

マイクロフォンアレイを用いた音環境の見守りによる 非日常音の検出システム

河本 満^{*,***}・浅野 太^{*,***}・車谷浩一^{**,***}

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

Mitsuru KAWAMOTO^{*,***}, Futoshi ASANO^{*,***} and Koichi KURUMATANI^{**,***}

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.

Key Words: 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。たとえば、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。本論文では、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（たとえば、参考文献^{10),13)}）、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（たとえば、参考文献^{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な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。Fig.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、

* 産業技術総合研究所 つくば市梅園 1-1-1
** 産業技術総合研究所 東京都江東区青海 2-41-6
*** 科学技術振興機構 CREST
* Advanced Industrial Science and Technology, 1-1-1, Umezono, Tsukuba
** Advanced Industrial Science and Technology, 2-41-6 Aomi, Koto, Tokyo
*** CREST, JST
(Received September 17, 2008)

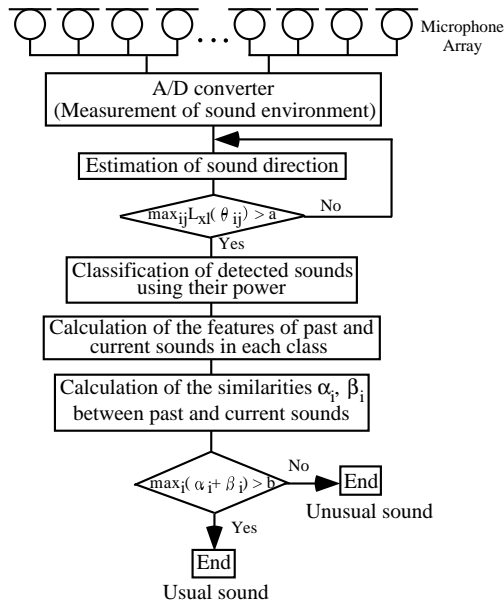


Fig. 1 Block diagram of detecting unusual sounds

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なう目的は、音がどちらの方向で発生したかを検出すること、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、(9)式を使ってパワーを計算する。計算されたパワー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 Fluctuation Pattern (FP)⁹⁾ 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

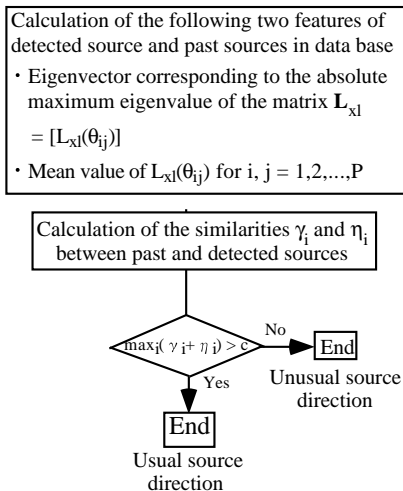
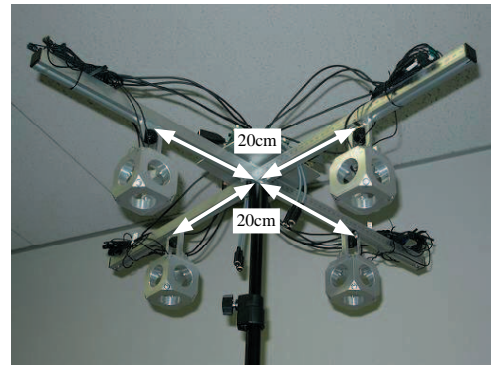


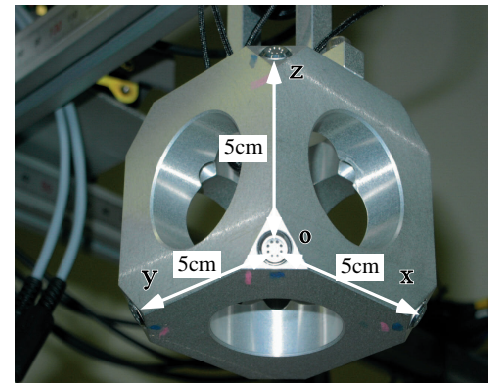
Fig. 2 Block diagram of detecting unusual source direction

