Vol.2016-MUS-111 No.48
2016/5/22

# 基底ベクトルの共有化に基づくNMFによるオーバラップ音響イベントの検出

石川 力[1]　山本 一公[1,a)]　中川 聖一[1]

**概要**：音響イベント検出（AED）はマルチメディアコンテンツの分析に重要な役割を果たすと考えられ，これまでも主に音声との音源分離や音声認識と同様の手法を用いて研究が行われてきた．しかし，ひとつのイベント種の中での音響バリエーションが大きいことや，イベント同士のオーバーラップが発生することによって，検出精度が低下することが問題となっている．本稿では，オーバーラップ音響イベントに焦点を当て，イベント間で基底を共有化したNMFによる音源分離手法を用いて音源分離を行い，DNNを組み合わせてイベントオーバーラップ区間において音響イベント検出を行う手法を提案する．IEEE D-CASE 2012のデータセットを用いて評価実験を行い，提案法により先行研究に対して絶対値で20%の検出性能改善を得ることができた．

**キーワード**：音響イベント検出，音源分離，DNN，NMF，基底ベクトル

## 1. はじめに

聴覚情景解析（Computational Auditory Scene Analysis: CASA）と呼ばれる，身の回りで発生する音への関心・理解の分野では，音声認識を利用した音響イベント検出 [1] や音環境推定 [2] に関する研究が行われている．この技術の応用として，騒音の音源特定 [3] や監視カメラの死角をカバーするための音入力の利用 [4] 等が検討されており，様々な応用が期待されている．

我々はこれまでに，オフィス環境における音響イベント検出，特にオーバーラップイベントの検出に関して研究を行った [5]．その際にNon-negative Matrxi Factorization (NMF) による音源分離後に音響モデルを利用した検出を行ったが，この時単に分離を行って検出を行うだけでは，分離を行わずに検出を行う結果と比較して，精度が悪化する結果となった．本稿では，この精度悪化の原因分析と共に，性能を改善するために音響イベントクラス間でNMFの基底ベクトルを共有する手法を提案する．実験では先行研究との比較を行い，本手法の有効性を示す．

## 2. 先行研究

CASAを題材にした競争型ワークショップとしてIEEE D-CASE [6] が開催されており，D-CASEの音響イベント検出タスクでは，オフィス環境を想定したトラックによる評価が行われている．評価はその中で2つのタスクに分けられており，実環境で音響イベントがオーバーラップしないように収録されたトラックであるOL (Office Live) による評価と，合成音による音響イベントのオーバーラップが含まれるトラックであるOS (Office Synthetic) による評価が設定されている．このタスクにおいて，Vuegenら [8] は音声認識で一般的に利用される特徴量であるMFCC (Mel-Frequency Cepstral Coefficients) とGMM (Gaussian Mixture Model) を利用して，OLでは43.4%，OSでは13.5%の$F$値を得ている．Gemmekeら [9] はExemplar-based NMFのアクティベーションを利用した識別でこのタスクに取り組んでおり，OLで31.4%，OSで21.3%の$F$値を得ている．これら2つの手法は，共に音響イベントがオーバーラップする条件下での検出精度は低く，イベントが重なることで検出性能が悪化することを示している．また，オーバーラップが生じないOLの条件においてもまだ十分な検出精度が得られているとは言い難いことから，検出・識別性能の改善が望まれると言える．

音響イベントの検出・識別精度が低くなる要因として，音響イベントが背景雑音に埋もれてしまう場合や，複数の音響イベントの重なりにより複雑なスペクトルパターンが生成され音響モデルとのミスマッチが生じることが考えられる．音響イベントの生成メカニズムは多様であり，それ

[1] 豊橋技術科学大学 情報・知能工学系
Dept. Computer Science and Engineering, Toyohashi University of Technology, Toyohashi, Aichi 441-8580, Japan
[a)] kyama@tut.jp

ぞれの音響イベント音は特異でスパースなスペクトルパターンを保持していることが期待されるが，実際には他の音響イベントと類似したスペクトルパターンを持つ場合も多いため，このことを考慮しないと十分な性能が得られないと考えられる．

