

擬音語と書体表現を用いた環境音の可視化

山本貴史 松原正樹 斎藤博昭

慶應義塾大学大学院

Visualization of Environmental Sounds using Onomatopoeia and Effective Fonts

Takashi Yamamoto Masaki Matsubara Hiroaki Saito

Keio University

{yamataka, masaki, hxs} @ nak.ics.keio.ac.jp

概要

我々は身の回りの環境音を通じて、周りの状況を把握し、危険や異常を感じている。さらに我々は、これらの環境音擬音語を用いて表現することで、他人にその場の状況や様子を伝えている。特に日本語では、擬音語の表現が多様であり、コミュニケーションにおいて非常に重要な役割を担っている。コンピュータが環境音を認識できるようになれば、マンマシンインタラクションはより自然なものとなる。

従来の研究では、擬音語や記号による検索や音色の類似度による表示など、環境音の検索を容易にすることが行われてきたが、環境音を視覚的に認識可能にするという研究はなかった。環境音の視覚的な認識は、耳の不自由な人間だけでなく、新しいコンテンツの創出や、環境音の視覚的な検索も可能となる。

そこで本論文では、環境音を擬音語へと変換し、さらに音の大きさや高さといった音響特徴をその書体へ反映させることで、本来、聴覚的に知覚される環境音を視覚的に認識可能にすることを目的とする。環境音は音声と発音原理が異なり、既存の音声認識手法ではうまく認識できない。そのため、擬音語への変換は、環境音のスペクトルの包絡を素性としたサポートベクタマシンと環境音の残響音の長さをを用いて行う。提案手法を用いることで、50種類以上の一音節の擬音語を創りだすことができ、さらにこれらの擬音語の組み合わせで、より多様な擬音語も創りだすことができる。また、明朝体やゴシック体などのフォントだけでなく、斜体、太字、文字の大きさや色の濃淡を組み合わせることで、環境音の特徴を表現する。

環境音と提案手法による表現の結びつけを行う実験の結果、全体で89%の正答率となり、提案手法により、環境音を適切な擬音語へと変換できるだけでなく、音響特徴を書体へ反映できていることが確かめられた。

Abstract

We are surrounded by various sounds. We can recognize what is going on through such environmental sounds and notice immediate danger or unusual events by crash sound or motor sound, for example. On the flip side, when we tell others the situation around us, we represent environmental sounds by onomatopoeia. Especially in Japan, onomatopoeia is very important to communicate with others, because there are a variety of onomatopoeias in Japanese. If computers can recognize environmental sounds and onomatopoeia, it would be more natural for human to interact with computers.

While environmental sound retrieval using onomatopoeia was pursued in previous research, there has been little work done which tries to transform environmental sounds to onomatopoeias. This paper proposes a method not only to do such transformation but also to reflect its sound effect such as loudness or sonority by using various fonts and changing their sizes.

Because the generation mechanism of environmental sounds is different from human voice, conventional speech recognition techniques can not be adopted. Thus, a support vector machine has been built in which spectrum contour is used as main features. Our system exploits the reverbs of environmental sounds in addition to SVM. Our method is capable of transcribing more than 50 kinds of onomatopoeia. Sound effects of loudness and the pitch level of sounds are expressed with various combination of normal fonts, italics and bold faces.

We evaluated how our system converts environmental sounds into onomatopoeias effectively and appropriately. 89% precision was attained, which proves effectiveness of our system.

1 はじめに

我々の身の回りには非常に多くの環境音が存在している。環境音とは、“衝突音”や“動作音”などの身の回りの音のことである。我々はこの環境音を通じて、周りの状況を把握し、危険や異常を感じている。そして我々は、これらの環境音を“チリン”や“パチパチ”といった擬音語を用いて表現している。Wakeらは、人間が他人に音を伝達する際には擬音語が頻繁に使われ、音源や音の印象による表現よりも確に音情報を伝えられるということを、聴取実験により示している[1]。このように我々は、環境音を擬音語を用いて表現し、他人にその時の状況や様子を伝えている。特に日本語においては、擬音語辞典[2]が出版されるなど、その表現が多様であり、コミュニケーションにおいて非常に重要な役割を担っている。また、環境音の擬音語への変換だけでなく、環境音を可視化させ、視覚的に認識可能とすることは、耳の不自由な人間や、なんらかの理由により音を聞くことができない場合に、音を伝える重要な手段となる。日本の漫画では、音情報やその場面の状況を表現するために、多様な擬音語が用いられており、漫画に、よりインパクトを与える重要な手段となっている。このように我々は、環境音を擬音語として表現しており、コンピュータに環境音を認識させることができれば、人間とコンピュータのインタラクションはより自然なものとなる。しかし、擬音語はその表現に個人差があり、ある人間が“チーン”と表現した音を別の人間が“キーン”と表現することもある。また、手を叩く音が、拍手の時には“パチパチ”と表現され、お参りの時には“パンパン”と表現されるように、同じ音でもその状況やイメージによって、擬音語表現が別のものとなることもあり、擬音語の表現は、様々な要因によって変わることがある。そのため複雑な問題となり、環境音の擬音語化や可視化はあまり研究がなされていない。

本稿では、環境音を擬音語へと変換し、さらに音の特徴や雰囲気を書体へ反映させることで、本来聴覚的に認識している環境音を、視覚的に認識可能にする手法を提案する。環境音はその発音原理が音声とは異なることから、既存の音声認識手法ではうまく認識することができず、環境音を擬音語へと変換するには別の手法が必要となる。問題を単純化するために、本稿では、音の発生状況やイメージ、表現の個人差は考慮しない。また、扱う環境音は単音であり、日本語の擬音語へ変換を行う。提案手法では、環境音のスペクトル包絡と残響時間から擬音語表現を決定する。また、人間の感じる音の特徴を書体表現と結びつけることで、音の特徴や雰囲気をも視覚的に認識できるようにする。

