AAMによる顔方位を考慮した発話認識

駒井　祐人†　　有木　康雄††　　滝口　哲也††

† 神戸大学大学院システム情報学研究科　〒 657–8501 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　〒 657–8501 兵庫県神戸市灘区六甲台町 1–1
E-mail: †komai@me.cs.scitec.kobe-u.ac.jp, ††{ariki,takigu}@kobe-u.ac.jp

あらまし　唇の動きから発話内容を認識するリップリーディングは，雑音に影響を受けないため，雑音環境下での認識が可能である．そのため，雑音環境下で頑健に発話認識を行う手法の一つとして，音声情報に唇動画像情報を併用して認識を行うマルチモーダル音声認識が注目され，研究が進められている．しかし，リップリーディングを行う際，顔が横を向いてしまうと，カメラから見たときの唇の形が変わってしまい，認識精度が大きく劣化してしまう．そのため，従来のリップリーディングは正面顔での発話に限定されていることが多かった．本研究では Active Appearance Model を用いることで，様々な角度の顔方位を正面に変換してリップリーディングを行う手法を提案する．実験の結果，顔方位が 30 ℃程度変わっても，飛躍的に認識率が改善されたことが確認できた．

キーワード　リップリーディング，Active Appearance Model，Combined パラメータ，, 顔方位推定, 顔方位変換

1. はじめに

近年，音声認識はパソコンや携帯電話のソフトとして広まり，キーボードからの入力に代わるハンズフリーな技術として注目を集めている．しかし，現在の音声認識技術は，雑音環境下では認識性能が著しく低下してしまうという問題点があり，音声認識の環境適応に向けて大きな課題となっている．

一方，人間は発話内容を理解する際，種々の情報を総合的に利用している．例えば，音声が聞き取り難い場合，発話者の唇の動きに注目して発話内容を理解しようとし，逆に，唇の動きと音声が一致しない場合には，唇の動きに影響されて発話内容を誤って理解してしまうこともある．これは，McGurk effect(マガーク効果) と呼ばれ，音韻知覚が，音声の聴覚情報のみで決まるのではなく，唇の動きといった視覚情報からも影響を受けることが報告されている．このように，人間による発話内容の理解には，唇の動き情報と音声情報の統合的利用が極めて重要である．

唇の動きから発話内容を読み取る技術は読唇 (リップリーディング) と呼ばれ，聴覚障害者のコミュニケーションの手段の一つとして用いられている．音声認識は雑音に弱く，雑音環境下では認識率が著しく低下するが，リップリーディングは雑音に影響されることがない．そこで，雑音環境下で頑健に音声認識を行う手法として，唇の動き情報と音声情報を統合したマルチモーダル音声認識が注目され，近年研究が進められている [1].

しかし，リップリーディングを行う際，顔が横を向いてしまうと，カメラから見たときの唇の形が変わってしまい，認識精度が大きく劣化してしまう．そのため，従来のリップリーディングは正面顔での発話に限定されていることが多かった．

そこで，本研究は，任意の角度の顔方位からの発話を，正面からの発話と同程度の精度で認識することを目的とする．

1.1 本論文の構成

本論文は次のように構成されている．2. で関連研究について述べ，3. で提案手法を述べる．4. で提案手法に対する有効性を確認する実験を行い，最後に 5. で本論文をまとめる．

2. 関連研究

リップリーディングを行う際には，まず唇領域の抽出が必要である．その後，唇の特徴量抽出を行う．

唇領域の抽出に関しては，色情報 [2] や，エッジ情報 [3], Haar-like 特徴量 [4] といったアピアランスベースの抽出法や，Snake [5]，Active Shape Models(ASM) [6]，Active Appearance Models(AAM) [7] といったモデルベースの抽出法などがある．

特徴量に関しても，主成分スコア [8] や DCT [9] といった画素ベースの特徴量や，唇の幅や高さ [10] といった輪郭ベースの特徴量，また，画素情報と輪郭情報の両方を用いたもの [7] [11] など多くの手法が用いられている．

