

令和 5 年 6 月 5 日現在

機関番号：12601

研究種目：挑戦的研究（萌芽）

研究期間：2020～2022

課題番号：20K21811

研究課題名（和文）音源定位技術を利用した霊長類の音声ネットワーク可視化と社会性の評価と検討

研究課題名（英文）Evaluation of primate vocal network visualization and sociality using sound source localization technology

研究代表者

香田 啓貴（Koda, Hiroki）

東京大学・大学院総合文化研究科・准教授

研究者番号：70418763

交付決定額（研究期間全体）：（直接経費） 4,900,000円

研究成果の概要（和文）：ヒト社会の動物モデルとして、コモンマーモセット家族集団を対象に、個体間の音声コミュニケーションの可視化と評価を、完全に非侵襲な方法で試みることを目的とした。特に、従来客観的評価が極めて困難であった音声のやり取りの可視化について、多チャンネルマイクロフォンアレイを制作し、ロボット聴覚技術を利用して挑戦した。先に実施した、音源定位環境の構築とその評価実験では、音声が雑音より大きく録音できるような環境では、正しく音源定位できた。つづいて実施したコモンマーモセットを対象にしたコミュニケーションの可視化実験でも、ペアの相互作用の様子を正しく可視化することに成功できた。

研究成果の学術的意義や社会的意義

先行研究で示唆されていた結果を追試できたこともあり、完全非侵襲な方法で評価できることが示され、この測定系の有用性が確認できたと言えるだろう。また、個体ペアの社会特性を反映するような行動データも収集でき、今後の社会ネットワーク解析において重要な知見になり得る可能性を提供できたと考えている。今後は、2個体間の操作実験に加え、他個体での相互作用の観察など、条件を整理して社会性の評価系の提案に貢献できる成果を示すことができたと考えられる。

研究成果の概要（英文）：The purpose of this study was to visualize and evaluate vocal communication between individuals in a common marmoset family group as an animal model of human society in a completely non-invasive manner. In particular, we challenged the visualization of vocal communication, which has conventionally been extremely difficult to evaluate objectively, by constructing a multi-channel microphone array and using robotic audition technology. In our first experiment, we constructed a sound source localization environment and evaluated it, showing that sound source localization is possible in an environment where sound can be recorded louder than the noise. In a subsequent communication visualization experiment with common marmosets, we succeeded in successfully visualizing the vocal interaction between pairs.

研究分野：実験心理学

キーワード：音声コミュニケーション 社会ネットワーク 霊長類

1. 研究開始当初の背景

ヒトは、家族や親族、地域や国家などといった様々な組織単位に所属し、階層的な社会構造を形成する。サル社会は、血縁関係を基盤としたヒト社会の基本構成集団から、社会集団(群れ)が定義できる。社会性の定量的把握は、ヒト以外の霊長類の場合、毛繕いなどの親和的な相互作用・行動の生起頻度を手がかりに社会的な関係性を定量化したり、他には敵対的攻撃交渉の2個体間での生起頻度を、毛繕いと同様の方法で定量し、その定量値から社会関係を把握する手法が主な方法である。親和性と関連性がある個体間距離も、社会関係を記述するのに重要な値として認識されてきた。

今回、我々は、サルの「会話」(= 音声コミュニケーション)を社会性指標として着目した。サルの群れの中での「会話」は、以下のようなやりとりの特徴を持っていることが知られている。例えば、群の中のサルAが発声をして、その発声Aのち一定時間内にサルBが発声が確認できたとする。そのAとBのやり取りを、「AとBの会話」が定義する。そうすることで、ヒトの会話もサルの「会話」も、二個体間の発声時間分布という基準のみを使って行動学的に定義できる。その上で、やり取りである「会話」の成立頻度を二個体間で算出できれば、音声コミュニケーションについての社会ネットワークが描けると考えられる。この音声の相互作用を基準とした社会性指標は、多くの霊長類社会で定義が可能であるため、有用な指標として考えられるが、その記録方式については、飼育条件では個体にマイクを装着するなどの多少の動物への負荷を前提とした記録が従来方式であった。

