人工知能学会研究会資料
JSAI Technical Report
SIG-Challenge-046-05 (11/09)

空間情報を用いた鳥の歌分析

小島 諒介 [1], 杉山　治 [2], 干場　功太郎 [1], 鈴木　麗璽 [2], 中臺 一博 [1,3]

Ryosuke KOJIMA[1], Osamu SUGIYAMA[2], Kotaro HOSHIBA[1], Reiji SUZUKI[3], Kazuhiro NAKADAI[1,4]

1. 東京工業大学, 2. 京都大学, 3. 名古屋大学,
4. （株）ホンダ・リサーチ・インスティチュート・ジャパン

1. Graduate School of Information Science and Engineering, Tokyo Institute of Technology,
2. Graduate School of Information Science, Nagoya University,
3. Honda Research Institute Japan Co., Ltd.

kojima@cyb.mei.titech.ac.jp, sugiyama@kuhp.kyoto-u.ac.jp,
hoshiba@cyb.mei.titech.ac.jp, reiji@nagoya-u.jp, nakadai@jp.honda-ri.com

Abstract

本稿では鳥の歌の自動分析システムの構築を目的とし，そのための空間情報を考慮した音源同定モデル提案する．鳥の歌を分析し，いつどこで何の鳥が鳴いているかを自動的に発見するシステムは野鳥研究や動物行動学研究において，観測の効率化･大規模化といった点から期待されている．しかし，このようなシステムは，異なる地点で同時に複数の野鳥が鳴いているといった状況に対応する必要があるといった困難がある．我々は，音源検出･定位･分離･同定といった，ロボット聴覚の技術を利用することで，鳥の歌分析フレームワークの構築を目指している．本稿では，位置情報に注目することで，これらの統合を可能にする確率モデル　Spatial-Cue-Based Probabilistic Model (SCBPM) を提案する．さらに，提案モデルを用いたアノテーション補助システムを構築し，実データを用いた評価実験を行った．その結果，鳥の歌識別のタスクにおいて，従来法より最大で 5%の識別率向上が確認できた．また，提案モデルが有効でない状況においても，従来法と変わらない性能が達成できることが確認できた．

1　はじめに

音環境理解は，環境中の音に注目することで，画像などの情報からでは得ることが難しい情報や障害物が多くオクルージョンが問題となるような環境で大きな手助けとなることから，環境モニタリングやロボットなどの分野で，注目されている．音環境理解における主な困難の一つは，環境中にある混合音の中から有用な情報を抽出することである．我々は音環境から 5W1H 情報 (When, Where, What, Who, Why, How) を抽出することを音環境理解と定義し，研究を行っている．これら 5W1H のうち，はじめの 4W は音環境中でのイベントに関する重要な情報であるため，特に，音源検出（When），音源定位（Where），音源分離（What），話者認識・音源同定（Who）として取り組まれている．これまでの 4W 情報抽出はそれぞれの情報に関して個別に抽出するものであり， 4W すべての情報を抽出する単純な手法はこれらをカスケード的に実行することである．つまり，収録した音から音源を検出･定位し，複数音源がある場合にはそれぞれ分離し，同定を行う手法である．このアプローチには二つの欠点がある．一つは，処理が多段になるため，誤差が蓄積してしまうことである．もう一つは，音源の位置と種類の関係性など 4W 間の相互依存を考慮していない点である．こういった課題を扱うため，音源定位と音源分離を同時に行う手法として BNP-MAP 法 [Otsuka 14] が提案されている．また，環境理解以外でも相互に依存した情報を扱うモデルは研究されており，例えば，関係データを扱う Stochastic Block Model （SBM） はネットワーク分析や関係クラスタリングで利用されている [Holland 83]．関係クラスタリングは，二つ以上の対象の関係を用いてクラスタリングする手法であり，同時に複数の音が存在する場合の相互依存関係を扱う問題とも関係が深い．

また，定位と同定の相互依存性，特に，いつどこで誰が話しているかという問題は話者ダイアライゼーションとして広く研究されている．例えば，マイクロホンアレイを用いて到達時間差を特徴量として補助的に利用することで話者のセグメンテーション精度が向上できることが知られている [Pardo 06]．これらの話者ダイアライゼーションの手法は会議室などの室内で，人の声を対象にしている．

本稿では，定位と同定の相互依存性を取り扱うために，鳥の歌分析を対象にした．鳥の歌の音源同定は音源の位置情報と深く結びついたタスクの一つである．なぜならば，野鳥は自分の縄張りを持っており，近くの個体は同一の個体であるといったこれらの情報を考慮しつつ，障害物の

