が高く、候補を選択するだけで誤り訂正することの有効性、効率性が示された。(B)の即時誤り訂正機能については、項目5.が他に比べると低い評価値となった。これは、発声途中から逐次表示される候補をチェックし、訂正処理を行うことに対して、被験者が一定の負担や慌ただしさを感じたためであるが、ある程度慣れることにより有効に使用できそうであるという意見は多かった(項目6.)。(C)の発話中休止機能についても、(B)と同様に機能を有効に使えたか(項目8.)、という点について評価値が低かった。この理由としては、今回の実験において、被験者が「言い淀む」ことに対して一定の抵抗を感じたことが挙げられる。特に女性にそのような意見が多かった。有効に使えた被験者においては、比較的長い文に対する入力が格段に楽になったとの意見を得た(項目9.)。最後に、項目11.の音声訂正システム全般については、8割以上の被験者が今後使いたいとの意見を得られ、以上のアンケート結果からも本インタフェースの有効性が確認できた。

### 4 まとめ

本稿では、音声認識による認識誤りをユーザによって効率的かつ容易に訂正できる「音声訂正」という音声入力インタフェース機能と、25名の被験者による評価実験について報告した。実験の結果、音声訂正により文入力が効率化され、被験者にとって今後も使いたいと思われるインタフェースであることがわかった。

今後は、未知語への対処を行い、より効率的な誤り訂正処理について検討する予定である。本研究は「音声補完シリーズ」[6]の第5弾に位置付けられるが、これから言い淀み以外の非言語情報も積極的に取り入れ、音声ならではの機能を持った、さらに使いやすい音声入力インタフェースを実現していきたいと考えている。

### 参考文献

- [1] 緒方, 後藤: WISS'2004 論文集, pp.47-52, 2004.
- [2] 緒方, 後藤: 情処研報 2004-SLP-54-54, pp.319-324, 2004.
- [3] J.Ogata, M.Goto: EuroSpeech'2005, (to appear).
- [4] L.Mangu, et al: CSL, Vol.14, No.4, pp.373-400, 2000.
- [5] 後藤 他: 信学論, Vol.J83-D-II, No.11, pp.2330-2340, 2000.
- [6] 後藤: 情処研報 2004-SLP-52-7, pp.41-46, 2004.