いるガードマンに送り、公共空間の見守りに利用される。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得提供することが可能である (Fig.2を参照のこと。Fig.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。



(a)



(b)

Fig. 3 Microphone array

2. 提案手法

2.1 音計測：マイクロフォンアレイ

音計測には、マイクロフォンアレイを使用する。音源定位が可能になるマイクロフォンアレイであれば、どのようなものを使用しても構わないが、ここでは、Fig.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。Fig.3のマイクロフォンアレイは、Nehoraiら⁸⁾によって提案されたベクトルセンサを参考にして作ったものなので、ここでは、Fig.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。Fig.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 (Fig.3(b)). したがって, 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. それぞれのマイクロフォンで観測した信号を $v_i^c(t)$ と記述する. ここで, 下付き i は, $i = 1, 2, 3, 4$ であり, それぞれ Fig.3(b) に示される o, 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,

$$y_1^c(t) = v_1^c(t) \quad (1)$$

$$y_i^c(t) = v_i^c(t) - v_1^c(t), i = 2, 3, 4. \quad (2)$$

である. つまり, Fig.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. このマイクロフォンアレイの特徴としては, 差分をとることから, 音源定位を行なうときに, 従来の円型のマイクロフォンアレイで差分をとらない場合に比べて, 低周波信号に対して, 音源定位の誤推定を起こしにくいという特徴があることを確認している.

2.2 音源方向推定アルゴリズム

音源の方向推定は, 周波数領域で行なわれる. 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 $\mathbf{y}(t, \omega)$ = $[Y_1^c(t, \omega), Y_2^c(t, \omega), \dots, Y_4^c(t, \omega)]^T$ 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 $y_i^c(t)$ 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.

$$\mathbf{y}(t) = \mathbf{A}\mathbf{s}(t) + \mathbf{n}(t), \quad (3)$$

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わしている. $\mathbf{s}(t) = [S_1(t), \dots, S_L(t)]^T$ 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わしている. 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わしている. $\mathbf{s}(t)$ と $\mathbf{n}(t)$ の共分散行列を, それぞれ, $E[\mathbf{s}(t)\mathbf{s}^H(t)] = \mathbf{K}_s = \text{diag}\{\gamma_1, \dots, \gamma_L\}$, $E[\mathbf{n}(t)\mathbf{n}^H(t)] = \sigma^2\mathbf{I}$ と定義する. $\text{diag}\{\dots\}$ は, $\{\dots\}$ を要素を持つ対角行列である. γ_l ($l = 1, 2, \dots, L$) は, $S_l(t)$ ($l = 1, 2, \dots, L$) 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, H は共役転置を表わしている.

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

$$\mathbf{y}(t) = \sum_{i=1}^L \mathbf{x}_i(t) = \mathbf{H}\mathbf{x}(t), \quad (4)$$

$\mathbf{x}_i(t) = \mathbf{a}(\theta_i)S_i(t) + \mathbf{n}_i(t)$, $\mathbf{x}(t) = [\mathbf{x}_1^T(t), \dots, \mathbf{x}_L^T(t)]^T$ は, ML 次元のベクトル, $\mathbf{H} = [\mathbf{I}, \dots, \mathbf{I}]$ は, $M \times ML$ 行列, $\mathbf{n}_i(t)$ は, $\sum_{i=1}^L \mathbf{n}_i(t) = \mathbf{n}(t)$ を満足するノイズで, その共分散行列は, $E[\mathbf{n}_i(t)\mathbf{n}_i^H(t)] = \frac{\sigma^2}{L}\mathbf{I}$. 本論文では, $\mathbf{x}_i(t)$ の共

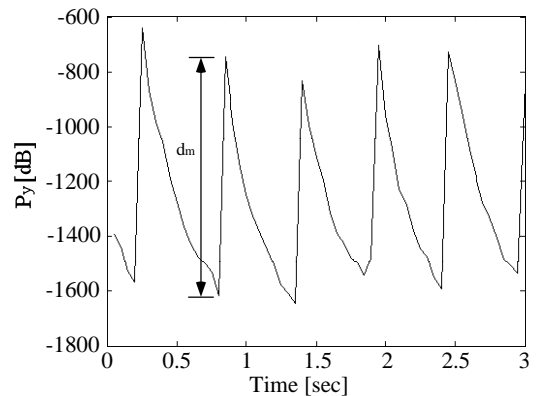


Fig. 4 A histogram of the power spectrums of a sound

分散行列を使って音源位置方向推定を行なう. しかしながら, その共分散行列 $\mathbf{C}_{x_i} = \frac{1}{N} \sum_{n=1}^N \mathbf{x}_i(n)\mathbf{x}_i^H(n)$ 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.

