

# アクティブ音響センシングを用いた 手のジェスチャー認識手法の検討

山田 知之<sup>†1,a)</sup> 棟方 渚<sup>†1,b)</sup> 小野 哲雄<sup>†1,c)</sup>

**概要:** 本研究では、振動スピーカとピエゾマイクを装着し、人体の周波数特性を利用するアクティブ音響センシングによる手のジェスチャー認識の手法について検討を行った。直感的に操作可能なインターフェースは、アートパフォーマンスやVRゲームなどの応用可能性を持つため、手のジェスチャー認識の研究は数多く行われている。しかし、提案手法によるジェスチャー認識の研究は、プロトタイピングが容易にも関わらず、今まで行われていない。本研究では4種類のジェスチャーを用い、装着位置を変えた3条件の元で実験を行った。結果として、ジェスチャーを変えることで、周波数特性が変化することを確認し、本手法の実現可能性があることを示した。また、複数名の被験者による認識精度の調査を行い、現状のシステムの問題点について検討した。

## Investigation of Hand Gestures Recognition Method using Active Acoustic Sensing

TOMOYUKI YAMADA<sup>†1,a)</sup> NAGISA MUNEKATA<sup>†1,b)</sup> TETSUO ONO<sup>†1,c)</sup>

### 1. はじめに

身振りや手振りを用いて入力を行うジェスチャーインターフェースに関する研究は、これまで数多く行われている。中でも、随意的に動かしやすい手のジェスチャーに関するものは、視覚を用いる必要がないことや、直感的な入力が可能であることから応用の幅が広く、日常生活やエンターテインメントに活用することが可能である。例えば、Leap Motion[1]はカメラを利用した3D空間内におけるジェスチャー認識を可能とする小型デバイスで、0.01mmと非常に高い精度でジェスチャーの認識を可能とする。このようなカメラベースのジェスチャー認識の研究は特に多く、他の手法と比べて精度の点で優れるが、カメラそのものを必要とすることや、センサーとなるカメラと手にある程度の距離を必要とすることから、利用できる状況が限ら

れてしまうという課題もある。

これに対し、手にセンサを装着することで、ユビキタなジェスチャー認識を行う研究も多く行われている。例えばDementyevらのWristFlex[2]は、手の甲の腱の動きをアレイ状に配置した圧力センサを用いて検出し、ジェスチャー認識を行っている。他にも、筋電や静電容量、手首のインピーダンスなどなど、様々なアプローチで研究が行われており、カメラベースのように、デバイスの場所に縛られないジェスチャー認識を実現している。これらは、装着するセンサの大きさや装着位置、消費電力、認識精度、製作コストが異なるため、単純に性能を比較することは難しい。また、実環境において発生するノイズに対するロバスト性も異なるため、様々な手法を検討することは非常に重要である。

そこで我々は、アクティブ音響センシングと呼ばれる技術を用いて、環境に縛られることなく、片手のみでのジェスチャー認識を行う手法を検討する。アクティブ音響センシングとは、自ら信号となる音を発信し、それに対する反応の違いを利用してセンシングを行う技術である。ジェス

<sup>†1</sup> 現在、北海道大学

Presently with Hokkaido University

a) yamada@complex.ist.hokudai.ac.jp

b) munekata@complex.ist.hokudai.ac.jp

c) tono@ist.hokudai.ac.jp

チャータ認識にアクティブ音響センシングを用いる利点として、周波数特性という高次元の特徴量を利用できること、スピーカーとマイクのみで実装でき、プロトタイプングが比較的容易であることが挙げられる。また、スピーカーを用いずに音を発することが可能であるため、タップなどの動作からスピーカーのオンオフを操作することも可能であると考えられる。日常生活における利用を考えた際、消費電力の問題を解決し得るため、重要な特徴である。

本手法は、1組のスピーカーとマイクを手や腕に装着し、スピーカーから発した高周波のスイープ音をマイクによって捉えることで、音が伝播した部分の周波数特性を計測する。ジェスチャーごとの周波数特性の違いから、ジェスチャー認識を行う。

本稿では、まず本手法の実現のために実装したシステムについて述べる。次に、実装したシステムを用いて実験を行い、実現可能性の有無と現状のシステムの問題点について議論する。

