

環境計測から得られる環境音データのサービス利活用への挑戦
—環境音データの分類法について—

Challenges for service application utilizing sound data obtained from environmental measurements

-On a classification of sound data-

河本 満† 幸島 明男† 車谷 浩一†
Mitsuru Kawamoto Akio Sashima Koichi Kurumatani

1. まえがき

サービス事業者の生産性向上、活性化の取り組みとして、近年、センサ技術を使いサービス現場の従業員の行動等を観測し、その観測結果に基づいて現場や業務改善を行う工学的アプローチ、いわゆるサービス工学が注目を集めている。これまでにサービス工学の分野では、「観測」、「分析」、「設計」、「適用」を軸に収集したデータを効率的、効果的に利活用するための手法を提案している。例えば、「観測」には、カメラ、RFID、生体センサなどのセンサを用いて、サービス現場での従業員の動線などから業務の無駄を分析し、業務改善が可能になることを示している[1]。

本研究では、環境計測によって得られる音データを取り上げる。センサにはマイクロフォンアレイを用いて、環境音を計測し、観測した環境音から、現在の音環境の状態を把握、その把握状況をサービスに応用することを目的とするサービス工学に関わる研究を取り扱う。この目的を達成するため、本論文では、「どこで」、「どのような」音が発生したかが把握できる手法を用いた未知の音環境から得られる環境音データの分類法を提案し、分類されたデータから現状を把握する考え方について報告する。現状把握の手法や把握された情報からサービスへの応用に関しては、まだ具体化されていないので、詳細は次の論文で報告することにする。

以下、2章では音環境計測に関する関連研究を述べる。3章では、本研究の問題設定、4章で提案手法、5章で実験結果を示し、6章でまとめを行う。

2. 関連研究

環境計測を取り扱っている従来研究には、大きく分けて2つあると思われる。1つは予め観測したい音が決められている場合、もう1つは決められていない場合である。前者の場合は、環境計測を通じて予め決められた音やそれ以外の音を区別する研究が最も代表的である[例えば、2]。応用例としては、異常や通常音など決められた音を検出することによって、異常状態を音で検知する例が示されている。

一方、後者は、未知の音環境における環境計測を通じて、どのような音が観測できるのか、計測過程で決められていく性質をもっている[3]。このことから、非日常音の検知の応用例が示されている。しかしながら、音データのさらなる2次利用を想定した場合は、音環境全体の把握をどうすべきかが重要な課題となるであろう。

本論文では、オンラインで計測される音データを一日単位でまとめ、その集合の中で相関関係を解析することにより、音データを分類・整理することで音環境全体が把握で

†(独) 産業技術総合研究所, AIST サービス工学研究センター都市空間サービス基盤研究チーム

きることを示される。

3. 問題設定

3.1 音環境のモデル化の可能性

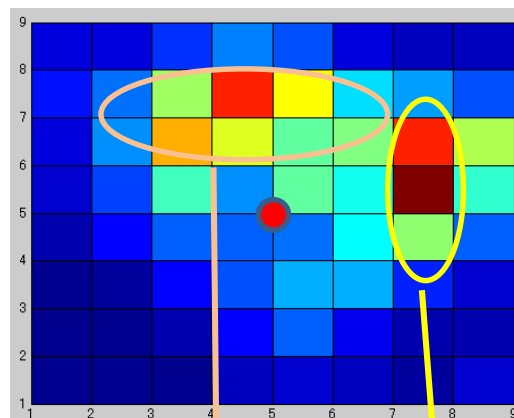


図1 音分布の一例

ある音環境におけるある一日の音源位置推定分布を図1に示す。図では、集合住宅周りの音分布を示しており、色の濃い部分では多くの音が発生したことを表している。同様の分布図を別の日で作成しても大体同じような傾向となる。また、平日の発生音源を時間経過に合わせて再生するとあるパターンが存在することが分った。つまり、集合住宅周りでは、朝、昼、晩の音発生パターンが平日ではほぼ同じということである。

