

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

河本 満^{†*} 浅野 太^{††*} 車谷 浩一^{*}

†産業技術総合研究所情報技術研究部門 〒305-8568 茨城県つくば市梅園 1-1-1 中央第2

‡産業技術総合研究所情報技術研究部門 〒135-0064 東京都江東区青海 2-41-6

*科学技術振興機構 CREST

E-mail: {†m.kawamoto, ††f.asano, ‡k.kurumatani}@aist.go.jp

あらまし 本研究では、複数のマイクロフォンを構造化したマイクロフォンアレイを用いて音環境を見守り、非日常音を自動的に検出し、危険状態を自動通報する「非日常音検出システム」を提案する。このシステムでは、日頃の日常時における音環境の計測結果を基に、日常的でない音（非日常音）とその方向を自動的に検出することを目的とする。非日常音を検出し、その結果が日常時にない何か異常な情報だとみなされた場合、公共空間・ビルディングを監視している管理者・警備員等に送ることで、公共空間での安全・安心を提供しようと考えている。提案手法の特徴は、異常音検出の従来法とは異なり、非日常音や音源を予め定義することなく、日常の計測を通じて観測した音の特徴から、非日常音とその方向を自動的に学習・検出するところにある。このことから、従来法で行っているような非日常音のデータ集めや、そのデータを使った非日常音のモデル化などが必要ない。また、カメラを使う見守りシステムと比較すると、本システムではカメラの画角に計測範囲を制限されることがなく、音が検出できる範囲の見守りが可能である。また、画像の死角に音源が存在する場合でも危険情報を検出することができる。また、実環境での実験によって、本システムの基本的有効性を検証した。

キーワード 空間見守りシステム、非日常音、危険状態、音環境、マイクロフォンアレイ

A Security Monitoring System of Detecting Unusual Sounds and Hazardous Situations by Sound Environment Measurement Using Microphone Arrays

Mitsuru KAWAMOTO^{†*} Futoshi ASANO^{††*} and Koichi KURUMATANI^{‡*}

† ITRI, AIST Central 2, 1-1-1 Umezono, Tsukuba, Ibaraki 305-8568 Japan

‡ ITRI, AIST 2-41-6 Aomi, Koto, Tokyo, 135-0064 Japan

*CREST, JST.

E-mail: {†m.kawamoto, ††f.asano, ‡k.kurumatani}@aist.go.jp

Abstract We propose an “unusual sound detection system” which can automatically detect unusual sounds and send hazardous situations to guards, monitoring sound environments with microphone arrays which have a lot of microphones. The objective of the system is to automatically detect unusual sounds and their directions, based on usual measurement results. If the detected unusual sound is regarded as abnormal information, our proposed system provides security and safety by sending the information to guards who watching public spaces. Differently from the conventional systems, the attractive property of the system is that the unusual sound and its direction can be detected without defining them in advance. Therefore we do not have such troublesomeness of the conventional methods that the features of the predefined unusual sounds must be modeled and training data corresponding to them must be collected. Moreover, unlike conventional systems using video cameras, our system is not limited by video camera angles. Experimental results show the validity of the proposed system.

Keyword Security and safety monitoring system, Unusual sounds, Hazardous situations, Sound environment, Microphone arrays

1. はじめに

信号処理技術は、音環境を理解する上で、必要な技

術である。例えば、音源分離は、複数の音が混合している混合信号から、ある特定の信号を分離して取り出

すことができるので、発生した音がどのような音なのかを理解するための前処理として必要な技術である。また、音源定位は、音の発生する方向はもちろんのこと、音が発生したかどうかの規範としても利用することができる。

本稿では、このような信号処理技術を使って、日頃の音環境計測から、日頃聞きなれない音、つまり、非日常音を検出し、検出した音を公共空間の安全・安心の見守りのために利用する“非日常音検出システム”を提案する。ここでは、この非日常音が、日常的でない何か危険なことが起こったことのサインであると考え、この音を警備員に伝えることで、公共空間の安全安心を守ろうと考えている。

このような非日常信号、つまり、異常音によって、安全・安心を提供しようと考えている従来のシステムには、大きく分けて2種類ある。1つは、検出したいいくつかの異常音を予め定義しておいて、その定義した異常音のみを検出するシステム（例えば、[9][12]）、もう1つは、正常な音を予め定義しておいて、その正常音とのずれを計算し、その計算結果によって異常音を検出するというシステムである（例えば、[3][4]）。前者のタイプは、いくつかのガウス分布を使ってモデル化を行うガウス混合モデル（Gaussian Mixture Model, GMM）などを使って、定義した異常音の特徴をモデル化する必要があるため、異常音のデータを予め集めておかなければならない煩わしさがある。後者のタイプは、正常な音を把握しなければならず、また、その正常音を、トレーニングデータを使ってモデル化しなければならない。しかしながら、公共空間において正常音とは一体何かを決めるのは非常に難しい問題である。

