Vol.2015-MUS-107 No.55
2015/5/24

視聴覚統合NMFによるカエル合唱音声の分析

糸山 克寿[1,a)]　坂東 宜昭[1]　粟野 浩光[1]　合原 一究[2]　吉井 和佳[1]

概要：本稿では，映像と音響信号に対して統合的に非負値行列因子分解 (NMF) を行うことでカエルなどの動物の合唱行動を分析する手法について報告する．カエルをはじめとした様々な動物は合唱（音声によるコミュニケーション）を行うことが知られており，各個体がどのように合唱に参加しているかを調べることはその生態の解明に重要である．空間的な音場を光に変換するデバイスであるカエルホタルを用いて，ビデオカメラで録画した映像およびモノラル音響信号に対して統合的に NMF を行うことで，各個体の鳴き声を分離抽出する．カエルホタルの輝度とパワースペクトルの振幅を NMF のアクティベーションとして共有させることで，スペクトル形状が類似した同種別個体の鳴き声を相異なる基底へと分解する．

1. はじめに

人間をはじめとして，音声によるコミュニケーションを行う生物は多種多様である．鳥の鳴き声は最もよく研究されている音声コミュニケーションのひとつであり，eBird Project [1] といった大規模な取り組みがなされている．ジュウシマツ (*Lonchura striata* var. *domestica*) などの鳥の歌には文法があり [2]，さらに後天的に文法を獲得する [3] ことが明らかになっている．カエルは水田などで身近に観察でき，童謡でも歌われているように鳴き声が特徴的な生物である．多数のカエルが鳴き交わす「合唱」の時空間的構造をとらえることで生態や行動戦略，その地域差の解明に大きく寄与することが期待される．そのためには，カエルの「合唱」を録音し，そこから個々のカエルの鳴き声を分離抽出する，音源分離が必要不可欠である．

マイクロフォンアレイの幾何的・統計的な特徴を利用したマルチチャネル音響信号に対する音源分離技術は複数話者音声やカエル，野鳥の鳴き声に適用されている．しかしながら，マイクロフォンアレイを用いた録音は必ずしも容易ではない．第一に，多数（おおむね 8 以上）のチャンネルを同時に扱える録音機器はモノラルやステレオのみを扱える機器に比べて相応に高価である．第二に，目的や環境に合わせた適切なアレイの設計は自明ではない．第三に，高性能・高分解能なアレイは相応の大きさがあるため，フィールドでの持ち運びや設置が容易ではない．したがって，比較的安価であり，高度な知識や技術を必要とせず，持ち運びが容易な大きさのデバイスでも録音可能な，モノラルやステレオの音響信号に対する音源分離は重要な課題である．

図 1　左：ニホンアマガエル *Hyla japonica*
右：シュレーゲルアオガエル *Rhacophorus schlegelii*

本稿では，モノラル音響信号の音源分離によく用いられる非負値行列因子分解 (non-negative matrix factorization; NMF) を応用した視聴覚統合 NMF とその実環境での実験結果について報告する．様々な種類のカエルが鳴き声でインタラクションを行う様子を解明するため，水田での複数のカエルの鳴き声を録音したモノラル音響信号から個々のカエルの鳴き声を分離抽出することが目的である．同種のカエルの鳴き声は基本的に類似したスペクトル構造をもつため，音響信号のみを用いた NMF では同種で異なる個体の鳴き声を分離することができない．また，水田では多数の生き物が鳴き声でコミュニケーションを行うため，鳴き声の分析対象外の（周囲の水田で鳴く）カエルやカエル以外の昆虫の鳴き声も同時に録音されてしまい，鳴き声分離抽出の妨げとなる．視聴覚統合 NMF では，音響信号のスペクトログラムと映像に対して同時に因子分解を行うことで，モノラル音響信号のみでは実現不可能な空間構造に基づく音源分離を実現する．さらに，スペクトログラムに対してのみ追加の因子を定義することで，分析対象・対象外の鳴き声を相異なる因子へとクラスタリングする．

[1] 京都大学 Kyoto University
[2] 同志社大学 Doshisha University
[a)] itoyama@kuis.kyoto-u.ac.jp

