赤外線距離センサ付き眼鏡型デバイスを用いた発話内容認識手法

五十嵐 雄也[1,a)] 双見 京介[1,2,b)] 村尾 和哉[1,c)]

概要：スマートグラスなどのアイウェアデバイスが今後一般普及するにあたって，多様な状況や人が利用できる入力手法の提供は重要である．無声発話を用いたサイレントスピーチインタラクション（SII）は，様々な状況や有声発話困難者へのハンズフリー入力手法として有用な可能性を持つが，アイウェア機器にSII 機能を簡便に適用できる手法は少ない．本研究では，赤外線距離センサを搭載したアイウェアデバイスを用いて，簡便なサイレントスピーチインタラクションを行う手法を提案する．提案手法はアイウェアデバイスに搭載された赤外線距離センサから発話に伴う顔の皮膚の動きを計測し，時系列のセンサデータに機械学習を適用することで，入力操作用の発話コマンド（例: 再生，停止，次，戻る）を認識する．提案手法をメガネフレームと耳掛けマイクパーツに適用して，頬と顎関節と顎の 3 カ所の動きを計測するプロトタイプデバイスを実装した．評価結果から，21 種類の発話コマンドを F 値 0.60，5 種の発話コマンドで F 値 0.86 の精度で認識できた．この結果は，提案手法がアイウェア機器にサイレントスピーチインタラクション機能を付与するために役立つ可能性を示した．

1. はじめに

情報機器のハンズフリー入力手法のひとつにサイレントスピーチ入力手法がある．音声認識はハンズフリー入力手法に用いられることが多いが，音声認識では公共の場での発声が周囲の人の迷惑となり，プライベートな情報が含まれることは発声することができない．また，バックグラウンドノイズなどが多い環境によっては音声認識の精度が低下する．

これらの課題に対処するため無声で発話するサイレントスピーチインタラクションに関する研究が活発に行われている．サイレントスピーチインタラクションは，有声ではない発話による音声インタラクション手段であり，ウェアラブルコンピューターなど，さまざまな状況での対話手段としてや発声困難者への支援技術としての可能性を持つ．また，音声認識と比べて発話内容が周囲の人に知られることがなく声を発さないため社会的受容性が高いことが示されている．これまでに，口唇の画像を取得し，画像認識によって発話内容を推定する手法 [10]，筋電図を使用して口腔付近の筋肉の状況から音声を推定する手法 [15]，加速度/角速度センサを皮膚に取り付けて発話内容を推定する手法などがある [1]．

一方，AR 用スマートグラスや音楽鑑賞用アイウェア等，多くのアイウェア機器が普及している．多くのアイウェア機器が今後普及すると想定されるが，アイウェア機器のみを用いてサイレントスピーチを行う手法はほとんど提案されていない．もし，アイウェア機器に簡易なセンサをつけるだけでサイレントスピーチが行える手法があれば，アイウェア機器の利用時に役立つと期待できる．

そこで本研究では，赤外線距離センサ付きのアイウェア機器を用いて，サイレントスピーチ内容の認識手法を提案する．提案手法では，発話の際に口に連動して動く顔（頬と顎）の皮膚の動きを眼鏡型デバイスに装着した赤外線距離センサにより測定し，発話内容の認識をする．提案手法を実装したプロトタイプデバイスを作成し，3 種類の発話方法で 21 種類の発話コマンドの認識精度を評価した．

以降では，2 節で関連研究を紹介し，3 節で提案手法を説明する．4 節で評価実験について述べ，5 節で有効なセンサについて検討し，6 節で発話内容を限定した際の精度を評価し，最後に 7 節で本稿をまとめる．

2. 関連研究

2.1 赤外線距離センサを利用したセンシング手法

赤外線距離センサを用いた手法が皮膚の動きの測定に利

[1] 立命館大学大学院情報理工学研究科
[2] Digital Spirits Teck
[a)] yuya.igarashi@iis.ise.ritsumei.ac.jp
[b)] futami@fc.ritsumei.ac.jp
[c)] murao@cs.ritsumei.ac.jp