我々の以前の研究 [5] では，音源分離を行った後に音響モデルによる識別行う手法を採用した．しかし，単純に分離を行った後に音響モデルによる識別を行った場合の結果は，分離を行わずに音響モデル単独で識別を行った結果と比較して，$F$ 値が絶対値で 16.1% 低下する結果となった (特徴量が MFCC の場合)．このことは，NMF による音源分離の性能が十分でないことを示している．この原因として，音源分離を行う際のクラス間の基底ベクトルの類似性が考えられる．D-CASE の音響イベント検出では，対象となる音源クラス（音響イベント）数は無音を含め 17 種類で，この中に音響的特徴が類似しているクラスや広い周波数帯域に渡って強いパワーを示すクラスが含まれている．そのため分離を行う際，各基底ベクトルに割り振られるアクティベーションの大きさに偏りが生じ，結果として検出性能も悪くなったと考えられる．

本稿では，この [5] における音響イベントがオーバーラップする条件での問題に対する原因分析と，NMF に対する検出率改善手法を検討する．具体的には，NMF で利用する基底ベクトルをクラス間で共有して分離を行う手法を提案する．全音響イベントに対する共通の基底ベクトルを，各クラスの基底ベクトルと対応付けることで，クラス間で類似した基底ベクトルが存在する問題を防ぐことができる．

## 3. 音響イベント検出について

本稿では，我々の以前の研究 [5] と同様に，音源分離を前処理として行い，後段の音響モデルで識別を行う音響イベント検出システムを用いる．図 1 にシステムの概要を示す．前段では，入力トラックに対して NMF を行うことによって各音響イベントクラスへと分離する．この処理は実際にイベント音のオーバーラップの有無にかかわらず，全体に対して行う．後段の音響モデルでは，Deep Neural Network(DNN) による識別モデルを用いて，分離されたクラストラックごとにフレームレベルの各クラスの識別結果を導出し，閾値処理を用いてイベント区間の検出を行う．また，複数の音響モデルの出力スコアを統合する手法（FUSION）も利用する．

### 3.1 NMF による音源分離

#### 3.1.1 従来手法

オーバーラップした音源同士の分離を行うため，NMF [12] を利用する．フレーム単位の周波数スペクトルを時系列で結合したものを行列 $S$ とし，これを各音源の基底ベクトルの集合行列 $W$ とアクティベーション行列 $H$ で

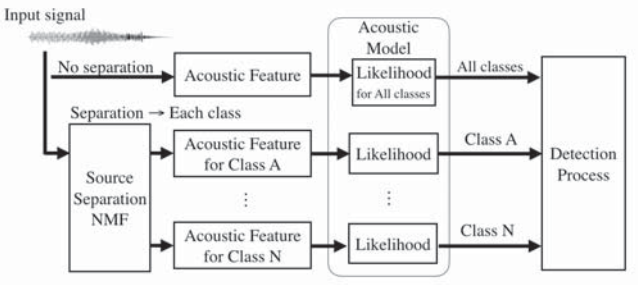


**図 1** 使用した音響イベント検出システム
**Fig. 1** Our aoustic event detection system

表すと，$S \simeq \hat{S} = WH$ という関係になる．周波数ビン数を $I(i = 1, 2, \cdots, I)$，フレーム数を $J(j = 1, 2, \cdots, J)$，基底ベクトル数を $K(k = 1, 2, \cdots, K)$ とすると，行列 $S$，$\hat{S}$ は $I \times J$ 行列，$W$ は $I \times K$ 行列，$H$ は $K \times J$ 行列となる．$W_n$ を各クラスの基底ベクトルの集合とすると，$W$ は $W = [W_1, W_2, \ldots, W_N]$ と表される（$N$ は音響イベントクラス数）．$W_n$ に対応するアクティベーションの集合を $H_n$ とし，$W$ を固定して各基底ベクトルに対する重み $H$ を推定することで $S \simeq \sum_n^N W_n H_n (n = 1, 2, \cdots, N)$ のように観測信号を $N$ 個の音源の和にした形で表すことができる．音源 $S_n$ を取り出す場合は分離結果よりフィルタを構成し，入力信号にかけることで音源別に推定することができる．

$$S_n \simeq \hat{S_n} = S \times \frac{W_n H_n}{\sum_m^N W_m H_m} \tag{1}$$

本稿では基底ベクトルの集合を作成するため，文献 [13] で提案したベクトル量子化（VQ）を用いたコードブック作成法を用い，各クラスのコードブック (コードベクトルの集合) として基底ベクトル集合を作成した．分離では無音を含めた全クラスに対する分離を行った．

