Exploring the form and functions of chimpanzee pant-hoots from basic evolutionary principles.

A DISSERTATION SUBMITTED TO THE FACULTY
OF THE UNIVERSITY OF MINNESOTA BY

Nisarg P. Desai

IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE OF DOCTOR OF
PHILOSOPHY

Dr. Michael L. Wilson

May 2022

Acknowledgements

First of all, I am deeply grateful to my advisor Dr. Michael Wilson for years of advice and support that made this dissertation possible. I have learned a lot from many discussions, seminars, and lectures with Mike that I have had the privilege to be a part of and also from spending time in the field, which I will cherish forever. His insight and patiently objective outlook towards science, and life in general, has made be a better scientist and a better human being. I would also like to thank my committee members: Drs. Ben Munson, Clarence Lehman, Kieran McNulty, and Martha Tappen. I thank Kieran for his spectacular teaching in the quantitative methods class that has played a key role in giving me the confidence to further explore and grow my quantitative skills. It set the trajectory for this dissertation, and will have a lasting impact on everything I do for the rest of my career. I thank Ben for discussions and readings that have also been crucial in guiding this dissertation and preparing me with the necessary skillset required for performing the acoustic analyses in all the chapters. I am also thankful to Ben for introducing me to Dr. Matthew Winn, who I thank for answering many of my silly questions and for all the help on acoustic analyses. I thank Clarence for many discussions and teachings that were important for this dissertation. Conversations with Clarence, and his lectures have the highest rate of insights-per-minute that I have ever encountered, and I always have my mind blown multiple times at each interaction with him. He never ceases to amaze me and gets me excited about being a scientist. I thank Martha for expanding my understanding about anthropology and for feedback on my prelims that helped me be a better anthropologist.

I have a lot of collaborators and colleagues to thank that have made this dissertation possible. I thank Drs. Kurt Hammerschmidt and Julia Fischer for hosting me in Germany and for providing advice and software tools for acoustic analyses. I thank Drs. Pawel Fedurek, and Katie Slocombe for sharing chimpanzee vocalizations and for constructive feedback on my manuscripts. I also thank Drs. Lisa O’Bryan, Charlotte Uhlenbroek, Adam Clark Arcadi, Hetty van de Rijt Plooij, and Frans Plooij for sharing chimpanzee vocalizations. I thank Drs. Elizabeth Lonsdorf, Joseph Feldblum, and Ian Gilby for sharing data, software, maintaining the Gombe database, and for advice on analyses. I thank Maggie Hoffman for sharing data on short notices. I want to thank the researchers and staff at the Gombe Stream Research Center, and especially Drs. Anthony Collins, Dismas Mwacha and Deus Mjungu for facilitating the logistics of data collection and for great company during my time at Gombe. Thanks to Ketura for keeping me fed at Gombe. Special thanks to Hashimu Salala and Nasibu Madumbi for their help with the field work, teaching me about chimps, and for all the fun Swahili conversations in the field.

I want to thank the community at the Department of Anthropology for years of academic, administrative, and financial support. Thanks to Kara Kersteter, Nora Last, Christina Wiencke, and especially Megan Whaley who have patiently answered all my questions and made my life as an international student so much easier. I would also like to thank the Department of Anthropology as well as the University of Minnesota for funding my research through various block grants, Graduate Research Partnership Program Fellowship, Interdisciplinary Doctoral Fellowship, and Doctoral Dissertation Fellowship. Thanks to Mike Wilson and the University of Minnesota for funding my

fieldwork through the Talle Faculty Research Award. Thanks to Drs. Ben Hayden and Jan Zimmermann for their generosity and support during my last semester.

I want to thank the undergraduate and graduate student community from anthropology and ecology, evolution, and behavior departments for years of friendship as well as collaboration. Thank you, Maud, Kristy, Carrie, Tony, Rebecca, Samuel, Somaye, and Rutger for making my time in grad school enjoyable. I also want to thank research interns Theresa Woodward, Rachel Pladson, Stephanie Olsen, Sydney Baum-Haines, Maggie Edge, Grant Piepkorn, Gretchen Remus, Rachel Aron, Joseph Holdreith, Lili Hagg, Kelley Hoiseth, Jordan Loy, Jonah Bacon, Annika Herdtle, and Marissa Glazos for helping me with the data processing. Thanks to Ava Fox and Brady MacKay for allowing me to supervise their senior theses and for help with the data processing. It has been a pleasure working with you all.

Lastly, I am immensely grateful to my parents Parimal and Geeta Desai for their unconditional love and support. There are simply no words in language, as complex as it is, to describe my love and gratitude for my sister Ruta. One of the best things about grad school was having breaks that allowed me to spend time with you.

Dissertation Abstract

Researchers have studied chimpanzee vocal communication extensively, focusing on evidence of parallels with human language. This approach has been effective in encouraging vocal communication research and providing some insights about the evolution of language. However, it has obscured our understanding of non-human animal communication by motivating researchers to adopt a problematic conceptual framework that uses complex linguistic phenomena as models for simpler primate vocal communication mechanisms. An approach focusing on basic evolutionary principles involves studying the intimate connection between form and function to obtain insights about the biological and evolutionary origins and mechanisms of traits. Such an approach, when employed for studying chimpanzee vocalizations, may be more fruitful in revealing fundamental factors that may shape their vocalizations.

This dissertation extends our knowledge of the forms and functions of chimpanzee vocal communication. I first explored different acoustical and statistical analysis methods for describing the form of vocalizations. Next, I studied connection of the form of chimpanzee vocalization, the pant-hoot, to its possible functions. Using audio recordings and behavioral data from two chimpanzee communities in Gombe National Park, Tanzania, and one chimpanzee community in Kibale National Park, Uganda, I tested if the variation in chimpanzee calls is explained primarily by (i) community membership, or (ii) by individual traits such as age, rank, and health, and (iii) if any of these acoustic cues predicted male mating success. Individual traits better explained the acoustic variation in pant-hoots than community membership. Acoustic variation also reflected male mating success. These findings suggest that sexual selection is a key evolutionary force shaping chimpanzee vocalizations.

Table of Contents

List of Tables

List of Figures

Dissertation Introduction

Vocal communication research focuses on obtaining insights about the forms, functions, and developmental mechanisms of vocal signals; their behavioral biology; and on obtaining comparative data to understand the evolutionary history of vocal communication in the human lineage. These goals are derived from traditional ethological theory that considers signals as stimuli that elicit certain behaviors in receivers. This conceptual framework, that evolutionary anthropologists commonly employ, assumes a close relationship between the form and function of signals. Indeed, much of the early research on vocal signals focused on this relationship (Marler, 1955, 1958, 1968). However, in recent decades, primate vocal communication research has mainly focused on comparisons with human language and on finding language-like properties in primate calls. This focus is in part due to the close phylogenetic relationship of non-human primates to humans, and in part inspired by Seyfarth, Cheney, and Marler’s seminal work on semantic properties of vervet monkey alarm calls (Seyfarth et al., 1980a, 1980b). This approach has been fruitful in inspiring extensive interest in empirical studies of primate vocal communication. Nevertheless, it has limited progress by inspiring the use of a teleological conceptual framework that employs complex linguistic phenomena as models for simpler vocal communication patterns of non-human primates; and by distracting from the foundational ethological conceptual framework that focuses on the intimate relationship between form and function (Owren & Rendall, 2001; Wheeler & Fischer, 2012). As chimpanzees are one of the two most closely related species to humans, many studies of chimpanzee vocal communication have especially focused on exploring language parallels.

With the goal to inspire and conduct primate vocal communication research with more robust philosophical foundations, this dissertation is my attempt to study chimpanzee vocal communication with a focus on basic evolutionary principles: exploring the form, and then studying the relationship between form and function. It is organized in three chapters. In the first chapter, I explored the form of chimpanzee vocalizations. I compared acoustical and statistical analysis techniques traditionally used in primatology with newer techniques employed in other fields such as speech recognition and machine learning to determine the best quantitative techniques to describe from and study the functions. In the next two chapters, I examined the functions of what is arguably the most complex chimpanzee vocalization, the pant-hoot. Inspired by Platt's strong inference framework (Platt, 1964), I proposed and tested two alternative hypotheses informed by contemporary literature. The first hypothesis, called *the vocal learning hypothesis*, is that the acoustic structure of pant-hoots reflects socially learned components that function to signal community membership. The second hypothesis, called *the sexual selection hypothesis*, is that the variation in acoustic structure of the pant-hoot calls reflects a combination of individual features including body size, age, rank, and health and predicts a male's mating success. While these hypotheses are not mutually exclusive, they provide a framework to study a vocalization that serves many socioecological functions in chimpanzee societies. Below, I describe the three chapters in more detail.

The first chapter focuses on studying the form of chimpanzee vocalizations through an exploration of new tools, examining their potential to improve our understanding of form. I initially conducted a pilot analysis focused on the utility of new acoustical analysis and machine learning techniques and presented my findings at the Acoustical Society of America meetings (Desai et al., 2018). To expand on these findings, I visited the Georg-August-Universität Göttingen, Germany in February-March 2020 to collaborate with Dr. Kurt Hammerschmidt, who has developed a software package, LMA, to study primate vocalizations (Schrader & Hammerschmidt, 1997). Researchers studying animal vocalizations typically begin with classifying calls into different categories. They may begin this classification by ear, with refinement or confirmation of categories using software packages such as LMA to describe the vocal repertoire of the species. In chimpanzees, researchers qualitatively described chimpanzee vocal repertoire as early as 1965 (Reynolds, 1965), and produced first systematic descriptions of chimpanzee vocal repertoire as early as 1968 (Goodall et al., 1968; Marler, 1969). Such work involves classifying vocalizations by listening to calls, inspecting spectrograms, and by observing the behavioral contexts in which the calls are produced. Such qualitative approaches to describe the form of vocalizations and classify the vocal repertoire of chimpanzees are useful and present an important baseline to inform chimpanzee vocal communication research. However, qualitative approaches are subject to human biases and are limited by human senses. For instance, humans can only identify and categorize call types that are in the audible range of frequencies of human hearing. This can potentially influence classifications as humans only perceive frequencies in a narrow range that human ears are capable of and do not perceive all frequencies equally. Quantitative approaches might help overcome such human limitations. As computers became mainstream and software to study vocalizations became more accessible, researchers have used a mix of qualitative and quantitative methods to study animal vocalizations (Hammerschmidt & Fischer, 1998; Winter et al., 1966, 1973). However, efforts to quantitatively describe the vocal repertoire of chimpanzees remain in early stages (Crockford, 2019). This presents a challenge in studying the functions of chimpanzee calls in a way that is replicable. Chimpanzee vocalizations are highly variable and often intergrade into different call types like screams, hoos, and hoots, which may be differentially meaningful in chimpanzee social life. A quantitative approach could help in describing the variation in pant-hoots more reliably by picking up on subtle differences in the vocalizations. As time has progressed, technological developments have enabled new ways to analyze and study animal vocalizations. For example, industrial software such as Avisoft SASLab Pro (Specht, 2004), or software developed by academics such as Raven Pro (Dugan et al., 2016), LMA (Fischer et al., 2013; Schrader & Hammerschmidt, 1997), Praat (Boersma & Weenink, 2018) and others are often used to study primate vocalizations. Recent developments in fields such as speech recognition and machine learning (Giannakopoulos, 2015; Hinton et al., 2012; Ittichaichareon et al., 2012) provide promising avenues to study primate vocalizations, yet these are relatively underutilized in primatology (Mielke & Zuberbühler, 2013; Pozzi et al., 2010). Speech recognition research has resulted in acoustic features known as mel frequency cepstral coefficients that incorporate human auditory perception (Ittichaichareon et al., 2012; Mermelstein, 1976). They account for the fact that some frequencies are better perceived than others by humans, and hence assist in better categorization of sounds, in a way that is more compatible with human hearing. These new techniques are promising, but we need to test whether they make a difference. Hence, in this chapter, I compared the performance of these techniques, including mel frequency cepstral coefficients for measuring chimpanzee vocalizations and machine learning techniques for analyzing chimpanzee vocalizations with that of existing approaches commonly used in primatology such as LMA for sound measurement, and discriminant functions analysis for statistical hypothesis testing. As my data collection was still in progress at the time of beginning this study, I used publically available archival recordings of chimpanzee vocalizations obtained by Dr. Hetty van de Rijt Plooij at Gombe National Park, Tanzania between 1972-73 (Plooij et al., 2015). For acoustic feature extraction, I used (i) pyAudioAnalysis, a software package written in Python (Giannakopoulos, 2015) for extracting mel frequency cepstral coefficients and some other acoustic features, and (ii) LMA for extracting commonly used acoustic features in primatology (Schrader & Hammerschmidt, 1997). I compared these acoustic features using some of the best performing machine learning techniques using the statistical software R. Lastly, I used the results obtained in this chapter to inform which acoustic features and statistical analysis techniques to use to study the functions of chimpanzee pant-hoots in the remaining chapters in the thesis.

In the second chapter, I tested *the vocal learning hypothesis*: chimpanzee pant-hoots provide reliable cues of community membership and function to advertise territory ownership. This was inspired from previous studies that have reported regional variation (‘dialects’) in the pant-hoot calls of chimpanzees (Crockford et al., 2004; Marshall et al., 1999; Mitani et al., 1992, 1999). These dialects have been proposed to reflect learned differences in acoustic structure, resulting from vocal matching of the pant-hoots of other community members. If chimpanzees can learn acoustic features from conspecifics, it suggests that chimpanzees have vocal learning capacity: the ability to voluntarily modify the acoustic structure of vocalizations, which is a key component of human spoken language. This hypothesis thus reflects the traditional approach that focused on the search for language parallels, but I take this approach to reassess and replicate previous findings for the following reasons. The group-specific differences in the acoustic structures reported in the previous studies are subtle and account for a relatively small proportion of the total acoustic variation in these calls (Mitani & Brandt, 1994). Additionally, other alternative explanations for the existence of geographic variation such as habitat acoustics, sound environment, and body size differences are not adequately explored (Mitani et al., 1999). Furthermore, compared to other species such as songbirds and humpback whales that can learn elaborate songs, and parrots that can mimic human speech, the vocal learning capacities of nonhuman primates appear much more limited, raising questions about how and why this capacity evolved in the human lineage (Vernes et al., 2021). Thus, while chimpanzee ‘dialects’ are widely accepted as evidence of vocal learning, I consider this matter far from settled, and in need of further study. As the reported existence of vocal learning capacity in one of our two closest living relatives suggests that some capacity for vocal learning may have existed in the last common ancestor of humans and chimpanzees, the existence of dialects and vocal learning capacities in chimpanzees remains interesting. Furthermore, these could suggest other functions of pant-hoots such as signaling territory ownership or facilitate social bonding (Fedurek, Schel, et al., 2013; Mitani & Gros-Louis, 1998; Wilson et al., 2007) Hence, I attempted to replicate the previous findings of dialects in chimpanzee pant-hoots to gain functional insights.

Several studies from the field (Arcadi, 1996; Crockford et al., 2004; Mitani et al., 1992) and captivity (Marshall et al., 1999) have reported the existence of regional variation ("dialects") in pant-hoots. In the first study of chimpanzee dialects, Mitani and colleagues reported differences between pant-hoots from two National Parks in Tanzania: Gombe and Mahale. They suggested that the differences may be an outcome of vocal learning (Mitani et al., 1992). However, the differences between the communities were subtle, with a subset of acoustic features showing community-specific differences. In another study, Mitani and Brandt found that community membership accounted for only 11% of variation in acoustic structure (Mitani & Brandt, 1994). Mitani later reassessed his findings by examining the pant-hoots from Mahale and Kibale National Parks in Tanzania. He pointed out that since Mahale and Kibale are relatively far from one another (~700 km), and genetically isolated, the acoustic differences may not necessarily represent vocal learning, but could instead represent genetic differences (Mitani et al., 1999). Additionally, other differences like habitat acoustics, sound environment, or body size might be more important in explaining the variation in such geographically distant communities. To test the extent to which the acoustic structure of calls specifically signals community membership, we need to record vocalizations from known individuals from neighboring communities, in order to control for other external differences like Mitani and colleagues mentioned. This still does not rule out genetic differences as an alternative explanation, which needs to be controlled for separately. Another field study of chimpanzee dialects compared individuals in three neighboring communities and one more distant community in Taï National Park, Côte d'Ivoire, and also controlled for genetic differences (Crockford et al., 2004). This study thus supports the view that social learning accounts for the acoustic differences among communities. However, sample sizes in this study were limited, with calls from only three individuals per group analyzed. A larger sample of individuals from different field sites across Africa are needed to be confident that differences in acoustic structure among communities are not the result of chance.

I visited Gombe first in the summer of 2016 to train Tanzanian field assistants Hashim Salala and Nasibu Madumbi to record chimpanzee calls from two neighboring communities at Gombe: Kasekela and Mitumba. After my return, I coordinated the data collection via Skype calls with Hashimu and revisited Gombe in the summer of 2017 to record more calls. Salala and Madumbi recorded vocalizations from August 2016 to December 2017. Furthermore, I collaborated with Dr. Pawel Fedurek and Dr. Katie Slocombe and obtained pant-hoot recordings from a geographically distant chimpanzee community, Kanyawara, from Kibale National Park, Uganda collected during 2010-2011 (Fedurek, Schel, et al., 2013). I tested for differences among these neighboring and a geographically distant community.

Whether chimpanzee pant-hoots have community-specific dialects or not, several fundamental questions about the form and functions of chimpanzee pant-hoots remain unexplored due to the focus on finding language parallels. Specifically, comparative considerations of the functions of loud calls in other species suggest that pant-hoots may provide cues of individual traits related to mate quality, such as body-size, age, rank, and health. Pant-hoots thus may function as sexually selected signals, providing information about male mate quality, rather than community membership. Hence, in the third chapter, I tested *the sexual selection hypothesis*: chimpanzee pant-hoots are sexually selected signals that indicate male competitive ability and mate quality, and the variation in pant-hoot acoustic structure can be explained by individual traits that also signal mate quality like age, rank, health, and predicts male mating success. Several aspects of pant-hoots indicate that they may be sexually selected signals of male competitive ability. Males produce pant-hoots much more frequently than females (Mitani & Nishida, 1993; Wilson et al., 2007); high-ranking males call more often than low-ranking males (Clark & Wrangham, 1994; Fedurek et al., 2014; Mitani & Nishida, 1993; Wilson et al., 2007); males with high testosterone levels produce more pant-hoots (Fedurek, Slocombe, et al., 2016); and males produce more pant-hoots in the presence of parous swollen females (Fedurek et al., 2014). Riede and colleagues suggested that pant-hoots signal male physical condition through the presence of non-linear phenomena (NLP)—producing calls in the upper frequency range results in NLP in the calls if the male is not in peak physical condition. Individuals that are young and healthy may have vocal folds that are better able to maintain the production of linear acoustics at high frequencies than unhealthy individuals (Riede et al., 2004, 2007). Insofar as chimpanzee pant-hoots are sexually selected signals of male mate quality and competitive ability, their acoustic structure, i.e., form, may be expected to correlate with other cues of male mate quality such as age, rank, health. Additionally, if females choose males based on these traits, the acoustic variation could function to signal mate quality and correlate with male mating success.

I used the recordings obtained from Kasekela and Mitumba communities in Gombe, and Kanyawara community in Kibale, obtained for the previous chapter and measured other acoustic features relevant for the sexual selection hypothesis. I measured the prevalence of non-linear phenomena in the climax-screams of the pant-hoots as they may signal physical condition (Riede et al., 2007). I first tested if the prevalence of non-linear phenomena was associated with the age of the individual. I collaborated with other researchers from Gombe and Kibale to obtain additional recordings and compile a longitudinal dataset of chimpanzees that allowed me to test the relationship with age more robustly. From Gombe, I collaborated with Dr. Charlotte Uhlenbroek, who recorded chimpanzee vocalizations between 1992-1993 (Uhlenbroek, 1996) and Dr. Lisa O'Bryan, who recorded chimpanzee vocalizations between 2012 and 2013 (O'Bryan, 2015). From Kibale, in addition to Dr. Pawel Fedurek and Dr. Katie Slocombe who I collaborated with for chapter 2, I collaborated with Dr. Adam Clark Arcadi, who recorded chimpanzee vocalizations during 1988 (Clark & Wrangham, 1993) and Dr. Michael Wilson, who recorded chimpanzee vocalizations between 1997-1998 (Wilson, 2002). I obtained the estimated or recorded birth dates of individual chimpanzees from the long-term databases at Gombe and Kibale to calculate their age at the time of recording. Next, I tested if the prevalence of non-linear phenomena and other acoustic features including the fundamental frequency and noise were explained by the rank and the health condition of Gombe chimpanzees, and if they predicted the mating success of males. I collaborated with Dr. Joseph Feldblum and Dr. Ian Gilby to obtain the data about rank and mating success, and Dr. Elizabeth Lonsdorf to obtain the data on the health condition of Gombe chimpanzees (Lonsdorf et al., 2018). I only tested Gombe chimpanzees for rank, health, and mating success as these data were not available from Kibale.

Ethical note

The research reported in this dissertation is based on data collected non-invasively from free-ranging chimpanzees. The Institutional Animal Care and Use Committee at the University of Minnesota did not require a review due to the purely observational nature of the research. Research at Gombe National Park was performed with approval from the Tanzania Wildlife Research Institute and the Tanzania Commission for Science and Technology and adhered to additional ethical guidelines set by the Jane Goodall Institute.

Research at Kibale National Park was approved by the Department of Psychology Ethics panel at the University of York and permission to conduct the study was granted by the Ugandan Wildlife Authority and the Ugandan National Council for Science and Technology. The study complied with the laws of Uganda.

Chapter 1

A comparison of quantitative classification techniques for studying chimpanzee vocalizations

1.1 Overview

Studying animal vocalizations requires classifying biologically meaningful call types and describing the vocal repertoire of a species. Quantitative tools for statistical classification and acoustic analysis have been evolving with developments in machine learning and speech recognition research as well as increasing computing power. There are two main methodological considerations in analyzing animal vocalizations: (i) choosing which acoustic features to use for measuring the vocalization, and (ii) choosing statistical modeling techniques for analysis. There is substantial variation in these choices both within and among different scientific fields that deal with sound analysis. In fields that deal with animal vocalizations, such as primatology, animal behavior, psychology, etc., researchers mainly measure acoustic features related to the fundamental and the peak frequencies of a defined call unit, energy distribution in the spectrogram and measures of the composition of different kinds of call units. While these features provide some useful insights about animal communication, they tend to have limitations in terms of their reliability and validity for biological questions. Developments in speech recognition research have resulted in acoustic features known as mel frequency cepstral coefficients

(MFCCs). MFCCs analyzed using new machine learning techniques have proven to be highly reliable for human speech-recognition. Their successes have prompted bioacoustics researchers to apply these tools to problems including individual-, species-, and call type identification, and vocal repertoire classification. However, researchers studying non-human primate vocalizations have only recently started adopting these approaches and none have applied them to study chimpanzee vocalizations. In this chapter, I tested if these new techniques provide improvements over commonly used acoustic feature extraction and statistical classification techniques in primatology. I analyzed archival recordings of vocalizations from Gombe National Park, Tanzania. I extracted acoustic features related to the fundamental and the peak frequency using the commonly used software LMA (LMA features), and MFCCs and other speech recognition acoustic features using a Python package called pyAudioAnalysis (speech recognition features). I tested the performance of different supervised (discriminant function analysis, random forests, and artificial neural networks) and unsupervised (k-means clustering, and principal components analysis) machine learning techniques, in terms of their precision, recall, and accuracy in call type identification. I found that the random forests performed the best overall with LMA features, and artificial neural networks performed the best overall with speech recognition features. However, neither kind of acoustic features were decidedly superior overall. Furthermore, for pant-hoots, discriminant function analysis applied on LMA features performed the best, implying that it is the best combination of techniques for other chapters in this dissertation. Combined these results indicate that speech recognition features or other machine learning techniques have comparable performance to the traditional techniques. For replication studies, employing traditional techniques is preferable as it facilitates comparisons with previous studies. However, the promising performance of 1-hidden layer artificial neural networks applied on speech recognition features using a small dataset warrants further investigation of deeper artificial neural networks applied on speech recognition features using larger datasets.

1.2 Introduction

Classifying vocal repertoire is the first step to studying vocal communication in animals. Before making any interpretations about the causes and functions of the variation in the vocalizations, it is imperative to have a proper understanding of the set of calls that a species produces in natural settings. Vocal repertoire serves as a common framework that describes the variation in vocalizations and allows researchers to conduct studies and communicate the observations about vocal communication behaviors. Researchers commonly describe vocal repertoire of a given species using descriptions of human-perceivable call types produced by the species. Call types are the fundamental unit that researchers use for framing research questions. For example, studies exploring functions and underlying behavioral ecology of chimpanzee vocal communication rely on descriptions of discrete call types such as ‘pant-hoots’, ‘screams’, ‘rough-grunts’, ‘pant-grunts’ and others. Marler studied chimpanzee ‘pant-hoot’ calls based on descriptions by Goodall and himself (Goodall et al., 1968; Marler, 1969). He researched individual specificity as well as sex differences in the pant-hoots (Marler & Hobbett, 1975). Clark and Wrangham found that the acoustic structure of the pant-hoots varied based on the social, and not ecological context, and found no evidence of an acoustically distinct food arrival pant-hoot (Clark & Wrangham, 1993). Fedurek and colleagues found that pant-hoots were more frequent in presence of parous swollen females, were given while consuming high-quality foods, facilitated male fusion in parties (Fedurek et al., 2014), and that pant-hoot chorusing is associated with male bonds (Fedurek, Machanda, et al., 2013). Mitani and colleagues found that pant-hoots encoded individual identity better than pant-grunts (Mitani et al., 1996). Studies on chimpanzee ‘screams’ have found that chimpanzees can distinguish different scream types based on the severity of aggression experienced by the caller (Slocombe et al., 2009), can modify scream type based on the composition of the audience (Slocombe & Zuberbühler, 2007), and can obtain social information about victims and aggressors based on the acoustics and context of the screams (Slocombe et al., 2010). Studies on food associated ‘rough-grunts’ have found that they are functionally referential i.e. chimpanzees produce acoustically distinct rough-grunts based on their preference of food (Slocombe & Zuberbühler, 2005). Additionally, rough-grunts are directed specifically at high ranking individuals (Schel et al., 2013), and also might exhibit vocal learning (Watson et al., 2015a). Laporte and Zuberbühler studied ‘pant-grunts’ and found that they are flexible vocal signals that vary based on the social situation; for instance, the alpha male received a higher proportion of pant-grunts from females and females gave more pant-grunts to other males when the alpha male was absent from the party (Laporte & Zuberbühler, 2010).

Some early studies looking at primate vocalizations used qualitative approaches for classifying vocal repertoires. Researchers classified vocalizations by ear or by visually inspecting spectrograms (see for example rhesus macaques (*Macaca mulatta*): (Rowell & Hinde, 1962); squirrel monkeys (*Saimiri sciureus*): (Winter et al., 1966); chimpanzees (*Pan troglodytes schweinfurthii*): (Goodall et al., 1968; Marler, 1969, 1976) and mountain gorillas (*Gorilla beringei beringei*): (Fossey, 1972)). In addition, describing vocal repertoires involved classifying the call types based on presumed function or the commonly observed contexts and behaviors during which the call is produced e.g. 'food calls', 'alarm calls', 'contact calls' etc. (de Waal, 1988; Hammerschmidt & Fischer, 1998). While these are immensely useful first steps to classification, such qualitative descriptions are susceptible to observer biases. Different human observers might classify call types differently based on individual human variation in hearing capacities and may be subject to biases such as confirmation bias, affecting the replicability. Because of a lack of standard quantitative descriptions of call types, researchers may not be able to accurately account for individual differences in call structures, which can lead to unreliable conclusions (Arcadi, 2005). Defining call types based on context hinders the ability to distinguish between responses elicited based on contextual cues and responses based on acoustics alone (Wheeler & Fischer, 2012). This may lead to erroneous conclusions about important issues. For example, in chimpanzee vocal communication, debates about important issues such as functional reference and vocal learning prevail. For instance, some researchers argue that food associated rough-grunts are functionally referential (Slocombe & Zuberbühler, 2005) and exhibit vocal learning (Watson et al., 2015a); whereas others argue that they may not be functionally referential or exhibit vocal learning and instead reflect arousal in captive contexts (Clay et al., 2012; Fischer et al., 2015); or in wild contexts, reflect the signaler's foraging or social intentions (O’Bryan et al., 2018). In any case, it is clear that there are multiple different potential call types that are getting lumped into ‘rough-grunts’ (O’Bryan, 2015; Slocombe & Zuberbühler, 2005), which is likely due to the substantial gradation observed in chimpanzee vocalizations (Marler, 1976). Graded vocalizations make it difficult to classify distinct call categories that researchers could use to communicate reproducible results.

Quantitative approaches to classification show promise in mitigating the problems associated with traditional qualitative approaches. Researchers have identified the need for quantitative approaches and have employed them for various studies exploring vocal communication in primates. For example, researchers have used a variety of different acoustic analysis and statistical classification techniques classifying vocal repertoires, including k-means clustering to classify graded call types in barbary macaques, *Macaca sylvanus* (Hammerschmidt & Fischer, 1998); discriminant function analysis to classify ten different call types of crowned lemurs, *Eulemur coronatus* (Gamba & Giacoma, 2007); a combination of hierarchical cluster analysis and discriminant function analysis for six call types of blue monkeys, *Cercopithecus mitis stulmanni*: (Fuller, 2014); cluster analysis to objectively describe vocal repertories of mountain gorillas, *Gorilla beringei beringei*, and western lowland gorillas, *Gorilla gorilla gorilla* (Hedwig et al., 2014); a combination of k-means clustering and support vector machines for describing the vocal repertoire of common marmosets, *Callithrix jacchus* (Agamaite et al., 2015); a combination of principal components analysis and discriminant function analysis for classifying eleven call types of Tibetan macaques, *Macaca thibetana* (Bernstein et al., 2016) and so on. Such quantitative techniques for call type classification could then facilitate further insights into the behavioral ecology of a species when behavioral and contextual information is incorporated along with the acoustic parameters. For example, when brought in contact with another population of their species, pygmy marmosets, *Cebuella pygmaea*, modified the acoustic structure of their contact call trill (Elowson & Snowdon, 1994) and Wied's black tufted-ear marmosets, *Callithrix kuhlii,* modified the acoustic structure of their phee calls (Rukstalis et al., 2003). Bouchet and colleagues quantitatively classified different call units of De Brazza's monkeys, *Cercopithecus neglectus,* and then concatenated these units into 10 different call combinations, of which 3 were found to be produced by both sexes and at all ages (Bouchet et al., 2012). Similarly, they concatenated call units of red-capped mangabeys, *Cercocebus torquatus,* into 8 call combinations that exhibit sex-differences in call production (Bouchet et al., 2010). Quantitative approaches can also be used to study the gradation within and among call types, which lets researchers make sense of vocalizations when it is hard to draw clear boundaries among call types (Keenan et al., 2013; Wadewitz et al., 2015). However, there remains substantial variation in the specific approaches employed by these different studies. Hence, we need a framework to evaluate the quality of quantitative research methods.

A research method's quality for a given research question can be evaluated based on three main epistemological constructs: validity, reliability, and objectivity (Bolarinwa, 2015). Validity is the extent to which a method measures what it claims to or intends to measure. Most measures in science are approximations to the actual entity that a researcher intends to study. The goal of a high-quality research method is to get as precise an approximation to reality as possible. Reliability is the extent to which a given research method reproduces the same results each time the method is repeated. It pertains to a method’s replicability (Bolarinwa, 2015). Figure 1 graphically depicts these concepts. A valid method is one that measures what it is intended to measure—the bull’s eye in Figure 1 (a)-(d). A reliable method is one that consistently produces similar results—depicted by the scatter of dots on the dartboards. A method can be neither reliable, nor valid (Figure 1 (a)), producing invalid and inconsistent results, rarely hitting the bull’s eye, and landing at unpredictable locations on the dartboard. It could be valid, but not reliable (Figure 1 (b)), hitting the bull’s eye often, but not consistently, risking erroneous results occasionally. It could be reliable, but not valid (Figure 1 (c)), producing consistent results at each iteration but never hitting the bull’s eye. Such methods carry a high risk of finding erroneous patterns that researchers may end up having confidence in because of the consistency of results. The goal is to develop and employ research methods that are both reliable and valid, consistently hitting the bull’s eye (Figure 1 (d)). Lastly, A high-quality research method should also facilitate objectivity. Objectivity is the quality of a method being independent of the observer or researcher employing the method. While humans are inherently biased creatures, a well-designed method facilitates objective inference.

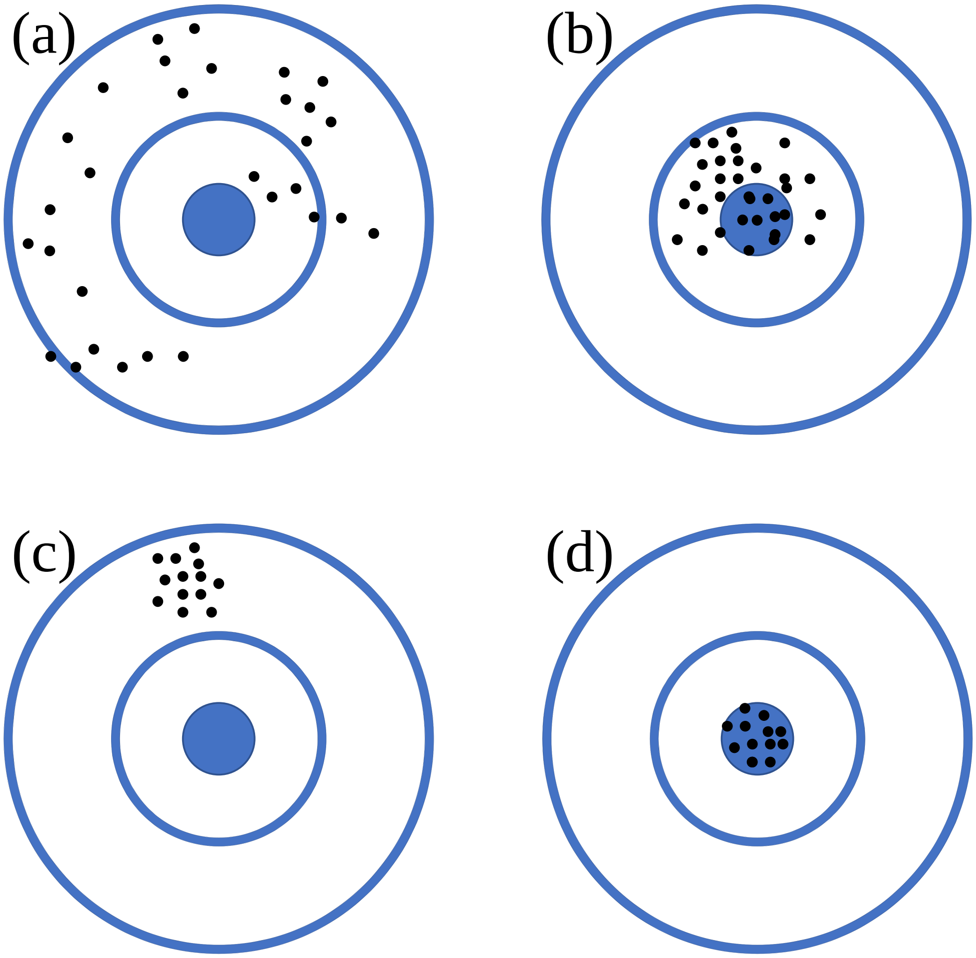


Figure 1: Graphical depiction of different kinds of research methods. A research method that is (a) neither reliable, nor valid; (b) somewhat valid, but not too reliable; (c) reliable, but not valid; and (d) reliable and valid (from: (Bolarinwa, 2015)).

Following these constructs, the quality of vocal communication research methodology can be evaluated based on the extent to which they employ reliable and valid: (i) acoustic features to measure the vocalizations, and (ii) statistical modeling techniques to analyze the vocalizations and make inference about biological questions. In primatology, commonly used acoustic features to represent primate vocalizations include measurements that broadly describe the call composition, frequency changes and temporal changes. These include features such as the number of units of a call type, the fundamental frequency, the peak frequency, and many other properties of the fundamental and peak frequencies such as the maximum, minimum, and range of frequencies, the shape of the first harmonic, harmonic to noise ratios, and so on. These acoustic features treat acoustic signals as linear soundwaves. In other words, these features assume equal importance of all frequencies in the sound, as well as linear transmission in time. Using such features presents a validity issue from a biological perspective as the auditory system of the animals is (i) not equally sensitive to all frequencies, (ii) perceives frequency changes proportionally rather than linearly, and (iii) perceives sounds differently based on proximal sounds that come before or after the sound in question. Acoustic features that are valid, i.e., biologically meaningful should take into account such auditory properties of sound receivers. This is because biological signals evolve to influence the intended signal receivers (Krebs & Dawkins, 1984), and not the human researcher studying the sound. Another related validity issue inherent in such approaches is that these measurements are made on call types that are defined by human listeners, using their human-specific auditory capacities. Since different species have different auditory capacities (e.g. (Kojima, 1990)), the discrete call types perceived by humans are unlikely to be the discrete call types perceived by another species. In case of chimpanzees, human listeners perceive many calls as graded and not discrete. Hence, one way to improve validity of acoustic features is to account for receiver perception.

Speech recognition researchers have identified a set of acoustic features called “mel frequency cepstral coefficients” (MFCCs) (Mermelstein, 1976) that aim to account for sound perception. The mel scale is related to the standard Hertz scale in a non-linear fashion and the relationship can be described as $F_{mel} = 2595 \times \log(1 + F_{Hz}/700)$. The mel scale represents the frequency perception of terrestrial vertebrates better than the Hertz scale. This is because a linear change in frequency does not represent a linear change in the perception of the pitch. Humans, for example, perceive the frequency linearly up to 1000 Hz, but logarithmically above 1000 Hz. Hence, the features from the Hertz scale would overestimate the importance of frequencies over 1000 Hz. MFCCs account for logarithmic perception of frequencies and also differentially attenuate higher and lower frequencies by filtering out frequencies that are poorly perceived and retaining frequencies that are perceived better. Figure 2 gives a graphical representation of how the filters are placed on a power spectrum to account for perception. These processing steps mimic mammalian hearing and account for some, but not all, aspects of perceptual transformations of the signal, providing the first steps in improving the validity of the acoustic features. MFCCs and features derived from MFCCs have proven to be highly reliable in human speech recognition tasks that involve identifying words (Ananthi & Dhanalakshmi, 2015; Ittichaichareon et al., 2012; Zolnay et al., 2005). However, we need to test whether MFCCs will perform as well in identifying the acoustics of the socio-indexical and phatic speech signals like chimpanzee vocalizations. Based on their success in classifying human speech, MFCCs are gaining popularity in bioacoustics for a variety of research problems (Albornoz et al., 2017; Clemins et al., 2005; Fedurek, Zuberbühler, et al., 2016; Xie et al., 2016). For example, a couple of studies have used MFCCs to classify frog and bird species using their vocalizations (Albornoz et al., 2017; Xie et al., 2016). Clemins and colleagues obtained reliable classifications of call types of African elephant (*Loxodonta africana*), and fairly reliable identification of the calling individual using MFCCs (Clemins et al., 2005). Fedurek and colleagues used MFCCs to identify individual callers and to understand whether sequential information was correlated with individual factors such as rank and age in chimpanzee pant-hoots (Fedurek, Zuberbühler, et al., 2016).

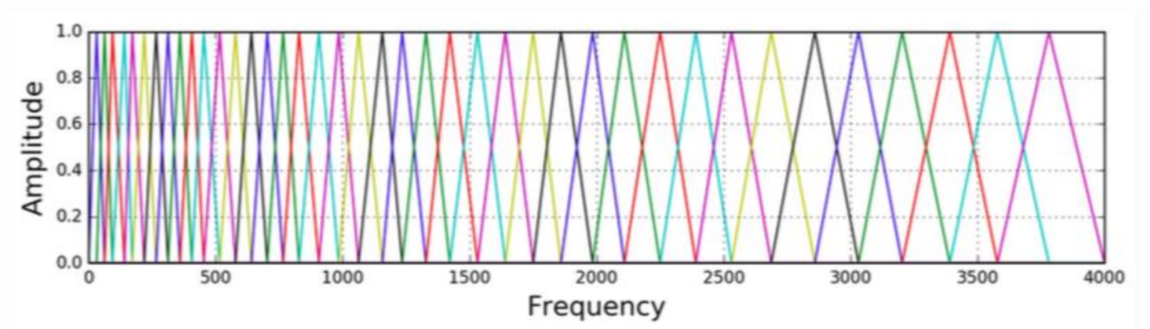


Figure 2: A step in the process of obtaining mel frequency cepstral coefficients. Filters are placed on the power spectrum such that they filter more of the poorly perceived frequencies and less of the better perceived frequencies (Singh, 2019).

When it comes to the statistical modeling techniques, the main quality issue tends to be that of reliability. The main goal of research involving the development of statistical modeling techniques is to develop techniques that are highly reliable, i.e. they produce consistently accurate results. Most of the recent developments in statistical modeling techniques relevant for vocal communication research have happened in the fields of artificial intelligence and machine learning. Machine learning research involves development and application of statistical models, known as machine learning models, with the primary goal of making predictions through classification (in case of discrete response variables) and regression (in case of continuous response variables). Machine learning models can either be supervised or unsupervised. In supervised classification, the models incorporate *a priori* labels for different categories (call types) and attempt to maximize among-category differences and reduce within-category differences, allowing reliable classification of different call types. The popular supervised methods that have been used for classifying vocal repertoires include discriminant function analysis (Gamba & Giacoma, 2007), multivariate analysis of variance (Owren et al., 1997), support vector machines (Xie et al., 2016) and others. In unsupervised classification, the models do not need *a priori* labels for categories, as they instead attempt to find different categories based on measures of similarities within the categories. Popular unsupervised methods used for vocal repertoire classification include principal components analysis (Gros-Louis et al., 2008) and k-means clustering (Hammerschmidt & Fischer, 1998). While these methods perform well in some cases giving high accuracy of classification, the results are not always unambiguous, as they rely on many assumptions about the structure of the data. Violations of assumptions could yield idiosyncratic rather than real differences, especially for data with inherent non-linearities, which are better suited for non-linear models. Machine learning techniques that make few assumptions about the data and perform better on data with inherent non-linearities include random forests and artificial neural networks. Random forest is a non-parametric ensemble technique that makes no assumptions about the properties of the data (Evans et al., 2011). This makes it useful for studying call types that could have non-linear boundaries in the multidimensional acoustic feature space. Bioacoustics researchers have found them to be reliable for problems such as species classification with vocalizations in birds and frogs (Albornoz et al., 2017; Xie et al., 2016) and for species and call type classification in bat echolocation calls (Armitage & Ober, 2010). Artificial neural network is another classification algorithm that works well for multidimensional non-linear data. Artificial neural networks can be both supervised and unsupervised. While artificial neural networks have existed for decades, they are only recently gaining popularity in their usage due to the rise of cheap computing power, which makes training computationally intensive artificial neural networks practically feasible and accessible. Artificial neural networks are gaining popularity in bioacoustics as well and researchers are increasingly finding that they perform better than traditional discriminant function analysis or multivariate analysis of variances in terms of accuracy of classification (Dawson et al., 2006; Mielke & Zuberbühler, 2013; Pozzi et al., 2010). While popular among bioacoustics researchers studying various animals, to my knowledge, only two studies have applied artificial neural networks to classify call types in primates (Mielke & Zuberbühler, 2013; Pozzi et al., 2010), and none have used them to classify chimpanzee vocal repertoire.

In this chapter, I extract commonly used acoustic features in primatology using a proprietary software called LMA (Schrader & Hammerschmidt, 1997). Acoustic features extracted using LMA are henceforth referred to as LMA features. I test their reliability in terms of evaluation metrics including precision, recall, F1 score and accuracy (see Methods for details) using three supervised machine learning techniques: discriminant function analysis, random forest, and 1-hidden layer artificial neural network. I used these methods to classify 5 different chimpanzee call types: barks, pant-hoots, rough-grunts, screams, and waa-barks; see Methods for more details. Similarly, I extract MFCCs and other speech recognition related acoustic features using an open-source

Python package called pyAudioAnalysis (Giannakopoulos, 2015). Acoustic features extracted using the pyAudioAnalysis package are henceforth referred to as speech recognition features. I test their performance using the same data, evaluation metrics, and machine learning techniques and compare their performance. Finally, I use unsupervised machine learning techniques: k-means clustering and principal components analysis and inspect the extent to which they suggest the existence of the call types that human listeners classified. I use the results from the comparison of acoustic features and machine learning techniques to decide which acoustical and statistical analysis techniques to employ in the next chapter, where I test for evidence of community-specific features of pant-hoots.

1.3 Methods

1.3.1 Study site and data source

Since my data collection and processing was still ongoing at the time of conducting this study in late 2017-early 2018, I used readily available archival recordings from Gombe National Park, Tanzania (Plooij et al., 2015). This archive contains chimpanzee recordings obtained by Dr. Hetty van de Rijt-Plooij between 1972-73. To ensure call-quality, certainty about the call types, and to avoid overrepresentation of any individual, I discarded recordings with poor recording quality, discarded chorused calls, randomly sampled recordings from a range of different individuals, and extracted units of

5 of the 39 chimpanzee call types described by Plooij et al. (2015) that I could personally identify with high certainty. This resulted in 281 recordings of five chimpanzee call types: bark (n=39), pant-hoot (n=67), rough-grunt (n=48), scream (n=46), and waa-bark (n=81). I annotated the call units using the acoustic analysis software Praat and trimmed them using a Praat script for further analysis.

1.3.2 Description of call types included in this study

For this study, I chose five different chimpanzee call types described by Plooij and colleagues based on my personal confidence in identifying them in the dataset. This created greater confidence about a call type as my way of choosing calls implied agreement about the call type from both me and Plooij et al. (2015). I did not attempt to include all call types described in previous qualitative classifications (Goodall et al., 1968; Marler, 1969; Plooij et al., 2015), because I could not identify them confidently, which introduces the same kind of bias that I hope to mitigate with a study like this. Hence, this study is aimed as a preliminary attempt at figuring out the best methods for tackling the difficult question of classifying chimpanzee vocal repertoire. The call types included in this study are described below.

1.3.2.1 Call descriptions

The fundamental frequency (F0) and the element duration are two important features to begin describing call types. The fundamental frequency is the lowest frequency of the periodic waves that make an element of a vocalization. Barks included short (<0.35s), high pitch (fundamental frequency (F0): ~900-1500 Hz), tonal (sometimes atonal), repeated elements (Figure 3 (a)). Pant-hoots are complex calls consisting of many different kinds of elements that range from low pitch grunt-like elements to high pitch bark-like and scream-like elements (Figure 3 (b)). They can be identified as distinct from other calls as four different kinds of elements often appear in the same order in pant-hoots, as opposed to other call types that consist of a repeated sequence of a single kind of element, as in barks, screams, rough-grunts, and waa-barks. Since pant-hoots are the primary focus of this dissertation, I describe them in more detail in chapter 2, section 2.3.4. Rough-grunts consist of short (<0.25s), low pitch (F0: ~200-500 Hz), noisy or tonal bark-like, repeated elements (Figure 3 (c)). They are often produced while feeding (Goodall, 1986; Marler, 1969; O’Bryan, 2015). Screams are longer (~0.35-1.3s), high pitch (F0: >1000 Hz), noisy calls, typically with many repeated elements (Figure 3 (d)). Lastly, like barks, waa-barks are typically shorter screams, but relatively longer than barks, and high pitch and noisy like screams (Figure 3 (e)). They are distinguished from barks and screams mostly based on their duration and noisiness, especially in the middle of the call unit.

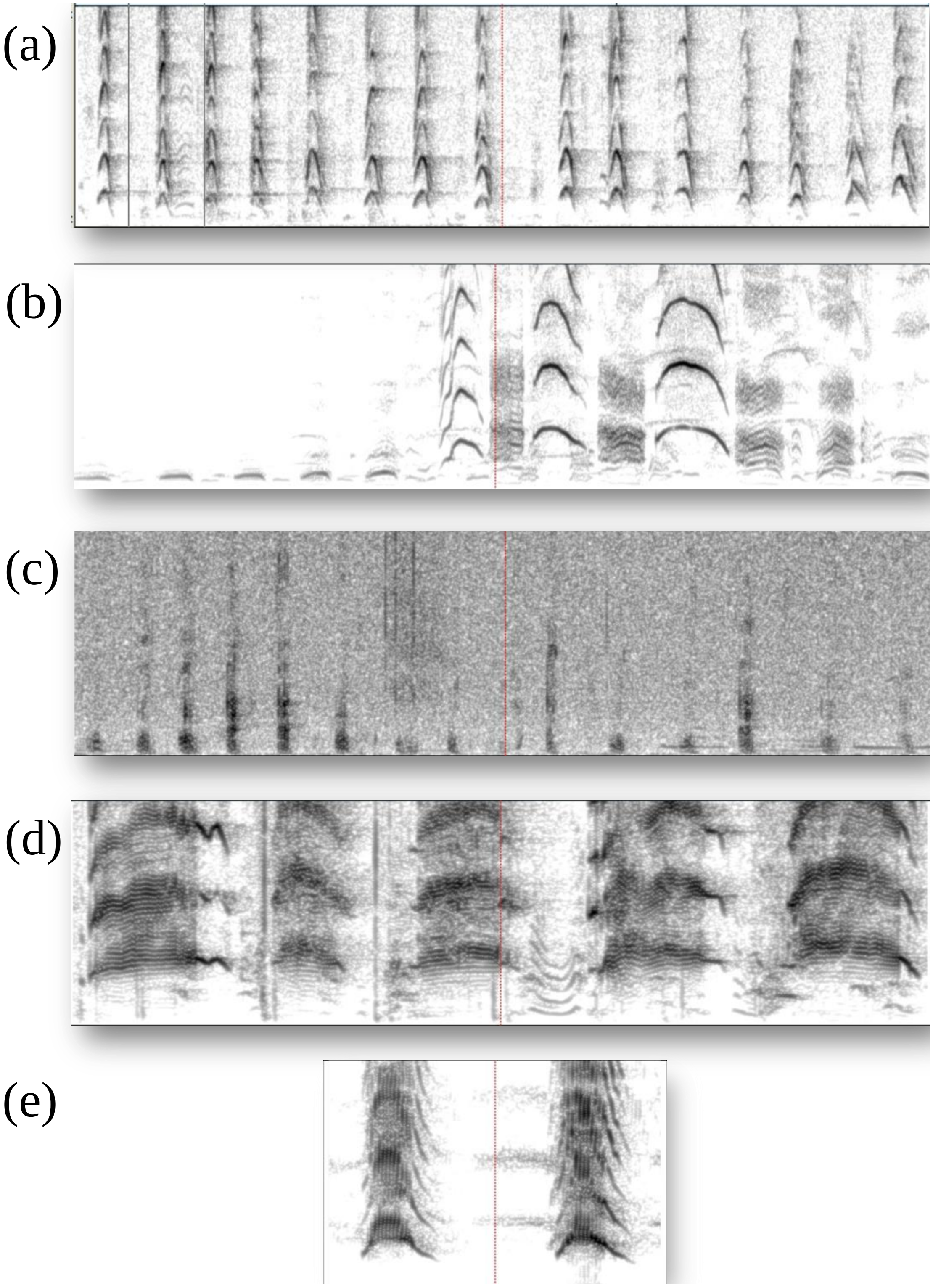


Figure 3: Typical spectrograms of call types included in this study. (a) Barks, (b) Pant-hoot, (c) Rough-grunts, (d) Screams, and (e) Waa-barks. Y-axis represents the frequency from 0-5000Hz, X-axis represents time from 0 to 7 s.

1.3.3 Acoustic feature extraction

1.3.3.1 LMA features

In March 2020, I visited the lab of primate vocal communication researcher, Dr. Kurt Hammerschmidt, at the German Primate Center at Georg-August-Universität Göttingen, Germany. Dr. Hammerschmidt developed an acoustic feature extraction software package, LMA (Fischer et al., 2013; Schrader & Hammerschmidt, 1997), which is widely used by primatologists and other bioacoustics researchers. With guidance from Dr. Hammerschmidt, I learned to use LMA and another bioacoustics software package, Avisoft SASLab Pro (Specht, 2004) that works with LMA. I used Avisoft SASLab Pro to construct LMA-readable spectrogram files (more details in Chapter 2, 2.3.5 Acoustic feature extraction), and used these files in LMA to extract a range of acoustic features including, but not limited to, the fundamental and the peak frequencies, and many other features representing properties of the fundamental and peak frequencies such as minimum, mean, and maximum fundamental and peak frequencies and so on. These acoustic features are listed in Table 13 in the Chapter 2 and I refer to them as LMA features.

1.3.3.2 Speech recognition features

I used an open-source Python library, pyAudioAnalysis for automatic extraction of acoustic features (Giannakopoulos, 2015). These include 13 mel frequency cepstral coefficients and a few spectral and energy related features that are listed in Table 1. The extracted features included the means and standard deviations of the features listed in Table 1, and the means and standard deviations of their deltas (difference between two successive frames for each of these features) over the duration of the call. This resulted in a total of 136 acoustic features for each call unit. I refer to them as speech recognition features.

Table 1 (reproduced from Giannakopoulos, 2015): Acoustic features extracted from pyAudioAnalysis, an open-source Python package for acoustic analysis.

