Vol.2019-GN-106 No.30
Vol.2019-CDS-24 No.30
Vol.2019-DCC-21 No.30
2019/1/25

生活音を用いた行動認識のための特徴量抽出についての一考察

荒川正規[†1]　大石優彦[†1]　撫中達司[†1]

概要：近年，社会の高齢化に伴い，高齢者の見守りや介護などを目的とした宅内の人間の行動を認識する研究が盛んに行われている．行動認識は高齢者の見守り以外にも子供の見守りや生活リズムを把握することでの健康管理などにも利用することができるなど重要な技術であると言える．行動認識の取組みとして，複数センサを使用する研究が多いが，本研究では人の行動により発生する生活音を認識手段として用いる．音の識別において，特徴量 MFCC(Mel-Frequency Cepstral Coefficient)が多くの研究で使用されてきたが，本研究では MFCC を生活音の特徴が低周波数に加え高周波数にも表れることを考慮した特徴量抽出を試みた．改良された特徴量を用いて機械学習 SVM(Support Vector Machine)により行動認識を行った結果をもとに，改良した特徴抽出方法を評価した結果につき報告する．

キーワード：行動認識，音の識別，特徴量，MFCC(Mel-Frequency Cepstral Coefficient)

1. はじめに

近年，社会の高齢化に伴い，一人暮らしの高齢者も増加している．家族と離れて暮らしているため，高齢者に病気やケガなど何か問題が起こった場合，家族はその異常に対してすぐに気が付くことができない．このような問題を IoT(Internet of Things)を用いて解決するために高齢者の見守りや介護などを目的とした宅内の人間の行動を認識する多くの研究が行われている[1][2][3]．また，行動認識技術は高齢者の見守り以外にも子供の見守りや生活リズムを把握することにより健康管理など様々な目的で利用することができる．

高齢者の身体能力や日常生活レベルを図るための重要な指標として用いられているものに ADL(Activities of Daily Living)と IADL (instrumental ADL)というものがある[4][5]．ADL とは日常生活における基本的な「食事，更衣，排泄，入浴，起立」などの動作のことを指す．IADL とは ADL よりも高い自立した日常生活をおくるための動作を指し,「料理・掃除・買い物などの家事，交通機関の利用，電話対応などのコミュニケーション，趣味」などの複雑な日常生活動作のことを指す．この２つの指標となる行動を認識することは高齢者の生活を見守るうえで重要と言える．そのため，本研究では ADL，IADL の行動を認識することを最終的な目的とする．

ここで，行動とはいくつかの事象の組み合わせであると考えられる．例えば，入浴を例に考えると，入浴をするために風呂場の扉を開ける，次に風呂場内でシャワーを使用する，あるいは水が流れる，最後に風呂場から出るために風呂場の扉を開ける，この 3 つの事象を時間を考察して組み合わせることで入浴をしたと考えることができる．本研究では行動をいくつかの事象により定義し,行動認識のための事象の発生を認証する．

行動認識の関連研究に赤外線センサや音センサ，人感センサなど複数のセンサを 1 つにまとめた複合センサを宅内に設置するものがある[6]．この研究で認識できる約 40 種類の行動のほとんどに対して音センサが有効であるという結果が示されており，行動認識に音センサを用いることは有用であると考えられる．

音の識別に関する関連研究では，人の行動認識や環境音を元にスマートフォンへ情報を提供するモバイルアプリケーションなどが報告されている[6][7][8]．これらの関連研究では特徴量として MFCC(Mel-Frequency Cepstral Coefficient)を用いている．しかしながら，MFCC は音声認識に適した特徴量である．そのため生活音の識別には必ずしも適していないため，生活音に適した特徴量を抽出するための MFCC を提案し，改良した MFCC を評価するために，家庭内で発生する生活音を用いて事象認証を行い，改良した特徴量抽出方法を評価するとともに，その結果から最終目的である行動認識の可能性について考察する．

2. 関連研究

行動認識の研究はこれまでもセンサやカメラなど様々なものを用いた手法が提案されている．カメラを用いた行動認識手法[9][10]も提案されているが宅内のようなプライベートな場所で用いる場合，プライバシーの問題があり，利用者には抵抗があるといわれている．

大内ら[7]はスマートフォンに内蔵されている加速度センサと音センサを用いて ADL と IADL の行動を対象とした行動認識手法を提案している．加速度センサで人の動きを認識し，音センサで録音を始め，録音したデータから排せつや掃除など 6 つの行動を認識する．音の識別をする手段として，特徴量である MFCC を抽出し，分類器としては機械学習の１つである SVM(Support Vector Machine)を用い

†1 東海大学　情報通信学研究科
Graduate School of Information and Telecommunication Engineering, Tokai University

ⓒ 2019 Information Processing Society of Japan

Vol.2019-GN-106 No.30
Vol.2019-CDS-24 No.30
Vol.2019-DCC-21 No.30
2019/1/25

ており，平均識別率は 90%以上となっている．しかし，ADL の行動のうち認識できる行動が排せつ 1 つのみである．また，スマートフォンを用いるため，人にセンサを持ってもらっている状態であり，スマートフォンを持ち込むことができない場所での行動が認識できない．例えば ADL の行動の 1 つである入浴を認識することができない．

