

## 音声と口唇の動き特徴を用いたマルチモーダル個人認証

佐藤 慶幸<sup>†</sup> 西田 眞<sup>†</sup> 西 健治<sup>‡</sup>

<sup>†</sup>秋田大学 <sup>‡</sup>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<sup>[1]</sup>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。また、個人認証を行う場合、登録者以外の人物を誤って受理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、未登録者を誤受理しないことが重要である。

そこで本研究では、口唇の動き特徴及び音声特徴を用い、各特徴に対する帰属度により予め本人候補者の絞り込みを行うことで未登録者の誤受理を低減可能な個人認証法を提案する。本稿では、被験者ごとに個別の発話内容により取得したデータを用いて実験を行い、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 発話動画像データ** 本研究では、CCDビデオカメラを用い、被験者 18 名が自身の氏名を発話した動画像データを取得した。このとき、雑音を付加しない場合と、電子協騒音データベース<sup>[2]</sup>（展示会場（ブース内））を用いて雑音を付加した場合の条件下において、それぞれ 9 回分の動画像データを取得した。なお、取得した動画像データの 3 回分をそれぞれ入力データ、教師データ、並びに参照データの 3 種類のデータとして解析に用いた。使用データの組み合わせを表 1 に示す。なお、一回の発話ごとに被験者自身がクリック操作し、発話開始及び発話終了のタイミングを決定することで、動画を取得した。

**2.2 音声データ及び時系列顔画像データ** 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。一方、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た。

**3. 個人認証手法** 本研究における個人認証手法の流れを図 1 に示す。個人認証は、以下に

示す(1)~(4)の手順により行われる。

(1)特徴量算出：音声データ及び時系列顔画像データから音声特徴として口唇の縦幅と横幅の比を、口唇の動き特徴として 16 次ケプストラム係数をそれぞれ算出する。

(2)帰属度算出：特徴量抽出後、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三角型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。

(3)候補者の絞り込み及び最終候補者推定：帰属度を用いてしきい値処理により候補者の絞り込みを行った。さらに、占有率（帰属度の総和に対

表 1 使用データパターン

発話回数	1~3	4~6	7~9
使用データパターン			
パターン1	入力	教師	参照
パターン2	参照	入力	教師
パターン3	教師	参照	入力

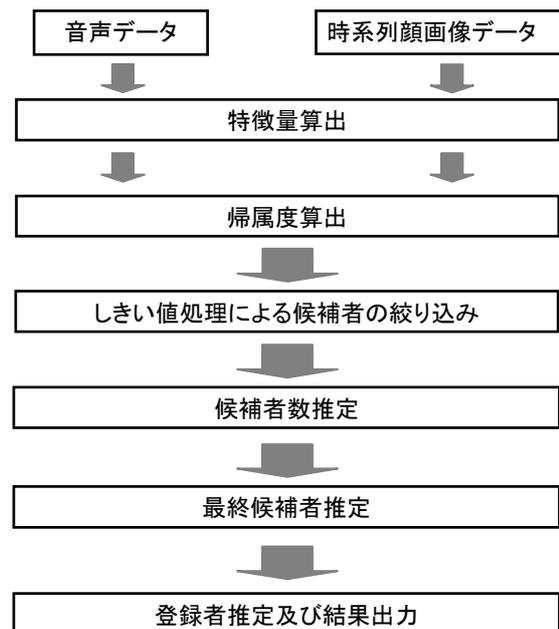


図 1 個人認証手法の流れ

A Multi-modal Individual Authentication by Using the Movement Features of Lips in Speech and the Speech Sound.

<sup>†</sup>Yoshiyuki Sato, Makoto Nishida (Akita Univ.)

<sup>‡</sup>Kenji Nishi (Alpha systems Inc.)