多い森林などの環境において，音源同定を行うタスクは環境モニタリングとしても挑戦的なタスクの一つである．

また，鳥研究においても，鳥の歌は縄張りの主張や求愛などの役割を持つことから興味深い対象の一つである [Catchpole 03]．鳥の歌研究には二つの方法があり，一つはよくコントロールされた環境での観測で，もう一つは実際のフィールドにおける観測である．前者は条件をコントロールした鳥の歌の性質の研究が可能であり，不要な条件を排除して実験を行うことができる．本稿では後者に主眼をおいており，この方法では，実際の環境下で鳥がどのように活動しているかを直接観測することができるため，注目されている方法である．

鳥の歌の同定は機械学習分野でも注目されており，いくつかのコンペティションで取り上げられており，様々な手法が提案されている [Briggs 13, Goëau 16]．これらのコンペティションで用いられているデータは 1 もしくは 2 チャンネルのマイクロフォンで収録された音である．3 つ以上のマイクロフォンで構成されたマイクロフォンアレイを用いた収録は，定位や分離の性能向上のために鳥の歌に利用できることが報告されている [Suzuki 16]．音源同定についても，定位や分離と深く結びついているため，マイクロフォンアレイを用いることが有効であると期待できるが，その性能評価については十分されていない．

先行研究　[小島 15]　として，音源位置を考慮した同定モデルが提案されている．この手法は，音源同定を行うための音響モデルを分離音ごとに独立なモデルとして学習した後，音源位置を考慮した関数を用いてそれらのモデルを結合するというアプローチを取っていた．しかし，このアプローチは，分離音ごとに独立に学習を行うため，音源位置を音響モデルに反映できなかった．これを解決するために，EM アルゴリズムによる相互依存の学習を可能にしたモデルも提案されている [小島 16]．しかし，この手法では一部のパラメータに関するパラメータ学習を行っておらず，また，[Kojima 16] では実験も一環境のみの限定的なものであった．本稿では，これらのパラメータについての学習を可能にし，異なる環境での実験を行い，モデルを評価した．

2　課題とアプローチ

我々は，マイクロフォンアレイを用いて収録した鳥の歌を自動的に分析し，鳥の研究を補助するシステムの構築を目指す．本稿では音源検出・定位・分離・同定をターゲットにし，音源同定においてこれらを統合するモデル Spatial-Cue-Based Probabilistic Model (SCBPM) を提案する．

モデルの構築にあり以下の 3 つの課題が挙げられる．

1. 音源定位情報を用いた音源同定のモデル化
2. 音源定位情報を考慮したモデルパラメータの学習法
3. 部分的なアノテーションを考慮した学習法

一つ目の課題は，野鳥観測を行うような森林や屋外では，木々などの障害物や地形による影響により，音源分離が十分な性能を発揮できないことに由来している．実際に分離音に他音源からの音（同時に鳴いている他の鳥の歌など）が漏れてしまい，識別がうまくいかない場合が多く存在した．一方で，MUSIC 法 [Schmidt 86] やその拡張手法は，屋外の騒音環境や野鳥が住処にするような森林であっても，ある程度の音源の到来方向を推定できることが報告されている [松林 15, Uemura 15]．そこで，分離音からの情報を補助するために，到来方向を含めた確率モデルを提案する．提案モデルは，分離音に関する確率分布と到来方向に関する確率分布の 2 つの分布から構成される．分離音に関する分布としては，音源同定のモデルとしてよく知られている 混合ガウス分布 (GMM: Gaussian Mixture Model) を用い，到来方向に関する分布には方向統計学でよく用いられる von Mises 分布を用いる．

二つ目は，構築したモデルのパラメータの学習をいかに行うかという課題である．提案モデルでは，同時刻の分離音は定位情報を通して相互に依存している．提案モデルは，隠れ変数を持つ確率モデルであるため，Expectation Maximization (EM) アルゴリズムで学習することが考えられるが，分離音間の相互依存性を考慮して学習する必要がある．そこで，本稿では新たに提案モデル上での EM アルゴリズムを導出し，この課題の解決を図る．

三つ目にデータのアノテーションに関する課題である．提案モデルの学習のためにはアノテーションが必要であるが，すべてのデータに対し人手でこれを行うは労力がかかる．そこで，一部のデータに人手でアノテーションをし，残りを推定するのが妥当である．すると，アノテーションの推定は大量のアノテーションされていないデータと少数のアノテーションデータから識別器を構成する問題となり，これは半教師あり学習問題として知られている．特に，上述の EM アルゴリズムでパラメータ学習をする場合には，アノテーションされていないデータを欠損値とみなすことにより，自然に半教師あり学習へと拡張することができる [Nigam 00]．