Gierad ら[6]は人間にセンサを装着するのではなく，センサを宅内に設置することで事象の認証を行う．この研究では加速度センサや音センサ，人感センサなどの複数のセンサを 1 つにまとめた複合センサを宅内に設置することで事象を認証する手法を提案しており，扉の開閉や電子レンジの使用など約 40 種類の事象を認証できる．検証実験では機械学習の AutoEncoder を用いており，キッチンでの事象の平均識別率は約 88.5%である．この研究でのほとんどの事象の認証に音センサが有効であったため，音センサを用いることは有用であると考えられる．しかし，この研究はアプリケーションは研究対象外としており，ADL，IADL については言及されていない．

また，佐々木ら[11]は宅内にセンサを設置する行動認識の手法として ECHONET Lite 対応家電を用いた手法を提案している．この手法は ECHONET Lite 対応家電の電源 ON/OFF 情報から機械学習を用いて掃除や料理，テレビの視聴など 9 つの行動を認識する．検証実験では機械学習のランダムフォレスト，ロジスティック回帰，LSTM の 3 つを用いており，それぞれの平均認識率は 85.5%，84.9%，81.5%となっている．しかし，この研究も ADL，IADL については言及しておらず，IADL の行動の 1 つである掃除の認識率はそれぞれの機械学習で 30%未満であった．また，ADL の行動の 1 つの排せつや歯磨きなど家電を利用することのない行動は認識が難しい．

そこで本研究ではカメラと比べプライバシーを守れ，宅内に設置することで多くの行動の認識に有用であると考えられる音センサを用いることで歯を磨いている音やシャワーの使用音，掃除機の使用音などの生活音を識別し事象を認証することで行動認識も試みる．

音の識別に関する関連研究では，人の行動認識や環境音を元にスマートフォンへ情報を提供するモバイルアプリケーションなどが報告されている[6][7][8]．これらの関連研究では特徴量として MFCC を用いている．しかしながら，MFCC は音声認識に適した特徴量であり生活音には必ずしも適していない．また，生活音は周波数の低い部分以外でも周波数成分の特徴が表れるため，本研究では生活音に適した特徴量を抽出するため，この特性を考慮した MFCC の改良方法を提示し，その評価結果につき報告する．

3. 生活音を用いた分析

人間の聴覚性質を基とした特徴量 MFCC は，音声認識で数多く使用されている．以下に MFCC の抽出手順を図 1 に示す．

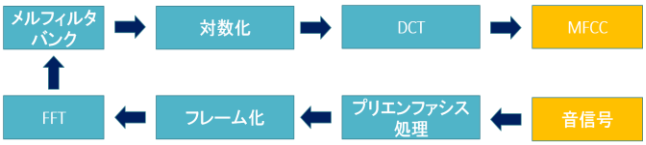


図 1 MFCC の抽出手順

音信号は，周波数が高くなるほど，振幅が小さくなり周波数成分が抽出しにくくなる．周波数成分が高い部分を抽出しやすくするためにプリエンファシス処理を行い，周波数成分を強調する．以下にプリエンファシス処理の式(1)を示す．

$$y(n) = s(n) - ps(n-1) \quad (1)$$

p の値は，0.97 が音声認識で頻繁に用いられる．s(n)のサンプル数 n を変数として，1 サンプル前の値と現在のサンプルの値の差分を取ることで全体の周波数を強調する．プリエンファシス処理を行った音信号をデジタル化する．

デジタル化された音信号から一定の長さを取り出すフレーム化を行う．本章の説明で使用しているフレームの長さは 0.8[s]である．以下にフレーム化されたデジタル信号を図 2 に示す．ここで図 2 の縦軸は音の強さを表す振幅で，横軸は時間を表している．

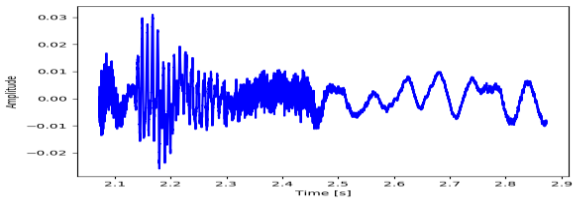


図 2 フレーム化されたデジタル信号

次に，フレーム化したデジタル信号に FFT（高速フーリエ変換）を行う．FFT を行うことでデジタル信号の横軸を，時間から音の高さを表す周波数に変換することができる．FFT により以下に示す図 3 の振幅スペクトルが求められる．図 3 の振幅スペクトルでは，縦軸を振幅，横軸で周波数を表している．

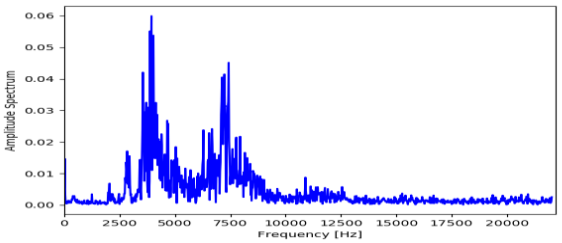


図 3 振幅スペクトル

振幅スペクトルをさらに，メルフィルタバンクを行い，より人間の聴覚情報に近い特徴を抽出する．図 4 は人間の聴覚性質を基に，低い周波数部分は繊細に，高い周波数部分は粗くフィルタを設定したメル尺度を用いたメルフィルタバンクである．またメルフィルタバンクのフィルタ数は 20 個である．

