

指さし訂正を伴う音声入力における候補提示手法の検討

池田 雄司[†] 梅澤 猛[†] 大澤 範高[†]

千葉大学大学院融合科学研究科[†]

1. はじめに

音声認識の誤りを完全になくすことは困難である。そのため、正確な文章入力に音声認識を用いる場合は、誤認識を訂正できるインターフェースが必要である。そこで、入力文章を発話した際に複数の単語の認識候補を提示し、その中から指さしによって意図した組み合わせを選択する文章入力手法に着目した。本研究では、空間効率がよく、効率的な入力可能な認識候補の提示手法を検討する。

2. 関連研究

提案する文章入力手法では音声認識結果の大量の候補から容易に適切な候補を選択できるように提示する必要がある。遠藤らの手法[1]では、音声認識における中間結果である単語トレリスを用いて単語の候補の認識時間を参照し、単語の境界を訂正する機能によって誤認識を訂正している。また、緒方らの手法[2]では、ユーザに音声認識結果の単語の候補を提示し、意図した単語の候補を選択するだけで訂正ができる。

しかし、これらの手法では、表示領域に単語の候補を表示する際の空間効率について考慮されておらず、表示すべき候補が増加した場合に1つの画面上に効率的に候補を提示できない場合がある。また、認識システムの出力の信頼度の高い単語候補のみで訂正が完結する場合の作業効率の向上が十分には考慮されていない。

3. 提案手法

提案手法では従来手法では考慮されていなかった以下の項目に着目し、改善をめざす。

- (1) 選択される可能性の高い単語候補の選択
- (2) 領域の縦方向が制限される場合の表示
- (3) 信頼度の高い単語候補で完結する場合の効率

3.1 単語候補の表示領域の大きさに対する処理

項目(1)を考慮し、ユーザが単語の候補を選択し易くするために、各単語候補の表示領域の高さを変える。そのために図1(a)のような音声認識の中間結果の単語候補のつながりを表すラティス構造を基にする。図1(a)の左端の空白領域を音声認識結果の発話開始記号とする。

H_i を単語候補*i*の表示領域の高さとし、 c_i を単語候補*i*と隣接する後続の単語候補の個数、集合 S_i を候補*i*と隣接し先行する単語候補の集合としたとき、 H_0 から隣接する後続の単語候補に高さを均等に分配する。次のように表せる。

$$H_0 = \text{Height}_{\max}$$

$$H_i = \sum_{j \in S_i} \frac{H_j}{c_j} \quad (i \neq 0)$$

ここで Height_{\max} は表示領域の縦の上限である。この処理を適用した結果を図1(b)に示す。

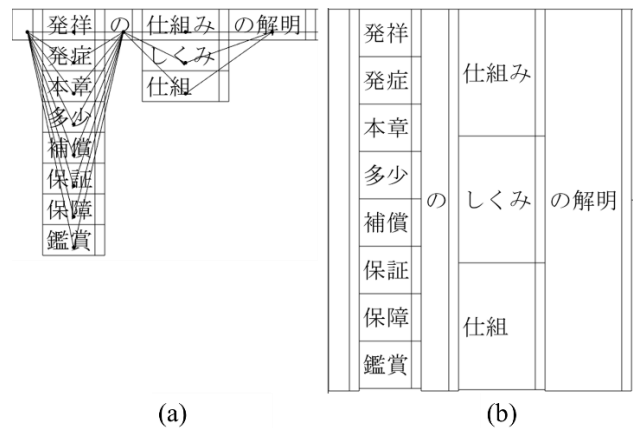


図1 単語候補の表示領域の大きさに対する処理

3.2 単語候補の列に対する処理

項目(2)の特定の単語の候補が表示領域の縦方向の制約によって表示しきれない場合は、候補選択のための操作回数が増加してしまう。そこで、表示領域を超えてしまう単語候補の列は分割し、横に並べて表示する。

A candidate selection method using fingertip tracking for speech-based text entry.

