

ウェアラブル端末・スマートフォンを利用した 生活行動認識の分散処理方式

大橋純^{†1} 須藤隆^{†1} 大内一成^{†1} 斉藤龍則^{†1}

スマートフォンのマイクから取得した音データを使用して生活行動を認識する技術が提案されている。従来技術では、例えば自宅内であっても常にスマートフォンを保持して生活しなければならず、より携帯性の高いウェアラブル端末を利用した生活行動認識技術の確立が望まれている。しかしながら、音を使用した行動認識手法は計算量が多く処理性能が低いウェアラブル端末で動作させることは困難だった。本論文では、計算量とデータ通信量を勘案し、音データの特徴量抽出までをウェアラブル端末で実行し、音特徴量を利用した生活行動認識をスマートフォンで実行する分散処理方式を提案し、分散処理の境界について妥当性を評価した。また特徴量抽出処理の間欠動作や音データ量の削減により、計算量や消費電力の削減が可能であることを示した。本システムで日常的な6行動を平均95%以上の精度で認識できることを確認した。

A Distributed Processing System of Living Activity Recognition Using a Wearable Device and a Smartphone

JUN OHASHI^{†1} TAKASHI SUDO^{†1}
KAZUSHIGE OUCHI^{†1} TATSUNORI SAITO^{†1}

An activity recognition method using a microphone installed in a smartphone was proposed. However, the method requires us to always put on a smartphone within the home, establishing activity recognition method by more portable wearable device is desired. Since the activity recognition using audio data requires a large amount of computation, it is difficult to process it by a wearable device. In consideration of the amount of data communication and computation, we propose a distributed processing system which a wearable device processes up to the acquisition of acoustic features and a smartphone recognize activities using acoustic features, and we verify the boundary of the distributed processing. We confirm to be able to reduce power consumption and amount of computation by reducing the amount of the sound data and the intermittent operation of the acoustic feature extraction. We confirm that our system can classify 6 usual activities with more than 95% accuracy on average.

1. はじめに

高齢化社会の進展に伴い、高齢者を対象とした見守りサービスが要求されている。高齢者の見守りサービスは、比較的検出が容易である安否確認だけでなく、生活行動パターンの変化から認知症の初期症状を検出するなどより高度な行動認識技術に基づいたサービスへの期待が高まっている。

行動認識の技術としては、スマートフォンやウェアラブル端末に搭載された加速度センサの情報を利用する手法が広く利用されている。この加速度センサを利用した行動認識手法により認識できる行動は、歩行や走行、自転車や自動車の乗車といった移動手段や、座る・立つといった姿勢が主であり、認識できる行動種別が限定的なため、高齢者を対象とした見守りサービスとしては十分ではない。

その解決策として、スマートフォンのマイクから取得した音の特徴量を使用して、屋内の生活行動を認識する技術

が提案されている[1]。音の情報を利用することにより、移動手段や姿勢だけでなく作業の内容が認識可能となる。しかしながら屋内、とくに自宅での生活行動認識を考えた場合、スマートフォンを常時携帯して行動することは想定しがたく、スマートフォンよりも常時携帯性に優れたウェアラブル端末を使用した生活行動認識技術の確立が望まれている。音を利用した生活行動認識は、加速度を使用した技術と比較して計算量が多いため、ウェアラブル端末に搭載することは困難であり、常時携帯することを考慮すると消費電力も重要な課題となる。

本論文では、計算量とデータ通信量を勘案し、音データの特徴量抽出までをウェアラブル端末で実行し、音特徴量を利用した生活行動認識をスマートフォンで実行する分散処理方式を提案する。そして、データ通信量・計算量・メモリ使用量から提案する分散処理の境界について妥当性を評価した。

また提案する分散処理方式のシステムを設計・試作し、特徴量抽出処理の間欠動作や音データ量の削減による認識性能への影響を評価し、認識対象に応じて計算量や消費電

^{†1} 株式会社 東芝
TOSHIBA CORPORATION.

力の削減が可能であることを示し、提案方式の有効性を確認した。

2. 加速度センサを利用した行動認識

ウェアラブル端末に搭載された加速度センサを利用した行動認識の研究として、複数の加速度センサを利用する研究[2]や、単一の加速度センサを利用する研究[3]がある。

文献[2]では、腕時計型ウェアラブル端末に搭載された6個の加速度センサおよび照度センサを使用し、階段を上る、階段を下る、歩行、走行、立つ、座るという6状態を認識する技術が提案されている。また文献[3]では、単一の3軸加速度センサにより座る、立つ、歩く、走るといった動作を認識する手法が提案されている。

いずれの研究もあくまで動作の認識であり、見守りサービスで重要となる作業状態を推定することはできないという課題がある。

3. マイクを利用した生活行動認識

3.1 処理概要

マイクを利用した生活行動認識の処理概要を図1に示す[1][4][5]。

加速度センサから取得したデータから、動作を「歩行」、「作業」、「静止」の3状態に分類する。「歩行」は階段の昇降や走行を含む移動状態である。「作業」は「歩行」以外の身体的な動作を伴う生活行動を実行していると想定される状態である。「歩行」「作業」のいずれでもない、身体を動かしていない状態を「静止」状態とする。

