TR－SLT－0006

データベースに基づく顔向き推定

Estimation of Facial Orientation based on Database

村井和昌
加藤秀和†

2002.2.26

近年，音声情報に加え，唇動画像の情報を利用するマルチモーダル音声認識は，雑音環境下での認識性能向上を目的として，多くの研究が行われている．雑音環境下での音声認識を困難にしている要因の一つとして，マイクから入力された音が音声か雑音かを識別できず，雑音をも認識しようとすることが挙げられる．一方，雑音がない環境下においても，認識して欲しい音声かそうでない音声かどうかを識別できないといった問題がある．

そこで本研究では，認識対象者が認識機を向いて話しているとき，認識機を作動させることを前提として，顔画像情報を利用した顔向き推定を目的とする．

本研究の顔向き推定は，予め用意しておいた複数の方向を向いた顔領域テンプレートの相関を算出することによって行う．

（株）国際電気通信基礎技術研究所
音声言語コミュニケーション研究所
〒619－0288 京都府相楽郡精華町光台二丁目 2 番地 2　TEL：0774－95－1301

Advanced Telecommunication Research Institute International
Spoken Language Translation Research Laboratories
2-2-2 Hikaridai Seika-cho Soraku-gun Kyoto 619-0288,Japan
Telephone:+81-774-95-1301
Fax　　:+81-774-95-1308
†奈良先端技術大学院大学 情報科学研究科

1 はじめに

近年，音声情報に加え，唇動画像の情報を利用するマルチモーダル音声認識は，雑音環境下での認識性能向上を目的として，多くの研究が行われている．雑音環境下での音声認識を困難にしている要因の一つとして，マイクから入力された音が音声か雑音かを識別できず，雑音をも認識しようとすることが挙げられる．一方，雑音がない環境下においても，認識して欲しい音声かそうでない音声かどうかを識別できないといった問題がある．

そこで本研究では，認識対象者が認識機を向いて話しているとき，認識機を作動させることを前提として，顔画像情報を利用した顔向き推定を目的とする．

本研究の顔向き推定は，予め用意しておいた複数の方向を向いた顔領域テンプレートとの相関を算出することによって行う．

以下，第2章で本研究の顔向き推定のアルゴリズムを述べた後，第3,4章で提案の顔向き推定の評価を行う．

2 顔向き推定のアルゴリズム

一般に顔の各器官は，水平方向の成分が多いことに着目し，本研究では，顔向き推定に顔の水平方向の特徴（輪郭）を利用した．顔向き推定は以下の手順による．

1) 予め複数人の正面，5°,10°,20°,45°の方向を向いた静止画像から水平方向の輪郭を抽出して顔領域を切り出したテンプレートを用意しておく．
2) 推定対象画像に対し，水平方向の輪郭を抽出する．
3) 2)の画像に対し，用意しておいたテンプレートでテンプレートマッチングを行い，顔を探索すると同時に，最も畳み込み値が大きくなるテンプレートを見つけ，その方向を推定方向とする．

2.1 輪郭抽出

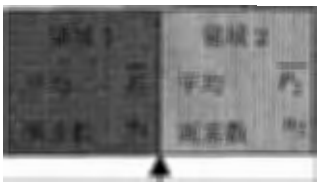
水平方向の輪郭抽出には，福井らによって提案された領域間の分離度に基づいた手法[2][3]を導入した．この手法では，領域と領域を最も分離する領域境界を輪郭として捉えるので，ノイズや照明の影響に対してロバストである．分離度は次式で与えられる．

$$\eta = \frac{{\sigma_b}^2}{{\sigma_T}^2}$$

$${\sigma_b}^2 = n_1(\overline{P_1} - \overline{P_m})^2 + n_2(\overline{P_2} - \overline{P_m})^2$$

$${\sigma_T}^2 = \sum_{i=0}^{N-1}(P_i - \overline{P_m})^2$$

ここで，N は矩形領域内の全画素数，n_1 は領域1に含まれる画素数，n_2 は領域2に含まれる画素数，P_i は位置 i における画像の特徴量，$\overline{P_m}$ は領域全体での画像の特徴量の平均値，$\overline{P_1}$ は領域1での画像の特徴量の平均値，$\overline{P_2}$ は領域2での画像の特徴量の平均値を示している．図2に示すように，マスクの形状，大きさを設定し，そのマスク内で上述した式を計算し，分離度を求める．本研究では輝度を画像の特徴量として利用し，マスクは水平方向の輪郭を抽出するため，図2の形状を90度回転させた形状とした．