顔方位を考慮した研究に関しては，方位毎に音響モデルを作成し，各々の方位ごとに認識する手法 [12] や，横顔から正面顔への変換行列を学習し，横顔を正面顔に変換して認識する手法 [13] などがある．しかし，これらの手法では，方位ごとにモデルを用意したり，方位ごとの変換行列が必要であり，任意の顔方位に対して認識することができないという問題点がある．

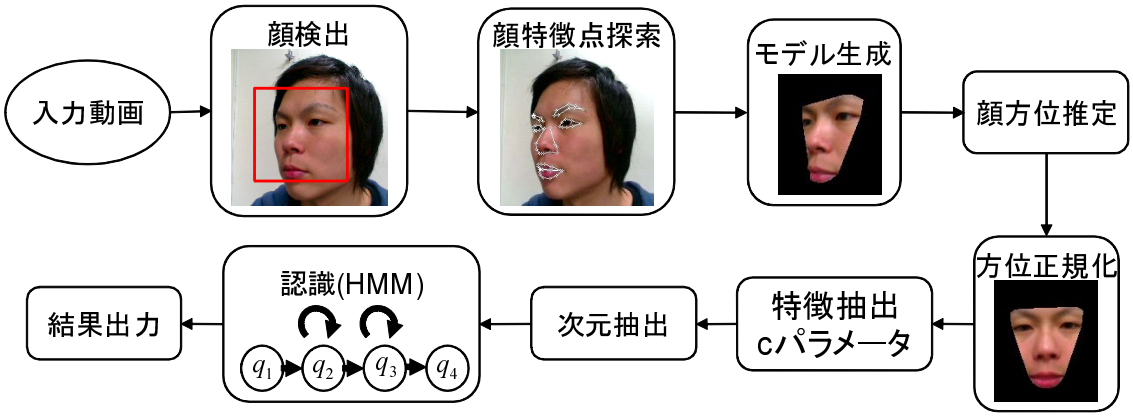


図 1　提案手法の流れ

3. 提 案 手 法

本研究では，Active Appearance Models(AAM) を用いることで，唇領域を自動的に抽出し，顔が横を向いてしまっていても，横顔を正面顔に変換することで，任意の横顔に対して，正面顔と同じように発話内容の認識を行う手法を提案する．

3.1 提案手法の流れ

図 1 に手法の流れを示す．まず，入力動画に対して Haar-like 特徴を用いた AdaBoost 法による顔領域検出を行う．これは AAM による特徴点探索では，特徴座標点の抽出精度が AAM の初期探索点に大きく依存するため，検出した顔領域を AAM の初期探索点として与えることで，特徴座標点の抽出精度が向上するためである．

次に検出した顔領域に対して AAM を適用する．AAM によって顔特徴点の探索を行い，入力画像に最も近いモデルのパラメータを生成する．生成したパラメータから，3.6 で述べる手法を用いて，発話者の顔方位を推定する．方位推定後，3.7 で述べる手法により，推定した角度の分だけ顔方位を正面に正規化することで，横顔を正面顔に変換する．その後，特徴量を抽出し，HMM で発話認識を行う．

特徴量は，唇の画素情報と輪郭情報の両方を含む，Combined パラメータ (**c** パラメータ) [7] [11] を用いる．本研究では，顔方位を正確に推定するため，顔領域の AAM を用いているが，顔領域の AAM を用いると，**c** パラメータには唇以外の顔情報も含まれてしまう．そのため，方位正規化後，**c** パラメータの中で唇の情報が集約されている次元を抽出し，認識を行う．認識はワード型 HMM を用いる．

3.2 AAM の構築

この節では，Active Appearance Models(AAM) について説明する．AAM は，特徴点の形状である shape と特徴点の輝度値である texture を主成分分析して部分空間を構成し，比較的低次元なパラメータにより顔モデルを表現する手法である．

