社団法人 情報処理学会 研究報告
IPSJ SIG Technical Report

2005－SLP－57 (1)
2005／7／15

音声訂正の評価

緒方 淳　　後藤 真孝

産業技術総合研究所

{jun.ogata,m.goto}@aist.go.jp

あらまし　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．音声訂正では，ユーザが音声入力を開始すると，認識結果を単語ごとに区切った表示と，区切られた各区間に対する他候補 (競合候補) が発話の最中から次々と画面に描画され，ユーザが競合候補の中から本来の正解を選択するだけで認識誤りを訂正可能にした．また，音声訂正では，ユーザが発声中であっても訂正処理が可能な「即時誤り訂正機能」と，ユーザが意図的に発声を休止し，認識処理を一時中断させることが可能な「発話中休止機能」を実現した．25 名の被験者による評価実験を行ったところ，音声訂正は使いやすく，効果的な音声入力インタフェースであることが確認された．

Evaluation of Speech Repair

Jun Ogata　　Masataka Goto

National Institute of Advanced Industrial Science and Technology (AIST)

Abstract　In this paper, we describe a speech input interface function, called "*Speech Repair*", which enables a user to easily correct recognition errors by selecting candidates. During the speech input, this function displays not only the typical speech-recognition result but also other competitive candidates. Each word in the result is separated by line segments and accompanied by other word candidates. A user who finds a recognition error can simply select the correct word from the candidates for that temporal region. Moreover, we introduced two additional functions: *immediate correction function* that enables the user to correct errors not only when the recognition process is complete but also whenever the user finds erroneous words, and *intentional suspension function* that enables the user to intentionally suspend and resume the recognition process. Experimental results with twenty-five subjects showed that the speech-repair function is easy to use and effective interface.

1　まえがき

計算機による音声認識は，認識誤りを避けることはできない．音声認識技術を改良してどんなに認識率を上げていったとしもて，人間にとって，常に明瞭で曖昧性のない発声をし続けることは極めて困難である以上，認識率は決して 100%にはならない．したがって，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．

そこで我々は，音声認識による認識誤りを，ユーザがより効率的で容易に訂正できる新たな音声入力インタフェース「音声訂正」を提案した [1]-[3]．音声訂正では，ユーザが音声入力を開始すると，認識結果を単語ごとに区切った表示が発話の最中から次々と画面に描画される．同時に，区切られた各区間の他候補 (競合候補) も常に列挙されていく．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，音声認識結果として信頼性が低い箇所ほど，多数の候補が表示される．ユーザはそれを見ながら，発話中あるいは発話終了後に正しい候補を選択するだけで訂正ができる．ここで重要なのは，わざわざユーザが誤認識箇所を発見して指摘しなくても，常に競合候補がリアルタイムにフィードバックされ続けていることである．これにより，従来研究のように誤認識箇所の発見，指摘，提示された候補の判断，選択といった手間をかけずに，いきなり候補を見て選択するだけで，効率良く訂正できる．さらに，こうして発話の最中に候補を選べるようになると，選択操作の間，音声認識器に一時的に待って欲しくなることがある．そこで，単に発話中に有声休止 (語中の任意の母音の引き延ばし) で言い淀むだけで発話を中断可能とし、その次の発話はあたかも中断前の発話が続いていたかのように入力できるようにした。

本稿では，これまでに提案した音声訂正機能につい

て述べ，音声訂正の有効性を確認する評価実験の結果を中心に報告する．以下，2章において音声訂正の概要とその機能について述べ，3章で音声訂正機能を持つ音声入力インタフェースを紹介する．次に，4章において音声訂正インタフェースの有効性を確認するための評価実験とその結果について述べる．最後に5章でまとめを述べる．

2 音声訂正

「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．一般的に，音声入力インタフェースにおいては，発話終了後にユーザが認識誤りを訂正するためには，主に以下の2つの手続きを取る必要がある．