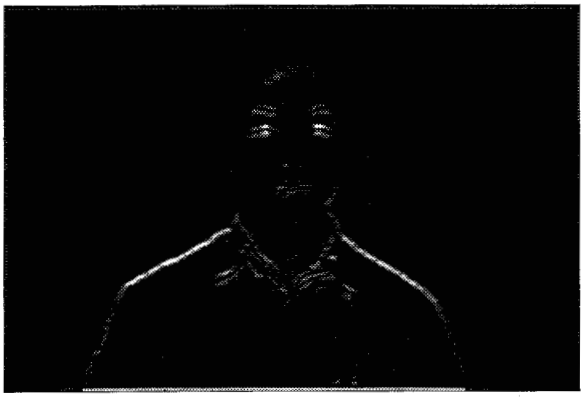
全体平均 $\overline{P_m}$
全体分散 σ_T
全画素数 $N = n_1 + n_2$

境界

図 2: 分離度

分離度による輪郭抽出は，輪郭強度が 0.0 ～1.0 に正規化されるため，この範囲をグレイスケールで表せば，輪郭抽出画像が得られる．以下，図 3 a.の画像に対し，水平方向の輪郭を抽出した画像を示す．

a. 元画像



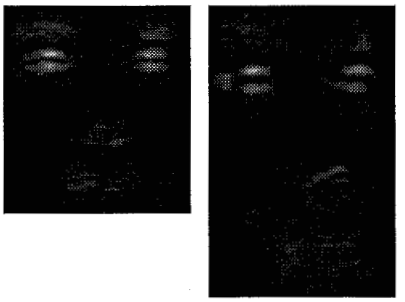
b. 水平方向の輪郭抽出画像

図 3: 分離度に基づく輪郭抽出

2.2 テンプレートの作成

黒幕の前に座った日本人男性 11 名を 5 方向（正面，5°,10°,20°,45°）から撮影した静止画（計 55 フレーム）を用いてテンプレートを作成した．カメラからの距離，照明条件は同一である．この画像に対し，両目瞳孔間の距離が 64 画素となるように拡大縮小，平行移動，回転を施し大きさ，位置を揃え，**2.1** の手法で水平方向の輪郭を抽出した後，方向ごとに輝度の平均をとり，顔領域を切り出すことにより作成した（図 4）．

