

# ヒアラブルへの入力を想定したユーザ定義型ジェスチャ調査とIMUセンサによる耳タッチジェスチャの認識

佐藤 優希菜<sup>\*1,a)</sup> 雨坂 宇宙<sup>\*1</sup> 山本 匠<sup>1</sup> 渡邊 拓貴<sup>2</sup> 杉浦 裕太<sup>1</sup>

**概要:** 近年、ライフログや音声サービスなどの様々な機能を搭載したイヤホン型デバイスである、ヒアラブルの研究が多く行われている。一方、既存のヒアラブル単体での操作手法では限られたコマンドしか指定できず、デバイスの操作手法の拡張が必要である。本研究では、ヒアラブルの多くに内蔵されているIMUを用いたハンズジェスチャ入力手法を提案する。まず、ジェスチャ認識実験にあたって、ユーザ定義型ジェスチャを調査する Elicitation study を行い、3つのインタラクション領域に対して2つのデバイスタイプの計6つの条件によるジェスチャ定義の傾向を調査した。ユーザ定義型ジェスチャのうち耳を使ったジェスチャの認識率の調査を行った結果、インイヤ型デバイスは91.6%、耳掛け型デバイスは76.6%であった。

## 1. はじめに

近年、イヤホン型のウェアラブルデバイス（ヒアラブル）がさらに注目されている。ヒアラブルとは従来の音楽鑑賞や通話などの機能だけでなく、様々な機能が付加された高機能なイヤホン型デバイスを指す。多くの研究や開発が進められており、今後はより多くのアプリケーションがヒアラブルに搭載されると予想される。一方、市販されているヒアラブルの多くはスマートフォンと連携して使用することが想定されているため、ヒアラブルの操作にはスマートフォンの画面をタッチする必要がある、ユーザビリティを損う。そのため、ヒアラブルのみを用いたデバイス操作手法が求められている。

ヒアラブルのための新たな操作手法は、主に手を使わない入力（ハンズフリー入力）と手を使って行う入力（ハンズ入力）にわかれる。ハンズフリー入力手法は、音声入力手法や慣性/気圧/筋電/赤外線/音響センシング [1], [2], [3], [4], [5] を用いた頭部ジェスチャ手法などがこれまで研究されてきた。音声入力の利用は公共の場での発声や騒音環境下での認識精度の低下などの課題がある。頭部ジェスチャの利用は、頭を振るなどの大きな頭部の動きを伴うジェスチャは周囲に与える印象の懸念や視線が外れてしまう危険性の課題がある。一方、顎をずらすなどの小さな頭部の動きで実現できるジェスチャは日常生活で行う動作と混同する課題がある。

一方、ハンズ入力手法 [6], [7], [8], [9] は触覚フィードバックによるデバイス操作を可能にし、ユーザーフレンドリーな入力方法である。Chen ら [10] は耳周辺をインタラクション領域として Gesture Elicitation Study を行い、ユーザ定義型ジェスチャを明らかにした。また、Rateau ら [11] はスマートウォッチとイヤホンを装着した状態のユーザ定義型ジェスチャを明らかにした。しかし、定義可能なジェスチャ種類に制限はなく、ヒアラブルのみを対象とした調査も行われていない。これらの研究には以下の2点の特徴がある。

**1. ユーザ定義型ジェスチャの一部が空中ジェスチャである:** 空中ジェスチャの認識は、Tamaki ら [12] や Metzger ら [13] によって提案されているが、カメラや外向き赤外線センサの搭載が必要である。したがって、空中ジェスチャの認識システムが社会実装されるまでには時間を要すると考えられる。一方で、タッチベースのジェスチャについては Xu ら [6] が、デバイスに内蔵されたマイクを使用して、耳元でタッチした際のジェスチャの音を記録し認識するシステムを提案し、市販製品での実験を行った。また、市販製品の一部では耳珠周辺のタップ動作時に発生する振動を加速度センサで測定することにより、ジェスチャ入力として利用できる製品 [14] もある。これらの研究や製品に示されているように、現在利用可能なジェスチャは限られており、今後の技術進展によって利用可能なジェスチャやインタラクション領域に制約が与えられた場合のユーザ定義型ジェスチャを調査することで、様々な入力技術に合わせた最適なジェスチャセットを明らかにする必要がある。

<sup>1</sup> 慶應義塾大学

<sup>2</sup> 北海道大学

<sup>a)</sup> sato-yukina@keio.jp

\*These authors contributed equally to this work.

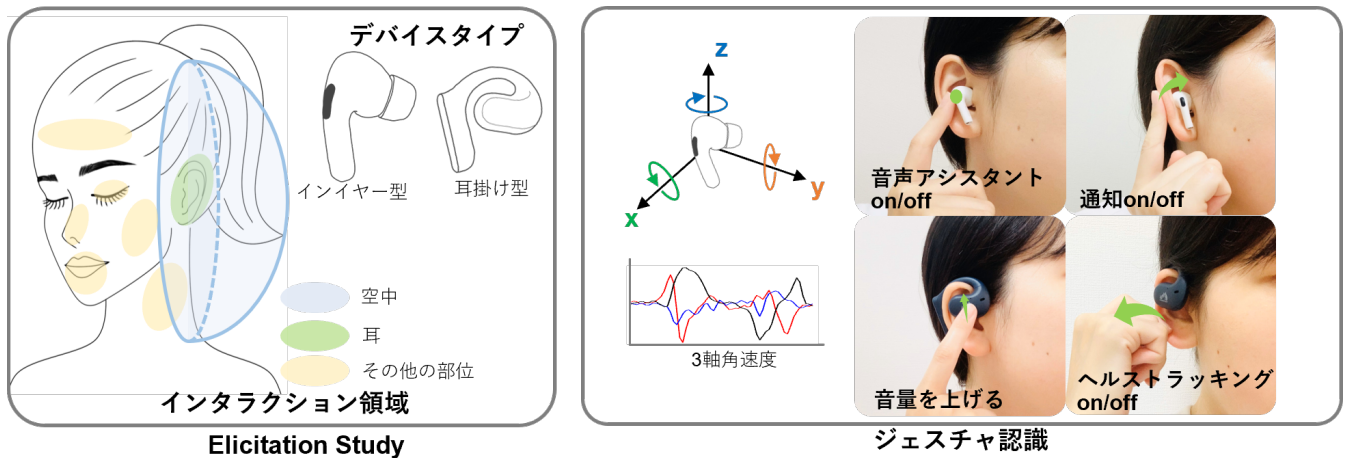


図 1 研究概要：3つのインタラクション領域と2つのデバイスタイプの計6条件でのユーザ定義型ジェスチャを調査する Elicitation study を行った。ユーザ定義型ジェスチャのうち耳を使ったジェスチャの認識率調査を行った。

2. ヒアラブルのみを利用したユーザ定義型ジェスチャが明らかにされていない：本研究は、ヒアラブルの操作をデバイス自身で行う利用環境を想定している。そのため、ユーザがヒアラブルのみを装着した状態のユーザ定義型ジェスチャを調査する必要がある。また、昨今のヒアラブルは形状や大きさが異なるデバイスが複数存在し、タッチできる箇所や面積が異なる。これらの差異がユーザ定義型ジェスチャにどのような影響を与えるかは明らかにされていない。