顔画像の各点の特徴点座標を並べた shape ベクトルを **s** と置き，学習画像に与えられたベクトル **s** を正規化することで，学習画像集合から平均形状 $\bar{\mathbf{s}}$ を求める．また，**s** の内部の texture を平均形状に正規化し，その輝度値を並べた texture ベクトルを **g** とすると，**s**，**g** は，式 (1)，(2) のように与えられる

$$\mathbf{s} = (x_1, y_1, \cdots, x_n, y_n)^T \tag{1}$$

$$\mathbf{g} = (g_1, \cdots, g_m)^T \tag{2}$$

ここで，x_i, y_i $(i \leqq n)$ は各特徴点の座標を表している．g_j $(j \leqq m)$ は，平均形状 $\bar{\mathbf{s}}$ に画像を正規化したときの $\bar{\mathbf{s}}$ 内部での各画素の輝度値であり，学習画像集合から平均輝度値 $\bar{\mathbf{g}}$ を求めることができる．**s**，**g** は，$\bar{\mathbf{s}}$，$\bar{\mathbf{g}}$ からの偏差を主成分分析して得られる固有ベクトル $\mathbf{P_s}$，$\mathbf{P_g}$ を用いて，式 (3)，(4) のように表すことができる．

$$\mathbf{s} = \bar{\mathbf{s}} + \mathbf{P_s b_s} \tag{3}$$

$$\mathbf{g} = \bar{\mathbf{g}} + \mathbf{P_g b_g} \tag{4}$$

$\mathbf{b_s}$，$\mathbf{b_g}$ はそれぞれ shape パラメータ，texture パラメータと呼ばれ，平均からの変化を表すパラメータであり，これらを変化させることで shape と texture を変化させることができる．また，shape と texture に相関があることから，$\mathbf{b_s}$ と $\mathbf{b_g}$ をさらに主成分分析することで，式 (5)，(6) のように表現できる．

$$\mathbf{b} = \begin{pmatrix} \mathbf{W_s b_s} \\ \mathbf{b_g} \end{pmatrix} = \begin{pmatrix} \mathbf{W_s P_s}^T(\mathbf{s} - \bar{\mathbf{s}}) \\ \mathbf{P_g}^T(\mathbf{g} - \bar{\mathbf{g}}) \end{pmatrix} = \mathbf{Qc} \tag{5}$$

$$\mathbf{Q} = \begin{pmatrix} \mathbf{Q_s} \\ \mathbf{Q_g} \end{pmatrix} \tag{6}$$

ここで，$\mathbf{W_s}$ は shape ベクトルと texture ベクトルの単位の違いを正規化する行列，**Q** は固有ベクトル行列，**c** は shape と texture の両方を制御するパラメータで combined パラメータと呼ばれる．**c** を用いて **s**，**g** を表現す

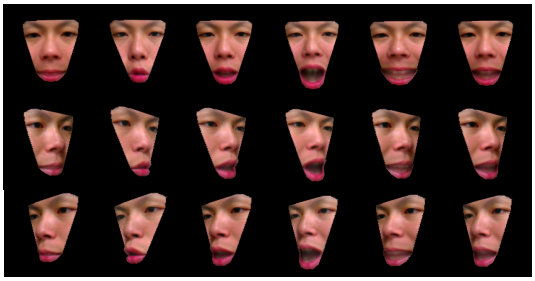
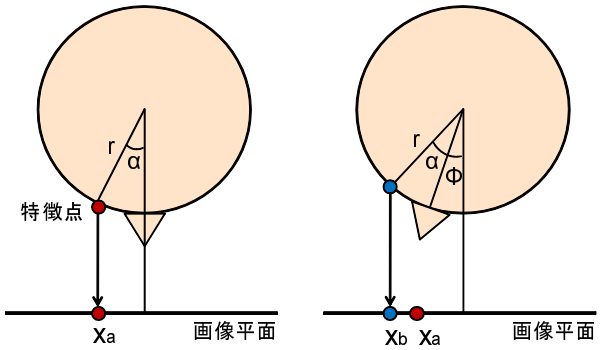


