

メガネ型デバイスを用いた 鼻呼吸ジェスチャによるハンズフリー入力手法の検討

小川 諒馬^{1,a)} 双見 京介^{1,b)} 村尾 和哉^{1,2,c)}

概要: 近年, ハンズフリー入力手法の研究が盛んに行われている。しかし, これまでの手法の多くは, 音声や体の動きを利用するため, 日常生活の中でいつでも利用することが困難であった。そこで, 本研究では, これらの問題点を解決するために, 鼻表面に設置された圧電素子を用いて呼吸を認識し, 取得した呼吸情報からデバイスの操作を想定したハンズフリー入力手法の提案を行う。提案手法では, メガネなどの鼻表面部分に触れる一般的な装着物に設置された圧電素子を用いて, 呼吸時に自然発生する呼吸音の振動をセンシングし, 取得した呼吸情報の音量から呼吸パターンを認識しデバイス进行操作することを目的として, プロトタイプデバイスを実装し, その有効性について検証を行った。呼吸パターンは, 呼吸回数, 時間間隔, 呼吸強度の要因に基づいて認識される。10人の参加者を対象とした評価実験において, 8種類の呼吸ジェスチャをメガネ型デバイスを用いて取得した。その結果, 82%のF値で8種類の鼻呼吸ジェスチャを認識できることを示した。この結果から, 本研究の手法は有効であると考えられる。私たちの研究は, ハンズフリー入力に鼻呼吸を使用する最初のウェアラブルセンシング技術を提供する。

1. はじめに

近年, 様々な用途, 目的のデバイスが普及している。日常生活の中での様々な行動, 動作はデバイスを介して行うものが多くなり, 人々は必ず何かのデバイスに触れている。それに伴って, デバイスの操作方法や入力手法に対する研究は活発に行われている。現在人々の身の回りにあるほとんどのデバイスの入力は指や手を用いたディスプレイタッチ入力やキーボード入力, ボタン操作などが用いられている。これらには手に荷物を持っている状態など手が埋まってしまっている状態では, 操作することができないという問題が存在する。そんな問題の解決策として, ハンズフリー入力手法が研究されている。

私たちの身の回りに一番普及しているハンズフリー入力手法としては音声入力 [21] があげられる。具体的には, AmazonのFire TV Stickリモコンに搭載されているAlexaやAppleのスマートフォンiPhoneに搭載されているSiriといったAIアシスタントと呼ばれる発話解析・認識インタフェースや, Google Homeなどのスマートホーム家電がある。これらは実際に家庭内で, 手が埋まっている場合や

手を離すことができない状態におけるハンズフリー操作を行うことができる。体のさまざまな部分を動かして行うハンズフリー入力も存在する。たとえば, 顔の表情 [3][5], 頭の動き [6][7], 口の動き [1][2], 目 [4][8] を使用するジェスチャ入力方法が存在する。

しかし, 入力手法に対する既存研究の中には日常生活において常時使用可能という点において課題を残しているものも多い。例えば, 音声を使用する入力方式には外部からのノイズや環境音, 第三者の音声などによって認識精度が低下するという問題点が存在する。また, 外で使用の際にも, 入力行為が不自然で目立ち, 入力行為とその内容が他人に知られてしまうリスクなど, 他者が存在する状況では入力ジェスチャを行うことが困難な場合がある。そのため, 発話音や体の動きのないウェアラブル機器を使ったハンズフリー入力方式が実現できれば, 日常生活での使用に支障はなく, そのような問題は少なくなる。

本研究では, ウェアラブル機器を用いた鼻呼吸によるハンズフリー入力方式を実現するために, メガネの鼻当て部分に設置した圧電素子を用いて, 鼻呼吸ジェスチャを認識する手法を提案する。本研究では, 鼻での呼吸情報に着目し, 鼻呼吸によって発生する振動から呼吸パターンを識別し, パターンに対応したデバイスの操作を行う手法の実現を目標とする。そのために, まず鼻呼吸パターンを認識する手法を提案する。提案手法では, 鼻呼吸を行なった際に

¹ 立命館大学大学院情報理工学研究科

² 科学技術振興機構さきがけ

a) ryoma.ogawa@iis.ise.ritsumei.ac.jp

b) futami@fc.ritsumei.ac.jp

c) murao@fc.ritsumei.ac.jp

発生する自然振動を音データとして、左右の鼻表面部分にマイクとして設置された小型の圧電素子から取得する。取得した音データを解析し、呼吸の音量（大きさ）や長さ、回数などから呼吸パターンを識別する。これにより、デバイスに対しての簡単な操作を行うことができる。プロトタイプシステムを実装した。具体的には、呼吸情報を取得するメガネ型デバイスを作成した。これはメガネの両鼻当て部分に圧電素子を設置したものである。評価実験では、提案手法による呼吸パターンの認識率をメガネ型デバイスを用いて評価する。提案手法が呼吸回数、時間間隔、呼吸強度の3つの要素を組み合わせた8種類の呼吸ジェスチャを認識できるかどうかを検証した。

本研究では、日常生活において常時行うことが可能なジェスチャとして、鼻での呼吸に着目した。呼吸を検出するためのデバイスの研究は数多く行われているものの、呼吸情報をデバイスへの入力として使用しているものは少ない。呼吸は人間が常時行っている生理現象の1つであり、日常のどんな場面でも人間全員が行なっているため、行動としての違和感も少なく、呼吸を行えないような状況にあることは極めて少ない。そのため、先ほど挙げた入力手法の問題点も解決していると考えられる。また、メガネ、スマートグラス、昨年から新型コロナウイルスの影響により生活必需品となっているマスク等、鼻表面部分へのセンサの設置は容易であるため、様々なウェアラブルデバイスに適用可能である。

