



超音波IDによる音声ログへの人物/位置情報付与

渡邊, 拓貴

寺田, 努

塚本, 昌彦

(Citation)

神戸大学大学院工学研究科・システム情報学研究科紀要, 8:44-51

(Issue Date)

2016

(Resource Type)

departmental bulletin paper

(Version)

Version of Record

(URL)

<https://hdl.handle.net/20.500.14094/81009699>



【研究論文】

超音波IDによる音声ログへの人物/位置情報付与

渡邊 拓貴^{1*}・寺田 努¹・塙本 昌彦¹

¹工学研究科電気電子工学専攻

(受付:November 18, 2016 受理:January 12, 2017 公開:January 20, 2017)

キーワード： 超音波，音声ログ，人物認識，位置認識，ウェアラブルコンピューティング

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のよう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。提案手法を実環境で運用したところ、デバイス装着位置の認識率は平均84.7%であり、超音波IDの認識率は従来手法では55.2%であったところ、改良手法では68.0%に向上した。

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。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できない。また、ウェアラブルカメラのような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、どこにどのよう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な場合には有効であるが、そのような操作が行えない状態(手が塞

がっている場合、デバイス操作が憚られる場合)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を集める必要がある。またマイクと、加速度センサやBluetoothなどの複数の情報と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。

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。本手法では、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、これらの問題点のうち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。

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

なお、本論文は文献[2]で発表した内容を元に、改良手法を実際のイベントで運用し、その結果と考察を追加した

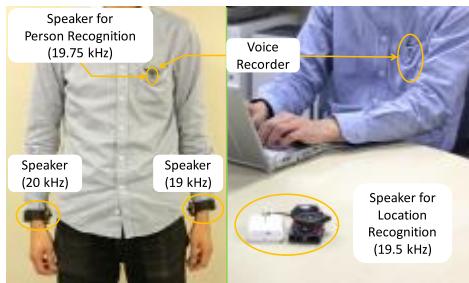


図 1 先行研究でのデバイス構成

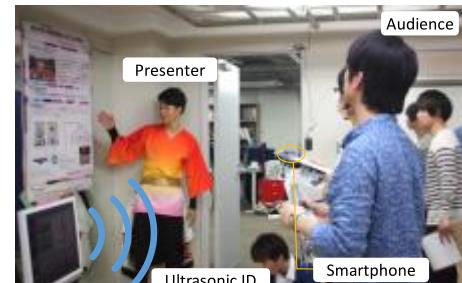


図 2 発表会場の様子

ものである。

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。その一つとして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が行われている[3]。この研究では、MFCC(Mel Frequency Cepstral Coefficient)など音声認識で使用される音響特徴量を主に用いる。大内らの研究[4]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。Choudhuryらの研究[5]では、赤外線センサ、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、これを用いた人間関係取得を議論している。近距離の端末同士の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[6]。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も提案されている[7]。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。

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[8]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[9]。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[10]。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、多数のデバイスへ一斉に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

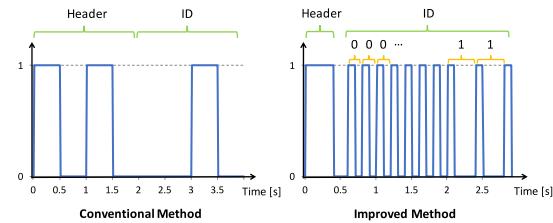


図 3 超音波 ID の一例

汎用的に利用できる。

本研究では、超音波でタグを付与するため、音声ファイル一つにすべてのデータが集約され、タグを付与するための他のデータを扱う必要がなく、タグ記録用の追加デバイスも必要ない。また、認識したい場所に小型の超音波スピーカを設置するだけでよいので、環境側の正確な位置にデバイスを設置する必要がない。提案手法で用いる音声の超音波領域へ情報を埋め込む手法は既存の技術、研究でも行われているが、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周囲環境(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は異なる。

