1J1-OS13-1

未登録語認識における音声による音韻認識誤り訂正手法

A Speech Interface for Correcting Misrecognized Phonemes in Out-of-Vocabulary Word Acquisition

住井 泰介[*1*2] Taisuke Sumii　岩橋 直人[*1] Naoto Iwahashi　船越 孝太郎[*3] Kotaro Funakoshi　中野 幹生[*3] Mikio Nakano　岡 夏樹[*2] Natsuki Oka

[*1]（株）国際電気通信基礎技術研究所 Advanced Telecommunication Research Labs　[*2]京都工芸繊維大学 Kyoto Institute of Technology

[*3]（株）ホンダ・リサーチ・インスティチュート・ジャパン Honda Research Institute Japan Co., Ltd.

Teaching new words to robot is critical in physical environments. In such cases, a method for correcting misrecognized phonemes is necessary. In this paper, we propose to use generalized posterior probability(GPP) as a confidence measure to generate more reliable phonemes from two sets of phonemes - first speech and repeated one. The experiment shows promising results.

1. はじめに

近年，ロボットを家庭等で使用するための研究が進展している．その一つに，ユーザとの対話を通してロボットに物体等の名前を獲得させる研究がある [中野 09]．その際，名前が未登録語の場合，音韻を学習させる必要があるが，現在の技術では音韻認識正解率は高々80%程度にとどまっており [中川 06]，認識誤りが生じた際にそれを訂正する方法が必要である．

音韻認識正解率を向上させる方法として，音声入力にボタン入力を併用し，探索空間を削減する方法 [Chung 02] 等がある．これは, ユーザが電話等を通じてシステムに自分の名前の綴り等を入力する場面を想定している．

一方，ロボットが相手の場合，ユーザビリティの観点から，ボタン操作が不要であることが望ましい．

本研究では，音声のみを用いて音韻列を訂正する方法を提案する．

なお，本研究では登録語と未登録語の判別法は議論しない．

2. 提案法

音韻訂正するには，“訂正すべき音韻の特定” と “訂正後の音韻の選択” が問題となる．これに対し，提案法では (I) ユーザに再度発話（訂正発話）してもらい (II) 訂正発話の認識音韻列と訂正前の認識音韻列をマッチングし (III) 各音韻のペアから，より信頼度の高い方を選ぶことで，訂正する．例えば，未登録語「ディスプレイ（/d/i/s/u/p/u/r/e/i/）」を「/j/i/s/u/p/u/r/e/i/」と誤認識した場合，これと訂正発話の認識音韻列「/d/i/s/o/p/u/r/e/i/」をマッチングし，j と d，u と o から，それぞれより信頼度の高い方を選ぶ．ユーザとシステムの対話例を図 1 に示す．

S:これはナニナニです，と述べてください.
U:これはフラバン茶です.
S:クラバンチャでいいですか？
U:違う違う.
S:正しくはナニナニです，と述べてください.
U:正しくはフラです.
S:フラバンチャでいいですか？
U:そうそう.

図 1: 対話例（S はシステム，U はユーザ）

2.1 マッチング

マッチングには，始終端フリー DP マッチングを使用する．また，訂正発話は間違えた音韻の周辺部分のみとし，訂正する必要のない音韻を間違った音韻に誤訂正することを起こりにくくする．その際，マッチング精度の向上を図るため，距離には，音韻認識の混同行列（表 1）を利用した値を用いる．すなわち，音声入力中の音韻 γ（「発話目的音韻」）が音韻 α（「認識結果音韻」）に認識される確率を $P(\alpha|\gamma)$ とすると，認識結果音韻 α の発話目的音韻と，認識結果音韻 β の発話目的音韻が一致する確率は,

$$P(\alpha,\beta) = \sum_{\gamma} P(\alpha|\gamma)P(\beta|\gamma)P(\gamma) \quad (1)$$

として求められるので（γ は発話目的であり得る任意の音韻），音韻 α，β 間の距離 d を次のように定める．

$$d(\alpha,\beta) = -\log P(\alpha,\beta) \quad (2)$$

α の挿入・脱落ペナルティは $d(\alpha,\phi)$ とする（表 1 参照）．本研究では，学習データに ATR ディジタル音声データベース・セット C を使用した．

		α				
		zh	ϕ	ng	a	b
γ	zh	7072	235	0	0	3
	ϕ	20	0	7783	9547	35
	ng	0	389	30807	5	1
	a	0	129	22	92229	0
	b	4	169	2	0	16235

表 1: 混同行列（抜粋）．γ は入力音声中の音韻，α は認識後の音韻．$\gamma = \phi$ は挿入誤りに，$\alpha = \phi$ は脱落誤りに対応．マッチングに利用する．

連絡先: 住井 泰介, 京都工芸繊維大学, e9600005@edu.kit.ac.jp

2.2 信頼度の算出

認識された音韻の一般化事後確率（GPP;Generalized posterior probability）[Soong 03] と認識正解率の間には，図 2 に示す関係が得られた．この関係を示す近似関数を事前に音韻の種類ごとに作成しておき，音韻の信頼度として，GPP とこの関数から得られる推定正解率を使用する．本研究では，学習データに 2.1 節と同じものを使用し，3 次の多項式近似とした．