一方、本提案システムは、従来法とは違って、検出したい非日常音は予め決めずに、日頃の計測を通じて、非日常音を検出することを目的としている。図1は、提案システムの処理流れ図を示している。まず、マイクロフォンアレイを使って、公共空間の環境音を計測する。計測した観測信号を用いて、音源の方向を推定する。音源の方向推定に関するアルゴリズムは、尤度関数を使う新しい方法を提案する。音源の方向推定を行う目的は、音がどちらの方向で発生したかを検出することと、公共空間で音が発生したか否かを判断するためでもある。次に、音源の方向推定により音が発生したと判断された音源からの音に対して、パワースペクトルを計算する。計算されたパワースペクトルのヒストグラムの変化の違いから音を分類する。分類された音は、分類されたグループ内の過去のすべての音に対して、類似度が計算される。このときの類似度の計算には、音声認識などで用いるケプストラムの1つである MFCC(Mel-Frequency Cepstrum Coefficient)と

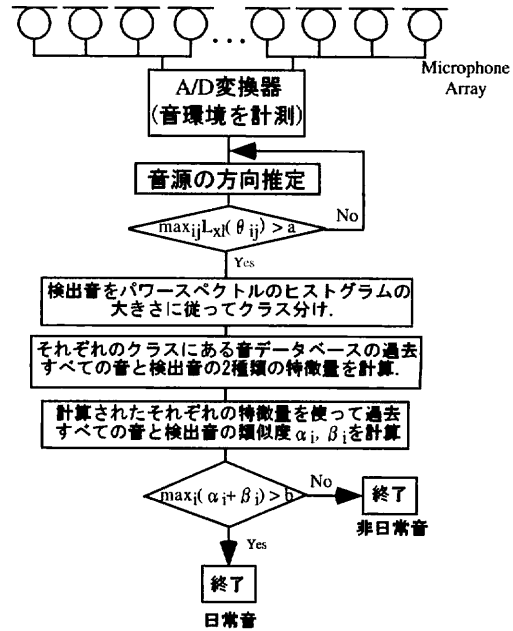


図1 非日常音検出の流れ図

Fig. 1 Block diagram of detecting unusual sounds

Fluctuation Pattern (FP)[8]の特徴量を用いる。類似度が高い音が過去に存在していれば、分類された音は、日常音と判断され、過去のどの音とも類似度が低ければ、非日常音と判断される。このようにして、非日常音を検出し、検出された非日常音は、公共空間を警備しているガードマンに送り、公共空間の見守りに利用される。

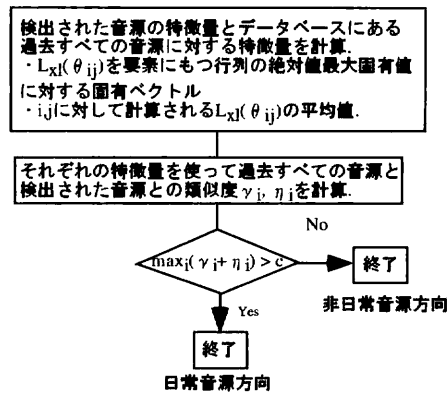


図2 非日常音源方向検出の流れ図

Fig. 2 Block diagram of detecting unusual sound direction

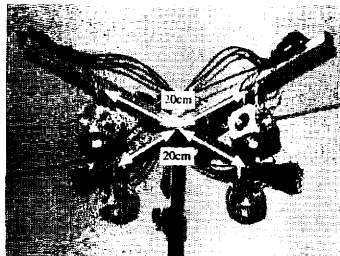
さらに、本提案システムは、音源の方向推定の際に計算される尤度関数の値を利用して、いつもは発生しないようなところからの音の方向情報、つまり、非日常の音の方向情報も、非日常音と似た類似度計算の考え方を使って提供することも可能である(図2を参照のこと。図2では、音源方向推定結果後に、“Yes”となった後の処理を示している)。

本提案システムの有効性は、簡単な実験とマイクロフォンアレイを産総研臨海副都心センターのロビーにおいて収録したデータを使った実験結果から検証する。

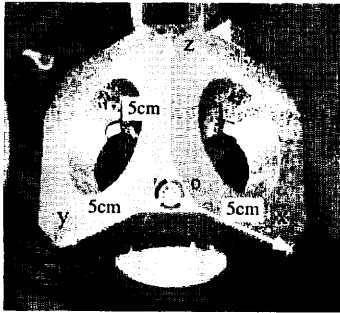
2. 提案手法

2.1. マイクロフォンアレイ

図3に示すマイクロフォンアレイを使って、公共空間の音環境を計測する。マイクロフォンアレイは、[7]に記述されているベクトルセンサを参考にして作られたので、ここでは、図3のマイクロフォンアレイのことをベクトルセンサと呼ぶことにする。図3(a)より、ベクトルセンサは、4個のキューブからなり、それぞれのキューブには4個のマイクロフォンが付けられている(図3(b))。従って、ベクトルセンサは、16個のマイクロフォンを持ったアレイとなっている。



