

EBGM を用いた唇の形状抽出による発話区間の検出

増田 健[†] 青木 政樹[†] 松田 博義[†] 滝口 哲也^{††} 有木 康雄^{††}

[†] 神戸大学大学院自然科学研究科
兵庫県神戸市灘区六甲台町 1-1

^{††} 神戸大学自然科学系先端融合研究環
兵庫県神戸市灘区六甲台町 1-1

E-mail: [†]{masudaken,masamax777,matsuda}@me.cs.scitec.kobe-u.ac.jp, ^{††}{takigu,ariki}@kobe-u.ac.jp

あらまし 音響信号と同時に目的発話者の顔画像を利用して、唇の形状を把握することで発話区間検出を行う手法を提案する。音声認識において、周囲の雑音、特に目的話者の近くで発話している人物の音声を誤って検出するという問題を防ぐために、顔画像より唇の形状を把握し、そのアスペクト比の時間変化量を求めることによって発話の有無を検出する。画像の撮影には、照明条件の変化にも対応できるように赤外線カメラを用いている。このため、提案手法ではグレースケール画像より抽出できる特徴量として一般的な輝度値を利用して、詳細な唇の形状を把握する手法として Elastic Bunch Graph Matching を採用した。実際に車内で撮影されたデータを用いて実験を行った結果、提案手法による適合率は音声判定のみの結果と比較して約 40% 向上した。

キーワード 発話区間検出, 唇領域抽出, Elastic Bunch Graph Matching

Voice activity detection by lip shape extraction using EBGM

Ken MASUDA[†], Masaki AOKI[†], Hiroyoshi MATSUDA[†], Tetsuya TAKIGUCHI^{††}, and
Yasuo ARIKI^{††}

[†] Graduate School of Science and Technology, Kobe University
1-1, Rokkodai, Nada, Kobe, Hyogo

^{††} Organization of Advanced science and Technology, Kobe University
1-1, Rokkodai, Nada, Kobe, Hyogo

E-mail: [†]{masudaken,masamax777,matsuda}@me.cs.scitec.kobe-u.ac.jp, ^{††}{takigu,ariki}@kobe-u.ac.jp

Abstract We propose a voice activity detection in a car by using shape of lip and acoustic signals. To prevent the wrong detection caused by the different speakers using speech processing technologies, the proposed system extracts the shape of lip and computes the aspect ratio. An infrared camera is used to cope with the change of lighting environment. Elastic Bunch Graph Matching is employed to extract the shape of lip from gray scale images. Experimental results show the proposed system improved the precision rate of voice activity detection by 40% compared to the method using acoustic signals recorded in a car.

Key words voice activity detection, lip area extraction, Elastic Bunch Graph Matching

1. はじめに

音声による意思伝達手法は、ボタンの操作などによる入力時の煩わしさがなく、基本的に両手が塞がっている状態であるドライバーなどにとって役に立つ。しかし、車内などのように狭く雑音の多い環境下では、音響信号のみを用いてドライバーの要求を認識しようとする、雑音や助手席の音声などの目的話者以外の音響信号によって誤った入力となされる可能性がある。

このような問題を解決するために、雑音を判別して除去する方法が考案されているが、検出された音声ドライバーによる音声なのか、それ以外の人間による音声なのかを音響信号のみから判別することは困難である。そこで、本論文では、この音声区間からドライバーの発話区間のみを検出することを目的として、唇の詳細な形状を把握し、そのアスペクト比の時間変動量を求めることによって、唇の動静判定処理を行うシステムの構築を提案する。

音響信号と同時にドライバーの顔を含む画像情報を用いることで、目的話者の発話区間を検出しようとする研究は盛んに行われている。しかし、その多くは音声と深く関る唇領域の抽出と形状の把握により、発話の有無を調べるというものである。唇領域を抽出するにあたって、RGB 値分布 [4]、HSV 値分布 [5]、エッジ抽出 [5]、動きベクトル [6] など様々な特徴が用いられている。また、目的領域を抽出する手法として、テンプレートマッチングによる抽出 [7]、SNAKE [8]、ヒストグラム比較によるアクティブ探索法 [9]、Boosting 法 [10] などよく知られている。