動作が「作業」であった場合、マイクから周囲の音を取得し、音を分析することにより作業内容を推定する。

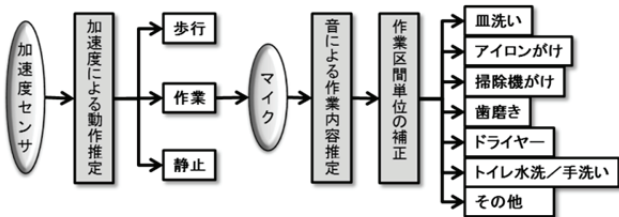


図1 生活行動認識処理の概要

3.2 加速度による動作推定

加速度による動作推定では、ウェアラブル端末に搭載した加速度センサから取得したデータを元に行動を分類する。ウェアラブル端末で処理することを想定しているため、低処理量・低消費電力の手法が要求されるが、胸ポケットに入れたスマートフォンの10bit 20Hzの3軸加速度センサを用い、1秒間の分散のみを用いる簡易な方法により、「歩行」、「作業」、「静止」の3行動を95%以上の精度で推定可能で

ある[4]。

3.3 音による作業内容推定

動作が「作業」の場合、マイクから周囲の音を取得し作業内容を推定する。音による作業内容推定の概要を図2に示す。

音による作業状態分類では、事前に作業者本人が分類対象の作業状態の音データを学習しておく。分類器には、高い汎化性能を持つサポートベクタマシン (SVM: Support Vector Machine) を用いる。事前学習は利用者の負担にならないよう短時間で実施することが必要であると考えており、学習時間は10秒間を想定している。

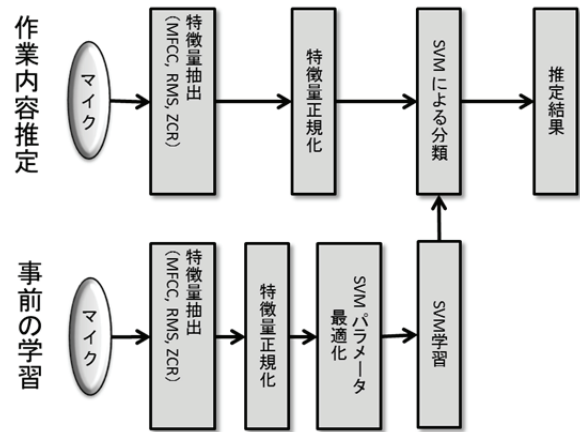


図2 音による作業内容推定

音データから抽出する特徴量としては、広く音声認識に使われているメル周波数ケプストラム係数 (MFCC: Mel-Frequency Cepstrum Coefficients) に加え、算出に際して計算量の比較的少ない二乗平均平方根 (RMS: Root Mean Square)、ゼロ交差率 (ZCR: Zero-Crossing Rate) を使用し、合計15次元 (MFCC: 13次元, RMS: 1次元, ZCR: 1次元) の特徴量とし、1秒毎にそれぞれの平均値を算出する。

抽出した合計15次元の特徴量は、それぞれでスケールが異なるため、学習データから取得した各特徴量の最大、最小値を用い±1で正規化を行う。

作業内容推定は、正規化された特徴量に対しSVMによる分類を実行し、作業状態の分類結果を1秒毎に得る。SVMライブラリにはLIBSVM[6]を利用し、SVMタイプはC-SVC、カーネル関数はRBF (radial basis function) を使用した。学習時にはLIBSVMが提供する最適パラメータ(C,)をヒューリスティックに探すGrid searchを用いて最適パラメータを取得し、そのパラメータで学習を行う。

加速度による動作推定により、動作が「作業」からそれ以外の動作に遷移したと判断された場合、作業区間単位の補正を行って「作業」の内容を推定する。区間補正では、作業区間内において50%以上の頻度で認識された行動をその作業区間の行動とする。50%以上の頻度で認識された行

動がなかった場合、その区間の行動は推定対象外(未学習)の作業である「その他」の行動として扱う。

3.4 従来システムの課題

スマートフォンを利用した生活行動認識技術[1][4]は、屋内、特に自宅で常時スマートフォンを携帯することは現実的ではないため、常時携帯性に優れたウェアラブル端末を使用したシステムの実現が望まれる。しかしながら、マイクを利用した生活行動認識技術は、計算量やメモリ使用量、消費電力の観点からすべての処理をウェアラブル端末で実行することは困難である。文献[5]では、ウェアラブル端末では加速度センサによる動作推定と、動作が「作業」の場合に音を Bluetooth™ の A2DP(Advanced Audio Distribution Profile)により PC に転送する処理のみ実行しているが、「作業」中に常時 PC と通信する必要があり、通信の消費電力の大きさにより長時間使用することは困難である。また通信時に A2DP では音を圧縮するため音質の劣化による認識精度への影響も懸念される。さらに電波状況の悪化等により通信が切断した場合、その区間の状態が認識できないといった課題がある。

4. 分散処理方式の提案

本章では、3.4 節で述べた課題を解決するため、ウェアラブル端末・スマートフォンを利用した生活行動認識の分散処理方式を提案する。本方式は、常時携帯するデバイスとして、マイクを搭載したウェアラブル端末を用いることにより携帯性が向上する。また、ウェアラブル端末でマイクから取得した音データを特徴量に変換した上でスマートフォンに送信することで、ウェアラブル端末での処理を省電力化する。