本研究では、ヒアラブルをセンシングデバイスとする場合のジェスチャ入力に着目し、インタフェースの制約やデバイスタイプの差異によってユーザが定義するジェスチャがどのように変化するか調査した。さらに、ヒアラブルと親和性の高いIMU(Inertial Measurement Unit)を用いてこれらのジェスチャを認識する手法を提案し、ジェスチャ認識実験を行った。IMUはライフログや空間オーディオ再現のために市販製品の多くに搭載されている。また、ヒアラブルの新機能を実現するための研究にもIMUが利用されている[15], [16], [17], [18]。したがって、IMUを用いたハンズジェスチャ認識は、IMUを有効活用している点で、筐体サイズが限られているヒアラブルの大きなメリットといえる。図1に本研究の概要を示す。本研究では3つのインタラクション領域に対して2つのデバイスタイプの計6条件にて Elicitation study を行いユーザ定義型ジェスチャを明らかにした。その後、ユーザ定義型ジェスチャのうち耳に触れるジェスチャ(耳タッチジェスチャ)をIMUにて認識するシステムを提案し、その認識性能を調査した。

## 2. 関連研究

### 2.1 ユーザ定義型ジェスチャの調査

ジェスチャ認識実験においてジェスチャセットを決定するプロセスは2つある。実験者によって設計したジェス

チャセットを利用する場合と、Elicitation Studyによって考案されたユーザ定義型ジェスチャを基にジェスチャセットを決定する場合である。Morrisら[19]は、HCI研究者によって定義されたジェスチャセットよりも、Elicitation studyによって定義されたジェスチャセットの方が直感的であることを明らかにしている[28]。ジェスチャ定義の研究は、顔[20]、足[21]、手[22], [23]、頭[24]、皮膚[25]など、さまざまな身体の部位を使用して実施されている。また、テレビ[26]、モバイル機器[27]、ヘッドマウントディスプレイ[28]、スマートウォッチ[29], [30]、スマートリング[29]、帽子[31]、マスク[32]など、様々なデバイスを使った研究も行われている。

耳に関する Elicitation study では、Chenら[10]が耳周辺のインタラクションに関する調査を行い、ユーザ定義型ジェスチャの調査と深い考察が行われている。一方で、実験参加者はジェスチャのインタラクション領域に制約がなく、イヤホン型デバイスを装着していなかった。これは純粋な耳へのインタラクションを調査する点では有効である。しかし、ヒアラブル向けのジェスチャ認識技術は開発中であり、利用可能なインタラクション領域は段階的に増えていくことが予想される。また、ヒアラブルへの入力を想定する場合、デバイスを装着した状態での調査が望ましい。Rateauら[11]は、ヒアラブルとスマートウォッチを同時に装着した状態での Elicitation study を行っているが、同様にインタラクション領域の制約が無い。また、デバイスがヒアラブルのみに設定されていない。

本研究では、ヒアラブルのみを装着した状態のユーザ定義型ジェスチャを調査した。また、3つのインタラクション領域の制約に対して2つのデバイスタイプの6条件において、ユーザ定義型ジェスチャがどのように変化するかを調査した。

## 2.2 ヒアラブル向けハンズジェスチャ入力

ハンズ入力に関する研究では、真鍋ら [7] は市販されているヘッドホンに簡単な回路を組み合わせてヘッドホンのタップ動作を認識できることを示した。Roman ら [9] は静電容量式センサを搭載した、耳掛け型デバイスによるタッチ入力システムを提案した。菊池ら [8] は反射型光センサをイヤホン背部に装着することで耳を引っ張るジェスチャを認識した。Xiu ら [6] はデバイス内蔵マイクを用いて頬や耳上で行うスワイプジェスチャ時に発生する音を取得することで、タッチジェスチャの認識に成功した。市販製品では、SONY の LinkBuds [14] が耳珠周辺のタップ時の振動を加速度センサを用いて認識し、デバイスへの入力に利用することができる。

空中ジェスチャでは、Metzger ら [13] が、近接センサを搭載した空中ジェスチャ認識システムを提案した。Tamaki ら [12] はカメラを搭載したイヤホン型デバイスでの空中ジェスチャ入力システムを提案した。これらの研究では、追加のセンサを必要とするため、市販のヒアラブルで利用可能ではないという制約がある。マイクによるジェスチャ認識は市販デバイスに実装されている [6] が、マスク装着時に利用できない制約や、騒音環境下でのジェスチャ抽出率や認識率の低下という課題がある。

本研究では、IMU によるジェスチャ入力システムを提案する。IMU は、市販されているヒアラブルにすでに内蔵されているため、実装コストの面で大きな制約はないと考える。騒音環境での認識率低下が生じないなどのメリットもある。さらに、既存研究に対する優位性だけでなく、IMU とマイクを用いたマルチモーダルなセンシングなどの協調的な利用も考えられる。このようなユースケースの可能性を明らかにするために、まず IMU 単体によるジェスチャ認識精度を調査する必要がある。そこで本研究では、ユーザ定義型ジェスチャを 2 種類のデバイスで調査することで、ジェスチャやデバイスごとの認識精度を調査した。

## 3. ユーザ定義型ジェスチャの設計

本実験の目的は、以下の 3 つの問いに答えることである。

- ユーザはどのようなジェスチャをヒアラブルの入力として好むのか
- インタフェースの制約やデバイスタイプの差異がユーザ定義型ジェスチャにどのように影響するか
- 既存のヒアラブルや耳に関する Elicitation study との何の共通点や相違点があるのか

本実験では、3 種類のインタフェース条件ごとにそれぞれ 2 種類のデバイスタイプの合計 6 条件に対して、同じタスク群のジェスチャ定義を行った。インタフェース条件は、これまでのヒアラブル向けジェスチャ入力手法を参考に、制限なし条件・タッチ条件・耳タッチ条件を設定した。制

表 1 タスク一覧

タスクグループ	タスク
画面操作	上下左右にスクロール ズームイン、アウト 最大化 / 最小化 ホーム画面に行く 次のアプリ、前のアプリ 進む、戻る
	再生 / 停止 音量を上げる、下げる 次の曲、前の曲
電話	電話を取る / 切る 電話を無視する 電話をかける マイク on / off スピーカ on / off
モード切替	音声アシスタント ボイスメモ カレンダーアプリの予定確認 ヘルストラッキング機能 通知

限なし条件では、空中ジェスチャやタッチジェスチャなど自由にジェスチャを定義できる。タッチ条件は顔周辺のいずれかに指先が触れることを条件としている。耳タッチ条件は耳に触れ、ヒアラブルが変位することを条件としている。また、いずれの条件もデバイスへのタッチジェスチャを定義可能としている。デバイスタイプは、一般的なデバイスタイプであるインイヤ型と耳掛け型の 2 種類を選択した。

### 3.1 実験概要

19 名の実験参加者（男性 11 名、女性 8 名、平均年齢：25.8 歳）を募集し調査を行った。参加者全員が右利きで、16 名が週に 1 回以上イヤホンを利用していた。そのうち 9 名は、イヤホンをタップするなどの操作経験があり、2 名は日常的にイヤホンでの操作を使用していた。実験の所要時間は 1~2 時間で、謝金として 3,000 円支払った。実験の実施には、慶應義塾大学理工学部の倫理委員会の許可（受付番号：2023-086）を得た。