しかし、カラー情報の利用は照明環境に左右されやすいため、夜間走行などが考えられる車内で使用するには適さない。従って、提案手法では、夜間でも使用可能な赤外線画像を用いる。赤外線画像はグレースケール画像であり、特徴量として得られる明度値から高速に唇領域を抽出可能な手法として AdaBoost 法や正規化相関法が存在する。これらの手法を用いることで、唇を含む方形領域の取得とフレーム間の画素値差分和の時間変動量を調べ、口の動静判定を行うことは可能である [16]。しかし、これらの手法によって唇の形状把握による発話内容の認識などにつなげることは困難である。そこで、唇の詳細な形状を把握するために、Elastic Bunch Graph Matching (EBGM) を採用した。さらに、発話に伴って唇の形状変化が大きくなることを考慮し、アスペクト比変動量の計測を行うことで、対象者の口が動く区間を判定する。最後に、音響信号を基に音声と雑音を判別して雑音区間を削除した音声区間を検出し、その結果と本手法の結果を統合することで、ドライバーの発話区間のみを検出する。

本論文の構成について述べる。2 節では、EBGM を用いて画像情報から唇の形状を把握し、アスペクト比変動量を求める手法について述べる。3 節では、GMM を用いた音声区間検出のアルゴリズムについて述べる。4 節では、画像情報と音響信号を用いた発話区間の検出結果の統合について述べる。5 節では、提案手法を用いた評価実験とその結果について述べる。

2. EBGM による発話区間の検出

2.1 Gabor Wavelet

Gabor Wavelet は Elastic Bunch Graph Matching の基礎であり、本節ではそのアルゴリズムについて述べる [12] [13]。

2.1.1 Gabor Wavelet

Gabor Wavelet は周波数を変化させることにより、全体的な特徴から局所的な特徴まで抽出することができる。また方向を変化させることで Wavelet の向きに対応した特徴を得ることができる。その様子を示したのが図 1 であり、これは画像を Gabor Wavelet で畳み込んだ時の Real part 画像と Magnitude 画像で、Magnitude 画像は Real part と Imaginary part の二乗和の平方根をとったものである。図の上の畳み込み結果の Magnitude 画像を見ると左顎が、下の画像を見ると右顎がよく反応していることがわかる。

Gabor Wavelet は式 (1) で与えられる、この Wavelet と顔の特徴点を畳み込むことにより特徴量を得る。この関数には、

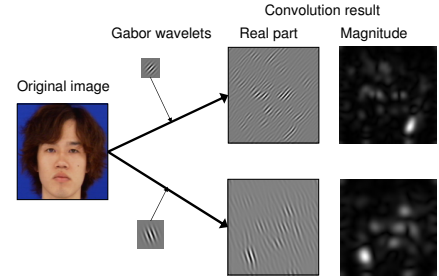


図 1 Two-dimensional wavelet convolution example

単純な波の周波数と方位の両方を表す波ベクトル \vec{k}_j と平滑化のためのガウス関数を含んでいる。

$$\psi_j(\vec{x}) = \frac{k_j^2}{\sigma^2} \exp\left(-\frac{k_j^2 x^2}{2\sigma^2}\right) \left[\exp(i\vec{k}_j \vec{x}) - \exp\left(-\frac{\sigma^2}{2}\right) \right] \quad (1)$$

$$\vec{k}_j = \begin{pmatrix} k_{jx} \\ k_{jy} \end{pmatrix} = \begin{pmatrix} k_\nu \cos \varphi_\mu \\ k_\nu \sin \varphi_\mu \end{pmatrix} \quad (2)$$