用できることは，先行研究で示されてきた．EarTouch は，イヤホンに赤外線距離センサを取り付けることで，耳を引っ張る方向が認識できることを示しており，これによるジェスチャ入力手法が提案されている [2]．また，赤外線距離センサの眼鏡への適用は多くある．例えば，笑顔の認識手法 [3] や，日常生活における 8 つの表情の認識手法 [4]，顔を擦るジェスチャ入力手法 [5] がある．ヘッドマウントディスプレイ（HMD）に赤外線距離センサを適用した手法もある．例えば，皮膚の動きをもとにしてタッチジェスチャを認識する手法 [6]，自分の表情をヴァーチャル空間におけるアバターの表情にマッピングする手法 [7]，自分の顔の動きをコンピューターグラフィックスモデルにマッピングしてアニメーション制作に利用する手法 [8] が提案されている．他にも，TongueInput では，マウスピースにセンサを仕込むことで，舌のジェスチャを認識する手法を提案している [9]．

2.2 サイレントスピーチインタラクション（SSI）

SSI は，ユーザの持つ発話能力を利用できるので，ジェスチャ入力と比較したとき新たなジェスチャコマンドを習得する必要がないなど，音声認識の利点を多く持つ．また，なんらかの理由で有声発話が困難な方でも利用できるので，さまざまな手法が提案・研究されている．画像方式としては，携帯電話のカメラにより話者の口唇の画像を取得し，発話内容を推定する手法 [10] があるが，カメラを顔の前に持つ必要があり，ユーザーの手が束縛されてしまう．音声方式としては，利用者の皮膚にマイクを取り付けつぶやき，発話を認識する手法 [11]. また，鼻に取り付けたピエゾ素子によって周りの環境音によって阻害される音声認識を補正する手法 [14] がある．ただし，正確に発話内容を認識するには実際に発声をするので周囲の人に会話を聞き取られてしまう可能性がある．ユーザに装着する方式としては，喉元に加速度/角速度センサをとりつけて，発話時の口の動きに伴って動く皮膚の動きから発話内容を推定する手法 [1]．筋電図（EMG）を使用して取得される口腔付近の筋肉の状況から音声を推定する手法 [15] があるが，筋電図を使用するには顔表面に電極を取り付ける必要があるため，装着が目立つ．SSI の社会的受容性について調査した研究 [16] では，SSI は無声で発話をするので公的な場でもプライベートな情報を守ることができ，周囲の人に目立ちにくいためユーザに受け入れられやすく有声でのインタラクションより精度が低くても許容できることが示されている．

2.3 口の状態認識手法

顔や頭部をセンサで測定することで口の状態を認識することはさまざまな手法で行われており，キャップに圧力センサを設置し，側頭筋の動きを測定することで日常生活の中での間食をモニタリングを行う手法 [13] や，首に赤外線センサを組み込んだデバイスを装着し，顎下部分に向けて照射することで顎の動きの変化をとり食事活動のモニタリングを行う手法 [12] が提案されているが，装着するものなので日常生活での違和感をなくすデザインが課題となる．

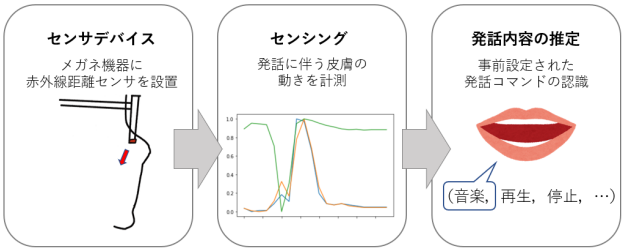

図 1 提案手法

3. 提案手法

本研究では，装着が容易であり装着時に周囲の人に目立たないデバイスがサイレントスピーチインタラクションでは有効であると考え，赤外線距離センサを取り付けた眼鏡型デバイスを作成する．