こうした中、近年、多チャンネルマイクロフォンアレイ(同時に複数の音源を同時録音するマイク装置)の技術進展や、その多チャンネル音源から音源位置を推定したり、目的の音源を分離する「音源推定・分離技術」といった機械学習技術の進展が目覚ましいものがあり、この録音装置と機械学習技術のみによる、完全な非侵襲の音源定位技術の動物音声コミュニケーションへの応用期待が高まっていた。

2. 研究の目的

こうした状況を受け、ヒト社会の動物モデルとして、コモンマーモセット家族集団を対象に、個体間の音声コミュニケーションの可視化と評価を、完全に非侵襲な方法で試みることを目的とした。特に、従来客観的評価が極めて困難であった音声のやり取りの可視化について、多チャンネルマイクロフォンアレイを制作し、ロボット聴覚技術(音源定位分離技 HARK や pyroomacoustics)を利用して挑戦することを目的とした。発声者(位置)を特定し、個体間相互作用(具体的には、集団内で「誰」と「誰」が頻繁に「会話する」か?血縁・母子間での「会話」はどの程度か?「会話」社会ネットワークで中心的な個体は定義できるか?など)の定量的評価を試みることを研究開始当初の目的とした。実際には、環境構築などでさまざまな検討項目が明らかになる中で、16チャンネルマイクロフォンアレイを自作し、その16チャンネル音源から音源分離・定位推定アルゴリズムを利用して、個体間の音声による相互作用の安定した定量化を図る技術実装とその将来性について模索し、次の研究へ発展する土台を形成することを目的とした。

3. 研究の方法

発声者特定のために準備した装置やアルゴリズムが、正しく発声者を特定しているかどうかを判定する評価系を事前準備した。そのために、京都大学旧霊長類研究所(現ヒト行動進化研究センター)で飼育されていた14頭のコモンマーモセットの夫婦(7組のオスメスペア)を、実験に使用した。この7組のペアは、通常は同じ家族で飼育されており、コドモを含めた家族集団で飼育されているものも含まれていた(実験ではペアのみを利用した)。

録音機材の中心となるのは音源定位のために必要な多チャンネル同時録音を可能とするマイクロフォンアレイである。市販のもので運用可能な安価な量販品は、ヒトの音声の処理を対象に制作されていたため、いくつか購入してその性能を確かめたが、高周波数帯域まで音声の範囲があるコモンマーモセットの音声では定位に必要な質を備えた音声録音ができなかった。そのため、まずコモンマーモセットの録音でも分析するのに十分な音質が録音できる16チャンネル同時録音可能なマイクロフォンアレイを制作した。実際には、システムズフロンティア社の提供するマイクアレイキット(RASP-ZX2-I2S)を購入し、円柱状の筐体(10cm半径 x 10cmの高さ)を3Dプリンタにより自作することで、360度にわたり簡潔で運用しやすい状態のマイクアレイ装置を製作した(Fig. 1)。このマイクアレイでは、32k Hzのサン

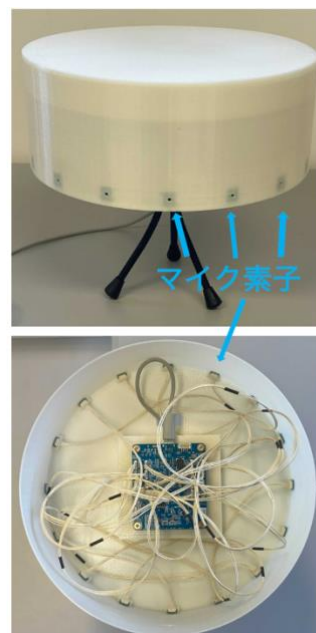


Fig. 1 マイクアレイ

プリングレートで録音できるように準備されており（これは、コモンマーモセットの発声がしばしば 10kHz を超えるため、多くの音声タイプを録音するために必要最低限のサンプリングレート値である）、コモンマーモセットの音声十分に録音できる状態であることを確認できた。