環境音の言語的可視化は、写真を撮影する際に、その場の音を認識し可視化結果を写真に合成することで、漫画のように音が可視化された写真にするなどといった新しいコンテンツの創出に応用可能である。また、異常音の擬音語表現の研究[3][4]や、環境音の可視化によるろう者の環境認識の研究[5]などもあり、危険音が鳴っていることや周囲の状況を耳の不自由な人間や音楽を聴いている人間などに視覚的に通知することも可能となってくる。以上のように、本研

究は、エンターテインメントの分野だけでなく、実用的な分野への応用も考えられる。

本稿の構成は以下の通りである。2節では、関連研究および本研究の位置付けを述べる。3節にて、本稿の目的を達成するための提案手法を説明する。4節にて、提案手法の有用性を示すための実験方法を、5節にて、その実験結果について述べる。6節で、実験結果について考察を行い、7節にて結論を述べる。最後に8節にて、今後の展望を述べる。

2 関連研究

環境音に関する研究および、書体イメージと音声イメージとの関連性に関する研究について述べる。

2.1 環境音の表現方法

環境音には様々な種類があるため、擬音語以外の様々な表現方法が提案されている。

小林らは音色記号を用いて環境音を表現している[6]。音色記号は音色を記号化するものであり、“擬音語”や、音色を音声の音韻表記と同等に記述することで、全く新しい記号体系よりもなじみ易いとされる“微視的音色”、そして、擬音語や微視的音色を複数含み、音色が類似したものを分類し、包括化して一つの記号とした“巨視的音色”を提案している。

石原らは書き起こしに基づく音声素グループを用いて環境音を表現している[7]。書き起こしに基づく音声素グループは音声素のあらゆる組み合わせを認めるものであり、音素集合は、学習用データの書き起こし(人間の付与した擬音語)中に十分な出現頻度を持つ音声素グループで構成される。このような音声素グループを用いることで、人間が環境音を擬音語にする際の個人差を吸収した表現が可能となるとされる。例えば、ある人間が“チーン”と表現した音を別の人間は“キーン”と表現するかもしれないが、このような音は音声素グループを用いた表現を用いることで、“/k-t//i://N/”と表現できる。

2.2 環境音検索

マルチメディアコンテンツを作成する上で音は非常に重要である。しかし、イメージに合う音を探すことは容易ではない。そのため、イメージ通りの環境音を検索するための研究がいくつかなされている。

Osakaらは環境音を視覚的に検索できるシステムを提案している[8]。2.1節で述べた巨視的音色を用いて環境音を表現し、それらを3次元表示をすることで、視覚的に環境音を検索できる。環境音の24次元のメル周波数ケプストラム係数(MFCC)を主成分分析を用いて3次元にすることで、表示を行っているため、似たような巨視的音色が近傍に配置されるようになっており、目的の環境音を検索する際に効果を発揮する。このインタフェースを用いることで、イメージ

通りの環境音を見つけるまでの検索時間は、リストから検索するのに比べて、74%の時間で検索できるとしている。

石原らは擬音語を用いて環境音を検索できるシステムを提案している [9]。多くの音検索システムでは、“音源名”や“音長”などでしか検索できないが、擬音語をクエリとして用いることが可能になれば、ユーザのイメージ通りの環境音を検索することが容易となる。さらに、環境音を擬音語で表現する際には個人差が出てくるが、この個人差を、2.1 節で述べた書き起こしに基づく音声素グループを用いることで吸収し、適切な検索結果が出力されるようにしている。また、環境音の認識では、MFCC を特徴量とした隠れマルコフモデルを構築するという音声認識で用いられる手法を応用している。

2.3 環境音からの特徴量抽出

環境音は音声とはその発音原理が異なり、その特徴として、音響波形の時間変化が大きいという非定常性や、音声の子音や母音にあたる波形が存在しないことが挙げられる。そのため、環境音を認識するには、音声認識手法とは異なる特徴量抽出法や認識手法を行う必要がある。

Chu らは環境音のような非定常信号の音源を同定する手法として、Matching-Pursuit(MP)[10]に Gabor 基底を用いて、時間周波数領域での分析窓内の特徴量を抽出することを提案している [11]。これにより従来は時間変化抽出のために短く設定されていた窓長を、より長くしても特徴抽出が行えるようになるとしている。

また、Yamakawa らは時間周波数領域での隣接窓間の特徴量抽出法について研究を行っており [12]、Chu らと同様 MP により抽出された特徴量の隣接窓間の時間変化を HMM でモデル化することにより、環境音の音源同定率が向上することを示した。

2.4 スペクトル構造と擬音語表現

比屋根らは、ガンマトーンを用いた擬似単発音についての認知実験を行い、中心周波数、残響時間、周波数ゆらぎに対するパラメータ依存性を定量的に解析した [13]。“カン”や“チン”などの衝突音では中心周波数により母音が変化し、残響時間が語尾変化して表されることや、衝突音に周波数ゆらぎを加えていくと、破裂音として認識されるようになることを示した。

2.5 書体と音のイメージの関係

野宮らは和文書体のイメージと音声イメージの関係について研究を行っている [14]。擬態語を様々な書体や文字の太さで表現し、どの表現が最も擬態語の音声イメージ（語感）に適合した表現かを調査することで、例えば、濁音のある表現は角ゴシック体で太字が

選択される傾向があるなど、音声イメージと書体表現に関連性があることを示した。また、明朝体には“鋭い”や“綺麗”といったイメージがあるなど、フォントや太字などのウェイトと形容詞にも関連性があることを示した。