提案手法の流れは**図 1** のようになる．発話をする際には，口の動きに合わせて顔の皮膚（頬，顎関節）が連動して動く．提案手法は，この皮膚の動きをもとに発話内容を認識する．まず，眼鏡に設置された複数の赤外線距離センサから赤外線を照射し，センサと皮膚の間の距離の変化を得る．次に，得たセンサ値の時系列データに DTW（Dynamic Time Warping）と kNN（k-Nearest Neighbor）を適用し，発話内容を推測する．

3.1 デバイスの実装

提案手法のプロトタイプを実装した．**図 2** にセンサデバイスを示す．赤外線距離センサは大きく分けて図 2 の A の 3 カ所に設置される．これは，提案手法を眼鏡機器のどの部分に適用することが効果的かを検証するためである．1 つ目は，頬の動きを取るために，眼鏡の下リム部分に図 2 の D のようにセンサを 6 個設置した．次に，図 2 の E に示すように，2 つ目を顎関節の動きを取るために，ヘッドセット用のマイクパーツの顎関節側にセンサを 3 個設置し，3 つ目を顎の動きを取るために，ヘッドセット用のマイクパーツの顎側にセンサを 3 個設置した．センサには，図 2 の B のように赤外線距離センサ（TPR-105F）を 3 個組み合わせた幅 5mm，長さ 30mm のもの利用した．提案手法は直接口の動きをセンシングすることで認識精度が高くなる可能性があるため，眼鏡機器を利用して口の動きをセンシングするためのアタッチメントとして，ヘッドセット用のマイクパーツを利用した．こういったマイクパーツは，眼鏡機器にアタッチメントとして付ける設計が可能で

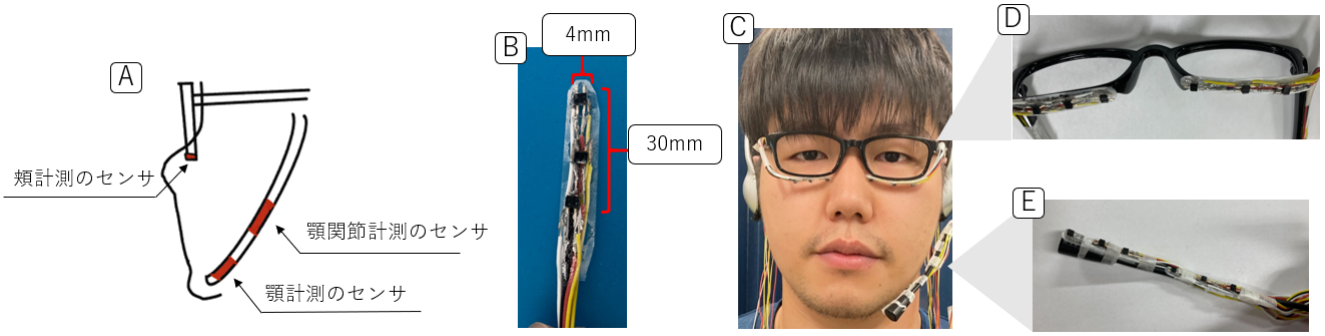

図 2　プロトタイプデバイス

ある．プロトタイプシステム全体は，センサデバイス，マイクロプロセッサ（Arduino），PC から構成される．

3.2　認識アルゴリズム

発話コマンドを識別するために，DTW と kNN を用いた．センサ値は，12 個のセンサ毎に正規化された．

詳細は以下の通りである．(1) 取得データと学習データの類似度を DTW で算出する．学習データには，あらかじめ用意されたすべてのジェスチャデータが含まれる．類似度はセンサごとに算出される．(2) 学習データから，kNN により取得データとの類似度が高いデータを選択する．選択されたデータのジェスチャラベルの割合から，取得したデータがどのジェスチャであるかの所属確立を算出する．例えば，kNN (k = 3) がジェスチャーラベル 1 の学習データを 3 つ選択した場合，ジェスチャラベル 1 の所属確率は 100 %である．この所属確率は，センサ毎に算出される．(3) 全センサの所属確率の総和が最も高いジェスチャラベルを，取得データの認識結果と判定する．例えば，センサの総数が 2 つで，センサ 1 とセンサ 2 のジェスチャラベル 1 の所属確率が 0.3，0.4 の場合，取得データのジェスチャラベル 1 の所属確率の和は 0.7（つまり全センサー合計値）である．

