

発話に伴う口唇の動き特徴と 雑音を含む音声信号を用いた個人識別

佐藤 慶幸[†] 西田 眞[†] 栗栖 怜史[‡] 西 健治[‡]

[†]秋田大学 [‡]株式会社アルファシステムズ

1. はじめに 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られる。

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する。具体的には、音声データに雑音を加えたデータを用いた実験を行い、提案手法の有用性について検討を加えたので報告する。

2. 使用データ

2.1 発話動画画像データ 本研究では、CCD ビデオカメラを用い、被験者 10 名、発話内容「あきたたろう」、発話回数 15 回として発話動画画像データを取得した。また、発話動画画像データを入力データ、教師データ、並びに参照データの 3 種類のデータとして使用した。なお、参照データとは後述するメンバーシップ関数作成に用いるデータである。

2.2 音声データおよび時系列顔画像データ 発話動画画像データを毎秒 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。一方、発話動画画像データから 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た。さらに、雑音が個人識別に与える影響について検討を加えるため、電子協騒音データベース^[2]（展示会場（ブース内））を用い、音声データと雑音を合成したものを解析に用いた。

3. 個人識別手法 提案手法における個人識別手法の流れを図 1 に示す。始めに、時系列顔画像データおよび音声データから口唇の動き特徴および音声特徴をそれぞれ抽出し、DP マッチン

グにより各データ間の距離（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った。具体的には、帰属度のしきい値 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、しきい値 (T_s : 音声特徴の占有率のしきい値, T_m : 口唇の動き特徴の占有率のしきい値) を越えるまで

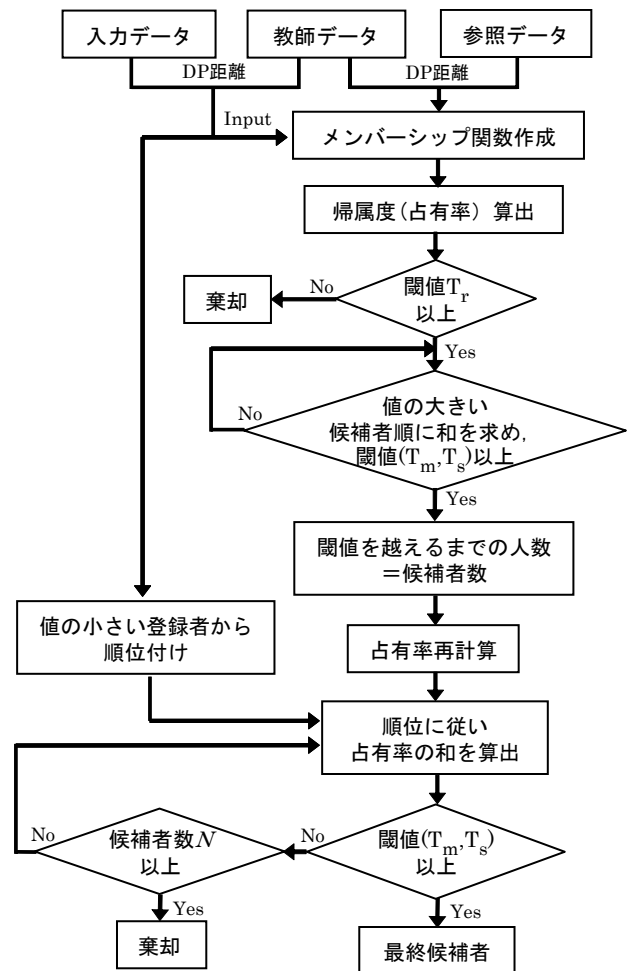


図 1 個人識別手法の流れ

An Individual Discernment Using the Movement Features of Lip in Speech and the Speech Sound with Noise.

[†]Yoshiyuki Sato, Makoto Nishida (Akita Univ.)

[‡]Satofumi Kurisu (Alpha systems Inc.)

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) を比較して最終候補者を推定した。最後に、各特徴を用いて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た。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

(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)~(4)によりシミュレーションを行った。

- (1) 被験者を5名ずつ (a~e, i~j) 登録者と未登録者に分ける。
- (2) 同一話者における発話時間長の変化³⁾を考慮し、表1に示すように、使用データの組み合わせを変更する。
- (3) 各しきい値 (T_r , T_s , T_m) の組み合わせを0.05区切りで変更 (0.00, 0.05, ..., 0.95, 1.00) する。
- (4) 全データパターンで共通のしきい値を設定した場合を“シミュレーション1”，データパターンごとに最適なしきい値を設定した場合を“シミュレーション2”とする。

5. 個人識別結果および検討

5.1 識別率を優先した場合の個人識別結果 識別率が最高となるように各しきい値を設定した場合に得られた結果の一例を表2に示す。提案手法により得られた識別率はシミュレーション1で91.7%，シミュレーション2で94.0%であった。これは、どちらの場合も全ての比較手法により得られた結果と比較して良好な結果であった。

表1 使用データパターン

発話回数	1~5	6~10	11~15
グループ化パターン			
パターン1	入力データ	教師データ	参照データ
パターン2	参照データ	入力データ	教師データ
パターン3	教師データ	参照データ	入力データ

5.2 誤受領の低減に重点を置いた個人識別結果

誤受領が最小となるように各しきい値を設定した場合に得られた結果の一例を表3に示す。提案手法により得られた識別率はシミュレーション1で84.0%，シミュレーション2で93.3%であった。これは、正答率が最高となるようにしきい値を設定した場合と同様に、どちらの場合も全ての比較手法により得られる結果と比較して良好な結果であった。

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。雑音を含む音声データを用いて実験を行った結果、提案手法により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。

以上のことは、提案手法が雑音の影響を受けにくく、高精度の個人識別システムを検討する上で有用であることを示唆するものである。

表2 識別結果の一例(1)

手法	正答	誤識別	誤棄却	誤受領	識別率 (%)
提案手法 シミュレーション1	275	0	22	3	91.7
提案手法 シミュレーション2	282	0	16	2	94.0
比較手法 (統合)	273	3	9	15	91.0
比較手法 (音声)	273	0	1	26	91.0
比較手法 (口唇)	217	13	27	43	72.3

表3 識別結果の一例(2)

手法	正答	誤識別	誤棄却	誤受領	識別率 (%)
提案手法 シミュレーション1	252	0	48	0	84.0
提案手法 シミュレーション2	280	0	20	0	93.3
比較手法 (統合)	234	0	66	0	78.0
比較手法 (音声)	203	0	97	0	67.7
比較手法 (口唇)	161	2	137	0	53.7

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]「電子協騒音データベース」，社団法人日本電子工業振興協会 (1996)
- [3]谷萩：「音声と画像のデジタル信号処理」，コロナ社 (1996)