4.1 システム概要

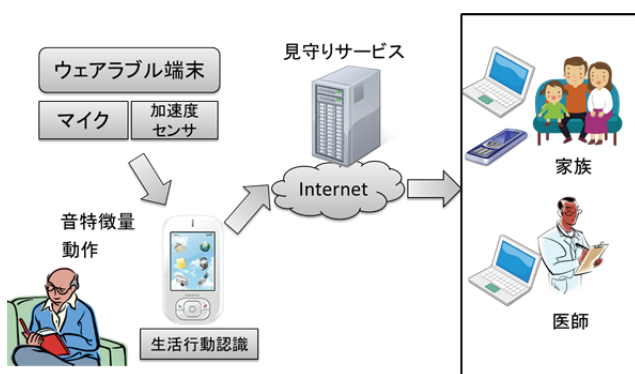


図 3 提案する分散処理方式のシステム概要

提案する屋内を想定した生活行動認識の分散処理方式のシステム概要を図 3 に示す。見守り対象者は衣服の胸近辺に加速度センサとマイクを備えたウェアラブル端末を装着し、加速度センサにより認識された歩行や作業、静止といった動作、動作が作業中の場合はマイクから取得した音の

特徴量を BLE(Bluetooth™ Low Energy)等の低消費電力な近距離無線技術により、宅内に置かれたスマートフォンに送信する。スマートフォンはウェアラブル端末から受け取ったデータから、見守り対象者の行動を認識し、インターネット上で運用される見守りサービスに認識結果の行動を送信する。見守りサービスでは、行動履歴から見守り対象者の安否や認知症の兆候の解析を行い、必要であれば家族や医師に通知する。家族や医師は PC やスマートフォンで見守りサービスにアクセスすることで、いつでも見守り対象者の状態を確認することができる。

4.2 処理の分散化

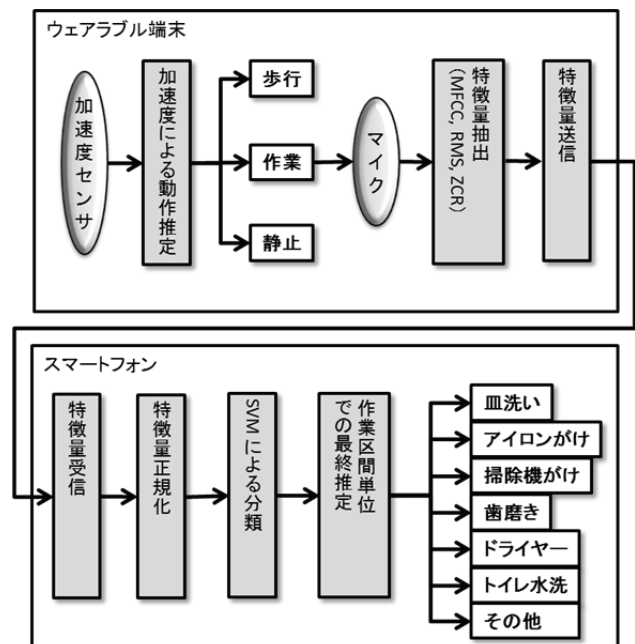


図 4 提案する分散処理方式

本節では図 1, 図 2 に示した生活行動認識処理について、ウェアラブル端末の計算量やデータ通信量を削減するという観点から、特徴量抽出処理までをウェアラブル端末で実行し、抽出した特徴量をスマートフォンに転送し、それ以降の処理をスマートフォンで実行する分散処理方式を提案する(図 4)。

提案方式において、ウェアラブル端末で最も計算量の大きい処理は特徴量抽出処理である。特徴量抽出処理の計算量は入力データ量に比例する。マイクから入力される音のデータ量は、サンプリングレート $F_s [Hz]$ 、量子化ビット数 $Q_b [bit]$ 、チャンネル数を C_n とすると

$$\text{入力データ量} = F_s \cdot Q_b \cdot C_n [bps]$$

で表せる。

一方ウェアラブル端末とスマートフォン間で通信されるデータは、動作を表すデータと動作が作業中の場合に送信される音特徴量データがある。動作を表すデータと比較してデータ量が大きい音特徴量のデータ通信量は、特徴量

のデータサイズを $A_s[bit]$ 、特徴量の次元数を A_n とすると、
 音特徴量データ通信量 = $A_s \cdot A_n[bps]$
 で表せる。

ここで、特徴量抽出処理を毎秒実行するのではなく、1秒実行して n 秒動作を停止するといった間欠動作も計算量やデータ量通信の削減に有効である。この動作間隔を $I_t[sec]$ とすると、作業中における音特徴量の平均データ通信量は

平均音特徴量データ通信量 = $A_s \cdot A_n / (I_t + 1)[bps]$
 になる。

本方式では音を特徴量に変換した上でスマートフォンに送信するため、音の圧縮による認識精度の低下は発生せず、通信データ量を削減することが可能となる。また電波状況の悪化等により通信が一次切断した場合でも、ウェアラブル端末のストレージに長時間保存しておくことができるため、通信状態が回復した際に再送することでデータロス回避することも可能である。