4.　評価 1．発話方法ごとの認識精度の評価

本実験では，21 種類の発話コマンドに対する提案手法の認識精度を評価した．加えて，複数の発話条件（無声発話，有声発話，誇張無声発話）における提案手法の認識精度を評価した．被験者は 13 名で，アジア人（男性 13 名，20 代前半，母国語は日本語）であった．

4.1　発話コマンド

発話コマンドは**表 1** に示す 21 種類を用意した．これらは先行研究 [1] を参考に，さまざまなデジタルデバイスを操作するために使用されるものとして選定した．

表 1　21 種類の発話コマンド．

音楽	はい	いいえ
スタート	停止	再生
ok	取り消す	メニュー
開く	閉じる	ホーム
次へ	戻る	回答
アレクサ	ミュート	左
右	音楽を再生する	音楽を停止する

表 2　発話者ごとの識別結果

発話者	有声発話			無声発話			誇張無声発話		
	P	R	F	P	R	F	P	R	F
1	0.93	0.93	0.93	0.96	0.95	0.95	0.99	0.99	0.99
2	0.75	0.70	0.70	0.55	0.53	0.51	0.78	0.76	0.75
3	0.66	0.63	0.62	0.59	0.52	0.49	0.84	0.83	0.82
4	0.61	0.57	0.56	0.71	0.68	0.67	0.75	0.70	0.68
5	0.87	0.84	0.84	0.91	0.90	0.89	0.92	0.91	0.90
6	0.64	0.59	0.59	0.66	0.64	0.64	0.07	0.07	0.07
7	0.49	0.44	0.43	0.35	0.34	0.31	0.74	0.68	0.67
8	0.48	0.46	0.43	0.63	0.58	0.56	0.77	0.73	0.72
9	0.57	0.51	0.51	0.63	0.56	0.55	0.84	0.80	0.79
10	0.38	0.39	0.36	0.35	0.32	0.31	0.36	0.36	0.33
11	0.75	0.70	0.69	0.50	0.45	0.43	0.69	0.66	0.64
12	0.66	0.65	0.62	0.85	0.81	0.81	0.86	0.83	0.82
13	0.85	0.81	0.81	0.75	0.75	0.73	0.81	0.79	0.79
平均	0.67	0.63	0.62	0.65	0.62	0.60	0.72	0.70	0.69

4.2　実験手順と評価方法

発話の条件は 3 種で，(1) 声を出す有声発話，(2) 声を出さない無声発話，(3) 口を大きく動かして声を出さない誇張無声発話であった．これらは，誇張無声発話において口を動かす程度は実験前に説明動画を用いて説明された．

実験タスクは，センサデバイスを装着し椅子に座った状態で，pc の画面に表示されるコマンドを意図的に発話するものであった．1 つの発話条件につき，21 種類の発話コマンドを 10 回ずつ行った．この際，同一の発話コマンドを連続して行わずに，21 種類の発話コマンドを 1 試行ずつ行った後で，次の試行を行った．発話条件の順番は，(1) 有声発話，(2) 無声発話，(3) 誇張無声発話で行った．取得データは，210 個の（発話コマンド 21 種× 10 試行×発話条件 3 種）であった．精度の評価には，個人内の測定デー

タをもとに 10 分割交差検証を行った．また，kNN の k の数値は 5 に設定した．これは，事前に 1 名のデータから k が 5 の際に認識精度が最も高くなったためである．