する各帰属度の割合)を用い、しきい値処理により候補者数の推定及び最終候補者の推定を行った。なお、帰属度のしきい値を  $T_r$ 、音声特徴の占有率のしきい値を  $T_s$ 、口唇の動き特徴の占有率のしきい値を  $T_m$  とし、各値を手動で設定した。なお、しきい値設定の詳細については4.2節で後述する。

(4)推定結果出力：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。一方、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** 音声特徴及び口唇の動き特徴を単独で用い、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を行った。

**4.2 シミュレーション条件** 本研究では、以下の条件(1)~(3)によりシミュレーションを行った。

- (1) 被験者を9名ずつ (a~i, j~r) 登録者と未登録者に分け、登録者と未登録者の入れ替えも行う。
- (2) 雑音を付加しないデータ間のシミュレーションをシミュレーション1、雑音を付加したデータ間のシミュレーションをシミュレーション2、雑音を付加したデータ及び雑音を付加しないデータを組み合わせたシミュレーションをシミュレーション3とした。
- (3) 各しきい値 ( $T_r$ ,  $T_s$ ,  $T_m$ ) の組み合わせを0.1区切りで変更する。すなわち (0.0, 0.1, ..., 0.9, 1.0) と設定した。

なお、本実験におけるシミュレーション回数は被験者数 (18名)、発話回数 (9回)、並びに登録者の入れ替えから、しきい値の組み合わせごとに324回となる。

**5. シミュレーション結果及び検討** 各シミュレーションにおいて、正答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及び表3にそれぞれ示す。

始めに、同条件で取得したデータを用いた場合 (シミュレーション1, シミュレーション2)、提案手法により得られた正答率はシミュレーション1で90.7%、シミュレーション2で80.6%であった。また、提案手法による誤受率はシミュレーション1で0.9%、シミュレーション2で1.5%であり、正答率では大きな差異が認められなかったシミュレーション2においても良好な結果が得られた。

次に、異なる条件で取得したデータを用いた場合 (シミュレーション3) に得られた結果と比較したところ、提案手法による正答率は71.3%と口唇の動き特徴を単独で用いた場合とほぼ同等であった。しかしながら、提案手法による誤受率は0.3%となり、口唇の動き特徴を単独で用いた場合と比較し良好な結果の得られていることが分かる。

以上の結果は、提案手法が未登録者の誤受理低減に有用であることを示唆するものである。

**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を予め絞り込んだ後に個人認証を行う手法を提案した。データ取得条件を変更して実験を行った結果、提案手法により未登録者の誤受率を低減可能であることが明らかとなった。

今後は、認証精度の向上を目的として個人認証手法について検討を加えるとともに、しきい値の自動設定処理についても検討を加える予定である。

表2 シミュレーション結果の一例(1)

シミュレーション	手法	正答	誤識別	誤棄却	誤受理
1	<b>提案手法</b>	<b>294</b>	<b>1</b>	<b>26</b>	<b>3</b>
	口唇	241	3	59	21
	音声	247	5	49	23
2	<b>提案手法</b>	<b>261</b>	<b>1</b>	<b>57</b>	<b>5</b>
	口唇	249	3	51	21
	音声	215	2	85	22
3	<b>提案手法</b>	<b>231</b>	<b>1</b>	<b>91</b>	<b>1</b>
	口唇	232	3	64	25
	音声	171	2	148	3

表3 シミュレーション結果の一例(2)

シミュレーション	手法	正答率(%)	誤受率率(%)
1	<b>提案手法</b>	<b>90.7</b>	<b>0.9</b>
	口唇	74.4	6.5
	音声	76.2	7.1
2	<b>提案手法</b>	<b>80.6</b>	<b>1.5</b>
	口唇	76.9	6.5
	音声	66.4	6.8
3	<b>提案手法</b>	<b>71.3</b>	<b>0.3</b>
	口唇	71.6	7.7
	音声	52.8	0.9

#### 参考文献

- [1] 佐藤慶幸・西田眞：「音声と発話に伴う口唇の動き特徴を用いた個人識別に関する検討」, 電学論 C, Vol.125-C, No.8, pp.1282-1289 (2005-8)
- [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)