

UltraSoundLog: 超音波IDによる音声ログへの位置情報付与

渡邊拓貴^{1,a)} 寺田 努^{1,2,b)} 塚本昌彦^{1,c)}

概要: ウェアラブルコンピューティング環境では、ユーザが装着している各種センサを用いてユーザの行動や状況をデータとして保存するライフログが可能になる。ライフログの一つに音声ログがあるが、長時間の音声ログはどこにどんなデータがあるのか判別できないため、有効に利用するためには音声データに対して適切にタグをつける必要がある。筆者らは先行研究において、超音波IDを音声ログに埋め込むことで音声データにタグ付けを行う手法を提案した。本研究では、この手法の実環境での使用について報告し、そこで得られた問題点である、録音デバイスの装着位置が認識精度に影響を及ぼす問題を解決する手法を提案する。具体的には、超音波スイープ信号の周波数応答により録音デバイスの装着位置を認識し、適切なしきい値での認識処理を施す。録音デバイス装着位置の認識率は平均 64.0%であり、装着位置によって認識する際のしきい値を最適なものに設定する処理を行ったところ、認識精度に改善が見られた。

UltraSoundLog: A Method for Embedding Location Information to Sound Log by Using Ultrasonic ID

WATANABE HIROKI^{1,a)} TERADA TSUTOMU^{1,2,b)} TSUKAMOTO MASAHICO^{1,c)}

Abstract: In wearable computing environments, various sensors enable to record user context as lifelog. Using sound-based lifelog, we have to attach tags to voice log appropriately since it is hard for us to extract the intended data from long-term sound log. We have proposed a method for tagging sound data by embedding ultrasonic ID. In this paper, we report the result of using our method in a real event, and we propose a method to resolve the issue that the worn position of a microphone affects the recognition accuracy. Our method recognizes a worn position by frequency response of ultrasonic sweep signals, and changes threshold for ID recognition for each positions. The accuracy of position recognition was 64.0%, and there were improvements in ID recognition using our method.

1. はじめに

近年のコンピュータ小型化に伴い、ユーザがコンピュータを常に身につけて生活するウェアラブルコンピューティングに対する注目が集まっている。ウェアラブルコンピューティング環境では、ユーザが装着している各種センサを用

いてユーザの行動や状況をデジタルデータとして保存するライフログが可能になる^{*1}^{*2}。これらのデバイスはユーザの活動量を記録し、解析したデータを健康管理などに用いている。ライフログは健康管理のみでなく、思い出を残すためや、自分の体験を他人と共有するため、行動パターンの解析から未来予測するなど、様々な利用の可能性が考えられる。一般的にライフログに利用されるセンサは加速度センサやカメラであり、加速度センサではユーザ自身の行動は取得できるが、ユーザの周囲の状況を認識することは

¹ 神戸大学大学院工学研究科
Graduate School of Engineering, Kobe University, 1-1
Rokkodai, Nada Kobe, Hyogo 657-8501, Japan

² 科学技術振興機構さきがけ
PRESTO, Japan Science and Technology Agency, 7 Goban-
cho, Chiyoda, Tokyo, 102-0076, Japan

a) hiroki.watanabe@stu.kobe-u.ac.jp

b) tsutomu@eedept.kobe-u.ac.jp

c) tuka@kobe-u.ac.jp

^{*1} Nike FuelBand: http://www.nike.com/jp/ja_jp/c/nikeplus-fuelband

^{*2} Up by Jawbone: <https://jawbone.com/up>

できない。また、Google Glass^{*3}のようなメガネ型のデバイスを用いれば、常時録画が可能であるが、カメラを常に人に向けることになり、プライバシーの問題が大きく、受け入れにはまだ時間を要すると考えられる。一方、音声だけでも人の会話、周囲の環境音のみならず、場の雰囲気なども記録することが可能である。このように音声ログはライフログとして有用であるといえるが、長時間記録した音声ログは、どこにどんなデータがあるのか把握できないため、適切にタグをつける必要がある。音声ログへのタグ付けとして一般的には、録音中に手動でタグをつけるものや、録音したデータを音響特徴量により解析することが考えられる。しかし、前者はタグ付けの操作が可能の場合には有効であるが、そういった操作が行えない状態(手が塞がっている、レコーダやスマートフォンを操作しにくい場面)や、単純に手動でのタグ付けを忘れてしまう場合が考えられる。後者は、様々な場面の学習データが十分に集まっている場合には、ユーザ周囲状況のタグ付けが可能であるが、ユーザ自身の詳細な動きや音の出にくい動きなどを認識することは困難である。またマイクと、加速度センサやBluetooth, NFCなどの複数の情報を組み合わせてタグ付けする手法も考えられるが、複数種類のデータを取り扱う必要があり、データ処理が煩雑になることが予想される。

