



部分遮蔽されたマイクを用いた超音波によるジェスチャ認識手法

渡邊, 拓貴
寺田, 努

(Citation)

ヒューマンインタフェース学会論文誌, 22(2):175-186

(Issue Date)

2020-05-25

(Resource Type)

journal article

(Version)

Version of Record

(Rights)

© 2020 ヒューマンインタフェース学会

本著作物の著作権は特定非営利活動法人ヒューマンインタフェース学会に帰属します。本著作物は著作権者であるヒューマンインタフェース学会の許可のもとに掲載するものです。ご利用に当たっては「著作権法」に従うことをお願い致します。

(URL)

<https://hdl.handle.net/20.500.14094/0100492903>



部分遮蔽されたマイクを用いた超音波によるジェスチャ認識手法

渡邊 拓貴^{*1} 寺田 努^{*2}

Using Partially Shielded Microphone for Improving Ultrasound-based Gesture Recognition

Hiroki Watanabe^{*1}, Tsutomu Terada^{*2}

Abstract – We propose a method to improve ultrasound-based in-air gesture recognition by altering the acoustic characteristics of a microphone. The Doppler effect is often used to recognize ultrasound-based gestures. However, increasing the number of gestures is difficult because of the limited information obtained from the Doppler effect. In this study, we partially shield a microphone with a 3D-printed cover. The cover alters the sensitivity of the microphone and the characteristics of the obtained Doppler effect. Since the proposed method uses a 3D-printed cover with a single microphone and speaker embedded in a device, it does not require additional electronic devices to improve gesture recognition. We design four different microphone covers and evaluate the performance of the proposed method on six gestures with eight participants. The evaluation results confirm that the recognition accuracy improves from 60.4% to 75.7% on average using the proposed method. We also discuss the velocity of the hand gesture and the effect of the cover for conventional usage of the microphone.

Keywords : gesture recognition, ultrasound, wearable computing, mobile computing

1. はじめに

近年のコンピュータ小型化に伴い、スマートフォンやスマートウォッチ等の小型デバイスを用いた、モバイル/ウェアラブルコンピューティング環境に注目が集まっている。これらのデバイスを操作するのに、主に用いられる手法はタッチ操作である。タッチ操作はマルチタッチやスワイプ等による直感的な操作が可能である。しかし、小型デバイスの画面上でのタッチ操作の問題の一つとして、操作を行う指自体で画面を覆ってしまうという問題点がある。さらに、タッチ操作時には画面に直接触れる必要があるが、料理中や機械の作業中のような手が濡れている/汚れている場合や、タッチパネルに反応しない手袋を装着している場合のように、画面に触れたくない/触れても反応しない状況がある。そのため、空中でのジェスチャが取得できれば、より便利である。

空中でのジェスチャ入力を可能にする手法として、追加のデバイスを用いる手法がある。赤外線距離センサをデバイス周囲に配置し、画面外でのジェスチャ認識を可能にする手法は複数提案されている^{[1]~[3], [15]}。しかし、この手法では、電力を消費する追加のデバイスが必要になる。追加デバイスがいらない手法として、デバイス内蔵のカメラを用いた手法も提案されてい

る^{[4]~[6]}。しかし、カメラを用いた手法では、一般的に計算コストが大きいことや、プライバシーの問題が考えられる。デバイス内蔵のセンサを用いたジェスチャ認識手法として、マイクとスピーカによる超音波を用いたジェスチャ認識手法も提案されている^{[7]~[9]}。これらの手法ではドップラー効果や複数のマイクへの到達時間差 (TDoA: Time Difference of Arrival) を利用しているが、ドップラー効果では対象が近づいているか/遠ざかっているかと、その速度しか得られないため、ジェスチャの種類によっては認識することは困難である。また、TDoAを利用するためには複数のマイクが必要であるが、スマートウォッチは複数のマイクを搭載していないことが一般的である。

本研究では、スマートウォッチ等のウェアラブルデバイスを考慮し、デバイスが1組のマイクとスピーカだけを内蔵している環境を想定する。文献^{[6], [10], [11]}で用いられているように、センサ部に物理的な機構を加えることで、取得できる情報量を増やす手法をマイクに適応する。具体的には、デバイスのマイク部分に、片側からの音は通常通り取得でき、もう片側からの音は取得しにくいような機構のカバーを装着することで、一つのマイクで取得できる情報量を増やすことを目的とする。従来、ドップラー効果を用いると、マイクに対して左右対称である動きを一つのマイクで識別することは困難であったが、提案手法によってこれらが区別できる。本研究では、マイクを部分的に遮蔽するためのカバーとして、half shield, hole, sponge,

^{*1}: 北海道大学大学院 情報科学研究所

^{*2}: 神戸大学大学院 工学研究科

^{*1}: Graduate School of Information Science and Technology, Hokkaido University

^{*2}: Graduate School of Engineering, Kobe University

directivity の 4 種類を提案した。予備実験を行なった結果、directivity カバーが最も提案手法に適していることが確認できた。提案手法の評価のため、directivity カバーを 4 種類のデバイスにおいて実装し評価を行なった結果、マイクカバー無しの状態と比較して、マイクカバーを装着することでジェスチャ認識の精度が平均で 60.4%から 75.7%に改善することが確認できた。本研究の貢献は以下の 3 点である。

- 電力消費の必要な追加デバイスが必要ない超音波によるジェスチャ認識改善手法を提案した。
- 4 種類のマイクカバーを考案し、directivity タイプが提案手法に最も適していることを明らかにした。
- Directivity マイクカバーが、マイクに対して左右対称の動作を含むジェスチャ認識の改善に有効であることを評価実験から示した。

なお、本論文は文献^[12]で発表した内容に、ジェスチャ速度に対する考察、およびマイクカバー装着による従来のマイク用途への影響の考察を加えて発展させ、全体的に補足を行なったものである。

以降、2 章で本研究に関連する研究について述べ、3 章で提案手法について述べる。4 章で実装を紹介し、5 章で評価実験について示す。6 章で考察を行い、最後に 7 章でまとめを行う。

2. 関連研究

2.1 様々なセンサを用いたインタラクション技術

様々なセンサを用いてデバイスとのインタラクションを拡張した研究は広く行われている。Harrison らは、腕に装着したマイクアレイを用いて、腕上のタップ位置を認識する手法を提案した^[13]。Ishikawa らは 4 つの距離センサを利用した空中ジェスチャによる入力手法を提案した^[14]。この手法では、センサの反応した順番がコマンドとして利用されている。SkinButtons はスマートウォッチ上でのインタラクション領域を拡張させる手法である^[15]。近接センサと小型のプロジェクタがスマートウォッチの横に装着され、肌上をインタラクション領域として利用できる。SideSight, HoverFlow は、スマートフォンの周囲に赤外線距離センサを装着することで、デバイス周囲での指の位置やハンドジェスチャを取得できるシステムである^[1]。^[2]。Manabe は単一のフォトリフレクタを用いて、マルチタッチの認識を可能にする手法を提案した^[10]。SenSkin は、フォトリフレクタアレイによって肌の変形を認識することで、肌上の入力を可能にする手法である^[16]。zSense は、赤外線センサを用いてウェアラブルデバイスでのジェスチャ入力を拡張する手法である^[3]。Tongue-in-Cheek は、ユーザの顔周辺に設置したレーダのドップラー効果を用いて顔関連のジェスチャを取