## 2. 関連研究

### 2.1 手のジェスチャー認識に関する研究

前述の通り、手のジェスチャー認識に関する研究は数多く存在している。

LinらのBackHand[3]は、WristFlexと同様、腕の動き取得しているが、手首ではなく手の甲においてセンシングを行っており、腕の動きの取得のために、アレイ状に配置した張力センサを用いているという点で異なる。手首の腕と比べて動きが大きいため、多くの種類のジェスチャーを認識することができるが、日常生活においては、手の甲はほぼ常に露出しているため、装着すること自体に抵抗を感じてしまう可能性がある。これに対し、提案手法は人体の周波数特性を利用するため、装着位置の比較的制約が少なく、手首など装着した場合でも、ジェスチャー認識を実現し得る。

ZhangらのTomos[4]は、電極をアレイ状に配置したりストバンドデバイスであり、電気インピーダンス・トモグラフィ（以下EIT）と呼ばれる技術により、人体のインピーダンスを計測することでジェスチャー認識を可能とする。しかし、EITの都合上、電極の位置の変化に非常に敏感であり、手首の輪郭の僅かな変化により、認識精度が下がってしまう。

SatoらのTouché[5]は、静電容量センシングを用いて物体をタッチセンサー化する技術であるが、これを人体に用いることで、両手でのジェスチャー認識を可能とする。この研究では、SFCS (Swept Frequency Capacitive Sensing) という技術を用いている。SFCSは1kHzから3.5MHzの間の200種類の周波数における静電容量の値から認識を行うものである。また、36mm×36mmの小さなセンシング基盤と物体を1つの電極を介してつなげるだけで機能する

ため、非常に手軽に利用できる。

### 2.2 音響センシングを応用した研究

音響を用いたセンシング技術に関する研究も多く行われている。yataniらのBody Scope[6]やRahmanらのBody Beat[7]は、首にマイクを取り付け、取得した音をスペクトログラムを用いて分析することで、食べる、飲む、喋るなどの動作の認識を可能とする。AmentらのThe Sound of One Hand[8]やDeyleらのHambone[9]は、手首にマイクを取り付け、タップや指を弾く動作の際に体内で発生する音を時系列的に解析することで、認識を行う。

これらは、センシングにマイクのみを用いるパッシブセンシングであり、スピーカーを必要としないため、比較の実装が容易であるが、体内で大きな音が発生する必要があるため、本研究で目的とするジェスチャー認識に応用するのは難しい。

一方、スピーカーとマイクを用いたアクティブセンシングも、様々な研究に応用されている。OnoらによるTouch & Activate[10]は、1組の振動スピーカーとピエゾマイクを物体に貼り付け、触れ方による周波数特性の変化を利用することで、手軽かつ安価に物体の把持状態認識を実現している。TakenakaらのActive Bone-Conducted[11]は、上腕に1つのスピーカーと2つのマイクを装着し、取得した音の解析を行うことで、肘の関節の角度を認識することができる。MujibiyaらのThe Sound of Touch[12]は、スピーカーとマイクを搭載した腕輪型のデバイスにより、上腕をタッチインターフェース化する研究である。この研究においては、35kHz、40kHz、45kHz、50kHzの4種類の超音波を用い、同腕輪か、別の手の任意の指に装着したマイクで音を取得・解析することで、上腕のどの部分に触れているかを認識する。GuptaらのSoundWave[12]は、PCに内蔵しているスピーカーとマイクを用い、ドップラー効果を応用することで、PCの周囲でのin-airジェスチャーの認識を行う。

## 3. 実装

### 3.1 システム概要

本研究において実装したシステムの概要を図1に示す。また、このシステムは信号の出力・取得を行うハードウェア部分と、取得した信号の処理・認識を行うソフトウェア部分に別れる。それぞれの詳細を以下の節にて述べる。

### 3.2 ハードウェア

実装したシステムのハードウェア部分の詳細を以下に示す。ハードウェアは、振動スピーカー・ピエゾマイク・ハイパスフィルタ・オーディオインターフェース・コンピュータからなる。

#### 振動スピーカー

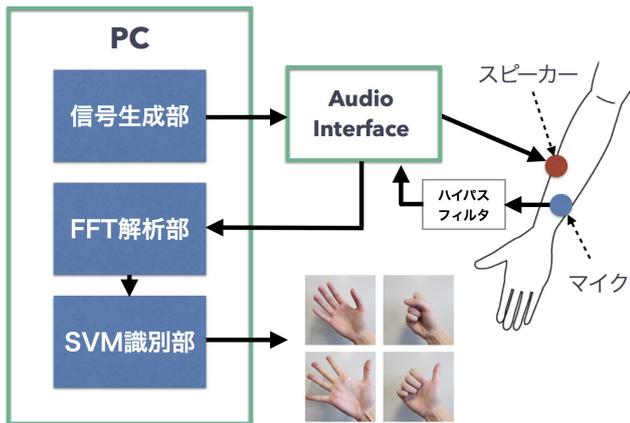


図 1 システム概要図

Fig. 1 System Configuration.

