

マルチステージ環境音識別法を用いた危険音の検出

林田 亘平^{1,a)} 小川 純平² 中山 雅人³ 西浦 敬信³ 山下 洋一³

概要：本研究では、我々の身の回りに存在する環境音から危機的状況を高精度に検知する方法について検討する。検出の対象となる非日常音の発生頻度が低い場合、従来の危険音検出では発生頻度の高い日常音から音響モデルを構築する。そして構築したモデルと観測信号との音響尤度と、あらかじめ設定しておいた閾値により危険を検出する。そのため、検出精度が閾値に依存する問題と、非日常音の中にも危険性の大小があるがそれを判断できない問題がある。これらの問題を解決するため、本研究では主観的危険度とマルチステージ環境音識別に基づく危険音検出法を提案する。提案法は、まず日常音のモデルに加えて複数の非日常音から構築したモデルを用いて非日常音を検出することで、閾値の問題を回避する。そして、検出された非日常音に対し主観的に決定した危険度毎に構築したモデルを用いて識別を行うことで、危険性の大小を識別する。実環境で収録した環境音を用いた評価実験の結果、提案法は従来法よりも非日常音の検出精度が約7%、危険度の識別精度が約7%向上したことを確認し、提案法の有効性を確認した。

1. はじめに

近年、日本では少子高齢化や核家族化が進行しており、高齢者が家族と離れた場所に一人で暮らす状況が増加している [1]。そのような一人暮らしの高齢者の危機的状況を早期に検出するために、家族から遠隔地に住む高齢者の生活を確認することが望まれている。そのためにビデオカメラを用いて高齢者の生活状況を記録し、記録した情報に対し画像処理を行うことで危機的状況を検出するシステムが研究されている [2], [3], [4], [5]。しかしながら、ビデオカメラにより観測した画像情報を用いるために、暗所や物体の重畳による死角、ビデオカメラの画角外において危険を検出できない問題がある。

この問題を解決するために、日常生活に存在するドアの開閉音などの環境音を利用した非日常音検出システムが研究されてきた [6], [7], [8], [9]。これらのシステムは、マイクロホンで観測した音響信号を用いて、環境音を識別することで危険を検出する。そのため、暗所や死角においても危機的状況を検出できる。従来の非日常音検出 [8] では、発生頻度の低い非日常音はモデルの学習に必要な音源数の取得が困難であるため、日常的に発生頻度の高い日常音から構築した音響モデルを用いて、観測信号と音響モデルから

計算した音響尤度の大きさから非日常音を検出する。従来法では日常音と非日常音の識別に音響尤度の大きさを用いるため、識別のための閾値の大きさにより非日常音の検出精度が変化する。また、最適な閾値の決定が困難という問題がある。さらに検出された非日常音の中にも、特に危険性の高い場合やそれほど危険性の高くない場合が存在するにもかかわらずそれを判断できないという問題があった。

そこで本研究ではこれらの問題を解決するために、主観的危険度とマルチステージ環境音識別に基づく危険音検出法を提案する。提案法は観測された信号に対し、異なる組み合わせの音響モデルを用いて識別を複数回行うマルチステージ環境音識別を行う。1度目の識別では、日常音のモデルに加えて複数の非日常音から構築したモデルを用いることで非日常音を高精度に検出する。発生頻度の低い非日常音については、非日常音として1つのモデルを構築することで、モデルの学習に必要な音源数を確保する。日常音と非日常音の両方の音響モデルを用いることで、従来法において決定の困難な閾値の問題を回避する。そして2度目の識別では、検出された非日常音に対して、主観的に決定した危険度毎に構築したモデルを用いて識別を行うことで、危険性の大小を識別する。2度目の識別によって、非日常音の中に存在する危険度の低い環境音や高い環境音を識別可能となる。

本研究では、まず2で環境音について述べる。3では従来の危険音検出法とその問題点について述べ、4では提案法について述べる。5では提案法の性能を評価するために実施した評価実験とその結果、及び考察について述べる。