3. 超音波IDの埋込手法

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ーを胸に、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、ユーザの居た場所認識用のスピーカを環境に、ユーザの周囲にいた人物認識用のスピーカをユーザ及び周囲の人物の胸に装着する。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にはそれぞれ固有の番号(ID)が割り振られ、超音波によりIDを定期的に発信している。ボイスレコーダー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

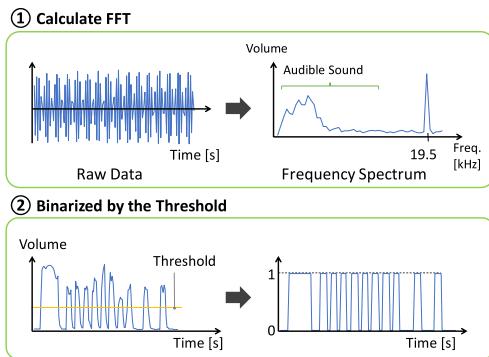


図 4 超音波 ID 認識の流れ

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、筆者らが先行研究[1]で提案した手法のうち、超音波IDで音声ログに情報を付与する手法のみに着目する。本論文では図2に示すような、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、コマーシャルデバイスを用いてこれらの人認識、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、コマーシャルデバイスでも先行研究で提案した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/PC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ともに、スピーカから固有の超音波IDを発信する。

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ので、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、信号の変調にはパルス位置変調(PPM: Pulse Position Modulation)を使用した。改良した超音波IDの一例を図3右に示す。この図において、出力が1となっている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。

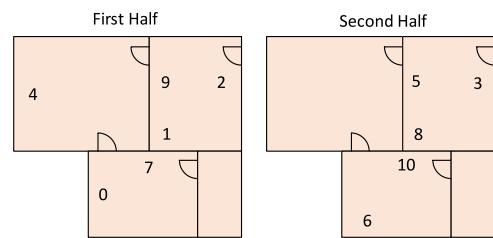


図 5 発表者の位置関係

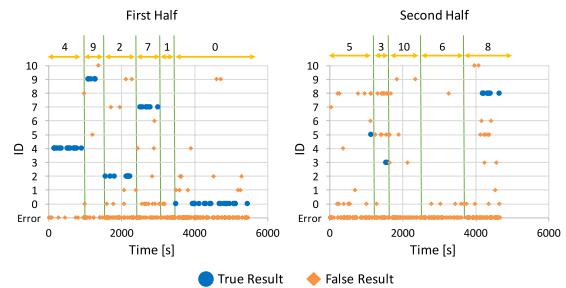


図 6 実験結果の一例

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を利用できるが、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。本フォーマットにより、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できる。

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4に示す。まず図4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4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、展示会での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60Hzをマージンとした。

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

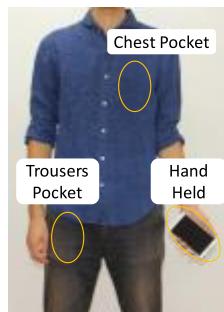


図 7 録音デバイスの装着位置

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2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を、Android端末の人はPCM録音を使用してもらった。音声ログを行った聴講者は12名であり、スマートフォンの装着位置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番号に対応している。図上の矢印は正解データを表しており、この範囲内に同じ番号の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。

ID4, 9, 2, 7, 0のように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、全体的にエラーが多く、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。

5. システムの改良

前節の予備実験により、認識エラーの主な原因是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だとわかった。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、本論文では展示会等での利用を想定しているため、前提としてユーザはスピーカから一定の距離にいる状況を想