振動スピーカは、物体に貼り付けることで、その物体に振動を伝え、音を出力するスピーカである。本システムでは、スライブ社の圧電バイモルフ型スピーカである OMR20F10-BP-310 を用いた。この部品を選択した理由は、厚さ 0.3mm、直径 21mm と薄く小型であり、人体に貼り付けることが可能なためである。

また、ノイズ対策のため、振動スピーカとオーディオインターフェースとを繋ぐケーブルを絶縁テープでカバーした。

#### ピエゾマイク

ピエゾマイクは、圧電素子を用いたマイクで、固体伝播音を取得することができる。振動スピーカと同様、薄く小型であり、人体に貼り付けることが可能であるため選択した。用いたのは、ARIO COMPANY の D27E40B で、厚さ 0.5mm、直径 27mm である。

振動スピーカと同様、ノイズ対策のため、ケーブルを絶縁テープでカバーした。

#### ハイパスフィルタ

ジェスチャー認識に用いる周波数帯は 20kHz から 37.5kHz であるので、日常生活において発生する衣擦れなどの低周波ノイズは必要ない。そのため、オーディオインターフェースに音を入力する前の段階でハイパスフィルタに通し、ノイズの低減を試みた。用いた抵抗は 180  $\Omega$  の金属皮膜抵抗、コンデンサは 0.047  $\mu\text{F}$  のマイラーコンデンサである。これにより、18812Hz 以下の信号を減衰させることができるが、現状全てのパーツがケーブルで繋がっているため、ハイパスフィルタを通った後も様々な理由でノイズが発生してしまい、不要な信号を全てカットすることはできていない。

#### オーディオインターフェース

振動スピーカ・ピエゾマイクとコンピュータを繋ぐ役割を果たす。一般的にピエゾマイクが生成できる電圧は非常に小さく、そのままでは信号解析に用いることができないため、信号を増幅する必要がある。そのため、アンプ機

能付きのオーディオインターフェースとして、Roland 社の UA-1000 を用いた。振動スピーカ・ピエゾマイクとは 6.3mm のモノラルプラグで、コンピュータとは USB で接続されている。

#### コンピュータ

信号出力・信号処理・機械学習を行うため、Apple 社の MacBook Pro (Retina, 15-inch, Mid 2014) を用いた。

### 3.3 ソフトウェア

ソフトウェアは、信号生成部・FFT 解析部・SVM 識別部の 3 つからなる。これらは基本的に C++ 言語及び openFrameworks を用いて作成したが、SVM 識別部の一部の Python で書かれたプログラムを使用している。

#### 信号生成

信号生成部では、振動スピーカで出力する音響信号を生成する。人体を伝播した後の周波数特性を調べるため、生成する音響信号は、解析に用いる周波数帯である 20kHz から 37.5kHz の音を含む sin 波のスイープ音とし、この音響信号の元となる音源は Audacity というアプリケーションを用いて作成した。生成時のサンプリング周波数は 96kHz としている。

具体的には、20ms の間に 20kHz から 37.5kHz まで上昇し、20ms で 20kHz まで戻るといった線形な変化を約 10 秒間繰り返すものとなっている。これをソフトウェアによりループさせて用いているため、約 10 秒毎に波形が少しぶれるが、この変化自体は一瞬のものであるため問題にはならないと判断した。この条件であれば、人間の可聴音域を超えるため聞こえることはなく、マイクと皮膚が擦れた時に発生するノイズや、体音の影響が少ないと考えられる。また、上限を 37.5kHz とした理由は、エイリアシング誤差の低減のためである。サンプリング定理より、サンプリング周波数の半分である 48kHz までは理論上原信号の復元が可能であるが、FFT アナライザに含まれるアンチエイリアシングフィルタの性能によっては、エイリアシング誤差が発生してしまう可能性があるため、余裕を持ってサンプリング周波数を 2.56 で割った値である 37.5kHz を用いている。

#### FFT 解析

FFT 解析部では、ピエゾマイクから取得した信号に対して FFT 解析を行い、周波数スペクトルに変換する。FFT を行うプログラムは、openFrameworks のアドオンである ofxFft を用い、サンプリング周波数は生成時の信号と同じく 96kHz、フレームサイズは 8192 とし、窓関数はハンニング窓とした。FFT 自体は毎秒 60 回行っているが、ノイズの影響により常に値がぶれているため、直前 10 回分の FFT の結果の平均を用いている。また、オーディオインターフェース由来のものであると思われるが、30kHz 前後の周波数帯において極端に大きいノイズが定期的に発生し、認識に影響をあたえることが多々あったため、実験 1 におい



図 2 ジェスチャーの一覧  
Fig. 2 Gesture list.

