

SVMを用いた身体動作・環境音の分類手法および

オノマトペ変換手法の提案

Onomatopoeic Text Generation from Physical State and Environmental Sounds using SVM

中 祐介† 吉田 直人† 河口 拓貴† 米澤 朋子†

Yusuke Naka Naoto Yoshida Hiroki Kawaguchi Tomoko Yonezawa

1. はじめに

人間は他者の存在や行動を理解し、気を配り、生活を営んでいる。日常生活の中には様々な生活音があり、周囲の人や物が発する環境音を聞き分けることで文脈や状況を理解し、他者とコミュニケーションを行う。また、人間は意識的・無意識的な身体動作を通じて、自身の状態を表すため、特に対面コミュニケーションにおいては、相手の状態や周辺の状況を考慮しながら話す内容や発言のタイミングを選択する。

しかし、近年、Twitter¹やSkype²のインスタントメッセージのような、ユーザの伝達情報がテキスト化されたコミュニケーションが多くの人に利用されるようになり、他者との対面的なコミュニケーションの頻度が少なくなっている。このようなコミュニケーションでは、ユーザの選別した情報のみが反映され、相手と空間的にも時間的にも一致していなくとも対話を行うことができ、その場限りでの情報伝達ではなく、時間を遡って確認することによってログとして蓄積された情報から長期的な状況理解が行える。ただし、このようなコミュニケーションはユーザの選択した意識的な情報のみが反映される場であり、他のユーザの周辺環境や相手の身体的状況を理解することが困難である。そのため、コミュニケーションのリアルタイム性の低下や、ユーザ間で齟齬が生まれ誤解を生む可能性がある。

本研究では、ユーザの「主体」情報である身体動作と、「周辺」情報である環境音をコミュニケーション内で提示することで、本来ユーザの意識的表現のみ含まれる文章であったテキストコミュニケーションに対し、無意識的な情報が含まれたテキストコミュニケーションの実現を目指す。リアルタイム性が備わったテキストコミュニケーションを実現することで、言葉だけでは伝わりづらい発言のタイミングや状況を伝達することが可能になると期待する。

しかし、身体動作や環境音といった異なった複数のモダリティは利用するコミュニケーションによって、適した提示表現を用いなければいけないと考える。例えば、テキストコミュニケーションに、身体情報は動画、環境音は音声情報として提示を行うと、伝達の遅延やプライバシーの過剰公開が発生する。このような問題の発生を軽減するためには、提示するモダリティを適した表現手法として統一することが考えられる。本研究では、テキストコミュニケーションにおいて提示が容易で、プライバシーの過剰公開を防ぐことを目的とし、身体情報と環境音の2つのモダリティをオノマトペのテキストとして反映させる。その目的を達成するため、本稿では身体動作と環境音の認識・分類、オノマトペのテキストへの変換手法を提案する。認識分類には、これまでに行ってきた聞き取り調査、評価実験を基に教師データを作成し、Support Vector Machine (以下 SVM) を用

いた。そして分類結果より、オノマトペのテキストとしてメディア変換を行うことで、環境音と身体動作のような様々なモダリティを、テキストコミュニケーション内で利用することに適したテキスト表現へと変換することが可能になると考えた。

2. 関連研究

2.1 SVMを用いた楽器同定手法

入力された音が何の楽器の音であるのかを推定する楽器同定を SVM によって行う試み[1][2]がある。北原らは、メル周波数ケプストラム係数 (以下 MFCC) と高速フーリエ変換 (以下 FFT) によって抽出した特徴量を SVM によって認識させ、楽器認識精度の比較を行った[1]。音のスペクトルにはスペクトル包絡と高音情報の2つの情報が含まれているが、MFCCによって、これらの情報を分離させ、その中でスペクトル包絡のみを利用することで、認識精度が向上することを示している。

また杉山は、自動採譜を行う際に楽器同定を行う事で、複数楽器の採譜をさせることを目指した[2]。この研究では、ケプストラムを使用し、ピアノとギターのカテゴリを可能とした。これらの研究から、ケプストラムによって求められるスペクトル包絡を用いることで、楽器情報を得ることができ、それによって認識率が向上することが示された。