ⓒ 2019 Information Processing Society of Japan

Vol.2019-GN-106 No.30
Vol.2019-CDS-24 No.30
Vol.2019-DCC-21 No.30
2019/1/25

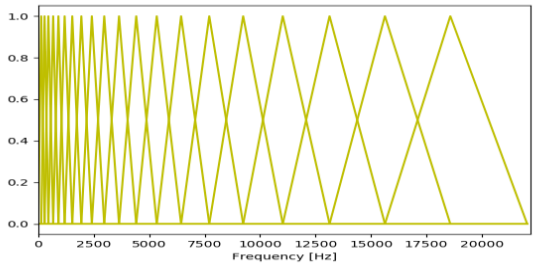


図 4 メルフィルタバンク

メルフィルタバンクを行うことで，人間の取得する聴覚情報に近いメル帯域スペクトルが得られる（図 5）．

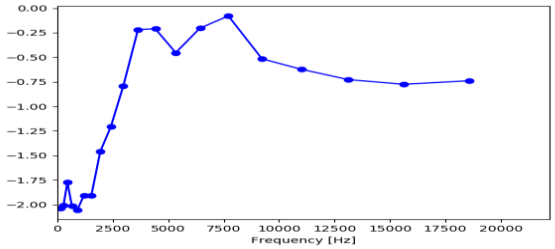


図 5 メル帯域スペクトル

メル帯域スペクトルは，メルフィルタバンクにより離散化されているので，離散コサイン変換(2)よりさらに周波数成分を求める．

$$Ci = \sqrt{\frac{2}{L}} \sum_{l=1}^{L} \log m(l) \cdot \cos\left\{\left(l - \frac{1}{2}\right)\frac{i\pi}{L}\right\} \qquad (2)$$

離散コサイン変換より求められた情報がケプストラムである（図 6 の左グラフ）．

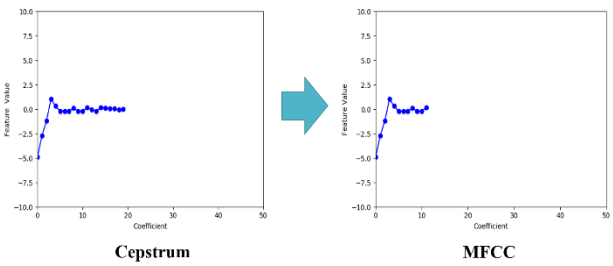


図 6 ケプストラムから MFCC の抽出

ケプストラムから音声の特徴が表れる低次元の 12 次元を抽出した情報が MFCC である（図 6 の右グラフ）．次元とはケプストラムの一つ一つの値のことを意味する．図 6 におけるケプストラムでの横軸である係数の小さいものを低次元と呼ぶ．図 6 の左図では，20 次元のケプストラムが示されており，そこから低次元となる 12 次元を抽出したものが右図の MFCC となる．しかし，音声以外の生活音には，周波数成分の特徴が高次元にも表れるため，より多くの次元を抽出する必要がある．

4. 高次元の周波数成分抽出を可能とする改良型 MFCC の提案

4. 1. 生活音を対象とした MFCC の課題

MFCC は人間の聴覚性質を基としており，音声認識で用いられる代表的な特徴量である．音声の特徴となる周波数成分は低い周波数部分に表れるため，その部分を抽出している．しかし，本研究の識別対象である生活音は，低い周波数部分だけではなく，高い周波数部分にも周波数成分の特徴が表れる． 以下の図 7 より音声と生活音の周波数成分の違いを，縦軸を音の強さを表し，横軸を周波数の高さを表す振幅スペクトルより示す．

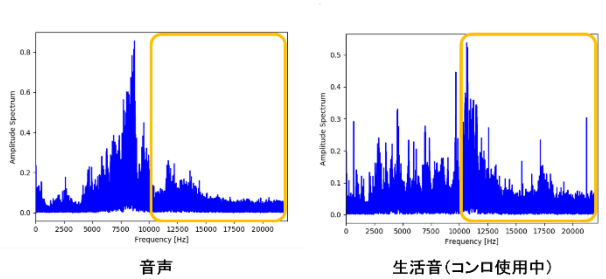


図 7 音声と生活音の振幅スペクトル

図 7 のように，音声は周波数が低い部分に大きく周波数成分の特徴があり，周波数が高い部分は特徴があまりない．一方，音声以外の生活音は周波数が低い部分と高い部分にも周波数成分の特徴が大きく表れることがわかる．したがって，生活音の場合では，特徴量の次元数を増やし，周波数の高い部分の周波数成分も抽出すれば識別精度が高まることが期待できる．

4. 2. MFCC の改良方法

音声以外の生活音では周波数成分の特徴が，周波数の低い部分から高い部分まで全体的に表れる．MFCC が抽出する周波数成分の範囲を周波数の低い部分だけではなく高い部分も加え，そこから重要となる特徴を抽出することにより識別精度を高めることができると仮説を立てた．特徴量の改良方法を以下に示す．