2.6 本研究の位置付け

これまでに環境音の検索や環境音を記号へと変換するという研究はなされてきたが、環境音を視覚的に認識可能にするという研究はなされていない。環境音を視覚的に認識できるようにするには、人間が聴覚から得ている情報を視覚的に得られる必要があり、どのような音なのかということが的確に伝わる表現方法を取るべきである。2.1 節で述べたように、環境音の表現方法には様々なものが提案されているが、Wake らの聴取実験 [1] から分かるように、音を伝えるには擬音語を用いて表現するのが効果的である。また、2.5 節で述べたように、音のイメージと書体のイメージには関連性があり、音の特徴を書体を用いることで伝えることができる。

そこで本研究では、環境音を擬音語へと変換し、さらに環境音の特徴を書体表現へと反映させ、本来、聴覚的に認識する環境音を視覚的にも認識可能にすることを目指す。これにより、なんらかの事情により環境音を知覚できない場合であっても、環境音の情報を得ることが可能となる。

3 提案手法

環境音を擬音語へと変換し、さらに環境音の特徴を書体表現へと反映させることで、環境音を視覚的に認識可能にする手法を提案する。

我々は、過去に同様の研究を行っている [15] が、本稿では、さらに音の迫力と鋭さの考慮、および、文字の太さと色の濃淡、フォントを書体の表現方法として活用することで、人間の音感の書体表現への反映の拡張を行っている。これらの詳細については後述するが、人間の音感と書体表現方法の関連付けを増やすことで、より多くの情報を書体として表現し、可視化することが可能となっている。

提案手法の処理順序を以下に示す。

1. 音響波形の分割
2. 擬音語への変換
3. 書体の決定
4. 可視化表現の表示

3.1 音響波形の分割

2.3 節で述べたように、環境音を既存の音声認識手法を用いて認識したのでは、うまく認識することができない。これは、環境音は音声に比べ、波形の時間変化が大きいためであり、音声認識における音素単位へ

の分割を環境音に用いると、人間が環境音を擬音語で表現する際よりも多くの音素に分割してしまう。また、環境音には子音や母音にあたる波形が明確に存在しないこともひとつの要因と考えられる。そこで、環境音を音素単位ではなく、擬音語の音節単位に分割する。人間が環境音の音響波形を一音節の擬音語にする単位は図 1 のような音響波形の一山であり [9][16]、この音節単位ごとに擬音語への変換を行う。また、Ishihara らはこの音節の構造は聴取者に依存しない傾向が強いことを聴取実験で示している [17]。

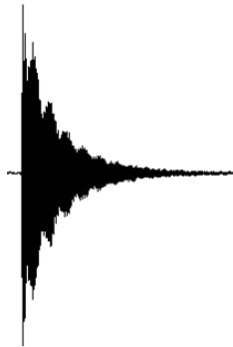


図 1: 擬音語の一音節に相当する音響波形の例

本手法では、音の大きさ（ラウドネス）の変化から環境音の鳴り始め位置を求め、この一山単位に分割する。環境音の鳴り始め位置は、ラウドネスが急激に上昇し始めた位置として求められるため、一定時間ごとにラウドネスを計算し、ラウドネスの変化量が閾値以上であれば、環境音の鳴り始めと判断する。ただし、ラウドネスの極大値が直前の極大値よりも小さい場合には、環境音の鳴り始め位置にはならない。このように音の鳴り始めを求めることで、一山単位に分割していく。

例えば、図 2 のような音響波形のラウドネスの変化と鳴り始め位置は、図 3 のようになる。図 3 において赤線が鳴り始め位置であり、この例では鳴り始め位置が二箇所存在する。よって図 2 のような音響波形は図 4 のように赤線で区切られる 3 区間に分割される。

ここでラウドネスは ITU-R BS.1770-1 に基づいて計算を行った。これは複数の音の相対的な音量を計算することができる国際基準である。単位は LKFS (Loudness, K weighted, relative to normal full scale) であり、0dBFS, 1kHz のサイン波が約 -3.01LKFS となる。

3.2 擬音語への変換

環境音を擬音語へと変換する。環境音の擬音語への変換手法はまだ確立されていないが、2.2 節で述べたように、石原らは MFCC を特徴量として隠れマルコフモデルを構築するという音声認識の手法を応用して

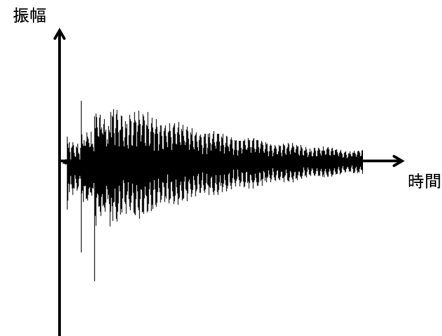


図 2: 音響波形の例

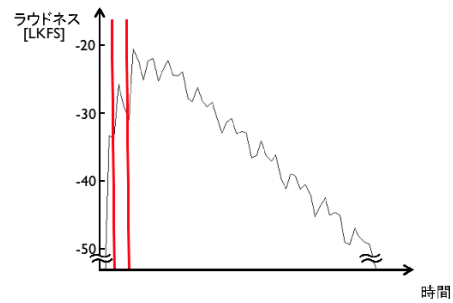


図 3: ラウドネスの変化と鳴り始めの位置

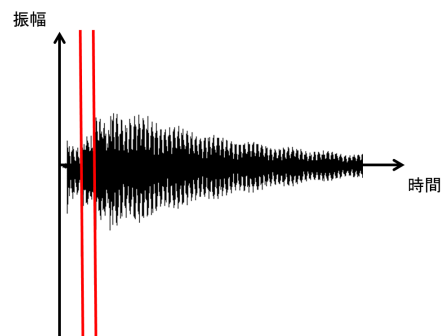


図 4: 音響波形の分割例

いる [9]。しかし、2.3 節で述べたように、環境音には音声の子音や母音にあたる波形が明確には存在しないため、音素単位に分割することが手作業でも非常に困難であり、音素の学習も困難である。