2. 関連研究

2.1 ハンズフリー入力手法

Lyons ら [1] は、小型カメラで撮影した口もとの画像から口の形状を認識することで、口の形状と手元のテンキーを使用した入力手法 MouthType を提案し、日本語入力と英語入力、2種類言語入力に対して実験を行なった。英語入力のエラー率は3.1%であり、日本語入力のエラー率は8.7%であった。どちらの言語の場合も、携帯電話におけるキーボード入力よりもエラー率は高くなったが、MouthTypeの方がかなり早く入力が可能であった。また、小口ら [2] は、同様に日本語の母音口唇形状認識を用いて、母音の入力列から推測したコマンドによるデバイス操作手法を提案している。また、これらの口唇形状認識技術には口がカメラに映る状態になくはならないという難点があるため、マスクのように、口元を隠すものの着用時に使用することができない。そのため、中尾ら [3] は口周りにつけた電極から、顔の下半分の筋肉の動きを認識し、舌、口、頬のジェスチャを用いたハンズフリー操作が可能で、マスク型ウェアラブルデバイス Make-a-Face を提案している。実際に、電車の中や立ち仕事、ジョギング中などでスマートフォンの操作を可能にしており、マスク型であるためヘッドマウントディスプレイを装着しながらの使用も可能である。デ

バイスの煩わしさなどを排除し、日常生活におけるデバイスの快適性を向上させている。真鍋ら [9] は、EOG センサーを使用して視線入力ジェスチャを認識するイヤホン提案した。雨坂ら [22] は、アクティブ音響センシングと外耳道の形状の変化を使用して、顔の表情のジェスチャを認識する方法を提案した。EarFieldSensing[10] は、イヤホンに取り付けられた電極（EMG）と外耳道の電界の変化を使用して、顔の動きのジェスチャを認識する手法である。谷口ら [11] は、外耳道型イヤホンの先端に取り付けられた赤外線距離センサーと外耳道底の動きを利用して、舌の動きを認識する方法を提案した。CanalSense[7] は、イヤホンに取り付けられた気圧計を使用して顔の動きを認識する手法である。TongueInput[12] は、赤外線距離センサー付きのマウスピースを使用して舌のジェスチャを認識する方法である。また、視線を用いてハンズフリー入力を行う手法 [4][8] や、頭部、表情を用いてハンズフリー入力を行う手法 [6][5] もある。

これらと比較して、本研究ではメガネのみで行うため、追加のデバイスを必要とせず日常生活に馴染みやすい。また、鼻呼吸を用いるため、外部に見た目ではバレにくく、行動の不自然さなども存在しない。

2.2 音検出および音認識を利用したインタラクション

呼吸情報は人間の心身を表す重要な生体情報として注目されている。その理由は、呼気圧、回数、音量など多くの情報を含んでいるためである。そのため、呼吸を常時検出・認識することができれば有益であると考えられ、呼吸を検出・認識するための研究は数多く行われている。また、音声や呼吸を含む動作などによって発生する音を利用したインタラクションの研究も行われている。

Chauhan ら [13] は、呼吸ジェスチャから得られた音声情報に基づいて、個人識別を行う行動生態認証システム BreathPrint を提案している。鼻の下 1-2cm の場所にイヤホンマイクを設置し、2回連続の素早い吸引、正常呼吸、深呼吸の3種類の呼吸ジェスチャを行い、得られた呼吸情報から個人識別を行う。3つのジェスチャ全てに対して94%以上の精度であり、ユーザー識別の手段として有望であるとされている。また、呼吸認識を用いた呼吸訓練を目的とした研究も存在する。Shih ら [14] は、ゲームを通じたバイオフィードバック型誘導呼吸訓練を行うアプリケーション Breeze を提案している。ユーザーからも高く評価され、臨床現場やセルフケア活動に利用できる可能性が示されている。呼吸が与える心理的影響に着目し、操作手法として提案されているものも存在する。また、呼吸を用いてゲームの操作を行う手法 [15][16] や、呼吸を様々な機器の入力インタフェースとする手法 [17][18][19] も存在する。また、鼻呼吸の活用先として本研究は、眼鏡型デバイスを用いて日常生活で使用可能なハンズフリー入力手法を提案している。