得する研究である^[17]。

上記のように、モバイル/ウェアラブル環境においてジェスチャ認識を拡張する研究は多く行われている。しかし、これらの研究ではジェスチャ認識技術を拡張するために、電力を消費する追加のデバイスが必要である。モバイル/ウェアラブル環境での限られた電力を考慮すると、電力消費のあるデバイスを追加することは望ましくない。本研究では、デバイス内蔵のマイクとスピーカを利用し、3D プリントされたマイクカバーによってジェスチャ認識を改善するため、電力を消費する追加のデバイスが必要ない。

Song らは、デバイス内蔵のカメラを用いることで空中のジェスチャを認識する手法を提案した^[4]。Digits は、手首に装着された赤外線カメラを用いて、ユーザの手の 3D 形状を認識するシステムである^[5]。Surround-See は、スマートフォンのフロントカメラに全方位ミラーを装着することで、ジェスチャ認識の範囲を拡張する研究である^[6]。しかし、カメラを用いた手法では、プライバシーの問題が懸念される。また、画像の処理は一般的に計算量が多いという問題がある。

Whoosh はデバイス内蔵のマイクを用いて、スマートウォッチの操作を息で行う手法である^[18]。Siri や Google Assistant のような音声認識技術を用いた音声アシスタントもデバイスの操作方法として利用できる^[19]。^[20]。しかし、可聴音は環境音に影響されやすいという問題がある。本研究では、超音波を利用するため環境音に影響されにくいという点でこれらの研究とは異なる。

2.2 超音波を用いたジェスチャ認識技術

Doplink, SurfaceLink, AirLink は超音波を用いることで、デバイス間でのインタラクションを可能にする手法である^{[21]~[23]}。本研究では、人とデバイスとのインタラクションに着目する点で異なる。

Chirp microsystems は、チップ上に実装された超音波レンジファインダによって、空中でのジェスチャを取得できる^[24]。しかし、この手法ではデバイスに特別なチップを搭載する必要がある。

Touch & Activate では、物体に一对のコンタクトマイクとコンタクトスピーカを装着することでタッチ入力を認識し、既存物体へのインタラクティブ性を付与している^[25]。Acoustruments では、スマートフォンのスピーカからマイクへと続く管のようなアタッチメントを装着し、スピーカから超音波を発信する。ユーザがアタッチメントに触れる位置によって伝播の特性が変化し、ユーザの操作が識別できる^[11]。これらの研究では、物体へのタッチインタラクションについて着目しており、本研究の目的とする空中でのジェスチャ認識とは異なる。

FingerIO と Strata は超音波エコーを用いた指先の

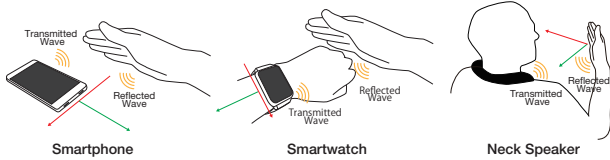


図1 想定環境

Fig.1 Assumed environments.

トラッキングシステムである [8], [26]. スマートフォン/ウォッチ型デバイスのスピーカから超音波を発信し、指先からの反射波の、二つのマイクへの到達時間差を用いて2次元のトラッキングを可能にしている. Sound-Trak も超音波を用いた指先のトラッキングシステムである [27]. ユーザは超音波を発信するリング上のデバイスを指に装着し、複数のマイクが搭載されたウェアラブルデバイスで超音波を受信することで、指先をトラッキングできる. しかし、上記の手法では認識のために複数のマイクが必要である.

SoundWave と AudioGest はデバイス内蔵のスピーカから超音波を発信し、反射波をデバイス内蔵のマイクで取得することでジェスチャ認識を可能にする手法である [7], [9]. これらの手法ではドップラー効果を利用しており、本研究はこれらの先行研究に基づいている. しかし、文献 [9] で述べているように、ドップラー効果を利用した手法では、ジェスチャの方向によっては識別することが難しい. したがって本研究では、部分遮蔽されたマイクを用いることで得られる特徴を増加させ、超音波を用いたジェスチャ認識を改良する手法を提案する. 提案手法は3D プリントされたマイクカバーをデバイスに装着するだけなので、電力消費のある追加デバイスが必要ない.

3. 提案手法

3.1 ドップラー効果によるジェスチャ認識

図1に示すように、本研究ではデバイス内蔵のスピーカから超音波を発信し、内蔵マイクで反射波を取得する環境を想定する. システムはドップラー効果により生じる周波数の遷移を利用して、空中でのハンドジェスチャを取得する. ドップラー効果は音源の移動によって生じ、観測者が停止している場合、以下の式で表せる.

$$f_r = f_t \left(\frac{c}{c-v} \right) \quad (1)$$

f_r は受信音の周波数, f_t は送信音の周波数, c は空気中での音速, v は音源の速度である. ただし、音源から観測者へ向かう向きを正とする. この式より、音源が観測者に近づいた時に受信音の周波数は上がり、遠ざかった時に下がることになる. 想定環境では、観測者はデバイス内蔵のマイク、音源はユーザの手とす

る. 実際の音源はデバイス内蔵のスピーカであるが、内蔵スピーカと内蔵マイクの位置は固定されているため、ドップラー効果は生じない. 本研究では、マイクで取得される超音波には、スピーカからマイクへとデバイス内部を通して伝わってくる直接音と、ユーザの手に反射して帰ってくる反射音の2種類がある. 音量に関しては直接音の方が大きい、ユーザのジェスチャによってドップラー効果が発生するのは反射音のみである. したがって、本研究では反射音を発生させるユーザの手を音源と考える. 発信する超音波の周波数は20kHzとした. これは、ほとんどの人間は20kHzの音を聞くことができないため、騒音とならないからである. また、既存のスマートフォンやスマートウォッチなどのスピーカも20kHzを発信できることが確認できたためである. さらに、超音波は環境音にほとんど含まれないため、影響を受けにくい. 本研究ではユーザのジェスチャ速度は一定であると想定した.

図2にジェスチャと対応するスペクトログラムを示す. この図に示すマイクとスピーカ構成においてスワイプを行うと、 v の $\cos\theta$ 成分によってドップラー効果が発生する. θ はジェスチャの進行方向と、マイクと手を結ぶ直線のなす角度である. 図2の左スワイプを例に考えると、 θ は手がマイクに近接すると大きくなり、遠ざかると小さくなる. 手が一定の速度で動いていると想定すると、ドップラー効果を引き起こす要素である $v\cos\theta$ は、マイクに遠い場所で大きく、マイクに近づくほど小さくなる. 手がマイクを通り過ぎた後には $v\cos\theta$ は負の値となり、ドップラー効果は逆方向へ生じる. マイクが手からの反射波を取得できない位置に到達すると、ドップラー効果は観測できなくなる. 図2に示すように、各ジェスチャによって得られるスペクトログラムが異なるため、この違いを利用してドップラー効果によりジェスチャを認識することが可能になる. しかし、単一のマイクでドップラー効果を用いた場合、ジェスチャによってはドップラー効果が似ているため認識できるジェスチャ数が制限される. 図2に示す左右スワイプは、似たドップラー効果を示すジェスチャの一例である. この図に示すように、ユーザの手がマイクに近づくときにドップラー効果はまず高周波側に発生し、マイクから通り過ぎる時は低周波側に発生する. ドップラー効果のみでは左右どちらからのスワイプも似たような特徴を示すため、これらのジェスチャを識別するのは難しい.