本稿では、SVM を用いて環境音と人体動作をそれぞれ分類する。データの特徴量に環境音の時間長と、岡本ら[3]と同様に環境音の発生時と終了時における簡略化したスペクトログラムのスペクトルを用いる。スペクトル包絡を用いることで、詳しく音の分類を行うことが出来るが、計算処理に時間を要するため、本研究の目的とするモバイル環境や組み込みシステム内での、テキストコミュニケーションへの環境音モダリティの付与に適さない。それに対し環境音の時間長と、環境音の発生時と終了時における簡略化したスペクトログラムのスペクトルを特徴量として用いることで、環境音を音素単位ではなく、一塊の音声として大まかに取得することが出来る。これによって類似する環境音が大まかに分類されてしまうが、計算処理が短くなり、テキストコミュニケーションのようなリアルタイム性が求

表1 簡易スペクトログラムとスペクトル包絡の違い
Tab.1 Difference between Simple Spectrogram and Spectrum Envelope

| | 簡易スペクトログラム | スペクトル包絡 |
|-----|------------|---------|
| 速度 | ○ | × |
| 正確性 | △ | ◎ |

† 関西大学, Kansai University

1: <https://twitter.com/>2: <https://www.skype.com/ja/>

められる場面においても利用が可能になると考える（表1）。

2.2 音響的關係性を考慮した環境音-擬音語変換システム

近年は、音声だけでなく日常における環境音も認識対象として対話システムへの応用を目指す研究がある[4][5][6][7]。

特に山川らは環境音がどのように鳴っているのかを認識するマンマシンインタラクションの実現を目指して、環境音の音響信号の擬音語による認識を行った[4][5]。これは人間の環境音の知覚と擬音語の認知の知見に基づいて選択した音響特徴を用いて、音響信号を擬音語として認識するシステムである。出力される擬音語は、大きく分類すると母音と子音の音素のグループの2種類で、さらに細かく、子音のグループの中で、類似している音素を1つの纏まりとして出力し、誤認識の軽減を目指した。

本研究においても、環境音の知覚実験を基に、環境音のクラス分けを行った。しかし、2.1項で述べたように類似する音が提示されたとしても、大幅な解釈のずれが無ければ内容の理解が可能であると考えられるため、山川らのように複数の音素を1つのクラスとして纏める手法は用いない。

2.3 テキストへの新たな情報付与によるコミュニケーションの変容

テキストに付与する情報によりコミュニケーションを変容させる試み[8][9]がある。

山中らは、テキストコミュニケーションと通常の対面式コミュニケーションとの違いについて、アクセント表現を付加したテキスト表現の有効性について検証を行った[8]。この研究では、ユーザが自らの発言中の単語を自身の感情に合わせてテキストの位置を上下にずらすことで、本来伝わりづらい書き手の意思を読み手に伝達することを提案した。また伊藤らは、テキストコミュニケーションにおいて心情を伝えるために用いられる顔文字に注目し、人間の感情を多く表出する、表情や視線の動きを視覚化させる擬人化エージェントを利用し、伝達される感情や雰囲気について検証を行った[9]。これらの研究により、テキスト表現に新たな要素を与えることによる、表現の増加、ユーザの理解度の向上の有効性を示している。

しかしアクセント表現を与えたテキストや、擬人化エージェントでは、ユーザの感情という意識的な情報のみしか新たに伝えることができず、無意識的な情報は伝えられない。本研究においても、テキスト表現に新たな要素を与えることにより表現の増加やユーザの理解度を向上させることをねらうと同時に、ユーザの身体情報や周辺の状況のような無意識的な情報の伝達もねらう。そのため、ユーザの発言を表したテキストとは別に、オノマトペ表現のテキストを追加して伝達することで、ユーザの無意識的な情報の伝達を目指す。

2.4 アンビエントコミュニケーションチャンネル

Miyajimaらは、我々の研究と同様に、空間を共有しない非対面コミュニケーションにおけるユーザの生活状況などの手がかり情報を他者に伝達できない問題に着目し、非言語コミュニケーションによるメッセージを用いて、相手に通話を行っているようなつながり感を与えるプランタ型の通信デバイスを作成した[10][11]。このデバイスは赤外線センサで取得した人の動きやタッチセンサによる接触の情報などの非言語情報を用いて、人の情報を互いに送り合う

新しいコミュニケーションチャンネルとして提案されたものである。この研究では提案デバイスを用い、家族内のコミュニケーションにおけるつながり感や安心感の有無の検証を行った。その結果、軽い合図などの短期的なつながりや、人間関係という長期的なつながりに対してデバイスは有効であり、言葉を伴わない新たなコミュニケーションチャンネルとして有用であるという結果を残している。