これらのことから、音環境のモデル化は可能であり、さらに、平日においては、一日分の音発生データだけで音環境のモデル化が可能になると考えることができる。

3.2 音環境計測

音環境計測を行うセンサには、図2に示すマイクロフォンアレイを用いる。4個のキューブから構成され、それぞれのキューブには4個のマイクロフォンが取り付けられて

いる。計16個のマイクロフォンから構成されるマイクロフォンアレイを使って音環境を計測する。

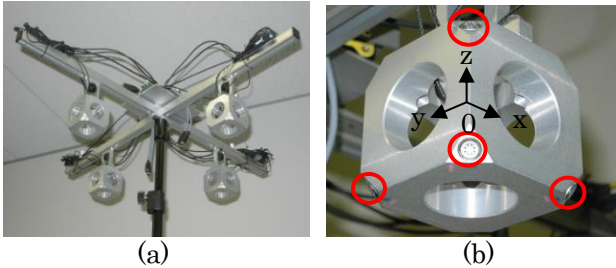


図2 (a)マイクロフォンアレイ
(b)キューブ(○:マイクロフォン)

音環境の計測場所は、屋外を想定している。図3は実際にショッピングセンター(二子玉川ライズ)にマイクロフォンアレイを設置した環境計測場所を示している。このような屋外で計測された音を対象にして「どこで」「どのような」に関する分析、分類問題を取り扱う。



図3 マイクロフォンアレイを設置した環境計測場所の例
(場所: 二子玉川ライズショッピングセンター)

3.3 観測信号

マイクロフォンアレイで観測された音は周波数領域で取り扱われる。すなわち、観測信号は短時間フーリエ変換される。このとき、得られる信号を $\mathbf{y}(t, \omega) = [Y_1(t, \omega), \dots, Y_M(t, \omega)]^T$ と定義する。 $Y_m(t, \omega)$ は周波数 ω 、時間 t における m 番目のマイクロフォンの出力を表している。記述の簡略化のため、これ以降、周波数 ω の表記は省略する。

4. 提案手法

本章では、図3に示すような音環境で発生した音源の位置推定法と分類法を提案する。2つの方法に共通して利用する数学的手法は非負値行列因子分解(NMF)である[4,5]。つまり、 $\mathbf{y}(t)$ のパワースペクトルにNMF法を適用し、得られた基底ベクトルと係数ベクトルを提案手法に利用する。

4.1 音源の位置推定法

音環境計測における音源位置推定において、発生音源の他に背景雑音を観測される。この背景雑音が音源位置推定の精度に大きく影響することを実験的に把握している。そ

こで、その影響をできるだけ排除するような音源位置推定を提案したいと考えている。そこで、その目的に対して、我々はNMF法によって得られた基底ベクトルと係数ベクトルに注目する。

ここで、音源位置推定を行いたい音源が以下のようなパワースペクトルになったとする。

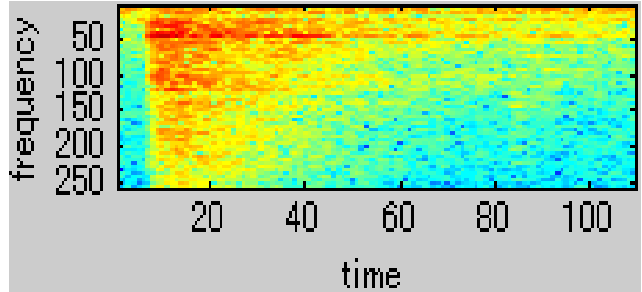


図4(a)空き缶を集めた袋の落下音のパワースペクトル

図4(a)では、約5sから40sの間が落下音のパワースペクトルを表している。その他は背景雑音のパワースペクトルとなっている。従って、音源位置推定には、できるだけ約5sから40sの区間の音を積極的に利用したい。このとき、図4(a)に示すパワースペクトルにNMF法(GaP-NMF[5])を適用すると基底ベクトルと係数ベクトルはそれぞれ図4(b)、図4(c)のようになる。