1. 認識結果の中から誤り箇所を探して指摘する．
2. 指摘した誤り箇所を訂正する．

音声訂正では，これらを一度の操作で効率的に行うことで，認識誤りを訂正する際のユーザへの負担を減らすことを目的としている．以下では，音声訂正により提供される主な3つの機能について説明する．なお，本稿では主に各機能の概要について述べ，詳細な実装方法については文献[1],[2]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，ユーザの発声が入力されると，図1上側に示すような結果が，音声入力開始と共に左から右へ順次表示されていく．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．図1のように，通常の認識結果が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，音声認識結果として信頼性の低い箇所ほど，多数の候補が表示される．そのため，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できる．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．

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(本来あるべきでない区間に余分な単語が挿入される誤り)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．つまり，上の方ほど音声認識結果として信頼性が高い候補であるので，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る．

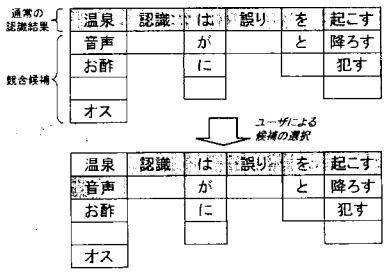


図1: 選択するだけで誤りの訂正ができる音声訂正インタフェース(「音声認識は誤りを起こす」という発声が誤認識された例)

以上の音声訂正の基本機能はシンプルだが，従来こうしたインタフェースは実現されていなかった．その理由としては，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(例えば単語グラフ，N-best文リスト)は非常に大規模なものとなり，図1のような効率的な候補提示が困難だったからである．それに対し音声訂正では，大規模な単語グラフを効率的な表現形式に圧縮した「confusion network[5]」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，このような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，音声認識器が最終的な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

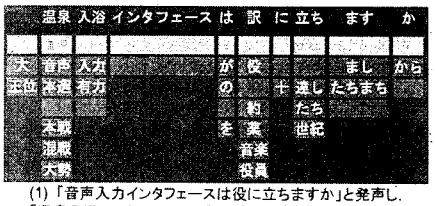
前節の即時誤り訂正機能を使っていると，発話中に正しい候補を選択している間，音声認識器に一時的に続きを言うのを待って欲しくなる場面が出てくる．しかし，通常の音声認識器による認識単位は，無音で区切られた一息で言える区間なので，むやみに発声を中断するとうまく認識されない問題があった．

これに対し，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い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，音声中の非言語情報の1つである有声休止 [6](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．有声休止は，人間同士の日常的な会話においては頻繁に発生し，人間にとってごく自然な行為といえる．そのため，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 音声訂正機能付き音声入力インタフェース

本研究にて構築した音声訂正機能付き音声入力インタフェースの動作状況として，図2に発話中休止機能を利用しない場合の表示画面を，図3に発話中休止機能を利用した場合の表示画面をそれぞれ示す．図1に相当する表示部分 (「候補表示部」と呼ぶ) の上に，さらに一行追加されているが，これは，候補を選択して訂正した後の最終的な音声入力結果を表示している．候補表示部では，現在選択されている単語の背景が着色される．何も選択していない状態では，候補表示部の最上段の最尤単語列が選択されている．ユーザが他の候補をクリックして選択すると，その候補の背景が着色されるだけでなく，画面最上部の最終的な音声入力結果も書き換えられる (選択操作で訂正した箇所だけ，文字の色を変えてわかりやすく表示している)．また，構築した音声訂正インタフェースでは，ユーザが選択した候補の時間情報や言語的な確率をもとに，その両隣の未選択候補に対する自動訂正機能も実装されている [2](例えば図2の (2))．