ては、正規化前に 30kHz 前後のスペクトルを示す部分も値を 0 に上書きした。ただ、この方法では、多くの特徴量を失うことになる上、他の周波数のノイズの影響を受けている可能性もあるので、実験 2 からは、認識に用いる範囲のスペクトルのうち、振幅の大きいものを 30 個削ることとした。これらの作業により、ノイズの低減を試みた。

最後に、FFT 解析後に得られるデータは 0Hz から 48kHz までの情報を含んでいるため、認識に用いる 20kHz から 37.5kHz 以外の値を全て 0 にした上で正規化を行った。

### SVM 識別

SVM 識別部では、機械学習アルゴリズムの 1 つである Support Vector Machine を用いて、 Piezo マイクから取得した信号の周波数特性と、手のジェスチャーの関係の学習及びリアルタイム識別を行う。機械学習に用いる特徴量は、教師データ・テストデータともに、FFT 解析後に 20kHz から 37.5kHz の以外の値を取り除いて正規化を行ったもので、float 型の 1708 次元配列である。

SVM による識別は 10 フレームに 1 回行っているが、1 回の結果から最終的な識別結果を決めてしまうと、ジェスチャーの途中で誤認識されたり、ばらつきが大きくなって判断しにくくなってしまうため、過去 9 回分の結果を記録し、その中で最も回数が多かったものを最終的な識別結果とした。ただし、結果のばらつきが激しく、1 つに確定できない場合は、識別不能とした。

このパートの実装のため、SVM のライブラリとして頻繁に用いられる LIBSVM を使用し、カーネルは RBF を選択した。また、SVM はパラメータの調整が認識精度に大きく影響を与えるため、学習データのサンプリングが終わった段階でグリッドサーチを行い、パラメータの最適化を行った。このグリッドサーチには、LIBSVM に付属する grid.py というプログラムを用いている。

## 4. 実験

### 4.1 動作の検証 (実験 1)

実装したシステムを用いて、本手法の動作の確認及び実現可能性の検討をする実験を行った。4 種類のジェスチャーを用い、装着位置を変更した 3 つの条件のもと、認識精度を調査した。被験者は 1 名のみである。実験を行う前に何度かシステムには触れており、扱い慣れていた。用いたジェスチャー及び装着位置は、図 2 と図 3 に示す。

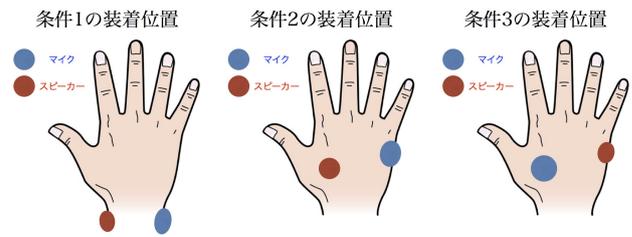


図 3 装着位置の一覧  
Fig. 3 Condition list.

#### 4.1.1 実験手順

実験は、以下の手順で行う。

- (1) 振動スピーカーと Piezo マイクを、サージカルテープを用いて装着する。
- (2) 各ジェスチャーを 3 回ずつ行い、学習データの取得、パラメータの最適化、学習を行う。
- (3) 予め決められた順番通りにジェスチャーを行い、結果を記録する。ジェスチャーの順番は、遷移 (Relax から Fist, Relax から Stretch など) の回数が概ね均等になるように設定している。1 つのジェスチャーにつき 25 回、合計 100 回ジェスチャーを行う。
- (4) 振動スピーカーとマイクの装着位置を変え、(1) から (3) の手順を繰り返す。

#### 4.1.2 実験結果

上記の実験により得られた結果を以下に示す。

##### 認識精度

各条件におけるジェスチャーの認識精度を図 4 に示す。

実験した 3 条件の中では、スピーカーとマイクを手手に装着した条件 1 が最も精度が高く、すべてのジェスチャーにおいて 80% 以上の精度で認識に成功した。一方、条件 1 に対してスピーカーとマイクの装着位置を入れ替えた条件 2 は最も精度が低く、Stretch ジェスチャーは全く認識することができなかった。手首にスピーカーとマイクを装着した条件 3 は、Stretch ジェスチャー以外は比較的高い精度でジェスチャー認識に成功した。