図 2　生成されるモデル画像の例



(a) 顔方位 = 0　　(b) 顔方位 = Φ

図 3　顔を頭上から見たときの概略図

ると式 (7)，(8) のようになる．

$$\mathbf{s(c)} \;=\; \bar{\mathbf{s}} \;+\; \mathbf{P_s W_s}^{-1}\mathbf{Q_s c} \tag{7}$$

$$\mathbf{g(c)} \;=\; \bar{\mathbf{g}} \;+\; \mathbf{P_g Q_g c} \tag{8}$$

このようにして，パラメータベクトル $\mathbf{c}$ を制御することによって，shape と texture を同時に扱い，顔の変化を表現することが可能となる．

3.3　AAM の探索

AAM の学習データとして，口の開閉や顔方位が含まれている画像を用いた場合，図 2 に示すように $\mathbf{c}$ を変化させる事により，多様な唇の動きが表現できることが分かる．$\mathbf{c}$ の抽出法としては，入力画像 $\mathbf{I}$ をアフィン変換させて得られる画像を $\mathbf{I(W(p))}$ とすると，AAM から生成した画像（これをモデル画像と呼ぶ）$\mathbf{g(c)}$ との誤差 $\mathbf{e}$ は式 (9) のようになり，$\mathbf{e}$ が最小となるように $\mathbf{c}$ と $\mathbf{p}$ を最急降下法によって求める．

$$\mathbf{e(c,\,p)} \;=\; \|\mathbf{g(c)} \;-\; \mathbf{I(W(p))}\|^2 \tag{9}$$

ただし，$\mathbf{p}$ はアフィン変換するための拡大縮小，回転，平行移動に関するパラメータであり，$\mathbf{W}$ はアフィン変換を実行する関数である．$\mathbf{c}$ の次元数は shape と texture の主成分分析の累積寄与率が 95% となるように計算しているため，特徴点の個数と学習画像の枚数によって可変である．

3.4　顔方位正規化のための回帰モデル

本研究では，Cootes らが提案している顔方位の正規化手法 [14] を用いる．

図 3 は，顔を半径 r の球体とみなし，頭部方向から見たときの概略図である．顔の中心から画像平面に垂線を引くと，垂線から角度 α にある顔特徴点は，図 3(a) のように，画像平面上の座標 x_a に射影される．また，顔を図 3(b) のように ϕ だけ動かしたとき，顔特徴点の位置は，画像平面上で x_b に射影される．このときの特徴点座標の変化を Δx とすると，

$$\begin{aligned}\Delta x \;&=\; x_b - x_a \\ &=\; r\sin(\phi+\alpha) - r\sin\alpha \\ &=\; r\sin\phi\cos\alpha + r\cos\phi\sin\alpha - r\sin\alpha \end{aligned} \tag{10}$$

となる．$\mathbf{c}$ パラメータはこの Δx と比例関係にあると考え，式 (10) の r，α を定数とみなせば，式 (11) のような回帰モデルを設定できる．

$$\mathbf{c} = \mathbf{c_0} + \mathbf{c_1}\cos\phi + \mathbf{c_2}\sin\phi \tag{11}$$

ここで，$\mathbf{c_0}$，$\mathbf{c_1}$，$\mathbf{c_2}$ は学習データから推定される回帰係数ベクトルである．

3.5　回帰係数の学習

前節で設定した式 (11) の回帰係数を学習する方法について説明する．学習画像の数を m 枚とし，それぞれの画像に対して AAM を適用して求めたパラメータ $\mathbf{c^i}$ と，人間が向いている方向の真の値 ϕ^i が与えられているとする (i≦k)．これらの学習用パラメータの組を用いて，最小二乗法で適切な $\mathbf{c_0}$，$\mathbf{c_1}$，$\mathbf{c_2}$ を求める．