図 2 左：カエルホタル（旧型），右：カエルホタル（新型）

実験では，ニホンアマガエル (*Hyla japonica*) とシュレーゲルアオガエル (*Rhacophorus schlegelii*)（**図 1**）のデータを用いて，個々のカエルの鳴き声を分離抽出する様子を示す．映像はカエルホタル [4,5] と新型カエルホタル [6]（**図 2**）の発光パターンをビデオカメラで撮影したもので，音響信号はそのビデオカメラのマイクロフォンで収録されたものである．

2. 関連研究

2.1 カエルの鳴き声による相互作用

オスのカエルの鳴き声による相互作用は，熱帯雨林，渓流，湖沼などさまざまな場所で観察される [7–9]．その空間的な配置は種や生息環境によって大きく異なっている．

ニホンアマガエル (*Hyla japonica*) は日本では最も一般的なカエルの種のひとつである [10]．オスのニホンアマガエルの鳴き声は，田植えのために水田に水が入れられた直後から初夏にかけて，主に夜の水田の縁で聞くことができる．室内実験では，2 匹のニホンアマガエルは主に逆相で同期して鳴く [11] ことが，3 匹のニホンアマガエルは 2 匹と 1 匹に分かれた逆相同期もしくは 1 匹ずつが分かれた三相同期で鳴く [12] ことが知られている．カエルホタル [4,5] を用いたニホンアマガエルの合唱分析では，近接した 2 匹のカエルが逆相で同期することで水田全体では大きく 2 つのグループに分かれて鳴く [5] ことが明らかになっている．さらに，カエルホタルに周波数フィルタを導入した新型のカエルホタルにより，ニホンアマガエルを含む複数種のカエルの発声行動の計測実験が行われてきている [6]．このように，一部のカエルに関しては鳴き声の時空間的な構造が解明されつつあるが，その他の多くの種類のカエルに関してはその生態や合唱行動などいまだ明らかになっていないものも多い．

2.2 音源分離

音源分離は音響信号処理における最も重要な課題の一つであり，マルチチャネル，シングルチャネルそれぞれの音響信号に対してさまざまな手法が報告されている．マルチチャネル信号処理においては，特にマイクロフォンアレイ（多数の同期されたマイクロフォン）を用いたものが主流である．ビームフォーミング [13] は音源から各マイクロフォンへの到達時間差を利用して信号を同期加算することで目的方向の音響信号を強調・抑圧する手法であり，計算量が少なく，ハードウェア化が容易である．独立成分分析 [14] およびその拡張である独立ベクトル分析 [15–17] は音源信号の統計的独立性および非ガウス性に基づいて相互に独立な音源信号を分離する手法である．これらの手法はアレイの幾何的配置などの事前情報を必要としないため自由なマイク配置で運用できる．アレイの伝達特性を表す共分散モデルのパラメータのベイズ推定による最適化に基づく手法 [18,19] は計算量は多いものの，優決定，劣決定を問わず音源数が未知の環境下でも適用可能であることが報告されている．これらの技術は人間の音声や音楽のみならず，カエルの鳴き声 [20] や野鳥の鳴き声 [21] の分析に応用されている．

シングルチャネル信号処理での音源分離は，特に音楽音響信号を対象としてさまざまな技術が開発されてきた．調波・非調波音分離 [22,23] は調波音のスペクトログラムは “横方向” に，打楽器音は “縦方向” に滑らかであるという異方性に着目してこれらの音を分離する．ロバスト主成分分析を用いた歌声分離 [24,25] は，ギター，ベース，キーボード，ドラムなどの伴奏音は低ランクのスペクトログラムで，歌声はスパースなスペクトログラムで表現されることに着目した手法である．これらの手法は基本的に事前情報が不要であるため，ジャンルを問わず幅広い楽曲に適用可能である．調波・非調波モデル [26] は楽器音の詳細なスペクトルモデルを構築することで複雑な楽曲からの個々の音源の抽出を行っている．近年では，非負値行列因子分解 [27] に基づく音源分離手法が多数報告されている．多くの楽器音のスペクトログラムは，時間的に不変なスペクトル基底とそのアクティベーションの時間的変動という少数の因子に分解できることに基づいている．NMF の因子分解モデルは非常にシンプルであるため，スパース制約の導入，ベイズ的事前分布の導入，基底スペクトル形状の調波構造への制約など，非常に多くの拡張が報告されている．