そこで本手法では、音響波形の音素単位への分割は行わず、スペクトル包絡と残響時間から擬音語へ変換する手法を提案する。また、本手法では、一音節の擬音語を語頭と語尾からなると定義し、別々に決定する。例えば、“チン”の語頭は“チ”であり、語尾は“ン”となる。

3.2.1 語頭の決定

本手法では、一音節の擬音語の語頭は環境音のスペクトル包絡から求められると仮定する。これは、2.4 節で述べたように、比屋根らの研究において、スペクトル構造を変化させた際に擬音語表現の子音が変化し、中心周波数で母音が変化した [13] ことから、人間が環境音を擬音語にする際にはスペクトル構造が関係していると仮定できるためである。

語頭の決定は、スペクトル包絡を素性としたサポートベクタマシンを用いて行う。語頭一つに対し、一つのサポートベクタマシンを構築する。今回は表 1 に示す 18 種類の語頭のサポートベクタマシンを構築した。

表 1: 語頭の種類

ヴィ, カ, ガ, ギ, コ, ゴ, シュ, ズ, チ,
チャ, ギ, ド, バ, パ, ピ, ブ, ポ

分割された音響波形中で最もラウドネスが大きな区間のスペクトル包絡を用いて、擬音語の語頭推定を行い、その結果をその音響波形の擬音語認識の語頭とする。

擬音語の中には、複数の音節からなるものも存在する。例えば、“チリン”や“カタッ”などが挙げられるが、このような擬音語は、同じような環境音が続き二音節目の語頭が変化している。比屋根らは、このような多重衝突音のスペクトルでは数 10 ~ 100ms 程度の時間間隔で同系統のスペクトルピークが並ぶ [13] としている。山内らも、20 ~ 90ms 周期の音の場合に [18] 二音節目の語頭が変化するとしている。そこで本手法では、100ms 以下の時間間隔で同一の語頭が現れた場合には語頭を変化させる。変化前の語頭と変化後の語頭を表 2 に示す。

3.2.2 語尾の決定

擬音語の語尾は環境音の残響時間から決定する。残響時間は、環境音の発生からラウドネスが 60LKFS 減衰するまでの時間と定義し、3.1 節で分割された区間において、ラウドネスのピークからそのピークと比較して 60LKFS 減衰するまでの時間を計算すること

表 2: 語頭の変化

変化前	変化後
カ	タ
ガ	タ
コ	ト
ゴ	ト
チ	リ
ヂ	リ
バ	タ
パ	タ
ピ	リ

で求める。区間の終わりまでに 60LKFS 減衰しない場合には、区間の終わりまでの時間とする。

比屋根らによれば、残響時間と語尾には表 3 のような極めて明瞭な関係が見られる [13]。そこで本手法では、残響時間から語尾を表 4 のように決定する。また、長音符の数は残響時間に比例して増える。

表 3: 比屋根らによる残響時間と語尾の関係

残響時間 (ms)	語尾表現
0 ~ 100	ッ
100 ~ 200	ン
300 ~	ー

表 4: 残響時間と語尾の関係

残響時間 (ms)	語尾表現
0 ~ 100	ッ
100 ~ 200	ン
200 ~	ー

ここで決定された語尾と 3.2.1 節で決定された語頭とを組み合わせ、一音節の擬音語を創る。組み合わせにより、50 種類以上の一音節の擬音語を創ることができ、さらに一音節の擬音語の組み合わせにより、複数音節の擬音語も創りだすことができる。

3.3 書体の決定

音の性質を決める基本的な要素には“大きさ”、“高さ”、“音色”がある。これらは音の三要素と呼ばれ、

聴覚的印象としての“音”が有する三つの側面のことである[19]．そのため、音の三要素を伝えることができれば、音を伝えることができる．しかし、人間が知覚する“音色”の性質は音の“大きさ”や“高さ”と比較すると、はるかに複雑なものであり、対応する物理量も一つではなく、周波数スペクトルや立ち上がり、減衰特性など様々なものが対応すると考えられている．そのため、“音色”を特徴量として数値化することは非常に難しい．本手法では、様々な物理量が関係する“音色”をそのまま考慮するのではなく、人間が音から感じる様々な印象を物理量と対応付けていくことで、最終的に“音色”を表現することを目指す．今回は“迫力と鋭さ”と“音の響き”を考慮する．本手法で用いる、人間の音感と音響特徴量と書体表現の関係を表5に示す．

表 5: 人間の音感と音響特徴量と書体表現の関係

人間の音感	音響特徴量	書体表現
音の大きさ	音圧レベル	文字の大きさ
音の高さ	周波数重心	文字の太さ 色の濃淡
迫力と鋭さ	音圧レベル 周波数	フォント 斜体
音の響き	反響音の有無	捨て仮名 ¹ の利用

人間が知覚する音の大きさは、“大きい–小さい”という尺度で表現できる一次元的な性質であり、音圧レベルと対応する．音の大きさは、音圧レベルと完全に比例しているわけではないが、等ラウドネス曲線を考慮することで、音圧レベルから人間の知覚する音の大きさを計算することが可能である．本手法では、人間が知覚する音の大きさを書体の大きさへ反映させて表現する．

人間が知覚する音の高さも、一般的には、“高い–低い”という心理的性質で表現できる一次元的な性質である．音の高さに対応する物理量は、純音の場合には周波数であり、周期的な複合音の場合には基本周波数となる．しかし、倍音成分などの影響で、人間が知覚する音の高さとは異なる場合もある．本手法では、人間の知覚する音の高さは、周波数スペクトルから周波数の重心を計算して求められた周波数重心を用いて、書体へと反映させる．野宮らの調査結果[14]から、“低い”イメージのある表現は“太字”、“高い”イメージのある表現は“細字”という相関があると言えるため、周波数重心が低い場合には太字表現を用いて表現する．また、長田らは、音高と色のマッピング実験から、音高が高い音ほど明度が高い色が対応付けられる[20]としているため、書体の色の濃淡にも反映させる．周波数重心が高くなるにつれて明るくし、低くな

