

音を用いた再生可能資源の自動分別に関する研究

荒川 正規^{*1}, 撫中 達司^{*2}

Study on automatic sorting of renewable resources using sound

by

Masaki ARAKAWA^{*1} and Tatsuji MUNAKA^{*2}

(received on Sep. 28, 2018 and accepted on Jan. 10, 2019)

あらまし

循環型社会を実現するための取り組みとして、東海大学高輪キャンパスではスマートゴミ箱を用いて再生可能資源が正しく分別されているか実態調査を行っている。本調査の結果、全体の30%が正しく分別されておらず、それらを回収員が手作業で分別をやり直している。この無駄な労働を無くす為、音を用いた自動分別を解決案とした。しかしながら、音を用いた自動分別を作製するにあたって、多くの研究で使用されてきた特徴量MFCC(Mel-Frequency Cepstral Coefficient)では目標の識別率に至らなかった。本論文では、再生可能資源の識別に適するMFCCの改良方法を提案し、その検証と考察を述べる。

Abstract

As an effort to realize a recycling-oriented society, Tokai University Takanawa Campus uses smart garbage boxes to demonstrate whether renewable resources are properly separated. As a result of this survey, it found that 30% of the garbage was sorted by mistake and the collector manually sorted them separately. In order to eliminate this unnecessary work, we propose automatic sorting using sound recognition. MFCC (Mel-Frequency Cepstral Coefficient) has been used as an algorithm in many studies for sound recognition. In this paper, we propose an improved method of MFCC suitable for identifying renewable resources and describe its verification and consideration.

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

Keywords: Identification of sound, a Feature value, MFCC(Mel-Frequency Cepstral Coefficient)

1. はじめに

循環型社会を実現するためにIoT(Internet of Things)を用いて取り組むべきであるとの考えのもと、東海大学高輪キャンパスではスマートゴミ箱 BigBelly™を用いて、再生可能資源であるビン、缶、ペットボトルを対象にどれだけ正しく分別されているか実態調査を行っている。本調査を2016年1月から2017年4月までの約1年半行った結果、70%は正しく分別をされていたが、30%は誤った分別をされていた¹⁾。この結果から、誤って分別をされた30%分の再生可能資源を回収作業員が一つ一つ手作業で分別をやり直さなければならない現状がわかった。無駄な労働を削減するために、自動で分別をする新たなスマートゴミ箱を作製すべきであるという考えから研究を開始した。

再生可能資源の自動分別の関連研究では、画像認識を用いた自動分別が報告されている²⁾。しかしながら、画像認識を用いた自動分別では確かに再生可能資源であるビン、缶、ペットボトルの分別は可能であるが、缶のアルミとスチールの分別をすることは難しい。そこで本研究では再生可能資源であるビン、アルミ缶、スチール缶、ペットボト

ルの自動分別を可能にする手段として、音(落下音)を用いる。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³⁾⁴⁾⁵⁾。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り再生可能資源の落下音には必ずしも適していない。そこで本研究では再生可能資源の落下音に適した特徴量を抽出するため、MFCCの改良を提案し識別率の向上を目的とする。

以下、第2章で、音を用いた再生可能資源の自動分別に関連する研究について述べる。第3章では、音を用いた分析としてMFCCについて述べる。第4章で、再生可能資源の落下音に対して識別率を高める為のMFCC改良方法について述べる。第5章にて、改良型MFCCの検証とその結果について述べる。第6章でまとめを述べる。

2. 関連研究

音の識別に関連する研究はこれまでも多数あり、人の生活行動認識や周りの状況から適切な情報をユーザーへ提供するために用いられている³⁾⁴⁾⁵⁾。参考文献3)では高齢者の見守りや健康状態などを知る為に加速度センサと音センサを使用している。加速度センサで人の動きを認識した上で、音センサは録音を開始する。録音したデータより、皿洗いや掃除機がけ、ドライヤなどの人の行動を識別している。音の識別をするための手段として、特徴量で

*1 情報通信学研究科情報通信学専攻 修士課程
Graduate School of Information and
Telecommunication Engineering, Course of Information
and Telecommunication Engineering, Master's program

*2 情報通信学研究科情報通信学専攻 教授
Graduate School of Information and
Telecommunication Engineering, Course of Information
and Telecommunication Engineering, Professor

