

音源定位技術が切り開くサルの生態と会話における未解決問題

Sound localizations reveal the unsolved issues of primate vocal communications, conversations and conservations

香田啓貴^{1*} 豊田有² Suchinda Malaivijitnond³⁴ 鈴木麗璽⁵
Hiroki Koda¹ Aru Toyoda² Suchinda Malaivijitnond³⁴ Reiji Suzuki⁵

¹ 京都大学霊長類研究所

¹ Primate Research Institute, Kyoto University

² 中部大学創発学術院

² Chubu University Academy of Emerging Science

³ チュラロンコン大学理学部

³ Faculty of Science, Chulalongkorn University

⁴ タイ国立霊長類センター

⁴ National Primate Research Center of Thailand

⁵ 名古屋大学大学院情報学研究科

⁵ Graduate School of Informatics, Nagoya University

Abstract: サル類は、群れを形成しその中でコミュニケーションする点で、ヒトとの類似性が高い。音声によるコミュニケーションを分析すると、発声者間のやり取りの時間規則性は、会話と類似する傾向が確認され、個体間の「会話」頻度の計測可能性は、音声コミュニケーションに基づいた社会ネットワーク分析の良い研究対象となる。本稿では、テナガザルの歌を例題としてマイクロフォンアレイとロボット聴覚技術を利用することで、サルの群れ社会のなかで運用される音声の社会生態学的な役割を明らかにする試みと将来性について論じる。

1 ヒトの会話とサルの「会話」

分類学上、ヒトは霊長類(サル類)の一種に分類される哺乳類である。脳の構造という点も含め身体設計において高い相同性を保持しながら多くの共通点が見られる。行動や心理現象、さらには群れと言った社会は、基本的に身体と生理などの生物基盤の上で実現されている現象である。そのため、行動や社会などの化石に残らない「人らしさ」については、サルの行動や脳機能、社会や生態の適応状況などの証拠と比較して、その進化と起源を推定する方法が有力となる。中でも、ヒトの会話は、個体間で言語活動を相互作用させる行為と言える。ここでいう言語活動とは、表象や意図を記号化する思考だとすれば、そのような記号変換行為はサルでは困難であるため、サルではヒトと同じ会話行為は見つからない。こうしたヒト固有な言語活動を排除すれば、共通成分としては、個体間の信号のやり取りが残る。そして、その信号のやり取りの規則性を客観的に解析すると、ヒトの会話と共通する規則が見えることが知られてきた [1]。

ヒトとサルの発声のやり取りのなかで、共通する規則性として客観的に定義できるのは、時間的規則性である [1]。古くは、Sacks と Schegloff, Jefferson が、ヒトの会話における話者交代現象を単純な時間的規則によって定義した [2]。二者間の発話のやり取り、時間の流れに着目すると、その発話の時間間隔 (inter-call intervals) の生起頻度の分布は、1 秒以内の短時間に最

頻値を示す。誰かが話したのち、「返事」や「相槌」は一定時間以内に応答する、という規則である。サルの発声にも同じ傾向が見える。サルは群れを形成し、空間的なまとまりを維持して、森の中を移動する。特定の意味を記号的に符号化し、情報をやり取りする発話とは異なり、サルの発声は個体間の空間的な位置確認を行い、凝集性を保つ生態学的な役割があると考えられている。二個体間の発声のやり取りを分析すると、類似した時間規則性が確認できる (図 1)。実際のところ、これらの現象は飼育室でくらすサルでも、声のやり取りが確認できる、同様の時間規則が確認できる [1]。こうして、サルの発声を検出し、時間間隔を評価することで、コミュニケーションを頻繁にとりあう関係性が客観的に評価できる。