本研究の目的は、あらゆるユーザ状況を超音波化し、音声情報に一元化して保存する、コンテキストの超音波化技術の確立である。この技術により、あらゆるユーザ状況が音声ファイル一つに記録でき、シンプルなデータ処理が可能になる。筆者らの先行研究[1]では、超音波を用いて音声へユーザの動きと超音波IDを埋め込む手法を提案した。提案手法では、会話音等の環境音、ジェスチャ、ユーザの居た場所、周囲に居た人のデータすべてがマイクのみで記録でき、録音した音声データファイル一つを解析するだけで、これらのユーザ状況が識別できる。スピーカは超音波を出しているだけでよく、またマイクは録音しているだけでよいので、データ通信のための無線/有線通信が必要ない。本研究では、コマーシャルデバイスにおいても使用できるシステム構築を目的としているが、先行研究[1]では、専用デバイスを用いていたため、コマーシャルデバイスで考え得る個々のデバイスの特性、装着位置が不定であること、電力消費等の問題点について考慮していなかった。そこで本論文では、装着位置の問題に着目し、実環境から得られたデータの分析に基づき、録音デバイスの装着位置を考慮したID認識を行うシステムを提案する。

なお、一般的には20kHz以上が超音波と呼ばれているが、本論文では成人には聞こえにくい周波数(18kHz以上)も超音波と定義する。

以降、2章では本研究に関連する研究について述べ、3章

で超音波IDについて述べる。4章で実イベントでの実験について示し、5章でシステムの改良を述べる。最後に6章でまとめを行う。

2. 関連研究

音声ログへのタグ付けには様々な手法が考えられる。録音データを音響解析し、音響特徴量により環境を推定する研究は多く行われている[2], [3], [4]。これらの研究では、MFCC(Mel Frequency Cepstral Coefficient)など音声認識で使用される音響特徴量のみを用いた認識という点で本研究と異なる。大内らの研究[5]では、加速度センサとマイクを組み合わせた行動認識が提案されている。加速度センサにより作業状態と認識されたときのみマイクを起動し、作業状態分類を行う計算量を抑えた手法をとっている。Choudhuryらの研究[6]では、人間関係を取得するために、赤外線センサとトランスミッタ、マイク、加速度センサを搭載したウェアラブルデバイスを開発し、センサデバイスでの人間関係取得可能性を議論している。近距離の端末同士の相対位置関係や近接性を取得する方法としてBluetoothを用いた研究も行われている[7], [8]。しかしこれらの手法を本研究に適用すると、音声に加えて他の種類のデータもタイムスタンプとともに取得し、それを後で音声と同期する必要がある。本研究ではこのデータを超音波で取得しているため、音声ファイル一つにすべてのデータが集約され、同期をとる必要がない。また、マイクで全てのデータを取得できるので追加のハードウェアが必要ない。本研究の目指すシンプルなファイル管理/データ処理という観点からこれらの手法は本研究には適していない。

超音波を用いてユーザの位置を追跡する手法は複数提案されている[9], [10]。ユーザの位置は、ユーザと環境側の決まったところに設置された受信機または送信機間とのドップラー効果や音の到達時間を利用して認識されている。この手法を用いればユーザが部屋のどの位置に居たのか正確に認識することができるが、環境側に正確にデバイスを設置する必要があり、設置コストが大きいと考えられる。本研究では小型の超音波スピーカを認識したい場所に置いておくだけでよいという環境を目標としている。

音波で情報通信するものとしては可聴域に音声を埋め込む、音響透かし技術[11]が用いられる。また、超音波帯域で情報通信を行う手法も存在し、様々な場面で用いられている。古くはテレビのリモコンの情報通信手段として用いられていた[12]。近年においては、NTTドコモが人間の耳には聞こえない高周波を用いたチェックイン機能Air Stamp^{*4}を提供している。また、他にもヤマハ株式会社のINFOSOUND^{*5}、Digital Solution Laboratoryの

^{*4} Air Stamp: https://www.nttdocomo.co.jp/info/news_release/2014/03/04_00.html

^{*5} INFOSOUND: <http://jp.yamaha.com/products/>

^{*3} Google Glass: <http://www.google.com/glass/start/>

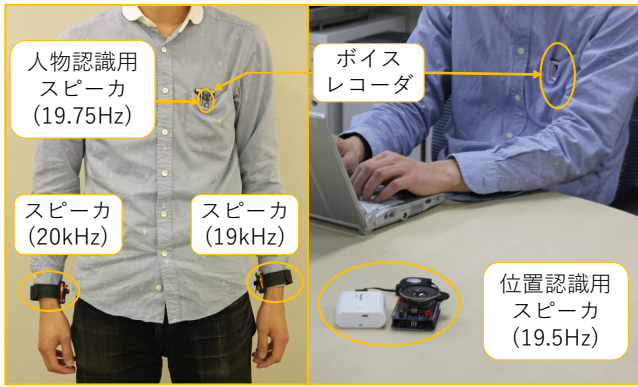


図 1 デバイスの構成

Sound Tag^{*6}, 近距離通信の Smart Sonic Communication^{*7}, Sonic Notify^{*8} などがあげられる。これらの技術を用いて、店舗のある商品の前に来た時に、その商品に関するクーポンや情報をユーザのスマートフォンに提示するなどの使い方がある。また、二つの違う周波数の超音波を組み合わせるとして信号としたもの (DTMF: Dual Tone Multi Frequency) を利用し、音楽作品への応用を目的とした研究も行われている [13]。これらの超音波通信で必要なものは一般的なスピーカとマイクだけであるので、既存の設備に追加しやすい点や、特別な設定の必要なく、多数のデバイスへデータ送信が可能である点が特徴的である。ウェアラブル環境において、ユーザはマイクとスピーカを常時装着していると考えられるので、この技術は汎用的に利用できる。

