

外耳道伝達関数による頭部状態認識手法

雨坂 宇宙^{1,†1,a)} 渡邊 拓貴^{1,b)} 杉本 雅則^{1,c)}

受付日 2019年10月23日, 採録日 2020年5月12日

概要：近年注目されているヒアラブルデバイスにおいて求められる機能の1つとして、手や視界を占有することのないデバイス操作機能があげられる。既存製品や既存研究では認識精度や認識できるジェスチャの種類、センサの追加コストなどの点で課題が残る。これらの課題の解決のために我々は首、顎、顔の状態（頭部状態）にともなって外耳道が変形することに着目し、外耳道伝達関数を測定、解析することで現在の頭部状態を認識する手法を提案した。提案手法は外耳道内部の音を取得できるマイクを利用するため、ノイズキャンセリング機能などとの併用が可能であり、ヒアラブルデバイスとの親和性が高い。また、デバイスの着脱や時間経過による装着具合の誤差を補正することで、認識精度の向上を実現した。11名の被験者に対して21種類の頭部状態の認識実験を行った結果、各被験者の分類器の平均認識精度は、未補正時で40.2%（F値）、補正時で62.5%（F値）となった。実際のアプリケーションでの利用を想定し、6種類の頭部状態の認識結果を行った結果、未補正時で74.4%（F値）、補正時で90.0%（F値）の認識精度が得られた。

キーワード：頭部状態認識, 超音波, 外耳道伝達関数, ヒアラブルデバイス

Ear Canal Transfer Function-based Facial Expression Recognition

TAKASHI AMESAKA^{1,†1,a)} HIROKI WATANABE^{1,b)} MASANORI SUGIMOTO^{1,c)}

Received: October 23, 2019, Accepted: May 12, 2020

Abstract: In this study, we propose a new input method for wearable computing using facial expressions. Facial muscle movements induce physical deformation in the ear canal. Our system utilizes such characteristics and estimates facial expressions using the ear canal transfer function (ECTF). Herein, a user puts on earphones with an equipped microphone that can record an internal sound of the ear canal. The system transmits ultrasonic band-limited swept sine signals and acquires the ECTF by analyzing the response. An important feature of the proposed method is that the microphone can also be used for other purposes, such as noise-canceling. Therefore, we consider that the proposed method is reasonable to be used in earphones. We investigated the performance of our proposed method for 21 facial expressions with 11 participants. Moreover, we proposed a signal correction method that reduces positional errors caused by attaching/detaching the device. The evaluation results confirmed that the f-score was 40.2% for the uncorrected signal method and 62.5% for the corrected signal method. We also investigated the practical performance of six facial expressions and confirmed that the f-score was 74.4% for the uncorrected signal method and 90.0% for the corrected signal method.

Keywords: facial expression recognition, ultrasound, ear canal transfer function, hearables

¹ 北海道大学
Hokkaido University, Sapporo, Hokkaido 060-0814, Japan
^{†1} 現在, 筑波大学
Presently with Tsukuba University
^{a)} amesaka@iplab.cs.tsukuba.ac.jp
^{b)} hiroki.watanabe@ist.hokudai.ac.jp
^{c)} sugi@ist.hokudai.ac.jp

1. はじめに

近年のウェアラブルコンピューティングの発展にともない、ヒアラブルデバイス [1] が着目されている。ヒアラブルデバイスとは、耳に装着するイヤホン型のウェアラブルコンピュータである。従来のイヤホンの用途である音楽鑑

賞を始め、スマートフォンと連携することによる音声アシスタントの利用や、搭載された各種センサを用いてユーザの生体情報や行動を認識できる。現在市販されているヒアラブルデバイスの多くはスマートフォンと連携して使用することを前提としており、デバイス进行操作する際にスマートフォンの画面に着目し、手で操作する必要がある。ヒアラブルデバイスにセンサを搭載し、デバイスへのタッチや頭部の動きによってコマンド操作できるデバイス [2], [3] も販売されているが、これらの製品ではヒアラブルデバイスを直接タッチする必要や、認識できる頭部のジェスチャは限られているという問題がある。ウェアラブルコンピューティングは、屋外環境や別の作業をしながらの使用が想定されるため、手や視界を占有することのないデバイス操作機能が求められている。音声アシスタントによるデバイス操作手法も存在するが、公共の場での発声や騒音による影響などの課題がある。

ヒアラブルデバイスでハンズフリー操作を実現する手法として気圧、赤外線、慣性、電極センサなどを用いた手法がある [4], [5], [6], [7]。しかし、気圧センサでは飛行機やエレベータ内などの急激に気圧が変化する環境で、認識精度が落ちる可能性がある。赤外線、慣性センサは、現時点では十分な認識精度を得られていない。また、電極センサを含めこれらの手法ではジェスチャ認識のための追加のセンサをイヤホンに組み込む必要がある。

本研究では首、顎、顔などの状態（頭部状態）によって、外耳道の形状が変化することに着目し、音響信号により頭部状態を認識する手法を提案する。頭部状態の認識を実現することで、手や視界を占有することなくデバイスを操作することが可能となる。具体的には、測定信号を用いたインパルス応答測定手法により外耳道内の帯域制限されたインパルス応答を求め、そのフーリエ変換により外耳道伝達関数を求める。頭部状態の変化にともなう外耳道の形状変化によって、得られる外耳道伝達関数が異なるため、その変化パターンを機械学習することで現在のユーザの頭部状態を認識する。提案手法では超音波帯域の音を利用するため、超音波帯域も取得可能なマイクを追加する必要があるが、市販製品でもノイズキャンセリングなどのために外耳道内部の音をマイクで取得できる製品が存在している [2], [8], [9]。したがって、超音波対応マイクの追加はヒアラブルデバイスと親和性が高く、ノイズキャンセリングなどの他の用途とも併用できる利点がある。また、デバイスの装着具合による外耳道伝達関数の変化（装着誤差）を低減するための信号補正手法を提案し、認識精度の向上を実現した。本研究の成果を以下にまとめる。

- 超音波信号による外耳道伝達関数の取得と頭部状態認識システムの提案
- 11名の被験者に対して、被験者ごとに21種類の頭部状態を未補正下で平均40.2% (F値)、補正下で平均

62.5% (F値) の精度で認識

- 11名の被験者に対して、被験者ごとに実用を考慮した6種類の頭部状態を未補正下で平均74.4% (F値)、補正下で平均90.0% (F値) の精度で認識

なお、本論文は文献 [10] で発表した内容に考察を加えて発展させたものである。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章で結論を述べる。