##### 周波数特性

本実験において最も認識精度が高かった条件 1 における

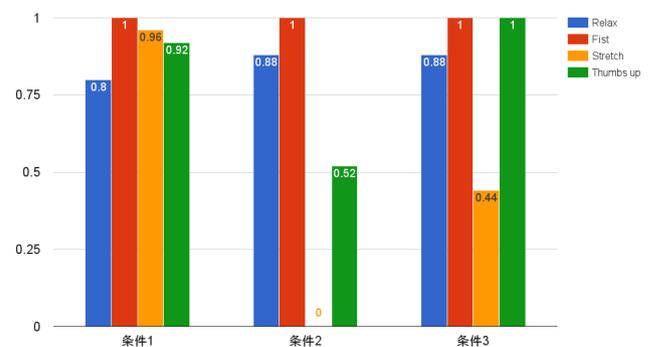


図 4 実験 1 における認識精度

Fig. 4 Recognition Accuracy on Experiment 1.

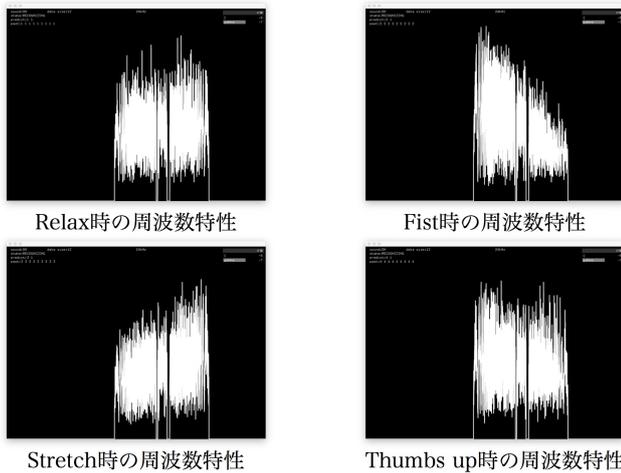


図 5 条件 1 における各ジェスチャーの周波数特性

Fig. 5 The Frequency Response of Each Gesture on Condition 1.

各ジェスチャー時の周波数特性を図 5 に示した。ジェスチャーごとの周波数特性の違いは一目瞭然であり、これに基づいて高い精度で認識に成功していることから、ジェスチャーにより周波数特性が変化することと、SVM による分類が正しく行えていることが示された。

各条件における混合行列

加えて、各条件におけるジェスチャーの認識精度の混同行列を図 6 に示す。これにより、ジェスチャー認識に失敗した時に、他のどのジェスチャーと誤認識していたか検証できる。ここで、他のジェスチャーに間違えられた確率を、被誤認識率と呼ぶこととする。

条件 1 については、そもそも全てのジェスチャーにおいて高い精度で認識できていたため、誤認識の仕方にはあまり傾向は見られなかったが、Fist ジェスチャーの被誤認識率が 0 であったことから、特に認識がしやすかったとわかる。これは、図 5 の各周波数特性を見比べることで理解できる。

条件 2 については、Stretch ジェスチャーや Thumbs up ジェスチャーを Relax ジェスチャーに誤認識している確率が高く、3 ジェスチャーの周波数特性にあまり違いがなかったことがわかる。一方、条件 1 と同様、Fist ジェスチャーの被誤認識率は 0 であった。

条件 3 については、Stretch ジェスチャーを Relax ジェスチャーや Fist ジェスチャーに誤認識している確率が高いが、Relax ジェスチャーと Stretch ジェスチャーを統合することで、3 種類のジェスチャー認識については可能になると考えられる。

4.2 評価実験 (実験 2)

実験 1 により、ジェスチャーによる周波数特性の変化と、SVM による学習・認識が出来ていることは確認できた。これに加え、現段階での本手法の実用性を検証するた

	Relax	Fist	Stretch	Thumbs up
Relax	0.8	0	0.08	0.12
Fist	0	1	0	0
Stretch	0	0	0.96	0.04
Thumbs up	0.08	0	0	0.92

条件1の混合行列

	Relax	Fist	Stretch	Thumbs up
Relax	0.88	0	0	0.12
Fist	0	1	0	0
Stretch	0.76	0	0	0.24
Thumbs up	0.48	0	0	0.52

条件2の混合行列

	Relax	Fist	Stretch	Thumbs up
Relax	0.88	0.08	0.04	0
Fist	0	1	0	0
Stretch	0.28	0.24	0.44	0.04
Thumbs up	0	0	0	1

条件3の混合行列

図 6 各条件における混同行列

Fig. 6 Confusion Matrix on Each Condition.