式 (11) を行列表現に変形すると，式 (12) になる．

$$\mathbf{c} \;=\; (\mathbf{c_0}\;\;\mathbf{c_1}\;\;\mathbf{c_2}) \begin{pmatrix} 1 \\ \cos\phi \\ \sin\phi \end{pmatrix} \tag{12}$$

式 (12) に対して，学習用パラメータとして与えられた $\mathbf{c^i}$ と ϕ^i を列に並べると，式 (13) のようになり，さらに計算すると式 (14) となる．このようにして得られる回帰係数は最小二乗法で得られるものと一致する．

$$\begin{aligned}&(\,\mathbf{c^0}\;\;\mathbf{c^1}\cdots\mathbf{c^k}) \;=\; \\ &(\,\mathbf{c_0}\;\;\mathbf{c_1}\;\;\mathbf{c_2}) \begin{pmatrix} 1 & 1 & \cdots & 1 \\ \cos\phi^1 & \cos\phi^2 & \cdots & \cos\phi^k \\ \sin\phi^1 & \sin\phi^2 & \cdots & \sin\phi^k \end{pmatrix}\end{aligned} \tag{13}$$

$$(\mathbf{c_0}\;\;\mathbf{c_1}\;\;\mathbf{c_2}) \;=\; (\mathbf{c^0}\;\;\mathbf{c^1}\cdots\mathbf{c^k})\mathbf{A}^+ \tag{14}$$

ただし，$\mathbf{A}^+$ は行列

$$\begin{pmatrix} 1 & 1 & \cdots & 1 \\ \cos\phi^1 & \cos\phi^2 & \cdots & \cos\phi^k \\ \sin\phi^1 & \sin\phi^2 & \cdots & \sin\phi^k \end{pmatrix}$$

の疑似逆行列である．

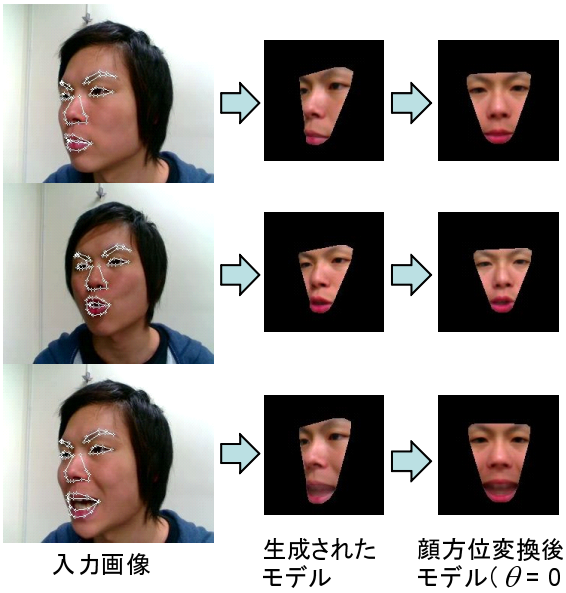


図 4　顔方位の変換例

3.6　顔方位推定

前節で求めた回帰係数と式 (11) から，顔方位を推定する方法について述べる．顔方位が未知の入力画像に対して AAM を適用し，特徴点探索によりパラメータ $\mathbf{c}'$ が得られたとき，顔の推定方向 ϕ は式 (11) を変形して，式 (15) のように求めることができる．

$$\begin{pmatrix} \cos\phi \\ \sin\phi \end{pmatrix} = \mathbf{B}^{+}(\mathbf{c}' - \mathbf{c_0}) \tag{15}$$

ただし，$\mathbf{B}^{+}$ は $(\mathbf{c_1}\ \ \mathbf{c_2})$ の疑似逆行列である．よって，式 (15) の $\cos\phi$，$\sin\phi$ から，推定角度 ϕ を式 (16) で求めることができる．

$$\phi = \tan^{-1}\left(\frac{\sin\phi}{\cos\phi}\right) \tag{16}$$

3.7　顔方位の正規化

前節で求めた推定角度から，顔方位を正面顔に正規化する方法について述べる．顔方位が未知の入力画像に対して AAM を適用し，特徴点探索によりパラメータ $\mathbf{c}'$ が得られ，式 (16) により，顔方位 ϕ が得られたとき，式 (17) の残差ベクトル $\mathbf{c_{res}}$ が考えられる．

$$\mathbf{c_{res}} = \mathbf{c}' - (\mathbf{c_0} + \mathbf{c_1}\cos\phi + \mathbf{c_2}\sin\phi) \tag{17}$$

