音声入力による情報検索システムのための音声認識誤り訂正方式†

4E－4

今井 裕志 (東洋大学)
井ノ上 直己 (KDD研究所)　橋本 和夫 (KDD研究所)　米山 正秀 (東洋大学)

1. まえがき

筆者らは、健常者とともに高齢者や障害者が容易にコンピュータを操作できる優しいユーザインタフェースの開発を目的に、音声によりコンピュータと対話しながら新聞記事の検索が行える情報検索システムの研究開発を行っている[1]。このようなシステムでは、一般に入力されたキーワードの正誤を確認しながらタスクを遂行する。筆者等は、利用者(特に高齢者)にとって、最も優しいインターフェース[2]の観点から、認識結果の確認方式として、認識誤り時だけ利用者が同じ音声を繰り返す方式を用いることにした。システムは、音声が繰り返し入力されたことを検出すると、認識結果が誤っていたと判断し、以前の認識結果を訂正する。本稿では、この繰り返し入力された音声を検出する方法について報告する。

2. 音声検索システム

検索対象は、1997 年 3 月 5 日までの約 10 年間分の朝日新聞記事約 180 万件であり、検索には市販の全文検索エンジンを用いている[3]。利用者は「セナ」「事故」「死亡」のように単語毎に区切って発声し、システムは 1 単語認識する毎に認識結果を画面表示するとともに音声合成装置を利用して認識結果を復唱する。システムが認識を誤った時に利用者は、同じ音声を繰り返し入力できる。例えば、図 1 に示すように、入力が「事故」に対し認識結果が「実行」と誤ると利用者は「事故」と再び入力し、システムは以前の出力結果「実行」を「事故」と書き換えて検索する。

3. 繰り返し音声識別手法

図 1 で、正しく認識された後に入力されたキーワードを異語音声、認識を誤った後に入力されたキーワードを同語音声と呼び区別する。ユーザの発声が異語音声か、同語音声かを識別するために以下の手法を試みた。

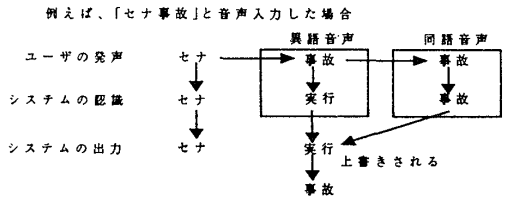

図 1　処理の流れ

(1) 認識結果の重なり度による識別手法

入力された音声に対して、認識候補を第 20 位まで求め、1 つ前の音声に対する認識候補との重なり度から判断する。

(2) 認識尤度差による識別手法

入力された音声に対して、1 つ前の音声に対する第 1 位の単語を含めた認識尤度を求め、この単語の尤度と認識候補中の第 1 位の候補の尤度との差から判断する。

(3) 音声波形の類似度による識別手法

入力された音声と 1 つ前の音声に対してフレーム毎にパワーを求め、DP マッチングを用いて計算した類似度から判断する。

4. 評価実験

4.1 評価用データ

評価実験では、31 名(男 16 女 15)の話者に、実際に新聞記事を検索してもらって、得られたデータ(異語音声 1240 語、同語音声 1446 語)を用いた。

4.2 評価方法

評価は、全ての同語音声中、正しく同語音声と判断された割合(recall)と、同語音声と判断されたデータの中で実際に同語音声であった割合(precision)で行う。

---

†Error Collection Method for Speech Input Information Retrieval System. By Yuji Imai (Toyo Univ.), Naomi Inoue, Kazuo Hashimoto (KDD R&D Lab.), and Masahide Yoneyama (Toyo Univ.)