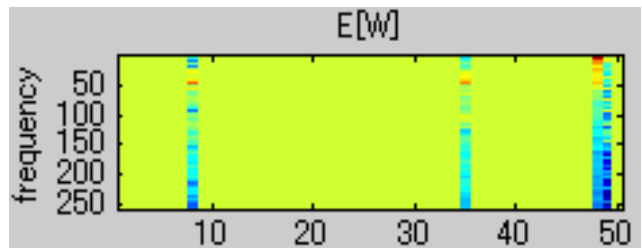


図4(b) 図4(a)のパワースペクトルの基底ベクトル

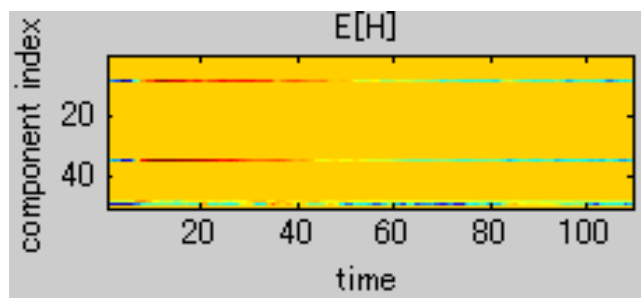


図4(c) 図4(a)のパワースペクトルの係数ベクトル

それぞれの図で縦線、横線に描かれている濃い色の部分がかつとも値が高い、利用したい重要な情報であることを表している。また、図4(b)の縦線は図4(a)のパワースペクトルで基底となる周波数成分を表しており、図4(c)の横線は、4本の縦線それぞれに対応した時間的変化を表している。従って、音源位置推定に利用したい信号の周波数成分は図4(c)から図4(b)の左から2本の基底ベクトルを参照すればよいことが分かる。このとき、左から3本目の基底ベクトルの濃い色の部分は背景雑音の周波数成分を表しているこ

とが分かる。これらのことから、濃い色の部分の周波数が低周波数となっている基底ベクトル以外の基底ベクトルを参照して、参照する基底ベクトルが持っている濃い色の周波数成分を選択的に利用することを考える。周波数選択方法としては、低周波数(約150Hz以下)は最初からカットし、150Hz以上で濃い色の周波数をもつ基底ベクトルを参照し、濃い色の周波数から順に選択すべき周波数を決める。このように決めた周波数のみを用いて音源位置推定を行う。音源位置推定法には周波数領域で動作する MUSIC 法[6]とする。

本手法を用いることにより、「どこで」に関する情報による音データの分類を行う。

4.2 音源の分類法

音源の分類法には Affinity Propagation 法(AP 法)[7]を用いる。AP 法の取扱いで重要な点は類似行列 \mathbf{M} をどのように作るかというところにある。この点に関して、従来にない環境音分類に適した類似行列を発見したところに新規な点であると考えている。つまり、環境音の分類を考えたとき、このような類似行列 \mathbf{M} を使えば、音環境内の妥当な分類が実行できることを示す。

まず、NMF 法をマイクロフォンアレイで計測した観測信号 $\mathbf{Y}_i(t)$ から得られる音源データ s_i ($i = 1, \dots, N$) のパワースペクトル \mathbf{P}_i に適用する。ここで、 N は、一日に音環境で発生した音源の数を表している。また、 s_i ($i = 1, \dots, N$) は音環境内で発生した音、例えば、電車の音、デジタルサイネージの音、子供のハシャギ声など、を表している。このとき、得られる基底ベクトルと係数ベクトルをそれぞれ \mathbf{F}_i 、 \mathbf{C}_i とすると、NMF 法によって再現される s_i のパワースペクトル $\hat{\mathbf{P}}_i$ は

$$\hat{\mathbf{P}}_i = \mathbf{F}_i \mathbf{C}_i, \quad (i = 1, 2, \dots, N) \quad (1)$$

となる。

次に、 \mathbf{F}_i を用いて、 s_j ($j = 1, \dots, N; j \neq i$) のパワースペクトル \mathbf{P}_j に対して NMF 法を適用する。このとき、得られる再現パワースペクトルは

$$\bar{\mathbf{P}}_{ij} = \mathbf{F}_i \mathbf{C}_j, \quad (i, j = 1, 2, \dots, N; j \neq i) \quad (2)$$

となる。ここで、 \mathbf{C}_j は s_j から得られた基底ベクトル \mathbf{F}_j を固定して NMF 法を用いたときに得られる係数ベクトルを表している。このとき、 $\hat{\mathbf{P}}_i$ と $\bar{\mathbf{P}}_{ij}$ の違いを以下の板倉-斉藤距離で計算する。

$$m_{ij} = -\log(\sum_{kl} ([\bar{\mathbf{P}}_{ij}]_{kl} / [\hat{\mathbf{P}}_i]_{kl}) - \log([\hat{\mathbf{P}}_i]_{kl})) \quad (3)$$