ここで，新しい角度 θ に顔方位を変換したい場合，式 (11)，残差ベクトル $\mathbf{c_{res}}$ より，式 (18) となる．

$$\mathbf{c_{new}} = \mathbf{c_0} + \mathbf{c_1}\cos\theta + \mathbf{c_2}\sin\theta + \mathbf{c_{res}} \tag{18}$$

式 (18) のパラメータ $\mathbf{c_{new}}$ を用いて，顔方位を変換した結果を図 4 に示す．

3.8　特徴量抽出

特徴量は，唇の詳細な形状と輝度値に関する情報を含んでいる $\mathbf{c}$ パラメータを用いる．本研究では，顔方位を正確に推定するため，顔領域の AAM を用いているが，顔領域の AAM を用いると，$\mathbf{c}$ パラメータには唇以外の顔情報も含まれてしまう．そのため，方位正規化後，$\mathbf{c}$ パラメータの唇の情報が集約されている次元を抽出する．

次元の抽出法は，$\mathbf{c}$ パラメータの次元の全組み合わせに対して正面方向の発話を認識し，最も認識精度が高かった次元の組み合わせを選択する．

4.　実　　験

提案手法の有効性を確認するため，以下の 4 つの実験を行った．

- 方位正規化を行わずに，正面顔と横顔のデータベースを用いて特徴量 ($\mathbf{c}$ パラメータ) を抽出する．正面顔で発話した特徴量を HMM で学習し，正面での発話と，横顔での発話に対して認識を行う．(実験 1)
- 3.7 で述べた顔方位の正規化を行い，正面顔と横顔のデータベースを用いて特徴量を抽出する．正面顔で発話した特徴量を HMM で学習し，正面での発話と，横顔での発話に対して認識を行う．(実験 2)
- 実験 2 と同様に顔方位の正規化を行い，正面顔と横顔のデータベースを用いて特徴量を抽出する．特徴量の唇の動きの情報が集約された次元を抽出し，その次元で学習，認識を行う．(実験 3)
- 方位正規化を行わずに，特徴量の唇の動きの情報が集約された次元を抽出し，その次元で学習，認識を行う．(実験 4)

実験 1 と実験 2 は，顔方位の正規化の効果を調べるために行い，実験 3 は唇の情報を含んだ次元を抽出する効果を調べるために行う．また, 実験 3 で抽出した次元に唇の情報のみが含まれているならば，顔方位を正規化する必要がないと考えられるため，実験 3 の手法の有効性を確認するため，実験 4 を行う．

4.1　実 験 条 件

本研究では，発話単語として ATR 音素バランス単語からランダムに選択した 10 単語を用いた．撮影機器は Logicool Qcam Orbit MP で，解像度は 320 × 240 画素，フレームレートは 30fps とした．カメラからの距離は約 60cm で固定し，特定話者 1 名に正面，左右 15 ℃，左右 30 ℃ではっきりとした口調で発話させた．また時期差は考慮しないため全て同一時期に撮影した．実験は，10 単語× 10 セットに対して leave-one-out 法を適用し，9 セットで学習，1 セットを認識して，10 セットの平均を認識率とした．発話数が少量であり，全音素を十分に学習することができないため，認識はワード型 HMM を用いた．HMM の状態数は 5，混合数は実験的に最も良い

2 混合とした.