Miyajimaらが新たな非言語コミュニケーションの実現をねらったのに対して、本研究でのねらいは、テキストを用いたコミュニケーション上への自動的なユーザ自身の状態や、周囲の状況の手がかり情報の追加によるコミュニケーションの擬似モダリティ付与である。ユーザの身体状態や、周辺状況の変化によって、ユーザの発言とは別に、オノマトペをテキストとして伝達することで、状況変化に合わせた発言タイミングなどを何気なく伝達するため、プランタ型デバイスに集約する上記の試みとは異なる。

3. 提案システム

3.1 システムコンセプト

本研究ではテキストコミュニケーションに、対面コミュニケーションで無意識的に伝達を行っているユーザの身体動作や状態、周辺の環境音の情報を、ユーザの意図的に行った発言とは別に自動的に反映させることを目指す。これによって、対面式のコミュニケーションのような、複数のモダリティが取得でき、相手の状況や状態に合わせたコミュニケーションが行えると考える。しかし、身体動作や環境音といった異なった複数のモダリティは、利用するコミュニケーションによって適した提示表現を用いる必要がある。例えば、テキストコミュニケーションに身体情報は動画、環境音は音声情報として提示を行うと、伝達の遅延やプライバシーの過剰公開が発生する。このような問題の発生を軽減するためには、提示するモダリティを適した表現手法として統一することが考えられる。本研究では、テキストコミュニケーションにおいて提示が容易で、プライバシーの過剰公開を防ぐことを目的とし、身体情報と環境音の2つのモダリティをオノマトペテキストとして反映させる。その目的を達成するため、本稿では身体動作と環境音の認識、分類、オノマトペテキストへの変換手法を提案する。

3.2 システム構成

従来のテキスト化されたコミュニケーションでは、ユーザが選択した意識的な情報を、テキストを介して伝達しあ

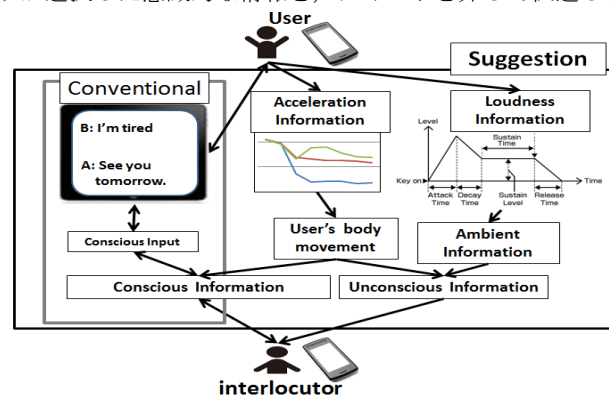


図1 提案コミュニケーションの構成
Fig.1 Configuration of the Proposed Communication

う。提案手法では、ここへユーザの無意識的な情報を自動的に反映させ、他のユーザに伝達することでリアルタイム性や状況の理解度を向上させたコミュニケーション（図1）が実現すると考える。

本研究で用いる無意識的な情報は、ユーザの主体情報を表す身体情報とユーザの周辺状況を表す環境音情報である。この2種類の情報を組み合わせることで、周辺の人物の存在や、ユーザの行動状況などを伝達することが可能となる。これらの情報は音声や映像のような表現手法を用いることで、対面式のコミュニケーションのような、より詳細に周辺状況や、身体状態変化の伝達を可能とする。しかし、ユーザのプライバシー保護の問題や、他者の私的情報の伝達による不快感の軽減も重視するため、テキスト表現、特にオノマトペによるテキスト表現を用い、問題解決を目指している。

本稿では、一般的に情報の表現手法が異なるモダリティを1種類のメディアへと変換するための、認識・分類システムを提案する。提案手法は、事前に行った調査を基に作成したデータを学習させた SVM によって分類を行った。SVM は環境音のような高次元の特徴量の分類にも適し、高速処理が行えるため、テキストコミュニケーションなどのネットワークを介したコミュニケーションでも利用できると考えた。

3.3 ユーザの身体動作認識・分類手法

本研究では、これまでにユーザが身体を動かした際に生じる加速度を測定し、身体動作・状態の検出を目指してきた [12] [13]。取得した加速度の値は、FFT 処理を用いて 0.78Hz 刻みで 8 つの特徴量として検出し、その中から、特徴量の変化が大きく表れる 2.34Hz~6.25Hz の 6 つの特徴量を取得した。3 軸において、これらの 6 つの特徴量と、各軸の加速度の値を加えた全 21 の特徴量を利用し、被介護