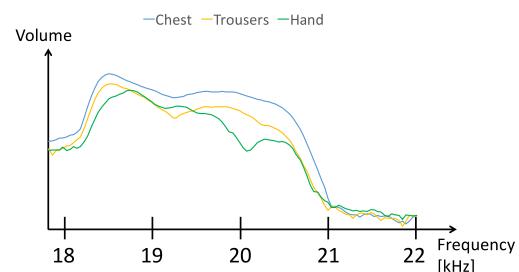


図 8 スイープ信号の周波数応答

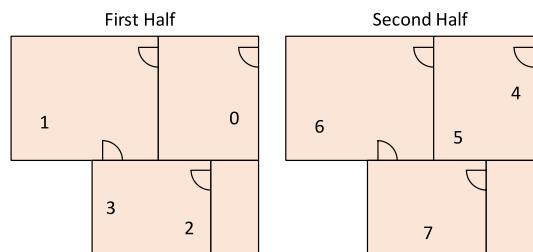


図 9 改良手法を用いたイベントでの発表者の位置関係

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。録音デバイスは装着位置ごとに異なるデバイス周囲環境を持っているため、取得できる超音波IDの音量が異なる。つまり、手持ちのような密閉度のない場合は十分に超音波IDを取得でき、ズボンポケットのような密閉度の高い場合は取得できる超音波IDが小さくなると考えられる。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。具体的には、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手法を考案する。

5. 1 装着位置認識手法

本研究の想定環境として、録音デバイス自身からも超音波IDを発信しているという特徴に着目し、自身の発信した超音波の応答から装着位置を認識する手法を提案する。具体的には、文献[11]を参考に、スイープ信号の周波数応答でデバイスの状態を認識する。装着位置(胸ポケット、ズボンポケット、手持ち等)によってデバイス周囲空間の密閉度等により音響特性が異なるため、各装着位置によって取得できるスイープ信号の周波数応答が異なると考えられる。この特性をデバイスの装着位置認識に利用する。

各装着位置による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発信し、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、手持ちの3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、0.1秒のうちに18kHzから22kHzまで変化するものを5回連続して発信した。FFT計算のためのウインドウサイズは、1回のFFTに少なくとも1回のスイープ信号(0.1秒)を含むように、8192(約0.19秒)とし、ウインドウが重ならないように8192ずつスライドした。使用した端末はSAMSUNG社のGALAXY Note3である。

表 1 取得音量と認識に用いたしきい値

被験者	装着位置	取得音量	用いたしきい値		
			胸	ズボン	手持ち
A	手持ち	35.2	6.0	1.5	10.0
C	ズボン	5.4			
B	手持ち	1.9	0.3	0.1	0.5
	ズボン	0.3			

表 2 装着位置の認識結果

被験者	装着位置	前後半	認識率 [%]
A	手持ち	前半	78.9
		後半	96.0
B	手持ち/ズボン	前半	96.5
		後半	61.5
C	ズボン	後半	90.4
平均			84.7

それぞれの状態のスイープ信号の周波数応答の平均を図8に示す。上記の3つの状態において、それぞれが異なる周波数応答を示している。これは、デバイス周囲空間の密閉度によって変化していると考えられる。つまり、胸ポケットのようにデバイス周囲に空間がある程度あり、かつ密閉されている場合には、スイープ信号の反射が多く、全周波数において大きな反応が得られている。ズボンポケットのように布地に密着している場合には反響が少ないため、胸ポケットよりも周波数応答が少なくなり、手持ちの場合にはスイープ信号は空中に拡散するため、19kHz付近以外では最も小さな周波数応答となった。

以上より、各デバイス周囲空間(装着位置)で取得できる周波数応答が異なるため、各装着位置をスイープ信号で識別可能だと考えられる。

5. 2 超音波IDの改良

前節の結果より、超音波IDのヘッダとしてスイープ信号を用いる。スイープ信号は、0.1秒間に18kHzから22kHzへと遷移するものを用い、これを5回連続で繰り返したものを作成した。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、3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。なお、本論文では取得したデータを観察して、3つの状態それぞれに適切と考えられるしきい値を設定した。

装着位置の認識を行わず、スイープ信号の反射波音量等を用いて直接しきい値を調整する手法も考えられるが、図8に示すように、胸ポケット装着時に反射音の音量は最大になり、手持ち時に最小になる。これは、取得できる超音波IDの音量の大小と対応していないため、この手法を用いるのは難しいと考えた。従って、本研究ではスイープ信号の周波数応答を用いて装着位置を識別し、それぞれに適したしきい値を用いて認識する手法を利用する。