タスクの一覧を表 1 に示す。関連研究 [10] を参考に画面操作、音楽再生操作、電話のタスクグループを決定した。さらに、ヒアラブルでの利用が想定される、音声アシスタントの起動やボイスメモの起動、音声によるカレンダーの確認、ヘルストラッキング機能の起動、通知のモード切替を行うグループを追加し、合計 32 種類とした。画面操作のタスクはスマートグラスやスマートウォッチなどのデバイスを操作することを想定しており、それ以外はすべてヒアラブルで完結するタスクである。まず、参加者に各タスク

の機能をより良く理解してもらうために、画面操作、音楽再生操作、電話のタスクを行った際の動作画面を録画した動画を観せ、タスクの具体的な動作結果を説明した。モード切替のタスクグループは口頭による説明を行った。結果として得られるジェスチャセットを社会的なシナリオにふさわしいものにするために、研究室内のオープンで快適な環境で実験を行った。

次に、ジェスチャを定義する際のルールを説明した。ジェスチャ定義のルールは、以下の通りである。

- 同一のタスクグループ内に同じジェスチャは割り当て不可
- スラッシュで区切られたジェスチャ同士（例：画面最大化/最小化）は状態の切替えタスクであるため、同じジェスチャを割り当て可能
- 実験中はいつでもジェスチャを変更可能
- 同じジェスチャを右手と左手で異なるタスクに割り当て可能
- 両手で行うジェスチャも定義可能

ルールの説明の後に、ジェスチャを定義する3つのインタフェース条件について説明した。実験は、制限なし条件、タッチ条件、耳タッチ条件の順で行った。デバイスは、各条件でインイヤ型デバイス、耳掛け型デバイスの順に行った。インイヤ型デバイスには AirPods Pro(Apple)、耳掛け型デバイスには HA-NP35TBK(Victor) を使用した。

### 3.2 解析結果

定義されたジェスチャのインタフェース箇所を空中、デバイス本体、耳、顔・首・胸の4種類に分類した。各条件でのインタフェース箇所の割合を図2にまとめた。空中では、指先のピンチイン、ピンチアウトや、手を開く、閉じる、電話のサイン、手をかざすなどのジェスチャが定義された。デバイス本体において定義されたジェスチャ位置と割合を図3、耳、顔・首・胸において定義されたジェスチャ位置と割合を図4に示す。

#### 3.2.1 インタフェースの制約による傾向

図2に示すように、制限なし条件ではいずれのデバイスタイプでも、空中ジェスチャ(平均42.9%)とデバイスへのジェスチャ(平均38.9%)が定義される割合が高かったが、顔・首・胸を使ったジェスチャ(平均2.9%)を定義するユーザは非常に少なかった。一方、タッチ条件では顔・首・胸を使ったジェスチャの定義割合が20%以上まで増加した。これは、空中へ定義されたジェスチャの多くが、顔・首・胸の部位に割り当てられたためである。特に、頬や首のジェスチャが好まれることが分かった。