#### 3.1.2 提案手法

NMF による音源分離において，各音源が特異でスパースなパターンを有している場合にはこの方法が有効であると考えられるが，分離対象のクラスが多い識別課題等については基底ベクトルのパターンがクラス間で類似する場合があり，分離時にある基底ベクトルのアクティベーションが類似した基底ベクトルのアクティベーションで置換されてしまうことが起こり得る．しかし，基底ベクトルをクラス間で共有することで，この問題を解決することができると考えられる．そこで，基底ベクトルをクラス間で共有し音源分離を行う手法を提案する．この手法の概要を 図 2 と図 3 に示す．3.1.1 節で述べた従来の NMF 手法では，行列 $W$ を構成する基底ベクトルは，クラスごとの音響パターンの集合から VQ により選択された代表ベクトルで構成される．本手法の基底ベクトル行列は，VQ を用いて作成した全クラスに共通なコードブックと，各クラス毎に求めたコードブックに対して，互いの要素ベクトルを対応付けることによって作成する．基底ベクトルとクラスとの対

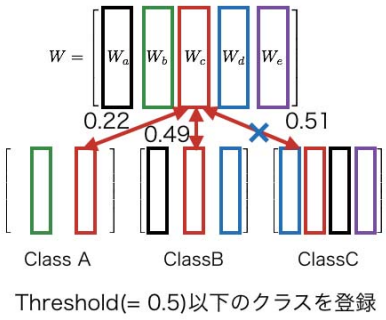


**図 2** 基底の共有 (1) –各クラスの基底ベクトル数が可変–
**Fig. 2** Shared bases (1) – number of bases for each class is variable –

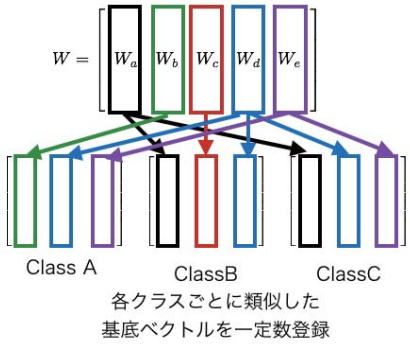


**図 3** 基底の共有 (2) –各クラスの基底ベクトル数が一定–
**Fig. 3** Shared bases (2) – number of bases for each class is constant –

応付けには，全クラス共通のコードブックからの基底ベクトル（図中の $W$ の基底ベクトル）と各クラス毎に作成したコードブックに含まれる基底ベクトルとの距離を利用する．クラスとの対応付けは，

(1) 基底ベクトルとクラス間の距離が一定値以下（図 2）．各クラスの基底ベクトル数は可変となる．
(2) 基底ベクトルと各クラスの基底ベクトルとの距離が近い基底ベクトルを近い順に一定数登録（図 3）．

の 2 つの方法を検討する．

なお，これに近い考え方の手法として，Komatsu らの手法 [14] が提案されている．本稿の提案手法と目的意識は同じであるが，本提案手法の方がより直接的に有用な共有基底を生成できると考えられる．