2. 関連研究

2.1 ヒアラブルデバイスに関する研究・製品

様々なセンサを用いて、ヒアラブルデバイスで頭部状態やジェスチャを認識する研究が行われている。Andoら [4] は、顔関連の動作による外耳道内部の気圧変化を気圧センサで取得し、ユーザごとに11種類の顔関連の動作を87.6%の精度で認識することに成功している。さらに、4段階の口の開け幅を87.5%の精度で認識することに成功している。Matthiesら [5] は電極センサを用いて外耳道内部から筋肉の動きを読み取り5種類の頭部状態を精度90.0% (座位状態)、85.2% (歩行状態) で認識することに成功している。Taniguchiら [6] はLEDとフォトトランジスタを使って舌の特定の動きを外耳道の変形から認識し、音楽プレーヤの操作を対象にユーザビリティの調査も行っている。Bedriら [7] は近接センサを用いて外耳道の変形を読み取り、心拍数、舌・顎の動作、まばたきを認識している。真鍋ら [11] 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るだけでヘッドホンをタップする動作を認識することに成功している。Laputら [12] はイヤホンに組み込まれたスピーカとマイクから反響音を取得し、イヤホンの着脱状態を認識している。製品では、DashPro [3] が慣性センサを用いて頭部ジェスチャを認識し、電話の応答や音楽プレーヤの曲変更などの簡単なハンズフリー操作を実現している。

上述した既存研究・製品は気圧、電極、光、慣性センサを用いて頭部状態認識を行っている。気圧センサを用いる手法では、高い精度で動作の認識を行っているが、飛行機やエレベータなどの急激に気圧が変化する環境で認識精度が落ちる可能性がある。光センサ、慣性センサを用いた手法では現段階において認識できる頭部状態の数や認識精度が十分とはいえない。また、電極センサを含めた既存研究では頭部状態認識のために追加のセンサが必要である。提案手法では、外耳道内部の超音波を取得できるマイクを追加する必要があるが、外耳道内部の音を取得できるマイクはノイズキャンセリングなどの他の用途とも併用できるため、ヒアラブルデバイスとの親和性が高い。

2.2 外耳道インパルス応答の測定と応用に関する研究

Hiipakkaら [13] は外耳道の音響特性などを調査してお

り, Swept-Sine 信号 [14] とマイク内蔵イヤホンを用いて外耳道インパルス応答を測定している. Akkermans ら [15] はプローブ信号より外耳道インパルス応答を測定し, 伝達関数より外耳道の個性を見出している. さらに Arakawa ら [16], [17] は Maximum Length Sequence 法 (MLS 法) [18] を用いて外耳道の個性を詳しく調査している.

本研究では Hiipakka らと同じ Swept-Sine 信号にてインパルス応答を測定している. これらの既存研究では可聴域のインパルス応答に着目している. 本研究では, 実用性を考慮し, ユーザに聞こえない超音波領域に帯域制限したインパルス応答の測定を行った. また, 既存研究は外耳道音響特性の個性に着目しているが, 頭部状態の変化にともなう音響特性の変化には着目されていなかった. 我々は頭部状態の変化による外耳道音響特性の変化を利用し, 現在の頭部状態を推定する手法を提案した. さらに, インパルス応答を基に測定信号を補正することでデバイスの装着誤差を低減している.

2.3 人体へ音波を適用した研究

Tan ら [19] はスマートフォンから発する音声信号を用いて発話による口周辺部の動きを読み取り, 口の動きによる個人認証システムを提案している. Watanabe ら [20] は腕と足にコンタクトスピーカとコンタクトマイクを装着し, 人体内部を伝播する音の変化から 21 種類の状態を認識することに成功している. Mujibiya ら [21] は皮膚を伝播する超音波を用いて体接触やハンドジェスチャの認識を行った. また, Takemura ら [22] は骨伝導マイクを用いて指のタップ位置検出と肘の角度推定を行っている.

上記の研究では人体へ音波を適用する点では本研究と同一であるが, 外耳道という人体の一部を空間ととらえ, インパルス応答によりその変化を認識する点が本研究と異なる. また, 測定信号を補正することでデバイスの装着誤差を低減させる手法は既存研究では行われていない.

3. 提案手法

3.1 システム構成

提案システム全体の流れを図 1 にまとめた. ユーザは外耳道内部の音を録音できるように設計されたマイク内蔵イヤホンを装着する. イヤホンから測定信号を再生し, 外耳道伝達関数を測定する. ユーザが顔の向きや表情を変えて外耳道の形状が変化するため, 外耳道側面や鼓膜からの反響が変化する. そのため, 得られる外耳道伝達関数も変化する (図 2). 外耳道伝達関数からそれぞれの頭部状態の特徴量を抽出し機械学習を行うことで, 頭部状態を認識する分類器を作成する. 本研究では, 測定信号を超音波領域に帯域制限することで, ユーザの聴覚への影響を少なくする. また実環境には可聴域の音が多く存在しており, ノイズ処理や測定信号の抽出の精度が認識精度に大き

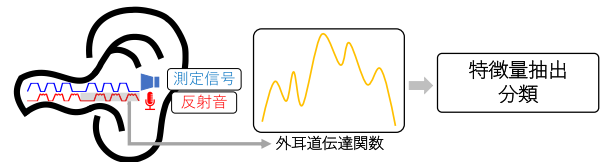


図 1 システム構成図

Fig. 1 System configuration.

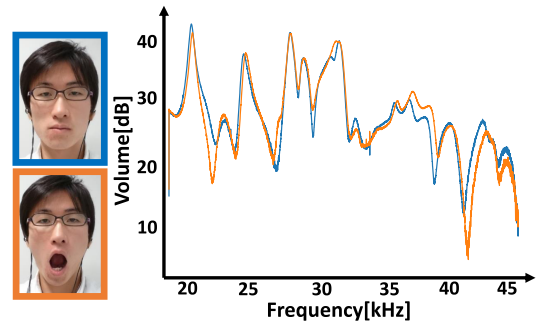


図 2 頭部状態による外耳道伝達関数の変化

Fig. 2 Changes in frequency spectrum of ECTF.

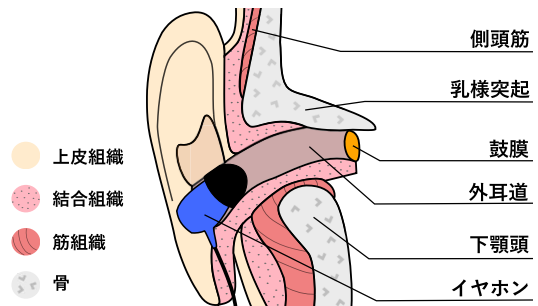


図 3 外耳道周辺部

Fig. 3 Surrounding structure of ear canal.