¹ 立命館大学大学院 情報理工学研究科
〒525-8577 滋賀県草津市野路東 1-1-1
² 立命館大学大学院 理工学研究科
〒525-8577 滋賀県草津市野路東 1-1-1
³ 立命館大学 情報理工学部
〒525-8577 滋賀県草津市野路東 1-1-1
a) cm012063@ed.ritsumeji.ac.jp

最後に6において本研究の結論と今後の課題について述べる。

2. 環境音

我々が日常生活において耳にする音は、電車や車が通過する音、虫や動物の鳴き声、機械の駆動音、電化製品や携帯電話の着信音や扉の開閉音など様々である。我々の日常生活を取り巻くこれらの音は、環境音と定義される。

環境音はこれまで雑音として扱われ、例えば音声認識や発話区間検出などの分野においては外乱要因として抑圧すべき対象として考えられてきた。また、環境音そのものに関する研究についても、自動車や空調機などの運転音や特定種類の定常音の音質評価 [10], [11], [12] や、音響機器や楽器の音色の分析 [13], [14], [15] など、感性情報分野での研究が殆どであった。しかし、人間は音声や音楽からだけでなく、環境音からも多くの情報を認識している。例えば、人間は落下音により危険を察知する、あるいは足音から人の気配を感じる事が可能である。

そのため、計算機の演算能力の発展に伴い、計算機も人間同様に日常環境音から生活環境を把握することの必要性が提唱されており [16]、計算機による高精度な生活環境の把握を実現するため、環境音識別に関する研究が注目されている。例えば、田中らによる靴音の解析 [17] がある。この研究は靴音から誰の靴音か識別することを目指し、靴音の音響的特徴について解析している。また、織田らは聴覚障害者支援のために、12種類の環境音に対して、機械が反応し使用者にイベント内容を提示する環境音識別法を提案している [18]。

このように現在の環境音識別法は、限定された環境音の中から目的とする音源を特定する研究が主流である。これらの環境音識別法は、識別したい音源が明確な場合には有効な方法である。しかしながら、発生頻度の低い非日常音を検出するような場合には、モデルの学習に十分な音源の取得が困難という問題がある。また、日常生活において環境音は無数に存在するため、実世界に存在する環境音を全て個々にモデル化することは不可能である。以上の理由から、環境音を用いた危険音検出を行う場合には、環境音を適切に分類する必要がある。

3. 従来の音響尤度を用いた危険音検出法

図1に従来の音響尤度を用いた危険音検出法の処理の流れを示す。従来の音響尤度を用いた危険音検出 [8] ではまず、日常音の音イベントの種類ごとにHMM(Hidden Markov Model)[19]を用いて音響モデル M を構築する。そして、構築したモデルごとにViterbiアルゴリズム [20] を用いて音響尤度を計算し、式(1)により音響尤度が最大となるモデル R を求める。

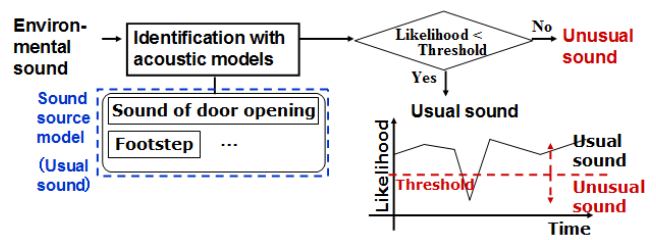


図1 音響尤度に基づく危険音検出の処理の流れ

Fig. 1 Processing flow for conventional abnormal sound detection with acoustic likelihood

$$R = \underset{M}{\operatorname{argmax}} P(\mathbf{U}, \mathbf{Q} | M). \quad (1)$$

ここで、 \mathbf{U} は観測信号の特徴ベクトル列を、 \mathbf{Q} は状態遷移確率 $P(\mathbf{U}, \mathbf{Q} | M)$ を最大にする経路を示す。音響尤度は識別対象の環境音がどの程度最尤なモデルと類似しているかを表すため、従来研究では最尤なモデル R の音響尤度が設定した閾値以上の場合は日常音、閾値未満の場合は危険性のある非日常音と識別することで危険を検出する。

