Vol.2013-UBI-39 No.9
2013/7/31

動作・人物・場所情報の超音波を用いた音声データへの埋め込み手法

渡邉 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，装着型センサを使った状況認識に注目が集まっている．一般的に用いられるセンサは加速度センサやマイクだが，前者は複数のセンサのデータを統合するために通信を行う必要があり，後者は音のみに頼っているため実際にそのユーザが関係している音なのかが分からない．そこで本研究では，超音波によってユーザの行動，周囲に居る人，現在居る場所などの情報を取得し，ボイスレコーダなどの音声記録に埋め込む手法を提案する．ユーザはマイクと超音波を発する小型スピーカを装着し，これらの距離を表す音量の変化と，ジェスチャの速度を示すドップラー効果を利用してジェスチャを認識する．また，環境や人に超音波 ID を発信する小型スピーカを装着することで，ユーザがどこにいたか，近くに誰がいたかという情報も同時に記録する．これにより，会話音等の環境音，ジェスチャ，ユーザのいた場所，会った人物のデータすべてがマイクのみで記録できる．提案手法では他者による環境音が無い場合，平均 86.6%の精度で認識でき，他者から発せられる環境音がある場合，平均 64.7%の精度で認識できた．

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高まっている．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，周りの環境などを認識することで，システムはその時のユーザの状況に合ったサービスを提供したり，日常生活のすべてをデジタルデータとして保存していくことを目指すライフログが可能になる．このような環境において，システムが適切なサービスを提供するためには，ユーザがどこで何をしているのかをセンサの値を用いてコンピュータに認識させる，状況認識技術が重要になる．一般的に状況認識に利用されるセンサは加速度センサ，マイク等であるが，装着型加速度センサの場合，ユーザ自身の動きは高い精度で認識できるが，どのような周辺環境においてユーザがその動作を行ったのかは認識できない．また，より複雑な行動を認識するには，複数のセンサを身体の各部に装着し，これらのデータを統合するために通信する必要がある．マイクを用いる手法では，環境音を解析することでユーザの周辺環境のコンテキストを認識できるが，取得した音がユーザに関係しているかどうかが分からない．そこで本研究では，ユーザがライフログサービス等のために音声記録用マイクを装着していることと，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，身体の部位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と組み合わせて状況認識を行う手法を提案する．また，環境や人に超音波 ID を発信する小型スピーカを装着することで，ユーザがどこに居たか，周囲に誰が居たかという情報も同時に記録できるため，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る．提案手法では，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，データ記録や処理のための通信が必要ない．

本研究では提案手法のプロトタイプを作成し，有効性を確かめるためにオフライン認識の評価実験を行った．ユーザ以外の他者や機器から発せられる環境音 (以降，他者環境音と呼ぶ) がある場合，コンテキストによっては認識精度が大幅に下がった．そこで，他者環境音が存在する場合には使用する特徴量を選択し，認識精度の向上を図った．

[1] 神戸大学大学院工学研究科
Graduate School of Engineering, Kobe University
[2] 科学技術振興機構さきがけ
PRESTO, Japan Science and Technology Agency
[a)] hiroki.watanabe@stu.kobe-u.ac.jp
[b)] tsutomu@eedept.kobe-u.ac.jp
[c)] tuka@kobe-u.ac.jp

また，人物認識，場所認識，シナリオに沿った連続するコンテキストについても評価を行った．

以降，2 章で本研究に関連する研究について述べ，3 章では認識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 章でまとめを行う．

2. 関連研究

2.1 加速度センサを用いた行動認識

加速度センサを用いてユーザの行動やジェスチャを認識する研究は幅広く行われている．Bao らの研究 [1] では，身体のそれぞれ別の箇所に装着された 5 つの 2 軸加速度センサのデータからジェスチャを検出する手法を提案している．村尾らの研究 [2] では，ボード上に配置された 9 種類の加速度センサと角速度センサを利用して 27 種類のジェスチャを認識し，センサの位置や数を変えることによって認識精度が変わることを示している．これらの研究では，会話のような，周辺環境の音に関係するコンテキストを認識するのは困難であると考えられる．さらに，それぞれのセンサのデータ統合のためにはセンサ間の通信が必要となる．携帯電話内蔵の加速度センサを用いた行動認識も行われている [3], [4]．多くの人々が個人のスマートフォンや携帯電話を持っているため，この手法を用いると日常生活行動の認識が現実的になるが，ユーザの手の動きなどの細かい動作は追加のセンサと組み合わせるなどしないと困難であると考えられる．

