

発話スピード変化のロバスト性を考慮した 口唇の動き特徴による個人識別

三浦 信 白澤 洋一 景山 陽一 西田 眞

秋田大学

1. はじめに

近年，入退室システムの構築や高度情報サービスの実現を目的とした個人識別法の重要性が認識されるようになり，漏洩や忘却といった人為的ミスにも堅固な個人識別法としてバイオメトリクス（生体認証）が注目されている．その中の一つに口唇の動き特徴，すなわち発話に伴う口唇の動き変化を特徴量とする識別法^[1]がある．

口唇の動き特徴は，()音声特徴など他の特徴量との併用が可能，()雑音下など音声特徴の使用が期待できない状況においても口唇の動き変化により個人識別が可能，()発話内容は可変可能であるため，他の生体情報と比較し特徴量の変更が容易である，等の利点を有している．しかしながら，個人ごとに発話内容や発話スピードは異なる場合や，発話スピードが一様であるとは限らない場合も生じる．このため，口唇の動き特徴を用いて精度良く個人識別を行うためには，識別のための規準を統一する必要がある．

本研究では，口唇の動き変化から得られる特徴量をグループ化することで，個人ごとに発話内容および発話スピードが異なる場合でも個人識別可能な手法の開発を目的とする．

2. 使用データ

本研究では同一内容を発話したデータを用いて実験を行った．すなわち，被験者 10 名(a-j)が単語「あきたたろう」を発話した時の動画を CCD ビデオカメラ(SONY 製：DCR-TRV900)により取得した．次に，得られた動画を 30 フレーム/秒の静止画像データに変換し，識別に用いた．

3. 個人識別法

3.1 特徴量算出処理

始めに，取得した各被験者の画像データから口唇の端点(上下左右の合計 4 点)を特徴点として自動抽出^[2]した．得られた上下および左右の座標値を用いて“ベクトルの大きさ”・“ベクトルの角度”を算出^[3]し，これを特徴量とした．また，発話スピード(発話フレーム数)も特徴量として用いた．

3.2 グループ化処理

発話スピード(発話フレーム数)は被験者ごとに異なる．また，同一被験者が同一内容を発話しても発話スピードは異なる．被験者 10 人を対象とした予備実験を行ったところ，同一人物が「あきたたろう」と 15 回発話したときに，平均 13.9 フレームの差異のあることが明らかとなった．

本研究では，特徴量となる口唇の動き変化の推移を保持し，かつ識別を行う上での規準を統一できるように，前後フレームの口唇の動き変化量に着目して発話データをグループ化した．具体的には，図 1 に示すようにグループ化によってフレームが余らないように 1 グループ内のフレーム数を設定した．次に，図 2 に示すようにグループ数とフレーム数の算出結果より想定される全てのグループ組み合わせにおいて，前後フレームの特徴量の変化量に着目し，グループごとの変化量を算出した．得られた変化量の中で最も顕著な値を持つ組み合わせを自動選択し，識別に用いた．

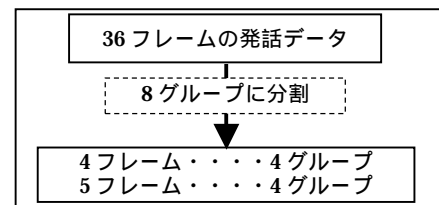


図 1 1グループ内のフレーム数の計算例

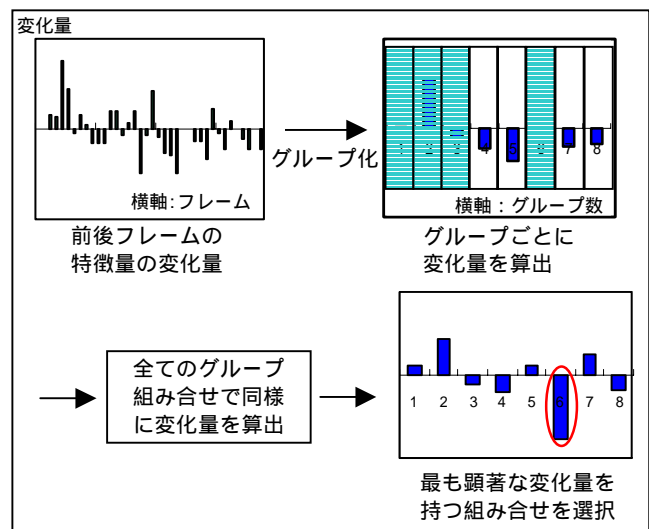


図 2 グループ組み合わせ選択のイメージ
(8グループの場合)

Robust personal identification using the lip motion features for the speech speed
Shin Miura, Yoichi Shirasawa, Yoichi Kageyama and Makoto Nishida (Akita Univ.)