従来の音響尤度に基づく危険音検出法は、モデル構築に必要な音源数の取得が容易な、発生頻度の高い日常音のみを対象に音響モデルを構築する。また、日常音と非日常音の識別は観測された信号と構築したモデルとの音響尤度を用いる。そのため、この方法による日常音の棄却精度と危険音の検出精度は、設定する閾値により大きく変化するが、この閾値の決定は一般的に困難という問題がある。また、非日常音の中にも危険性の大小があるがそれを判断できない問題がある。

4. 主観的危険度とマルチステージ環境音識別に基づく危険音検出法の提案

従来の音響尤度に基づく危険音検出法の問題点を解決するために、本研究では主観的危険度とマルチステージ環境音識別に基づく危険音検出法を提案する。提案法の処理の流れを図2に示す。提案法は観測された信号に対し、異なる組み合わせの音響モデルを用いて識別を複数回行うマル

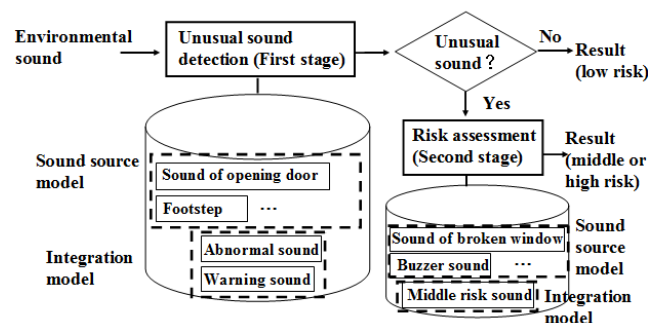


図2 主観的危険度とマルチステージ環境音識別に基づく危険音検出の処理の流れ

Fig. 2 Processing flow for proposed multi-stage abnormal sound detection based on subjective impression.

チステージ環境音識別を行う。

まず1度目の識別では、日常音から構築した音響モデルに加えて、数種類の非日常音から構築した音響モデルを用いることで、日常音と非日常音を識別する。発生頻度の低い非日常音については、非日常音として1つのモデルを構築することで、モデルの学習に必要な音源数を確保する。観測信号は、日常音か非日常音に識別されるため、提案法では従来法と異なり閾値の大きさに検出精度が依存する問題を回避できる。

そして2度目の識別では、検出された非日常音に対し、主観的危険度に基づき決定した危険度毎に構築した音響モデルを用いて識別を行うことで非日常音の危険性の大きさを識別する。本研究における主観的危険度については4.1で述べる。1度目の識別において検出された非日常音に対し、危険度毎に構築した音響モデルを用いた識別を行うことで、非日常音の危険性を識別可能となる。

提案法は観測された信号に対し、異なる組み合わせの音響モデルを用いて識別を複数回行うマルチステージ環境音識別を行うことで、非日常音を高精度に検出するとともにその危険性を識別することで、危険音の高精度な検出の実現が期待できる。

4.1 環境音の主観的危険度

従来、環境音は発生頻度の高い日常音と発生頻度の低い非日常音の2つのカテゴリに大別されて識別が行われてきた。しかしながら、発生頻度の低い非日常音の中にも危険性には差があるため、非日常音の中でも危険性の大きさを把握することが望ましい。そこで、本研究では環境音の主観的危険度を提案し、環境音を分類する。主観的危険度は、環境音を危険度小・中・大に分類する。環境音の主観的危険度の概要図を図3に示す。危険度小 (Low risk) の環境音とは、日常的に発生する危険性の低い環境音のことを示す。危険度中 (Middle risk) の環境音とは、日常的には発生しないが、危険性が低い環境音を示す。危険度大 (High risk) の環境音とは、日常的には発生せず、危険事象により発生する環境音を示す。この環境音の主観的危険度を用いることで高精度な危険音検出の実現を目指す。

