mIBP に基づくベイジアン半教師付き音響イベントダイアライゼーション*

◎大石康智 (NTT), 持橋大地, 松井知子 (統数研), 中野允裕,
亀岡弘和, 泉谷知範, 柏野邦夫 (NTT)

1 はじめに

膨大な音や映像のメディアデータが身の回りにあふれる中，これらのデータを自在に検索して活用するためには，付随するテキストデータに頼るだけではなく，それぞれの中身を表す情報を，音や映像自体から自動的に引き出す技術が必要不可欠である。音響信号から，人に認識されうる音の事象，例えば，話声や歌声などの音声をはじめ，動物の鳴き声，楽器音，環境音，効果音などを時間的に書き起こす音響イベントダイアライゼーションもその一つの技術であり (Fig. 1)，現在盛んに研究が行われている [1–4]。

本研究では，音響イベントダイアライゼーションにおける 2 つの課題に取り組む。1 つ目の課題は音響イベントの重なりが十分に考慮されなかった点である。多くの研究では，複数の重なった音響イベント（例えば，音楽+音声や雑音+音声）を一つの音響カテゴリとみなし，そこから抽出されたメル周波数ケプストラム係数 (MFCC) などの音響特徴量の分布をガウス混合モデル (GMM) や隠れマルコフモデル (HMM) で学習して，これらの音響カテゴリを検出した [5–7]。音環境によっては様々な音響イベントが複雑に重ね合わさっているため，音響カテゴリの数だけモデルを学習しなければならず，コストがかかる。文献 [8, 9] では，あらかじめ音響信号を複数のトラックに音源分離し，トラックごとに音響特徴量を抽出してイベントを検出した。ただし，音環境に合わせてトラック数を手動で調整する必要がある。

2 つ目の課題は，数千時間の書き起こしデータを用いる音声認識に比べ，音響イベントの音響的特徴を学習するためのラベル付データベースが少ない点である。さらに，新しい未知の音響イベントを検出したい場合，新たなラベル付データが必要となる。このようなデータスパースネスや未知の音響イベント検出問題に対処するために，教師なし，または半教師あり学習の枠組みを導入することは有望である [10–12]。

これらの課題に取り組むために，我々は文献 [13] において，非負値行列因子分解 (NMF) をベースとし，2 つのノンパラメトリックベイズ法を内包する，音響イベントダイアライゼーションのための生成モデルを提案した。これは，音響イベントが複数の音響要素（音声における音素，楽器における単音など）から構成されると想定し，NMF が教師なしでこれらの要素の重ね合わせを表現する。そして，ノンパラメトリックベイズ法により，モデルは潜在的に無限のパラメータを持ち，入力信号に応じて自律的に必要な分のパラメータを用いて，各音響要素を表現するように振る舞う（モデル選択問題が回避される）。評価実験より，音響要素の音響的特徴と，それらの各時刻におけるアクティベーション（on か off か）を，スライスサンプリングによって推論可能であることを確認した。

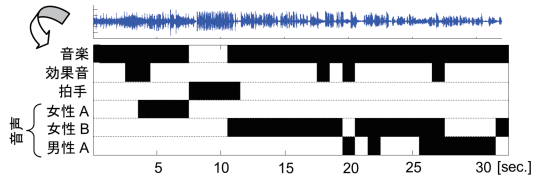

Fig. 1 音響イベントダイアライゼーション

本稿では，ベイズロジスティック回帰を利用して，音響要素の各時刻におけるアクティベーションから，その時刻に音響イベントラベルを付与するモデルへと拡張する。これは，あらかじめ一部の時刻にイベントラベルを付与し，大半の時刻にはラベルが付与されない状況の下で，半教師あり学習によって，時刻全体にラベル付けを行う枠組みである。すべてのモデルパラメータはギブスサンプラーによって推論される。実際の Podcast 音源を利用した評価実験では，提案モデルがベースライン手法よりも性能良く，イベントラベルを付与できることを示す。