ここで、 $[X]_{ij}$ は行列 X の (i, j) 要素を表している。式(3)で得られる値は板倉-斉藤距離に自然対数をとったものであり、この値を類似行列 \mathbf{M} の要素に用いる。

さらに、類似行列 \mathbf{M} の要素として次の2つの値も用いることにする。図5は s_i の再現パワースペクトル $\hat{\mathbf{P}}_i$ を表している。ここで、図5の縦軸は周波数成分、横軸は時間を表している。

時間軸方向 t (図5の横軸) に従って周波数成分を足していき得られたベクトルを $\hat{\mathbf{p}}_i^t$ ($i = 1, 2, \dots, N$) とする。また、周波数軸方向 f (図5の縦軸) に従って時間軸の値を足していき得られたベクトルを $\hat{\mathbf{p}}_i^f$ ($i = 1, 2, \dots, N$) とし、そのベクトルに対してパワースペクトル計算を行って、得られた値を $p_s[\hat{\mathbf{p}}_i^t]$ と表す。このとき、 $\hat{\mathbf{p}}_i^t$ は発生音源の全体の周波数パターンが表現され、 $p_s[\hat{\mathbf{p}}_i^f]$ は発生音源の波形パターンが表現されると考えられる。これら2つの値を使って、そ

れぞれの相関値を計算し、それらの値を類似行列の要素として用いる。

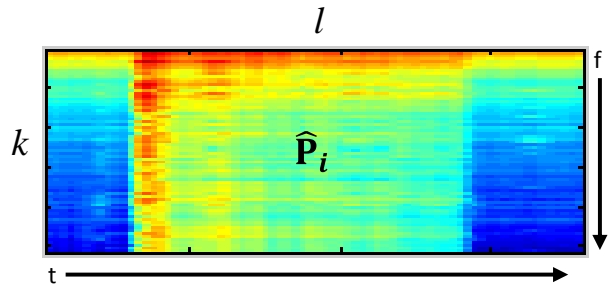


図5 再現パワースペクトル

つまり、

$$m'_{ij} = -(\hat{\mathbf{p}}_i^t \cdot \hat{\mathbf{p}}_j^t) / \|\hat{\mathbf{p}}_i^t\| \|\hat{\mathbf{p}}_j^t\| \quad (4)$$

$$m''_{ij} = -(p_s[\hat{\mathbf{p}}_i^f] \cdot p_s[\hat{\mathbf{p}}_j^f]) / \|p_s[\hat{\mathbf{p}}_i^f]\| \|p_s[\hat{\mathbf{p}}_j^f]\| \quad (5)$$

ここで、 $(\mathbf{a} \cdot \mathbf{b})$ は \mathbf{a} ベクトルと \mathbf{b} ベクトルの内積を表している。また、 $\|\mathbf{x}\|$ は \mathbf{x} のユークリッドノルムを表している。

従って、ここで用いる類似行列 \mathbf{M} は

$$[\mathbf{M}]_{ij} = m_{ij} + m'_{ij} + m''_{ij} \quad (6)$$

となる。

この類似行列 \mathbf{M} に AP 法を適用することによって、音環境内の代表音を検出される。ここでは、この代表音によって、「どのような」に関する情報による音データ分類を実行する。

また、検出した代表音と現在の発生音源とを比較することで、「どのような」音、つまり、環境内でいつも発生している音か否かの現状を把握することも考えている。把握方法に関しては、代表音に対して NMF 法を適用し、得られた基底ベクトルに対して式(2)の計算を行い、代表音の $\hat{\mathbf{P}}_i$ に対して、代表音の \mathbf{F}_i を用いて算出された現在の発生音源の $\bar{\mathbf{P}}_{ij}$ との違いを式(3)から計算し、その距離に応じて識別結果を算出することにより、上述の現状を把握することを考えている。詳細は後述の論文を参照のこと。