(a)



(b)

図3 ベクトルセンサ

Fig. 3 Vector sensor

それぞれのマイクロフォンで観測した信号を $v_i^c(t)$ と書くことにしよう。ここで、下付き“i”は、 $i = 1, 2, 3, 4$

であり、それぞれ図3(b)に示される x, y, z のマイクロフォンで観測される信号を表わしている。また、上付きの $c = 1, 2, 3, 4$ は、それぞれの4つのキューブを区別する番号である。それぞれのキューブに対して、観測信号として使われる信号は、

$$y_i^c(t) = v_i^c(t) \quad (1)$$

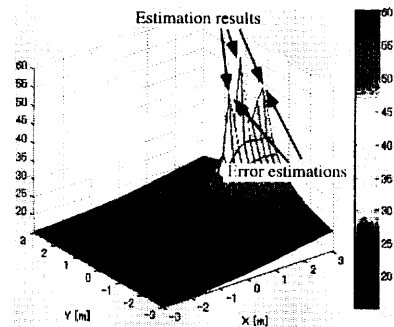
$$y_i^c(t) = v_i^c(t) - v_i^c(t), \quad i = 2, 3, 4 \quad (2)$$

である。つまり、図3(b)の o で観測される信号はそのまま用いるが、他の x, y, z で観測される信号はそのまま用いることはせずに、 o で観測される信号との差分を用いる。

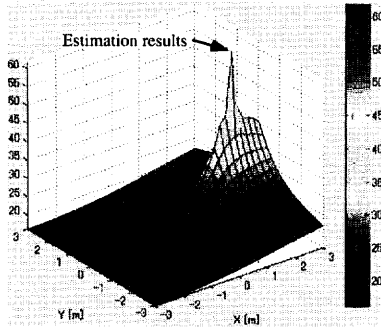
使用するベクトルセンサは、従来からあるような、例えば、円形のマイクロフォンアレイと異なる構造をもっている。この構造により、低周波信号に対する音源定位に関しては、円形のそれよりも推定誤差が現れにくいことが以下の説明から分かる。マイクロフォン i と位置 k 間の伝達特性は、周波数領域では以下の式で表わすことができる。

$$H_{ik}(\omega) = (1/\alpha_{ik}) \exp(-j\omega \tau_{ik}) \quad (3)$$

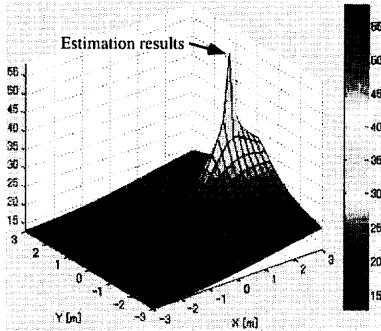
α_{ik} はマイクロフォン i と位置 k との距離を表わしている。 τ_{ik} は、 α_{ik}/C で計算される遅れを表わしている。ここで、 C は音の速度である。図4は、(2,0)の位置に音源があり、その音源から250Hzの正弦波が発生しているときの、音源定位の結果を表わしている。図の z 軸方向の値は、サブスペース法の1つである MUSIC[10]で計算される空間スペクトルの値である。図中、ピークが推定された音源位置を表わしている。円形のマイクロフォンアレイで得られた推定結果である図4(a)をみると、3つのピークがあり、そのうちの1つは、推定したい音源の位置であるが、残りの2つは、誤った推定位置である。これは、250Hzに対する式(3)の $H_{ik}(\omega)$ の値と誤って推定した位置における250Hz付近の伝達特性が似ているためである。つまり、音源からの信号が低周波信号の場合、円形のアレイでは、場所によって、似たような伝達特性を持つ場合があるので、誤った位置推定を引き起こすことになる。



(a) Circular microphone array



(b) Vector sensor (without differences)



(c) Vector sensor (with differences)

図 4 音源位置推定結果

Fig. 4 Estimation results of the sound localization

一方、本ベクトルセンサは、構造が特殊なため、円型のようなことは起こりにくくなっている。事実、図 4(b)を見ると図 4(a)にみられる誤った位置推定は見られず、ピークは 1 つとなっている。ただし、図 4(b)の結果は、差分をとらない観測信号を使った推定結果である。さらに、式(1)、(2)を使った推定結果をみると(図 4(c))、図 4(b)の結果よりもさらにピークが際立っていることがわかる。従って、本ベクトルセンサでは、音源からの音が低周波信号による音源位置の誤推定は起こりにくくなっていることがわかる。つまり、反響や残響は、主に低周波信号の影響が高周波信号のそれよりも強いと言われているので、本ベクトルセンサは、反響や残響の影響に鈍感に音源定位を行えることが期待できる。

