

# 発話に伴う口唇の動き特徴における区間分割 およびコマンド識別に関する検討

齋藤 歩<sup>†</sup> 高橋 毅<sup>†</sup> 景山 陽一<sup>†</sup> 百瀬 篤史<sup>†</sup> 石井 雅樹<sup>‡</sup> 西田 眞<sup>‡</sup>  
秋田大学<sup>†</sup> 秋田県立大学<sup>‡</sup>

## 1. はじめに

近年、情報通信システムの発展に伴い、セキュリティ確保を目的とした個人認証技術の重要性が増している。このため、生体情報を用いた個人認証に関する研究が盛んに行われ、既に実用化されつつある。しかしながら、生体情報を用いる各種システムは、個人認証までの利用に留まり、認証後はマウスやキーボードといった従来のコマンド入力インタフェースを採用しているのが現状である。

一方、これまでの研究において、発話に伴う口唇の動き特徴は、個人固有および発話内容固有の特徴を保持すること、並びに個人認証やコマンド識別に有用であることが明らかになっている<sup>[1][2]</sup>。また、発声の有無や心情変化に起因する特徴量の変化について明らかにすることは、実用システムを構築する上で有用である。そこで本研究では、発話に伴う口唇の縦幅、横幅、面積、アスペクト比、発話フレーム数の変動に着目し、有声発話と無声発話の相違点について検討を加えた。

## 2. 実験方法および使用データ

### 2.1. 実験環境と取得方法

日常一般的と考えられる蛍光灯下(照度約500~700lx)において、発話動画像(30fps)を取得した。被験者は無声と有声でそれぞれ特定のコマンドを発話した。データ取得の流れを以下に示す。

- (1)安静(5分)
- (2)各コマンドを6回ずつ発話(無声)
- (3)安静(2分)
- (4)各コマンドを6回ずつ発話(有声)

なお、①各コマンドを6回発話する度に30秒の休憩をとること、②発話の前後は口を閉じることを制約条件としている。

また、本研究は「秋田大学手形地区におけるヒトを対象とした研究に関する倫理規程第6条第2項」に基づいて倫理審査の申請を行い、承認を得た研究計画の下、被験者の同意を得てデータを取得している。

### 2.2. 使用コマンド

表1に検討に用いた6種類のコマンドを示す。コマンド1は被験者にとって最も馴染みのあると考えられる本人の氏名であり、コマンド2~4

表1 使用コマンド

コマンド1	「被験者自身の氏名」
コマンド2	あきたうめこ(a/ki/ta/u/me/ko)
コマンド3	おおだてけんしろう(o/o/da/te/ke/n/shi/ro/u)
コマンド4	かげやまよういち(ka/ge/ya/ma/yo/u/i/chi)
コマンド5	じょうほうこうがつか(zyo/u/ho/u/ko/u/ga/xtu/ka)
コマンド6	しゃしんしゅう(sya/shi/n/syu/u)

は日本語の母音「ア」、「イ」、「ウ」、「エ」、「オ」をすべて含むコマンドである。また、コマンド5およびコマンド6の選定に関して、破裂音、伸ばし音、撥音「ん」などを含むコマンドの発話は発声の有無に影響されやすいと仮定し、5つのコマンドを対象として予備検討を行った。具体的には、各コマンドを無声と有声でそれぞれ発話したデータを目視により評価した。その結果、「じょうほうこうがつか」および「しゃしんしゅう」において無声と有声の発話における変化が大きいことを認めため、これらの発話内容をそれぞれコマンド5およびコマンド6とした。

### 2.3. 使用データA

CCDビデオカメラ(SONY製DCR-VX2100)を用い、被験者a~gの7名を対象としてデータを取得した。なお、3日間取得したデータを使用データAとした。

### 2.4. 使用データB

CCDビデオカメラ(Point Grey Research製Grasshopper)を用い、被験者cを除く6名を対象として、新たに10日分のデータを取得した。本研究ではこれを使用データBとした。

## 3. 特徴量の算出

発声の有無が口唇の動きに与える影響を明らかにするため、①発話フレーム数Fの増加率 $R_F$ 、②横幅の累積差分値 $VaX$ 、③縦幅の累積差分値 $VaY$ 、④面積 $S_i$ 、並びに⑤縦幅と横幅のアスペクト比 $A_i$ の5つの特徴量をそれぞれ算出し、検討に用いた。さらに、使用データAの発話フレーム数を50フレームに統一し、面積とアスペクト比についてフレーム毎に平均と分散を算出した(以下、フレーム毎の平均、フレーム毎の分散とそれぞれ表記する)。各特徴量の算出式を(1)式~(7)式に示す。

$$R_F = \frac{100}{F_{vocalized}} \times (F_{non\_vocalized} - F_{vocalized}) \quad \dots(1)$$

$$VaX = \sum_{i=1}^{F-1} |raX_{i+1} - raX_i| \quad \dots(2)$$

$$VaY = \sum_{i=1}^{F-1} |raY_{i+1} - raY_i| \quad \dots(3)$$

$$S_i = diX_i \times diY_i \quad \dots(4)$$

$$A_i = diX_i / diY_i \quad \dots(5)$$

A study of lip motion feature segmentation for command recognition system.