2 サル研究での様々な問題

霊長類研究者は、発声のやり取りの時間的規則の会話との類似性に着目し、個体間の会話関係についての研究を実施してきた。例えば、飼育下でのリスザルの研究によると、親和的な個体(個体間の距離が近いなど)どうしほど、音声コミュニケーションの頻度が高いという [3]。一方で、野生ニホンザルの会話分析では、普段毛づくろいなどの身体接触が少ない個体同士でむしろ音声コミュニケーションが頻繁であるという、逆の傾向も報告されている [4]。飼育下のビグミーマーモセットの研究では、鳴き交わす個体間で、音声の特徴が似てくるといふ報告がある [5]。母子間のやり取りという、特別な関係性に着目した会話分析も研究例として存在

*連絡先: 京都大学霊長類研究所
〒484-8506 愛知県犬山市官林 41-2
E-mail: koda.hiroki.7a@kyoto-u.ac.jp

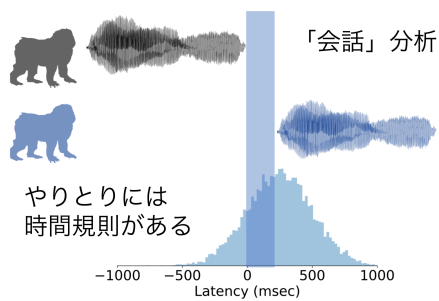


図 1: サルの発声の時間間隔には一定の最頻値が存在する。この値を境界として、音声に対する反応かどうかを客観的に定義することができる。

する。飼育下のコモンマーモセットの母子間の音声コミュニケーションの研究では、会話頻度とコドモの音声発達との間に相関関係が見られる(より「会話」する母子では、コドモの音声素早く発達する)[6]。母子が同時に発声する時間的規則が、発達に伴い変化するという報告も、複数のサル種で存在する[7, 8]。このように、発声者間のやり取りの頻度、すなわち、「会話頻度」は、霊長類の社会性を評価するようないい指標の一つであると言える。一方で、このような「会話頻度」が十分に調べられてきたとは言えないのも事実である。一連の先行研究の背景には、発声者情報の特定が必要不可欠な技術となる。しかし、これには方法的に極めて困難な作業であった。例えば、野生下の研究では、発声者の特定は観察者の技術と判断に依存する。典型的でわかりやすく、目視で確認もできる状況でのみ、2者間での発声やり取りを記載できるが、多くの観察データは発声者特定が困難であり、分析に利用できるデータは限られる。飼育下でのやり取りについては、マイクロフォンを首輪などの形で装着させる方法で特定することが手段の一つとして考えられるが、動物への負担などの面で、必ず採用できるとは限らない。音声コミュニケーションに基づく、社会ネットワークの評価は、サルの社会維持や音声発達という面で、重要性が認識されているが、方法の困難さから十分な評価がされていない。

3 応用問題としての類人猿歌の音源位置特定

発声者を特定し会話を可視化することは、重要な問題でありながら、未だ十分な形で取り組まれているとは言えない。サルの群れ社会での音声は、彼らの空間移動や集団凝集性と関わる媒体として生態学的に機能しているのであれば、群れ内と群れ間での音声コミュニケーションの全体像を把握することは、彼らの集団行動を理解する上で極めて重要であろう。

そこで、テナガザル類の歌に着目した我々の近年の取り組みについて紹介する。主に、マイクロフォンアレイによる録音とロボット聴覚技術 HARK(Honda Research Institute Japan Audition for Robots with Kyoto University)[9, 10]を用いた、テナガザル歌の音源定位問題についての予備的な報告をする。

3.1 生物学的背景

テナガザルが会話可視化の良い例題として選んだ理由となった、いくつかの行動学的な特徴について述べる。テナガザルは、主に東南アジアに生息するサルの仲間である。ヒトに近縁なサル類として類人猿が知られているが、テナガザルも類人猿に分類される[11]。ただし、チンパンジーやゴリラ、オランウータンといった大型類人猿とは異なり、体重は5kgから10kg程度と小柄なことから、小型類人猿と言われる。小柄な体格は、樹上生活によく適応している。ほとんどの時間を熱帯雨林の樹上30mの場所で暮らし、枝から枝に腕渡ししながら落下せず暮らしている。社会も特徴的である。群れの中心は、オトナメスとオトナオスの夫婦であり、その夫婦で1km²程度の広さの縄張りを防衛する。その夫婦には、2年から3年おきにコドモを1頭が生まれる。コドモは8歳ごろまでは、その家族で暮らし、性成熟後群れを移出し、新しい家族を形成していく。このように、夫婦に数頭のコドモが共にある、核家族を形成する。家族関係は、敵対的とされる。つまり、縄張りを家族間で防衛する。