2.2 様々なセンサを用いた行動認識

行動認識には様々な種類のセンサが用いられている．Pirkl らの研究 [5] では，磁界を用いたジェスチャ認識の手法が研究されている．Starner らの研究 [6] では，ホームオートメーションシステムを手のジェスチャで操作するためのデバイスが提案されている．デバイスはペンダントのような形をしており，搭載された小型カメラによってジェスチャを認識している．Mattmann らの研究 [7] では，上半身の姿勢を検出するために，ひずみセンサを利用している．身体に密着するような服の背中部分にセンサが複数取り付けられており，身体の動きを認識している．Naya らの研究 [8] では，看護師の仕事認識システムが提案されている．このシステムでは，加速度センサによって看護師の動きが認識されており，赤外線センサにより場所が認識されている．Ward らの研究 [9] では両腕の手首と上腕部に装着した 2 個の 3 軸加速度センサとマイクによって，木工作業での連続するコンテキスト 9 種類 (のこぎり引き，ハンマー打ち，ドリル，サンドがけなど) を判別している．これらの研究では，複数装着されたセンサのデータを統合するために，データ蓄積または処理用のデバイスと通信をする必要がある．

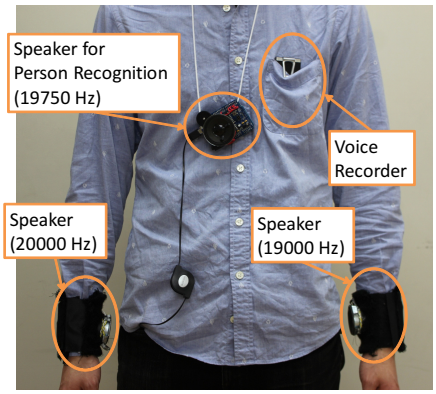

図 1 デバイスの構成

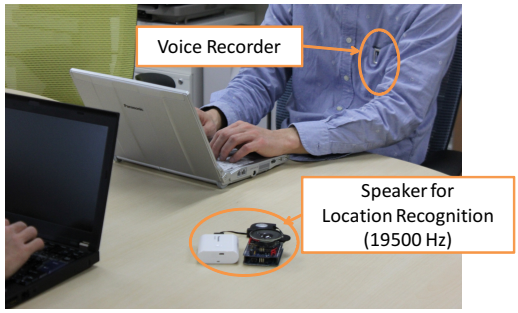

図 2 位置認識用スピーカの配置の様子

2.3 超音波を用いた行動認識

行動認識に超音波も用いられており，超音波を用いてユーザの位置を追跡する手法が提案されている [10], [11]．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．他のセンサと組み合わせることによって手の動きを認識するシステムも提案されている [12]．また，ドップラー効果を利用し，手のジェスチャを取得するシステムも提案されている [13]．これらの研究では，超音波は位置認識や，決まった場所でのジェスチャ取得にしか用いられていない．

3. 認識手法

本論文ではユーザのコンテキスト，位置，周囲に居た人物の 3 つを対象として認識している．筆者らの先行研究 [14] では，超音波を用いたジェスチャ認識と環境音認識の認識結果を組み合わせて，他者環境音あり/無しの場合のユーザのコンテキストのみを評価した．本論文ではコンテキストの認識に新たにドップラー効果を特徴量として使用し，認識手法を変更した．また，他者環境音が存在する場合に認識率が低下してしまうため，そのような環境下での新たな認識手法を提案した．さらに超音波 ID を発信する小型スピーカを用いて，位置と人物の認識を行った．

本研究の想定環境では，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机の上に，人

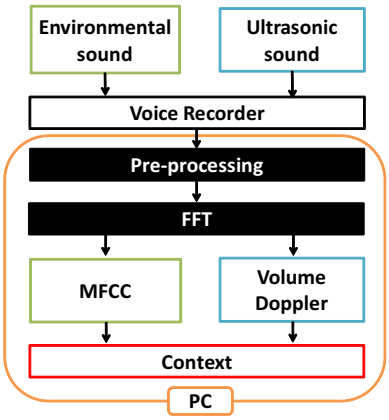

図 3 認識処理の流れ

物認識用のスピーカをユーザの胸に装着する．図 1 と図 2 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．本論文では，左手首のスピーカには 19000Hz を，右手首のスピーカには 20000Hz を使用した．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ボイスレコーダで同時に取得する．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，超音波を用いることとした．表 1 に示すように，ユーザは装着するデバイスによって，認識する対象を選択できる．マイクから取得した音を記録する際のサンプリング周波数は 44.1kHz とした．本章で述べる認識の処理の流れを図 3 に示す．以下，それぞれの処理について詳細に説明する．