4.3 結果

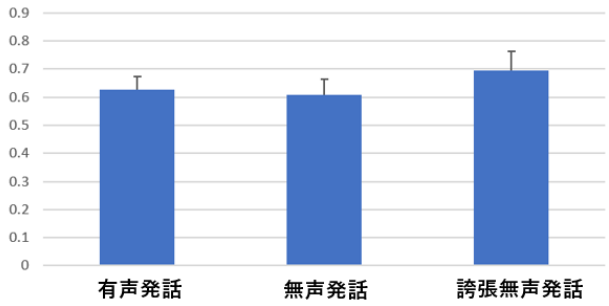

図 3 評価 1 の結果．発話条件ごとの認識結果．

発話条件ごとの識別結果を**図 3** に示す．個人ごとの結果を**表 2** に示す．認識精度は，誇張無声発話，有声発話，無声発話の順で低下した．誇張無声発話の適合率，再現率，F 値の平均値が最も高い値となり，F 値は 0.69 となった．次に，有声発話の平均値が高く，F 値は 0.62 となった．無声発話は 3 種類の発話方法の中で平均値が最も低く，F 値は 0.60 となった．

4.4 考察

発話条件ごとの精度の傾向について述べる．無声発話と有声発話の分類精度には大きさな差はなかったが，無声発話の方が精度が少し低くなった．この理由としては，無声での発話に慣れていなかったためだと考える．この点については，データを計測する際に被験者に十分に無声発話の練習を行わせることで認識精度が高くなると想定できる．誇張無声発話の方が無声発話よりも分類精度が高くなった．この理由としては，口を大きく動かすことで，センサ値の変動が大きくなり得られたデータの波形が発話コマンドごとに特徴的に現れたからだと考えられる．したがって，提案手法の認識精度を高めるために，ユーザ側が認識されやすい発話を行うことは有効な手段と考えられる．

結果は，提案手法の認識精度に個人差があることを示した．例えば，無声発話の平均 F 値は 0.60 となったが，F 値 0.8 以上の被験者は 3 名（発話者 1 は F 値 0.95，発話者 5 は F 値 0.89，発話者 12 は F 値 0.81），F 値 0.5 以下の被験者は 4 名であった（発話者 3 は F 値 0.49，発話者 7 は F 値 0.31，発話者 10 は F 値 0.31，発話者 11 は F 値 0.43）．有声発話や誇張無声発話に関しても，精度が高い被験者と低い被験者がいた．この個人差の原因には次のことが考えられる．(1) まず，作成したプロトタイプデバイスでは被験者の顔の特徴によっては顔の皮膚の動きをセンサで測定することが難しかったためだと考える．これについては，センサデバイスのセンサ位置が顔の適切な位置にあたるように個人ごとに調整することで対策できると考えられる．(2) 次に，発話時に顔の皮膚の動きの程度が少ない被験者は認識精度が低下したと考えられる．これについては，顔の皮膚に動きが出るように発話を被験者が行えれば，認識精度が向上すると考えられる．

また，次の点も見て取れた．一部の被験者は有声発話よりも無声発話の方が認識精度が高くなった．例えば，発話者 8 や発話者 12 の無声発話の F 値は有声発話と比べて約 0.13〜0.18 高くなった．この理由としては，慣れていないため無声発話においては有声発話よりも意識して口を動かしたためと考えられる．

5. 評価 2．発話コマンドを減らした場合の認識精度

評価 1 では 21 種類の発話コマンドを利用したが，本評価では発話コマンドを減らすことで推定精度が上がるかを評価した．本実験では評価 1 の無声発話のデータを利用した．

発話コマンドには次の 2 パターンを利用した．

- (1)5 種の発話コマンドを利用した．内容は「音楽」，「再生」，「停止」，「次へ」，「戻る」とする．ハンズフリー入力のための先行研究 [17] では，単純なハンズフリー入力には 5 種類程度のコマンド認識が必要かつ十分であることが示されています．例えば，メディアプレーヤ（音楽，動画，静止画など）の操作においては，再生，停止，進む，戻るなど 5 種類程度のコマンドで十分である．
- (2)18 種の発話コマンドを利用した．これらは，発話者ごとに評価実験 1 の 21 種の発話コマンドから F 値が低い下位 3 個の発話コマンドを除いたものである．