<sup>†</sup>Ayumu Saito, Tsuyoshi Takahashi, Yoichi Kageyama, Atsushi Momose and Makoto Nishida(Akita Univ)

<sup>‡</sup>Masaki Ishii (Akita Prefectural Univ.)

$$VaS = \sum_{i=1}^{F-1} |raS_{i+1} - raS_i| \quad \dots(6)$$

$$VA = \sum_{i=1}^{F-1} |A_{i+1} - A_i| \quad \dots(7)$$

ここで、 $F_{non\_vocalized}$  および  $F_{vocalized}$  は、無声発話および有声発話における発話フレーム数、 $diX_i$  および  $diY_i$  は、 $i$  フレームにおける口唇の横幅および縦幅であり、 $raX_i$  および  $raY_i$  は、初期フレーム( $diX_1$ ,  $diY_1$ )を基準に正規化した横幅および縦幅である。また、面積も同様に正規化した値(以下、 $raS_i$  と表記する)を求めた。なお、 $VaS$  および  $VA$  は、 $raS_i$  および  $A_i$  の累積差分である。

#### 4. 実験結果および検討

##### 4.1. 口唇の動作量に関する検討

フレーム毎の平均、フレーム毎の分散を用いて、無声発話および有声発話における口唇の動作量について検討した。コマンド 2、3 における  $raS_i$  および  $A_i$  のフレーム毎の平均の最大値と最小値、およびその差を表 2(a)、(b)に、 $raS_i$  と  $A_i$  のフレーム毎の分散の差を表 3 にそれぞれ示す。なお、コマンド 1 は被験者毎に発話内容が異なるため、本稿では検討に用いてない。

フレーム毎の平均の差は、 $raS_i$ 、 $A_i$  ともに有声発話よりも無声発話の方が大きい。また、他のコマンドにおいても同様の結果を得た。

フレーム毎の分散の差は、フレーム毎の平均の差と同様に  $raS_i$ 、 $A_i$  ともに有声発話よりも無声発話の方が大きい。また、他のコマンドにおいても同様の結果を得た。このことから、無声発話は有声発話よりも口を大きく動かす傾向があると予想される。

##### 4.2. 被験者のグループ分け

使用データ A における  $R_F$  の平均値  $\overline{R_F}$  を表 4 に示す。 $\overline{R_F}$  の値に着目すると、7 名の被験者は正の値を有する被験者群(a、b、d、e)と負の値を有する被験者群(c、f、g)の二組に大別可能であることがわかる。そこで、 $\overline{R_F} \geq 15$  となる被験者 a、b、d、e をグループ A、 $\overline{R_F} < 0$  となる被験者 c、f、g をグループ B とし、以後の検討を行った。

##### 4.3. 特徴量の経時変化に関する検討

使用データ A における  $F$ 、 $VaX$ 、 $VaY$ 、 $VaS$ 、 $VA$  の 5 つの特徴量について、取得日別に各特徴量の平均値の平均をそれぞれ算出した結果を表 5 に示す。被験者別、コマンド別に比較を行った結果、グループ A において全体的に各特徴量の値が減少する傾向を認めた。一方、グループ B では変化が見られなかった。このことは、グループ A に属する被験者は時間経過に伴い、特徴量の変動する可能性があることを示唆している。

次に、使用データ B を対象として、10 日間にわたって特徴量の経時変化を調査した。具体的には、グループ A、B から被験者 e、f をサンプルとして抽出し、各特徴量の経時変化について検討を加えた。グループ B に属する被験者 f は、4.2 節の結果と同様に、4 日目以降も経時的な変化が認められなかった。一方、グループ A に属する被験者 e は、4.2 節で示した傾向とは異なり、特徴量の経時変化が認められなかった。また、

有声発話を基準とした特徴量の増加率に着目すると、10 日間すべてにおいて、被験者 e は正の値、被験者 f は負の値となり、グループ固有の特徴を保持し続けた。この結果から、グループ A においては、経時変化の有無によって被験者をさらに区別できると考える。

以上より、有声発話と無声発話では特徴量の大きさが異なり、この差異から被験者を二組に大別できる。また、この特徴は経時変化を起こさないと考えられる。今後は他の被験者についても検討を加え、各グループにおける特徴量の経時変化の傾向を明らかにする予定である。

表 2 フレーム毎の平均値  
(a)正規化した面積  $raS_i$

	コマンド 2		コマンド 3	
	無声	有声	無声	有声
最大値	154	152	156	151
最小値	89	93	96	95
差	65	60	60	56

(b)アスペクト比  $A_i$

	コマンド 2		コマンド 3	
	無声	有声	無声	有声
最大値	1.883	1.855	1.814	1.851
最小値	1.172	1.213	1.148	1.195
差	0.710	0.642	0.666	0.656

表 3 フレーム毎の分散の差

コマンド	$raS_i$		$A_i$	
	2	3	2	3
無声	319	226	0.020	0.011
有声	280	192	0.017	0.007

表 4 被験者毎の  $\overline{R_F}$

被験者	a	b	c	d	e	f	g
$\overline{R_F}$	21.2	37.6	-5.7	17.4	21.4	-4.9	-1.8

表 5 各グループにおける VA の取得日別平均

	グループ A		グループ B	
	無声	有声	無声	有声
1 日目	3.770	3.153	2.741	2.455
2 日目	3.667	2.920	2.546	2.284
3 日目	3.324	2.733	2.626	2.292

#### 参考文献

- [1] 白澤,三浦,西田,景山,栗栖:「口唇の動き特徴を用いた個人識別に関する検討」,映像情報メディア学会誌, Vol.60, No.12, pp.1964-1970(2006)
- [2] 佐藤,景山,西田:「口唇の動き特徴を用いた非接触コマンド入力インタフェースの提案」,電気学会論文誌 C, Vol.129, No.10, pp.1865-1873(2009)