2 非負値行列因子分解に基づく音響イベントダイアライゼーション

音響イベントの重なりを表現するために，可変基底型 NMF[14] を利用する。これは各音響要素の基底スペクトルが時間的に遷移するように拡張した NMF である。音響要素の音色を特徴付けるため，振幅スペクトログラムをメルフィルタバンク処理して得られた出力値行列 $\boldsymbol{Y} = (Y_{\omega,t})_{\Omega\times T} \in \mathbb{R}^{\geq 0,\Omega\times T}$ に NMF を適用する。ここで，$\omega = 1,\ldots,\Omega$ はメルフィルタのインデックス，$t = 1,\ldots,T$ は分析フレームのインデックス，$d = 1,\ldots,D$ は音響要素のインデックスを表す。このとき，各音響要素の基底スペクトルは時刻 t にある一つの状態 $Z_{d,t} \in \mathbb{N}$ を取ると見なし，

$$Y_{\omega,t} = \textstyle\sum_d C_{\omega,t,d},\ C_{\omega,t,d} \sim \mathrm{Poisson}(H_{\omega,d}^{(Z_{d,t})} U_{d,t}) \quad (1)$$

と表現する。これは Fig. 2 に示す集合で構成される生成モデルである。まず，$\boldsymbol{H}_d = (H_{\omega,d}^{(k)})_{\Omega\times K_d}$ は音響要素 d における K_d 個の基底スペクトルを表現する。$\boldsymbol{U} = (U_{d,t})_{D\times T}$ は，音響要素の on/off を表現する 2 値系列からなるアクティベーション集合を表現する。

*Bayesian Semi-supervised Audio Event Diarization based on Markov Indian Buffet Process. by OHISHI, Yasunori (NTT), MOCHIHASHI, Daichi, MATSUI, Tomoko (The Institute of Statistical Mathematics), NAKANO, Masahiro, KAMEOKA, Hirokazu, IZUMITANI, Tomonori, KASHINO, Kunio (NTT)

このとき，音響信号の音量はすべて基底スペクトルに含まれて表現される。そして，$\boldsymbol{C}_d = (C_{\omega,t,d})_{\Omega\times T}$ は音響要素 d のメルフィルタバンク出力値に相当し，$C_{\omega,t,d}$ は $H^{(k)}_{\omega,d}$, $U_{d,t}$, $Z_{d,t}$ をパラメータとするポアソン分布から生成される。最終的に音響信号はこれらの音響要素の和で表現される。各音響イベントはこれらの音響要素の組合せで表現されることを想定する。

文献 [13] では，$\boldsymbol{U}$ と $\boldsymbol{Z}$ の事前分布として，ノンパラメトリックベイズ法である Markov Indian Buffet Process (mIBP) と Chinese Restaurant Process (CRP) を導入した。これにより，生成モデルは潜在的に無限のパラメータを持ち，入力信号に応じて自律的に必要な分のパラメータを用いて，音響要素を表現するように振る舞う。結果として，モデル化の際の音響要素数の調整から解放されることが確認できた。

本稿では，各音響要素のアクティベーションから，ベイズロジスティック回帰に基づく半教師あり学習の下で，音響イベントラベルを生成する。これらの導入方法を順番に説明する。

2.1 mIBP から生成されるアクティベーション行列

mIBP を $\boldsymbol{U}$ の事前分布に導入することで，次の 2 つの特性を満たす [15]。(1) 行数（音響要素数）は任意に大きいサイズを想定する。(2) 各列（フレーム）の on/off は音響要素の継続時間（発音時間）を表現するよう，一次マルコフ連鎖に従って生成される。

具体的には，各音響要素ごとに状態遷移行列

$$\boldsymbol{W}^{(d)} = \begin{bmatrix} 1-a_d & a_d \\ 1-b_d & b_d \end{bmatrix} \tag{2}$$

を用意する。つまり，$0 \to 0$ の遷移確率は $1-a_d$，$0 \to 1$ は a_d，$1 \to 0$ は $1-b_d$，$1 \to 1$ は b_d とし，これらの状態遷移によって，0 と 1 からなる音響要素 d のアクティベーション系列が生成される。すべての音響要素のアクティベーション系列からなる行列 $\boldsymbol{U}$ は

$$p(\boldsymbol{U}|\boldsymbol{a},\boldsymbol{b}) = \prod_{d=1}^{D} (1-a_d)^{c_d^{00}} a_d^{c_d^{01}} (1-b_d)^{c_d^{10}} b_d^{c_d^{11}} \tag{3}$$