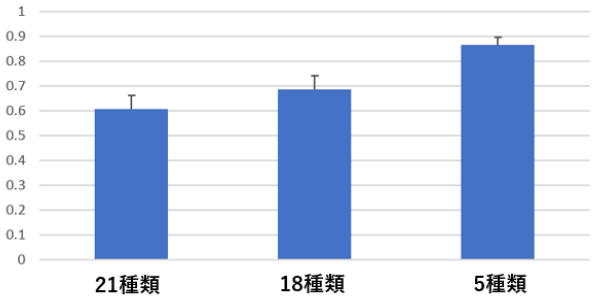

図 4 評価 2 の結果．発話コマンドを減らした場合の認識結果．

5.1 結果と考察

結果を**図 3** に示す．個人ごとの結果を表 3 に示す．5 種の発話コマンドの F 値は 0.86 であり，評価 1 の 21 種の発

表 3 発話コマンドを減らした場合の識別結果 (P: Precision、R: Recall、F: F-value)

発話者	21 種類			18 種類			5 種類		
	P	R	F	P	R	F	P	R	F
1	0.96	0.95	0.95	0.98	0.98	0.98	0.98	0.98	0.98
2	0.55	0.53	0.51	0.70	0.65	0.63	0.84	0.80	0.78
3	0.59	0.52	0.49	0.61	0.57	0.55	0.89	0.86	0.86
4	0.71	0.68	0.67	0.80	0.75	0.75	0.94	0.94	0.93
5	0.91	0.90	0.89	0.96	0.95	0.95	1.00	1.00	1.00
6	0.66	0.64	0.64	0.78	0.76	0.77	0.94	0.94	0.94
7	0.35	0.34	0.31	0.45	0.40	0.38	0.86	0.80	0.79
8	0.63	0.58	0.56	0.73	0.68	0.67	0.88	0.86	0.85
9	0.63	0.56	0.55	0.71	0.65	0.64	0.89	0.84	0.84
10	0.35	0.32	0.31	0.42	0.37	0.35	0.68	0.66	0.63
11	0.50	0.45	0.43	0.55	0.50	0.49	0.82	0.80	0.79
12	0.85	0.81	0.81	0.88	0.87	0.87	0.94	0.94	0.93
13	0.75	0.75	0.73	0.86	0.84	0.84	0.93	0.92	0.91
平均	0.65	0.62	0.60	0.73	0.69	0.68	0.89	0.87	0.86

話コマンドの F 値は 0.60 と比べると，27 %精度が上がった．また，18 種の発話コマンドの F 値は 0.68 であり，評価 1 の 21 種の発話コマンドの F 値と比べ 8 %上がった．この結果は，発話コマンドを減らすことで提案手法の認識精度が上がることを示した．また，簡易ハンズフリー入力法を検討した先行研究では，顔や視線の入力ジェスチャーを 5〜7 種類認識し，F 値が 0.85〜0.9 [17] であることが示されている．これらを踏まえると，提案手法は先行研究と同程度の認識精度を有していると考えられ，提案手法は簡易なハンズフリー入力手法として活用できると考えられる．

6. 評価 3. 有効なセンサ位置の評価

プロトタイプデバイスでは，頬，顎関節，顎の 3 カ所を測定しているが，提案手法をより少ないセンサ箇所で利用できれば便利である．そこで，評価 3 では，センサデバイスの測定部位を減らした場合に提案手法の認識精度がどの程度変化するかを検討した．

測定箇所を次の 3 種に分ける．3 種は，(1) 眼鏡の下リムを利用した頬のセンサ位置，(2) マイクパーツを利用した顎関節のセンサ位置，(3) マイクパーツを利用した顎のセンサ位置である．そして，評価 1 と同じ 21 種類の発話コマンドの無声発話のデータに対して提案手法の評価を行った．