### 3.2 音響モデルの出力値（スコア）の統合

音響イベントの識別は，DNN 出力ユニットの値を利用して行う．出力ユニット $j$（クラス $j$）の値 $O_j$ は次式から導出される．

$$O_j = \sum_{i=1}^{N} W_{ij} h_i + C_j \qquad (2)$$

式中の $h_i$ は直前の隠れ層のユニット $i$ の値，$W_{ij}$ は出力

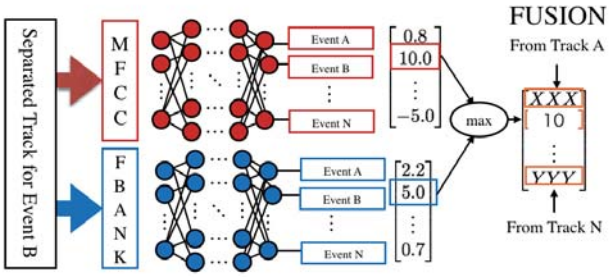


**図 4** 出力値の統合
**Fig. 4** Score fusion

ユニット $j$（クラス $j$）と隠れ層のユニット $i$ にかかる重み，$C_j$ は出力ユニットのバイアス，$N$ は隠れ層のユニット数を指す．この $O_i$ を用いて，出力ユニット数（クラス数）が $M$ の時，次式で出力ユニット $i$ の事後確率が求まる．

$$X_i = \frac{exp(O_i)}{\sum_j^M exp(O_j)} \qquad (3)$$

本稿では，$O_i$ そのものの値と事後確率 $X_i$ の値とを場合によって使い分ける．

また，本稿では図 4 に示すように，MFCC と対数メルフィルタバンク特徴（FBANK）の 2 種類の特徴量で学習した DNN 音響モデルを用意し，2 つの音響モデルの出力値から最大値を求め，音響イベントの検出に利用する手法（FUSION）についても検討する．値を統合することによって複数の音響モデルの傾向が結果に反映され，音響イベント同士のオーバーラップ区間における同時検出精度が改善すると考えている．

### 3.3 音響イベントの検出

(a) DNN による手法

音響イベントの検出は，NMF によって入力トラックを音響イベントクラスに分離，式 (1) で再構成した各クラスの音響イベントスペクトルから求めた特徴量を DNN に入力し，DNN の出力層の各ユニットの値を利用することによって行う．検出の方法を示すための例として，図 5 に DNN の出力層の各ユニットの値を時系列にプロットしたものを示す．図中の左の縦軸には無音 (sil) クラスの事後確率 $X_{sil}$，右の縦軸には各イベントクラスの DNN からの出力値 $O_j$，横軸は時間フレームを示している．無音区間の判別とイベントクラスの検出のために，無音クラスとイベントクラスに対する閾値を設ける．無音クラスの検出に用いる値は，DNN 音響モデルからの事後確率（式 (3)）である．音響イベントクラスの検出には，音響モデルからの値（式 (2)）を直接利用する．無音クラスのみ事後確率を利用する理由は，スコアの変動による識別への影響を抑えるためである．また，スコアの急峻な変動を抑えるために，前後のフレームの値を利用してスコアの平滑化を行っている．平滑化するためのフレーム数は OL の開発セットを用

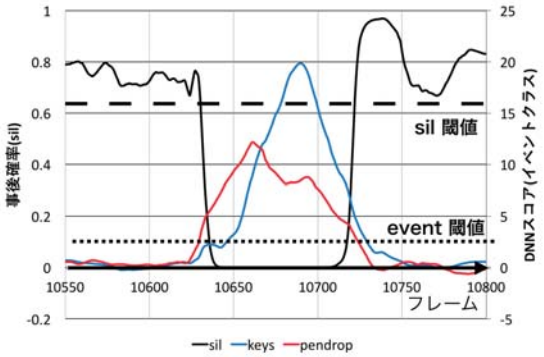


**図 5** 音響イベントの検出と閾値
**Fig. 5** Acoustic event detection and threshold

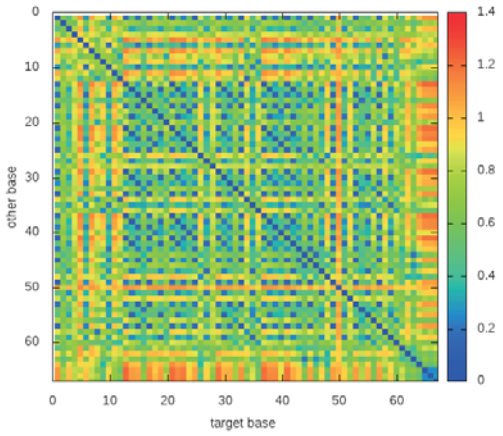


**図 6** 基底ベクトル間の距離
**Fig. 6** Distance between basis vectors within class and across classes

いて調整し，実験では前後 9 フレームの値から移動平均を算出している．分離されたトラックから特徴量を抽出，音響モデルからスコアを取得する際は，DNN の対象クラスに対応する出力ノードからのみスコアを取得し検出に利用する（すなわち，DNN は 1 つだけ学習している）．