$$\mathbf{C}_{x_i} := E[\mathbf{C}_{x_i} | \mathbf{C}_y; \hat{\mathbf{K}}_y] = \hat{\mathbf{K}}_{x_i} - \hat{\mathbf{K}}_{x_i}(\hat{\mathbf{K}}_y)^{-1}\hat{\mathbf{K}}_{x_i} + \hat{\mathbf{K}}_{x_i}(\hat{\mathbf{K}}_y)^{-1}\mathbf{C}_y(\hat{\mathbf{K}}_y)^{-1}\hat{\mathbf{K}}_{x_i}, \quad (5)$$

$$\hat{\mathbf{K}}_{x_i} = \hat{\gamma}_i \mathbf{a}_i \mathbf{a}_i^H + \frac{\sigma}{L} \mathbf{I}, \quad (6)$$

$$\hat{\mathbf{K}}_y = \sum_{l=1}^L \hat{\mathbf{K}}_{x_l}, \quad (7)$$

$\hat{\gamma}_i$ は, $(\mathbf{a}_i^H \mathbf{C}_y \mathbf{a}_i) / |\mathbf{a}_i|^4$ 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.

$$L_{x_i}(\theta_{ij}) = \exp\left(-\frac{1}{2} \text{tr}\left[\mathbf{C}_{x_i}(\theta_{ij})\hat{\mathbf{K}}_{x_i}^{-1}(\theta_{ij})\right]\right). \quad (8)$$

θ_{ij} は, 仮想的に決めた二次元平面の音源の位置を表わしている. 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, (8) 式を使って, すべての θ_{ij} に対して, $L_{x_i}(\theta_{ij})$ を計算し, もしある閾値よりも大きな値を示す位置 $\hat{\theta}_{ij}$ に対する $L_{x_i}(\hat{\theta}_{ij})$ があれば, $\hat{\theta}_{ij}$ の位置, または, $\hat{\theta}_{ij}$ 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

(8) 式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, 以下の式を使って得られる音のパワーを使って分類する.

$$P_y(t) = \sum_{\omega} 10 \log_{10} |Y_1^c(t, \omega)|^2. \quad (9)$$

ここで, $|\cdot|^2$ は, \cdot の 2 乗ノルムを表わしている. また, Y_1^c は, どのキューブでもよいが, 差分をとらない信号を使用することを意味している. この分類により, 非日常音の検出対象となる音を制限する. 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かる時間が削減でき, 効率よく非日常音を検出できるようにするために, ここでは, 検出された音をクラス分けし, クラスに属し

た音のみ非日常音検出器にかける方法をとる(ここで示すクラス分けの方法は一例であり, 他にも多くの方法が考えられることに注意すべきである). **Fig.4** は, ある検出された音に対する $P_y(t)$ の値を時間変化と共に図示したものである. 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, ここでは3つのクラスに分類する (**Table 1** 参照, ただし, **Table 1** に示す分類法は一例であり, 他のカウント数で, いくつのクラスに分類するかは任意であることに注意すべきである. **Table 1** のクラス分けは, 主に突発的な音の検出を意識して作成したものである). ここで, l_1, l_2, l_3 は, 適当な正の値である. **Table 1** に示すクラス分けは, 1 秒間隔毎に観

Table 1 Classification

	C1	C2	C3
Class1	1	0	3 以下
Class2a	2 以上	0 以上	0 以上
Class2b	1	1	2 以下
Class3	0	2 以下	3 以下

測される音を使って実行される. **Table 1** のクラス分けで分類された音の例に関しては, 3.3 節で言及する. なお, すべての音を対象として非日常音の検出を行いたい場合は, クラス分けをせずに, すべての音に対して後述する類似度計算を行えばよい.