4.3 特徴量抽出の省電力化

ウェアラブル端末は常時使用するため、ウェアラブル端末で実行する処理はできるだけ省電力化することが望まれる。4.2 節に消費電力に相関のある計算量やデータ通信量に影響するパラメータを示した。なかでも、 F_s と Q_b を削減することで、計算量を大きく削減することができ省電力化に対する効果が大きい。文献[1][4][5]では F_s は16k、 Q_b は16に固定されているが、認識精度に影響しない範囲で削減可能であると期待できる。

また I_t を大きく取ることで、大幅な省電力化が見込めるが、認識精度とのトレードオフが想定される。ただし認識対象である行動の区間長を超えない一定範囲内では認識精度への影響が軽微であると期待できる。

これらの特徴量抽出のパラメータを認識精度が劣化しない範囲で小さくすることで、省電力化すると期待できる。

5. 評価

提案する生活行動認識の分散処理方式について評価を実施した。5.1 節では、提案手法において、ウェアラブル端末とスマートフォンの処理境界を特徴量抽出処理に設定しているが、本方式の妥当性について評価を行う。5.2 節では電力消費やメモリ使用量に影響するパラメータを変更した場合の認識精度を評価する。5.2.1 節では評価環境を示す。5.2.2 節では、省電力化に寄与する F_s と Q_b を削減した場合の認識精度について評価する。5.2.3 節では予備実験によりメモリ使用量に影響することを確認したウィンドウ幅を削減した際の認識精度を評価する。5.2.4 節では、 I_t を変更して認識精度を評価し、認識精度への影響が軽微である最長間隔を検証する。5.3 節ではウェアラブル端末向けに最適化したプログラムを利用して、提案方式がウェアラブル端末

で実現可能か、認識精度、計算量、メモリ使用量の観点から検討する。

5.1 分散処理の境界の評価

本節では4.2 節で提案した処理境界が妥当であるか、ウェアラブル端末でのデータ通信量や計算量、メモリ使用量の観点から評価する。

データ通信量の観点では、音データの F_s 16k、 Q_b が16、 I_t が0の場合、データ量は256kbpsである。特徴量のデータサイズ A_s が32、次元数 A_n を15とすると特徴量抽出により480bpsまでデータ量を圧縮することができる。特徴量正規化ではデータ量は削減されないが、行動種別が6種類の場合SVMによる分類により、3bpsまで削減可能である。さらに、5.2.4 節で述べる今回評価した6行動の平均区間長が82.7sec(表4)であることから、作業区間単位での最終推定によりデータ通信量は平均0.048bpsまで圧縮される。

一方計算量・メモリ量の観点では、SVMによる分類は計算量やメモリ使用量が大きく、消費電力の観点からもウェアラブル端末ではなくスマートフォンで実行することが望ましい。そのため、特徴量抽出後もしくは特徴量正規化後が処理境界の候補となる。特徴量正規化は計算量の大きい処理ではないが、本処理によりデータ通信量は削減されないため、ウェアラブル端末で実行する必要性はない。

本評価により、提案方式である特徴量抽出後に処理境界を設定するのが最適であることを確認した。

5.2 認識性能の評価

5.2.1 評価環境

一般のリビングを模した評価施設を用い、食器洗い、アイロンがけ、掃除、歯磨き、ヘアドライヤー、トイレ(水洗区間)の6行動を認識対象とした被験者評価を実施した。被験者は30代から60代の男女22名であり、内訳を表1に示す。

表1 被験者の内訳

	30代	40代	50代	60代	計
男性	0	3	3	5	11
女性	4	3	2	2	11
計	4	6	5	7	22

評価にはAtmel社の組み込み向け評価機であるSAM4L-EK[7]にMEMSマイクを接続した試作機(図5,表2,表3)を使用した。防水での使用を想定し、マイクの防水対策として広く使用されている防水透湿性素材であるゴアテックスを音孔に貼りつけ防水環境での音響を模擬した。

評価にあたって、マイクは実験用に用意した上着の胸ポケットにクリップで固定した(図5)。この状態で全6行動を F_s 16k、 Q_b 16、 C_n 1のPCMデータとして録音した。

被験者には予め録音対象の行動とその実行順について説明した後、各自のやり方で各行動を逐次実行させた。

表 2 試作機のスペック

CPU	ARM Cortex™-M4
CPU クロック	最大 48MHz
メモリ	32kByte
ストレージ	USB メモリ

表 3 マイクのスペック

マイク種別	MEMS マイク
型版	ADMP441(InvenSense)
感度(dBFS)	-26
帯域(Hz)	60~15k
S/N(dB)	61
サンプリングレート(Hz)	8k・16k・24k・48k
量子化ビット数(bit)	24