ここで、 $k_\nu = 2^{-\frac{\nu+2}{2}} \pi$ 、 $\varphi_\mu = \mu \frac{\pi}{8}$ であり、 $\nu = 0, 1, 2, 3, 4$ の 5 空間周波数、 $\mu = 0, 1, 2, 3, 4, 5, 6, 7$ の 8 方位の Wavelet が得られる。

2.1.2 Jet

Gabor Wavelet による畳み込みの特徴量を Jet と呼ぶ。これは特徴点の複素 Gabor Wavelet 係数の集合であり、これらと比較することにより特徴点位置の推定を行う。Jet のイメージ図を図 2 に示す。

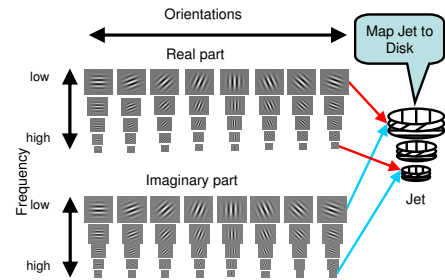


図 2 Gabor Wavelet

今回は 8 方位、5 空間周波数を持った Real part と Imaginary part の計 80 の wavelet を使い、Real part と Imaginary part から生成された 40 の複素係数を Jet は持っている。Jet は式 (3) のように書くことができる。

$$\mathcal{J}_j = a_j \exp(i\phi_j) \quad (3)$$

ここで、 \vec{x} は顔特徴点座標 (x, y) で $a_j(\vec{x})$ は複素係数の Magnitude、 $\phi_j(\vec{x})$ は複素係数の Phase、 j は 0 ~ 39 の計 40 個である。

2.2 Jet Similarity

2つの Jet, J と J' の相関を考える．2つの Jet の位置は、画像上で \vec{x} と \vec{x}' であり、位置の差ベクトルは、

$$\vec{d} = \vec{x} - \vec{x}' = \begin{pmatrix} dx \\ dy \end{pmatrix} \quad (4)$$

である．

ここで、2つの Jet の大きさと同位相に関する類似度を考える．即ち、

$$S_D(J, J') = \frac{\sum_{j=0}^{N-1} a_j a'_j \cos(\phi_j - \phi'_j)}{\sqrt{\sum_{j=0}^{N-1} a_j^2 \sum_{j=0}^{N-1} a'^2}} \quad (5)$$

この式を用い J の Jet にもっとも類似した J' を見つけるには、 $\phi_j - \phi'_j = \vec{k}_j \vec{x} - \vec{k}_j \vec{x}' = \vec{k}_j (\vec{x} - \vec{x}') = \vec{k}_j \vec{d}$ であることに注目し、位相差 $\phi_j - \phi'_j - \vec{k}_j \vec{d}$ を考慮に入れると、

$$S_D(J, J') = \frac{\sum_{j=0}^{N-1} a_j a'_j \cos(\phi_j - (\phi'_j + \vec{d} \vec{k}_j))}{\sqrt{\sum_{j=0}^{N-1} a_j^2 \sum_{j=0}^{N-1} a'^2}} \quad (6)$$

となり、すべての j に対して位相だけでなく、大きさも含めた類似度が最大になる \vec{d} を求めればよい．

2.2.1 Displacement Estimation

式 (6) で微小変位ベクトル \vec{d} は明確に定義されていない．ここでは \vec{d} の推定手法について述べる．まず初期探索点の場所の Jet の類似度を $dx = dy = 0$ とし式 (6) で計算し、その後、初期探索点の右上、右下、左下、左上の点における Jet の類似度と初期探索点の Jet の類似度を比較し、類似度が高いほうを次の探索開始点として、さらに右上、右下、左下、左上と探索を続けていく．また探索点が周りの点よりも高い類似度の場合はステップ幅を $1/2$ にしてさらに同じことを繰り返し、局所的な探索に入る．これを 50 回繰り返し、一番高い類似度の点を推定される特徴点位置として抽出する．