3　カスケード法

ここでは，収録した音から音源を検出・定位・分離・同定のための単純なカスケード手法を説明する．図 1 は，マイクロフォンアレイで収録した音を検出・定位・分離・同定の順に実行して分析するカスケード手法を示している．この手法ではそれぞりの処理は独立して行われる．この節の以降ではこれらの処理についての詳細を述べる．

音源検出・定位では，音源数，音源の位置，音源がアク

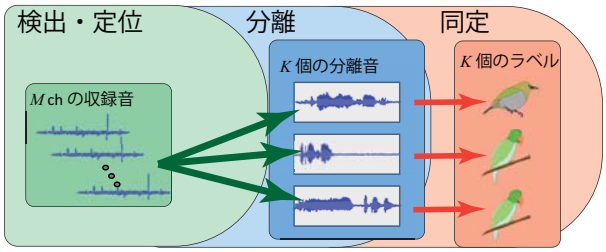

図 1: 音源検出・定位・分離・同定のカスケード法

ティブになっている区間の推定を行う．これらは MUSIC 法を用いて達成される．MUSIC 法ではそれぞれの方向 θ ごとに MUSIC パワー $P(\theta)$ と呼ばれる信号部分空間のパワーを計算する．

$$P(\theta) = \frac{||\mathbf{a}^H(\theta)\mathbf{a}(\theta)||}{\sum_{i=L+1}^{M} ||\mathbf{a}^H(\theta)\mathbf{e}_i||}$$

ただし，$\mathbf{a}(\theta)$ は予め計測しておくステアリングベクトル，$\mathbf{e}_i$ はマイクロフォンアレイのチャネル間の相関行列を固有値展開し，固有値の大きいもの順に並べた固有ベクトルとそのインデックス i である．L は MUSIC 法のパラメータであり，音源候補の数，M はマイクロフォンの数である．MUSIC 法では信号部分空間が雑音部分空間と直交することを利用しており，音源のある方向で $P(\theta)$ は高い値となる．この時，スレッショルドを超えるもののみを検出し，そのピーク値を見つけることにより，音源の方向を推定する．また，各時間フレームごとに方向を推定し，フレーム間での方向の差がある閾値以下であれば同一の音源とみなし，時間的に音源を追跡することで，音源がアクティブになっている区間を推定することができる．これらのスレッショルドの値も MUSIC 法のパラメータである．

音源分離では混合音から目的の音源の音を抽出する．音源分離の手法としてはビームフォーミング法などがよく知られている．GHDSS(geometric highorder decorrelation-based source separation) 法は分離音間の高次無相関性を考慮してビームフォーミングを拡張した手法であり，方向性ノイズに強い．我々は，屋外の様々な方向性ノイズから野鳥の歌を分離するため，GHDSS 法を用いた．

音源同定では GMM を用いた音響モデルを利用する．このモデルでは，一つの音源クラスは複数のサブクラスを持ち，各時刻でのある音源からの音は，それらの中から確率的に選択されると仮定する．より具体的には，周波数スペクトルから計算した音響特徴量が多変量ガウス分布に従うとし，一つの音源クラスであってもサブクラスの数だけ周波数スペクトルのパターンを表現できる．本研究では，周波数スペクトルを時間方向に主成分分析（PCA）により次元圧縮をしたベクトルを音響特徴量とした．

このようにモデル化すると，入力の音響特徴量 $\mathbf{x}$ は以

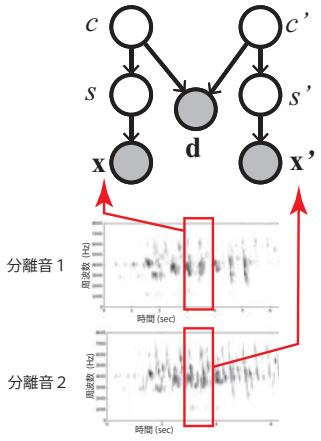

図 2: 提案モデルのベイジアンネットワーク表現（同時刻の音源が二つの場合）:観測変数 $\mathbf{x}$，$\mathbf{x}'$ は同時刻の別の分離音から計算される音響特徴量であり，観測変数 $\mathbf{d}$ はそれらの到来方向のベクトル．

下の混合ガウス分布に従うこととなる．

$$p(\mathbf{x}, s_{cj}, c) = \mathcal{N}_{cj}(\mathbf{x})p(s_{cj} \mid C = c)p(C = c)$$