に従う。ここで，$\boldsymbol{a} = \{a_1,\ldots,a_D\}$, $\boldsymbol{b} = \{b_1,\ldots,b_D\}$ とし，$c_d^{00}, c_d^{01}, c_d^{10}, c_d^{11}$ はそれぞれの遷移回数を表す。さらに，a_d と b_d の事前分布をこれらの共役性から，$a_d \sim \mathrm{Beta}(\theta_\mathrm{a}/D, 1)$, $b_d \sim \mathrm{Beta}(\theta_\mathrm{b}^{(0)}, \theta_\mathrm{b}^{(1)})$ とする。そして，式 (3) を $D \to \infty$ としてモデルの複雑度を無限に拡張するために，stick breaking construction[16] を適用する。ここで，$\boldsymbol{a}$ を $a_{(1)} > a_{(2)} > \ldots > a_{(D)}$ のように順序付けると，$D \to \infty$ における $a_{(d)}$ の生成過程は $\nu_{(d)} \sim \mathrm{Beta}(\theta_\mathrm{a}, 1)$, $a_{(d)} = \prod_{d'=1}^{d} \nu_{(d')}$ に従う。一方，d 番目に大きい $a_{(d)}$ に対応する変数を $b_{(d)}$ とすると，$b_{(d)} \sim \mathrm{Beta}(\theta_\mathrm{b}^{(0)}, \theta_\mathrm{b}^{(1)})$ となる。mIBP では頻繁にアクティベートされる音響要素ほど，小さいインデックス d に割り当てられる。

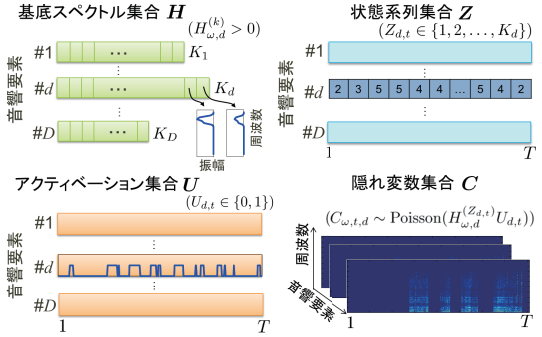

Fig. 2　音響要素を構成する基底スペクトル集合，その状態系列集合，アクティベーション集合

2.2 CRP から生成される時変なスペクトル特性

音響要素のスペクトルの時間変化を表現するために，基底スペクトルの状態数は固定でなく，音響信号から自動的に決定されることが望ましい。中野らは可変基底型 NMF にディリクレ過程を導入した無限状態スペクトルモデルを提案し，音楽音響信号を楽器の単音ごとに分解できることを示した [14]。この枠組みを音響イベントダイアライゼーションに適用する。

音響要素 d の状態系列 $Z_{d,1},\ldots,Z_{d,T}$ はそれぞれ，$1,\ldots,K_d$ の中の離散的な値をとる。状態数を $K_d \to \infty$ として，$\boldsymbol{Z}_{d,\backslash t} = \{Z_{d,1},\ldots,Z_{d,t-1},Z_{d,t+1},\ldots,Z_{d,T}\}$ が与えられたときの $Z_{d,t}$ の条件付き確率は

$$p(Z_{d,t}=k|\boldsymbol{Z}_{d,\backslash t},\theta_\beta^{(d)}) = \begin{cases} \dfrac{n^{(k)}_{d,\backslash t}}{(T-1+\theta_\beta^{(d)})} & (n^{(k)}_{d,\backslash t} > 0) \\ \dfrac{\theta_\beta^{(d)}}{(T-1+\theta_\beta^{(d)})} & (k = K_{\backslash t,+}+1) \end{cases} \tag{4}$$

と書ける。ここで，$n^{(k)}_{d,\backslash t}$ は $Z_{d,t'} = k$ $(Z_{d,t'} \in \boldsymbol{Z}_{d,\backslash t})$ を満たす t' の個数を表す。また，$K_{\backslash t,+}$ は $n^{(k)}_{d,\backslash t} > 0$ となるクラスの個数である。これが CRP と呼ばれ，ディリクレ過程の一構成法を与える。この過程は，各時刻に用いられる状態が，他の時刻に多く用いられている状態ほど使われやすくなる性質がある。また，新しい状態が用いられやすくなるか否かは正のパラメータ $\theta_\beta^{(d)}$ によって調整される。