2.3 EBGM

本節では、Elastic Bunch Graph Matching について述べる．

2.3.1 Graph

グラフとは、各特徴点位置から Jet を抜き出したものを言う．グラフのイメージ図を図 3 に示す．この抽出された顔領域におけるグラフを利用することで、目的領域の形状把握が可能になる．

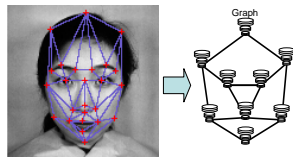


図 3 Jet extracted from facial feature points

2.3.2 Bunch Graph

Bunch Graph とは複数人からグラフを抽出し、それを束 (Bunch) 状にしたグラフのことである．この Bunch Graph を元に対象の特徴点位置の探索を行う．Bunch Graph は束にした

グラフの情報を保持し、探索の時に束にしてあるものと探索点での Jet を比較したのち、類似度の一番高い Jet を採用して位置決めをする．Bunch Graph により様々な状況において特徴点探索ができるようになる．例えば、髭の生えた人の顎周りや、目を瞑っている人の目の位置などである．つまり Bunch Graph には様々な人を学習させておく必要がある．Bunch Graph のイメージ図を図 4 に示す．

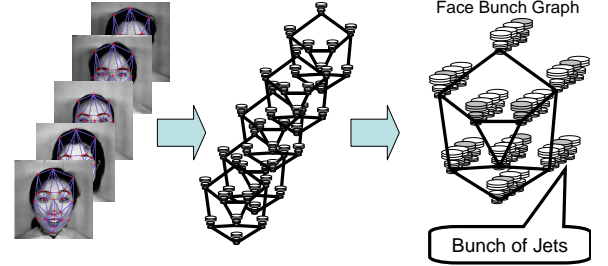


図 4 Face Bunch Graph collected from Face Graphs

2.3.3 Elastic Bunch Graph Matching

手動で特徴点を指定した Bunch Graph は特徴点位置探索に使われる．Elastic Bunch Graph Matching の流れ図を図 5 に示す．まず始めに、特徴点を検出したい画像を入力すると、そこに Bunch Graph が貼り付き、そこから 2.2.1 項で示した手法を用いて特徴点位置の局所探索に入る．位置の推定が完了した後、Graph が抽出される．

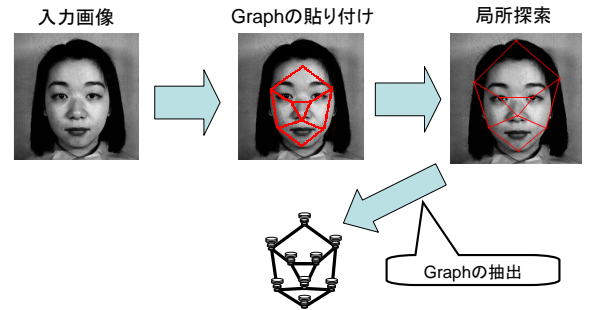


図 5 Elastic Bunch Graph Matching procedure

2.4 唇の形状認識と発話区間の検出

唇の輪郭に沿って抽出した特徴点と、鼻腔、耳、顎の特徴点を用いて、Bunch Graph を作成する．この Bunch Graph を用いて EBGM を行うことで、人間の顔の下半分から唇の詳細な形状を検出することが可能となる．図 6 に抽出する特徴点のモデルを、図 7 に検出結果の一例を示す．抽出された唇の形