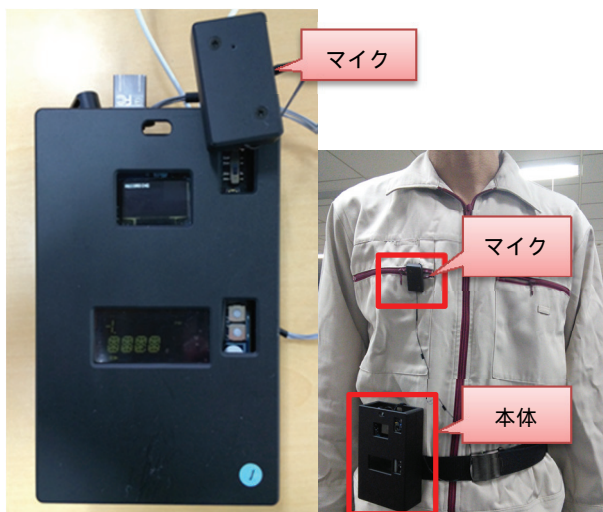


図 5 試作機の外観

加速度センサにより作業区間が適切に分割されていることを仮定し、手作業により各 6 行動とそれ以外の音声区間に分離した。その上で各行動の任意の 10 秒間を学習用として切り出し、それ以外の区間を認識対象とした。トイレ(水洗区間)のみ 1 行動の所要時間が短く 1 回の行動で学習区間と認識区間を確保できないため、学習用、認識用として計 2 回録音を行った。抽出する音特徴量のデータサイズ A_s は 32、次元数 A_n は 15 とした。

この音データに対し、文献[4]で開発した Android™ アプリケーションを使用して、パラメータ変更時の性能評価を行った。

5.2.2 サンプリングレートと量子化ビット数

音声のパラメータのうち、計算量への影響が大きい F_s および Q_b を変更して評価を実行した。

各 F_s の評価データは、音声編集ソフト Adobe Audition CS6[8]を使用し、 F_s 16k の元データを 8k、6k、4k へとダウンサンプリングすることで生成した。

Q_b の変更は、15~9 に関しては評価プログラムのデータ入

力処理において、 Q_b 16 の入力データに対し下記変換を実行することで擬似的に生成した。

$$a_n = \left\lfloor \frac{a_{16}}{2^{16-n}} \right\rfloor \quad (a_n \text{は量子化ビット数 } n \text{ の PCM サンプル})$$

生成した各 F_s 、 Q_b における 22 名の評価データに対する 6 行動の平均認識精度を図 6 に、区間補正を行った場合の 6 行動の平均認識精度を図 7 に示す。

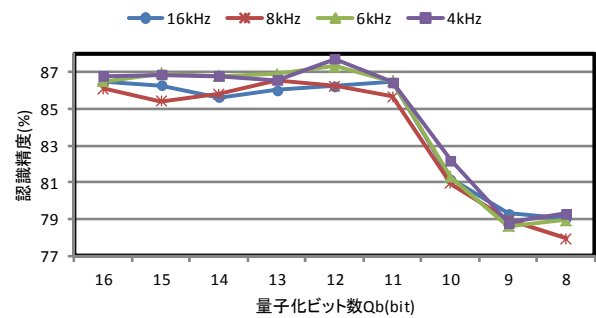


図 6 サンプリングレート・量子化ビット数ごとの 6 行動の平均認識精度

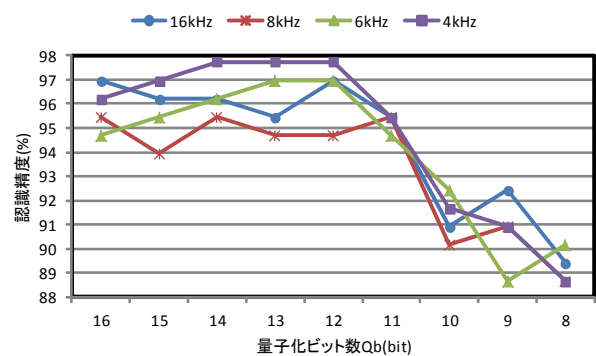


図 7 サンプリングレート・量子化ビット数ごとの 6 行動の平均認識精度 (区間補正後)

F_s に関しては認識性能への影響が軽微であることがわかる。 F_s 4k が計算量を勘案すると最適解という結果となった。今回の 6 行動に関しては、2kHz 以下の周波数成分が識別に有効であるためだと考える。

また Q_b に関しては 12 までは認識精度に対する影響が小さい。 Q_b を 12 としてもプログラム内部では 16bit データとして処理するため、計算量やメモリ使用量を大きく削減することはできないが、ハードウェアの要求スペックの観点で見ると高精度な AD コンバータを必要とせず一般的な汎用 AD コンバータで対応可能である。

5.2.3 ウィンドウ幅

特徴量抽出処理で使用している MFCC では、スペクトルを得るため、高速フーリエ変換(FFT: Fast Fourier Transform)

を実行する。5.3 節で述べるウェアラブル端末向けに最適化した特徴量抽出処理では、高速化のため入力データに応じてメモリ上に動的生成するテーブルを使用しており、このテーブルのサイズがウィンドウ幅に比例する。そこでメモリ使用量を削減するため、ウィンドウ幅を縮小した場合の認識精度への影響を評価した。