3.2 カバーデザイン

提案手法の基本的なアイデアは、マイクカバーを装着することでマイクの音響特性を変化させることである. 音響レンズの分野では、マイク/スピーカ部に物理的機構を加えることでマイク/スピーカの指向性や距離による感度を変化させる手法が提案されてい

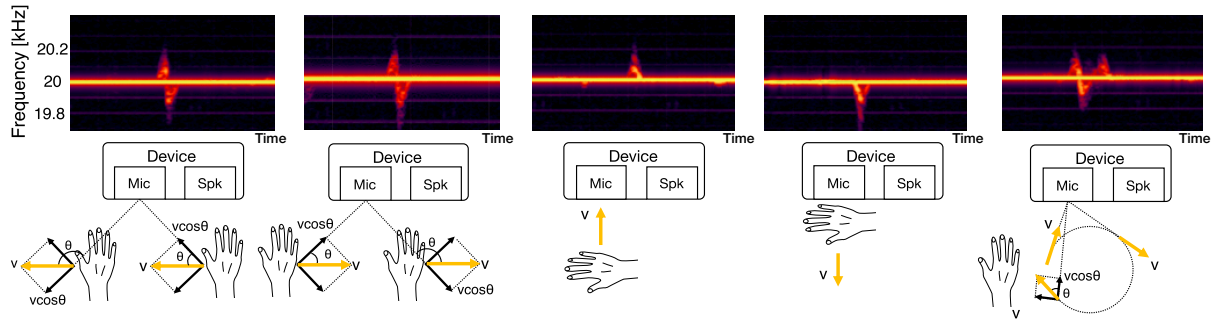


図2 ジェスチャと対応するスペクトログラム
Fig. 2 Gestures and corresponding spectrograms.

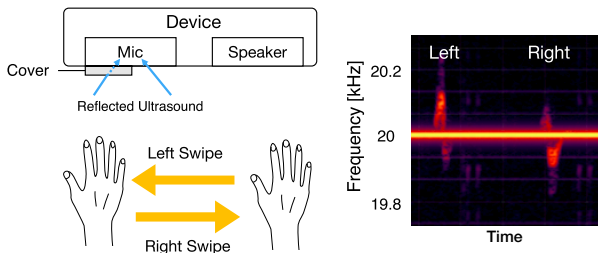


図3 マイクを部分遮蔽した時のジェスチャとスペクトログラム
Fig. 3 Gestures and spectrograms with microphone cover.

る [28], [29]. 提案手法は、これらの手法の応用だといえる。図3左に示すように、マイクカバーがマイクの片方を覆うことで、カバー側からの音量を減少させ、カバーされていない側からの音量は通常通りに取得できる。マイクカバー正面での一定速度の左右スワイプを考えると、マイクから遠いほどドップラー効果による周波数シフトは大きく、マイクに近い場所では周波数シフトは小さくなる。また、マイクで得られるユーザの手からの反射波の音量は、マイクから遠いほど小さく、マイクに近いほど大きい。図3左のようにマイクカバーを装着することで、マイク右側では通常通りのドップラー効果が得られる。一方、マイク左側では、マイク付近で生じるドップラーシフトが小さく音量が大きい反射波は、マイクカバーによって減衰しても取得できるが、マイク遠方で生じるドップラーシフトが大きく音量が小さい反射波は、マイクカバーによって減衰し、確認できなくなる。したがって、図3右に示すように、マイクカバーを装着することで、左スワイプと右スワイプを行なった際に得られるドップラー効果に変化が生じ、これらを識別できる。

マイクを部分的に遮蔽するカバーの実現方法として、図4に示す4種類のマイクカバーを考案した。これらのマイクカバーは、マイク左右で得られる音量を変化させることを考慮し、考案した。

- Half Shield: マイクの半分がカバーによって完全に隠されているマイクカバーである。このカバーにより、マイク右側からの音量は通常通りに取得しながら、マイク左側からの音量を減少させることができると考えた。
- Hole: マイク左側部分に直径 1mm の複数の穴を開けたカバーである。Half shield では、マイク左側を完全に遮蔽しており、マイク左側からの音量が十分に取得できない可能性を考え、カバー左側部分に複数の穴を設けた。このタイプのカバーでは、half shield よりもマイク左側からの音量が得られると考えた。
- Sponge: マイク左側が吸音性の物質で遮蔽されているカバーである。本研究では吸音性物質としてスポンジを利用した。スポンジをマイク左側に設置することで、左側からの音量を減少させ、右側からの音量を通常通りに取得することを考えた。
- Directivity: 管のようなカバーによってマイクの指向性を変化させるカバーである。カバーは右側からの音をより取得でき、マイク正面方向からの音も取得できるように、右側に 45 度の角度をつけた。上記3つのカバーデザインと異なり、directivity カバーでは左側からの音量を減少させるだけでなく、右側からの音量を増加させるような構造になっている。したがって、マイク左右間で得られる音量の差をより顕著にできると考えた。

考案したカバーのうち、提案手法に適したものを調査するために予備実験を行なった。利用したデバイスは Huawei P9 Lite¹ であり、対象としたジェスチャはマイク正面での左スワイプと右スワイプである。ジェスチャの識別は、発生したドップラー効果の低周波側と高周波側のどちらの変化量が大きいことによって識別した。被験者は男性 1 名であり、各マイクカバーにおいて各ジェスチャを 15 回行なった。

図5と表1に予備実験の結果を示す。この図は、ドッ

1: <https://consumer.huawei.com/jp/phones/p9-lite/>

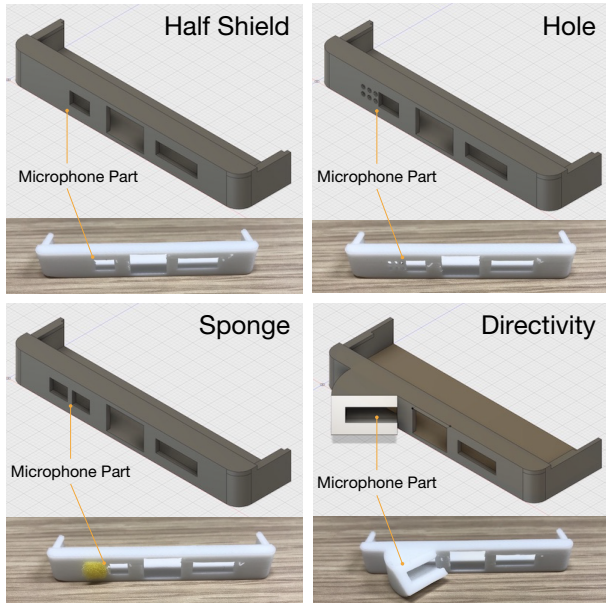


図4 カバーデザインの候補
Fig. 4 Cover design candidates.