以上の機能を持つ音声訂正インタフェースは，訂正のための競合候補を生成する音声認識部，発話中休止機能のための有声休止を検出する有声休止検出，訂正処理等のインタフェース全体の状態管理を行うインタフェース管理部の主に3つの構成要素により実現される．これらは，効率的に負荷分散が可能で，拡張性が高くなるように，ネットワークプロトコル RVCP (Remote Voice Control Protocol)[7] を用いて複数のプロセス群として実装されている．



(1)「音声入力インタフェースは役に立ちますか」と発声し，「温泉入浴インタフェースは訳に立ちますか」と認識された．

(2) 競合候補を選択することで，誤りを訂正．この場合，ユーザはたった2回クリックするだけで全誤りを訂正できた(「入力」は「音声」を選択したときに自動修正された)．

図 2: 発話中休止機能を利用しない場合の画面表示例 (「音声入力インタフェースは役に立ちますか」という文を発声)

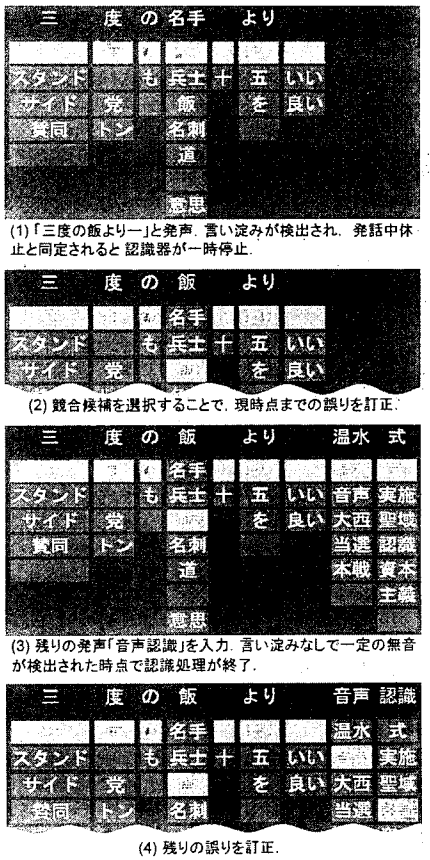
4 評価実験

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により，以下の点について調査を行った．

- 音声訂正により，通常の音声入力 (最尤単語列のみを出力する音声認識) に比べて，文入力がどの程度効率的になるか．
- 比較的長い文やうろ覚えの文を入力する際に，発話中休止機能が有効に働くかどうか．
- 音声訂正インタフェース全体に対してどのような印象を受けたか．

今回の実験では，次節で述べるように，評価の目的に応じた3つの課題を被験者に対して行ったが，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．

本実験で使用した音声認識システムは，音響モデルには，新聞記事読み上げコーパス JNAS から学習した triphone モデルを，言語モデルには，CSRC ソフトウェア 2000 年度版 [8] の中から，新聞記事テキストより学



(1)「三度の飯より一」と発声．言い淀みが検出され．発話中休止と同定されると 認識器が一時停止．

(2) 競合候補を選択することで．現時点までの誤りを訂正．

(3) 残りの発声「音声認識」を入力．言い淀みなしで一定の無音が検出された時点で認識処理が終了．

(4) 残りの誤りを訂正．

図 3: 発話中休止機能を利用した場合の画面表示例 (「三度の飯より音声認識」という文を発声)

習された 20000 語の bigram をそれぞれ用いた．また，本インタフェースでは，音声認識器の認識アルゴリズムとして，back-off 制約 N-best 探索手法 [9] を用いることで，リアルタイムに競合候補を生成，提示することが可能となっている．なお，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．すなわち，本実験における通常の音声入力による認識結果は，音声訂正の競合候補の中の最尤単語列 (最上部に表示される単語列) のみを表示した場合に相当する．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．