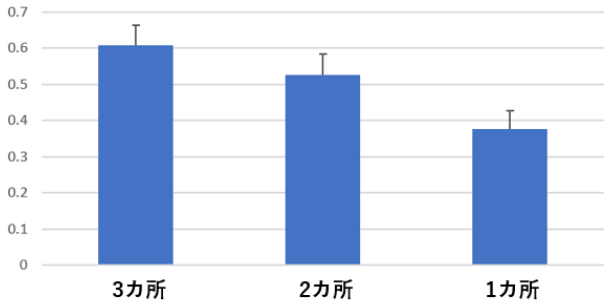

図 5 評価 3 の結果．測定部位を減らした場合の認識結果．

表 4 測定部位ごとの識別結果

発話者	3 カ所（頬，顎関節，顎）			2 カ所（頬，顎関節）			1 カ所（頬）		
	P	R	F	P	R	F	P	R	F
1	0.96	0.95	0.95	0.89	0.87	0.87	0.65	0.62	0.61
2	0.55	0.53	0.51	0.48	0.47	0.45	0.47	0.43	0.43
3	0.59	0.52	0.49	0.45	0.41	0.37	0.23	0.25	0.22
4	0.71	0.68	0.67	0.67	0.62	0.61	0.41	0.41	0.38
5	0.91	0.90	0.89	0.85	0.83	0.83	0.55	0.54	0.52
6	0.66	0.64	0.64	0.59	0.55	0.54	0.25	0.26	0.23
7	0.35	0.34	0.31	0.27	0.26	0.24	0.12	0.17	0.13
8	0.63	0.58	0.56	0.51	0.48	0.46	0.31	0.34	0.31
9	0.63	0.56	0.55	0.53	0.46	0.44	0.38	0.35	0.33
10	0.35	0.32	0.31	0.24	0.24	0.23	0.18	0.20	0.18
11	0.50	0.45	0.43	0.42	0.36	0.34	0.27	0.25	0.24
12	0.85	0.81	0.81	0.79	0.75	0.75	0.75	0.67	0.65
13	0.75	0.75	0.73	0.71	0.67	0.66	0.71	0.64	0.63
平均	0.65	0.62	0.60	0.57	0.54	0.52	0.41	0.39	0.37

6.1 結果と考察

図 5 に，1 か所パターン（頬），2 カ所パターン（頬＋顎関節），3 カ所パターン（頬＋顎関節＋顎），の識別結果を示す．**表 4** に個人ごとの結果を示す．

3 カ所パターン（頬＋顎関節＋顎）の f 値が 0.61，2 カ所パターン（頬＋顎関節＋顎）の f 値が 0.52，1 カ所パターン（頬＋顎関節＋顎）の f 値が 0.37 となった．

7. まとめ

本研究では，容易に装着が可能なウェアラブルデバイスによるサイレントスピーチインタラクションが可能かを調査するために，赤外線距離センサ付き眼鏡型デバイスを用いて発話内容を認識する手法を検討した．本手法は，眼鏡の下リム部分とヘッドセットの頬側面部分と顎側面部分に設置された赤外線距離センサから顔の皮膚までの距離の変化をもとに，発話に伴って起こる顔の皮膚の動きを測定することで，発話内容の推定を行った．提案手法を実現するためのプロトタイプデバイスを作成し，先行研究を参考に 21 種類の発話コマンドを選定し，有声発話，無声発話，誇張無声発話の 3 種類の発話方法で認識精度を評価した．有声発話の平均 F 値は 0.62，無声発話の平均 F 値は 0.60，誇張無声発話の平均 F 値は 0.69 となった．5 種類の発話コマンドでは無声発話で平均 F 値は 0.86 となった．今後はセンサの設置位置や角度の細かな調整が可能なデバイスを作成することで認識精度が向上するかの検証と分類手法を変えて大量のデータによる推定を行う予定である．

謝辞

本研究の一部は，JSPS 科研費 JP19K20330 の助成によるものである．

参考文献

[1] Rekimoto, J., Nishimura, Y. (2021, February). Derma: Silent Speech Interaction Using Transcutaneous Motion

Sensing. In Augmented Humans Conference 2021 (pp. 91-100).

[2] Kikuchi, T., Sugiura, Y., Masai, K., Sugimoto, M., Thomas, B. H. (2017, September). EarTouch: turning the ear into an input surface. In Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services (pp. 1-6).

[3] Fukumoto, K., Terada, T., Tsukamoto, M. (2013, March). A smile/laughter recognition mechanism for smile-based life logging. In Proceedings of the 4th Augmented Human International Conference (pp. 213-220).