く影響するが, 超音波ではそれらの影響が少ないという利点がある.

3.2 認識原理

顎は顎運動時に下顎頭 (図 3) の動きと外耳道のひずみに相関関係があることを報告している [23]. また外耳道上部には側頭骨があり, それに属する側頭筋周辺には表情筋が密集しており眼球の運動が外耳道の形状に影響すると考える. よって本研究では, 顎・首・眼周辺の運動を基に頭部状態を定義し, 認識を行う. 図 4 に選出したジェスチャを示す. 本研究では, どのジェスチャが提案手法に有効なのかを調査するために, 先行研究 [5] で用いられたものを基に 21 種類の頭部状態を定義した.

3.3 外耳道伝達関数測定原理

測定信号を用いた伝達関数の測定原理を図 5 に示す. 図 5 中の線形システム H の前に周波数特性 $S(k)$ (k : 離散周波数番号) を持った信号合成フィルタ S と, 後ろに逆特性 $1/S(k)$ を持つ逆フィルタ $1/S$ を加えた測定系を考



図 4 頭部状態

Fig. 4 Facial expressions.



図 5 伝達関数測定原理

Fig. 5 Measuring principle of transfer functions.

える．インパルス信号 $\delta(k)$ の周波数特性は 1 なのでフィルタに入力したときの出力は，フィルタ特性と同じ周波数特性 $S(k)$ を持った測定信号となる．その測定信号 $S(k)$ を特性 $H(k)$ を持った線形システムに入力すると，出力は $H(k) \cdot S(k)$ となる．この出力に逆フィルタ $1/S(k)$ を入力すると，出力は伝達関数 $H(k)$ となる．本研究では信号合成フィルタで Swept-Sine 信号を作成し，外耳道を線形システムと考え測定信号を入力し，その出力信号に逆フィルタを適用し外耳道伝達関数を測定する．

3.4 Swept-Sine 法

インパルスを時間軸上に引き伸ばした Swept-Sine 信号を測定信号として用いる手法を Swept-Sine 法 (SS 法) [14] と呼び，インパルス応答の測定に広く利用されている．周波数領域での Swept-Sine 信号 $SS(k)$ は以下のように表される．

$$SS(k) = \begin{cases} \exp(j\alpha k^2) & (0 \leq k \leq N/2) \\ \exp(-j\alpha(N-k)^2) & (N/2 \leq k \leq N-1) \end{cases}$$

ただし， $\alpha = 4m\pi/N^2$ ， j は虚数， k は離散周波数番号， N は離散データ数， m はパルスの引き延ばし係数である．

$SS(k)$ を逆フーリエ変換することで Swept-Sine 信号 $ss(n)$ を得る．また逆フィルタ $SS^{-1}(k)$ は $SS(k)$ の複素共役で求められる．インパルス応答はすべての周波数の音を含むため，測定信号発信時にユーザにも聞こえ，不快となる．そこで，本研究では測定信号の帯域制限を行う [24]．サンプリングレートを f_s として 0 Hz から f_0 Hz の帯域制限を行うには $0 \leq k \leq (f_0 N / f_s)$ と $N - (f_0 N / f_s) \leq k \leq N$ の範囲の $SS(k)$ の振幅値を 0 にすることで実現する．なお，帯域制限を行うことで逆フィルタとの畳み込みは完全

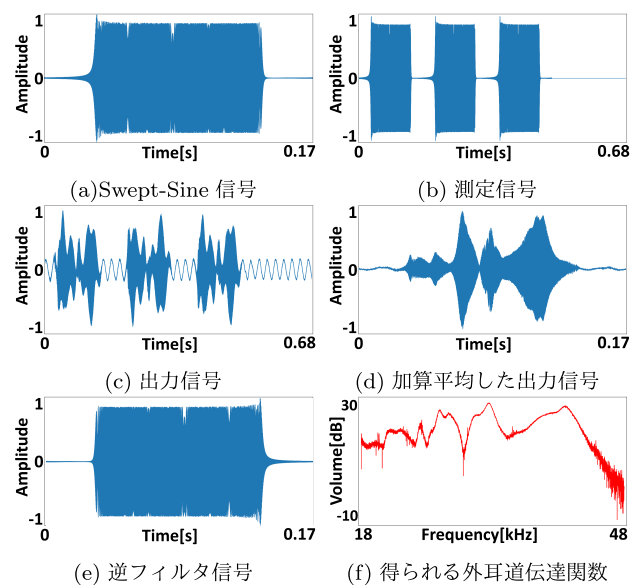


図 6 円状畳み込みの原理を用いたインパルス応答測定

Fig. 6 The procedure of ECTF measurement.

なインパルスとはならず，本研究で測定するインパルス応答や伝達関数は厳密にはインパルス応答や伝達関数とはならないが，便宜上，本論文ではインパルス応答・伝達関数と表記する．

3.4.1 円状畳み込みの原理を用いたインパルス応答測定法

円状畳み込みの原理を用いることで，測定信号長をインパルス応答の長さよりも長く設定すれば，高 SN 比のインパルス応答を取得できる．Swept-Sine 信号 (図 6(a)) を用いた測定手順は以下のとおりである．

- (1) Swept-Sine 信号を同期加算の回数分だけ隙間なく並び，1 周期分の無音区間を末尾に追加する (図 6(b))．
- (2) 測定信号を再生し，その反響音を録音することで出力信号 (図 6(c)) を取得する．
- (3) 出力信号を Swept-Sine 信号長で切り出し加算平均する (図 6(d))．
- (4) 加算平均した出力信号と逆フィルタ信号 (図 6(e)) のフーリエ変換を要素ごとに掛け算する．

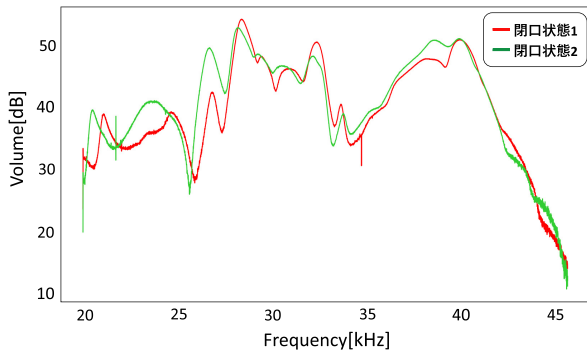


図 7 デバイス着脱による外耳道伝達関数の変化 (未補正)

Fig. 7 The change in ECTF for re-wearing the device (Uncorrected).

上記の操作を外耳道に対して適用することで外耳道伝達関数を取得する (図 6(f)).

