Vol.2019-MBL-90 No.7
Vol.2019-UBI-61 No.7
2019/3/4

外耳道伝達関数を用いた頭部状態認識手法

雨坂 宇宙[1,a)]　渡邉 拓貴[1,b)]　杉本 雅則[1,c)]

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一つとして，手や視界を占有することのないデバイス操作機能が挙げられる．既存の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，頑健性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) に伴って外耳道が変形することに着目し，外耳道インパルス応答を測定することで頭部状態を認識する手法を提案した．提案手法ではマイク内蔵型のイヤホンでの利用を想定し，イヤホンから測定信号を発信する．反射音をマイクで取得し解析することで，外耳道伝達関数を求める．外耳道伝達関数から特徴量として LFCC を抽出し，サポートベクタマシンにて分類器を作成した．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率の向上を実現した．6 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率は未補正時で 47.2%(F 値)，補正時で 58.9%(F 値) となった．また，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なった結果，補正時で 86.6%(F 値) の認識精度が得られた．

キーワード：頭部状態認識，ジェスチャ認識，ウェアラブルコンピューティング，外耳道インパルス応答

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伴い，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バイオメトリクスや行動，状態を認識することができる．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，デバイスを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，デバイスの操作時に視界や片手が占有されることは避けるのが望ましい．

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 [4–7]．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 (図 1) 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．頭部状態の変化に伴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で認識することで現在のユーザの頭部状態を認識する．本研究で利用するセンサはマイクとスピーカだけであり本来のイヤホンの用途で使われるセンサのみで構成されている．市販の製品で外耳道内部の音を録音できるイヤホンも存在するため，ジェスチャ認識のための追加デバイスが必要ない [8, 9]．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インパルス応答の取得と頭部

[1] 北海道大学大学院情報科学研究科
[a)] amesaka@ist.hokudai.ac.jp
[b)] hiroki.watanabe@ist.hokudai.ac.jp
[c)] sugi@ist.hokudai.ac.jp

状態認識システムの提案

- 6 人の被験者に対して，被験者ごとに 21 種類の頭部状態を未補正下で平均 47.2% (F 値)，補正下で平均 58.9% (F 値) の精度で認識
- 6 人の被験者に対して，被験者ごとに音楽プレーヤの操作を考慮した 6 種類の頭部状態を補正下で平均 86.6% (F 値) の精度で認識

2. 関連研究

2.1 ヒアラブルデバイスを用いた研究

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．安藤ら [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，ユーザ毎に 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．更に，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．Denys ら [5] は電極を用いて外耳道内部から筋肉の動きを読み取り 5 種類の表情を精度 87.5%で認識することに成功しており，また言葉や目の動きなどの細かな顔の動きの認識を試みている．谷口ら [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．音波を使ったヒアラブルデバイスの例として，矢野ら [10] は外耳道内のインパルス応答を求め個人認証を試み，分類誤差 1%以下を実現している．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．

上述した既存研究は気圧，電極，光，慣性センサを用いて表情認識を行っている．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なっているが，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表情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，マイク内蔵型のスピーカ (イヤホン) の利用を想定するため，追加のセンサが必要ない．

2.2 人体へ音波を適用した研究

Tan ら [13] はスマートフォン本体から発する音声信号を用いて口周辺部からユーザの発話パターンを読み取り口の動きだけで個人認証を行った．渡邉ら [14] 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．また，竹村ら [15] は骨伝導マイクを用いて食事の咀嚼回数の計測を行っている．

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捉え，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．

2.3 外耳道インパルス応答の測定

Hiipakka ら [16] は外耳道の音響特性などを調査しており，Swept-Sine 信号 [17] 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 [18] はプローブ信号より外耳道のインパルス応答を求め，伝達関数より外耳道の個人性を見出している．更に矢野ら [19] は MLS 法（Maximum Length Sequence）[20] を用いて外耳道の個人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．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う．また，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具合の誤差を低減している．

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．ユーザは外耳道からの音を録音できるように設計したマイク付きイヤホンを装着する．イヤホンからは測定信号を再生し，インパルス応答を測定する．図 2 に示すように，ユーザが顔の向きや表情を変えることで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し，インパルス応答のフーリエ変換 (伝達関数) も変化する．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，現在の頭部状態を推測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溢れており，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．本研究で使用した超音波の音量は，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．