#### 3.2.2 デバイスタイプの差異

インイヤ型デバイスと比較して、耳掛け型デバイスはデバイスへのジェスチャの割合がいずれのインタフェース条件でも高かった。例えば、インイヤ型デバイスでは、

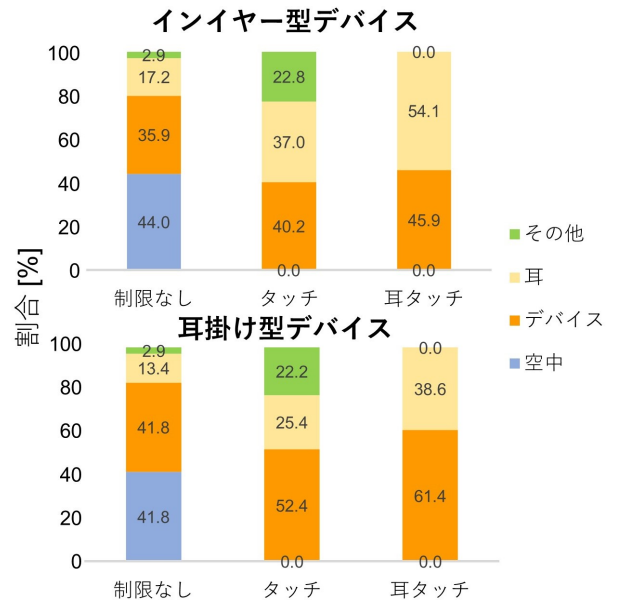


図2 ジェスチャ定義されたインタフェース領域の割合

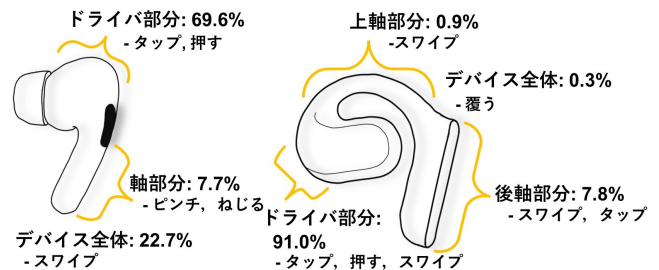


図3 デバイス上のジェスチャ位置と割合

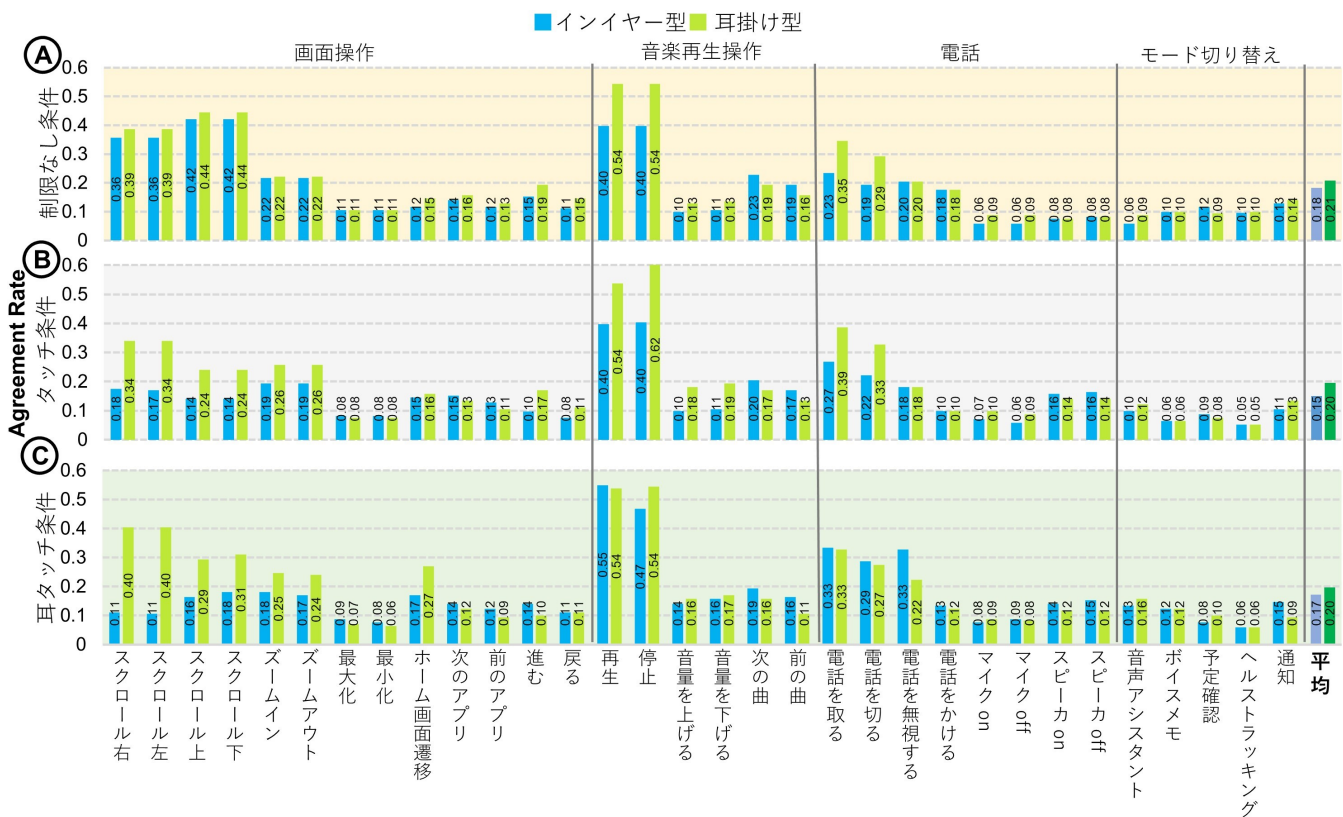
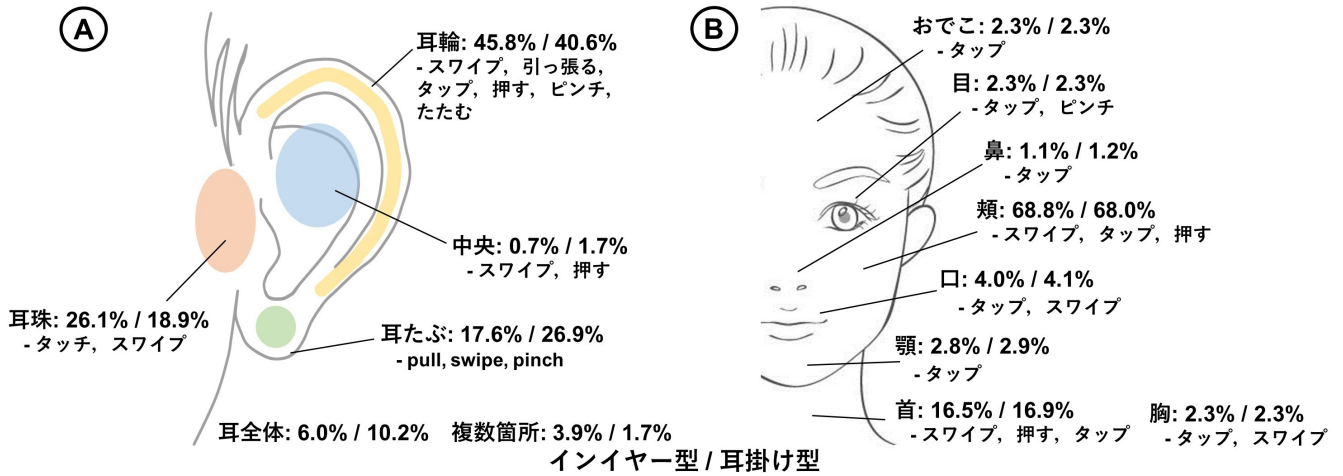
スワイプジェスチャを複数個所に設定するユーザはいなかったが、耳掛けでは、ドライバ部分、上部、背部のステム部分にそれぞれスワイプジェスチャを定義するユーザがいた。このように、耳掛け型デバイスの筐体サイズが大きいため、タッチ可能な部位が多くなることで、デバイス本体への定義割合が増加した。

#### 3.2.3 Agreement Rates

Vatavu と Webbrock らによって提案されている Agreement Rates [33] を算出した。Agreement Rates は以下の定義式によって算出される。

$$AR(r) = \frac{|P|}{|P|-1} \sum_{P_i \subseteq P} \left( \frac{|P_i|}{|P|} \right)^2 - \frac{1}{|P|-1}$$

$P$  は操作  $r$  に対する提案の集合、 $|P|$  は  $P$  の要素数、 $P_i$  は  $P$  から同じ提案についての部分集合である [10]。図5に、各条件におけるそれぞれのタスクに対する AR を示す。全条件の平均が最も高かったタスクは、音楽再生グループにおける停止タスクで、平均0.50であった。全条件の平均が最も低かったタスクは、アプリケーショングループのヘルストラッキングタスクで、平均は0.07であった。ARは非常に高い一致率 ( $AR > 0.5$ )、高い一致率 ( $0.5 > AR > 0.3$ )、中程度の一致率 ( $0.3 > AR > 0.1$ )、低い一致率 ( $AR < 0.1$ ) の



4つに分類される。各タスクのARのスコア分布割合は、2.7%(AR ≥ 0.5), 22.5%(0.5 > AR ≥ 0.3), 50.9%(0.3 > AR ≥ 0.1), 23.9%(AR < 0.1)であった。ARが0.1以上のタスクは全体の約80%であった。

### 3.3 ユーザ定義型ジェスチャ

ARが概ね0.1(76.1%)以上であったため、AR導出時に最も定義率が高かったジェスチャを各タスクのユーザ定義型ジェスチャとした。耳タッチ条件で定義されたユーザ定義型ジェスチャを表2に示す。太字になっているジェスチャはデバイスタイプの差異によって、定義されたジェス

チャが異なったものである。

#### 3.3.1 ユーザがヒアラブルの入力として好むジェスチャ

ARが最も高く、最も定義されたのはデバイスへのタップであった。これは、最も直感的かつ多くのヒアラブルでデバイスのタップ機能がすでに利用可能であるためと考えられる。また、対称的なタスクや方向を示すタスクは、動作も対称的や方向を示すジェスチャが定義された。この傾向はすべての条件において同じであった。