め、被験者を増やして実験を行った。実験の内容は概ね実験 1 と同様であるが、FFT 解析部におけるノイズ処理の方法を変更した。これに加え、スピーカー・マイクの大きさに対しての手の大きさが結果に影響すると考えたため、寸法項目一覧 [13] に基づき、最大手幅、手厚：第 5 中手骨頭と手首周りの長さを計測した。また、装着作業は全て実験者である筆者が行った。実験時間は 1 人につき約 1 時間半であった。

4.2.1 実験結果

本実験において得た結果を以下に示す。

認識精度

それぞれの被験者の各条件における認識率を図 7 に示す。

実験 1 と比較すると認識精度が低下している。これについて考えられる理由として、被験者のシステムに対する慣れが挙げられる。実験 1 においては、被験者は実験の前にある程度システムを利用していたが、本実験における被験者は、システムを全く利用したことのない状態で実験を行っていたため、このような差が生まれたと考えられる。

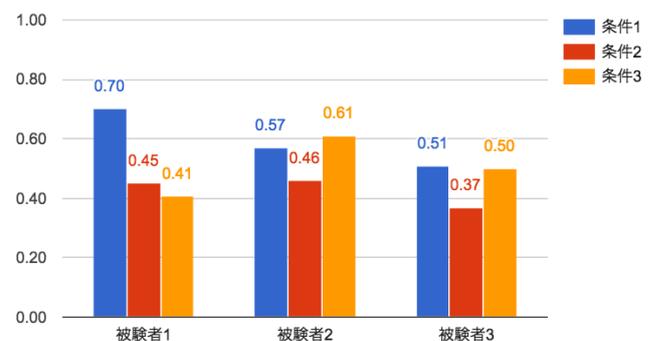


図 7 実験 2 における認識精度

Fig. 7 Recognition Accuracy on Experiment 2.

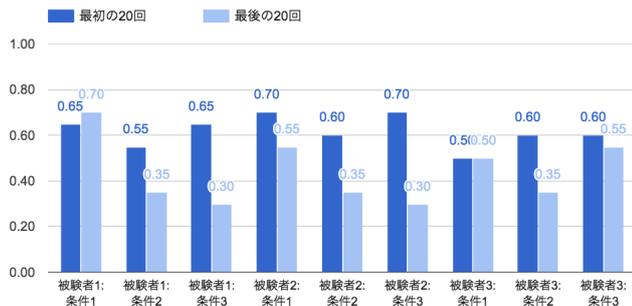


図 8 実験前半と後半の認識精度の比較

Fig. 8 Comparison of Recognition Accuracy between Former and Later of Experiment 2.

また、実験の前半と後半で認識精度の差が観察された。そこで、100回の試行の内、最初の20回と最後の20回の認識精度を比較した。その結果を図8に示す。

本実験の最初の20回のみを抽出すると、装着するのは1組のスピーカーとマイクのためのシンプルなシステムであり、被験者はこのシステムに慣れていないにも関わらず、平均して62%の精度でジェスチャーの認識が出来ていた。一方で、最後の20回の認識精度は、平均して約44%まで低下している。これも、被験者のシステムへの慣れの有無が関係していると考えられる。

現状のシステムでは、サージカルテープを用いてスピーカー・マイクの固定を行っているが、何度もジェスチャーを行うことで、少しずつ装着位置がずれたり、テープが剥がれたりしてしまう。本実験の3人の被験者は、このシステムに慣れていなかったため、スピーカー・マイクの状態の変化に気付かず、このような結果となったのではないかと考えられる。また、100回という試行回数が適切でなく、集中力が低下してしまった結果、無意識の内にジェスチャーの形が僅かに変化してしまった可能性も考えられる。システムに慣れることによって、これらの問題に対処することは出来るが、安定した装着方法について検討したり、ジェスチャーの形の僅かな変化を許容できるようなシステムにしたりする必要がある。

また、実験1では条件1が最も認識精度が高く、条件2が最も低かったが、この実験における被験者3人に関しては、結果が異なり、個人差が見られた。

#### 条件1における各被験者の混合行列

次に、条件1における各被験者ごとの混合行列を、図9に示す。行の値の和が1にならない場合があるが、これは識別不能と判定されたためである。

被験者1についてはThumbs upジェスチャー以外の3ジェスチャーが、被験者2についてはFistジェスチャーとStretchジェスチャーが、被験者3についてはStretchジェスチャーとThumbs upジェスチャーの認識精度が高くなった。この結果より、同一条件でも、被験者によって認識しやすいジェスチャーが異なることがわかる。