・第 1 段階

（目的）

- 音信号の抽出対象を拡大
- 再生可能資源の落下音の周波数成分を強調
- 高次元の周波数成分も抽出

（手段）

- フレーム化する範囲を調節する
- FFT のサンプル数を変更する
- プリエンファシス係数を変更する
- メルフィルタバンクのフィルタ数を増やす

・第 2 段階

（目的）

- 重要ではない次元を削減

（手段）

- 主成分分析 PCA により次元数を最適化する

ⓒ 2019 Information Processing Society of Japan

Vol.2019-GN-106 No.30
Vol.2019-CDS-24 No.30
Vol.2019-DCC-21 No.30
2019/1/25

改良方法は 2 段階によって構成される. 1 段階目にて, 元の音信号に存在する周波数成分の特徴が表れるよう強調し, 周波数が高い部分まで特徴を抽出するよう次元数を増やす. フレーム化する範囲を調節した結果, フレームの長さを 3.5[s]とした. FFT のサンプル数は 50000 とし, プリエンファシス係数は p=0.99 に設定した. また, 抽出する次元数は元が 20 個であったが, 45 個まで増やした. 2 段階目より, 分類器の SVM では特徴量の次元数が多すぎると識別率が下がる為, 主成分分析 PCA(Principal Component Analysis)を用いて重要な特徴だけを抽出している. PCA で抽出する次元数は 33 個に設定した.

4. 3. 改良型 MFCC の検証

識別の対象である水道を使用しているときの音信号を用いて改良型 MFCC が目的とするものとなっているかどうかを MFCC との比較により検証する. はじめに, 振幅スペクトルの抽出計算で扱う FFT のサンプル数とプリエンファシス係数を増やしたことで, 周波数成分が更に強調されることが確認できる(図 8).

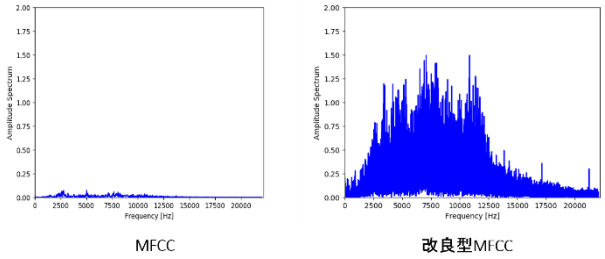


図 8 MFCC と改良型 MFCC の振幅スペクトル

次に, 人間の聴覚をモデルとしたメルフィルタバンクの比較を図 9 に示す. 図 9 で示すように, 周波数の高い部分にも繊細にフィルタが設定されていることが分かる.

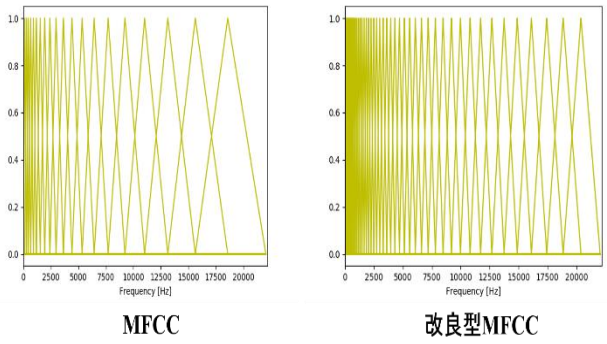


図 9 メルフィルタバンクの比較

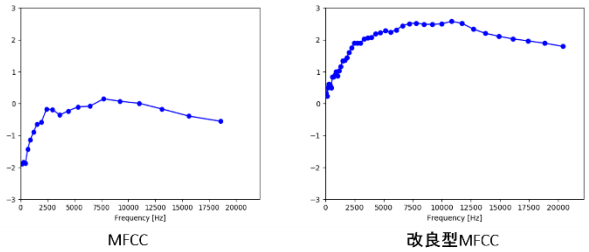


図 10 メル帯域スペクトルの比較

図 10 は振幅スペクトルにメルフィルタバンクを掛けて抽出されたメル帯域スペクトルである. メル帯域スペクトルはメルフィルタバンクにより離散化されているため変換をする必要がある. そこで改良型 MFCC では, 通常の MFCC で用いられる離散コサイン変換ではなく, 離散サイン変換より周波数成分を抽出する. 図 11 に, メル帯域スペクトルの抽出までは同じ処理を行った音信号より, 離散コサイン変換と離散サイン変換による特徴抽出の違いを示す. 離散サイン変換を用いることでより大きな特徴を抽出していることが確認できる.

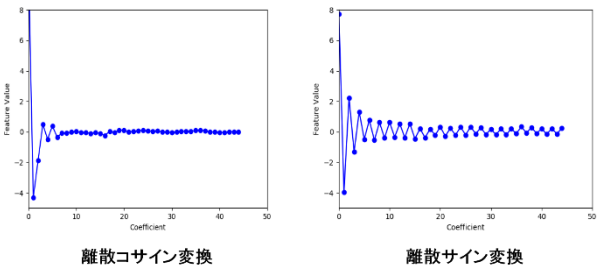


図 11 離散変換の違い

離散サイン変換を行ったケプストラムと通常のケプストラムの比較を図 12 に示す. 音信号から抽出した次元数や特徴の大きさが違うことが分かる.