ただし，s_{cj} は音源 c の j 番目のサブクラス ($\sum_j p(s_{cj} \mid C = c) = 1$)，$\mathcal{N}(\cdot)$ は多変量ガウス分布である．ここで，音源の種類 C を確率変数とおき，アノテーション済みデータの場合には固定値とすることで，EM アルゴリズムにより半教師あり学習を行う．また，モデルの構築後は MAP(Maximum A Postriori) 推定により，音源の同定を行うことができる．

4 提案法

前節の音源同定のモデルでは音源位置に関することは考慮されていなかった．ここでは音源位置を利用できるようにモデルを拡張し（4.1 節），そのモデルのパラメータ学習について述べる（4.2 節）．

4.1 提案モデル

前述の GMM による音響モデルでは分離音ごと独立にモデル化していた．したがって，時刻 t，分離音 k_t ごとに独立であった．提案モデルは各分離音間の依存性を導入し，これを拡張したものになっている．提案モデルのベイジアンネットワーク表現を図 2 に示す．ここで，時刻 t の音源 k_t の方向 $\mathbf{d}_t = d_{t,1}, d_{t,2}, ..., d_{t,k_t}, ..., d_{t,K_t}$ $(0 \leq d_{t,k_t} < 2\pi, 1 \leq k_t \leq K_t)$ は MUSIC 法を用いて計算できる．この時，3 節で述べたように時刻 t の音源数 K_t も決定する．また，各分離音の音響特徴量 $\mathbf{x}_{k_t}$ は GHDSS 法を用いて計算できる．そのため，図 2 ではこれらを観測として表現している．以下では，時刻 t は簡単のために

具体的には，以下のように記述される．

$$p(\mathbf{x},\mathbf{d},\mathbf{s},\mathbf{c}) = p(\mathbf{d}\mid\mathbf{c})\prod_{k=1}^{K}\mathcal{N}_{s_{c_k}}(\mathbf{x}_k)p(s_{c_k}\mid c_k)p(c_k) \tag{1}$$

$$p(\mathbf{d}\mid\mathbf{c}) = \prod_{c_i=c_j,i\neq j} p(d_i,d_j\mid c_i=c_j) \prod_{c_i\neq c_j,i\neq j} p(d_i,d_j\mid c_i\neq c_j) \tag{2}$$

$$p(d_i,d_j\mid c_i=c_j) = f(d_i-d_j;\kappa_1) \tag{3}$$

$$p(d_i,d_j\mid c_i\neq c_j) = f(d_i-d_j+\pi;\kappa_2) \tag{4}$$

$$f(d;\kappa) = \frac{\exp(\kappa\cos(d))}{2\pi I_0(\kappa)} \tag{5}$$

ただし， $f(d;\kappa)$ は von Mises 分布であり，$I_0(\kappa)$ は 0 次の変形ベッセル関数である．また，κ は分布の集中度を表すパラメータである ($\kappa \geq 0$)．式 (3) で定義される $p(d_i,d_j\mid c_i=c_j)$ に注目すると，この確率値は二つの音源の位置が近く，かつ，二つの音源が同じクラスに属している時に高い値をとる．一方，式 (4) で定義される $p(d_i,d_j\mid c_i\neq c_j)$ に注目すると，この確率値は二つの音源の位置が遠く，かつ，二つの音源が異なるクラスに属している時に高い値をとる．同時刻に二つ以上の音源がある場合 ($K_t>2$) を考慮するために，$p(\mathbf{d}\mid\mathbf{c})$ は式 (2) のようにすべての音源間の組み合わせによって定義されている．

このモデルを用いて音源のクラスを推定するときには，同時刻の他の音源のクラスを考慮する必要がある．したがって，音源クラスの MAP 推定は以下のようになる．

$$\mathbf{c}^* = \underset{\mathbf{c}}{\operatorname{argmax}}\, p(\mathbf{x}\mid\mathbf{c})p(\mathbf{d}\mid\mathbf{c})p(\mathbf{c}) \tag{6}$$

4.2 提案モデルのパラメータ学習

本節では提案モデルにおける EM アルゴリズムについて説明する．まず，音響特徴量 $\mathbf{x}$ に対応するクラス c が与えられた場合，つまり教師あり学習の場合，図 2 より，ベイジアンネットワークの性質から，c は他の音源クラス c' と独立に計算することができ，従来の GMM による音響モデルのパラメータ学習と同様に学習を行うことができる．しかし，部分的なアノテーションの場合，つまり，半教師あり学習を行う場合には，c と c' が独立とはならず，$\mathbf{x}$ ごとに独立に学習することもできない．以下ではクラス c, c' がアノテーションされていない場合について説明する．