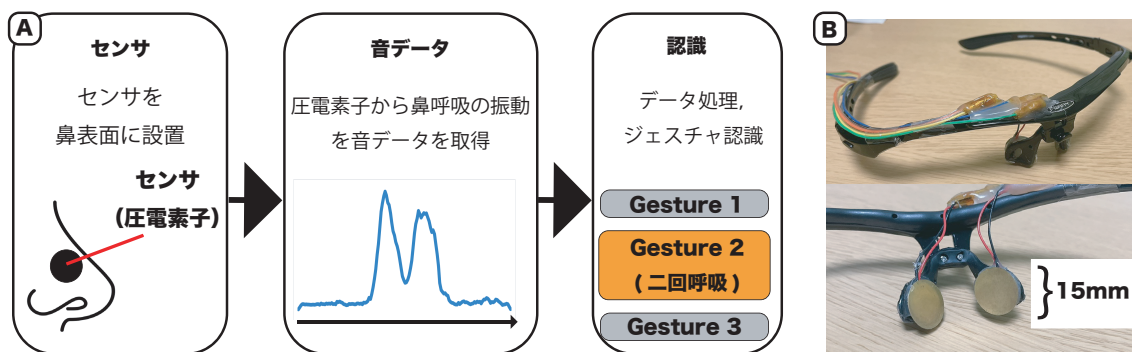


図 1 提案手法の概要とプロトタイプデバイス

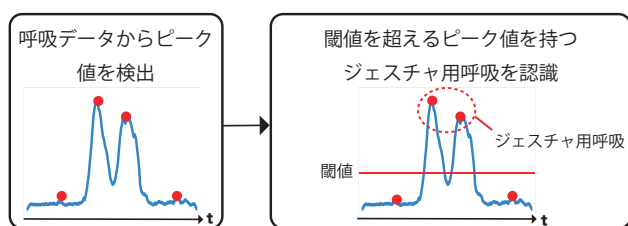


図 2 ジェスチャ用呼吸の認識

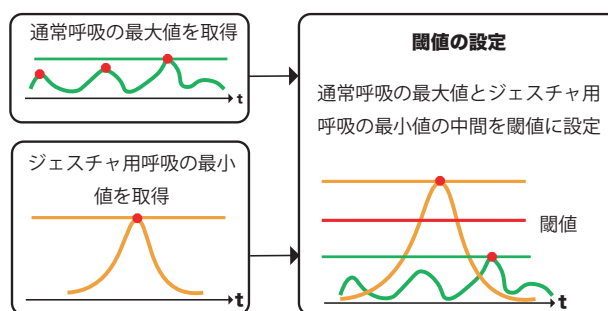


図 3 閾値の設定

3. 提案手法

提案手法の概要を図 1 (A) に示す。ユーザーが装着するメガネ型デバイスの両鼻当て部分に圧電素子を設置する。左右の鼻表面に接している 2 つの圧電素子をマイクとし、それぞれの圧電素子から鼻呼吸によって発生する呼吸音の自然振動を音声信号として取得する。音声処理プロセスでは、メガネ型デバイスに設置した 2 つの圧電素子から取得した音声信号から特徴量として音量を抽出する。認識プロセスでは、呼吸回数、時間間隔、呼吸強度から呼吸ジェスチャの分類を行う。メガネの鼻当て部分に圧電素子を設置し、鼻表面に触れさせるようにした理由を以下に示す。先行研究に、Hector[20] らが提案している V-Speech というものが存在する。スマートグラスの鼻パッド部分に設置された振動センサを用いて発話時の鼻骨の振動を取得することで、騒音下でも発話音声を認識しやすくするというものである。この先行研究の結果を踏まえると、同様のセンサ位置に圧電素子を設置することで、音声ではなく鼻呼吸の振動を取得することは可能であろうと考えた。

3.1 音データの処理

音データの処理では、取得した音データから特徴量として音量を抽出し、音量の時系列データを得る。次に RMS (Root Mean Square) を行う。その次に、音量の時系列データを滑らかなにするための移動平均と、サンプル数を減らすためのリサンプリングを行う。サンプリング周波数が 48000 (Hz) で 1 秒間に 48000 サンプルあるため、4800 サンプルごとに移動平均をし、1/1000 倍のリサンプリングを行う。

3.2 ジェスチャの認識

このシステムでは、振動から得る音量の変化を基に、ユーザーの鼻呼吸ジェスチャを認識する。鼻呼吸の吸い始め (吐き始め) に音量が上がるため、波形の音量も上がる。そして、吸い終わり (吐き終わり) に音量が下がるため、波形の音量も下がる。この流れを一呼吸とする。その結果、鼻呼吸を行うと、図 1 の音データ部分の波形に示すように、音量のピークが生成される。

3.2.1 ジェスチャ用呼吸の認識

ジェスチャ用呼吸を認識する方法を図 2 に示す。(1) まず、1 つの呼吸ジェスチャの音量の波形の最大値 (ピーク値) を求める。これは、python のライブラリを用いて求めた。波形の勾配値を微分値の増減から取得し、この勾配値に基づいて、呼吸波形のピーク値を検出する。図 2 では、ピーク値を赤い点で示している。(2) 次に、最大値として検出された点のうち、閾値を超えている点を、意図的な呼吸として認識する。図 2 では、閾値を赤い線で示している。(3) この閾値を超えるピーク値がない場合は、通常呼吸であると判断する。

3.2.2 ジェスチャ用呼吸認識のための閾値設定

閾値設定の説明を図 3 に示す。この閾値は、ジェスチャ用呼吸が通常呼吸と誤認識されないように設定している。閾値の設定には、ユーザーごとの通常呼吸データとジェスチャ用呼吸データの両方を使用する。閾値は、通常呼吸の最大値 (図 3 の緑線) とジェスチャ用呼吸のピーク値 (図 3 のオレンジ線) の中央値 (図 3 の赤線) に設定する。ジェ

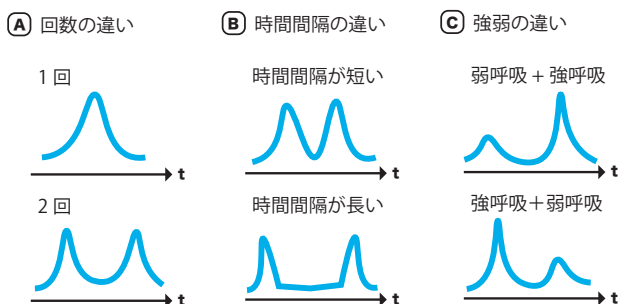


図 4 呼吸の波形の例

スチャ用呼吸データに複数のジェスチャ呼吸が含まれている場合、複数のジェスチャ呼吸のピーク値の最小値が使用される。また、閾値はユーザーごとに設定した。この閾値設定は、認識システムを使用する前に行う。