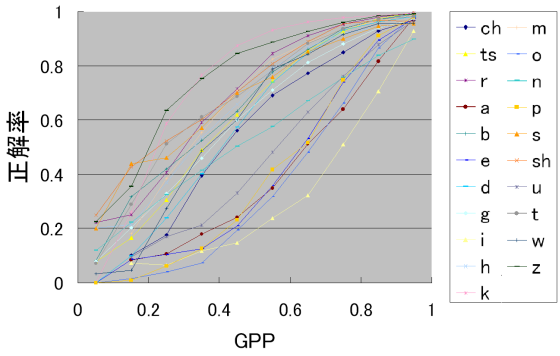

図 2: 音韻の種類ごとの GPP と正解率

2.3 音韻の選択

各マッチング・ペアから，推定正解率の高い方を正しい音韻として選ぶ．マッチングの相手がいない音韻は，推定正解率が 0.5 を上回れば正解，下回れば誤りとみなす．訂正法を図 3 に示す．

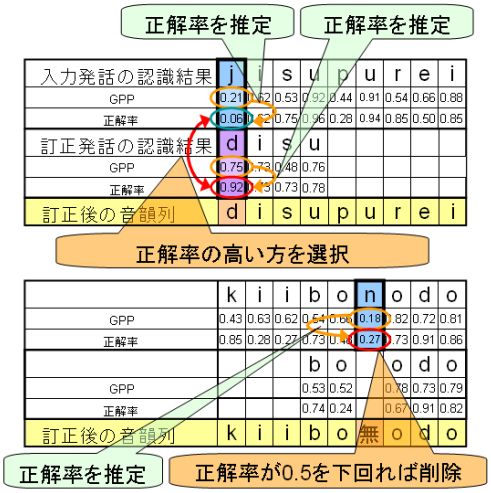

図 3: 推定正解率に基づく音韻の選択

なお，繰り返し訂正し続ける場合，それまでの訂正結果はどれも正しい音韻列ではないと分かっているので，新たな訂正結果はそれらとは一致させないようにする．

3. 実験

3.1 実験条件

被験者 1 名，単語数 40 で実験を行った（音韻認識器は ATRASR[Nakamura 06]．特徴量は 25 次元 MFCC．16kHzPCM，16 ビット量子化）．用いた単語の一部を表 2 に示す．ただし，母音の長さの訂正は行わず，長母音と短母音を同一視して訂正と結果の集計を行った．

単語	正解音韻列
フラバン茶	/f/u/r/a/b/a/N/ch/a/
ケビン・ベーコン	/k/e/b/i/N/b/e/e/k/o/N/
リモコン	/r/i/m/o/k/o/N/

表 2: 実験に使用した単語（抜粋）

3.2 結果

結果を図 4 に示す．訂正しない場合の単語正解率 8%が，1 回以下の訂正を許すことで 40%，2 回以下許すことでで 60%に向上した．3 回以下の訂正を許した場合は，66%にとどまった．

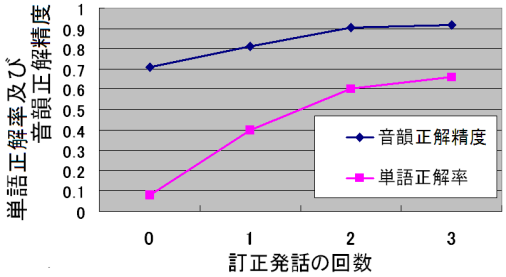

図 4: 実験結果．単語正解率は全音韻が正解した単語の割合．音韻正解精度は $(C_N - C_S - C_D - C_I)/C_N$（ただし C_N:真の音韻数，C_S:置換音韻数，C_D:脱落音韻数，C_I:挿入音韻数）

4. 結論

本稿では，未登録語音韻認識誤りの音声のみによる訂正手法を提案し，それが有望な手法であることを示した．今後，被験者の数を増やした実験を行う予定である．

参考文献

[中野 09] 中野 幹生, 能勢 隆, 田口 亮, 水谷 了, 中村 友昭, 船越 孝太郎, 長谷川 雄二, 鳥井 豊隆, 岩橋 直人, 長井 隆行: 自然な対話の中で物体の名前を覚えるロボット, 第 23 回人工知能学会全国大会講演論文集 (2009)

[中川 06] 中川 聖一: 自然な連続会話音声認識:その挑戦と限界, フェロー&マスターズ未来技術研究会資料 (2006)

[Chung 02] Chung, G. and Seneff, S.: Integrating Speech with Keypad Input for Automatic Entry of Spelling and Pronunciacion of New Words, *Proc. ICSLP2002*

[Soong 03] Soong,F.K.,*et al*: 連続音声認識候補受理/リジェクションのためのワードスポッティング仮説検証手法, IEICE technical report. Speech Vol.103,No.520,pp.41-46(2003)

[Nakamura 06] Nakamura, S.,*et al*: The ATR multilingual speech-to-speech translation system, *IEEE Trans. on Audio,Speech,and Language Processing*,Vol.14,No.2,pp.365-376(2006)