3.1 音データの前処理

音の信号は低周波数成分が大きく，周波数が大きくなるにつれて次第に振幅スペクトルが小さくなっていくという特徴がある．この周波数の偏りを修正するために，以下の式に基づき高域を強調する処理を行う．N はサンプル数，x_n は高域強調処理前の n 番目の音データ $(n = 1, \cdots, N)$，x'_n は処理後の n 番目の音データとする．

$$x'_n = x_n - 0.97x_{n-1} \tag{1}$$

また，音データに処理を行う際に切り出した波形は，波形の始まりと終わりが不連続であり，フーリエ変換を使う際に不都合が生じる．そのため，切り出した波形に窓関数をかけ合わせて，切り出した境界部分を滑らかにする．本論文ではハミング窓を使用した．これらの処理を行ったデータを環境音と超音波に分離するため，高速フーリエ変換 (FFT: Fast Furier Transform) によって周波数スペクトルを求める．

3.2 ユーザ状況認識手法

一般的にセンサデータや音響データから状況の認識を行う際には，得られた値をそのまま使うのではなく，挙動を効率的に把握するために特徴量抽出と呼ばれる処理を行う．本研究では，環境音の特徴量としてメル周波数ケプストラム係数 (MFCC: Mel Frequency Cepstral Coefficient) を用いた．MFCC は人間の聴覚上重要な周波数成分を強調した特徴量であり，音声認識で一般的に利用される特徴量である [15]．得られた 20 次元の値のうち，一般的に低次成分 12 次元の値をとる．ジェスチャの特徴量は超音波の音量の平均と分散を用いた．音量 $x(T)$ の過去 10 サンプルの平均 $\mu(T)$ および分散 $\sigma^2(T)$ を以下の式に基づいて計算する．T は時刻を表す．得られる特徴量は 19000Hz と 20000Hz のそれぞれの平均，分散であるため，4 次元の特徴量となる．

$$\mu(T) = \frac{1}{10}\sum_{t=T-9}^{T} x(t) \tag{2}$$

$$\sigma^2(T) = \frac{1}{10}\sum_{t=T-9}^{T} \{x(t) - \mu(T)\}^2 \tag{3}$$

また，ジェスチャの動きの速さを考慮するためにドップラー効果を利用する．動きの速さによって周波数が変化するので，19000Hz と 20000Hz それぞれの前後 50Hz で最も音量の高い周波数を記録しておき，これの過去 10 サンプルの平均と分散を計算し，4 次元の特徴量として使用する．

以上の合計 20 次元の特徴量はスケールが異なり対等に扱うことができないため，次式に従い正規化し，20 次元の特徴量 $Z(T) = (z_1(T), \cdots, z_{20}(T))$ (平均 0，分散 1) を得る．ここで，M および S は $X(T) = (x_1(T), \cdots, x_{20}(T))$ の各成分の平均および標準偏差である．T は時刻を表す．

$$Z(T) = \frac{X(T) - M}{S} \tag{4}$$

認識方法には k 最近傍法 (k-NN: k-nearest neighbor algorithm) を用いた．それぞれのコンテキストに対して正規化した 20 次元の特徴量を事前に計算し，学習データ $Z_i = (z_{i1}, \cdots, z_{ij}, \cdots, z_{i20})$ と未知のデータの 20 次元特徴量の値 $Z = (z_1, \cdots, z_j, \cdots, z_{20})$ とのユークリッド距離 d_i を次式に従い計算し，すべての学習データのうち未知のデータとの距離が近いもののラベル i を認識結果とする (k = 1)．

$$d_i = \sqrt{\sum_{j=1}^{20}(z_{ij} - z_j)^2} \tag{5}$$

3.3 位置と人物の認識手法

位置認識用に環境に設置されたスピーカや，人物認識用に人に装着されたスピーカからは超音波 ID が発せられる．本論文では，位置認識用に用いられる周波数は 19500 Hz，環境認識用に用いられる周波数は 19750 Hz とした．ID はヘッダ (ex. 1010) とメイン ID の部分からなっており，そ

表 1　デバイスの選択と認識可能対象の組合せ

Worn device	Environment	Gesture	Location	Person
Recorder	○			
Recorder + Wrists[1]	○	○		
Recorder + Location[2]	○		○	
Recorder + Person[3]	○			○
Recorder + Wrists + Location	○	○	○	
Recorder + Wrists + Person	○	○		○
Recorder + Wrists + Location + Person	○	○	○	○

[1]Speakers for gesture recognition [2]Speaker for location recognition [3]Speaker for person recognition

れぞれのスピーカに異なる ID が割り当てられている．スピーカは 1 の区間だけ発音し，0 の区間は発音しない．システム側では，対応する周波数の値がしきい値を超えたときのみ，1 と認識するようにし，それ以外の場合は 0 と認識する．ヘッダを認識したときのみ，後に続くメイン ID の部分を認識する．本論文ではメイン ID を 4 ビットとした．それぞれの場所や人物にはそれぞれ違うメイン ID が割り当てられたスピーカが装着されており，これらと取得した ID とを比較して位置や人物を認識できる．メイン ID を長くすることで認識可能数を増やすことができる．ID の 1 パルスの長さは 0.5 秒とした．