INDEX	NAME	DESCRIPTION
1	Zero Crossing rate	The rate of sign-changes of the signal during the duration of a particular frame.
2	Energy	The sum of squares of the signal values, normalized by the respective frame length.
3	Entropy of Energy	The entropy of sub-frames' normalized energies, it can be interpreted as a measure of abrupt changes.
4	Spectral Centroid	The center of gravity of the spectrum.
5	Spectral Spread	The second central moment of the spectrum.
6	Spectral Entropy	Entropy of the normalized spectral energies for a set of sub-frames.
7	Spectral Flux	The squared difference between the normalized magnitudes of the spectra of the two successive frames.
8	Spectral Rolloff	The frequency below which 90% of the magnitude distribution of the spectrum is concentrated.
9-21	MFCCs	13 Mel frequency Cepstral Coefficients form a cepstral representation where the frequency bands are not linear but distributed according to the mel-scale.

22-33	Chroma Vector	A 12-element representation of the spectral energy where the bins represent 12 equal tempered pitch classes
34	Chroma Deviation	The standard deviation of the 12 chroma coefficients

1.3.4 Statistical classification

I performed all statistical analyses in R version 4.1.3. For supervised classification, I used the R package ‘caret’ that provides tools to streamline the process of performing analysis with machine learning models (Kuhn et al., 2020). This included tools to perform data partitions into training, validation, and testing; performing cross-validated model fitting for a wide range of machine learning techniques; and evaluating the models with a range of evaluation metrics. For unsupervised analysis, I used the base R function kmeans() for k-means clustering, and princomp() for principal components analysis.

While there are large number of high performing machine learning techniques employed for bioacoustics research, I chose three supervised techniques that I have found to either be very commonly used (discriminant function analysis) or have found to have high performance for a wide range of problems (random forests), or both (artificial neural networks). For a pilot analysis, I had also tested two other techniques, k-nearest neighbors, and support vector machines, but I remove them from this analysis as they did not stand out in any way in my pilot analysis, and are less likely to be of interest to primatologists.

I applied these three supervised techniques: discriminant function analysis, random forests, and artificial neural networks; and two unsupervised techniques: k-means clustering, and principal components analysis on each of the two sets of acoustic features: LMA features, and speech recognition features.

1.3.4.1 Descriptions of supervised techniques

Discriminant function analysis is one of the most commonly used statistical classification techniques in bioacoustics (Armitage & Ober, 2010; Crockford et al., 2004; Xie et al., 2016). The method involves using many numeric features that measure different properties of each observation and incorporating *a priori* labels of different classes to which the observations belong. This information is used to perform transformations to the multidimensional space such that the variance among classes is maximized relative to the variance within classes. This results in a multidimensional space where the observations from different classes are at the maximum distance, and the classes can be discriminated based on the relative distances of observations in this space, known as the Mahalanobis distance (Neff & Marcus, 1980). This method is also popular because one can calculate an F-statistic in the multidimensional space and perform statistical hypothesis tests about whether the classes can be distinguished based on their features. The model assumes homogeneous variances among classes and multivariate normality among the features. The method tends to often be robust to violations of these assumptions, but correlated features could lead to underperformance. Hence, I used a version of the discriminant function analysis that penalizes features, often correlated, that add less value to discriminating among classes, known as the penalized discriminant analysis, that is known to improve performance (Hastie et al., 1995).

Random forest belongs to a class of algorithms known as tree-based ensemble models. Tree-based models rely on algorithms that construct decision trees based on the features to make decisions about what class to classify a given observation. These can be thought of as if-else statements on values of a given set of features that eventually lead to classifying an observation to a specific class. For a given training set, one can construct many decision trees that result in similar predictive performance. Ensemble techniques rely on the idea that while a specific decision tree may not have generalizable performance, averaging results from a set of many decision trees (known as ensembles) could improve predictive performance on new data. Random forest is one such algorithm that includes building many decision trees, and averaging the predictions from a subset of these trees to result in high predictive performance. Random forest is a powerful technique since it non-parametric, i.e., it makes no assumptions about the structure of the data or the relationships among classes. The intuitive structure of the algorithm used (the decision trees) makes it highly interpretable and easy to communicate the results. Furthermore, one can easily remove some features and test the extent to which removing a given feature reduces the performance, allowing an easy way for feature selection. This makes it highly versatile for a wide range of applications, and is gaining popularity in bioacoustics studies (Albornoz et al., 2017; Armitage & Ober, 2010; Valletta et al., 2017; Xie et al., 2016).

Artificial neural networks attempt to mimic the functioning of the human brain. They are structured like the human brain with many layers of simulated neurons and connections among the neurons and the layers. The first layer is the input layer that consists of the values of the input variables, the features, of the entity we are interested in classifying. These values are fed into the next layer, known as hidden layer, which consists of many units, known as neurons, that perform mathematical transformations on the inputs. Connection of each input unit is associated with a weight that changes an input value to an output value that goes into the associated neuron before applying transformations. An activation function is applied to this output value, and it determines whether the value will be fed into the next hidden layer. If fed, the value will again be associated with a weight in the next hidden layer, and an activation function will again be applied to this weighted value. There can be one or many hidden layers in an artificial neural network and a multilayer artificial neural network is called a multilayer perceptron. The output from the last hidden layer goes into the output layer, where an activation function is again applied to output a value. For each observation, this output value is the probability of belonging to a particular category, and the highest probability category is considered to be the predicted category. Artificial neural networks are among the best performing models for a wide range of classification problems. Much of the recent progress in the fields of machine learning and artificial intelligence is attributable to artificial neural networks, and in particular, artificial neural networks involving many hidden layers known as deep neural networks. They are gaining popularity in bioacoustics (Armitage & Ober, 2010; Bermant, 2021; Bermant et al., 2019) and to my knowledge, two studies have applied them for studying primate vocalizations as well

(Mielke & Zuberbühler, 2013; Pozzi et al., 2010). Here, I use a single hidden layer neural network due to a lack of enough data to train deep neural networks.

1.3.4.2 Descriptions of unsupervised techniques

K-means clustering involves partitioning a dataset of n observations into k clusters using features of each observation in the dataset, where k ≤ n. The clustering process involves specifying k and then determining k clusters such that the within-cluster variances are minimized. Within-cluster variances are the squared Euclidean distances in the multidimensional space of features, also known as within-cluster sum of squares. We classify a given observation in the cluster whose mean, defined as the cluster centroid, is closest to that observation.

Principal component analysis involves determining n-axes in an n-dimensional space of n-features such that each axis is in the direction of the line that best fits the data, while also simultaneously being orthogonal to every other axis. We determine the best fitting line by minimizing the squared distances of each point from the line. Determining the axes this way ensures that each axis is uncorrelated with every other axis, and lies in the direction that explains the maximum variance in the data under these constraints. These axes are known as the principal components. Hence, the first axis, the principal component 1 is the axis that explains the maximum variance in the data, and the i-th axis, or the i-th principal component is the axis that is orthogonal to the first i-1 axes that maximizes variance. One can use the principal component analysis as an unsupervised way for classification by determining an observation’s nearest neighbors, or by utilizing similarity measures such as the Euclidean distances on the first few principal components.

1.3.4.3 Model fitting process

The general process of performing analysis with supervised machine learning models involves (i) using labeled data i.e., data with labeled ground truth, to ‘train’ the machine learning model, and (ii) using labeled data that was not used to train the model, to ‘test’ the model’s predictive capabilities on new data. Testing a model on data that it has never seen before is quite important as most statistical techniques have a risk of overfitting to the data that they were trained on and not generalizing to new data, thus giving a false sense of high predictive performance. Hence, before fitting any machine learning models, we need to keep separate some data to test the performance of the model. This test set should only be used once for testing the generalizability, because using it more than once involves the risk overfitting on the test data (Hastie et al., 2009). However, the process of training a machine learning model often involves training and testing multiple times to tune the set of model parameters that are most likely to result in a model with high predictive performance. Hence, one can first split the data into two sets: the training set, and the testing set, and then, further split the training set into training set, and ‘validation’ set. With these three sets of data, the process of training goes as follows: we first choose an initial set of parameters to train the model on the training set and test its performance on the validation set. The validation set acts as a temporary test set. Then we change the initial parameters and see if we can improve on the initial validation accuracy and keep repeating the process until we don’t see much improvement in the performance with changing parameters. Since this process risks overfitting on the validation set, we take an extra step to mitigate the risk by taking many different splits of the training and the validation sets from the main training set. Then, the training process described above is repeated on each of the new splits of training and validation sets. This process is known as cross-validation (Hastie et al., 2009). Once we have exhausted all the splits of training and validation sets and reach the point of diminishing returns with changing parameters, we call the model trained and then do a final test of predictive performance on the test set that we had kept separate. I used the createDataPartition() function from the caret R package to create training and testing sets that are well balanced in terms of the inclusion of all the call types that I wanted to train the models for. I used 70% of my data for the training set and 30% for the testing set. Before training, I standardized (z-scored) the training data by mean centering and scaling by standard deviation so as not have any feature dominate due to a difference in order of magnitude. I trained all models with cross-validation on the 70% of the data used for training, and report the performance of all the supervised models on the 30% of the data that is the test set, based on the evaluation metrics described in the next section.

For unsupervised K-means clustering, I included the entire dataset (without splitting it into training, validation, or test sets), standardized (z-scored) it by mean centering and scaling with the standard deviation so as not to have the difference in orders of magnitude affect the clustering. I fitted nine different k-means models with k ranging from two to ten clusters. I determined the optimal number of clusters in the data based a commonly used heuristic known as the scree test (Cattell, 1966). The heuristic involves determining the optimal number of clusters based on where the ‘elbow’ lies in the scree plot of the ratio of the within-cluster sum of squares to the total sum of squares vs. the number of clusters in the k-means model. This ratio tends to be lower the more clusters we choose, but after a certain number of clusters, the reduction in the ratio is slower with each additional cluster added in the analysis. The number of clusters beyond which this happens is the point from where we have diminishing returns for adding more clusters and is represented by the aforementioned elbow in the scree plot. I looked at the classifications form the k-means model with the optimal number of clusters, as well as a k-means model with 5 clusters, which is the number of classes I had in dataset. I report the performance measured from both based on the evaluation metrics described in the next section.

Lastly, I performed principal components analysis as an unsupervised way to visualize the classes in the multidimensional space on the axes of maximum explained variances. I visualized the results by plotting the 68% normal-data ellipses, i.e., ellipses containing 68% of the data points on the first and the second principal components. This number is chosen as 68% of samples from normally distributed data are contained within 1 standard deviation of the data.

1.3.4.4 Evaluation metrics

There are many ways to evaluate a classification model’s performance based on the goals of the research question. Commonly used metrics to evaluate a classification model’s performance are based on the type-I and type-II error rates, and correct prediction rates (Table 2).

Table 2: Truth table defining type I and type II errors in a model’s predictions for a given class in a classification model.

		Ground Truth	
Prediction from the model		Positive	Negative
	Positive	True Positive (TP)	Type-I error or False Positive (FP)
	Negative	Type-II error or False Negative (FN)	True Negative (TN)

Depending on which of type-I or type-II errors or overall performance is more important for a given research question, one can choose specific metrics to evaluate the model. These are described in the Table 3. The metric that estimates what percentage of the original calls labeled as a particular type by the human observer are classified by the model as that type is known as the sensitivity, or recall. It is the ability of a model to detect true positives. Mathematically, it is the conditional probability that the model predicts a particular call type, given that it is labelled as that call type by the human observer. In other words, what fraction of the calls that I labelled as say pant-hoots in the original sample were classified by the model as pant-hoots? The metric that estimates what percentage of calls classified as a particular call type were labelled by me as that type is known as positive predictive value, or precision. Mathematically, it is the conditional probability that a particular call actually has the label that the model classified it into. In other words, what fraction of the calls classified by the model as say pant-hoots, were also labelled by me as pant-hoots? It gives an estimate of how well the model recognizes a particular call type. Precision and recall are metrics of a model's ability to detect the presence of a given call type. There are also metrics that estimate a model's ability to correctly detect the absence of a given call type. Those are the negative predictive value, and specificity. Since the goal of this study is to assess a model's ability to detect the presence of call types, the metrics of interest for this study are the precision and recall. Hence, I don't discuss negative predictive value and specificity further and only include their mathematical definitions in Table 3. However, I still included a metric that estimates the overall performance of the model across all call types. This metric is known as the accuracy. The accuracy of the model is the percentage of calls classified to be the same as their labels. This metric could also be defined at the level of a call type. Lastly, I included a metric for the overall (average) performance of a model in detecting a given call type. That metric is the F1 score, which is the harmonic mean of Precision and Recall. While I included call type level classification accuracies in the results, I used the F1 score for inference about a model's quality. This is because the F1 score it is a more valid metric compared to accuracy for a model's ability to detect a call type as it is independent of which other call types are included. Call type level accuracy can overestimate a model's performance for some call types, especially when there are unequal sample-sizes in the dataset.

These metrics give a sense of the various strengths or weaknesses of a particular technique. Recall values are important when it is important to ensure that a given call type is detected, even at the cost of incorrectly detecting some other call types. This metric would be important for building detectors of a given call type that could be used to automatically extract that call type from a large dataset. Precision values are important to ensure that a given call type is correctly detected. This could be important for studies that involve asking functional questions about a given call type. However, when one does not have narrow definitions of a given call type and they are mostly defined by human-listeners, the F1 score provides a good measure to ensure that most of the variation in a call type is captured while being as precise about the definition of call type as possible.

Table 3: Commonly used evaluation metrics for a classification technique. The metrics of interest are indicated in bold. TP = true positive, FP = false positive, TN = true negative, and FN = false negative.

Evaluation Metric	***Definition***
Sensitivity (Recall)	**P(model predicts bark|human predicts bark)** **TP/(TP+FN)**
Specificity	P(negative prediction|negative truth) TN/(TN+FP)
Positive Predictive Value (Precision)	**P(human predicts bark|model predicts bark)** **TP/(TP+FP)**
Negative Predictive Value	P(negative truth|negative prediction) TN/(TN+FN)
F1-score	**Harmonic mean of precision and recall**
Accuracy	**P(model and human predict the same)** **(TP+TN)/(TP+TN+FP+FN)**

1.4 Results

1.4.1 Performance of supervised machine learning techniques on LMA features

The average performance of the three models when applied to LMA features varied based on different call types. Barks were the worst detected, whereas pant-hoots and screams were best detected (Table 4; Barks: mean F1 score = 47.3, SD = 11.4, range = [34.8, 57.1], Pant-hoots: mean F1 score = 81.3, SD = 3.6, range = [77.8, 85], Rough-grunts: mean F1 score = 78.3, SD = 6.1, range = [72.7, 84.8], Screams: mean F1 score = 81.2, SD = 7, range = [74.1, 88], Waa-barks: mean F1 score = 74.7, SD = 2.6, range = [72.7, 77.6]). Of the three techniques, random forests had the highest overall accuracy (78.0%, 95% CI = [67.5, 86.4]), followed by discriminant functions analysis (73.2%, 95% CI = [62.2, 82.4]), which comes close to artificial neural networks (72.0%, 95% CI = [60.9, 81.3]). Random forest was overall the best model to detect four out of five call types as it had the highest F1 score for barks, rough-grunts, screams and waa-barks. The best model for detecting pant-hoots was discriminant functions analysis with an F1 score of 85. Table 4 below summarizes the precision, recall, F1 score, and accuracy values for the three unsupervised techniques when applied on LMA features:

Table 4: Summary of the precision, recall, F1 score and accuracy values for the three supervised techniques applied on LMA acoustic features.

Discriminant Functions Analysis					
Metric	**Bark**	**Pant-hoot**	**Rough-grunt**	**Scream**	**Waa-bark**
Precision	55.6	85	100	71.4	64.5
Recall	45.5	85	57.1	76.9	83.3

F1 score	50	85	72.7	74.1	72.7
Class Accuracy	69.9	90.1	78.6	85.6	82.2
Overall Accuracy [95% CI]	73.2 [62.2, 82.4]				
Random Forests					
Precision	60	87.5	73.7	91.7	76
Recall	54.5	70	100	84.6	79.2
F1 score	57.1	77.8	84.8	88	77.6
Class Accuracy	74.5	83.4	96.3	91.6	84.4
Overall Accuracy [95% CI]	78.0 [67.5, 86.4]				
Artificial Neural Networks					
Precision	33.3	88.2	70.6	78.6	77.3
Recall	36.4	75	85.7	84.6	70.8
F1 score	34.8	81.1	77.4	81.5	73.9
Class Accuracy	62.5	85.9	89.2	90.1	81.1
Overall Accuracy [95% CI]	72.0 [60.9, 81.3]				

1.4.2 Performance of supervised machine learning techniques on speech recognition features

Analogous to LMA features, the average performance of the three models when applied to speech recognition features also varied based on different call types. Like LMA features, barks were the worst detected, whereas pant-hoots and screams were best detected (Table 5; Barks: mean F1 score = 45.5, SD = 17.8, range = [25, 57.1], Pant-hoots: mean F1 score = 82.9, SD = 4.5, range = [78.9, 87.8], Rough-grunts: mean F1 score = 67.7, SD = 5.7, range = [61.5, 72.7], Screams: mean F1 score = 87.8, SD = 3.6, range = [84.6, 91.7], Waa-barks: mean F1 score = 74.2, SD = 6.4, range = [70.2, 81.6]). Of the three techniques, artificial neural networks had the highest overall accuracy (76.8%, 95% CI = [66.2, 85.4]), followed by random forests (74.4%, 95% CI = [63.6, 83.4]), followed by discriminant functions analysis (70.7%, 95% CI = [59.6, 80.3]). There was no obviously best model to detect most call types as the highest F1 score varied by the call type. Random forests had the highest F1 score for barks and rough-grunts, discriminant functions analysis for screams and, artificial neural networks had the highest F1 score for waa-barks and pant-hoots. Table 5 summarizes the precision, recall, F1 score, and accuracy values for the three unsupervised techniques when applied on speech recognition features:

Table 5: Summary of the precision, recall, F1 score, and accuracy values for the three supervised techniques applied on speech recognition acoustic features.

Discriminant Functions Analysis					
Metric	**Bark**	**Pant-hoot**	**Rough-grunt**	**Scream**	**Waa-bark**
Precision	40	83.3	66.7	100	60.6
Recall	18.2	75	71.4	84.6	83.3
F1 score	25	78.9	69	91.7	70.2
Class Accuracy	57	85.1	82	92.3	80.5
Overall Accuracy [95% CI]	70.7 [59.6, 80.3]				
Random Forests					
Precision	60	84.2	63.2	100	70.8
Recall	54.5	80	85.7	76.9	70.8
F1 score	57.1	82.1	72.7	87	70.8
Class Accuracy	74.5	87.6	87.7	88.5	79.4
Overall Accuracy [95% CI]	74.4 [63.6, 83.4]				
Artificial Neural Networks					
Precision	54.5	85.7	66.7	84.6	80

Recall	54.5	90	57.1	84.6	83.3
F1 score	54.5	87.8	61.5	84.6	81.6
Class Accuracy	73.8	92.6	75.6	90.9	87.4
Overall Accuracy [95% CI]	76.8 [66.2, 85.4]				

1.4.3 A comparison of LMA and speech recognition features

Assessing which technique had the highest metric for a given call type, metric, and acoustic feature combination reveals that random forests work well with LMA features, and artificial neural networks work well with speech recognition features (Table 6). For LMA features, random forests had the highest precision, recall, F1-score or accuracy in 14 out of 21 combinations, followed by discriminant functions analysis that had the highest metrics in 5 out of 21, followed by artificial neural networks that had the highest metrics in 3 out of 21 combinations (Table 6). For speech recognition features, artificial neural networks had the highest metrics in 12 out of 21 combinations, followed by random forests that had the highest metrics in 8 out of 21 combinations followed by discriminant functions analysis that had the highest metrics in 6 out of 21 combinations (Table 6).

In terms of the acoustic features, there was no obviously better set that performed consistently well over another. Moreover, the best performing acoustic features varied by the call type. Both kinds of acoustic features performed similarly (less than 5% difference in F1 score) for barks (maximum F1 score = 57.1 for both), pant-hoots (maximum F1 score for LMA features = 85 and maximum F1 score for speech recognition features = 87.8), screams (maximum F1 score for LMA features = 88 and maximum F1 score for speech recognition features = 91.7), and waa-barks (maximum F1 score for LMA features = 77.6 and maximum F1 score for speech recognition features = 81.6). LMA features performed better for rough-grunts as compared to speech recognition features (maximum F1 score for LMA features = 84.8 and maximum F1 score for speech recognition features = 72.7).

Table 6: A comparison of LMA acoustic features and speech recognition acoustic features.

LMA acoustic features					
Metric	**Bark**	**Pant-hoot**	**Rough-grunt**	**Scream**	**Waa-bark**
Maximum Precision	60 (RF)	88.2 (ANN)	100 (DFA)	91.7 (RF)	77.3 (ANN)
Maximum Recall	54.5 (RF)	85 (DFA)	100 (RF)	84.6 (RF & ANN)	83.3 (DFA)
Maximum F1 score	57.1 (RF)	85 (DFA)	84.8 (RF)	88 (RF)	77.6 (RF)
Maximum Class Accuracy	74.5 (RF)	90.1 (DFA)	96.3 (RF)	91.6 (RF)	84.4 (RF)
Maximum Overall Accuracy	78 (RF)				
Speech recognition acoustic features					
Maximum Precision	60 (RF)	85.7 (ANN)	66.7 (DFA & ANN)	100 (DFA & RF)	80 (ANN)
Maximum Recall	54.5 (RF & ANN)	90 (ANN)	85.7 (RF)	84.6 (DFA & ANN)	83.3 (DFA & ANN)
Maximum F1 score	57.1 (RF)	87.8 (ANN)	72.7 (RF)	91.7 (DFA)	81.6 (ANN)

Maximum Class Accuracy	74.5 (RF)	92.6 (ANN)	87.7 (RF)	92.3 (DFA)	87.4 (ANN)
Maximum Overall Accuracy	76.8 (ANN)				

1.4.4 Unsupervised k-means clustering with LMA features

Figure 4 shows the scree plot for the k-means clustering applied to the standardized LMA features with k ranging from 2 to 10. The y-axis is the ratio of within-cluster sum of squares to the total sum of squares. A lower ratio represents better clusters, but the trade-off is with the number of clusters as we can keep lowering the ratio by adding more clusters. Hence, an optimal number of clusters is the point at which there is an elbow in the scree plot. In this case, there isn’t a clearly visible elbow, but a closer look reveals 3 clusters as the point at which there’s a slight elbow.

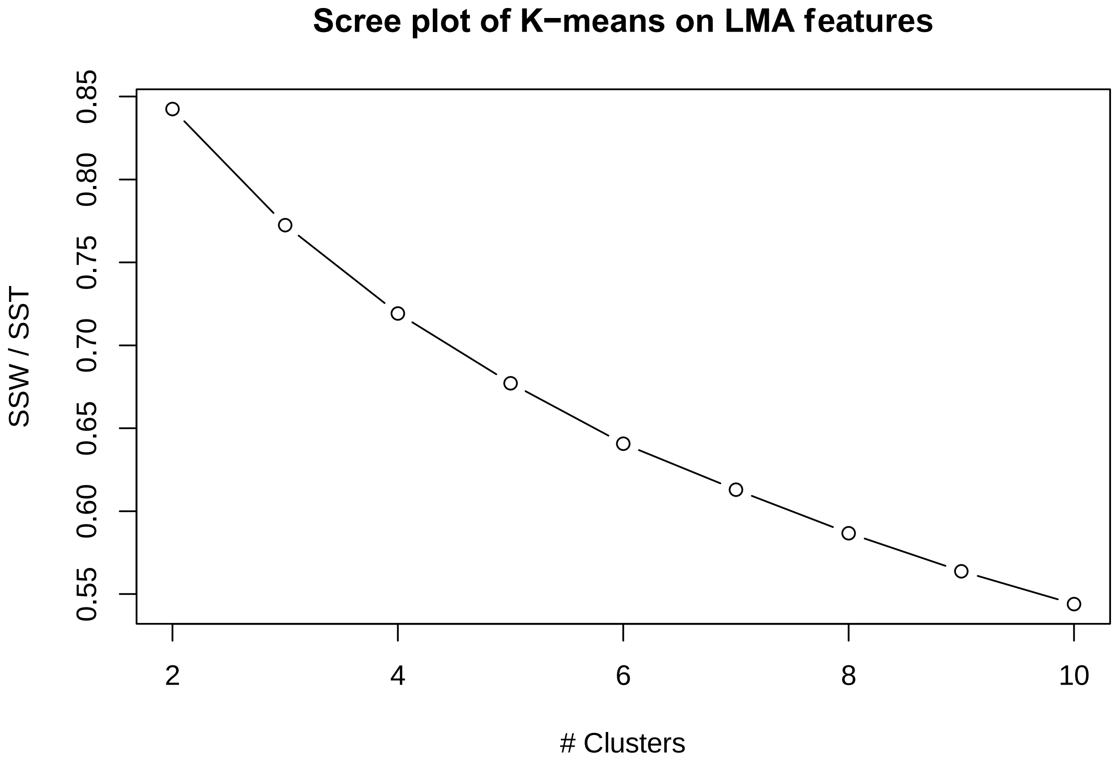


Figure 4: Scree plot of k-means clustering applied to standardized LMA features with k ranging from 2 to 10. SSW = within-cluster sum of squares, SST = total sum of squares.

The elbow is at 3 clusters where cluster 1 corresponds to screams, 2 to pant-hoots, and 3 to waa-barks. The rough-grunts are getting mainly classified into waa-barks followed by pant-hoots (Table 7). In the 5-cluster case, there seem to be clearer clusters for pant-hoots (cluster 5), rough-grunts (cluster 1), screams (cluster 4), and waa-barks (cluster 2). However, barks are mostly getting classified with screams (cluster 4) and in cluster 3 which also includes rough-grunts. In the at 6 or 7 cluster cases, the pant-hoots, rough-grunts, and screams often have a cluster they dominate in, but barks and waa-barks get mixed up with other call types. This might be a result of the gradation between barks and waa-barks and points towards the possibility that there may be mutiple types of barks and waa-barks.

Table 7: Results from k-means clustering on LMA features with k = 3, 5, 6, and 7. Rows indicate how many of a given call type were classified into a given cluster ID in a k-means analysis.

Cluster ID	**Bark**	**Pant-hoot**	**Rough-grunt**	**Scream**	**Waa-bark**
K = 3 clusters					
1	19	1	6	43	10
2	12	60	18	0	8
3	8	6	24	3	63
K = 5 clusters					
1	1	1	25	1	2
2	3	5	4	3	59
3	11	4	12	0	3
4	17	1	0	42	10
5	7	56	7	0	7
K = 6 clusters					
1	17	1	1	40	6
2	11	4	12	0	3
3	0	0	0	5	38
4	1	1	24	0	3
5	10	61	11	0	31
6	0	0	0	1	0
K = 7 clusters					
1	11	4	12	0	3
2	1	40	5	0	1
3	1	1	24	0	2
4	0	0	0	5	20
5	0	0	0	1	0
6	9	21	6	1	49
7	17	1	1	39	6

Comparing the optimal number of clusters based on the scree plot (i.e., 3 clusters), with the number of call types in the data as I classified, the 5-cluster k-means appear to perform better overall. The F1 scores for the each of the call types were higher in the 5-cluster case (Table 8). The best identified call types are pant-hoots, screams, and waa-barks, with pant-hoots being the best identified.

Table 8: Summary of the precision, recall, and accuracy values for K-means clustering applied on LMA acoustic features, and choosing the number of clusters as either 5 or 3.

K-means (5 clusters)					
Metric	**Bark**	**Pant-hoot**	**Rough-grunt**	**Scream**	**Waa-bark**
Precision	36.7	72.7	83.3	60	79.7
Recall	28.2	83.6	52.1	91.3	72.8
F1 score	31.9	77.8	64.1	72.4	76.1
Class Accuracy	60.2	86.9	75	89.7	82.7
Overall Accuracy [95% CI]	68.7 [62.9, 74.1]				
K-means (3 clusters)					
Precision	NA	61.2	NA	54.4	60.6
Recall	0	89.6	0	93.5	77.8
F1 score	NA	72.7	NA	68.8	68.1
Class Accuracy	50	85.9	50	89.1	78.6
Overall Accuracy [95% CI]	59.1 [53.1, 64.9]				

1.4.5 Unsupervised k-means clustering with speech recognition features

Figure 5 shows the scree plot for the k-means clustering applied to the standardized speech recognition features with k ranging from 2 to 10. Compared to LMA features, the elbow in this scree plot is more clearly visible at 3 clusters. Cluster 1 predominantly contains waa-barks but overlaps with screams and barks, cluster 2 mainly corresponds to pant-hoots, and cluster 3 predominantly contains rough-grunts, but overlaps with barks, waa-barks, and screams (Table 9). In the 5-cluster case, there seem to be clearer clusters for waa-barks (cluster 2), pant-hoots (cluster 4), and screams (cluster 5). However, in contrast to the 3-cluster case, rough-grunts seem to be forming two dominant clusters (1 and 3) that have some overlaps with barks and waa-barks (Table 9). This suggests the existence of two distinct kinds of rough-grunts. This pattern is also supported in the 6-cluster case where the rough-grunts form two distinct clusters (2 and 6) that overlap with barks and waa-barks, whereas pant-hoots (cluster 3), screams (cluster 4), and waa-barks (cluster 1) form their own distinct clusters. In the 7-cluster case too, the rough-grunts from two distinct clusters (1 and 4), which get better differentiated from barks and waa-barks. The waa-barks form two distinct clusters too (3 and 6) that show some overlaps with pant-hoots and barks. The pant-hoots (cluster 2) and screams (cluster 7) form distinct clusters of their own, whereas barks show overlaps with waa-barks and screams (Table 9).

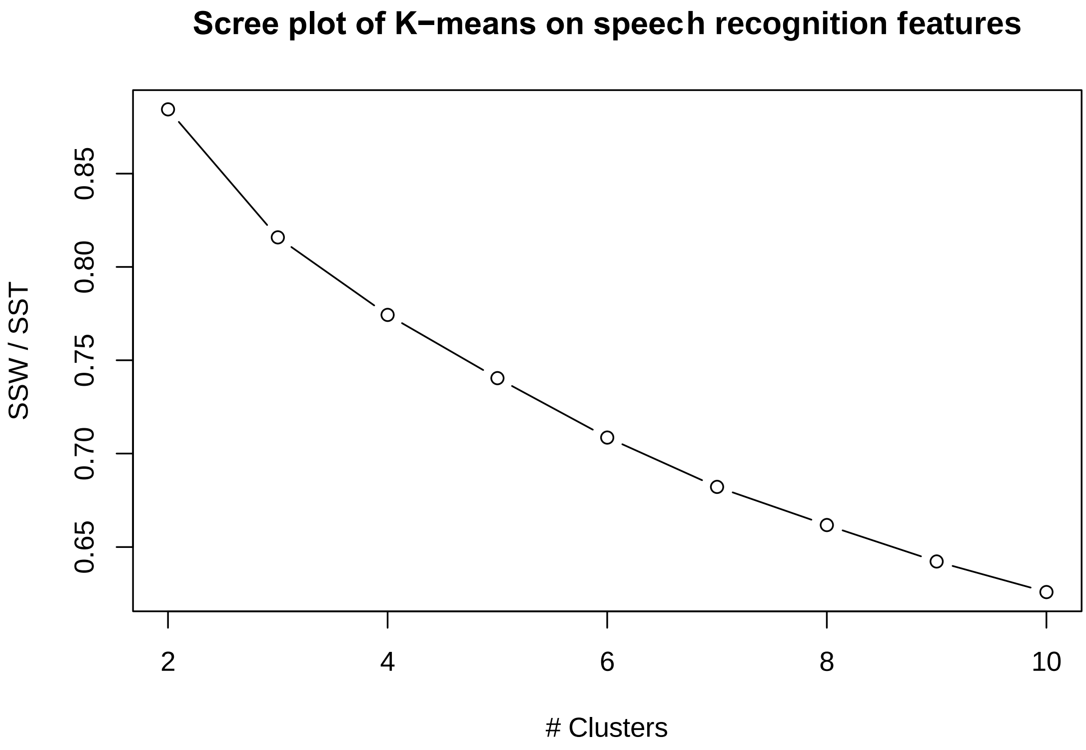


Figure 5: Scree plot of k-means clustering applied to standardized LMA features with k ranging from 2 to 10. SSW = within-cluster sum of squares, SST = total sum of squares.

Table 9: Results from k-means clustering on speech recognition features with k = 3, 5, 6, and 7. Rows indicate how many of a given call type were classified into a given cluster ID in a k-means analysis.

Cluster ID	**Bark**	**Pant-hoot**	**Rough-grunt**	**Scream**	**Waa-bark**
K = 3 clusters					
1	14	4	1	35	57
2	1	57	1	0	9
3	24	6	46	11	15
K = 5 clusters					
1	13	4	20	2	4
2	11	3	1	5	47
3	8	3	26	2	11
4	1	56	1	0	8
5	6	1	0	37	11
K = 6 clusters					
1	11	9	1	3	47
2	8	3	17	2	9

3	1	49	5	0	7
4	6	0	0	39	11
5	0	1	5	0	2
6	13	5	20	2	5
K = 7 clusters					
1	4	2	13	2	6
2	0	43	3	0	3
3	4	17	0	0	22
4	2	4	21	0	0
5	13	0	9	1	9
6	8	0	1	6	28
7	8	1	1	37	13

Unlike LMA features, comparing the optimal number of clusters based on the scree plot (i.e., 3 clusters), with the number of call types in the data as I classified, the 5-cluster case doesn’t perform better in all cases. Instead, the performance varies by the call type. The F1 score for rough-grunts was higher in the 3-cluster case and for waa-barks was higher in the 5-cluster case. The F1 score for pant-hoots was similar in both cases (Table 10). The best identified call types are pant-hoots, rough-grunts, and waa-barks, with pant-hoots being the best identified, and even better identified than LMA features.

Table 10: Summary of the precision, recall, and accuracy values for K-means clustering applied on speech recognition acoustic features, and choosing the number of clusters as either 5 or 3.

K-means (5 clusters)					
Metric	**Bark**	**Pant-hoot**	**Rough-grunt**	**Scream**	**Waa-bark**
Precision	30.2	84.8	52	67.3	70.1
Recall	33.3	83.6	54.2	80.4	58
F1 score	31.7	84.2	53.1	73.3	63.5
Class Accuracy	60.5	89.5	71.9	86.4	74
Overall Accuracy [95% CI]	63.7 [57.8, 69.3]				
K-means (3 clusters)					
Precision	NA	83.8	45.1	NA	51.4

Recall	0	85.1	95.8	0	70.4
F1 score	NA	84.4	61.3	NA	59.4
Class Accuracy	50	90	85.9	50	71.7
Overall Accuracy [95% CI]	56.9 [50.9, 62.8]				

1.4.6 Unsupervised principal components analysis with LMA and speech recognition features

Figure 6 incudes the plot of the data on the first two principal components after performing principal components analysis on the LMA features using the correlation matrix. The waa-barks, screams and pant-hoots form relatively distinct clusters on the principal component 1, but the barks and rough-grunts, and the pant-hoots and rough-grunts overlap substantially. On the principal component 2, pant-hoots, rough-grunts, screams, and barks overlap substantially and none of the call types seem to form a distinct cluster. This suggests that at least the first two principal components are not likely to yield the distinct call types that I classified.

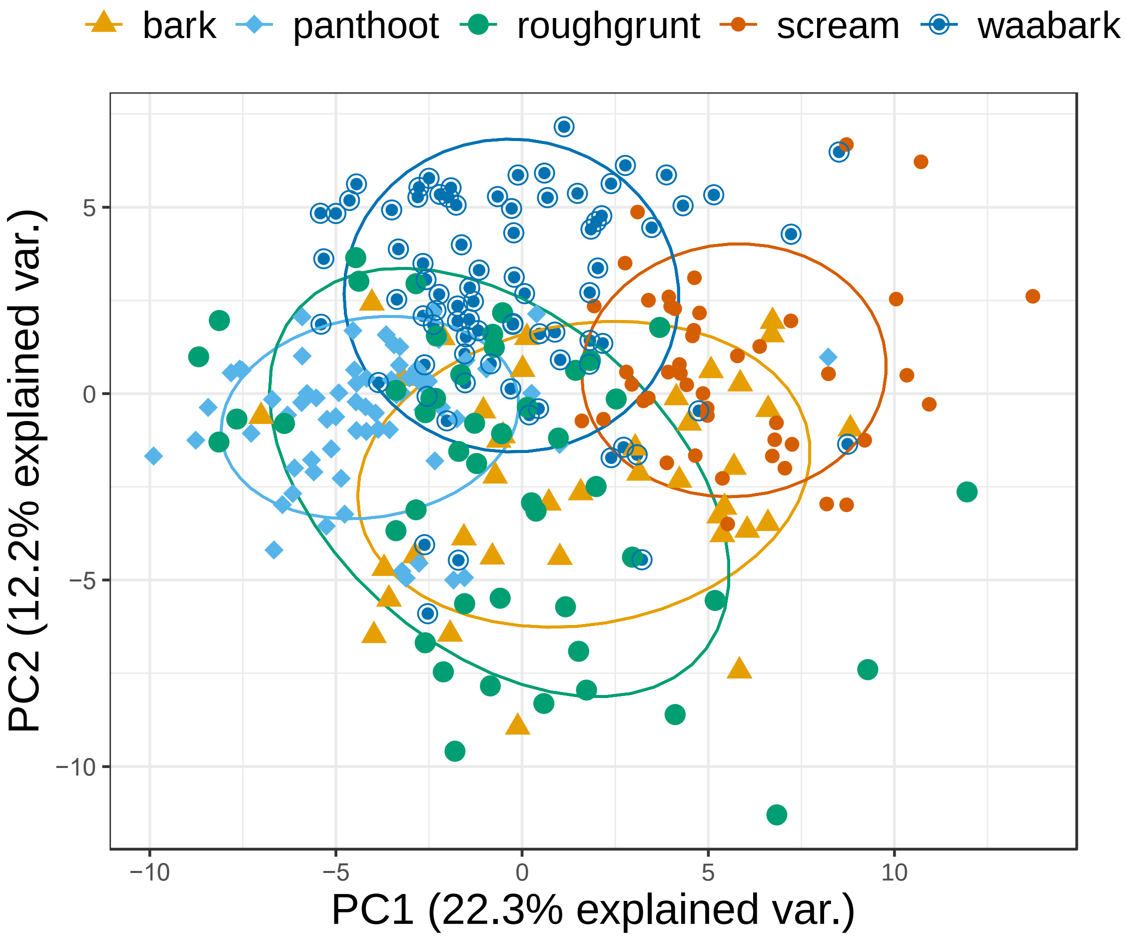


Figure 6: The data plotted on PC1 and PC2 of the principal components analysis performed on the LMA features using the correlation matrix.

Contrary to results from the LMA features, principal components analysis on the speech recognition features results in principal components that show a much better distinction in the multidimensional space (Figure 7). The pairs of waa-barks and rough-grunts, and the rough-grunts and the pant-hoots form relatively distinct clusters on the principal component 1, and the pant-hoots, screams and a cluster containing waa-barks, barks, and rough-grunts form relatively distinct clusters on the principal component 2, with the 3-call cluster showing some overlap with the screams. All call types but the rough-grunts overlap substantially on the principal component 1. Rough-grunts also show two clusters on principal component 1. Additionally, the barks overlap substantially with the waa-barks, and with the screams on both the principal components. The barks also show some overlap with the rough grunts on the principal component 1. However, the overall distinction of call types is better than the first two principal components of LMA features.

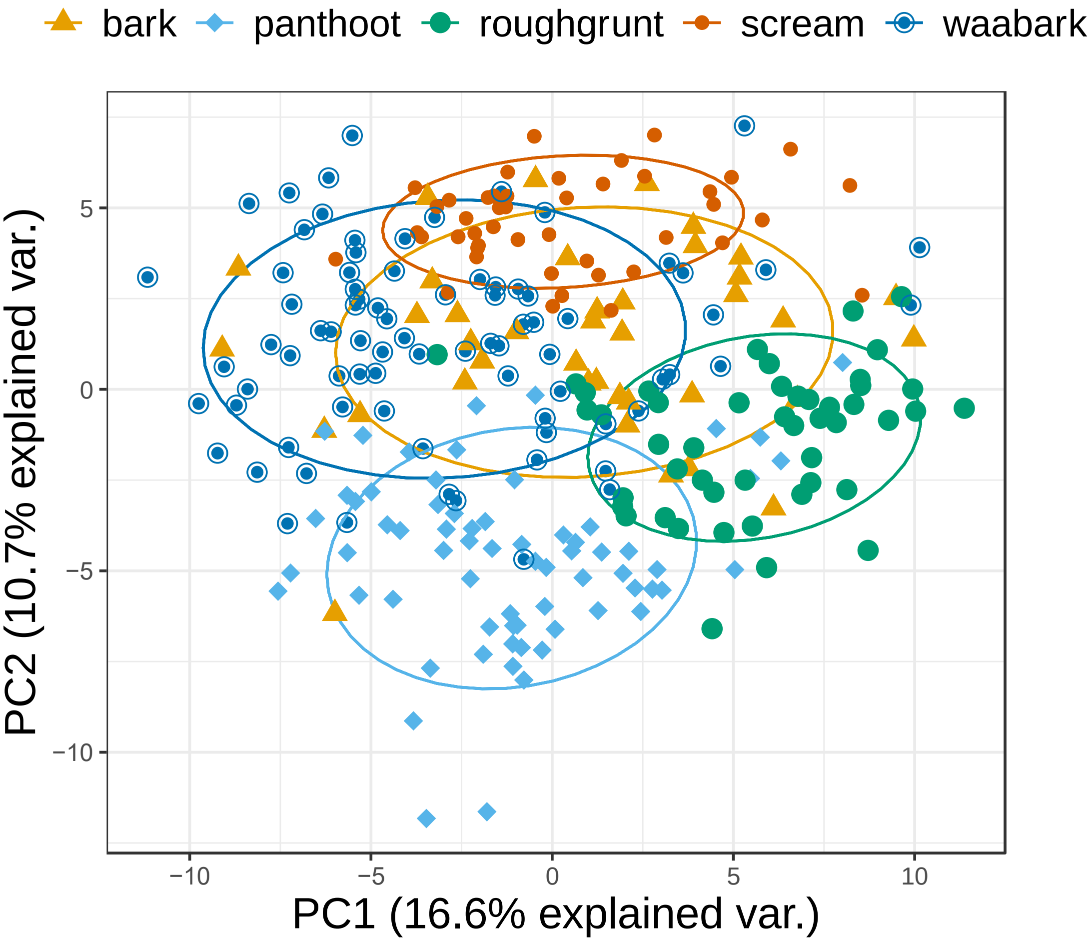


Figure 7: The data plotted on PC1 and PC2 of the principal components analysis performed on the speech recognition features using the correlation matrix.

1.5 Discussion

I compared the classification performance of three different supervised machine learning techniques (discriminant functions analysis, random forests, and artificial neural networks) when applied to two different sets of acoustic features (LMA features and speech recognition features). Newer classification techniques including random forests and artificial neural networks applied on speech recognition or LMA features overall performed better than the traditional approach involving the application of discriminant functions analysis on LMA features. However, discriminant functions analysis applied on LMA features performed the best for pant-hoots in terms of the F1 score (i.e., in terms of both precision and recall), meaning that discriminant functions analysis was the best technique to (i) detect most of the pant-hoots in a given dataset (recall), and (ii) detect pant-hoots more precisely while not incorrectly detecting other call types as pant-hoots (precision). For other call types and LMA features, random forests had the highest F1 score and overall accuracy. For speech recognition features, the performance of different classification techniques varied based on call type. In terms of F1 score, for pant-hoots, artificial neural networks performed the best. Artificial neural networks also performed the best for waa-barks. Discriminant functions analysis performed the best on screams, and random forests on barks and rough-grunts. Artificial neural networks had the highest overall accuracy. When comparing the two different kinds of acoustic features, the best combination of acoustic features and classification technique varied by call type. Both kinds of acoustic features performed similarly (within 5 percentage points of the maximum F1 score) for barks, screams, waa-barks, and pant-hoots. The performance was substantially higher for LMA features only for rough-grunts. For the call type of interest for the rest of this dissertation (pant-hoots), both kinds of acoustic features had similar best performance, but the classification technique that worked the best for LMA features was discriminant functions analysis, whereas for speech recognition features was artificial neural networks.

Next, I applied two unsupervised techniques (k-means clustering and principal components analysis) that attempt to find clusters in the data without *a priori* knowledge of call types. In unsupervised clustering using the k-means clustering algorithm, LMA features had a higher overall accuracy compared to speech recognition features. However, speech recognition features better detected pant-hoots than LMA features in the k-means clustering. In unsupervised principal components analysis, the speech recognition features resulted in more distinct clusters on the first two principal components than the LMA features. Furthermore, k-means clustering with 6 or 7 clusters suggested the existence of at least two kinds of rough-grunts, barks, and waa-barks demonstrating the utility of unsupervised methods to detect call types that may not be apparent to human listeners.

These findings demonstrate that the supervised machine learning techniques commonly used in artificial intelligence research (random forests and artificial neural networks) when applied to acoustic features commonly used in speech recognition research (MFCCs and other acoustic features) perform similarly to the commonly used supervised machine learning technique in primatology (discriminant functions analysis) applied on acoustic features commonly used in primatology (LMA features). However, it seems likely that artificial neural networks applied to speech recognition features would have advantages when working with larger datasets. Previous studies have found that artificial neural networks have superior performance when trained on large datasets, using many more hidden layers**,** rather than the single hidden later used for this study (Bermant et al., 2019; Hinton et al., 2012; LeCun et al., 2015). Speech recognition features were better at separating graded call types, such as rough-grunts, as demonstrated by their superior performance in the unsupervised principal components analysis. This suggests that may be promising in identifying calls that are not be audible to human listeners. These findings provide a starting methodological framework for future studies of quantitative vocal repertoire classification in chimpanzees.

My findings provide further justification for the choice of discriminant functions analysis applied to LMA or LMA-like features (the fundamental frequency, the peak frequency and their other features) in the studies of vocal repertoire classification and vocal communication in primates (Bernstein et al., 2016; Crockford et al., 2004; Fischer et al., 2013; Fuller, 2014; Gamba & Giacoma, 2007; Gros-Louis et al., 2008; Hammerschmidt & Fischer, 1998, 2019; Pozzi et al., 2010). This is because they performed similarly to other machine learning techniques applied on speech recognition features. The classification accuracies I achieved are relatively low compared to other studies of primate vocal repertoire classification. For example, Hammerschmidt and Fischer (2019) obtained 96.1% cross-validated classification accuracy using discriminant functions analysis for chacma baboons **(***Papio ursinus***)** and 97.3% for Guinea baboons **(***Papio papio***)** for 6 major call types (Hammerschmidt & Fischer, 2019); Fuller (2013) obtained 83.4% cross-validated classification accuracy using discriminant functions analysis for 6 major call types produced by adult male blue monkeys (*Cercopithecus mitis stulmanni*) (Fuller, 2014); Gamba & Giacoma (2007) obtained a cross-validated classification accuracy of 91.9% using discriminant functions analysis on 10 call types produced by crowned lemurs (*Eulemur coronatus*) (Gamba & Giacoma, 2007); Bernstein et al. (2016) obtained 88.8% cross-validated classification accuracy using discriminant functions analysis on 11 call types produced by Tibetan macaques (*Macaca thibetana*) (Bernstein et al., 2016). Whereas I obtained 73.2% cross-validated classification accuracy using discriminant functions analysis on LMA features and the highest I obtained was 78% using random forests on LMA features. However, this might be an outcome of the kinds of call types included in my analysis. I included barks, waa-barks, and screams, which intergrade with one another, and also rough-grunts, which include multiple different graded call types as suggested by the unsupervised k-means clustering, as well as previous studies (O'Bryan, 2015; Slocombe & Zuberbühler, 2005). I achieved comparable or better accuracies than achieved in other studies for graded or similar call types within species, or similar call types among species. For example, when looking at grunts that tend to be similar among species, Hammerschmidt & Fischer (2019) obtained 64.2% cross-validated classification accuracy in identifying three species of baboons (olive (*Papio anubis*), chacma, and Guinea baboons) using discriminant functions analysis LMA features, whereas I obtained 78.6% accuracy on rough-grunts using discriminant functions analysis and even higher 96.3% using random forests on LMA features. Fuller (2014) obtained 68% accuracy for blue monkey *ka-train*-units and 60% for *kas*-units that are similar sounding. For similar sounding barks and waa-barks, I obtained 69.9% and 82.2% accuracies respectively using discriminant functions analysis, and best performance of 74.5% and 84.4% respectively using random forests on LMA features. Bernstein et al. (2016) obtained accuracies of 100% of tonal screams, and 87% for noisy screams using discriminant functions analysis in Tibetan macaques. I obtained 85.6% accuracy on tonal/atonal screams using discriminant functions analysis, which could be improved to 91.6% using random forests on LMA features.

The superior performance of artificial neural networks on speech recognition features matches expectations from human speech recognition literature (Barua et al., 2014; Das & Parekh, 2012; Moselhy & Abdelnaiem, 2013; Srinivasan et al., 2014). Artificial neural networks work well for applications in bioacoustics (Armitage & Ober, 2010; Xie et al., 2016) as well and have been shown to work well for primates whether applied to speech recognition features (Mielke & Zuberbühler, 2013) or not (Pozzi et al., 2010). The overall highest accuracy of 76.8% that I obtained by using artificial neural networks on speech recognition features was much lower than other studies of primate vocal repertoire classification that used similar methods. Mielke & Zuberbühler (2013) obtained an accuracy of 98% for blue monkey calls known as pyows and hacks, and Pozzi et al. (2010) obtained an average accuracy of 93.7% for seven call types in black lemurs (*Eulemur macaco macaco*). Mielke & Zuberbühler (2013) also found that artificial neural networks applied on speech recognition features outperformed discriminant functions analysis in identifying individuals based on a single pyow call or from a sequence of pyows. This finding is partly consistent with mine as I found that artificial neural networks applied to speech recognition features outperform discriminant functions analysis in three (barks, pant-hoots, and waa-barks) of the five call types analyzed. However, notably, discriminant functions analysis outperformed for two (rough-grunts and screams) call types, suggesting that Mielke & Zuberbühler's finding might be because they only looked at one discrete call type. Similarly, Pozzi et al. (2010) found that artificial neural networks outperformed discriminant functions analysis in terms of overall classification accuracy (93.7% vs. 83.7%), but the performance difference varied based on call types with some call types better classified by discriminant functions analysis and others by artificial neural networks. This is similar to my results where I found that artificial neural networks when applied to speech recognition features outperformed discriminant functions analysis in terms of overall accuracy (76.8% vs. 70.7% respectively) but the performance varied by call type as noted above. Notably, this was not true for LMA features where both techniques performed similarly in terms of overall accuracy (73.2% for discriminant functions analysis vs. 72% for artificial neural networks). However, despite the similar performance of artificial neural networks to discriminant functions analysis, my study remains limited in its ability to evaluate the full potential of artificial neural networks. This is because I used a neural network with only one hidden layer on a small sample size. Most of the success attributed to the applications of artificial neural networks involve using networks with many hidden layers, known as deep neural networks on large datasets (Hinton et al., 2012; LeCun et al., 2015; Mikolov et al., 2011). Given that artificial neural networks already achieve similar performance with only one hidden layer and a small dataset, a promising future direction is to use deep neural networks with speech recognition features on larger datasets of primate vocalizations (Zhang et al., 2018).

My results from unsupervised classification using k-means clustering and principal components analysis demonstrate that the call types can be classified with relatively high accuracy even without *a priori* knowledge of call types. This makes the unsupervised techniques potentially useful for eliminating human biases. Some previous studies of primate vocal repertoire classification have identified this potential and used call types resulting from unsupervised classification for vocal repertoire classification (Fuller, 2014; Hammerschmidt & Fischer, 1998). For example, Fuller (2014) used a combination of principal components analysis and hierarchical cluster analysis to confirm human classified call types before applying discriminant functions analysis for call type classification. Hammerschmidt & Fischer (1998) used k-means clustering to identify clusters of call types in the acoustic feature space, which they then used to make biological inferences about age, sex, and context differences in Barbary macaques. For future studies, I argue that this is the better approach to take for studying animal vocalizations—instead of relying on human identified call types, a less biased approach would involve clusters identified using unsupervised approaches that don't rely completely on human hearing. The quality of clusters could be improved in terms of validity by using acoustic features that incorporate perception, such as MFCCs, which could be further modified to incorporate species-specific perceptual parameters.