2.2. 音源方向推定アルゴリズム

音源の方向推定は、周波数領域で行われる。観測信号は、短時間フーリエ変換によって周波数領域に変換される。このとき、周波数領域における観測信号ベクトルを $\mathbf{y}(t, \omega) = [Y_1^1(t, \omega), Y_2^1(t, \omega), \dots, Y_4^4(t, \omega)]^T$ と表わす。 $Y_i^c(t, \omega)$ は、キューブ c のマイクロフォン i で観測した信号の短時間フーリエ変換後の信号を表わしている。 T は転置を意味する記号である。以後、記述の簡

単化のため、 ω は消去して記述する。観測信号ベクトルと音源からの信号との関係は、周波数領域において、以下の式でモデル化できる。

$$\mathbf{y}(t) = \mathbf{A}s(t) + \mathbf{n}(t) \quad (4)$$

ここで、 \mathbf{A} は、 $\mathbf{A} = [\mathbf{a}_1, \dots, \mathbf{a}_L]$ と定義され、信号源とマイクロフォン間の伝達特性行列を表している。 $\mathbf{s}(t) = [S_1(t), \dots, S_L(t)]^T$ は、周波数領域における入力信号、 $\mathbf{n}(t) = [N_1(t), \dots, N_M(t)]^T$ は、周波数領域におけるノイズを表している。ノイズはガウスノイズを仮定している。 L は入力信号の数、 M はマイクロフォンの数を表している。ここで、 $\mathbf{s}(t)$ と $\mathbf{n}(t)$ の共分散行列を、それぞれ、

$$E[\mathbf{s}(t)\mathbf{s}^H(t)] = \mathbf{K}s = \text{diag}\{\gamma_1, \dots, \gamma_L\}$$

$$E[\mathbf{n}(t)\mathbf{n}^H(t)] = \sigma \mathbf{I}$$

と定義する。 $\text{diag}\{\dots\}$ は、 $\{\dots\}$ を要素に持つ対角行列である。 γ_l 、 $l = 1, 2, \dots, L$ は、 $S_l(t)$ のパワースペクトル、 σ はノイズ $\mathbf{n}(t)$ のパワー、 \mathbf{I} は単位行列、 \mathbf{H} は共役転置を表している。

ここで、出力ベクトル $\mathbf{y}(t)$ を以下のように変形する。

$$\mathbf{y}(t) = \sum_{l=1}^L \mathbf{x}_l(t) = \mathbf{H}\mathbf{x}(t) \quad (5)$$

$\mathbf{x}_l(t) = \mathbf{a}(\theta_{ij})S_l(t) + \mathbf{n}_l(t)$ 、 $\mathbf{x}(t) = [\mathbf{x}_1^T(t), \dots, \mathbf{x}_L^T(t)]^T$ は、 ML 次元のベクトル、 $\mathbf{H} = [\mathbf{I}, \dots, \mathbf{I}]$ は $M \times ML$ 行列、 $\mathbf{n}_l(t)$ は、 $\sum_{l=1}^L \mathbf{n}_l(t) = \mathbf{n}(t)$ を満足するノイズで、その共分散行列は、 $E[\mathbf{n}_l(t)\mathbf{n}_l^H(t)] = (\sigma/L)\mathbf{I}$ 。本稿では、 $\mathbf{x}_l(t)$ の共分散行列を使って音源位置方向推定を行う。しかしながら、その共分散行列 $\mathbf{C}_{x_l} = (1/N) \sum_{n=1}^N \mathbf{x}_l(n)\mathbf{x}_l^H(n)$ は直接計算ができない。そこで、以下の条件付期待値で推定することにする。

$$\mathbf{C}_{x_l} = E[\mathbf{C}_{x_l} | \mathbf{C}_y, \mathbf{K}'_y] = \mathbf{K}'_{x_l} - \mathbf{K}'_{x_l}(\mathbf{K}'_y)^{-1}\mathbf{K}'_{x_l} + \mathbf{K}'_{x_l}(\mathbf{K}'_y)^{-1}\mathbf{C}_y(\mathbf{K}'_y)^{-1}\mathbf{K}'_{x_l} \quad (6)$$

$$\mathbf{K}'_{x_l} = \gamma'_l \mathbf{a}_l \mathbf{a}_l^H + (\sigma/L)\mathbf{I} \quad (7)$$

$$\mathbf{K}'_y = \sum_{l=1}^L \mathbf{K}'_{x_l} \quad (8)$$

γ'_l は、 $(\mathbf{a}_l^H \mathbf{C}_y \mathbf{a}_l) / |\mathbf{a}_l|^4$ で推定できる。 \mathbf{C}_y は $\mathbf{y}(t)$ の共分散行列である。