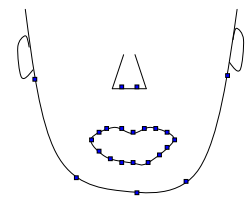


図 6 An example of feature points for EBGM



図 7 Lip shape extraction using EBGMM

状より，上下端，左右端を求め，唇の縦幅 (Lip_{Height}) と横幅 (Lip_{Width}) の長さを算出する．この結果を用いて式 (7) からアスペクト比 ($Aspect\ Ratio$) を計ることにより，図 8 に示すように，画像内における対象者の唇サイズの変化に対して頑健な値を得ることが可能となる．

$$Aspect\ Ratio = \frac{Lip_{Height}}{Lip_{Width}} \quad (7)$$

この値は，図 8 に示すように口を開くと大きくなり，閉じると

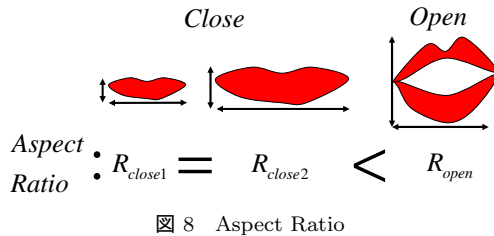


図 8 Aspect Ratio

小さくなる．人間の発話には唇の動きが伴うという前提から，発話時にアスペクト比の変動量が大きくなると予想される．そこで，連続するフレーム間でのアスペクト比の差分を求めると，唇の動き量を求めた．図 9 に，発話時と非発話時の，アスペクト比のフレーム間における変動量の例を示した．横軸はフレーム番号，縦軸は，各フレームでのアスペクト比と，直前フレームでのアスペクト比の差分値を表す．本研究では，アス

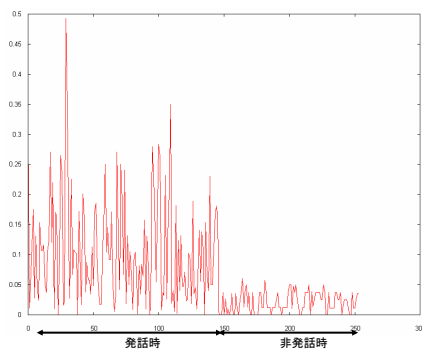


図 9 Change of Aspect Ratio

ペクト比の変動量が閾値を超えた場合を，対象者の発話区間として検出する．また，あくびなどのように，発話を伴わない唇の動きを発話として誤検出することを防ぐために，変動量が閾値を超える回数が一定以上である音声区間を対象者の発話区間として決定することとする．

3. GMM による音声区間検出

3.1 音声特徴抽出

音声特徴抽出には MFCC (Mel Frequency Cepstral Coefficients) を用いる．MFCC は各時刻におけるメル周波数スペクトルを DCT により直交変換 (ケプストラム化) した静的な特徴である．時間的に動的な特徴も含めるため，各フレームから得られた MFCC 音声特徴ベクトルを，前後の 2 フレームとあわせる．それにより得られる計 5 フレームからなるベクトルを，一つの音声特徴ベクトルとして扱う．ただしこれらの手順は，5 フレームごとにブロック単位で行うのではなく，1 フレームごとに窓をシフトさせることによって行う．この手法により，尤度計算の際に計算コストが大きくなるが，雑音と音声をより明確に判別することが可能になる [11]．

3.2 尤度計算

前述した音声特徴に対して，GMM (Gaussian Mixture Model: 混合正規分布) を用いて尤度計算を行う．GMM とは，複数の正規分布の重みつき和をとることにより，複数の山を持たせた分布のことであり，次式で定義される．

$$f(o) = \sum_{w=1}^W \lambda_w N(o; \mu_w, \Sigma_w) \quad (8)$$

$$\sum_{w=1}^W \lambda_w = 1 \quad (9)$$

ここで，データの次元数を D とした場合， W は混合数， $o \in R^D$ はデータのベクトル， $\mu_w \in R^D$ は w 番目の分布に属する平均ベクトル， $\Sigma_w \in R^{D \times D}$ は分散共分散行列であり， $N(o; \mu_w, \Sigma_w)$ は正規分布である．