EM アルゴリズムにおいては，データセット中のサブクラス s の出現確率の期待値を計算する必要があり，以下のように表現される．

$$N_s = \sum_t\sum_{k_t} p(s_{t,k_t}=s,\mathbf{X},\mathbf{d}) \tag{7}$$

ただし，s_{t,k_t} は時刻 t の音源 k_t に関するサブクラスを表す確率変数とし，$\mathbf{X}$ は時刻 t の音響特徴量全ての集合とする．$p(s_{t,k_t}=s,\mathbf{X},\mathbf{d})$ は提案モデル上で計算することができる． ただし，図 2 より，ベイジアンネットワークの性質から，$p(s_{t,k_t}=s,\mathbf{X},\mathbf{d})$ は音源 k_t だけでなく，時刻 t におけるそのほかの音源と独立に決定することはできない．具体的に，$p(s_{t,k_t}=s,\mathbf{X},\mathbf{d})$ を計算する方法を示す．まず，ここでは簡単のため時刻 t に 2 つの音源のみがあるとして，それぞれ，音源 k_t，k'_t，音響特徴量 $\mathbf{x}$, と $\mathbf{x}'$（$\mathbf{X}=\{\mathbf{x},\mathbf{x}'\}$），音源方向 d と d' が与えられた場合を考える．すると，音源 k_t のサブクラス s に関する確率 $p(s,\mathbf{X},\mathbf{d})$ は以下のように表現される．

$$p(s,\mathbf{X},\mathbf{d}) = \sum_{c,c'} p(d,d'|c,c')p(\mathbf{x}|s)p(s|c)p(c)p(\mathbf{x}'|c')p(c') \tag{8}$$

ただし，$p(\mathbf{x}'|c')=\sum_{s'}p(\mathbf{x}'|s')p(s'|c')p(c')$ とする．二つ以上の音源がある場合，$p(\mathbf{x}|c)$ を何度も計算する必要があるため，予め依存しているフレーム全てに対して $p(\mathbf{x}|c)$ を計算し，テーブルを作っておくことで，高速に計算することができる．また， $p(s|\mathbf{x})$ は s に関する多変量ガウス分布となり，それ以外の確率は定義より与えられる．

von Mises 分布のパラメータ κ_1, κ_2 についても EM アルゴリズムで決定可能である．これらのパラメータの更新式は，混合 von Mises 分布のパラメータ推定 [Banerjee 05] と類似したものになる．κ_1 の更新値 $\kappa_1^{(new)}$ を決める式は以下のようになる．

$$U_{c=c'} = \sum_t\sum_{x,x'}\sum_{c=c'} cos(d-d')p(c\mid d,x_t)p(c'\mid d',x'_t)$$

$$V_{c=c'} = \sum_t\sum_{x,x'}\sum_{c=c'} p(c\mid d,x_t)p(c'\mid d',x'_t)$$

$$\kappa_1^{(new)} = A^{-1}\left(\frac{U_{c=c'}}{V_{c=c'}}\right)$$

ここで，$U_{c=c'}$ と $V_{c=c'}$ はモデル上ですべての同時刻の音イベントについて $c=c'$, $x=x'$ 可能な組み合わせを計算する．$A(x)$ は以下のように定義される．

$$A(x) = \frac{I_1(x)}{I_0(x)}$$

ただし，$I_0(x)$ と $I_1(x)$ はそれぞれ 0 次と 1 次の変形ベッセル関数とする．$A(x)$ の逆関数 $A^{-1}(x)$ は以下の近似式で計算可能である [Sra 12].

