発話に伴う音声信号と口唇の動き特徴を用いた個人識別アルゴリズム

佐藤 慶幸　景山 陽一　西田 眞
秋田大学

1. はじめに　近年研究が行われているマルチモーダル認証技術[1]の中で，画像と音声の組み合わせは非接触でデータの取得が可能であるため，ユーザへの負担が軽いことなどの利点を有している．

一方，個人識別を行う場合，登録者以外の人物を誤受領した場合にデータの再入力を求めることは不可能であるため，未登録者を誤受領しないことが重要であると考える．これまでの研究[2]で，音声と画像に重み付けをして組み合わせることにより，誤受領数を減少可能であることを明らかにしているが，最適な重み付けの自動設定には至っていないのが現状である．

そこで本研究では，各特徴に対して重み付けすることなく，未登録者の誤受領を低減可能な個人識別手法を提案し，その有用性について検討を加えた．

2. 使用データ

2.1 発話動画像データ　本研究では，通常の室内環境下（蛍光灯による照明）かつ雑音の少ない状況下で CCD ビデオカメラ（SONY 製：DCR-TRV900）を用い，被験者 10 名（a～j）が単語「あきたたろう」と 15 回ずつ発話した発話動画像データを取得した．次に，発話動画像データを 5 回分ずつ 3 グループに分け，それぞれ教師データ，参照データ，並びに入力データとした．ここで，参照データとは後述するメンバーシップ関数の作成に用いるデータを意味している．

2.2 音声データおよび時系列顔画像データ　ビデオカメラで撮影した発話動画像データから 32kHz, 16bit の音声データを取得し，16kHz にダウンサンプリングしたデータを音声データとして解析に用いた．一方，発話動画像データを毎秒 30 フレームの静止画像に変換し，時系列顔画像データとした．

3. 個人識別アルゴリズム

3.1 特徴抽出　音声データから得られる特徴量として，離散フーリエ変換によりケプストラム係数[3] を抽出し，音声特徴として解析に用いた．また，図 1 に示すように，時系列顔画像データからはフレーム毎に口唇の横幅 diX および縦幅 diY を取得し，これらを要素とするベクトルの大きさ $|\overrightarrow{diZ}|$ を時系列顔画像データ 1 フレームから得られる口唇の動き特徴とした．

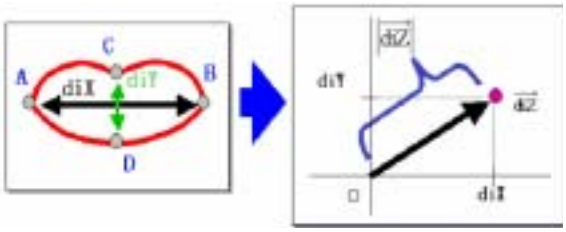


図 1　時系列顔画像から得られる特徴量

3.2 提案手法の概要　提案手法における個人識別アルゴリズムの流れを図 2 に示す．始めに，同一話者における発話時間長の伸縮[4]の影響を低減するため，音声特徴および口唇の動き特徴それぞれについて DP マッチングを用い，各データ間の距離（以下，DP 距離と略記する）を算出した．次に，同一登録者の教師データと参照データの間で算出した DP 距離の平均と標準偏差を用いて作成したメンバーシップ関数に従い，各登録者に対する入力データの帰属度を算出した．続いて，帰属度と占有率（帰属度の割合を表す値）を用いて候補者数の推定および最終候補者の推定を行う．具体的には，帰属度の閾値 T_r を越えた登録者のみを候補者として占有率の和を求め，閾値（T_s：音声特徴の占有率の閾値，T_m：口唇の動き特徴の占有率の閾値）を越えるまでの人数を候補者数 N とした．さらに，占有率が上位 N 名までの候補者を対象として占有率を再計算し，占有率の和と閾値（T_s，T_m）の比較により最終候補者を推定した．

音声特徴および口唇の動き特徴から推定された最終候補者に共通の候補者が含まれている場合，本人であると推定して出力し，共通の候補者が存在しない場合は棄却した．

4. 提案手法の評価

4.1 比較手法　入力データと教師データから算出した統合類似度により個人識別を行う手法[2]（以下，比較手法（統合）と略記する）により得られた結果との比較を行った．また，各特徴を単独で用い，DP 距離が最小となる登録者を本人と推定する手法により得られた結果との比較も行った．なお，本研究では，各特徴に均等に重み付けをした場合に得られた結果を比

An algorithm for individual discernment using sound signal and movement features of the lip in speech.
Yoshiyuki Sato, Yoichi Kageyama, Makoto Nishida (Akita Univ.)