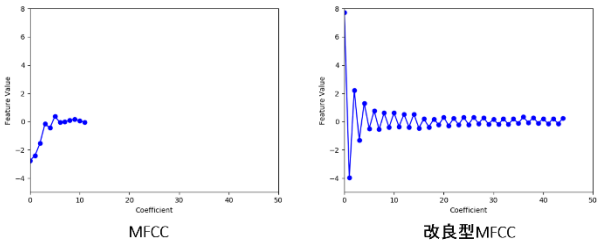


図 12 従来の MFCC と改良型 MFCC の比較

なお, 改良型 MFCC では, 周波数の低い部分から高い部分まで周波数成分を抽出したが, その分抽出された次元数が多く識別時に余分となる特徴量が表れる. 余分な特徴量を削減し, より識別率を上げるため PCA を用いる. PCA は, 多次元の特徴量から重要な部分を抽出する主成分分析である.

ⓒ 2019 Information Processing Society of Japan

Vol.2019-GN-106 No.30
Vol.2019-CDS-24 No.30
Vol.2019-DCC-21 No.30
2019/1/25

5. 事象認識の評価結果と考察

5.1. 検証の環境設定

図 13 は，改良型 MFCC の評価を行ったソフトウェア構成を示したものである．

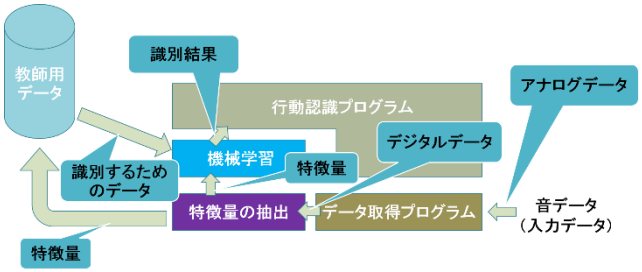


図 13 ソフトウェア構成

評価では，生活音を取得する為の音センサと分析を行う PC を使用した．入力データを取得する音センサとして，サンワサプライ株式会社の MM-MCU01BK を使用した．入力データを A/D 変換よりデジタルデータに変換する．その際のサンプリングレートは 44.1[kHz] に設定した．そのデジタルデータより特徴量を抽出する．特徴量とは教師ありの機械学習で識別するデータの特徴を適量的に表現する値である．抽出された特徴量から，システムに格納された教師用データを用いて機械学習アルゴリズムによりどの生活音であるかを識別する．

5.2. 評価対象とする生活音の条件

本研究では一人住まいの高齢者を対象として，食事，入浴，整容などの行動を指す ADL と，料理・掃除など自立性の高い日常生活をおくるための行動を指す IADL を認識するための事象を音データから抽出することとする．

（1）行動認識を行う対象となる場所

図 14 は想定した部屋の間取り図と測定対象を示している．ADL，IADL である入浴，整容，料理，掃除などから発生する音を識別するためにキッチン，リビング，洗面所の 3 ヵ所を対象として選択した．

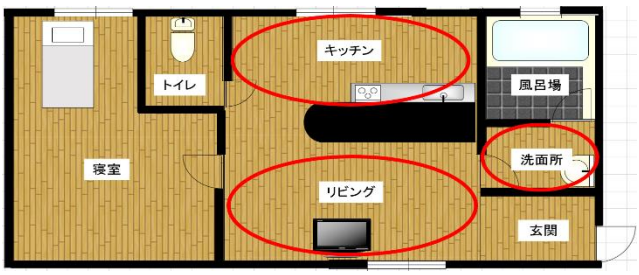


図 14 想定した部屋の間取りと測定対象

（2）対象とする家屋の種類

本評価では，環境による生活音の違いを考慮するために 2 軒の世帯を対象とした．A 宅を一軒家とし，B 宅はマンションとした．表 1 ，表 2 に世帯ごとで収集した音データの一覧を示す．

表 1 A 宅で収集した音一覧

録音した場所	識別する音のラベル	詳細	全データ数
キッチン	Cookinng Stove	コンロを使用	284
	Range	レンジを使用	160
	Water	水道を使用	148
	Exhaust Fan	換気扇を使用	110
	Vacuum Cleaner	掃除機を使用	120
	Cookinng Stove + Water	コンロと水道を同時に使用	60
	Cookinng Stove + Range	コンロとレンジを同時に使用	60
	Cooking Stove + Exhaust Fan	コンロと換気扇を同時に使用	60
	Not Sound	何も使用していない	100
リビング	TV	テレビを使用	200
	Vacuum Cleaner	掃除機を使用	100
	Not Sound	何も使用していない	140
洗面所	Washing Machine	洗濯機を使用	366
	Water	水道を使用	120
	Brushing teeth	歯を磨いている	198
	Shaving	電気シェーバーを使用	160
	Shower	シャワーを使用	280
	Hairdryer	ドライヤーを使用	110
	Exhaust Fan	換気扇を使用	100
	Bathroom Door	風呂場のドアを開く	40
	Vacuum Cleaner	掃除機を使用	60
	Not Sound	何も使用していない	100