提案手法では、超音波 ID を音声ログへ同時に録音し、様々な情報を埋め込む。音声へ情報を埋め込む、超音波領域で情報通信するという点では既存の技術、研究と同じであるが、本研究では、端末自身の発する超音波の周波数応答から端末の装着位置を推定し、装着位置に応じた認識処理を行うという点で既存の研究と異なる。

3. 超音波 ID の埋め込み手法

3.1 これまでの研究

筆者らの先行研究 [1] の想定環境では、ユーザはボイスレコーダを胸に、ジェスチャ認識用のスピーカを両手首に、自分に向けて装着する。また、位置認識用のスピーカを環境に、人物認識用のスピーカをユーザの胸に装着する。図 1 にデバイスの構成を示す。ジェスチャ認識の際に両手の動きの違いを認識するため、両手首のスピーカの周波数は違う周波数に設定する。音量は音源からの距離に応じて変化するという特徴と音源の動きの速さによって周波数が変

化するドップラー効果に着目し、両手首に装着した小型スピーカから出力した超音波の音量、ドップラー効果による周波数ピーク値の変化、環境音の音響特徴量を組み合わせるとして状況認識を行う手法を提案した。また、位置認識用のスピーカ、人物認識用のスピーカからの超音波は ID 化されており、それぞれ固有の ID を持っている。ボイスレコーダは周囲からの環境音、両手からの超音波、環境側からの超音波、人物からの超音波を同時に取得する。スピーカからの音は、可聴域の音だと不快であるとともに、環境音とスピーカからの音を分離する必要があるため、超音波を用いた。提案手法では、マイク一つであらゆるユーザ状況が取得でき、録音した音声データファイルひとつを解析するだけで、これらのユーザ状況が識別できる。また、超音波受信機 (ボイスレコーダ) と超音波発信機 (小型スピーカ) が一人の身体の内側で完結しているため、ジェスチャ取得のために環境側の正確な位置に受信機または送信機を設置する必要がない。さらに、スピーカは超音波を発信するだけ、ボイスレコーダは音を記録するだけなので、データ通信のための有線/無線通信が必要ない。

3.2 想定環境

本論文では展示会のような、会場に複数の展示や発表がある場で音声ログを行い、後から特定の発表者との会話や展示での音声を振り返る環境を想定する。また、商業デバイスを用いてこれらの人物認識、場所認識を行う。手首内側にスピーカのあるスマートウォッチ (e.g. Galaxy Gear) を使用すれば商業デバイスでもジェスチャについての評価が可能であるが、本論文では、実イベントでの展示物/発表者の位置と話の内容の記録に焦点をあて、ジェスチャについては考慮しない。

音声記録用のマイクとしてスマートフォンのマイクを、超音波 ID 発信用のスピーカとしてスマートフォンのスピーカを使用する。ユーザのスマートフォンは音声ログを行うと同時に、スピーカから固有の超音波 ID を発信する。また、本論文で使用するサンプリング周波数はすべて 44.1kHz とし、フーリエ変換時のサンプル数は 1024 とした。

3.3 超音波 ID のフォーマット

筆者らが先行研究 [1] で用いた超音波 ID は再生に時間がかかり、またエラーへの対処をしていなかったため、本研究では文献 [1] とは異なる超音波 ID フォーマットを用いる。具体的には、信号の変調には既存手法と同様に PPM (Pulse Position Modulation) を使用し、通信方式は赤外線リモコンの通信方式を参考にした^{*9}。改良した超音波 ID の一例を図 2 右に示す。この図において、出力が 1 となっている

soundsignage/infosound/

*6 スマートフォン連動ソリューション「Sound Tag」: http://ds-lab.co.jp/solu_soundtag.html

*7 Smart Sonic Communication スペシャルサイト: <http://ssc-connect.com/>

*8 Sonic Notify: <https://sonicnotify.com/>

*9 リモコンフォーマット: http://akizukidenshi.com/download/k4174_format.pdf

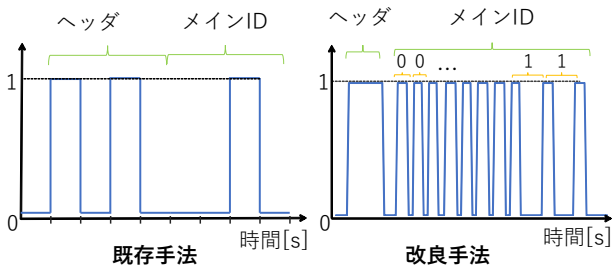


図 2 超音波 ID の一例

区間でスピーカは発音し、0 の区間では発音しない。超音波 ID はヘッダとメイン ID で構成されている。ヘッダは一定期間の超音波の出力で表され、メイン ID は超音波の細かい立ち上がりと無音期間の長さで 0 と 1 を表している。メイン ID の 0 と 1 の組み合わせの違いで、それぞれの超音波 ID の違いを認識している。なお本論文では、ヘッダの発音期間は 0.4 秒、メイン ID までの無音期間は 0.2 秒、メイン ID は、0.1 秒の発音と 0.1 秒の無音で 0 を、0.1 秒の発音と 0.3 秒の無音で 1 を表している。