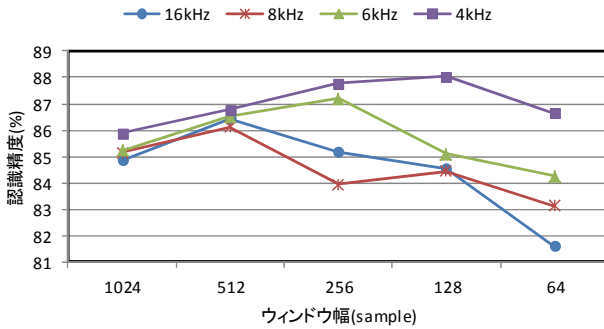


図 8 サンプルレート・ウィンドウ幅ごとの 6 行動の平均認識精度

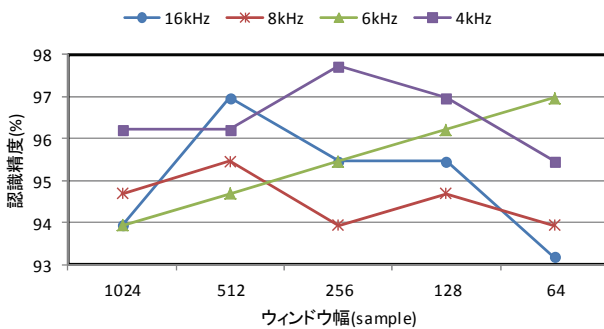


図 9 サンプルレート・ウィンドウ幅ごとの 6 行動の平均認識精度 (区間補正後)

図 8, 図 9 に F_s ・ウィンドウ幅ごとの 6 行動の平均認識精度を示す。区間補正を行う前の認識精度のピーク付近に区間補正後のピークが来ると考えるが、 F_s 6k のみ両ピークが大きく解離する結果となった。これは今回の評価結果がたまたま区間補正により正解となる境界条件下にあるデータが多かったためであり、被験者数を増やせば他のデータと同様補正前の認識精度のピーク付近に補正後のピークがくるものとする。

ウィンドウ幅の最適値は、 F_s が 16k あるいは 8k の場合は 512, 6k の場合は 256, 4k の場合は 256 あるいは 128 となった。

5.2.4 間欠動作

低消費電力化に対して大きな効果が期待できる特徴量抽出処理の間欠動作に関する評価を実施した。

1 秒間特徴量抽出処理を実行し n 秒間動作を停止するという間欠動作について、 I_t n を 0~20 秒まで 1 秒単位で変化させた。 I_t n の認識精度は、 $n+1$ 秒中の任意の 1 秒を特

徴量抽出処理にあてそれぞれについて認識精度を算出し、これらの平均値を用いた。ウィンドウ幅 512, Q_b 16, F_s は 16k・8k・6k・4k のそれぞれについて評価した。評価結果を、図 10, 図 11 に示す。 I_t を 7 に設定しても、 F_s が 16k・4k の場合に区間補正後の 6 行動の平均認識精度は 95% を超えることがわかる。

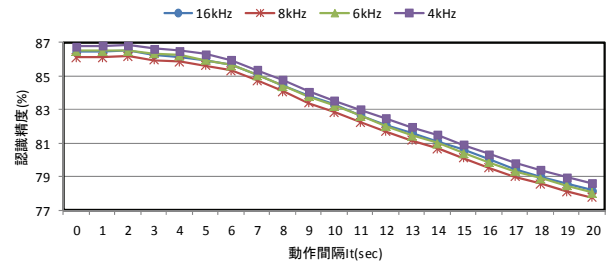


図 10 サンプルレート・動作間隔ごとの 6 行動の平均認識精度

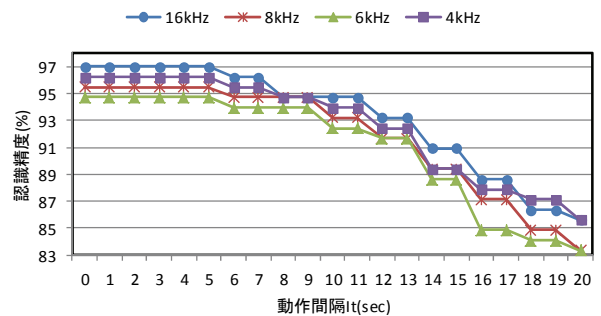


図 11 サンプルレート・動作間隔ごとの 6 行動の平均認識精度 (区間補正後)

表 4 6 行動の平均区間長

行動	平均区間長(sec)
皿洗い	135.1
アイロンがけ	141.7
掃除機がけ	84.1
歯磨き	61.2
ヘアドライヤー	65.0
トイレ(水洗)	9.2
6 行動平均	82.7

今回の 6 行動の被験者 22 人の平均区間長を表 4 に示す。最も所要時間の短いものがトイレ(水洗区間)で 9.2 秒であった。区間長が短いと I_t を広げた際に区間全体が認識対象外となってしまう場合があり、認識精度への影響が大き

いと考える．間欠動作時の区間長と認識精度の影響を確認するため，トイレ（水洗区間）とそれ以外の 5 行動平均の認識精度を評価した．評価結果を図 12 に示す．

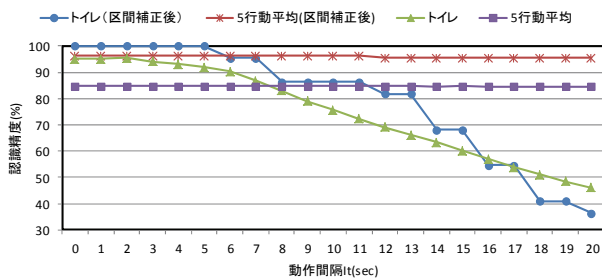


図 12 動作間隔ごとのトイレ・その他 5 行動平均の認識精度

評価パラメータは，ウィンドウ幅 512, Q_b 16, F_s 16k である．トイレ以外の 5 行動に関しては，平均区間長が I_t に対して十分に大きいため，認識精度への影響は非常に軽微であると考えられる．一方トイレ（水洗区間）に関しては，平均区間長に近い I_t 8 の近辺から区間補正後の認識精度が落ち込むものの， I_t 7 までは区間補正後の 6 行動の平均認識精度が 95%以上であることがわかる．