4.1 実験方法

音声訂正の機能の有効性を確認するため，各被験者に対して，以下の 3 つの課題を順に実施した．

課題 1

まず，キーボードとマウスを利用して，以下の 3 つの文を入力した．

1. これらの地域の中学では非行が少ない
2. この病院のベッド数は十五床です
3. お年寄りからの注文にも備え二千個分の材料を確保

なお，課題 1 は，以降の音声による文入力の際の訂正処理を，被験者がスムーズに行えるよう，本実験でのキーボード，マウスを利用した日本語入力環境 (かな漢字変換) に慣れることを主な目的として実施した．

課題 2

次に，音声を利用した文入力として，上記の 3 文を以下の条件でそれぞれ入力した．

(1) 通常の音声入力を使用して文入力
(2) 音声訂正を使用して文入力

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．ただし，本課題では，音声訂正の文入力に対する基本性能，効率性を評価するため，音声訂正の発話中休止機能については被験者に対して一切説明はしなかった (あくまでも被験者が発話中休止機能について知らされていないだけであり，機能自体はシステムに含まれていた)．また，課題 2 において，(1) と (2) の条件の順番は，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．

課題 3

最後に，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．このような機能は，例えば，ユーザが一息では発声しきれないような比較的長い文を入力する際などに特に有効に働くと考えられる．また，我々は，本機能の実際の利用状況として，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

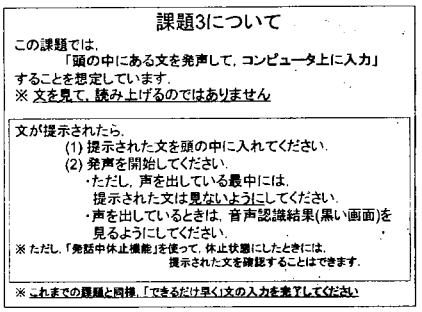


図 4: 課題 3 における被験者に対する指示文

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 (実際には，図 4 に示す指示文を被験者に提示)，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，以下の 3 つの文の入力を行った．

1. これから先の高いレベルを目指すにはもう少し読みの裏付けが必要だ
2. 九十四年度当初は予算と同規模で景気対策を理由に二年連続で高水準となった
3. 全国のコンビニエンスストア 書店 私鉄 地下鉄の駅などで販売 年間講読者への宅配も行う予定

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．また，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 (5 項目 7 段階評価) を実施した．

4.2 実験結果

表 1 に，課題 2 における，通常音声入力を使用したときの認識率，音声訂正を使用したときの訂正前・訂正後の認識率をそれぞれ示す．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，課題 2 における各入力手段ごとの，文を入力し終えるまでの 1 文あたりの平均入力時間を表 2 に示す．なお，課題 1 におけるキーボードでの平均入力時間は 23.65 秒であった (ただし，課題

表 1: 課題 2 における認識率

	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

表 2: 各入力手段の平均入力時間

	平均入力時間 (sec.)
通常音声入力	14.41
音声訂正	9.94

1 ではタイプ入力する練習も兼ねているので，入力時間に関する厳密な比較対象にはならない)．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．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．また，課題 2 の実験に際し，音声訂正に関して，被験者はたかだか 2 つの文を練習として入力しただけであり，これより音声訂正の基本機能は，入念な練習は不要で，インタフェースとしても直観的であったといえる．課題 3 の実験では，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，課題 3 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 5 に音声訂正使用後のアンケート結果を示す．図中，(A) は音声訂正の基本機能について，(B) は即時誤り訂正機能について，(C) は発話中休止機能について，図 (D) は音声訂正システム全般についてのアンケート結果となっており，各項目とも-3〜+3 の 7 段階尺度での評定を行った．評定値が全体的に高い傾向にあったのは，項目 1., 4., 7. で，音声訂正，各機能は直観的でわかりやすいインタフェースであったことがわかる．(A) の音声訂正の基本機能については，項目 1., 2., 3. ともに評定値が高く，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定値となった．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．しかしながら，この点に関しては，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.)．また，今回の実験においては，即時誤り訂正機能により，多くの被験者は発声をしながらの訂正処理は難しいながらも，発声が終了した瞬間に，認識処理が実行中であっても即座に訂正処理に移行でき，最終的な文の入力を比較的早く完了させることができていた．(C) の発話中休止機能についても，(B) と同様