My study provides a proof-of-concept methodological framework for analysis of primate vocalizations using different kinds of acoustic features and machine learning techniques. However, several limitations remain in the design that could be improved in future studies. First, this analysis is on call type that are classified by humans (Plooij and me). Hence, I am working with the assumption that these intuitive classifications are meaningful, and then testing the performance of different methods on these intuitive classifications. To test this assumption, we need perceptual experiments on chimpanzees, such as playbacks to truly test if chimpanzees are perceiving these call types discretely. Thus, an important future research direction is to combine perceptual experiments along with unsupervised classifications to have a more objective understanding of the vocal repertoire. Second, my dataset includes a small sample size of identifiable call types from only one geographical population of chimpanzees. A more robust and widely applicable analysis should incorporate a larger sample of vocalizations from multiple field sites across Africa to sufficiently capture the variation in chimpanzee vocal repertoire and to come up with a common set of names for call types that could be used across all studies of chimpanzee vocal communication. Third, I only included five call types based on my own classifications. Previous classifications of chimpanzee vocal repertoire have included as many as 39 call types (Plooij et al., 2014, 2015). Future analyses should incorporate a greater number of call types and could decide number of call types using unsupervised classification techniques or employ a hierarchical classification scheme such as that described by Crockford (2019) that categorized call types into a few major classes and other calls as variants of major classes. Ideally, these should be validated using perceptual experiments as noted in the first point. Fourth, my procedure for measuring acoustic features involved averaging acoustic features measured at small overlapping windows of time, over the duration of the call. While this is the standard approach in bioacoustics, averaging over the duration results in loss of some temporal properties of the call, which may be important for the perception of the call. So, in future analyses, classification techniques that can model time-series data, such as Hidden-Markov Models should be compared with aforementioned deep neural networks to ensure capturing temporal information (Ananthi & Dhanalakshmi, 2015; Hinton et al., 2012). In conclusion, it appears that the supervised machine learning techniques result in a reasonably high accuracy and will be useful in modelling with a greater number of call types. The unsupervised methods are also separating call types well and may also give insights about the gradation in the call types that may not be immediately apparent to human listeners.

1.6 Conclusions

I performed a broad analysis of different acoustical analysis and statistical analysis techniques with the goals of (i) establishing a methodological framework for chimpanzee vocal repertoire classification, and (ii) informing methods for testing other hypotheses in this dissertation. The first goal remains important broadly for the chimpanzee vocal communication research community as the chimpanzee vocal repertoire still remains poorly understood in quantitative terms (Crockford, 2019). I identified this to be a major hurdle for other aims of my dissertation as answering functional questions about chimpanzee vocalizations first requires a good understanding of the variation in chimpanzee vocalizations. I aimed to study the functions of the variation in chimpanzee pant-hoots, but I was struck by the naturally occurring variation in this complex call. There was substantial apparent variation in the composition of different call elements in the pant-hoots, both within-, and among-individuals. Such high variation could make statistical inference about functions highly sensitive to sampling variation, especially with the small sample sizes that are common in vocal communication research. While the quality of data is always more important than the analysis technique employed, it was clear to me that with the limited sample sizes that I would have for my dissertation, I needed as high-quality analysis techniques as possible to ensure best inference. Hence, I explored approaches taken by different disciplines such as speech recognition and artificial intelligence to see if I could improve on existing methods. I chose acoustical analysis and statistical classification techniques that have been found to be highly reliable in these fields. This exercise was relevant for the next chapter as there I test whether chimpanzees have vocal dialects in the pant-hoots, a question that a few previous studies have tried to answer. I had to balance the trade-off between choosing new techniques that may perform better than existing techniques and choosing the existing techniques that may underperform but have the advantage of making my study comparable to the previous studies of chimpanzee dialects. In this study, I found that supervised machine learning techniques such as random forests and artificial neural networks with speech recognition features only slightly improved the performance over existing techniques employed in primatology such as LMA features and discriminant functions analysis, especially for the call type of my interest, the pant-hoots. It was clear that this slight performance advantage was not enough to sacrifice the ability to compare my results with previous findings. Hence, for subsequent chapters, particularly for chapter 2, I chose to use LMA features and discriminant functions analysis, to ensure direct comparability with previous studies.

Chapter 2

Does vocal-learning influence the variation in chimpanzee pant-hoots via learned acoustic cues of community membership?

2.1 Overview

Vocal learning, the ability to voluntarily modify the acoustic structure of vocalizations based on social cues, is a fundamental feature of speech in humans (*Homo sapiens*). While vocal learning is common in taxa such as songbirds and whales, the vocal learning capacities of nonhuman primates appear more limited. Intriguingly, evidence for vocal learning has been reported in chimpanzees (*Pan troglodytes*), for example in the form of regional variation ('dialects') in the 'pant-hoot' calls. This suggests that some capacity for vocal learning may be an ancient feature of the *Pan-Homo* clade. Nonetheless, reported differences have been subtle, with inter-community variation representing only a small portion of the total acoustic variation. To gain further insights into the extent of regional variation in chimpanzee vocalizations, I performed an analysis of pant-hoots from chimpanzees in the neighboring Kasekela and Mitumba communities at Gombe National Park, Tanzania, and the geographically distant Kanyawara community at Kibale National Park, Uganda. I did not find any statistically significant differences between the neighboring communities at Gombe or among geographically distant communities. Furthermore, I found differences among individuals in all communities. Hence, the variation in chimpanzee pant-hoots reflected individual differences, rather than group differences. Thus, I did not find evidence of dialects in this population, suggesting that extensive vocal learning emerged only after the lineages of *Homo* and *Pan* diverged.

2.2 Introduction

Vocal learning underlies the human capacity for speech. The desire to understand the evolution of this capacity motivates much of the research into vocal learning in other animals (Fitch, 2010). Over time, the definition of vocal learning has evolved as researchers have identified several nuances in vocal learning ability across animals. Janik & Slater (2000) defined vocal production learning broadly, as “signals modified in form as a result of experience with those of other individuals.” (Janik & Slater, 2000). Other researchers have focused on more specific aspects, such as the ability to voluntarily modify and learn new vocalizations through imitation (Fitch, 2010). Regardless of the particular definition used, it is clear that vocal learning has evolved independently multiple times in animals (Vernes et al., 2021). For example, songbirds (Passeriformes) (Cunningham & Baker, 1983) and humpback whales (*Megaptera novaeangliae*) (Garland et al., 2011) learn elaborate songs. Parrots (Psittaciformes) can mimic human speech, and distinguish group members from drifters based on learned vocalizations (Bartlett & Slater, 1999; Hile & Striedter, 2000).

In comparison to birds and whales, the vocal learning capacities of nonhuman primates appear much more limited (Fischer & Hammerschmidt, 2020). Little evidence exists for active learning of new vocalizations by non-human primates (Tyack, 2020), although recent studies indicate that orangutans (*Pongo spp.*) can acquire a voiceless signal (whistle) in captivity (Wich et al., 2009) and produce novel voiced vocalizations in controlled settings (e.g., using a membranophone (Lameira & Shumaker, 2019)). Some non-human primates have been reported to engage in vocal learning through modifying the acoustic structure of vocalizations based on auditory feedback and imitation. Takahashi and colleagues (2015) found that in common marmosets (*Callithrix jacchus*), parental feedback influences the rate of vocal development (Takahashi et al., 2015). Marmosets (*Callithrix spp.*) exhibit geographical variation in their vocalizations in the wild (de la Torre & Snowdon, 2009) as well as population specific acoustic structure across call types in captivity (dialects) (Zürcher & Burkart, 2017). Sugiura reported that Japanese macaques (*Macaca fuscata*) match some of the acoustic features of recorded ‘coo’ calls during a playback experiment (Sugiura, 1998). Fischer and colleagues (2020) reported vocal convergence in the grunts of male Guinea baboons (*Papio papio*) as individuals that interacted more frequently with one another exhibited greater resemblance than the grunts of males that interacted less frequently (Fischer et al., 2020).

Much of the literature on vocal learning in animals focuses on dialects, defined as regional variation in vocal production (Nowicki & Searcy, 2014). When such variation is learned, it may signal membership in the local population (as in songbirds (Cunningham & Baker, 1983)), or membership in a particular social group, as in orcas (*Orcinus orca*) (Filatova et al., 2012). Studies of social birds and mammals have found that learned signals of group membership can benefit individual signalers in two main ways: (i) by eliciting affiliative interactions from group members and mates and/or (ii) by advertising group membership to rivals during agonistic interactions, such as during territory defense. For example, in birds, group-specific calls appear to (i) help maintain social bonds among group members, as in budgerigars, *Melopsittacus undulatus* (Farabaugh et al., 1994; Hile & Striedter, 2000); (ii) facilitate territory defence by helping individuals identify flock members and focus aggression on foreign callers, as in black-capped chickadees, *Parus atricadpillus* (Nowicki, 1983). Researchers have inferred similar functions in social mammals. For example, several species of toothed whales (Odontocetes) appear to use vocal dialects to facilitate spatial group cohesion and maintain social relationships (Janik, 2014; Tyack & Sayigh, 1997). Spatial cohesion in group-living species facilitates maintaining social bonds, finding mates, and defending territories (Janik & Slater, 1998).

Although vocal data from all great apes is informative for understanding the evolution of language (Lameira & Call, 2020), historically, researchers interested in the origins of human language have particularly focused on the vocal behavior of chimpanzees (*Pan troglodytes*), given that they are one of the two living species most closely related to humans (Fedurek & Slocombe, 2011). Several studies from the field (Arcadi, 1996; Crockford et al., 2004; Mitani et al., 1992) and captivity (Marshall et al., 1999) have found evidence for regional variation (dialects) in chimpanzee ‘pant-hoot’ calls, which has been proposed to result from vocal learning (Crockford et al., 2004; Marshall et al., 1999). Pant-hoots of males that spend more time together are more similar, and the acoustic features of their calls converge when chorusing together (Mitani & Brandt, 1994), suggesting a possible mechanism for the convergence of acoustic properties within groups (Fedurek, Schel, et al., 2013; Mitani & Gros-Louis, 1998). Call convergence has also been reported for chimpanzee rough-grunt calls (Watson et al., 2015a), but see (Fischer et al., 2015) and (Watson et al., 2015b). Chimpanzees live in groups with fission-fusion dynamics, in which individuals travel in subgroups (known as 'parties') of varying size, and they communicate over long distances using vocalizations, often in noisy environments. Thus, vocal dialects potentially facilitate spatial group cohesion, and territorial defense during intergroup encounters.

Chimpanzee pant-hoots are structurally complex loud calls with a relatively consistent temporal patterning. The typical pattern consists of a sequence of four kinds of sound elements over a duration range of 2-20s. Each sequence of similar elements is called a phase and so the pant-hoots typically have four phases (see Methods for details). Of the four phases, one (climax) phase is the loudest, and can be heard most clearly over long distances. However, pant-hoots exhibit considerable acoustic variation within and among individuals (Fedurek, Schel, et al., 2013; Marler & Hobbett, 1975). The variation is not only limited to frequency properties of elements such as fundamental frequency, peak frequency, etc., but also involves variation in the number and presence/absence of different elements and phases (Appendix; Figure 29-Figure 41). Chimpanzees use pant-hoots in a variety of intra-community and inter-community contexts. In intra-community contexts, chimpanzees use pant-hoot calls to communicate with members of their own community over long distances (Goodall, 1986). Pant-hoots may function to communicate the caller's location to allies and associates within their own community (Goodall, 1986; Mitani & Brandt, 1994). Further, pant-hoots play a role in facilitating social bonds as affiliative partners chorus more together (Fedurek, Machanda, et al., 2013) and play a role in regulating grouping dynamics by attracting allies and potential mates to the caller's location (Fedurek et al., 2014; Mitani & Nishida, 1993; Wrangham, 1977). In inter-community contexts, interactions often involve hearing — and sometimes responding to — pant-hoots from callers that are hundreds of meters away, far out of view (Wilson et al., 2012). The long-distance nature of pant-hoots allows chimpanzees to use pant-hoots to advertise territory ownership (Wilson et al., 2007), and to signal numerical strength to members of neighboring communities during agonistic intergroup encounters (Herbinger et al., 2009; Wilson et al., 2001, 2012). Individual callers might thus benefit from encoding community-specific cues. Playback experiments have demonstrated that chimpanzees can distinguish stranger pant-hoots from those of familiar individuals (Herbinger et al., 2009) and that they are sensitive to numerical strength during intergroup encounters, being more likely to respond to simulated intruders when they are in parties with more males (Wilson et al., 2001). Hence, community-specific dialects could play a role in cooperative defense by signaling community membership. While genetic similarity could lead to community-specific vocalizations, socially learned signals of group membership might be useful in cases where not all group members are close genetic kin.

Despite these reasons for thinking that vocal dialects would benefit chimpanzees, current evidence raises several questions about the extent to which chimpanzees have socially learned signals of group membership. In the first study of chimpanzee dialects, Mitani and colleagues reported differences between Gombe and Mahale pant-hoots and suggested that they may be an outcome of vocal learning (Mitani et al., 1992). However, the differences among the communities were subtle compared to differences observed in songbirds (Cunningham & Baker, 1983) or whales (Garland et al., 2011). Mitani observed geographical differences in the composition for only one of the four phases of the pant-hoot call — the build-up, and in frequency properties of another phase — the climax (Mitani et al., 1992). Mitani & Brandt (1994) later found that in a principal components analysis of acoustic structure, community membership accounted for only 0-11% of the variance on the principal components, compared to within-individual factors (48-79% of the variance) and between individual factors (17-52 % of the variance). Mitani further reassessed his findings, pointing out that since Gombe and Mahale are far from one another (~160 km) and likely genetically isolated, the acoustic differences may not necessarily represent vocal learning, but instead could represent genetic differences and/or body size (Mitani et al., 1999). Additionally, other environmental factors like habitat acoustics and/or sound environment might be more important in explaining the variation in such geographically distant communities.

In addition to assessing whether pant-hoots signal group membership, researchers have studied the acoustic structure of pant-hoots produced in different contexts, such as traveling, feeding, group fusion, and arrival at food sources (Clark & Wrangham, 1993, 1994; Fedurek, Zuberbühler, et al., 2016; Goodall, 1986; Mitani & Nishida, 1993; Notman & Rendall, 2005; Uhlenbroek, 1996; Wrangham, 1977). Some studies reported an association of some properties of the letdown phase of the pant-hoots with the context (Clark & Wrangham, 1993; Fedurek, Zuberbühler, et al., 2016; Notman & Rendall, 2005). Notman & Rendall (2005) and Uhlenbroek (1996) reported an association of the tonal structure of the climax scream element of the pant-hoots with the context of the production. While this variation provides information about context to receivers, Notman & Rendall (2005) argued that these differences are unlikely to be an outcome of vocal learning and are more likely to reflect arousal states of chimpanzees when calling. In any case, the context of the call production is a covariate that may need to be controlled for when testing for group differences (refer to the methods and the Directed Acyclic Graph in Figure 9). Finally, as Marler & Hobbett (1975) noted previously, pant-hoots are individually distinctive. Signaling individual identity, rather than group membership, might therefore be the primary function of these calls.

To test the extent to which the acoustic structure of pant-hoots specifically signals community membership and arises out of vocal learning via auditory feedback, three questions need to be answered: (i) Do the calls contain reliable acoustic cues of community membership that might allow chimpanzees to distinguish extra-community pant-hoots based on those cues alone, rather than through familiarity with the calls of particular individuals? (ii) Do chimpanzees from neighboring communities have more distinct pant-hoots than those from geographically distant communities? Greater differences among neighboring communities compared to geographically distant communities would indicate that chimpanzees are actively modifying the acoustic structure of pant-hoots to differentiate their calls from those of neighbors. (iii) Does community membership explain vocal similarity better than genetic relatedness? Crockford et al. (2004) addressed all three of these questions by comparing genotyped individuals in three neighbouring communities and one more distant community in Taï National Park, Côte d’Ivoire. They found that neighboring communities differed from one another more than they differed from the distant community, despite neighboring communities inhabiting adjacent areas of similar continuous forest environment, which supports the view that chimpanzees learned to produce an acoustic structure distinct to their own community. This study thus supports the view that vocal learning accounts for the acoustic differences among communities. However, sample sizes in this study were small, with calls from only three individuals per group analyzed, raising the possibility that the findings are a statistical artifact resulting from small sample size. While it is well known that small sample sizes may led to false negatives, the possibility of small sample sizes leading to false positives is less well acknowledged. A small sample size with noisy data could artificially exaggerate effect sizes and lead to false positives (Loken & Gelman, 2017). Hence, more studies are needed to replicate these findings in order to have more confidence in the results.

As a step towards re-evaluating the role of vocal learning in chimpanzee calls, we recorded pant-hoot calls from two neighboring chimpanzee communities in Gombe National Park, Tanzania and the geographically distant Kanyawara community of chimpanzees in Kibale National Park, Uganda. The objective of this study is to assess the extent to which variation in the acoustic structure of the pant-hoots can be explained by community membership. To that end, I test two hypotheses. My first hypothesis is: the acoustic structure of pant-hoots contains features that provide reliable cues of community membership. In line with Crockford et al (2004), if vocal learning is shaping the acoustic structure of pant hoots into community-specific dialects, we would expect to find greater differences in the structure of calls in the two neighboring Gombe communities, compared to the geographically distant Kanyawara community. The second hypothesis is: the acoustic structure of pant-hoots contains cues of individual identity more than community identity. While these are not mutually exclusive hypotheses (i.e. one or both or neither could be supported), they provide a framework for my research questions.

2.3 Methods

2.3.1 Subjects and study sites

I studied chimpanzees at two study sites: Gombe National Park, Tanzania and Kibale National Park, Uganda. In Gombe, I studied two neighboring communities: Kasekela and Mitumba. In Kibale, we studied the chimpanzees of the Kanyawara community. Gombe is located in western Tanzania, along the shore of Lake Tanganyika (4°40′S, 29°38′E). At the time of the study, Gombe had three contiguous communities of chimpanzees, two of which (Kasekela and Mitumba) were well habituated and were followed nearly every day, throughout the day as part of the long-term research at Gombe. Kibale is located in western Uganda (0°33’N, 30°21’E). I analyzed calls recorded as a part of a previous chimpanzee vocal communication study at Kanyawara (Fedurek, Schel, et al., 2013). Following initial observations by Isabirye-Basuta in 1983-1985 (Isabirye-Basuta, 1988), the Kanyawara chimpanzees have been studied continuously since 1987 (Emery Thompson et al., 2020; Wrangham et al., 1992).

For this study, I included male chimpanzees ≥14 yr, by which age male chimpanzees at Gombe have reached approximately adult body mass (Pusey et al., 2005), and exhibit a marked increase in rate of pant-hoot production (Pusey, 1990).

<u>2.3.2 Data collection</u>

Dr. Michael Wilson and I trained two Tanzanian field assistants, Nasibu Zuberi Madumbi and Hashim Issa Salala, to conduct focal follows and record chimpanzee vocalizations at Gombe. They used a Sennheiser ME66 shotgun microphone with K6 power module and a Marantz PMD661 MKII audio recorder. They recorded the vocalizations with a 96 kHz sampling frequency and a 16-bit amplitude resolution. They conducted focal follows of individual males with the goal of recording as many calls as possible from the focal male, throughout the day. In addition to recording calls from the focal target, they also opportunistically recorded as many other calls as possible from known individuals to obtain the maximum number of calls. For each recording, they noted additional information including caller behavior, context, location, and party composition. Here, the recordings were obtained in traveling (caller travelling), feeding (caller feeding or arriving at a feeding site), displaying (caller displaying) and resting (caller resting —not travelling, feeding, or displaying) contexts. If pant-hoots provide any information about food, an individual could produce them when they see food and also when consuming food. Hence, a pant-hoot given when arriving at a patch with visible food was considered feeding context. Furthermore, in situations where multiple contexts overlapped, we included the highest priority context based on the following hierarchy (travel > feed > display > rest). To ensure sufficient sample sizes and consistency with recordings from Kanyawara, I limited analysis for context differences to calls recorded in traveling and feeding contexts and only included individuals with at least 3 calls recorded in both contexts. While the field assistants recorded all call types from both males and females, here I focus on pant-hoots from males, because (1) pant-hoots have been the focus of previous dialect studies; (2) they can be heard from far away, making them plausible signals of community membership, and (3) males produce pant-hoots more often than females (Wilson et al., 2007).

From July 2016 to December 2017 the team recorded a total of N = 1252 calls (N = 884 from Kasekela and N = 368 from Mitumba). I reviewed these recordings and found that N = 723 (N = 481 from Kasekela and N = 242 from Mitumba) were of sufficiently high quality for acoustic analyses. These recordings consisted of a variety of calls including pant-hoots, pant-grunts, rough-grunts, waa-barks, and screams. Of the pant-hoots in these recordings, some were choruses (where multiple individuals pant-hoot together), and not all were from identified individuals. Choruses that had overlapping elements from multiple callers were excluded, as such overlap makes it harder to extract meaningful acoustic features from known individual callers. Further, to optimize both the number of recordings per individual and the total number of individuals included in the analyses, I excluded individuals that had fewer than 8 pant-hoot call recordings. Based on this criterion, I excluded two individuals from the Kasekela community: Ferdinand (FE) and Gimli (GIM). While high-ranking males usually call most frequently (Wilson et al., 2007), the highest-ranking male at the start of this study, FE, was overthrown in October

2016, after which we were unable to record any more pant-hoots from him. In Mitumba, in July 2017, the alpha male Edgar (EDG) killed one of the adult males Fansi (FAN) (Massaro et al., 2021). Prior to this, we were able to record enough calls from FAN for some analyses. These selection criteria yielded a total of 214 pant-hoots (N = 128 from Kasekela and N = 86 from Mitumba) from 11 individuals (N = 6 males from Kasekela and N = 5 males from Mitumba) for acoustic analysis (Table 11).

At Kanyawara, Dr. Pawel Fedurek recorded chimpanzee calls using a Sennheiser ME67 shotgun microphone and a Marantz Professional PMD661 solid-state recorder. He recorded with a 44.1 kHz sampling frequency and a 16-bit amplitude resolution. He obtained the recordings during continuous sampling of focal individuals (October 2010 – September 2011). In addition to recording all calls from the focal individual, he recorded any other vocal interactions between the focal and other individuals in the focal party. For each recording, he noted the identity of the caller who started a vocal bout, the identities of any other callers in a vocal bout, and the context of the vocalizations. He obtained the recordings in traveling (caller travelling) and feeding (caller feeding or arriving at a feeding site with visible food). The aforementioned selection criteria yielded 111 calls from 7 Kanyawara males for acoustic analysis (Table 11).

2.3.3 Potential sampling biases

I evaluate the sources of bias using the STRANGE framework (Webster & Rutz, 2020). STRANGE stands for Social background; Trappability and self-selection; Rearing history; Acclimation and habituation; Natural changes in responsiveness; Genetic make-up; and Experience. In terms of social background and self-selection, we know that high-ranking males call more frequently (Wilson et al., 2007), so they are more likely to be sampled (Table 11). I attempted to avoid overcontribution from any particular individual in the statistical analyses by performing multiple permutations on balanced and randomized subsets of the data (see Statistical analysis), but some bias towards individuals that call more might have been introduced due to needing a minimum number of recordings from each individual (see Data collection). Furthermore, the chimpanzee community sizes included in this study (Kasekela ~50 individuals, Mitumba ~30 individuals (Wilson et al., 2020), and Kanyawara ~54 individuals) are close to the mean community size of 44 individuals observed in long-term studies of wild chimpanzees (Wilson et al., 2014). In terms of rearing history, acclimation, and habituation, the chimpanzees at both Gombe and Kanyawara are wild and were well habituated for many years at the time of recording. Additionally, they were not subjected to any invasive testing as these are both observational studies, thus mitigating any potential biases from acclimation, habituation, and experience. Natural changes in responsiveness due to seasons or timing could be sources of bias as chimpanzees are more likely to call in the mornings (Wilson et al., 2007) and can vary with season depending on fruit availability (personal observation). While the field assistants followed the chimpanzees throughout the day and in all seasons, the sample is likely to contain more recordings from the mornings and from the wet season. Lastly, 3 out of 6 individuals at Kasekela were close kin (two brothers: FU and FND and their father: SL, Table 11) and none of the other individuals at any of the communities included were known to be close kin. If calls of genetically related chimpanzees are more similar, then calls of Kasekela individuals might appear different from other communities due to genetic similarity.

Table 11: Number of pant-hoots by each individual in the two contiguous Kasekela and Mitumba communities at Gombe National Park, Tanzania and one geographically distant Kanyawara community at Kibale National Park, Uganda included in this study.

National Park	Community	Individual	Age at beginning (years)	Total pant-hoots	Pant-hoots with climax screams	Pant-hoots with build ups	Pant-hoots per context
Gombe	Kasekela (N = 128)	Fundi (FND)	16	11	11	6	Feed: 1 Travel: 3
		Faustino (FO)	26	20	19	12	Feed: 7 Travel: 5
		Fudge (FU)	19	33	33	27	Feed: 8 Travel: 6
		Sheldon (SL)	33	15	12	14	Feed: 1 Travel: 6
		Sampson (SN)	19	38	35	34	Feed: 22 Travel: 5
		Zeus (ZS)	22	11	8	7	Feed: 1 Travel: 9
	Mitumba (N = 86)	Edgar (EDG)	27	45	41	24	Feed: 27 Travel: 9
		Fansi (FAN)	14	8	8	3	Feed: 2 Travel: 1
		Kocha (KOC)	15	16	16	12	Feed: 4 Travel: 6
		Lamba (LAM)	14	9	9	5	Feed: 3 Travel: 3
		Londo (LON)	15	8	8	6	Feed: 4 Travel: 0
Kibale	Kanyawara (N = 111)	Big Brown (BB)	44	8	8	6	Feed: 7 Travel: 1
		Eslom (ES)	15	18	18	10	Feed: 9 Travel: 9
		Kakama (KK)	25	21	21	20	Feed: 9 Travel: 12
		Makokou	28	14	14	14	Feed: 7

		(LK)					Travel: 7
		Twig (PG)	22	10	10	6	Feed: 3 Travel: 7
		Stout (ST)	55	14	14	11	Feed: 7 Travel: 7
		Lanjo (TJ)	15	26	26	9	Feed: 23 Travel: 3

2.3.4 The pant-hoot call

The pant-hoot is a complex call composed of multiple elements. Researchers typically divide pant-hoots into four phases, each of which consists of one or more acoustically similar elements: (i) the introduction — inhaled and exhaled tonal elements (fundamental frequency F0: 300-600 Hz), (ii) the build-up — shorter but more frequent exhaled tonal elements and noisy inhaled elements (F0: 200-500 Hz), (iii) the climax — loud tonal screams (F0: 800-2000 Hz) but often including other elements such as hoos and barks, (iv) the letdown — short, build-up-like exhaled elements, decreasing in F0 (Figure 8). Chimpanzees do not always produce all four of these phases when giving pant-hoot calls. Sometimes during the call, chimpanzees hit tree buttresses with their feet (and rarely with their hands), producing drum-like sounds (Arcadi & Wallauer, 2013).

Distinguishing these pant-hoot phases can be difficult as the elements vary substantially in their acoustic structure within each phase. To address this ambiguity and to distinguish systematically among these phases, I proceeded as follows. I identified the exhaled elements in all phases as the elements that reached relatively higher maximum frequencies compared to elements preceding and succeeding them. To distinguish between the introduction and the build-up phase, I defined the start of the build-up as the first exhaled element of markedly shorter duration compared to the previous elements. The build-up consisted of a series of elements with a similarly short duration. Next, to distinguish between the build-up and the climax, I defined the start of the climax as the first exhaled element with a fundamental frequency greater than 500 Hz (see ‘500 Hz’ rule Mitani et al., 1999). Next, to distinguish between the climax and the letdown, I defined the end of the climax as the last tonal scream element. In cases where the climax phase did not include screams, I marked the end of the climax as the first element of a reduced fundamental frequency. The letdown phase consisted of a series of these elements of a lower fundamental frequency. Since the different elements of the pant-hoots could often be difficult to distinguish and are subject to observer bias, I only describe the climax in terms of scream and non-scream elements that I found relatively easy to distinguish.

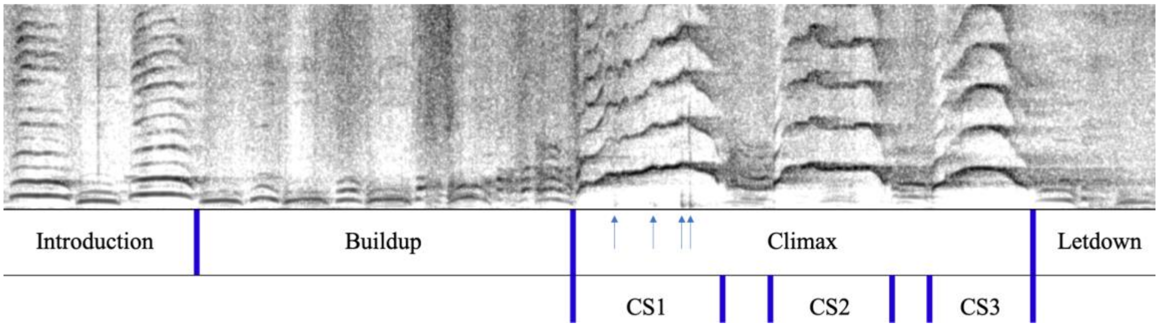


Figure 8: Spectrogram of a typical pant-hoot with the four phases (introduction, buildup, climax, and letdown) labeled. The three climax-screams are labeled as CS1, CS2, and CS3, and the arrows represent instances of buttress drumming.

2.3.5 Acoustic feature extraction

Given the structure of the pant-hoots described above, acoustic analysis of pant-hoots could be performed by measuring acoustic features in different ways. I extracted two main categories of acoustic features from the spectrogram representations of the calls: structural features and spectral features. Structural features describe the composition of the elements in different phases of the pant-hoots and their temporal patterning. For these, I selected 25 acoustic features similar to those used in previous studies of chimpanzee dialects (Crockford et al., 2004; Mitani et al., 1992, 1999) (Table 12). Spectral features quantify the frequency and tonal structure of individual elements from the power spectrum. I measured these from selected specific elements: one build-up element (24 features), and one climax element (25 features). I used semi-automatic measurements of acoustic features—a process involving manually chosen call elements used in automatic feature extraction—using Avisoft-SASLab Pro v. 5.2 (Specht, 2004) and LMA (Fischer et al., 2013; Schrader & Hammerschmidt, 1997) (Table 13).

I extracted the acoustic features as follows. First, I measured structural features from the pant-hoot phases by visually inspecting spectrograms of entire pant-hoots using Praat version 6.1.15. I considered each phase separately and measured a set of acoustic features from each phase (Table 12). I present the visual summaries of these structural features of the pant-hoots from the three communities in the supplementary materials (Appendix; Figure 29-Figure 41). Next, for the semi-automatic extraction of acoustic features, I chose one element from the build-up phase and one element from the climax phase. From the build-up phase, I chose the middle element in case of an odd number of build-up elements, and the element immediately preceding the middle of the build-up in case of an even number of elements (Mitani et al., 1999). From the climax phase, I chose the scream that reached the highest fundamental frequency in the spectrogram. To obtain appropriate frequency and time resolutions, I down-sampled the sampling frequency to 24 kHz using Avisoft-SASLab Pro, resulting in a frequency range of 12 kHz. Next, using Avisoft-SASLab Pro, I created spectrograms with an FFT length of 1024 points, frame size of 100%, and Hamming window with an overlap of 93.75%. This resulted in a frequency resolution of 23 Hz, and a time resolution of 2.7 ms, which is sufficient to reveal tonal properties and extract acoustic features from build-up and climax elements. I then imported the spectrograms in LMA and extracted acoustic features (listed in Table 13) using the harmonic cursor tool. I did not extract additional acoustic features from elements in the introduction and the letdown phases as the introduction was not always recorded fully, and the letdown exhibited high variability in the type of elements, making comparison among letdowns difficult. I attempted to be consistent with previous studies by including as many acoustic features used in previous studies as possible (Table 12 and Table 13). However, some acoustic features used in previous studies could not be measured using the software packages available to us. Nevertheless, I consider that the acoustic features I used should encompass the relevant range of variation in chimpanzee pant-hoots, without loss of generality. To further facilitate comparisons with previous studies, I report means and standard deviations of acoustic features found to have community specific differences in previous studies in the appendix (Table 34).

*Table 12: Structural acoustic features manually measured using Praat v 6.1.15 that were used in this study. I also indicate which features were used in other studies of chimpanzee dialects. Categorical variables are marked with *. Only numeric variables were used in the multivariate analyses including the PCAs and pDFAs since those techniques do not handle categorical variables.*

***Drumming related variables were not included in the multivariate analysis due to small sample sizes. However, descriptive plots are included in the appendix.*

Structural acoustic features used in this study	***Part of the pant-hoot***	***Crockford et al. 2004***	***Mitani et al. 1999***	***Mitani et al. 1992***
Duration of the call (from build-up to letdown phases) (s)	Entire call	No	No	No
Presence of introduction phase*	Introduction	Yes	No	No
Presence of build-up phase*	Build-up	Yes	No	No
Number of build-up exhalation elements	Build-up	Yes	No	No
Number of build-up elements in the first half of the build-up	Build-up	Yes	No	No
Number of build-up elements in the second half of the build-up	Build-up	Yes	No	No
Duration of build-up phase (s)	Build-up	Yes	Yes	No
Rate of build-up phase (elements/s)	Build-up	Yes	Yes	Yes
Rate of first half of build-up phase (elements/s)	Build-up	Yes	No	No
Rate of second half of build-up phase (elements/s)	Build-up	Yes	No	No
Build-up acceleration (rate of second half – rate of first half)	Build-up	Yes	No	No
Presence of climax phase*	Climax	Yes	No	No
Total number of climax elements (including screams and non-scream elements)	Climax	Yes	No	No
Number of screams in climax	Climax	Yes	No	No
Proportion of climax elements that are screams	Climax	Yes	No	No

Duration of climax phase (s)	Climax	No	No	No
Presence of letdown phase*	Letdown	Yes	No	No
Number of elements in letdown phase	Letdown	Yes	No	No
Structural acoustic feature(s) NOT used in this study				
Number of introduction elements	Introduction	Yes	No	No
Duration of introduction element	Introduction	No	Yes	No
Drumming related features**	Drumming	Yes	No	No

Table 13: Semi-automatically measured acoustic features using LMA from the selected build-up and climax elements compared with other studies. Some acoustic features were not used in this study as they were not measured by the version of LMA available to me.

Acoustic feature(s) used in this study	***Crockford et al. 2004***	***Mitani et al. 1999***	***Mitani et al. 1992***
Duration of the element (ms)	No (for build-up); Yes (for climax)	Yes	Yes
Start, end, maximum, minimum, and mean fundamental frequency F0 (Hz)	No (for build-up); Yes (for climax)	No (start and end); Yes (maximum, minimum, and mean)	No
Frequency range of F0 (Maximum F0 – Minimum F0) (Hz)	No	Yes	Yes
Tonality measures: mean and maximum frequency difference between the original F0 curve and the floating average curve (Hz)	No (for build-up); Yes (for climax)	No	No
Location of maximum F0 relative to the duration ([1/duration]*location)	No (for build-up); Yes (for climax)	No	No

Factor of linear trend of F0 (measures if the F0 is rising, falling, or flat on average)	No (for build-up); Yes (for climax)	No	No
Mean and maximum deviations between F0 and linear trend line (Hz)	No (for build-up); Yes (for climax)	No	No
Start, end, maximum, minimum, and mean peak frequencies (Hz)	No (for build-up); Yes (for climax)	No	No
Peak frequencies with maximum and minimum amplitude (Hz)	No (for build-up); Yes (for climax)	No	No
Locations of maximum and minimum peak frequencies relative to the duration ([1/duration]*location)	No	No	No
Maximum difference between peak frequency values in successive time segments (Hz)	No (for build-up); Yes (for climax)	No	No
Mean and maximum wiener entropy coefficient (0-1; 1=noise)	No	No	No
Acoustic feature(s) NOT used in this study			
Slope of F0 from start to maximum (Hz/ms)	Yes	No	No
Slope of peak frequency from start to maximum (Hz/ms)	Yes	No	No
Maximum F0 start F0 (Hz) and Maximum F0 minimum F0 (Hz)	Yes	No	No
F0 at midpoint of introduction element (Hz)	Yes	Yes	No
Peak frequency at midpoint of inhaled elements (Hz)	Yes	No	No
Peak frequency at midpoint of exhaled elements (Hz)	Yes	No	No
Peak frequency of inhaled – peak frequency of exhaled elements (Hz)	Yes	No	No
Ratio of F1/F2, the first and second formant frequencies	No	Yes	No
Bandwidth (Hz)	No	Yes	No

<u>2.3.6 Statistical analysis</u>

To confidently make conclusions about whether chimpanzees have community-specific dialects that are an outcome of vocal learning, we need to control for confounding factors, which I examine using Directed Acyclic Graphs (DAGs) (McElreath, 2018). DAGs portray our assumptions of causal relationships among variables. Based on these biologically informed assumptions, DAGs allow us to (i) identify confounding causal paths that may cause spurious statistical associations between variables, (ii) identify causal paths that may mask real causal relationships. Thus, DAGs allow us to make biologically informed decisions about which confounding variables to control in our statistical analyses.

Figure 9 portrays my assumed causal relationships as follows: the geographical location of a community can affect the environmental conditions (because environmental features such as forest structure may cause locations to vary in habitat acoustics), community identity (because communities are in part defined based on geographical proximity), and genetics of the chimpanzees (because geographically closer chimpanzees are more likely to be genetically related). Genetic similarity and environmental conditions such as habitat acoustics and sound environment can in turn affect the acoustic structure of vocalizations (Mitani et al., 1999). Furthermore, genetics affects individual identity, and individual identity may affect both acoustic structure and community identity, because communities are defined based on a group of individuals that live within the same territory. Lastly, context may affect acoustic structure (Fedurek, Zuberbühler, et al., 2016; Notman & Rendall, 2005; Uhlenbroek, 1996). I used this DAG, to identify minimally sufficient adjustment sets for assessing the relationship of interest, i.e., community identity and acoustic structure. A minimally sufficient adjustment set of variables is a list of variables that are sufficient to control for estimating, in an unbiased way, the statistical association of two variables in a DAG. The adjustmentSets function in the R-package dagitty prints a list of all minimally sufficient adjustment sets (Textor et al., 2016). This package identified the sets {Individual Identity, Geographical Location} and {Environment, Genetics, Individual Identity} as the minimally sufficient adjustment sets to assess the relationship between community identity and acoustic structure. Hence, I need to either control for individual identity and geographical location, or environment, genetics, and individual identity. Since I did not measure environmental variables or genetics, I could not control for those. However, I could obtain an unbiased association between community identity and acoustic structure by controlling for individual identity and geographical location. I therefore controlled for geographical location by testing for differences between calls from neighboring communities and compared them with the geographically distant Kanyawara community. Next, I controlled for individual identity using the permuted Discriminant Functions Analysis (pDFA) procedure. The pDFA procedure is used to test for differences in a factor of interest (a.k.a. test factor) while controlling for a confounding factor (a.k.a. control factor) (Mundry & Sommer, 2007). I needed to control for individual identity not only to close confounding “backdoor” pathways, but also to account for the non-independence of data points due to there being multiple recordings from the same individual. I describe the pDFA procedure in more detail in the next section. Lastly, while context is not a confound opening any “backdoor” paths based on my DAG, context is a precision covariate that may affect the relationship of interest (community ID to acoustic structure) (Laubach et al., 2021). Controlling for a precision covariate could improve the precision of my model estimates and prevent any masking of the relationship of interest (*ibid.*). Hence, I control for context for each set of acoustic features for which there exist differences between contexts.

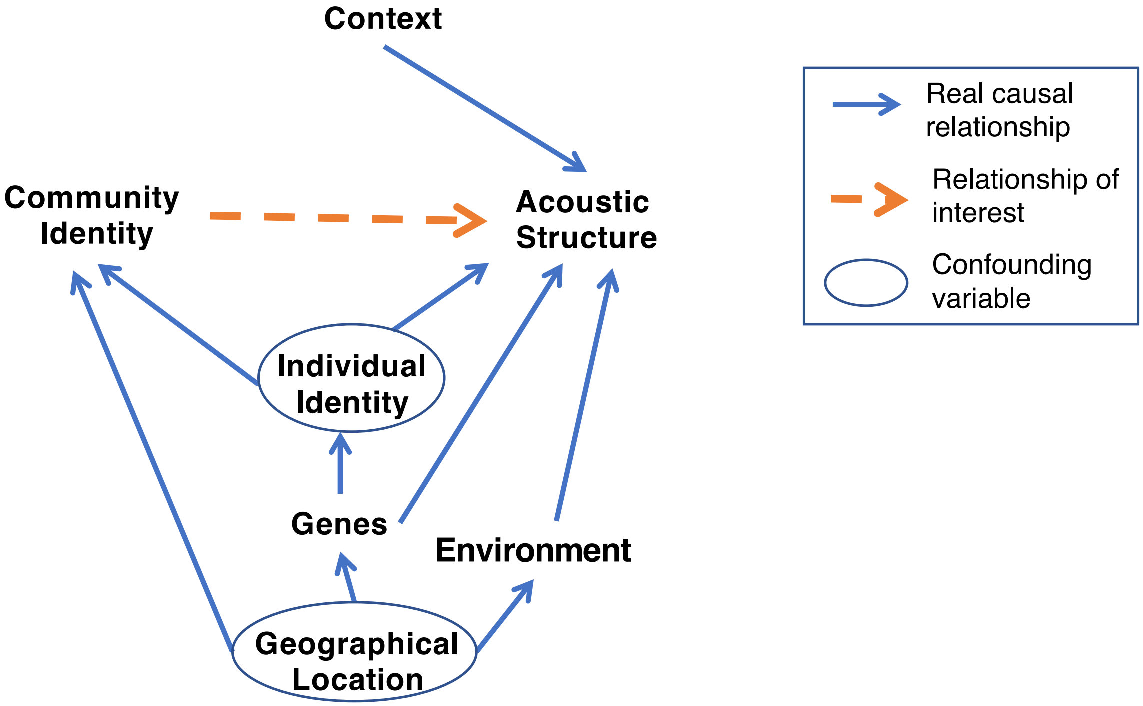


Figure 9: Directed Acyclic Graph (DAG) of the assumed causal relationships among relevant variables.

2.3.7 Description of the pDFA procedure

The pDFA procedure improves upon traditional Discriminant Functions Analysis (DFA) technique by allowing two-factor designs. The traditional DFA technique handles only one factor at a time and is known to inflate group differences in two-factorial designs with a confounding factor (individual identity is among the most common confounding factors in similar studies (Mundry & Sommer, 2007). The pDFA procedure allows two-factor designs by performing a permutation test on the classification accuracy of DFA. The permutation test tests if the observed classification accuracy of DFA is significantly higher than expected by chance while accounting for the accuracy inflating effect of a confounding factor. The procedure works as follows. Firstly, the procedure samples a specified number of balanced and randomized datasets from the original dataset. It randomizes the labels of the test factor based on the combinations of different categories of test and control factors. It performs the balancing such that there is the same number of observations from each category of the test factor. Next, it performs a traditional DFA on each of these randomized datasets and obtains a distribution of classification accuracies of these DFAs. The observations left out due to balancing are used for cross-validation to obtain the out-of-sample, cross-validated classification accuracies. The distribution of classification accuracies of randomized datasets describes the probabilities of obtaining particular classification accuracies using a traditional DFA just based on chance. The expected value from this distribution is compared with the classification accuracy obtained on the original dataset (observed classification accuracy) to obtain a p-value for the permutation test. In other words, this distribution provides an estimate of the inflation in classification accuracy of the test factor caused by the confounding effect of the control factor.

2.3.8 Analysis steps

I performed the pDFAs with 1000 permutations (i.e., 1000 randomized datasets including the original dataset) in each of my analyses and used an alpha level of $\boldsymbol{\alpha} = 0.05$ on the cross-validated classification accuracy to infer a significant difference. The test factors of interest in this study were context, community identity, and individual identity. The control factors were context and individual identity, depending on the analysis. For different test factors in the pDFA, I had to consider the data designs to ensure proper randomization and balancing of the permuted datasets. In this study, two design situations occurred: crossed and nested. A crossed design occurs when all the categories of the test factor are recorded in all categories of the control factor. So, a pDFA with crossed design could only be used in testing for differences in context by only including individuals recorded in both contexts. While testing for differences in community identity and individual identity, a nested design occurs. In a nested design, the categories of the control factor are nested within categories of the test factor, or the categories of the test factor are nested within some other factor known as the restriction factor. Since individual identity is nested within community identity, a nested design occurs when testing for differences in the community or individual identities.

I performed all statistical analyses in R version 4.0.2 using Rstudio version 1.3.1093. The pDFAs were carried out using a set to R functions provided by R. Mundry. These functions implement the pDFA procedure and are built on top of the lda function in the R package MASS (Ripley et al., 2020) that is used to perform traditional DFAs. I performed four different analyses for different kinds of acoustic features: (i) structural features (Table 12), (ii) build-up element features (Table 13), (iii) climax scream features (Table 13), and (iv) all features combined (Table 12 and Table 13). I tested for differences in context, community identity, and individual identity in each of these four types of acoustic features. For each kind of acoustic feature set, I tested for context before performing other analyses to determine whether context was a precision covariate (Laubach et al., 2021) that needed to be statistically controlled for in the subsequent analyses (refer to the DAG logic in Methods). If I found statistically significant differences in the context in any acoustic feature set, I controlled for the context by stratifying the data and considering calls from only one context at a time in separate analyses (Crockford et al., 2004). If I did not find any significant effect of context on an acoustic feature set, I did not need to control for context. To control for geographical differences, I performed two separate analyses for each kind of acoustic feature set. Following Crockford et al. (2004), I first investigated the acoustic structure of pant hoots from the two neighboring communities of Gombe, where maximal differences were expected. In order to then compare the two Gombe communities to a geographically distant community, I ran pDFAs including all three communities. To control for individual differences, I used individual identity as the control factor in each of the pDFAs when testing for context and community identity. When testing for individual differences using individual identity as the test factor, I used community identity as the restriction factor. When a restriction factor is added in the pDFA, the randomization process is done while accounting for the fact that the test factor is nested within the restriction factor. To avoid overfitting, I ensured that only as many, or fewer acoustic features are used to perform the DFAs as there are observations in the category of the test factor with the fewest observations. To avoid multicollinearity and to reduce the number of acoustic features while accounting for most variation contained in different acoustic features, I used Principal Components Analysis (PCA) on the acoustic features. I used the scores of each observation on the principal components as the features to be used in the DFAs. To choose the number of principal components to include, I used two heuristics. First, I used as many principal components as there were number of observations in that category of the test factor with the fewest observations or as many principal components that explained 90% of the variation, whichever was smaller. Limiting the number of principal components to those that explained 90% of the variation allowed me to avoid including too many components of little explanatory power when including many more components was possible in the pDFA design. Second, since no heuristic is perfect in all circumstances (Jolliffe, 2002), I used an additional heuristic to ensure the stability of results. I verified the consistency of the results of the pDFAs over different numbers of principal components selected using Cattell’s scree test (Cattell, 1966). Using this heuristic, I chose the number of principal components by identifying the “elbows” in the scree plot of variances explained against the number of the principal component.

For each acoustic feature set used in the analyses, I chose a subset of recordings based on the following criteria. For structural features, I first removed acoustic features that had too many missing values for sufficient statistical power. These mainly included acoustic features related to drumming, as only 18% of the recorded pant-hoots had drumming (Table 12; appendix: Figure 40 and Figure 41). After that, I removed categorical features that indicated the presence or absence of the four phases as categorical features are not handled by DFAs. Next, I removed the cases that had missing values in any of the remaining 14 acoustic features. While the categorical features were eliminated, the information contained in them was included in other features that indicated the number of elements in each phase. A value of 0 in those features would indicate absence of a phase, whereas a non-zero value would indicate the presence. For the build-up feature set, I only included pant-hoots that included the build-up phase. Similarly, for the climax feature set, I only included pant-hoots that included the climax phase. And lastly, while including all the acoustic features simultaneously (structural, build-up, and climax features), I only included pant-hoots that had both build-up and climax phases.

The pDFA is the omnibus test that warrants further post hoc tests in case it revealed significant differences. To perform post-hoc tests, I used the *repDFA* function written by C. Neumann (Berthet et al., 2017; Neumann, 2020) followed by Generalized Linear Mixed Models (GLMMs). *repDFA* function allows us to identify the key variables that discriminate the test factor. This function creates 1000 balanced datasets in crossed designs, re-runs 1000 DFAs and records the variable that had the highest coefficient on the first and second linear discriminant functions in each of those DFAs. Variables that have the highest coefficient in many of those permutations are arguably the most important in discriminating the test factor. For nested designs, I modified Neumann’s function and wrote a new function called *repDFA_nested* (available on GitHub). This function modified *repDFA* such that it created balanced datasets by randomly sampling with replacement the same number of recordings for each individual in the analysis. Next, I tested the significance of the individual variables identified with *repDFA* and *repDFA_nested* using GLMMs. I controlled for individual identity in the GLMMs by including individual IDs as random intercepts and adjusted the p-values for multiple comparisons using Benjamini–Hochberg adjustment method.

2.3.9 Data sharing statement

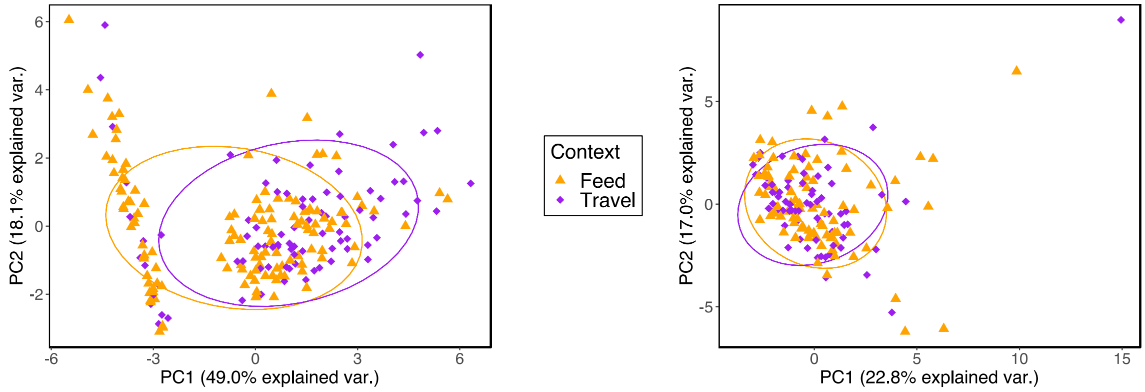
The R code and data for the analyses are available from GitHub at https://github.com/desai-nisarg/Gombe-dialects. Audio recordings from Gombe are available from Dr. Michael Wilson and from Kanyawara available from Dr. Pawel Fedurek at reasonable request.

2.4 Results

2.4.1 Differences in pant-hoots between contexts

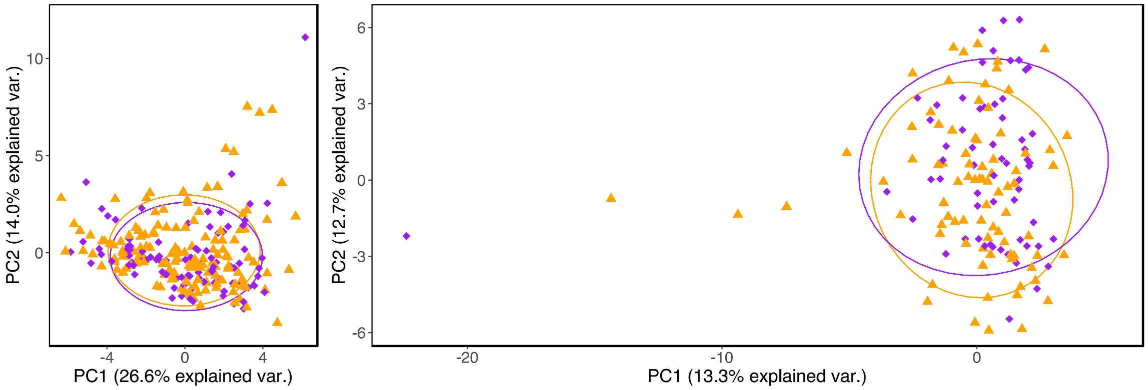
To ascertain if I needed to control for context in our main analyses regarding acoustic differences in pant hoot structure as a function of community, I started by examining if context affected any of our acoustic feature sets. I found a statistically significant difference in the structural acoustic features (Table 12) between feeding and traveling contexts after controlling for individual identity (pDFA with structural features, observed classification accuracy: 63.9 % vs. expected by chance: 51.7 %; p = 0.044, Table 14). Using the repDFA function, I identified principal component 6 (PC6) to be the best discriminator of the contexts in 716 out of the 1000 DFAs followed by principal component 1 (PC1) in 191 out of the 1000 DFAs. PC6 loaded most heavily on the number of letdown elements and build-up acceleration had the second highest loading. PC1 loaded most heavily on the number of build-up elements but other features of the build-up had comparable loadings. These features were: duration of the build-up phase, rate of the build-up phase, number of elements in the first and second half of the build-up, and rate of the first and the second half of the build-up. Further, the principal components plot made by performing principal components analysis on the structural features shows a distinct band of calls given mostly in feeding contexts (Figure 10 (a)). Principal component 2 (PC2) explained the most variance in this band and it loaded most heavily on the number of climax elements. These were pant-hoots with a greater than average number of climax elements, and which did not have a build-up phase. I performed significance tests for the highest loading features of these three important components using Poisson GLMMs and controlled for individual identity by including it as a random effect in the models. All three of these acoustic features were statistically significantly different between contexts. Travel pant-hoots had a (i) greater number of letdown elements, β (Travel) = 0.57, Benjamini–Hochberg adjusted p-value = 3.8e-07; (ii) greater number of build-up exhalation elements, β (Travel) = 0.34, Benjamini–Hochberg adjusted p-value = 4.8e-06; and (iii) lower number of climax elements, β (Travel) = -0.27, Benjamini–Hochberg adjusted p-value = 1.8e-03, compared to feeding pant-hoots (Table 15; appendix: Figure 24 to 26). Hence, I conclude that the number of elements in the build-up, the climax, and the letdown phases potentially encoded contextual information.