プラー効果によって生じた低周波側の変化量を横軸、高周波側の変化量を縦軸としてプロットしている。つまり、左スワイプと右スワイプの分布が大きく離れているほどマイクカバーの性能が高いといえる。図5aが示すように、マイクカバー無しの場合では左スワイプと右スワイプの分布が混在しており、識別するのが難しいとわかる。マイクカバー無しの場合、認識精度は56.7%であった。図5b, dが示すように、half shieldとspongeも顕著な効果は見られなかった。認識精度はそれぞれ43.3%と50.0%であった。図5c, eが示すように、holeとdirectivityではマイクカバーの効果が確認できた。Holeカバーでは、左スワイプと右スワイプの分布に一定の乖離が見られるが、我々の想定した動きと異なっていた。今回、左側を遮蔽したため、例えば左スワイプの場合はマイクに近づく際のドップラー効果(高周波側)の変化量が大きく、マイクから遠ざかる際のドップラー効果(低周波側)の変化量が小さくなると思った。しかし、図5cに示すように、今回の予備実験では逆の現象が見られた。そのため、holeカバーでは他のカバーとは異なり、ドップラー効果の高周波側の変化量が大きければ右スワイプ、逆であれば左スワイプとして認識を行い、認識精度は90%であった。一方、directivityカバーはマイクの特性に想定通りの顕著な効果を与えており、左スワイプと右スワイプの分布に大きな乖離が確認できた。また、表1が示すように、directivityカバーは認識精度が100%であり、今回用いたマイクカバーの中で最も良い結果であった。本研究では、マイクカバーによって超音波によるジェスチャ認識が改善できるかに着目しているため、予備実験の結果から、最も性能の高かったdirectivityカバー

表1 予備実験の認識率 [%]

Table 1 Recognition rates of preliminarily experiment [%].

Cover	w/o	Half Shield	Hole	Sponge	Directivity
Ave.	56.7	43.3	90.0	50.0	100

を以降の調査では用いる。Holeカバーの詳細な特性解析については今後の課題とする。

3.3 認識手法

3.3.1 前処理

認識処理の流れを図6に示す。認識手法の一部は文献[9]の手法を参考にした。モバイル/ウェアラブルデバイスでは、得られる超音波の音量に変動があり、反響音の音量も小さいことが確認されている。また、本研究で用いたネックスピーカ(JBL SoundGear²)からの超音波が不安定であり、発信する超音波の周波数が変化している場合が確認できた。そこで、本研究では前処理によってこれらの不安定なマイク/スピーカの特性を補正する。

まず、マイクからの入力に対して、高速フーリエ変換(FFT: Fast Fourier Transform)を行い周波数スペクトルを得る(図6A)。サンプリング周波数は44.1kHzとし、FFTを計算する際のウィンドウは重複無しで2048サンプルとした。発信される超音波の不安定性を補正するため、周波数ピーク周囲の音を抜き出す。本研究では、18kHzから22kHzの範囲で最大値を取る周波数を検出し、そのピーク周波数の上下20ビンずつを取り出す。デバイスに対するジェスチャの速さは約3.9m/sであるため[7]、生じるドップラー効果は約218Hzである。本研究の設定では、周波数の分解能が約21.5Hz(44100/2048)であるため、ピーク周波数上下の11ビンを抽出すれば十分であるが、余裕を持って20ビンを抽出した。

モバイル/ウェアラブルデバイスにおけるスピーカ/マイクでは、超音波領域の出力/入力音量が不安定になる場合が確認できた。したがって、この時間による不安定さを除去するために、各フレームの周波数スペクトルに対して、以下の式に従い正規化を行なった(図6B)。

$$f_n(T) = \frac{f(T)}{f_{max}(T)} \quad (2)$$

$f(T)$ は時刻 T における FFT の結果、 $f_n(T)$ は正規化された FFT の結果、 $f_{max}(T)$ は $f(T)$ の最大値を示す。

超音波反響音の音量の小ささを補正するために変化量を強調した(図6C)。以下の式に示すように、正規化された周波数スペクトルを1フレーム前の周波数スペクトルから減算し自乗することで周波数スペクトル

2: <https://www.jbl.com/JBLSOUNDGEARBLK.html>

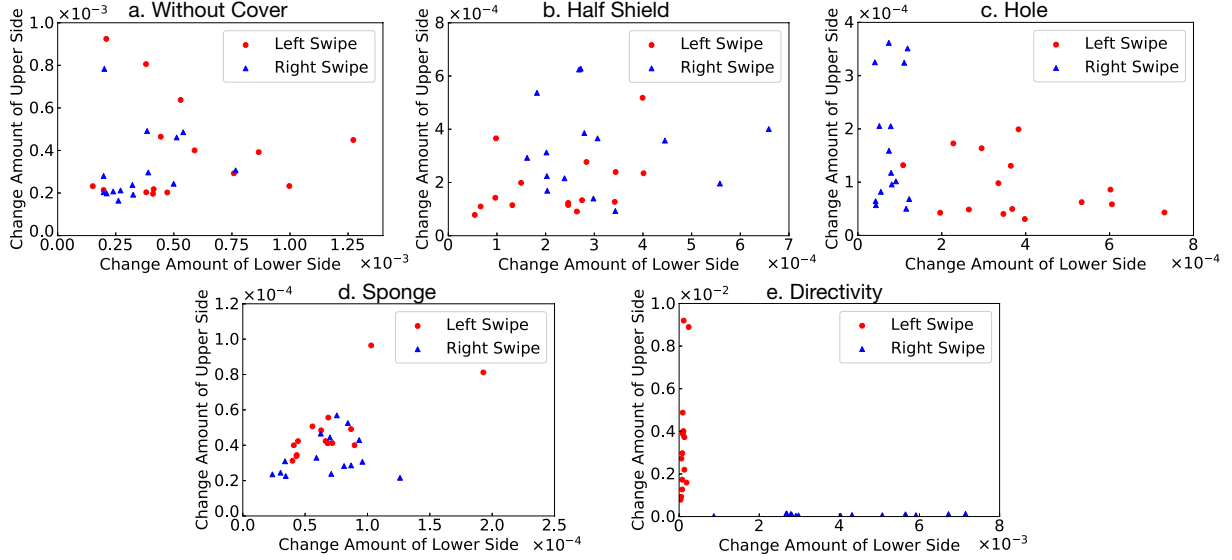


図5 予備実験の結果
Fig.5 Results of preliminary experiment.