ある MFCC を抽出し、分類器では機械学習の SVM(Support Vector Machine)を用いており、検証結果では識別率が平均 90% 以上となっている。

再生可能資源の自動分別については、画像センサや光、金属、磁気センサを組み合わせる研究が報告されている²⁾⁶⁾⁷⁾。参考文献 2) では、再生可能資源であるビン、缶、ペットボトルの三種類を対象に画像の識別を用いて 98.3% という識別率となっている。参考文献 7) では再生可能資源であるアルミ缶、スチール缶、ペットボトルの三種類を対象に光、金属、磁気センサを組み合わせる研究を行っているが具体的な識別率については言及されていない。

また、ヒューマノイドロボットが再生可能資源であるアルミ缶、スチール缶、ビン、ペットボトルの四種類を対象に画像センサと音センサを用いて分別を行う研究が報告されている⁸⁾。画像の識別より、ビン、缶、ペットボトルを分別し、音の識別より画像で識別できないアルミ缶、スチール缶を分別している。音の識別ではケプストラム特徴量を抽出し、分類器で HMM(Hidden Markov Model)を用いている。具体的には、音データに対してフレームの長さ 30[ms] のケプストラム特徴量を抽出し、10[ms] ずつずらし新たな特徴量を抽出している。このように繰り返し抽出された特徴量には、とりの特徴量と 20[ms] 重なるため時系列な依存関係がある。これら時系列の依存関係を持つ特徴量を用いて HMM より分類を行う。識別率はアルミ缶 93.3%、スチール缶 96.7%、ペットボトル 93.3% という検証結果であった。

本研究では、再生可能資源であるビン、アルミ缶、スチール缶、ペットボトルの落下した音を識別するため、音センサを使用する。音を識別するための特徴量において、再生可能資源の落下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、この特性を考慮した MFCC の改良により識別率を高めることを目的とする。

3. 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 Fig. 1 に示す。

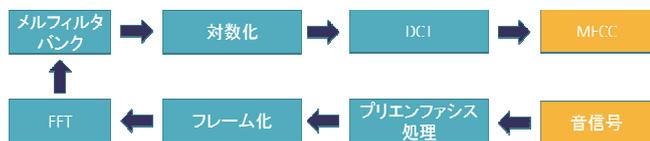


Fig. 1 MFCC Extraction Procedure

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アンファシス処理を行い、周波数成分を強調する。以下にプリア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。プリア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。フレームの長さは 0.6[s] である。以下にフレーム化されたデジタル信号を Fig. 2 に示す。ここで Fig. 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。

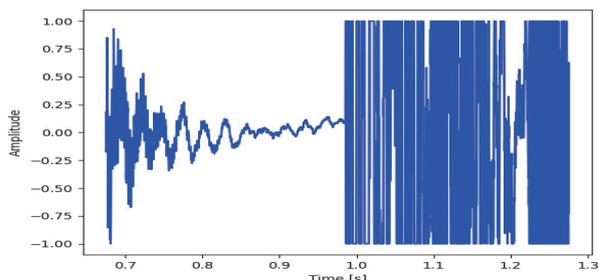


Fig. 2 Framed digital signal

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、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 Fig. 3 の振幅スペクトルが求められる。Fig. 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。

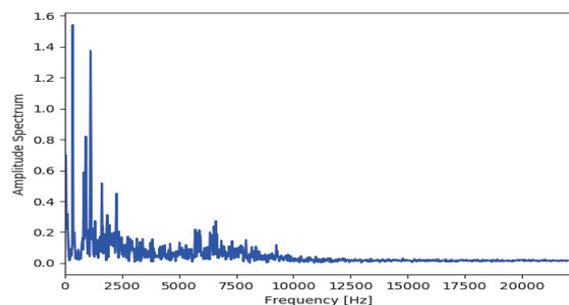


Fig. 3 Amplitude spectrum

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。以下に示す Fig. 4 は人間の聴覚性質を基に、低い周波数部分は細かく、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。

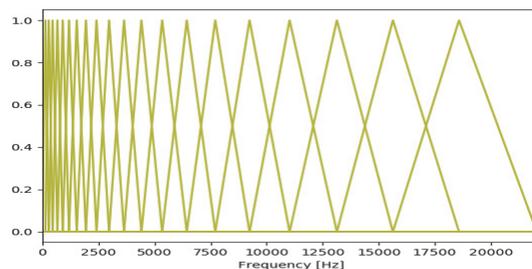


Fig. 4 Mel Filter Bank

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(Fig. 5).