るにつれて暗くする．図5に周波数重心の高い音の表示例を、図6に周波数重心の低い音の表示例を示す．

チーン

図 5: 周波数重心の高い音の表示例

ゴン

図 6: 周波数重心の低い音の表示例

栗山らは、表6に示すように、音圧レベルが大きく、周波数が低くなるほど“迫力のある”音色に、音圧レベルが小さく、周波数が高くなるほど“もの足りない”音色になり、音圧レベルが大きく、周波数が高くなるほど“鋭い”音色に、音圧レベルが小さく、周波数が低くなるほど、“鈍い”音色になるとしている[21]．また、野宮らはフォントとそのイメージには表7のような関係があると言っていることから、本手法では、表8のように書体表現を決定する．

表 6: 音圧レベルと周波数と音色を表す形容詞の関係

音圧レベル	周波数	形容詞
大きい	低い	迫力のある
小さい	高い	もの足りない
大きい	高い	鋭い
小さい	低い	鈍い

音の響きは、振幅包絡から求める．反響音が存在するときには音が響いていると仮定し、反響音の有無を調べることで音の響きを書体表現へと反映させる．音響波形の振幅の包絡に音の鳴り始め位置を含む包絡のピークよりも小さなピークが複数回検出された際に、反響音が存在するとする．反響音が存在した場合には、長音符を捨て仮名に変換することで音の響きを表現する．どの捨て仮名に変換するかは長音符の直前の母音によって決定される．反響音が存在し、その擬音語表現が“チーン”である場合には、“チン”と表現を変化させる．

以上のようにして音響特徴が書体表現に反映される．

3.4 可視化表現の表示

認識結果である可視化表現を表示する際に、横座標で環境音の時間情報を、縦座標で環境音の高さを表現する．可視化表現の表示例を図7に示す．

¹ アイウエオのような小書きの文字のこと

表 7: フォントとそのイメージ

音色を表す形容詞	書体表現
明朝体	堅い, 綺麗, 鋭い
ゴシック体	太い, 堅い, 遅い

表 8: 音圧レベルと周波数と書体表現の関係

音圧レベル	周波数	書体表現
大きい	低い	ゴシック体
小さい	高い	明朝体
大きい	高い	明朝体, 斜体
小さい	低い	ゴシック体

認識された環境音の可視化表現は右端から左に流れるように表示される。これにより、環境音の発生したタイミングや間隔を認識することができる。さらに、Y座標で音の高さを表現しており、楽譜のように、高い音は低い音に比べて上部に表示される。

また、ウェブカメラとマイクを用いて、録音された音を即時認識しつつ、ウェブカメラの映像を出力することも可能であり、図 8 のように、その状況を視覚的に認識できる。

以上のようにして認識結果を表示することで、さらに環境音の特徴を視覚的に認識できるようにしている。

4 実験方法

提案手法により、環境音が適切な擬音語に変換できていること、また、書体表現に環境音の特徴が反映されていることを確かめるために、ユーザによるアンケート評価とマッチングテストを行った。

4.1 ユーザによるアンケート評価

環境音認識システムを実際に使用してもらい、鳴らした音の特徴が出力結果に反映されていたかどうかをユーザに評価してもらった。評価してもらう項目を以下に示す。

- 適切な擬音語に変換されていたか
- 音の大きさが視覚的に認識できたか
- 音の高さが視覚的に認識できたか
- 音の雰囲気を感じることができたか

アンケート評価において、ユーザにはシステムが採用している環境音の音響特徴と書体表現の関係はあらかじめ教示していない。

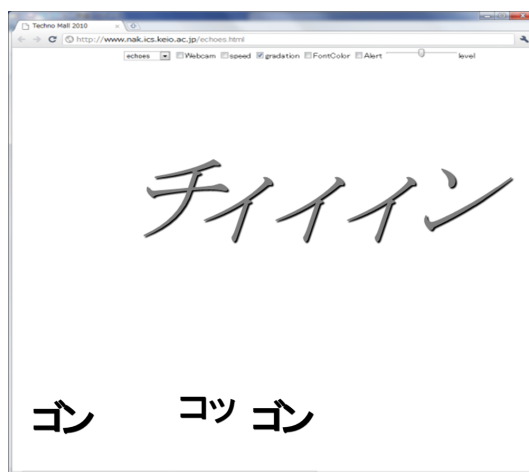


図 7: 可視化表現の表示例



図 8: ウェブカメラを用いた例

4.2 環境音とシステムの出力のマッチングテスト

被験者に、環境音とそれに対する環境音認識システムの出力結果の可視化表現との結びつけを行ってもらう。

実験のルールを以下に示す。

- 複数の音とそれに対するシステムの出力結果がランダムな順に並べられる
- 被験者は音を何度でも聴くことができる
- それぞれの音に対して必ずどれかの表現を選択する

マッチングテストにおいて、被験者には本システムが採用している環境音の音響特徴と書体表現の関係はあらかじめ教示していない。

本稿では擬音語表現の個人差に焦点を当てているわけではないため、今回の実験には“該当無し”とい

う選択肢を設けなかった．これは，例えば，“チーン”とも聞こえるが“キーン”という表現の方がより適切だという被験者が，“該当無し”を選択してしまうというような擬音語表現の違和感による判断を防ぐためである．

今回は 1 つの設問につき，4 つの音とそれに対するシステムの出力結果を表示し，結びつけを行ってもらった．この実験で，本システムが擬音語への変換や音響特徴の書体表現へ反映ができていないか調べるために，以下のような意図の設問を各設問につき複数作成した．