簡易的なエラーチェック機能として、偶数パリティチェックを実装している。これは、ID 中の 1 の数が偶数個になるように、ID の末尾に 0 または 1 を付け加える手法で、受信側では受け取った ID に奇数個の 1 がある場合にはエラーと判断できる。

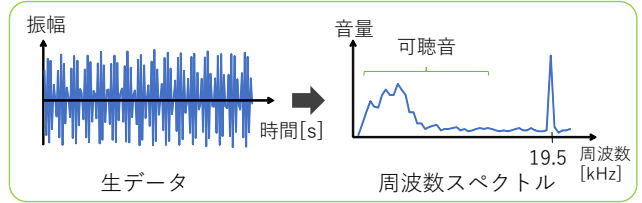
本論文では 19.5kHz の 1 チャンネルのみを使用し、メイン ID の長さは 8 ビットとしているため 256 個の ID を識別できるが、他の帯域も使用することやメイン ID のビット数を増やすことで、より多くの超音波 ID を生成することが可能である。既存手法と比べて短い時間で、より多くの情報を伝達することが可能になった。

3.4 超音波 ID の認識処理

超音波 ID の認識処理の流れを図 3 に示す。まず図 3 上段に示すように、録音した音声データには超音波と環境音が混在しているため、高速フーリエ変換 (FFT: Fast Fourier Transform) を行い、超音波部分の音量を取り出す。次に図 3 下段に示すように、19.5kHz 付近の音量の時間変化を取得し、しきい値以上であれば 1、それ以下であれば 0 で表す。前節で述べた ID フォーマットをもとに 0, 1 の組み合わせを取得し、あらかじめ登録している ID リストと比較することで取得した ID を認識する。

音源に近づく/遠ざかる動きにより、ドップラー効果で周波数が変化することが予想される。日常の歩行速度を時速 4km と仮定すると、ドップラー効果により約 60Hz の周波数変化が生じる。本論文では、デモ発表形式の実イベント利用を想定しており、デモや展示を見ながらの移動はこの歩行速度より遅いと考えられる。従って、本論文の認識処理では、19.5kHz 前後の約 50Hz をマージンとした。

① 生データを周波数スペクトルに変換



② 超音波音量の時間変化をしきい値で二値化

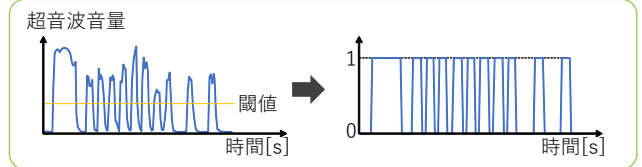


図 3 超音波 ID 認識処理の流れ

4. 実イベントでの運用

実イベントで 3 章に示す超音波 ID を用いた。なお、このイベントはもとより開催が決定していたものであり、本システムの実験のために開催したものではない。イベント会場はデモ形式の発表会場であり、発表者は前半 6 人、後半 5 人の合計 11 人であった。聴講者は約 30 名であり、自由に発表を聞きながら会場を回る。会場の様子を図 4 に、発表者の位置関係を図 5 に示す。図 5 の番号は、発表者の ID 番号を表している。各デモ発表者は発表を行いながら、それぞれ固有の超音波 ID の音声ファイルを自分の PC で再生し続ける。音声ファイルの長さは約 30 秒であり、超音波 ID が一回だけ再生されるようになっている。すなわち、この音声ファイルをループ再生することで約 30 秒間に一回超音波 ID を発信することになる。聴講者は各自のスマートフォンの録音アプリケーションで音声を録音しながら会場を回ってもらった。録音アプリケーションには、iOS 端末の人は PCM 録音 Lite^{*10} を、Android 端末の人は PCM 録音^{*11} を使用してもらった。音声ログを行った聴講者は 12 名であった。スマートフォンの装着位置は胸ポケットを推奨したが、胸ポケットが無い場合には任意の位置で行ってもらった。録音のサンプリング周波数は 44.1kHz であり、1024 サンプルごとに FFT を行った。録音したデータの解析には、マイクロソフト社の Visual C# を用いて開発した解析ソフトウェアを使用した。また、解析に用いた PC は lenovo 社の X240s である。

図 6 に実験結果の一部を示す。縦軸の 0 から 10 はそれぞれ発表者の ID 番号に対応しており、-1 はエラーチェックで検知されたもの。-2 はエラーチェックを通過したが、今回用意していない ID と認識された場合を表している。図下の矢印は正解データを表しており、この範囲内に同じ

^{*10} PCM 録音 Lite: <https://itunes.apple.com/jp/app/pcm-lu-yin-lite/id439572045?mt=8>

^{*11} PCM 録音: <https://play.google.com/store/apps/details?id=com.kohei.android.pcmrecorder&hl=ja>