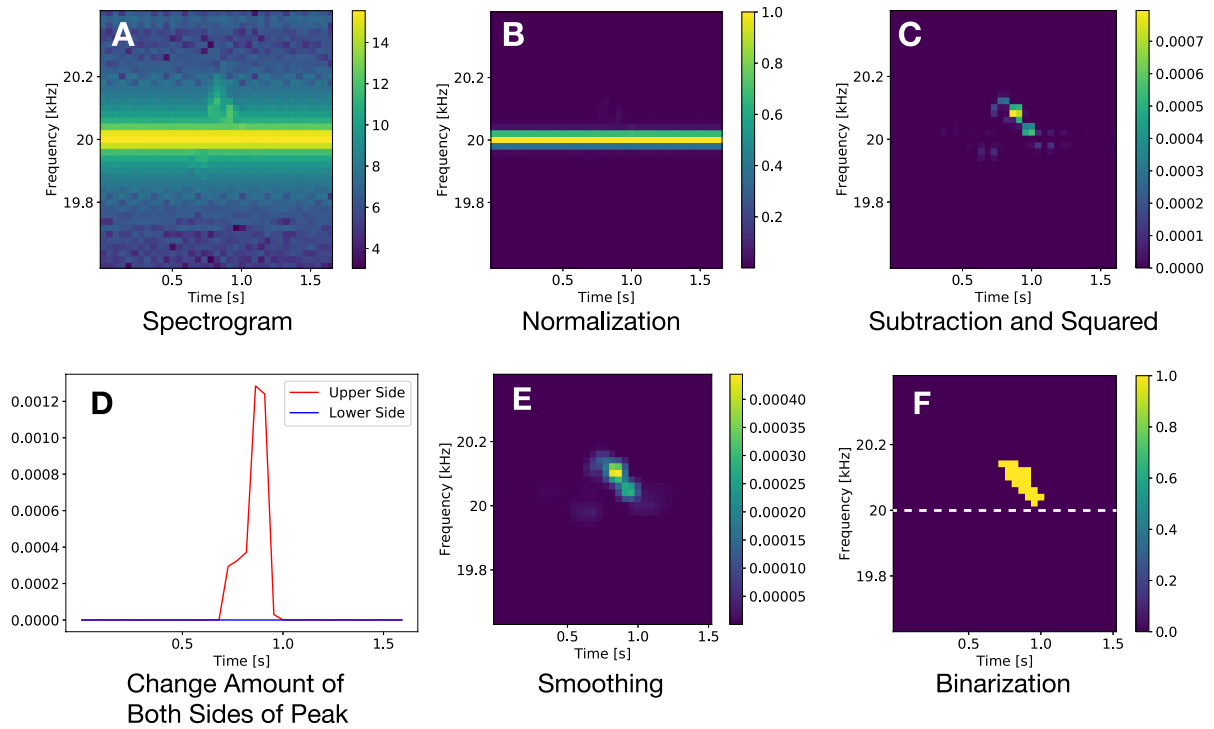


図6 認識処理
Fig.6 Recognition process.

における変化量を強調する。 $f_{em}(T)$ は強調後の周波数スペクトルである。

$$f_{em}(T) = (f_n(T) - f_n(T-1))^2 \quad (3)$$

3.3.2 特徴量抽出

本研究では、周波数ピーク周辺の変化量とスペクトログラムから特徴量を抽出した。

ピーク周波数周辺の変化量は、図 6D から得る。図 6D は、図 6C の各フレームにおける周波数ピーク上

側での変化量の合計、および周波数ピーク下側での変化量の合計をそれぞれ時系列でプロットしたものである。この時系列データから、ピーク周波数上側/下側の最大値、ピーク周波数上側/下側の平均値、ピーク周波数上側/下側の分散値、ピーク周波数上側/下側の積分値、ピーク周波数上側と下側の平均値の比、ピーク周波数上側と下側の積分値の比の合計 10 種類の特徴量を抽出する。

スペクトログラムからの特徴量を得るために、まず

スペクトログラムからジェスチャ関連部分を抽出する。以下の式で表すガウシアンフィルタによってスペクトログラムを平滑化する (図 6E)。

$$G(x, y) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right) \quad (4)$$

x は水平方向の原点からの距離, y は垂直方向の原点からの距離を表す。 σ^2 は、ガウス分布の分散である。図 6F に示すように、平滑化したスペクトログラムをしきい値で 2 値化した後に、ピーク周波数 (20kHz) を境に二つの領域に分ける。この二つの領域から、上部/下部領域の幅、上部/下部領域の高さ、上部/下部領域の面積、上部と下部領域の幅の比、上部と下部領域の高さの比、上部と下部領域の面積比の計 9 種類の特徴量を抽出する。しかし、2 値化時に下部領域に反応が見られない場合、0 が除数になり比が計算できない。したがって、その場合には 1 を除数とした。

上記の特徴量抽出により、得られた計 19 種類の特徴量を利用して機械学習を行う。認識器はランダムフォレストを利用した。

4. 実装

4.1 マイクカバー

3.2 節で述べたカバーデザインに従って、マイクカバーを実装した。実装したカバーを図 7 に示す。カバー全体の厚みは 1.5mm としたが、実際にマイクを遮蔽する部分は超音波を十分遮断するように 3mm の厚さとした。使用したデバイスは、スマートフォン (Huawei P9 Lite)、スマートウォッチ (Asus ZenWatch 2³, Huawei Watch⁴)、ネックスピーカ (JBL SoundGear) である。図 7 上に示すように、P9 Lite はデバイス底面の右側にスピーカを、デバイス底面左側と上面に二つのマイクを内蔵している。本研究では、マイク単体での超音波ジェスチャ認識に着目したため、底面のマイクだけを利用した。ZenWatch 2 は、左側面上部にスピーカを、左側面下部にマイクを内蔵している。Huawei Watch のマイクとスピーカはデバイス背面に位置している。しかし、Huawei Watch の内蔵マイクでは超音波領域の音を取得できないことが確認できたため、本研究では追加のマイク (Audio-Technica AT9931PC) で代用した。図 8 に外付けマイク用に実装したマイクカバーを示す。directivity カバーと同様の効果を得るため、マイクに対して 45 度の角度をつけた管状のカバーを作製した。SoundGear のスピーカはデバイスの両辺に、マイクはデバイス先端に位置している。しかし、SoundGear のマイクも超音波を取得できないことが確認できたため、図 8 に示す追加のマイクとマイクカバーで代用した。

マイクカバー作製に利用した 3D プリンタは PP3DP UP Plus 2 である。3D プリントの素材には ABS 樹脂を、3D モデルの作成には Autodesk 社の Fusion 360 を利用した。

4.2 ソフトウェア

スマートフォン/スマートウォッチ用に、20kHz の超音波の再生と録音を同時に可能なソフトウェアを作成した。記録された音声ファイルは内蔵ストレージに保存される。4.1 節で述べたように、Huawei Watch と SoundGear は内蔵マイクで超音波が録音できなかったため、外付けのマイクを装着した。外付けのマイクは PC (Apple MacBook Pro) に接続され、マイクから得られたデータは Adobe Audition によって記録した。取得したデータの解析には Python 3 を利用した。