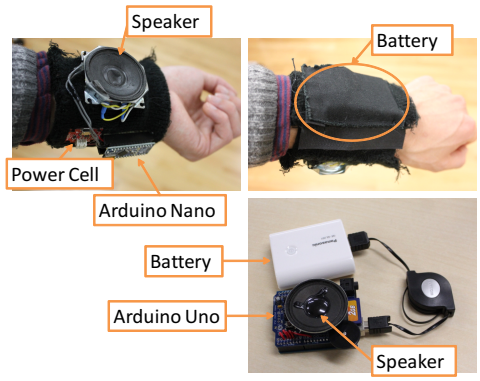

図 4　デバイス外観

4. 実装

提案手法に基づきプロトタイプを実装する．両手首に装着されるデバイスは，図 4 の上側に示すように，スピーカ，Arduino Nano ver. 3.0，リチウムポリマバッテリ (3.7 V，2000 mAh)，充電/昇圧器から構成されている．これらは着脱が可能なようにリストバンドに縫い付けられている．Arduino Nano で生成した矩形波をスピーカを通じて再生している．位置認識，人物認識に用いられるデバイスは，図 4 の下に示すように，スピーカ，Arduino Uno R3，Wave Shield ver. 1.1，リチウムイオンバッテリ (5V，5400 mAh) から構成されている．音声編集/録音ソフトの Audacity を用いて超音波 ID を生成し，これを SD カードに記録し，Wave Shield を通じて再生される．使用したボイスレコーダは Sony 社の ICD-TX50 であり，録音形式は LPCM(44.1 kHz，16 bit) である．取得したデータを解析するためのソフトウェアはマイクロソフト社の Visual C#を用いて開発した．使用した PC はパナソニック社の Let's Note CF-S9LYYKDS である．

5. 評価実験

5.1 他者環境音無しの場合

日常生活を想定し，他者環境音が無い場合のラベルとして着席，歩行，ランニング，食事，歯磨き，掃除機がけ，手洗い，タイピング，会話の 9 種類を想定する．食事はフォークを使ってスパゲッティを食べる場合，歯磨きは電動のものではなく一般的な歯ブラシ，会話は立った状態での会話を想定している．まず 20 次元全ての特徴量を用いた場合のコンテキスト判別精度を求め，次に特徴量の選択による認識精度の変化を求めた．被験者は 10 人であり，サンプリング周波数は約 10Hz である．被験者はそれぞれのラベルの行動を 45 秒間行い，学習データには被験者本人の各ラベルの 50 サンプルを用いた．評価用データには各ラベルの学習データに使用しなかった任意の 100 サンプルを用いた．評価実験で使用した掃除機は Dyson 社の DC26 である．また，今回用いたプロトタイプのスピーカではノート PC でのタイピングは困難であったため，ノート PC にキーボードを接続し，タイピングのデータを取得した．使用したキーボードは DELL 社の SK-8115 である．認識には k-NN 法 (k = 5) を用い，特徴量を計算するためのウインドウサイズは，環境音認識は 4096，ジェスチャ認識は 10 とした．

被験者 10 人の平均の実験結果を表 2 に示す．MFCC を特徴量として用いた場合，平均の認識率は 74.4%，音量を用いた場合 78.7%，ドップラー効果を用いた場合 48.7%であった．MFCC のみを特徴量として用いた場合，定常的に音を発しないランニング，食事，会話などは認識率が良くなかった．一方，着席や掃除機などの特徴的な音を持つコンテキストでは認識率が良かった．音量を特徴量としたとき，歯磨きは高い認識率を得られた．これは，歯磨きの時にはユーザの利き手だけがマイクに近づき，特徴的な

表 2　他者環境音無しの場合の認識精度 [%]

Combination of feature values	Sitting	Walking	Running	Eating	Typing	Brushing	Washing	Cleaning	Talking	Average
MFCC	90.6	80.4	60.6	63.4	70.1	66.6	77.6	100	60.5	74.4
Volume	85.1	77.3	68.4	64.0	81.4	95.5	79.9	73.8	82.8	78.7
Doppler	97.5	60.1	62.7	41.2	24.8	41.2	39.1	48.0	23.7	48.7
$M^1+ V^2$	93.2	90.5	68.2	77.3	86.1	78.3	86.6	100	70.7	83.4
$V + D^3$	88.7	67.8	69.8	70.2	91.6	84.1	80.0	69.6	89.2	79.0
M + D	97.7	76.6	73.6	70.2	82.6	67.3	82.3	100	66.1	79.6
All	98.9	82.8	76.0	82.0	91.8	78.5	88.0	100	81.5	86.6