者の状態・行動の分類、認識を SVM によって行った（図2）。

本稿では、ネットワークを介したコミュニケーションにおける利用を目指し、ユーザが所持する携帯端末による加速度の数値の取得を行う。さらに、加速度による各周波数帯の周期性を学習データにするのみでは、ユーザの身体動作・行動を詳細に分類できないため、加速度と同様に、携帯端末に搭載されている角速度センサにより、角速度情報の数値を利用する（図3）。これにより、しゃがみこみ動作や起き上がり動作など、より複雑な動作を認識・分類することができると考える。

3.4 ユーザの周辺情報としての環境音認識・分類手法

本研究では、マイクロフォンを用いることで、ユーザの周囲の環境音を取得し、評価実験を基に設定された環境音の長さにより、擬音語テキストの要素である長音符「ー」、促音「ッ」、撥音「ン」、有無の決定を行った。また、時間軸方向でサンプル音を分析し、アタック、サスティン、リリースの長さの測定を行い、先行研究 [13] の知覚実験結果より設定した閾値（表2、3）に基づき、オノマトペの語尾や音韻を決定した。

本稿では、事前に行った知覚実験の結果を基にクラス分けした“t”, “k”, “b”, “d”, “g”, “z”, “j”, “p”, “s” の 9 種類の子音として環境音を分類、認識を行う。認識には、環境音の時間長と、環境音の発生時と終了時における簡略化したスペクトログラムのスペクトルを特徴量とし抽出し学習させ、SVM によって認識を行う。先行研究 [13] では、環境音の長さによる擬音語テキストの要素である長音符「ー」、促音「ッ」、撥音「ン」有無の決定というルールベースの分類手法を行ってきたが、知覚実験の結果を基に環境音を分類するコーパスベースの分類手法を加える。これにより環境音の分類が大まかになって

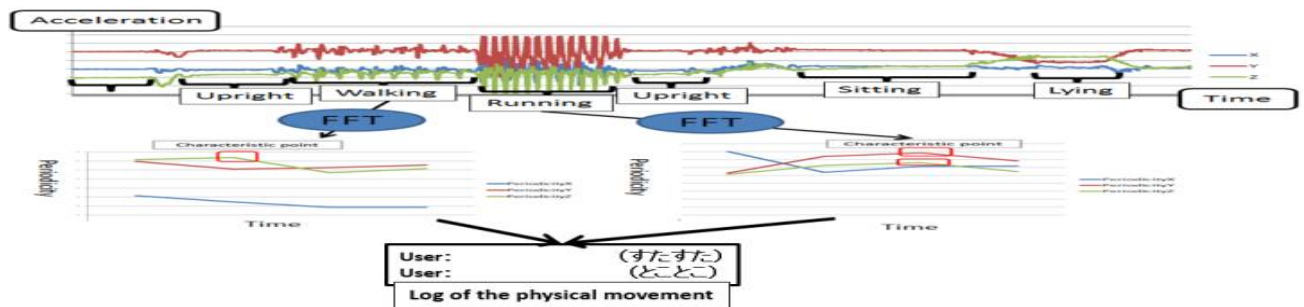


図2 加速度データ取得とFFT解析によるオノマトペテキスト出力
Fig.2 Acceleration data and the FFT analysis

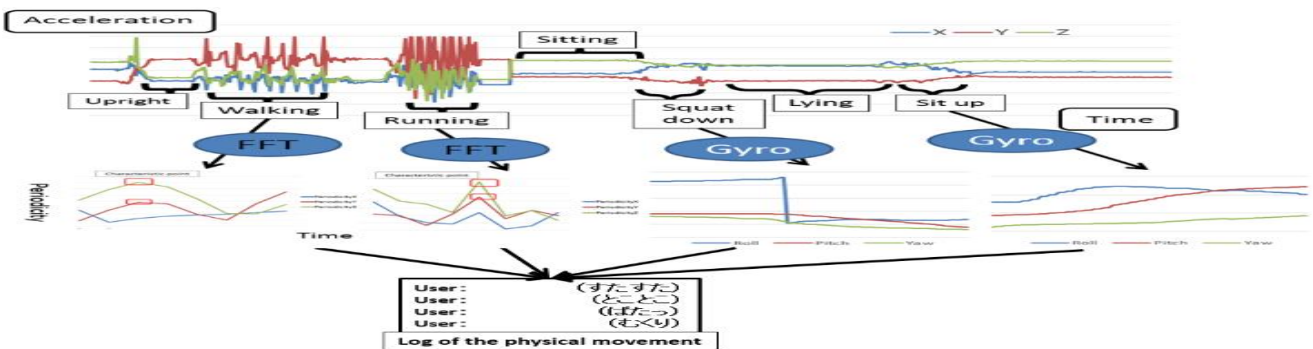


図3 角速度データと加速度データによるオノマトペテキスト出力
Fig.3 Recognition of Physical Movements with Angular Velocity Data and Acceleration Data