Further, I found no differences in the contexts in other types of acoustic features after controlling for individual identity among the contiguous communities or all communities taken together. Cross-validated p-values for pDFA performed on all communities with (i) build-ups: p = 0.38, (ii) climax screams: p = 0.34, and (iii) all acoustic features simultaneously: p = 0.37. Cross-validated p-values for pDFA performed on communities within Gombe with (i) structural features: p = 0.29, (ii) build-ups: p = 0.38, and (iii) climax screams: p = 0.39 (Table 14). Figure 10 (b-d) show the overlap between contexts in these acoustic features in a multidimensional space. Given that context is confounding only when it has a significant effect (refer to DAG logic in Methods), I controlled for context when testing for differences in structural features alone and not when testing for differences in other types of acoustic features in the subsequent analyses.



(a) Structural features

(b) Build-up features



(c) Climax scream features

(d) All acoustic features

Figure 10: Principal components plots with the 68% normal data ellipses containing 68% of the data points included for each context. (a) Principal Components Analysis performed on structural features. Pant-hoots given in different context separate over PC1. (b) Principal Components Analysis performed on acoustic features of the selected build-up element. (c) Principal Components Analysis performed on acoustic features of the selected climax element. (d) principal Components Analysis performed on all acoustic features simultaneously from all three communities. (b), (c), and (d) reveal strong overlap between contexts.

Table 14: Summary of the results from the pDFAs with context as the test factor and individual identity as the control factor for different types of acoustic features. I indicate the number of individuals recorded in both feeding and traveling contexts, the range of

number of calls per individual and the total number of calls considered for each of the analyses.

Acoustic features used	***Control factor***	***Number of individuals included in both contexts***	***Median number of calls per individual in each context*** ***(Range)***	***Total number of calls used***	***Observed cross-validated classification accuracy*** ***(Expected value)***	***P-value for cross-validated classification accuracy***
Communities from Gombe						
Structural (Table 2)	Individual	5	Feed: 7 (3-25) Travel: 4 (3-9)	82	63 (55.9)	0.291
Build-up (Table 3)	Individual	5	Feed: 6 (3-20) Travel: 4 (3-7)	66	52.8 (50.9)	0.382
Climax (Table 3)	Individual	6	Feed: 7.5 (3-26) Travel: 5.5 (3-7)	100	50.1 (49.4)	0.394
Entire call (Table 2 and 3)	Not performed due to low sample sizes of individuals recorded in both contexts					
All communities						
Structural (Table 2)	Individual	11	Feed: 7 (3-25)	183	63.9 (51.7)	0.044*

			Travel: 6 (3-12)			
Build-up (Table 3)	Individual	9	Feed: 6 (3-20) Travel: 6 (3-12)	121	51.9 (49.6)	0.376
Climax (Table 3)	Individual	12	Feed: 7.5 (3-26) Travel: 6.5 (3-12)	203	52.2 (50.2)	0.34
Entire call (Table 2 and 3)	Individual	6	Feed: 6 (3-9) Travel: 6.5 (3-12)	77	50.8 (47.7)	0.37

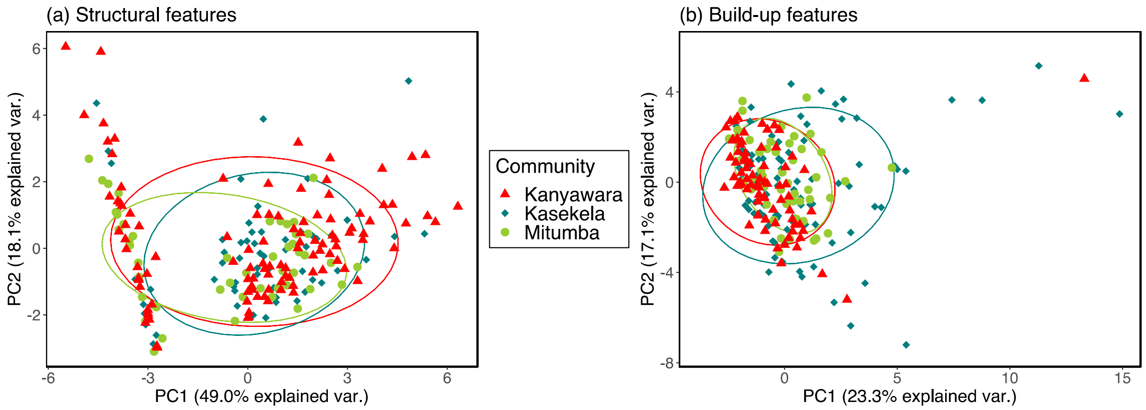
Table 15: Structural acoustic features showing differences between contexts.

Acoustic variable	Context	$\bar{x} \pm SD$	β (Travel)	p-value
Number of letdown elements	Feed	1.26 ± 1.36	0.57	3.8e-07
	Travel	2.34 ± 2.59		
Number of buildup exhalation elements	Feed	3.21 ± 3.2	0.34	4.8e-06
	Travel	5.31 ± 4.13		
Number of climax elements	Feed	3.18 ± 2.34	-0.27	1.8e-03
	Travel	2.63 ± 2.34		

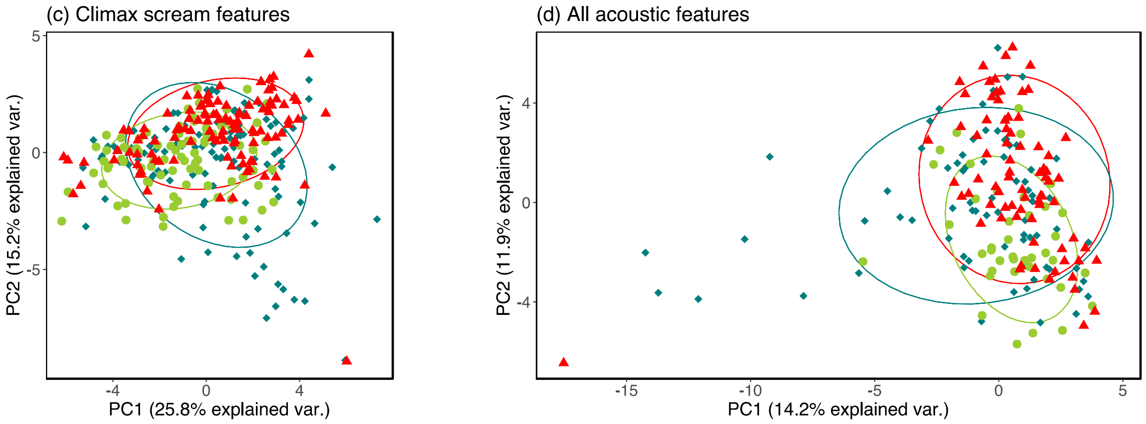
2.4.2 Differences in pant-hoots among communities of chimpanzees

Figure 11 (a-d) show the clusters of the three communities in multidimensional spaces of the structural features, build-ups, climaxes, and all features taken simultaneously. These show strong overlap among communities, suggesting a lack of community-level differences that is confirmed by the pDFAs, with one exception: a statistically significant difference in acoustic features of the climax scream among the communities when I included the geographically distant Kanyawara community in the analysis (pDFA on climaxes of all communities, observed classification accuracy: 54 % vs. expected: 40.8 %; p = 0.016, Table 16). These features did not differ statistically between the two neighboring communities at Gombe, but the relatively low p-value indicates these features may warrant further investigation (observed classification accuracy: 70.7 % vs. expected: 58 %; p = 0.089, Table 16). Furthermore, these features did not differ between the pair Kasekela-Kanyawara (observed classification accuracy: 88 % vs. expected: 77.1 %; p = 0.18) or between the pair Mitumba-Kanyawara (observed classification accuracy: 68.1 % vs. expected: 57.2 %; p = 0.13). Hence, I did not use the *repDFA_nested* function to test which acoustic features were important in the 3-community analysis. Considering that the DFA could be sensitive to outliers (Mundry & Sommer, 2007), I checked for the consistency of the results after removing outliers. The patterns remain similar after the removal of outliers (See appendix for details; Figure 27 and 28 (a-b)).

Additionally, I observed no differences between the contiguous communities, or all communities taken together in the structural features (controlled for individual identity and context), acoustic features of the build-ups, or all acoustic features considered together (Table 16).


(a) Structural features

(b) Build-up features


(c) Climax scream features

(d) All acoustic features

Figure 11: Principal Components plots with the 68% normal data ellipses containing 68% of the data points included for each community. (a) Principal Components Analysis performed on structural features. (b) Principal Components Analysis performed on acoustic features of the selected build-up element. (c) Principal Components Analysis performed on acoustic features of the selected climax element. Kasekela and geographically distant Kanyawara communities separate to some extent over PC2. (d) Principal Components Analysis on all acoustic features simultaneously from all three communities. (a), (b), and (d) reveal strong overlap among communities.

Table 16: Summary of the results from the pDFAs with community identity as the test factor for different types of acoustic features. I indicate the control factor, the number of individuals from each community, the range of number of calls per individual and the total number of calls considered for each of the analyses. I used context as a control factor only in case of structural features since there was a difference between contexts only in structural features.

Acoustic features used	***Control factor***	***Number of individuals included per community***	***Median number of calls per individual in each community (Range)***	***Total number of calls used***	***Observed cross-validated classification accuracy (Expected value)***	***P-value for cross-validated classification accuracy***
Communities from Gombe						
Structural (Table 2)	Individual (Calls included from both contexts)	Kasekela: 6 Mitumba: 3	Kasekela: 8 (4-21) Mitumba: 7 (4-34)	103	53 (56)	0.639
Structural (Table 2)	Individual (Calls only from feed context)	Kasekela: 3 Mitumba: 3	Kasekela: 7 (5-18) Mitumba: 3 (3-25)	61	63.84 (58.77)	0.368
Structural (Table 2)	Individual (Calls only from travel context)	Kasekela: 6 Mitumba: 2	Kasekela: 4 (3-6) Mitumba: 6.5 (4-9)	39	44.41 (51.34)	0.725
Build-up (Table 3)	Individual	Kasekela: 6 Mitumba: 4	Kasekela: 13 (6-33) Mitumba: 9 (5-24)	146	58.5 (53.5)	0.255

Climax (Table 3)	Individual	Kasekela: 6 Mitumba: 5	Kasekela: 15.5 (8-34) Mitumba: 9 (8-41)	199	70.7 (58)	0.089
Entire call (Table 2 and 3)	Individual	Kasekela: 5 Mitumba: 5	Kasekela: 11 (6-28) Mitumba: 5 (3-19)	115	59.6 (54.6)	0.272
All communities						
Structural (Table 2)	Individual (Calls included from both contexts)	Kasekela: 6 Mitumba: 3 Kanyawara: 7	Kasekela: 8 (4-21) Mitumba: 7 (4-34) Kanyawara: 14 (8-26)	212	35.4 (39.7)	0.729
Structural (Table 2)	Individual (Calls only from feed context)	Kasekela: 3 Mitumba: 3 Kanyawara: 7	Kasekela: 7 (5-18) Mitumba: 3 (3-25) Kanyawara: 7 (3-23)	126	37.54 (41.37)	0.655
Structural (Table 2)	Individual (Calls only from travel context)	Kasekela: 6 Mitumba: 2 Kanyawara: 6	Kasekela: 4 (3-6) Mitumba: 6.5 (4-9) Kanyawara: 7 (3-12)	82	32.5 (36.28)	0.676
Build-up (Table 3)	Individual	Kasekela: 6 Mitumba: 4	Kasekela: 13 (6-33) Mitumba: 9 (5-24)	222	45 (37.6)	0.08

		Kanyawara: 7	Kanyawara: 10 (6-20)			
Climax (Table 3)	Individual	Kasekela: 6 Mitumba: 5 Kanyawara: 7	Kasekela: 15.5 (8-34) Mitumba: 9 (8-41) Kanyawara: 14 (8-26)	310	54.0 (40.8)	0.016*
Entire call (Table 2 and 3)	Individual	Kasekela: 5 Mitumba: 5 Kanyawara: 7	Kasekela: 11 (6-28) Mitumba: 5 (3-19) Kanyawara: 10 (6-20)	191	51.3 (42.2)	0.079

2.4.3 Differences in pant-hoots among individuals

I observed statistically significant differences among the individuals in the structural features, acoustic features of the climax screams, and all acoustic features taken simultaneously. This was true when all communities were taken together as well as when the geographically adjacent communities of Gombe were assessed separately (Table 17). However, the individuals could not be separated based on acoustic features of the selected build-up elements in any setting (pDFA on build-up features of all communities: p = 0.18, and Gombe: p = 0.15; Table 17).

Figure 12 (a-c) show the differences among individuals of the three communities in the multidimensional space all acoustic features taken simultaneously. In Kasekela, calls from the individuals FND, FU and the pair FO and SL separate over PC2. While calls from FO, SL, and SN overlap, SN could be differentiated to some extent on PC1 (Figure 12 (a)). In Mitumba, while calls from the individuals EDG and LAM overlap, they could be differentiated from KOC, LON, and FAN from a combination of PC1 and PC2 values (Figure 12 (b)). In Kanyawara, calls from the individuals BB, ES, and TJ separate from LK on PC1 and from PG and KK on PC2 (Figure 12 (c)).

For the structural features, the *repDFA_nested* function identified principal component 2 (PC2) and principal component 7 (PC7) to have the highest loadings on both discriminant function 1 (PC2 higher than PC7) and discriminant function 2 (PC7 higher than PC2). Combined they had the highest loadings on discriminant function 1 in 619 out of 1000 DFAs and on discriminant function 2 in 423 out of 1000 DFAs. Top three acoustic features with the highest loadings on PC2 were the number of climax elements, build-up to letdown duration, and the duration of climax. And on PC7, the number of climax screams, build-up to letdown duration, and the duration of climax loaded the highest. For the climax features, the *repDFA_nested* function identified principal component 1 (PC1) to have the highest loading on discriminant function 1 in 860 out of 1000 DFAs and principal component 3 (PC3) to have the highest loading on discriminant function 2 in 842 out of 1000 DFAs. The top three acoustic features with the highest loadings on PC1 were mean F0, maximum F0, and frequency range of F0. On PC3, the top three acoustic features were minimum F0, minimum peak frequency, and start frequency of F0. Lastly, when all features were taken together, PC2 loaded the highest on 945 out of 1000 DFAs. The top three acoustic features that loaded the highest on PC2 were maximum F0, mean F0, and frequency range of F0. This confirmed the findings above and also suggested that the strongest individual signal was in the acoustic features of the fundamental frequency in the climax scream. I do not report the results from the GLMMs for these acoustic features as differences among specific individuals are not of general interest. However, I observed statistically significant differences in some (but not all) pairs of individuals in each of these acoustic features suggesting that some individuals could be identified with more certainty than others. I can see this reflected in the low classification accuracies in the pDFAs (Table 17).

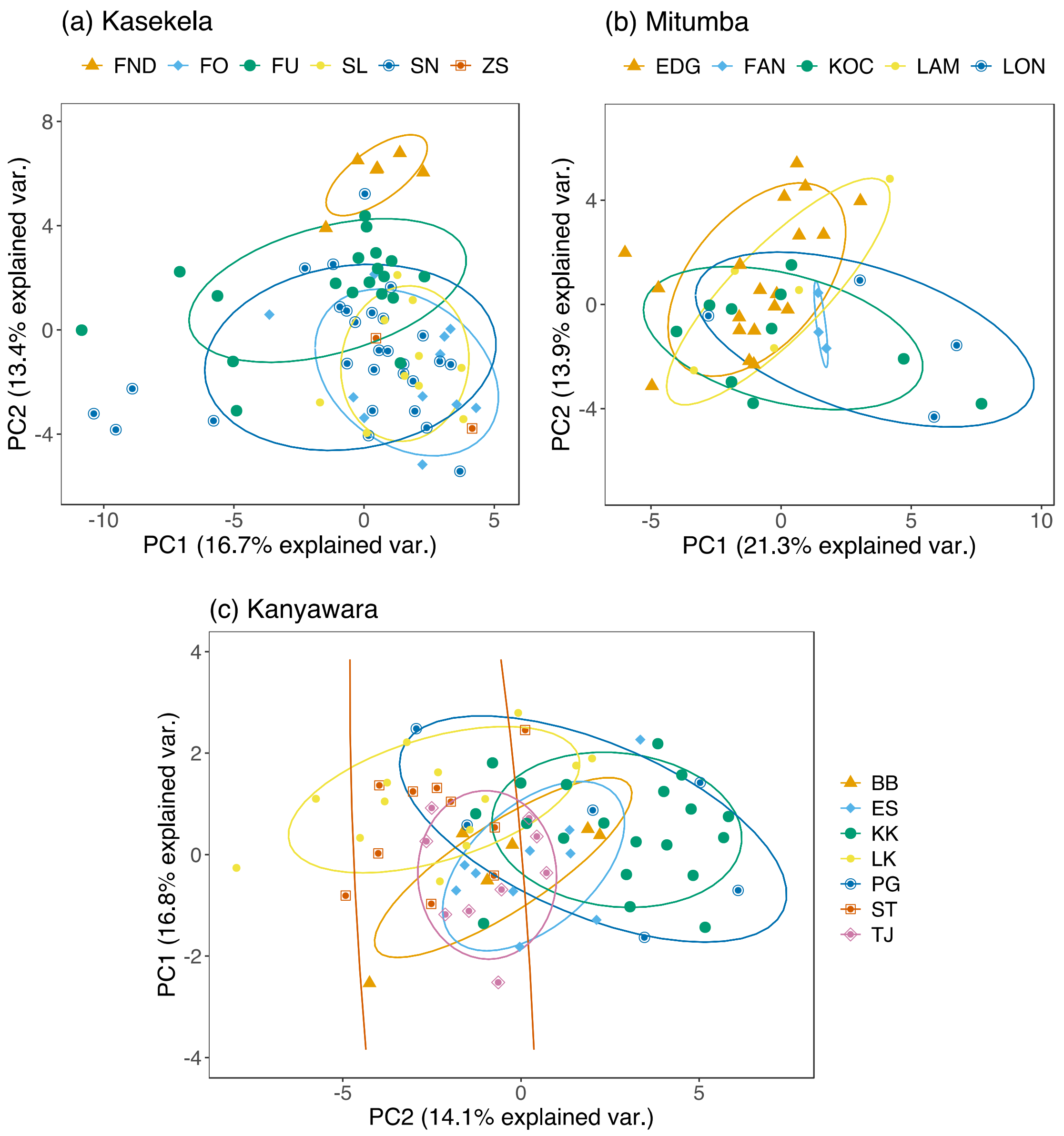


Figure 12: Principal Components plots with the 68% normal data ellipses containing 68% of the data points included for each individual. Principal Components Analysis was performed on the structural features as well as features of the selected build-up and climax elements simultaneously from the three communities. The 68% normal data ellipses revealed a lower overlap compared to community identity and context. Plot for (a) Kasekala. Some individuals formed distinct clusters over PC2. (b) Mitumba. Some individuals formed distinct clusters over a combination of PC1 and PC2. (c) Kanyawara. Some individuals formed distinct clusters over PC2 and others over PC1.

Table 17: Summary of the results from the pDFAs with individual identity as the test factor for different types of acoustic features. I used community ID as the restriction factor except when using context as a control factor. I indicate the number of individuals included, the range of number of calls per individual and the total number of calls considered for each of the analyses.

Acoustic features used	***Control or restriction factor***	***Number of individuals included***	***Median number of calls per individual (Range)***	***Total number of calls used***	***Observed cross-validated classification accuracy*** ***(Expected value)***	***P-value for cross-validated classification accuracy***
Communities from Gombe						
Structural (Table 2)	Community (restriction factor)	11	13 (4-41)	171	24 (10.3)	0.001*
Structural (Table 2)	Context (control factor)	Not performed due to low sample sizes of individuals recorded in both contexts				
Build-up (Table 3)	Community (restriction factor)	10	12 (5-33)	146	16.1 (12.1)	0.15
Climax (Table 3)	Community (restriction factor)	11	12 (8-41)	199	24.2 (13.2)	0.006*
Entire call (Table 2 and 3)	Community (restriction factor)	10	10 (3-28)	115	23.5 (13.2)	0.024*
All communities						

Structural (Table 2)	Community (restriction factor)	18	13.5 (4-41)	280	19.5 (6.9)	0.001*
Structural (Table 2)	Context (control factor)	7	14 (9-34)	119	35.8 (24.7)	0.043*
Build-up (Table 3)	Community (restriction factor)	17	11 (5-33)	222	10.6 (8.2)	0.18
Climax (Table 3)	Community (restriction factor)	18	14 (8-41)	310	20.1 (9.4)	0.001*
Entire call (Table 2 and 3)	Community (restriction factor)	17	10 (3-28)	191	14.4 (7.6)	0.007*

2.5 Discussion

My analysis of multiple acoustic features of chimpanzee pant-hoots found that pant-hoots could not be distinguished reliably based on the community identity, but instead reflected individual identity and potentially encoded some contextual information. The pant-hoots differed among the communities in only one type of acoustic features (the acoustic features of the climax scream), and only when I included the geographically distant Kanyawara community in the analysis. However, I did not find statistically significant pairwise differences in post-hoc comparisons. Most importantly, I did not observe a statistically significant difference in the climax screams of the geographically adjacent communities of Gombe. I also did not observe any differences among the communities in either the structural features, the build-up features, or when taking all the acoustic features simultaneously. The pant-hoots differed most substantially among individuals, irrespective of the inclusion of the geographically distant community in our analyses. The acoustic features of the climax scream element and the structural acoustic features distinguished the individuals, whereas the acoustic features from the build-up element alone did not. My findings indicate that individual differences are more prominent than group differences in the acoustic structure of chimpanzee pant-hoots.

We found that the context of the vocalization could be identified from some structural acoustic features but not from any other kind of acoustic features. Within the structural features, the number of climax elements was higher in feeding contexts and the number of letdown elements as well as build-up elements was higher in traveling contexts. My results support the findings of Clark & Wrangham (1993), Fedurek et al. (2016), and Notman & Rendall (2005) in finding an association of the letdown phase with the context of the pant-hoot. Their observations of greater number of pant-hoots with letdown components in traveling contexts are consistent with my findings of observing a greater number of letdown components in traveling contexts. However, I did not have sufficiently detailed behavioral data to distinguish food arrival pant-hoots separately, and hence, I could not confirm the finding of Clark & Wrangham (1993), that a higher proportion of pant-hoots with letdowns occurred in the context of arrival at a food source. I further observed two more differences that have not been reported previously. First, I found that pant-hoots given in feeding contexts had more climax elements. Second, I observed a higher number of build-up elements in travel context. Furthermore, I found no differences between the contexts in other acoustic features that describe the tonal properties of the build-up and climax elements. Uhlenbroek (1996) described different types of pant-hoots based on their tonal and spectral properties: a ‘wail-like’ pant-hoot is a pant-hoot with clear harmonic structure and a power spectrum with clear peaks; a ‘roar-like’ pant-hoot is a noisy pant-hoot lacking a clear harmonic structure and a more evenly distributed power spectrum (Uhlenbroek, 1996). Notman & Rendall (2005) found that pant-hoots given in traveling contexts were more ‘roar-like’ and those given in feeding contexts were more ‘wail-like.’ Since I found no context differences in the acoustic features related to the tonal properties, fundamental frequency, noise, or peak frequency, I could not confirm the findings from either Uhlenbroek (1996) or Notman & Rendall (2005). My results indicate that a more fine-grained differentiation of contexts while recording pant-hoots may be needed to distinguish arrival pant-hoots as well as pant-hoots from other contexts such as resting, grooming, displaying. Additionally, my findings suggest that future studies should pay special attention to the structural features whenever the context of pant-hoot production is relevant to the analysis.

In contrast to communities and contexts, I found substantial differences among individuals. Individuals differed in structural features and in climax scream features, but not in build-up element features. When all features were taken together, I observed the strongest differences in the climax scream features. The temporal properties that revealed greatest individual distinctiveness were duration of the climax phase, duration from build-up to letdown, and number of climax elements and screams. The spectral acoustic features showing the greatest individual differences were acoustic features related to the fundamental frequency F0. Specifically, the start, minimum, maximum, and mean F0, frequency range of F0, and minimum peak frequency were the features with the strongest individual level signal. Some of these acoustic features that correlated with individual differences were consistent with those identified by Crockford et al. (2004): maximum F0 and minimum peak frequency. Additionally, my results are consistent with results from previous studies by Mitani and colleagues (Mitani et al., 1999; Mitani & Brandt, 1994). Mitani & Brandt (1994) found that the principal component that explained the most variance among individuals loaded most highly in acoustic features of the fundamental frequency F0 including, start, minimum, maximum, and mean F0. Similarly, Mitani et al. (1999) found significant individual differences in the minimum, maximum, and mean F0, and the frequency range of F0.

My findings contrast with those of previous studies looking at community-specific acoustic differences in pant hooting (Crockford et al., 2004; Marshall et al., 1999; Mitani et al., 1992). In the first study reporting vocal dialects in chimpanzees, Mitani et al. (1992) found differences in geographically distant communities of Gombe and Mahale National Parks. Mitani and colleagues (1999) subsequently reassessed these findings but still found differences between geographically distant and Mahale and Kibale National Parks. In contrast to Crockford et al. (2004), I did not observe any significant differences among neighboring communities. The consistencies and inconsistencies of my results with previous studies reveal several insights and raise new questions. Consistent with previous studies, my study confirms the individual distinctiveness of chimpanzee pant-hoots in both spectral and temporal properties. My study also found some differences in the temporal properties of pant-hoots given in feeding and traveling contexts, confirming the possibility of some contextual encoding. In terms of community-specific differences, I could not confirm previous studies. My failure to find evidence for community-specific signatures could be an outcome of pant-hoots being innate calls, could reflect features peculiar to Gombe chimpanzees, or alternatively, it may be the case that previous findings of differences among chimpanzee communities resulted from statistical artifacts. Vocalizing vertebrates often exhibit both innate and learned vocalizations. For example, allopatric populations of humpback whales share temporally stable call types (Fournet et al., 2018), doves (*Streptopelia sp.*) show little geographic variation (De Kort et al., 2002), and gibbon (*Hylobates sp.*) loud calls show no evidence of vocal learning (Brockelman & Schilling, 1984). Such instances of a lack of learned signals could be explained by genetic similarities and hybridization. In chimpanzees, several community-specific peculiarities can lead to differential selection pressures for community-specific vocalizations. For example, (i) a recent history of intergroup violence could lead to a greater selection pressure for community-specific vocalizations to facilitate identifying own community vs. neighbors. There is a history of lethal intergroup violence in Gombe (Wilson et al., 2004), Kibale (Watts et al., 2006), as well as in Taï chimpanzees studied by Crockford et al. (2004) (Boesch et al., 2008). However, Gombe chimpanzees have experienced a higher rate of inter-community killings (Boesch et al., 2008; Wilson et al., 2004), suggesting that the selection for community-specific vocalizations should be at least as strong as that for Taï chimpanzees, if not higher. (ii) Stability of hierarchy and strength of affiliative bonds in the community promote vocal convergence (Fedurek, Machanda, et al., 2013; Mitani & Brandt, 1994; Mitani & Gros-Louis, 1998) and thus could create positive selection pressure for community-specific vocalizations. In Gombe, within-community bonds are likely stronger in the Kasekela community, which has more maternal brothers (Bray & Gilby, 2020) compared to Mitumba, which has fewer brothers and higher within-community violence (Massaro et al., 2021, 2022). More data are needed to accurately test if social bonds affect vocal convergence across field sites. (iii) A larger community size may lead to a greater selection pressure for community-specific signatures as it becomes more difficult to keep track of individuals. All communities in this study and in Crockford et al. (2004) were relatively small, so this is less likely to explain the discrepancies in the results. Furthermore, while Crockford and colleagues attempted to control for confounding factors, their sample size of only three individuals per community increases the possibility that apparent differences could emerge by chance. As evidenced from a simulation study (Loken & Gelman, 2017), noisy data with small sample sizes could lead to false positives. My study included a slightly greater number of individuals per community (5-7 individuals per community compared to 3 individuals per community in Crockford et al. (2004)) and hence should have detected any differences among communities that were similar in effect size to those reported by Crockford et al. (2004). However, because my sample size remains modest, I could have failed to detect differences if the effect size at Gombe is lower than that at Taï, and hence we cannot rule out the potential for false negatives either. Further, neither my study, nor Crockford et al. (2004) controlled for individual-level factors such as age, body size, health condition, and rank that could influence the acoustic structure. In addition, no studies have been able to quantitatively control for other factors such as the influence of habitat differences and sound environments that Mitani et al. (1999) suggested could be important. Hence, I argue that firm conclusions regarding chimpanzee vocal learning ability require further study, ideally with a larger number of sampled individuals per community. Furthermore, reanalyses of existing data with different methods such as Bayesian inference is another potential avenue for future research.

My results reinforce the importance of replicating findings in animal behavior research. A key feature of scientific discovery is seeking results that are consistently reproducible (Burman et al., 2010; Johnson, 2002; Lamal, 1990; Popper, 1959). In recent decades, analyses of studies in several scientific disciplines, including fields as diverse as psychology and medicine, have found that most scientific findings fail to be reproduced by subsequent studies, leading to what has been called the replication crisis (Ioannidis, 2005; Wiggins & Chrisopherson, 2019). One factor contributing to this crisis is that studies replicating existing findings are rarely conducted, and are implicitly discouraged through reviewer bias against them (Neuliep & Crandall, 1993). Given that field studies in animal behavior typically have smaller sample sizes than studies in psychology or medicine, it is likely that the field of animal behavior is in even greater need of replication to test the validity of previous results with sufficient sample sizes (Johnson, 2002). Within animal behavior, the need for replication may be particularly acute for species such as chimpanzees, for which field conditions make it challenging to obtain sample sizes sufficient to be confident in results. Long-term data from multiple field sites have proven essential for providing sufficient sample sizes for a range of topics (e.g., culture: (Whiten et al., 1999); reproductive cessation: (Emery Thompson et al., 2007); lethal aggression: (Wilson et al., 2014)). Such collaboration across long-term studies will be essential for answering questions about vocal communication as well.

Chapter 3

Does sexual-selection influence the variation in chimpanzee pant-hoots via acoustic cues of male quality?

3.1 Overview

In many species, animals produce conspicuous vocalizations to advertise their quality to prospective mates. Sexual selection theory predicts that acoustic structure may provide cues of individual traits that signal mate quality such as body size, age, physical condition or health. Given the importance of male body size in mate choice and competition for mates, cues of body size such as low pitch and high formant frequency dispersion may be particularly important. Chimpanzee pant-hoot vocalizations are loud complex calls, consisting of high and low pitch elements, that are particularly interesting from the perspective of sexual selection. Their complexity and sexual dimorphism points towards the possibility that they may be sexually selected. Previous studies have suggested that pant-hoots may play a role in attracting females to male callers. However, the high-pitched, tonal components of pant-hoots present something of a puzzle from a sexual selection standpoint, as they are unlikely to provide cues of large body size. A previous study found that pant-hoots are difficult to produce, and individuals may be reaching their physiological limits in producing them, thus providing honest signals of physical condition. The highest pitch elements in the pant-hoots–that require vocalizing at peak physiological capacity–exhibit certain distortions in the acoustic structure that are known as non-linear phenomena (NLP). Hence, producing high pitch vocalizations with few NLPs might provide honest signals of superior physical condition. In this chapter, I study whether the proportion of NLPs, the pitch (F0max) and noise in the climax-scream elements in the pant-hoots contain cues of markers of mate quality including age, rank, and health (SIVcpz status), and if they predicted male mating success. I found a non-linear relationship between the proportion of NLPs and age–young and old individuals exhibited higher proportions of NLPs compared to individuals in their prime mating age. Next, at a population level, I observed a lower proportion of NLPs and lower noise measures in higher ranking individuals, but after controlling for the effect of age, the noise exhibited no relationship with rank, and proportion of NLPs had a significant positive effect, possibly indicating the attention-grabbing role of NLPs for high-ranking individuals. In terms of cues of health, I observed a lower proportion of NLPs and maximum fundamental frequency, and higher maximum noise in SIVcpz positive individuals. However, only the association with noise was statistically significant. Lastly, I observed that a lower proportion of NLPs predicted higher mating success even beyond the effects of rank and age on mating success. Collectively, these findings provide support for the sexual selection hypothesis as NLPs and noise provide reliable cues to mate quality in terms of the age of the caller, some cues in terms of rank and health, and the proportion of NLPs predict the mating success of males. This study adds to previous findings supporting the sexually selected nature of chimpanzee pant-hoots and warrants further exploration of cues of mate quality in the acoustic structure of the pant-hoots.

3.2 Introduction

Animals may choose mates based on various factors, including cues of genetic and phenotypic quality (Andersson & Simmons, 2006). The specific features indicating mate quality vary across species, but the general theme is that individuals possess features that directly or indirectly indicate their genotypic or phenotypic quality. One way by which traits reliably signal quality is if they are costly to produce, because costly signals make cheating more difficult (Zahavi, 1975). In many animals, vocalizations can function as costly signals of quality; if calls are complex or difficult to produce, they can reliably indicate phenotypic and/or genetic quality. For example, larger male Antarctic leopard seals (*Hydrurga leptonyx*) can produce more underwater calls with longer calling bouts than smaller males, thus providing reliable cues of body size (Rogers, 2017). Producing vocalizations underwater is difficult for a mammal as they cannot breathe and must produce vocalizations while holding their breath. Vocalizations signal body size and female quality also in Asian particoloured bats (*Vespertilio sinensis*) (Zhao et al., 2018). In primates, vocalizations signal quality by providing cues of dominance and male fighting ability as in male chacma baboon (*Papio ursinus*) ‘wahoo calls’ (Kitchen et al., 2003, 2013), crested macaques (*Macaca nigra)* (Neumann et al., 2010), Thomas langurs (*Presbytis thomasi)* (Wich et al., 2003), and chimpanzees (Riede et al., 2007). Vocalizations may also indicate age in many animals, for example red deer (*Cervus elaphus*) (Reby & McComb, 2003), giant panda (*Ailuropoda melanoleuca*) (Charlton et al., 2009), European ground squirrels (*Spermophilus citellus*) (Schneiderová et al., 2015), North Atlantic right whales (*Eubalaena glacialis*) (McCordic et al., 2016), and many primates, for example De Brazza's monkeys (*Cercopithecus neglectus*) (Bouchet et al., 2012), pig-tailed langur (*Simias concolor*) (Erb et al., 2013), chacma baboons (Fischer et al., 2004), white-handed gibbons (*Hylobates lar*) (Barelli et al., 2013), and chimpanzees (Fedurek, Zuberbühler, et al., 2016). Calls may also reflect correlates of health such as caller fatigue in pig-tailed langur (Erb et al., 2013) and stamina in chacma baboons (Fischer et al., 2004).

Vocalizations could be subject to sexual selection not only in the wider animal kingdom but also in humans. Acoustic features could influence mate quality perception, mating success, as well as sexual orientation (Suire et al., 2018, 2020). Studies also support the possibility that male voice may be a costly signal of phenotypic quality. Arnocky and colleagues found that masculine vocal characteristics were associated with a marker of immunocompetence (salivary immunoglobulin-A) as well as self-reported health status (Arnocky et al., 2018). Voice could be a reliable indicator of upper body strength as American undergraduates could reliably assess upper body strength in male voices of linguistically and culturally diverse populations of Andean herder-horticulturalists, Tsimane of Bolivia, and Romanian and US college students (Sell et al., 2010). Hodges-Simeon and colleagues found that Bolivian adolescent males with lower voices were in better energetic condition and had higher testosterone levels (Hodges-Simeon et al., 2015). Suire and colleagues reported that women found lower pitch and higher intonation vocal patterns more attractive in men (Suire et al., 2019). Association of lower pitch with mating success in men is not limited to Western cultures (Hodges-Simeon et al., 2010; Suire et al., 2019), but is also observed in Amazonian forager-horticulturists (Rosenfield et al., 2020).

Given the evidence of sexual selection in animals as well as humans, it is particularly interesting to study the vocalizations of one of our closest evolutionary relatives: chimpanzees. Chimpanzees exhibit a diverse vocal repertoire with call types ranging from soft grunts to loud hoos and screams (Goodall, 1986; Marler, 1969). Of the range of chimpanzee vocalizations, the pant-hoot has received considerable research attention for its sexual function. This is because chimpanzee pant-hoots are loud, acoustically complex calls that are given in a range of different contexts and likely serve a variety of different functions in chimpanzee societies (Clark & Wrangham, 1994; Fedurek et al., 2014; Fedurek, Slocombe, et al., 2016; Mitani & Brandt, 1994). They are conspicuous calls that can be heard over long distances due to their loudness. They are produced in many different contexts such as while arriving at food patches and feeding (Clark & Wrangham, 1993), traveling, during displays of dominance, resting, fighting, intergroup encounters, etc. (Desai et al., 2021; Fedurek, Machanda, et al., 2013; Wilson et al., 2007). Given this variety of production, it is clear that they serve multiple functions. For example, pant-hoots may play a role in attracting females or allies (Reynolds & Reynolds, 1965; Wrangham, 1977), signaling territory ownership (Crockford et al., 2004; Mitani et al., 1992; Wilson et al., 2007), signaling status (Clark, 1993; Wilson et al., 2001) or physical condition (Riede et al., 2004, 2007). Several properties of pant-hoots indicate that may be sexually selected signals of male mate quality and competitive ability. For example, males produce pant-hoots more frequently than females (Arcadi, 1996); high-ranking males produce pant-hoots more often than low-ranking males (Clark, 1993; Clark & Wrangham, 1994; Mitani & Nishida, 1993; Wilson et al., 2001); males produce pant-hoots in the peripheries of their territory, potentially to signal territory ownership (Wilson et al., 2007); males with high testosterone produce more pant-hoots (Fedurek, Slocombe, et al., 2016); and males give more pant-hoots in presence of parous swollen females (Fedurek et al., 2014). These results point towards the possibility that chimpanzee pant-hoots may be sexually selected signals of male mate quality.

Researchers have studied various properties of the acoustic structure of animal vocalizations to assess their role as costly signals of mate quality. For example, in birds, vocal repertoire size, calling rate, trill rate, and pitch reflect the cost of production and hence, the quality of the mate (Araya-Ajoy et al., 2009; Ballentine et al., 2004; Catchpole, 1996) and pitch reflects body size in owls (Hardouin et al., 2007). In mammals, call rate in leopard seals (Rogers, 2017), and formants in red deer stags (Reby et al., 2005) and rhesus macaques (Fitch, 1997) indicate body size and male competitive ability, and syllable duration indicates female body size in particoloured bats (Zhao et al., 2018). In humans, low voice pitch may reflect dominance (Wolff & Puts, 2010), mate quality (Hodges-Simeon et al., 2015), and mating success (Suire et al., 2018). Similarly, other acoustic features such as breathiness, loudness, and articulation rate may predict mate quality and mating success as well (Suire et al., 2018, 2019). Given the evidence for the attractiveness of low pitch in heterosexual humans and other animals, pant-hoots exhibit an important puzzling property: they not only contain low-pitch elements, but also extraordinarily high pitch elements (up to 2000Hz) (Desai et al., 2021). These high pitch elements are the scream-like tonal elements in the climax phase of the pant-hoots (see Methods in chapter 2, Figure 8). They’re also the loudest elements that could be heard from long distances in their habitat and may play a role in attracting females to the location of the calling male (Wrangham, 1977). If pant-hoots are subject to sexual selection and contain properties that are attractive to the females, we need to explain the prevalence and the function of these high pitch elements.

One hypothesis that could explain the how sexual selection may occur in pant-hoots is that vocalizations may contain cues of an individual’s health and physical condition (Fitch et al., 2002; Riede et al., 2004, 2007). Pant-hoots (and pant-hoot climax-screams in particular) exhibit what are known as non-linear phenomena (NLP) in the acoustic structure (Riede et al., 2004, 2007). NLPs are disruptions in the tonality of a call that are visible in the spectrogram representations of vocalizations. NLPs are an outcome of the physics of vocal fold vibrations. The vocal folds act as coupled oscillators during vocal production wherein the movements of one of the folds affect the movements of the other and together exhibit complex non-linear dynamics. Coupled oscillator non-linear dynamics involve a range of phenomena; from periodic limit cycles that manifest as harmonic, tonal vocalizations, to deterministic chaos that resembles, but is separate from, noise (Fitch et al., 2002; Tokuda, 2018; Wilden et al., 1998). The described NLPs in mammal vocalizations include frequency jumps, subharmonics, biphonation, and deterministic chaos (Fitch et al., 2002; Riede et al., 2004, 2007; Tokuda, 2018) (See Methods for more details). Given that chimpanzee pant hoots are produced through the vibration of vocal folds that are susceptible to NLPs, producing loud, high-pitch, and tonal climax-screams may be physiologically costly. Hence, the prevalence of such NLPs in the pant-hoot calls may signal poorer physical condition of the caller (Riede et al., 1997, 2004, 2007).

Studies support this possibility. Riede and colleagues found that the presence of non-linear phenomena in chimpanzee pant-hoot climax-screams were associated with the highest frequency in the first harmonic (Riede et al., 2007). This suggests that the individuals are vocalizing at the upper limits of the stability of vocal folds, and the occurrence of non-linear phenomena represents the loss of stability. Individuals that are young and healthy may have vocal folds that are better able to maintain the production of linear acoustics at high frequencies than older or unhealthy individuals (Riede et al., 2007). NLPs are commonly observed in other mammal vocalizations and researchers have proposed several functions of the NLPs that suggest the possibility of sexual selection. NLPs could indicate physical condition, age, body size, stress levels, and arousal (Blumstein & Récapet, 2009; Charlton et al., 2017; Digby et al., 2014; Fitch et al., 2002; Marx et al., 2021; Serrano et al., 2020; Stathopoulos et al., 2011). Serrano and colleagues found that NLPs indicated individual distinctiveness and chaos in particular indicated body size in Darwin’s frog (Serrano et al., 2020). NLPs may help defend territories and mates. Digby and colleagues found that kiwi calls produced in territorial contexts contained more subharmonics (Digby et al., 2014). Blumstein and Récapet conducted playbacks of marmot alarm calls with and without NLP (white noise) added to them and found that NLP increased the responsiveness to alarm calls as marmots foraged for less time following alarm calls with NLP (Blumstein & Récapet, 2009). In koalas, females often reject male copulation attempts by producing rejection calls in response to male bellows during breeding season. Other males use these rejection calls as cues to approach the females, thus inciting male-male competition. Charlton and colleagues found that female rejection calls that contained more subharmonics, less biphonation, and higher harmonics-to-noise ratios incited greater looking responses among male koalas (Charlton et al., 2017). Similarly, red deer male calls that contained harsh roars elicited greater female attention (Reby & Charlton, 2012). NLPs could also provide indicators of age. Marx and colleagues found that NLPs in dog whines were associated with age and stress levels (Marx et al., 2021). In humans, Stathopoulos and colleagues found lower signal to noise ratio in younger and older individuals and higher signal to noise ratio in middle-aged individuals (Stathopoulos et al., 2011). However, to my knowledge, only two studies (Riede et al., 2004, 2007) have looked at NLPs in chimpanzee pant-hoot screams, and only one of them (Riede et al., 2007) studied the function of NLPs in the climax-screams.

Furthermore, if pant-hoots are sexually selected because they reflect male mate quality and physical condition, they should be associated with other markers of physical condition such as age, rank, and health. Additionally, they should also predict reproductive success for the individual. This correlation could be a direct causation if females choose to mate with males with attractive pant-hoots. On the other hand, it could be indirect if males with attractive pant-hoots also possess other traits that directly increase reproductive success such as male competitive ability or aggressive temperament. The former relies on female choice, the extent of which is disputed in the literature. Studies have found mixed evidence of female choice in chimpanzees. On one hand, many studies report evidence of female choice e.g. females mate with high-ranking males during the periovulatory period of cycling when conception is most likely (Matsumoto-Oda, 1999). Also, females show differential mate preferences during periovulatory and non-periovulatory periods in order to strategically confuse paternity

(Pieta, 2008). Additionally, Stumpf and Boesch found that female proceptivity and resistance are directly and inversely related to male reproductive success irrespective of male rank, indicating strong female choice (Stumpf & Boesch, 2006); and male coercion was rare and exhibited by less preferred males, and did not effectively constrain female mate choice (Stumpf & Boesch, 2010). On the other hand, some recent studies have found that male coercion rather than female choice explains male reproductive success. Muller and colleagues report that females mate with more aggressive males and even the apparently choosy behaviors of females (e.g. copulatory approaches) are correlated with aggression in males (Muller et al., 2011). In Gombe, greater sexual coercion results in more paternities (Feldblum et al., 2014).

In this study, I aim to extend Riede and colleagues' previous work on studying the prevalence and functions of NLPs in chimpanzee pant-hoot screams. Riede and colleagues described the prevalence of different NLPs in the climax-screams including frequency jumps, subharmonics, biphonation, and deterministic chaos. They also studied NLPs' function by testing if the F0max (the maximum value of F0 in a given call segment) and F0mean (the mean value of F0 in a given call segment) of climax-screams exhibiting NLPs had cues of age and rank (Riede et al., 2007). While they did not find a statistically significant association of F0max or F0mean in climax-screams with NLPs and age or rank, they argued that it might be due to a small sample size and found preliminary evidence that rank may be associated with F0max as a high-ranking male in their sample had the highest F0max, which got lower after the male lost his high rank. Hence, a further analysis is required to be confident in whether NLP measures and F0 measures provide cues of age. For instance, no study has tested for the differences in prevalence of NLPs by age. Given that age is an important marker of physical condition and mate quality, and that NLPs provide cues of physical condition (Riede et al., 2007), then the occurrence of NLPs should vary with age. In this chapter, I leverage a longitudinal dataset of pant-hoot vocalizations to study both within and between individual differences in the prevalence of NLPs by age. Furthermore, an individual's attractiveness may vary by age in a non-linear fashion, i.e. a male may be more attractive in peak adult age compared to adolescent or old ages. Hence, I study the relationship of NLPs with age using polynomial regressions to allow for non-linear relationships. Furthermore, while F0max and F0mean were not associated with age or rank in Riede et al.'s study, they may me associated with other markers of male mate quality including health, and mating success. Additionally, the prevalence of NLPs may be associated with rank, health, and mating success, which Riede et al. did not test. Hence, I test for these associations in my dataset. Lastly, I also include measures of noise (mean and maximum noise) as it provides a broader measure of NLPs compared to proportions of subharmonics, biphonation, and deterministic chaos.

3.3 Methods

3.3.1 Subjects and study sites

I obtained recordings of chimpanzee pant-hoots from two study sites: Gombe National Park, Tanzania, and Kibale National Park, Uganda. From Gombe, we obtained recordings from two communities: Kasekela and Mitumba. From Kibale, I obtained recordings from the Kanyawara community. Gombe is located along the shore of Lake Tanganyika (4°40′S, 29°38′E) in western Tanzania. Kibale is in western Uganda (0°33’N, 30°21’E).

To obtain a longitudinal dataset to assess the relationship with age more robustly, I collected recordings of pant-hoots obtained during different studies of chimpanzee vocal communication from both sites. From Gombe, we obtained recordings from the following years (i) Kasekela: 1992-1993 (Uhlenbroek, 1996), 2012-2013 (O’Bryan, 2015), 2016-2017 (Desai et al., 2021); and (ii) Mitumba: 2016-2017 (Desai et al., 2021). From the Kanyawara community in Kibale, we obtained recordings from the following years: 1988 (Clark & Wrangham, 1993), 1997-1998 (Wilson, 2002; Wilson et al., 2007), 2010-2011 (Fedurek, Schel, et al., 2013). Collectively, I obtained recordings from N=8 males from Kasekela, N=5 males from Mitumba and N=10 males from Kanyawara. Figure 13 describes the number of pant-hoots from each male and his age at recording. In total, I included N=359 pant-hoot calls and N=877 climax-screams in the analyses.

For assessing the relationship of proportion of NLPs and other acoustic features including F0max, F0mean, mean and maximum noise, to other markers of male mate quality including rank, health (SIV status), and mating success, I only used the recordings from the two Gombe communities, Kasekela and Mitumba. This is because the data about these markers were only available from Gombe.

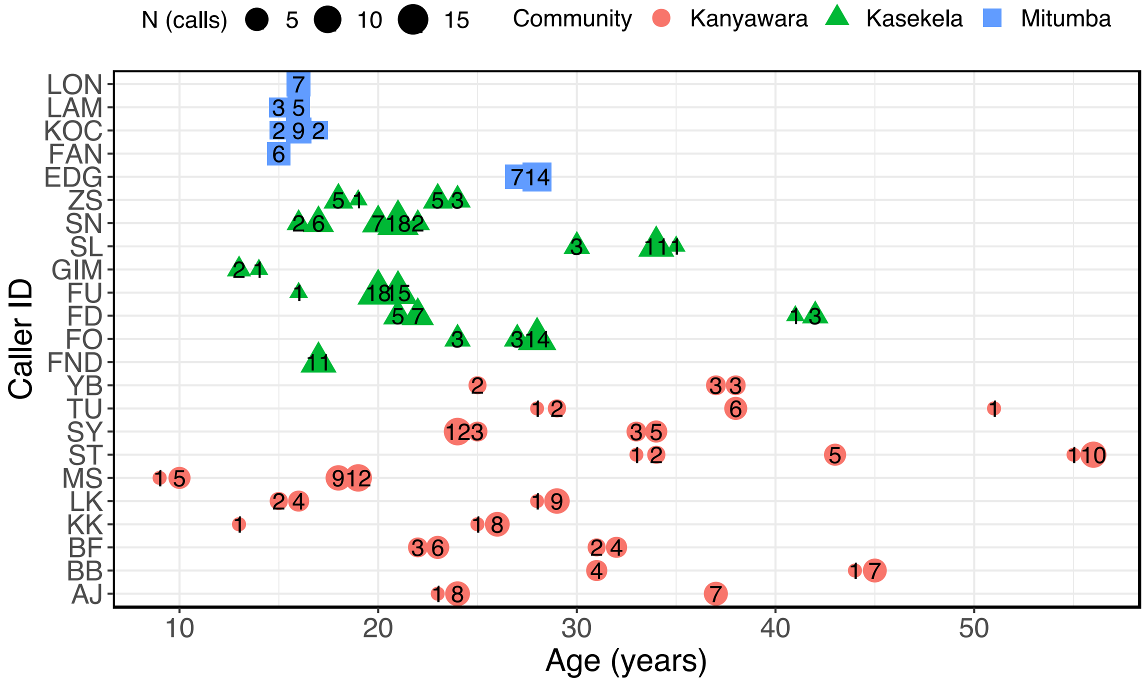


Figure 13: Number of pant-hoots included from each male at different ages.

3.3.2 Data collection

3.3.2.1 Audio recordings of chimpanzee vocalizations

I collected recordings of chimpanzees at Gombe from three studies (Desai et al., 2021; O'Bryan, 2015; Uhlenbroek, 1996). In the first study, Uhlenbroek and two Tanzanian researchers Issa Salala and Karoli Alberto recorded vocalizations from chimpanzees of the Kasekela community. They obtained recordings during a period of 18 months, from May 1992 to October 1993. They performed two-point, all-day focal follows (simultaneous follows of two males at different locations in the forest) and recorded all pant-hoots they heard using a Sony Professional Walkman recorder and either Sony ECM 909 or Aiwa multidirectional clip-on microphones. For more details on the methods, see (Uhlenbroek, 1996).

In the second study, O'Bryan, assisted by Tanzanian field assistants Kassimu Sadick and Sadiki Haruna, obtained recordings from the Kasekela community during 8 hour focal follows of 10 adult male chimpanzees. She recorded during February-July 2012 and January-June 2013. She used a Marantz PMD 620 recorder and a Sennheiser ME66 shotgun microphone with K6 power module. For more details on the methods, see (O'Bryan, 2015).

In the third study, two Tanzanian field assistants Hashim Issa Salala and Nasibu Zuberi Madumbi, trained by Dr. Michael Wilson and me obtained recordings from the Kasekela and Mitumba communities. They obtained recordings from July 2016 to December 2017 during 8-hour focal follows using a Marantz PMD661 MKII audio recorder and a Sennheiser ME66 shotgun microphone with K6 power module. For more details on methods, see Methods in chapter 2.

From Kanyawara, I collected recordings from three studies (Fedurek, Schel, et al., 2013; Riede et al., 2004; Wilson, 2002). In the first study, Arcadi and field assistant Francis Mugurusi obtained recordings from the Kanyawara community between May 1988 and December 1989. They used a Marantz PMD430 cassette recorder, Sennheiser ME80 directional microphone with K3U power module, and CrO2 cassette tapes (Clark & Wrangham, 1993; Riede et al., 2004).

In the second study, Wilson obtained recordings from the Kanyawara community from June 1997 to November 1998. He obtained recordings during all-day party follows using a Sony TCD-D8 DAT audio recorder and a Sennheiser ME66 shotgun microphone with K6 power module (Wilson, 2002).

In the third study, Fedurek recorded vocalizations from the Kanyawara community from October 2010 to September 2011. He obtained recordings during all-day focal follows using a Marantz Professional PMD661 solid-state recorder and a Sennheiser ME67 shotgun microphone (Fedurek, Schel, et al., 2013).

3.3.2.2 Data on individual male quality traits

I obtained the data on markers of male quality including age, rank, health, and mating success from the Gombe chimpanzee database that includes detailed demographic data (1963-present) (Strier et al., 2010), dominance ranks (males: 1978-2017) (Foerster et al., 2016); mating success (males: 1978-2017) (Wilson et al., 2020); SIVcpz infection status (2000-present) (Keele et al., 2009), and systematic health data, including clinical signs (2005-present) (Lonsdorf et al., 2018).