[1]MFCC [2]Volume [3]Doppler

値を取ったためだと考えられる．ドップラー効果のみでは，これらのコンテキストでは良い認識率を得られなかった．2 種類の特徴量を組み合わせた場合，平均の認識率が 13.4%上昇した．これは，それぞれの特徴量が認識率の低い部分を補いあったためだと考えられる．3 種類全ての特徴量を用いた場合，平均の認識は 86.6%であった．以上より，特徴量を組み合わせることは有効であるといえ，他者環境音がない場合には 3 種類全ての特徴量を使用して認識するべきであると考えられる．

5.2　他者環境音ありの場合

他者環境音がある場合についても評価を行った．他者環境音には，特徴的な環境音だと考えられる歯磨き，掃除機がけ，手洗い，タイピング，会話の 5 つを選択し，これらの環境音のもとで表 3 に示すような組合せの行動を評価した．この環境音のもとでのランニングは考えられにくいため，本論文では考慮しない．被験者は他者環境音無しの場合と同じであり，それぞれの行動を約 45 秒間行った．サンプリング周波数は約 10Hz である．学習データは他者環境音無しの時と同じものを用いた．

まず，全ての特徴量を用いた認識を行ったところ，表 4 に示すように，全体的に認識率が下がっており，平均の認識率は 57.3%であった．これは，他者環境音が認識に悪影響を及ぼしているからだと考えられる．特に，掃除機の音は他のコンテキストの音と比べて十分大きいため，他者環境音が掃除のとき，ユーザの本来のコンテキストの音をかき消してしまい，特徴量に MFCC を含んでいる場合の認識率が大幅に下がっている．

そこで，3 種類の特徴量を用いた認識結果と音の特徴量である MFCC を除いた特徴量で認識した結果が食い違う場合は，他者環境音があるときと想定し，他者環境音を排除するために，MFCC を除いた，音量，ドップラー効果の 2 種類の特徴量のみを用いた場合の認識結果を用いる．この改良手法の認識の流れを図 5 に，認識結果を表 5 に示す．認識率が下がったコンテキストもあるが，特に認識率が低かった，他者環境音が掃除の場合においては 45%程の認識率の大幅な改善がみられ，全体の平均では 7.4%の改善がみられた．

表 3　他者環境音とユーザの行動の組合せ

User's action	Environmental sound of others
Sitting	Typing, Washing, Brushing, Cleaning, Talking
Walking	Typing, Washing, Brushing, Cleaning, Talking
Running	-
Eating	Typing, Talking
Typing	Typing, Washing, Brushing, Cleaning, Talking
Brushing	Typing, Washing, Brushing, Cleaning, Talking
Washing	Typing, Brushing, Cleaning, Talking
Cleaning	Typing, Washing, Brushing, Talking
Talking	Typing, Washing, Brushing, Cleaning, Talking

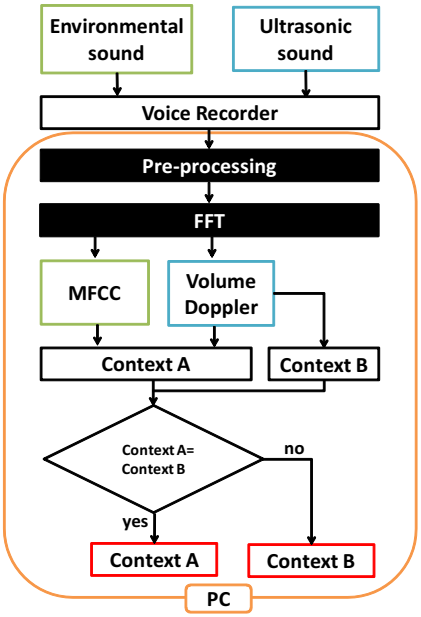

図 5　他者環境音がある場合の認識処理の流れ

5.3　位置と人物の認識

位置認識，人物認識についても評価を行った．5 つの部屋の机の上にそれぞれ異なる ID を発する超音波スピーカを設置する．被験者は一人である．被験者は机の前の椅子

表 4 他者環境音がある場合の認識精度 [%]