表2 時間軸に関する擬音語要素の分類
Tab.2 Classification of the Onomatopoeia Element about Temporal Axes

| 要素 | 用いる情報 | 擬音語文字 | 時間長 |
|-----|------------|-------|----------|
| 長音符 | sustainの長さ | 「ー」 | 150ms 以上 |
| 促音 | sustainの長さ | 「ッ」 | 70ms 以下 |
| 撥音 | releaseの長さ | 「ン」 | 750ms 以上 |

表3 音高・音色による擬音語要素の分類
Tab.3 Classification of Onomatopoeia Elements According to Tone and Pitch

| 要素 | 用いる情報 | 擬音語 |
|-----|----------|------|
| “i” | pitchの高さ | 「イ」 |
| “u” | pitchの高さ | 「ウ」 |
| “o” | pitchの高さ | 「オ」 |
| “p” | 音色・ノイズ量 | 「パ」行 |
| “b” | 音色・ノイズ量 | 「バ」行 |
| “d” | 音色・ノイズ量 | 「ダ」行 |

しまうが、本研究では、環境音をオノマトペとして分類、テキストコミュニケーションでの利用を目的としているため、認識精度ではなく、認識、分類処理の速度を重視した。また、類似したオノマトペとして出力されたとしても、大よそ同様の意味としてユーザ同士で理解し合えることが可能であると考えられる。

3.5 オノマトペ変換手法を用いたテキストコミュニケーションへのマルチモダリティの付与

本研究では、ユーザの主体・周辺情報を詳細に伝達することが可能な映像や音声といった表現を代替するテキスト表現として、特にオノマトペによるテキスト表現を用い、情報の概略を伝達することを目指す。オノマトペは端的にテキスト化された非言語情報で、大まかに状況を理解させることができるため、手軽に情報を取得できる。さらに、他の表現手法と比較すると、情報量が抽象化されているため、周辺情報の取得によるプライバシー問題をある程度軽減することが期待される。その反面、同じオノマトペの異なる場面での使い分けや、複数のニュアンスが考えられるため、一般化することが困難である。しかしこれらの問題は、単一でオノマトペを利用の際に起こることが主であり、複数のオノマトペ表示による状況の前後内容によって軽減されると考える。

我々はこれまでに、ユーザの周辺の環境音、身体状態・行動を表すオノマトペのテキストを提示したテキストコミュニケーションにおける他ユーザの状況の理解度、オノマトペのテキストの有効性、コミュニケーションの楽しさの

表4 簡易スペクトログラムを用いた学習データ数
Tab.4 The Number of Training Data with Simple Spectrogram.

| | | | | | | | | |
|------|------|------|------|------|------|------|------|------|
| t:21 | k:82 | b:54 | z:34 | j:79 | p:26 | s:28 | g:23 | d:34 |
|------|------|------|------|------|------|------|------|------|

表5 簡易スペクトログラムを用いたテストデータ数
Tab.5 The Number of Test Data with Simple Spectrogram.

| | | | | | | | | |
|------|------|------|------|------|------|------|------|------|
| t:21 | k:81 | b:53 | z:34 | j:78 | p:25 | s:27 | g:22 | d:34 |
|------|------|------|------|------|------|------|------|------|

表6 スペクトル包絡を用いた学習データ数
Tab.6 The Number of Training Data with Spectrum Envelope.

| | | | | | | | | |
|------|------|------|------|------|------|------|------|------|
| t:20 | k:50 | b:50 | z:30 | j:50 | p:25 | s:25 | g:20 | d:25 |
|------|------|------|------|------|------|------|------|------|

表7 スペクトル包絡を用いたテストデータ数
Tab.7 The Number of Test Data with Spectrum Envelope.

| | | | | | | | | |
|------|------|------|------|------|------|------|------|------|
| t:20 | k:50 | b:50 | z:30 | j:50 | p:25 | s:25 | g:20 | d:25 |
|------|------|------|------|------|------|------|------|------|