(b) NMF を用いた各クラスのアクティベーションの重みの和で判定する手法

NMF のみで検出を行う場合は，NMF により求まるアクティベーション行列の各クラスに対応する基底ベクトルのアクティベーションの和と閾値により検出を行う．

## 4. 以前の研究における問題の調査

我々が以前行った研究 [5] における NMF によるイベント音分離の低性能問題について原因分析するため，各クラスの基底ベクトル間距離について調査した．図 6 に各音響イベントクラスのコードブックに含まれる基底ベクトル間の距離を示す．基底ベクトル間の距離にはユークリッド距離を用いた．この図では，各クラスを 4 個の基底ベクトルで表現し，クラス順に並べてある．図中の色は基底ベクトル間の距離を示しており，赤に近い色であれば基底ベクトル間が離れており，青ければ近いことを示している．当然，対角線に沿った $4 \times 4$ の小行列（クラス内の基底ベクトル間）の距離は小さくなっているが，それ以外の基底ベクトル間の距離も全体として緑から青が多くを占めており，基底ベクトル間の距離が近いものが多いことを示している．このことから，NMF でのイベント音分離時にクラス間の混同が起こっており，このために分離精度が低いのだと考えられる．

## 5. 実験

本節では，IEEE D-CASE [6] TASK2 OS のテストトラックを利用し，音響イベントのオーバーラップが含まれるトラックにおける連続音響イベント検出実験を行い，提案手法の性能評価を行う．

データベースでは，16 クラス (alert, clearthroat, cough, doorslam, drawer, keyboard, keys, knock, laughter, mouse, pageturn, pendrop, phone, printer, speech, switch) の音響イベントが定義されている．学習セットは，各クラス 20 のデータサンプルで構成されており，学習セット全体で収録時間は 15 分程度，サンプリングレート/量子化ビット数は 44.1kHz/24bit で収録されている．OS のテストトラックは 12 トラック収録されており，各トラック 2 分で構成されている．検出対象フレーム数は 99,981 フレームで，そのうち 15,180 フレームは音響イベントのオーバーラップが観測されるフレームである．

### 5.1 実験条件

#### 5.1.1 NMF

従来法の NMF の基底行列は，フレーム長 20ms，シフト幅 10ms，プリエンファシス $1 - 0.97z^{-1}$ としてフーリエ変換を行い，線形振幅スペクトルを求め，振幅正規化，VQ を行い各クラスのコードブック（基底ベクトル群）として作成した．コードブック作成時の距離尺度はユークリッド距離とし，各クラスのコードブックのサイズは 4 とした．分離の際の距離尺度は KL-divergence を使用した．

提案手法では，各クラスのコードブックだけでなく，全クラスのデータの集合からもコードブックを作成した．全クラスのコードブックサイズは実験では 64 に固定し，各クラスのコードブックサイズを 4 もしくは 8 として，全クラス共通の基底ベクトルと各クラスの基底ベクトルとの距離を求めた上でクラスとの対応付を行った．基底とクラスとの対応付にはユークリッド距離を用いて，(1) 距離に対する閾値による判定，及び，(2) クラス毎に全体の基底ベクトル集合から近い基底を上位 4 もしくは 5 個選択して割り当てる，それぞれにより実施した．(1) の閾値は各クラスのコードブックサイズが 4 の場合，$0.35, 0.40, 0.45, 0.50$ とし，各クラスのコードブックサイズが 8 の場合は，$0.40, 0.45, 0.50, 0.55$