図 4 発表会場の様子

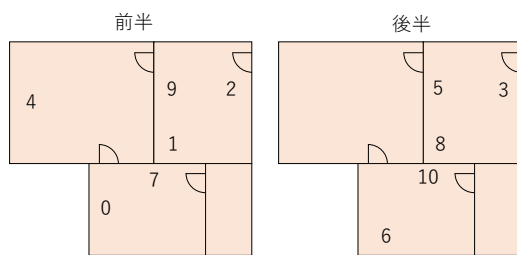


図 5 発表者の位置関係

番号の ID が存在していたら正しく認識できていることを示している。

正解データの通りに ID 番号の変遷が見られる箇所もあるが、全体的に -1 と -2 が多く、認識できていない状況が多く見られた。これは、一人の発表者に聴講者が集まりすぎて、音源から聴講者が遠かったこと、音源と聴講者との間に別の聴講者がおり、マイクへの超音波を妨げていたこと、聴講者のスマートフォンの装着位置により、マイク入力に十分な音量が確保できていなかったことなどが原因であると考えられる。また、同部屋に複数の ID が存在するときは互いに干渉して認識できないことや、同部屋の他の ID が交互に認識されることが確認された。これは、各 ID の発信間隔を被らないようランダムに設定することや、本論文では 19.5kHz のみを使用しているが、他の帯域も使用すること、発信する超音波 ID の音量を干渉しないように調節することなどで改善できると考えられる。

5. システムの改良

前節の予備実験により、スマートフォンに十分な音量の超音波が到達していないことが主な問題であるとわかった。この原因として、音源との距離が離れている、録音デバイスの装着位置が不定である等の問題点があげられるが、本論文では、前提としてユーザはスピーカから一定距離にいる状況を想定し、まずはスピーカから一定距離にいる人に正しく情報を伝えることを目指す。音源とユーザとの距離を一定距離に設定した場合の問題点としては、録音デバイスの装着位置が不定であることがあげられる。音は高周波になるほど指向性が鋭くなる特性があるため、スマート

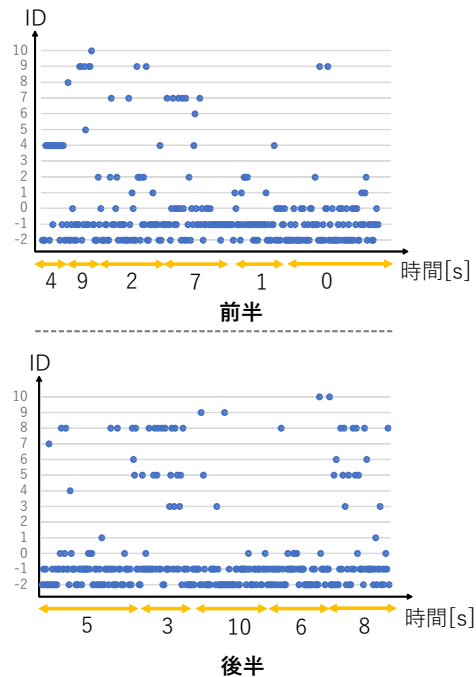


図 6 実イベント実験の結果の一部

フォンの装着位置によって、超音波のマイク入力の特徴が変化し、認識精度に影響を及ぼす。そこで、本論文では録音デバイスの装着位置による音響特性の問題に取り組む。録音端末の装着位置を検出し、その状態に応じた適切な処理を行う改良システムを考案する。

端末の把持状態を、スイープ信号の周波数応答から認識している研究 [14] がある。この手法から着想を得て、スイープ信号の周波数応答で端末の状態を認識することができるのではないかと考えた。

スイープ信号の周波数応答を確認するための実験を行った。スマートフォンからスイープ信号を数回発し、同じスマートフォンで同時に録音した。スマートフォンの装着位置は図 7 に示すように、胸ポケット、ズボンの前ポケット、ズボンの後ろポケット、手持ちの 4 つの状態を想定し、ポケットに入れる際にはマイク部分を上側にした。スイープ信号は、0.5 秒のうちに 18kHz から 22kHz まで変化するものを用いた。使用した端末は SAMSUNG 社の GALAXY Note3 である。

それぞれの状態のスイープ信号の 5 回分の平均周波数応答を図 8 に示す。上記の 4 つの状態において、それぞれが若干異なる周波数応答を示している。従って、それぞれの状態をスイープ信号で識別できる可能性がある。

以上の結果より、超音波 ID のヘッダとしてスイープ信号を用いることにした。スイープ信号の周波数応答によってスマートフォンの装着位置が取得できるので、その状態に応じた認識処理を行う。すなわち、4 つの状態それぞれに対応したしきい値を用意し、現在の端末の状態に適したしきい値で認識を行う。本論文では取得したデータを見



図 7 録音デバイスの装着位置

て、4つの状態それぞれに適切と考えられるしきい値を手動で決定した。

また、予備実験より、偶数パリティチェックが有効でないことが判明したので、誤り訂正符号としてIDの後尾に4ビットのハミング符号を付与し、誤り訂正ができるように改良した。