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, その類似度にしたがって, 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, **Table 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 divergence⁹⁾で計算する. ここで, MFCC と FP, それぞれの特徴量に対して求めた類似度を d_{MFCC} , d_{FP} と記述する. さらに, 求めた類似度は, それぞれの値の桁を大体そろえるために, 以下の変換を行なう.

$$\hat{d}_{MFCC} = \exp\left(-\frac{d_{MFCC}}{\alpha_1}\right), \hat{d}_{FP} = \exp\left(-\frac{d_{FP}}{\alpha_2}\right) \quad (11)$$

このとき, $\hat{d}_{MFCC} + \hat{d}_{FP}$ 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 $\hat{d}_{MFCC} + \hat{d}_{FP}$ 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 $\hat{d}_{MFCC} + \hat{d}_{FP}$ 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. したがって, 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 との比較を行ない, 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, $L_{xl}(\theta_{ij})$ を用いる. まず, $L_{xl}(\theta_{ij})$, $i, j = 1, \dots, P$ を要素にもつ正方行列を考える. ここで, その正方行列を \mathbf{L}_{xl} 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 \mathbf{L}_{xl} と新たに検出された音の \mathbf{L}_{xl} に対して適用される. 新たに検出された音の \mathbf{L}_{xl} と過去のすべての音の \mathbf{L}_{xl} の類似度を, 計算された固有ベクトルを使い, ペア毎に下式で求める.

$$d_{Ev} = |\mathbf{v}_c - \mathbf{v}_p|^2 \quad (12)$$

ここで, \mathbf{v}_c は, 新たに検出された \mathbf{L}_{xl} の固有ベクトル, \mathbf{v}_p は, 過去に検出された \mathbf{L}_{xl} の固有ベクトルを表わしている. 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.

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

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, 以下の式を使って類似度を計算する.

$$d_{EL} = |\hat{L}_c - \hat{L}_p| \quad (13)$$

ここで, \hat{L}_c, \hat{L}_p は, それぞれ新たに検出された音と過去に検出された音の \hat{L}_{xl} を表わしている. このとき, 非日常音検出の場合と同様の理由で, 以下の変換を行なう.

$$\hat{d}_{Ev} = \exp(-\alpha_3 d_{Ev}), \hat{d}_{EL} = \exp(-\alpha_4 d_{EL}) \quad (14)$$

したがって, $\hat{d}_{Ev} + \hat{d}_{EL}$ 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 $\hat{d}_{Ev} + \hat{d}_{EL}$ 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_{xi} 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 \hat{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 \hat{L}_{xi} を計算.
4. 計算された \hat{L}_{xi} を特徴量として、すべてのペアの類似度 \hat{d}_{EL} を計算.
5. $\hat{d}_{Ev} + \hat{d}_{EL}$ 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。