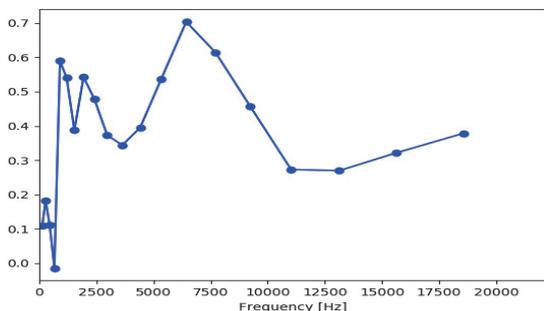


Fig. 5 Logarithmized mel band spectrum

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。

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

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 (Fig. 6 の左グラフ)。

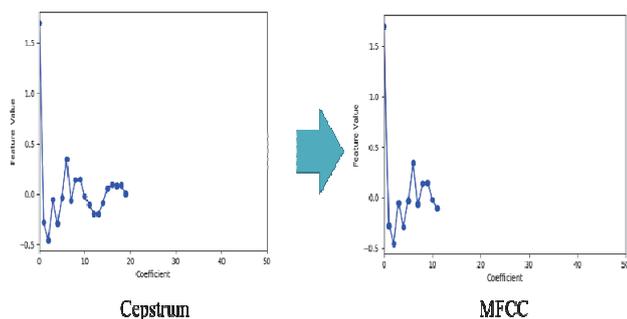


Fig. 6 Extraction of MFCC from cepstrum

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(Fig. 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。Fig. 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。Fig. 6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MFCCとなる。しかし、再生可能資源の落下音には、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版MFCCの提案

4.1. 再生可能資源の落下音を対象とした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再生可能資源の落下音は、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Fig. 7より音声と再生可能資源の落下音の周波数成分の違いを、縦軸を音の強さを表す振幅、横軸を周波数の高さを表す振幅スペクトルより示す。

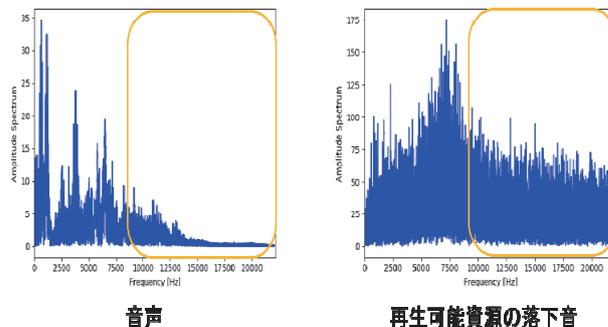


Fig. 7 Amplitude spectrum of voice and falling sound of renewable resources

Fig. 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、再生可能資源の落下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、再生可能資源の落下音の場合では、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率が高まることが期待できる。

4.2. MFCCの改良方法

再生可能資源の落下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率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

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1.4[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

識別の対象であるアルミ缶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(Fig. 8)。

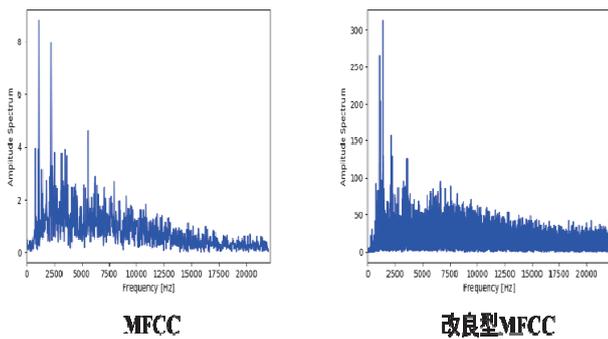


Fig. 8 Amplitude spectrum of MFCC and improved MFCC

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 Fig. 9 に示す。Fig. 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。