3.2 認識原理

祁ら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 [21]．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．よって今回の実験では，顎・首・眼周辺の運動を基に 21 種類の頭部状態 (図 4) を定義し，認識を行う．

Vol.2019-MBL-90 No.7
Vol.2019-UBI-61 No.7
2019/3/4

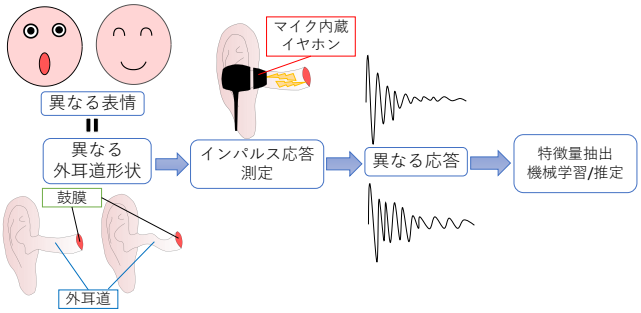


図 1　システム構成図

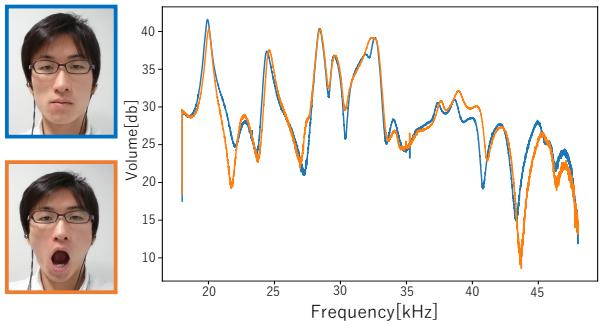


図 2　振幅スペクトルの変化

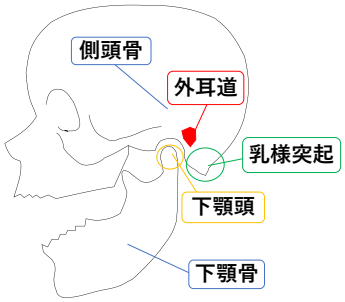


図 3　顎の構造

3.3　インパルス応答測定原理

測定信号を用いたインパルス応答の測定原理を図 5 に示す．図 5 中の線形系 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時の出力は，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．この測定方法で得られる $H(k)$ を逆フーリエ変換すると，インパルス応答 $h(n)$ を得ることができる．本研究では信号合成フィルタで Swept-Sine 信号を作成し，外耳道を線形系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インパルス応答を測定する．

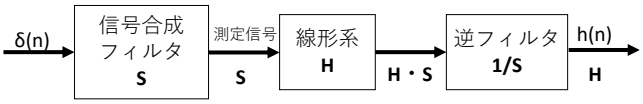


図 5　インパルス応答測定原理

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) [17] 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．インパルス応答は全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，不快となりうる．そこで，本研究では測定信号の帯域制限を行う [22]．サンプリングレートを f_s として 0Hz から f_0Hz の帯域制限は $0 \leq k \leq (f_0 N/f_s)$ 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は厳密なインパルス応答とはならないが，便宜上，本論文ではインパルス応答と表記する．

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通りである．

(1) Swept-Sine 信号を同期加算の回数分だけ隙間なく並べ，1 周期分の無音区間を末尾に追加する (図 6(b))．
(2) 測定信号の出力結果 (図 6(c)) を Swept-Sine 信号長で切り出し加算平均する (図 6(d))
(3) 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
(4) 演算結果を逆フーリエ変換することでインパルス応答を得る (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インパルス応答はデバイスの着脱や時間経過に伴う装着具合の差異によっても変化するため，同じ頭部状態でも測定ごとに応答が異なる．認識率の向上のためには装着具合による誤差が小さいほうが望ましい．そこで閉口状態時に測定信号を補正することで，装着