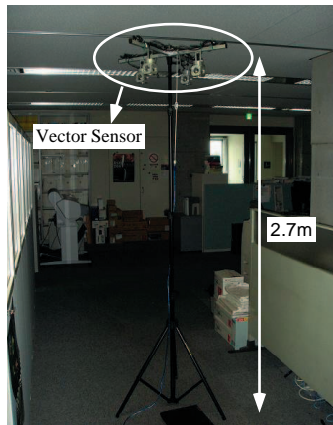


Fig. 5 Vector sensor

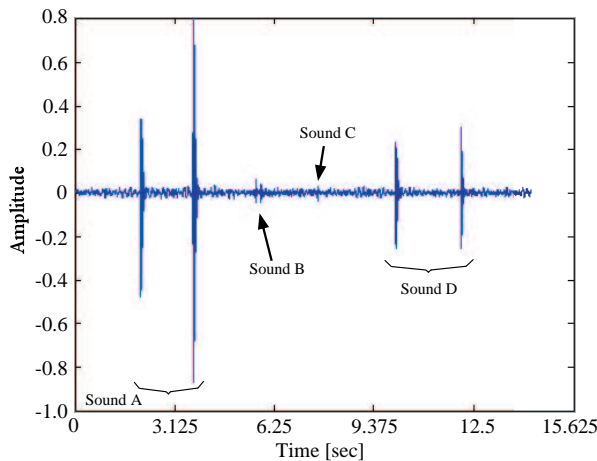


Fig. 6 One of the sounds observed by the vector sensor

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 (Fig.5). 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なった。音は、4種類の音からなり、金属のプレートを下した音 (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とは異なる音), である。4種類の音は、時間的に連続に発生させた (Fig.6). (10) 式のパラメータ l_1, l_2, l_3 は、それぞれ、600, 250, 50とした。Fig.1とFig.2における閾値 a, b, c は、テストデータ (Fig.7) を使い、(8) 式の $L_{xi}(\theta_{ij})$ の値や類似度の値を基にして、 $a = 0.4, b = 1.5, c = 1.5$ とした。ただし、(11) 式の α_1, α_2 は、それぞれ、70, 1000, (14) 式の α_3, α_4 は、それぞれ、50000, 30とした。このとき、 \hat{d}_{MFCC} と \hat{d}_{Ev} の値から、それぞれの値が大体 0.55 より大きければ、似た音と判断できることをテストデータを使った実験で確かめたので、0.55 が類似度 60% となる換算を行ない、後述する表において、Similarity のパーセントが 60% 以上であれば、日常音と判断できるようにした。また、 $0.55 \geq \hat{d}_{MFCC}, \hat{d}_{Ev} > 0.25$ のときは、Attention!、 $0.25 \geq \hat{d}_{MFCC}, \hat{d}_{Ev}$ のときは、Warning!! を表示するようにした。これは、Attention! で、日常音、日常音源方向と判断できそうだが、ちょっと注意が必要だということを表わし、Warning!! で、非日常音、非日常音源方向であると判断できそうだとことを表わそうとしたものである。つまり、類似度に応じて、非日常音、非日常音源方向に関する注意度の度合を言葉で表現しようと試みたものである。ここで、Fig.7 のテストデータの Sound E, Sound F, Sound G は、Fig.6 のそれぞれの Sound とは異なる音で、Sound E は人の声、Sound F は物を叩く音、Sound G は物を落とす音である。

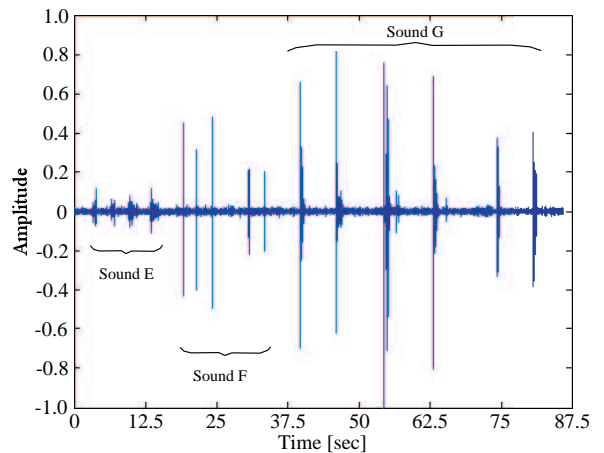


Fig. 7 One of the sounds observed by the vector sensor

(8), (9), (10) 式によって、5つの音が検出された。ここで、検出された音は、すべて Class1 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、 $\hat{d}_{MFCC} + \hat{d}_{FP}$ で類似度を計算した結果を **Table 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。Table 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。

Table 2 Similarity results (Sound)

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

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。

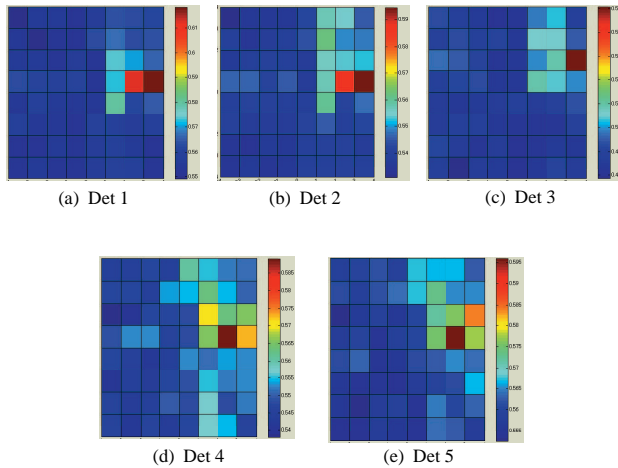


Fig. 8 The estimation results of the sound source direction

3.2 実験 2

Fig. 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 $L_{xi}(\theta_{ij})$, $i, j = 1, \dots, 8$ 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、 $L_{xi}(\theta_{ij})$, $i, j = 1, \dots, 8$ 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。