#### 3.3.2 インタフェースの制約やデバイスタイプの差異による影響

インタラクション領域の制約がユーザ定義型ジェスチャ

表 2 耳タッチ条件のユーザ定義型ジェスチャ

タスクグループ	タスク	インイヤー型デバイス	耳掛け型デバイス
画面操作	スクロール右/左	デバイスを前/後ろにスワイプ	デバイスを前/後ろにスワイプ
	スクロール上/下	デバイスを上/下にスワイプ	デバイスを上/下にスワイプ
	ズームイン	<b>耳輪上部と耳たぶをたたむ</b>	<b>デバイス上でピンチイン</b>
	ズームアウト	<b>耳輪をピンチアウト</b>	<b>デバイス上でピンチアウト</b>
	最大化	耳輪上部を上引っ張る	耳輪上部を上引っ張る
	最小化	耳たぶを下引っ張る	耳たぶを下引っ張る
	ホーム画面に行く	デバイスをタップ	デバイスをタップ
	次のアプリ	<b>耳輪を下にスワイプ</b>	<b>デバイスをダブルタップ</b>
	前のアプリ	<b>耳輪を上スワイプ</b>	<b>デバイスをダブルタップ (左耳)</b>
	進む	<b>耳珠を後ろにスワイプ</b>	<b>耳たぶを後ろに引っ張る</b>
戻る	<b>耳珠を前にスワイプ</b>	<b>耳たぶを前に引っ張る</b>	
音楽再生操作	再生 / 停止	デバイスをタップ	デバイスをタップ
	音量を上げる / 下げる	<b>耳輪を上/下にスワイプ</b>	<b>デバイスを上/下にスワイプ</b>
	次の曲	デバイスをダブルタップ	デバイスをダブルタップ
	前の曲	デバイスをダブルタップ (左耳)	デバイスをダブルタップ (左耳)
電話	電話を取る / 切る	デバイスをタップ	デバイスをタップ
	電話を無視する	デバイスを長押し	デバイスを長押し
	電話をかける	デバイスを後ろにスワイプ	デバイスを後ろにスワイプ
	マイク on / off	耳たぶを下引っ張る	耳たぶを下引っ張る
	スピーカ on / off	デバイスをダブルタップ (左耳)	デバイスをダブルタップ (左耳)
モード切り替え	音声アシスタント	デバイスを長押し	デバイスを長押し
	ボイスメモ	デバイスをタップ	デバイスをタップ
	カレンダーアプリで予定確認	耳輪を下にスワイプ	耳輪を下にスワイプ
	ヘルストラッキング通知	<b>耳珠を後ろにスワイプ</b> <b>耳全体を前にたたむ</b>	<b>耳輪中部を後ろに引っ張る</b> <b>耳たぶを下引っ張る</b>

に与える影響は、画面操作グループの上下左右にスクロールするタスクにおいて顕著であった。方向を示すタスクの平均 AR は各インタラクション条件で 0.42(制限なし条件), 0.18(タッチ条件), 0.21 (耳タッチ条件) とばらつきがあった。ほとんどの方向指示タスクは、手や指のスワイプを伴うジェスチャが定義されることが多かったため、広い空間で定義される傾向があった。そのため、制限なし条件では、最も自由度の高い空中ジェスチャが方向指示タスクに用いられると考えられる。しかし、空中ジェスチャが利用できない場合は、デバイス、耳 (主に耳珠と耳輪)、顔・首・胸にジェスチャの定義が分散され、AR が低下した。

デバイスタイプの差異による影響について説明する。耳掛け型デバイスの AR がインイヤー型デバイスよりも高い傾向がある。これは、デバイス面積が大きくなったことでデバイスへのジェスチャ定義が好まれたことと、耳輪や耳珠を使ったジェスチャが避けられたことの 2つの原因によると考えられる。また、同じタスクでデバイスタイプによってジェスチャが異なるペアの多くは、インイヤー型デバイスでは耳を使ったジェスチャ、耳掛け型デバイスではデバイスへのジェスチャが定義されていた。インイヤー型デバイスの耳を使ったジェスチャでは、耳輪、耳たぶ、耳珠、耳全体の変形といった変形パターンに関してもバラ

ス良く分布していた。一方、耳掛け型デバイスでは、耳輪と耳たぶに対するジェスチャが多く定義された。これらのジェスチャは、デバイスタイプに依存しない普遍的なユーザ定義ジェスチャである可能性が高いが、確認するためにはさらに異なるデバイスタイプで調査する必要がある。

### 3.3.3 既存の Elicitation study との共通点や相違点

本研究では、ヒアラブルのみでの入力を想定したユーザ定義型ジェスチャを調査した。デバイスへのタップやスワイプ関連のジェスチャが音楽の再生・停止や方向指示のタスクにおいて、最も AR が高い結果となった。この結果は、デバイスを装着した状態の Elicitation Study [11] と一致するが、デバイスを装着せずに行った Elicitation study [10] の結果とは異なった。つまり、デバイス装着の有無がユーザ定義型ジェスチャに大きく影響することが示唆された。また、AR の平均は本研究が 0.20(SD=0.03) であり、関連研究 [10], [11] の 0.21 と非常に似た結果となった。これらの結果は、耳周辺に関するインタラクションでは、デバイスタイプや装着状態に関係なく、多くのユーザが同程度の直感を持ってジェスチャを定義していることが示された。

## 4. IMU によるジェスチャ認識

ヒアラブルに内蔵されている IMU を用いたジェスチャ

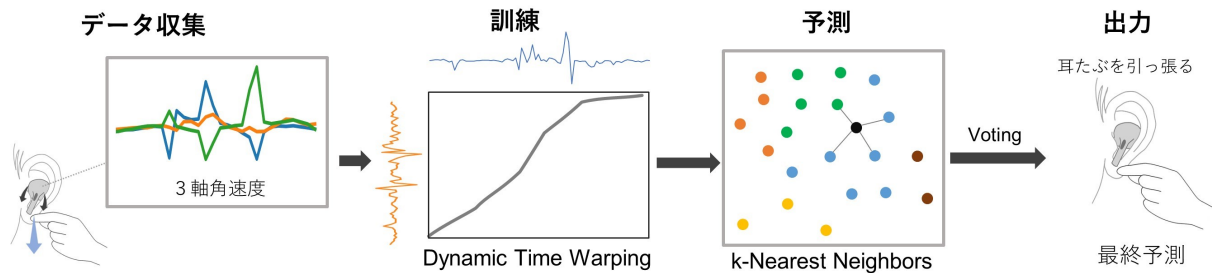


図 6 システムの概要

認識手法を提案する。本手法では、ヒアラブルと親和性の高い IMU を用いることで、実装コストを削減し IMU の効率的な利用を可能にする。安定した環境で座った状態での実験を行い、本手法のベースライン認識精度を調査した。

#### 4.1 認識システム

認識システムの概要を図 6 に示す。ユーザは IMU が内蔵されているヒアラブルを装着し、ジェスチャを行った。ジェスチャによってイヤホンにかかる荷重の方向や強さが異なるため、得られる IMU データも異なる。認識システムは、DTW (Dynamic Time Warping) 距離をメトリックとした k-NN (k-Nearest Neighbor) によってジェスチャの予測をした。

まず、事前に収集したジェスチャデータ間の DTW 距離を計算し、時系列データの類似度を学習させた。そして、k-NN で未知のジェスチャデータと DTW 距離が最も近い k 個の学習データを選択し、多数決で未知データのジェスチャラベルを予測した。この実験では、データの自動切り出しは行わず、データ計測の開始と終了は手動で行った。予備実験にて、角速度データのみで認識精度が最も高かったため、加速度と重力加速度のデータは使用しなかった。

#### 4.2 実装

本研究では、インイヤー型と耳掛け型の 2 種類のデバイス (図 7) を用いてジェスチャ認識実験を行った。インイヤー型デバイスの IMU データの取得には、Apple の AirPods Pro を使用した。AirPods Pro からセンサデータを取得できる API を使用し、IMU データを取得するアプリを Swift を用いて実装した。耳掛け型デバイスは、IMU が取得可能なデバイスを発見することができなかったため、デバイス筐体のみ Victor の HA-NP35TBK を使用し、内部に 9 軸センサ BNO005 を配置したプロトタイプデバイスを作製した。センサデータは有線接続した Arduino Uno を介してシリアル通信にてラップトップ PC (ASUS: ROG FLOW) へと送信し、データを収集した。各デバイスのデータ測定のサンプリングレートは約 30Hz であった。機械学習および耳掛け型デバイスの PC 側のデータ収集のプログラムは python3.7 で実装した。