2.3 ロジスティック回帰に基づくマルチラベリング

ベイズロジスティック回帰 [17] を利用して，アクティベーション行列から音響イベントラベルを生成する。ここでは，教師あり学習の下でトピックモデルを構成する sLDA[18] を参考にする。フレーム t における音響要素のアクティベーションをまとめて $\boldsymbol{U}_t = [U_{1,t},U_{2,t},\ldots,U_{D,t}]^\mathrm{T}$ と表現すると，音響イベント l のラベル $\boldsymbol{X}_l = \{X_{l,1},X_{l,2},\ldots,X_{l,T}\}$ $(X_{l,t} \in \{0,1\})$ の尤度関数は，

$$p(\boldsymbol{X}_l|\boldsymbol{U},\boldsymbol{w}_l) = \prod_{t=1}^{T} \exp(\boldsymbol{w}_l^\mathrm{T}\boldsymbol{U}_t X_{l,t}) \sigma(-\boldsymbol{w}_l^\mathrm{T}\boldsymbol{U}_t) \tag{5}$$

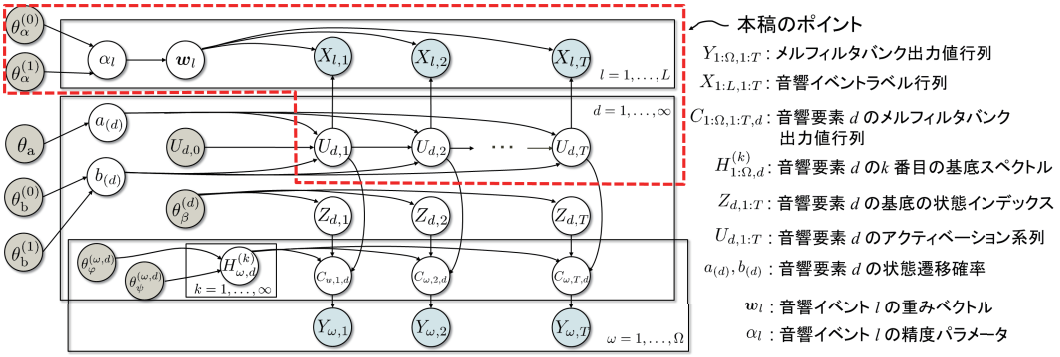

Fig. 3　音響イベントダイアライゼーションモデルのグラフィカル表現

と書ける。ここで，$\boldsymbol{w}_l = [w_{l,1}, w_{l,2}, \ldots, w_{l,D}]^{\mathrm{T}}$ は重みベクトルであり，$\sigma(\cdot)$ はシグモイド関数 $\sigma(a) = 1/(1+\exp(-a))$ を表す。さらに，単純な等方ガウス事前分布 $\boldsymbol{w}_l|\alpha_l \sim \mathcal{N}(\boldsymbol{0}, \alpha_l^{-1}\boldsymbol{I}_D)$ と共役超事前分布 $\alpha_l \sim \mathrm{Gamma}(\theta_\alpha^{(0)}, \theta_\alpha^{(1)})$ を導入する。$\theta_\alpha^{(0)}, \theta_\alpha^{(1)}$ は超パラメータであり，$\boldsymbol{I}_D$ は $D \times D$ の単位行列を表す。

音響イベントダイアライゼーションモデルのグラフィカル表現を Fig. 3 に示す。これは音響特徴行列 $\boldsymbol{Y}$ とラベル行列 $\boldsymbol{X}$ を共に生成するモデルといえる。ここで，基底スペクトル集合 $\boldsymbol{H}$ の事前分布として，ガンマ分布 $H_{\omega,d}^{(k)} \sim \mathrm{Gamma}(\theta_\varphi^{(\omega,d)}, \theta_\psi^{(\omega,d)})$ を利用する。また，$\theta_\alpha^{(0)}, \theta_\alpha^{(1)}, \theta_{\mathrm{a}}^{(0)}, \theta_{\mathrm{a}}^{(1)}, \theta_{\mathrm{b}}^{(0)}, \theta_{\mathrm{b}}^{(1)}, \theta_\beta^{(d)}, \theta_\varphi^{(\omega,d)}, \theta_\psi^{(\omega,d)}$ はすべて超パラメータである。