音源位置方向を推定するために、 θ_{ij} に対して以下の関数を計算する。

$$L_{x_l}(\theta_{ij}) = \exp(-(1/2)\text{tr}[\mathbf{C}_{x_l}(\theta_{ij})\mathbf{K}'_{x_l}^{-1}(\theta_{ij})]) \quad (9)$$

θ_{ij} は、仮想的に決めた二次元平面の音源の位置を表している。つまり、二次元平面をある間隔でメッシュに区切り、その交点に音源があると仮定する。ここでは、 $i, j = 1, \dots, P$ とする。音源が存在するか否かは、式(9)を使って、すべての θ_{ij} に対して、 $L_{x_l}(\theta_{ij})$ を計算し、もしある閾値よりも大きな値を示す位置 θ'_{ij} に対する $L_{x_l}(\theta'_{ij})$ があれば、 θ'_{ij} の位置、または θ'_{ij} の方向に音源があると判断される。ただし、閾値は適当な値に設定する。

2.3. 音分類法

式(9)を使って検出された音は、非日常音か否かを判

別する前に、パワースペクトルのヒストグラムを使って、分類することで、非日常音の検出対象となる音を制限する。図5は、検出されたある音のパワースペクトルのヒストグラムを示している。分類をするときに調べる値は、極大と極小との差、つまり、図中の dm の値で、検出した音の中で、極小から極大の変化がある毎にその差を計算し、計算された値が、3種類

$$\{C1: dm > l_1\} \{C2: l_1 \geq dm > l_2\} \{C3: l_2 \geq dm > l_3\} \quad (10)$$

のどの範囲に当てはまるかを調べ、当てはまった数をカウントし、それぞれの範囲でカウントされた合計の数によって、3つのクラスに分類する(表1参照)。

表1 クラス分け
Table 1 Classification

	C1	C2	C3
Class 1	1	0	3 以下
Class 2a	2 以上	0 以上	0 以上
Class 2b	1	1	2 以下
Class 3	0	2 以下	3 以上

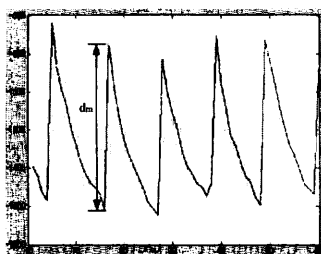


図5 音のパワースペクトルに対するヒストグラム
Fig. 5 A histogram of the power spectrums of a sound

表1から主に検出対象音は、突発的な音としている。しかしながら、3章で示される実験では、Class3には音声なども分類されるようである。すべての音を対象として非日常音を検出することもできるが、後述する類似度計算にかなりの時間がかかってしまうため、ここでは、類似度の計算時間の大幅削減を図るべく、予めクラス分けをして検出する音を制限する。

2.4. 非日常音の検出法

非日常音は、それぞれのクラス、Class1, Class2, Class3の中で検出される。それぞれのクラスで最初に検出された音は、まず、そのクラスを代表する音と考え、その後、それぞれのクラスに新たに入ってくる音に対して、過去の音と類似度を計算し、その類似度に従って、新たに入ってきた音は、非日常音か日常音かを決定する。

非日常音の検出法は、表1のクラスに新たに入った

音とそのクラスの過去のすべての音との類似度で判断する方法を提案する。ここでは、MFCCとFluctuation Pattern (FP)分析で得られた値を特徴量と考えている。

FP分析とは、周波数帯域毎の音の大きさを解析するもので、突発的な音の特徴量を調べるには必要な値である。新たに入った音と過去の音に対するMFCC, FP, それぞれの特徴量のすべてのペアに対して類似度を対称 Kullback-Leibler divergence で計算する。ここで、MFCCとFP, それぞれの特徴量に対する類似度の値を d_{MFCC} , d_{FP} と記述する。このとき、 $d_{MFCC+d_{FP}}$ の値がある閾値よりも大きな値となれば、そのペアは、似たような音であり、新たに入った音は、過去に似た音がある、つまり、日常音だと判断される。逆に、新たに入った音とすべての過去の音のペアに対する $d_{MFCC+d_{FP}}$ の値が閾値よりも小さければ、新たに入った音は、非日常音であると判断される。このようにして、非日常音は検出される。まとめると、

1. 検出された音の特徴量(MFCC, FP)を計算。同様に過去の音の特徴量も計算
2. 計算された検出音の特徴量と過去の音の特徴量すべてのペアに対して類似度を計算。
3. 類似度の値 $d_{MFCC+d_{FP}}$ によって、非日常音か日常音かを判断。