表 2 B 宅で収集した音一覧

録音した場所	識別する音のラベル	詳細	全データ数
キッチン	Cookinng Stove	コンロを使用	64
	Range	レンジを使用	60
	Water	水道を使用	40
	Exhaust Fan	換気扇を使用	40
	Not Sound	何も使用していない	40
リビング	TV	テレビを使用	100
	Vacuum Cleaner	掃除機を使用	100
	Not Sound	何も使用していな	100
洗面所	Washing Machine	洗濯機を使用	100
	Water	水道を使用	40
	Brushing teeth	歯を磨いている	40
	Shower	シャワーを使用	60
	Hairdryer	ドライヤーを使用	30
	Not Sound	何も使用していない	40

表 1 の A 宅では識別する音のラベルを 22 種類とした．また，表 2 の B 宅では識別する音のラベルは 14 種類である．収集した音データは 2 世帯合計で 3930 個である．今回は，収集したデータをラベルごとに全データを半分ずつ分け教師用データと検証用データに格納し，評価を行う．

5.3. 改良型 MFCC の評価方法

5.2 節で設定した条件に基づき，以下の観点から改良型 MFCC の評価を行う．

(1) 改良型 MFCC の優位性評価
(2) 評価に用いる機械学習アルゴリズムの選択
(3) 環境による生活音の違いについての評価
(4) 教師用データについての評価
(5) 通常の生活環境を想定した混合する生活音での評価

まず(1)では従来の MFCC と比較し，改良型 MFCC の優位性について，識別精度を検証する．(2)では(3)以降の評価に用いる機械学習アルゴリズムを選択するために，改良型 MFCC を用いた機械学習アルゴリズムを検証する．検証で用いた機械学習アルゴリズムは，SVM，Random Forest，KNN(K Nearest Neighbor), ロジスティック回帰の 4 種類である．(3)では，異なる世帯で録音した音データから，世帯の違いによる識別精度を検証する．(4)では教師用データの作成方法につき検証する．検証方法は，次の三段階とした．はじ

ⓒ 2019 Information Processing Society of Japan

Vol.2019-GN-106 No.30
Vol.2019-CDS-24 No.30
Vol.2019-DCC-21 No.30
2019/1/25

めに A 宅で録音された音データを教師用データとし，B 宅で録音された音データを検証用データとして検証を行う．次に，逆の B 宅で録音された音データを教師用データとし，A 宅で録音された音データを検証用データとして検証を行う．最後に A 宅と B 宅の両世帯で録音した音データを教師用データと検証用データに分けて検証する．最後に(5)では，実際の生活環境における評価を目的として，混合した生活音の検証を行う．(1)から(4)までは，発生する音が単独で発生することを想定として検証を行ったが，人が生活する中では，純粋な音が重なり合った混合音が多く発生する．そのため，混合音での検証を行った．

5.4. 評価結果

(1) 改良型 MFCC の優位性評価

A 宅のキッチン，リビング，洗面所の 3 ヵ所で取得した音データで検証を行った．図 15，図 17 左グラフ，図 18 は改良型 MFCC の結果，図 16，図 17 右グラフ，図 19 は従来の MFCC の結果を混同行列で示したものである．表3，表4は，これらの結果を表で整理したものである．結果として，キッチンでの識別精度は改良型 MFCC と従来の MFCC では差がないが，リビングと洗面所では約 10%の差がある．3 ヵ所を平均した識別精度も約 10%の有意な差がある．