3.5 測定信号の修正によるインパルス応答の補正

3.4 節の手順で得られる外耳道伝達関数はデバイスの着脱や時間経過によっても生じる装着誤差によっても変化するため、同じ頭部状態でも測定ごとに外耳道伝達関数が異なるという問題がある (図 7)。これらの変化は頭部状態による外耳道伝達関数の変化と混同される可能性があり、認識精度向上のためには装着誤差は最小限にする必要がある。そこで閉口状態時に測定信号を補正することで、装着誤差を低減する手法を提案する。外耳道伝達関数とそのときの測定信号、出力信号の周波数特性をそれぞれ $H(k)$, $X(k)$, $Y(k)$ (k : 離散周波数番号) とすると以下の関係が成り立つ。

$$X(k)H(k) = Y(k) \quad (1)$$

3.5 節冒頭で述べたとおり、デバイスの装着誤差によって $H(k)$ は変化する。装着誤差を含んだ場合の外耳道伝達関数を $H'(k)$ とする。ここで、測定信号 $X(k)$ による新たな出力を $Y'(k)$ とすると同様に

$$X(k)H'(k) = Y'(k) \quad (2)$$

と表せる。このとき、測定信号に補正を加えて出力を $Y(k)$ にするような補正測定信号 $X'(k)$ は式 (1), (2) より

$$X'(k) = \frac{Y(k)}{H'(k)} = \frac{Y(k)}{Y'(k)} X(k) \quad (3)$$

と表せる。 $Y(k)$ を一意に設定し、閉口状態時に式 (3) による信号の補正を与えることで装着誤差が低減でき、頭部状態の変化による外耳道伝達関数の変化を効率的に測定できると考えられる。本研究では事前に各被験者から取得した閉口状態での外耳道伝達関数の加算平均を $Y(k)$ として利用する。図 8 が実際にデバイスの装着誤差を補正し、得られた外耳道伝達関数である。図 7 と図 8 を比較すると、同じ頭部状態のときに得られる外耳道伝達関数が近くなっており、補正されていることが確認できる。

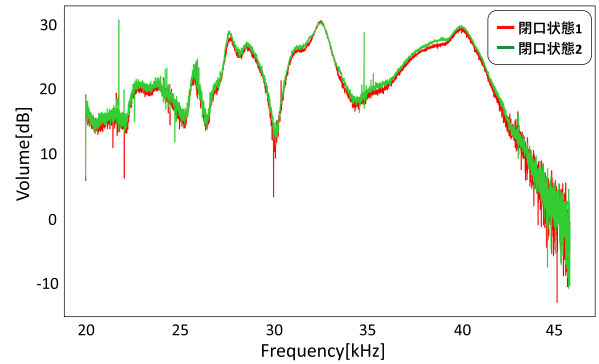


図 8 デバイス着脱による外耳道伝達関数の変化 (補正)

Fig. 8 The change in ECTF for re-wearing the device (Corrected).

3.6 特徴量抽出

取得した音声信号をそのまま機械学習に使用すると、特徴量の次元数が多く学習効率が悪い。特徴量抽出を行う。本研究は特徴量として人間の音声認識によく用いられるメル周波数ケプストラム係数 (MFCC: Mel-Frequency Cepstrum Coefficients) の線形バージョンである LFCC (Linear-Frequency Cepstrum Coefficients) [25] を用いる。これらの違いは、音声データの周波数スペクトルの対数表示に対して、MFCC ではメル周波数領域において等間隔なメルフィルタバンクを適用するが、LFCC では周波数領域において等間隔な線形フィルタバンクを適用する。メル周波数は人間の声の特徴を考慮して設計されているため、本研究には適さないと考え LFCC を採用した。LFCC を特徴量として最大値、最小値で正規化を行い機械学習にかけ、分類器を作成する。分類器には k 近傍法やランダムフォレストなども考えられるが、予備実験にて最も精度の高かったサポートベクタマシン (SVM: Support Vector Machine) を用いた。分類器を基に現在の外耳道伝達関数から頭部状態を予測する。

4. 実装

提案するシステムは外耳道内部に音声信号を再生し、その反響音を取得するためのハードウェアの実装と頭部状態認識アルゴリズムのソフトウェアの実装に分かれる。以下でそれぞれの実装の詳細を述べる。

4.1 外耳道反響音取得デバイス

外耳道の反響音を取得するために市販のイヤホンと小型 MEMS マイクを用いてデバイスを作製した。市販のマイク内蔵型イヤホン [2], [8], [9] を用いることも考えられるが、これらのデバイス内蔵マイクの周波数特性では超音波領域の音を取得できないため、本研究では自作のデバイスを用いた。外部の騒音の影響を小さくし、反響音を最大限に取得するためにイヤホンは耳穴に挿入するタイプのカナル型イヤホンを使用した。また超音波を測定信号に使用す

るため、イヤホンとマイクの対応周波数帯域も考慮し、イヤホンに Xiaomi の Pro HD ハイレゾ対応（再生周波数帯域：20–48 kHz）、マイクに knowles の SPU0410LR5H（録音周波数帯域：100–80 kHz）を用いた。マイクは反響音を効率良く録音できるようにイヤホンの音声再生部分に密着するように取り付けた（図 9）。また、これらの作業を行うにあたってイヤホンの外耳道挿入部分を取り除いたため、新たに 3D プリンタでパーツを作り直した。音声入出力時の DA/AD 変換にオーディオインタフェース（Komplete Audio 6）を使用し、マイクの電源に Sunhayato の DK-910 を使用した。また、PC は Thinkpad X270 を使用した。再生、録音ともにサンプリングレートは 96 kHz で行う。再生・録音デバイスの構成図を図 10 にまとめた。

4.2 頭部状態認識アルゴリズム

本研究で使用する Swept-Sine 信号は 18 kHz 以下を帯域制限した、18 kHz から 48 kHz のアップスイープ信号（16,384 サンプル）を使用した。また、SN 比向上のために同期加算を行う必要がある。本研究で使用する Swept-Sine 信号が 1 周期で約 0.17 秒であり、分類できる時間間隔と SN 比のバランスを考慮し、同期加算回数は 3 回とした。無音区間を追加するため、測定信号は全体で約 0.68 秒となる。実験では測定信号（図 6(b)）を連続で 7 回再生し、1 回の測定で 7 個の外耳道伝達関数を取得した。3.4 節の手順で得られる外耳道伝達関数は Swept-Sine 信号長と同じ 16,384 サンプルとなるので LFCC 抽出時のフーリエ変換

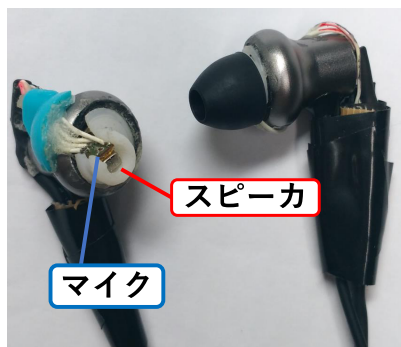


図 9 実験デバイス

Fig. 9 Implemented speaker and microphone.