3　パラメータの推論

mIBP は，スライスサンプリングと動的計画法を組合せることでモデルパラメータを効率的に推論できるが [15]，本稿では，計算コストを小さくするよう，打ち切り stick breaking construction の下で（D を大きな値に固定して），パラメータを推論する。紙面の都合上，$\boldsymbol{U}$ の推論方法のみ記述する。$\boldsymbol{U}$ の各行は互いに独立と想定し，Forward-filtering backward-sampling アルゴリズム [19] を利用して，行ごとに $U_{d,1}, \ldots, U_{d,T}$ を推論する。まず，$t = 1, \ldots, T$ に対して，

$$
\begin{aligned}
&p(U_{d,t}|Y_{:,1:t}, C_{:,1:t,d}, Z_{d,:}, H_{:,d}^{(:)}, X_{:,1:t}) \\
&\propto p(Y_{:,t}, C_{:,t,d}|U_{d,t}, Z_{d,:}, H_{:,d}^{(:)}) \textstyle\prod_l p(X_{l,t}|U_{d,t}, \boldsymbol{U}_{\setminus d,t}, \boldsymbol{w}_l) \\
&\quad \textstyle\sum_{U_{d,t-1}} p(U_{d,t}|U_{d,t-1}) \\
&\qquad p(U_{d,t-1}|Y_{:,1:t-1}, C_{:,1:t-1,d}, Z_{d,:}, H_{:,d}^{(:)}, X_{:,1:t-1})
\end{aligned}
$$

を再帰的に計算する。ここで，$X_{l,t}$ の尤度関数はラベルが付与されるフレームに対してだけ計算される。

次に，$p(U_{d,T}|Y_{:,1:T}, C_{:,1:T,d}, Z_{d,:}, H_{:,d}^{(:)}, X_{:,1:T})$ から $U_{d,T}$ をサンプリングする。そして $t = T-1, \ldots, 1$ に対して，$U_{d,t+1}$ が与えられた下で，

$$
\begin{aligned}
&p(U_{d,t}|U_{d,t+1}, Y_{:,1:t}, C_{:,1:t,d}, Z_{d,:}, H_{:,d}^{(:)}, X_{:,1:t}) \\
&\propto p(U_{d,t}|Y_{:,1:t}, C_{:,1:t,d}, Z_{d,:}, H_{:,d}^{(:)}, X_{:,1:T}) p(U_{d,t+1}|U_{d,t})
\end{aligned}
$$

に従って，$U_{d,t}$ を後方から順番にサンプリングすれば $U_{d,1}, \ldots, U_{d,T}$ が求まる。これはラベルが付与される少量のフレームを手がかりとして $\boldsymbol{U}$ 全体を推論するため，半教師あり学習の枠組みと言える。その他のパラメータは，その事後確率に基づくギブスサンプリングによって推論されるが，方法の詳細は割愛する。

4　評価実験

英語学習用 Podcast の音響信号（計 35 分のうち，はじめの 5 分間）を用いて，提案法の性能を評価する。音響信号は 16kHz サンプリングで 16 ビット量子化されたものである。Fig. 4 の上図は，100ms ごとに手動でラベル付けされた音響イベントラベルを示す。黒色が音響イベントの on を，白色が off を表す。この音響信号は “音楽”，“効果音”，“電話のベル音”，5 名の “音声” からなる 8 種類の音響イベントを含む。音響信号は，フレームシフト長 100ms，フレーム長 100ms，ハニング窓を用いてフレームに分割され，短時間フーリエ変換によって振幅スペクトログラムに変換される。各フレームの振幅スペクトルをフィルタバンク処理して得られる出力値 (24 個の値) を音響特徴量行列とする ($\boldsymbol{Y}$ は 24×3000 の行列である)。

Fig. 4 に示すように，ラベル付与データとして 3 種類，(1)50 秒，(2)100 秒，(3)150 秒を用意した。そして，後半の 150 秒に対して, 提案法の性能を評価する。評価尺度として，イベントラベルの予測分布を利用する。$\boldsymbol{U}_t$ とラベル付与データが与えられた下で，フレーム t の音響イベントラベル $X_{l,t}$ の予測分布は，

$$
\begin{aligned}
&p(X_{l,t} = 1|\boldsymbol{U}_t, \boldsymbol{X}_{l_N}, \boldsymbol{U}_N) \simeq \sigma\Big(\mu_{l,t} \Big/ \sqrt{1 + \pi\sigma_{l,t}^2/8}\Big), \\
&\quad \mu_{l,t} = \boldsymbol{w}_{l_{\mathrm{MAP}}}^{\mathrm{T}} \boldsymbol{U}_t, \quad \sigma_{l,t}^2 = \boldsymbol{U}_t^{\mathrm{T}} \boldsymbol{\Sigma}_l \boldsymbol{U}_t, \\
&\boldsymbol{\Sigma}_l^{-1} = \alpha_l \boldsymbol{I}_D + \textstyle\sum_{t=t_1}^{t_N} X_{l,t}(1 - X_{l,t}) \boldsymbol{U}_t \boldsymbol{U}_t^{\mathrm{T}} \qquad (6)
\end{aligned}
$$

