DISS. ETH NO. 29067

TOWARDS AUTOMATED QUANTIFICATION OF VOCAL COMMUNICATION DURING SOCIAL BEHAVIORS IN SONGBIRDS

A thesis submitted to attain the degree of

DOCTOR OF SCIENCES of ETH ZURICH

(Dr. sc. ETH Zurich)

presented by

TOMAS TOMKA

MSc ETH in Biotechnology, ETH Zurich

born on 19.02.1992

citizen of Basel, Switzerland

accepted on the recommendation of

Prof. Dr. Richard Hahnloser

Prof. Dr. Dina Lipkind

Prof. Dr. Benjamin Grewe

2023

Abstract

Vocalizations are produced by highly specialized motor gestures and regulate social interactions in many species. Vocal learners, such as songbirds or humans, acquire their vocal repertoire through cultural transmission with an intriguing computational efficiency. The zebra finch, a highly social songbird, has been a model organism of outstanding importance for our understanding of vocal learning. Primarily reductionist research has generated valuable insights into molecular, neural, and behavioral aspects of vocal learning. But the combinatorial effect of social factors on cultural transmission remains largely unknown. Today, the field is transitioning towards more holistic inquiries at the social level, using big data paradigms to uncover systemic principles. However, multiple challenges need to be solved to enable conclusive longitudinal studies of entire animal groups.

Reliable vocal detection in large-scale sound data has been a longstanding problem and has served as playground for many machine learning efforts, but benchmark animal datasets with labelled vocalization boundaries are scarce. Creation of such datasets requires tedious screening for vocalizations that have been missed with machine-based approaches. The challenge of faithfully annotating vocal data aggravates when studying interactive behaviors, due to overlap of individual vocalizations and noises from animal interactions. Additionally, contextualization of vocal interactions with relatively rare and brief non-vocal events, such as copulations, previously required strenuous and time-consuming inspection of video data. Lastly, correlation-based hypotheses need to be tested for causality, which requires experimental control over individual social interactions. We tackle these challenges in threefold manner.

First, we introduce a benchmark dataset of vocal segments from single zebra finches at different developmental stages. We test how well zebra finch vocalizations can be retrieved as vocal neighbors of each other in spectrographic space, using different distance measures. Interestingly, the Spearman distance outperforms other popular distance measures such as the cosine and Euclidean distances. We find excellent performance for adults (F1 score of 0.93 ± 0.07) using 50 labelled examples (templates), but not for juveniles (F1 score of 0.64 ± 0.18), which produce highly variable vocalizations. For juveniles, the retrieval is improved when searching with equally sized overlapping template slices (F1 score of 0.72 ± 0.10), compared to searches with entire templates. As an addition to a growing array of computational tools for vocal communication research, our vocal retrieval method is useful to proofread human- or computer-annotated datasets.

Secondly, we introduce a dataset of interacting mixed-sex zebra finch couples engaging in copulations. We have found that animal-borne wireless sensors, which have been originally introduced to assign vocalizations to individuals, are highly suitable for automated copulation detection. We have observed that the female radio transmitter's carrier frequency is modulated by the physical mounting of the flying male. Copulation attempts are detected by joint occurrence of this modulation and male wing flaps. Annotating vocal and non-vocal behaviors, we find behavioral signatures signaling solicited copulations roughly 25-30 s in advance: for instance, with frequent female nest/whine calls, or changes in courtship song tempo and composition. Monitoring, or even predicting, copulations based on behavioral signatures could benefit animal caretaking and wildlife conservation programs.

Thirdly, our group has developed a system for real-time control of vocal interactions among separately housed and digitally connected animals. We have characterized vocal interactions between pairs of connected birds by the cross-covariance function and we have shown that birds engaged in reliable vocal interactions constrained by the imposed network topology. Our system and analysis could be applied in the

future to probe detailed causal relationships in vocal interactions among songbird couples or during vocal learning in juvenile birds.

Taken together, our main contribution is to democratize access to large-scale curated zebra finch datasets, which can be used in the future to train machine-based solutions to detect vocalizations or predict reproductive behaviors. Additionally, we provide a computational tool for proofreading existing datasets, and a system to manipulate vocal interactions in real-time. With these efforts, we aim to accelerate systemic insights into the structure, development, and function of vocal expressions – and positively impact human coexistence with animal wildlife.

Zusammenfassung[3]

Vokalisationen werden durch hochspezialisierte motorische Bewegungen erzeugt und regeln soziale Interaktionen bei vielen Arten. Sprech- oder Gesangslerner, wie Singvögel oder Menschen, erwerben ihr vokales Repertoire durch kulturelle Weitergabe mit einer verblüffenden rechnerischen Effizienz. Der Zebrafinke, ein äußerst sozialer Singvogel, ist ein Modellorganismus von herausragender Bedeutung für unser Verständnis des vokalen Lernens. Primär reduktionistische Forschung hat wertvolle Erkenntnisse über molekulare, neuronale und verhaltensbezogene Aspekte des Gesangslernens erbracht. Die kombinatorische Auswirkung sozialer Faktoren auf die kulturelle Weitergabe ist jedoch noch weitgehend unbekannt. Heute geht dieses wissenschaftliche Feld zu ganzheitlicheren Untersuchungen auf sozialer Ebene über, um in grossen Datensätzen systemische Prinzipien aufzudecken. Um aussagekräftige Längsschnittstudien ganzer Tiergruppen zu ermöglichen, müssen jedoch noch zahlreiche Herausforderungen gelöst werden.