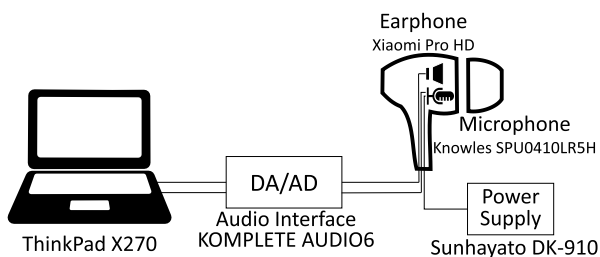


図 10 デバイス構成図

Fig. 10 Device configuration.

は 16,384 サンプルで行い、線形フィルタバンクは 20 分割で行った。学習に用いる特徴量は直流成分である 1 次元目を除いた 19 次元を特徴量とし、外耳道伝達関数は両耳から取得するため、特徴量の次元数は計 38 次元となる。測定信号の補正時に、一意に定める必要のある $Y(k)$ は事前に被験者から取得した閉口状態での外耳道伝達関数の平均値とした。外耳道伝達関数の測定・特徴量抽出・機械学習は Python 3.6 で実装した。

5. 評価実験

5.1 未補正信号による識別実験

5.1.1 実験環境

実験は 23–26 歳の 11 名のボランティア（男：10 名、女：1 名）に参加してもらった。データ測定は我々の研究室の学生部屋で座った状態で行った。周りの人には静かにしてもらうなどの騒音対策は行っていない。実験は 3 日間に分けて行った。最初に、図 4 に示す頭部状態を被験者に口頭で説明し、すべての頭部状態を再現できるようにする。そしてデバイスを耳に装着してもらい、被験者によって安定する装着方法が異なったため、装着方法は通常の掛け方（普通掛け）と耳の後ろからコードを通し耳上部から装着する方法（耳掛け）の 2 種類から、デバイスが安定している装着方法を被験者に選択してもらった（図 11）。測定手順は、まず被験者に測定する頭部状態を再現させ、維持してもらう。維持状態を確認したら測定信号を再生し、録音を行う。この測定をすべての頭部状態でランダムな順番で行ったものを 1 セットとし、デバイスの着脱をセットごとに行った。測定 1 日目に 4 セット、測定 2 日目に 4 セット、同様に測定 3 日目に 4 セットのデータ測定を行い、各被験者 1,764 個分の外耳道伝達関数（7 外耳道伝達関数 × 21 状態 × 12 セット）を集めた。本実験で使用した測定信号の音量は、使用するイヤホンで音楽を聴いて、適切であった音量レベルに設定した。

5.1.2 実験結果

評価実験で取得したデータを使用して認識精度を確かめた。本研究では各被験者ごとの測定データを用いて、各



(a) 普通掛け (b) 耳掛け

図 11 デバイスの装着方法

Fig. 11 The manner of wearing device.

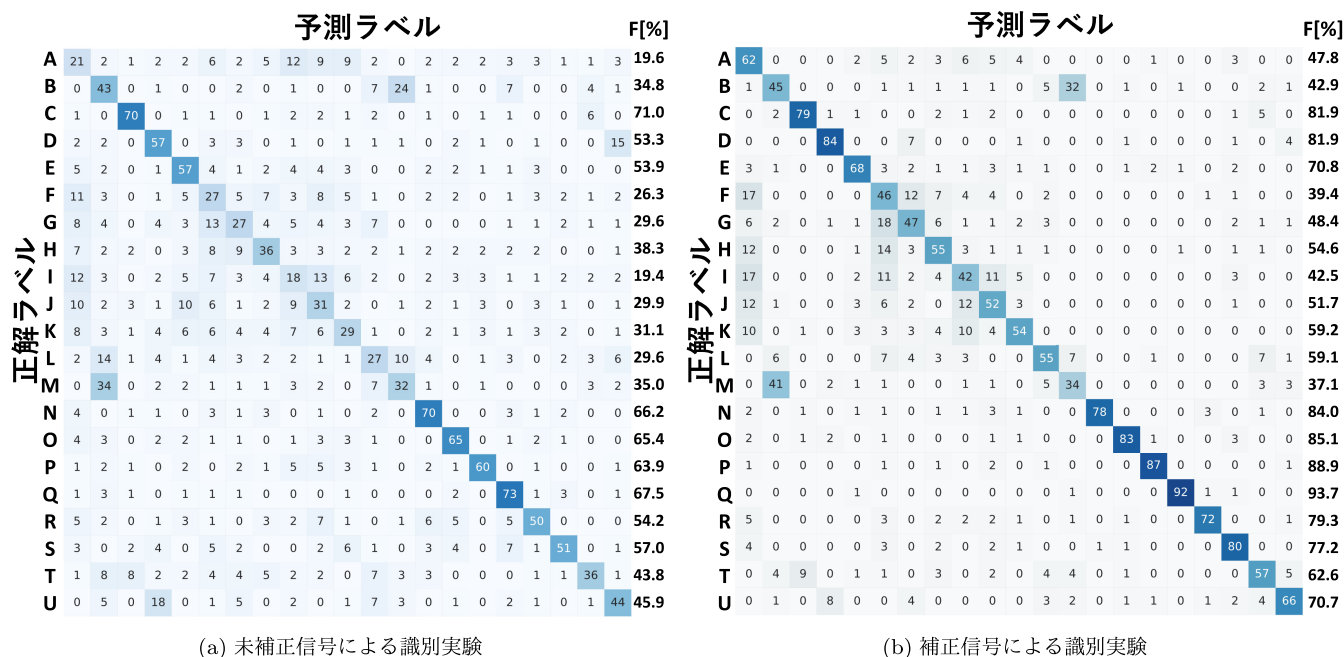


図 12 混同行列 [%]

Fig. 12 Confusion matrix of all facial expressions.

表 1 すべての頭部状態での各被験者の認識精度 [%]

Table 1 Recognition accuracy of all facial expressions [%].