3.2.3 3つの要素に基づく呼吸ジェスチャの認識

このジェスチャ用呼吸検出アルゴリズムを利用して、次の3つの要素の異なる呼吸ジェスチャを認識する。各認識の流れを以下に示す。以下のフローでは、呼吸ジェスチャは最大2回の呼吸で構成される。

1. 呼吸回数の異なるジェスチャの認識: 本アルゴリズムでは、回数の違いから呼吸ジェスチャを認識する。これは例えば、同一時間内の呼吸回数(例: 1回か2回)でジェスチャを区別する際に使用する。図4(A)に例を示す。これは、次のアルゴリズムから認識する。(1)一定の時間幅におけるジェスチャ用呼吸の数をカウントする。ジェスチャ用呼吸は、「ジェスチャ用呼吸の認識」で前述した方法を用いて認識する。一定の時間幅は4秒とする。

2. 呼吸の時間間隔の異なるジェスチャの認識: 本アルゴリズムでは、呼吸と呼吸の時間間隔の違いから呼吸ジェスチャを分類する。これは例えば、時間間隔が短いジェスチャと、時間間隔が長いジェスチャを区別する際に使用する。図4(B)に例を示す。これは、次のアルゴリズムから認識する。(1)一定の時間幅においてジェスチャ用呼吸が2個以上検出された場合、前の呼吸の最大値の時間から次の呼吸の最大値の時間までの時間間隔を求める。(2)この時間間隔が一定の閾値以上の場合、時間間隔が長いジェスチャと判定し、時間間隔が一定の閾値未満の場合、時間間隔が短いジェスチャと判定する。一定の閾値は1秒とする。

3. 呼吸の音量の強弱の異なるジェスチャの認識: 本アルゴリズムでは、連続する呼吸の強弱の違いから呼吸ジェスチャを分類する。これは例えば、強い呼吸の次に弱い呼吸を行う呼吸ジェスチャと、弱い呼吸の次に強い呼吸を行う呼吸ジェスチャ、強弱差がない呼吸ジェスチャの3つのパターンを区別する際に使用する。図4(C)に例を示す。これは、次のアルゴリズムから認識する。(1)一定の時間幅においてジェスチャ用呼吸が2個以上検出された場合、次の呼吸の最大値から前の呼吸の最大値を割った商を求め



図 5 プロトタイプシステム

る。(2)この商が閾値以上もしくは以下の場合、異なる音量の呼吸が連続して行われたジェスチャと判定する。後述の評価では、以下のように閾値を設定した。今回は、商が1.8以上であれば強呼吸→弱呼吸の順と判定し、商が5/9以下であれば弱呼吸→強呼吸の順と判定する。それ以外の場合、強度に差のない連続呼吸からなるものと判断する。

4. 実装

4.1 ハードウェア

実装したプロトタイプシステムを図5に示す。プロトタイプシステムは2つの圧電素子を左右の鼻当て部分に設置したメガネ型デバイス、前処理やジェスチャ認識などを行うNote PC (Lenovo ThinkPad X1 Carbon)で構成される。またメガネ型デバイスとNote PCはオーディオインターフェース (Steinberg UR22mk)を介して接続されている。今回は鼻部分にセンサを当てるデバイスとしてメガネを用いている。オーディオインターフェースは、圧電素子から得られた振動データをデジタルデータに変換するためのADC (analog to digital converter)として使用する。圧電素子 (FGT-15T-6.0A1W40)は、直径15mm、厚さ0.3mm以下である。オーディオケーブルの中の導線と圧電素子を接続し、圧電素子を両鼻当て部分に設置し、ケーブルをメガネのフレームに沿わせている。交代制鼻閉などによる呼吸量の変化にも対応するため、圧電素子を二つ用いて、両鼻の呼吸を取れるようにした。メガネ型デバイスを図1(B)に示す。

4.2 ソフトウェア

本研究では2つのソフトウェアを使用している。一つ目はデバイスから音声を取得するための音声編集ソフトウェアであるAudacity、二つ目は処理プロセス、呼吸パターンの認識を行うソフトウェアである。こちらのソフトウェアはPython言語で実装した。

5. 評価1

本章では、提案手法の有効性を検証するために、呼吸ジェスチャ8種類に対する提案手法の認識精度を評価した。被験者は男女10名(男性9名、女性1名)、大学生、平均年齢22歳(最大23歳、最小19歳)だった。