Other's environmental sound	Feature values	Sitting	Walking	Eating	Typing	Brushing	Washing	Cleaning	Talking	Average
Typing	MFCC	17.8	65.3	17.1	72.3	29.8	59.8	100	54.5	52.1
	Volume	37.7	53.6	32.6	55.9	75.2	58.8	50.7	60.5	53.1
	Doppler	89.6	52.2	45.6	23.8	29.7	31.8	43.4	13.2	41.2
	M[1] + V[2]	16.9	80.7	32.3	67.0	48.9	76.7	100	62.9	60.7
	V + D[3]	59.4	59.5	40.9	66.5	67.2	59.2	53.3	77.7	60.5
	M + D	47.0	68.6	27.4	91.5	38.1	64.0	100	57.4	61.7
	All	50.9	74.3	37.8	83.7	50.7	82.1	99.9	71.9	68.9
Washing	MFCC	17.6	57.3	-	43.6	17.6	-	99.3	50.5	47.7
	Volume	43.1	55.1	-	45.1	77.4	-	36.7	61.8	53.2
	Doppler	78.3	57.7	-	24.0	20.8	-	44.0	16.6	40.3
	M + V	23.5	75.0	-	44.6	39.6	-	99.5	62.8	57.5
	V + D	57.2	63.4	-	55.8	63.7	-	48.5	82.8	61.9
	M + D	54.4	68.5	-	66.6	19.6	-	99.3	54.9	60.5
	All	71.4	73.9	-	68.5	41.8	-	99.2	73.4	71.3
Brushing	MFCC	43.3	71.5	-	58.1	34.9	68.8	100	54.7	61.6
	Volume	58.9	60.2	-	44.0	67.7	74.5	38.1	68.1	58.8
	Doppler	95.7	59.1	-	28.7	33.7	40.0	45.0	10.2	44.6
	M + V	58.5	85.0	-	53.1	51.9	84.8	100	65.4	71.2
	V + D	70.4	66.8	-	62.6	59.8	75.2	51.2	79.5	66.5
	M + D	74.2	73.1	-	83.4	41.2	74.6	99.9	55.2	71.6
	All	82.5	77.9	-	78.9	51.6	90.7	99.8	73.5	79.3
Cleaning	MFCC	0.0	0.0	-	0.0	0.0	0.9	-	2.6	0.6
	Volume	41.8	67.0	-	47.8	76.4	63.2	-	59.4	59.3
	Doppler	65.1	48.1	-	24.4	40.4	34.6	-	5.3	36.3
	M + V	0.0	0.1	-	1.6	0.1	1.7	-	8.4	2.0
	V + D	39.6	56.2	-	59.5	73.4	63.0	-	29.2	53.5
	M + D	0.0	6.2	-	1.4	0.0	4.4	-	2.2	2.4
	All	0.6	16.6	-	11.2	1.9	5.6	-	7.2	7.2
Talking	MFCC	16.9	32.6	17.8	17.2	11.8	46.4	99.9	64.6	38.4
	Volume	59.0	67.8	37.9	57.5	76.7	80.9	41.4	60.5	60.2
	Doppler	88.2	55.9	42.0	26.1	26.5	39.1	50.7	11.0	42.4
	M + V	23.2	46.2	30.2	20.3	29.0	72.5	99.9	76.5	49.7
	V + D	59.7	62.9	45.2	66.0	66.3	76.9	55.0	84.5	64.6
	M + D	28.1	56.5	30.8	37.4	17.1	54.9	99.9	68.1	49.1
	All	34.6	64.0	37.6	45.1	34.5	79.4	99.7	84.3	59.9

[1]MFCC [2]Volume [3]Doppler

に 30 秒間座り，次の部屋に移動する．これを 2 セット繰り返し，位置認識の精度を確認する．実験を行った 10 箇所のうち，10 箇所を認識することができたが，そのうち 1 箇所では，初めは正確に認識できていたが，途中から違う部屋と認識してしまっていた．これは，ID 取得の途中でスピーカの向きや音量の不足などの何らかの影響で音が十分に届かず，誤認識を続けたままだったからだと考えられる．また，同室の 5 箇所で同時にこれらを用いた場合，それぞれの ID が干渉して，正確に認識することができなかった．よって，人物認識などのように，互いの音が干渉しあう距離で用いるときは，ID の発信タイミングをずらすなど工夫をするべきだと分かる．

人物認識については，3 人の被験者で行った．超音波 ID を発するスピーカを胸に装着した 3 人のユーザの内，一人がボイスレコーダを胸に装着し，約 2 分間向かい合って会話をした結果，約 1 分間ほどで 3 人全員の被験者を認識できていた．本システムでは，同時に認識する人数が増えるほど全員を認識するのに時間がかかってしまう．今回は 1

表 5　従来手法と改良手法との認識精度の比較

Sound emitted from others	Accuracy [%]	
	Conventional method	Revised method
Typing	68.9	64.0
Washing	71.3	67.2
Brushing	79.3	74.8
Cleaning	7.2	53.6
Talking	59.9	63.8
Average	57.3	64.7

パルスの長さを 0.5 秒と長く設定したため，その傾向が顕著であった．1 パルスの長さを短くすることで認識までの時間を短くできると考えられるので，今後適切なパルスの長さを調査していく必要がある．