正面　　5°　　10°



20°　　45°

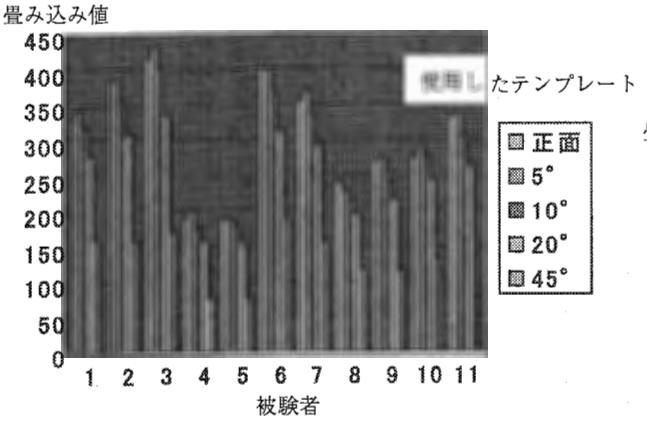
図 4: 顔向き推定テンプレート

3 実験結果

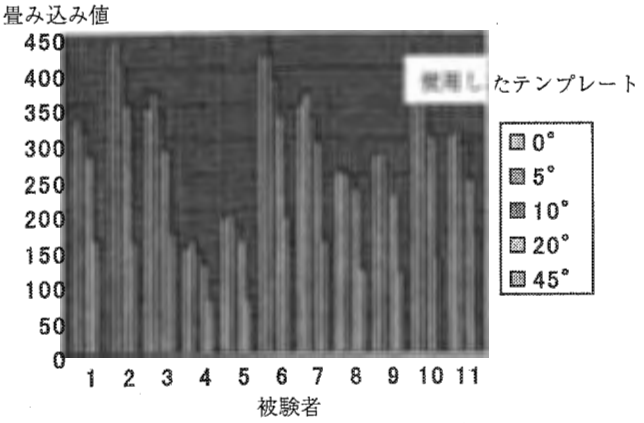
テンプレートの作成に用いた静止画（計 55 フレーム）に対して実験を行った．使用したテンプレートは実験用の画像を除く 10 名分の画像から作成した．

結果，顔の探索は，100%の検出率が得られ

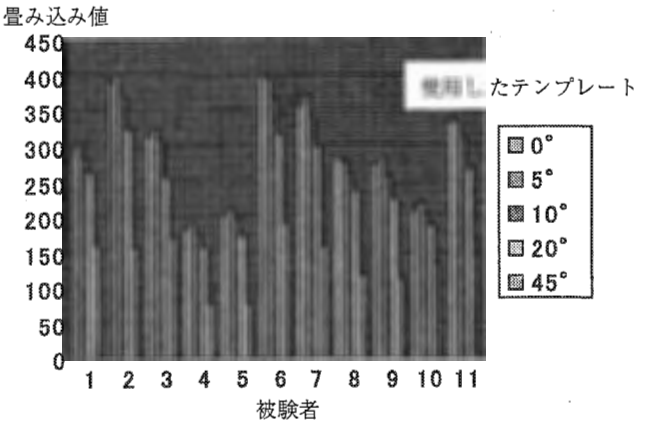
た．方向推定結果を図５に示す．図５は，推定対象画像に対する各テンプレートとの畳み込み値を示している．すなわち，畳み込み値が最も大きなテンプレートが推定した方向であることを明示している．図５より，ほとんど全ての被験者でどの方向に対しても5°の方向を推定していることがわかる．



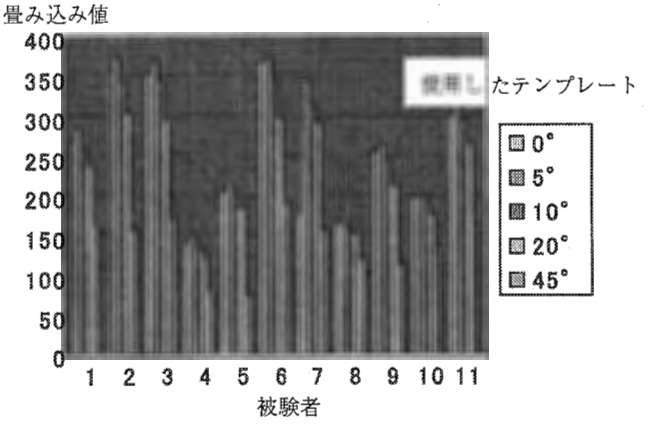
a. 正面



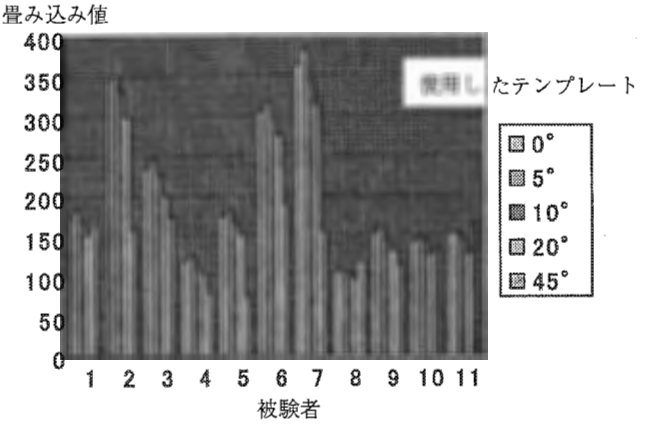
b. 5°



c. 10°



d. 20°



e. 45°

図 5: 顔向き推定の結果

4 考察

図５のように，顔向き推定がうまくできてい

ないのは，作成したテンプレートに問題があるのではないかと考えた．つまり，テンプレートごとに畳み込み値が大きくなりやすいものとそうでないものとがあるのではないかと考えた．そこで，テンプレート自身の畳み込み値を求めてみた（図 6）．