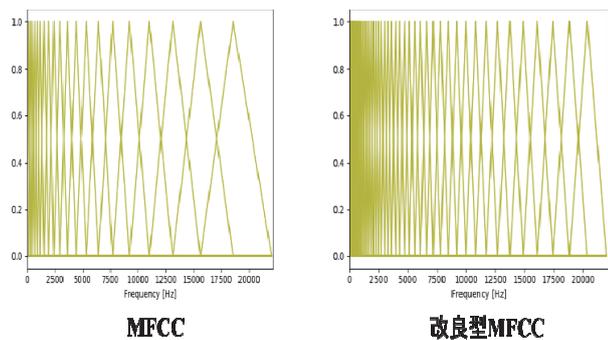


Fig. 9 Comparison of Mel Filter Banks

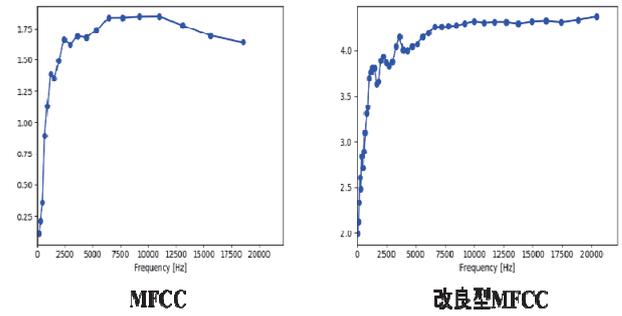


Fig. 10 Comparison of logarized Mel band spectrum

Fig. 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。Fig. 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。

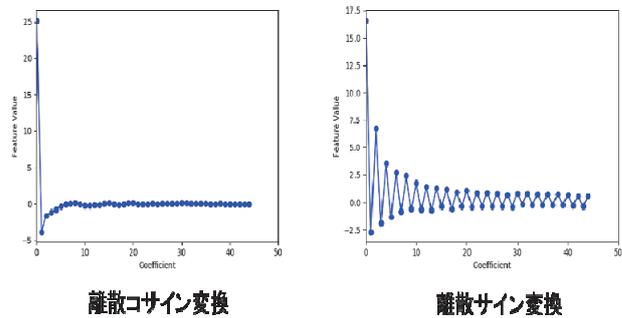


Fig. 11 Difference of discrete transformation

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 Fig. 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。

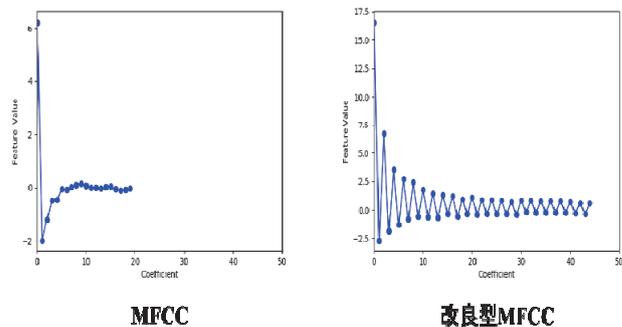


Fig. 12 Cepstrum of MFCC and improved MFCC

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。以下に、従来の MFCC と PCA によって抽出された改良型 MFCC を可視化した 3 次元プロットを Fig. 13, Fig. 14 に示す。