被験者	(a) 未補正信号			(b) 補正信号		
	P ¹	R ²	F ³	P	R	F
P1 ^N	56.1	61.6	58.7	82.5	84.7	83.6
P2 ^N	25.7	37.1	30.4	39.1	41.3	40.2
P3 ^N	48.8	54.5	51.5	59.5	63.0	61.2
P4 ^N	40.9	49.0	44.6	47.2	50.3	48.7
P5 ^N	44.4	50.1	47.1	65.5	69.2	67.3
P6 ^N	48.2	54.0	50.9	51.2	54.5	52.8
P7 ^E	24.0	29.3	26.4	59.4	63.4	61.3
P8 ^N	34.3	40.8	37.3	62.6	69.0	65.6
P9 ^N	25.2	31.4	28.0	67.4	69.2	68.3
P10 ^N	60.3	63.8	62.0	51.2	54.5	52.8
P11 ^N	24.4	29.3	26.6	61.2	64.5	62.8
平均	37.7	43.1	40.2	60.8	64.2	62.5

¹ 適合率 ² 再現率 ³ F 値 ^N 普通掛け ^E 耳掛け

被験者ごとに分類器を作成して、その性能を確かめる。機械学習は Leave-One-Set-Out-CV で行った。訓練データはさらに 5 分割の交差検証を行い、F 値が最も高かったパラメータをハイパーパラメータとした。交差検証に用いたカーネル関数は線形カーネル ($C: 10, 100, 1,000, 10,000$) と RBF カーネル ($C: 10, 100, 1,000, 10,000/\gamma: 0.01, 0.001, 0.0001$) の 2 種類である。未補正信号による識別実験での被験者の認識精度の平均を表 1 (a) にまとめた。未補正信号による識別実験で最も認識精度が高い被験者は F 値 62.0%，低い被験者は F 値 26.4%であった。未補正信号による識別実験での各頭部状態ごとの認識精度を図 12 (a)

にまとめた。混同行列のラベルは図 4 の写真右上のアルファベットに対応し、数値は各行のサンプル数を基準に百分率換算している（小数点以下切り捨て）。最も認識精度の高い頭部状態は「顎を右にずらす」で F 値 71.0%，低い頭部状態は「両目を閉じる」で F 値 19.4%となった。

5.2 補正信号による識別実験

5.2.1 実験環境

補正信号による識別実験では最初に閉口状態にて 3.5 節で述べた測定信号の補正処理を行ってから頭部状態を再現し、維持してもらう。測定信号は各頭部状態の外耳道伝達関数を取得する前に毎回補正を行う。測定 1 日目に 3 セット、測定 2 日目に 3 セット、同様に測定 3 日目に 3 セットのデータ測定を行い、各被験者 1,323 個分の外耳道伝達関数（7 外耳道伝達関数 × 21 状態 × 9 セット）を集めた。それ以外は未補正信号による識別実験と同様である。

5.2.2 実験結果

補正信号による識別実験で取得したデータのみを使用して、未補正信号による識別実験と同様の機械学習を行い、認識精度を確かめた。補正信号による識別実験での被験者の認識精度の平均を表 1 (b) にまとめた。補正信号による識別実験で最も認識精度が高い被験者は F 値 83.6%，低い被験者は F 値 40.2%であった。補正信号による識別実験の方が全体の精度平均は 22.3%（F 値）高かった。補正信号による識別実験での全体の各頭部状態ごとの認識精度を図 12 (b) にまとめた。最も認識精度の高い頭部状態は「上を向く」で F 値 93.7%，低い頭部状態は「舌を出す（開口）」で F 値 37.1%となった。上記の実験より、測定信号の補正

表 2 認識精度の高い頭部状態上位 7 種類 [%]

Table 2 Top 7 facial expressions with regard to recognition accuracy [%].

順位	頭部状態 (F 値)	
	(a) 未補正信号	(b) 補正信号
1	右顎 (71.0)	上向き (93.7)
2	上向き (67.5)	下向き (88.9)
3	右首傾 (66.2)	左首傾 (85.1)
4	左首傾 (65.4)	右首傾 (84.0)
5	下を向く (63.9)	右顎 (81.92)
6	左向き (57.0)	左顎 (81.91)
7	右向き (54.2)	右向き (79.3)

による装着誤差の低減は認識精度を向上させることが確認できた。

5.3 アプリケーションを考慮した頭部状態認識

5.1 節と 5.2 節の実験では、どの頭部状態が提案手法に有効であるかを調査するために 21 種類の頭部状態を調査した。しかし、デバイスの操作や入力インタフェースとしての実用性を考えると認識精度が 62.5% (補正信号による識別実験) では十分とはいえない。そこで、想定する利用シーンと、そのときに必要な頭部状態を絞ることで提案手法の実用性を評価する。利用シーンとしては、ハンズフリーでの操作が求められる音楽プレーヤーやコンテンツリーダーの操作を想定する。5 種類の頭部状態をこれらのアプリケーションのコマンドとして選定し、認識精度の調査を行った。頭部状態の選定にあたっては、認識精度、視線を逸らす必要の有無、公共の場での使用の 3 点を考慮した。表 2 に各評価実験での認識精度の高い上位 7 種類の頭部状態をまとめた。表 2 から上記の選定基準を考慮して、「右顎、左顎、右首傾、左首傾」の 4 種類の頭部状態を選定した。次に図 12 より、B-開口と M-舌開口はそれぞれを混同する傾向があることが分かる。M-舌開口は公共の場での使用が難しいため選定対象から外し、B-開口を最後の 5 種類目の頭部状態とする。その場合、それぞれの頭部状態を混同することがなくなるため、B-開口の認識精度は上がると考えられる。以上の考察より、5 種類「開口、右顎、左顎、右首傾、左首傾」の頭部状態をコマンドとして選定し、アプリケーションの非操作状態として A-閉口状態を加えた 6 種類のデータから再度分類器を作成し、認識精度の調査を行った。

各被験者ごとの認識精度の平均を表 3 にまとめる。未補正信号による取得データでの認識精度は平均 74.4%、補正信号による取得データでの認識精度は平均 90.0%であった。未補正信号下では P10 のみ 85%以上の認識精度を示した。一方で、補正信号下では多くの被験者 (9/11 名) が 85%以上の認識精度を示した。また図 13 に 6 種類の各頭部状態の認識精度をまとめた。未補正信号下では B-開口

表 3 6 種類の頭部状態での各被験者の認識精度 [%]

Table 3 Recognition accuracy of 6 facial expressions [%].