[†]Yuji Ikeda, Takeshi Umezawa, Noritaka Osawa, Graduate School of Advanced Integration Science, Chiba University

3.3 信頼度の高さに応じた単語候補の並び替え

項目(3)を考慮し、音声認識の誤認識を訂正するシステム全体でユーザの工数を減らすために、音声認識の中間結果のN-best単語列を用いてコスト関数を定義し、これが最小となる単語候補の列内の並び方で配置する。発話に対して第1候補の文章から第n候補の文章が出力され、その中に正しい文章があると仮定する。また、コスト関数 $f = \sum_i^n a_i x_i$ と定義する。ここで、第i候補の文章の重みを、 $a_i = \frac{\sum_{j=i}^n R_j}{\sum_{j=1}^n R_j}$ とし、 x_i を第i候補の文章の単語候補を最初から最後まで順に選択する際に必要な最短の移動距離、 R_i を第i候補の文章の信頼度とする。信頼度は0から1の範囲で、数値が1に近いほど、認識システムが確信をもって認識結果を出力したことを示す。

4. 実験

ユーザがマイクに対して発話した後、スクリーンに候補が提示され、それを空中にある仮想タッチパネルを利用して指で選択してゆく実験によって評価をした。指のモーションキャプチャには Leap Motion を用い、人差し指の操作によって候補の選択を行った。

4.1. 被験者実験による操作性の調査

提案手法の適用の有無による操作性の違いを2種類の条件で評価した。条件Aでは文章の最初から最後まで単語候補をすべて選択し、条件Bでは第1候補では不適切な場合のみ選択する。各条件とも20文に対して操作を行った。単語候補の選択回数が最小でない場合は誤操作として判定した。被験者8人で実験を実施した。

実験結果を表1に示す。平均移動距離は、1つの候補の選択にかかる移動距離の平均であり、平均操作時間は、単語候補の選択を開始してから正しい文章を作成できるまでの所要時間の平均である。平均誤操作率は、20文のうち誤操作があった文の割合である。誤操作時の平均移動距離、平均操作時間の割合は、誤操作がなかった場合に対する、誤操作と判定した場合の移動距離と操作時間の割合である。

提案手法の適用によって、条件Aで操作時間は0.3秒短縮された(表1条件A参照)。一方、条件Bで提案手法を利用すると操作時間が長くなった(表1条件B参照)。また、被験者に対して5段階リッカート尺度の質問紙調査を行った。提案手法の適用があると、質問文「意図した候補を選択できた」、「候補の表示が分かりやすかった」、「次に選択したい候補を素早く見つける

ことができた」では肯定的評価が増え、「使い易かった」の評価はあまり変化しなかった。また、「候補を選択する際に注意を要した」の肯定的評価が減り、候補を注視する負担が軽くなったといえる。

表1の誤操作率が低くなった要因として、提案手法によって隣接する単語候補が多い単語候補の表示面積が大きくなったため、ユーザが選択する可能性が高い候補を見つけやすくなったことが挙げられる。また、同じ要因で、ユーザが間違った候補を選択してしまった場合でも、次に選択したい候補を素早く見つけやすくなり、誤操作時の操作時間が短縮された(表1条件B参照)。また、指先の移動距離は提案手法の適用によって増加したが、その影響は大きくなかった。

表1 被験者実験の結果

	条件A		条件B	
	無し	有り	無し	有り
提案の適用				
平均移動距離(pixel)	850	873	352	402
平均操作時間(s)	4.35	4.05	1.77	1.80
平均誤操作率(%)	15.9	11.8	13.9	8.5
誤操作時の平均移動距離の割合(%)	226	175	228	138
誤操作時の平均操作時間の割合(%)	247	218	220	146

4.2. 候補提示手法の空間効率の調査

提案手法の適用の有無による空間効率の評価を20文を基に1860通りの文字サイズと表示領域のアスペクト比の組み合わせを用いて行った。提案手法を適用すると、ウィンドウ面積に対する単語候補表示面積の割合の平均が、20.5%から59.7%に増加した。しかし、単語候補総数に対する表示できた単語候補数の割合の平均は、提案手法を適用すると、95.5%から94.8%に低下した。

5. まとめ

本研究では、音声認識結果の単語候補提示手法を提案した。実験によって評価した結果、選択される可能性が高い候補の表示領域を大きくしたため、候補を見つけやすくなったことを確認した。しかし、空間効率の向上は十分にできておらず、その改善が今後の課題である。

参考文献

- [1] 遠藤拓, 寺田実. "音声入力における対話的候補選択手法." インタクション pp. 195-196, (2003).
- [2] 緒方淳, 後藤真孝. "音声訂正: 選択操作による効率的な誤り訂正が可能な音声入力インタフェース (音声言語)." 情報処理学会論文誌 Vol. 48.1, pp. 375-385, (2007).