1において、特徴量の1つとして、MFCCを用いたが、最近の論文[14]で、突発音、特に足音に対して、MFCC以外で、線形予測係数(Linear Prediction Coefficient, LPC)ケプストラムも類似度を計算するときの特徴量として有効であることが報告されている。従って、今後の課題として、LPCケプストラムも特徴量として使い、MFCCとの比較を行い、類似度計算により適した特徴量を検討していく予定である。

2.5. 非日常音源方向の検出法

非日常音源方向の検出には、 $L_{x1}(\theta_{ij})$ を用いる。まず、 $L_{x1}(\theta_{ij})$, $i, j = 1, \dots, P$ を要素にもつ正方行列を考える。ここで、その正方行列を L_{x1} と書くことにする。この行列に固有値分解を適用し、絶対値最大固有値に対する固有ベクトルを計算する。この固有ベクトルの計算は、過去に検出されたすべての音の L_{x1} と新たに検出された音の L_{x1} に対して適用される。新たに検出された音の L_{x1} と過去のすべての音の L_{x1} の類似度を、計算された固有ベクトルを使い、ペア毎にユークリッド距離で表現する。求めた距離を d_{Ev} と記述する。さらに、以下の値を過去に検出されたすべての音と新たに検出された音に対して計算する。

$$L'_{x1} = (1/P^2) \sum_{i,j=1}^P \log L_{x1}(\theta_{ij})$$

この値に対しても、過去のすべての音と新たに検出された音とのペア毎に、類似度を計算する。計算された

値を d_{EL} とおく。このとき、非日常音の検出と同様に、 $d_{Ev} + d_{EL}$ がある閾値よりも大きければ、新たに検出された音源方向は、日常音源方向であり、どのペアの $d_{Ev} + d_{EL}$ もある閾値より小さければ、新たに検出された音源方向は、非日常音源方向であると判断される。このようにして、非日常音源方向を検出する。まとめると、

1. 新たに検出された音と過去のすべての音の L_{x1} を用いて、絶対値最大固有値に対する固有ベクトルを計算。
2. 固有ベクトルを特徴量とし、すべてのペアの類似度 d_{Ev} を計算。
3. 新たに検出された音と過去のすべての音の L'_{x1} を計算。
4. 計算された L'_{x1} を特徴量として、すべてのペアの類似度 d_{EL} を計算。
5. $d_{Ev} + d_{EL}$ を使って、非日常音源方向か日常音源方向かを判断。

3. 実験結果

3.1. 実験1

本節では、提案した非日常音検出法と非日常音源検出法の有効性を簡単な計測実験で得られた観測信号を用いて検証する。

ベクトルセンサは、部屋に置き、マイクロフォンの部分が天井近くになるように、三脚で固定し設置した(図6)。三脚から約1.5メートル離れたところで、物を落としたり、置いたりすることによって、音を発生させた。その音をベクトルセンサで計測し、計測した音を使って実験を行った。音は、4種類の音からなり、金属のプレートを落とした音(Sound A)、陶器のコップを置く音(Sound B)、雑誌を置く音(Sound C)、陶器のコップを置く音(Sound D、Sound Bとは異なった音)、である。4種類の音は、時間的に連続に発生させた(図7)。式(10)のパラメータ l_1, l_2, l_3 は、それぞれ、600、250、50とした。

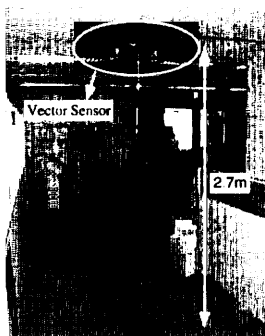


図6 ベクトルセンサ
Fig. 6 Vector sensor

式(9)、(10)によって、5つの音が検出された。ここで、検出された音は、すべて Class 1 に分類された。Sound B の音は、3つのクラスに分類されなかった。検出された音を Det1, Det2, ..., Det5 とし、 $d_{MFCC} + d_{FP}$ で類似度を計算した結果を表2に示す。左列に検出された音、右列に類似度の結果を示している。表2から同じグループの音が続けて発生した場合は、2番目の音は、1番目の音と類似していると判断されている。

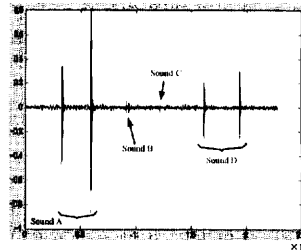


図7 実験1で使用した音
Fig. 7 Sounds used at Example 1

表2 類似度の計算結果(音)
Table 2 Similarity results (Sound)

Detected sounds	Similar sound
Det1 (Sound A)	First bang
Det2 (Sound A)	Det1 (Similarity:62%)
Det3 (Sound C)	Warning!!
Det4 (Sound D)	Warning!!
Det5 (Sound D)	Det4 (Similarity:61%)