5.1 改良システムを用いた実験

改良したシステムを用いて、評価を行った。図9に示す部屋に、3つの音源を置き、ユーザは装着したスマートフォンで音声ログをする。図9中の数字は音源のID番号を表しており、ID番号に対応する超音波IDを図9の記号の向きに発信している。また、ユーザのスマートフォンからもID0の超音波IDを発信している。超音波IDは約30秒間に一回発信される。ユーザは一つの音源の前に約5分間滞在し、図9に示す矢印の順番に移動した。滞在中は超音波IDの方向を向いて立った状態である。ユーザのスマートフォンの装着位置は、胸ポケット、ズボンの前ポケット、ズボンの後ろポケット、手持ちの4箇所について行った。各装着状態において、録音したデータ中の任意のスweep信号ヘッダ5つを学習データとし、認識手法には最近傍法を用いた。認識手法としてはSVM(Support Vector Machine)なども考えられるが、認識精度に大きな違いが見られなかったため、今回は最もシンプルな最近傍法を採用した。録音用のアプリケーションは予備実験で使用したものと同一PCM録音アプリを用い、サンプリング周波数は44.1kHzである。被験者は1名であり、録音に使用したスマートフォンはSAMUSUNG社のGALAXY Note3である。音源に使用したデバイスは、ID1がSAMSUNG社のGALAXY Note3、ID2がマウスコンピュータ社のPC、ID3がApple社のMacBook Airである。

端末の装着位置認識の精度を表1に示す。この表が示すように、超音波sweep信号の周波数応答を用いた場合、胸ポケットは9割近くの精度で、前ポケットと手持ちは約7割の精度で認識ができたが、後ろポケットは2割程度の認識率であり、平均で64.0%の認識率であった。録音デバ

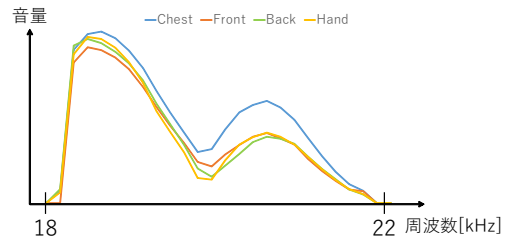


図 8 sweep信号の周波数応答

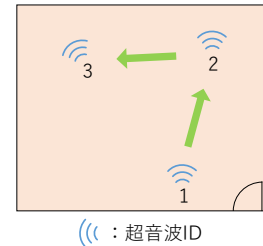


図 9 音源の設置箇所とユーザの滞在順

表 1 端末装着位置の認識結果

Input	Output				Accuracy[%]
	Chest	Front	Back	Hand	
Chest	27	1	0	1	93.1
Front	0	20	3	6	70.0
Back	2	13	7	9	22.6
Hand	0	7	1	20	71.4

イスの装着位置は一度装着したらしばらくは同じ場所であると考えられるので、ある程度の期間の結果から多数決をとることで、正確な装着位置に確定できると考えている。

解析した録音データから胸ポケット、前ポケット、後ろポケット、手持ちの4状態におけるID認識のための最適なしきい値をそれぞれ決定し、この4つのしきい値の平均の値を平均しきい値として求める。図10最上段に常に平均しきい値を用いた、しきい値調整なし(従来手法)の認識結果を示す。また、装着位置認識で得られた結果に合わせて、その装着位置の最適しきい値に切り替えるしきい値調整(提案手法)を用いた結果を図10の2段目に、常に最適しきい値を用いた場合の結果を図10の3段目に示す。図下の矢印は、正解データを表しており、グラフは左から順に胸ポケット、前ポケット、後ろポケット、手持ちの結果を表している。これらの結果をまとめたものを表2に示す。実験では各ID音源の前に5分(300秒)滞在しており、各IDは30秒に一回発信されている。従って、表中の数字の最大値は10となる。

ID0が定期的に認識されているが、これは録音端末自身のIDである。認識エラーはすべて-2で表している。図中に見られる-2は、ほとんどがID干渉によって認識できていない結果であった。

表2に示すように、従来手法で認識した結果より、最適しきい値で認識した結果の方が4つの状態全てにおいて良

表 2 各手法に対する ID の認識数

装着位置	ID	従来手法	提案手法	最適しきい値
胸ポケット	1	10	10	10
	2	4	8	8
	3	4	4	4
前ポケット	1	3	3	4
	2	8	9	10
	3	0	4	4
後ろポケット	1	0	4	6
	2	4	0	1
	3	0	0	2
手持ち	1	6	6	6
	2	4	7	7
	3	6	4	4

い結果となっており、装着位置によってしきい値を変化させる手法が有効であるといえる。また、提案手法の結果は装着位置認識の精度と似た傾向が見られ、胸ポケットの時は最適しきい値と同レベルの認識精度であり、前ポケットと手持ちの時にも認識精度に改善が見られた。しかし、装着位置認識精度自体が低かった後ろポケットに関しては、しきい値調整がうまく適用されず、改善が見られなかった。また、仮に装着位置認識精度がよく、常に後ろポケットにおける最適しきい値で認識できたとしても、他の装着箇所と比べると認識精度がよくないことが表 2 から分かる。従って、提案手法において後ろポケットに録音デバイスを装着することはふさわしくないといえる。