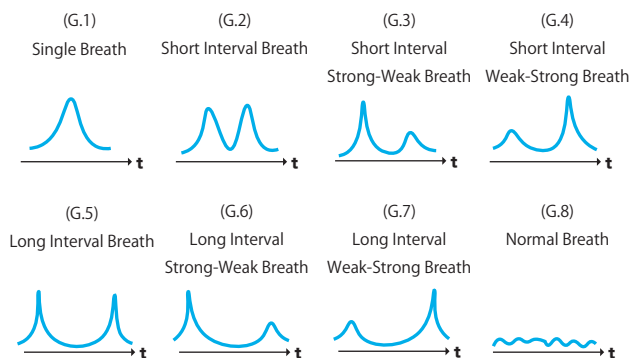


図 6 ジェスチャの種類

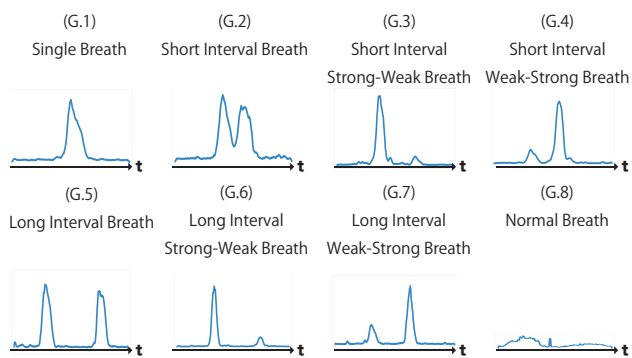


図 7 実際のジェスチャ波形例

呼吸ジェスチャの種類

本実験では、8種類の鼻呼吸ジェスチャを用意した。鼻呼吸ジェスチャ8種類のイメージを図6に示す。Single Breath (G.1)は呼吸を一回、Short Interval Breath (G.2)は同程度の強さの間隔が短い呼吸、Short Interval Strong-Weak Breath (G.3)は1回目を強く2回目を弱く、音量に相対差をつけた短い間隔の呼吸、Short Interval Weak-Strong Breath (G.4)はG.3の強弱が逆の呼吸、Long Interval Breath (G.5)は同程度の強さの間隔が長い呼吸、Long Interval Strong-Weak Breath (G.6)はG.3と同じ強弱で間隔が長い呼吸、Long Interval Weak-Strong Breath (G.7)はG.4と同じ強弱で間隔が長い呼吸、Normal Breath (G.8)は通常時の鼻呼吸の切り出しである。切り出す間隔は6秒である。また、実際に取得したジェスチャ波形の例を図7に示す。

実験手順

まず、実験で行うジェスチャの説明を行った。具体的には、手本のジェスチャ例として実験監督者が事前に録音した呼吸ジェスチャの音声を被験者に聞かせた。そして、実験監督者の説明と指導を受けて、被験者は呼吸ジェスチャを練習した。練習時間は最大30秒とした。また、呼吸ジェスチャに関して、吸う吐くの指定はせず、被験者にとってやりやすい方を選択してもらった。各ジェスチャ内での、吸う吐くは統一してもらった。評価1のデータ収集は次のように行なった。

表 1 評価1 被験者ごとの精度 (R: Recall, P: Precision, F: F-measure)

subjects	R	P	F
Sub.1	0.94	0.94	0.94
Sub.2	0.38	0.35	0.31
Sub.3	0.80	0.82	0.79
Sub.4	0.92	0.93	0.92
Sub.5	0.89	0.90	0.89
Sub.6	0.86	0.87	0.86
Sub.7	0.97	0.98	0.97
Sub.8	0.86	0.87	0.86
Sub.9	0.89	0.90	0.89
Sub.10	0.81	0.82	0.81
Average	0.83	0.84	0.82

	F[%]								
G.1	66	5	4	1	0	1	1	2	78%
G.2	5	67	2	6	0	0	0	0	82%
G.3	9	8	57	3	0	3	0	0	75%
G.4	4	7	1	68	0	0	0	0	80%
G.5	3	0	0	0	60	8	9	0	79%
G.6	9	0	0	2	3	66	0	0	80%
G.7	2	0	0	1	8	0	69	0	85%
G.8	1	0	0	0	0	0	0	79	98%
	1	2	3	4	5	6	7	8	

図 8 評価1. 混同行列

まず、6秒の通常呼吸 (G.8) を10試行行った次に、G.1~G.7の各ジェスチャを10試行ずつ行ったこの実験中、被験者は椅子に座って実験を行った。全評価実験終了後に、やりづらいジェスチャのヒアリング調査を行った。データセットは、8種類の呼吸ジェスチャ×10試行で構成され、呼吸ジェスチャのサンプル数は被験者一人当たり80であった。3章で述べた通常呼吸とジェスチャ用呼吸のデータを用いて、閾値を決定するための学習データに、10試行のデータのうち2試行のデータを利用した。そして、残りの8試行のデータに対して、提案手法の認識精度を評価した。

5.1 結果

被験者ごとの認識結果を表1に示す。表1には、Recall, Precision, F値を示す。図8に被験者全体の混同行列を示す。私たちの認識アルゴリズムは、すべての被験者で平均82%のF値を達成した。これは、提案された方法を使用して鼻呼吸ジェスチャを認識できることを示している。また、G.8の通常呼吸の精度が高いことから、ジェスチャ用呼吸と通常呼吸との認識はできていると考える。

一方、実験結果は、いくつかのジェスチャを実行するのが難しいことも示した (図 8)。

1つ目として、短い間隔の呼吸が長い間隔の呼吸に比べて精度が低い傾向にある。これは、呼吸の時間間隔が長いジェスチャ (G.6, G.7) は、呼吸の時間間隔が短いジェスチャ (G.3, G.4) よりも F 値が高いことからわかる。これらから、呼吸の時間間隔が長いジェスチャの方が実行しやすいことがわかる。2つ目に、呼吸強度の変化に関しては、強弱の順の方が弱強の順よりも精度が低い傾向にある。これは、強弱の順の呼吸ジェスチャ (G.3, G.6) よりも、弱強の順の呼吸ジェスチャ (G.4, G.7) の方が F 値が高いことからわかる。実際一部の被験者は、この 2 点を満たした G.3 が最も難しいとコメントしていた。

6. 評価 2

提案手法の認識精度を低下させ得ると考えられる複数の条件において、提案手法の認識精度を評価する。今回は、条件として再装着、身体振動を採用した。評価 1 同様、被験者は男女 10 名 (男性 9 名, 女性 1 名), 大学生, 平均年齢 22 歳 (最大 23 歳, 最小 19 歳) であった。