4.3 実験結果

4.3.1 3つの手法によるそれぞれの評価結果

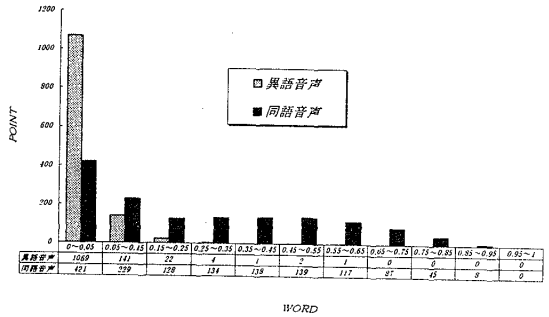

図 2 認識結果の重なり度

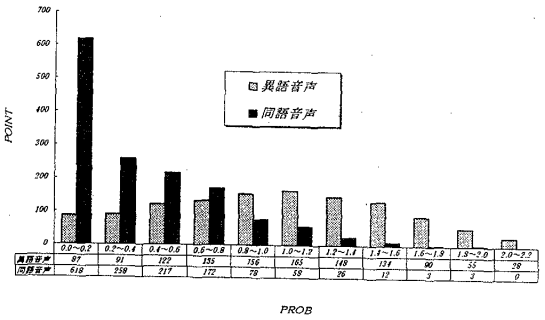

図 3 認識尤度差

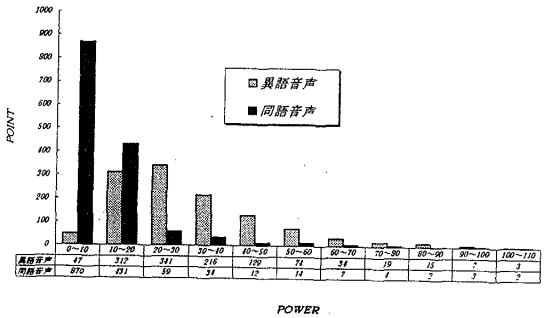

図 4 音声波形の類似度

手法(1)～(3)に対する同語音声と異語音声の分布を図2～4に示す。横軸は、各手法のパラメータ(重なり度、尤度差、類似度)であり、縦軸はデータ数である。図2では、ほとんどの異語音声が重なり度0～0.15の範囲に含まれ、ばらつきが小さいことが、図3では、異語音声・同語音声ともに広い範囲に分布していることが、図4では、ほとんどの同語音声が類似度0～20の範囲に含まればらつきが小さいことが分かる。図2のように異語音声のばらつきが小さい場合閾値を適切に設定することで、高いprecisionが得られ、図4では、逆に、高いrecallが得られると考えられる。

4.3.2 組み合わせによる評価結果

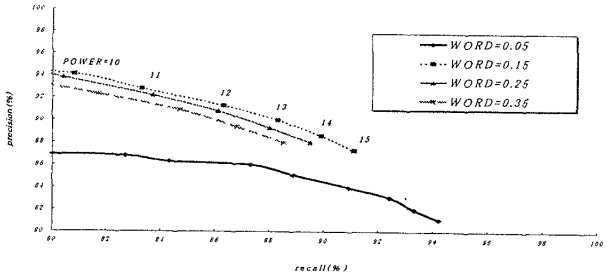

図 5 recall、precisionの関係

前節の結果から高い性能を得るためには、手法(1)と手法(3)を組み合わせることが良いと考えられる。手法(1)および手法(3)において閾値を設定し同語音声と異語音声を識別した。その結果を図5に示す。図5では、手法(1)における閾値(WORD)を0.05、0.15、0.25、0.35とし、それぞれに対し手法(3)における閾値(POWER)を変化させた時のrecall-precision曲線を示している。その結果、WORD=0.15の時が最も性能が良く、表1に示す通りWORD=0.15、POWER=14で、recall、precisionともに約90%の性能を得ることができた。

表 1 識別率

threshold	recall	precision
WORD=0.15 && POWER=14	89.9%	88.6%

5. まとめ

本論文では、入力された音声に対し、システムが認識を誤った時にだけ、再度同じ音声を入力して、上書き訂正する方式を提案した。繰り返し音声を識別するため、3つの手法を評価した結果、認識結果の重なり度と音声波形の類似度を組み合わせることにより、高い性能が得られることが分かった。

本研究は通信・放送機構(TAO)からの受託研究「高齢者・障害者のための機能代行・支援システム技術の研究開発」の一環として実施した。

参考文献

[1] 井ノ上他:"情報検索タスクにおける音声対話の分析",音講論集 3-1-12,pp109-110,Sept. 1997.

[2] 口ノ町他:"高齢者に親和性のあるヒューマンインターフェース",情処研報 SLP20-6,1998-2.

[3] 松下電気産業:"インターネット高速全文検索登録ソフトウェア PanaSearch / SSW マニュアル ",1996.