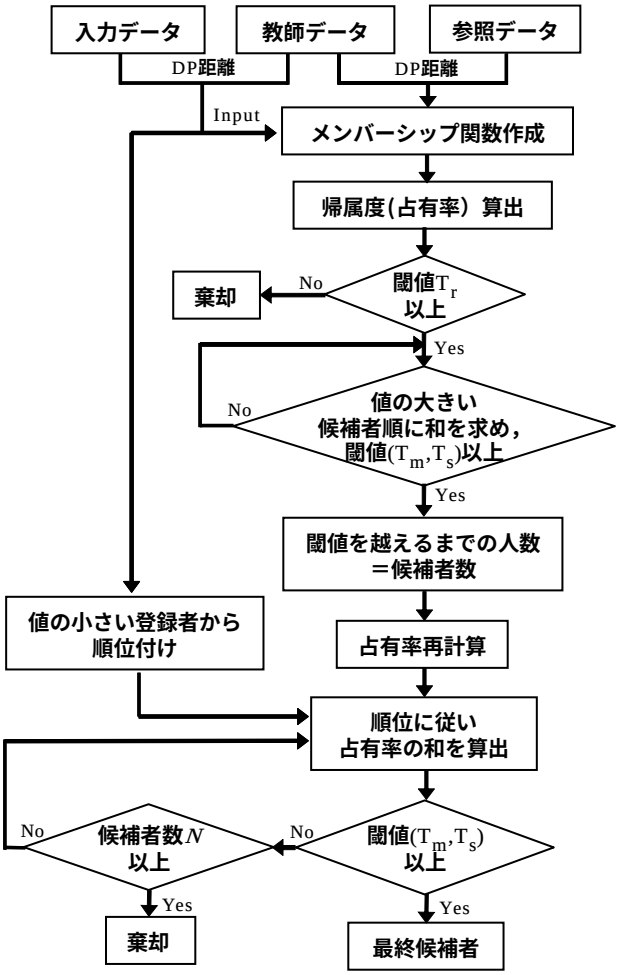


図2　個人識別アルゴリズムの流れ

較に用いた.

4.2　シミュレーション条件　本研究では，以下の条件(1), (2)によりシミュレーションを行った．(1) 被験者 10 名全員（a～j）を登録者とした場合
(2) 被験者を 5 名ずつ（a～e, i～j）登録者と未登録者に分けた場合

4.3　識別結果の評価基準　本研究では，以下に示す基準を用いて識別結果を評価した.

① 正答：登録者の入力データを正しく推定する.
② 誤識別：登録者の入力データを他の登録者として推定する.
③ 誤棄却：登録者の入力データを未登録者として推定し，棄却する.
④ 誤受領：未登録者の入力データを登録者として推定し，受領する.

5. 実験結果および検討

5.1 被験者全員を登録者とした場合の識別結果　実験結果の一例を表 1 に示す．提案手法では 86.0%（150 例中 129 例正答）の識別率が得られ，音声特徴のみを用いた場合および音声特徴と口唇の動き特徴を統合した場合よりも識別率は下回るものの，口唇特徴のみを用いた場合よりも精度良く識別している．また，本人を誤りとする誤棄却は増加したものの，誤識別を行うことなく本人を推定することが可能であった.

5.2 被験者を登録者および未登録者とした場合の識別結果　実験結果の一例を表 2 に示す．提案手法の識別率は 93.0%（300 例中 279 例正答）であり，被験者全員を登録者した場合と同様，音声特徴のみを用いた場合および音声特徴と口唇の動き特徴を統合した場合よりも識別率は低下した．しかしながら，口唇特徴のみを用いた場合よりも精度良く識別されている．一方，誤受領が認められないのは提案手法のみである．このことは，高精度の個人識別システムを検討する上で重要であると判断される.

6. まとめ

本研究では，各登録者に対する帰属度を用いて候補者数を推定した後に個人識別を行う手法を提案した．提案手法を用いた個人識別では，音声特徴および口唇の動き特徴を組み合わせるときに重み付けを考慮することなく，未登録者の誤受領を低減可能であることが明らかとなった.

表 1　個人識別結果の一例(1)

手法	正答	誤識別	誤棄却	正答率(%)
提案手法	**129**	**0**	**21**	**86.0**
比較手法（統合）	149	1	0	99.3
比較手法（音声）	150	0	0	100.0
比較手法（口唇）	114	36	0	76.0

表 2　個人識別結果の一例(2)

手法	正答	誤識別	誤棄却	誤受領	正答率(%)
提案手法	**279**	**0**	**21**	**0**	**93.0**
比較手法（統合）	286	0	9	5	95.3
比較手法（音声）	293	0	1	6	97.7
比較手法（口唇）	218	17	36	29	72.7

参考文献

[1]前田他：「顔画像特徴，歩行画像特徴および音声特徴の統合による個人識別」，信学論，D-Ⅱ，J79-D-Ⅱ，No4，pp.600-607 (1996-4)
[2]佐藤他：「音声と発話に伴う口唇の動き特徴を用いた話者認識に関する検討」，第 65 回情報処理学会全国大会，1G-3，pp.2-43 (2003)
[3]谷萩：「音声と画像のディジタル信号処理」，コロナ社 (1996)
[4] 古井：「ディジタル音声処理」，東海大学出版会 (1985)