Fig. 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、**Table 3** のような結果を得た。**Fig. 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

Table 3 Similarity results (Sound position)

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

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。**Fig. 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 (**Fig. 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。



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。(10) 式による分類結果は、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も確認した。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。たとえば、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なう。部屋 (W885×D585×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、カメラの

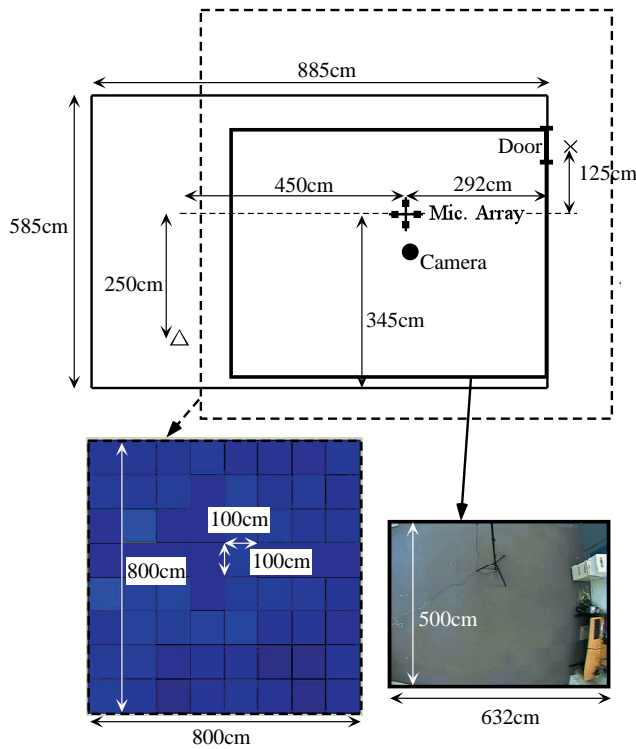


Fig. 10 Layout of a laboratory

撮影可能な範囲は約 500cm×632cm となる。Fig.10 において、実線で描かれた外枠の四角が部屋の大きさを表わしている、その内側の実線の四角がカメラの映像範囲を表わしている。したがって、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、カメラの近くにベクトルセンサを置いた場合、予め音源定位を行なうと設定した範囲は、四角い破線の部分である。(実験 1 とは、音を発生させる場所の高さが異なった設定となっていることに注意。実験 1 では、天井から約 1.3m、本章では、天井から約 2.6m)。破線の四角い範囲内であれば、たとえドアが開まった状態で部屋の外に音源があったとしても (Fig.10 の×印が音源)、音源の場所を推定することができる (Fig.11(a) の濃い赤色の部分が音源である確率が一番高いことを表わしている)。また、破線の四角い範囲外に音源があったとしても (Fig.10 中の△)、音源△の方向で音が発生したことがわかる音源推定結果を得ることができる (Fig.11(b) の濃い赤色の部分が推定結果、Fig.10 中の△の位置関係から、△の音源の方向が推定できていることがわかる)。

したがって、見守りを目的とした場合、本システムは、カメラの映像には映らないような場所からの音でも、その音が観測できれば、音に関する危険を察知できる可能性があると考えられるので、カメラを使用するとき苦勞する設置場所選び、つまり、死角を無くすように設置するには、どの場所にカメラを設置すればよいかという問題が、カメラほど苦勞