- 擬音語表現での判断を意図した設問（設問 A）
- 音の大きさでの判断を意図した設問（設問 B）
- 音の長さでの判断を意図した設問（設問 C）
- 音の響きでの判断を意図した設問（設問 D）

設問 A，B，C，D それぞれの例を図 9 と 10，11，12，13 に示す．

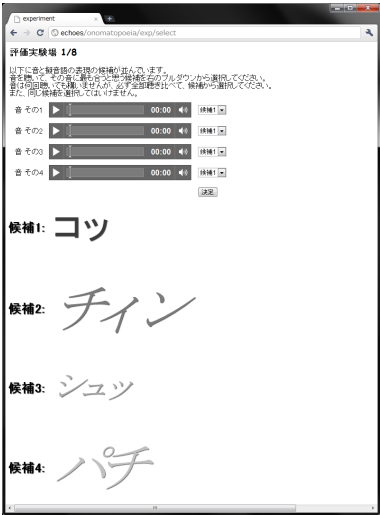


図 9: 実験の設問 A の例

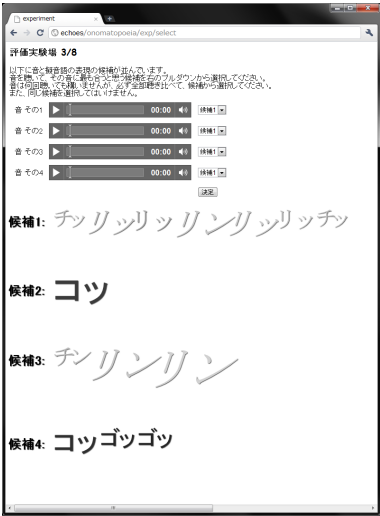


図 10: 実験の設問 A の例 2

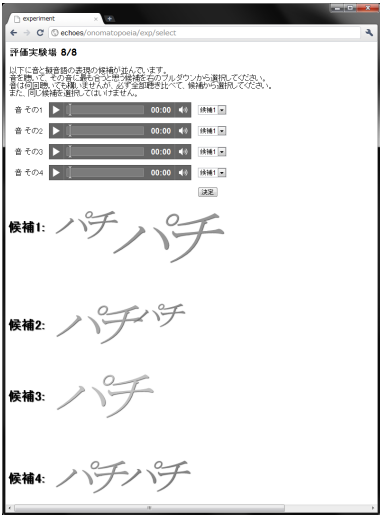


図 11: 実験の設問 B の例

5 実験結果

ユーザによるアンケート評価の集計結果を表 9 と表 10 に示す．被験者数は 28 名である．

環境音とシステムの出力のマッチングテストの結果を表 11 に示す．

6 考察

表 9 を見ると，ほぼ半数のユーザが“うまく認識されていた”と回答しており，“ほぼうまく認識されていた”を合わせると約 9 割に達する．本手法では，擬

表 9: 思ったような擬音語に認識されたか

選択肢	回答の割合
うまく認識されていた	46%
ほぼうまく認識されていた	43%
許容できる範囲の認識だった	7%
うまく認識されなかった	0%
無回答	4%

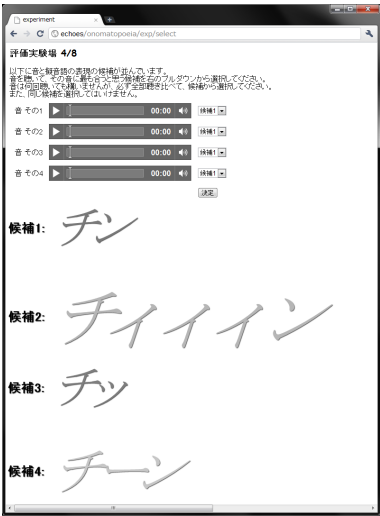


図 12: 実験の設問 C の例

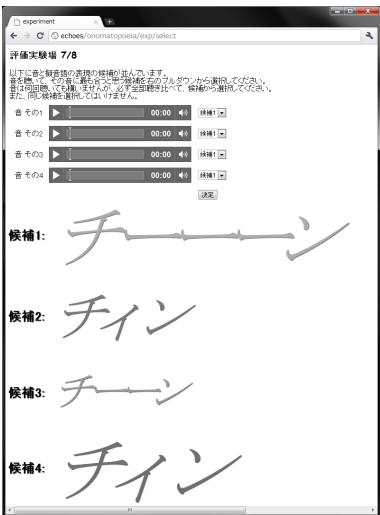


図 13: 実験の設問 D の例

表 10: 音の特徴が視覚的に認識できたか

	できた	できない	どちらでもない
大きさ	93%	0%	7%
高さ	57%	25%	18%
雰囲気	75%	7%	18%

表 11: マッチングテストの正答率

	正答率
設問 A	99%
設問 B	98%
設問 C	100%
設問 D	61%
全体	89%

音語表現の個人差は考慮していないため、擬音語表現に違和感を覚えるユーザも存在した可能性もあるが、この結果から、全く見当違いな擬音語表現になってしまったということはなかったと言える。

表 10 から、音の大きさについて、9 割以上のユーザが視覚的に“認識できた”と回答しており、本手法を用いることで“音の大きさ”を視覚化できていることが分かる。また、“音の雰囲気”についても、7 割以上のユーザが視覚的に“認識できた”と回答しており、“音の迫力と鋭さ”と“音の響き”しか考慮されていないにも関わらず、高い値となった。しかし、“音の高さ”については約半数のユーザしか“認識できた”と回答せず、25%のユーザは“認識できない”と回答した。本手法では、“音の高さ”を“文字の太さ”や“色の濃淡”、“表示位置”で表現したが、“文字の太さ”や“色の濃淡”は、視覚的に違いが認識しにくかったことが理由として考えられる。また、“表示位置”が環境音においては直感的ではないとも考えられる。しかし、今回は、単純に周波数重心を求め、“音の高さ”としており、人間の感じる“音の高さ”が適切に抽出されているのかという問題も挙げられ、一概に表現方法が適切ではなかったとは言えない。“音の高さ”については、より適切な特徴量やより効果的な表現方法を模索する必要があるが、本手法により、環境音の特徴を視覚的に認識できると言える。