3.3 発話フレーム数から得られる特徴量

発話フレーム数を特徴量として用いるためには、全体の発話フレーム数および個々の単語(例：「あ」)の発話フレーム数が異なる場合もあることを考慮する必要がある。そこで、各ベクトルのグループを構成する発話フレーム数の組み合わせ(以後、それぞれ大きさフレーム数、角度フレーム数と表記する)も特徴量として識別に用いた。

3.4 識別処理

発話内容や発話スピードが異なる場合、最適なグループ数は一意的には設定できない。また、予備実験の結果、各被験者の発話データは、本人に対してグループ数 4~10 のほぼ全ての場合で高い一致度を持つことが明らかとなった。そこで、4~10 の範囲でグループ数を設定し、全てのグループについて一致度を求めて識別を行った。識別処理の流れを図 3 に示す。始めに、各被験者の発話データを 4~10 グループにグループ化し、被験者ごとの教師データを作成した。次に、同様にグループ化された入力データに対し、ファジィ推論法^[3]を用いて一致度を求めた。最後に、得られた各グループ数における一致度に対して max-min 法を用い、各被験者への最終的な一致度を算出した。

4. シミュレーション

4.1 シミュレーションデータ

同一人物の発話データ 9 回分のうち 6 回分を教師データ、3 回分を実験データとして用いた。さらに、教師データと実験データを入れ替えて、30 回の識別処理を 3 回、合計 90 回のシミュレーションを行った。

4.2 比較に用いた手法

本研究で提案する個人識別法の有用性を検証するため、以下の 2 つの手法と比較を行った。

手法 a：グループ化を行わず、発話データごとにユークリッド距離を算出し最短距離法により識別処理を行う手法。

手法 b：データの時系列変化を伸縮して比較できる DP マッチングを用いて識別処理を行う手法。

5. 実験結果およびまとめ

各特徴量を用いて得られた識別結果を表 1 に示す。各ベクトルから得られる特徴量に加えて、大きさフレーム数および角度フレーム数を特徴量とした提案手法において最も良好な識別が行われ、最大で 93%の識別率(90 回中 84 回成功)が得られた。提案手法はグループ化した発話フレーム数による特徴量を使用しているため、発話フレーム数全体での特徴量のみを使用した場合と比較し、各被験者の特徴がグループごとに顕著に表れたと

推測される。また、大きさフレーム数および角度フレーム数から得られた特徴量に加えて発話時の全フレーム数を特徴量とした場合には、識別率はわずかながら低下した。これは、発話フレーム数全体から得られる特徴量が、グループ化した発話フレーム数から得られる特徴量を相殺したため、識別率の低下をもたらしたものと推測される。

提案手法と比較手法による識別結果を表 2 に示す。本研究で提案する手法は、手法 a, b よりも良好に識別していることが分かる。この結果は、グループ化によって算出された特徴量が、個別の発話内容を対象とする個人識別を行う上で有用であることを示唆している。

参考文献

- [1]根田, 西田, 石井, 佐藤: 「口唇の動き特徴の個人識別法への適用」, 電学論 C, Vol.120-C, No.5, pp.765-766 (2000)
- [2]白澤, 西田: 「あいまいさを考慮した色彩情報による口唇抽出アルゴリズム」, 電学論 C, Vol.123-C, No.4, pp.756-764 (2003)
- [3]水本: 「ファジィ理論とその応用」, サイエンス社 (1992)

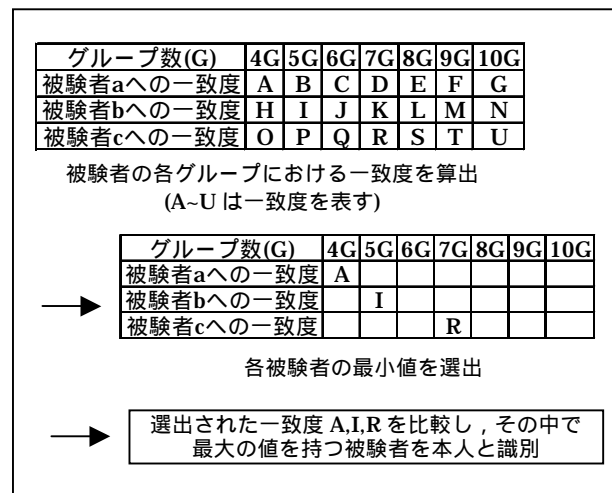


図 3 識別処理の流れ

表 1 使用特徴量別での識別結果

	標準偏差×1	×2	×3	×4	×5
角度+大きさ+					
大きさフレーム数+角度フレーム数	86%	93%	93%	92%	90%
角度+大きさ+フレーム数	87%	82%	84%	86%	86%
角度+大きさ+フレーム数+					
大きさフレーム数+角度フレーム数	82%	93%	93%	91%	90%

表 2 提案手法と比較に用いた手法による識別結果

	識別率
提案手法	93%
手法a (ユークリッド距離)	80%
手法b (DPマッチング)	82%