

論文 / 著書情報
Article / Book Information

論題(和文)	韓国人の日本語発声に対する発音誤りの自動検出・評定法
Title(English)	
著者(和文)	小俣 修一, 岩野 公司, 古井 貞熙
Authors(English)	shuichi omata, Koji Iwano, SADAOKI FURUI
出典(和文)	日本音響学会 2003年秋季講演論文集, Vol. , No. 2-8-8, pp. 259-260
Citation(English)	, Vol. , No. 2-8-8, pp. 259-260
発行日 / Pub. date	2003, 9

1 はじめに

近年、留学生を始めとする日本語学習者が増加しており、日本語の発音の自律学習システムの必要性が高まっている。そこで、音声認識技術を用いて、発音誤りを自動検出し、その発音を評価した上で、発音の矯正方法を提示するといった、発音自律学習システムが提案されている [1]。

本研究では、音声認識技術を用いた、日本語自律学習システムのための音素発音誤りの検出手法についての検討を行う。一般の音声認識では、音響特徴量として声道の形状特徴を反映したケプストラム情報を用いており、これまで提案されている発音誤り検出手法においても、同様の特徴量を利用している [2, 3]。しかしこの情報だけでは、調音構造が似た有声子音と無声子音の発音誤りを高精度に判断することは難しい。そこで、ケプストラム情報だけでなく、声帯の振動情報を併用した発音誤り検出手法を提案する。

2 使用データベース

本研究では、文部科学省科学技術研究費特定領域研究「メディア教育利用」で作成された日本語音声 DB [4] 中の韓国人母語話者 (32 名) 分を使用した。各話者は、日本語の難音音素ミニマルペア 115 単語を発声しており、これを実験対象のデータとした。なお、本 DB には 41 名の日本人による同単語の発声も手本用として含まれている。

3 韓国語母語話者の発音誤り傾向の分析

発音誤り検出に先立ち、韓国語母語話者がどのような音素発音誤りを起こしやすいか分析を行った。なお、今回の実験では、子音の置換誤りのみを対象とした。まず、日本語を母語とする 5 人の被験者に、全ての単語発声について「どのような日本語音素列に聞こえるか」という観点での書き起こしを依頼した。次に、得られた書き起こしの音素列と読み上げリストの音素列との比較を行い、発音の置換誤りが生じやすい音素を選択した。音素の選択に際し、読み上げテキスト中の音素の出現数にはばらつきがあるため、仮説検定を利用した。具体的には、「その音素の発音を間違える確率が 10% 以下になる」とした検定を、全ての子音について行い、有意水準 5% で仮説が却下され、発音誤りが 10% より大きくなると判断された音素を選出した。その結果、表 1 に挙げる 7 種類の音素の発音が、特定の音素に誤りやすいという結果を得た。このうちの 5 つは、閉鎖音の有声子音から同一調音構造の無声子音への発音誤りである。この誤り分析の結果は、[3] での韓国語母語話者日本語発音誤り分析と同様の結果であった。

4 発音誤り検出手法

提案する発音検出手法の流れを図 1 に示す。得られた 7 つの誤りやすい音素を含む単語の読み上げ発声を入力とし、正しく発音された場合と、誤って (誤りやすい音素で) 発音された場合を仮定して、それぞれの尤度を算出し、尤度比をスコアとする。

スコアには、声道の特徴を反映するケプストラム情報に基づくスコア (ケプストラムスコア: S_c) と声帯の振動情報を反映する有声・無声の度合いを表すスコア (有声・無声スコア: S_p) を用意し、それぞれを融合して最終的なスコアとする。

表 1. 韓国語母語話者の発音誤り傾向。

正	→	誤
/b/	→	/p/
/d/	→	/t/
/g/	→	/k/
/bj/	→	/pj/
/ɕj/	→	/tʃ/
/z/	→	/ɕj/
/ts/	→	/tʃ/