被験者	(a) 未補正信号			(b) 補正信号		
	P ¹	R ²	F ³	P	R	F
P1 ^N	73.4	79.6	76.4	96.5	96.0	96.2
P2 ^N	76.1	80.6	78.3	67.5	72.8	70.0
P3 ^N	79.3	84.7	81.9	94.2	96.0	95.1
P4 ^N	76.5	80.6	78.5	92.8	90.5	91.6
P5 ^N	70.7	74.2	72.4	88.2	90.5	89.3
P6 ^E	75.8	81.0	78.3	78.1	82.3	80.1
P7 ^N	70.5	76.0	73.1	99.5	99.5	99.5
P8 ^N	61.9	70.8	66.1	87.7	90.7	89.2
P9 ^N	60.3	68.8	64.3	92.1	92.9	92.5
P10 ^N	86.7	89.9	88.3	99.6	99.5	99.5
P11 ^N	55.8	61.5	58.5	86.1	87.3	86.7
平均	71.7	77.2	74.4	89.3	90.7	90.0

¹ 適合率 ² 再現率 ³ F 値 ^N 普通掛け ^E 耳掛け

予測ラベル	未補正信号による識別実験						予測ラベル	補正信号による識別実験					
	A	B	C	D	N	O		A	B	C	D	N	O
A	64	5	7	8	9	5	A	93	1	1	0	2	1
B	5	87	0	4	0	1	B	4	87	3	0	1	3
C	5	4	83	0	5	0	C	1	3	90	1	2	1
D	9	6	1	77	1	4	D	4	0	0	91	1	2
N	10	1	4	2	79	0	N	7	2	1	0	89	0
O	11	5	5	3	0	78	O	3	2	1	2	0	90

図 13 6 種類の頭部状態での混同行列

Fig. 13 Confusion matrix of six facial expressions.

以外の頭部状態の認識精度が 85%を下回ったのに対して、補正信号下ではすべての頭部状態が 85%以上の認識精度であった。以上の結果より、補正信号下では多くの被験者が各頭部状態をコマンドとして使える可能性が高いことが分かった。

6. 考察

6.1 各頭部状態の認識精度

矢野らは外耳道音響特性による個人認証の際に音響測定に含まれる計測誤差の原因は

- (a) 背景雑音や電氣的ノイズなどの雑音性誤差
- (b) イヤホンの装着具合などに起因する観測揺らぎ
- (c) ユーザの動きによるアーチファクト
- (d) 温度や気圧などの環境変動

の 4 つに分類されると述べている [17]。本研究では (c) を利用して頭部状態を認識し (a), (b), (d) が認識誤差の原因になると考えられる。(a) による誤差の低減には回路構成などのハードウェアによる対策とノイズ処理などのソフトウェアによる対策を行う必要がある。(d) による誤差に関して、気圧の変化は音速に影響せず無視できると考えられ

る。一方、気温の変化は音速に影響するため、今後の調査が必要であるが、提案手法の信号補正を応用することで認識精度の低下は抑えられると考えられる。最も大きい誤差要因は (b) によるものだと考えられ、その対策として本研究では測定信号に補正を加える手法を提案し、実験を行った。認識精度は向上したため、提案手法は有効だと考えられる。

頭部状態による認識精度の差を図 12 から確認すると、C、D や N-S のような首や顎を動かす動作が含まれる頭部状態は認識精度が高く、F-K のような目や頬の動作が含まれる頭部状態では認識精度が低かった。その理由は動作の大きい頭部状態の方が外耳道の変形がより大きく、周波数応答も大きく変化するため、認識精度が高くなったためだと考えられる。舌を出す（開口）状態の認識精度が低かった理由は開口状態と混同されやすかったためと考えられる。

6.2 測定信号の自動補正

本研究では測定信号の補正を行うことで認識精度の向上を実現したが、条件として一意の頭部状態（本研究では閉口状態）のときに補正を行う必要がある。本研究では閉口状態を確認して手動で補正を行ったが、実用性を考慮すると自動で補正タイミングを検出する必要がある。補正タイミングの自動検出には、現在の頭部状態が閉口状態か非閉口状態であるかを認識する必要がある。その場合、システムの誤認識にはユーザが閉口状態のときに非閉口状態と誤認識してしまう場合と、非閉口状態のときに閉口状態と誤認識してしまう場合の 2 種類の誤認識が存在する。前者に関しての誤認識は、信号の補正が行われただけであり、システムとしては大きな問題にはならない。一方、後者に関しての誤認識は非閉口状態時に信号の補正が行われるため、装着誤差が増える可能性が高く、認識精度を著しく低下させると考えられる。そのため、閉口状態を Positive、非閉口状態を Negative と考えた場合の TNR (True Negative Rate) は 100% に近い精度であることが求められる。

実際に実用を考慮した 6 種類の頭部状態に対して閉口状態と非閉口状態の 2 グループに分割し、再度認識精度の調査を行った。調査には 5.3 節と同じ分類器と特徴量を用いた。調査の結果、TPR (True Positive Rate) は 79.7%、TNR は 93.7% であった。TPR は、適切な頻度で信号の補正を行うために十分な認識精度と考えられる。TNR は 90% 以上ではあるが、経時的に補正判定を行う場合、誤った補正を行う可能性が高くなる。たとえば、非閉口状態が 7 秒続く場合、4.2 節より認識は 0.68 秒ごとに行われるため、補正判定は 10 回発生する。10 回すべてで正しい判定を行う確率は約 52.2% であり、連続する非閉口状態の際には誤った補正を行う可能性が高い。測定信号を自動で補正するためには、TNR を向上させる必要がある。また、信号の誤補正が起こった場合の適切な対処手法も考察する必要がある。

6.3 実用環境での認識精度

本実験では座位状態にて各被験者の外耳道伝達関数を取得した。今後は実用を考慮し、歩行時や音楽鑑賞時に外耳道伝達関数を取得し、認識精度の調査を行う必要がある。

また、今回の実験では複数の頭部状態を組み合わせた状態や詳細な頭部状態などを認識していない。たとえば、口を開けたまま右を向くなどの状態や、多段階の口の開け方の認識も頭部状態と考え得る。これらの詳細な調査を進め、実環境での提案手法の有効性を確認する必要がある。さらに、実環境ではデバイスのコマンドに使用する頭部状態を日常生活でも使用するため、デバイスの非操作モードと操作モードを切り替える必要がある。コマンドモードを切り替える専用のコマンドとして、歯同士の接触音や複数の頭部状態を組み合わせるなどの日常生活では意図せず出にくいものを採用することを考えている。

また、4.2 節で述べたように提案手法の 1 回の認識にかかる時間は約 0.68 秒であるが、ユーザが実際に頭部状態をコマンドとして使用する場合の適切な反応時間やユーザビリティについて調査を行う必要がある。