2.3 マルチモーダル信号処理

映像や楽譜などにより欠落した情報を補完したり曖昧性をなくすことで，信号処理の精度を高める試みが報告されている．音声認識においては，音声に加えて唇の映像やジェスチャを用いたマルチモーダル音声認識 [28] が報告されている．音楽音響信号処理においては，楽曲の拍を検出するタスクであるビートトラッキングにおいて，ギタリストの映像を用いて複雑なビートを認識する手法 [29]，ダンサーの映像を用いて雑音下でビートを認識する手法 [30] などが報告されている．音源分離タスクにおいても，楽譜 [26,31] やハミング [32] を事前情報に用いた音源分離が報告されており，複雑な楽曲の分離や高精度な分離が実現されている．

3. 視聴覚統合 NMF

音響信号の振幅スペクトログラムに対して，非負値行列因子分解 (non-negative matrix factorization; NMF) [33] を適用する．ただし，スペクトログラムにそのまま NMF を適用すると以下の 2 点の問題が起こる．

(1) 同種の複数の個体の鳴き声が一つのスペクトル基底にクラスタリングされ，複数個体の鳴き声を分離できない．同じ種類のカエルは鳴き声の特徴が強く類似しており，NMF はこれをできるだけ少ない数の基底で表現しようとするためである．

(2) 目標のカエルのみの鳴き声を抽出できない．水田には目標（インタラクションの観察対象）のカエルの他にも多数のカエルや昆虫などが鳴き声を発しており，これらすべての鳴き声が重畳して録音されるためである．

これらの問題を以下のように解決する．

(1) 音響信号と同期した，カエルホタルの発光を録画した映像を用いる．NMF のアクティベーション行列を音響信号と映像とで共有させ，目標のカエルだけを抽出し，同種の異個体を分離する．

(2) 音響信号側にのみ，少数の因子ペアを追加して NMF する．ホタルが捕捉していないカエルの鳴き声をこれらの因子でトラップし，共有アクティベーションにこれらの鳴き声成分が混入することを防ぐ．

2 つの非負値行列：モノラル音響信号のスペクトログラム $Y^A \in \mathbb{R}_+^{M\times N_A}$ とカエルホタルの映像 $Y^V \in \mathbb{R}_+^{M\times N_V}$ （前処理 [5] により各ホタルの発光時系列を抽出済み）を観測データとする．$M > 0$ はスペクトログラムとカエルホタル映像のフレーム数，$N_A > 0$ はスペクトログラムの周波数ビン数，$N_V > 0$ カエルホタルの数である．音響信号は映像のフレームレートに合わせて短時間分析がなされているものとする．なお，音源分離対象の（カエルホタルの範囲内に存在する）カエルの個体数 $K_S > 0$ は既知とする．

観測された 2 つの非負値行列に対する因子分解を以下で定義する．

$$Y^A \approx H^S U^A + H^D U^D, \quad Y^V \approx H^S U^V \tag{1}$$

$H^S \in \mathbb{R}_+^{M\times K_S}$ は分離対象のカエルの鳴き声アクティベーション（**音声と映像で共有**），$H^D \in \mathbb{R}_+^{M\times K_D}$ は分離対象外のカエルなどの鳴き声アクティベーション（**音声のみ**），$U^A \in \mathbb{R}_+^{K_S\times N_A}$ は分離対象のカエルの鳴き声スペクトル基底，$U^D \in \mathbb{R}_+^{K_D\times N_A}$ は分離対象外のカエルなどのスペクトル基底，$U^V \in \mathbb{R}_+^{K_S\times N_V}$ は分離対象のカエルのカエルホタル発光パターンである．$K_D > 0$ は目標以外の鳴き声をトラップするための因子の数である．

因子分解の良し悪しを評価するための目的関数 L を Kullback-Leibler divergence (KLD) を用いて以下で定義する．

$$\begin{aligned}
L &= \alpha_A D_I(Y^A||H^S U^A + H^D U^D) + \alpha_V D_I(Y^V||H^S U^V) \\
&= \alpha_A \sum_{m=1}^{M}\sum_{n=1}^{N_A}\Big(Y^A_{m,n}\log\frac{Y^A_{m,n}}{(H^S U^A + H^D U^D)_{m,n}} \\
&\qquad - Y^A_{m,n} + (H^S U^A + H^D U^D)_{m,n}\Big) \\
&\quad + \alpha_V \sum_{m=1}^{M}\sum_{n=1}^{N_V}\Big(Y^V_{m,n}\log\frac{Y^V_{m,n}}{(H^S U^V)_{m,n}} \\
&\qquad - Y^V_{m,n} + (H^S U^V)_{m,n}\Big)
\end{aligned} \tag{2}$$