図 7 インイヤー型デバイスと耳掛け型デバイス

#### 4.3 評価実験

本実験ではインイヤー型デバイスと耳掛け型デバイスにて、それぞれ 11 名ずつの実験を行った。5 名が両方のデバイス実験に参加し、残りは異なる参加者であるため、延べ人数は 16 名であった (男: 11 名, 女: 5 名, 年齢: 平均 26.8 歳)。ただし、耳掛け型実験では、1 名のデータ測定に不備があったため有効なデータは 10 名分となった。1 デバイスあたりの実験時間は約 1 時間で、謝金として 3,000 円を支払った。実験参加者は全員右利きで、14 名が週に 1 回以上イヤホンを利用していた。本実験の実施には、慶應義塾大学理工学部の倫理委員会の許可を得た。

まず、実験参加者にジェスチャの説明を行った。参加者は右手を使ってすべてのジェスチャをするよう指示した。各ジェスチャについて参加者が正しく行えるように簡単な練習を行った。その後、ディスプレイに表示されたジェスチャを実行してもらい、その間のセンサデータを取得した。ジェスチャはランダムに表示され、全てのジェスチャの測定を行った。これを 1 ラウンドとし、デバイスの着脱を挟んで合計 12 ラウンド測定した。

表 2 の耳タッチ条件のユーザ定義型ジェスチャから耳を使ったジェスチャ (図 8) を選択し、認識率を評価した。ただし、デバイススタップ (左耳) などはデバイススタップ (右耳) と同一動作のジェスチャであるため省略した。インイヤー型デバイスでは 9 種類、耳掛け型デバイスでは 6 種類のジェスチャのデータ測定を行った。本研究で使用した AirPods Pro や HA-NP35TBK には、感圧センサを利用したデバイスへのタップなどのジェスチャ入力機能が実装されている。今後も利用できるジェスチャは増加し、ピンチイン/アウトを含めたデバイスへのユーザ定義型ジェス

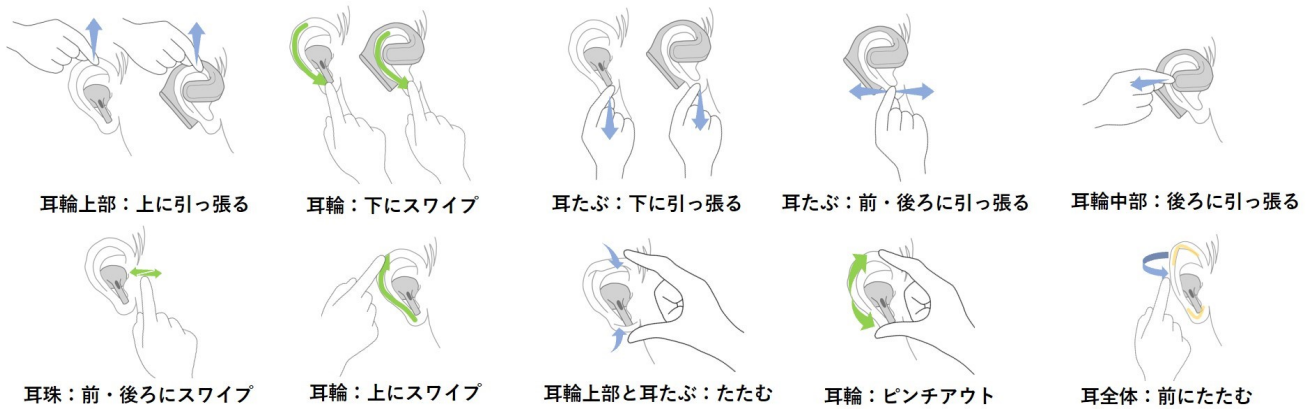


図 8 ジェスチャセット

		インイヤ型デバイス										耳掛け型デバイス											
		F値 [%]										F値 [%]											
正解ラベル	耳輪上部：上に引っ張る	90.8	0.0	0.0	3.1	0.0	0.0	0.8	3.8	1.5	88.1	正解ラベル	耳輪上部：上に引っ張る	78.2	10.9	3.4	2.5	2.5	2.5	2.5	2.5	2.5	77.5
	耳輪：上にスワイプ	1.5	88.5	0.8	0.0	2.3	0.8	2.3	3.8	0.0	91.4		耳輪：下にスワイプ	5.8	80.8	1.7	1.7	4.2	5.8	70.1			
	耳輪：下にスワイプ	0.0	0.0	93.1	0.8	0.8	0.0	2.3	3.1	0.0	95.7		耳たぶ：下に引っ張る	5.0	8.3	72.5	4.2	7.5	2.5	73.5			
	耳たぶ：下に引っ張る	13.1	3.1	0.0	74.6	1.5	0.0	2.3	4.6	0.8	82.8		耳たぶ：後ろに引っ張る	4.2	15.0	11.7	58.3	3.3	7.5	70.6			
	耳珠：後ろにスワイプ	0.0	3.1	0.0	0.0	90.1	1.5	3.8	1.5	0.0	92.2		耳たぶ：前に引っ張る	4.2	7.5	3.3	0.0	82.5	2.5	82.5			
	耳珠：前にスワイプ	0.0	0.0	0.0	0.8	1.5	95.4	0.0	2.3	0.0	96.5		耳輪中部：後ろに引っ張る	0.0	7.5	1.7	0.8	0.8	89.2	85.1			
	耳輪上部と耳たぶ：たたむ	0.0	0.0	0.0	0.0	0.0	0.0	96.2	3.8	0.0	91.6												
	耳輪：ピンチアウト	0.0	0.0	0.0	0.0	0.0	0.0	3.1	96.9	0.0	88.0												
	耳全体：前にたたむ	1.5	0.8	0.0	0.0	0.0	0.0	0.0	0.0	97.7	97.7												
			予測ラベル										予測ラベル										

図 9 識別結果

チャのほとんどは感圧センサを用いて認識可能となると考えられる。本実験では耳を使ったジェスチャの認識率を調査することにより、感圧センサと組み合わせた場合の本手法の精度を評価した。

#### 4.4 結果

図 9 にジェスチャセットの認識精度をまとめた混同行列を示す。混同行列内の数値は precision の百分率を示す。全体平均認識率は、インイヤ型デバイスが 91.6%、耳掛け型デバイスが 76.6%であった。最も認識率の高いジェスチャは、インイヤ型デバイスでは耳全体を前にたたむ (97.7%) で、耳掛け型デバイスでは耳輪中部を前に引っ張る (89.2%) であった。また、最も認識率の低いジェスチャは、インイヤ型デバイスでは耳たぶを下に引っ張る (74.6%) で、耳掛け型デバイスでは耳たぶを後ろに引っ張る (58.3%) であった。両デバイスにおいて耳たぶを使ったジェスチャの認識精度が低かったが、耳たぶの変位によるデバイスの変位が他の部位と比べて小さかったことや、人によって耳たぶの大きさが異なるためジェスチャのしやすさに差が出たことなどが要因として考えられる。