本論文では、2つの提案手法を用いて得られる「どこで」と「どのような」に関する情報に従い、環境音の分類を行う。

5. 実験結果

本章では、提案する音源の位置推定法と分類法についてそれらの有効性を実験により示す。

5.1 音源の分類法



図6 マイクロフォンアレイ設置場所

本節では、提案手法の 1 つである位置推定法に関する実験結果を示す。本実験は、図 6 に示す集合住宅の 3F のベランダにマイクロフォンアレイを設置し、集合住宅周辺の音環境が計測する空間で行われた。実験に、用いた音源は、複数の空き缶を入れた袋を人が胸の高さくらいから落とす落下音を用いた。音源場所は図 6 の①、②、③である。その落下音をマイクロフォンアレイで観測し、その観測信号を用いて、落下した場所を推定する実験を行った。比較のために、周波数選択を行わない MUSIC を用いた。それぞれの実験結果を表 1 に示す。

表 1 位置推定実験結果

Positions	The number of misestimate times		
	①	②	③
The conventional MUSIC	2	1	4
The MUSIC with our proposed technique	0	0	1

3 か所のそれぞれの場所で 10 回程度落下音を発生させ、そのとき、誤った推定結果を示したときの回数を数えた。ここで、本マイクロフォンアレイで位置が把握できる場所は、図中の白色のメッシュ内の位置である。従って、②は落下位置が推定できるが、①や③は方向のみ推定できていけばよいものとした。ただし、①に関しては、左端の真ん中付近、③に関しては右端の真ん中付近を正解とした。表 1 から提案手法の方が精度良く位置推定ができていていることが分かる。従って、周波数選択は屋外における環境音位置推定では利用すべき重要な技術であると考えることができる。

5.2 音源の分類法

音源分類の提案法について、その実験結果を示す。まず、以下の 20 個の音源を用いて分類の実験を行った。

[s1, s2, s3, s4, s5, s6, s7, s8, s9, s10, s11, s12, s13, s14, s15, s16, s17, s18, s19, s20]
 = [Cn, Can, Can, Can, Cn, Cn, Cn, Can, Cn, Cn, Cn, Can, CC, Can, Can, K, CC, K, Cn, CnCC], (7)

ここで、Cn は車の走行音、Can は複数の空き缶を入れた袋の落下音、CC はものを叩いたような単発音、K はカラスの鳴き声、CnCC は Cn と CC 両方が混在した音を表している。また、これらの音は、図 6 に示す音環境で観測された音である。

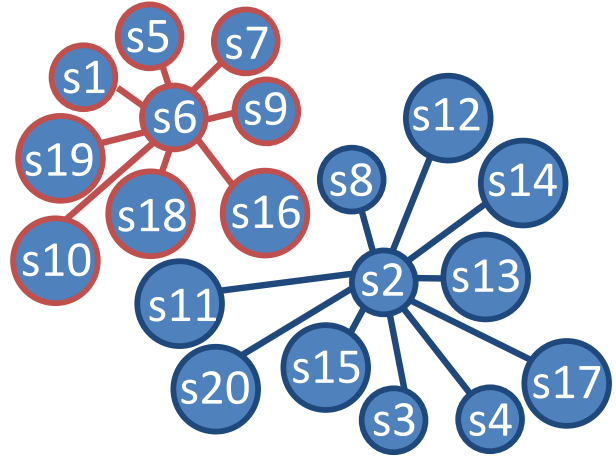
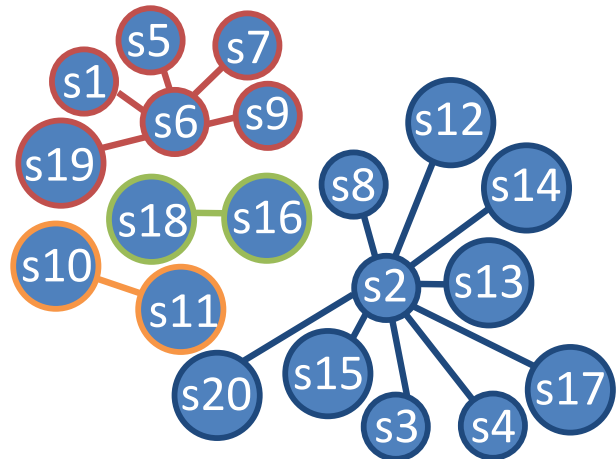
本実験において、類似行列を計算する際、その対角要素 m_{ii} は、 m_{ii} に係る m_{ij} 以外の列ベクトルと行ベクトルのそれぞれの中央値の平均値に係数 α をかけた値を用いた。すなわち、列ベクトルを $\mathbf{m}_{\cdot i} = [m_{1i}, m_{2i}, \dots, m_{i-1i}, m_{i+1i}, \dots, m_{Ni}]^T$ 、行ベクトルを $\mathbf{m}_{i\cdot} = [m_{i1}, m_{i2}, \dots, m_{i-1i}, m_{i+1i}, \dots, m_{iN}]$ とすると類似行列 \mathbf{M} の対角要素 m_{ii} は以下の値とした。