I calculated daily dominance ranks from the data on pant-grunt encounters. The data included each instance of a pant-grunt observed, the ID of the individual producing the pant-grunt, and the ID of the individual receiving the pant-grunt. In a given dyadic interaction, the individual who produced a pant-grunt was considered the subordinate individual in that interaction ("loser" of that interaction), and the individual who received the pant-grunt was considered the dominant individual in that interaction ("winner" of that interaction). Such data that include "winners" and "losers" of dyadic encounters could be used to calculate Elo scores, which could be used as a measure of the position in the dominance hierarchy. I used the *eloratingopt* function in the EloOptimized R-package to calculate the Elo scores (Foerster et al., 2016). I used the EloCardinal measure as the measure for the dominance rank. EloCardinal is the expected number of dyadic encounters won by an individual scaled as the percent of the total number of ranked individuals present on the day of calculating the measure (Foerster et al., 2016). I used this measure to test whether acoustic features of a calls predicted the dominance ranks of individuals.

In the health data, there were six clinical signs recorded on a weekly basis for each individual: abnormal weight, visible skin conditions, wounds, lameness, abnormal feces, and respiratory infection symptoms including, runny nose and cough (Lonsdorf et al., 2018). Additionally, I obtained the data about the SIVcpz status of each individual including the date at which an individual became SIVcpz positive (Keele et al., 2009). Among the six clinical signs of health that were available, the most relevant sign that could affect the vocalizations was the presence of respiratory infection symptoms. Over the durations of recording the vocalizations at Gombe from Uhlenbroek, O'Bryan, and my studies, only three Mitumba individuals, EDG, LAM, and LON showed respiratory infection symptoms over only a duration of three weeks. There were no recordings from this time period in the dataset so I could not test for the differences in the acoustic features of a sick vs. a healthy chimpanzee. However, I tested for the differences in the acoustic features of three SIVcpz positive chimpanzees, FO, ZS, and GIM compared to SIVcpz negative chimpanzees.

For estimating mating success, I had to determine the appropriate time resolution of the data, i.e., the time period for which the acoustic features of a call could be considered for testing the relationship with mating success. Considering a time period of one month led to elimination of many recordings of vocalizations (leaving only 71 recordings from 9 individuals) as there would be no matings with a parous, swollen female with the male whose recording I had, in the month in which the recording was obtained. Hence, I considered a time period of a quarter of a year (1st quarter: January to March, 2nd quarter: April to June, 3rd quarter: July to September, and 4th quarter: October to December). I calculated the number of matings an individual male obtained with parous females who exhibited sexual swellings during a given quarter of a year. Next, I calculated the total time that each male was observed when a parous, swollen female was present in the party during that quarter. Using these two estimates, I obtained the number of matings per 24h of observation within a quarter and used it as the estimate of mating success of an individual. I used this measure to test whether acoustic features of calls recorded in a given quarter predicted the mating success of the males in that quarter.

<u>3.3.3 Call selection</u>

Pant-hoots consist of four distinct acoustic phases, each defined by the similarities of the elements within each phase. These phases are: (i) introduction, (ii) build-up, (iii) climax, and (iv) letdown. Figure 8 shows a spectrogram of a typical pant-hoot from an adult male. For this study, I focused on the climax phase, because the climax phase is the most relevant phase for testing the sexual selection hypothesis due to the following properties. First, the climax phase consists of the loudest elements in the vocal repertoire of chimpanzees and is the phase that can be heard most easily over long distances. Second, an analysis of pant-hoot climax-screams suggested that the individuals producing the climax-screams were vocalizing at the peak capacity of their vocal folds and instability in vocal folds resulted in NLPs (Riede et al., 2007). Hence, I only considered pant-hoots with climax-scream elements for the analyses. Furthermore, I only included pant-hoots for which I was confident that the entire climax phases were recorded.

3.3.4 Acoustic analysis

As opposed to chapters 1 & 2 wherein I used all the acoustic features I could extract, for this study, I only chose a few biologically meaningful acoustic features informed based on previous studies: the proportion of climax scream duration that contains NLPs (Riede et al., 2007), the maximum frequency in the first harmonic/fundamental frequency F0 (F0max), the mean frequency in the first harmonic/fundamental frequency F0 (F0mean) (Riede et al., 2007), and a perceptually relevant measure of noisiness (from pure tone to white noise)—the Wiener entropy coefficient, a.k.a spectral flatness (Johnston, 1988) maximum (noise_max) and mean (noise_mean).

I used Praat version 6.2.04 to produce spectrograms and annotate each climax-scream to identify and annotate the NLPs within them. I obtained the durations of different NLPs (subharmonics, biphonation, and deterministic chaos), the number of frequency jumps, and durations of climax-screams by visually inspecting the spectrograms. To ensure the best view of the spectrogram for detecting NLPs and measuring their durations, I used a frequency range of 0 to 7000 Hz, a window length of 0.02 seconds, and a dynamic range of 55 dB. I obtained other acoustic features including F0max, F0mean, and mean and maximum Wiener entropy coefficient (noise_max and noise_mean) using LMA using the process described in the Methods in chapter 2. I observed a high correlation between F0max and F0mean (Pearson's $\rho = 0.82$), and between noise (max) and noise (mean) (Pearson's $\rho = 0.85$). Hence, I only considered one variable from these pairs: F0max and noise_max for the analyses.

3.3.5 Identifying non-linear phenomena

Frequency jumps manifest as abrupt changes in the F0 during an otherwise tonal segment of a call (Figure 14, left). They are caused by regime instability, i.e. unpredictable jumps in the oscillations of the vocal folds (Riede et al., 2004, 2007; Wilden et al., 1998). Subharmonics manifest as additional harmonics over the identifiable F0 and its integer multiples (Figure 14, middle). These typically appear at integer fractional values of F0, such as at F0/2, F0/3, etc. and include their own harmonics at integer multiples (Tokuda, 2018; Wilden et al., 1998). Biphonation manifests as either two independent fundamental frequencies or as sidebands along the original F0 and its harmonics (Figure 14, middle). It could be an outcome of asynchronous vibrations of the vocal folds—when the two folds vibrate at similar frequencies that are unrelated—causing them to move in and out of phase frequently (Riede et al., 2004; Wilden et al., 1998). Such asynchrony between the folds could be because of anatomical asymmetry of the larynx, or could be due to a pathological condition affecting the size of vocal folds (Riede et al., 2004). Lastly, deterministic chaos manifests as noise-like segments in a call (Figure 14, right). However, deterministic chaos is separate from random noise and instead represents dynamics of a deterministic system that never returns to the same state (Tokuda, 2018).

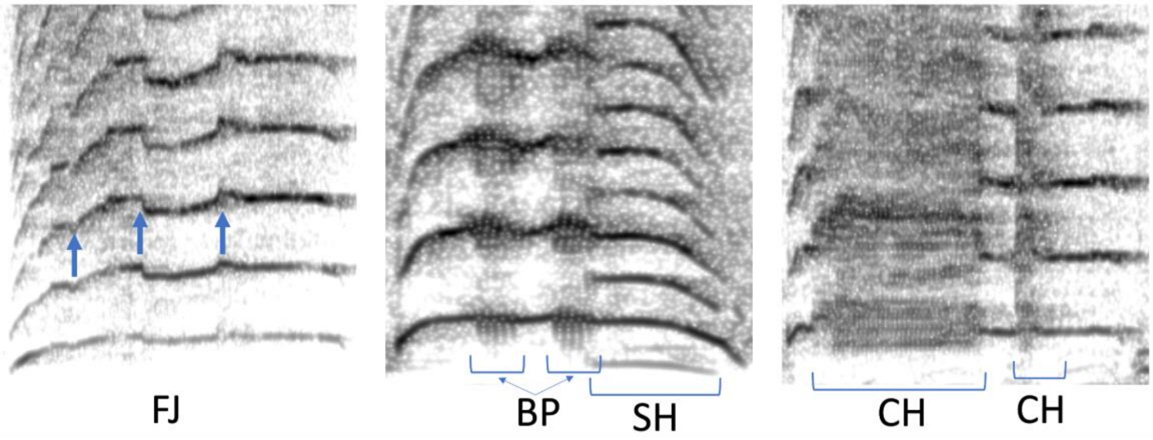


Figure 14: Spectrograms of three climax-screams exhibiting different non-linear phenomena. Climax-scream on the left exhibits three frequency jumps (FJ), on the middle exhibits sub-segments of biphonation (BP) and subharmonics (SH) and on the right exhibits sub-s sub-segments of deterministic chaos (CH).

3.3.6 Statistical analyses

3.3.6.1 Causal assumptions

To determine the control variables for obtaining unbiased estimates for the strengths of the relationships of interest in this study, I use a Directed Acyclic Graph (DAG) that specifies my assumptions of causal relationships among the variables, and the relationships of interest that I want to test (Figure 15). Age directly affects rank as males climb up in hierarchy as they mature, and then later start losing status as they get older and get challenged by younger males. Rank directly affects mating success as higher ranking males have greater access to fertile females. Age also directly affects health as younger males would be healthier and would have had less exposures to pathogens such as SIVcpz. Health could affect mating success as healthier males would be able to better compete for access to fertile females and would be more likely to be fertile themselves. Hence, age affects mating success indirectly via rank and health. Age may also have a direct effect on mating success as a male's fertility may be affected by age.

Starting with these causal assumptions, I determine the control variables for the relationships of interest using the *adjustmentSets* function from the dagitty R package (Textor et al., 2016). For the first relationship of interest: between the acoustic features (proportion on NLPs, F0max, or noise_max) and age, there is no confounding variable that needs to be added as a control variable.

For the next relationship of interest: between the acoustic features and rank, the only confounding variable to control for was {age} for the two acoustic features (proportion of NLPs and noise_max) for which I found a significant relationship with age (see Results sections 3.4.4 and 3.4.6), whereas no control variables needed to be included when testing for the relationship of F0max and rank as the relationship of F0max with age was not significant (see Results section 3.4.5).

For the next relationship of interest of acoustic features with mating success, the control variables varied for different models. For testing the relationship between the proportion of NLPs and mating success, the valid sets of control variables to include were either {age and rank}, or {health and rank} for the proportion of NLPs. This was because the relationships of the proportion of NLPs with age and rank were significant (see Results section 3.4.4). I chose {age and rank} as the control variables since the health measures were too sparse. For testing the relationship with noise_max and mating success, either only {age} or {rank and health} were the valid sets of control variables as the relationship of noise_max with age was significant (see Results section 3.4.6) but the relationship of noise_max and rank was not significant (see Results section 3.4.6). For testing the relationship of F0max and mating success, no control variables were needed as neither age nor rank had a significant relationship with F0max (see Results section 3.4.5).

Lastly, for testing the relationship of acoustic features (proportion of NLPs and noise_max) with SIVcpz status, the only control variable to include was age as that was sufficient to close all backdoor paths. For testing the relationship of F0max and SIVcpz status, no control variables were needed.

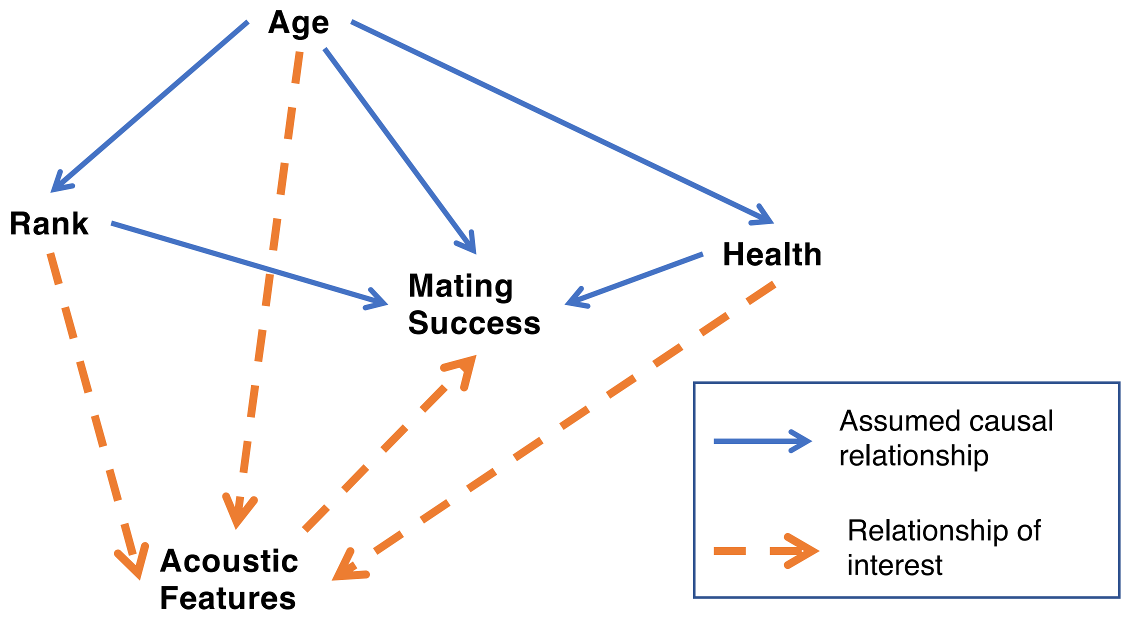


Figure 15: A DAG specifying the assumed causal relationships and relationships that are being tested in this study.

3.3.6.2 Models for proportion of NLPs

For modeling the proportion of NLPs in a pant-hoot as a function of age, mating success, rank, or health, I performed Bayesian binomial polynomial regressions. I also tested for the relationship of the proportion of NLPs with community ID to test whether it was a factor that needed to be controlled for using random intercepts in addition to individual ID for the relationships of interest. Since I aimed to model a response variable (the proportion of climax-scream duration exhibiting NLPs) that is a proportion, the suitable maximum entropy distribution to model it is a binomial distribution in an aggregated binomial regression (McElreath, 2018). A maximum entropy distribution is a distribution that contains theoretically the maximum amount of information for a given data generating process, making it the most plausible distribution that could describe a given dataset and the most conservative choice for modeling (*ibid.*). The binomial distribution has two parameters: the number of Bernoulli trials, and the number of successes among the trials. The proportion of interest in this study was the duration of subharmonics + biphonation + deterministic chaos observed in the climax-screams in a pant-hoot divided by the total duration of climax-screams in the pant-hoot. I did not include frequency jumps in this measure as they cannot be represented as the proportion of duration of climax-scream. To include the denominator so as to include the maximum available information in the dependent variable, I modeled it by using the total duration of NLPs (subharmonics + biphonation + deterministic chaos) observed in all climax-screams in each pant-hoot as the number of successes in the Bernoulli trials, with the total climax-scream duration in a pant-hoot as the total number of Bernoulli trials.

First, I modeled the relationship of the proportion on NLPs with age. Since age was the most basic independent variable that an individual has no control over, I tested for the association of age with an acoustic feature before any of the other individual traits. This allowed me to decide whether to control for the linear or non-linear effect of age in testing other relationships of an acoustic features with an individual male’s traits (see DAG in section 3.3.6.1). For the association with age, I also looked at individual NLPs separately and fit the same set of models for their proportions. The proportions considered for those models were the duration of subharmonics or biphonation or deterministic chaos observed in the climax-screams in a pant-hoot divided by the total duration of climax-screams in the pant-hoot. For mating success, rank, and health, I only considered the total proportion of NLPs and not the individual NLPs due to smaller sample sizes.

I fitted four polynomial models to represent different relationships with age: linear, quadratic, cubic, and 4th-degree polynomial. For each model, I fitted varying intercepts (a.k.a. random intercepts) for each individual chimpanzee to account for individual-level variation and the non-independence of observations due to multiple calls being included from the same individual. The mathematical structure of the 4th-degree polynomial model is specified below. Other models follow the same structure except with fewer polynomial terms.

$\text{NLP duration} \sim \mathit{Binomial}(\text{Scream duration}, p_i)$

$\text{logit}(p_i) = \boldsymbol{\alpha}_{\text{CALLER}[i]} + \boldsymbol{\beta}_1 * \text{Age} + \boldsymbol{\beta}_2 * \text{Age}^2 + \boldsymbol{\beta}_3 * \text{Age}^3 + \boldsymbol{\beta}_4 * \text{Age}^4$

$\boldsymbol{\beta}_j \sim \mathit{Normal}(0, 1) \text{ for } j = 1, 2, 3, 4$

$\boldsymbol{\alpha}_i \sim \mathit{Normal}(\bar{\alpha}, \boldsymbol{\sigma}_{\boldsymbol{\alpha}}) \text{ for } i = 1, 2, 3, . . . 23$

$\bar{\alpha} \sim \mathit{Normal}(0, 1.5)$

$\boldsymbol{\sigma}_{\boldsymbol{\alpha}} \sim \mathit{Exponential}(1)$

Here, p_i is the probability of observing a given duration of NLPs in climax-screams of observed duration. $\boldsymbol{\alpha}_{\text{CALLER}[i]}$ is the varying intercept for the i-th individual chimpanzee (I had N=23 total individuals) and $\bar{\alpha}$ is the mean intercept for the population of chimpanzees. $\boldsymbol{\sigma_\alpha}$ is the standard deviation of the intercepts and $\boldsymbol{\beta}_j$ is the coefficient of the j-th polynomial term. I standardized the Age (in months) before fitting all models and calculating the polynomial terms. I used weakly informative priors for the fixed effects ($\boldsymbol{\beta}_j$ and $\boldsymbol{\sigma_\alpha}$ coefficients) and an adaptive prior for $\boldsymbol{\alpha}_i$ that allows the model to learn the prior for each intercept from the data (McElreath, 2018). Weakly informative priors allow us to constrain the coefficients to biologically meaningful values while allowing the information in the data to dominate the estimation of the parameters. Standardizing the independent variables allowed me to use a mean of 0 for the priors of the independent variables. For the standard deviation component, I used an exponential prior that better resembles the distributions of variances and helps the models to converge (McElreath, 2018).

Similarly, to test for differences among communities, I used the same model with varying intercepts for different individuals, except with community ID as the only dependent variable. Since I found no effect of community ID (see Results section 3.4.2), I did not use community ID as a control variable in any of the models.

I fitted each model using the Hamiltonian Monte Carlo (HMC) method — a fast algorithm to estimate posterior distributions using the Markov chain Monte Carlo (MCMC) algorithm — implemented in RStan v.2.21.3. To fit the models, I used the *ulam* function in the rethinking R package v.2.21 (McElreath, 2013/2022) that utilizes the HMC algorithm implemented in RStan. I used R v.4.1.2 to perform all the analyses. Such Bayesian models are generative, i.e. they produce joint posterior distributions for each parameter. This allows us to sample from these posterior distributions for statistical inference and predictions. I could use the samples from the posterior to describe the posterior distributions in various ways including by calculating posterior means and standard deviations for each parameter. Furthermore, since I had the entire distributions, I could calculate credible intervals in various ways. Here, I used 89% highest posterior density intervals (HPDI). HPDI is the narrowest interval in the posterior distribution that contains a specified probability mass (here 89%). I followed McElreath (2018) in using an 89% interval to prevent readers from confusing it with the 95% confidence intervals used in frequentist inference as well as to discourage p-value like binary inference (McElreath, 2018). The direction, magnitude, and HPDIs of the parameters provide useful information about their effects. Additionally, plotting posterior predictions provides further insights into what the model expects the larger population of chimpanzees to be like. Hence, I reported both: parameter estimates, as well as plots of posterior predictions. I ensured that the models converged properly by visually inspecting the trace plots of Markov chains, and by ensuring that the Gelman-Rubin convergence diagnostic (Rhat) remained close to the value of 1 (Gelman & Rubin, 1992). Rhat for all the estimates in all of the models were either 1, 1.01, or 1.02.

For testing if the proportion of NLPs predict rank, I considered quadratic and linear polynomial models. Since I found that age predicted the proportion of NLPs quadratically (see Results section 3.4.4), I controlled for the quadratic effect of age when testing for the relationship with mating success or rank (see DAG in section 3.3.6.1). This allowed me to assess whether rank predicts the proportion of NLPs above and beyond age. The equation modeled for the quadratic rank relationship was of the form $\text{logit}(p_i) = \boldsymbol{\alpha}_{\text{CALLER}[i]} + \boldsymbol{\beta}_1 * \text{EloCardinal} + \boldsymbol{\beta}_2 * \text{EloCardinal}^2 + \boldsymbol{\beta}_3 * \text{Age} + \boldsymbol{\beta}_4 * \text{Age}^2$, and the linear model did not include a term for EloCardinal^2. The specifications of priors remain the same as the age models.

For testing if the proportion of NLPs predict mating success, the quadratic effect of age (see Results section 3.4.4) and the linear effect of rank (see Results section 3.4.4) needed to be controlled for (see DAG in section 3.3.6.1). The mathematical specification of the quadratic equation modeled for the mating success relationship was of the form $\text{logit}(p_i) = \boldsymbol{\alpha}_{\text{CALLER}[i]} + \boldsymbol{\beta}_1 * \text{Mating\_success} + \boldsymbol{\beta}_2 * \text{Mating\_success}^2 + \boldsymbol{\beta}_3 * \text{EloCardinal} + \boldsymbol{\beta}_4 * \text{Age} + \boldsymbol{\beta}_5 * \text{Age}^2$, and the linear model did not include the quadratic term for mating success. The specifications of priors remain the same as the age models.

When testing for the difference between SIVcpz positve and negative chimpanzees, the Markov chains did not converge. Hence, I fitted binomial mixed models using Template Model Builder (TMB) and maximum likelihood estimation implemented in the glmmTMB R package (Magnusson et al., 2017). Since there was no reason to expect a non-linear relationship with SIVcpz status, I only fitted a linear model. Similar to previous models, I included age and age^2 as control variables in the model (see DAG in section 3.3.6.1) and controlled for the individual ID using random intercepts for different individuals.

I compared the polynomial models using the Widely Applicable Information Criterion (WAIC) for Bayesian models using the *compare* function in the rethinking R package. WAIC provides an estimate of the out-of-sample performance of a model while making no assumptions about the posterior, thus making it more general and reliable than the traditional Akike Information Criterion (McElreath, 2018). I used the standard error of the difference in WAIC to compare the relative performance of each pair of models and determine the overall best model.

3.3.6.3 Models for F0max

For modeling F0max as a function of age, mating success, rank, or health, I first performed Bayesian polynomial regression. The suitable maximum entropy distribution for F0max was the normal distribution and hence I attempted to model it using the normal distribution (McElreath, 2018). However, none of the Markov chains converged. Hence, I used the updated *lmer* function that provides p-values using the Satterthwaite approximation for degrees of freedom from the lmerTest R-package (Kuznetsova et al., 2015) for the F0max models. I first tested the relationship of F0max with age with quadratic and linear polynomial models. Since I observed no relationship with age, I tested for the relationship with rank and mating success without including age as a control variable (see DAG in section 3.3.6.1). Also, since there was no relationship with rank either, the model of relationship with mating success did not include rank. Similarly, no control variables were added in the model comparing F0max of SIVcpz positive and negative individuals.

3.3.6.4 Models for noise_max

For modeling noise_max as a function of age, mating success, rank, or health, the suitable maximum entropy distribution was the beta distribution as noise_max is a proportion with unknown denominator. Beta regressions are better performed using the glmmTMB R package and hence, I used the glmmTMB function with beta family of distributions. I used no control variables when testing the relationship with age, used linear and quadratic terms of age as controls when testing the relationship with rank and SIVcpz status as age had a quadratic relationship with noise_max (see Results section 3.4.6), and used only linear and quadratic terms for age as controls when testing for the relationship with mating success as I found no relationship between noise_max and rank (see Results section 3.4.6 and DAG in section 3.3.6.1). I used the *model.sel* function in the MuMIn R package, to compare models using AICc (Barton, 2010).

3.4 Results

3.4.1 Prevalence of NLPs

NLPs were present in the majority of the pant-hoots from all communities. In the entire dataset (N = 359 calls and N = 877 climax-screams), I observed at least one NLP in 85% of the calls and 70% of the climax-screams. In particular, I observed frequency jumps in 51% of the calls and 31% of the climax-screams, subharmonics in 31% of the calls and 17% of the climax-screams, biphonation in 38% of the calls and 26% of the climax-screams, and deterministic chaos in 38% of the calls and 22% of the climax-screams. When looking at within-individual variation (N=23 individuals), I observed a mean of 86% calls with at least one NLP (SD=11, range = [68,100]), and 73% climax-screams with at least one NLP (SD=14, range = [41,96]) (Table 18). Frequency jumps were the most common kind of NLP observed, followed by deterministic chaos, biphonation, and subharmonics in similar proportions (Table 18). I summarize the proportion of calls and climax-screams with different NLPs for each individual at different ages in appendix Table 35; and for different communities in appendix Table 36. Furthermore, at the population level, I observed more than one kind of NLP in 51% of the calls and 22% of the climax-screams.

Table 18: Mean percent of pant-hoot calls and climax-screams within the pant-hoots exhibiting a given non-linear phenomenon, its standard deviation, and range.

Unit	At least one non-linear phenomenon $\bar{x}$ ± SD [range]	Frequency Jumps $\bar{x}$ ± SD [range]	Subharmonics $\bar{x}$ ± SD [range]	Biphonation $\bar{x}$ ± SD [range]	Deterministic Chaos $\bar{x}$ ± SD [range]
Calls (N = 359)	86 ± 11 [68,100]	52 ± 23 [10,86]	33 ± 22 [0,83]	37 ± 25 [0,91]	43 ±24 [0,100]
Climax screams (N = 877)	73 ±14 [41,96]	34 ± 18 [6,75]	21 ± 19 [0,75]	25 ±18 [0,69]	24 ±14 [0,51]

3.4.2 Community differences in proportion of NLPs in climax-screams

I observed no differences among the communities in the proportion of climax-scream duration that exhibits NLPs, as the credible interval of pairwise differences for each community pair included 0 (Table 19). The three communities had differences in the age distribution of the individuals included with Kanyawara being the oldest on average and Mitumba being the youngest on average (Kanyawara $\bar{x}$ = 29.6 years, SD = 10.8, range = [8.6, 56.0]; Kasekela $\bar{x}$ = 22.0 years, SD = 5.5, range = [13.4, 41.9]; Mitumba $\bar{x}$ = 20.2 years, SD = 5.9, range = [14.8, 28.3]). In the main regression model, the patterns of NLP proportions matched this age distribution. I observed the highest relative proportion of NLPs in Mitumba that included the youngest individuals ($\boldsymbol{\beta}$ (Mitumba) = 0.12, 89% HPDI: [-0.49, 0.68]), followed by Kanyawara that included the oldest individuals ($\boldsymbol{\beta}$ (Kanyawara) = -0.06, 89% HPDI: [-0.65, 0.52]), and relatively lowest proportion of NLP in Kasekela that included individuals of mean age 22.0 years ($\boldsymbol{\beta}$ (Kasekela) = -0.18, 89% HPDI: [-0.73, 0.42]; Table 37). This suggests a non-linear relationship of the proportion of NLPs to age. I summarize the estimates from the regression model in the appendix (Table 37).

Table 19: Pairwise contrasts using posterior estimates of the coefficients for each community in the binomial regression and their 89% credible intervals.

Contrast	**Mean**	**SD**	**CI lower (5.50%)**	**CI upper (94.50%)**
$\boldsymbol{\beta}$[Kanyawara]- $\boldsymbol{\beta}$[Kasekela]	0.121555	0.365922	-0.42591	0.684317

$\boldsymbol{\beta}$[Kanyawara]-$\boldsymbol{\beta}$[Mitumba]	-0.17545	0.374274	-0.74712	0.415285
$\boldsymbol{\beta}$[Kasekela]-$\boldsymbol{\beta}$[Mitumba]	-0.297	0.403797	-0.93948	0.329551

3.4.3 Within-individual variation in NLPs over an individual's lifespan

In general, I observed that the proportion of NLPs reduces from adolescence till after adulthood (~ages < 25), stays relatively stable after that (~25 < ages < 35), and increases during late adulthood (~ages > 35) (Figure 16). This pattern broadly applies to specific NLPs but is more pronounced in deterministic chaos (Figure 42) and subharmonics (Figure 43) as compared to frequency jumps (Figure 44) and biphonation (Figure 45). This again suggests a non-linear relationship between age and prevalence of NLPs. Hence, I compared linear, quadratic, cubic, and 4-th degree polynomial relationships of proportion of total scream duration that consists of NLPs with age using WAIC. I summarize the mean percent of climax-scream duration that consists of NLPs for each individual in Table 20 and for each individual at different ages in the appendix Table 38.

Table 20: Mean percentages of the climax-scream duration that contains specific NLPs (subharmonics (SH), biphonation (BP), deterministic chaos(CH)), mean number of frequency jumps (FJ) per climax-scream, and mean percent of scream duration consisting of all NLPs (subharmonics + biphonation + deterministic chaos) and their ranges for each individual.

Community	*Caller*	*Number of calls*	*Mean Age*	*Mean # of FJs per climax scream (Range)*	*Mean % of climax screams duration that is SH (Range)*	*Mean % of climax screams duration that is BP (Range)*	*Mean % of climax screams duration that is CH (Range)*	*Mean % of climax screams duration that is NLP (Range)*
Kanyawara	AJ	16	30	1 (0,2)	12 (0,51)	4 (0,31)	7 (0,36)	22 (0,51)
Kanyawara	BB	12	40	0 (0,2)	9 (0,26)	16 (0,100)	5 (0,31)	30 (0,100)
Kanyawara	BF	15	26	1 (0,2)	6 (0,42)	10 (0,42)	6 (0,29)	23 (0,64)
Kanyawara	KK	10	25	0 (0,2)	0 (0,5)	7 (0,26)	23 (0,67)	31 (0,67)
Kanyawara	LK	16	24	0 (0,1)	15 (0,67)	1 (0,18)	24 (0,72)	41 (0,84)
Kanyawara	MS	27	17	0 (0,1)	5 (0,40)	11 (0,55)	5 (0,61)	20 (0,86)
Kanyawara	ST	19	49	0 (0,1)	4 (0,22)	15 (0,60)	17 (0,100)	36 (0,100)
Kanyawara	SY	23	28	0 (0,1)	3 (0,12)	8 (0,28)	5 (0,31)	16 (0,50)
Kanyawara	TU	10	37	0 (0,1)	0 (0,0)	19 (0,62)	8 (0,32)	26 (0,62)
Kanyawara	YB	8	34	0 (0,2)	0 (0,0)	10 (0,44)	4 (0,23)	13 (0,55)
Kasekela	FD	16	27	0 (0,1)	8 (0,25)	12 (0,59)	3 (0,17)	23 (0,59)
Kasekela	FND	11	17	0 (0,2)	2 (0,12)	36 (0,58)	9 (0,23)	47 (0,70)
Kasekela	FO	20	27	0 (0,2)	3 (0,24)	1 (0,10)	3 (0,18)	7 (0,28)
Kasekela	FU	34	20	0 (0,1)	2 (0,18)	2 (0,27)	3 (0,43)	7 (0,43)
Kasekela	GIM	3	14	0 (0,1)	27 (0,49)	3 (0,9)	25 (11,42)	55 (11,84)
Kasekela	SL	15	33	1 (0,2)	4 (0,17)	0 (0,0)	24 (0,100)	28 (0,100)

Kasekela	SN	35	20	1 (0,3)	2 (0,47)	11 (0,48)	12 (0,81)	26 (0,85)
Kasekela	ZS	14	21	1 (0,2)	10 (0,52)	0 (0,0)	4 (0,20)	14 (0,52)
Mitumba	EDG	21	28	1 (0,2)	3 (0,27)	11 (0,55)	6 (0,57)	20 (0,57)
Mitumba	FAN	6	15	1 (0,2)	34 (0,53)	2 (0,12)	18 (0,39)	54 (24,83)
Mitumba	KOC	13	16	1 (0,3)	5 (0,25)	2 (0,22)	14 (0,36)	20 (0,45)
Mitumba	LAM	8	15	0 (0,1)	0 (0,0)	22 (0,56)	0 (0,0)	22 (0,56)
Mitumba	LON	7	16	1 (0,2)	17 (0,47)	11 (0,42)	2 (0,15)	30 (0,60)

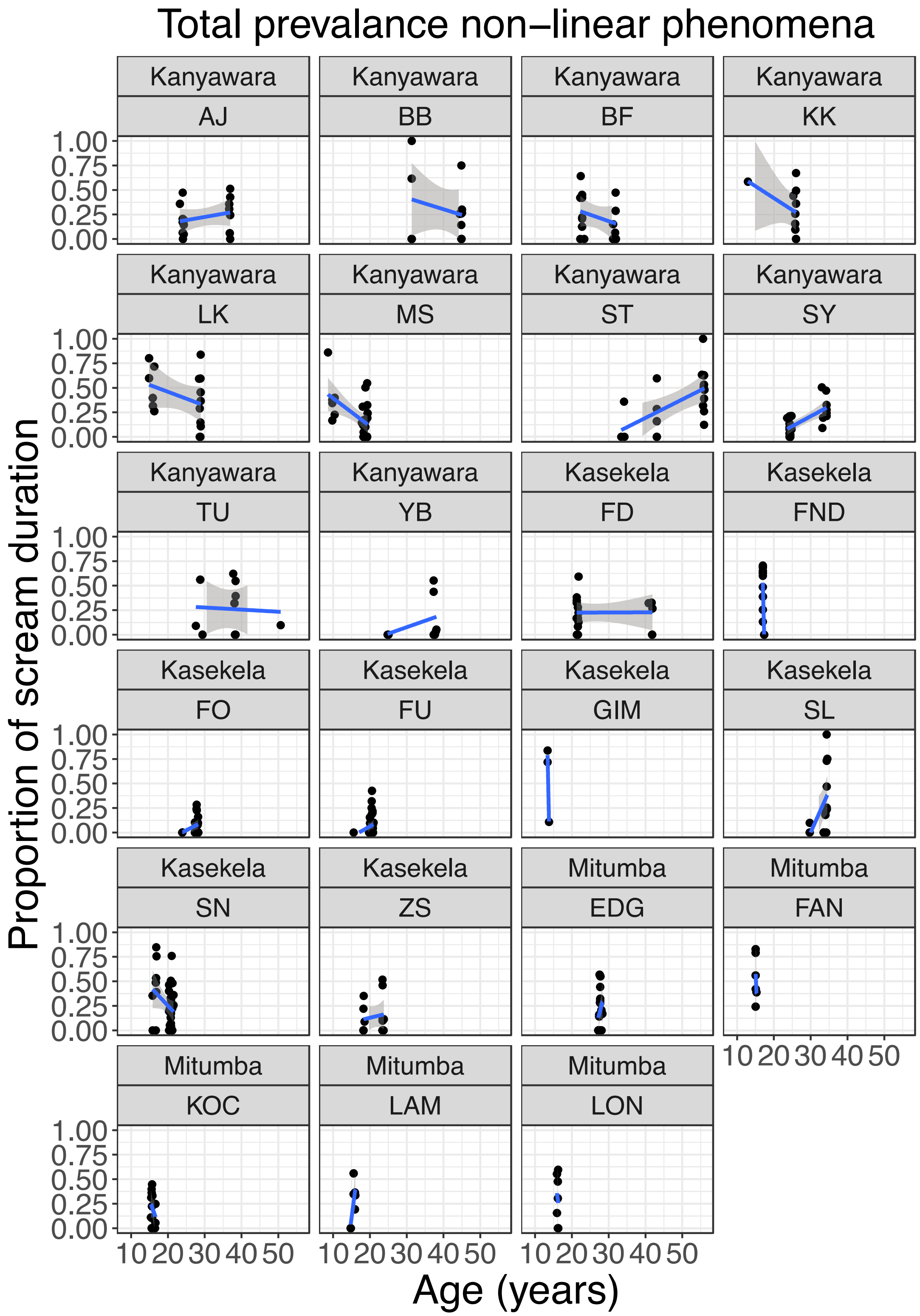


Figure 16: Within-individual longitudinal variation in proportion of total climax-scream duration that consists of NLPs (subharmonics + biphonation + deterministic chaos).

3.4.4 Association of the proportion of NLPs with individual traits

3.4.4.1 Proportion of NLPs as a function of age

Consistent with the pattern observed above, the polynomial regressions suggest a non-linear relationship between age and the proportion of NLPs in climax-screams. The proportion of NLPs in climax-screams tends to be relatively higher in adolescent and old individuals as compared to adults in their prime mating years. While a crude comparison of WAICs suggests that the 4th degree polynomial was the best (Table 39), this might be an outcome of adding terms and overfitting. Hence, I infer the best model based on the standard error of the difference in the WAIC values between a pair of models (the light gray lines in Figure 17). Based on that, the quadratic polynomial was best supported as it represents a significant improvement over the linear model (their **Δ**WAIC lies outside of the SE of **Δ**WAIC). In contrast, the cubic and 4th degree polynomials have WAICs that lie within the standard error of the difference in WAICs (Figure 17).

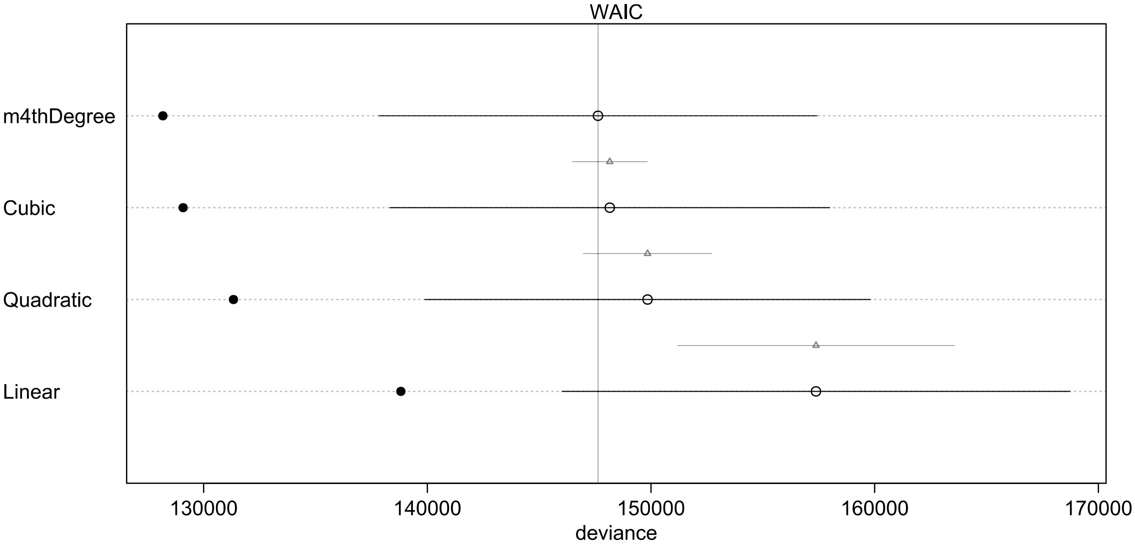


Figure 17: Comparison of four binomial polynomial regression models assessing the relationship of age and proportion of NLPs in climax-screams. The filled black dots represent in-sample deviance. The open dots represent the WAIC values, which are a theoretical estimate of the out-of-sample deviance. The black line segments represent the standard error of the WAIC values. The gray line segments represent the standard error of the difference in the WAIC values.

I report the estimates from the best (quadratic) model in Table 21, and the posterior predictions of the relationship between age and proportion of NLPs in Figure 18. The best supported quadratic model was an upward-facing parabola (as $\boldsymbol{\beta}_2 > 0$; Table 21, Figure 18). I include the estimates from other models and their posterior predictions in the appendix (Table 40, Table 41, and Table 42 and Figure 46, Figure 47, and Figure 48). The general pattern observed in the best model (high prevalence in adolescents and old individuals and low prevalence in adult males in their peak mating years) is reflected in all non-linear polynomials (Figure 46, Figure 47, and Figure 48), bolstering the confidence in the pattern.

Table 21: Parameter estimates from the quadratic polynomial model assessing the relationship of age and proportion of NLPs in climax-screams, their standard deviations (SD), 89% highest posterior density intervals (HPDI), the effective sample size (N_eff), and the Gelman-Rubin convergence diagnostic (Rhat).

Coefficient	Mean	SD	HPDI lower (5.50%)	HPDI upper (94.50%)	N_eff	Rhat
β_1	-0.34	0.01	-0.35	-0.33	1048.79	1
β_2	0.3	0	0.29	0.3	1315.14	1
α[AJ]	-1.38	0.01	-1.4	-1.36	3740.9	1
α[BB]	-1.09	0.02	-1.12	-1.06	2034.04	1
α[BF]	-1.07	0.01	-1.09	-1.05	4616.43	1
α[EDG]	-1.46	0.02	-1.49	-1.44	4218.38	1
α[FAN]	-0.56	0.02	-0.6	-0.52	2052.24	1
α[FD]	-1.51	0.02	-1.53	-1.48	4607.18	1
α[FND]	-0.79	0.02	-0.81	-0.76	2023.65	1
α[FO]	-2.44	0.02	-2.47	-2.41	4431.03	1
α[FU]	-2.96	0.02	-2.99	-2.93	2838.06	1
α[GIM]	-0.43	0.04	-0.49	-0.36	2188.23	1
α[KK]	-0.97	0.02	-1	-0.93	3957.96	1
α[KOC]	-1.85	0.02	-1.88	-1.82	2142.38	1
α[LAM]	-1.96	0.03	-2	-1.92	2152.44	1
α[LK]	-0.76	0.02	-0.78	-0.73	3437.05	1
α[LON]	-1.48	0.03	-1.53	-1.44	2908.81	1
α[MS]	-1.92	0.02	-1.94	-1.89	1563.64	1
α[SL]	-1.34	0.02	-1.38	-1.31	3313.66	1
α[SN]	-1.29	0.01	-1.3	-1.27	1932.59	1

𝜶[ST]	-2.01	0.03	-2.05	-1.96	1816.42	1
𝜶[SY]	-1.66	0.01	-1.68	-1.64	5038.16	1
𝜶[TU]	-1.58	0.02	-1.62	-1.54	2897.08	1
𝜶[YB]	-2.45	0.03	-2.5	-2.4	3175.08	1
𝜶[ZS]	-2.26	0.02	-2.3	-2.23	3608.54	1
$\bar{\alpha}$	-1.52	0.15	-1.74	-1.29	3192.99	1
σ	0.67	0.11	0.52	0.88	2602.27	1

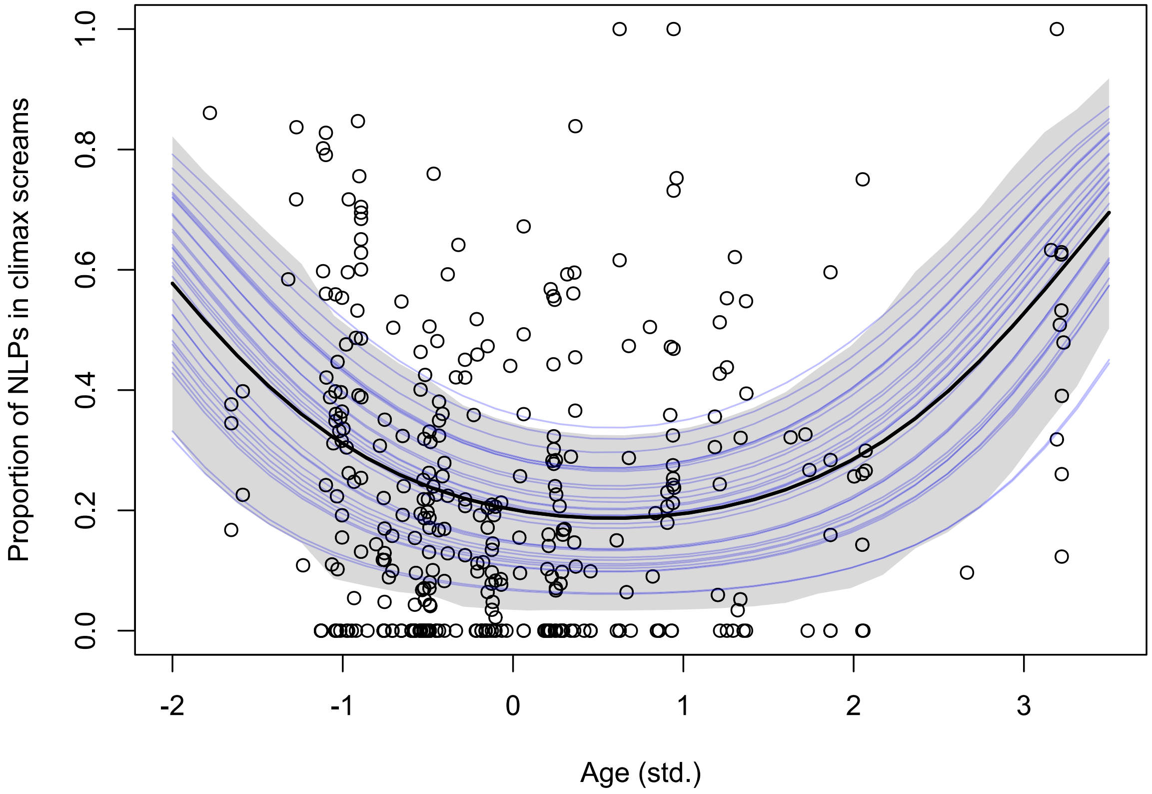


Figure 18: Posterior predictions simulated from the estimated joint posterior distribution from the quadratic model fitting the relationship of the proportion of NLPs in climax-screams to age (standardized). Dots represent raw data, the solid black line represents the predicted average chimpanzee, blue lines represent 30 random chimpanzees simulated from the posterior, and the gray shade represents the 89% HPDI of the population of chimpanzees.

When considering specific NLPs separately, the proportion of deterministic chaos and subharmonics showed a similar pattern, whereas the proportion of biphonation was unrelated to age. I report the model comparison and posterior predictions from the best model in each case in the appendix (Figure 49-54).

3.4.4.2 Proportion of NLPs as a function of rank

I found that the linear model better explained the relationship of the proportion of NLPs with rank as compared to the quadratic model (Figure 19). I specify the estimates from the linear model in Table 22. After accounting for the quadratic effect of age, rank had a positive relationship with the proportion of NLPs ($\boldsymbol{\beta}$ (Rank) = 0.07, SD = 0.03, 89% HPDI = [0.03, 0.11]).

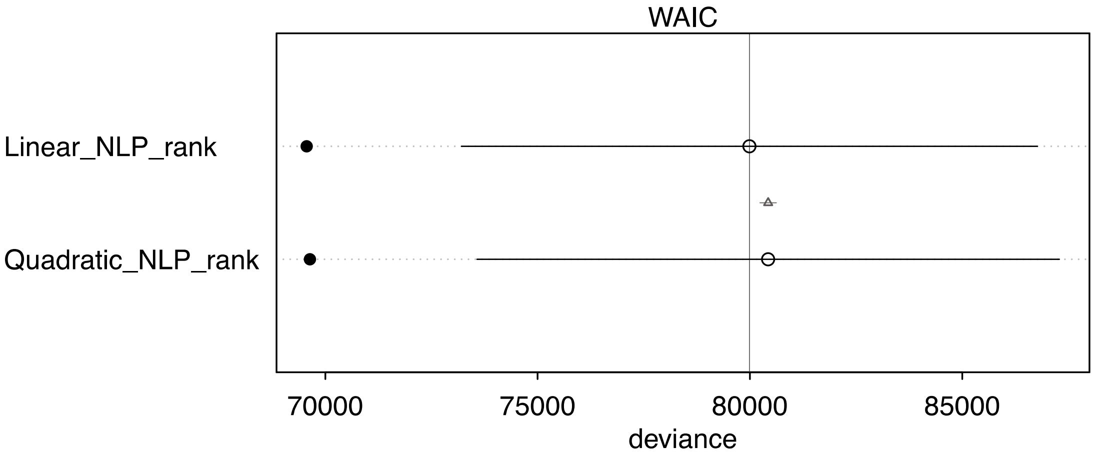


Figure 19: Comparison of linear and quadratic regression models assessing the relationship of rank and proportion of NLPs in climax-screams. The filled black dots represent in-sample deviance. The open dots represent the WAIC values, which are a theoretical estimate of the out-of-sample deviance. The black line segments represent the

standard error of the WAIC values. The gray line segments represent the standard error of the difference in the WAIC values.

Table 22: Parameter estimates from the linear model assessing the relationship of rank and proportion of NLPs in climax-screams, their standard deviations (SD), 89% highest posterior density intervals (HPDI), the effective sample size (N_eff), and the Gelman-Rubin convergence diagnostic (Rhat).

Coefficient	**Mean**	**SD**	**HPDI lower (5.50%)**	**HPDI upper (94.50%)**	**N_eff**	**Rhat**
β (Rank)	0.07	0.03	0.03	0.11	313.87	1.02
β (Age)	-0.7	0.03	-0.75	-0.64	301.51	1.02
β (Age^2)	0.26	0.01	0.25	0.28	305.1	1.02
α[EDG]	-1.18	0.02	-1.21	-1.14	820.83	1.01
α[FAN]	-0.96	0.04	-1.03	-0.9	440.61	1.01
α[FD]	-1.55	0.03	-1.61	-1.5	382.92	1.02
α[FND]	-1.08	0.05	-1.16	-1	309.41	1.02
α[FO]	-2.14	0.02	-2.18	-2.11	1027.42	1
α[FU]	-3	0.04	-3.05	-2.94	345.42	1.02
α[GIM]	-1.03	0.06	-1.12	-0.93	439.81	1.01
α[KOC]	-2.15	0.03	-2.21	-2.1	513.13	1.01
α[LAM]	-2.37	0.05	-2.45	-2.29	363.54	1.02
α[LON]	-1.61	0.06	-1.7	-1.52	471.84	1.02
α[SL]	-1.08	0.03	-1.13	-1.03	692.2	1
α[SN]	-1.28	0.01	-1.3	-1.26	877.32	1.01
α[ZS]	-2.2	0.03	-2.25	-2.15	561.66	1.01
$\bar{\alpha}$	-1.63	0.2	-1.94	-1.33	1140.92	1

σ	0.7	0.15	0.5	0.96	1393.19	1

3.4.4.3 Proportion of NLPs as a function of SIVcpz status

I found no difference in the proportion of NLPs between SIVcpz positive and SIVcpz negative chimpanzees after accounting for the quadratic effect of age ($\boldsymbol{\beta}$ (SIVcpz Positive) = -0.18, SE = 0.41, 95% CI = [-0.98, 0.62], Table 23).

Table 23: Coefficients, their standard errors, and their 95% confidence intervals (CI) from the binomial regression assessing the relationship of SIVcpz status and the proportion of NLPs in climax screams.

	Estimate ($\boldsymbol{\beta}$)	**Std. Error**	**CI (2.5%)**	**CI (97.5%)**
Intercept	-1.59	0.2	-1.98	-1.2
SIVcpz (Positive)	-0.18	0.41	-0.98	0.62
Age	-0.62	0.02	-0.65	-0.58
Age^2	0.24	0.01	0.22	0.25

3.4.4.4 Proportion of NLPs as a function of mating success

I found that while the quadratic model had slightly lower WAIC values, they were not significantly different than the linear model in explaining the relationship of the proportion of NLPs with mating success (Figure 20). Hence, conservatively, the linear model is the better model to explain the relationship. I specify the estimates from the linear model in Table 24. After accounting for the quadratic effect of age, and the linear effect of rank as observed in previous sections, mating success had a negative relationship with the proportion of NLPs ($\boldsymbol{\beta}$ (Mating) = -0.42, SD = 0.02, 89% HPDI = [-

0.45, -0.39]). This suggests that individuals exhibiting lower proportion of NLPs had higher mating success while keeping their age and rank constant.

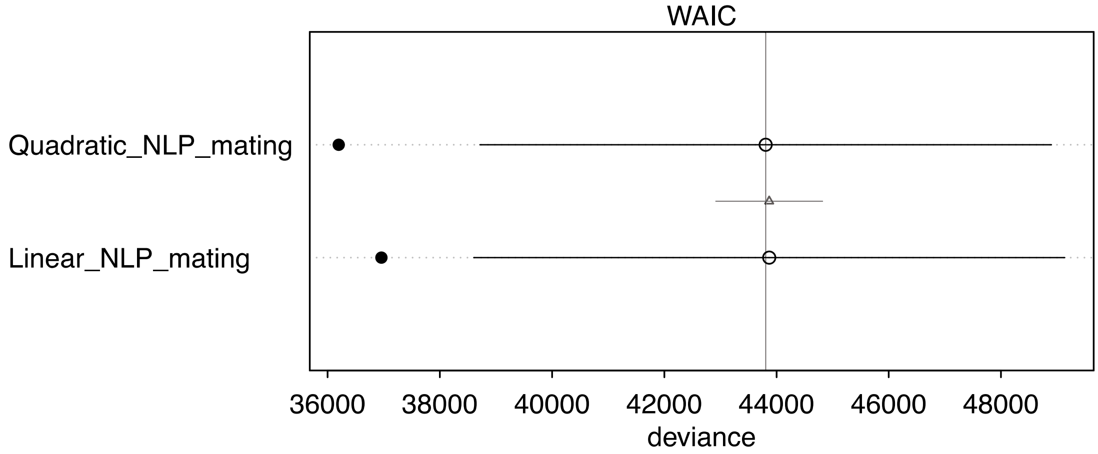


Figure 20: Comparison of linear and quadratic binomial regression models assessing the relationship of mating success and proportion of NLPs in climax-screams. The filled black dots represent in-sample deviance. The open dots represent the WAIC values, which are a theoretical estimate of the out-of-sample deviance. The black line segments represent the standard error of the WAIC values. The gray line segments represent the standard error of the difference in the WAIC values.