インイヤ型デバイスでは 91.6%という高い認識精度が得られており、IMU を用いたジェスチャ認識が可能であることがわかった。一方、耳掛け型デバイスの認識率はインイヤ型デバイスと比べて低かった。これは、耳掛け型デバイスでは 6 種類のジェスチャの認識精度を調査したが、

その内耳輪を 2 方向に引っ張るジェスチャ、耳たぶを 3 方向に引っ張るジェスチャと似たジェスチャが 5 種類あったため、これらのジェスチャ間の誤認識率が増加したことが要因と考えられる。また、耳掛け型デバイスは、耳に接している面積が大きく安定した装着ができるため、耳の変位に対して得られるデバイスの変位が小さかったことも要因として考えられる。

## 5. 議論・今後の展望

### 5.1 ジェスチャセット

本研究では、3 条件のインタフェース条件に対して 2 種類のデバイスタイプの計 6 条件において、ヒアラブルへの入力に着目したユーザ定義型ジェスチャを調査した。特に、デバイスタイプの差異はユーザ定義型ジェスチャに影響を与えることが分かった。ヒアラブルの形体は、本研究で調査したインイヤ型や耳掛け型以外にも、一体型 [34] やクリップ型 [35] など様々である。このような多くのデバイスタイプでのユーザ定義型ジェスチャの知見を収集することで、デバイスタイプのデザイン指針を明らかにすることが、本研究の展望として可能性が広がる。また、実験参加者の中には、すでにイヤホンへのタップジェスチャなどを利用していることがある、もしくは知識として知っている参加者がいた。このような事前知識の有無や、年齢、文化などの差異が及ぼす影響を今後はさらに調査する必要がある。



## 5.2 ジェスチャ認識

特徴量の改善, 深層学習の利用などによってさらなる認識率の向上が期待できる. また, ジェスチャ検出機能を実装し, 歩行時や運動時など日常動作中のジェスチャの検出や認識精度を評価する必要がある. 日常動作でのジェスチャには, 動作ノイズが含まれる. 認識精度の低下を防ぐために, 動作ノイズを取り除くための手法 [36], [37] が多く提案されており, これらの手法の利用も検討する必要がある. 一方で, 本手法はヒアラブルを利用するため, 左右のデバイスから慣性データを同時に取得できる特徴がある. したがって, 一方はジェスチャ+動作成分, もう一方は, 動作成分のみを含む慣性データが得られる. これらの左右のデータを利用する方式はイヤホン特有のノイズ除去アルゴリズムであり, 今後研究を進める予定である. Elicitation study と同様に一体型やクリップ型のデバイスでのジェスチャ認識精度も調査し, 提案手法の適用可能性を明らかにする予定である.

Elicitation study ではデバイスの面積が広い耳掛け型デバイスの AR のほうが高かった. 一方, デバイス面積が大きくなることで, 類似したジェスチャが定義されることが多くなり, ジェスチャ認識率は低下するという課題も生じることが分かった. 今後は, 実際の認識システムをユーザに提示することで, 直感性やユーザビリティを調査する予定である. これによって認識精度がユーザビリティにどのような影響を与えるかが明らかになり, 実生活に最適なジェスチャセットや認識システムの開発が可能になると考える.

## 6. 結論

本研究では, ヒアラブルへの入力を想定したハンズジェスチャの Elicitation study を行い, ユーザの好むジェスチャや制約・条件による定義の傾向の変化を明らかにした. 決定したユーザ定義型ジェスチャのうち耳を使ったジェスチャに対して, IMU によるジェスチャ認識実験を行った. 実験の結果, インイヤ型デバイスでの 9 種類のジェスチャ認識率が 91.6%, 耳掛け型デバイスでの 6 種類のジェスチャ認識率が 76.6%であった.

**謝辞** 本研究の一部は, JST さきがけ (課題番号: JPMJPR2134, JPMJPR2138) および JSPS 科研費 (課題番号: JP23KJ1884) の支援を受けたものである.

## 参考文献

- [1] Ando, T., Kubo, Y., Shizuki, B. and Takahashi, S.: CanalSense:Face-Related Movement Recognition System Based on Sensing Air Pressure in Ear Canals, *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology*, UIST '17, ACM, pp. 679–689 (2017).
- [2] 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*, Vol. 18, No. 3 (2018).
- [3] Matthies, D. J. C., Strecker, B. A. and Urban, B.: EarFieldSensing: A Novel In-Ear Electric Field Sensing to Enrich Wearable Gesture Input Through Facial Expressions, *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, CHI '17, ACM, pp. 1911–1922 (2017).
- [4] Amesaka, T., Watanabe, H. and Sugimoto, M.: Facial Expression Recognition Using Ear Canal Transfer Function, *Proceedings of the 23rd International Symposium on Wearable Computers*, ISWC '19, ACM, pp. 1–9 (online), DOI: 10.1145/3341163.3347747 (2019).
- [5] Bedri, A., Byrd, D., Presti, P., Sahni, H., Gue, Z. and Starner, T.: Stick It in Your Ear: Building an In-ear Jaw Movement Sensor, *Adjunct Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2015 ACM International Symposium on Wearable Computers*, UbiComp/ISWC '15 Adjunct, ACM, pp. 1333–1338 (2015).
- [6] Xu, X., Shi, H., Yi, X., Liu, W., Yan, Y., Shi, Y., Mariakakis, A., Mankoff, J. and Dey, A. K.: EarBuddy: Enabling On-Face Interaction via Wireless Earbuds, *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, ACM, pp. 1–14 (online), DOI: 10.1145/3313831.3376836 (2020).
- [7] Manabe, H. and Fukumoto, M.: Headphone Taps: Tap Control for Regular Headphones, *Journal of Information Processing*, Vol. 55, No. 4, pp. 1334–1343 (2014).
- [8] Kikuchi, T., Sugiura, Y., Masai, K., Sugimoto, M. and Thomas, B. H.: EarTouch: Turning the Ear into an Input Surface, *Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services*, MobileHCI '17, New York, NY, USA, ACM, (online), DOI: 10.1145/3098279.3098538 (2017).
- [9] Lissermann, R., Huber, J., Hadjakos, A., Nanayakkara, S. and Mühlhäuser, M.: EarPut: Augmenting Ear-Worn Devices for Ear-Based Interaction, *Proceedings of the 26th Australian Computer-Human Interaction Conference on Designing Futures: The Future of Design*, OzCHI '14, New York, NY, USA, ACM, pp. 300–307 (online), DOI: 10.1145/2686612.2686655 (2014).
- [10] Chen, Y.-C., Liao, C.-Y., Hsu, S.-w., Huang, D.-Y. and Chen, B.-Y.: Exploring User Defined Gestures for Ear-Based Interactions, *Proceedings of the ACM on Human-Computer Interaction*, Vol. 4, No. ISS (online), DOI: 10.1145/3427314 (2020).
- [11] Rateau, H., Lank, E. and Liu, Z.: Leveraging Smartwatch and Earbuds Gesture Capture to Support Wearable Interaction, *Proceedings of the ACM on Human-Computer Interaction*, Vol. 6, No. ISS (online), DOI: 10.1145/3567710 (2022).
- [12] Tamaki, E., Miyaki, T. and Rekimoto, J.: Brainy Hand: An Ear-Worn Hand Gesture Interaction Device, *CHI '09 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '09, New York, NY, USA, ACM, pp. 4255–4260 (online), DOI: 10.1145/1520340.1520649 (2009).
- [13] Metzger, C., Anderson, M. and Starner, T.: FreeDigiter: A Contact-Free Device for Gesture Control, *Eighth International Symposium on Wearable Computers*, Vol. 1, pp. 18–21 (online), DOI: 10.1109/ISWC.2004.23 (2004).
- [14] SONY: LinkBuds, <https://www.sony.jp/headphone/>