$$m_{ii} = \alpha(\text{median}(\mathbf{m}_{\cdot i}) + \text{median}(\mathbf{m}_{i\cdot}))/2 \quad (i = 1, 2, \dots, N) \quad (8)$$
 ただし、 $\text{median}(\mathbf{x})$ はベクトル \mathbf{x} の中央値を表している。

ここで、 α を変えることによって、代表音の数が変化することに注意したい。つまり、 $\alpha = 1.0$ の場合、分類結果は、図 7 (a) のようなかなり大雑把な分類となるが、 α の

値を $\alpha = 0.95$ 程度にすれば、妥当な相関図が得られるようになる。

図 7 (b) を見ると、代表音 (s2, s6, s11, s18) に対して類似な音が集まる相関図となっている。図 7 (b) は NMF 法の初期値を変えると若干代表音の数が増減するが大体似たような相関図を得ることができる。 $\alpha = 0.9$ 程度までであれば、データを 1/4~1/5 程度までに圧縮でき、妥当なデータ分類が可能であることを確認した。

図 7 (a) 発生音の相関図 ($\alpha = 1.0$ の場合)図 7 (b) 発生音の相関図 ($\alpha = 0.95$ の場合)

しかしながら、 α の値を小さくし過ぎると、ほとんどのすべての音が代表音となり、データ分類の機能を果たさなくなるので注意が必要である。本実験では $\alpha = 0.8$ よりも小さな値を設定したとき、ほとんどすべての音が代表音となる傾向を示した。



図 8 マイクロフォンアレイ設置場所(二子玉川ライズ)

上述した実験結果を踏まえて、次に図 8 に示すショッピングセンター(二子玉川ライズ)で計測した一日(平日)の環境音の分類を試みた。その結果、代表音からこのショッピングセンターの音環境は、電車の音、子供のハシャギ声、その他雑音(例えば、作業音、天候音(風音)など)で表現されていると考えることができそうだとということが分かった。また、一日の音環境の時間変化も同時に把握できることが分かった。

5.3 実験結果可視化法

提案手法により得られたデータ(位置情報、分類情報)を可視化する一つの方法を紹介する。ここで、可視化するデータは、図 8 に示したショッピングセンターで計測した観測信号から得られるものとする。ここで、図 9(a)に示す画面に位置推定結果と分類結果が表される。すなわち、図 9(a)はショッピングセンターの音環境空間を表している。すなわち、紫色のメッシュ部分が位置推定可能な領域を表しており、図 8 では、中央の通路の部分にそれが該当する。メッシュ外で発生した音源に関しては、発生した方向によりそれに近い外側のメッシュに反応がある。一つのメッシュの一边は約 5 m である。青丸はマイクロフォンアレイの設置位置を示しており、建物の 2 階(地上から約 9 m)の天井に設置してある(図 8 参照)。

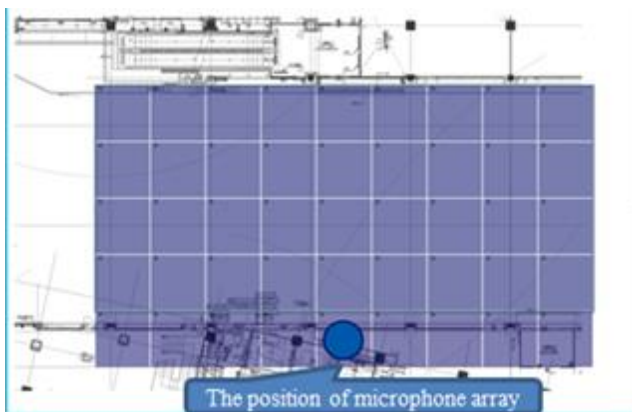


図 9(a) データ可視化の一例: 音環境空間図

可視化の例としては、図 9(b)に示すような音源位置推定で推定した位置と分類法で得られた代表音との類似度計算により得られた識別結果を表示することが可能である。ここで、図 9(b)では、子供のクリップアートの位置に子供のハシャギ声が発生したことを表している。