5. 評価実験

5.1 実験環境

提案手法を評価するための実験を行なった。本研究では、ユーザがモバイル/ウェアラブルデバイス进行操作したいが、手が汚れて/濡れており画面をタッチできない状況や、手袋を装着しており画面にタッチしても反応しない状況を想定している。これらの状況は料理中にレシピを見るときや、機械のメンテナンス中にマニュアルを見るとき、冬の屋外で音楽アプリケーションを操作するときなどが考えられる。これらの状況を考慮し、評価するジェスチャとして図 9 に示す 6 種類のジェスチャを選定した。マイクの正面を横切るジェスチャ (gesture 1, 2)、マイクに近づく/遠ざかるジェスチャ (gesture 3, 4)、マイク正面で時計回り/反時計回りを描くジェスチャ (gesture 5, 6) の 6 種類である。

評価したデバイスは、4.1 節で述べた Huawei P9 Lite (D1)、Asus ZenWatch 2 (D2)、Huawei Watch (D3)、JBL SoundGear (D4) の 4 つである。被験者は 22 歳から 28 歳の男性 8 名であり、全員右利きであったためジェスチャは右手で行われた。ジェスチャの速度はできるだけ一定に保つよう、被験者に伝えた上で実験を行なった。被験者はモニタが設置された机の前に座り、モニタに表示される指示にしたがってジェスチャを行なった。各ジェスチャはランダムに 20 回表示され、カバー無し/ありの状態でそれぞれ合計 120 個のデータを取得した。以上より、本実験では合計で 7,680 個のデータを収集した (6 ジェスチャ × 20 回 × 8 被験者 × 4 デバイス × 2 状態 (カバー無し/あり))。収集したデータから、各被験者の各デバイスにおいて 10 分割交差検証で評価を行なった。つまり、収集した 120 個のデータをクラスの偏りが無いように 10 分割する。一つの分割をテストデータとし、残りで学習を行い、評価する。この操作を 10 回行い、得られた認識精度の平均を求めた。図 1 に示すように、評価実験

3: https://www.asus.com/us/ZenWatch/ASUS_ZenWatch_2-WI501Q/

4: <https://consumer.huawei.com/us/wearables/watch/>



図7 実装したマイクカバー
Fig. 7 Implemented cover.



図8 外付けマイクのカバー
Fig. 8 Implemented cover of external microphone.

中はスマートフォンは机の上に、スマートウォッチは左手首に、ネックスピーカは首にかけられた状態で実験を行なった。

5.2 結果

表2に評価実験の結果を示す。この表に示すように、マイクカバーを装着することで、全てのデバイスにおいてジェスチャ認識の精度が向上していることが確認できる。提案手法による改善は、P9 Liteで73.1%から87.1%，ZenWatch 2で68.9%から74.9%，Huawei Watchで45.3%から64.3%，SoundGearで54.3%から76.6%であり、全体の認識精度は平均で60.4%から75.7%に改善した。

図10に全てのデバイスにおけるマイクカバーあり/

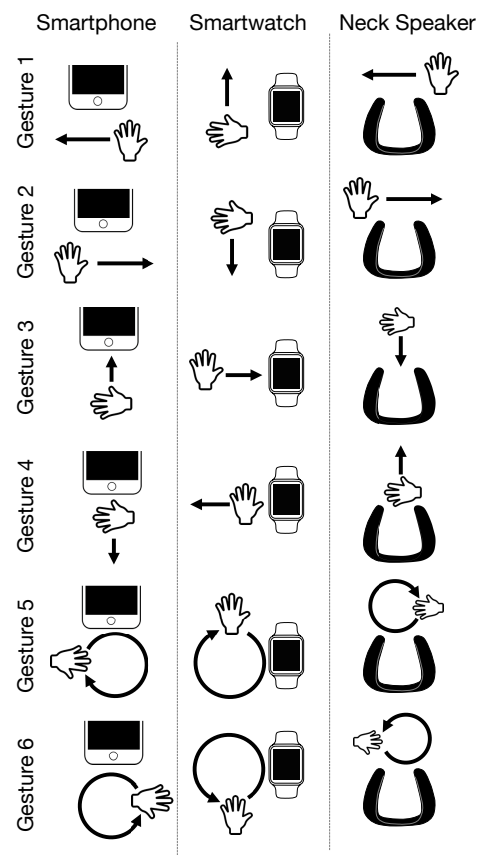


図9 評価したジェスチャ
Fig. 9 Evaluated gestures.

部分遮蔽されたマイクを用いた超音波によるジェスチャ認識手法

表 2 カバー無し/ありの場合の認識精度 [%]

Table 2 Results of evaluation without and with a cover [%].

P ¹	D1		D2		D3		D4	
	w/o	w/	w/o	w/	w/o	w/	w/o	w/
A	77.5	80.8	73.3	76.7	55.0	87.5	83.3	86.7
B	71.7	75.0	66.7	71.7	40.0	54.2	66.7	69.2
C	66.7	91.7	75.8	80.8	46.7	68.3	28.3	85.8
D	70.8	89.2	68.3	76.7	47.5	60.0	69.2	77.5
E	74.2	83.3	73.3	80.8	52.5	56.7	35.0	75.8
F	61.7	89.2	67.5	69.2	38.3	65.0	57.5	68.3
G	80.8	95.8	52.5	60.0	42.5	66.7	21.7	72.5
H	81.7	91.7	73.3	83.3	40.0	55.8	72.5	76.7
Ave.	73.1	87.1	68.9	74.9	45.3	64.3	54.3	76.6

¹Participants

無しの場合の混同行列を示す。この図より、マイクカバーを用いることで認識のばらつきが全体的に減少していることが確認できる。マイクカバーを装着することで、gesture 1 で 50.8% から 65.5%，gesture 2 で 50.8% から 68.8%，gesture 3 で 68.0% から 83.8%，gesture 4 で 70.8% から 83.4%，gesture 5 で 63.6% から 77.3%，gesture 6 で 58.4% から 74.8% への認識精度の改善が確認できた。また、gesture 2 と 6 に特に大きな改善が確認できた。これは、ジェスチャから得られるドップラー効果が、マイクカバー無しでは gesture 2 は gesture 1 に、gesture 6 は gesture 5 に似ていたからだと考えられる。したがって、今回評価したジェスチャにおいて、マイクカバーによってジェスチャ間で得られる特徴に差が生じ、認識精度の向上につながったと考える。

6. 考察

6.1 Hole カバーの特性

予備実験で用いた hole カバーが想定と異なる効果をもたらした。カバーの構造によっては、音の干渉等によって振幅を強める構造になっていた可能性がある。本研究では directivity カバーによって片側からの音量を減少させ、もう片側からの音量を増やす設計を採用したが、それ以外の方法でも取得音量を変化させる構造が実現可能であるのか、今後の調査が必要である。

6.2 ジェスチャ速度

本研究では、ユーザのジェスチャ速度を一定と想定した。提案手法はドップラー効果の変化量を認識に利用しているため、ジェスチャ中に速度が大幅に変化すると誤認識する可能性がある。例えば、左スワイプの際にマイク正面まではゆっくりと近づき、マイクを通り過ぎてから大幅に速度をあげると、右スワイプと誤認識する可能性がある。評価実験の際にははできるだけ一定の速度でジェスチャを行うことを被験者に伝えたが、実際には一定の速度を保つことは難しいと考えられる。しかし、その環境においても評価実験ではマイ