と近似する。ここで，ラベルが付与されたフレームを $t_1, \ldots, t_N$ とし，これらのフレームの音響要素のアクティベーションを $\boldsymbol{U}_N = \{\boldsymbol{U}_{t_1}, \boldsymbol{U}_{t_2}, \ldots, \boldsymbol{U}_{t_N}\}$，音響イベント l のラベルを $\boldsymbol{X}_{l_N} = \{X_{l,t_1}, X_{l,t_2}, \ldots, X_{l,t_N}\}$ とする。上式では，シグモイド関数とガウス分布のたたみ込み積分を近似するために，プロビット関数を利用した。$\boldsymbol{w}_{l_{\mathrm{MAP}}}$ は MAP（最大事後確率）解を表し，実際はラベルが付与されたフレームを利用して推論される [17]。式 (6) の $p(X_{l,1501} = 1|\cdot), p(X_{l,1502} = 1|\cdot), \ldots, p(X_{l,3000} = 1|\cdot)$ と正解ラベルとを比較する

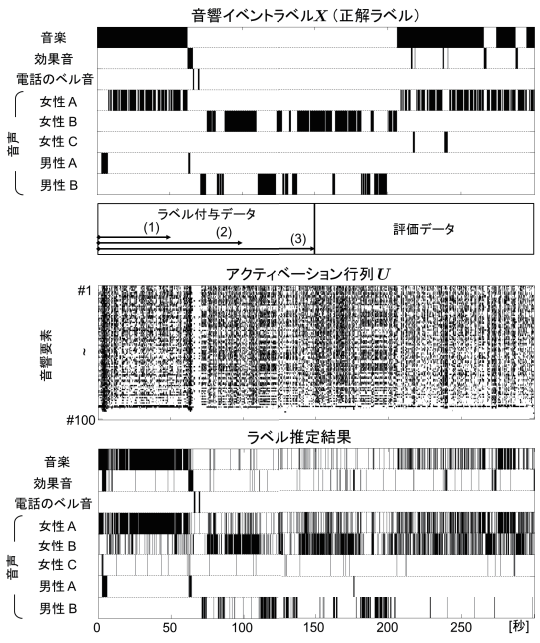

Fig. 4　手動で付与したイベントラベル（上図），アクティベーション行列（中図），ラベル推定結果（下図）

ことで計算される ROC 曲線の下側面積 (AUC) を評価尺度とする。パラメータの初期値を脚注に示す[1]。

Fig. 4 は，行列 $\boldsymbol{U}$ とラベル付与データ (3) を利用したときのラベル推定結果を示す。$\hat{X}_{l,t} = \mathbb{I}(p(X_{l,t} = 1|\cdot) > 0.5)$ の基準に基づいて，各フレームに音響イベントラベルを付与するか否かを判定した。ここで，$\mathbb{I}(A)$ はインジケータ関数であり，条件 A が真であれば 1 を，偽であれば 0 を返す関数である。アクティベーション行列のロジスティック回帰によって推定されるラベル付結果が，正解ラベルと全体的に近い結果が得られたことがわかる。また，女性話者の区別が難しいこともわかる。Tab. 1 では，ラベル付与データの長さに対するラベル付性能を評価する。“電話のベル音” と “男性 A の音声” は評価データに含まれていないため，これらは評価しない。データの長さが増えるにつれて，各イベントのラベル付性能を表す AUC の値が向上することがわかる。

Tab. 2 はベースライン手法と提案法の性能を比較する。どちらもラベル付与データ (3) を使用した結果である。ベースライン手法では，ラベル付与区間の音響特徴量（24 個のメルフィルタバンク出力値）の分布を GMM で学習し，評価データの各フレームに対して各音響イベントラベルの事後確率を計算する。そして，5 点移動平均フィルタを利用して，事後確率を平滑化し，平均 AUC の値を計算した。提案法はベースライン手法と同等以上の性能が得られることがわかった。ただし注目すべきは，(1) や (2) のようにラベル付与データが少ない場合，GMM は効果的にその特徴量の分布を学習できない。しかしながら，提案法は，たとえ少量のラベルデータであったとしても，ラベル付けされていない区間を活用することによって，全体のイベントラベルを推定できると考えられる。