5. 評価実験

提案する主観的危険度とマルチステージ環境音識別に基づく危険音検出法の危険音検出精度を評価した。ここでは2種類の実験により検出性能を評価した。まず、音響尤度に基づく従来法と提案法を用いて、環境音から非日常音を検出する実験により、非日常音の検出精度を評価した。そして、検出した非日常音の危険度を識別する実験により、提案法の危険度識別精度を評価した。

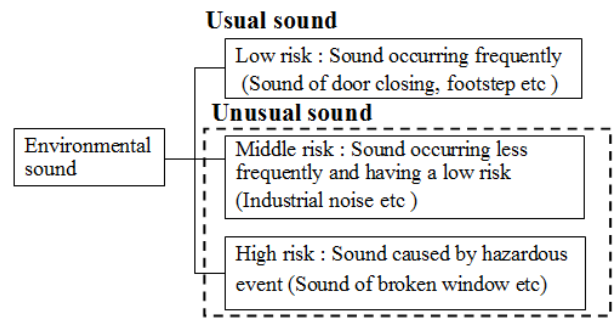


図3 環境音の主観的危険度

Fig. 3 Subjective risk of environmental sounds.

5.1 非日常音の検出精度に関する評価実験

ここでは、音響尤度に基づく従来法と提案法の非日常音検出精度を評価した。

5.1.1 実験条件

実験条件を表1に示す。なお、モデルの学習および評価には、技術研究組合新情報処理開発機構 (RWCP: Real World Computing Partnership) が作成した環境音データベース [21], [22] の環境音 10 種類と実環境において収録した環境音 20 種類を併用し、合計 30 種類を使用した。収録した環境音を表2に示す。30 種類の環境音の内、日常音が 10 種類、非日常音が 20 種類である。日常音の音響モデルの構築には、10 種類の日常音を 10 サンプルずつ使用した。また、非日常音の音響モデルの構築には、6 種類の非日常音を 10 サンプルずつ使用した。評価には、30 種類の環境音を各 5 サンプル、計 150 サンプルを用いた。

検出の精度は式 (2) により計算した日常音検出率と非日

表1 非日常音検出実験の条件

Table 1 Experimental conditions for unusual sound detection.

Sampling	48 kHz
Quantization	16 bit
Feature vector	33 orders (16 orders MFCC + 16 orders Δ MFCC + 1 order Δ Power)
Number of state	5 states
Training data	Conventional method : 100 samples (Usual sounds 100 samples) Proposed method : 160 samples (Usual sounds 100 samples + Unusual sounds 60 samples)
Test data	150 samples (Usual sounds 50 samples + Unusual sounds 100 samples)
Number of models	Conventional method : 10 models Proposed method : 12 models

常音検出率で評価した.

$$A = 100 \frac{1}{N} \sum_{i=1}^N G(i), \quad (2)$$

$$G(i) = \begin{cases} 1 & (\text{correct}) \\ 0 & (\text{incorrect}) \end{cases} \quad (3)$$

ここで, A は検出率, i は環境音のインデックス, N は環境音の総数を示す. また, $G(i)(i = 1, \dots, N)$ は危険音検出の正否により, 1か0の値を返す関数である.

以上の条件で, 音響尤度に基づく従来法と提案法の非日常音検出精度を評価した. ここで, 音響尤度に基づく従来法における非日常音検出のための閾値は, 非日常音の検出精度が100%となる値とした.

5.1.2 実験結果

図4に従来法の各環境音に対する音響尤度を, 従来法と提案法の日常音検出率と非日常音検出率を表3に示す. 実験結果より, 従来法の音響尤度を用いた検出法と比較して, 提案法は検出率が約7%向上したことを確認した. 以上の結果から提案法の非日常音検出における有効性を確認した.

5.2 非日常音の危険度識別に関する評価実験

ここでは, 非日常音検出実験において検出した非日常音を用いて, 従来法と提案法の危険度識別精度を評価した.

5.2.1 実験条件

実験条件を Table 4 に示す. 学習と識別には, 5.1 の評価実験で用いた 20 種類の非日常音を用いた. 従来法では,

表 2 収録した環境音

Table 2 Environmental sounds by recording.