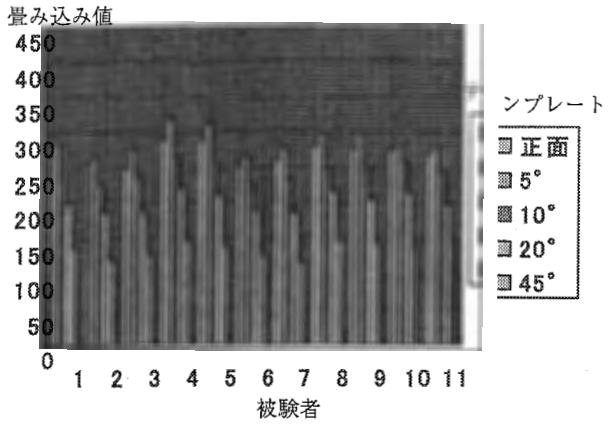


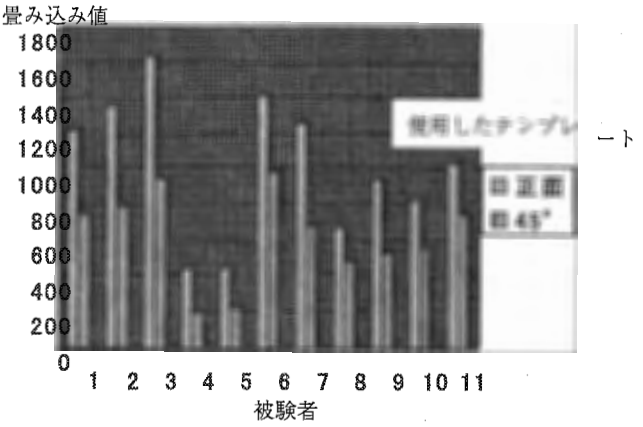
図 6: テンプレートの特性

図 6 をみると，テンプレートの特性は均一ではなく，全ての被験者において 5° のテンプレートが畳み込み値が大きくなりやすくなっていることがわかる．推定結果がほとんど 5° を推定したのは，この特性が大きく影響していたと考えられる．そこで，図 5 の結果をこのテンプレートの特性で正規化を行ったところ，表 1 に示す推定率が得られた．

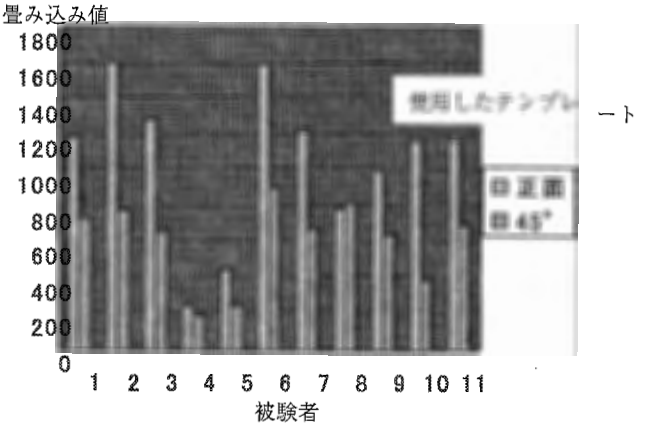
表 1: 顔向き推定率

	正規化前	正規化後
正面	1/11	0/11
5°	9/11	0/11
10°	0/11	0/11
20°	0/11	11/11
45°	0/11	11/11
平均	18.2%	40%