ここで、Similarity のパーセントは、 d_{MFCC} の値を基に算出し、60%以上であれば、そのペアは似た音であるとした。「Warning!!」は、非日常音を表わしている。表から、異なったグループの音が初めて発生した場合には、Warning となっており、今までに発生していない音、つまり、非日常音の検出がうまくいっていることがわかる。

3.2. 実験2

図8は、Det1 から Det5 を検出したときの、音源の方向推定結果を表わしている。図中、 8×8 のそれぞれのマスに $L_{x1}(\theta_{ij})$, $i, j = 1, \dots, 8$ の値を表示している。濃い赤色の部分に音源がある可能性が高いことを表わしている。ここで、ベクトルセンサの位置は、図中の真ん中にあるとしている。1つのマスの一辺は、約50cmで、 $L_{x1}(\theta_{ij})$, $i, j = 1, \dots, 8$ の値で、大体4m四方の範囲内であれば、音源の位置がわかるようになっている。また、その範囲外で音が発生したとしても、その方向は大体わかるようになっている。

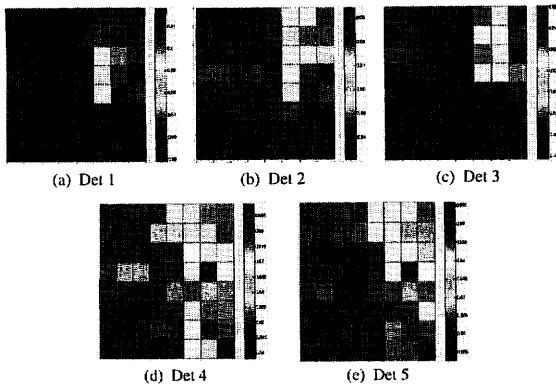


図 8 音源方向推定結果

Fig. 8 The estimation results of the sound source direction

図 8 の結果をもとに、非日常音源方向の検出法を試したところ、表 3 のような結果を得た。

表 3 類似度の計算結果 (音源)

Table 3 Similarity results (Sound position)

Detected sounds	Similar sound position
Det1	First bang
Det2	Det1 (Similarity: 72%)
Det3	Warning!!
Det4	Warning!!
Det5	Det4 (Attention!)

図 8 から、似たような位置に濃い赤色のマスがある音源方向推定結果が続くと、2 番目には、Similarity が表示されているのがわかる (Det2 の音源方向推定と Det1 の音源方向推定の関係)。また、過去に発生した音の場所以外の場所で初めて音が発生した場合、つまり、濃い赤色のマスが過去にない場所で発生した場合、Warning が表示され、非日常音源方向として検出されていることがわかる (Det3 と Det4 の音源方向推定結果)。Det5 の音源方向推定結果については、Det4 の音源方向と似ているのだが、ちょっと注意が必要であると判断されて、Attention! の表示が出力されている。Warning と Attention の違いは、類似度のパーセントによる。パーセントの大小関係は、Attention > Warning である。類似度のパーセントは、 d_{Ev} の値を基に計算される。表 2 の場合と同様で、60%以上の類似度があれば、そのペアは、日常音源方向と判断される。Warning は、非日常音源方向を表わして、Attention は、日常音源方向と判断してもよいが、ちょっと注意した方がよいということを表わしている。ここで、非日常音の検出の場合も同様の意味で Attention が出力される場合がある

ことに注意。図 8 から、非日常音源方向の提案検出法は、妥当な結果を与えていると考えることができる。

3.3. 実験 3

産総研臨海副都心センターのロビーにベクトルセンサを置き (図 9)、ロビーにおける音環境を計測し、その観測信号を使って、提案した非日常音の検出法を検証してみた。



図 9 ロビーに設置されたベクトルセンサ

Fig. 9 Vector sensor in a lobby

実験に使用したデータは、午前 9 時半くらいから午後 12 時半くらいまでに収録された環境音である。式 (10) による分類結果は、Class1 には主に突発的な音が、Class2 には、足音や音量が大きい音声など、Class3 には、音声やざわついた音などが、多く分類された。また、それぞれのクラスにおいて、類似度からグルーピングされた日常音には、類似した音が集まっていることを確認した。また、非日常音に関しては、過去に発生していないような音が Warning として、選ばれているようである。例えば、何かのブザーの音、子供の叫び声、奇声など。それらの音が危険な音として判断してよいかどうかについては、今後、詳しい検討が必要である。

4. 従来手法との比較

本章では、カメラを使う従来法との比較を行う。部屋 (W885×D588×H264(cm)) の天井に、CCTV カメラ (Panasonic WV-CF46) を設置したとしよう。この場合、画角は約 450cm×550cm となる (図 10)。従って、この部屋の見守りで、死角がないようにするためには、少なくとも 2 台のカメラが必要になる。一方、我々が提案したシステムでは、部屋の隅にベクトルセンサを置き、その対角線上の部屋の隅で床上 50cm くらいから金属のプレートを落とした場合、その落下した音は容易にベクトルセンサで観測できた。つまり、見守りを目的とした場合にカメラでは 2 台以上必要な部屋に対