3.3 音声区間検出

音声の各フレームから得られた尤度値に対し，式 (10) により，尤度比検定を行う．

$$L(x) = \log \frac{P(x | speech)}{P(x | noise)} \quad (10)$$

ここで， $P(x | speech)$ は音声尤度， $P(x | noise)$ は非音声尤度である． $L(x)$ の値は，以下のように前後のフレーム間で平滑化しておく．

$$L(x_i) = \sum_{j=i-\frac{N}{2}}^{i+\frac{N}{2}} L(x_j) \quad (11)$$

こうして得られた $L(x)$ の値が，ある閾値 θ 以上であれば音声，以下であれば非音声フレームとする．さらに，得られた区間のうち，短い区間を削除することによって，最終的な音声，非音声区間を得る．

4. 音響情報と画像情報の統合

本節では，音響信号より検出した音声区間と，2 節で述べた唇の形状認識手法を用いて，実際に発話区間を検出する手法について述べる．図 10 に，システム全体の処理の流れを示す．

音響信号より検出された音声区間内 (図 10-(b)) でのみ，唇

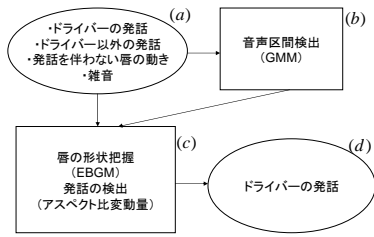


図 10 Integration System

の動き検出 (図 10-(c)) を行うこととする。各音声区間で算出したアスペクト比の変動量の結果、唇が動いていると判定されたフレームが存在する音声区間をドライバーによる発話区間、唇が動かなかったと判定された音声区間を非発話区間とした (図 10-(d))。図 11 に、縦軸にフレーム間アスペクト比変動量、横軸をフレーム番号としたときの判定の例を示す。また、閾値 ζ は、適合率が最大になるように設定し、各フレームにおいて出力されるアスペクト比変動量が ζ 以上であれば唇は動いていると判定し、 ζ 未満であれば唇は動いていないと判定する。

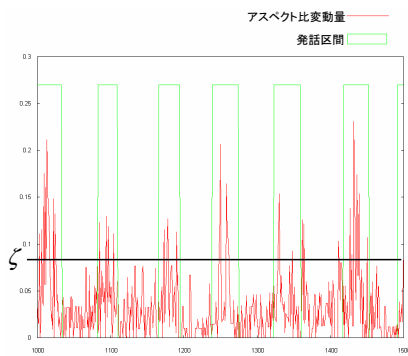


図 11 Voice activity detection with change of Aspect Ratio

5. 評価実験

5.1 実験条件

EBGM 用の Bunch Graph を作成するため、SoftPia Japan database [15] より 142 名の正面顔領域を含む写真を利用して、特徴点位置を手動で取得した。図 12 に、学習画像の例を示す。また、EBGM 学習用の特徴点の数は 23 個とした。各特徴点の位置を、図 6 に示す。提案手法を用いて実験を行うにあたり、

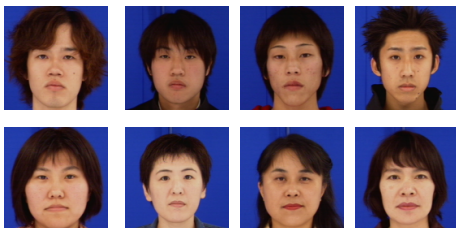


図 12 Learning Images

テストデータとして日本人男性 1 名と女性 1 名の顔画像を含む動画を使用した。昼間時におけるアイドリング状態の車内に

て、日本全国の地名 100 単語を発声してもらい、撮影と録音を行った。音響信号におけるノイズの 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。また、夜間時でも使用可能にすることを考え、赤外線カメラを用いている。動画を使用する際には、ファイルサイズの問題から DivX-Codec [14] を用いて圧縮した。動画の撮影は、図 13 に示すように、車内のドライバー席の正面位置に取り付けられたカメラを用いて行った。図 14 に、EBGM による唇形状の把握結果を示す。