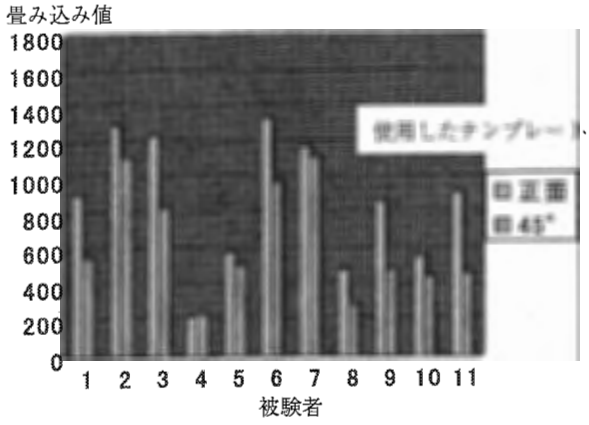
表 1 のように，正規化を行うことで正規化を行う前の 2 倍の推定率が得られた．しかし，20° ,45° の推定率が 100%なのに対し，正面～10° の推定率が 0%になっている．その原因として二つのことが考えられる．一つはテンプレートをみてわかるように，鼻と口の位置の個人差が大きくて相関が正しくとれないこと，もう一つは正面～20° のテンプレートのパターンに大差がないことである．事実，正面～10° の推定率が 0%になっているもののこれらは全て 20° 以内を推定していた．そこで，正面～20° を正面とし，正面と 45° の目領域テンプレートを利用して，この二方向で顔向き推定を行った．その結果，平均 87%（正面:44 フレーム中 42 フレーム，45° :11 フレーム中 6 フレーム）の推定率が得られた（図 7）．



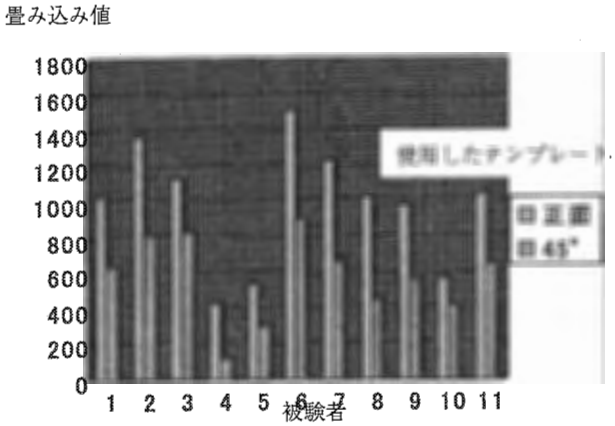
a. 正面



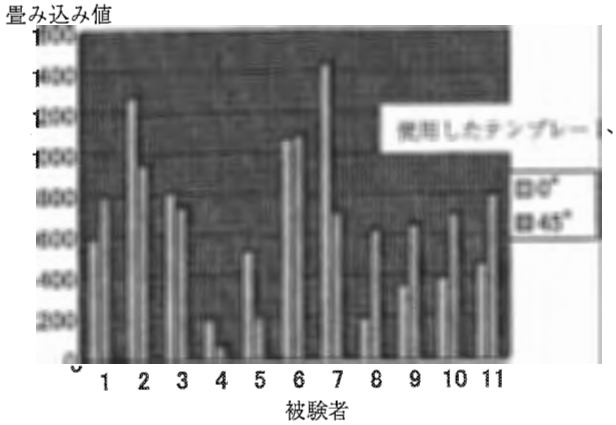
b. 5°



c. 10°



d. 20°



e. 45°

図 7: 2 方向の目領域テンプレート
を利用した顔向き推定の結果

5 結論

顔の水平方向の特徴を用いて予め用意しておいた顔画像テンプレートとのマッチングによる顔向き推定を行った．その結果，正面～20° 内で推定をすることは困難であったが 45° 程度の精度であれば，推定可能であることがわかった．ただ，本研究から正面か 45° かに推定される閾値がわからないため，今後これを調べる必要がある．

また，このシステムを音声認識の作動スイッチに用いるにはより精度の高い顔向き推定が要求される．テンプレートを作成する際のデータ量などを評価し，適切なデータ量で作成したテンプレートを使うことで性能の改善が見込まれる．

参考文献

[1] 村井和昌，中村哲，”話者の顔画像を併用した発話検出”，日本音響学会 2001 年秋期研究発表会予稿集，pp.23-24

[2] 山口修，福井和広，”分離度特徴を用いた顔画像解析”，情報処理学会第 52 回全国大会(2)，pp.187-188，1996

[3] 福井和広，”領域間の分離度を用いた輪郭抽出”，情報研報，CV88-2，pp.9-16，1994