提案手法では、最終的には一定のウィンドウサイズで認識結果の多数決を出し、この時間帯はこの ID が近くにいた、という情報が出せればよいと考えているので、認識精度は 6 割ほどあれば実現できると考えている。後ろポケット以外の 3 つの装着状態それぞれにおいて、ID 認識の合計が約 6 割の認識率であるので、この認識精度が有効であるといえる。実際に今回の結果を、自分の端末 ID である 0 を除き、またエラーである -2 を直近の多数決の結果と同じにするようにウィンドウサイズ 5 で多数決を出した結果、表 3 のような結果になった。表の結果は、正解数/各 ID 音源の周囲滞在期間中に認識した ID 数で表している。グラフで表したものを図 10 最下段に示す。厳密な認識精度としては、胸ポケット装着時以外は高くないが、図 10 最下段に示すように、胸ポケットと前ポケットに関しては ID の変遷とおおよその時間は把握することができる。後ろポケットと手持ちに関しては、ユーザの動き通りの ID の変遷は得られなかったが、まとまった ID1, 2, 3 が交互に出現し、ユーザの周囲に 3 つの ID が同時に存在していることが把握できる。以上より、装着位置によってしきい値を切り替える手法は有効であるといえる。

5.2 考察

端末の位置推定には加速度センサ等を用いることが一般

表 3 多数決後の ID 認識数

ID	1	2	3
胸ポケット	8/10	9/12	5/16
前ポケット	2/2	9/9	5/7
後ろポケット	5/12	4/9	1/10
手持ち	12/15	4/10	1/17

的であるが、本研究では音声ログへのタグ付けのために超音波 ID を発信しているため、これを利用して認識することができないかと考え、スイープ信号の周波数応答のみで端末の位置推定を行った。本提案と加速度センサを組み合わせることによって、より効果的な端末の位置推定が期待できる。本論文では、超音波を受信する側の処理の工夫のみを考慮したが、超音波 ID を発信するスピーカ側にも状況により音量を調整するなどの工夫を加えることが可能である。また、プライバシー保護のため、他人に不正に取得された ID が解析されないように、ID を暗号化するなどの処理をする必要がある。本論文ではマイク部分は常に上側になるように設定したが、マイク部分が下になる場合についても調査を行う必要がある。

6. まとめ

本論文では、超音波 ID による音声データへのタグ付け手法を実環境で運用し、装着位置による認識精度の変化を明らかにした。また、超音波スイープ信号の周波数応答により録音端末の装着位置を推定する手法を提案し、4 箇所を平均 64.0% の認識率で認識した。さらに装着位置認識の結果に合わせて認識処理のしきい値を装着位置ごとの最適なしきい値に変化させることにより、認識精度の改善を確認した。今後の課題として、超音波 ID の暗号化や、端末の状態によってリアルタイムに超音波 ID の音量を変化させる手法を実装する予定である。

謝辞 本研究の一部は、科学技術振興機構戦略的創造研究推進事業(さきがけ)によるものである。ここに記して謝意を表す。

参考文献

- [1] H. Watanabe, T. Terada, and M. Tsukamoto: A Method for Embedding Context to Sound-based Life Log, *Journal of Information Processing*, Vol. 22, No. 4 (Oct. 2014).
- [2] 大石康智: あらゆる音の検出・識別を目指して: 音響イベント検出研究の現在と未来, 日本音響学会研究発表会講演論文集, pp. 1521-1524 (2014).
- [3] T. Heittola, A. Mesaros, A. J. Eronen, and T. Virtanen: Audio Context Recognition Using Audio Event Histograms, *Proc. of the 18th European Signal Processing Conference (EUSIPCO2010)*, pp. 1272-1276 (2010).
- [4] M. Rossi, S. Feese, O. Amft, N. Braune, S. Martis, and G. Troster: AmbientSense: A Real-Time Ambient Sound Recognition System for Smartphones, *International Conference on Pervasive Computing and Communications Workshops (PERCOM Workshops)*, pp.