して、本システムは、1台で見守りは可能になる。さらに、死角の多い場所をカメラで見守ろうとする場合、死角をなくするためにカメラの台数は必然的に増え、カメラの設置場所選びにも苦勞するであろう。本システムは、音が計測できれば、死角が多い場所であったとしても、1台でその空間の見守りはできるであろう。

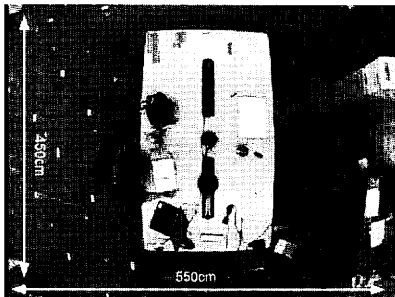


図 10 カメラの画角

Fig. 10 Range of camera

5. おわりに

本稿では、音環境をマイクロフォンアレイで計測し、その計測結果を利用して、公共空間を見守ることを目的としたシステムを提案した。提案システムは、日頃からの音環境計測を通じて、日頃発生しない音、聴きなれない音、つまり非日常音を検出することを目的としている。このことが、従来法にはない新規な点である。つまり、非日常音に関して、予め定義する必要がなく、非日常音のモデル化も事前に行う必要もなく、計測をしながら、日常音、非日常音が検出できる。

今後は、さらに実証実験を行うことによって、非日常音検出が、公共空間の見守りに有効かどうかの検証を行い、より便利に働く、見守りシステムの構築を目指す。また、マイクデバイスや電子回路のコンパクト化を通して、実空間での実利用を想定したデバイスの開発を進め、また非常時の緊急通報システムの自動起動などの研究を展開する予定である。

文 献

- [1] F. Asano and H. Asoh, "Sound Source Localization and Separation Based on the EM Algorithm," Proc. SAPA 2004.
- [2] R. Cai, L. Lu, H.-J. Zhang, L.-H. Cai, "Highlight Sound Effects Detection in Audio Stream," Proc. of ICME03, Vol. III, pp. 37--40, 2003.
- [3] Y.-K. Choi, K.-M. Kim, J.-W. Jung, S.-Y. Chun, and K.-S. Park, "Acoustic Intruder Detection System for Home Security," IEEE Trans. Consumer Electronics, Vol. 51, No. 1, pp. 130--138, Feb. 2005.
- [4] C. Clavel, T. Ehrette, and G. Richard, "Events Detection for An Audio-Based Surveillance System,"

Proc. of ICME2005, pp. 1306--1309, 2005..

- [5] A. Dufaux, L. Besacier, M. Ansoorge, F. Pellandini, "Automatic Sound Detection and Recognition for Noisy Environment," Proc. of EUSIPCO 2000, September 2000.
- [6] G. Guo and S. Z. Li, "Content-Based Audio Classification and Retrieval by Support Vector Machines," IEEE Trans. Neural Networks, Vol. 14, No. 1, pp. 209--215, 2003.
- [7] A. Nehorai and E. Paldi, "Acoustic Vector-Sensor Array Processing," IEEE Trans. Signal Processing, Vol. 42, No. 9, pp. 2481--2491, 1994.
- [8] E. Pampalk, "Computational Models of Music Similarity and their Application to Music Information Retrieval", Doctoral Thesis, Vienna University of Technology, Austria, March 2006.
- [9] A. R. A.-E.-Quran, R. A. Goubran, A. D. C. Chan, "Security Monitoring Using Microphone Arrays and Audio Classification," IEEE Trans. Instrumentation and Measurement, Vol. 55, No. 4, pp. 1025--1032, Aug. 2006.
- [10] R. O. Schmidt, "Multiple emitter location and signal parameter estimation," IEEE Trans. Antennas Propag., Vol. AP-34, No. 3, pp. 276-280, 1986.
- [11] A. Temko and C. Nadeu, "Classification of meeting-room acoustic events with support vector machines and variable-feature-set clustering," Proc. of ICASSP2005, Vol. V, pp. 505--508.
- [12] M. Vacher, D. Istrate, L. Besacier, J. F. Serignat, and E. Castelli, "Sound Detection and Classification for Medical Telesurvey," Proc. of the 2nd International Conf. on Biomedical Eng., pp. 395--399, 2004.
- [13] A. Itai, H. Yasukawa, "Footstep Classification Using Simple Speech Recognition Technique," Proc. of 2008 IEEE International Symposium on Circuits and Systems, pp. 3234--3237, 2008.