$$A^{-1}(x) \approx \frac{x(2-x^2)}{1-x^2}$$

κ_2 の更新式も $c=c'$ を $c\neq c'$ とすることにより，同様に計算可能である．

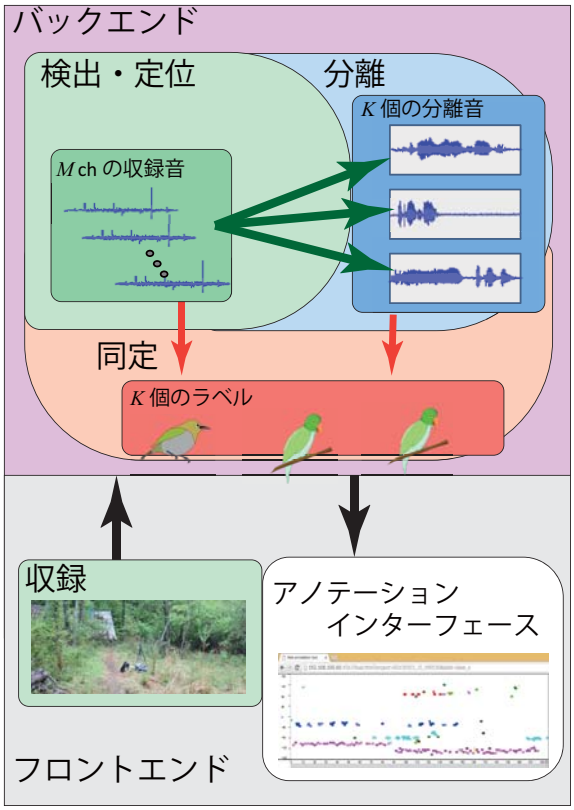

図 3: 提案モデルを用いたプロトタイプシステム

5 プロトタイプシステム

提案モデルを評価するためのプロトタイプシステムを構築した．このシステムは図 3 に示したようにバックエンドとフロントエンドからなる．バックエンドは音源検出・定位・分離・同定を行い．音源検出・定位（MUSIC 法）・分離（GHDSS 法）には HARK[1] [奥乃 10] をもちいた．フロントエンドは，収録とアノテーションからなり，マイクロフォンアレイには 7 つのマイクロフォンからなる Microcone [2] を用いた．収録は図 4 に示すように，三脚に固定し，ラップトップ PC に接続して収録を行った．アノテーションインターフェースでは，図 4　に示すように縦軸に音源方向，横軸に時間，色がアノテーションしたラベルを表すものとした．本システムでは一部にラベルをつけると残りのラベルに自動的にラベルの候補が半透明で着色されるようになっており，アノテータは予測されたラベルが正しいかどうかを確認・修正することができるようになっている．

6 評価実験

SCBPM を評価するために，データセット (A)(B) の二つの異なるデータセットを用意した．

[1]Honda Research Institute Japan Audition for Robots with Kyoto University

[2]http://www.dev-audio.com/

図 4: 収録システム

表 1: データセット (A): ラベルとイベントの数，色は図 6 と対応．ヒヨドリ（A）とヒヨドリ（B）は異なる個体のヒヨドリで，歌い方の特徴が異なるため，別ラベルとした．

ラベル	イベント数	色
キビタキ	5	red
メジロ	7	cyan
ヒヨドリ (A)	12	blue
ヒヨドリ (B)	13	yellow
その他	17	green

データセット (A) は 2013 年 5 月 5 日の朝（晴天）に愛知県都市部の公園で収録した [Suzuki 15]．この 1 分間のデータを対象として，3 節 で述べた MUSIC 法を適用することで， 54 のイベントが抽出された（図 6）．この時，MUSIC 法のパラメータは，一つのイベントが鳥の歌の一フレーズになるように選択した．これらのイベントに対し，分離音を手掛かりに，表 1 のラベルを用いて，人手でアノテーションを行ったものを正解として，データセットを作成した．データセット (B) は 2013 年 5 月 9 日の朝（晴天）にアメリカのカリフォルニア州の針葉樹とナラの混合樹林で収録した．収録した 4 分間に 140 のイベントを抽出した（図 7 ）．データセットは表 2 に示した 8 つのラベルを用いた．

図 5 と 図 8 はそれぞれデータセット (A)(B) に関して，カスケード法と SCBPM を用いた提案システムでの結果を比較したものである．これは，収録時間を 10 等分し，ラベルありの区間とラベルなしの区間に分けて，ラベルなしの区間のイベントを予測した際の正答率で評価を行った．これらのグラフの横軸（アノテーション率）は全体のうちのラベルありの区間の割合である．

ここで用いた，音響特徴量・モデルパラメータは以下のように設定した．まず，16-bit 16kHz サンプリングの分離音から窓幅 80（オーバーラップは 40 サンプル）として