ケプストラムスコア S_c 算出のため、音響 HMM を用いて、正しい発音時の音素列と、誤った発音時の音素列に対する強制切り出しを行う。正しい音素、誤った音素それぞれの HMM に対する対数尤度を求め、フレーム平均した後、

$$S_c = (\text{正しい音素 HMM に対する対数尤度}) - (\text{誤った音素 HMM に対する対数尤度}) \quad (1)$$

としてスコアを得る。

有声・無声スコア S_p の算出では、相関法に基づく基本周波数抽出法の Entropic 社の get_f0 [5] で有声・無声の判別に用いられる正規化自己相関関数のピーク値 (以後、ac_peak と表記する) を利用する。ac_peak は 0 ~ 1 の連続値となる。ケプストラムスコアの算出時に強制切り出しによって得られた音素境界情報を元に、正しい音素、誤った音素区間における ac_peak のフレーム平均を求め、それぞれの場合の尤度を算出する。尤度の計算のため、予め、日本人 41 名の単語発声データを用いて、各音素のフレーム平均 ac_peak の確率分布を求めておき、この分布と照合することによって、正しい音素、誤った音素に対する対数尤度を求める。最終的には、

$$S_p = (\text{正しい音素の ac_peak に対する対数尤度}) - (\text{誤った音素の ac_peak に対する対数尤度}) \quad (2)$$

によってスコアを得る。なお、本実験で用いる ac_peak の確率分布は、データ量が十分でないことから、0 ~ 1 を 5 等分した区間に対する頻度分布として表現した。

融合スコア S_f は、得られた 2 つのスコアを重み付けして足し合わせて求める。

$$S_f = \lambda_c S_c + \lambda_p S_p \quad (\lambda_c + \lambda_p = 1) \quad (3)$$

λ_c, λ_p はケプストラムスコア、有声・無声スコアに対する重み係数である。得られた融合スコアを、閾値と比較することによって、対象の音素が正しく発音されているか、誤って発音されているかを判断する。

5 発音誤り検出実験

ケプストラムスコアの算出に用いる音響特徴量は、MFCC 12 次元、 Δ MFCC 12 次元、 Δ 対数パワーの計 25 次元である。強制切り出しで使用する音素 HMM には、連続音声認識コンソーシアム 2000 年度版ソフトウェアの高精度音響モデル (性別依存の

* A mispronunciation detection method in a Japanese pronunciation learning system for Korean

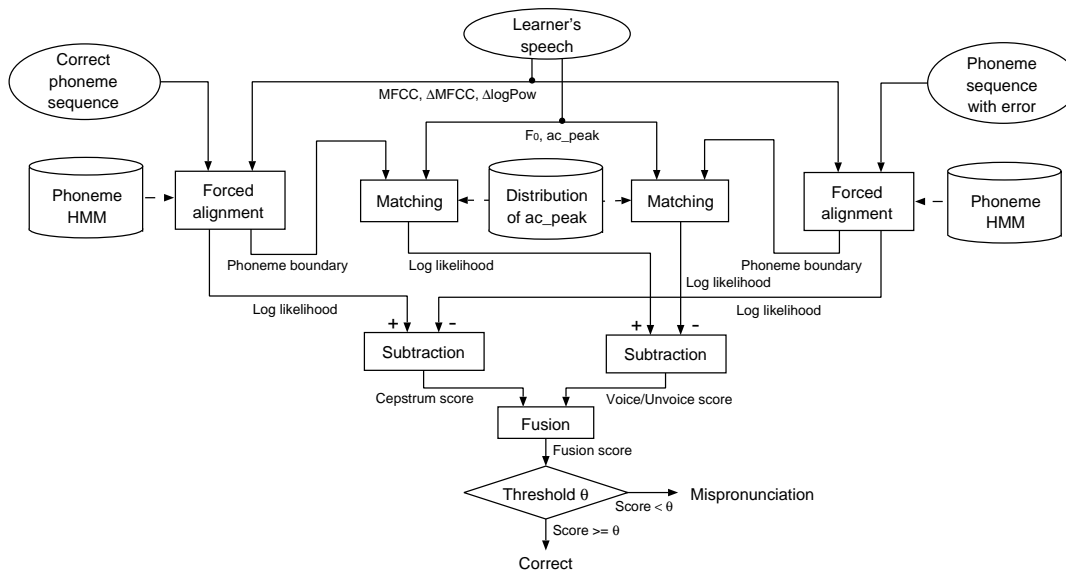


図 1. 発音誤り検出システム .

表 2. 正解発音と誤り発音の数 .

読み上げ音素	正解発音	誤り発音
/b/	/b/ 379	/p/ 51
/d/	/d/ 158	/t/ 21
/g/	/g/ 227	/k/ 30
/bj/	/bj/ 54	/pj/ 16
/ɕ/	/ɕ/ 239	/tɕ/ 32
/z/	/z/ 71	/ɕ/ 18
/ts/	/ts/ 164	/tɕ/ 12

表 3. 等誤り率 (%) の比較 .