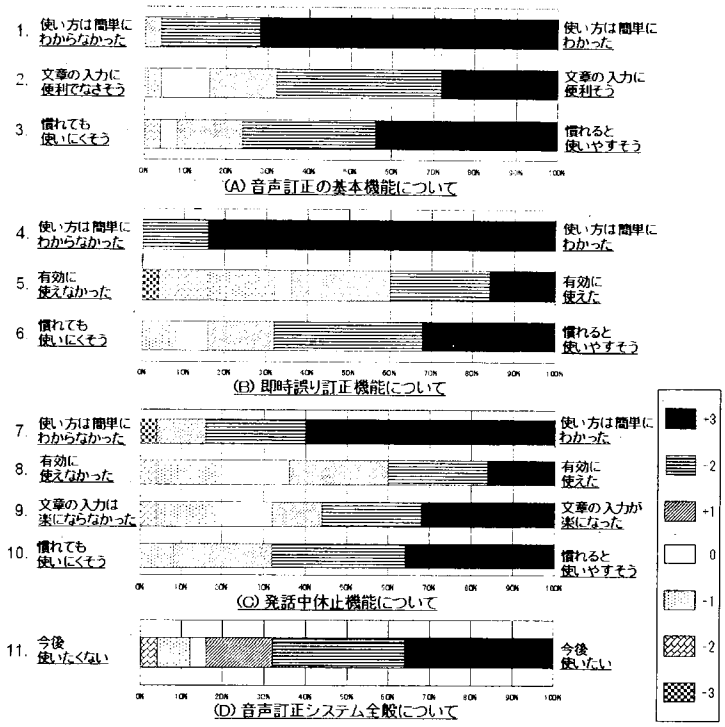


図 5: アンケートの集計結果

に機能を有効に使えたか (項目 8.)，という点について評定値が低かった．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，本実験のような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．

5 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」[10] 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: “音声訂正: 認識誤りを選択操作だけで訂正ができる新たな音声入力インタフェース”, WISS'2004 論文集, pp.47-52, 2004.

[2] 緒方，後藤: “音声訂正 : “CHOICE” on Speech”, 情処研報 2004-SLP-54-54, pp.319-324, 2004.

[3] J.Ogata, M.Goto: “Speech Repair: Quick Error Correction Just by Using Selection Operation for Speech Input Interfaces”, EuroSpeech'2005, (to appear).

[4] C-M.Karat, et al: “Patterns of Entry and Correction in Large Vocabulary Continuous Speech Recognition Systems”, Proc. CHI'99, pp.568-575, 1999.

[5] L.Mangu, et al: “Finding Consensus in Speech Recognition: Word Error Minimization and Other Applications of Confusion Network” Computer Speech and Language,Vol.14, No.4, pp.373-400, 2000.

[6] 後藤 他 : “自然発話中の有声休止箇所のリアルタイム検出システム”, 信学論, Vol.J83-D-II, No.11, pp.2330-2340, 2000.

[7] 後藤 他. “音声補完: 音声入力インタフェースへの新しいモダリティの導入,” コンピュータソフトウェア, Vol.19, No.4, pp.10-21, 2002.

[8] 河原 他: “連続音声認識コンソーシアム 2000 年度版ソフトウエアの概要と評価”, 情処研報, 2001-SLP-38-6, 2001.

[9] 緒方，有木: “大語彙連続音声認識における最ゆう単語 back-off 接続を用いた効率的な N-best 探索法”, 信学論, Vol.84-D-II, No.12, pp.2489-2500, 2001.

[10] 後藤: “非言語情報を活用した音声インタフェース”, 情処研報 2004-SLP-52-7, pp.41-46, 2004.