する問題とはならなくなるであろう。また、カメラよりも広範囲の見守りが可能になるので、設置するシステムの台数もカメラよりは少なくすむようになるであろう。

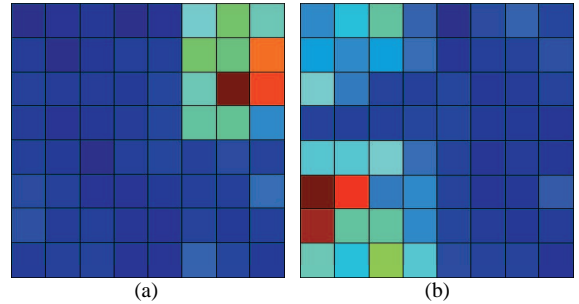


Fig. 11 Localization results

5. おわりに

本論文では、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なう必要もなく、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なうことによって、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ない、より便利に働く、見守りシステムの構築を目指す。また、マイクロフォンデバイスや電子回路のコンパクト化を通じて、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。さらに、集まった日常音と日常音源方向の特徴量を学習により更新できるようにすることで、すべての特徴量を保存しなくてもよいようにし、データベースのコンパクト化を図り、過去の特徴量すべてを使って類似度を計算するのではなく、更新後の特徴量のみを使うことで、より効率よく非日常音や非日常音源方向が検出できるようにする方法も検討する。

参考文献

- 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," Proce. of ICME03, III, 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, 51-1, 130/138(2005)
- 4) C. Clavel, T. Ehrette, and G. Richard, "Events Detection for An Audio-Based Surveillance System," Proce. of

- ICME2005, 1306/1309(2005)
- 5) A. Dufaux, L. Besacier, M. Ansonge, F. Pellandini, "Automatic Sound Detection and Recognition for Noisy Environment," *Proce. of EUSIPCO 2000*(2000)
 - 6) F. J. Fahy, *Sound Intensity*, Elsevier Applied Science(1998)
 - 7) G. Guo and S. Z. Li, "Content-Based Audio Classification and Retrieval by Support Vector Machines," *IEEE Trans. Neural Networks*, **14**-1, 209/215(2003)
 - 8) A. Nehorai and E. Paldi, "Acoustic Vector-Sensor Array Processing," *IEEE Trans. Signal Processing*, **42**-9, 2481/2491(1994)
 - 9) E. Pampalk, "Computational Models of Music Similarity and their Application to Music Information Retrieval", Doctoral Thesis, Vienna University of Technology, Austria(2006)
 - 10) 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*, **55**-4, 1025/1032(2006)
 - 11) R. O. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Trans. Antennas Propag.*, **AP-34**-3, 276/280(1986)
 - 12) A. Temko and C. Nadeu, "Classification of meeting-room acoustic events with support vector machines and variable-feature-set clustering," *Proce. of ICASSP2005*, **V**, 505/508(2005)
 - 13) M. Vacher, D. Istrate, L. Besacier, J. F. Serignat, and E. Castelli, "Sound Detection and Classification for Medical Telesurvey," *Proce. of the 2nd International Conf. on Biomedical Eng.*, 395/399(2004)
 - 14) A. Itai, H. Yasukawa, "Footstep Classification Using Simple Speech Recognition Technique," *Proce. of 2008 IEEE International Symposium on Circuits and Systems*, 3234/3237(2008)
 - 15) 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," 第19回ユビキタスコンピューティングシステム研究発表会論文集, 情報処理学会研究報告, **2008-66**, 19/26(2008)

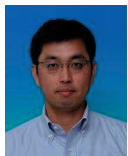
車谷 浩一



1989年 東大大学院工学系研究科博士課程修了。工学博士。同年、電子技術総合研究所入所。2001年 産業技術総合研究所へ出向。情報技術研究部門マルチエージェントグループ長。センサーネットワーク、ユビキタスコンピューティング、測位システム、マルチエージェント社会シミュレーション等の研究に興味を持つ。

[著者紹介]

河本 満 (正会員)



1997年九工大大学院工学研究科博士後期課程修了, 同年理化学研究所バイオ・ミメティックコントロール研究センターフロンティア研究員。2000年年島根大学総合理工学部講師, 2002年同大学助教授, 2004年4月から2005年3月英国リバプール大学に訪問研究員として滞在。2005年より産業技術総合研究所研究員となり, 現在に至る。博士(工学)。ブラインド信号処理, マイクロホンアレイを使った音響信号処理などの研究に従事。

浅野 太



1986年東北大学工学部電気工学科卒業, 91年から95年東北大学電気通信研究所助手。95年から2001年電子技術総合研究所研究員, 2002年から産業技術総合研究所(旧電総研)グループリーダー, アレイ信号処理, 実環境音声認識などの研究に従事。