条件 1. 再装着条件

本条件では、デバイス再装着のずれによる提案手法の精度の変化を評価する。この条件からは、センサデバイス再装着後の提案手法の利用の際に、学習データを取得しなおす必要があるかがわかる。

手順: 評価 1 同様、実験前に各ジェスチャの説明を行い、デバイスを装着した。データ収集は次のように行なった。まず、デバイスを再装着する。次に、G.1~G.7 の各ジェスチャを 5 試行ずつ行う。そして、再びデバイスを再装着する。同様に、再び G.1~G.7 の各ジェスチャを 5 試行ずつ行う。以上のように、再装着を 2 回行い、計 10 試行分のデータを得た。この実験中も被験者は椅子に座って実験を行った。検証においては、評価 1 と被験者は同一であるため、評価 1 の際に取得した被験者ごとの閾値と同様のものを使用し認識を行う。そして、2 回再装着した 10 試行のデータに対して、提案手法の認識精度を評価した。

条件 2. 身体振動条件

続いて、本条件では、歩行程度の身体振動による提案手法の精度の変化を評価する。

手順: 評価 1 同様、実験前に各ジェスチャの説明を行い、デバイスを装着した。データ収集は次のように行なった。まず、その場で歩行 (足踏み) を行いながら、G.1~G.7 の各ジェスチャを 5 試行ずつ行った。歩行スピードは時速 2 マイル程度とした。以上のように、歩行 (足踏み) 状態を行い、5 試行分のデータを得た。検証においては、こちらも評価 1 と被験者は同一であるため、評価 1 の際に取得した被験者ごとの閾値と同様のものを使用し認識を行う。そして、5 試行のデータに対して、提案手法の認識精

表 2 評価 2 被験者ごとの精度 (R: Recall, P: Precision, F: F-measure)

	再装着条件			身体振動条件		
	R	P	F	R	P	F
Sub.1	0.89	0.86	0.87	0.95	0.95	0.95
Sub.2	0.62	0.67	0.57	0.63	0.63	0.61
Sub.3	0.76	0.84	0.75	0.78	0.78	0.75
Sub.4	0.85	0.90	0.84	0.88	0.90	0.87
Sub.5	0.97	0.97	0.96	0.90	0.92	0.90
Sub.6	0.73	0.80	0.76	0.58	0.46	0.51
Sub.7	0.94	0.95	0.93	0.73	0.78	0.70
Sub.8	0.92	0.94	0.91	0.75	0.78	0.76
Sub.9	0.93	0.94	0.93	0.93	0.94	0.92
Sub.10	0.94	0.95	0.94	0.88	0.91	0.87
Average	0.86	0.88	0.84	0.80	0.81	0.78

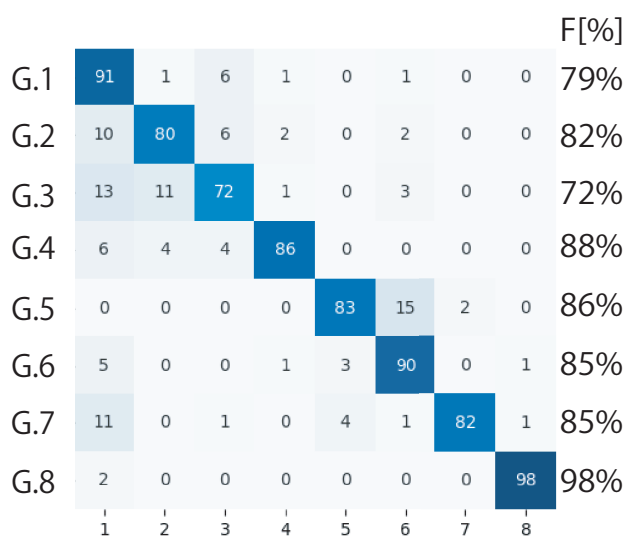


図 9 評価 2. 混同行列 (再装着条件)

度を評価した。

6.1 結果

被験者ごとの認識結果を表 2 左に示す。表には、Recall, Precision, F 値を示す。図 9 に被験者全体の混同行列を示す。

まず、再装着条件の結果について述べる。提案した認識アルゴリズムは、すべての被験者で平均 84% の F 値を達成した。各被験者において、評価 1 の結果に比べて多少精度は前後したものの、全体平均精度においては評価 1 との差が少なかった。また、被験者ごとの精度の順には大きな違いがなかったため、提案された方法を使用して、再装着によるズレが存在しても鼻呼吸ジェスチャを認識できることを示している。

また評価 1 同様、いくつかのジェスチャを実行するのが難しいことも示した (図 9)。評価 1 同様、呼吸の時間間隔が長いジェスチャ (G.5, G.6) は、呼吸の時間間隔が短いジェスチャ (G.2, G.3) よりも F 値が高い傾向にあった。G.5 は G.6 と誤認識されることが多かった。これは、呼吸

	F[%]								
G.1	47	0	1	1	0	0	0	1	89%
G.2	1	39	6	3	0	0	1	0	80%
G.3	4	5	29	6	1	4	1	0	57%
G.4	2	1	3	39	0	2	3	0	77%
G.5	0	0	1	0	43	4	2	0	79%
G.6	0	1	1	2	6	40	0	0	78%
G.7	2	2	2	0	9	2	33	0	70%
G.8	0	0	0	0	1	0	0	49	98%
	1	2	3	4	5	6	7	8	