彼らは、「歌」を歌うことで知られている[12]。ここでいうテナガザルの「歌」とは、連続的な発声のことである。多くのサル類の発声は、「単発」であり、連続的な発声はあまり見られない。しかし、テナガザルは、系列的な規則性を伴った発声をする(図2)。この系列規則性は、種に特異的とされる。すなわち、種に応じて歌の系列的な特長があることが知られている。さらに、この歌の系列的な特長は、雌雄差があることがわかっている。すなわち、オスの歌とメスの歌が存在する。核家族を形成するテナガザルの家族の中で、夫婦ではデュエットとしてやり取りをすることが知られている。夫婦間のデュエットの生態学的な機能としては、二つ提唱されている。一つは、夫婦間の絆の強化とされる。よく唱和することで、夫婦の関係性を強化すると考えられている。もう一つは、縄張り防衛機能である。森林内では、複数の家族が連続的に暮らしている状況となる。家族は縄張りを持つが、お互いの家族デュエットを歌い合い、縄張りの防衛に役立っていると考えられている。このデュエットは、多くのテナガザルで毎朝早朝に見られる。大変興味深いことに、歌は一旦開始されると2時間以上継続することが多い。さらに、歌は群れの縄張り防衛の役割を果たすことから、音は2km程度と、かなり長距離まで到達する。まとめると、以下のような特徴があると言える。

- 歌は家族単位で長時間連続的に鳴く。
- 歌はオスとメスで特徴があり音響的に区別できる。
- 歌は種によって特徴が異なり歌の特徴から種を区別できる。
- 歌は2キロ程度届くため、遠くからの歌による観測が可能である。

こうした、生物学的な特徴は音源定位を扱う問題として、大変有望な可能性を有していると考えられた。そこで、我々は、まずテナガザルが飼育されている集団ケージを対象に、歌の可視化可能性について検討した。

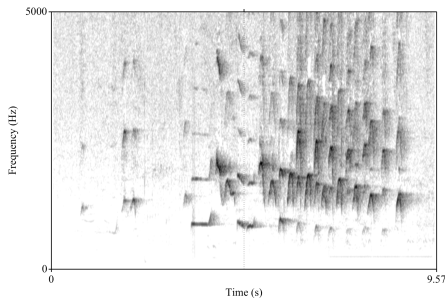


図 2: テナガザルの歌のサウンドスペクトログラム。図は、テナガザルの一種であるボウシテナガザルのオスの歌。こうした歌声が、長時間にわたって観察できる。

3.2 対象

対象は、タイ国 Krabok Koo Wildlife Breeding Center(以下、センターとする)で飼育されているテナガザルを対象に実施した。録音は、2019年1月に実施した。センターでは、3m x 10m x 5m 程度のケージに、一つの家族が飼育されるケージを配置してテナガザルを飼育していた。そして、そのケージは、10個が連続した形で配置され、それぞれに家族(1-4頭が暮らしている)が飼育されていた。そこで、10連ケージの一つを対象し、歌の発声時の音源定位を試みた(図3)。



図 3: 10 連ケージを正面から撮影。正面に、マイクロフォンアレイが三脚にのせて設置されている。

3.3 装置

音源定位に利用する録音には市販されている 8ch マイクロフォンアレイ (TAMAGO-03, System-In-Frontier Inc. Tokyo, Japan) を利用した。図のように、10 連ケージに対し、3 台を挟み込むように配置した(図4)。録音には Raspberry Pi 3 を利用し、スマートフォンの Wi-Fi テザリングを利用して簡易ネットワーク環境を構築した上で、時刻同期と録音制御を ssh を経由して制御・実施した。