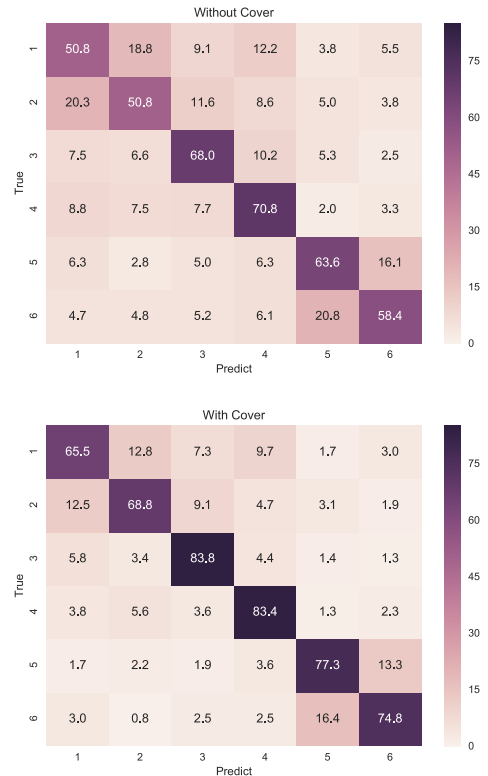


図 10 カバー無し/ありの場合の混同行列 [%]

Fig. 10 Comparison of confusion matrices without and with a cover [%].

クカバーによってジェスチャ認識に改善が見られたため、提案手法は有効と考える。

6.3 内蔵マイク/スピーカの性能

本研究で用いた Huawei Watch と SoundGear の内蔵マイクでは超音波が取得できないことが確認できたため、超音波を取得するのに外付けのマイクを追加する必要があった。外付けのマイクを装着することで、提案手法により認識精度が向上することが確認できたため、超音波を取得できる内蔵マイクであれば認識精度が向上することが確認できた。今後の機器の発展により内蔵マイクの性能が向上すれば、デバイス内蔵のマイクとスピーカのみで提案手法を実装できると考える。また、スマートウォッチの内蔵スピーカでは、超音波と共に可聴音のノイズも同時に発信していることが確認できた。提案手法は、超音波を利用しているため認識に悪影響は及ぼさないが、システム利用時に可聴音のノイズが発生するのはユーザにとっては不快となりうる。一方、スマートフォンとネックスピーカは超音波のみを発信し、可聴音のノイズは確認できなかった。したがって、今後の機器の発展によりスマートウォッチのノイズの問題は解消されると考える。

6.4 センシング領域の限定

本研究の提案手法では、ジェスチャが認識できる領域が限定される。例えば、図9に示すように、スマートフォンの下側やスマートウォッチの左側、ネックスピーカの正面などに限られる。提案手法では、マイクカバーの左右での感度の違いを利用しているため、これらの領域の外でのジェスチャを認識するのは難しい。領域を限定せずにジェスチャを認識するには、手の位置を認識する必要があるが、TDoAのような位置認識手法を用いるには、複数のマイクを利用する必要がある。多くのウェアラブルデバイスは複数のマイクを内蔵していないため、手の位置認識を行うのは難しい。本研究の想定環境では、センシング領域が限定されることは避けられないと考える。

6.5 周囲環境の影響

本稿では、マイクカバーを装着することで認識精度が向上するののかという点に着目したため、周囲の人の動きや環境音によるジェスチャの誤検出については考慮していない。また、同一のデータセット内における評価のため、環境の異なる未知のデータセットに対する評価は考慮できていない。周囲の異なる環境や、異なるデータセットに対してマイクカバーがどのように影響するのか今後の調査が必要である。

6.6 マイクカバー装着による影響

マイクカバーを装着することで、電話や音声認識システム、音声メモなどの通常のマイクの利用に影響を与える可能性がある。そこで、マイクカバーのあり/無しによってマイクの周波数特性がどのように変化するのかを調査した。スピーカから20Hzから22.05kHzへと190秒間で遷移するスイープ信号を発信し、デバイス内蔵のマイクで音を取得した。スピーカにはFostexのFT200Dを用いた。スピーカとマイクの距離は5cmとし、カバー無しの場合とありの場合に取得したスイープ信号をフーリエ変換し、周波数特性を比較した。結果を図11に示す。マイクカバーを装着することで、どのマイクにおいても5kHzから20kHz付近に周波数特性の変化が確認できた。しかし、日常的に用いられる人の音声の周波数は500Hzから4kHz程度であるため、マイクカバーによって従来の音声を利用するシステムの妨げにはならないと考えられる。

7. おわりに

本研究では、マイクカバーを装着しマイクの音響特性を変化させることで、超音波によるジェスチャ認識を改善する手法を提案した。提案手法では、3Dプリントされたマイクカバーとデバイス内蔵のマイクとスピーカのみを利用するため、ジェスチャ認識改善のために電力を消費する追加のデバイスが必要ない。4種類のマイクカバーを提案し、予備実験より directivity

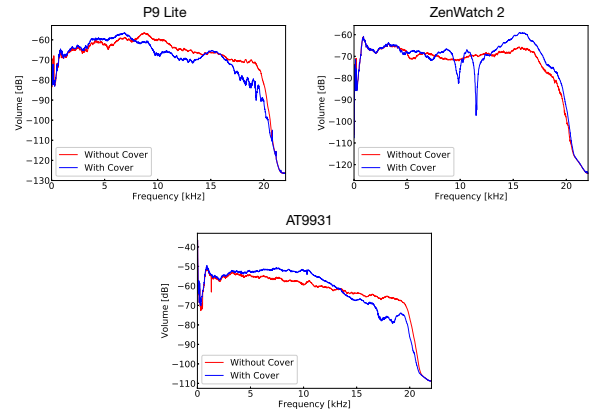


図11 マイクカバー無し/ありの場合のマイク特性の比較

Fig. 11 Comparison of microphone characteristics without and with a cover.

カバーが最適であることを確認した。8人の被験者、6種類のジェスチャに対して4種類のウェアラブルデバイスで評価実験を行なった結果、提案手法によって平均で60.4%から75.7%への認識精度の向上が確認できた。

謝辞

本研究はJSPS 科研費 JP18K18084の助成を受けたものである。

参考文献

- [1] Butler, A., Izadi, S., Hodges, S.: SideSight: Multi-“Touch” Interaction Around Small Devices; Proceedings of the 21st Annual ACM Symposium on User Interface Software and Technology (UIST '08), pp. 201–204 (2008).
- [2] Kratz, S., Rohs, M.: HoverFlow: Expanding the Design Space of Around-device Interaction; Proceedings of the 11th International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI '09), pp. 4:1–4:8 (2009).
- [3] Withana, A., Peiris, R., Samarasekara, N., Nanayakkara, S.: zSense: Enabling Shallow Depth Gesture Recognition for Greater Input Expressivity on Smart Wearables; Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15), pp. 3661–3670 (2015).
- [4] Song, J., Sörös, G., Pece, F., Fanello, S. R., Izadi, S., Keskin, C., Hilliges, O.: In-air Gestures Around Unmodified Mobile Devices; Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology (UIST '14), pp. 319–329 (2014).
- [5] Kim, D., Hilliges, O., Izadi, S., Butler, A. D., Chen, J., Oikonomidis, I., Olivier, P.: Digits: Free-hand 3D Interactions Anywhere Using a Wrist-worn Gloveless Sensor; Proceedings of the 25th Annual ACM Symposium on User Interface Software and Technology (UIST '12), pp. 167–176 (2012).
- [6] Yang, X.-D., Hasan, K., Bruce, N., Irani, P.:

- Surround-see: Enabling Peripheral Vision on Smartphones During Active Use; Proceedings of the 26th Annual ACM Symposium on User Interface Software and Technology (UIST '13), pp. 291–300 (2013).
- [7] Gupta, S., Morris, D., Patel, S., Tan, D.: Sound-Wave: Using the Doppler Effect to Sense Gestures; Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '12), pp. 1911–1914 (2012).
- [8] Nandakumar, R., Iyer, V., Tan, D., Gollakota, S.: FingerIO: Using Active Sonar for Fine-Grained Finger Tracking; Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16), pp. 1515–1525 (2016).
- [9] Ruan, W., Sheng, Q. Z., Yang, L., Gu, T., Xu, P., Shangguan, L.: AudioGest: Enabling Fine-grained Hand Gesture Detection by Decoding Echo Signal; Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '16), pp. 474–485 (2016).
- [10] Manabe, H.: Multi-touch Gesture Recognition by Single Photoreflector; Proceedings of the Adjunct Publication of the 26th Annual ACM Symposium on User Interface Software and Technology (UIST '13), pp. 15–16 (2013).
- [11] Laput, G., Brockmeyer, E., Hudson, S. E., Harrison, C.: Acoustuments: Passive, Acoustically-Driven, Interactive Controls for Handheld Devices; Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15), pp. 2161–2170 (2015).
- [12] Watanabe, H., Terada, T.: Improving Ultrasound-based Gesture Recognition Using a Partially Shielded Single Microphone; Proceedings of the 2018 ACM International Symposium on Wearable Computers (ISWC '18), pp. 9–16 (2018).
- [13] Harrison, C., Tan, D., Morris, D.: Skinput: Appropriating the Body As an Input Surface; Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '10), pp. 453–462 (2010).
- [14] Ishikawa, T., Horry, Y., Hoshino, T.: Touchless input device and gesture commands; Proceedings of the IEEE International Conference on Consumer Electronics (ICCE '05), pp. 205–206 (2005).
- [15] Laput, G., Xiao, R., Chen, X. A., Hudson, S. E., Harrison, C.: Skin Buttons: Cheap, Small, Low-powered and Clickable Fixed-icon Laser Projectors; Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology, UIST '14, pp. 389–394 (2014).
- [16] Ogata, M., Sugiura, Y., Makino, Y., Inami, M., Imai, M.: SenSkin: Adapting Skin As a Soft Interface; Proceedings of the 26th Annual ACM Symposium on User Interface Software and Technology (UIST '13), pp. 539–544 (2013).
- [17] Goel, M., Zhao, C., Vinisha, R. and Patel, S. N.: Tongue-in-Cheek: Using Wireless Signals to Enable Non-Intrusive and Flexible Facial Gestures Detection; Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15), pp. 255–258 (2015).
- [18] Reyes, G., Zhang, D., Ghosh, S., Shah, P., Wu, J., Parnami, A., Bercik, B., Starner, T., Abowd, G. D., Edwards, W. K.: Whoosh: Non-voice Acoustics for Low-cost, Hands-free, and Rapid Input on Smartwatches, Proceedings of the 2016 ACM International Symposium on Wearable Computers (ISWC '16), pp. 120–127 (2016).
- [19] Apple: Siri; available from <https://www.apple.com/ios/siri/> (accessed 2019-6-4).
- [20] Google: Google Assistant; available from <https://assistant.google.com> (accessed 2019-6-4).
- [21] Aumi, M. T. I., Gupta, S., Goel, M., Larson, E., Patel, S.: DopLink: Using the Doppler Effect for Multi-device Interaction; Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '13), pp. 583–586 (2013).
- [22] Goel, M., Lee, B., Islam Aumi, M. T., Patel, S., Borriello, G., Hibino, S., Begole, B.: SurfaceLink: Using Inertial and Acoustic Sensing to Enable Multi-device Interaction on a Surface; Proceedings of the 32Nd Annual ACM Conference on Human Factors in Computing Systems (CHI '14), pp. 1387–1396 (2014).
- [23] Chen, K.-Y., Ashbrook, D., Goel, M., Lee, S.-H., Patel, S.: AirLink: Sharing Files Between Multiple Devices Using In-air Gestures; Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '14), pp. 565–569 (2014).
- [24] Chirp Microsystems: Chirp Technology; available from <http://www.chirpmicro.com/technology.html> (accessed 2019-6-4).
- [25] Ono, M., Shizuki, B., 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), pp. 31–40 (2013).
- [26] Yun, S., Chen, Y.-C., Zheng, H., Qiu, L., Mao, W.: Strata: Fine-Grained Acoustic-based Device-Free Tracking; Proceedings of the 15th Annual International Conference on Mobile Systems, Applications, and Services (MobiSys '17), pp. 15–28 (2017).
- [27] Zhang, C., Xue, Q., Waghmare, A., Jain, S., Pu, Y., Hersek, S., Lyons, K., Cunefare, K. A., Inan, O. T., Abowd, G. D.: SoundTrak: Continuous 3D Tracking of a Finger Using Active Acoustics; Proc. ACM Interact. Mob. Wearable Ubiquitous Technol., Vol. 1, No. 2, pp. 30:1–30:25 (2017).
- [28] Memoli, G., Chisari, L., Eccles, J. P., Caleap, M., Drinkwater, B. W., Subramanian, S.: VARI-SOUND: A Varifocal Lens for Sound; Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19), pp. 483:1–483:14
- [29] 高橋義典, 小林真萌, 田中 潤: 音響レンズによる遅延和を利用した距離選択收音; 日本音響学会誌, Vol. 72, No. 12, pp. 761–763 (2016).

(2019 年 7 月 3 日受付, 12 月 2 日再受付)

著者紹介

渡邊 拓貴



1989 年生。2012 年神戸大学工学部電気電子工学科卒業。2014 年同大学大学院工学研究科博士前期課程修了。2017 年同大学大学院工学研究科博士後期課程修了。同年より北海道大学大学院情報科学研究院助教、現在に至る。博士(工学)。ウェアラブルコンピューティング、ユビキタスコンピューティングの研究に従事。ACM、情報処理学会の各会員

寺田 努 (正会員)



1974 年生。1997 年大阪大学工学部情報システム工学科卒業。1999 年同大学大学院工学研究科博士前期課程修了。2000 年同大学院工学研究科博士後期課程退学。同年より大阪大学サイバーメディアセンター助手。2005 年より同講師。2007 年神戸大学大学院工学研究科准教授。2018 年同大学院工学研究科教授、現在に至る。博士(工学)。ウェアラブルコンピューティング、ユビキタスコンピューティングの研究に従事。ACM、IEEE、ヒューマンインタフェース学会等、7 学会の会員。