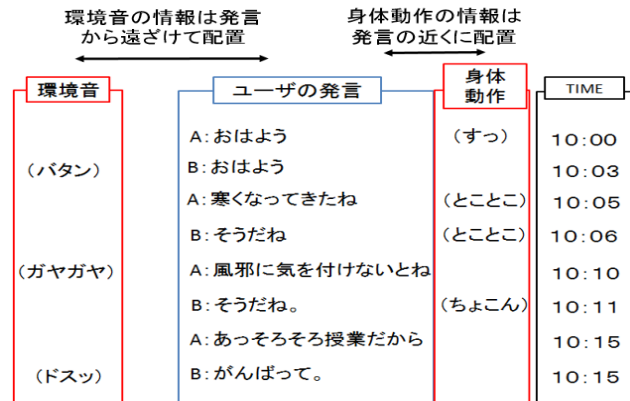


図4 アンビエント情報テキストを用いたコミュニケーション

Fig.4 Text Communication with Ambient Information

増加を検証し、有意性を得た [12].

本システムでは身体動作に関するオノマトペと環境音に関するオノマトペという2種類の異なるオノマトペをテキストコミュニケーションに付与し反映させるために、情報が混在し混乱を招いてしまう。そのため、身体的な情報はユーザの発言により近く、環境音は少し遠ざけて配置した。これにより、身体情報と環境音情報が混在することが無くなるため、状況を想像した際に混乱を招くことを減少させる(図4)。

4. SVMを用いた環境音の認識率の検証

4.1 実験の目的

3節で述べたユーザの周辺情報として利用する環境音の、認識・分類を行うシステムの認識率の検証を行う。環境音の時間長と、環境音の発生時と終了時における簡略化したスペクトログラムのパワースペクトルを特徴量とし抽出し、学習させSVMによって認識させた際の認識率と正解率の高さ、誤認識された際には、誤認識された子音が類似しているものとして分類されているかを検証する。これによって、正解、または類似している子音として出力されなかったものは、どのように認識されたかを確かめ、今後どのように学習、認識させるかを検討することを目的とする。

また同様の環境音のスペクトル包絡を学習させた認識機に環境音を認識させ、その認識結果と比較を行う。

4.2 実験の設定

実験設定

SVMを用いた環境音の認識実験用の音源には、事前に行

った知覚実験を基に用意した“t”，“k”，“b”，“d”，“g”，“z”，“j”，“p”，“s”の9種類の子音の音を用意した。録音した音源の半数を学習用，残り半数をテスト用とし，認識実験を行った。各音源の時間長と簡易スペクトログラムを特徴量として学習させたデータ数とテスト数は表4，表5，スペクトル包絡の特徴量として学習させたデータ数とテスト数は表6，表7とおりである。

実験仮説

SVM は学習データが増える毎にクラス間のマージンが小さくなるため，異なるクラスとして誤認識されることがあるが，本研究においては，“t”と“k”，“z”と“j”のような類似する子音が誤認識されたとしても，かけ離れたものでなければ，おおそ認識はできていると考える。また，子音が類似していない場合であっても，学習データが少ないクラスは，多いクラスとして誤認識されてしまうことがあると推測する。

4.3 実験結果

簡易スペクトログラムを特徴量として用いた場合の，学習データのクラス（列）に対するテストデータの認識結果（行），認識正解数，認識誤り数，全データ数より求めた認識率と正解率を表8に，同様にスペクトル包絡を用いた場合を表9示す。ここでの認識率と正解率は式(1)と式(2)によって求められる。

$$\text{認識率} = \text{認識正解数} / \text{全データ数} \quad (1)$$

$$\text{正解率} = (\text{認識正解数} - \text{認識誤り数}) / \text{全データ数} \quad (2)$$

表8より，簡易スペクトログラムと環境音の時間長を特

徴量として利用した場合，“b”，“j”，“p”，“s”の認識率がおおよそ50%を超える高いものとなった。しかし，認識率が，低い結果となった子音も見られ，“z”に関しては，認識されないという結果となった。また，認識率が高かった“b”や“p”に関しても正解率は大きく低下していることがわかる。また，スペクトル包絡を特徴量として学習させた場合の認識制度はかなり高いものであった。