Die zuverlässige Erkennung von Vokalisationen in großen Datenmengen ist seit langem ein Problem und diente als Spielwiese für viele Bemühungen im Bereich des maschinellen Lernens; jedoch gibt es nur wenige Referenzdatensätze mit abgegrenzt-annotierten Tierlauten, um maschinelle Lösungen zu testen. Die Erstellung solcher Datensätze erfordert ein mühsames Suchen nach fehlenden Vokalisationen, die bei maschinenbasierten Ansätzen übersehen wurden. Die Herausforderung, vokale Daten zuverlässig zu annotieren, verschärft sich bei der Untersuchung interaktiver Verhaltensweisen, da sich individuelle Vokalisationen überschneiden können oder von Geräuschen, die durch Tierinteraktionen verursacht werden, überdeckt werden. Darüber hinaus erforderte die Kontextualisierung von vokalen Interaktionen mit relativ seltenen und kurzen nicht-vokalen Ereignissen, wie z. B. Kopulationen, bisher eine anstrengende und zeitaufwändige Inspektion von Videodaten. Und schließlich müssen korrelationsbasierte Hypothesen auf Kausalität geprüft werden, was eine experimentelle Kontrolle über einzelne soziale Interaktionen erfordert. Wir gehen diese Herausforderungen auf dreifache Weise an.

Erstens stellen wir einen Referenzdatensatz mit Vokalsegmenten von einzelnen Zebrafinken in verschiedenen Entwicklungsstadien vor. Wir testen, wie gut Zebrafink-Vokalisationen als akustische Nachbarn voneinander im spektrographischen Raum unter Verwendung verschiedener Abstandsmaße wiedergefunden werden können. Interessanterweise übertrifft die Spearman-Distanz andere populäre Distanzmaße, wie die Kosinus-Distanz oder die euklidische Distanz. Wir erhalten ausgezeichnete Resultate für erwachsene Tiere (F1-Wert von 0.93 ± 0.07), indem wir mit 50 annotierten Beispiel-Vokalisationen (Schablonen) suchen, aber nicht für Jungtiere (F1-Wert von 0.64 ± 0.18), die sehr variable Vokalisationen produzieren. Bei Jungtieren verbessern sich die Resultate durch das Suchen mit gleich großen, sich überlappenden Schablonen-Scheiben (F1-Wert von 0.72 ± 0.10) im Vergleich zur Suche mit ganzen Schablonen. Als Ergänzung zu einer wachsenden Anzahl von computergestützten Werkzeugen für die Erforschung der vokalen Kommunikation ist unsere Methode nützlich, um von Menschen oder Computern annotierte Datensätze zu überprüfen.

Zweitens stellen wir einen Datensatz von kopulierenden gemischt-geschlechtlichen Zebrafinkenpaaren vor. Wir haben herausgefunden, dass die von Tieren getragenen Funksensoren, die ursprünglich eingeführt wurden, um Vokalisationen einzelnen Individuen zuzuordnen, sehr gut für die automatische

[3] I have used DeepL (Kutylowski, 2017) for an initial translation of my original "Abstract" written in English. I modified the automatic German translation with manual corrections.

Kopulationserkennung geeignet sind. Wir haben beobachtet, dass die Trägerfrequenz des weiblichen Funksenders durch das physische Besteigen seitens des fliegenden Männchens moduliert wird. Kopulationsversuche werden durch das gemeinsame Auftreten dieser Modulation und der Flügelschläge des Männchens erkannt. Bei der Analyse von vokalem und nicht-vokalem Verhalten finden wir Verhaltenssignaturen, die auf eine einvernehmliche Kopulation etwa 25-30 Sekunden im Voraus hinweisen: z.B. häufige Nest- und Heulrufe der Weibchen oder Veränderungen in Tempo und Zusammensetzung des Balzgesangs. Die Registrierung oder gar die Vorhersage von Kopulationen auf der Grundlage von detektierten Verhaltenssignaturen könnte für Tierpflege- und Wildtierschutzprogramme von Nutzen sein.

Drittens hat unsere Gruppe ein System für die Echtzeitkontrolle von vokalen Interaktionen zwischen getrennt untergebrachten und digital verbundenen Tieren entwickelt. Wir haben vokale Interaktionen zwischen jeglichen zwei Vögeln durch die Kreuzkovarianzfunktion charakterisiert und gezeigt, dass die Vögel zuverlässige vokale Interaktionen durchführen, die durch die vorgegebene Netzwerktopologie eingeschränkt werden. Unser System und unsere Analyse könnten in Zukunft eingesetzt werden, um detaillierte kausale Zusammenhänge bei vokalen Interaktionen zwischen Singvogelpaaren oder während des Gesangslernens bei Jungvögeln zu untersuchen.

Insgesamt besteht unser Hauptbeitrag darin, den Zugang zu großen kuratierten Zebrafink-Datensätzen zu demokratisieren. Diese können in Zukunft verwendet werden, um maschinelle Lösungen zur Erkennung von Vokalisationen oder zur Vorhersage des Fortpflanzungsverhaltens zu trainieren. Darüber hinaus stellen wir ein computergestütztes Werkzeug zum Korrekturlesen bestehender Datensätze und ein System zur Echtzeit-Manipulation von vokalen Interaktionen bereit. Mit diesen Bemühungen wollen wir ganzheitliche Erkenntnisse über die Struktur, Entwicklung und Funktion von Vokalisationen beschleunigen – und das Zusammenleben von Mensch und Tier positiv beeinflussen.