Table 24: Parameter estimates from the linear binomial regression model assessing the relationship of mating success and proportion of NLPs in climax-screams, their standard deviations (SD), 89% highest posterior density intervals (HPDI), the effective sample size (N_eff), and the Gelman-Rubin convergence diagnostic (Rhat).

Coefficient	**Mean**	**SD**	**HPDI lower (5.50%)**	**HPDI upper (94.50%)**	**N_eff**	**Rhat**
β (Mating)	-0.42	0.02	-0.45	-0.39	1420.26	1
β (Rank)	-0.14	0.03	-0.19	-0.09	1029.4	1
β (Age)	-0.92	0.13	-1.13	-0.71	677.72	1.01
β (Age^2)	3.47	0.08	3.34	3.59	938.81	1
α[EDG]	-12.29	3.43	-18.86	-8.29	688.05	1

α[FD]	-1.54	0.05	-1.62	-1.47	732.78	1.01
α[FND]	-6.34	0.19	-6.64	-6.05	741.91	1.01
α[FO]	-6.15	0.18	-6.44	-5.86	803.52	1.01
α[FU]	-3.43	0.07	-3.54	-3.32	661.25	1.01
α[GIM]	-14.71	0.38	-15.33	-14.12	892.68	1
α[SL]	-20.66	0.6	-21.61	-19.72	883.87	1
α[SN]	-2.64	0.07	-2.75	-2.55	855.3	1
α[ZS]	-2.28	0.07	-2.39	-2.17	815.43	1.01
$\bar{\alpha}$	-2.49	1.32	-4.56	-0.32	1910.6	1
σ	6.79	1.37	4.89	9.14	1517.74	1

3.4.4.5 Summary of the observed associations of the proportion of NLPs with individual traits

I plotted the raw data with OLS linear or quadratic smoothing functions that correspond to the observed associations of the proportion of NLPs in climax screams with different individual traits in the raw data in Figure 21. Age had a statistically significant concave-up, quadratic relationship with the proportion of NLPs (Figure 21 (a)). While rank appears to have a negative relationship with the proportion of NLPs (Figure 21 (b)), I observed a significant positive relationship after controlling for the quadratic effect of age. SIVcpz positive individuals seem to have a lower proportion of

NLPs but the difference was not statistically significant (Figure 21 (c)). Lastly, Mating success had a significant negative linear relationship with the proportion of NLPs, meaning that the individuals that exhibited a higher proportion of NLPs had lower mating success (Figure 21 (d)).

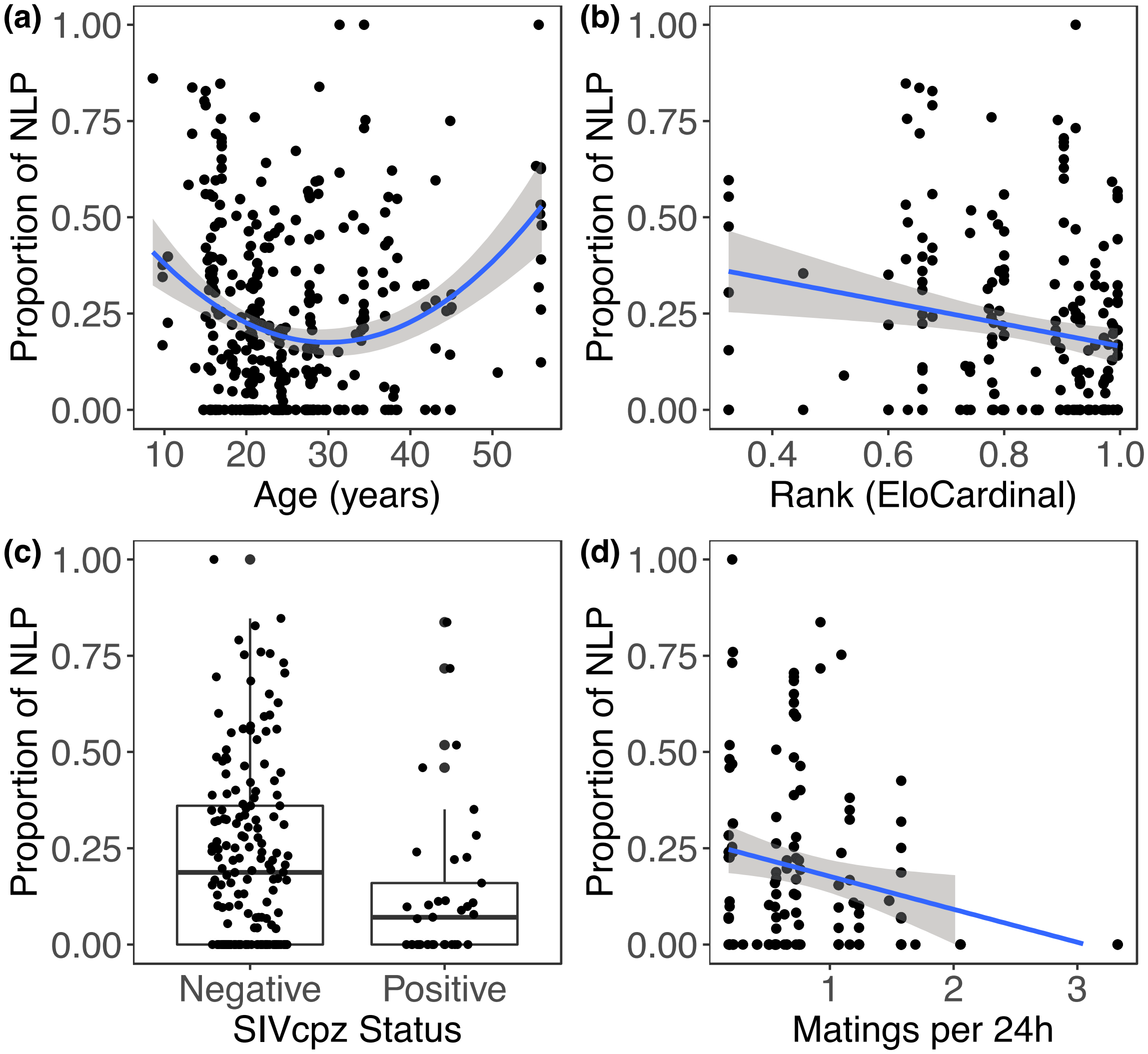


Figure 21: Association of the proportion of NLPs with individual traits including (a) age, (b) rank, (c) SIVcpz status, and (d) mating success.

3.4.5 Association of the F0max with individual traits

3.4.5.1 F0max as a function of age

I found no association of F0max with age either in the quadratic or the linear model as the 95% confidence of all coefficients of age included 0 (Table 25).

Table 25: Coefficients, their standard errors, and their 95% confidence intervals (CI) from the linear mixed effects models assessing the quadratic and linear relationships between F0max and age.

Quadratic Model				
	Estimate (β)	**Std. Error**	**CI (2.5%)**	**CI (97.5%)**
Intercept	1250.98	60.24	1136.71	1365.59
Age	-35.51	58.2	-145.85	74.87
Age^2	-8.06	28.55	-62.22	46.09
Linear Model				
Intercept	1240.59	46.91	1149.16	1332.69
Age	-47.86	37.71	-121.29	25.75

3.4.5.2 F0max as a function of rank

I found no association of F0max with rank either in the quadratic or the linear model as the 95% confidence of all coefficients of rank included 0 (Table 26).

Table 26: Coefficients, their standard errors, and their and their 95% confidence intervals (CI) from the linear mixed effects models assessing the quadratic and linear relationships between F0max and rank.

Quadratic Model				
	Estimate (β)	**Std. Error**	**CI (2.5%)**	**CI (97.5%)**
Intercept	1226.09	48.79	1132.85	1318.88

Rank	-3.98	52.36	-103.43	95.31
Rank^2	7.92	24.38	-38.34	54.2
Linear Model				
Intercept	1233.58	41.78	1150.85	1315.97
Rank	-16.59	32.68	-80.74	46.54

3.4.5.3 F0max as a function of SIVcpz status

I found no association of F0max with age either in the quadratic or the linear model as the 95% confidence of all age coefficients included 0 (Table 27).

Table 27: Coefficients, their standard errors, and their 95% confidence intervals (CI) from the linear mixed effects models assessing the quadratic and linear relationships between F0max and SIVcpz status.

	Estimate (β)	**Std. Error**	**CI (2.5%)**	**CI (97.5%)**
Intercept	1248.95	45.34	1160.01	1336.72
SIVcpz (Positive)	-58.18	99.24	-248.32	140.1

3.4.5.4 F0max as a function of mating success

I found no association of F0max with mating success either in the quadratic or the linear model as the 95% confidence of all coefficients of mating success included 0 (Table 28).

Table 28: Coefficients, their standard errors, and their 95% confidence intervals (CI) from the linear mixed effects models assessing the quadratic and linear relationships between F0max and mating success.

Quadratic Model				
	Estimate (β)	**Std. Error**	**CI (2.5%)**	**CI (97.5%)**
Intercept	1285.7	58.19	1167.38	1395.38

Matings per 24h	26.49	47.71	-66.48	135.95
Matings per $24h^2$	-18.03	17.41	-56.26	16.81
Linear Model				
Intercept	1257.9	57.96	1140.45	1373.86
Matings per 24h	-9.82	34.74	-76.85	57.25

3.4.5.5 Summary of the observed associations of the F0max with individual traits

I plotted the raw data with OLS linear or quadratic smoothing functions that correspond to the observed associations of the F0max in climax screams with different individual traits in the Figure 22. Age appears to have a negative relationship with the F0max (Figure 22 (a)), but the association was not statistically significant. Rank appears to have no relationship with the F0max (Figure 22 (b)) and that was also observed in the non-significant statistical model. I observed no difference in the F0max between SIVcpz positive and negative individuals and the difference was also not statistically significant (Figure 22 (c)). Lastly, Mating success appears to be slightly higher for individuals with higher F0max (Figure 22 (d)), but the association was not statistically significant.

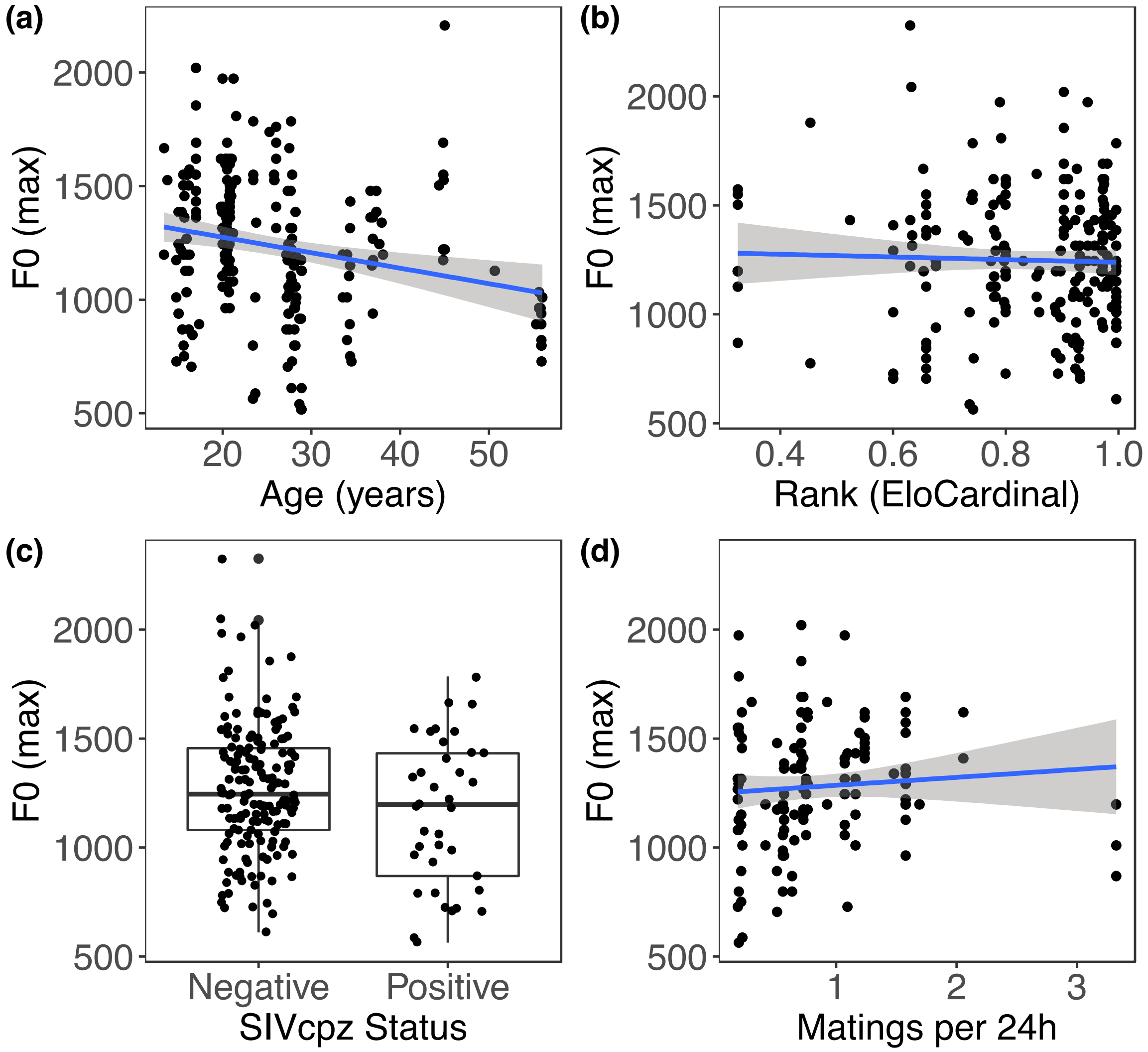


Figure 22: Association of the F0max with individual traits including (a) age, (b) rank, (c) SIVcpz status, and (d) mating success.

3.4.6 Association of the Wiener entropy (noise_max) with individual traits

3.4.6.1 Noise_max as a function of age

The quadratic and linear beta mixed effects regression models assessing the association of noise_max and age resulted in significant quadratic and the linear coefficients of age (Table 29). The quadratic model contained a concave-up parabola ($\boldsymbol{\beta}$ (Age^2) > 0). The linear model revealed a negative association, with meaning that individuals exhibited lower noise as they got older. The model comparison of the quadratic and linear model suggests that the quadratic model was the better model to describe the relationship (Table 30).

Table 29: Coefficients, their standard errors, and their 95% confidence intervals (CI) from the beta mixed effects regression models assessing the quadratic and linear relationships between Wiener entropy and age.

Quadratic Model				
	Estimate ($\boldsymbol{\beta}$)	**Std. Error**	**CI (2.5%)**	**CI (97.5%)**
Intercept	-0.12	0.08	-0.27	0.03
Age	-0.31	0.08	-0.47	-0.16
Age^2	0.09	0.04	0.02	0.17
Linear Model				
Intercept	-0.01	0.07	-0.15	0.12
Age	-0.17	0.06	-0.29	-0.06

Table 30: Comparison of the linear and quadratic beta regressions assessing the relationship of Wiener entropy with age using log likelihood, AICc, and model weights.

	Intercept	**β (Age)**	**β (Age^2)**	**df**	**logLik**	**AICc**	**Δ AICc**	**weight**
Quadratic	-0.12	-0.31	0.09	5	95.3	-180.32	0	0.86
Linear	-0.01	-0.17		4	92.41	-176.62	3.7	0.14

3.4.6.2 Noise_max as a function of rank

After controlling for the quadratic effect of age, neither the quadratic, nor the linear beta mixed effects regression models assessing the association of noise_max and rank resulted in significant coefficients for rank (Table 31), suggesting a lack of relationship between noise_max and rank.

Table 31: Coefficients, their standard errors, and their 95% confidence intervals (CI) from the beta mixed effects regression models assessing the quadratic and linear relationships between Wiener entropy and rank.

Quadratic Model				
	Estimate (β)	**Std. Error**	**CI (2.5%)**	**CI (97.5%)**
Intercept	0.09	0.1	-0.11	0.29
Rank	-0.17	0.09	-0.35	0.01
Rank^2	-0.05	0.04	-0.13	0.03
Age	-0.09	0.1	-0.28	0.1
Age^2	0.11	0.06	-0.01	0.22
Linear Model				
Intercept	-0.01	0.11	-0.21	0.2
Rank	-0.07	0.09	-0.24	0.1
Age	-0.15	0.12	-0.39	0.08
Age^2	0.14	0.06	0.02	0.27

3.4.6.3 Noise_max as a function of SIVcpz status

After controlling for the quadratic effect of age the linear beta mixed effects regression model assessing the difference in noise_max between SIVcpz positive and SIVcpz negative males revealed that SIVcpz positive individuals exhibited greater noise in their climax screams (Table 32; $\boldsymbol{\beta}$ (SIVcpz (Positive)) = 0.29, SE = 0.15, 95% CI = [0, 0.59])

Table 32: Coefficients, their standard errors, and their 95% confidence intervals (CI) from the beta mixed effects regression model assessing the difference in noise_max between SIVcpz positive and SIVcpz negative chimpanzees.

	Estimate ($\boldsymbol{\beta}$)	**Std. Error**	**CI (2.5%)**	**CI (97.5%)**
Intercept	-0.08	0.09	-0.26	0.1
SIVcpz (Positive)	0.29	0.15	0	0.59
Age	-0.23	0.07	-0.37	-0.08
Age^2	0.17	0.05	0.07	0.26

3.4.6.4 Noise_max as a function of mating success

After controlling for the quadratic effect of age, neither the quadratic, nor the linear beta mixed effects regression models assessing the association of noise_max and mating success resulted in significant coefficients for mating success (Table 33), suggesting that noise_max does not predict mating success.

Table 33: Coefficients, their standard errors, and their 95% confidence intervals (CI) from the beta mixed effects regression models assessing the quadratic and linear relationships between Wiener entropy and mating success.

Quadratic Model				
	Estimate (β)	**Std. Error**	**CI (2.5%)**	**CI (97.5%)**
Intercept	-0.05	0.11	-0.26	0.17
Matings per 24h	-0.08	0.1	-0.27	0.12
Matings per 24h^2	0.01	0.04	-0.06	0.08
Age	0.05	0.09	-0.13	0.23
Age^2	0.01	0.05	-0.09	0.11
Linear Model				
Intercept	-0.03	0.09	-0.21	0.16
Matings per 24h	-0.06	0.06	-0.18	0.07
Age	0.06	0.08	-0.1	0.22
Age^2	0.01	0.05	-0.09	0.1

3.4.5.5 Summary of the observed associations of the Wiener entropy with individual traits

I plotted the raw data with OLS linear or quadratic smoothing functions that correspond to the observed associations of the Wiener entropy (noise_max) in the climax screams with different individual traits in the Figure 23. Like the proportion of NLPs, age had a statistically significant concave-up, quadratic relationship with the noise_max. Rank and mating success appear to have negative linear relationships with the noise_max, meaning that the individuals that exhibited a more noise were lower ranking and had lower mating success. However, these relationships were not statistically significant after controlling for the quadratic effect of age. Lastly, I observed that SIVcpz positive individuals exhibited significantly greater noise in their climax screams than SIVcpz negative individuals.

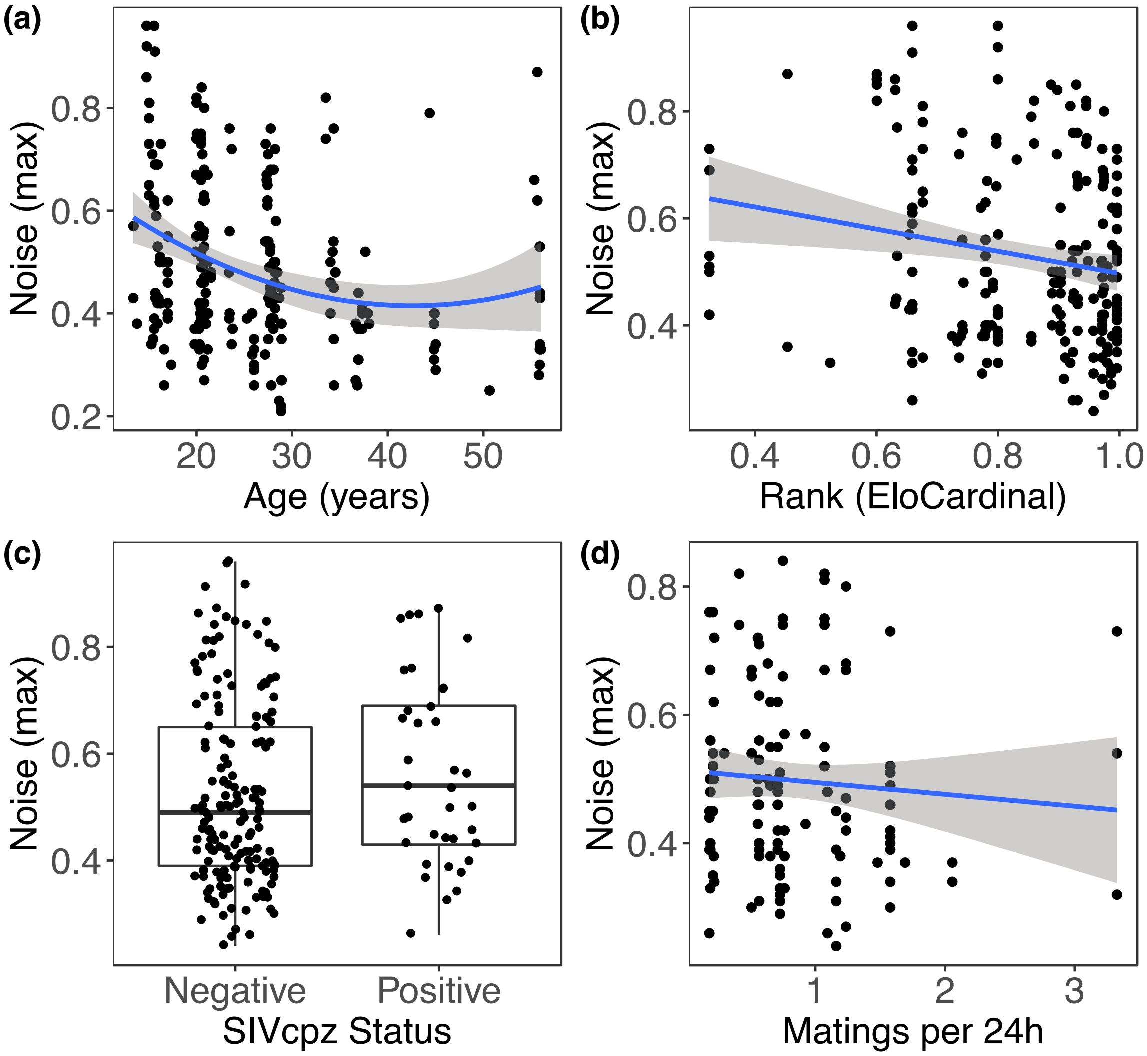


Figure 23: Association of the Wiener entropy (noise_max) with individual traits including (a) age, (b) rank, (c) SIVcpz status, and (d) mating success.

3.5 Discussion

I documented the prevalence of four kinds of non-linear phenomena (NLPs): frequency jumps, subharmonics, biphonation, and deterministic chaos in pant-hoot climax screams. I measured the proportion of total climax-scream duration exhibiting subharmonics, biphonation, and deterministic chaos and tested their relationship with the age of the individual. I found a non-linear relationship between the proportion of climax-scream duration exhibiting NLPs and the age of the individual. The proportion of NLPs was relatively high during adolescence and early adulthood until about the age of 25 years after which the proportion of NLPs remained low until about the age of 35 years beyond which the proportion of NLPs began to increase again. This pattern held for deterministic chaos and subharmonics, but not for biphonation, which showed no relationship with age. Furthermore, I tested for the association of F0max and noise in the climax screams with age. Noise had a non-linear relationship with age that was similar to the above-mentioned relationship of the proportion of NLPs with age, except that the increase in the noise in later years wasn’t as stark as the increase in the proportion of NLPs. F0max appeared to reduce with age but did not have a statistically significant association with age. Overall, these findings reveal within-individual changes in the longitudinal dataset, demonstrating that chimpanzee vocalizations have a developmental trajectory and do not remain fixed after maturity. Next, I tested if these acoustic features reflected other male quality traits such as rank and health and if they predicted male mating success. I found that F0max was not associated with rank, and after controlling for the effect of age, noise had a positive relationship with rank. In terms of health, SIVcpz positive individuals exhibited greater noise in the climax screams and no other acoustic feature predicted SIVcpz status. Lastly, I observed that the individuals with lower noise in their calls and higher F0max appeared to have greater mating success, but the associations were not statistically significant. However, males with lower proportion of NLPs had statistically significantly greater mating success, providing strong evidence for the sexual selection hypothesis.

NLPs occurred at higher rates in the dataset than reported previously–in 85% of the calls compared to 52% reported previously (Riede et al., 2004). In the light of current findings, this might reflect the difference in the age distribution of the individuals included in the samples. The dataset included more individuals (N = 23, compared to N = 12 in (Riede et al., 2004, 2007)) and more samples from young and old individuals that are likely to have calls with more NLPs (N = 17 individuals under the age of 25 in our dataset compared to N = 8 individuals in Riede et al., (2004); and N = 6 individuals over the age of 35 compared to N = 2 individuals in Riede et al., (2004)). This bolsters the confidence in my observed pattern–the non-linear relationship with age. Frequency jumps were the most commonly occurring NLP, whereas subharmonics, biphonation, and deterministic chaos were prevalent in similar proportions (Table 18). This again is contrary to Riede et al., (2004)'s finding of biphonation being the most common NLP. This might also be an outcome of the difference in the age distribution of the individuals included in the study. With more young and old individuals included in the sample, we observed no relationship between the proportion of biphonation and age. Hence, Riede et al. 's finding may represent the pattern observed in adults (between ages ~25 and ~35) as middle-aged individuals appear to have a relatively lower proportion of subharmonics and deterministic chaos and a higher proportion of biphonation. A higher relative proportion of biphonation in middle-aged individuals compared to younger and older individuals also occurs in North Atlantic right whales (*Eubalaena glacialis*) (Root-Gutteridge et al., 2018).

These results challenge the notion that non-human primates and mammals have stereotyped vocalizations beyond physical maturity (Hammerschmidt et al., 2000). Chimpanzee males reach social and sexual maturity–when they begin to show a reduction in juvenile behaviors and begin siring offspring–at around the age of 12 (Pusey, 1990). Furthermore, they reach close to their adult body mass by age 14 (*ibid.*). I observed a reduction in the proportion of NLPs until around the age of 25, which is well beyond physical maturity. Furthermore, I observed changes in late adulthood after around the age of 35. The reduction in proportions of NLPs from adolescence to early adulthood resembles the pattern observed in sperm whale clicks wherein younger individuals exhibited more noisy components than older individuals (Watkins et al., 1988) and in Siberian and Neva Masquerade cats wherein younger cats exhibited greater occurrence of frequency jumps and biphonation (Magiera et al., 2020). Furthermore, the observed non-linear relationship with age resembles the observed pattern in North Atlantic right whales (Root-Gutteridge et al., 2018). Root-Gutteridge and colleagues observed a non-linear relationship of deterministic chaos with age wherein it first reduced with increasing age before increasing again in late adulthood (*ibid.*). However, opposite to our findings, they observed an increase in subharmonics with age that appeared to reduce in older adults. Similarly, for biphonation, they observed an initial increase with age and subsequent reduction in older adults as opposed to no relationship of biphonation with age in our data. These differences might simply reflect idiosyncratic species differences or might reflect different functional roles that these different NLPs play in whale and chimpanzee vocal communication. For instance, researchers have reported that noisy vocalizations, and hence, potentially deterministic chaos, may be related to stress levels or arousal in some species (Manteuffel et al., 2004; Stoeger et al., 2011; Viljoen et al., 2015), but body size in others (Serrano et al., 2020). Similarly, subharmonics may function to convey aggression in the territorial calls of spotted kiwis (Digby et al., 2014) or function to increase the auditory impact of calls in some other species (Charlton et al., 2017; Digby et al., 2014; Townsend & Manser, 2011).

Researchers have proposed several hypotheses to explain the function of NLPs in animal vocalizations: (i) individual recognition hypothesis: NLPs may function to facilitate individual identification by increasing vocal distinctiveness (Volodina et al., 2006; Wilden et al., 1998). Evidence for this hypothesis remains mixed. Several studies have reported that NLPs facilitate individual distinctiveness (e.g. biphonation facilitates individual recognition in dhole (*Cuon alpinus*) (Volodina et al., 2006), and NLPs facilitate vocal distinctiveness in Holstein-Friesian heifers (Green et al., 2019), Darwin's frog (*Rhinoderma darwinii*) (Serrano et al., 2020) and Altai pika (*Ochotona alpina*) (Volodin et al., 2018). However, some other studies have not found such evidence (e.g. NLPs did not provide evidence for individual distinctiveness kiwi calls (Digby et al., 2014), and only moderately predicted individual distinctiveness in infant African elephant roars (Stoeger et al., 2011)). (ii) auditory impact hypothesis: NLPs in calls may increase the auditory impact on call receivers as calls with NLPs are unpredictable and hard to ignore (Fitch et al., 2002). Many studies support this possibility. For example, subharmonics increase the auditory impact of spotted kiwi calls in territorial contexts (Digby et al., 2014), of female koala rejection calls in mating contexts (Charlton et al., 2017), and meerkat alarm calls (Townsend & Manser, 2011). However, Riede and colleagues argued that the patterns of NLPs observed in chimpanzee pant-hoot climax-screams were not consistent with the auditory impact hypothesis (Riede et al., 2007). This is because the auditory impact hypothesis predicts that individuals should increase the prevalence of NLPs to increase their auditory impact, and chimpanzee pant-hoot climax-screams exhibit a lower proportion of NLPs compared to other screams that chimpanzees produce. While the relatively lower proportion of NLPs in climax-screams compared to other screams does not support the auditory impact hypothesis, given the high absolute prevalence of NLPs in my dataset, it is possible that NLPs increase the auditory impact of pant-hoots. Furthermore, I observed that after controlling for age, higher ranking individuals had a greater proportion of NLPs. This suggests that higher ranking individuals may be able to increase the impact of their calls with NLPs to attract mates, allies, and to sustain their positions in the hierarchy. Nevertheless, I still observed chimpanzees exhibiting a relatively lower proportion of NLPs in middle-age, supporting Riede and colleagues' argument. (iii) vocalizer condition hypothesis: NLPs may signal the physical condition of the caller as NLPs are an outcome of unstable vibrations of the vocal folds, which may be an outcome of pathological conditions of the caller (Fitch et al., 2002; Riede et al., 1997). Studies support the possibility that vocalizations are affected by the pathological condition of the caller. For instance, reports indicate that human voice is affected by fatigue (Welham & Maclagan, 2003), age (Hacki & Heitmüller, 1999; Heylen et al., 1998), and disease (Heylen et al., 1998; Ikeda et al., 1999). Furthermore, NLPs in voice may be a useful diagnostic tool for detecting pathological conditions not only in humans (Herzel et al., 1994), but also in non-human animals (Riede & Stolle-Malorny, 1999). Pattern of deterministic chaos observed in North Atlantic right whales is also consistent with this hypothesis (Root-Gutteridge et al., 2018). Riede and colleagues found that the prevalence of NLPs in chimpanzee pant-hoot climax-screams was associated with their F0 maxima, suggesting that the individuals producing the climax-screams were vocalizing at the peak capacity of their vocal folds (Riede et al., 2007). Hence, producing high-pitch climax-screams with minimal NLPs might indicate physical vigor and is consistent with the vocalizer condition hypothesis. My results provide further strong support for the vocalizer condition hypothesis as chimpanzee males in their peak adult years were able to produce climax-screams with fewer NLPs, less noise and were able to achieve greater mating success. Furthermore, I found that SIVcpz positive individuals that often suffer various pathological conditions including $CD4^+$ T-cell depletion, higher death rates, and lower fertility (Keele et al., 2009) exhibited greater noise in their vocalizations. Nevertheless, this hypothesis still remains unexplored in most species and the strong evidence in this study provides a promising avenue for future studies, especially in species where calls may be sexually selected. Collectively, my results strongly support the vocalizer condition hypothesis and provide some support for the auditory impact hypothesis.

My observations of the (non-statistically significant) pattern of higher mating success of individuals exhibiting a higher F0max and lower noise, along with the finding of a statistically significant relationship of higher mating success of individuals exhibiting lower proportion of NLPs provides evidence of female choice. The extent of female choice in chimpanzees remains unclear with studies reporting both the existence (Matsumoto-Oda, 1999; Pieta, 2008; Stumpf & Boesch, 2006, 2010) and the lack of female choice in chimpanzees (Feldblum et al., 2014; Muller et al., 2011). I fitted a quadratic model for predicting mating success from acoustic features to test for the possibility that we may see a linear increase or decrease in mating success for a given acoustic feature due to female choice, but there might be sexually coercive males who may be able to achieve higher mating success irrespective of their acoustic features (Feldblum et al., 2014). Since I found that the linear model was better than the quadratic model for the association of the proportion of NLPs to mating success, it provides more support for female choice over male coercion. Furthermore, I found that the proportion of NLPs predicted mating success above and beyond male age and rank, supporting Stumpf & Boesch's and Pieta's findings that females may be able to avert unwanted matings despite male solicitation and higher dominance of males, and strategically mate with preferred males when conception is most likely (Pieta, 2008; Stumpf & Boesch, 2006). My findings demonstrate that studying vocalizations may be another promising avenue for future studies of sexual selection and mate choice in primates.

The consistencies of my results with previous studies (Magiera et al., 2020; Riede et al., 2007; Root-Gutteridge et al., 2018; Watkins et al., 1988) bolster the confidence in the observed findings. However, certain limitations still remain in the study design. While my sample includes more individuals than any previous similar study and includes multiple individuals with longitudinal samples, the number of adolescents and old aged individuals are limited. Hence, the possibility exists that the pattern I observed is an outcome of the relatively lower sample sizes for individuals in these age groups. Future studies should attempt to include more individuals, especially from the young and old age groups. Next, I measured the NLPs via visual inspection and labeling of spectrograms. While this is unlikely to cause any identification issues for frequency jumps, subharmonics, and biphonation, there is a possibility that some of the deterministic chaos instances were inaccurately detected. Deterministic chaos resembles noise in the spectrograms and might require special software such as TISEAN (Hegger et al., 1999) to detect separately from noise (Tyson et al., 2007). Hence, I might have overestimated the occurrence of deterministic chaos in some cases. However, it is worth noting that TISEAN might also overestimate the occurrence of deterministic chaos (Tyson et al., 2007). My analysis that included 3 NLPs (subharmonics, biphonation, and deterministic chaos) together should be robust and reliable. Lastly, I only looked at high-frequency elements in the pant-hoots—the climax screams, and the low-frequency elements such as the build-ups and letdowns may include information about mate quality as well, especially of body size, as it has been found to be associated with lower frequency in many species. Future studies should attempt to include acoustic features from these low-frequency elements.

Dissertation Conclusion

Language makes humans profoundly different from all other species. Questions about how and why language evolved remain largely unanswered, but recent work focused on a comparative approach has begun to shed light on some key areas (Fitch, 2010). As one of the two living species most closely related to humans, chimpanzees provide valuable comparative data for the study of the evolution of human vocal communication. This phylogenetic proximity has led several studies to explore language-like features in chimpanzee vocalizations. Despite substantial efforts to find language parallels in chimpanzee vocal communication, the evidence for such parallels remains slim. Furthermore, this approach presents conceptual concerns as it involves employing complex linguistic phenomena as models for arguably much simpler communication mechanisms of primate vocal communication (Owren & Rendall, 2001), whereas, scientific progress relies on using simpler models to explain complex phenomena. One of the best candidates for a language parallel in chimpanzee vocal communication concerns dialects, which potentially represent learned signals of community membership (Crockford et al., 2004; Marshall et al., 1999; Mitani et al., 1992). However, several alternative explanations that are closer to basic evolutionary principles, such as the variation in acoustic structure reflecting basic individual traits such as age, rank, and health, are only getting attention recently (Fedurek, Slocombe, et al., 2016; Fedurek, Zuberbühler, et al., 2016). In this dissertation, I attempted to study chimpanzee vocal communication using a more basic biological framework that obtains insights by studying the intimate connections of form and function. I explored new methodologies with a goal of finding ways to consistently describe the ‘form’ of vocalizations, and leveraged the resulting methodological insights to study the connection of form to function. Using audio recordings and behavioral data from two chimpanzee communities in Gombe National Park, Tanzania, and one chimpanzee community from Kibale National Park, Uganda, I tested the extent to which variation in the acoustic structure of pant-hoots is explained by community membership (and thus potentially learned signals), compared to individual traits that do not reflect vocal learning, such as age, rank, and health. Furthermore, I tested if the variation reflected adaptation by testing if some acoustic features of pant-hoots predicted mating success. This dissertation thus extends our knowledge of the forms and functions of chimpanzee vocal communication.

Chimpanzee pant-hoots have received considerable attention in studies of chimpanzee vocal communication due to their acoustic complexity and ubiquity in chimpanzee societies (Clark & Wrangham, 1994; Fedurek, Zuberbühler, et al., 2016; Mitani & Nishida, 1993). Pant-hoots are loud, conspicuous calls that can be heard over long distances. To study the variation in pant-hoots effectively, a quantitative description of the ‘form’ of pant-hoots is necessary. In other words, what range of acoustic variation constitutes a pant-hoot. Goodall and Marler provided the first detailed systematic descriptions of the vocal repertoire of chimpanzees (Goodall et al., 1968; Marler, 1969, 1976). However, more work needs to be done to mitigate human biases and improve the replicability of findings. While classifying a graded vocal repertoire such as that of chimpanzees remains challenging in absence of perceptual experiments such as playbacks, a standard protocol for classification based on quantitative analysis of acoustic parameters can make studies more replicable. Additionally, since animals vocalize to communicate with members of their own species, we should consider perceptual abilities of call receivers when studying their calls. However, conventional call classification approaches do not incorporate hearing abilities of the study species. Hence, in the first chapter, I integrated approaches from speech recognition and machine learning research and used acoustic features that incorporate auditory processing including mel frequency cepstral coefficients and emerging techniques in statistical classification such as random forests and artificial neural networks. I compared the performance of these new acoustic features and statistical classification techniques with that of traditional acoustical features commonly used in primatology such as acoustic features of the fundamental and peak frequencies, and measures of tonality and the statistical classification technique—discriminant functions analysis. I found some preliminary evidence that acoustic features used in speech recognition such as mel frequency cepstral coefficients and advanced statistical classification techniques such as random forests and deep artificial neural networks may be more valid and reliable for studying chimpanzee vocalizations in general. However, their performance improvement over traditional techniques used in primatology, especially for the vocalization of my interest (pant-hoots), was not large enough to warrant their use for other chapters in my dissertation that needed to use comparable techniques to previous studies.

In the next two chapters, I studied the ‘functions’ of the pant-hoots by testing two alternative hypotheses. The first hypothesis, called *the vocal learning hypothesis* is that the community-specific variation in chimpanzee pant-hoots functioned to signal territory ownership and reflected vocal learning capacity of chimpanzees. Vocal learning, which is the ability to learn and voluntarily modify acoustic structures of vocalizations is a fundamental characteristic of human speech. It has evolved independently in several lineages of birds and mammals but appears rare in non-human primates. The vocal learning hypothesis has been proposed and evaluated by several previous studies (Crockford et al., 2004; Marshall et al., 1999; Mitani et al., 1992, 1999). However, the reported differences among chimpanzee communities in these studies are subtle and the extent to which vocal learning explains the variation in pant-hoots is not well understood (Mitani et al., 1999). Hence, a replication of these findings was warranted. Furthermore, as mentioned above, this hypothesis employs a problematic conceptual framework that adopts complex linguistic phenomena as models for simpler primate vocal communication, and ignores several fundamental questions about the form, functions, and other sources of variation in chimpanzee vocalizations. Hence, I proposed another alternative hypothesis, called *the sexual selection hypothesis*: that the variation in the ‘form’ of pant-hoots reflects basic individual male traits that signal male quality and competitive ability such as age, rank, and health and functions to improve the mating success of chimpanzee males. I used a comparative approach in proposing this hypothesis. Loud calls in other primate species have been shown to provide cues of individual traits related to mate quality, such as body size, health, rank, and age (Erb et al., 2013; Fischer et al., 2004). Furthermore, the prevalence of non-linear phenomena such as subharmonics, biphonation, and deterministic chaos in the pant-hoots may reflect the difficulty of producing these loud calls and may signal the physical condition of the caller (Riede et al., 1997, 2004, 2007). Based on that, pant-hoots may provide less information about community membership, and more information about male mate quality, and thus may be shaped primarily by sexual selection, rather than by intergroup competition.

To test the vocal learning hypothesis, I compared two neighboring chimpanzee communities at Gombe: Kasekela and Mitumba, and one geographically distant chimpanzee community at Kibale: Kanyawara. I used traditional acoustic analysis and statistical classification techniques to test for differences in the acoustic structure of the pant-hoots among these communities. I found no differences among these communities in either (i) structural acoustic features that focus on the global compositional properties of the pant-hoots such as the number of different kinds of elements, presence or absence of different phases etc., or (ii) spectral acoustic features that focus on the acoustic features measured in the power spectrum and spectrogram of individual build-up or climax elements. Hence, I could not replicate previous findings of community-specific differences in the pant-hoots and did not find evidence in support of the vocal learning hypothesis.

To test the sexual selection hypothesis, I focused on three relevant acoustic features that may reflect a male's competitive traits: the proportion of pant-hoot climax screams that contains non-linear phenomena (proportion of NLPs), the maximum frequency in F0 (F0max), and maximum Wiener entropy, which is a measure of noisiness of a climax scream (noise_max). I tested if these acoustic features reflected male traits including age, rank, health, and whether they predicted male mating success. Consistent with the sexual selection hypothesis, I found that the proportion of NLPs varied quadratically with age, wherein males in their prime mating age (~25-35 years) had lower proportion of NLPs as compared to younger and older males. Similarly, noise_max also varied quadratically with age and males in their peak mating ages exhibited less noise in their climax screams. In further support of the sexual selection hypothesis, I found that the proportion of NLPs predicted mating success of males and males exhibiting lower proportion of NLPs had greater success even after controlling for rank and age. Furthermore, noise_max provided some information about health as SIVcpz positive chimpanzees exhibited greater noise in their vocalizations compared to SIVcpz negative chimpanzees after controlling for age. Next, I found that while at a population level, higher ranking males had lower proportion of NLPs, after controlling for age, higher ranking males exhibited a higher proportion of NLPs. This pattern appears inconsistent with the sexual selection hypothesis, but it may reflect other functions of NLPs such as attracting attention of conspecifics. Lastly, I found that while F0max appeared to reduce with age and lead to slightly higher mating success, the association of F0max with either age, rank, health, or mating success was not statistically significant.

In conclusion, this dissertation provides a methodological framework for future studies of chimpanzee vocal repertoire classification. My findings suggest that as chimpanzee vocal communication research matures further, providing larger and larger samples of recordings of chimpanzee vocalizations, the use of newer acoustic features that incorporate receiver perception such as mel frequency cepstral coefficients and advanced statistical classification techniques such as random forests and deep artificial neural networks could provide more valid and reliable results describing the acoustic ‘form’ of the chimpanzee vocal repertoire. In terms of function, my findings provide better support for the sexual selection hypothesis and do not provide support for the vocal leaning hypothesis. I found no evidence of community-specific dialects that have been believed to reflect chimpanzees' vocal learning ability. On the other hand, I found that certain acoustic features of chimpanzee pant-hoots such as the prevalence of non-linear phenomena and noisiness, reliably signal an individual's age and rank, may provide some signals of an individual's health and even predict their mating success. This provides evidence that pant-hoots are sexually selected signals of male mate quality and competitive ability, and females may choose mates based on the acoustic properties of male vocalizations. Absence of dialects in chimpanzee pant-hoots suggests that we cannot conclude that vocal learning was shared by the last common ancestor of chimpanzees and humans. Instead, since I found that individual traits rather than vocal learning better explained the variance in chimpanzee calls, it suggests that vocal learning in humans likely resulted from selective pressures specific to the human lineage. My findings demonstrate that returning to studying primate vocal communication with a basic biological focus on form and function may provide more fundamental insights, bolstering the sentiment in (Owren & Rendall, 2001).

Bibliography

Agamaite, J. A., Chang, C.-J., Osmanski, M. S., & Wang, X. (2015). A quantitative acoustic analysis of the vocal repertoire of the common marmoset (*Callithrix jacchus*). *Journal of the Acoustical Society of America*, *138*(5), 2906–2928. https://doi.org/10.1121/1.4934268

Albornoz, E. M., Vignolo, L. D., Sarquis, J. A., & Leon, E. (2017). Automatic classification of Furnariidae species from the Paranaense Littoral region using speech-related features and machine learning. *Ecological Informatics*, *38*(Supplement C), 39–49. https://doi.org/10.1016/j.ecoinf.2017.01.004

Ananthi, S., & Dhanalakshmi, P. (2015). *SVM and HMM modeling techniques for speech recognition using LPCC and MFCC features*. 519–526.

Andersson, M., & Simmons, L. W. (2006). Sexual selection and mate choice. *Trends in Ecology & Evolution*, *21*(6), 296–302. https://doi.org/10.1016/j.tree.2006.03.015

Araya-Ajoy, Y., Chaves-Campos, J., Kalko, E. K. V., & DeWoody, J. A. (2009). High-Pitched Notes during Vocal Contests Signal Genetic Diversity in Ocellated Antbirds. *Plos One*, *4*(12), e8137. https://doi.org/10.1371/journal.pone.0008137

Arcadi, A. C. (1996). Phrase structure of wild chimpanzee pant hoots: Patterns of production and interpopulation variability. *American Journal of Primatology*, *39*(3), 159–178. https://doi.org/10.1002/(SICI)1098-2345(1996)39:3<159::AID-AJP2>3.0.CO;2-Y

Arcadi, A. C. (2005). Language Evolution: What Do Chimpanzees Have to Say? *Current Biology*, *15*(21), R884–R886. https://doi.org/10.1016/j.cub.2005.10.020

Arcadi, A. C., & Wallauer, W. (2013). They wallop like they gallop: Audiovisual analysis reveals the influence of gait on buttress drumming by wild chimpanzees (*Pan troglodytes*). *International Journal of Primatology*, *34*(1), 194–215.

Armitage, D. W., & Ober, H. K. (2010). A comparison of supervised learning techniques in the classification of bat echolocation calls. *Ecological Informatics*, *5*(6), 465–473. https://doi.org/10.1016/j.ecoinf.2010.08.001

Arnocky, S., Hodges-Simeon, C. R., Ouellette, D., & Albert, G. (2018). Do men with more masculine voices have better immunocompetence? *Evolution and Human Behavior*, *39*(6), 602–610. https://doi.org/10.1016/j.evolhumbehav.2018.06.003

Ballentine, B., Hyman, J., & Nowicki, S. (2004). Vocal performance influences female response to male bird song: An experimental test. *Behavioral Ecology*, *15*(1), 163–168. https://doi.org/10.1093/beheco/arg090

Barelli, C., Mundry, R., Heistermann, M., & Hammerschmidt, K. (2013). Cues to Androgens and Quality in Male Gibbon Songs. *PLOS ONE*, *8*(12), e82748. https://doi.org/10.1371/journal.pone.0082748

Bartlett, P., & Slater, P. J. B. (1999). The effect of new recruits on the flock specific call of budgerigars (*Melopsittacus undulatus*). *Ethology Ecology & Evolution*, *11*(2), 139–147. https://doi.org/10.1080/08927014.1999.9522832

Barton, K. (2010). MuMIn: Multi-model inference. R package version 0.13. 17. *Http://CRAN. R-Project. Org/Package= MuMIn*.

Barua, P., Ahmad, K., Khan, A. A. S., & Sanaullah, M. (2014). *Neural network based recognition of speech using MFCC features*. 1–6.

Bermant, P. C. (2021). BioCPPNet: Automatic bioacoustic source separation with deep neural networks. *Scientific Reports*, *11*(1), 23502. https://doi.org/10.1038/s41598-021-02790-2

Bermant, P. C., Bronstein, M. M., Wood, R. J., Gero, S., & Gruber, D. F. (2019). Deep Machine Learning Techniques for the Detection and Classification of Sperm Whale Bioacoustics. *Scientific Reports*, *9*(1), 12588. https://doi.org/10.1038/s41598-019-48909-4

Bernstein, S. K., Sheeran, L. K., Wagner, R. S., Li, J.-H., & Koda, H. (2016). The vocal repertoire of Tibetan macaques (*Macaca thibetana*): A quantitative classification. *American Journal of Primatology*, *78*(9), 937–949. https://doi.org/10.1002/ajp.22564

Berthet, M., Neumann, C., Mesbahi, G., Cäsar, C., & Zuberbühler, K. (2017). Contextual encoding in titi monkey alarm call sequences. *Behavioral Ecology and Sociobiology*, *72*(1), 8. https://doi.org/10.1007/s00265-017-2424-z

Blumstein, D. T., & Récapet, C. (2009). The Sound of Arousal: The Addition of Novel Non-linearities Increases Responsiveness in Marmot Alarm Calls. *Ethology*, *115*(11), 1074–1081. https://doi.org/10.1111/j.1439-0310.2009.01691.x

Boersma, P., & Weenink, D. (2018). Praat: Doing phonetics by computer [Computer program]. Version 6.1.15. *Retrieved February*, *3*, 2018.

Boesch, C., Crockford, C., Herbinger, I., Wittig, R., Moebius, Y., & Normand, E. (2008). Intergroup conflicts among chimpanzees in Taï National Park: Lethal violence and the female perspective. *American Journal of Primatology*, *70*(6), 519–532. https://doi.org/10.1002/ajp.20524

Bolarinwa, O. A. (2015). Principles and methods of validity and reliability testing of questionnaires used in social and health science researches. *Nigerian Postgraduate Medical Journal*, *22*(4), 195.

Bouchet, H., Blois-Heulin, C., & Lemasson, A. (2012). Age- and sex-specific patterns of vocal behavior in De Brazza's monkeys (*Cercopithecus neglectus*). *American Journal of Primatology*, *74*(1), 12–28. https://doi.org/10.1002/ajp.21002

Bouchet, H., Pellier, A.-S., Blois-Heulin, C., & Lemasson, A. (2010). Sex differences in the vocal repertoire of adult red-capped mangabeys (*Cercocebus torquatus*): A multi-level acoustic analysis. *American Journal of Primatology*, *72*(4), 360–375. https://doi.org/10.1002/ajp.20791

Bray, J., & Gilby, I. C. (2020). Social relationships among adult male chimpanzees (*Pan troglodytes schweinfurthii*): Variation in the strength and quality of social bonds. *Behavioral Ecology and Sociobiology*, *74*(9), 112. https://doi.org/10.1007/s00265-020-02892-3

Brockelman, W. Y., & Schilling, D. (1984). Inheritance of stereotyped gibbon calls. *Nature*, *312*(5995), 634–636. https://doi.org/10.1038/312634a0

Burman, L. E., Reed, W. R., & Alm, J. (2010). A call for replication studies. *Public Finance Review*, *38*(6), 787–793.

Catchpole, C. K. (1996). Song and female choice: Good genes and big brains? *Trends in Ecology & Evolution*, *11*(9), 358–360.

Cattell, R. B. (1966). The Scree Test For The Number Of Factors. *Multivariate Behavioral Research*, *1*(2), 245–276. https://doi.org/10.1207/s15327906mbr0102_10

Charlton, B. D., Watchorn, D. J., & Whisson, D. A. (2017). Subharmonics increase the auditory impact of female koala rejection calls. *Ethology*, *123*(8), 571–579. https://doi.org/10.1111/eth.12628

Charlton, B. D., Zhihe, Z., & Snyder, R. J. (2009). The information content of giant panda, *Ailuropoda melanoleuca*, bleats: Acoustic cues to sex, age and size. *Animal Behaviour*, *78*(4), 893–898. https://doi.org/10.1016/j.anbehav.2009.06.029

Clark, A. P. (1993). Rank differences in the production of vocalizations by wild chimpanzees as a function of social context. *American Journal of Primatology*, *31*(3), 159–179. https://doi.org/10.1002/ajp.1350310302

Clark, A. P., & Wrangham, R. W. (1993). Acoustic analysis of wild chimpanzee pant hoots: Do Kibale Forest chimpanzees have an acoustically distinct food arrival pant hoot? *American Journal of Primatology*, *31*(2), 99–109. https://doi.org/10.1002/ajp.1350310203

Clark, A. P., & Wrangham, R. W. (1994). Chimpanzee arrival pant-hoots: Do they signify food or status? *International Journal of Primatology*, *15*(2), 185–205. https://doi.org/10.1007/BF02735273

Clay, Z., Smith, C. L., & Blumstein, D. T. (2012). Food-associated vocalizations in mammals and birds: What do these calls really mean? *Animal Behaviour*, *83*(2), 323–330. https://doi.org/10.1016/j.anbehav.2011.12.008

Clemins, P. J., Johnson, M. T., Leong, K. M., & Savage, A. (2005). Automatic classification and speaker identification of African elephant (*Loxodonta africana*) vocalizations. *The Journal of the Acoustical Society of America*, *117*(2), 956–963. https://doi.org/10.1121/1.1847850

Crockford, C. (2019). Why does the chimpanzee vocal repertoire remain poorly understood?-and what can be done about it. In *The chimpanzees of the Taï forest: 40 years of research* (pp. 394–409). Cambridge University Press.

