

音声認識信頼度を用いた 誤認識修正支援エディタ¹

遠藤 拓、ナイジェル ワード
東京大学 工学部 機械情報工学科

1 概要

音声認識技術を用いたテキストエディタ（音声入力エディタ）は入力速度が速いものの、誤認識の発見と修正に時間がかかる。本研究では誤認識修正の効率を向上させるために、誤認識である可能性の高い部分をユーザに示すことの有効性を検証する。

2 はじめに

発声という行為は人間にとって日常的に行われるものであり、音声入力はキーボード操作に比べて疲労が少ない。また、入力速度という観点から見てもキーボードより速い。しかし、ディクテーションによる文書作成では誤認識の修正に時間がかかるので、さほど早くはないのが実情である（表1）[1,2]。

| 入力方法 | 時間 (min) |
|------------|-------------|
| キーボード入力 | 6'09 ~ 8'25 |
| 音声入力 (修正有) | 4'35 ~ 5'12 |
| 音声入力 (修正無) | 1'55 ~ 2'05 |

表 1: 約 700 字の日本語の文章入力における音声入力とキーボード入力の比較

連続音声認識のディクテーションによる文書作成は以下の手順で行われる。

- ユーザが音声で入力
- 認識結果が画面に表示される
- ユーザが誤認識を発見
- ユーザが誤認識部分にカーソルを移動
- ユーザが誤認識を修正

B は音声認識エンジンによって行われ、C、D、E はユーザがテキストエディタを用いて行う。

¹Displaying speech recognizer confidence information to support user correction of misrecognitions, Taku Endo and Nigel Ward, Department of Mechano-Informatics, School of Engineering, University of Tokyo

発見した誤認識部分の修正 (E) の方法に関しては、

- 音声での修正
- キーボードとマウスを用いた修正
- ペン入力での修正

など、様々な方法が試みられている [1]。

しかし、表示された認識結果の中から誤認識を見つける (C) に関してはあまり研究されていない。

3 仮説

単語ごとの尤度スコアを認識の信頼度とする。その値の小さい単語をユーザに示すことで誤認識発見速度は向上するという仮説を立て、それを検証する。

4 実験

この仮説の有効性を確認するために以下のような実験を行なった。

4.1 実験方法

あらかじめ用意された文書（日本語 410 字、120 単語）を被験者に読み上げさせる。その文章に一定の確率で誤りを含ませたものを、さも音声認識ソフトで入力したかのようなタイミングでエディタに表示する。被験者はその文章を修正し、文書を完成させる。文章中の誤り部分が強調表示されることで文書完成時間が短縮されることを示す（図 1）。

同じ内容の文書を使ったので実験を重ねると被験者が内容に慣れてしまった。それによる文書完成時間の短縮を差し引くため、被験者に誤り強調表示がある場合とない場合の実験を交互に 3 回行なった。

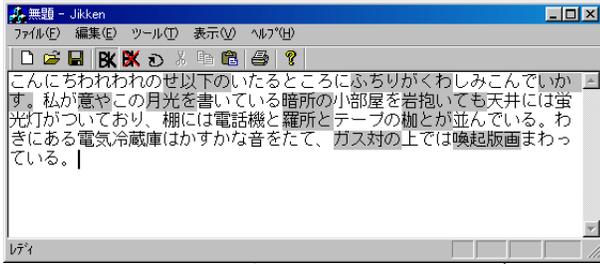


図 1: 誤り部分が強調表示されたエディタ

4.2 結果

実験の結果(表 2)、誤認識部分の強調表示がある場合は、強調表示がない場合に比べて文書作成時間が平均 8 % 短縮された。これにより、誤認識部分の強調表示が誤認識修正に有効であると確認できた。

| 被験者 | 回 | 誤り強調 | 文書完成時間 (sec) | 平均短縮率 |
|-------|---|------|--------------|---------|
| 被験者 1 | 1 | 有 | 408 | 11.11 % |
| | 2 | 無 | 378 | |
| | 3 | 有 | 264 | |
| 被験者 2 | 1 | 無 | 396 | 4.13 % |
| | 2 | 有 | 371 | |
| | 3 | 無 | 378 | |
| 被験者 3 | 1 | 有 | 384 | 8.76 % |
| | 2 | 無 | 411 | |
| | 3 | 有 | 366 | |

表 2: 強調表示の有無による文書作成速度の違い

4.3 実験の問題点

今回の実験では誤り部分をすべて強調表示したが、実際の音声認識を用いたシステムでは誤認識部分を完璧に示すことは困難である。誤りではない部分が強調表示されることもあり得る。これにより、ユーザの文書完成時間に生じる影響も実験をして調べなければならない。

4.4 実験によって得られたその他の知見

今回の実験では誤り部分を強調表示するだけが、その強調表示された部分へ Tab キーでジャンプする機能を追加すると、修正 (D、E) がより速くなるだろうと被験者に指摘された。

5 今後

今回の実験結果から誤認識部分をユーザに示すことが有効であることは確認できた。現在、音声認識エンジン Julius([3]) を用いて単語ごとの認識信頼度を表示することもできている。今後は実際にその信頼度を用いたテキストエディタを作成して実験を行い、仮説の有効性を調べる予定である。また、今回の実験に用いたエディタに更に (4.4) で示した機能を追加して再度実験を行ない、修正速度の効率が改善されるか調べる予定である。

参考文献

- [1] C-M.Karat, C. Halverson, D. Horn and J.Karat: Patterns of Entry and Correction in Large Vocabulary Continuous Speech Recognition Systems. In *Proc. CHI '99*, pp.568-575, 1999.
- [2] B. Suhm, B. Myers and A. Waibel: Model-based and Empirical Evaluation of Multi-modal Interactive Error Correction. In *Proc. CHI '99*, pp.584-591, 1999.
- [3] T.Kawahara, T.Kobayashi, K.Takeda, N.Minematsu, K.Itou, M.Yamamoto, T.Utsuro, and K.Shikano: Sharable software repository for Japanese large vocabulary continuous speech recognition. In *Proc. Int'l Conf. Spoken Language Processing (ICSLP)*, pp.3257-3260, 1998.