3.4 予備的分析

図5は、10 連ケージの個体の鳴き声を定位するために設置した、同じ側にある2つのマイクロフォンアレイの録音を用いて予備的分析を行った例である。2019年1月に行った録音のうちの日分、早朝からの約8時間の録音を採用した。

図5(a)は片方のマイクロフォンアレイ(図中印のついたもの)を用いて音源到来方向を推定した例であり、図5(b)は、そのうち約2分のスペクトログラム、

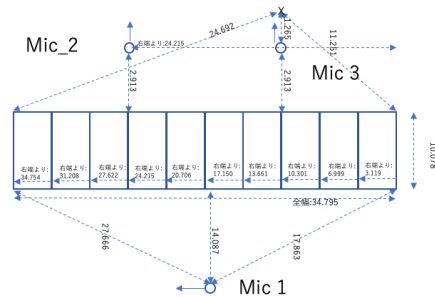


図 4: 10 連ケージとマイク配置の位置関係の模式図。レーザー測位により、マイク位置とケージのは位置関係を測位した。単位は m。

MUSIC スペクトル、定位音源情報(矩形)を示している。MUSIC スペクトルの算出には鳥類の鳴き声音源定位ツール HARKBird[13] を利用しテナガザルの鳴き声のみを抽出しやすいように 500-1800Hz の周波数域を対象とした。音源定位情報は、各時刻におけるスペクトルのピーク(図中の点)をクラスタリングして算出した。これは、現環境では MUSIC スペクトルのノイズや揺れが大きいいため、時間と方位の空間においてばらつきながらも集中して存在するピークを音源として抽出することを狙った。その結果、同図下段のように、異なる方位で交互に定位される音源が抽出され、そのパターンが上段のスペクトログラムにおける鳴き声の周波数パターンと概ね対応する状況を確認した。これは、前述の個体間における会話を抽出できることを示唆している。同時に、図5(a)からは、時刻によって複数の方向で音源が繰り返し定位され、上記のような会話が断続的に生じている可能性を示している。

図5(c)は、両マイクロフォンアレイが同時に定位した音源の方向から三点測量の要領で位置を定位できた音源を、時間経過に伴って色を変えながらプロットしたものである。ケージ内で音源が集中する領域が複数見られ、一部のケージ内の個体が頻りに鳴き声を発したことが示唆される。その他、ケージを挟んで反対側に設置したマイクロフォンアレイからは、10 連ケージの個体に加え、隣接するケージの個体の鳴き声も観測され、ケージ間での相互作用の分析の可能性も示唆された。また、音源定位結果の重複状況から一個体による単独の鳴き声の抽出が可能であることも示唆された。しかし、ノイズの大きい環境である上に鳴き声が複雑であり、個体間の間隔もごく小さいため、高い精度で個体・集団レベルの鳴き声相互作用を抽出するには更なる工夫が必要な状況にあるといえる。

4 将来に向けた展望

飼育下とはいえ、東南アジア地域の不安定なネットワーク環境および電源環境でのマイクロフォンアレイによる録音は、様々な困難に直面する。短期間での実装には、安価なりチウムイオン電池やラズベリーパイに代表される簡便な録音装置開発環境の登場、またスマートフォンでも実装可能な簡易ネットワークなど技術面での運用のしやすさの向上が関与している。今回の取り組みでは、3 台の録音装置を用い、同期録音を試みたが、10 台程度の環境構築も不可能ではない。実際、筆