5. 考察

本稿での認識結果において，子音毎の認識率に大きな差があったのは，認識機に学習をさせたデータ数に差があったことが関係していると考えられる。“k”や“j”の学習データ数と“s”や“t”を比較するとおおよそ2倍以上の差が存在する。このことが認識精度を大きく下げる結果となった。この問題は，今後更なる知覚実験を基に，学習データを追加することで，クラス間の学習データ数の差を減少させることで，解決する可能性があると考えられる。また，“k”という破裂音が“j”という摩擦音として誤認識された理由として，簡略化されたスペクトログラムの高周波数帯のパワースペクトルがどちらも強く，類似していたことから，誤認識率が高くなったと考えられる。今後このような問題を減少させるには，新たな特徴量を抽出し，学習させる。もしくは，誤認識の原因になっている特徴量を省くことが必要になると考える。また，スペクトル包絡を特徴量として学習させた結果と比較すると，簡易スペクトルグラムを用いた場合の認識制度はかなり低いといえるが，今回の実験においては，学習時間，認識時間を計測していなかったため，一概にスペクトル包絡を特徴量として用いることは

表8 提案パラメータによる認識結果
Tab.8 Recognition Result with Simple Spectrogram

| 認識結果 \ クラス | “t” | “k” | “b” | “d” | “g” | “z” | “j” | “p” | “s” |
|------------|------|-------|-------|-------|------|------|-------|-------|-------|
| “t” | 1 | 5 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| “k” | 5 | 30 | 11 | 0 | 0 | 22 | 7 | 1 | 2 |
| “b” | 8 | 2 | 29 | 27 | 8 | 1 | 7 | 2 | 0 |
| “d” | 1 | 5 | 5 | 7 | 0 | 0 | 0 | 2 | 0 |
| “g” | 0 | 3 | 2 | 0 | 2 | 1 | 2 | 0 | 0 |
| “z” | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| “j” | 0 | 22 | 5 | 0 | 9 | 9 | 61 | 1 | 2 |
| “p” | 5 | 14 | 1 | 0 | 0 | 0 | 0 | 16 | 0 |
| “s” | 0 | 0 | 0 | 0 | 3 | 2 | 1 | 3 | 24 |
| 認識率 | 4.8% | 37.0% | 54.7% | 20.6% | 9.1% | 0.0% | 78.2% | 64.0% | 88.9% |
| 正解率 | 0.0% | 0.0% | 9.4% | 0.0% | 0.0% | 0.0% | 56.4% | 28.0% | 77.8% |

表9 スペクトル包絡パラメータによる認識結果
Tab.9 Recognition Result with Spectrum Envelope

| 認識結果 \ クラス | t | k | b | d | g | z | j | p | s |
|------------|-------|-------|-------|------|-------|-------|------|-------|-------|
| t | 19 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| k | 1 | 46 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| b | 0 | 1 | 48 | 0 | 0 | 1 | 0 | 0 | 0 |
| d | 0 | 0 | 0 | 25 | 0 | 0 | 0 | 8 | 0 |
| g | 0 | 0 | 0 | 0 | 17 | 3 | 0 | 0 | 0 |
| j | 0 | 1 | 0 | 0 | 1 | 1 | 50 | 0 | 2 |
| p | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 15 | 0 |
| s | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 0 | 23 |
| 認識率 | 95.0% | 92.0% | 96.0% | 100% | 85.0% | 76.7% | 100% | 60.0% | 92.0% |
| 正解率 | 90.0% | 84.0% | 92.0% | 100% | 70.0% | 53.3% | 100% | 20.0% | 84.0% |

できない。また、今後 k-近傍法のような別の認識機との認識速度、認識率の比較を行うことで、本研究の目的とするモバイル環境や組み込みシステム内での、テキストコミュニケーションへの環境音モダリティの付与に適した手法を利用することができると思われる。

6. おわりに

本研究では、テキストコミュニケーションのような、ユーザの選別した情報のみが反映され、ユーザが意識して伝達を行わない主体情報である身体動作や周辺状況の情報である環境音を提示することで、対面式のコミュニケーションのようなリアルタイム性を出すことを目指した。複数のモダリティを異なる表現手法で反映させることは、伝達速度の低下や、過剰な情報公開という問題が発生する。これを解決するために、統一された表現手法として伝達することが必要となる。本稿では、解決手法としてオノマトペテキストとして表現手法を統一させることを目的とし、SVMを用いて身体動作と環境音をオノマトペテキストに分類するシステムを提案した。教師データの特徴量には本稿では、環境音の時間長と、環境音の発生時と終了時における簡略化したスペクトログラムのパワースペクトルを特徴量として用い、“t”、“k”、“b”、“d”、“g”、“z”、“j”、“p”、“s”の9種類の子音として学習させた。

また、この環境音を学習した認識機を用い、認識率を測定する実験を行った。結果として、学習データの差によって、あまり制度の高い認識率は得られなかったが、今後新たな特徴量を利用することや、学習データをさらに追加させることで、誤認識を度減少させ、オノマトペのテキストへと変換させるための子音の分類が可能となると考える。