表 11 を見ると、擬音語での判断を意図した設問（設問 A）では、99%の正答率となり、ほぼすべての被験者が環境音とシステムによる可視化表現を結びつけた。本手法では環境音の微妙な違いを擬音語の表現ではなく、書体表現に反映させることにしているため、今回の実験の設問 A は、擬音語表現がすべて異なる組み合わせになっており、比較的容易なマッチングになっていたとも言えるが、本手法により、人間が音と可視化表現を結びつけることが可能なレベルで環境音を擬音語に変換できていると言える。ただし、今回の実験では“該当無し”が選択肢になく、被験者が必ずしも違和感のない状況で結びつけを行えたかどうか疑問であり、今後、擬音語表現の個人差などを考慮していくためには、本手法の擬音語への変換で、どこまで違和感なく表現できるかを調査していく必要がある。

同様に、音の大きさでの判断を意図した設問（設問 B）では 98%、音の長さでの判断を意図した設問（設問 C）では 100%の正答率となり、本手法により、“音の大きさ”と“音の長さ”を書体表現に反映できてい

と言える。しかし、音の響きでの判断を意図した設問（設問 D）では、正答率が 61% となり、約半数の被験者がシステムによる可視化表現ではない選択肢を選択する結果となった。このような結果になった理由として、“音の響き” が、“音の大きさ” や “音の長さ” に比べると音としての違いが分かりにくいのではないかと我々は考えている。また、“音の響き” の感じ方や書体からのイメージにある個人差だけでなく、特徴量の抽出方法も要因として考えられ、より万人に共通する表現方法や特徴量抽出方法を検討していく必要がある。

7 結論

問題を単純化したとはいえ、環境音を擬音語へと変換し、さらに音響特徴を書体へ反映させることで、本来聴覚的に認識している環境音を、視覚的に認識することが可能になることが確かめられた。本手法を用いることで、特に“音の大きさ” や “音の長さ” を効果的に表現できたが、一方で、“音の高さ” や “音の響き” など、万人に対して効果的に表現できたとは言えない要素も存在した。

8 今後の展望

8.1 環境音認識の改善

“音の高さ” や “音の響き” については本手法で十分に表現できたとは言えないため、より適切な特徴量やより適切な表現方法を検討していく必要がある。また、音の三要素のひとつである“音色”は、そのすべてを考慮しているわけではないので、メル周波数ケプストラム係数などの特徴量を利用し、書体へと反映させていきたいと考えている。さらに、本稿では触れていないが、濁音が含まれる擬音語表現や大きな音や高い音は危険音と仮定し、文字の色を赤くするといった機能も実装しており、濃淡だけでなく色を用いた表現も検討していきたいと考えている。

また、本稿では、問題の単純化のために、音の発生状況やイメージ、表現の個人差は考慮していないが、これらを考慮することができれば、擬音語表現はより適切で違和感のないものにすることができる。しかし、そのためには音響情報だけでなく、様々な要素を考慮する必要がある、非常に難しいタスクとなる。

8.2 研究の応用

環境音を視覚的に認識できるという利点を活かした以下のような応用例が考えられる。

- 新しいコンテンツの創出
- 危険音通知支援
- 視覚的な環境音検索

環境音の言語的可視化により、音を文字として残すことが可能となる。例えば、写真を撮る際に、その場の音を認識し可視化することで、漫画のように音が可視化された写真を撮ることができる。また、デジタル放送に伴い、多くの番組において字幕を表示することが可能となったが、字幕だけでなく、可視化された音を効果的に画面上に配置することができれば、よりインパクトのある映像を楽しむことができるかもしれない。さらに、近年では、スマートフォンやタブレット端末などの携帯端末の流行により、拡張現実感を体験することのできる機会がますます増えている。3.4 節で述べたような身の回りの音を観るという体験も近い将来当たり前になるのではないだろうか。

また、危険な音を認識した際に、単に“危険音でした”と通知するよりは、“どのような音がしたのか”ということまで伝えることのできるシステムの方が、よりその場の状況を伝えることができると言える。このような危険音通知システムは、耳の不自由な人間に対してだけでなく、音楽を聴きながら携帯端末を操作するような、周りの状況に注意が向かなくなっている際にも有効だと考えられる。危険音や異常音に特化することにより、音の微妙な違いも表現できるようになる可能性もある。“方向” や “距離感” などの情報も提示することができれば、より良い支援となる。

さらに、擬音語による環境音検索から、より発展した視覚的な環境音検索も可能となる。人間は一度に複数の音を聴くことは得意ではないが、複数の文字列を認識することは可能なため、従来では、音をひとつずつ聴いて違いを聞き分けて目的の音を探していたが、音を可視化することができれば、目的の環境音を検索し易くできる。

このように様々な応用例が考えられるが、本稿のシステムが効果的に導入できるのは、その性質上、ディスプレイなどが伴っており、表示スペースが確保できる場合である。逆に、高い可読性が求められる場合や、ロボットの読み上げに用いる場合など、書体表現が必要なく、テキスト情報のみが必要とされる場合には、環境音の擬音語化でも十分だと言える。しかし、本稿の手法の表現方法を書体ではなく、人間の声の抑揚などに置き換えることができれば、読み上げの際に大きな効果を発揮すると考えられる。

