

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いように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)

3 誤り訂正の全体図

誤り訂正の流れを図1に示す。

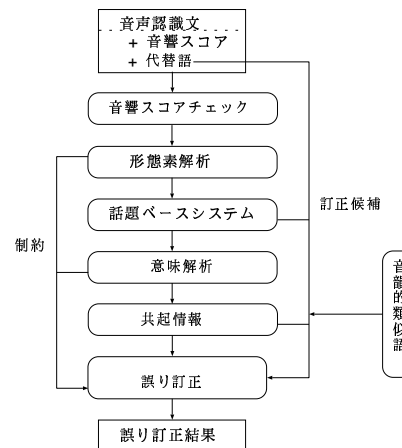


図1: 言語処理手順の全体図

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。

* Error correction of speech recognition data using language information

† Fumikazu Otsubo, Hiroshi Karasawa

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

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を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であって、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 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%以上的一致	50%未満的一致	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。

4.4 意味解析

形態素的・構文的には正しい文でも、意味的に不整合な物が含まれている可能性があるため意味制約情報を用いてこれらを排除するのが目的である。具体的には用言に対する深層格フィルターとなる概念とその深層格に対応する深層格を動詞毎にまとめておき、認識文中に用言が現れると深層格データを取り出し意味制約の整合を図る。

意味制約情報 (深層格データ) は EDR 共起辞書 [5] と EDR 日本語動詞共起パターン副辞書から抽出したものを利用した。表に深層格データの EDR 日本語辞書に対する格品詞の網羅率を示し、図4に深層格データの例を示す。

なお図中の網羅率は EDR 日本語単語辞書中の各品詞のエントリー数に対する、今回抽出された各品詞の割合で求めた。

$$\text{網羅率} = \frac{\text{今回抽出した各品詞数}}{\text{日本語単語時暑中の各品詞のエントリー数}}$$

表1：深層格データの EDR 日本語辞書に対する各品詞の網羅率

品詞	日本語辞書中のユニークな概念数	ユニークな概念数	網羅率 (%)
動詞 (共起辞書)		10554	
動詞 (副辞書)		13160	
動詞 (総数)	30569	20247	66
形容詞	1572	575	37
形容動詞	5172	918	18
計	37313	21740	-
用言 (品詞間の重複なし)	36501	21543	59

用言概念 深層格ラベル 深層格フィルター概念 後置詞
 061c7d agent 0f5020 は
 061c7d agent 1012f7 が

図4：深層格データ

深層格データにより意味制約を充足する深層格フィルターが文中にあるのならば、確らしい語としてスコアを上げ確定語とし、充足しないのならば訂正候補の中から制約を充足するものをさがしスコアを上げる。

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