次に、このマイクを利用して安定して音源定位ができる録音・行動実験環境の構築を行った。実験は分析の障害となりうる背景雑音や音声以外の音雑音が低減できる防音室（150cm W×170cm D×220cm H）で、マイクを含めた行動実験環境を準備した。防音室内に、二つのテーブル（約 80cm の高さ）を配置し、その上に動物個体を配置するためのメッシュケージ（キャリングでき飼育室から実験室に移送が可能である）を配置できるようにした。そのメッシュケージに、対象となる実験個体が入れた。そのメッシュケージのちょう間に、設計したマイクロフォンアレイを配置して、音声の音源定位（音源の到来方位を予測）できた場合、どのケージ個体の音声だったのかを判定できるように工夫された。マイクは、防音室の外に設置されたコンピューターによって録音の開始・保存など管理が自動的に行われた。また、録音時の状況を記録するために、3 台のネットワークカメラ（TS-NS210, TS-NA220, I-O DATA 社）を設置し、音声コミュニケーション時の動画を同時記録した。

録音した音声は、データストレージに保存され、録音後、オフライン解析をして、音源定位と音声のやり取りの可視化を行った。分析は大まかには二段階に区分され、順次実行することで (Fig. 2), 定位を行った。

第一段階として、音声区間の検出を行った。音声区間検出には公開されて利用可能だった深層学習の公開コード (DeepAudioSegmenter, DAS) を利用した。DAS は、事前に十分な量の音声区間教師データをヒトが準備することで、新規な音声に対して音声区間を高精度にラベル付けするためのツールである (区間検出に加えて、音声の分類も実施でき、その音声分類についても事後的に検討した)。

次の段階として、検出された音声区間についての音源到来方向の推定を行った (音源定位)。主に二つの公開ツール (HARK あるいは pyroomacoustics) を利用して音源定位を実施したが、今回は実装が比較的簡単な pyroomacoustics (python で提供されているツール) を利用した。また、pyroomacoustics の中でも、MUSIC (Multiple Signal Classification) 法を演算手法としては選択した。この時点までで、うまく進めば、区間検出と音源定位 (発声者推定) まで実施できたとみなすことができる。

こうした環境構築をしたのち、実際の実験は二つ実施した。まず 1 つは、テーブルに配置されたメッシュケージに代わりに、2 台のスピーカを配置し、スピーカから事前に録音したコモンマーモセットの音声を、さまざまな状況下 (さまざまな音圧レベルの背景雑音条件) で再生し、上述のアルゴリズムでどの程度判定可能かどうかを検討した、音源定位測定の評価実験である。もう一つは、その評価実験での成績に基づいて、実際のコモンマーモセットの二個体ペアの音声のやり取りを定量化した、ペア観察実験であった。

4. 研究成果

1) 測定系評価実験

音源定位測定の評価実験では、白色雑音をノイズとして準備し、音声とノイズの重複率 (0% ~ 100% まで段階的に操作) とノイズの大きさ (音声の最大音圧レベルに対して、どれぐらいの音圧レベルに設定したか、-30 dB ~ 30dB まで段階的に操作) の二次元に対して条件を設定し、ノイズと音声を再生しながら、音源定位推定した到来方角と、実際の正解方位 (実験者が設定しているため、正解方位が既知) との比較を実施して、その成績を評価した。各条件で複数回の音声を再生し、その正答率を計算するために、合計で 15313 回の再生実験を実施した。到来方向と正解方位とのズレが +/-5 度未満の時、正しく定位できたと定義して、その正答率を確認したところ、重複率が 75% 以下であれば成績はどのような音圧レベルであっても 90% 以上の高い精度で正しい音源位置を推定することができた。逆に、成績が悪かったのは、100% の重複率でノイズの大きさが 0dB 以下の状況、即ち、白色雑音の音圧レベルが音声よりも大きかった時に成績が悪くなった。興味深いことに、白色雑音の音圧レベルより音声の方が大きい状況では、100% の重複率で