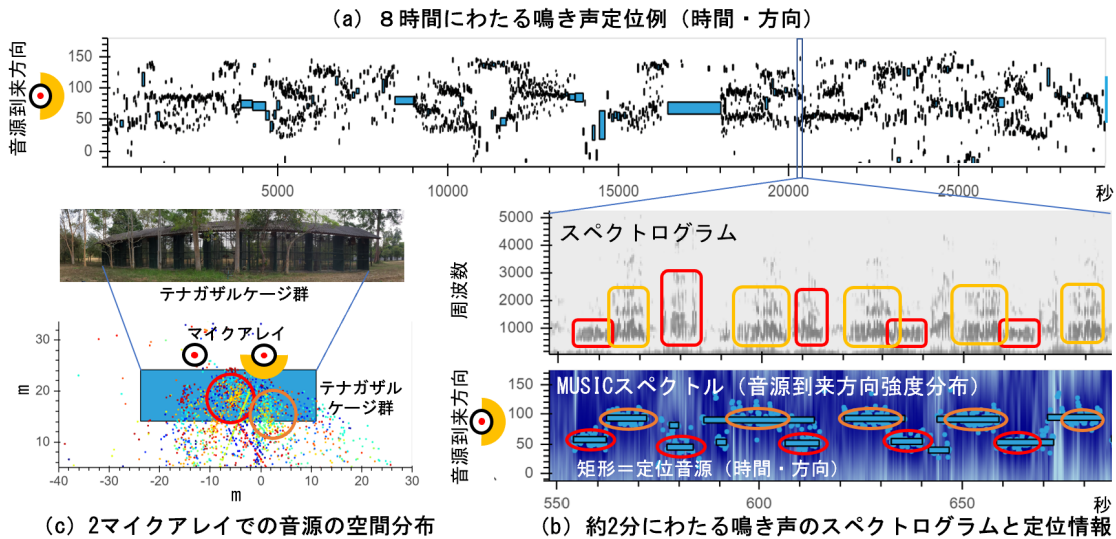


図 5: HARK で予測した音源到来方位を利用して 10 連ケージ内の個体の鳴き声を抽出した一例. 予備段階で多くの雑音等が含まれる状況に基づく結果だが, 明瞭な歌信号の特定も数多く存在する.

者たちの研究グループは, 10 台以上のマイクロフォンアレイによる同期録音を試み, 一部データを得ている. ケージという場所の情報明らかである環境でもあり, 環境計測を合わせて進めることで, 音源定位の精度向上, 最終的にテナガザル家族間の会話の可視化の実現が期待できる.

さらに, テナガザル固有の問題にも, こうした音源定位技術は将来有望な技術の一つと認識している. 筆者は 10 年ほど前に, インドネシア・スマトラ島にて, 長期間に渡り, 生態調査をおこなっていた. その調査では, 毎朝早朝に, 森に調査仲間とともに出かけ, 尾根に登り歌の聞こえる方位と時刻を, GPS と録音機を用いて, ひたすら地図上に記録していた. 観察者の計測位置を GPS により取得し, 歌の聞こえる時刻・方位を時計とコンパスによって記録した. 複数人の調査データ (歌の時刻・方位データ) を持ち合わせ, 地図上でその音源を定位し, 日々テナガザル家族の「歌地図」を手で作成していた (図 6). 歌の一記録から縄張り面積の推定などを実施したが, 以前, こうした作業の自動化は遠い将来と考えていた. 技術向上とりわけ環境構築上の簡易化と安価は, テナガザルという絶滅危惧種の社会と生態を解き明かす手段として有望であろうと期待される.

謝辞

本稿執筆にあたり, 文部科学省・新学術領域研究「共創言語進化」(#4903, 17H06380, 17H06383), および JST 戦略的創造研究推進事業 (CREST) 「人間と情報環境の共生インタラクション基盤技術の創出と展開」(「脳領域/個体/集団間のインタラクション創発原理の解明と適用」17941861 #JPMJCR17A4) からの支援を一部いただいた. また, 京都大学の森田堯博士と中部大学松田一希博士から様々な助言をいただいた.