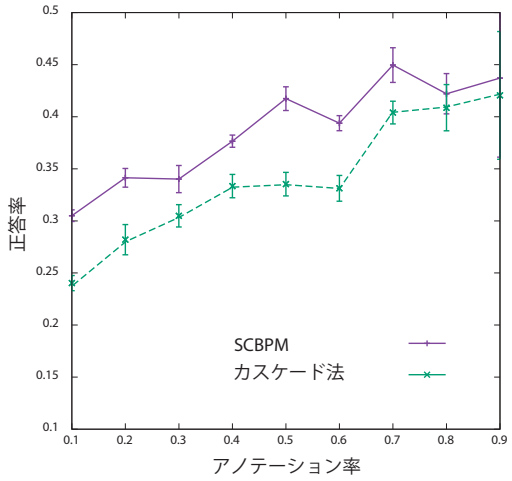

図 5: データセット (A) に関する正答率の比較

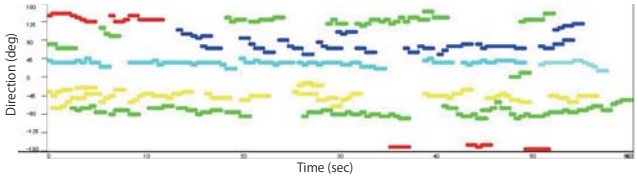

図 6: データセット (A) のイベント．ラベルは表 1 と対応．

STFT により周波数スペクトルを計算した[3]．さらに，10 フレームのステップ幅で 100 フレーム分を 1 ブロック 4100 次元のベクトルとみなし，PCA により 32 次元へと次元削減し，利用した．また，この 1 ブロックごとにクラスを推定し，最終的にイベント内の全てのブロックの多数決によってそのイベントのクラスを決定した．GMM の混合数は 30 とした．半教師あり学習ではラベルありデータの重みをとラベルなしデータの重みを設定した [Nigam 00]．ラベルありデータの重みを 1.0 とし，ラベルなしデータの重みはデータセット (A) に関しては 0.1，データセット (B) に関しては 0.001 とした．

図 5 と 図 8 からこの図から全てのアノテーション率において，SCBPM が良い正答率であることがわかる．また，データセット (B) に比べデータセット (A) の正答率がカスケード法，SCBPM ともに低いことがわかる．これはデータセット (A) のほうが同時に歌っている鳥の数が多く（図 6），完全な分離が困難であったためと考えられる．提案法は，特に，このような分離の性能が低い場合において，比較的良い性能であり，最大でおよそ 5%の差が確認できた．一方，データセット (B) では同時に歌っている鳥の数が少なく（図 7），いくつかのイベントについては他の鳥の方向と重複してしまっている．このような状況では提案法があまりよく働かないが，そのような場合

[3]これらのパラメータは人間が．鳥の歌を識別する際によく利用するパラメータセットの一つと同じものを用いた．

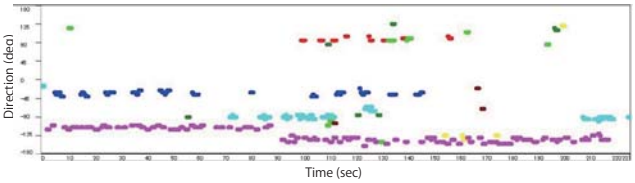

図 7: データセット (B) のイベント．ラベルは表 2 と対応．

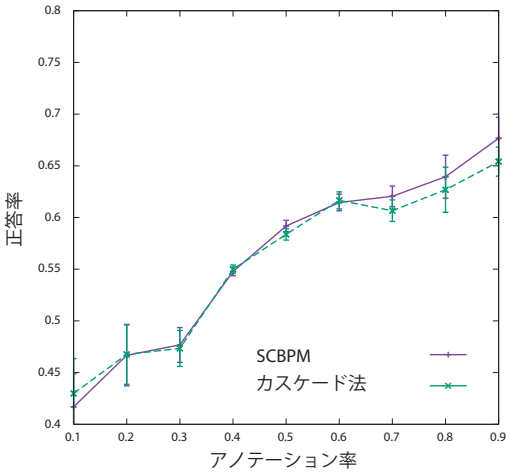

図 8: データセット (B) に関する正答率の比較

でも従来法と同等の性能が確認できた．

7 まとめ

定位情報を利用した音源同定モデルである SCBPM を提案し，モデルパラメータを分離音間の依存性を考慮しつつ学習する EM アルゴリズムを導入した．また，そのモデルを用いて音源検出・定位・分離・同定を行うアノテーションシステムのプロトタイプを構築し，実際のフィールドにおける鳥の歌のデータを用いて提案法の有効性を示した．今後は，より大規模なデータを用いたパラメータ学習を行い，音源同定の正答率を上昇させることや本システムのスケーラビリティを評価することが課題である．また，専門家に鳥の歌の分析システムを利用してもらい，評価を得ることで新たな課題の抽出を行いたいと考えている．

謝辞

本研究におけるフィールドでの収録手順や現状について助言を頂いた名古屋大学の松林志保氏に感謝する．
本研究は、JSPS 科研費 24220006, 16H02884, 16K00294 および，JST ImPACT タフロボティクスチャレンジの助成をうけた．

参考文献

[Banerjee 05] Banerjee, A., Dhillon, I. S., Ghosh, J., and Sra, S.: Clustering on the unit hypersphere using von Mises-Fisher distributions, *Journal of Machine Learning Research*, Vol. 6, No. Sep, pp. 1345–1382 (2005)