ここで，$\alpha_A > 0$ と $\alpha_V > 0$ はそれぞれ音響信号と映像の KLD に対する重み（信頼度）である．この目的関数 L に対して，各因子 H^S, H^D, U^A, U^V, U^D に応じて設計された補助関数に基づいて各因子の更新式を導出する．ここでは設計した補助関数と更新式の導出過程については省略する．

$$H^S_{m,k} \leftarrow H^S_{m,k}\frac{\hat{H}^S_{m,k}}{\alpha_A\sum_{n=1}^{N_A}U^A_{k,n} + \alpha_V\sum_{n=1}^{N_V}U^V_{k,n}} \tag{3}$$

$$\hat{H}^S_{m,k} = \alpha_A\sum_{n=1}^{N_A}\frac{Y^A_{m,n}U^A_{k,n}}{(\hat{H}^S U^A + H^D U^D)_{m,n}} \tag{4}$$

$$+ \alpha_V\sum_{n=1}^{N_V}\frac{Y^V_{m,n}U^V_{k,n}}{(\hat{H}^S U^V)_{m,n}} \tag{5}$$

$$H^D_{m,k} \leftarrow H^D_{m,k}\frac{\sum_{n=1}^{N_A}\frac{Y^A_{m,n}U^D_{k,n}}{(H^S U^A + \hat{H}^D U^D)_{m,n}}}{\sum_{n=1}^{N_A}U^D_{k,n}} \tag{6}$$

$$U^A_{k,n} \leftarrow U^A_{k,n}\frac{\sum_{m=1}^{M}\frac{Y^A_{m,n}H^S_{m,k}}{(H^S\hat{U}^A + H^D U^D)_{m,n}}}{\sum_{m=1}^{M}H^S_{m,k}} \tag{7}$$

$$U^D_{k,n} \leftarrow U^D_{k,n}\frac{\sum_{m=1}^{M}\frac{Y^A_{m,n}H^D_{m,k}}{(H^S U^A + H^D\hat{U}^D)_{m,n}}}{\sum_{m=1}^{M}H^D_{m,k}} \tag{8}$$

$$U^V_{k,n} \leftarrow U^V_{k,n}\frac{\sum_{m=1}^{M}\frac{Y^V_{m,n}H^S_{m,k}}{(H^S\hat{U}^V)_{m,n}}}{\sum_{m=1}^{M}H^S_{m,k}} \tag{9}$$

4. 実験

視聴覚統合 NMF を用いて個々のカエルの鳴き声を抽出する実験を行った．島根県隠岐の島壇鏡の滝付近の水田で収録した 2 つのデータを用いた．カエルホタルを水田の畦に並べ，その発光パターンをビデオカメラで撮影するとともにその鳴き声をビデオカメラのマイクロフォンで録音した．音響信号はサンプリングレート 48kHz で録音し，16kHz にダウンサンプリングした．映像は 59.94fps，full HD で録画した．

データ 1　2014 年 6 月 1 日 21 時 55 分．旧型のカエルホタル 9 台を用いた．カエルホタルを並べた付近では 3 匹のシュレーゲルアオガエルが鳴いていることを確認した．**図 3** に音響信号のスペクトログラムと映像からカエルホタルの発光状態を抽出したものを示す．