	Relax	Fist	Stretch	Thumbs up
Relax	1	0	0	0
Fist	0.04	0.84	0	0.12
Stretch	0	0.04	0.96	0
Thumbs up	0	1	0	0

被験者1の混合行列

	Relax	Fist	Stretch	Thumbs up
Relax	0.08	0.76	0.16	0
Fist	0	0.92	0.08	0
Stretch	0.16	0.08	0.76	0
Thumbs up	0	0.48	0	0.52

被験者2の混合行列

	Relax	Fist	Stretch	Thumbs up
Relax	0.32	0.04	0.04	0.52
Fist	0	0.04	0	0.96
Stretch	0.28	0	0.72	0
Thumbs up	0	0	0	0.96

被験者3の混合行列

図 9 条件1における各被験者の混合行列

Fig. 9 Confusion Matrix of Each Subject on Condition 1.

この理由として、被験者間のジェスチャーの違いが考えられる。本実験に用いた4つのジェスチャーのうち、特にRelaxジェスチャーとThumbs upジェスチャーについて、被験者ごとに形が僅かに異なることが実験を通して観察された。手の力を抜いた状態であるRelaxジェスチャーの形の違いは特に顕著で、手を開いてStretchジェスチャーに近くなっている被験者と、指を丸めFistジェスチャーに近くなっている被験者が存在した。Thumbs upジェスチャーについても、親指の角度が被験者間で僅かに異なっていた。

今後も各被験者ごとに訓練データを取得すれば、被験者間のジェスチャーの違いから認識精度が低下することはないが、実用化のため、個人のジェスチャーの特徴と、認識精度の傾向にどのような関係があるか調査する必要がある。

## 5. 議論

### 5.1 提案手法による手のジェスチャー認識の実現可能性

実験1により、ジェスチャーにより周波数特性が変化したことが示された。この変化をジェスチャー認識に用いることで、提案手法は十分に実現可能性があると考えられる。実際に、条件1については、全てのジェスチャーについて、80%を超える精度で認識できた。しかし、実験2の結果から、現状のシステムは、まだ実用的とは言えないため、以下に示すような多数の問題点を改善し、実験室環境における精度の向上、実環境下での利用に向けたシステムの改善を行っていく必要がある。

### 5.2 現状のシステムの問題点

#### 装着方法

現状のシステムでは、サージカルテープを用いてスピー

カーとマイクを装着しているが、人によってはすぐ剥がれてしまったり、実験2で起こったように、ジェスチャーを行うことで少しずつずれてしまう可能性がある。装着に時間がかかるため、容易に装着可能であり、かつ安定した装着方法を検討しなければならない。

### ノイズ対策

現状すべてのシステムがケーブルで繋がれている。そのため、フィルタを通したあとも、僅かなフィルタの動きによってノイズが発生してしまう。また、振動スピーカーから発生する振動も、ケーブルを伝って僅かにマイクに届いてしまうため、無線化を行うことでノイズの低減が可能である。また、今後実環境での実験を行うにあたり、より多くのノイズに対抗する必要がある。ノイズそのものを取り除くことも重要であるが、ノイズの中でも正しく認識するようなシステムを実装しなければならない。

### 他動作時のジェスチャー認識

現状のシステムでは、座った状態でのジェスチャー認識しか想定していない。実環境では、他の動作中にもジェスチャー認識が必要になる場面が想定され、その場合の周波数特性は座っている状態での周波数特性とは異なると考えられるため、この点に関しても対策が必要となる。

### ジェスチャー遷移と認識精度

今回の実験において、ジェスチャーの遷移回数ができるだけ均一になるように順番を決めた。データ数が充分でないため予測の域を出ないが、遷移前のジェスチャーの種類によって、僅かに異なったジェスチャーを意図せず行ってしまうと考えている。例えば、Fist ジェスチャーから Thumbs up ジェスチャーに遷移する場合と、Stretch ジェスチャーから Thumbs up ジェスチャーに遷移する場合とでは、親指の力の入り具合が異なる可能性がある。

現状この点については対策を取っていない。遷移の順番も考慮した上で訓練データを取ることで、ある程度改善が期待できるが、ユーザの負担も同時に増えてしまうため、より良い改善策を考える必要がある。