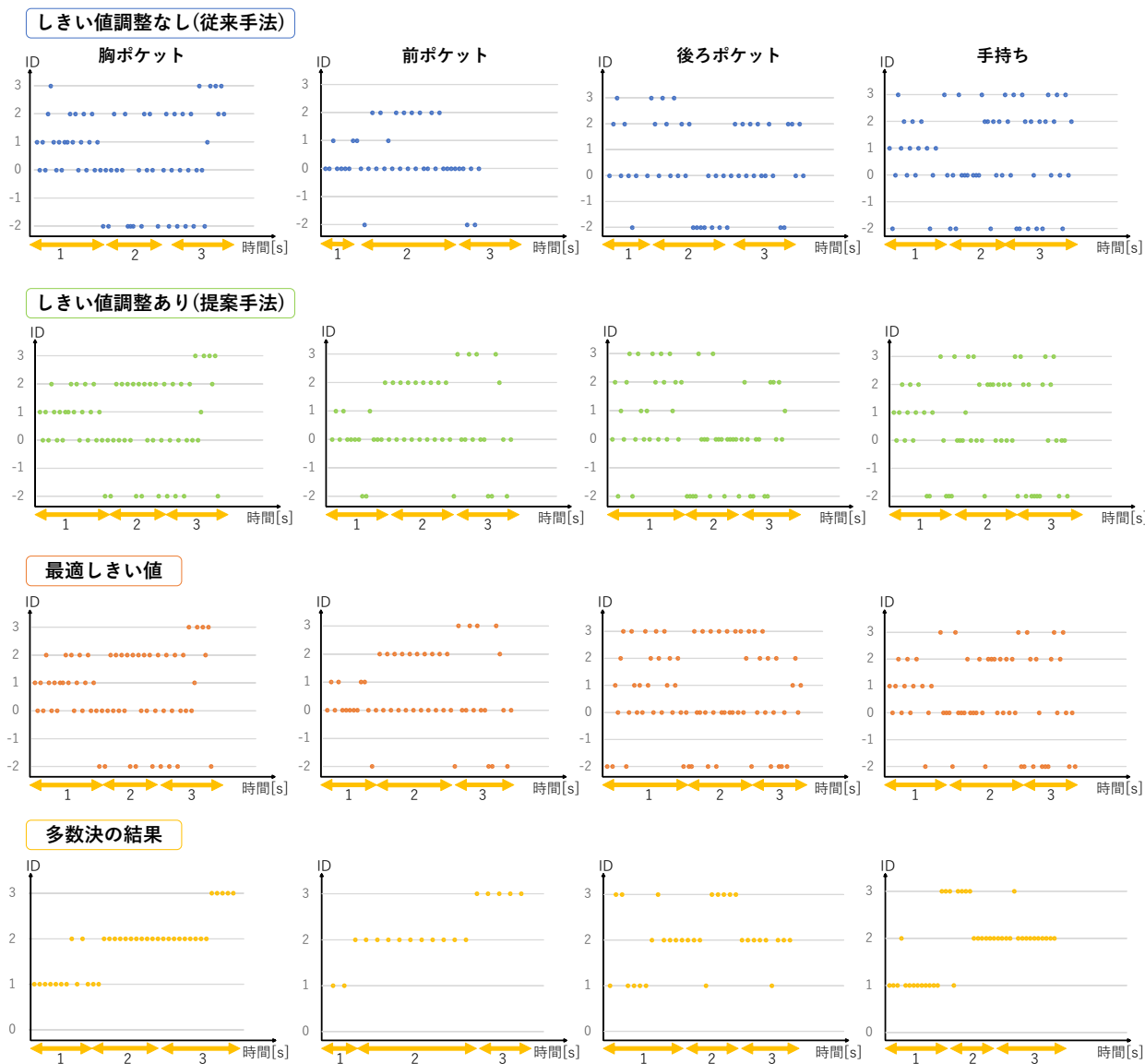


図 10 実験結果

- 230–235 (2013).
- [5] 大内一成, 土井美和子: 加速度と音による家庭内ユーザ状況認識の可能性検討, 情報処理学会マルチメディア, 分散, 協調とモバイルシンポジウム (DICOMO2010) 論文集, Vol. 2010, pp. 508–515 (2010).
- [6] T. Choudhury and A. Pentland: The sociometer: A wearable device for understanding human networks, *the 5th ACM Conference on Computer-Supported Cooperative Work and Social Computing (CSCW '02) Workshop: Ad hoc Communications and Collaboration in Ubiquitous Computing Environments* (2002).
- [7] T. M. T. Do and D. Gatica-Perez: Groupus: Smartphone proximity data and human interaction type mining, *Proc. of the 15th International Symposium on Wearable Computers (ISWC '11)*, pp. 21–28 (2011).
- [8] 田岡康裕, 納谷 太, 野間春生, 小暮 潔, 李 周浩: Bluetoothの電波強度を用いたユーザの位置推定手法, 情報処理学会研究報告 ユビキタスコンピューティングシステム (UBI), Vol. 2008, No. 66, pp. 147–152 (2008).
- [9] A. Ward, A. Jones, and A. Hopper: A New Location Technique for the Active Office, *IEEE Personal Communications*, Vol. 4, No. 5, pp. 42–47 (1997).
- [10] H. L. Muller, M. McCarthy, and C. Randell: Particle Filters for Position Sensing with Asynchronous Ultrasonic Beacons, *Proc. of the 2nd International Workshop on Location- and Context-Awareness (LoCA '06)*, pp. 1–13 (2006).
- [11] 松岡保静, 中島悠輔, 吉村 健: 可聴帯域における音波情報伝送技術: 音響 OFDM, 電子情報通信学会技術研究報告, Vol. 106, No. 125, pp. 25–29 (2006).
- [12] 福島斉彦: Remote Control of Television Receiver, *テレビジョン*, Vol. 14, No. 1, pp. 34–38 (1960).
- [13] 平林真美, 江島和臣: 高可聴域音による音声 ID の観客参加型音楽作品への応用, *インタラクシオン 2014*, インタラクティブ発表 (2014).
- [14] M. Ono, B. Shizuki, and J. Tanaka: Touch & activate: adding interactivity to existing objects using active acoustic sensing, *Proc. of the 26th Symposium on User Interface Software and Technology (UIST '13)*, pp. 31–40 (2013).