Crockford, C., Herbinger, I., Vigilant, L., & Boesch, C. (2004). Wild Chimpanzees Produce Group-Specific Calls: A Case for Vocal Learning? *Ethology*, *110*(3), 221–243. https://doi.org/10.1111/j.1439-0310.2004.00968.x

Cunningham, M., & Baker, M. (1983). Vocal Learning in White-Crowned Sparrows—Sensitive Phase and Song Dialects. *Behavioral Ecology and Sociobiology*, *13*(4), 259–269. https://doi.org/10.1007/BF00299673

Das, B. P., & Parekh, R. (2012). Recognition of isolated words using features based on LPC, MFCC, ZCR and STE, with neural network classifiers. *International Journal of Modern Engineering Research*, *2*(3), 854–858.

Dawson, M. R. W., Charrier, I., & Sturdy, C. B. (2006). Using an artificial neural network to classify black-capped chickadee (*Poecile atricapillus*) call note types. *The Journal of the Acoustical Society of America*, *119*(5), 3161–3172. https://doi.org/10.1121/1.2189028

De Kort, S. R., Den Hartog, P. M., & Ten Cate, C. (2002). Diverge or merge? The effect of sympatric occurrence on the territorial vocalizations of the vinaceous dove *Streptopelia vinacea* and the ring-necked dove *S. capicola. Journal of Avian Biology*, *33*(2), 150–158. https://doi.org/10.1034/j.1600-048X.2002.330205.x

de la Torre, S., & Snowdon, C. T. (2009). Dialects in pygmy marmosets? Population variation in call structure. *American Journal of Primatology*, *71*(4), 333–342. https://doi.org/10.1002/ajp.20657

de Waal, F. B. M. (1988). The Communicative Repertoire of Captive Bonobos (*Pan Paniscus*), Compared To That of Chimpanzees. *Behaviour*, *106*(3), 183–251. https://doi.org/10.1163/156853988X00269

Desai, N. P., Fedurek, P., Slocombe, K. E., & Wilson, M. L. (2021). Chimpanzee pant-hoots encode information about individual but not group differences. *BioRxiv*, 2021.03.09.434515. https://doi.org/10.1101/2021.03.09.434515

Desai, N. P., Lehman, C., Munson, B., & Wilson, M. (2018). Supervised and unsupervised machine learning approaches to classifying chimpanzee vocalizations. *The Journal of the Acoustical Society of America*, *143*(3), 1786–1786. https://doi.org/10.1121/1.5035847

Digby, A., Bell, B. D., & Teal, P. D. (2014). Non-linear phenomena in little spotted kiwi calls. *Bioacoustics*, *23*(2), 113–128. https://doi.org/10.1080/09524622.2013.829755

Dugan, P. J., Klinck, H., Roch, M. A., & Helble, T. A. (2016). Raven x high performance data mining toolbox for bioacoustic data analysis. *ArXiv Preprint ArXiv:1610.03772*.

Elowson, A. M., & Snowdon, C. T. (1994). Pygmy marmosets, *Cebuella pygmaea*, modify vocal structure in response to changed social environment. *Animal Behaviour*, *47*(6), 1267–1277. https://doi.org/10.1006/anbe.1994.1175

Emery Thompson, M., Jones, J. H., Pusey, A. E., Brewer-Marsden, S., Goodall, J., Marsden, D., Matsuzawa, T., Nishida, T., Reynolds, V., Sugiyama, Y., & Wrangham, R. W. (2007). Aging and Fertility Patterns in Wild Chimpanzees Provide Insights into the Evolution of Menopause. *Current Biology*, *17*(24), 2150–2156. https://doi.org/10.1016/j.cub.2007.11.033

Emery Thompson, M., Muller, M. N., Machanda, Z. P., Otali, E., & Wrangham, R. W. (2020). The Kibale Chimpanzee Project: Over thirty years of research, conservation, and change. *Biological Conservation*, *252*, 108857. https://doi.org/10.1016/j.biocon.2020.108857

Erb, W. M., Hodges, J. K., & Hammerschmidt, K. (2013). Individual, Contextual, and Age-Related Acoustic Variation in Simakobu (*Simias concolor*) Loud Calls. *PLOS ONE*, *8*(12), e83131. https://doi.org/10.1371/journal.pone.0083131

Evans, J. S., Murphy, M. A., Holden, Z. A., & Cushman, S. A. (2011). Modeling species distribution and change using random forest. In *Predictive species and habitat modeling in landscape ecology* (pp. 139–159). Springer.

Farabaugh, S. M., Linzenbold, A., & Dooling, R. J. (1994). Vocal plasticity in budgerigars (*Melopsittacus undulatus*): Evidence for social factors in the learning of contact calls. *Journal of Comparative Psychology (Washington, D.C.: 1983)*, *108*(1), 81–92.

Fedurek, P., Donnellan, E., & Slocombe, K. E. (2014). Social and ecological correlates of long-distance pant hoot calls in male chimpanzees. *Behavioral Ecology and Sociobiology*, *68*(8), 1345–1355. https://doi.org/10.1007/s00265-014-1745-4

Fedurek, P., Machanda, Z. P., Schel, A. M., & Slocombe, K. E. (2013). Pant hoot chorusing and social bonds in male chimpanzees. *Animal Behaviour*, *86*(1), 189–196. https://doi.org/10.1016/j.anbehav.2013.05.010

Fedurek, P., Schel, A. M., & Slocombe, K. E. (2013). The acoustic structure of chimpanzee pant-hooting facilitates chorusing. *Behavioral Ecology and Sociobiology*, *67*(11), 1781–1789. https://doi.org/10.1007/s00265-013-1585-7

Fedurek, P., & Slocombe, K. E. (2011). Primate Vocal Communication: A Useful Tool for Understanding Human Speech and Language Evolution? *Human Biology*, *83*(2), 153–173. https://doi.org/10.3378/027.083.0202

Fedurek, P., Slocombe, K. E., Enigk, D. K., Thompson, M. E., Wrangham, R. W., & Muller, M. N. (2016). The relationship between testosterone and long-distance calling in wild male chimpanzees. *Behavioral Ecology and Sociobiology*, *70*(5), 659–672. https://doi.org/10.1007/s00265-016-2087-1

Fedurek, P., Zuberbühler, K., & Dahl, C. D. (2016). Sequential information in a great ape utterance. *Scientific Reports*, *6*, 38226. https://doi.org/10.1038/srep38226

Feldblum, J. T., Wroblewski, E. E., Rudicell, R. S., Hahn, B. H., Paiva, T., Cetinkaya-Rundel, M., Pusey, A. E., & Gilby, I. C. (2014). Sexually Coercive Male Chimpanzees Sire More Offspring. *Current Biology*, *24*(23), 2855–2860. https://doi.org/10.1016/j.cub.2014.10.039

Filatova, O. A., Deecke, V. B., Ford, J. K. B., Matkin, C. O., Barrett-Lennard, L. G., Guzeev, M. A., Burdin, A. M., & Hoyt, E. (2012). Call diversity in the North Pacific killer whale populations: Implications for dialect evolution and population history. *Animal Behaviour*, *83*(3), 595–603. https://doi.org/10.1016/j.anbehav.2011.12.013

Fischer, J., & Hammerschmidt, K. (2020). Towards a new taxonomy of primate vocal production learning. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *375*(1789), 20190045. https://doi.org/10.1098/rstb.2019.0045

Fischer, J., Kitchen, D. M., Seyfarth, R. M., & Cheney, D. L. (2004). Baboon loud calls advertise male quality: Acoustic features and their relation to rank, age, and exhaustion. *Behavioral Ecology and Sociobiology, 56*(2), 140–148. https://doi.org/10.1007/s00265-003-0739-4

Fischer, J., Noser, R., & Hammerschmidt, K. (2013). Bioacoustic Field Research: A Primer to Acoustic Analyses and Playback Experiments With Primates. *American Journal of Primatology*, *75*(7), 643–663. https://doi.org/10.1002/ajp.22153

Fischer, J., Wegdell, F., Trede, F., Dal Pesco, F., & Hammerschmidt, K. (2020). Vocal convergence in a multi-level primate society: Insights into the evolution of vocal learning. *Proceedings of the Royal Society B: Biological Sciences*, *287*(1941), 20202531. https://doi.org/10.1098/rspb.2020.2531

Fischer, J., Wheeler, B. C., & Higham, J. P. (2015). Is there any evidence for vocal learning in chimpanzee food calls? *Current Biology*, *25*(21), R1028–R1029. https://doi.org/10.1016/j.cub.2015.09.010

Fitch, W. T. (1997). Vocal tract length and formant frequency dispersion correlate with body size in rhesus macaques. *The Journal of the Acoustical Society of America*, *102*(2), 1213–1222. https://doi.org/10.1121/1.421048

Fitch, W. T. (2010). *The Evolution of Language*. Cambridge University Press.

Fitch, W. T., Neubauer, J., & Herzel, H. (2002). Calls out of chaos: The adaptive significance of nonlinear phenomena in mammalian vocal production. *Animal Behaviour, 63*(3), 407–418. https://doi.org/10.1006/anbe.2001.1912

Foerster, S., Franz, M., Murray, C. M., Gilby, I. C., Feldblum, J. T., Walker, K. K., & Pusey, A. E. (2016). Chimpanzee females queue but males compete for social status. *Scientific Reports*, *6*, 35404.

Fossey, D. (1972). Vocalizations of the mountain Gorilla (*Gorilla gorilla beringei*). *Animal Behaviour*, *20*(1), 36–53. https://doi.org/10.1016/S0003-3472(72)80171-4

Fournet, M. E. H., Jacobsen, L., Gabriele, C. M., Mellinger, D. K., & Klinck, H. (2018). More of the same: Allopatric humpback whale populations share acoustic repertoire. *Peerj*, *6*, e5365. https://doi.org/10.7717/peerj.5365

Fuller, J. L. (2014). The vocal repertoire of adult male blue monkeys (*Cercopithecus mitis stulmanni*): A quantitative analysis of acoustic structure. *American Journal of Primatology*, *76*(3), 203–216. https://doi.org/10.1002/ajp.22223

Gamba, M., & Giacoma, C. (2007). Quantitative acoustic analysis of the vocal repertoire of the crowned lemur. *Ethology Ecology & Evolution*, *19*(4), 323–343. https://doi.org/10.1080/08927014.2007.9522555

Garland, E. C., Goldizen, A. W., Rekdahl, M. L., Constantine, R., Garrigue, C., Hauser, N. D., Poole, M. M., Robbins, J., & Noad, M. J. (2011). Dynamic Horizontal Cultural Transmission of Humpback Whale Song at the Ocean Basin Scale. *Current Biology*, *21*(8), 687–691. https://doi.org/10.1016/j.cub.2011.03.019

Gelman, A., & Rubin, D. B. (1992). Inference from Iterative Simulation Using Multiple Sequences. *Statistical Science*, *7*(4), 457–472. https://doi.org/10.1214/ss/1177011136

Giannakopoulos, T. (2015). pyAudioAnalysis: An Open-Source Python Library for Audio Signal Analysis. *PLOS ONE*, *10*(12), e0144610. https://doi.org/10.1371/journal.pone.0144610

Goodall, J. (1986). *The chimpanzees of Gombe: Patterns of behavior*. Belknap Press of Harvard University Press.

Goodall, J., Jay, P., & Holt, R. (1968). A preliminary report on expressive movements and communication in the Gombe Stream chimpanzees. *Primates: Studies in Adaptation and Variability*, 313–374.

Green, A., Clark, C., Favaro, L., Lomax, S., & Reby, D. (2019). Vocal individuality of Holstein-Friesian cattle is maintained across putatively positive and negative farming contexts. *Scientific Reports*, *9*(1), 18468. https://doi.org/10.1038/s41598-019-54968-4

Gros-Louis, J. J., Perry, S. E., Fichtel, C., Wikberg, E., Gilkenson, H., Wofsy, S., & Fuentes, A. (2008). Vocal Repertoire of *Cebus capucinus*: Acoustic Structure, Context, and Usage. *International Journal of Primatology*, *29*(3), 641–670. https://doi.org/10.1007/s10764-008-9263-8

Hacki, T., & Heitmüller, S. (1999). Development of the child's voice: Premutation, mutation. *International Journal of Pediatric Otorhinolaryngology, 49,* S141–S144. https://doi.org/10.1016/S0165-5876(99)00150-0

Hammerschmidt, K., & Fischer, J. (1998). The Vocal Repertoire of Barbary Macaques: A Quantitative Analysis of a Graded Signal System. *Ethology, 104*(3), 203–216. https://doi.org/10.1111/j.1439-0310.1998.tb00063.x

Hammerschmidt, K., & Fischer, J. (2019). Baboon vocal repertoires and the evolution of primate vocal diversity. *Journal of Human Evolution*, *126*, 1–13. https://doi.org/10.1016/j.jhevol.2018.10.010

Hammerschmidt, K., Newman, J. D., Champoux, M., & Suomi, S. J. (2000). Changes in rhesus macaque 'coo'vocalizations during early development. *Ethology*, *106*(10), 873–886.

Hardouin, L. A., Reby, D., Bavoux, C., Burneleau, G., & Bretagnolle, V. (2007). Communication of Male Quality in Owl Hoots. *The American Naturalist*, *169*(4), 552–562. https://doi.org/10.1086/512136

Hastie, T., Buja, A., & Tibshirani, R. (1995). Penalized discriminant analysis. *The Annals of Statistics*, *23*(1), 73–102.

Hastie, T., Tibshirani, R., Friedman, J. H., & Friedman, J. H. (2009). *The elements of statistical learning: Data mining, inference, and prediction* (Vol. 2). Springer.

Hedwig, D., Hammerschmidt, K., Mundry, R., Robbins, M. M., & Boesch, C. (2014). Acoustic structure and variation in mountain and western gorilla close calls: A syntactic approach. *Behaviour, 151*(8), 1091–1120. https://doi.org/10.1163/1568539X-00003175

Hegger, R., Kantz, H., & Schreiber, T. (1999). Practical implementation of nonlinear time series methods: The TISEAN package. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, *9*(2), 413–435. https://doi.org/10.1063/1.166424

Herbinger, I., Papworth, S., Boesch, C., & Zuberbühler, K. (2009). Vocal, gestural and locomotor responses of wild chimpanzees to familiar and unfamiliar intruders: A playback study. *Animal Behaviour, 78*(6), 1389–1396. https://doi.org/10.1016/j.anbehav.2009.09.010

Herzel, H., Berry, D., Titze, I. R., & Saleh, M. (1994). Analysis of Vocal Disorders With Methods From Nonlinear Dynamics. *Journal of Speech, Language, and Hearing Research*, *37*(5), 1008–1019. https://doi.org/10.1044/jshr.3705.1008

Heylen, L., Wuyts, F. L., Mertens, F., De Bodt, M., Pattyn, J., Croux, C., & Van de Heyning, P. H. (1998). Evaluation of the vocal performance of children using a voice range profile index. *Journal of Speech, Language, and Hearing Research: JSLHR*, *41*(2), 232–238. https://doi.org/10.1044/jslhr.4102.232

Hile, A. G., & Striedter, G. F. (2000). Call Convergence within Groups of Female Budgerigars (*Melopsittacus undulatus*). *Ethology*, *106*(12), 1105–1114. https://doi.org/10.1046/j.1439-0310.2000.00637.x

Hinton, G., Deng, L., Yu, D., Dahl, G. E., Mohamed, A., Jaitly, N., Senior, A., Vanhoucke, V., Nguyen, P., & Sainath, T. N. (2012). Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. *IEEE Signal Processing Magazine*, *29*(6), 82–97.

Hodges-Simeon, C. R., Gaulin, S. J. C., & Puts, D. A. (2010). Different Vocal Parameters Predict Perceptions of Dominance and Attractiveness. *Human Nature*, *21*(4), 406–427. https://doi.org/10.1007/s12110-010-9101-5

Hodges-Simeon, C. R., Gurven, M., & Gaulin, S. J. C. (2015). The low male voice is a costly signal of phenotypic quality among Bolivian adolescents. *Evolution and Human Behavior*, *36*(4), 294–302. https://doi.org/10.1016/j.evolhumbehav.2015.01.002

Ikeda, Y., Masuda, T., Manako, H., Yamashita, H., Yamamoto, T., & Komiyama, S. (1999). Quantitative evaluation of the voice range profile in patients with voice disorder. *European Archives of Oto-Rhino-Laryngology*, *256*(1), S51–S55. https://doi.org/10.1007/PL00014154

Ioannidis, J. P. A. (2005). Why Most Published Research Findings Are False. *PLOS Medicine*, *2*(8), e124. https://doi.org/10.1371/journal.pmed.0020124

Isabirye-Basuta, G. (1988). Food Competition Among Individuals in a Free-Ranging Chimpanzee Community in Kibale Forest, Uganda. *Behaviour*, *105*(1–2), 135–147. https://doi.org/10.1163/156853988X00485

Ittichaichareon, C., Suksri, S., & Yingthawornsuk, T. (2012). *Speech recognition using MFCC. 9.*

Janik, V. M. (2014). Cetacean vocal learning and communication. *Current Opinion in Neurobiology*, *28*, 60–65. https://doi.org/10.1016/j.conb.2014.06.010

Janik, V. M., & Slater, P. J. B. (1998). Context-specific use suggests that bottlenose dolphin signature whistles are cohesion calls. *Animal Behaviour*, *56*(4), 829–838. https://doi.org/10.1006/anbe.1998.0881

Janik, V. M., & Slater, P. J. B. (2000). The different roles of social learning in vocal communication. *Animal Behaviour*, *60*(1), 1–11. https://doi.org/10.1006/anbe.2000.1410

Johnson, D. H. (2002). The Importance of Replication in Wildlife Research. *The Journal of Wildlife Management*, *66*(4), 919–932. https://doi.org/10.2307/3802926

Johnston, J. D. (1988). Transform coding of audio signals using perceptual noise criteria. *IEEE Journal on Selected Areas in Communications*, *6*(2), 314–323. https://doi.org/10.1109/49.608

Jolliffe, I. T. (Ed.). (2002). Choosing a Subset of Principal Components or Variables. In *Principal Component Analysis* (pp. 111–149). Springer. https://doi.org/10.1007/0-387-22440-8_6

Keele, B. F., Jones, J. H., Terio, K. A., Estes, J. D., Rudicell, R. S., Wilson, M. L., Li, Y., Learn, G. H., Beasley, T. M., & Schumacher-Stankey, J. (2009). Increased mortality and AIDS-like immunopathology in wild chimpanzees infected with SIVcpz. *Nature*, *460*(7254), 515.

Keenan, S., Lemasson, A., & Zuberbühler, K. (2013). Graded or discrete? A quantitative analysis of Campbell’s monkey alarm calls. *Animal Behaviour*, *85*(1), 109–118. https://doi.org/10.1016/j.anbehav.2012.10.014

Kitchen, D. M., Cheney, D. L., Engh, A. L., Fischer, J., Moscovice, L. R., & Seyfarth, R. M. (2013). Male baboon responses to experimental manipulations of loud “wahoo calls”: Testing an honest signal of fighting ability. *Behavioral Ecology and Sociobiology*, *67*(11), 1825–1835. https://doi.org/10.1007/s00265-013-1592-8

Kitchen, D. M., Seyfarth, R. M., Fischer, J., & Cheney, D. L. (2003). Loud calls as indicators of dominance in male baboons (*Papio cynocephalus ursinus*). *Behavioral Ecology and Sociobiology*, *53*(6), 374–384. https://doi.org/10.1007/s00265-003-0588-1

Kojima, S. (1990). Comparison of Auditory Functions in the Chimpanzee and Human. *Folia Primatologica*, *55*(2), 62–72. https://doi.org/10.1159/000156501

Krebs, J. R., & Dawkins, R. (1984). Animal signals: Mind-reading and manipulation. *Behavioural Ecology: An Evolutionary Approach, 2nd Edn (Ed. JR Krebs & NB Davies)*, 380–402.

Kuhn, M., Wing, J., Weston, S., Williams, A., Keefer, C., Engelhardt, A., Cooper, T., Mayer, Z., Kenkel, B., & Team, R. C. (2020). Package 'caret.' *The R Journal*, *223*, 7.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2015). Package 'lmertest.' *R Package Version*, *2*(0), 734.

Lamal, P. A. (1990). On the Importance of Replication. *Journal of Social Behavior and Personality*, *5*(4), 31–35.

Lameira, A. R., & Call, J. (2020). Understanding Language Evolution: Beyond Pan-Centrism. *BioEssays*, *42*(3), 1900102. https://doi.org/10.1002/bies.201900102

Lameira, A. R., & Shumaker, R. W. (2019). Orangutans show active voicing through a membranophone. *Scientific Reports*, *9*(1), 12289. https://doi.org/10.1038/s41598-019-48760-7

Laporte, M. N. C., & Zuberbühler, K. (2010). Vocal greeting behaviour in wild chimpanzee females. *Animal Behaviour*, *80*(3), 467–473. https://doi.org/10.1016/j.anbehav.2010.06.005

Laubach, Z. M., Murray, E. J., Hoke, K. L., Safran, R. J., & Perng, W. (2021). A biologist's guide to model selection and causal inference. *Proceedings of the Royal Society B: Biological Sciences*, *288*(1943), 20202815. https://doi.org/10.1098/rspb.2020.2815

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444. https://doi.org/10.1038/nature14539

Loken, E., & Gelman, A. (2017). Measurement error and the replication crisis. *Science*, *355*(6325), 584–585. https://doi.org/10.1126/science.aal3618

Lonsdorf, E. V., Gillespie, T. R., Wolf, T. M., Lipende, I., Raphael, J., Bakuza, J., Murray, C. M., Wilson, M. L., Kamenya, S., Mjungu, D., Collins, D. A., Gilby, I. C., Stanton, M. A., Terio, K. A., Barbian, H. J., Li, Y., Ramirez, M., Krupnick, A., Seidl, E., … Travis, D. A. (2018). Socioecological correlates of clinical signs in two communities of wild chimpanzees (*Pan troglodytes*) at Gombe National Park, Tanzania. *American Journal of Primatology*, *80*(1), e22562. https://doi.org/10.1002/ajp.22562

Magiera, A., Penar, W., & Klocek, C. (2020). *Occurrence of nonlinear phenomena in Siberian and Neva Masquerade cats? Vocalisation*. 2767–2770.

Magnusson, A., Skaug, H., Nielsen, A., Berg, C., Kristensen, K., Maechler, M., van Bentham, K., Bolker, B., & Brooks, M. E. (2017). GlmmTMB: generalized linear mixed models using template model builder. *R Package Version 0.1*, *3*.

Manteuffel, G., Puppe, B., & Schön, P. C. (2004). Vocalization of farm animals as a measure of welfare. *Applied Animal Behaviour Science*, *88*(1), 163–182. https://doi.org/10.1016/j.applanim.2004.02.012

Marler, P. (1955). Characteristics of some animal calls. *Nature*, *176*(4470), 6–8.

Marler, P. (1958). Bird songs and mate selection. *Animal Behaviour*, *6*(3–4), 254–254.

Marler, P. (1968). Aggregation and dispersal: Two functions in primate communication. *Primates: Studies in Adaptation and Variability*.

Marler, P. (1969). Vocalizations of wild chimpanzees. An introduction. *Proceedings of the Second International Congress of Primatology, Atlanta 1968*, *1*, 94–100.

Marler, P. (1976). Social organization, communication and graded signals: The chimpanzee and the gorilla. In *Growing points in Ethology* (pp. 239–280).

Marler, P., & Hobbett, L. (1975). Individuality in a Long-Range Vocalization of Wild Chimpanzees. *Zeitschrift Für Tierpsychologie*, *38*(1), 97–109. https://doi.org/10.1111/j.1439-0310.1975.tb01994.x

Marshall, A. J., Wrangham, R. W., & Arcadi, A. C. (1999). Does learning affect the structure of vocalizations in chimpanzees? *Animal Behaviour*, *58*(4), 825–830. https://doi.org/10.1006/anbe.1999.1219

Marx, A., Lenkei, R., Pérez Fraga, P., Bakos, V., Kubinyi, E., & Faragó, T. (2021). Occurrences of non-linear phenomena and vocal harshness in dog whines as indicators of stress and ageing. *Scientific Reports*, *11*(1), 4468. https://doi.org/10.1038/s41598-021-83614-1

Massaro, A. P., Gilby, I. C., Desai, N., Weiss, A., Feldblum, J. T., Pusey, A. E., & Wilson, M. L. (2022). Correlates of individual participation in boundary patrols by male chimpanzees. *Philosophical Transactions of the Royal Society B*, *377*(1851), 20210151.

Massaro, A. P., Wroblewski, E. E., Mjungu, D. C., Boehm, E. E., Desai, N. P., Foerster, S., Rudicell, R. S., Hahn, B. H., Pusey, A. E., & Wilson, M. L. (2021). Female monopolizability promotes within-community killing in chimpanzees. *Nature Portfolio*. https://doi.org/10.21203/rs.3.rs-163673/v1

Matsumoto-Oda, A. (1999). Female choice in the opportunistic mating of wild chimpanzees (*Pan troglodytes schweinfurthii*) at Mahale. *Behavioral Ecology and Sociobiology*, *46*(4), 258–266. https://doi.org/10.1007/s002650050618

McCordic, J. A., Root-Gutteridge, H., Cusano, D. A., Denes, S. L., & Parks, S. E. (2016). Calls of North Atlantic right whales *Eubalaena glacialis* contain information on individual identity and age class. *Endangered Species Research*, *30*, 157–169. https://doi.org/10.3354/esr00735

McElreath, R. (2018). *Statistical rethinking: A Bayesian course with examples in R and Stan*. Chapman and Hall/CRC.

McElreath, R. (2022). *Rethinking* [R]. https://github.com/rmcelreath/rethinking (Original work published 2013)

Mermelstein, P. (1976). Distance measures for speech recognition, psychological and instrumental. *Pattern Recognition and Artificial Intelligence*, *116*, 374–388.

Mielke, A., & Zuberbühler, K. (2013). A method for automated individual, species and call type recognition in free-ranging animals. *Animal Behaviour*, *86*(2), 475–482. https://doi.org/10.1016/j.anbehav.2013.04.017

Mikolov, T., Deoras, A., Povey, D., Burget, L., & Černocký, J. (2011). Strategies for training large scale neural network language models. *2011 IEEE Workshop on Automatic Speech Recognition Understanding*, 196–201. https://doi.org/10.1109/ASRU.2011.6163930

Mitani, J. C., & Brandt, K. (1994). Social factors influence the acoustic variability in the long-distance calls of male chimpanzees. *Ethology*, *96*(3), 233–252. https://doi.org/10.1111/j.1439-0310.1994.tb01012.x

Mitani, J. C., & Gros-Louis, J. (1998). Chorusing and Call Convergence in Chimpanzees: Tests of Three Hypotheses. *Behaviour*, *135*(8/9), 1041–1064.

Mitani, J. C., Gros-Louis, J., & Macedonia, J. (1996). Selection for acoustic individuality within the vocal repertoire of wild chimpanzees. *International Journal of Primatology*, *17*(4), 569–583. https://doi.org/10.1007/BF02735192

Mitani, J. C., Hasegawa, T., Gros-Louis, J., Marler, P., & Byrne, R. (1992). Dialects in wild chimpanzees? *American Journal of Primatology*, *27*(4), 233–243. https://doi.org/10.1002/ajp.1350270402

Mitani, J. C., Hunley, K., & Murdoch, E. (1999). Geographic variation in the calls of wild chimpanzees: A reassessment. *American Journal of Primatology*, *47*(2), 133–151. https://doi.org/10.1002/(SICI)1098-2345(1999)47:2<133::AID-AJP4>3.0.CO;2-I

Mitani, J. C., & Nishida, T. (1993). Contexts and social correlates of long-distance calling by male chimpanzees. *Animal Behaviour*, *45*(4), 735–746. https://doi.org/10.1006/anbe.1993.1088

Moselhy, A. M., & Abdelnaiem, A. A. (2013). LPC and MFCC performance evaluation with artificial neural network for spoken language identification. *International Journal of Signal Processing, Image Processing and Pattern Recognition*, *6*(3), 55.

Muller, M. N., Thompson, M. E., Kahlenberg, S. M., & Wrangham, R. W. (2011). Sexual coercion by male chimpanzees shows that female choice may be more apparent than real. *Behavioral Ecology and Sociobiology*, *65*(5), 921–933. https://doi.org/10.1007/s00265-010-1093-y

Mundry, R., & Sommer, C. (2007). Discriminant function analysis with nonindependent data: Consequences and an alternative. *Animal Behaviour*, *74*(4), 965–976. https://doi.org/10.1016/j.anbehav.2006.12.028

Neff, N. A., & Marcus, L. F. (1980). *A survey of multivariate methods for systematics*.

Neuliep, J. W., & Crandall, R. (1993). Reviewer Bias Against Replication Research. *Journal of Social Behavior and Personality*, *8*(6), 21–29.

Neumann, C. (2020). *Cfp: Christof's functions package v. 0.1.17.* [R]. https://github.com/gobbios/cfp

Neumann, C., Assahad, G., Hammerschmidt, K., Perwitasari-Farajallah, D., & Engelhardt, A. (2010). Loud calls in male crested macaques, *Macaca nigra*: A signal of dominance in a tolerant species. *Animal Behaviour*, *79*(1), 187–193. https://doi.org/10.1016/j.anbehav.2009.10.026

Notman, H., & Rendall, D. (2005). Contextual variation in chimpanzee pant hoots and its implications for referential communication. *Animal Behaviour*, *70*(1), 177–190.

Nowicki, S. (1983). Flock-specific recognition of chickadee calls. *Behavioral Ecology and Sociobiology*, *12*(4), 317–320. https://doi.org/10.1007/BF00302899

Nowicki, S., & Searcy, W. A. (2014). The evolution of vocal learning. *Current Opinion in Neurobiology*, *28*, 48–53. https://doi.org/10.1016/j.conb.2014.06.007

O'Bryan, L. (2015). *The Socioecology of Chimpanzee Foraging and Food-associated Calling Behavior*. http://conservancy.umn.edu/handle/11299/175282

O'Bryan, L., Munson, B., & Wilson, M. (2018). Chimpanzee food-associated calls display low referential potential in a wild population. *The Journal of the Acoustical Society of America*, *143*(3), 1788–1788. https://doi.org/10.1121/1.5035852

Owren, M. J., & Rendall, D. (2001). Sound on the rebound: Bringing form and function back to the forefront in understanding nonhuman primate vocal signaling.

Evolutionary Anthropology: Issues, News, and Reviews, *10*(2), 58–71. https://doi.org/10.1002/evan.1014

Owren, M. J., Seyfarth, R. M., & Cheney, D. L. (1997). The acoustic features of vowel-like grunt calls in chacma baboons (*Papio cyncephalus ursinus*): Implications for production processes and functions. *The Journal of the Acoustical Society of America*, *101*(5), 2951–2963. https://doi.org/10.1121/1.418523

Pieta, K. (2008). Female Mate Preferences Among *Pan troglodytes schweinfurthii* of Kanyawara, Kibale National Park, Uganda. *International Journal of Primatology*, *29*(4), 845. https://doi.org/10.1007/s10764-008-9282-5

Platt, J. R. (1964). Strong Inference. *Science*, *146*(3642), 347–353.

Plooij, F. X., van de Rijt-Plooij, H., Fischer, M., & Pusey, A. (2014). Longitudinal recordings of the vocalizations of immature Gombe chimpanzees for developmental studies. *Scientific Data*, *1*. https://doi.org/10.1038/sdata.2014.25

Plooij, F. X., van de Rijt-Plooij, H., Fischer, M., Wilson, M. L., & Pusey, A. (2015). An archive of longitudinal recordings of the vocalizations of adult Gombe chimpanzees. *Scientific Data*, *2*, 150027. https://doi.org/10.1038/sdata.2015.27

Popper, K. R. (1959). *The logic of scientific discovery* (p. 480). Basic Books.

Pozzi, L., Gamba, M., & Giacoma, C. (2010). The use of Artificial Neural Networks to classify primate vocalizations: A pilot study on black lemurs. *American Journal of Primatology*, *72*(4), 337–348. https://doi.org/10.1002/ajp.20786

Pusey, A. E. (1990). Behavioural Changes at Adolescence in Chimpanzees. *Behaviour*, *115*(3/4), 203–246.

Pusey, A. E., Oehlert, G. W., Williams, J. M., & Goodall, J. (2005). Influence of ecological and social factors on body mass of wild chimpanzees. *International Journal of Primatology*, *26*(1), 3–31.

Reby, D., & Charlton, B. D. (2012). Attention grabbing in red deer sexual calls. *Animal Cognition*, *15*(2), 265–270. https://doi.org/10.1007/s10071-011-0451-0

Reby, D., & McComb, K. (2003). Anatomical constraints generate honesty: Acoustic cues to age and weight in the roars of red deer stags. *Animal Behaviour*, *65*(3), 519–530. https://doi.org/10.1006/anbe.2003.2078

Reby, D., McComb, K., Cargnelutti, B., Darwin, C., Fitch, W. T., & Clutton-Brock, T. (2005). Red deer stags use formants as assessment cues during intrasexual agonistic interactions. *Proceedings of the Royal Society B: Biological Sciences*, *272*(1566), 941–947. https://doi.org/10.1098/rspb.2004.2954

Reynolds, V. (1965). Some behavioral comparisons between the chimpanzee and the mountain gorilla in the wild. *American Anthropologist*, *67*(3), 691–706.

Reynolds, V., & Reynolds, F. (1965). Chimpanzees of the Budongo Forest. In *Primate behavior, ed. I DeVore, New York, Holt, Rinehart and Winston* (pp. 368–424).

Riede, T., Arcadi, A. C., & Owren, M. J. (2007). Nonlinear acoustics in the pant hoots of common chimpanzees (*Pan troglodytes*): Vocalizing at the edge. *The Journal of the Acoustical Society of America*, *121*(3), 1758–1767. https://doi.org/10.1121/1.2427115

Riede, T., Owren, M. J., & Arcadi, A. C. (2004). Nonlinear acoustics in pant hoots of common chimpanzees (*Pan troglodytes*): Frequency jumps, subharmonics, biphonation, and deterministic chaos. *American Journal of Primatology*, *64*(3), 277–291. https://doi.org/10.1002/ajp.20078

Riede, T., & Stolle-Malorny, A. (1999). The Vocal Change of a Kitten with Craniocerebellar Trauma—A Case Study. *Bioacoustics*, *10*(2–3), 131–141. https://doi.org/10.1080/09524622.1999.9753426

Riede, T., Wilden, I., & Tembrock, G. (1997). Subharmonics, biphonations, and frequency jumps-common components of mammal vocalization or indicators for disorders? *Zeitschrift Fur Saugetierkunde*, *62*, 198–203.

Ripley, B., Venables, B., Bates, D. M., Hornik, K., Gebhardt, A., Firth, D., & Ripley, M. B. (2020). *Package 'MASS.'*

Rogers, T. L. (2017). Calling underwater is a costly signal: Size-related differences in the call rates of Antarctic leopard seals. *Current Zoology*, *63*(4), 433–443. https://doi.org/10.1093/cz/zox028

Root-Gutteridge, H., Cusano, D. A., Shiu, Y., Nowacek, D. P., Van Parijs, S. M., & Parks, S. E. (2018). A lifetime of changing calls: North Atlantic right whales, *Eubalaena glacialis*, refine call production as they age. *Animal Behaviour*, *137*, 21–34. https://doi.org/10.1016/j.anbehav.2017.12.016

Rosenfield, K. A., Sorokowska, A., Sorokowski, P., & Puts, D. A. (2020). Sexual selection for low male voice pitch among Amazonian forager-horticulturists. *Evolution and Human Behavior*, *41*(1), 3–11. https://doi.org/10.1016/j.evolhumbehav.2019.07.002

Rowell, T. E., & Hinde, R. A. (1962). Vocal Communication by the Rhesus Monkey (*Macaca mulatta*). *Proceedings of the Zoological Society of London*, *138*(2), 279–294. https://doi.org/10.1111/j.1469-7998.1962.tb05698.x

Rukstalis, M., Fite, J. E., & French, J. A. (2003). Social Change Affects Vocal Structure in a Callitrichid Primate (*Callithrix kuhlii*). *Ethology*, *109*(4), 327–340. https://doi.org/10.1046/j.1439-0310.2003.00875.x

Schel, A. M., Machanda, Z., Townsend, S. W., Zuberbühler, K., & Slocombe, K. E. (2013). Chimpanzee food calls are directed at specific individuals. *Animal Behaviour*, *86*(5), 955–965. https://doi.org/10.1016/j.anbehav.2013.08.013

Schneiderová, I., Schnitzerová, P., Uhlíková, J., Brandl, P., Zouhar, J., & Matějů, J. (2015). Differences in alarm calls of juvenile and adult European ground squirrels (*Spermophilus citellus*): Findings on permanently marked animals from a semi-natural enclosure. *Zoo Biology*, *34*(6), 503–512. https://doi.org/10.1002/zoo.21233

Schrader, L., & Hammerschmidt, K. (1997). Computer-Aided Analysis of Acoustic Parameters in Animal Vocalisations: A Multi-Parametric Approach. *Bioacoustics*, *7*(4), 247–265. https://doi.org/10.1080/09524622.1997.9753338

Sell, A., Bryant, G. A., Cosmides, L., Tooby, J., Sznycer, D., von Rueden, C., Krauss, A., & Gurven, M. (2010). Adaptations in humans for assessing physical strength from the voice. *Proceedings of the Royal Society B: Biological Sciences*, *277*(1699), 3509–3518. https://doi.org/10.1098/rspb.2010.0769

Serrano, J. M., Penna, M., & Soto-Azat, C. (2020). Individual and population variation of linear and non-linear components of the advertisement call of Darwin's frog

(*Rhinoderma darwinii*). *Bioacoustics*, *29*(5), 572–589. https://doi.org/10.1080/09524622.2019.1631214

Seyfarth, R. M., Cheney, D. L., & Marler, P. (1980a). Vervet monkey alarm calls: Semantic communication in a free-ranging primate. *Animal Behaviour*, *28*(4), 1070–1094. https://doi.org/10.1016/S0003-3472(80)80097-2

Seyfarth, R. M., Cheney, D. L., & Marler, P. (1980b). Monkey responses to three different alarm calls: Evidence of predator classification and semantic communication. *Science*, *210*(4471), 801–803. https://doi.org/10.1126/science.7433999

Singh, T. (2019, June 18). MFCC's Made Easy. *Medium*. https://medium.com/@tanveer9812/mfccs-made-easy-7ef383006040

Slocombe, K. E., Kaller, T., Call, J., & Zuberbühler, K. (2010). Chimpanzees Extract Social Information from Agonistic Screams. *PLOS ONE*, *5*(7), e11473. https://doi.org/10.1371/journal.pone.0011473

Slocombe, K. E., Townsend, S. W., & Zuberbühler, K. (2009). Wild chimpanzees (*Pan troglodytes schweinfurthii*) distinguish between different scream types: Evidence from a playback study. *Animal Cognition*, *12*(3), 441–449. https://doi.org/10.1007/s10071-008-0204-x

Slocombe, K. E., & Zuberbühler, K. (2005). Functionally Referential Communication in a Chimpanzee. *Current Biology*, *15*(19), 1779–1784. https://doi.org/10.1016/j.cub.2005.08.068

Slocombe, K. E., & Zuberbühler, K. (2007). Chimpanzees modify recruitment screams as a function of audience composition. *Proceedings of the National Academy of Sciences*, *104*(43), 17228–17233. https://doi.org/10.1073/pnas.0706741104

Specht, R. (2004). Avisoft SAS-Lab Pro version 5.2. *Avisoft Bioacoustics, Berlin*.

Srinivasan, V., Ramalingam, V., & Arulmozhi, P. (2014). Artificial neural network based pathological voice classification using MFCC features. *International Journal of Science, Environment and Technology*, *3*(1), 291–302.

Stathopoulos, E. T., Huber, J. E., & Sussman, J. E. (2011). Changes in Acoustic Characteristics of the Voice Across the Life Span: Measures From Individuals 4–

93 Years of Age. *Journal of Speech, Language, and Hearing Research*, *54*(4), 1011–1021. https://doi.org/10.1044/1092-4388(2010/10-0036)

Stoeger, A. S., Charlton, B. D., Kratochvil, H., & Fitch, W. T. (2011). Vocal cues indicate level of arousal in infant African elephant roars. *The Journal of the Acoustical Society of America*, *130*(3), 1700–1710. https://doi.org/10.1121/1.3605538

Strier, K. B., Altmann, J., Brockman, D. K., Bronikowski, A. M., Cords, M., Fedigan, L. M., Lapp, H., Liu, X., Morris, W. F., & Pusey, A. E. (2010). The Primate Life History Database: A unique shared ecological data resource. *Methods in Ecology and Evolution*, *1*(2), 199–211.

Stumpf, R. M., & Boesch, C. (2006). The efficacy of female choice in chimpanzees of the Taï Forest, Côte d’Ivoire. *Behavioral Ecology and Sociobiology*, *60*(6), 749–765. https://doi.org/10.1007/s00265-006-0219-8

Stumpf, R. M., & Boesch, C. (2010). Male aggression and sexual coercion in wild West African chimpanzees, *Pan troglodytes verus*. *Animal Behaviour*, *79*(2), 333–342. https://doi.org/10.1016/j.anbehav.2009.11.008

Sugiura, H. (1998). Matching of acoustic features during the vocal exchange of coo calls by Japanese macaques. *Animal Behaviour*, *55*(3), 673–687. https://doi.org/10.1006/anbe.1997.0602

Suire, A., Raymond, M., & Barkat-Defradas, M. (2018). Human vocal behavior within competitive and courtship contexts and its relation to mating success. *Evolution and Human Behavior*, *39*(6), 684–691. https://doi.org/10.1016/j.evolhumbehav.2018.07.001

Suire, A., Raymond, M., & Barkat-Defradas, M. (2019). Male Vocal Quality and Its Relation to Females’ Preferences. *Evolutionary Psychology*, *17*(3), 1474704919874675. https://doi.org/10.1177/1474704919874675

Suire, A., Tognetti, A., Durand, V., Raymond, M., & Barkat-Defradas, M. (2020). Speech Acoustic Features: A Comparison of Gay Men, Heterosexual Men, and Heterosexual Women. *Archives of Sexual Behavior*, *49*(7), 2575–2583. https://doi.org/10.1007/s10508-020-01665-3

Takahashi, D. Y., Fenley, A. R., Teramoto, Y., Narayanan, D. Z., Borjon, J. I., Holmes, P., & Ghazanfar, A. A. (2015). The developmental dynamics of marmoset monkey vocal production. *Science*, *349*(6249), 734–738. https://doi.org/10.1126/science.aab1058

Textor, J., van der Zander, B., Gilthorpe, M. S., Liśkiewicz, M., & Ellison, G. T. (2016). Robust causal inference using directed acyclic graphs: The R package 'dagitty.' *International Journal of Epidemiology*, *45*(6), 1887–1894.

Tokuda, I. T. (2018). Non-linear dynamics in mammalian voice production. *Anthropological Science*, *126*(1), 35–41. https://doi.org/10.1537/ase.171130

Townsend, S. W., & Manser, M. B. (2011). The function of nonlinear phenomena in meerkat alarm calls. *Biology Letters*, *7*(1), 47–49. https://doi.org/10.1098/rsbl.2010.0537

Tyack, P. L. (2020). A taxonomy for vocal learning. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *375*(1789), 20180406. https://doi.org/10.1098/rstb.2018.0406

Tyack, P. L., & Sayigh, L. S. (1997). 11 Vocal learning in cetaceans. *Social Influences on Vocal Development*, 208.

Tyson, R. B., Nowacek, D. P., & Miller, P. J. O. (2007). Nonlinear phenomena in the vocalizations of North Atlantic right whales (*Eubalaena glacialis*) and killer whales (*Orcinus orca*). *The Journal of the Acoustical Society of America*, *122*(3), 1365–1373. https://doi.org/10.1121/1.2756263

Uhlenbroek, C. (1996). *The structure and function of the long-distance calls given by male chimpanzees in Gombe National Park.* University of Bristol.

Valletta, J. J., Torney, C., Kings, M., Thornton, A., & Madden, J. (2017). Applications of machine learning in animal behaviour studies. *Animal Behaviour*, *124,* 203–220. https://doi.org/10.1016/j.anbehav.2016.12.005

Vernes, S. C., Janik, V. M., Fitch, W. T., & Slater, P. J. B. (2021). Vocal learning in animals and humans. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *376*(1836), 20200234. https://doi.org/10.1098/rstb.2020.0234

Viljoen, J. J., Ganswindt, A., Reynecke, C., Stoeger, A. S., & Langbauer, W. R. (2015). Vocal stress associated with a translocation of a family herd of African elephants (*Loxodonta africana*) in the Kruger National Park, South Africa. *Bioacoustics*, *24*(1), 1–12. https://doi.org/10.1080/09524622.2014.906320

Volodin, I. A., Matrosova, V. A., Frey, R., Kozhevnikova, J. D., Isaeva, I. L., & Volodina, E. V. (2018). Altai pika (*Ochotona alpina*) alarm calls: Individual acoustic variation and the phenomenon of call-synchronous ear folding behavior. *The Science of Nature*, *105*(7), 40. https://doi.org/10.1007/s00114-018-1567-8

Volodina, E. V., Volodin, I. A., Isaeva, I. V., & Unck, C. (2006). Biphonation May Function to Enhance Individual Recognition in the Dhole, *Cuon alpinus*. *Ethology*, *112*(8), 815–825. https://doi.org/10.1111/j.1439-0310.2006.01231.x

Wadewitz, P., Hammerschmidt, K., Battaglia, D., Witt, A., Wolf, F., & Fischer, J. (2015). Characterizing Vocal Repertoires—Hard vs. Soft Classification Approaches. *PLOS ONE*, *10*(4), e0125785. https://doi.org/10.1371/journal.pone.0125785

Watkins, W. A., Moore, K. E., Clark, C. W., & Dahlheim, M. E. (1988). The Sounds of Sperm Whale Calves. In P. E. Nachtigall & P. W. B. Moore (Eds.), *Animal Sonar: Processes and Performance* (pp. 99–107). Springer US. https://doi.org/10.1007/978-1-4684-7493-0_11

Watson, S. K., Townsend, S. W., Schel, A. M., Wilke, C., Wallace, E. K., Cheng, L., West, V., & Slocombe, K. E. (2015a). Vocal Learning in the Functionally Referential Food Grunts of Chimpanzees. *Current Biology*, *25*(4), 495–499. https://doi.org/10.1016/j.cub.2014.12.032

Watson, S. K., Townsend, S. W., Schel, A. M., Wilke, C., Wallace, E. K., Cheng, L., West, V., & Slocombe, K. E. (2015b). Reply to Fischer et al. *Current Biology*, *25*(21), R1030–R1031. https://doi.org/10.1016/j.cub.2015.09.024

Watts, D. P., Muller, M., Amsler, S. J., Mbabazi, G., & Mitani, J. C. (2006). Lethal intergroup aggression by chimpanzees in Kibale National Park, Uganda. *American Journal of Primatology*, *68*(2), 161–180. https://doi.org/10.1002/ajp.20214

Webster, M. M., & Rutz, C. (2020). How STRANGE are your study animals? *Nature*, *582*(7812), 337–340. https://doi.org/10.1038/d41586-020-01751-5

Welham, N. V., & Maclagan, M. A. (2003). Vocal fatigue: Current knowledge and future directions. *Journal of Voice: Official Journal of the Voice Foundation*, *17*(1), 21–30. https://doi.org/10.1016/s0892-1997(03)00033-x

Wheeler, B. C., & Fischer, J. (2012). Functionally referential signals: A promising paradigm whose time has passed. *Evolutionary Anthropology: Issues, News, and Reviews*, *21*(5), 195–205. https://doi.org/10.1002/evan.21319

Whiten, A., Goodall, J., McGrew, W. C., Nishida, T., Reynolds, V., Sugiyama, Y., Tutin, C. E. G., Wrangham, R. W., & Boesch, C. (1999). Cultures in chimpanzees. *Nature*, *399*(6737), 682–685. https://doi.org/10.1038/21415

Wich, S. A., Swartz, K. B., Hardus, M. E., Lameira, A. R., Stromberg, E., & Shumaker, R. W. (2009). A case of spontaneous acquisition of a human sound by an orangutan. *Primates*, *50*(1), 56–64. https://doi.org/10.1007/s10329-008-0117-y

Wich, S. A., Van der Post, D. J., Heistermann, M., Möhle, U., Van Hooff, J., & Sterck, E. H. M. (2003). Life-phase related changes in male loud call characteristics and testosterone levels in wild Thomas langurs. *International Journal of Primatology*, *24*(6), 1251–1265.

Wiggins, B. J., & Chrisopherson, C. D. (2019). The replication crisis in psychology: An overview for theoretical and philosophical psychology. *Journal of Theoretical and Philosophical Psychology*, *39*(4), 202–217. https://doi.org/10.1037/teo0000137

Wilden, I., Herzel, H., Peters, G., & Tembrock, G. (1998). Subharmonics, Biphonation, and Deterministic Chaos in Mammal Vocalization. *Bioacoustics*, *9*(3), 171–196. https://doi.org/10.1080/09524622.1998.9753394

Wilson, M. L. (2002). *Imbalances of power: How chimpanzees respond to the threat of intergroup aggression (Pan troglodytes).*

Wilson, M. L., Boesch, C., Fruth, B., Furuichi, T., Gilby, I. C., Hashimoto, C., Hobaiter, C. L., Hohmann, G., Itoh, N., Koops, K., Lloyd, J. N., Matsuzawa, T., Mitani, J. C., Mjungu, D. C., Morgan, D., Muller, M. N., Mundry, R., Nakamura, M., Pruetz, J., … Wrangham, R. W. (2014). Lethal aggression in *Pan* is better explained by adaptive strategies than human impacts. *Nature*, *513*(7518), 414–417. https://doi.org/10.1038/nature13727

Wilson, M. L., Hauser, M. D., & Wrangham, R. W. (2001). Does participation in intergroup conflict depend on numerical assessment, range location, or rank for wild chimpanzees? *Animal Behaviour*, *61*(6), 1203–1216. https://doi.org/10.1006/anbe.2000.1706

Wilson, M. L., Hauser, M. D., & Wrangham, R. W. (2007). Chimpanzees (*Pan troglodytes*) Modify Grouping and Vocal Behaviour in Response to Location-Specific Risk. *Behaviour*, *144*(12), 1621–1653.

Wilson, M. L., Kahlenberg, S. M., Wells, M., & Wrangham, R. W. (2012). Ecological and social factors affect the occurrence and outcomes of intergroup encounters in chimpanzees. *Animal Behaviour*, *83*(1), 277–291. https://doi.org/10.1016/j.anbehav.2011.11.004

Wilson, M. L., Lonsdorf, E. V., Mjungu, D. C., Kamenya, S., Kimaro, E. W., Collins, D. A., Gillespie, T. R., Travis, D. A., Lipende, I., Mwacha, D., Ndimuligo, S. A., Pintea, L., Raphael, J., Mtiti, E. R., Hahn, B. H., Pusey, A. E., & Goodall, J. (2020). Research and conservation in the greater Gombe ecosystem, Uganda: Challenges and opportunities. *Biological Conservation*, *252*, 108853. https://doi.org/10.1016/j.biocon.2020.108853

Wilson, M. L., Wallauer, W. R., & Pusey, A. E. (2004). New Cases of Intergroup Violence Among Chimpanzees in Gombe National Park, Tanzania. *International Journal of Primatology*, *25*(3), 523–549. https://doi.org/10.1023/B:IJOP.0000023574.38219.92

Winter, P., Handley, P., Ploog, D., & Schott, D. (1973). Ontogeny of squirrel monkey calls under normal conditions and under acoustic isolation. *Behaviour*, *47*(3), 230–239.

Winter, P., Ploog, D., & Latta, J. (1966). Vocal repertoire of the squirrel monkey (*Saimiri sciureus*), its analysis and significance. *Experimental Brain Research*, *1*(4), 359–384. https://doi.org/10.1007/BF00237707

Wolff, S. E., & Puts, D. A. (2010). Vocal masculinity is a robust dominance signal in men. *Behavioral Ecology and Sociobiology*, *64*(10), 1673–1683. https://doi.org/10.1007/s00265-010-0981-5

Wrangham, R. W. (1977). Feeding behaviour of chimpanzees in Gombe national park, Tanzania. In *Primate ecology* (pp. 503–538).

Wrangham, R. W., Clark, A. P., & Isabirye-Basuta, G. (1992). *Female social relationships and social organization of Kibale Forest chimpanzees. In (T. Nishida, WC McGrew, P. Marler, M. Pickford & FBM de Waal, Eds) Topics in Primatology. Vol. 1: Human Origins*. Tokyo: University of Tokyo Press.

Xie, J., Towsey, M., Zhang, J., & Roe, P. (2016). Acoustic classification of Australian frogs based on enhanced features and machine learning algorithms. *Applied Acoustics*, *113*(Supplement C), 193–201. https://doi.org/10.1016/j.apacoust.2016.06.029

Zahavi, A. (1975). Mate selection—A selection for a handicap. *Journal of Theoretical Biology*, *53*(1), 205–214. https://doi.org/10.1016/0022-5193(75)90111-3

Zhang, Y.-J., Huang, J.-F., Gong, N., Ling, Z.-H., & Hu, Y. (2018). Automatic detection and classification of marmoset vocalizations using deep and recurrent neural networks. *The Journal of the Acoustical Society of America*, *144*(1), 478–487. https://doi.org/10.1121/1.5047743

Zhao, X., Jiang, T., Gu, H., Liu, H., Sun, C., Liu, Y., & Feng, J. (2018). Are aggressive vocalizations the honest signals of body size and quality in female Asian particoloured bats? *Behavioral Ecology and Sociobiology*, *72*(6), 96. https://doi.org/10.1007/s00265-018-2510-x

Zolnay, A., Schluter, R., & Ney, H. (2005). *Acoustic feature combination for robust speech recognition*. *1*, I–457.