また、一般的なウェアラブルセンサに共通する問題として、以下の項目についても検討する必要がある。

- 手の大きさや体型、ジェスチャーの形などの個人差
- 使用するセンサ数と装着時の負荷

## 6. 終わりに

我々は、1組の振動スピーカーとピエゾマイクを用いてアクティブ音響センシングを行うことにより、手のジェスチャー認識を行う手法を検討し、その実現可能性の調査のため、実験を行った。結果として、ジェスチャーによって周波数特性が変化することを確認し、本手法に実現可能性があることを示した。また、複数名の被験者による認識精度の調査を行い、現状のシステムの問題点について検討した。

## 参考文献

- [1] Leap Motion, available from <https://www.leapmotion.com/> (accessed 2016-07-11).
- [2] Dementyev, A. and Paradiso, J. A.: WristFlex: Low-power Gesture Input with Wrist-worn Pressure Sensors, *Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology*, UIST '14, New York, NY, USA, ACM, pp. 161–166 (online), DOI: 10.1145/2642918.2647396 (2014).
- [3] Lin, J.-W., Wang, C., Huang, Y. Y., Chou, K.-T., Chen, H.-Y., Tseng, W.-L. and Chen, M. Y.: Back-Hand: Sensing Hand Gestures via Back of the Hand, *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology*, UIST '15, New York, NY, USA, ACM, pp. 557–564 (online), DOI: 10.1145/2807442.2807462 (2015).
- [4] Zhang, Y. and Harrison, C.: Tomo: Wearable, Low-Cost Electrical Impedance Tomography for Hand Gesture Recognition, *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology*, UIST '15, New York, NY, USA, ACM, pp. 167–173 (online), DOI: 10.1145/2807442.2807480 (2015).
- [5] Popyrev, I., Harrison, C. and Sato, M.: Touch&Gesture Sensing for the Real World, *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*, UbiComp '12, New York, NY, USA, ACM, pp. 536–536 (online), DOI: 10.1145/2370216.2370296 (2012).
- [6] Yatani, K. and Truong, K. N.: BodyScope: A Wearable Acoustic Sensor for Activity Recognition, *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*, UbiComp '12, New York, NY, USA, ACM, pp. 341–350 (online), DOI: 10.1145/2370216.2370269 (2012).
- [7] Rahman, T., Adams, A. T., Zhang, M., Cherry, E., Zhou, B., Peng, H. and Choudhury, T.: BodyBeat: A Mobile System for Sensing Non-speech Body Sounds, *Proceedings of the 12th Annual International Conference on Mobile Systems, Applications, and Services*, MobiSys '14, New York, NY, USA, ACM, pp. 2–13 (online), DOI: 10.1145/2594368.2594386 (2014).
- [8] Amento, B., Hill, W. and Terveen, L.: The Sound of One Hand: A Wrist-mounted Bio-acoustic Fingertip Gesture Interface, *CHI '02 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '02, New York, NY, USA, ACM, pp. 724–725 (online), DOI: 10.1145/506443.506566 (2002).
- [9] Deyle, T., Palinko, S., Poole, E. S. and Starner, T.: Hambone: A Bio-Acoustic Gesture Interface, *Proceedings of the 2007 11th IEEE International Symposium on Wearable Computers*, ISWC '07, Washington, DC, USA, IEEE Computer Society, pp. 1–8 (online), DOI: 10.1109/ISWC.2007.4373768 (2007).
- [10] Ono, M., Shizuki, B. and Tanaka, J.: Touch & Activate: Adding Interactivity to Existing Objects Using Active Acoustic Sensing, *Proceedings of the 26th Annual ACM Symposium on User Interface Software and Technology*, UIST '13, New York, NY, USA, ACM, pp. 31–40 (online), DOI: 10.1145/2501988.2501989 (2013).
- [11] Takemura, K., Ito, A., Takamatsu, J. and Ogasawara, T.: Active Bone-conducted Sound Sensing for Wearable Interfaces, *Proceedings of the 24th Annual ACM Symposium Adjunct on User Interface Software and Technology*, UIST '11 Adjunct, New York, NY, USA, ACM, pp. 53–54 (online), DOI: 10.1145/2046396.2046419 (2011).
- [12] Mujibiya, A., Cao, X., Tan, D. S., Morris, D., Patel,

S. N. and Rekimoto, J.: The Sound of Touch: On-body Touch and Gesture Sensing Based on Transdermal Ultrasound Propagation, *Proceedings of the 2013 ACM International Conference on Interactive Tabletops and Surfaces, ITS '13*, New York, NY, USA, ACM, pp. 189–198 (online), DOI: 10.1145/2512349.2512821 (2013).

- [13] AIST 日本人の手の寸法データ, Digital Human Research Center, AIST (オンライン), 入手先 (<https://www.dh.aist.go.jp/database/hand/data/list.html>) (参照 2016-07-11).