図 10 評価 2. 混同行列 (身体振動条件)

の間隔を長くすると、前の呼吸強度を覚えておらず、次の呼吸をするときに目的の強弱をつけづらくなってしまったためだと考えられる。また、短い間隔の呼吸は、二回呼吸しているつもりでも呼吸が連続して認識されてしまい、G.1の一回呼吸と誤認識されてしまうジェスチャ (G.2, G.3, G.4, G.5, G.6) が多かった。F 値を見ると一回呼吸の G.1 の値が低く見られる。他の連続するジェスチャの弱呼吸などが認識されないと、その呼吸ジェスチャは一回呼吸と誤認識されてしまう。そのため、G.1 と誤認識されるジェスチャが多く、どうしても F 値が低くなってしまふ。しかし、正解率で見れば他のジェスチャよりもいい精度であることがわかる。また、評価 1 同様、G.3 が最も認識しづらかった。

被験者ごとの認識結果を表 2 右に示す。表 2 には、Recall, Precision, F 値を示す。図 10 に被験者全体の混同行列を示す。次に、身体振動条件の結果について述べる。私たちの認識アルゴリズムは、すべての被験者で平均 78% の F 値を達成した。これは、提案された方法を使用して歩行状態でも呼吸ジェスチャを認識できることを示している。しかし、座っている状態の評価実験に比べると F 値は低くなっている。これは、足を動かしながら呼吸のリズムを変更することが、着席時に比べて難しいことが理由と考えられる。実際、被験者から「足を動かしながら強弱をつける呼吸は難しい」という意見があった。歩行などの体を動かしている状態では一定のリズムで呼吸している場合が多く、そのリズムや強度を変更することは着席時に比べて難しいことが原因だと考えられる。

7. 考察、今後の課題

評価実験の結果は、眼鏡の鼻当ての圧電素子を使用して鼻呼吸を認識できることを証明した。また、提案手法の実現可能性も確認した。提案手法は、視力矯正用の眼鏡、VR ヘッドマウントディスプレイ、AR グラスなどのスマート

グラス、鼻当てを備えたあらゆるウェアラブルデバイスに広く使用できる。アイウェアデバイスを用いて、簡単に鼻呼吸を認識できる技術が普及すれば、鼻呼吸を使った新しいサービスへの応用が期待できる。具体的には、呼気量や頻度に基づいた、鼻づまりや呼吸の深さなどの健康情報の定量化などが期待できる。

全評価実験を通して、提案手法では 80% 前後の F 値を得た。被験者の中で、Sub.2 は他の被験者に比べて精度が低い傾向にあった。この理由は、被験者 sub.2 にとって鼻呼吸ジェスチャ自体が苦手であったためと考えられる。実際、評価実験後のアンケート調査において、Sub.2 は「呼吸ジェスチャ全体が苦手である」と回答した。普段から人の大半は通常時の鼻呼吸を行なっていると思われるが、強弱をつけたり、呼吸間隔を変えるなどの呼吸コントロールには、呼吸を使用した経験 (スポーツなど) によって影響を受ける場合があるのかもしれないと推測される。この点も調査していく予定である。

提案された手法の精度を向上させる 1 つの方法は、実行しやすいジェスチャのみを使用することである。8 種類のジェスチャを用意しましたが、約 5 種類のジェスチャでさまざまなアプリケーションのハンズフリー入力が可能であるため、先行研究 [22] では 5 種類のジェスチャを用いて評価した。したがって、将来的には、今回の評価で用いた呼吸回数、時間間隔、呼吸強度の 3 要素の限定などからジェスチャの数を減らして評価を行い、新しいジェスチャを検討する予定である。

また一部の被験者より、デバイスの装着に関して、センサを設置しているメガネの鼻当て部分に違和感を感じるとの意見があった。本デバイスはスポーツサングラス型のメガネに圧電素子を設置しているため、個人によって鼻当て部分の調節を行うことができない。そのため、適切な位置に圧電素子が設置されない場合があり、音量が思ったように取れないことがあったと考えられる。将来的には、個人ごとに適切な位置にセンサーを設置する方法・デバイスを設計し、鼻呼吸ジェスチャが苦手なユーザー向けのジェスチャを設計する予定である。また、今後は評価実験の際に NASA-TLX などの主観的尺度を計測したり、呼吸ジェスチャの外部へのノイズとしての音量を測定する実験を行うことも考えている。ジェスチャに対する意見などの主観的な評価を得ることで、ジェスチャの限定や新デバイスの実装の際に活用できると推測される。

8. おわりに

本研究では、鼻呼吸情報を用いたハンズフリー入力手法の提案を行なった。そのために、圧電素子を使用したメガネタイプのデバイスを実装し、呼吸回数、時間間隔、呼吸強度を使用して 8 種類の鼻呼吸ジェスチャを設計し、認識を行い評価した。10 人の被験者の評価結果は、平均 82% の

F 値で 8 種類の鼻呼吸ジェスチャを認識できることを示した。今後は被験者を増やして評価を行う。また実験では、8 種類のジェスチャを実装したメガネ型デバイスを用いて認識を行ったが、今後は様々なデバイスの実装と呼吸ジェスチャの限定を行う。本実験ではできるだけ静かな場所でデータの取得を行なったが、今後は騒音のある場所、人の会話などのノイズがある場所での精度の変化などを評価する。本研究は、ハンズフリー入力に鼻呼吸を使用する最初のウェアラブルセンシング技術を提供する。