	S_c only	S_p only	S_f (λ_c)
/b/	17.5	22.8	15.1 (0.5)
/d/	18.5	17.2	15.0 (0.2)
/g/	10.8	27.8	10.0 (0.7)
/bj/	37.2	20.0	20.0 (0.0)
/ɕ/	24.0	30.0	24.0 (1.0)
/z/	12.5	44.8	11.0 (0.7)
/ts/	8.0	44.4	8.0 (1.0)

不特定話者 triphone モデル) を, MLLR によって話者適応したものを用いる. DB 中には, 各韓国語母語話者が発声した ATR 音素バランス文 50 (または 53) 文が存在しており, これを適応データとした. 適応の際に, 読み上げリストの音素列を教師として使用した. ただし, この文発声も発音ミスを含み, 読み上げリストの通りの発音ではない可能性がある. 発声に音素の発音誤りの傾向 (例えば, /b/ → /p/ など) が含まれていると, 音素の遷移が話者適応によって適応後のモデルに反映されてしまい, 適応モデルを用いて正誤の発音の識別を行うことが困難となる. そこで, 大局的に話者性のみの適応化を行うため, MLLR のクラス数 を 1 とした.

発音誤り検出実験の対象データとしては, 表 1 に示す 7 つの誤りやすい音素を含む単語の読み上げ音声を用いた. ただし, 雑音が比較的多かった男性 3 名, 女性 2 名のデータは除いた. このうち, 日本語母語話者の評価において, 5 人中 4 人以上が正解としたものを「正解発音」, 5 人中 4 人以上が対立する誤りやすい音素で発音されたと判断したものを「誤り発音」として実験に使用し, その他の音素に置換されて発音されたもの, 評価が分かれたものは使用しなかった. 実験対象となった正解発音, 誤り発音を含む単語の数を表 2 に示す.

融合スコアの発音誤り検出性能を等誤り率 (EER) によって評価した. 表 3 に実験結果を示す. 閾値, 融合重みは音素種ごとに最適化した. 融合法における最適なケプストラムスコア重み λ_c の値を括弧内に示す. 比較のため, ケプストラムスコアのみ ($\lambda_c = 1, \lambda_p = 0$), 有声・無声スコアのみ ($\lambda_c = 0, \lambda_p = 1$) を用いた時の結果も併せて示す. /b/, /bj/, /d/, /g/, /ɕ/ の同一調音構造の有声・無声子音の対立では, 有声・無声スコアを併用することにより, 等誤り率が平均で 3.8% 削減された.

6 まとめ

本研究では韓国語母語話者の日本語発音における子音の発音置換誤り検出において, 同一調音構造の有声子音と無声子音の発音誤りを高精度に判別するため, 従来まで用いられていたケプストラム情報に加え, 有声・無声の度合情報を併せて利用した発音誤り検出手法を提案した.

韓国語母語話者による日本語単語発音を用いて, 発音誤り傾向の分析を行い, 発音の置換誤りの生じやすい音素を調べたところ, 7 種類の音素が挙げられ, そのうち 5 つが, 有声子音から同一調音構造の無声子音に発音される誤りであった. これらの音素を含む単語発音を用いて, 提案法の評価実験を行ったところ, 有声・無声情報を利用することによる発音誤り検出性能の向上を確認した.

今後の課題としては, ac_peak の確率分布の高精度化, 文章発声や他の母語話者の発音における本手法の有効性の検討, などが挙げられる.

参考文献

- [1] 中川聖一, 牧野正三, 壇辻正剛, “音声言語処理技術を用いた語学学習システム,” 音響誌, vol.59, no.6, pp.337-344 (2003-6).
- [2] 河合 剛, 石田 朗, 広瀬啓吉, “2 言語間の音響モデルを用いた音声認識による非母語発音誤りの検出と発音評価,” 音響誌, vol.57, no.9, pp.569-580 (2001-9).
- [3] 権 五杓, 鈴木基之, 伊藤彰則, 牧野正三, “音声認識を利用した韓国人のための日本語発音学習システム,” 信学技報, SP2002-164, pp.19-24 (2003-1).
- [4] 峯松信明, 仁科喜久子, 中川聖一, “外国語学習用読み上げ音声データベース,” 音響誌, vol.59, no.6, pp.345-350 (2003-6).
- [5] Entropic research laboratories, esps/waves+ with ensig5.3, reference release (1998).