として実験を行った．

#### 5.1.2 特徴量

特徴抽出時に，スペクトルサブトラクション (SS) 法 [15] を用いて，背景雑音の抑圧処理を行う．減算は周波数ビンレベルで行い，SS の減算係数は 2.0，フロアリング係数は 0.01 とした．学習時はラベルに示された雑音区間から，テストトラックではトラックの先頭 100 フレームから雑音のスペクトルを推定し，SS に用いた．

特徴量は音声認識で一般的に利用される MFCC 及び FBANK を使用する．特徴量の抽出条件は，フレーム長 20ms，シフト幅 10ms（NMF に同じ），MFCC として 12 次元のケプストラム係数と対数パワー，$\Delta, \Delta\Delta$ の 1 フレームあたり合計 39 次元の特徴ベクトルを抽出した．この時のフィルタバンクのチャネル数は 33 である．FBANK を特徴量として用いる場合は，チャネル数を 45 とし，$\Delta, \Delta\Delta$ を求めて，合計 135 次元の特徴ベクトルを抽出した．

#### 5.1.3 音響モデル

DNN 音響モデルの学習には，D-CASE の学習セットに対して独自に収集した室内雑音を重畳したデータセットを作成し，これを用いてマルチコンディション学習を行った．作成したデータの SNR は $10, 15, 20$dB である．また，D-CASE で公開されている OL の開発セットも併せて学習に利用した．

DNN には，当該時刻フレームに前後 3 フレーム分の情報を加えた計 7 フレームを与えて，学習と識別を行う．入力層は MFCC の場合 273 ユニット，FBANK の場合は 945 ユニットで，隠れ層は 5 層（$512 \to 256 \to 128 \to 64 \to 32$ ユニット），出力層は 17 ユニット（16 音響イベントクラス + 無音）とした．活性化関数には Rectified Linear Unit を使用し，事前学習は行わずに教師あり学習のみ行っている．

#### 5.1.4 評価指標

評価は D-CASE Challenge に基づき，フレームベースの $F$ 値で評価する．フレームベースの評価は 10ms 毎の判定となる．評価に利用する $F$ 値は，

$$F[\%] = 2 \cdot \frac{Precison \cdot Recall}{Precision + Recall} \tag{4}$$

$$Precison[\%] = \frac{C}{E}, \quad Recall[\%] = \frac{C}{GT} \tag{5}$$

として定義されており，式中の $C$ は正解数，$E$ は検出総数，$GT$ はリファレンスラベル数を示している．

評価は先行研究 [5], [6], [8], [9], [10] の結果とフレームベースの $F$ 値で比較を行う．また，オーバーラップ区間における $Recall$ を併せて調査する．ここでは音響イベントが 2 つ以上同時に発生している区間をオーバーラップ区間と定義し，オーバーラップ区間での $Recall$ を $R_{overlap}$ とする．$F$ 値の評価については全区間が対象となる．

検出のための各閾値は 12 トラック共通の値を使用し，結果では 12 トラックの平均値を評価に用いる．

**表 1** 先行研究の評価
**Table 1** Results of revious works

| | Baseline[6] | DHV[10] | GVV[9] | VVK[8] |
|---|---|---|---|---|
| $F$ [%] | 12.8 | 18.7 | 21.3 | 13.5 |

**表 2** NMF のみ または DNN のみの結果 [5]
**Table 2** Results for only NMF and only DNN from [5]

| | NMF のみ | DNN のみ | | |
|---|---|---|---|---|
| | | MFCC | FBANK | FUSION |
| $R_{overlap}$ [%] | 9.5 | 14.8 | 30.0 | 28.3 |
| $F$ [%] | 14.4 | 27.3 | 33.0 | 36.5 |

### 5.2 結果

表 1，2 に先行研究，及び，我々の以前の研究結果を示す．我々の以前の研究（表 2）では，NMF による分離を行わずに DNN で検出を行うことで，文献 [9] の結果を上回る性能を示している．しかしながら，NMF 単独で分離を行った場合は，NMF で求められた各クラスのアクティベーションの値の和で判定した方法では，$F$ 値が 14.4% と低く，満足な性能が得られていない．NMF の元々の分離精度が低いことが原因であると考えられる．

