

電子音楽モデル追加による 構内雑音環境下音声区間検出法の改良

江連 哲[†] 堀内 靖雄[†] 黒岩 眞吾[†]
[†] 千葉大学大学院融合科学研究科

1. はじめに

実環境下での音声認識において、雑音による認識精度の低下を抑制する取り組みは多い。しかし、構内アナウンスのような音声雑音として存在する場合、抑制することは難しく、挿入誤りにより音声認識精度が大幅低下する原因となる。音声雑音を考慮した音声区間検出を行った研究も近年取り組まれている[1]。先行研究[2]ではアナウンスモデルを用いてアナウンス区間を検出する手法を提案した。しかし、構内アナウンスには、アナウンスモデルでは検出することが出来ない音楽区間が存在しており、同じく挿入誤りの原因となっている。そこで本研究では、電子音楽モデル追加によって音楽区間を検出する手法を提案する。

2. 電子音楽モデルの追加

本研究では[2]で提案した、GMMを用いてアナウンス音声区間を事前に雑音区間として検出し、棄却する手法に対し、新たに電子音楽モデル(音楽 GMM)を追加することにより、構内アナウンス中の音楽区間の検出、棄却を行う手法を提案する。構内アナウンスに含まれる音楽はアナウンス音声とは異なり認識対象の音声と同じ特徴を持たないことから、電子音楽からモデルを作成できると考えた。そこで構内アナウンス中の音楽の構成音域に対応するために1オクターブ間の8音を往復する電子音4オクターブ分の電子音楽を作成しモデル学習に用いる。各GMMに対する入力フレーム尤度を比較し、音声/非音声の判別を行う。

3. 評価実験

CENSREC1-C[4]を用い音声区間検出実験及び、検出した音声区間に対する連続数字音声認識実験を行った。

4. 実験条件

音声区間検出のための各GMMの学習には表1のデータを用い、GMM混合数は40とした。評価音声には、表2に示す音声とアナウンスをSNR 0dB, 10dB, 20dBになるように重畳し、実環境を模擬した音声を用いた。音声区間検出の評価には誤棄却率と誤受理率が等しくなる誤り率である等誤り率を用いた。音声認識の評価には単語正解精度を用いた。音声認識のための音響モデルは状態数18、混合数20のHMM /zero/~/kyu/ と無音を用意した[3]。音声、無音GMMを用いた場合をベースラインとし、アナウンスGMMを加えた場合を従来手法とする。

表 1. 学習データ CENSREC1[3]

①通常音声	男女 55 名の数字読み上げ 8440 発話
アナウンス	①を 5m 離れたスピーカから流し録音
音楽	電子音楽 2400 秒

表 2. 評価データ CENSREC1-C[4]

音声	男女 55 名の数字読み上げ 312 発話
アナウンス	JR 西千葉駅スピーカから 16m で録音

表 3. 音声区間検出結果(等誤り率)

	0dB	10dB	20dB
ベースライン	24.1%	13.4%	6.1%
従来手法[2]	17.5%	7.1%	4.3%
提案手法	14.6%	6.5%	3.4%

表 4. 音声認識結果(単語正解精度)

	0dB	10dB	20dB
ベースライン	44.84%	75.33%	91.29%
従来手法[2]	49.47%	81.80%	95.73%
提案手法	50.24%	84.72%	97.13%

5. 実験結果

表3に音声区間検出実験結果を示す。また、表4に音声認識結果を示す。音楽GMMの追加により構内アナウンスに含まれる音楽による挿入誤りを抑制することができ、従来手法と比較して音声区間検出性能が向上し、結果として音声認識精度が向上した。

6. まとめ

本研究では、構内アナウンスに対し頑健な音声区間検出手法に音楽モデルを追加することで、頑健性の向上を行った。今後は様々な構内アナウンスに対する実験、区間検出の際の最適閾値推定手法の検討を行う。

参考文献

- [1] Shigeki Matuda, Naoya, et al. "Speaker Dependent Voice Activity Detection Robust to Background Speech Noise", Interspeech, Sep. 2012
- [2] 紺野遼輔ほか, "構内アナウンス環境下における音声認識のための音声区間検出法," 電子情報通信学会学生ポスターセッション予稿集, pp.136, Mar. 2013.
- [3] NAKAMURA Satoshi, et al. "AURORA-2J: An evaluation framework for Japanese noisy speech recognition" IEICE transactions on information and systems E88-D(3), pp. 535-544, Mar.2005.
- [4] 北岡教英ほか, "雑音下音声区間検出評価基盤の構築" 情処研報, 2006-SLP-63-1, Oct. 2006