また、4章での実験より、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。

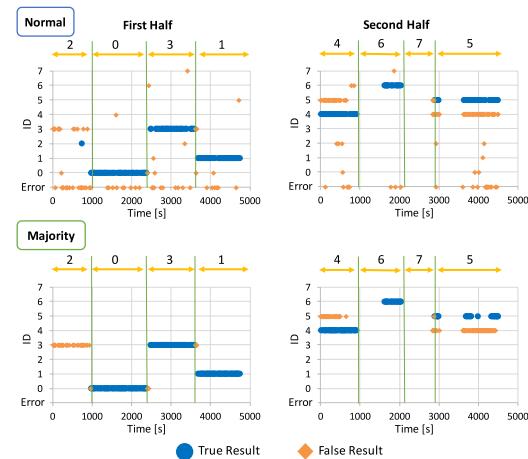


図 10 認識結果の一例

5. 3 改良手法の実イベントでの運用

改良手法の効果を確認するために、実際のデモ発表イベントで改良手法を用いた。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4人、後半4人の合計8人であった。聴講者は約40名であり、自由に発表を聞きながら会場を回る。会場を一通り見て回るのには、前半、後半ともにそれぞれ約75分かかった。発表者の位置関係を図9に示す。図9中の数字は各発表者のID番号を表している。各デモ発表者は発表を行なながら、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15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15秒間に一回超音波IDを発信することになる。4章で行った実験と比べて超音波IDの発信間隔を短くすることで、超音波ID認識の機会を多くしている。聴講者は各自のスマートフォンで、スイープ信号を発信すると同時に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。なお、本イベントでは、聴講者の音声ログに発表者の情報をタグ付けすることが目的であるので、聴講者自身にIDは割り振っていない。従って、聴講者のスマートフォンから再生される超音波情報はスイープ信号のみで、ID情報は含まれていない。録音アプリケーションは、iOS端末の人はAir Recorderを、Android端末の人はPCM録音を使用してもらった。デバイスの装着位置は胸ポケット、前ポケット、手持ちの3箇所を対象とする。認識のためのしきい値はデバイスごとに異なるため、録音データを観察し、設定した。録音のサンプリング周波数は44.1kHzである。FFT計算のためのウインドウサイズは、装着位置認識のためのスイープ信号を解析する時のみ8192とし、超音波IDを解析する際には1024とした。被験者として12名の男女に改良手法を用いた録音を行ってもらったが、被験者の大半はiOS端末であり、Air Recorderは端末がスリープ状態に入ると停止してしまうため、イベント中に操作ミスでスイープ信号の再生を止めてしまった被験者が多く、改良手法のデータを解析できたのは12名のうち3名であった。また、3名のうち1名は後半だけのデータであった。

表 3 実イベントでの超音波 ID 認識結果

被験者	デバイス	前後半	多数決処理	従来手法			改良手法				
				認識ID数	正解数	正解率 [%]	認識ID数	正解数	正解率 [%]		
A	iPhone 5	前半	なし	239	141	59.0	213	145	68.1		
			あり	239	192	80.3	213	175	82.2		
		後半	なし	197	67	34.0	201	91	45.3		
			あり	197	82	41.6	201	97	48.3		
B	iPhone 6	前半	なし	144	63	43.8	102	51	50.0		
			あり	144	70	48.6	102	64	62.8		
		後半	なし	152	58	38.2	143	60	42.0		
			あり	152	67	44.1	143	100	69.9		
C	iPhone 5	後半	なし	85	29	34.1	65	28	43.1		
			あり	85	52	61.2	65	50	76.9		
平均			なし	163.4	71.6	41.8	144.8	75.0	49.7		
			あり	163.4	92.6	55.2	144.8	97.2	68.0		