音源番号	音源の内容
1	アラーム音
2	瓶を金属台に殴りつける音
3	紙を破る音
4	ビニール袋を擦り合わせる音
5	椅子の落下音
6	椅子を蹴る音
7	拍手の音
8	紙を丸める音
9	ドアを閉じる音
10	ブザー音
11	ドアをロックする音
12	火災報知音
13	足音
14	ガラスコップの割れる音
15	金属台の落ちる音
16	金属台と椅子の衝突音
17	電子レンジ音
18	緊急地震速報音
19	ガラスコップの衝突する音
20	ガラスコップを机に置く音

表 3 日常音と非日常音の検出率

Table 3 Detection rates for normal and abnormal sounds.

Conventional method		
	Identification rate	
Usual sound	35/50	(70 %)
Unusual sound	100/100	(100 %)
Total	135/150	(90 %)
Propose method		
	Identification rate	
Usual sound	45/50	(90 %)
Unusual sound	100/100	(100 %)
Total	145/150	(96.7 %)

20種類の非日常音の中から危険性の大きい環境音(危険度大の環境音)7種類を用いて音響モデルを構築し, 観測信号から計算した音響尤度に基づき危険度を識別する. また提案法では, 従来法で使用する音響モデルに加えて, 非日常音の中でも危険性の低い環境音(危険度中の環境音)6種類を統合したモデルを用いて危険度を識別する. 本研究における危険度の大小は, 一人暮らしの学生の生活環境音を収録した予備実験の結果に基づき決定し, 日常的に頻繁に発生する環境音を危険度小, 日常的に発生しない環境音で危険性が低い環境音を危険度中, 危険事象により発生する環境音を危険度大とした. 音響モデルの構築には, 1種類の環境音につき10サンプルずつを用いた. また評価には, 非日常音20種類を5サンプルずつ, 計100サンプルを使用した. 環境音100サンプルに対して, 音響尤度を用いた従来法と, 提案法をそれぞれ用いて危険度の識別を行った. ここで, 従来法の危険度識別のための閾値は, 学習環境音(危険度大の環境音)の危険度識別精度が100%となる値とした. 危険度の識別精度は式(2)により計算した危険度識別率により評価した.

5.2.2 実験結果

図5に従来法の各環境音に対する音響尤度を, 各方法の危険度識別率を表5に示す. 実験結果より, 音響尤度を用いた従来法と比較して, 提案法の識別率が約7%向上したことを確認した. 以上の結果から提案法の危険度識別にお

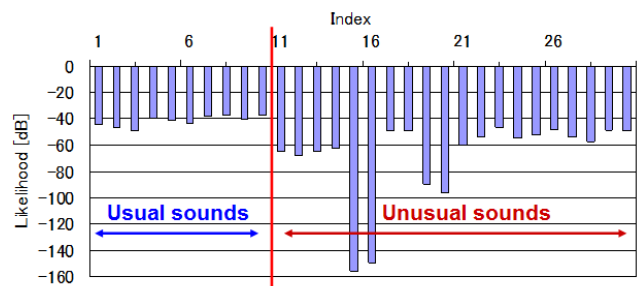


図 4 従来法における各環境音に対する音響尤度(1st ステージ)
Fig. 4 Acoustic likelihood for environmental sounds with conventional method in 1st stage.

表 4 危険度識別実験の条件

Table 4 Experimental conditions for risk identification.

Sampling	48 kHz
Quantization	16 bit
Feature vector	33 orders (16 orders MFCC + 16 orders Δ MFCC + 1 order Δ Power)
Number of state	5 states
Training data	Conventional method : 70 samples (High risk sounds 70 samples) Proposed method : 130 samples (High risk sounds 70 samples + Middle risk sounds 60 samples)
Test data	100 samples (Unusual sounds 100 samples)
Number of models	Conventional method : 7 models Proposed method : 9 models