5.4　行動シナリオ認識

日常生活での実用性を確かめるために，表 6 に示すようなシナリオに沿った日常生活行動の認識を評価した．被験者は 3 人の男性であり，録音時間は約 10 分間，移動は全て歩行とした．行動を行う一人のユーザはボイスレコーダを胸に，超音波スピーカを両手首に装着し，残り二人の被験者は person A，person B として，人物認識用 ID を発するスピーカを胸に装着する．位置認識用の超音波スピーカは図 6 に示すように 3 箇所に設置する．図 7 は行動シナリオの正解データを，図 8 は行動シナリオの実験結果を表している．認識結果は過去 10 秒間の認識結果の多数決を表している．図 8 が示すように，コンテキストはほとんど正確に認識されている．しかし，Room B でのタイピングが認識されていなかった．これは，Room B で使用したキーボードと学習データに使用したキーボードの種類が違ったためだと考えられる．Room A で用いたキーボードは学習で用いたものと同じであり，認識できていた．位置認識に関しては，5 箇所のうち，4 箇所が正確に認識できていた．ユーザが person A と会話しているときに位置情報が記録されていなかったが，これは音量が十分でなかったためだと考えられる．Room B での行動は全てスピーカの近くで行われたが，person A との会話は Room A に設置されたスピーカから離れた場所で行われた．そのため，超音波が十分にレコーダに届かず，位置 ID を認識しなかったと考えられる．より正確に認識するには，スピーカの音量を大きくするか，音が被らない範囲で複数のスピーカを設置するべきである．

person A と person B はどちらも正しく認識されていた．ユーザが掃除機を使用しているときに，位置や人物の誤認識が起こることがあった．これは，掃除機の音が位置認識，人物認識にしている 19500Hz と 19750Hz を含んでいるからである．これを解決するために，掃除機に含まれる該当周波数の音量では誤認識しないよう適切なしきい値を予備実験から設定するか，より高い，掃除機の音に含まれていないような周波数を位置と人物認識用の周波数に設定するなどの方法が考えられる．

表 6　行動シナリオ

# of context	Context	Location	Environmental sound of others
1	Typing	Room A	Talking
2	Eating	Room B	Typing
3	Sitting		
4	Talking with A	Room A	
5	Cleaning		
6	Talking with B	Room B	
7	Typing		Typing
8	Brushing	Room C	
9	Washing		

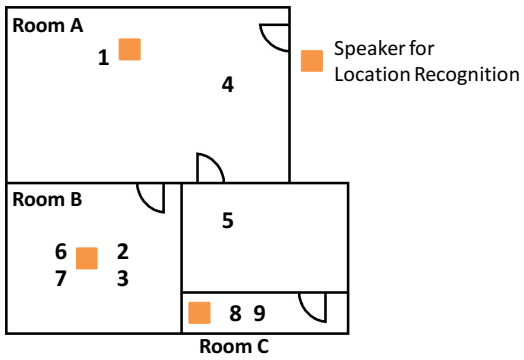

図 6　スピーカの配置とコンテキストの行われた位置

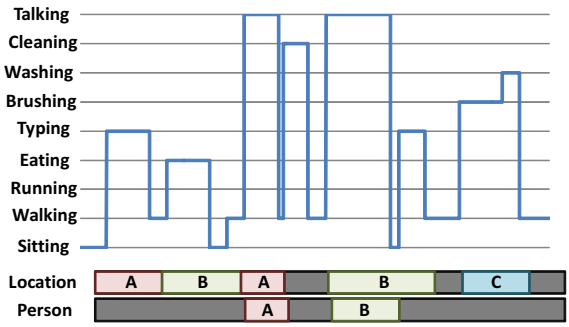

図 7　行動シナリオの正解データ

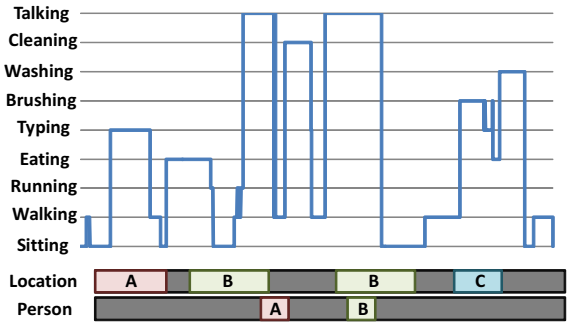

図 8　行動シナリオの実験結果

6. 考察

従来の音声ログではトイレにいるときなどのような，人に聞かれたくない音まで記録してしまうプライバシの問題があったが，提案手法では記録した音声に適切なタグ付けをすることにより，そういった音声を削除するか暗号化することで解決できると考えられる．本論文では一般的なボイスレコーダを用いてオフライン解析を行ったが，スマートフォンのような，録音とデータ解析ができるデバイスを用いればオンラインの認識も可能であると考えられる．20000Hz 付近の音は人間にはほとんど知覚できないが，犬や猫などの人間以外の動物は人間に比べて幅広い可聴領域を持っているので，本論文で用いたような周波数はペットにとって不快なものとなる可能性がある．ペットなど動物の近くで使用する場合には音量の設定や，使用する周波数の選択を慎重にする必要がある．超音波は壁に反射するため，壁が近くにあるような狭い場所では反射が誤認識を引き起こす可能性がある．本実験は十分に広い部屋で行われたため，そのような影響は見られなかった．本論文では，身体で一番動きがある両手にスピーカを，安定している場所として胸にレコーダを装着したが，どの装着箇所が最も良いのか，装着数を増やす又は減らすとどうなるのか等の調査の余地がある．