表1に超音波IDの取得音量と、用いたしきい値を示す。取得音量は、超音波IDの各パルスの平均値を表している。同じ録音デバイスであった被験者A, Cについては共通のしきい値を用い、被験者Bには別のしきい値を設定した。表1に示すように、同デバイス内であれば装着位置ごとに密閉度が異なり取得できる音量が異なるため、装着位置によってしきい値を切り替える手法が有効であるといえる。被験者Bのデバイスでの取得音量は他二人と比較するとスケールが異なるが、同様に装着位置によって取得音量が異なるため、しきい値の切り替え手法が有効だと考えられる。

認識結果の一例として、図10に被験者Aが改良手法を用いた場合の認識結果を示す。矢印とその上の数字は正解のIDと期間を表している。上の二つのグラフは多数決処理を行っていない認識結果であり、下二つは多数決処理を行った場合の認識結果である。ID認識の結果が多数決処理後でもエラーであった場合は、最後に認識したIDに置き換えている。また、ID7はどの被験者においても検出されなかつたため、ID7の発表者は、メディアプレイヤーを誤って閉じた、スピーカのボリュームがゼロになっていたなどの、何らかの理由で超音波IDを発信できていなかったと考えられる。従って、本イベントではID7については考慮していない。

表2に装着位置の認識精度を示す。この表に示すように、装着位置の認識精度は平均で84.7%であった。表3にしきい値を一定にして認識を行う手法(従来手法)と、装着状態に応じてしきい値を切り替える手法(改良手法)を用いた場合の超音波IDの認識精度を示す。表の認識ID数は認識したIDの合計数を、正解数は正しく認識できたID数を、多数決あるいは過去60秒間の認識結果による多数決処理を行った結果の認識精度を示している。60秒間の多数決処理を行うと、認識結果に30秒ほどの遅れが想定されるが、これはイベント時間(約75分間)と比べて十分小さいため、問題ないと考える。表3に示すように、従来手法においても改良手法においても多数決処理によってID認識精度は向上している。従来手法と改良手法のID認識精度を比較すると、従来手法では平均55.2%の認識率であったところを、改良手法では、平均68.0%の認識率となり、改良手法による超音波ID認識率の向上が確認できた。結果の平均に着目すると、認識ID数は減少し、

正解数が増加している。つまり、改良手法によって余分なID認識を減少させ、音声ログへより正確にタグ情報を付与できたことを示している。本研究の想定環境では、長時間の音声ログに大まかな目安として超音波IDでタグを付与し、さらに細かい部分は実際に音声ログを聞きながら絞り込んでいくことを想定している。実際の録音データのタイムラインと超音波IDの情報を見ながら主観的に判断したところ、70%程度の超音波ID認識精度があれば、想定環境での利用には問題ないと考える。改良手法を用いた場合の認識精度は68.0%であり、70%を下回っているが、本実験での認識率低下の原因の一つは、図9、10に示すように同部屋に複数の超音波IDが存在する際の超音波IDの混信だといえる。この問題は、超音波IDの発信間隔をランダムにすること、他の周波数も用いて超音波IDを生成することで回避でき、認識精度のさらなる向上は可能だと考える。従って、本実験の環境では68%の認識精度で、想定する用途に利用できると考える。

5. 4 考察

改良手法の実イベントでの利用時には、超音波IDの発信間隔を従来の30秒から15秒へと短くすることで、ID認識の機会を多く設けた。一方、同部屋に複数の超音波IDが存在する場合は、より多くの超音波IDの混信が生じる結果となつた。超音波ID混信の解決手法としては、超音波IDの発信間隔をランダムに設定すること、今回用いた19.5kHz以外も用いて超音波IDを生成することを考えている。

提案手法で用いるスイープ信号は、身につけている服や体型によって変化する可能性がある。提案手法では、それぞれのユーザ自身から学習用のデータを取得しているため、学習データはこれらの情報を含んだデータである。

被験者A, Cと被験者Bとの取得音量のスケールの違いの原因としては、被験者Bが録音アプリケーションの入力最大音量設定のスライダを誤って操作してしまったこと、あるいはデバイスごとのマイク特性の違いが考えられる。本研究では認識のためのしきい値を、データを観察した上でデバイス個々に手動で設定した。今後はマイクの特性を考慮し、自動でしきい値を設定するような仕組みを実装することを予定している。

