

言語情報による音声認識文の誤り訂正*

大坪 文和, 唐澤 博†

山梨大学 工学部‡

E-mail:{otsubo, karasawa}@jewel.yamanashi.ac.jp

1 はじめに

現在の音声認識は発話者とマイクとの距離が2 cm程度の接話型マイクを使うことによって高い認識率を得ているが、マイクとの距離が離れるにしたがって音響処理だけでは処理しきれなくなり認識率が下がってしまう。本研究では設置型マイクを用いてマイクと離れたところから発話を行い、音声認識により得られた誤りを含む音声入力データを音声認識エンジンより得られる音響スコアや、言語情報を用いて誤り訂正する。

2 研究の位置付け

音声認識率をあげる研究には音響処理によるものと言語情報を用いた処理がある。音響処理ではマイクロホンアレー [1] を用いたものや、音声認識アルゴリズムの改良 [2] などがある。言語情報処理では小野・唐澤らが開発した、誤り位置がノイズ記号として明示されているテキストを補完するシステム [3] 等がある。

本研究では音響処理部の認識文をそのまま扱い誤り訂正処理を行うため誤り位置が特定できない文に対して誤り訂正を行う。このため確からしい箇所、誤りが起きているであろう箇所を音響スコアや言語情報を用いて推定しつつ、誤りであろうか箇所に入るべき語の情報を集め、これらの情報やその他の知識、情報を用いて誤りと思われる箇所の誤り訂正を行う。

開発環境を表1に示す。

表1: 開発環境

音声処理用 PC	IBM Aptiva E6j(Windows2000)
音声認識ソフト	IBM ViaVoice98
マイク	SONY ECM-HS1
コンパイラ	MicroSoft Visual C++ 6.0
SDK	IBM ViaVoice SDK 3.0
言語処理用 PC	SunMicroSystem ULTRA10(SOLARIS8)

* Error correction of speech recognition data using language information

† Fumikazu Otsubo, Hiroshi Karasawa

‡ Yamanashi University, 4-3-11 Takeda, Kofu, Yamanashi 400-8511, Japan

3 誤り訂正の全体図

誤り訂正の流れを図1に示す。

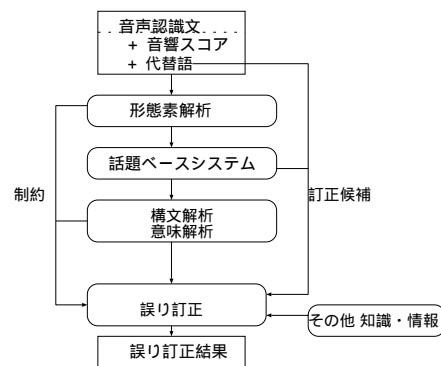


図1: 言語処理手順の全体図

4 用いた手法

4.1 音響スコア

音声認識エンジンが音声をテキストにデコードする際にスコア付けをする。ViaVoiceの音声認識エンジンは単語使用モデルを使用して、その単語の前後の文脈に基づくスコアと音響スコアを組み合わせる [4]。スコアの信頼性の調査結果を表2に示す。

表2: 音響スコアと正解率

スコア	正答率 (%)	全体における割合 (%)
11 ~ 20	100	5
1 ~ 10	99	24
0	84	5
-1 ~ -10	86	42
-11 ~ -20	60	17
-21 ~ -31	30	6
-31 ~ -41	0	1

調査の結果音響スコア1以上のものはほぼ正解だということになったため、これらの認識語は確定語と見な

し、またスコア -30 以下のものは誤りの可能性が高いため誤り箇所とし以降の処理を行う。

4.2 話題

対話に一貫性 (coherence) がある場合、対話者間に共通の話題が設定されることが多い。対話はお互いが今何の話題について話しているのかを理解しながら、それ以前に対話をしてきた話題に基づいて発言していると考えられる。そこで話題を特定するシステムである話題ベースシステムを用いて対話履歴から動的に話題を特定し、図2の様にその話題を特徴づける名詞群を得る。図2の/は形態素の切れ目を示し、() カッコ内の語は話題を特徴づける語である。

昨日 / は / (新作) / の / (映画) / を / 見 / た / よ / .
 へえ / 何 / を / み / た / の / ?
 新作 / の / (コメディ) / だよ / .

図2：話題「映画」

これらの話題によって指示されている語群は現在発言された文中に含まれている確率が高いといえ何処が正しいか分からない音声認識文の中にあって、確らしいと考えられることから、確定語とみなし以降の処理を行う。

4.3 隣接可能性

形態素の接可能性行列をもちいて形態素的な接続可能性をチェックし認識文の誤り箇所を推定する。以下のような条件で場合分けをし、図3ようにテキストを分類する。

- 隣なった前後の語の隣接を見て、両方から接続可能なら WhiteZone とする
- 確定語も WhiteZone とする
- 一方が可能で、もう一方が不可能なら GrayZone とする
- 両方が不可能ならば BlackZone とする
- 確定語の前後の語が隣接不可能ならば BlackZone とする
- スコアが -30 以下の語も BlackZone とする

WhiteZone は認識があっている可能性が高いが、逆に BlackZone, GrayZone は誤認識された可能性が高い。ここで Black, GrayZone に入るべき品詞を覚えておき後で誤り訂正を行うときの制約とする。

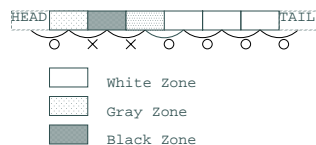


図3：隣接可能性チェック

4.4 音韻的類似語による誤り訂正

音声入力をしてどのような誤りパターンが多いかを調査したところ誤認識している部分の多くが音韻的に近いものだった。誤認識部を子音、母音に分けた場合の正解との一致度を表3に示す。

表3：誤りパターン調査

50 %以上の一致	50 %未満の一致	脱字
71 %	21 %	8 %

そこで誤認識語は音韻的に正解に近いという事から音韻的類似語から誤り訂正候補を探る。

EDR 日本語辞書 [5] に登録されている語から抽出した 14 万語を総当たりで DP マッチングにより文字列間距離コストを計算し、各語に対し類似度の高い語 100 語からなる音韻的類似語辞書を作成した。これにより誤りと思われる語から音韻的類似語辞書を引き得られる語を品詞的、構文的、概念的に適合するかをチェックし、推定語とする。

4.5 課題

現在誤り訂正をするためのモジュールを作成をしている段階なので、今後これらを統合し全体として一つの誤り訂正システムを作成していくことが課題となる。また新たな誤り訂正手法を模索していかなければならない。

参考文献

- [1] 安達理沙, マイクロホアレーを用いた実環境音声の認識, 山梨大学卒業論文 (2001)
- [2] 甘利, 中川, 鹿野, 東倉: 音声, 聴覚と神経回路網モデル: オーム社 (1990)
- [3] 小野, 唐澤: 補完候補の優先順位の決定, 情報処理学会第 48 回全国大会講演論文集, 4R-3, pp.181-182 (1994.3)
- [4] ナショナル・ランゲージ・サポート, SMAPI Developer's Guide, 日本 IBM 株式会社
- [5] 日本電子化辞書研究所, EDR 共起辞書, 日本電子化辞書研究所