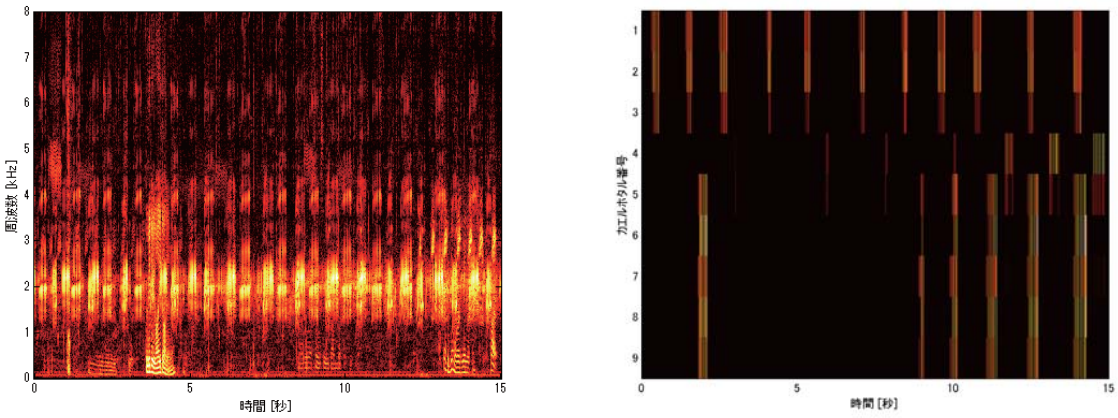

図 3　データ 1．左：音響信号（スペクトログラム）と右：映像（カエルホタルの輝度）

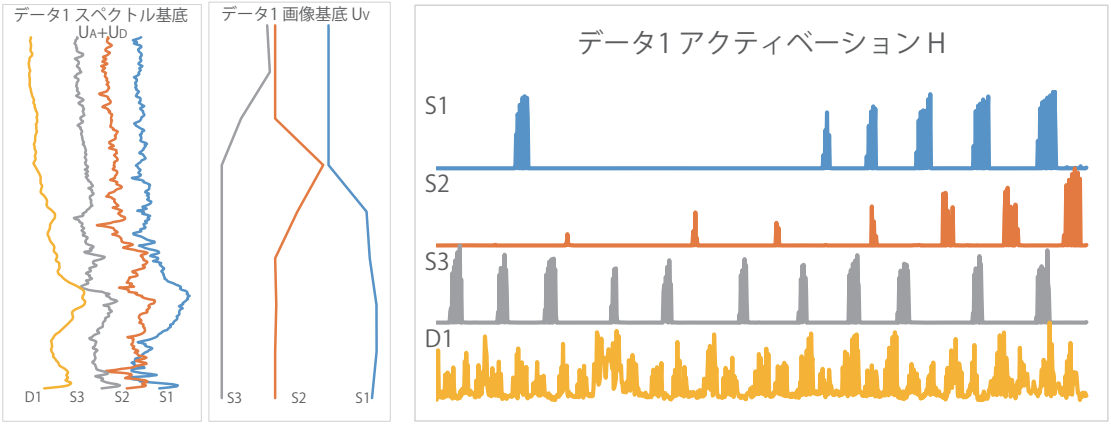

図 4　データ 1 の分析結果

データ 2　2014 年 6 月 3 日 23 時 16 分．旧型と新型のカエルホタル 12 台ずつ，合計 24 台を用いた．新型のカエルホタルはシュレーゲルアオガエルの鳴き声にのみ反応するように周波数フィルタを調整した．前処理 [5] のエラーにより 1 つのカエルホタルが 2 つに誤認識されている（12 番と 13 番）．カエルホタルを並べた付近では 2 匹のニホンアマガエルと 2 匹のシュレーゲルアオガエルが鳴いていることを確認した．**図 5** に音響信号のスペクトログラムと映像からカエルホタルの発光状態を抽出したものを示す．

図 4 および**図 6** に各データに対して視聴覚統合 NMF を適用した結果を示す．データ 1 については，画像基底 U^V より，カエルホタル番号 1–2，4–5, 7–9 の付近でカエルが 1 匹ずつ鳴いている様子が観察できる．アクティベーション H^S, H^D からは 3 匹のカエルが位相をずらして鳴いており，さらに背景雑音として多数のカエルなどが常に鳴いている様子が観察できる．