今回の実験で用いたテストデータの収録では、車内にドライ

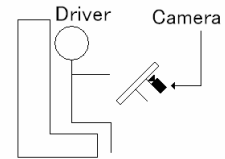


図 13 Shoot Environment

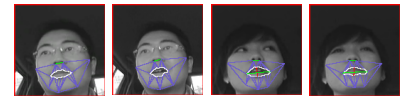


図 14 Examples of EBGM Results with test data

バーしか存在していなかったため、周囲環境からの雑音はあってもドライバー以外の人間の発話は存在していなかった。従って、検出された各単語の発話区間の間に、ドライバー以外の発話を擬似的に挿入し、ドライバーの発話区間検出実験を行った。図 15 の (b) に、提案手法によって検出した音声区間の例を示す。横軸はフレーム番号である。また、図の上段 (a) はドライバーのみが発声したデータに対して、手で切り出した音声区間である。下段 (c) はドライバーの発声の間にドライバー以外の発声を挿入したデータに対して、3 節で述べた音声処理のみによって音声区間を検出した結果の例である。この (b) で示す結果が (a) に近づくほど、提案手法による発話区間検出の精度が高くなっていると言える。なお、(c) に示した音声処理のみに

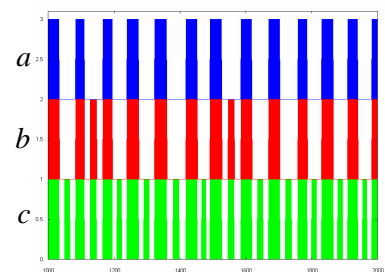


図 15 Results

- (a) Collect Voice
- (b) Proposed System
- (c) Speech processing

よる検出結果は、全音声区間検出数 (Detect All) が 200、ドライバーの真の発話区間の検出数 (Detect True) は 100 であっ

表 1 Results using proposed System

Speaker ID	Detect All	Detect True	Recall	Precision
Man	106	100	100	94.33
Woman	118	100	100	84.75

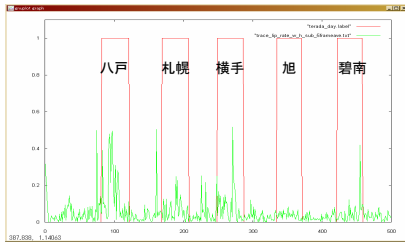


図 16 Changes of Aspect Ratio with test data



図 17 EBGM Error

たことから，正解率 (Recall) は 100% ，適合率 (Precision) は 50% であった．

正解率，適合率は以下の式で求めた．

$$\text{正解率} = \frac{\text{Detect True}}{100} \times 100(\%)$$

$$\text{適合率} = \frac{\text{Detect True}}{\text{Detect All}} \times 100(\%)$$

5.2 提案手法による実験結果

表 1 に，提案手法によって発話区間を検出した結果を示す．今回の実験では閾値を手動で設定した．図 16 に示すように，ドライバーの音声区間検出を行う際，“旭”などのように，唇の動きが小さい単語などではアスペクト比変動量が少なくなり，発話区間を検出することが困難になる．また，今回 EBGM の Bunch Graph 作成に使用した画像は，人の正面顔に限定しているため，発話区間検出の対象者が正面を向いていない場合に，グラフの検出を失敗していることがあった．例を図 17 に示す．各方向を向いた顔画像を用いて，Bunch Graph を作成することで，この問題を解決することができると考えられる．

6. ま と め

本研究では，車内という限定された環境下において，赤外線映像と音響信号を用いることにより，周囲からの雑音とドライバー以外の音声を除去し，効果的にドライバーの音声のみを検出するシステムの提案を行った．また，実験結果より提案手法の有効性が証明できた．唇のアスペクト比変動量の閾値を動的に決定しながら，このシステムのリアルタイム性を高めることが今後の課題である．

7. 謝 辞

本研究を進めるにあたり，富士通テン株式会社の古賀健太郎氏には，シミュレーションデータの提供をはじめとして大変お世話になりました．深く感謝いたします．また，本論文に使用