表 6 マルチステージ環境音識別法の評価実験結果

Table 6 Total experimental results of multi-stage identification

Conventional method	
	Identification rate
First stage	35/50 100/100 (70 % 100 %)
Second stage	80/100 (80 %)
Total	115/150 (76.7 %)
Propose method	
	Identification rate
First stage	45/50 100/100 (90 % 100 %)
Second stage	87/100 (87 %)
Total	132/150 (88 %)

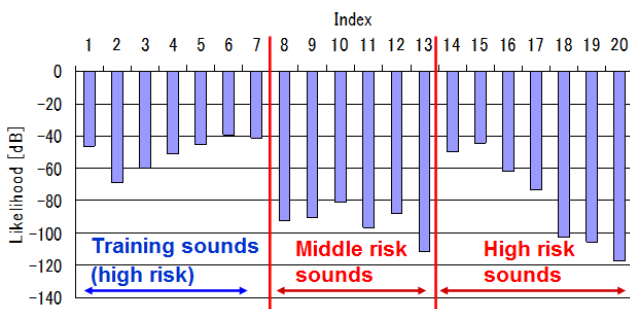


図 5 従来法における各環境音に対する音響尤度 (2nd ステージ)

Fig. 5 Acoustic likelihood for environmental sounds with conventional method in 2nd stage.

表 5 危険度識別率

Table 5 Experimental results of risk identification.

Conventional method	
	Identification rate
Total	80/100 (80 %)
Propose method	
	Identification rate
Total	87/100 (87 %)

ける有効性を確認した。

5.3 考察

表 3 より、音響尤度を用いた従来法と比較して、提案法では検出率が約 7% 向上していることが確認できる。この結果から、非日常音数種類を統合して構築した音響モデルを用いることで、適切に非日常音を検出できたとはいえる。これは、非日常音を統合することで、非日常音に共通する特徴が反映されたモデルを構築できたため、高精度に非日

常音を検出できたと考えられる。また、提案法では閾値を決定する必要がないため、非日常音の検出に適した方法であるといえる。

表 5 より、音響尤度を用いた従来法と比較して、提案法では危険度の識別率が約 7% 向上していることが確認できる。この結果から、非日常音の危険度識別においても危険度中を統合して構築した音響モデルを用いることで適切に危険度を決定できると考えられる。

表 6 に非日常音検出実験と危険度識別実験の結果を統合した結果を示す。非日常音の検出精度と危険度の識別精度のどちらにおいても、提案法は従来法よりも精度が向上していることを確認できる。以上の結果から、最初に非日常音を高精度に検出した後に、その危険度を識別する提案法は、環境音から危険を高精度に検出できるため、危険音検出に有効であるといえる。

本研究では、評価実験において従来と同様の音響特徴量を用いたために、危険音検出のために最適な音響特徴量については検討を行っていない。そのため、今後は危険音検出のための最適な音響特徴量の検討を行うことで、更なる検出精度向上を目指す必要がある。

6. おわりに

本研究では、我々の身の回りに存在する環境音を用いた危機的状況の高精度な検出の実現を目的として、観測した音響信号から発生頻度の低い非日常音を検出した後に、主観的危险度に基づき構築したモデルを用いて危険度を識別するマルチステージ環境音識別法を提案した。実環境で収録した環境音を用いた評価実験の結果、従来の音響尤度を用いた危険音検出法と比較して提案法は非日常音の検出精度が約 7%、非日常音の危険度識別精度が約 7% 向上したことを確認し、提案法の有効性を確認した。今後は、高齢者の危機的状況を高精度に検出するために、危険音検出に最適な音響特徴量を検討する。