さらに、ユーザの身体動作は加速度センサを用い、FFT処理により行動の周期性の特徴量を検出し、それらの値をSVMによってユーザの状態や行動と結び付けることを検討した。加速度情報のみでは、身体動作の違いは分類できるが、変化を詳細に分類することは困難であった。そのため、角速度情報を加速度同様に取得し、特徴量として利用することで、より細かな分類が行えると考える。

今後は、環境音と身体動作、状態の組み合わせによるクラス分けなども考慮しながら学習データを作成し、オノマトペテキストに変換する手法、オノマトペの表現手法の効果、システムによって提示されるオノマトペテキストと実際の状況の比較による確からしさの検証を行う予定である。

謝辞

本研究は、科研費 25700021 および科研費 24300047 の助成の一部を受け実施したものである。

参考文献

- [1] 北原 聡志, 甲藤 二郎:Support Vector Machine を用いた楽器同定における特徴量の比較, 電子情報通信学会総合大会講演論文集 2005 年_情報・システム(1), 125, 2005-3-07.
- [2] 杉山幸恵:SVM を用いた楽器音の識別, 函館工業高等専門学校卒業論文, 2006 年度.
- [3] 岡本 尚紀, 米澤 朋子, 山添 大丈, 安部 伸治, 服部 文夫, 萩田 紀博:簡略化した音情報によるプライバシー保護型の生活状況識別;研究報告. MVE, マルチメディア・仮想環境基礎, Vol.110, No.35, pp.7-12, (2010).

- [4] 山川 暢英, 北原 鉄朗, 高橋 徹, 尾形 哲也, 奥乃 博:擬音語と環境音の音響的関係性を考慮した環境音-擬音語変換システム, 人工知能学会全国大会, 1C2-OS4b-4, (2011).
- [5] 山川 暢英, 北原 鉄朗, 高橋 徹, 駒谷 和範, 尾形 哲也, 奥乃 博:環境音から擬音語への自動変換における特徴量抽出法の検討, 全国大会講演論文集第 72 回平成 22 年(2), “2-257”-“2-258”, 2010-03-08.
- [6] Eronen, A. J., Peltonen, V. T., Tuomi, J. T., Klapuri, A. P., Fagerlund, S., Sorsa, T., Lorho, G., and Huopaniemi, J.: Audio-based context recognition, IEEE Trans. on ASLP, Vol. 14, No. 1, pp. 321-329 (2006).
- [7] Chu, S., Narayanan, S., and Kuo, C.: Environmental Sound Recognition With Time-Frequency Audio Features, IEEE Trans. on ASLP, Vol. 17, No. 6, pp. 1142-1158 (2009).
- [8] 山中 啓祐, 鈴木 悟, 高橋 利幸, 森口 友也, 後藤 清豪, 大東和 忠幸, 大西 雅宏, 野口 尚吾:アクセント付加による豊かな表現が可能なテキストコミュニケーションの実現, インタラクション 2010, Vol. SB, No. 17, (2010).
- [9] 伊藤 淳子, 宗森 純:擬人化エージェントを介したチャットにおける顔文字と対話雰囲気との関連性の分析, 情報処理学会研究報告. GN, グループウェアとネットワークサービス, Vol. GN, No. 67, pp. 127-132, (2010).
- [10] Asami Miyajima, Yoshihiro Itoh, Masako Itoh, Takumi Watanabe: “Tsunagari-kan” communication: Design of a new telecommunication environment and a field test with family members living apart, INTERNATIONAL JOURNAL OF HUMAN-COMPUTER INTERACTION, Vol. 19, Issue 2, p. 253-276, (2005).
- [11] 宮島 麻美, 伊藤 良浩, 渡邊 琢美:社会実証実験によるなり感通信の効果の検証と分析, インタラクション 2003 論文集, 情報処理学会シンポジウムシリーズ, Vol. 2003, No. 7, pp. 271-277, (2003).
- [12] 中 祐介, 木村 圭介, 中谷 友香梨, 米澤 朋子:擬音語・擬態語テキストを利用した持続的コミュニケーションチャンネルの提案;電子情報通信学会技術研究報告, Vol. 113, No. 283, pp. 69-72, (2013).
- [13] 中 祐介, 木村 圭介, 中谷 友香梨, 米澤 朋子:テキストコミュニケーションへの環境音擬音語付与手法の提案;ヒューマンインタフェース学会研究会, Vol. 15, No. 6, pp. 51-56, (2013).