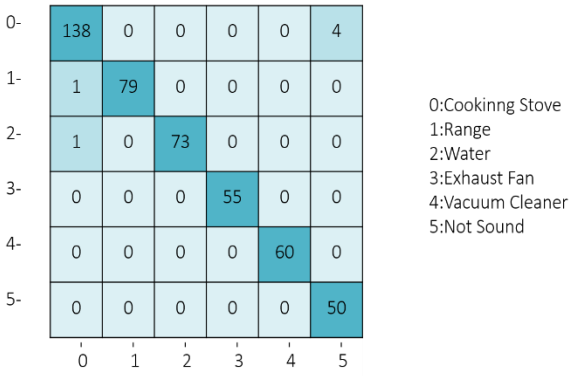


図 15　A 宅のキッチンでの改良型 MFCC による混同行列

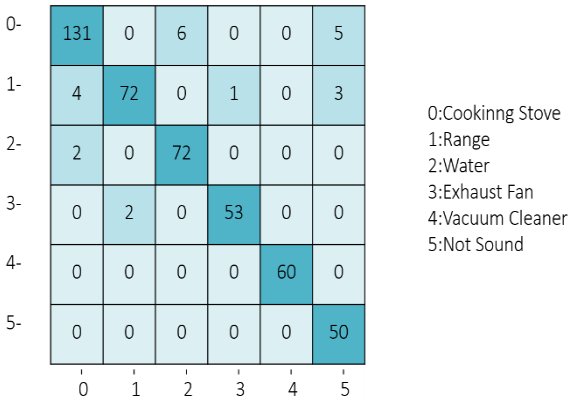


図 16　A 宅のキッチンでの従来の MFCC による混同行列

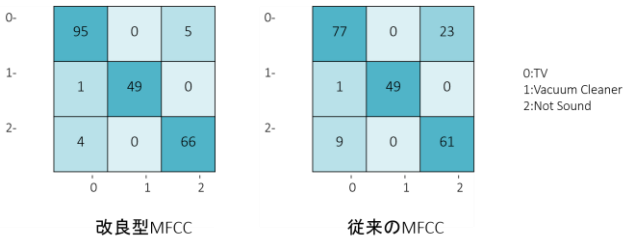


図 17　A 宅のリビングでの改良型 MFCC と従来の MFCC による混同行列

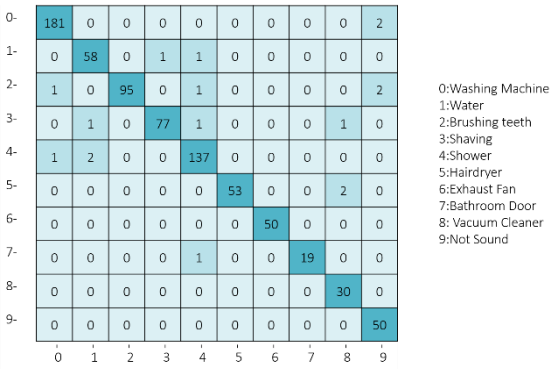


図 18　A 宅の洗面所での改良型 MFCC による混同行列

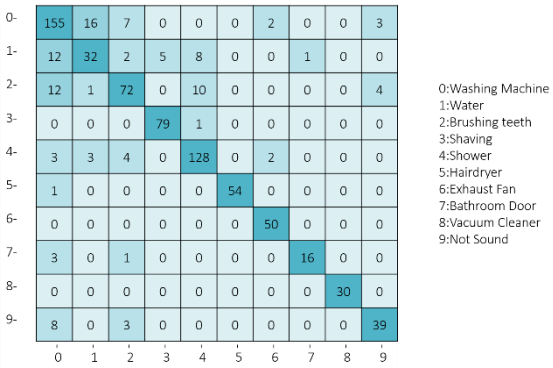


図 19　A 宅の洗面所での従来の MFCC による混同行列

表 3　A 宅での改良型 MFCC による識別精度の結果

録音場所	キッチン	リビング	洗面所	3ヵ所平均
識別精度	98.69%	95.45%	97.78%	97.30%

表 4　A 宅での従来の MFCC による識別精度の結果

録音場所	キッチン	リビング	洗面所	3ヵ所平均
識別精度	95.01%	85.00%	85.39%	88.46%

ⓒ 2019 Information Processing Society of Japan

Vol.2019-GN-106 No.30
Vol.2019-CDS-24 No.30
Vol.2019-DCC-21 No.30
2019/1/25

表 5　機械学習アルゴリズムごとの識別精度の結果

機械学習	SVM	Random Forest	KNN	ロジスティック回帰
平均識別精度	97.78%	96.34%	95.95%	97.13%

表 6　B 宅での改良型 MFCC による識別精度の結果

録音場所	キッチン	リビング	洗面所	3ヵ所平均
識別精度	99.18%	94.66%	92.22%	95.35%

表 7　A 宅=教師用データ，B 宅=検証用データの結果

録音場所	キッチン	リビング	洗面所	3ヵ所平均
識別精度	33.60%	65.33%	61.29%	53.40%

表 8　B 宅=教師用データ，A 宅=検証用データの結果

録音場所	キッチン	リビング	洗面所	3ヵ所平均
識別精度	33.66%	83.18%	63.88%	60.24%

表 9　A＋B=教師用データ，A，B=検証用データの結果

録音場所	キッチン	リビング	洗面所	3ヵ所平均
識別精度	98.66%	94.05%	97.43%	96.71%

（2）評価に用いる機械学習アルゴリズムの選択

次に本評価に用いる機械学習アルゴリズムを選択するための検証を行う．表 5 は，洗面所で取得した音データより検証を行った結果を示している．検証で用いたアルゴリズムのパラメータは，それぞれグリッドサーチより選定した．選定したパラメータは，SVM では C パラメータが 100，gamma パラメータを 0.03 である．Random Forest では，決定木の数を 1000，決定木の深さを 10，乱数を制御するパラメータを 10 とした．KNN では，K の数を 2 とした．ロジスティックス回帰では，C パラメータを 20 とした．本評価結果では，識別精度が最も高く，パラメータの調節が容易である SVM をこれ以降採用する．

（3）環境による生活音の違いについての評価

次に，録音場所が異なる場合での，改良型 MFCC の識別精度を検証する．表 6 は B 宅のキッチン，リビング，洗面所の 3 ヵ所の音データから検証した結果である．表 3 の A 宅の結果，表 6 の B 宅の結果より，二箇所の環境では精度に大きな差がなく．改良型 MFCC は異なる環境においても適用可能であるといえる．