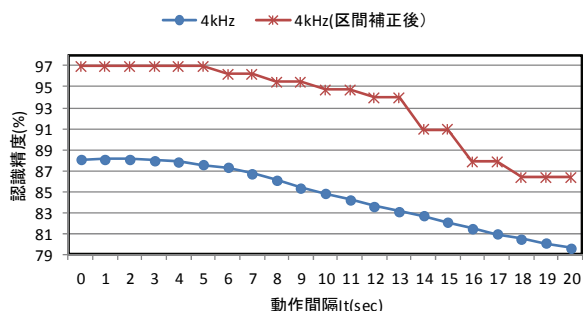


図 13 4kHz での動作間隔ごとの 6 行動の平均認識精度

今回の 6 行動に関して最適パラメータと考える F_s が 4k, ウィンドウ幅 128 で動作間隔を変更した場合の 6 行動の平均認識精度を図 13 に示す． Q_b は 16 とした．本パラメータでは， I_t を 9 に設定しても区間補正後の平均認識精度が 95% 以上であることがわかる．その場合， A_s が 32, A_n が 15 では，音特徴量の平均データ通信量は 48bps であり，低消費電力である BLE を使用して間欠的に送信することが可能となり省電力化が見込まれる．

5.3 ウェアラブル端末向け実装の性能評価

音の特徴量抽出処理 (MFCC,RMS,ZCR) がウェアラブル端末上で動作可能か評価を行った．

スマートフォン向けの特徴量抽出処理を ARM CortexTM-M4 を対象とした固定小数点プログラムに移植し，

各種最適化により計算量，メモリ使用量を削減した．

表 5 特徴量抽出の RAM・ROM 使用量

ウィンドウ幅	RAM 使用量(Byte)	ROM 使用量(Byte)
512	5328	25680
256	2768	22948
128	1488	21544

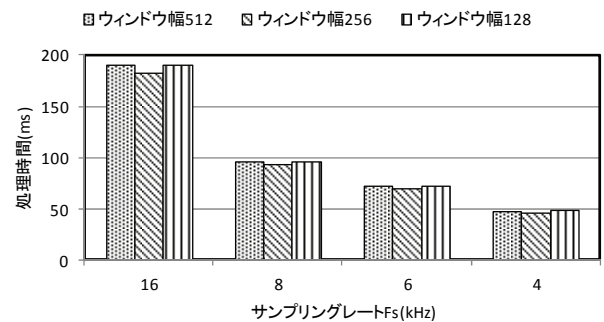


図 14 ウィンドウ幅・サンプリングレートごとの特徴量抽出処理の処理時間

Q_b は 16, F_s が 16k での特徴量抽出処理の RAM および ROM 使用量を表 5 に示す．本実装では RAM および ROM の使用量はウィンドウ幅に依存するが， F_s には依存しない． Q_b を 8 にすることにより RAM および ROM 使用量の削減が可能であるが，認識精度への影響が大きいことが評価の結果わかったため，検討対象外とした．RAM 使用量に関しては最大でも 5kByte ほど，ROM に関しても 25kByte ほどで済み，ウェアラブル端末への搭載が十分に可能である．

図 14 に約 1 秒分の音データに対して特徴量抽出処理を実行した場合の処理時間を，ウィンドウ幅， F_s ごとに示す．評価は試作機(表 2)を使用し，CPU クロック 48MHz で動作させ処理時間を計測した．最大でも 190ms 程度で 1 秒分のデータの特徴量抽出処理が実行できることがわかる．認識精度の高い F_s 4k に関してはウィンドウ幅 128 で 48.96ms で処理できており，CPU クロックを下げても十分処理できることから，ダウンスロークによる消費電力の削減も可能である．

ARM CortexTM-M4 を対象とした特徴量抽出処理の認識精度について評価した結果を図 15, 図 16 に示す．凡例はウィンドウ幅を表す数字と特徴量抽出処理の種類を表す英字から構成され，英字の fixed が ARM CortexTM-M4 向け固定小数点版特徴量抽出処理を使用した場合の，float が従来の AndroidTM 向け浮動小数点版特徴量抽出処理を使用した場合の 6 行動の平均認識精度である．ウィンドウ幅 512・256・128, F_s 16k・8k・6k・4k それぞれについて評価した．ウェアラブル端末向けのプログラム最適化による認識性能