データ 2 については，画像基底 U^V より因子 S1 と S3 がシュレーゲルアオガエル，因子 S2 と S4 がニホンアマガエルであると推測できる．データ 2 では，あらゆる周波数に反応する旧型カエルホタルとシュレーゲルアオガエルの鳴き声の周波数のみに反応する新型カエルホタルを交互に並べている．S1 と S3 は旧型と新型の両方のカエルホタルに反応が見られるためシュレーゲルアオガエル，S2 と S4 は反応が見られるカエルホタルとそうでないものが交互に現れているためニホンアマガエルであるといえる．アクティベーション H^S からは S1, S3 と S2, S4 とで時間的密度に違いがみられ，このことからもこれらが違う種のカエルであることが推測される．アクティベーション H^D をみると，データ先頭からしばらくの間は値が大きくなっておりその後は値が小さくなっていることから，ビデオカメラの真下などにもカエルがいたと推測される．

一方で，スペクトル基底 U^A には各因子の間で大きな違いが現れなかった．得られた因子から音響信号を復元して聴取してみたところ，鳴き方の時間的構造は確かにカエルの種の違いを反映していたが，音色からそれぞれの種の特徴をつかむことは困難であった．目的音に対する背景雑音の音量が大きく分離音の音色的特徴が背景音にマスクされていたことなどが原因であると考えられる．

5. おわりに

本稿では，野外での複数種のカエルの鳴き声を録音したモノラル音響信号から各カエル個体の鳴き声を分離抽出するために視聴覚統合 NMF を開発し，隠岐の島で収録した

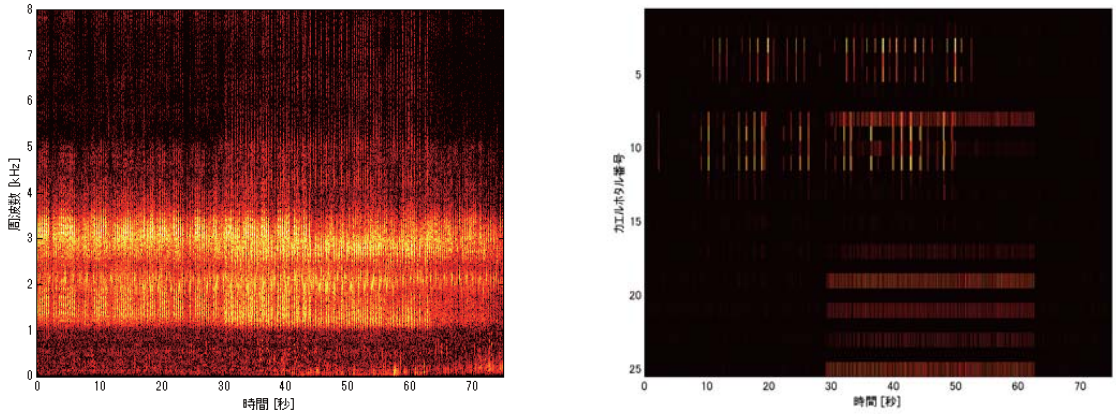

図 5 データ 2．左：音響信号（スペクトログラム）と右：映像（カエルホタルの輝度）

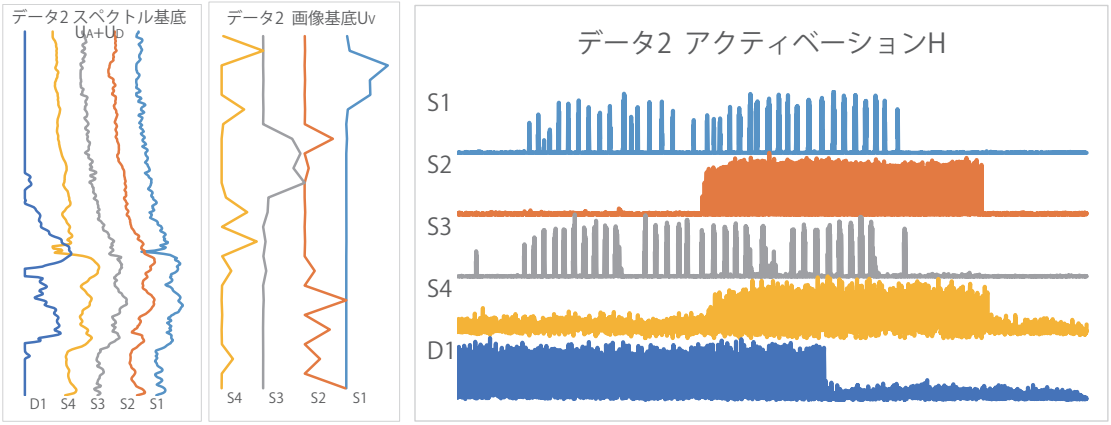

図 6 データ 2 の分析結果

データに対して適用した．映像データに対する NMF はカエルホタルの情報を適切にクラスタリングしておりデータの分析においては一定の有用性があると考えられる．一方で鳴き声の分離においては，音色の再現性に課題が残されている．今後は提案法の音源分離性能向上および他のデータ分析手法との組み合わせによる効率的なデータ分析フレームワークの開発に取り組む予定である．