参考文献

- [1] Lyons, M. J., Chan, C. H., and Tetsutani, N, MouthType: text entry by hand and mouth, CHI EA '04: CHI '04 Extended Abstracts on Human Factors in Computing Systems, pp. 1383–1386 (2004).
- [2] Koguchi, Y., Oharada, K., Takagi, Y., Sawada, Y., Shizuki, B., and Takahashi, S, A Mobile Command Input Through Vowel Lip Shape Recognition, In International Conference on Human-Computer Interaction, pp. 297–305 (2018).
- [3] Takuro Nakao, Yun Suen Pai, Megumi Isogai, Hideaki Kimata, and Kai Kunze, Make-a-face: a hands-free, non-intrusive device for tongue/mouth/cheek input using emg, In ACM SIGGRAPH 2018 Posters, pp. 1–2 (2018).
- [4] Augusto Esteves, Eduardo Velloso, Andreas Bulling, and Hans Gellersen, Orbits: Gaze Interaction for Smart Watches using Smooth Pursuit Eye Movements, In Proceedings of the 28th Annual ACM Symposium on User Interface Software and Technology, pp. 457–466 (2015).
- [5] Katsutoshi Masai, Kai Kunze, Yuta Sugiura, Masa Ogata, Masahiko Inami, and Maki Sugimoto, Evaluation of Facial Expression Recognition by a Smart Eyewear for Facial Direction Changes, Repeatability, and Positional Drift, ACM Transactions on Interactive Intelligent Systems (TiiS), 7(4), pp. 1–23 (2017).
- [6] Yan, Y., Yu, C., Yi, X., and Shi, Y, Headgesture: Hands-free input approach leveraging head movements for hmd devices, Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 2(4), pp. 1–23 (2018).
- [7] Ando, T., Kubo, Y., Shizuki, B., and Takahashi, S, Canalsense: Face-related movement recognition system based on sensing air pressure in ear canals, In Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology, pp. 679–689 (2017).
- [8] Lutz, O. H. M., Venjakob, A. C., and Ruff, S, SMOOVS: Towards calibration-free text entry by gaze using smooth pursuit movements, Journal of Eye Movement Research, 8(1), (2015).
- [9] Manabe, H., Fukumoto, M., and Yagi, T., Conductive rubber electrodes for earphone-based eye gesture input interface. Personal and Ubiquitous Computing, 19(1), pp. 143–154 (2015).
- [10] Matthies, D. J., Strecker, B. A., and Urban, B., Earfield-sensing: A novel in-ear electric field sensing to enrich wearable gesture input through facial expressions. In Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, pp. 1911–1922 (2017).
- [11] Taniguchi, K., Kondo, H., Kurosawa, M., and Nishikawa, A. Earable TEMPO: a novel, hands-free input device that uses the movement of the tongue measured with a wearable ear sensor. Sensors, 18(3):733 (2018).
- [12] Hashimoto, T., Low, S., Fujita, K., Usumi, R., et al., 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 (2018).
- [13] Chauhan, J., Hu, Y., Seneviratne, S., Misra, A., Seneviratne, A., and Lee, Y, BreathPrint: Breathing acoustics-based user authentication, In Proceedings of the 15th Annual International Conference on Mobile Systems, Applications, and Services, pp. 278–291 (2017).
- [14] Shih, C. H., Tomita, N., Lukic, Y. X., Reguera, Á. H., Fleisch, E., and Kowatsch, T, Breeze: Smartphone-based acoustic real-time detection of breathing phases for a gamified biofeedback breath training, Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 3(4), pp. 1–30 (2019).
- [15] Tennent, P., Rowland, D., Marshall, J., Egglestone, S. R., Harrison, A., Jaime, Z., ... and Benford, S, Breathalising games: understanding the potential of breath control in game interfaces, In Proceedings of the 8th international conference on advances in computer entertainment technology, pp. 1–8 (2011).
- [16] Sra, M., Xu, X., and Maes, P, Breathvr: Leveraging breathing as a directly controlled interface for virtual reality games, In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, pp. 1–12 (2018).
- [17] Evreinov, G., and Evreinova, T, " Breath-Joystick"-Graphical Manipulator for Physically Disabled Users, Proc. of the ICCHP2000, pp. 193–200 (2000).
- [18] Yamamoto, M., Ikeda, T., and Sasaki, Y, Real-time analog input device using breath pressure for the operation of powered wheelchair, In 2008 IEEE International Conference on Robotics and Automation, pp. 3914–3919, (2008).
- [19] Marshall, J., Rowland, D., Rennick Egglestone, S., Benford, S., Walker, B., and McAuley, D, Breath control of amusement rides, In Proceedings of the SIGCHI conference on Human Factors in computing systems, pp. 73–82 (2011).
- [20] Héctor A. Cordourier Maruri, Paulo Lopez-Meyer, Jonathan Huang, Willem Marco Beltman, Lama Nachman, and Hong Lu, V-Speech: Noise-Robust Speech Capturing Glasses Using Vibration Sensors, Proc. ACM Interact. Mob. Wearable Ubiquitous Technol, 2(4), pp. 180 (2018).
- [21] He, Jibo and Chaparro, Alex and Nguyen, Bobby and Burge, Rondell and Crandall, Joseph and Chaparro, Barbara and Ni, Rui and Cao, Shi, Texting while driving: Is speech-based texting less risky than handheld texting?, Proceedings of the 5th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, pp. 124–130 (2013).
- [22] Amesaka, T., Watanabe, H., and Sugimoto, M., Facial expression recognition using ear canal transfer function, In Proceedings of the 23rd International Symposium on Wearable Computers, pp. 1–9 (2019).