した顔画像データは財団法人ソフピアジャパンから使用許諾を受けたものです．権利者に無断で，複写，利用，配布などを行うことは禁じられています．

文 献

- [1] 村井和昌，中村哲，“画像と音声情報の併用による雑音に頑強な発話検出”，情報処理学会研究報告，音声言語情報処理 37-10 pp.55-60 (2001.7.14)
- [2] 村井和昌，野間啓介，熊谷建一，松井知子，中村哲，“口周囲画像による頑強な発話検出”，情報処理学会研究報告，音声言語情報処理 37-13 pp.73-78 (2000.12.21)
- [3] 川戸慎二郎，内海章，桑原和宏，“あなたの顔をインターフェースに実時間処理で目，鼻，口を入力デバイスに使う”，画像の認識・理解シンポジウム (MIRU2004)，pp.I-107-108，(2004.7)
- [4] 中田康之，安藤護俊，“色抽出法と固有空間法を用いた読唇処理”，信学論，D-II Vol.J85-D-II No.12 pp.1813-1822 (2002.12)
- [5] 吉永智明，田村哲嗣，岩野公，古井貞照，“横顔画像から抽出した口唇角度情報を用いたマルチモーダル音声認識”，日本音響学会 2004 年秋季研究発表会講演論文集，vol.1，3-1-19，pp.147-148 (2004-9)．
- [6] 菊池稔，木村元，小林重信，“遺伝的アルゴリズムを用いた赤外線画像からの移動目標検出システム”，信学論，D-II Vol.J84-D-II No.10 pp.2224-2233 (2001.10)
- [7] 横川勇仁，船曳信生，東野輝夫，小田政志，森悦秀，“Deformable Template マッチング法による唇輪郭抽出法の改良と歯科医療応用を目的とした評価”，信学論，D-II Vol.J86-D-II No.8 pp.1177-1185 (2003.8)
- [8] 須賀弘道，羽鳥好律，樽松明，“SNAKE を用いた顔画像からの構成部品の輪郭抽出”，信学論，A Vol.J79-A No.2 pp.298-301 (1996.2)
- [9] 柏野邦夫，ガビン スミス，村瀬洋，“ヒストグラム特徴を用いた音響信号の高速探索法-時系列アクティブ探索法-”，信学論，D-II，Vol.J82-D-II No.9 pp.1365-1373 (1999.9)
- [10] Robert E.Schapire，Yoram Singer，“Improved Boosting Algorithm Using Confidence-rated Predictions”，*Machine Learning*，37(3)：297-336，1999
- [11] Norbert Binder，Konstantin Markov，Rainer Gruhn，Satoshi Nakamura：“SPEECH NON-SPEECH SEPARATION WITH GMMS”，日本音響学会講演論文集 2001 年 10 月，pp141-142．
- [12] David S. Bolme “ELASTIC BUNCH GRAPH MATCHING ”In partial fulfillment of the requirements for the Degree of Master of Science Colorado State University Fort Collins，Colorado，Summer 2003
- [13] Laurenz Wiskott，Jean-Marc Fellous，Norbert Kruger，Christoph von der Malsburg “Face Recognition by Elastic Bunch Graph Matching ”IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE，VOL. 19，NO. 7，pp775-779，JULY 1997
- [14] DivX.Inc，“http://www.divx.com/ ”
- [15] Michael J. Lyons，Shigeru Akamatsu，Miyuki Kamachi，Jiro Gyoba Proceedings“ Coding Facial Expressions with Gabor Wavelets，”Third IEEE International Conference on Automatic Face and Gesture Recognition，April 14-16 1998，Nara Japan，IEEE Computer Society，pp. 200-205.
- [16] 増田 健，松田 博義，井上 淳一，有木 康雄，滝口 哲也，古賀健太郎，“唇領域の動静判定と音声・雑音判定の統合に基づく発話区間の検出 ”画像の認識・理解シンポジウム (MIRU2006)，pp934-939，(2006.7)