参考文献

[1] Kelling, S., Lagoze, C., Wong, W.-K., Yu, J., Damoulas, T., Gerbracht, J., Fink, D. and Gomes, C.: eBird: A Human / Computer Learning Network to Improve Biodiversity Conservation and Research, *AI magazine*, Vol. 34, No. 1, pp. 10–20 (2013).

[2] 岡ノ谷一夫：さえずり言語起源論 新版 小鳥の歌からヒトの言葉へ，岩波書店 (2010).

[3] Abe, K. and Watanabe, D.: Songbirds Possess the Spontaneous Ability to Discriminate Syntactic Rules, *Nature Neuroscience*, Vol. 14, pp. 1067–1074 (2011).

[4] Mizumoto, T., Aihara, I., Otsuka, T., Takeda, R., Aihara, K., and Okuno, H. G.: Sound Imaging of Nocturnal Animal Calls in Their Natural Habitat, *Journal of Comparative Physiology A*, Vol. 197, No. 9, pp. 915–921 (2011).

[5] Aihara, I., Mizumoto, T., Otsuka, T., Awano, H., Nagira, K., Okuno, H. G. and Aihara, K.: Spatio-Temporal Dynamics in Collective Frog Choruses Examined by Mathematical Modeling and Field Observations, *Scientific Reports*, Vol. 4, No. 3891, pp. 1–8 (2014).

[6] Mizumoto, T., Awano, H., Aihara, I., Otsuka, T. and Okuno, H. G.: Sound Imaging System for Visualizing Multiple Sound Sources from Two Species, *Tenth International Congress of Neuroethology* (2012).

[7] Gerhardt, H. C. and Huber, F.: *Acoustic Communication in Insects and Anurans*, The University of Chicago Press (2002).

[8] Wells, K. D.: *The Ecology and Behavior of Amphibians*, The University of Chicago Press (2007).

[9] Narins, P. M., Feng, A. S., Fay, R. R. and Popper, A. N.(eds.): *Hearing and Sound Communication in Amphibians*, Springer (2006).

[10] 前田憲男，松井正文：日本カエル図鑑，文一総合出版 (1999).

[11] Aihara, I.: Modeling Synchronized Calling Behavior of Japanese Tree Frogs, *Physical Review E*, Vol. 80, No. 1, pp. 1–7 (2008).

[12] Aihara, I., Takeda, R., Mizumoto, T., Otsuka, T., Takahashi, T., Okuno, H. G. and Aihara, K.: Complex and Transitive Synchronization in a Frustrated System of Calling Frogs, *Physical Review E*, Vol. 83, No. 3, pp. 1–5 (2011).

[13] Veen, B. D. V. and Buckley, K. M.: Beamforming: A Versatile Approach to Spatial Filtering, *IEEE ASSP Magazine*, Vol. 5, No. 2, pp. 4–24 (1988).

[14] Hyvärinen, A., Karhunen, J. and Oja, E.: *Independent Component Analysis*, John Wiley & Sons (2001).