表 3，4 に提案手法の結果を示す．NMF による分離を行った後に DNN で検出を行う場合の $F$ 値は，表 2 の DNN のみの結果と比較して，絶対値で MFCC の場合に最大 10.3%，FBANK の場合に最大 4.4% 改善した．また，FUSION の結果では最大 41.3% の $F$ 値が得られた．提案手法で高い $F$ 値を示した条件は，各クラスのコードブックのコードブックサイズを 8，基底ベクトル間の距離を 0.45 以下の時にクラスとして割り当てた場合である．この時各クラス平均して 11 種類の基底ベクトルが割り当てられており，最も多い基底の種類で構成されていたクラスは laughter で基底数は 17 個，最も少ない基底で構成されていたクラスは alert, switch で基底数 5 個であった．実験から，誤ったアクティベーションの分配を防げたことで再現性が向上し検出精度の改善につながったと考えられる．また，近い基底ベクトルを一定数選択する手法では，表 4 の NMF-DNN の結果と比較して高い $F$ 値が得られていることが分かる．しかし，閾値処理で基底ベクトルにクラスを割り当てる結果と比較すると，閾値処理の方が高い性能が得られた．

## 6. まとめ

本稿では，[5] で行った NMF による音源分離に関する問題に対する調査と，音響イベントがオーバーラップする条件での検出精度の改善を検討した．従来法の NMF では音響イベントクラスごとの基底ベクトルの集合を利用し，音源分離を行っていた．しかしながら，基底ベクトルがクラスをまたいで類似性を示すため，十分な分離精度が得ら

**表 3** 共有化法 -F 値-
**Table 3** Sharing method -F-measure-

| 共有化法 - 閾値処理 - | | | |
|---|---|---|---|
| Threshold | cb_size | MFCC | FBANK |
| 0.35 | 4 | 36.7 | 35.9 |
| 0.40 | 4 | 36.2 | 35.4 |
| 0.45 | 4 | 35.7 | 33.8 |
| 0.50 | 4 | 36.3 | 36.4 |
| 0.40 | 8 | 29.6 | 35.3 |
| 0.45 | 8 | 34.2 | 37.4 |
| 0.50 | 8 | 32.0 | 36.0 |
| 0.55 | 8 | 37.6 | 36.5 |
| 共有化法 -近い基底ベクトルを一定数選択 - | | | |
| Select | cb_size | MFCC | FBANK |
| 4 | 4 | 30.7 | 31.2 |
| 4 | 8 | 30.7 | 31.2 |
| 5 | 4 | 29.6 | 34.0 |
| 5 | 8 | 28.1 | 24.8 |

**表 4** 従来法と提案手法の比較
**Table 4** Comparison between conventional and proposed methods

| Method | Feature | *R_overlap*[%] | *F*[%] |
|---|---|---|---|
| DNN[5] | FUSION | 28.3 | 36.5 |
| NMF-DNN[5] | MFCC | 8.8 | 10.4 |
| | FBANK | 23.1 | 20.6 |
| | FUSION | 32.1 | 17.2 |
| 共有化法 Threhold = 0.45 cb_size = 8 | MFCC | 31.4 | 34.2 |
| | FBANK | 48.7 | 37.4 |
| | FUSION | 39.9 | 41.3 |
| 共有化法 Threhold = 0.55 cb_size = 8 | MFCC | 33.4 | 37.6 |
| | FBANK | 36.3 | 36.5 |
| | FUSION | 39.6 | 40.5 |

れていないと考えられた．このことを確認するために，基底ベクトル間の距離を調査し，多くの音響イベントのコードブック（基底の集合）で類似した基底ベクトルを有していることが分かった．

この問題に対処するため，基底ベクトルをクラス間で共有する手法を提案し，NMF の後段に DNN を用いる手法において，[5] で良い性能を示した DNN のみの結果と比較して，$F$ 値の絶対値で 4.4% の改善が得られた（特徴量が FBANK の場合）．また，複数の音響モデルの出力を統合する FUSION が最も高い性能を示し，$F$ 値で 41.3% を得た．これは，先行研究で最も良かった [9] に対して絶対値で 20% の改善，[5] での最も良かった FUSION の結果に対して絶対値で 4.8% の改善となっており，提案手法の効果が示された．