図 4　頭部状態

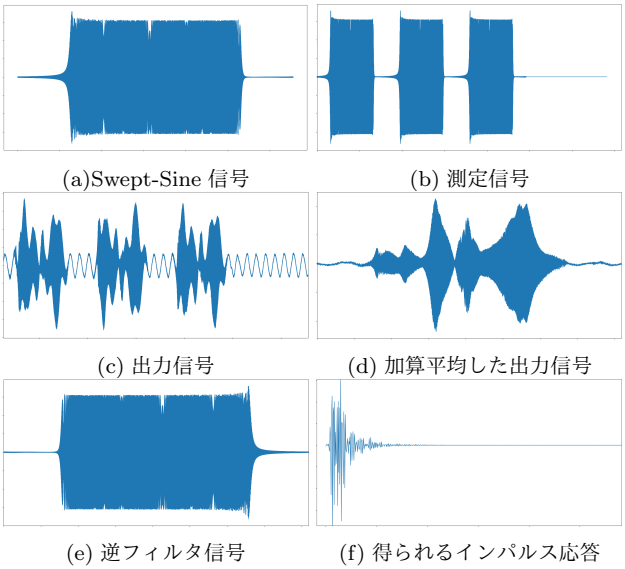


図 6　円状畳み込みの原理を用いたインパルス応答測定

具合による誤差を低減する手法を提案する．閉口状態時に得られるインパルス応答の伝達関数とその時の測定信号，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．

$$X(k)H(k) = Y(k) \tag{1}$$

3.4 節冒頭で述べた通りデバイスの装着具合の誤差によって $H(k)$ は変化する．装着具合の誤差を含んだ場合の閉口状態の伝達関数を $H'(k)$ 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

$$X(k)H'(k) = Y'(k) \tag{2}$$

と表せる．この時，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 (1)(2) より

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

と表せる．$Y(k)$ を一意に設定し，閉口状態時に (3) による信号の補正を与えることで装着具合の誤差が低減でき，頭部状態の変化によるインパルス応答の変化を効率的に測定できると考えられる．

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ため，特徴量抽出を行う．今回は特徴量として人間の音声認識によく用いられる MFCC (Mel-Frequency Cepstrum Coefficients) の線形バージョンである LFCC (Linear-Frequency Cepstrum Coefficients) [23] 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応するが，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．本研究では，分類器に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インパルス応答から頭部状態を予測する．

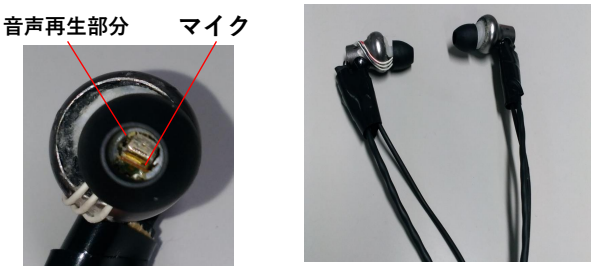
Vol.2019-MBL-90 No.7
Vol.2019-UBI-61 No.7
2019/3/4

4. 実装

今回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 [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:20–48kHz)，マイクに knowles の SPU0410LR5H (録音周波数帯域:100–80kHz) を用いた．マイクは反響音を効率よ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 7(a))．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．完成したデバイスが図 7(b) となる．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，録音共にサンプリングレートは 96kHz で行う．再生・録音デバイスの構成図を図 8 にまとめた．



(a) 音声再生・録音部分　　(b) デバイス全体図

図 7 実験デバイス

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 18kHz 以下を帯域制限した，18kHz から 48kHz のアップスイープ信号 (16,384 サンプル) を使用する．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 回とする．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 個のインパルス応答を取得する．3.4 節の手順で得られるインパルス応答は Swept-Sine 信号長と同じ 16,348 サンプルとなるので LFCC 抽出時のフーリエ変換は 16,348 ポイントで行い，線形フィルタバンクは 20 分割で行う．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，インパルス応答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外耳道伝達関数の平均値とした．インパルス応答測定・特徴量抽出・機械学習は Python 3.6 で実装した．

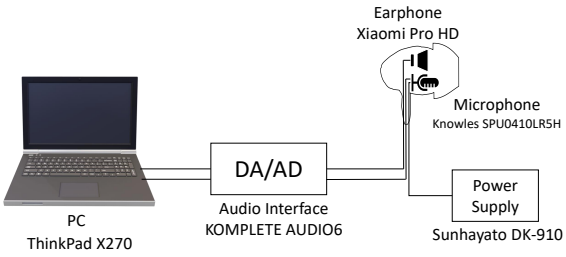


図 8 デバイス構成図

5. 評価実験

5.1 未補正実験

実験はボランティアで 6 人の 23–25 歳の男性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，全ての頭部状態を再現できるようにする．そして実験デバイスを耳に装着させる．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う (図 9)．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．この測定を全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 12 セット分のデータ (7 インパルス応答 × 21 状態 × 12 セット) を集めた．