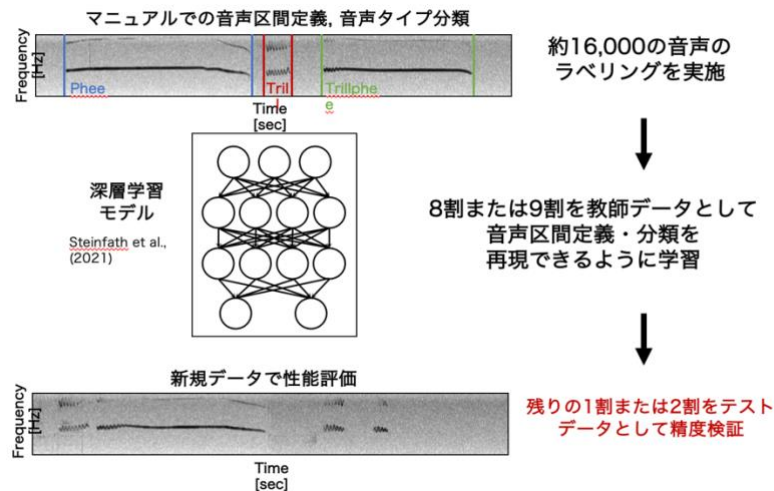


Fig.2 測定系の模式図。区間検出を深層学習で実施し、定位を行う。

も良い定位成績が見られた。また、再生する音声の種類にはよらず、一貫してこの成績であった。以上のことから、音声は雑音より大きく録音できるような環境では、正しく音源定位できることがわかった。今回の実験での録音環境は、防音室での環境であるため、この定位環境で十分に定位できる環境が備わっていると考えられた。

2) ペア観察実験

環境構築と評価実験で、音源定位の精度が十分に確認できたため、7ペアの音声交換の観察実験に移行した。実験では、さらに2条件を設定し、2-1)ペア観察視覚的確認可能条件、2-2)ペア観察視覚的遮断条件、の二つを設定した。視覚的遮断条件では、メッシュケージの間にカーテンを設置し、お互いの音声を確認できるが視覚的に確認できない状況を作り、音声の頻度や、利用される音声の種類の変動に変化が確認できるかどうかを検討した。

視覚可能条件と視覚遮断条件それぞれで、1ペアあたり300分以上の観察を行ったところ、全体で14000以上の音声区間を検出し、13500以上の音声の個体情報が決定できた。そのやり取りを分析した結果、3つのことが明らかとなった。1) ペアのうち、どちらかが主体となって発生するような非対称なやり取りが多いこと、2) 非対称性に、性差のような偏りはないこと（メスだけが多くなるとの傾向はないこと）、3) 視覚遮断条件によって、Phee call から Trill call に利用される音声タイプが変化すること、が確認できた。3) については、先行研究で示唆されていた結果であり、それがよく再現され、この完全非侵襲な方法で評価できることが示され、この測定系の有用性が確認できると言えるだろう。一方で、1) と 2) については、ペアの特性を反映するような行動データが収集できたといえ、今後の社会ネットワーク解析において重要な知見になり得る可能性を提供できたと考えられる。2個体間の操作実験に加え、他個体での相互作用の観察など、条件を整理して社会性の評価系の構築につながる有用なデータを収集できると言えるだろう。

5. 主な発表論文等

〔雑誌論文〕 計1件（うち査読付論文 1件 / うち国際共著 0件 / うちオープンアクセス 0件）

1. 著者名 香田啓貴	4. 巻 79
2. 論文標題 発声の霊長類的基盤:情動と社会集団における音声交換とその役割についての考察 -群れ社会での音声の運用に着目して-	5. 発行年 2023年
3. 雑誌名 日本音響学会誌	6. 最初と最後の頁 34-40
掲載論文のDOI (デジタルオブジェクト識別子) 10.20697/jasj.79.1.34	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

〔学会発表〕 計2件（うち招待講演 2件 / うち国際学会 0件）

1. 発表者名 香田啓貴
2. 発表標題 コミュニケーションの未来的カタチ
3. 学会等名 第10回超異分野学会（招待講演）
4. 発表年 2021年

1. 発表者名 香田啓貴
2. 発表標題 霊長類の群れ社会での音声交換と音声の情動
3. 学会等名 日本音響学会2023年春季研究発表会（招待講演）
4. 発表年 2023年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
研究分担者	鈴木 麗壘 (Suzuki Reiji) (20362296)	名古屋大学・情報学研究科・准教授 (13901)	

6. 研究組織（つづき）

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
研究 分 担 者	中村 克樹 (Nakamura Katsuki) (70243110)	京都大学・ヒト行動進化研究センター・教授 (14301)	

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関