今後は，全クラス共通の基底ベクトルに対応する各クラスの基底ベクトルを使用して音響イベントを再構成する手法，音源分離としてオートエンコーダを利用した多クラスへの音源分離，基底ベクトルの作り方としてケプストラムスムージングを利用すること等を検討していきたい．

## 参考文献

[1] A. Temko, et al., “Comparison of sequence discriminant support vector machines for acoustic event classification,” in *Proc. IEEE ICASSP 2006*, pp.721–724, 2006.
[2] S. Deng, et al., “Robust minimum statistic project coefficients feature for acoustic environment recognition,” in *Proc. IEEE ICASSP 2014*, pp.8282–8266, 2014
[3] J. Salmon and J. P. Bello, “Unsupervised feature learning for urban sound classification,” in *Proc. IEEE ICASSP 2015*, pp.171–175, 2015.
[4] R. Radhakrishnan, et al., “Audio analysys for surveillance applications,” in *Proc. 2005 IEEE WASPAA*, pp.158–161, 2005.
[5] 石川他, “オフィス環境における音響イベントのオーバーラップ区間でのイベント検出,” 日本音響学会 2016 年春季研究発表会講演論文集, 3-P-8, pp.169–172, 2016.
[6] D. Stowell, D. Giannoulis, E. Benetos, M. Lagrange, and M. D. Plumbley, “Detection and classification of acousitc scenes and events,” in *IEEE Trans. Multimedia*, vol.17, pp.1733–1746, 2014.
[7] N. Moreau, et al., “Data collection for the CHIL CLEAR 2007 evaluation campaign,” in *Proc. LREC’08*, pp.28–30, 2008.
[8] L. Vuegen, et al., “An MFCC - GMM approach apploach for event detection and classification,” in *IEEE AASP Challenge: Detection and Classification of Acoustic Scenes and Events*, 2013. [Online]. Available: `http://c4dm.eecs.qmul.ac.uk/sceneseventschallenge/abstracts/OL/VVK.pdf`
[9] J. F. Gemmeke, et al., “An exemplar-based NMF approach to audio event detection,” in *IEEE AASP Challenge: Detection and Classification of Acoustic Scenes and Events*, 2013. [Online]. Available: `http://c4dm.eecs.qmul.ac.uk/sceneseventschallenge/abstracts/OL/GVV.pdf`
[10] A. Diment, et al., “Sound event detection for office live and office synthetic AASP challenge,” in *IEEE AASP Challenge: Detection and Classification of Acoustic Scenes and Events*, 2013. [Online]. Available: `http://c4dm.eecs.qmul.ac.uk/sceneseventschallenge/abstracts/OL/DHV.pdf`
[11] X. Zhou, et al., “HMM-based acoustic event detection with AdaBoost feature selection,” in *Classification of Events, Activities and Relationships Evaluation and Workshop*, 2007.
[12] D. D. Lee, H. S. Seung, “Algorithms for non-negative matrix factorization”, in *NIPS*, pp.556–562, 2000.
[13] 仲野他, “音楽重畳音声の音声認識のための NMF による音楽除去の高速化及び VQ 手法の改善,” 日本音響学会 2012 年春季研究発表会講演論文集, 1-P-18, pp.165–168, 2012.
[14] T. Komatsu, et al., “Acoustic event detection based on non-negative matrix factorization with mixtures of local dictionaries and activation aggregation,” in *Proc. ICASSP 2016*, pp.2259–2263, 2016.
[15] S. F. Boll, “Suppression of acoustic noise in speech using spectral subtraction,” in *IEEE Trans. ASSP*, Vol.27, No.2, pp.113–120, 1979.
[16] 安齊他, “劣決定音源分離のための分離音声のケプストラムスムージング”, 日本音響学会誌, Vol.68, No.2, pp.74–85, 2012.