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．本論文では各被験者ごとにそれぞれの分類器を作成して，その性能を確かめる．機械学習はセットごとの交差検証を用いた．訓練データはさらに 5 分割の交差検証を行い F 値に基づくハイパラメータを求めた．未補正実験での

被験者の認識精度の平均を表 1(a) にまとめた．未補正実験で最も認識精度が高い被験者は F 値 58.7%, 低い被験者は F 値 30.4%であった．未補正実験での各頭部状態毎の認識精度を図 10(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，各行ごとに正規化を行った．最も認識精度の高い頭部状態は「顎を左にずらす」で F 値 77.2%，低い頭部状態は「両目を閉じる 」で F 値 17.4%となった．

(a) 普通掛け　(b) 耳掛け

図 9　デバイスの装着方法

5.2　補正実験

補正実験では最初に閉口状態にて 3.5 節で説明し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 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 9 セット分のデータ (7 インパルス応答 × 21 状態 × 9 セット) を集めた．それ以外は未補正実験と同様である．

取得したデータを使用して，補正実験と同様の処理を行い認識精度を確かめた．補正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．補正実験の方が全体の精度平均は 11.7%(F 値) 高かった．補正実験での全体の各頭部状態毎の認識精度を図 10(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 90.4%，低い頭部状態は「舌を出す（開口） 」で F 値 33.5%となった．

表 1　認識精度の平均 [%](N:普通掛け,E:耳掛け)

	(a) 未補正実験			(b) 補正実験		
被験者	適合率	再現率	F 値	適合率	再現率	F 値
A (N)	56.1	61.6	58.7	82.5	84.7	83.6
B (N)	25.7	37.1	30.4	39.1	41.3	40.2
C (N)	48.8	54.5	51.5	59.5	63.0	61.2
D (N)	40.9	49.0	44.6	47.2	50.3	48.7
E (N)	44.4	50.1	47.1	65.5	69.2	67.3
F (E)	48.2	54.0	50.9	51.2	54.5	52.8
平均	44.0	51.1	47.2	57.5	60.5	58.9

5.3　アプリケーションを考慮した頭部状態認識

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率が 58.9%（補正実験）では十分とはいえないが，認識する頭部状態と利用シーンを絞ることで提案手法の実用性を評価する．今回は音楽プレーヤの操作を考える．一般的な音楽プレーヤの操作には音楽の再生/一時停止・早送り・早戻し・次の曲へ移動・前の曲へ移動・音量を上げる・音量を下げる操作の 7 種類が存在する．早送りと次の曲へ移動，早戻しと前の曲へ移動の操作をそれぞれ同じ頭部状態で割り当て，状態が維持された場合は早送り又は早戻し操作を行い，状態が維持されなかった場合は次の曲へ移動又は前の曲へ移動操作を行うとして 5 種類の頭部状態を選定し，実用性を確認する．頭部状態の覚えやすさや見た目の奇異さから舌を出すなどの状態は選定から外した．また視界の影響を考慮し，目を閉じる状態や顔を上下左右に向ける状態は選定から外した．認識精度を考慮して，次の 5 種類「開口，右顎，左顎，右首傾，左首傾」の頭部状態を選定し，非操作状態として閉口状態を加えた 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．各被験者毎の認識精度の平均と標準偏差を表 2 にまとめる．全体の認識率として F 値が 86.6%となった．被験者 B に関しては 66.8%と実用的な精度とはいえないが，他の被験者に関しては 80%以上の認識精度を実現しており，実用的な精度であるといえる．

表 2　音楽プレーヤ操作を想定した認識精度の平均 (標準偏差)[%]