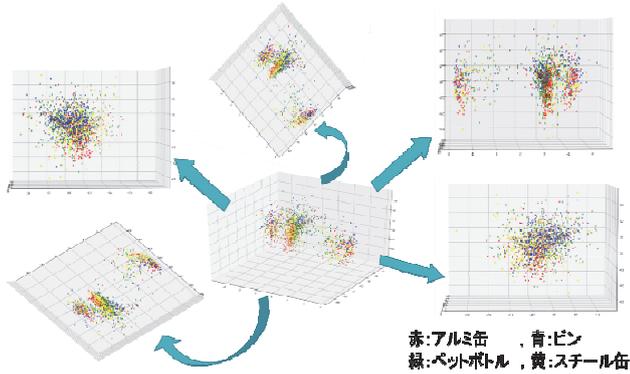


Fig. 13 Visualization of MFCC

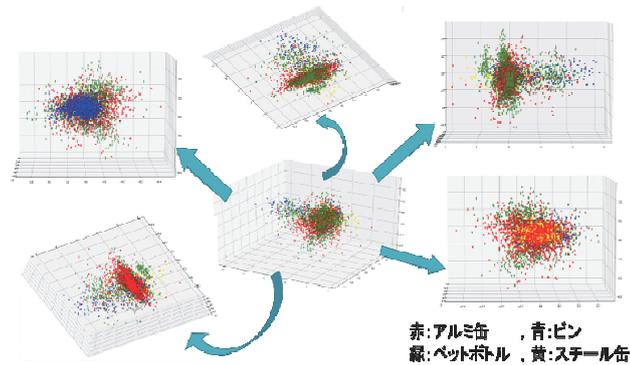


Fig. 14 Visualization of improved MFCC

Fig. 13, Fig. 14 は、多次元の特徴量を 3 分割し、それぞれ X 軸、Y 軸、Z 軸に合わせた 3 次元プロットである。また、これらの 3 次元プロットは、教師用データの特徴量にした値を使用している。赤色プロットがアルミ缶、青色プロットがビン、緑色プロットがペットボトル、黄色プロットがスチール缶を表している。Fig. 13 は通常の MFCC であるが、識別対象それぞれのプロットが分散していることが確認できる。Fig. 14 の改良型 MFCC では、全体的にプロットが分散せず、また識別対象それぞれのプロットが種類ごとに一定の範囲に収まっているため特徴量が上手く抽出されていると考える。

5. 検証結果と考察

Fig. 15 は、改良型 MFCC の評価を行ったソフトウェア構成と示したものである。

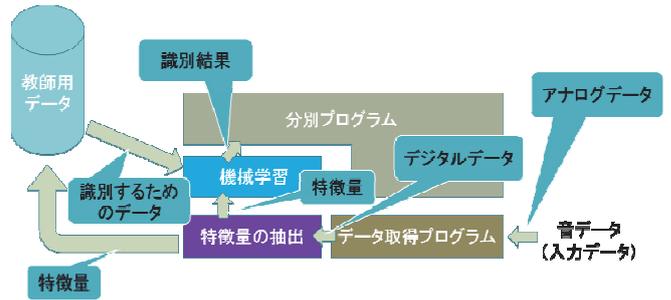


Fig. 15 Software configuration

本研究では、実際に再生可能資源の落下音を取得する為の音センサと落下地点にアルミ複合板を設置した原理検証用ゴミ箱を用意した。ここでアルミ複合板を設置した理由は、耐久性が高く、複数回再生可能資源を落下させたとしても、落下音の変化が少ないためである。再生可能資源がアルミ複合板に落下した時に発生する衝突音を入力データとして取得する。また、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-MCU03BK 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。本研究でも教師ありの機械学習である SVM を用いるため、特徴量により識別率の良し悪しが左右される。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 SVM よりどの再生可能資源の落下音であるかを識別した。

True label	Alcan-	84	1	2	13
	Bin-	1	75	14	10
	Pet-	1	8	83	2
	Stcan-	15	8	8	69
		Alcan-	Bin-	Pet-	Stcan-

平均識別率 = 77.75%

Fig. 16 Verification result by MFCC

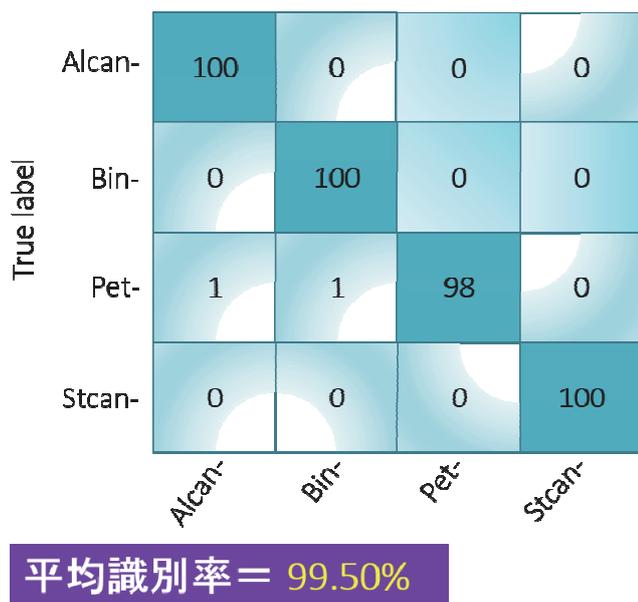


Fig. 17 Verification result by improved MFCC

教師用データを再生可能資源であるビン、アルミ缶、スチール缶、ペットボトルの4種類に対しそれぞれ100個ずつ用意し合計400個を使用した。また検証用データにも同様の量を用意した。教師用、検証用データとして録音した再生可能資源の種類は、ビン3種類、アルミ缶7種類、スチール缶4種類、ペットボトル7種類である。