表 2: データセット (B): ラベルとイベントの数．色は図 7 と対応．

ラベル	イベント数	色
Pacific-Slope Flycatcher	7	green
Spotted Towhee	8	red
Nashville Warbler	12	blue
Black-Headed Grosbeak	10	cyan
Orange-Crowned Warbler	4	yellow
Cassin's Vireo	90	magenta
Unknown bird song	6	dark green
Others	3	dark red

[Briggs 13] Briggs, F., Raich, R., Eftaxias, K., Lei, Z., and Huang, Y.: The ninth annual MLSP competition: overview, in *IEEE International workshop on machine learning for signal processing, Southampton, United Kingdom., Sept*, pp. 22–25 (2013)

[Catchpole 03] Catchpole, C. K. and Slater, P. J.: *Bird song: biological themes and variations*, Cambridge University Press (2003)

[Goëau 16] Goëau, H., Glotin, H., Vellinga, W.-P., Planqué, R., and Joly, A.: LifeCLEF Bird Identification Task 2016, in *CLEF working notes 2016* (2016)

[Holland 83] Holland, P. W., Laskey, K. B., and Leinhardt, S.: Stochastic blockmodels: First steps, *Social networks*, Vol. 5, No. 2, pp. 109–137 (1983)

[Kojima 16] Kojima, R., Sugiyama, O., Suzuki, R., Nakadai, K., and Taylor, E. C.: Semi-Automatic Bird Song Analysis by Spatial-Cue-Based Integration of Sound Source Detection, Localization, Separation, and Identification, in *Intelligent Robots and Systems (IROS), 2016 IEEE/RSJ International Conference on*IEEE (2016)

[Nigam 00] Nigam, K., McCallum, A. K., Thrun, S., and Mitchell, T.: Text classification from labeled and unlabeled documents using EM, *Machine learning*, Vol. 39, No. 2-3, pp. 103–134 (2000)

[Otsuka 14] Otsuka, T., Ishiguro, K., Sawada, H., and Okuno, H. G.: Bayesian nonparametrics for microphone array processing, *T-ASLP*, Vol. 22, No. 2, pp. 493–504 (2014)

[Pardo 06] Pardo, J. M., Anguera, X., and Wooters, C.: Speaker diarization for multiple distant microphone meetings: mixing acoustic features and inter-channel time differences, *Proceedings of the Ninth International Conference on Spoken Language Processing*, pp. 2194–2197 (2006)

[Schmidt 86] Schmidt, R. O.: Multiple emitter location and signal parameter estimation, *IEEE Transactions on Antennas and Propagation*, Vol. 34, No. 3, pp. 276–280 (1986)

[Sra 12] Sra, S.: A short note on parameter approximation for von Mises-Fisher distributions: and a fast implementation of I s (x), *Computational Statistics*, Vol. 27, No. 1, pp. 177–190 (2012)

[Suzuki 15] Suzuki, R. and Cody, M. L.: Complex systems approaches to temporal soundspace partitioning in bird communities as a self-organizing phenomenon based on behavioral plasticity, in *Proc. of the 20th International Symposium on Artificial Life and Robotics*, pp. 11–15 (2015)

[Suzuki 16] Suzuki, R., Matsubayashi, S., Nakadai, K., and Hiroshi, O. G.: Localizing bird songs using an open source robot audition system with a microphone array, in *Proceedings of Interspeech 2016*, pp. 2026–2030 (2016)

[Uemura 15] Uemura, S., Sugiyama, O., Kojima, R., and Nakadai, K.: Outdoor Acoustic Event Identification using Sound Source Separation and Deep Learning with a Quadrotor-Embedded Microphone Array, in *ICAM2015*, pp. 329–330, JSME (2015)

[小島 15] 小島 諒介, 杉山 治, 鈴木 麗璽, 中臺 一博 : 音源アノテーション補助のための音源位置を考慮した同定モデル, in *RSJ2015*, 日本ロボット学会 (2015)

[小島 16] 小島 諒介, 杉山 治, 鈴木 麗璽, 中臺 一博 : 音源位置を考慮した音源同定のための確率モデルとその学習, in *RSJ2016*, 日本ロボット学会 (2016)

[奥乃 10] 奥乃 博, 中臺 一博 : ロボット聴覚オープンソースソフトウエア HARK, 日本ロボット学会誌, Vol. 28, No. 1, pp. 6–9 (2010)

[松林 15] 松林 志保, 鈴木 麗璽, 小島 諒介, 中臺 一博 : 複数のマイクロフォンアレイとロボット聴覚ソフトウェア HARK を用いた野鳥の位置観測精度の検討, 第 43 回 AI-Challenge 研究会, pp. 54–59, 人工知能学会 (2015)