7. 結論

本研究では、頭部状態の変化にともなって外耳道の形状が変化することに着目し、外耳道伝達関数を測定することで頭部状態を認識する手法を提案した。さらに測定信号補正手法によって認識率の向上を実現した。プロトタイプデバイスを実装し、評価実験を行った結果、未補正信号による識別実験で 40.2% (F 値)、補正信号による識別実験で 62.5% (F 値) の精度で認識できることを確認した。また、実際の利用シーンを考慮し、認識する頭部状態を 6 種類に限定したところ、未補正信号による識別実験で 74.4% (F 値)、補正信号による識別実験で 90.0% (F 値) の認識精度が得られた。

謝辞 本研究は JSPS 科研費 JP18K18084, JP19H04222 の助成を受けたものである。

参考文献

- [1] 古谷 聡, 越仲孝文, 大杉孝司: ヒアラブル技術によるヒューマン系 IoT ソリューションの取り組みと展望 (デジタルビジネスを支える IoT 特集) - (お客様に価値を提供する IoT ソリューション), NEC 技報 = NEC technical journal, Vol.70, No.1, pp.47–51 (2017).
- [2] AirPodsPro: Apple (2019), available from (<https://www.apple.com/jp/airpods-pro/>).
- [3] DashPro: Bragi (2014), available from (<https://www.bragi.com/>).
- [4] Ando, T., Kubo, Y., Shizuki, B. and Takahashi, S.: CanalSense: Face-Related Movement Recognition System Based on Sensing Air Pressure in Ear Canals, *Proc. 30th Annual ACM Symposium on User Interface Software and Technology, UIST '17*, pp.679–689, ACM (2017).
- [5] 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, *Proc. 2017 CHI Conference on Human Factors in Computing Systems, CHI '17*, pp.1911–1922, ACM (2017).
- [6] 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).
- [7] 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 Proc. 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proc. 2015 ACM International Symposium on Wearable Computers, UbiComp/ISWC'15 Adjunct*, pp.1333–1338, ACM (2015).
- [8] inCore: ナップエンタープライズ (1983), 入手先 <http://www.incore.jp/>.
- [9] h.ear in NC: SONY (2016), available from <https://www.sony.jp/headphone/products/MDR-EX750NA/>.
- [10] Amesaka, T., Watanabe, H. and Sugimoto, M.: Facial Expression Recognition Using Ear Canal Transfer Function, *Proc. 23rd International Symposium on Wearable Computers, ISWC'19*, pp.1–9 (2019).
- [11] 真鍋宏幸, 福本雅朗: Headphone Taps: 通常のヘッドホンへのタップ入力, 情報処理学会論文誌, Vol.55, No.4, pp.1334–1343 (2014).
- [12] Laput, G., Chen, X.A. and Harrison, C.: SweepSense: Ad Hoc Configuration Sensing Using Reflected Swept-Frequency Ultrasonics, *Proc. 21st International Conference on Intelligent User Interfaces, IUI '16*, pp.332–335, ACM (2016).
- [13] Hiipakka, M.: Measurement apparatus and modelling techniques of ear canal acoustics, *Espoo: Helsinki University of Technology* (2008).
- [14] 佐藤史明: Swept-Sine 法に基づく音響伝播測定, 音響学会誌, Vol.63, No.6, pp.322–327 (2007).
- [15] Akkermans, T.H.M., Kevenaar, T.A.M. and Schobben, D.W.E.: Acoustic Ear Recognition, *Advances in Biometrics*, Zhang, D. and Jain, A.K. (Eds.), pp.697–705, Springer Berlin Heidelberg (2005).
- [16] Arakawa, T., Koshinaka, T., Yano, S., Irisawa, H., Miyahara, R. and Imaoka, H.: Fast and accurate personal authentication using ear acoustics, *Proc. Asia-Pacific Signal and Information Processing Association Annual Summit and Conference, APSIPA '16*, pp.1–4, IEEE (2016).
- [17] Yano, S., Arakawa, T., Koshinaka, T., Imaoka, H. and Irisawa, H.: Improving Acoustic Ear Recognition Accuracy for Personal Identification by Averaging Biometric Data and Spreading Measurement Error over a Wide Frequency Range, *IEICE Trans. Electronics*, Vol.J100-A, pp.161–168 (2017).
- [18] Borish, J. and Angell, J.B.: An Efficient Algorithm for Measuring the Impulse Response Using Pseudorandom Noise, *J. Audio Eng. Soc.*, Vol.31, No.7/8, pp.478–488 (1983).
- [19] Tan, J., Wang, X., Nguyen, C.-T. and Shi, Y.: SilentKey: A New Authentication Framework Through Ultrasonic-based Lip Reading, *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, Vol.2, No.1, pp.36:1–36:18 (2018).
- [20] Watanabe, H., Terada, T. and Tsukamoto, M.: Gesture Recognition Method Utilizing Ultrasonic Active Acoustic Sensing, *Journal of Information Processing*, Vol.25, pp.331–340 (2017).
- [21] 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, *Proc. 2013 ACM International Conference on Interactive Tabletops and Surfaces, ITS '13*, pp.189–198 (2013).
- [22] Takemura, K., Ito, A., Takamatsu, J. and Ogasawara, T.: Active Bone-conducted Sound Sensing for Wearable Interfaces, *Proc. 24th Annual ACM Symposium Adjunct on User Interface Software and Technology, UIST '11 Adjunct*, pp.53–54, ACM (2011).
- [23] 祁 君容: 顎運動時に起こる外耳道のひずみと下顎頭運動の相関関係, 博士論文, 松本歯科大学 (2016).
- [24] 橘 秀樹, 矢野博夫: 環境騒音・建築音響の測定, chapter 5.2 インパルス応答の測定方法, コロナ社 (2004).
- [25] Lei, H. and Gonzalo, E.L.: Mel, linear, and antmel frequency cepstral coefficients in broad phonetic regions for telephone speaker recognition, *INTERSPEECH 2009, 10th Annual Conference of the International Speech Communication Association, Brighton, United Kingdom, September 6–10, 2009, ISCA 2009*, pp.2323–2326 (2009).


雨坂 宇宙 (学生会員)

2018 年北海道大学工学部情報エレクトロニクス学科卒業。2020 年同大学大学院情報科学研究科修士課程卒業。同年筑波大学理工情報生命学術院システム情報工学研究群博士後期課程入学, 現在に至る。


渡邊 拓貴 (正会員)

2012 年神戸大学工学部電気電子工学科卒業。2017 年同大学大学院工学研究科博士後期課程修了。博士 (工学)。同年より北海道大学大学院情報科学研究科助教, 現在に至る。


杉本 雅則 (正会員)

1990 年東京大学工学部航空学科卒業。1995 年同大学大学院工学系研究科博士課程修了。博士 (工学)。同年学術情報センター (現, 国立情報学研究所) 助手。1999 年東京大学助教授。2012 年北海道大学教授となり, 現在に至る。