7. まとめ

本論文では超音波を用いた，音のみによる行動認識を提案した．他者環境音がない場合，86.6%の認識率で，他者環境音がある場合，改良手法を用いた場合 64.7%で認識できた．また，位置認識，人物認識，行動シナリオについての評価も行った．今後の課題として，スピーカやデバイスの小型化や認識アルゴリズムの改良を予定している．また，超音波と人間の健康問題との関係がまだ明らかでないので，超音波による人間の健康問題や心理問題についての調査や，本システムによる認識可能数の限界の調査も行っていく予定である．

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) および文部科学省科学研究費補助金基盤研究 (A)(20240009) によるものである．ここに記して謝意を表す．

参考文献

[1] Bao, L. and Intille, S. S.: Activity recognition from user-annotated acceleration data, *Pervasive Computing*, Springer, pp. 1–17 (2004).

[2] Murao, K., Terada, T., Yano, A. and Matsukura, R.: Evaluating gesture recognition by multiple-sensor-containing mobile devices, *Wearable Computers (ISWC), 2011 15th Annual International Symposium on*, IEEE, pp. 55–58 (2011).

[3] Iso, T. and Yamazaki, K.: Gait analyzer based on a cell phone with a single three-axis accelerometer, *Proceedings of the 8th conference on Human-computer interaction with mobile devices and services*, ACM, pp. 141–144 (2006).

[4] Duong, T. V., Bui, H. H., Phung, D. Q. and Venkatesh, S.: Activity recognition and abnormality detection with the switching hidden semi-markov model, *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, Vol. 1, IEEE, pp. 838–845 (2005).

[5] Pirkl, G., Stockinger, K., Kunze, K. and Lukowicz, P.: Adapting magnetic resonant coupling based relative positioning technology for wearable activitiy recogniton, *Wearable Computers, 2008. ISWC 2008. 12th IEEE International Symposium on*, IEEE, pp. 47–54 (2008).

[6] Starner, T., Auxier, J., Ashbrook, D. and Gandy, M.: The gesture pendant: A self-illuminating, wearable, infrared computer vision system for home automation control and medical monitoring, *Wearable Computers, The Fourth International Symposium on*, IEEE, pp. 87–94 (2000).

[7] Mattmann, C., Amft, O., Harms, H., Troster, G. and Clemens, F.: Recognizing upper body postures using textile strain sensors, *Wearable Computers, 2007 11th IEEE International Symposium on*, IEEE, pp. 29–36 (2007).

[8] Naya, F., Ohmura, R., Takayanagi, F., Noma, H. and Kogure, K.: Workers' routine activity recognition using body movements and location information, *Wearable Computers, 2006 10th IEEE International Symposium on*, IEEE, pp. 105–108 (2006).

[9] Ward, J. A., Lukowicz, P., Troster, G. and Starner, T. E.: Activity recognition of assembly tasks using body-worn microphones and accelerometers, *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, Vol. 28, No. 10, pp. 1553–1567 (2006).

[10] Ward, A., Jones, A. and Hopper, A.: A new location technique for the active office, *Personal Communications, IEEE*, Vol. 4, No. 5, pp. 42–47 (1997).

[11] Muller, H. L., McCarthy, M. and Randell, C.: Particle filters for position sensing with asynchronous ultrasonic beacons, *Location-and Context-Awareness*, Springer, pp. 1–13 (2006).

[12] Ogris, G., Stiefmeier, T., Junker, H., Lukowicz, P. and Troster, G.: Using ultrasonic hand tracking to augment motion analysis based recognition of manipulative gestures, *Wearable Computers, 2005. Proceedings. Ninth IEEE International Symposium on*, IEEE, pp. 152–159 (2005).

[13] Gupta, S., Morris, D., Patel, S. and Tan, D.: SoundWave: using the doppler effect to sense gestures, *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems*, ACM, pp. 1911–1914 (2012).

[14] 渡邉拓貴，寺田　努，塚本昌彦：超音波を用いたジェスチャ検出と環境音検出を組み合わせた状況認識手法，*DICOMO2012*, pp. 157–164 (2012).

[15] 大内一成ほか：加速度と音による家庭内ユーザ状況認識の可能性検討，*DICOMO2010*, pp. 508–515 (2010).