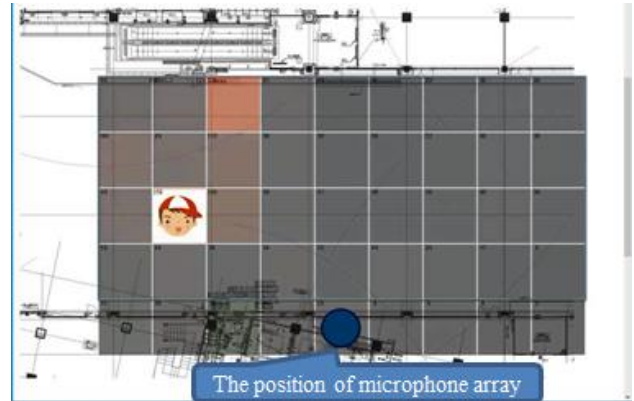


図 9(b) データ可視化の一例: 「どこで」、「どのような」音が発生したかの識別結果

6. おわりに

本論文では、音環境におけるビックデータに関して、そのデータの分類方法を提案した。提案分類方法では、「どこで」、「どのような」音が発生したかが把握できるようにデータが整理される。また、分類法には、音源位置推定法と AP 法を用いた。両方に共通して NMF 法を用い、基底ベクトルと係数ベクトルの情報を有効利用することにより、音源位置推定法に関しては周波数選択を、AP 法に関しては類似行列計算を効果的に行うアルゴリズムを提案した。

本研究では、提案手法を利用したサービス提供が最終目的であるので、今後は、2つの提案法から得られる解析結果をサービス現場で活用できるようにするためのシステムづくりを行う。具体的には、提案分類法で音環境における代表音を検出できることや、検出した代表音から音環境モデルも構築できることから、日ごろの音(代表音)との違いを検出することによって、非日常的な音の識別を行い、いつもと違うことへの「気づき」が与えられるような仕組みづくりや、提案分類法でイベント時の拍手などの喝采音とイベント音との区別が可能になることも分かっているので、これらの音から「賑わい度」を検知することによって、賑わいに関する客観指標が与えられる仕組みづくりなどを行っていこうと考えている。

謝辞

本研究の一部は、科学研究費補助金(基盤研究 C、研究課題番号: 25330379)の援助を受けた。また、東京農工大の田中希武氏(現コニカミノルタ)に研究協力を受けた。

参考文献

- [1] サービス工学研究開発事業(平成 21、22 年度)、サービス工学研究開発分野(平成 23 年度)、<https://unit.aist.go.jp/cfsr/contents/meti-top.htm>.

- [2] A. Ito, A. Aiba, M. Ito, and S. Makino, "Detection of Abnormal Sound Using Multi-stage GMM for Surveillance Microphone," Proc. of 2009 Fifth International Conference on Information Assurance and Security, pp. 733-736, 2009.
- [3] M. Kawamoto, F. Asano, K. Kurumatani, Y. Hua, "A System for Detecting Unusual Sounds from Sound Environment Observed by Microphone Arrays," International Journal of Computer Information Systems and Industrial Management Applications (IJCISIM), Vol. 1, pp. 232-238, 2009.
- [4] M. D. Hoffman, D. M. Blei, P.R. Cook, "Bayesian Nonparametric Matrix Factorization for Recorded Music," in *Proc. of the International Conference on Machine Learning (ICML)*, Haifa, 2010.
- [5] MATLAB code for GaP-NMF, <http://www.cs.princeton.edu/~mdhoffma>.
- [6] R. O. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Trans. Antenna Propag.*, Vol. AP-34, No. 3, pp. 276-280, March 1986.
- [7] B.J.Frey and D. Dueck, "Clustering by Passing Messages Between Data Points," *SCIENCE*, Vol. 315, 16 Feb. 2007.