本研究では、ユーザは常に音源と一定の距離を保った位

置にいることを想定して、デバイスの装着位置による音量の変化のみに着目したが、音量は音源との距離によっても変化する。今後は、スイープ信号で装着位置を推定した後に、取得できる超音波音量によりユーザと音源との距離を推定した上で、これを考慮した認識を行えないか調査を行う。

6. まとめ

本研究では、超音波IDにより音声ログへ人物/場所情報を埋め込む手法を実環境のイベントで運用した結果、デバイスの装着位置によって取得できる超音波音量が異なり、超音波IDの認識精度が減少することを明らかにした。そこで、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、装着位置によって超音波ID認識のためのしきい値切り替えを行う改良手法を提案した。改良手法を実環境のイベントで運用したところ、装着位置の認識率は84.7%であり、超音波IDの認識率は従来手法では55.2%であったところ、改良手法では68.0%に向上した。

References

- 1) H. Watanabe, T. Terada, and M. Tsukamoto: Ultrasound-based Movement Sensing, Gesture-, and Context-recognition, in Proc. of the 17th International Symposium on Wearable Computers (ISWC 2013), pp. 57–64 (2013).
- 2) 渡邊拓貴, 寺田 努, 塚本昌彦: UltraSoundLog: 超音波IDによる音声ログへの位置情報付与, インタラクション2015論文集, pp. 136–143 (2015).
- 3) 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 Workshop (PERCOM Workshops), pp. 230–235 (2013).
- 4) 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).
- 5) 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 2002) Workshop: Ad hoc Communications and Collaboration in Ubiquitous Computing Environments (2002).
- 6) 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 2011), pp. 21–28 (2011).
- 7) 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).
- 8) Y. Lei, Y. Soon, F. Zhou, Z. Li, and H. Lei: A Robust Audio Watermarking Scheme Based on Lifting Wavelet Transform and Singular Value Decomposition, Signal Processing, Vol. 92, No. 9, pp. 1985–2001 (2012).
- 9) YAMAHA INFOSOUND, <http://www.yamaha.com/news/release/2012/20120613a.html?from=global search>.
- 10) M. Hirabayashi and M. Shimizu: Cryptone: Interaction between Performances and Audiences with Inaudible DTMF Sounds, SIGGRAPH Asia 2012 Emerging Technologies (2012).
- 11) 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 2013), pp. 31–40 (2013).

[Research Paper]

Person/Location-aware Sound Log System Using Ultrasonic ID

Hiroki WATANABE¹, Tsutomu TERADA¹, Masahiko TSUKAMOTO¹

¹*Graduate School of Engineering, Department of Electrical and Electronic Engineering*

Key words: Ultrasound, Sound Log, Person Recognition, Location Recognition, Wearable Computing

In wearable computing environments, various sensors enable the recording of user context as a life log. For enhancing the experience in a museum or exhibition, a voice recorder is useful for recording users' comments and provided explanations for the showpieces. However, to utilize the sound log after the event, the voice recorder should attach appropriate tags to the sound data for extracting intended data from the long-term sound log. In a previous study, we have proposed an activity and context recognition method where the user carries a neck-worn receiver comprising a microphone and small speakers on his/her wrists that generate ultrasound. Our system also recognizes the location of the user and the people whom are near the user by ultrasonic ID signals generated from speakers placed in rooms and on people. In this study, we focus on embedding location/person IDs to sound logs by using ultrasonic IDs. We actually used the proposed method during a real event, and we confirmed that the wearing position of the microphone affects the accuracy of ultrasonic ID recognition because the acquired volume of ultrasound is different at each position. We propose a new method for recognizing the position of a microphone to improve the accuracy of ID recognition, and we actually use the improved method at another real event. Evaluation results confirmed that the accuracy of position recognition was 84.7 %, and the accuracy of ID recognition was 68.0 % for the proposed method, improved from 55.2 % for the conventional method.