に対する影響は非常に軽微であり、同等の認識精度がでていることがわかる。

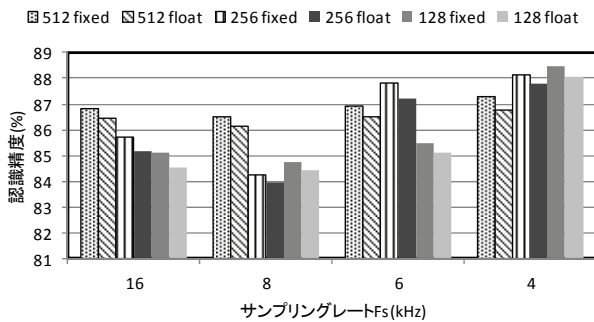


図 15 固定・浮動小数点版特徴量抽出処理の 6 行動の平均認識精度

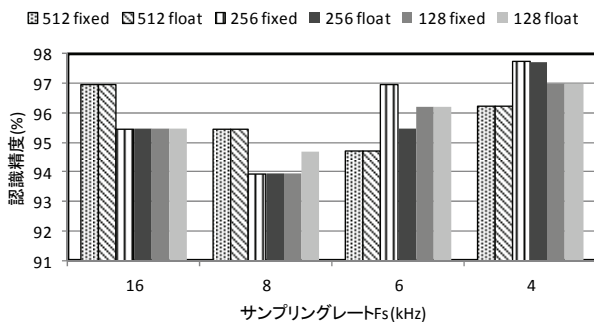


図 16 固定・浮動小数点版特徴量抽出処理の 6 行動の平均認識精度 (区間補正後)

本検討により RAM・ROM 使用量, 計算量ともウェアラブル端末に搭載可能であり, ウェアラブル端末に搭載することによる認識精度の劣化も非常に軽微であることが確認できた。特に今回の評価で最も認識精度のよかった F_s 4k, ウィンドウ幅 128 の特徴量抽出処理に関しては計算量が約 2.35MIPS, RAM 使用量 1488Byte と非常に軽量であり, 様々なウェアラブル端末への搭載が期待できる。

6. まとめ

本論文では, ウェアラブル端末とスマートフォンを使用する生活行動認識の分散処理方式の検討と評価を行った。評価の結果, ウェアラブル端末とスマートフォンの処理境界を特徴量抽出後とする提案方式がデータ通信量, 処理量の観点で最適であることを確認した。またウェアラブル端末の省電力化に有効であるサンプリングレートと量子化ビット数の削減, 間欠動作に関して評価し, サンプリングレート削減と間欠動作の有効性を確認した。

ウェアラブル端末で実行する処理のうち最も計算量・メモリ使用量の大きい特徴量抽出処理が認識精度を維持したまま搭載可能であることがわかり, 提案システムの実現性を確認した。

今回の 6 行動に関しては, 計算量・消費電力と認識精度を考慮した音の特徴量抽出処理の最適パラメータは, サンプリングレート 4kHz, ウィンドウ幅 128, 間欠動作時の動作間隔 9 秒で, 区間補正後の 6 行動平均認識精度が 95%以上, 音特徴量の平均データ通信量は 48bps となることを確認した。また量子化ビット数に関しては, 12bit まで削減しても認識精度への影響が軽微であることを確認した。

最適パラメータで ARM Cortex™-M4 を搭載したウェアラブル端末により特徴量抽出処理を実行した場合, 1 秒分の音データの特徴量抽出処理の計算量は約 2.35MIPS, RAM 使用量 1488Byte と非常に軽量である。消費電力に関しては間欠動作により認識精度を下げることなく大きく削減可能であり, 様々なウェアラブル端末への搭載が期待できる。

今後の課題として, 今回検討した最適パラメータの汎用性について検討する必要がある。例えば, 会話という行動を認識しようと考えた場合, サンプリングレートが 4kHz では十分な認識精度がでない可能性がある。本システムの使用用途としては様々な行動を学習により認識する汎用性も重要であるため, 今回の 6 行動以外に関しても十分に検討を行う必要がある。

今後は本方式での生活行動認識について, 実用化に向けた検討を進めていく。

参考文献

- 1) K. Ouchi and M. Doi. Indoor-Outdoor Activity Recognition by a Smartphone. In Adjunct Proc. UbiComp 2012. pp.600-601(2012).
- 2) Maurer, U., Smailagic, A., Siewiorek, D.P. and Deisher, M.: Activity Recognition and Monitoring Using Multiple Sensors on Different Body Positions, International Workshop on Wearable and Implantable Body Sensor Networks(BSN 2006), pp.113-116, IEEE(2006).
- 3) 倉沢央, 川原圭博, 森川博之, 青山友紀: センサ装着場所を考慮した 3 軸加速度センサを用いた姿勢推定手法, 情報処理学会 研究報告, 2006-UBI-11(3), pp.15-22(2006).
- 4) 大内一成, 土井美和子: スマートフォンによる屋内外生活行動センシング, マルチメディア, 分散, 協調とモバイル(DICOMO2012) シンポジウム, pp.173-179(2012).
- 5) 大内一成, 土井美和子: 携帯電話搭載センサによるリアルタイム生活行動認識システム, 情報処理学会論文誌, Vol.53, No.7, pp.1675-1688(2012).
- 6) Chang, C.-C. and Lin, C.-J.: LIBSVM: A library for support vector machines (2001). Software available from <http://www.csie.ntu.edu.tw/~cjlin/libsvm/>
- 7) Atmel SAM4L-EK <http://www.atmel.com/tools/sam4l-ek.aspx>
- 8) Adobe Audition, <https://creative.adobe.com/ja/products/audition/>