（4）教師用データについての評価

教師用データの作成に手間がかかることが課題の一つであるため，教師用データについて検証を行う．表 7，表 8 に A 宅と B 宅で教師用データをそれぞれと交換した場合の検証結果を示す．結果として，異なった世帯で作成した教師用データを使用した場合，識別精度が低いことから，教師用データはそれぞれの環境で作成する必要がある．また，表 9 では A 宅と B 宅の両方の世帯のデータを教師用データとして検証を行った結果を示しており，教師用データを作成しても認識率に大きな差はないことから，共通の教師用データの作成にかかる労力を考慮すると，本研究で提案する音の識別アルゴリズムは，使用する世帯ごとに教師用データを用いる必要がある．

(5) 通常の生活環境を想定した混合する生活音での評価

図 20 は A 宅のキッチンで純粋な音と混合音を加えて検証した結果を示したものである．



図 20　A 宅キッチンでの混合音を含んだ混同行列

この条件での識別精度は 93.28%であったが，コンロと水道を同時に使用している時の混合音は識別精度 36.67%といちじるしく低下する．このことから，実際の生活環境に近い条件での識別精度はまだ十分ではなく，識別精度をあげるには，ノイズキャンセルなどの方法により単一音にする方法の検討が必要であると考える．一方，ADL，IADL の行動を認識することを目的とした場合，コンロと水の同時使用は“食事”を作っている行動を認識する事象だと考えると，これらを区別する必要はなく，混合音を教師用データとする方法もある．しかし，テレビを見ながら，食事を作るようなシーンを想定した場合，この状況での生活音から食事を作っていることが判断できるかが重要であり，前提としている一人暮らしの高齢者の生活において，重複して発生する行動と，それに関する事象がどのようなものなのか，その際に発生する生活音が正しく識別できるのかの検討が必要である．

6. まとめ

本研究では最終目標である ADL，IADL の行動の認識のために生活音に適した特徴量を抽出するために MFCC の改良を提案し，改良型 MFCC を評価する為に，家庭内で発生する生活音を用いて検証を行った．

単一の生活音は，高い精度で識別できることが確認できた．しかし，複数の生活音が組み合わさった混合音の識別精度は十分な精度には至らなかった．混合音の識別精度を高めることができれば，ADL，IADL の行動認識に適用できると考える．今後は，生活音を用いて精度の高い行動認識をするために，純粋な音だけでなく，混合音の識別精度

ⓒ 2019 Information Processing Society of Japan

Vol.2019-GN-106 No.30
Vol.2019-CDS-24 No.30
Vol.2019-DCC-21 No.30
2019/1/25

を高める特徴量の抽出方法を考え,その検証を行っていく.

また，複数の事象の組み合わせによって行動を認識するという仮説の検証をさらに進めるとともに，ADL，IADL にて重要とされている行動の事象の組み合わせを定義し，検証していく予定である.

参考文献

[1] 石井 瞭， 君野 敬祐， 井上 雅裕， “認知症早期発見のための行動センシングセンサシステムの研究”， 情報処理学会第 78 回全国大会， 2016
[2] 中原 啓太， 山口 弘純， 東野 輝夫， “移動型センサーを想定した家庭内行動検出手法”， DICOMO2016， 2016
[3] 玉水 一柔， 榊原 誠司， 佐伯 幸郎，中村 匡秀， 安田 清， “変化時の行動記録と環境センシングに基づく機械学習を用いた宅内行動認識”， 電子情報通信学会技術研究報告 Vol.117 ，no.250， 2017
[4] 総論参考資料｜厚生労働省，https://www.mhlw.go.jp/www1/topics/kenko21_11/s1.html
[5] 自立生活の指標：日常生活動作（ADL）とは｜健康長寿ネット，https://www.tyojyu.or.jp/net/kenkou-tyoju/kenkou-undou/jiritu.html
[6] Gierad Laput，Yang Zhang，Chris Harrison，“Synthetic Sensors: Towards General-Purpose Sensing”，2017
[7] 大内 一成， 土井 美和子， “スマートフォンを用いた生活行動認識技術”， 一般論文， 東芝レビュー vol．68 No．6 (2013)
[8] 津田 貴彦， 中西 恭介， 松山 みのり， 西村 竜一， 山田 順之介， 河原 英紀， 入野 俊夫， “モバイル携帯端末を用いた環境音収集とその認識手法の検討”， 情報処理学会研究報告， Vol．2013-MUS-99 No．18 2013/5/11
[9] Jesse Hoey，James J Little．“Value-directed human behavior analysis from video using partially observable markov decision processes”，IEEE transactions on pattern analysis and machine intelligence，Vol．29，No．7，pp.1118–1132，2007
[10] Loren Fiore，Duc Fehr，Robot Bodor，Andrew Drenner，Guruprasad Somasundaram，Nikolaos Papanikolopoulos，“Multi-camera human activity monitoring． Journal of Intelligent and Robotic Systems”，Vol.52，No.1，pp.5–43，2008
[11] 佐々木 渉， 藤原 聖司， 諏訪 博彦， 藤本 まなと，荒川 豊， 木村 亜紀， 三木 智子， 安本 慶一 ，“ECHONET Lite 対応家電と人感センサの時系列データ分析に よる宅内行動認識”，SIG-SAI-031 2018/03/01

ⓒ 2019 Information Processing Society of Japan