D-14-7

AdaBoost と音声・唇 GMM による発話区間検出

松田博義 [1]　増田健 [2]　滝口哲也 [3]　有木康雄 [4]　神谷　昌宏 [5]
Hiroyoshi Matsuda　Ken Masuda　Tetsuya Takiguchi　Yasuo Ariki　Masahiro Kamiya

神戸大学工学部　情報知能工学科 [1−4]，富士通テン株式会社　開発本部 [5]

1　まえがき

車内のように狭く雑音が多い環境で，音響信号からドライバーの発話区間検出をしようとすると，雑音だけでなく助手席の音声など目的話者以外の音声を検出してしまうといった問題が生じる．この問題を解決するためには，(1) 音響信号を基に音声と雑音を判定する，(2) ドライバーの音声のみを検出する，いった 2 つの処理が必要である．(1) については音声と雑音の GMM を用いて音声区間検出 [1] を行う方法が一般的である．(2) については個人識別を行う方法も考えられるが，ドライバーを固定してしまうため，本論文では顔画像からドライバーの唇領域を検出し，開閉を判定することにより，ドライバーの発話区間のみを検出することを提案する．以下では，まず GMM による音声区間検出，Adaboost[2] を用いた唇領域検出，GMM を用いた開閉判定について述べる．次にそれらの統合手法，最後に有効性を確認する実験について述べる．

2　GMM による音声区間検出

男性，女性の音声および雑音の 3 種類の学習データを用いて，3 種類の GMM を作成しておき，テストデータに対して尤度比

$$L(x) = \log \frac{P_s(x_i|Model_{male\ or\ female\ speech})}{P_n(x_i|Model_{noise})} \quad (1)$$

を計算し，結果が閾値 θ 以上であれば音声，そうでなければ非音声と判定する．

GMM の学習では，音声データをケプストラム分析し，前後 2 フレームのデータを含めて，ひとつの特徴量ベクトルとする．男性，女性，雑音の 3 種類の GMM を用いて (1) 式の尤度比を計算し，男性と女性のうち尤度比の高かった方を採用する．実際には尤度 $L(x)$ を時間的に平滑化した後，閾値 θ で判定を行う．最後に，検出された音声区間からある一定時間以下のものを削除することにより，最終的な音声区間を得る．

3　AdaBoost と GMM による唇開閉区間検出

本研究では，カスケード型 AdaBoost を用いることで高速かつ高精度な唇領域検出を行う．唇領域以外の検出を抑えるため，開始 30 フレームを利用して，次の方法により唇領域を確定した．始めに顔画像を N 個のブロックに分割しておく．各フレームで検出された複数の唇候補領域の各重心を求め，この重心が属するブロックに 1 点を投票していく．30 フレーム終了時点で最大投票を得たブロックを求め，このブロックに重心が存在する候補領域の内，ブロック内で最も中央に近い候補領域を唇領域として決定する．以降のフレームでは，直前に検出された領域の最近傍候補を唇領域として出力する．

次に，GMM による開閉判定を行う．上述の方法で検出された唇領域を M 個のブロックに分割し，各ブロックについて画素値の平均を求め，M 次元特徴ベクトルを得る．予め複数人の開いた唇画像と閉じた唇画像を基にオフラインで 2 つの GMM を作成しておく．この GMM に，得られた特徴ベクトルを入力することにより，(1) 式と同様にして尤度比を計算し，唇画像の開閉を判断する．

表 1　区間検出の実験結果

	音声区間検出結果	唇開閉検出統合結果
正答率	0.951	0.940
適合率	0.476	0.534

4　音声区間検出と唇開閉検出の統合

音声が検出された区間内でドライバーの唇開閉判定を行い，ドライバー以外の発話やノイズ等の誤検出を抑える (図 1)．

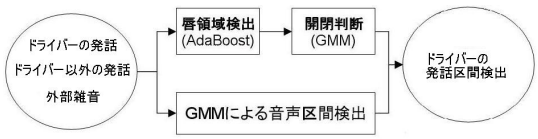

図 1　発話区間検出システム

5　実験

室内において，蛍光灯の下で撮影された映像を用いて実験を行った．カメラは夜間走行時も考慮して赤外線カメラを使用した．話者は男性 1 名である．発話内容は日本全国の地名，50 単語である．SN 比はカットオフ周波数 200Hz のハイパスフィルタをかけて 10dB〜20dB であった．各単語発話の間に別の人物の単語発話を挿入し，それをサイドシート，もしくは後部座席に座っている人が発話した音声とした．音声区間検出結果，唇開閉区間検出結果との統合結果の平均を表 1 に示す．

6　おわりに

ドライバーの発話区間の検出精度を高めるために，Adaboost を用いた唇領域の検出，さらに GMM を用いた開閉判定を行った．学習の改善，並びに使用特徴量の改良による精度向上を今後の課題としたい．

参考文献

[1] Norbert Binder, Konstantin Markov, Rainer Gruhn, Satoshi Nakamura, “ Speech Non-Speech Separation With GMMs ”, 日本音響学会論文集，pp.141-142，2001 年 10 月

[2] 勞　他, “ 高速全方向顔検出 ”, 画像の認識・理解シンポジウム（MIRU2004），pp.II-271-276，2004 年 7 月