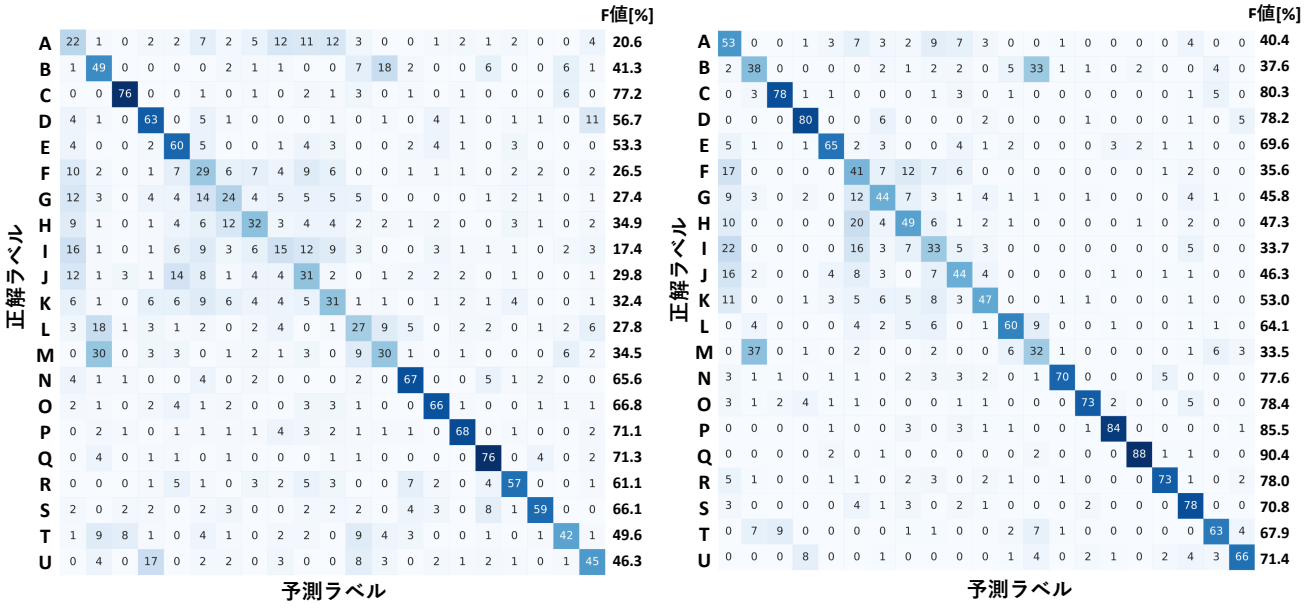
被験者	適合率	再現率	F 値
A	96.5 (4.9)	96.0 (5.8)	96.2 (5.3)
B	64.1 (7.0)	69.8 (4.8)	66.8 (5.7)
C	94.2 (4.8)	96.0 (5.1)	95.1 (4.9)
D	92.8 (6.0)	90.5 (15.3)	91.6 (8.6)
E	88.2 (7.8)	90.5 (8.5)	89.3 (8.1)
F	78.1 (11.5)	82.3 (9.5)	80.1 (10.4)
平均	85.7 (7.0)	87.5 (8.2)	86.6 (7.5)

6.　考察

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). 背景雑音や電気的ノイズ等の雑音性誤差
b). イヤホンの装着具合等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). 温度や気圧等の環境変動

の 4 つに分類されると述べている [10]．今回の研究では (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) による誤差の低減は対策が難しいが，利用シーンを考慮すれば影響は限



(a) 未補正実験　　(b) 補正実験

図 10　混同行列

定的と考える．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，本提案手法は有効であると考えられる．実際に被験者 A の 1 セット目から 7 セット目までの閉口状態の未補正時 (図 11) と補正時 (図 12) の振幅スペクトルを示す．図 12 を見ると，音量の大小はあるが，図 11 よりも概形は同じ形になっている．特徴量抽出時には正規化を行うため，グラフの概形が同じであれば，音量の大小は問題ないといえる．また被験者 A の 1 セット目から 7 セット目までの閉口状態の LFCC の分散を確認した．未補正時では各次元の分散の平均が 5.1×10^{-2} であったが，補正時には 6.2×10^{-3} に減少していることが確認できた．

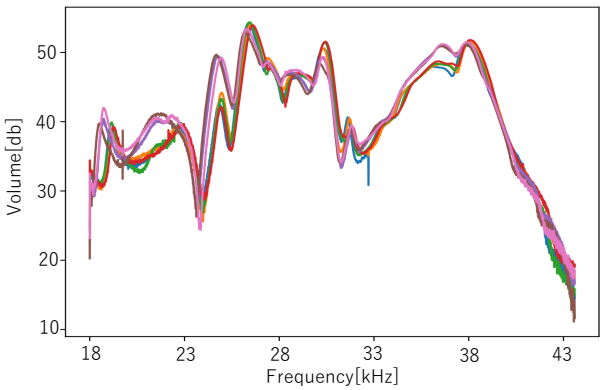


図 11　デバイス着脱による振幅スペクトルの変化（未補正）

認識精度は大きな個人差があり，補正実験では最大 43.4%（F 値）の差があった．この原因は個人によって頭部状態の変化が外耳道の変形に影響しにくいためである．認識精度の低い被験者 B の各頭部状態における振幅スペクトルの