また、近年、擬音語・擬態語は認知科学や心理学の分野においても様々なトピックで扱われるようになってきており、これらの分野の研究と連携を強めていくことで、表現の個人差や曖昧さの要因、さらには新たな方向性の研究を見出していきたいと考えている。

謝辞

本研究では、RWCP の実環境音声・音響データベースの非音声音ドライソース [22] を利用させていただいた。

参考文献

- [1] Sanae Wake and Toshiyuki Asahi. Sound Retrieval with Intuitive Verbal Descriptions. *IE-ICE transactions on information and systems*, No. 11, pp. 1568–1576, 2001.
- [2] 山口仲美. 暮らしの言葉 擬音・擬態語辞典. 講談社, 2003.
- [3] 田中基八郎, 松原謙一郎, 佐藤太一. 異音の表現における擬音語の検討: 衝突音等の単発音やうなり音の場合. 日本機械学会論文集 C 編, Vol. 61, No. 592, pp. 4730–4735, 1995.
- [4] 田中基八郎, 松原謙一郎, 佐藤太一. 機械の異常音の擬音語表現. 日本音響学会誌, Vol. 53, No. 6, pp. 472–482, 1997.
- [5] 高宮浩平, 岡本誠. 環境音の可視化によるろう者の環境認識. デザイン学研究. 研究発表大会概要集, No. 56, pp. 110–111, 2009.
- [6] 小林洋平, 小坂直敏. 楽曲製作のための音色理論の構築に向けて. 日本音響学会講演論文誌, pp. 3–7–15, 2005.
- [7] 石原一志, 坪田康, 奥乃博. 日本語の音節構造に着目した環境音の擬音語への変換. 電子情報通信学会技術研究報告, Vol. SP2003-38, pp. 19–24, 2003.
- [8] Naotoshi Osaka, Yoshinori Saito, Shinya Ishitsuka, and Yasuhiro Yoshio. AN ELECTRONIC TIMBRE DICTIONARY AND 3D TIMBRE DISPLAY. *Proceeding of the International Computer Music Conference*, pp. 9–12, 2009.
- [9] 石原一志, 駒谷和範, 尾形哲也, 奥乃博. 環境音を対象とした擬音語自動認識: 擬音語表現における音素決定曖昧性の解消. 人工知能学会論文誌, Vol. 20, pp. 229–236, 2005.
- [10] S.G. Mallat and Zhifeng Zhang. Matching pursuits with time-frequency dictionaries. In *IEEE Transactions on Signal Processing*, Vol. 41, pp. 3397–3415, 1993.
- [11] Selina Chu, Shrikanth Narayanan, and C.-C. Jay Kuo. ENVIRONMENTAL SOUND RECOGNITION USING MP-BASED FEATURES. In *Proc, ICASSP*, Vol. 17, pp. 1–4, 2008.
- [12] Nobuhide Yamakawa, Tetsuro Kitahara, Toru Takahashi, Kazunori Komatani, Tetsuya Ogata, and Hiroshi G. Okuno. Effects of modelling within- and between-frame temporal variations in powerspectra on non-verbal sound recognition. *Proceedings of International Conference on Spoken Language Processing (Interspeech 2010)*, pp. 2342–2345, 2010.
- [13] 比屋根一雄, 澤部直太, 飯尾淳. 単発音のスペクトル構造とその擬音語表現に関する検討. 電子情報通信学会技術研究報告, Vol. SP97-125, pp. 65–72, 1998.
- [14] 野宮謙吾, 渡辺静香. 和文書体の視覚イメージと音声イメージの関係. 日本デザイン学会研究発表大会, Vol. 50, No. 6, pp. 11–18, 2004.
- [15] 山本貴史, 松原正樹, 斎藤博昭. ECHOES: 環境音から擬音語への変換と効果的な表現方法の提案. エンタテインメントコンピューティング 2010, 2010.
- [16] Peter Ladefoged. *A Course In Phonetics*. Harcourt Brace College Publishers, 1993.
- [17] Kazushi Ishihara, Yasushi Tsubota, and Hiroshi G. Okuno. Automatic Transformation of Environmental Sounds into Sound-Imitation Words Based on Japanese Syllable Structure. *Proceeding of the Eighth European Conference on Speech Communication and Technology*, pp. 3185–3188, 2003.
- [18] 山内勝也, 岩宮眞一郎. 振幅変調音の擬音語表現とサイン音としての機能イメージ. 日本音響学会誌, Vol. 60, No. 7, pp. 358–367, 2004.
- [19] 岩宮眞一郎. 音色の感性学. 音色・音質の評価と創造. コロナ社, 2010.
- [20] 長田典子, 岩井大輔, 津田学, 和氣早苗, 井口征士. 音と色のノンバーバルマッピング—色聴保持者のマッピング抽出とその応用—. 電子情報通信学会論文誌, No. 11, pp. 1219–1230, 2003.
- [21] 栗山譲二, 二井 (岩宮) 眞一郎, 北村音一. 純音の音色の因子分析的研究. 日本音響学会講演論文集 (春季), pp. 657–658, 1979.
- [22] RWCP 実環境音声・音響データベース (RWCP-SSD). <http://tosa.mri.co.jp/sounddb/index.htm>.



山本 貴史

2009 年慶應義塾大学理工学部情報工学科卒業. 同大学院理工学研究科前期博士課程修了.



松原 正樹

2006 年慶應義塾大学大学院理工学研究科前期博士課程修了. 現在, 同後期博士課程在学中. 音楽情報処理, 認知科学に関する研究に従事. 情報処理学会, 日本音響学会, 日本認知科学会, 人工知能学会, 先端芸術創作学会各会員.



斎藤 博昭 (正会員)

慶應義塾大学工学部数理工学科卒業. 現在, 同大理工学部情報工学科准教授. 工学博士. 自然言語処理, 音声言語理解などに興味を持つ. 言語処理学会, 日本音響学会, 電子情報通信学会各会員.