謝辞 本研究の一部は科研費の助成を受けた。

参考文献

- [1] 立山龍彦, “高齢化社会の諸問題,” 東海大学出版会, 1991.
- [2] 関弘和, 堀洋一, “高齢者モニタリングのためのカメラ画像を用いた異常動作検出,” 電気学会論文誌 D, vol.122, no.2, pp.182–188, 2002.
- [3] 林健太郎, 羽下哲司, 関真規人, 笹川耕一, “映像監視における人物位置行動の検出技術,” 情報処理学会論文誌, vol.47, no.9, pp.12–20, 2006.
- [4] 磯和之, 山下勝也, 数藤恭子, 秦泉寺久美, 小島明, 仲澤斉, “遠隔映像モニタリングシステムを用いたオフィス状況監視,” 電子情報通信学会技術研究報告, vol.107, no.348, pp.37–42, 2007.
- [5] W. Yao-Dong, T. Takeshi and I. Idaku, “HFR-Video-Based machinery surveillance for high-speed periodic operations,” *Journal of System Design and Dynamics*, vol.5, no.6, pp.1310–1325, 2011.
- [6] 三木一浩, 西浦敬信, 中村哲, 鹿野清宏, “HMM による環境音識別の評価,” 日本音響学会春季研究発表会講演論文集, pp.15–16, 2000.
- [7] M. Cowling, “Comparison of techniques for environmental sound recognition,” *Pattern Recognit. Lett.*, vol. 24, no. 15, pp. 2895–2907, 2003.
- [8] K.-M. Kim, J.-W. Jung, S.-Y. Chun, and K.-S Park, “Acoustic intruder detection system for home security,” *IEEE Transactions on Consumer Electronics*, vol.51, no. 1, pp.130–138, 2005.
- [9] 河本満, 浅野太, 車谷浩一, “マイクロフォンアレイを用いた音環境の見守りによる非日常音と危険状態の検出システム,” 電子情報通信学会技術研究報告, vol.108, pp.19–26, 2008.
- [10] 橋本竹夫, “音質空間の次元とその定量化,” *Journal of the Japan Society for Simulation Technology*, vol.13, no.1, pp.25–33, 1994.
- [11] 横山一喜, 浅原康之, 戸井武司, “EV の音質評価と内燃機関車両との位置付け,” 環境工学総合シンポジウム講演論文集, vol.17, pp.111–114, 2007.
- [12] 難波精一郎, 桑野 園子, “音の評価のための心理学的測定法,” コロナ社, 1998.
- [13] 難波精一郎, “音色の定義を巡って,” 日本音響学会誌, vol.49, no.11, pp.823–831, 1993.
- [14] 石光俊介, 阪本浩二, 荒井貴行, 好美敏和, 菅原啓太郎, 佐々木勝弘, “音響機器評価に関する基礎検討,” 環境工学総合シンポジウム講演論文集, vol.17, pp.95–98, 2007.
- [15] 井上壮太, 降旗建治, “エレキギターの音響特性,” 電子情報通信学会技術研究報告, vol.104, no.379, pp.25–30, 2004.
- [16] 奥乃博, “コンピュータサイエンスから見た聴覚の情景分析,” 日本音響学会誌, vol.50, no.12, pp.1017–1022, 1994.
- [17] 田中基八郎, 巽敏寛, 千田剛, 藤野隆, 渡邊鉄也, 佐藤太一, 戸田富士夫, “靴音の解析,” 日本機械学会論文集 (C), vol.67, no.657, pp.1303–1308, 2001.
- [18] 織田修平, 青木真理子, 古家賢一, 片岡章俊, “雑音環境下における報知音伝達システムとその有効性,” 電子情報通信学会論文誌 D, vol.J90-D, no.10, pp.2765–2774, 2007.
- [19] R. J. Elliott, L. Aggoun, and J. B. Moore, “Hidden markov models: estimation and control,” Springer-Verlag, 1995.
- [20] 北研二, “言語と計算-4 確率的言語モデル,” 東京大学出版会, 1999.
- [21] 比屋根一雄, 中村哲, 浅野太, 遠藤隆, “RWCP 実環境音声・音響データベース,” 電子情報通信学会総合大会講演論文集, no.1, pp.255, 1998.
- [22] S. Nakamura, K. Hiyane, F. Asano, T. Nishiura, and T. Yamada, “Acoustical sound database in real environments for sound scene understanding and hands-free speech recognition,” *Proc. 2nd. International Conference on Language Resources and Evaluation (LREC2000)*, vol.2, pp.965–968, 2000.