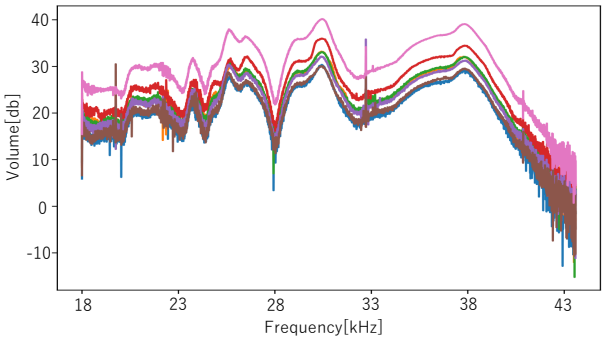


図 12　デバイス着脱による振幅スペクトルの変化（補正）

変化を観察したところ，変化が小さいことが確認できた．頭部状態による認識精度の差を図 10 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．

本研究では，超音波領域の音を利用するため自作のデバイスを利用した．今後の機器の発展により，マイク内蔵型のイヤホンを用いて，追加のデバイスなく頭部状態を認識できると考えられる．

今後の課題として，本研究では測定信号の補正を行うことで認識率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）の時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．

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おらず，例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を行っていない．これらの詳細な調査を進め，デバイスのコマンド操作だけでなくユーザの行動記録が可能であるか考える必要がある．

アプリケーションを考慮した実験において，本研究では認識精度，頭部状態の覚えやすさ，見た目の奇異さを考慮して，利用する頭部状態を選択した．実用を考えたときにどのような頭部状態を採用するべきかなどのユーザビリティの調査も，実際のアプリケーションを作成して行う必要がある．

7. 結論

本研究では，頭部状態の変化に伴って外耳道の形状が変化することに着目し，外耳道内のインパルス応答を測定することで頭部状態を認識する手法を提案した．更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，評価実験を行なった結果，未補正実験で 47.2%(F 値)，補正実験で 58.9%(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，86.6%(F 値) の認識率を得られた．

謝辞 本研究は JSPS 科研費 JP18K18084 の助成を受けたものです．

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] AirPods: Apple, https://www.apple.com/jp/airpods/.
[3] DashPro: Bragi, 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, *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology*, UIST '17, pp. 679–689 (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, *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, CHI '17, pp. 1911–1922 (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 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, pp. 1333–1338 (2015).
[8] inCore: ナップエンタープライズ, http://www.incore.jp/.
[9] h.ear in NC: SONY, https://www.sony.jp/headphone/products/MDR-EX750NA/.
[10] 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 Errorsover a Wide Frequency Range, *IEICE Transactions on Electronics*, Vol. J100-A, pp. 161–168 (2017).
[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, *Proceedings of the 21st International Conference on Intelligent User Interfaces*, IUI '16, pp. 332–335 (2016).
[13] 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).
[14] Watanabe, H., Terada, T. and Tsukamoto, M.: Gesture Recognition Method Based on Ultrasound Propagation in Body, *Proceedings of the 13th International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services*, MOBIQUITOUS 2016, pp. 288–289 (2016).
[15] 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, pp. 53–54 (2011).
[16] Hiipakka, M.: Measurement apparatus and modelling techniques of ear canal acoustics, *Espoo: Helsinki University of Technology* (2008).
[17] 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).
[18] 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.), Berlin, Heidelberg, Springer Berlin Heidelberg, pp. 697–705 (2005).
[19] Yano, S., Hokari, H. and Shimada, S.: A Study on Personal Difference in the Transfer Functions of Sound Localization Using Stereo Earphones, *Audio Engineering Society Convention 106* (1999).
[20] 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).
[21] 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).
[22] 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.
[23] Lei, H. and Gonzalo, E. L.: Mel, linear, and antimel 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).