AAM のモデルに与えた特徴点は，両目，両眉にそれぞれ 8 点，鼻に 11 点，外側の唇輪郭点に対して 12 点，内側の唇輪郭点に対して 8 点の合計 63 点である．モデルの学習画像を 90 枚用意した結果，**c** の次元数は 12 次元となり, 次元抽出を行った結果,**c** の次元数は 2 次元となった．この 2 次元の **c** と **c** の $\Delta, \Delta\Delta$ 係数，計 6 次元を最終的に画像特徴量として使用した.

4.2 実験結果

表 1　認識率 (%)(学習はすべて正面顔での発話)

	正面	左 15 ℃	左 30 ℃	右 15 ℃	右 30 ℃
実験 1	99	14	11	17	10
実験 2	98	43	11	38	18
実験 3	98	96	54	89	47
実験 4	98	95	50	77	35

表 1 に実験結果を示す．表 1 の正面は, 正面顔から発話したときの認識率, 左 15 ℃, 左 30 ℃は発話者が左方向を向いて発話したときの認識率, 右 15 ℃, 右 30 ℃は発話者が右方向を向いて発話したときの認識率を示している.

実験 1 は, 方位正規化をせずに認識をした結果であり, 正面では高い認識率を得ているのに対し，正面以外では認識率が大きく低下している．これは, 横顔で発話したときに, カメラから見たときの唇の形が変わってしまうためであると考えられる.

実験 2 は方位正規化をした後の認識率であり, 実験 1 と比べて, 左 15 ℃と右 15 ℃の認識率が 20 ポイントから 30 ポイント程度改善されている．しかし, 左 30 ℃や右 30 ℃の認識率は, 実験 1 とあまり変わらない．これは, 顔方位の角度が大きくなると,AAM の特徴点抽出の精度が悪くなることが原因の一つであると考えられる．また, 特徴量の **c** パラメータに唇以外の顔の情報が多く含まれているため, 顔方位を変換した際, 唇以外の顔の情報が少しでも崩れてしまうと, 認識率に悪影響を及ぼすことも原因であると考えられる.

実験 3 は方位正規化後, 唇の動きの情報を含んだ次元を抽出して認識を行った結果である．実験 2 と比べて, 左 15 ℃と右 15 ℃の認識率が 50 ポイント程度改善されており, 正面での認識率と同程度となっている．また, 左 30 ℃と右 30 ℃の認識率は実験 2 と比べて 20 ポイントから 30 ポイント程度改善されており, 提案手法の有効性が確認できる．しかし, 左 30 ℃と右 30 ℃の認識率は未だに正面の認識率に及ばない．これは, 顔方位の角度が大きくなると,AAM の特徴点抽出の精度が悪くなることが原因であると考えられる．AAM の顔方位がある画像への正確な適用問題は, 本研究の今後の課題の一つである.

実験 4 は, 方位正規化を行わずに, 唇の動きの情報を含んだ次元を抽出して認識を行った実験である．実験 3 と比べて, 正面以外の認識率がわずかに低下している.

AAM は主成分分析を行ってモデルを構築するため, モデルの主成分には，学習画像の変動の大きい動きが集約される．学習データとして，口の開閉や顔方位が含まれている画像を用いた場合，モデルには，唇の動きが集約された主成分や顔方位の変動が集約された主成分が含まれる．ある主成分に唇の動きのみが集約されるならば, 顔方位の正規化を行わない実験 4 の結果は実験 3 の結果と等しくなるが, 実際の実験 4 の認識率は, 実験 3 より低下している．すなわち, 唇の動きが集約された主成分にもわずかに方位の成分が含まれているということであり, 方位正規化を行ってから次元抽出を行う, 実験 3 の手法の有効性が確認できる.

5. まとめ

本研究では,AAM によって横顔からの発話を正面顔からの発話に変換し認識を行う手法を提案した．実験結果は,15 ℃程度の方位の変化であれば, 正面顔からの発話と同程度の認識精度を得ることができた．30 ℃程度の方位の変化が加わると, 正面顔からの発話と同程度までには及ばなかったが, 方位変換前に比べ, 40 ポイント程度, 認識率を改善することができた.