[15] Kim, T., Lee, I. and Lee, T.-W.: Independent Vector Analy-

sis: Definition and Algorithms, *Fortieth Asilomar Conference on Signals, Systems and Computers, 2006 (ACSSC '06)*, pp. 1393–1396 (2006).
[16] Kim, T.: Real-Time Independent Vector Analysis for Convolutive Blind Source Separation, *IEEE Trans. Circ. Syst. I: Regular Papers*, Vol. 57, No. 7, pp. 1431–1438 (online), DOI: 10.1109/TCSI.2010.2048777 (2010).
[17] Anderson, M., Fu, G.-S., Phlypo, R. and Adalı, T.: Independent Vector Analysis: Identification Conditions and Performance Bounds, *IEEE Trans. Signal Process.*, Vol. 62, No. 17, pp. 4399–4410 (2014).
[18] Otsuka, T., Ishiguro, K., Sawada, H. and Okuno, H. G.: Bayesian Nonparametrics for Microphone Array Processing, *IEEE/ACM Trans. Audio, Speech and Lang. Process.*, Vol. 22, No. 2, pp. 493–504 (2014).
[19] Otsuka, T., Ishiguro, K., Yoshioka, T., Sawada, H. and Okuno, H. G.: Multichannel Sound Source Dereverberation and Separation for Arbitrary Number of Sources Based on Bayesian Nonparametrics, *IEEE/ACM Trans. Audio, Speech and Lang. Process.*, Vol. 22, No. 12, pp. 2218–2232 (2014).
[20] Bando, Y., Otsuka, T., Aihara, I., Awano, H., Itoyama, K., Yoshii, K. and Okuno, H. G.: Recognition of In-Field Frog Chorusing Using Bayesian Nonparametric Microphone Array Processing, *2015 AAAI Workshop*, pp. 2–6 (2015).
[21] 鈴木麗璽，Taylor, C. E.，Cody, M. L.：野鳥の歌コミュニケーション理解への試み，人工知能学会研究会資料 SIG-Challenge-B302-05，pp. 22–27 (2013).
[22] Ono, N., Miyamoto, K., Kameoka, H. and Sagayama, S.: A Real-time Equalizer of Harmonic and Percussive Components in Music Signals, *ISMIR2008*, pp. 139–144 (2008).
[23] FitzGerald, D.: Harmonic/percussive Separation Using Median Filtering, *DAFx-10*, pp. 1–4 (2010).
[24] Huang, P.-S., Chen, S. D., Smaragdis, P. and Hasegawa-Johnson, M.: Singing-voice Separation from Monaural Recordings Using Robust Principal Component Analysis, *ICASSP2012*, pp. 57–60 (2012).
[25] Ikemiya, Y., Yoshii, K. and Itoyama, K.: Singing Voice Analysis and Editing Based on Mutually Dependent F0 Estimation and Source Separation, *ICASSP2015*, pp. 574–577 (2015).
[26] 糸山克寿，後藤真孝，駒谷和範，尾形哲也，奥乃 博：楽譜情報を援用した多重奏音楽音響信号の音源分離と調波・非調波統合モデルの制約付パラメータ推定の同時実現，情処論，Vol. 49, No. 3, pp. 1465–1479 (2007).
[27] Smaragdis, P. and Brown, J. C.: Non-negative Matrix Factorization for Polyphonic Music Transcription, *WASPAA2003*, pp. 177–180 (2003).
[28] Chan, A. D. C., Englehart, K., Hudgins, B. and Lovely, D. F.: Hidden Markov Model Classification of Myoelectric Signals in Speech, *IEEE Engineering in Medicine and Biology Magazine*, Vol. 21, No. 5, pp. 143–146 (online), DOI: 10.1109/MEMB.2002.1044184 (2002).
[29] Itohara, T., Otsuka, T., Mizumoto, T., Lim, A., Ogata, T. and Okuno, H. G.: A Multimodal Tempo and Beat-tracking System Based on Audiovisual Information from Live Guitar Performances, *EURASIP J. Audio, Speech, Music Process.*, Vol. 6, pp. 1–17 (2012).
[30] 大喜多美里，坂東宣昭，池宮由楽，糸山克寿，吉井和佳：ダンス共演ロボットのためのマルチモーダルビートトラッキング，情報処理学会第 77 回全国大会講演論文集 5S-05 (2015).
[31] Ewert, S., Pardo, B., Müller, M. and Plumbley, M. D.: Score-Informed Source Separation for Musical Audio Recordings, *IEEE Signal Processing Magazine*, Vol. 31, No. 3, pp. 116–124 (2014).
[32] Smaragdis, P. and Mysore, G. J.: Separation by "Humming": User-guided Sound Extraction from Monophonic Mixtures, *WASPAA2009*, pp. 69–72 (2009).
[33] Lee, D. D. and Seung, H. S.: Algorithms for Non-negative Matrix Factorizaton, *Advances in Neural Information Processing Systems 13 (NIPS 2000)*, MIT Press, pp. 556–562 (2000).