[4] 正井克俊, 杉浦裕太, 尾形正泰, 稲見昌彦, 杉本麻樹. (2016). AffectiveWear: 装着者の日常的な表情を認識する眼鏡型装置. 日本バーチャルリアリティ学会論文誌, 21(2), 385-394.

[5] Masai, K., Sugiura, Y., Sugimoto, M. (2018, February). Facerubbing: Input technique by rubbing face using optical sensors on smart eyewear for facial expression recognition. In Proceedings of the 9th Augmented Human International Conference (pp. 1-5).

[6] 山下幸輝, 菊地高史, 正井克俊, 杉本麻樹, 杉浦裕太. (2018). CheekInput: 頬をタッチサーフェースとする頭部装着型ディスプレイへの入力手法. ヒューマンインタフェース学会論文誌, 20(3), 311-320.

[7] 鈴木克洋, 中村文彦, 大塚慈雨, 正井克俊, 伊藤勇太, 杉浦裕太, 杉本麻樹. (2017). AffectiveHMD: 組み込み型光センサを用いた表情認識とバーチャルアバターへの表情マッピング. 日本バーチャルリアリティ学会論文誌, 22(3), 379-389.

[8] 浅野直生, 正井克俊, 杉浦裕太, 杉本麻樹. (2018). 反射型光センサを用いた眼鏡型ウェアラブルデバイスによる顔表情パフォーマンスキャプチャ. 日本バーチャルリアリティ学会論文誌, 23(3), 197-206.

[9] Hashimoto, T., Low, S., Fujita, K., Usumi, R., Yanagihara, H., Takahashi, C., ... Sugiura, Y. (2018, September). TongueInput: Input Method by Tongue Gestures Using Optical Sensors Embedded in Mouthpiece. In 2018 57th Annual Conference of the Society of Instrument and Control Engineers of Japan (SICE) (pp. 1219-1224). IEEE.

[10] Sun, K., Yu, C., Shi, W., Liu, L., Shi, Y. (2018, October). Lip-interact: Improving mobile device interaction with silent speech commands. In Proceedings of the 31st Annual ACM Symposium on User Interface Software and Technology (pp. 581-593).

[11] Nakajima, Y., Kashioka, H., Shikano, K., Campbell, N. (2003, April). Non-audible murmur recognition input interface using stethoscopic microphone attached to the skin. In 2003 IEEE International Conference on Acoustics, Speech, and Signal Processing, 2003. Proceedings.(ICASSP'03). (Vol. 5, pp. V-708). IEEE.

[12] Zhang, S., Zhao, Y., Nguyen, D. T., Xu, R., Sen, S., Hester, J., Alshurafa, N. (2020). Necksense: A multi-sensor necklace for detecting eating activities in free-living conditions. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 4(2), 1-26.

[13] Zhou, B., Lukowicz, P. (2020, September). Snacap: snacking behavior monitoring with smart fabric mechanomyography on the temporalis. In Proceedings of the 2020 International Symposium on Wearable Computers (pp. 96-100).

[14] Maruri, H. A. C., Lopez-Meyer, P., Huang, J., Beltman, W. M., Nachman, L., Lu, H. (2018). V-speech: Noise-robust speech capturing glasses using vibration sensors. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 2(4), 1-23.

[15] Kapur, A., Kapur, S., Maes, P. (2018, March). Alterego: A personalized wearable silent speech interface. In 23rd International conference on intelligent user interfaces (pp. 43-53).

[16] Laxmi Pandey, Khalad Hasan, and Ahmed Sabbir Arif. 2021. Acceptability of Speech and Silent Speech Input Methods in Private and Public. In Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21). Association for Computing Machinery, New York, NY, USA, Article 251, 1–13.

[17] Amesaka, T.; Watanabe, H.; Sugimoto, M. Facial Expression Recognition Using Ear Canal Transfer Function. In Proceedings of the 23rd International Symposium on Wearable Computers, London, United Kingdom, 9-13, September, 2019; pp. 1–9.