本研究では, 特定話者 1 名で,10 単語発話という小タスクの実験であった．今後の課題としては，単語数を増やした大タスクでの実験, 複数名での認識，顔方位のある画像への AAM の正確な適用, 自然な口調に対する認識，時期差の検討，などが挙げられる．また今回の実験はデータ数の数から，ワード型 HMM を選択したが，データ数を増やしサブワード型 HMM を用いて，学習データにない単語での実験なども課題として挙げられる.

文　献

[1] G. Potamianos, C. Neti, J. Luettin, and I. Matthews, “Audio-Visual Automatic Speech Recognition：An Overview”, In：Issues in Visual and Audio-Visual Speech Processing, G. Bailly, E. Vatikiotis-Bateson, and P. Perrier (Eds.), MIT Press (In Press), 2004.

[2] Wang Lirong, Xu Jing, Zhao Yanyan, “Research of Visual Features Detection and Tracking Methods about Audio-Visual Bimodal Speech Recognition”, 2010 International Forum on Information Technology and Applications(IFITA), pp.204-207, 2010.

[3] Ye-Peng Guan , “Automatic extraction of lip based on wavelet edge detection”, Proceedings of the Eighth International Symposium on Symbolic and Numeric Algorithms for Scientific Computing(SYNASC ’06), pp.125-132, 2006.

[4] Lirong Wang, Xiaoli Wang, Jing Xu, “Lip Detection and Tracking Using Variance Based Haar-like Features and Kalman filter”, 2010 Fifth International Conference on Frontier of Computer Science and Technology, pp.608-612, 2010.

[5] Matthew Ramage, Euan Lindsay, “WRAPPING SNAKES FOR IMPROVED LIP SEGMENTATION”, 2009 IEEE International Conference on Acoustics, Speech and Signal Processing(ICASSP ’09), pp.1205-1208, 2009.

[6] K.L. Sum, WH. Lau, S.H. Leung, Alan WC. Liew

and K. W Tse, “A new optimization procedure for extracting the point-based lip contour using active shape model”, 2001 IEEE International Conference on Acoustics, Speech, and Signal Processing(ICASSP ’01), pp.1485-1488, 2001.

[7] 駒井祐人，宮本千琴，滝口哲也，有木康雄，“唇領域のAAMを用いた発話認識における画像特徴量の音素解析”，画像の認識・理解シンポジウム，MIRU2010，IS3-31，pp.1771-1778，July, 2010.

[8] M.J.Tomlinson, M.J.Russell, and N.M.Brooke, “Integrating audio and visual information to provide highly robust speech recognition”, 1996 IEEE International Conference on Acoustics, Speech, and Signal Processing(ICASSP ’96), pp.821-824, 1996.

[9] He Jun , Zhang Hua, “Research on Visual Speech Feature Extraction”, 2009 International Conference on Computer Engineering and Technology, pp.499-502, 2009.

[10] Takami Yoshida, Kazuhiro Nakadai, “Audio-Visual Speech Recognition System for a Robot”, International Comference on Auditory-Visual Speech Processing(AVSP2010), pp.8-13, 2010.

[11] Yuxuan Lan, Barry-John Theobald, Richard Harvey, Eng-Jon Ong and Richard Bowden, “Improving Visual Features for Lip-reading”, International Comference on Auditory-Visual Speech Processing(AVSP2010), pp.142-147, 2010.

[12] 山口健，山本俊一，駒谷和範，尾形哲也，奥乃博，“多方向の唇画像を利用した音声認識”, JSAI2004 (人工知能学会全国大会), 1E2-02, pp.1-4, 2004.

[13] Patrick Lucey, Gerasimos Potamianos, Sridha Sridharan, “An Extended Pose-Invariant Lipreading System”, International Comference on Auditory-Visual Speech Processing(AVSP2007), 2007.

[14] T.F.Cootes, K Walker, C.J.Taylor, “View-based Active Appearance Models”, Image and Vision Computing 20(2002), pp.657-664, 2002.