[1]音響要素の打ち切り数は $D = 100$ とした。$\boldsymbol{C}, \boldsymbol{H}, \boldsymbol{U}$ の初期値は GaP-NMF[20] を利用して決定する。推定された $H_{1:\Omega,d}$ は d 番目の音響要素の初期基底スペクトルとする（初期状態数は $K_d = 1$ とする）。一方，$\boldsymbol{U}$ は，その中央値よりも大きい要素は 1，平均値よりも小さい要素は 0 に二値化して初期値とする。超パラメータの推論も可能であるが，本稿では簡単のため $\theta_{\mathrm{a}}=1$, $\theta_{\mathrm{b}}^{(0)}=10$, $\theta_{\mathrm{b}}^{(1)}=1$, $\theta_{\beta}^{(d)}=1$, $\theta_{\varphi}^{(\omega,d)}=1$, $\theta_{\psi}^{(\omega,d)}=1$ ($d = 1,\ldots,D, \omega = 1,\ldots,\Omega$), $\theta_{\alpha}^{(0)}=1$, $\theta_{\alpha}^{(1)}=1000$ に固定した。各パラメータの更新回数は 1000 回とした。

Table 1　ラベル付与データの長さに対する性能比較

ラベル付与データの長さ		50 秒	100 秒	150 秒
音楽		0.542	0.735	**0.769**
効果音		0.298	0.382	**0.683**
音声	女性 A	0.647	0.766	**0.769**
	女性 B	0.605	0.647	**0.744**
	女性 C	0.437	0.466	**0.813**
	男性 B	0.560	0.935	**0.962**
平均 AUC		0.514	0.655	**0.790**

Table 2　提案法とベースライン法 (GMM) の比較

	提案法	ベースライン法 (GMM)
平均 AUC	**0.790**	0.734

5　まとめと今後の展開

音響信号の中で複雑に重ね合わさった様々な音響イベントを時間的に書き起こすために，これまで提案した無限生成モデルにベイズロジスティック回帰を導入して，音響イベントのマルチラベリング方法を提案した。アクティベーション行列を更にスパースにする工夫を施すこと，大規模データベースを利用して提案法の妥当性を評価すること，パラメータ推論における計算コストを削減することが今後の課題である。

参考文献

[1] A. Temko *et al.*, *Pattern Recogn. Lett.*, vol.30, no.14, pp. 1281–1288, 2009.
[2] T. Butko *et al.*, in *Proc. ICASSP 2011*.
[3] 佐々木ほか, 情処研報, 2011–SLP87–6, 2011.
[4] 井本ほか, 音講論集, 2-Q-32，pp. 975–976, 2012.
[5] K. Sumi *et al.*, in *Proc. INTERSPEECH 2009*.
[6] A. Mesaros *et al.*, in *Proc. EUSIPCO 2010*.
[7] A. Misra, in *Proc. INTERSPEECH 2012*.
[8] T. Heittola *et al.*, in *Proc. CHiME 2011*.
[9] M. Espi *et al.*, in *Proc. ICASSP 2012*.
[10] Z. Zhang *et al.*, in *Proc. ICASSP 2012*.
[11] B. Byun *et al.*, in *Proc. INTERSPEECH 2012*.
[12] S. Chaudhuri *et al.*, in *Proc. INTERSPEECH 2012*.
[13] 大石ほか, 音講論集, 1-P-22，pp. 775–778, 2012.
[14] M. Nakano *et al.*, in *Proc. ICASSP 2011*.
[15] J. V. Gael *et al.*, in *Proc. NIPS 2008*.
[16] Y. W. Teh *et al.*, in *Proc. NIPS 2007*.
[17] D. Spiegelhalter *et al.*, Networks, vol. 20, pp. 579–605, 1990.
[18] D. M. Blei *et al.*, in *Proc. NIPS 2007*.
[19] S. L. Scott, *JASA*, vol.97, pp. 337–351, 2002.
[20] M. Hoffman *et al.*, in *Proc. ICML 2010*.