- products/LinkBuds/ (2022). (2023.09.13. accessed).
- [15] Röddiger, T., Wolfram, D., Laubenstein, D., Budde, M. and Beigl, M.: Towards Respiration Rate Monitoring Using an In-Ear Headphone Inertial Measurement Unit, *Proceedings of the 1st International Workshop on Earable Computing*, EarComp '19, New York, NY, USA, ACM, pp. 48–53 (online), DOI: 10.1145/3345615.3361130 (2020).
- [16] Bi, H., Sun, Y., Liu, J. and Cao, L.: SmartEar: Rhythm-Based Tap Authentication Using Earphone in Information-Centric Wireless Sensor Network, *IEEE Internet of Things Journal*, Vol. 9, No. 2, pp. 885–896 (online), DOI: 10.1109/JIOT.2021.3063479 (2022).
- [17] Mollyn, V., Arakawa, R., Goel, M., Harrison, C. and Ahuja, K.: IMUPoser: Full-Body Pose Estimation Using IMUs in Phones, Watches, and Earbuds, *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, CHI '23, New York, NY, USA, ACM, (online), DOI: 10.1145/3544548.3581392 (2023).
- [18] Verma, D., Bhalla, S., Sahnani, D., Shukla, J. and Parnami, A.: ExpressEar: Sensing Fine-Grained Facial Expressions with Earables, *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, Vol. 5, No. 3 (online), DOI: 10.1145/3478085 (2021).
- [19] Morris, M. R., Wobbrock, J. O. and Wilson, A. D.: Understanding Users' Preferences for Surface Gestures, *Proceedings of Graphics Interface 2010*, GI '10, CAN, Canadian Information Processing Society, p. 261–268 (2010).
- [20] Masai, K., Kunze, K., Sakamoto, D., Sugiura, Y. and Sugimoto, M.: Face Commands - User-Defined Facial Gestures for Smart Glasses, *2020 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 374–386 (online), DOI: 10.1109/ISMAR50242.2020.00064 (2020).
- [21] Fukahori, K., Sakamoto, D. and Igarashi, T.: Exploring Subtle Foot Plantar-Based Gestures with Sock-Placed Pressure Sensors, *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI '15, New York, NY, USA, ACM, p. 3019–3028 (online), DOI: 10.1145/2702123.2702308 (2015).
- [22] Chan, E., Seyed, T., Stuerzlinger, W., Yang, X.-D. and Maurer, F.: User Elicitation on Single-Hand Microgestures, *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, CHI '16, New York, NY, USA, ACM, p. 3403–3414 (online), DOI: 10.1145/2858036.2858589 (2016).
- [23] Sharma, A., Roo, J. S. and Steimle, J.: Grasping Microgestures: Eliciting Single-Hand Microgestures for Hand-held Objects, *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, New York, NY, USA, ACM, p. 1–13 (online), DOI: 10.1145/3290605.3300632 (2019).
- [24] Yan, Y., Yu, C., Yi, X. and Shi, Y.: HeadGesture: Hands-Free Input Approach Leveraging Head Movements for HMD Devices, *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, Vol. 2, No. 4 (online), DOI: 10.1145/3287076 (2018).
- [25] Weigel, M., Mehta, V. and Steimle, J.: More than Touch: Understanding How People Use Skin as an Input Surface for Mobile Computing, *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '14, New York, NY, USA, ACM, p. 179–188 (online), DOI: 10.1145/2556288.2557239 (2014).
- [26] Vatavu, R.-D.: User-Defined Gestures for Free-Hand TV Control, *Proceedings of the 10th European Conference on Interactive TV and Video*, EuroITV '12, New York, NY, USA, ACM, p. 45–48 (online), DOI: 10.1145/2325616.2325626 (2012).
- [27] Ruiz, J., Li, Y. and Lank, E.: User-Defined Motion Gestures for Mobile Interaction, *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, New York, NY, USA, ACM, p. 197–206 (online), DOI: 10.1145/1978942.1978971 (2011).
- [28] Serrano, M., Ens, B. M. and Irani, P. P.: Exploring the Use of Hand-to-Face Input for Interacting with Head-Worn Displays, *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '14, New York, NY, USA, ACM, p. 3181–3190 (online), DOI: 10.1145/2556288.2556984 (2014).
- [29] Gheran, B.-F., Vanderdonckt, J. and Vatavu, R.-D.: Gestures for Smart Rings: Empirical Results, Insights, and Design Implications, *Proceedings of the 2018 Designing Interactive Systems Conference*, DIS '18, New York, NY, USA, ACM, p. 623–635 (online), DOI: 10.1145/3196709.3196741 (2018).
- [30] Malu, M., Chundury, P. and Findlater, L.: Exploring Accessible Smartwatch Interactions for People with Upper Body Motor Impairments, *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, New York, NY, USA, ACM, p. 1–12 (online), DOI: 10.1145/3173574.3174062 (2018).
- [31] Dierk, C., Carter, S., Chiu, P., Dunnigan, T. and Kimber, D.: Use Your Head! Exploring Interaction Modalities for Hat Technologies, *Proceedings of the 2019 on Designing Interactive Systems Conference*, DIS '19, New York, NY, USA, ACM, pp. 1033–1045 (online), DOI: 10.1145/3322276.3322356 (2019).
- [32] Yamamoto, T., Masai, K., Withana, A. and Sugiura, Y.: Masktrap: Designing and Identifying Gestures to Transform Mask Strap into an Input Interface, *Proceedings of the 28th International Conference on Intelligent User Interfaces*, IUI '23, New York, NY, USA, ACM, p. 762–775 (online), DOI: 10.1145/3581641.3584062 (2023).
- [33] Vatavu, R.-D. and Wobbrock, J. O.: Formalizing Agreement Analysis for Elicitation Studies: New Measures, Significance Test, and Toolkit, *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI '15, New York, NY, USA, Association for Computing Machinery, pp. 1325–1334 (online), DOI: 10.1145/2702123.2702223 (2015).
- [34] SHOKZ: OPENRUN, <https://jp.shokz.com/products/openrun> (2023). (2023.09.13. accessed).
- [35] Ambie: Ambie, <https://ambie.co.jp/soundearcuffs/> (2017). (2023.09.13. accessed).
- [36] Song, Z., Cao, Z., Li, Z., Wang, J. and Liu, Y.: Inertial motion tracking on mobile and wearable devices: Recent advancements and challenges, *Tsinghua Science and Technology*, Vol. 26, No. 5, pp. 692–705 (online), DOI: 10.26599/TST.2021.9010017 (2021).
- [37] Kang, P., Li, J., Fan, B., Jiang, S. and Shull, P. B.: Wrist-Worn Hand Gesture Recognition While Walking via Transfer Learning, *IEEE Journal of Biomedical and Health Informatics*, Vol. 26, No. 3, pp. 952–961 (online), DOI: 10.1109/JBHI.2021.3100099 (2022).