検証結果は、従来のMFCCを特徴量として用いた場合が77.75%(Fig. 16)、改良型MFCCの場合が99.50%(Fig. 17)であった。SVMでは、scikit-learnをライブラリとして使用し、パラメータはグリッドサーチより調節を行い、最も識別率の高い組合せを使用している。SVMで使用したパラメータは、Cパラメータとgammaパラメータである。グリッドサーチにより探索した値は、Cパラメータを1, 10, 100, 1000, 10000とし、gammaパラメータではデフォルト値で探索を行った。

Fig. 17の改良型MFCCによる検証結果では、識別対象であるアルミ缶、スチール缶、ビンの3種類は検証用データ100個に対し、全て正しく識別したが、ペットボトルでは2回誤った識別をした。2回の誤った識別は、1回ずつアルミ缶とビンであった。今後より識別率を高めるためには、アルミ缶とビンの教師用データを増やすことや、アルミ缶とビンの落下音からまだ抽出されていない重要な特徴を分析する必要があると考える。例えば文献8)ではアルミ缶、ペットボトルなどの対象物をつぶす際の音を元に戻ろうとする音が特徴的であるという指摘があり、これらを参考にすることも検討していきたいと考えている。また時系列な依存関係を持つ特徴量による識別手法についても、本研究の手法との比較、評価を行う予定である。

6. まとめ

本研究では再生可能資源であるビン、アルミ缶、スチール缶、ペットボトルの自動分別に対して特徴量MFCCの改

良を行い、識別率の向上することを目的とした。検証結果より、従来のMFCCより識別率を約20%高い99.50%まで改善することができた。しかしながら、検証では限られた形状の再生可能資源から教師用、検証用データを構成している為、新たな形状の再生可能資源に対して識別が必ずしも可能ではないと考える。また、本研究の提案を自動分別ゴミ箱として実用化する場合、再生可能資源を識別することに加えて、再生可能資源以外の識別機能も必要である。

今後は、実際に自動分別ゴミ箱として使用し、識別率、環境性、実用性について検証する。実用化に向けて考慮すべき評価軸としては機能性、効率性、安全性、耐久性などがある。機能性については、ゴミ箱内のゴミの量が溜まった時に、管理者へゴミ箱内の状況を自動で知らせる機能や、蓄積されたゴミを圧縮する機能などを検討していく。また、効率性については実際に収集した回収資源の分量を学習し、4種類の再生可能資源を収納するための最適な容量を決め、ゴミ箱の容量を変更する方法を検討する必要がある。また安全性については、ゴミ箱内に異物が混入した場合、どう対処をするのかが非常に大きな課題となる。さらに耐久性については、室外での設置などの条件で、マイクなどのセンサや、認識を行うための機器（今回はラズベリーパイを使用）をどの様にして品質を保つかを検討しなければならない。これらを含め実用化に向けては今後も検討を行っていく。

また、再生可能資源の自動分別以外への適用、例えば人の行動認識や環境音の認識などへも本研究で提案したMFCCが有効であるかどうかを検証し、音を用いたIoTプラットフォームとするべく検討を進めていきたい。

参考文献

- 1) 撫中 達司, 大島 幸司, “再生可能資源IoT回収システムの実証実験結果について”, マルチメディア, 分散, 協調とモバイルシンポジウム, 2017
- 2) 板垣 弦矢, 撫中 達司, “選別一体化を目的とした再生可能資源の自動分別機試作による実用性の検討とその評価”, マルチメディア, 分散, 協調とモバイルシンポジウム, 2018
- 3) 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
- 4) Gierad Laput, Yang Zhang, Chris Harrison, “Synthetic Sensors: Towards General-Purpose Sensing”, 2017
- 5) 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
- 6) A. T. Garcia, O. R. Aragon, O.L. Gandara, F. S. Garcia, L. E. Gonzalez-Jimenez: Computacion y Sistemas, Intelligent Waste Separator, 2015
- 7) 自動回収機について | 株式会社誠宇ジャパン, http://www.seiujapan.com/?mid=auto_recycle
- 8) 稲邑 哲也, 古城 直樹, 畑尾 直孝, 得津 覚, 藤本 純也, 園田 朋之, 岡田 慧, 稲葉 雅幸, “飲料缶・ボトル類を目と手と耳で分別廃棄するヒューマノイド行動の実現”, 日本ロボット学会誌, Vol.25 No.6, 2007