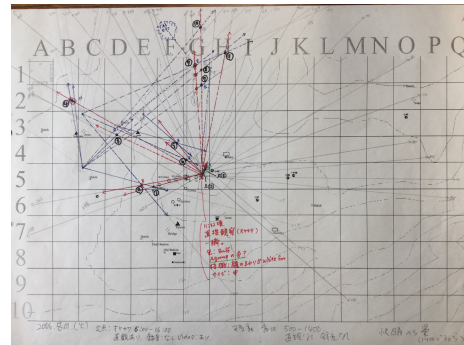


図 6: 筆者らの研究チームで自作した地図の上に, 音の定位の記録を描き, 交点を求め音源場所を推定している. 当時は, 自由に利用可能な衛星地図も少なく, 位置計測などは研究者自身で実施した. 図の中で見られるたくさんの線分は, 音が聞こえた「方位」を書き込んだものである.

参考文献

- [1] Stephen C Levinson. Turn-taking in human communication - origins and implications for language processing. *Trends in Cognitive Sciences*, 20(1):6–14, jan 2016.
- [2] Harvey Sacks, Emanuel A. Schegloff, and Gail Jefferson. A simplest systematics for the organization of turn-taking for conversation. *Language*, 50(4):696–735, dec 1974.
- [3] Nobuo Masataka and Maxeen Biben. Temporal rules regulating affiliative vocal exchanges of squirrel monkeys. *Behaviour*, 101(4):311–319, jan 1987.
- [4] Masazumi Mitani. Voiceprint identification and its application to sociological studies of wild

- Japanese monkeys (*Macaca fuscata yakui*). *Primates*, 27(4):397–412, oct 1986.
- [5] A. Margaret Elowson and Charles T. Snowdon. Pygmy marmosets, *Cebuella pygmaea*, modify vocal structure in response to changed social environment. *Animal Behaviour*, 47(6):1267–1277, jun 1994.
- [6] D Y Takahashi, A R Fenley, Y Teramoto, D Z Narayanan, J I Borjon, P Holmes, and A A Ghazanfar. The developmental dynamics of marmoset monkey vocal production. *Science*, 349(6249):734–738, 2015.
- [7] Hiroki Koda, Alban Lemasson, Chisako Oyakawa, Rizaldi, Joko Pamungkas, and Nobuo Masataka. Possible role of mother-daughter vocal interactions on the development of species-specific song in gibbons. *PLoS ONE*, 8(8):e71432, aug 2013.
- [8] A. Lemasson, L. Glas, S. Barbu, A. Lacroix, M. Guilloux, K. Remeuf, and H. Koda. Youngsters do not pay attention to conversational rules: is this so for nonhuman primates? *Scientific Reports*, 1(1):22, dec 2011.
- [9] Kazuhiro Nakadai, Toru Takahashi, Hiroshi G. Okuno, Hirofumi Nakajima, Yuji Hasegawa, and Hiroshi Tsujino. Design and implementation of robot audition system ‘HARK’ - Open source software for listening to three simultaneous speakers. *Advanced Robotics*, 24(5-6):739–761, jan 2010.
- [10] Kazuhiro Nakadai, Hiroshi G. Okuno and Takeshi Mizumoto. Development, deployment and applications of robot audition open source software HARK. *Journal of Robotics and Mechatronics*, 27:16–25, Feb 2017.
- [11] Susan Lappan and Danielle June Whittaker. *The gibbons: new perspectives on small ape socioecology and population biology*. Springer, 2009.
- [12] Hiroki Koda. Gibbon songs: Understanding the evolution and development of this unique form of vocal communication. In Ulrich H. Reichard, Hirohisa Hirai, and Claudia Barelli, editors, *Evolution of Gibbons and Siamang*, pages 347–357. Springer, 2016.
- [13] Shinji Sumitani, Reiji Suzuki, Naoaki Chiba, Shiho Matsubayashi, Takaya Arita, Kazuhiro Nakadai and Hiroshi G. Okuno. An integrated framework for field recording, localization, classification and annotation of birdsongs using robot audition techniques – HARKBird 2.0. In *Proceedings of 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP2019)*, pages 8246–8250. IEEE, 2019.