Zürcher, Y., & Burkart, J. M. (2017). Evidence for Dialects in Three Captive Populations of Common Marmosets (*Callithrix jacchus*). *International Journal of Primatology*, *38*(4), 780–793. https://doi.org/10.1007/s10764-017-9979-4

Appendix

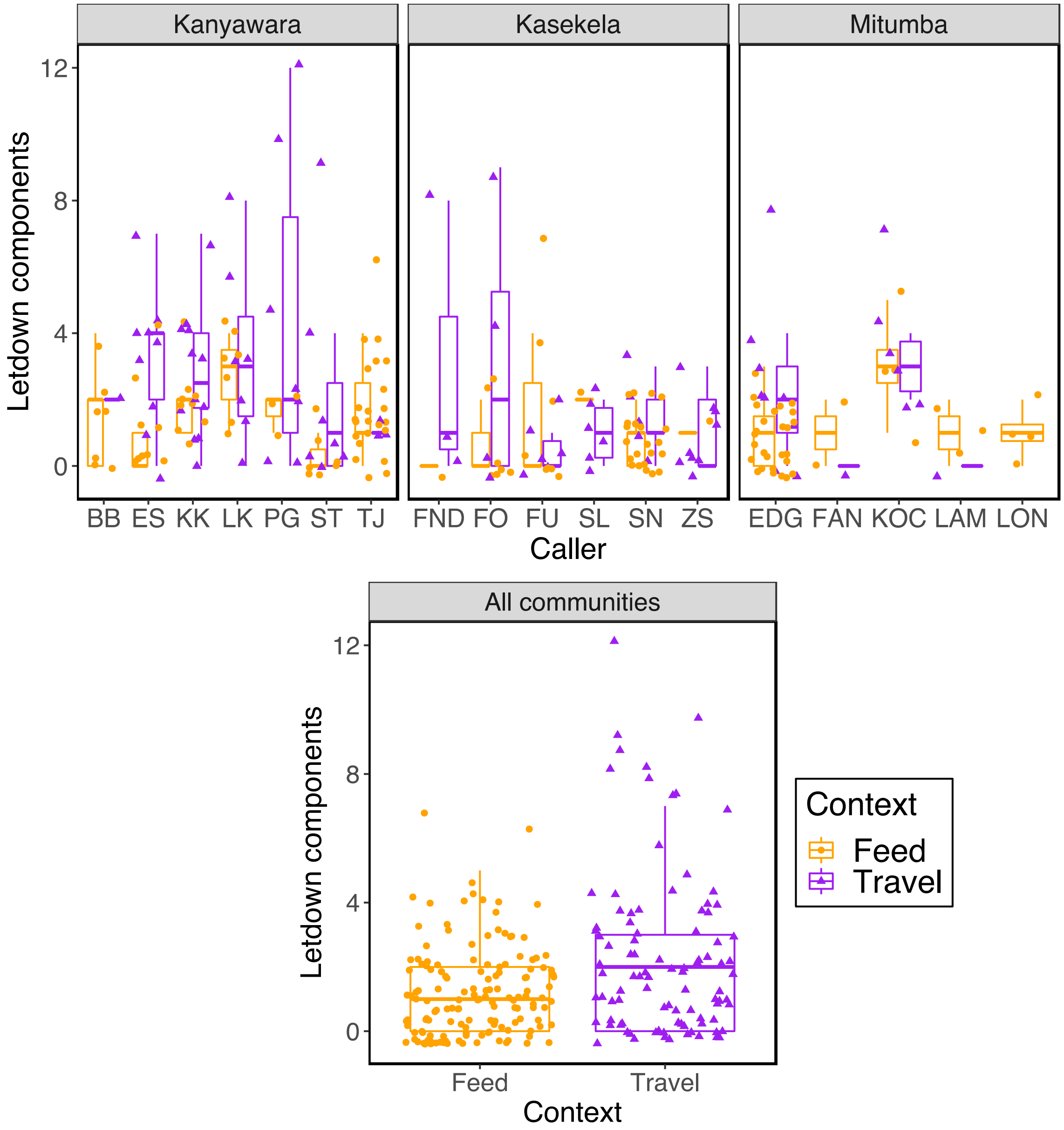


Figure 24: Differences in the number of letdown components between contexts at individual and community levels.

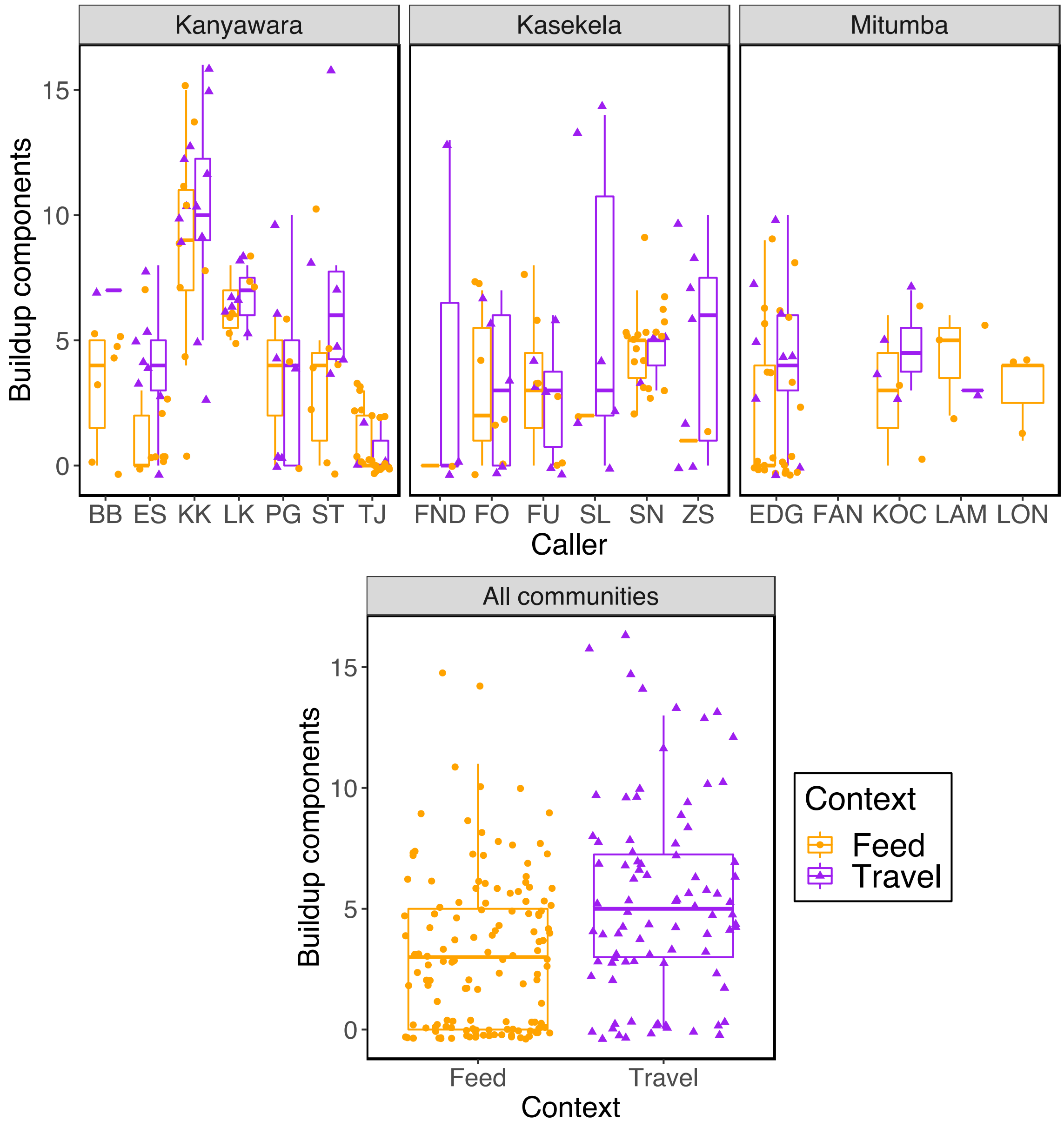


Figure 25: Differences in the number of build-up components between contexts at individual and community levels.

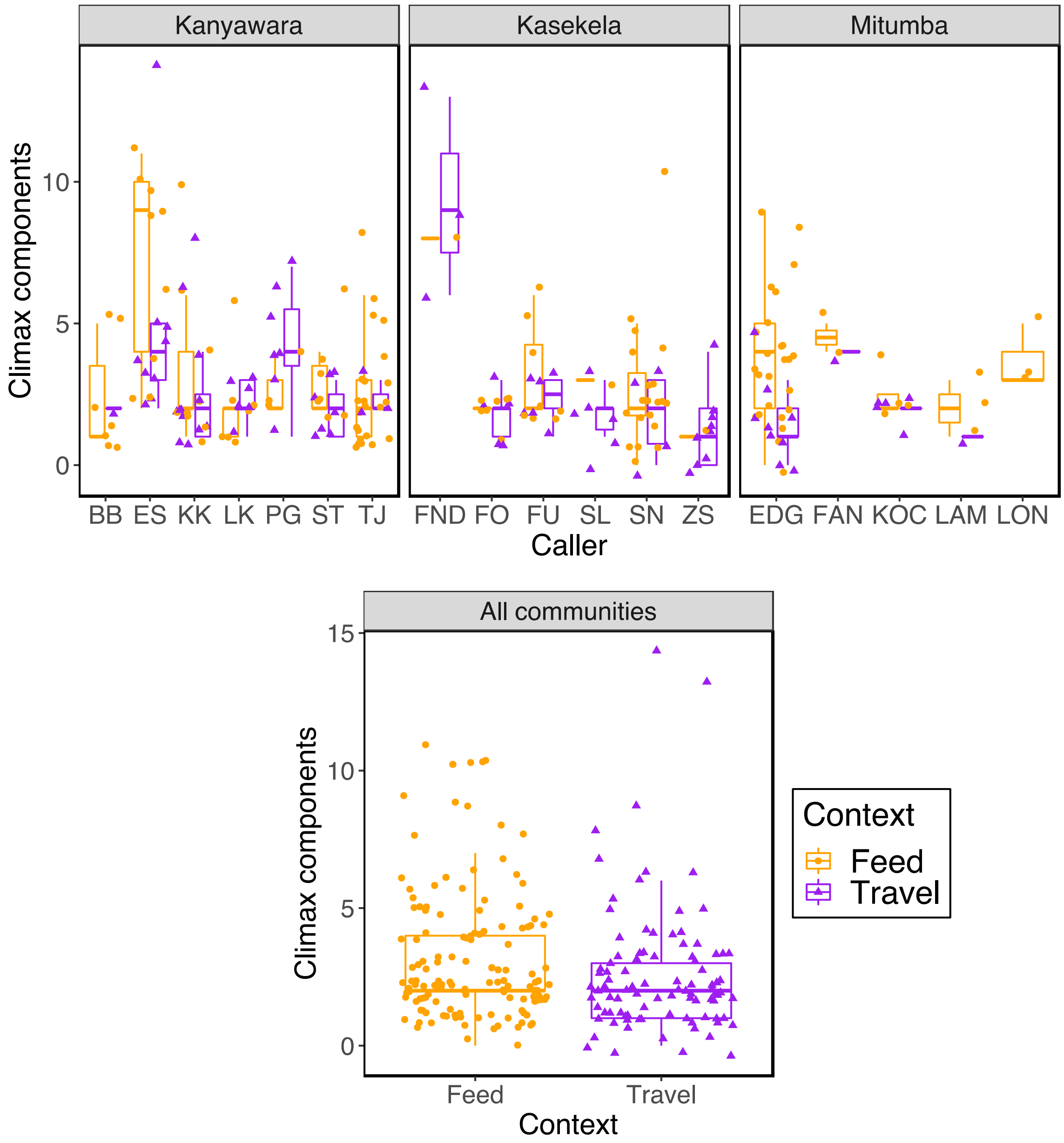


Figure 26: Differences in the number of climax components between contexts at individual and community levels.

Performing the analyses in Table 16 after removing outliers (points outside 2SD from mean on PC1 or PC2):

i) For climax screams: the 3-community analysis p-value changed from 0.016 (in Table 16) to 0.019. The 2-neighboring community analysis p-value changed from 0.089 to 0.076 (in Table 16).

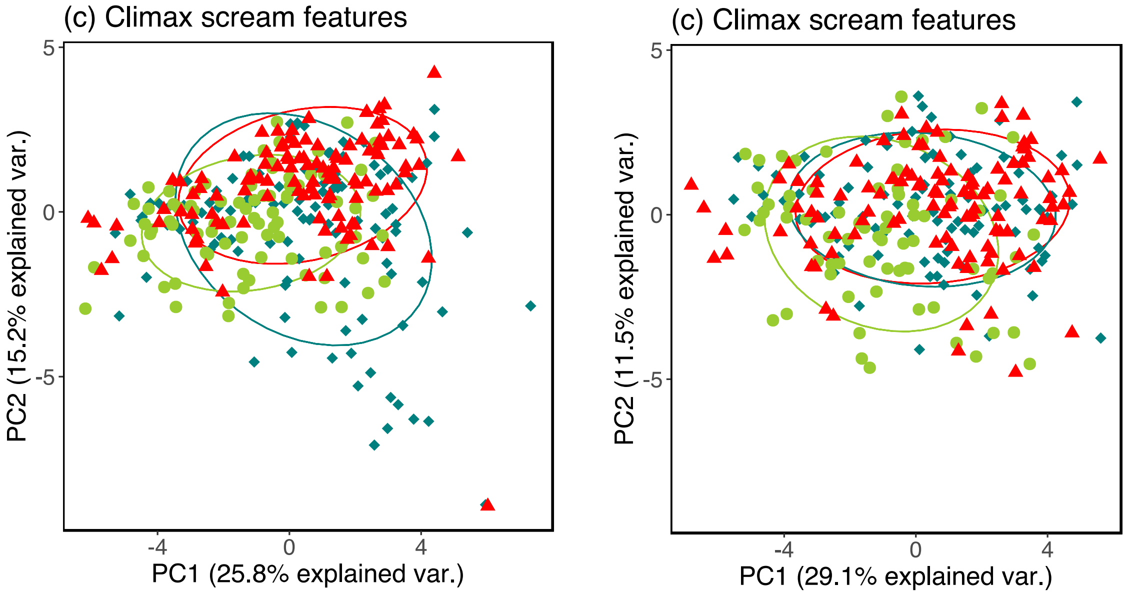


Figure 27: Climax scream features (Figure 11 (c) in the chapter 2) with (left) and without (right) outliers.

ii) For entire calls: the 3-community analysis p-value changed from 0.079 (in Table 16) to 0.089. The 2-neighboring community analysis p-value changed from 0.272 (in Table 16) to 0.215.

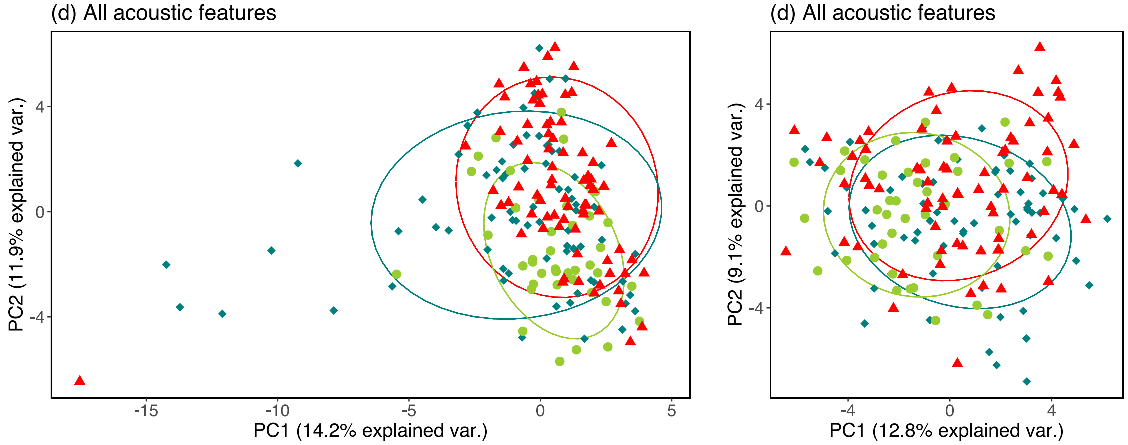


Figure 28: Acoustic features from entire calls (Figure 11 (d) in the chapter 2) with (left) and without (right) outliers.

Table 34: Mean ± SD of acoustic features found to have group differences in Crockford et al. (2004)[a], Mitani et al. (1999)[b], Mitani et al. (1992)[c]:

Acoustic variable	Context	Community		
		Kanyawara	Kasekela	Mitumba
Climax: scream duration[a]	Feed	781.14 ± 249.02	988.11 ± 237.22	680.56 ± 189.45
	Travel	677.75 ± 192.09	864.56 ± 223.24	497.89 ± 196.51
Climax: maximum peak frequency[a]	Feed	2308.23 ± 1003.87	3308.15 ± 2042.71	2194.85 ± 1384.31
	Travel	2359.54 ± 1548.16	3152.52 ± 2008.58	1454.53 ± 967.32
Climax: minimum peak frequency[a]	Feed	554.02 ± 164.16	504.59 ± 200.6	545.03 ± 216.51
Climax: mean tonal quality[a]	Travel	124.43 ± 47.75	91.79 ± 34.52	96.41 ± 51.29
Climax: maximum tonal quality[a]	Feed	463.48 ± 195.96	370.28 ± 133.33	308.74 ± 131.3

Climax: minimum F0[a]	Travel	656 ± 131.35	637.28 ± 179.35	594.12 ± 116.94
Climax: maximum F0[b]	Feed	1328.43 ± 345.35	1180.21 ± 235.17	1082.72 ± 303.94
	Travel	1293.59 ± 302.2	1256.03 ± 363.49	1091.29 ± 270.06
Climax: mean F0[b,c]	Feed	1108.35 ± 267.16	970.74 ± 199.7	884.69 ± 259.74
	Travel	1067.63 ± 272.65	998.24 ± 279.03	864.47 ± 207.09
Climax: frequency range of F0[c]	Feed	649.2 ± 256.29	561.18 ± 198.22	490.82 ± 193.71
	Travel	637.59 ± 209.71	618.76 ± 258.78	497.18 ± 234.3
Buildup: phase duration[b]	Feed	1.38 ± 1.48	2.12 ± 1.39	1.3 ± 1.57
	Travel	2.57 ± 1.69	2.07 ± 2.02	2.18 ± 1.35
Build-up: element duration[b,c]	Feed	301.06 ± 67.02	361.86 ± 103.44	342.6 ± 89.1
	Travel	280.85 ± 59.1	340.85 ± 71.58	345.65 ± 87.72
Build-up: rate[b,c]	Feed	1.35 ± 1.21	1.73 ± 0.91	1.06 ± 1.12
	Travel	2.19 ± 0.99	1.49 ± 1.09	1.75 ± 0.81
Build-up: rate of first half[a]	Feed	1.24 ± 1.15	1.49 ± 0.77	1.03 ± 1.11
Build-up: acceleration rate[a]	Travel	-0.02 ± 0.4	0.1 ± 0.38	0.08 ± 0.2
Build-up: number of exhalation elements[a]	Travel	6.36 ± 4.28	4.2 ± 4.17	4.36 ± 2.68

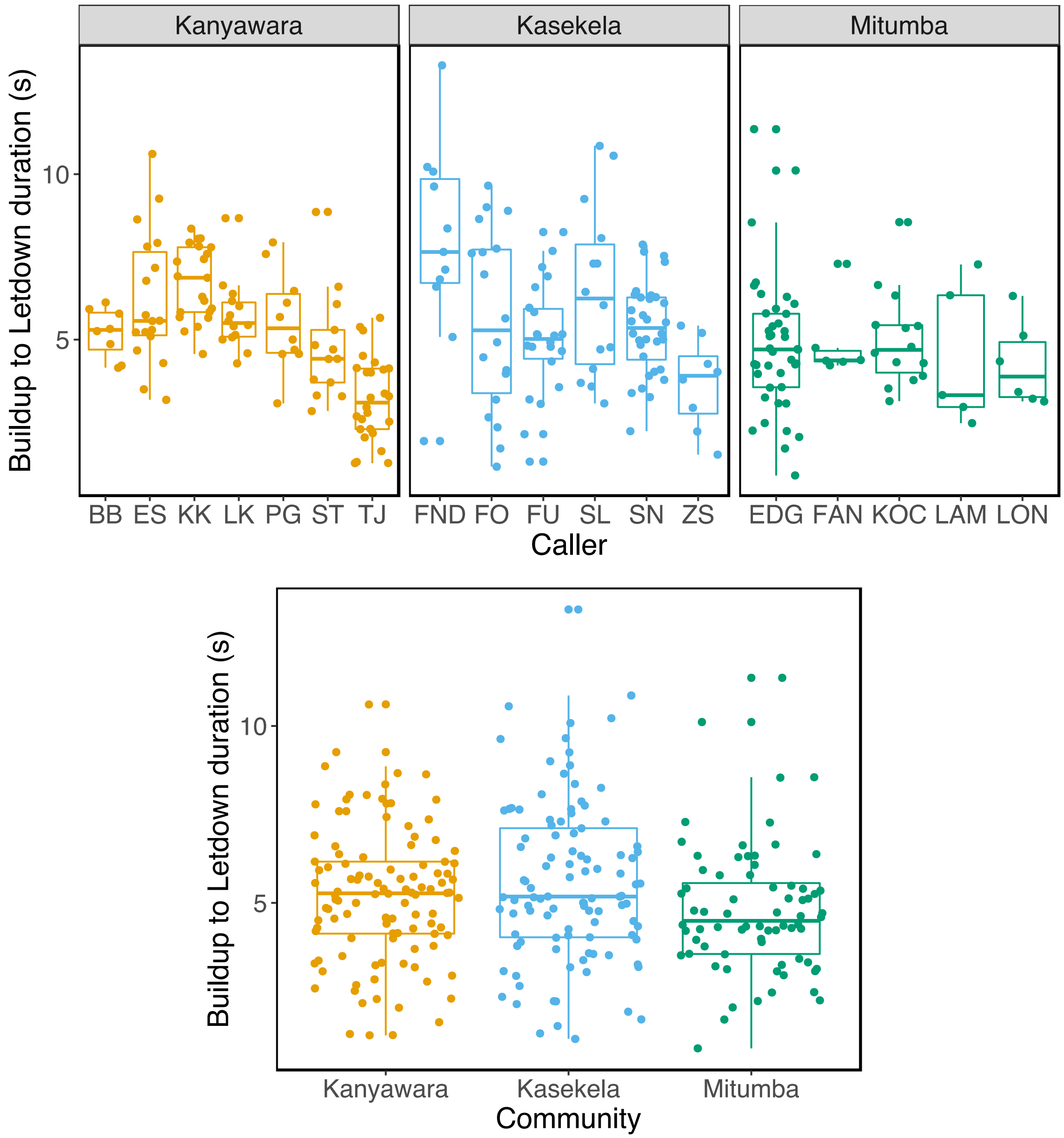


Figure 29: Buildup to letdown duration at individual and community levels.

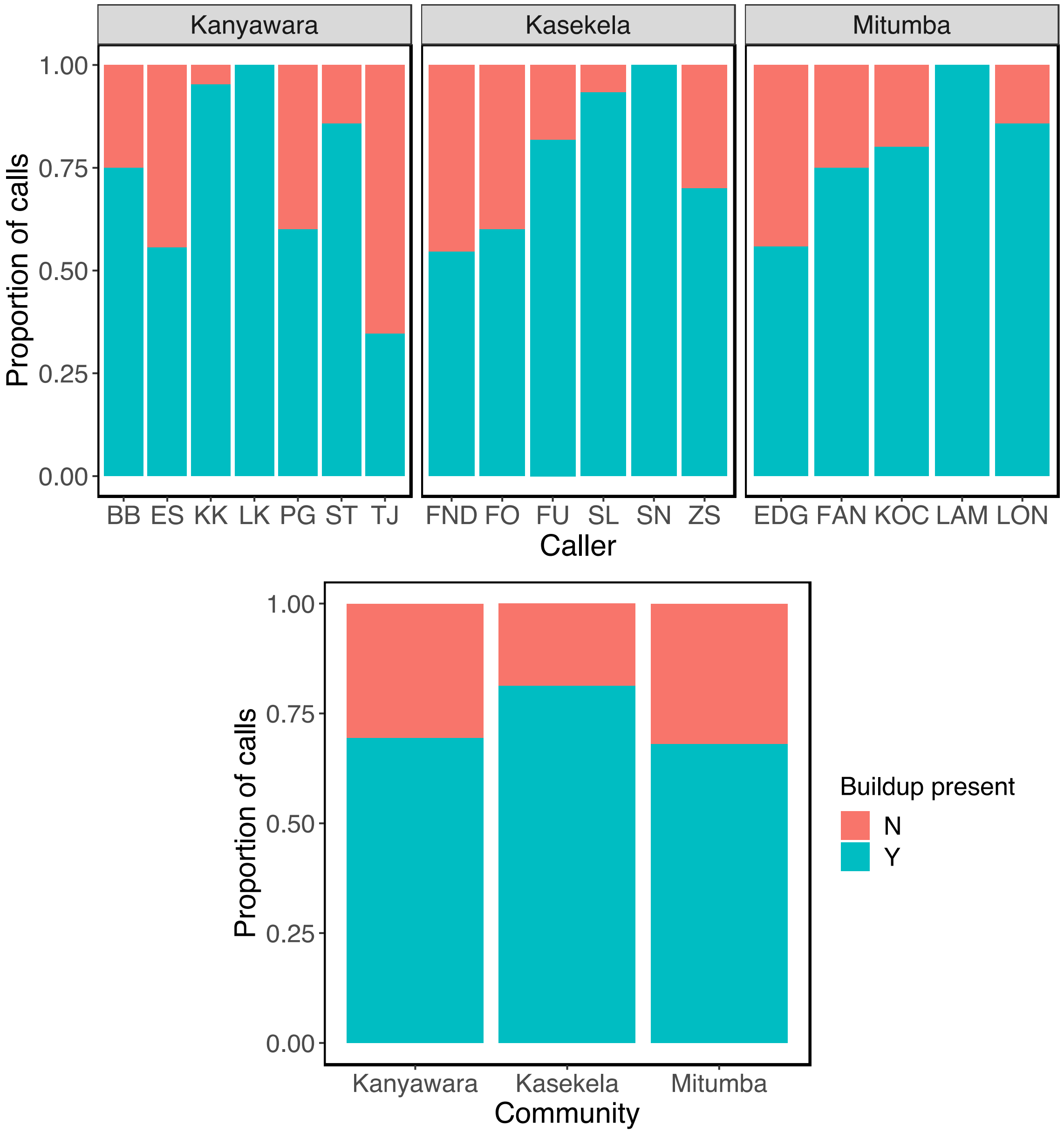


Figure 30: Proportion of calls with build-up at individual and community levels.

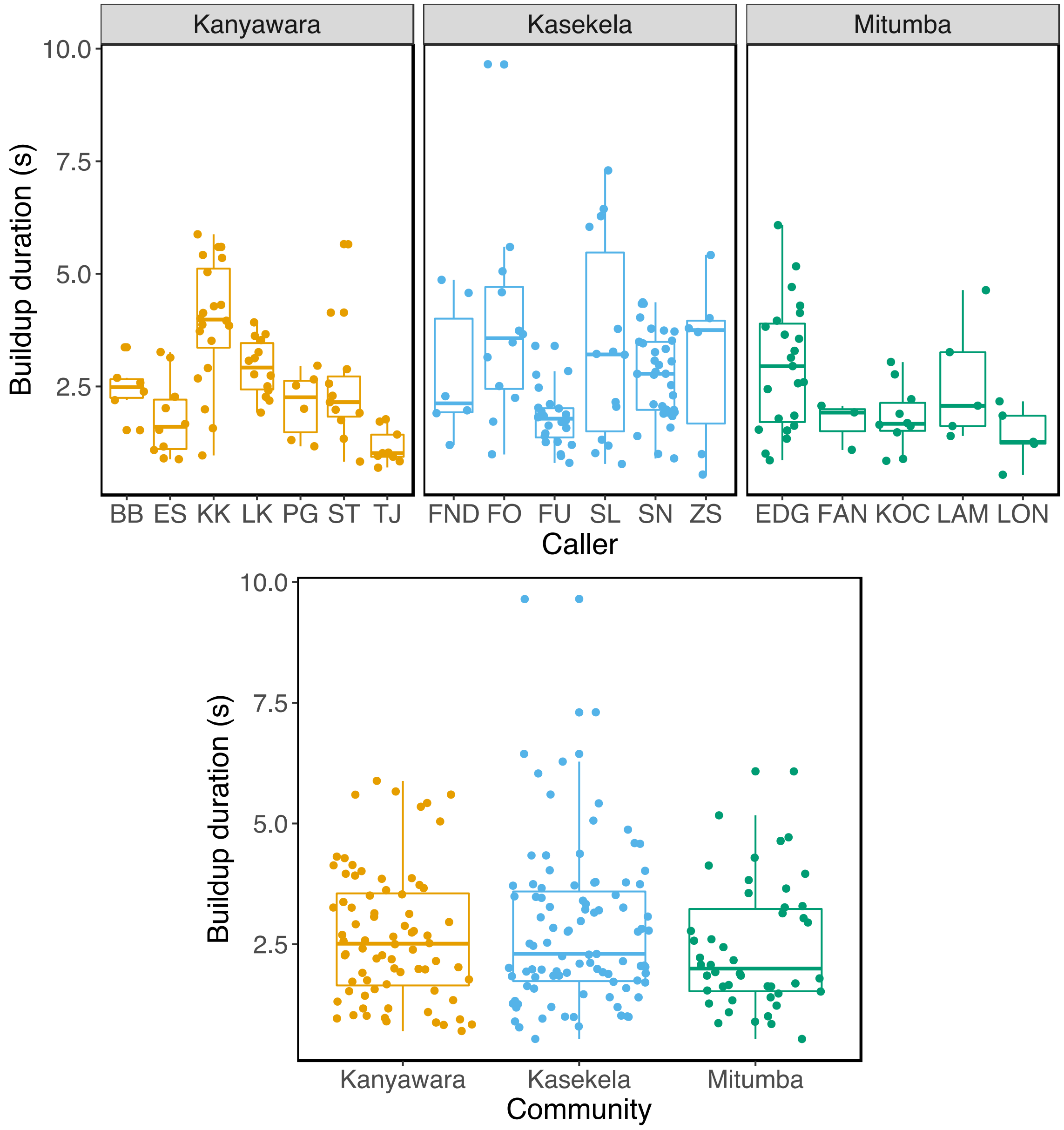


Figure 31: Build-up duration at individual and community levels.

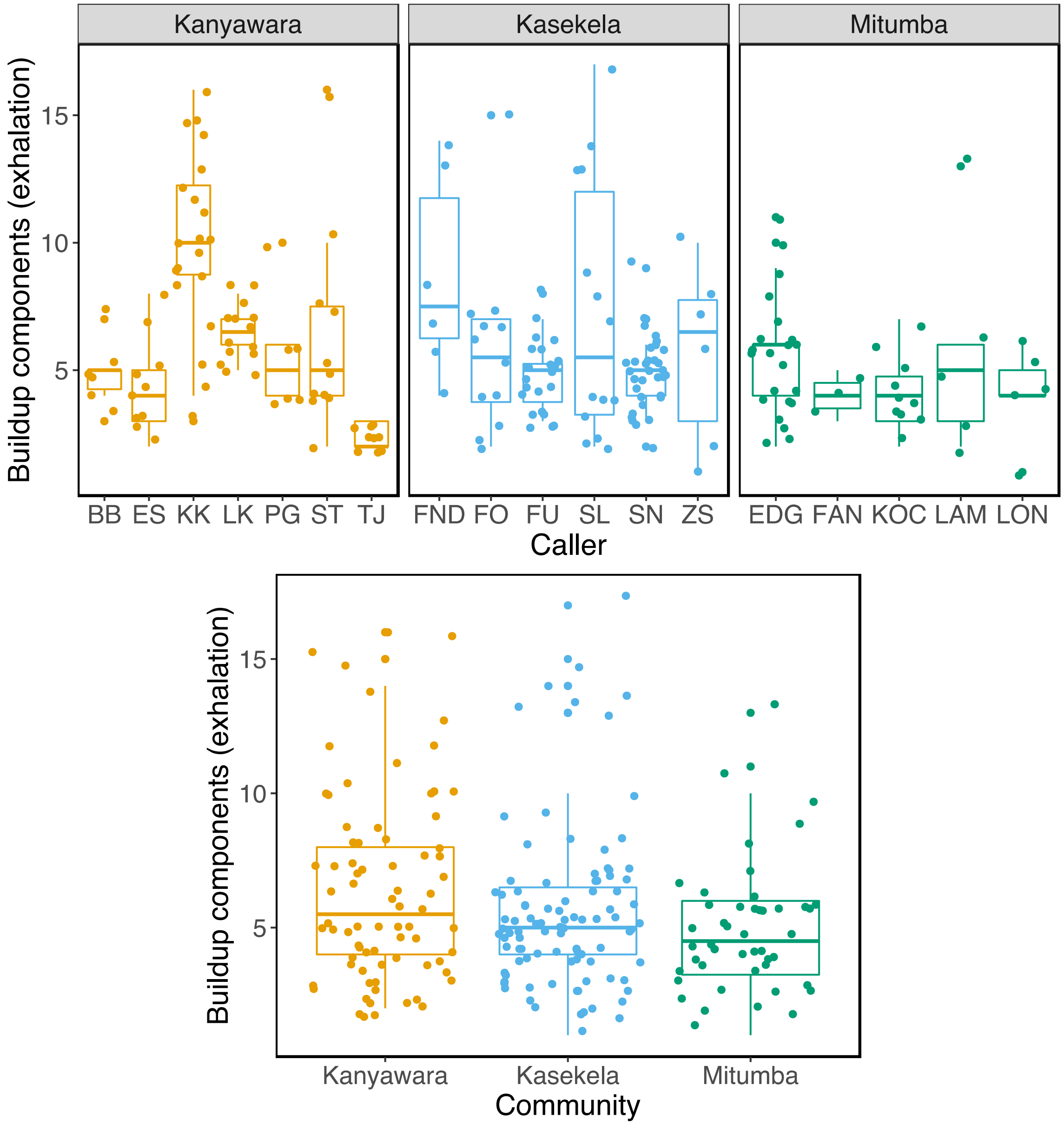
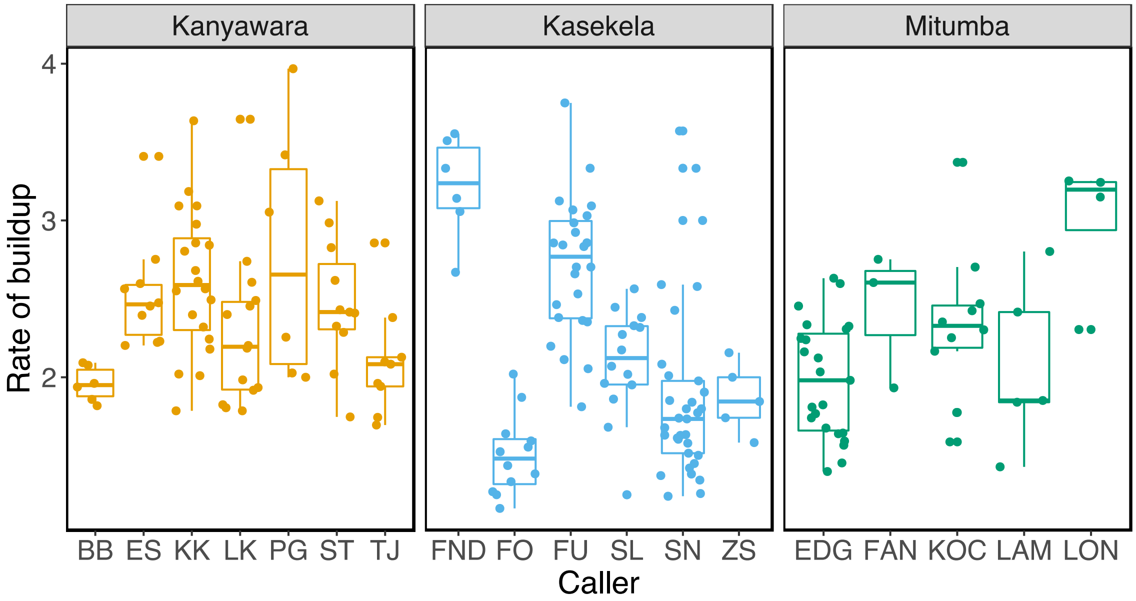


Figure 32: Number of build-up exhalation elements at individual and community levels.



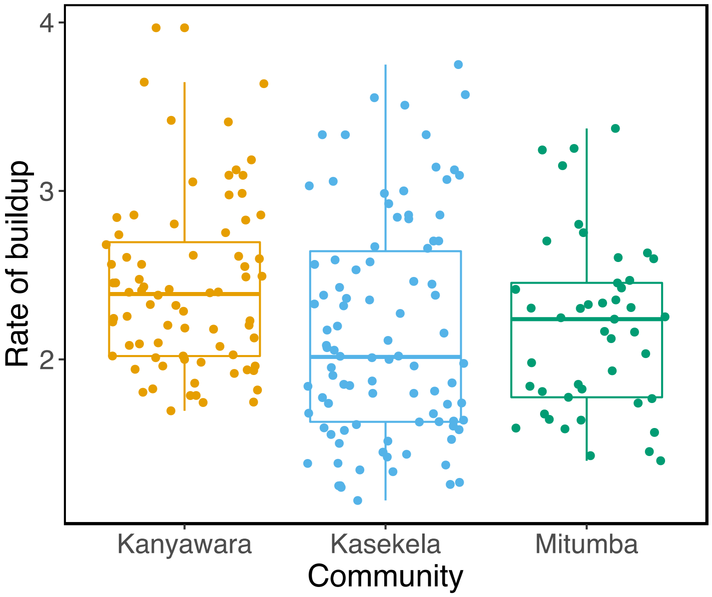


Figure 33: Rate of build-up at individual and community levels.

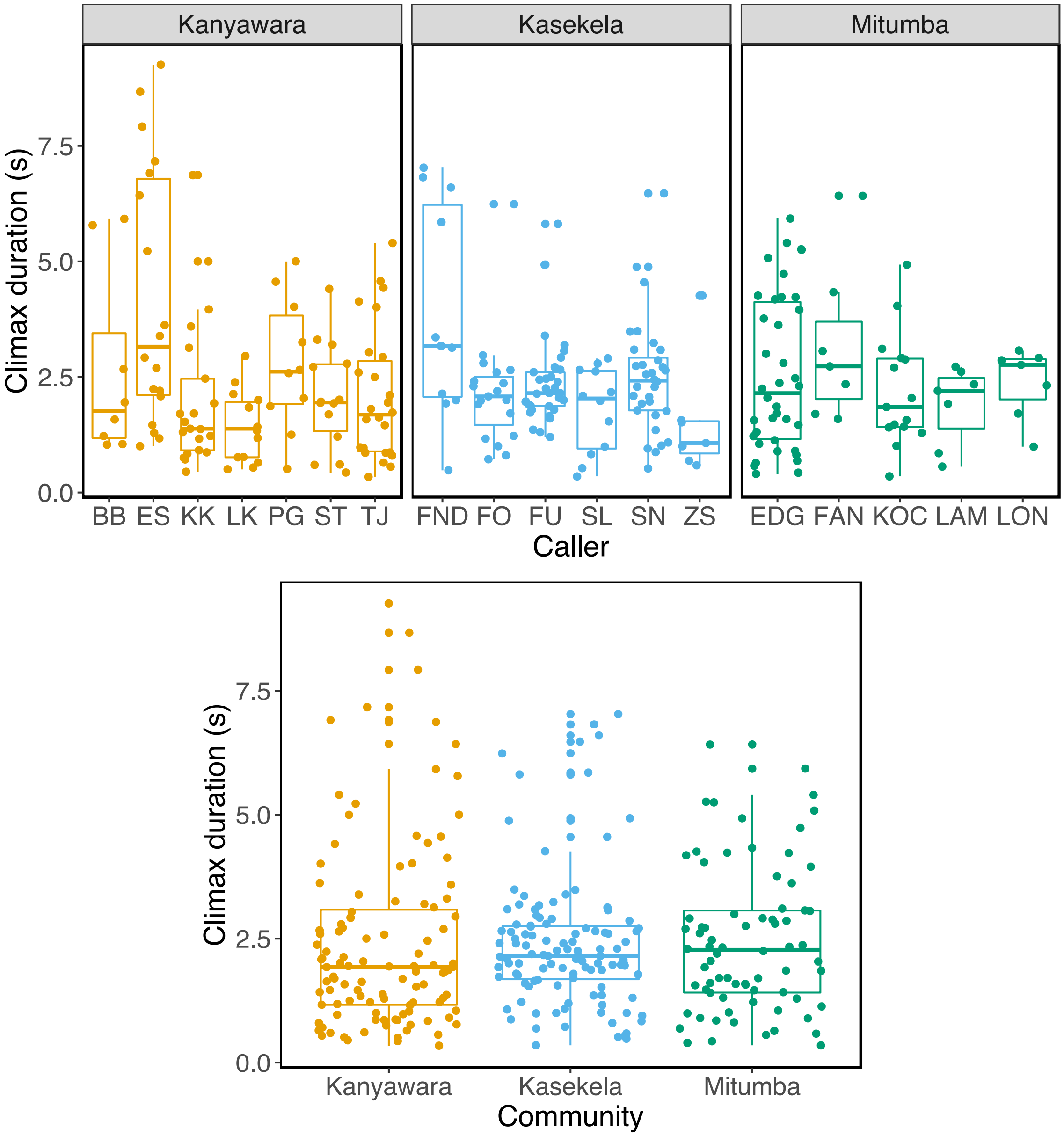


Figure 34: Climax duration at individual and community levels.

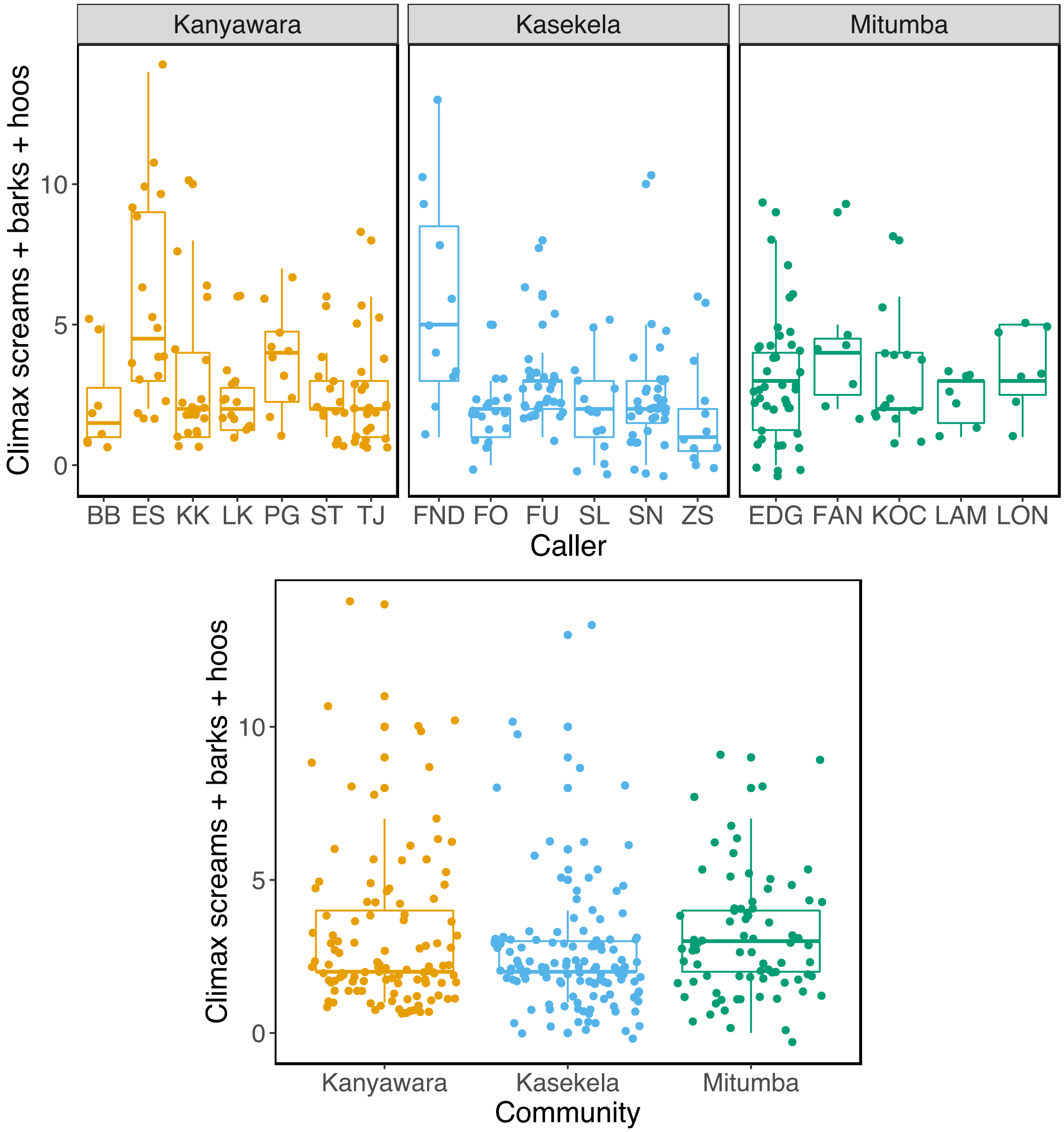


Figure 35: Number of climax components at individual and community levels.

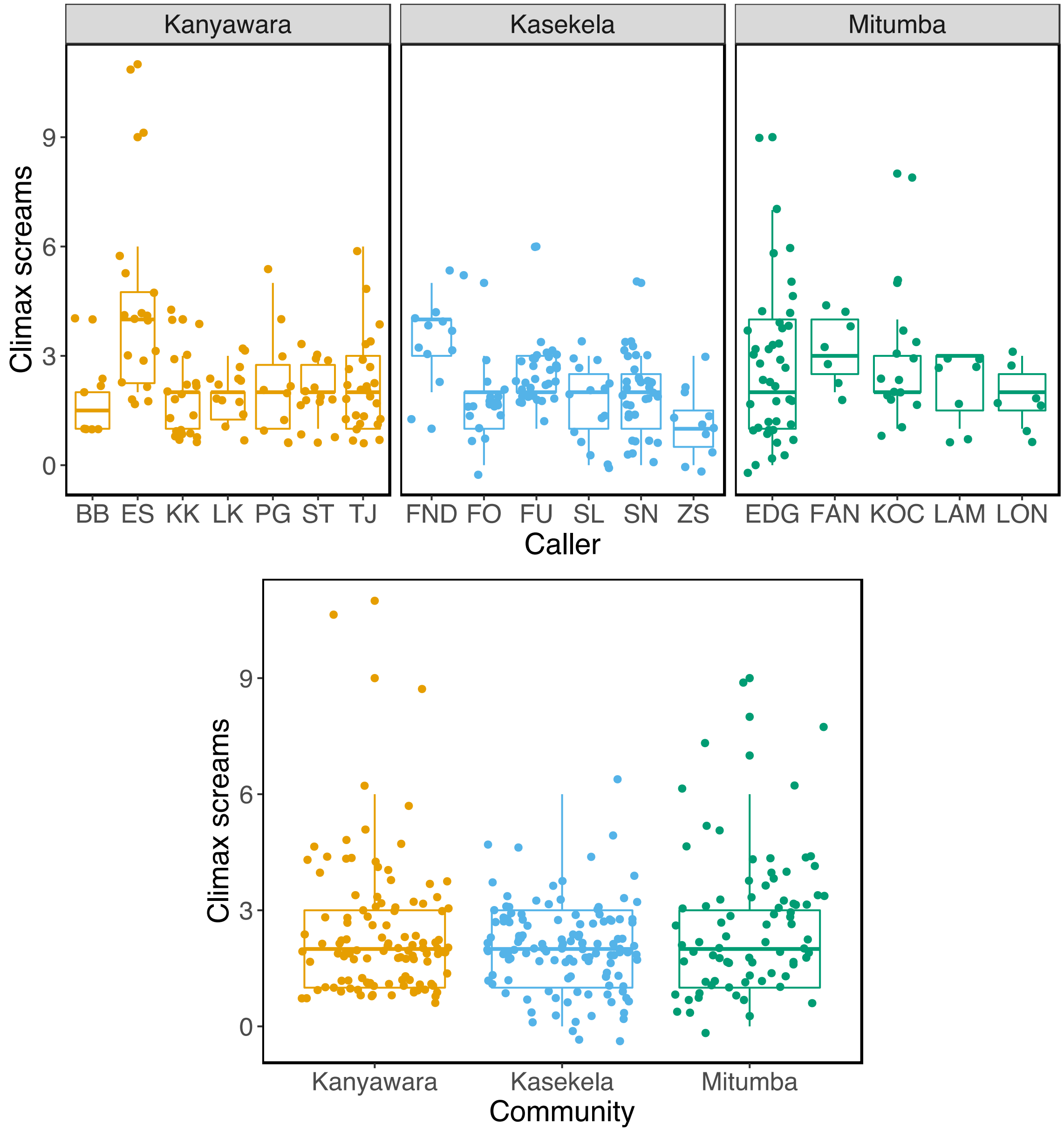


Figure 36: Number of climax screams at individual and community levels.

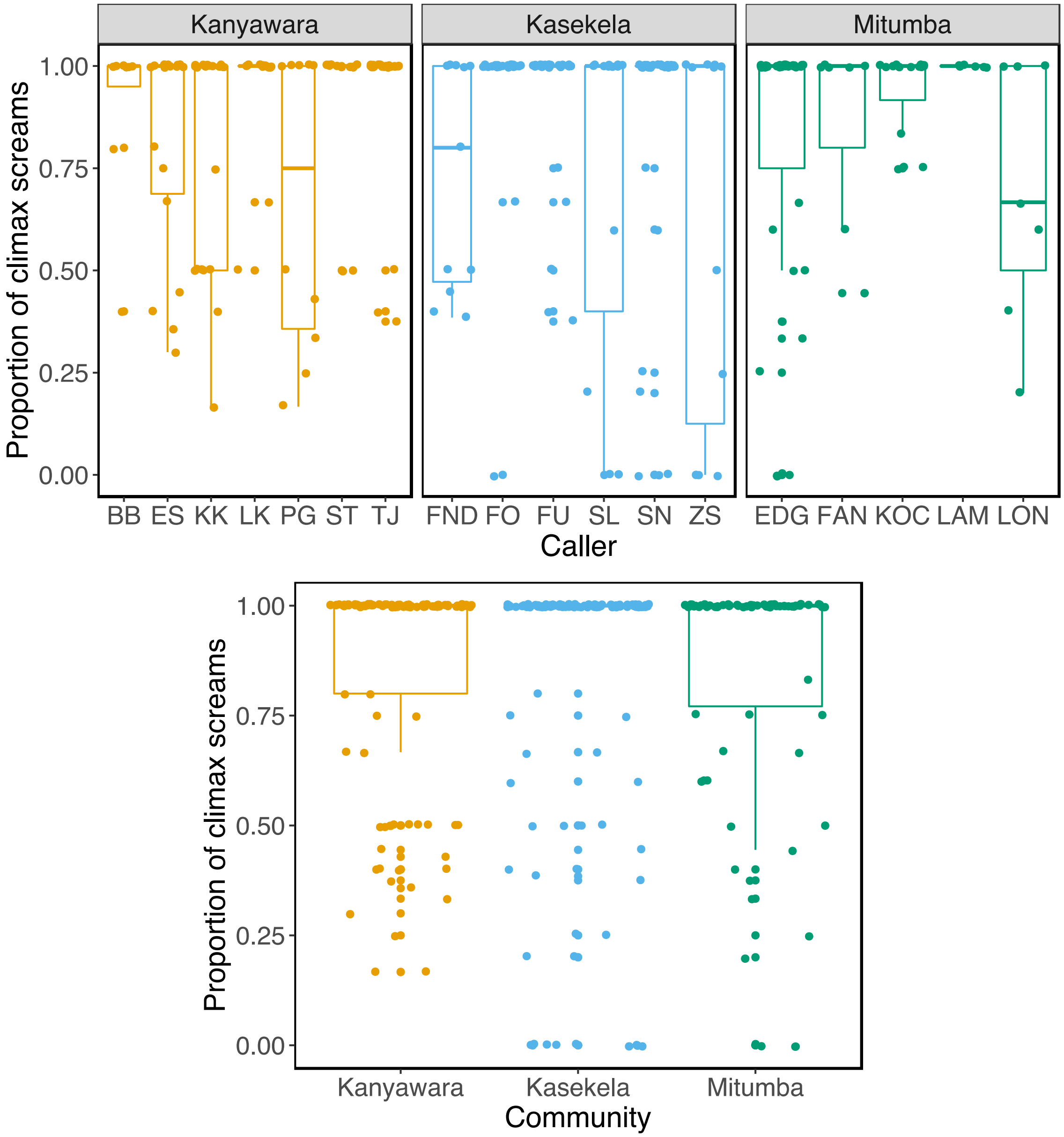


Figure 37: Proportion of climax components that are screams at individual and community levels.

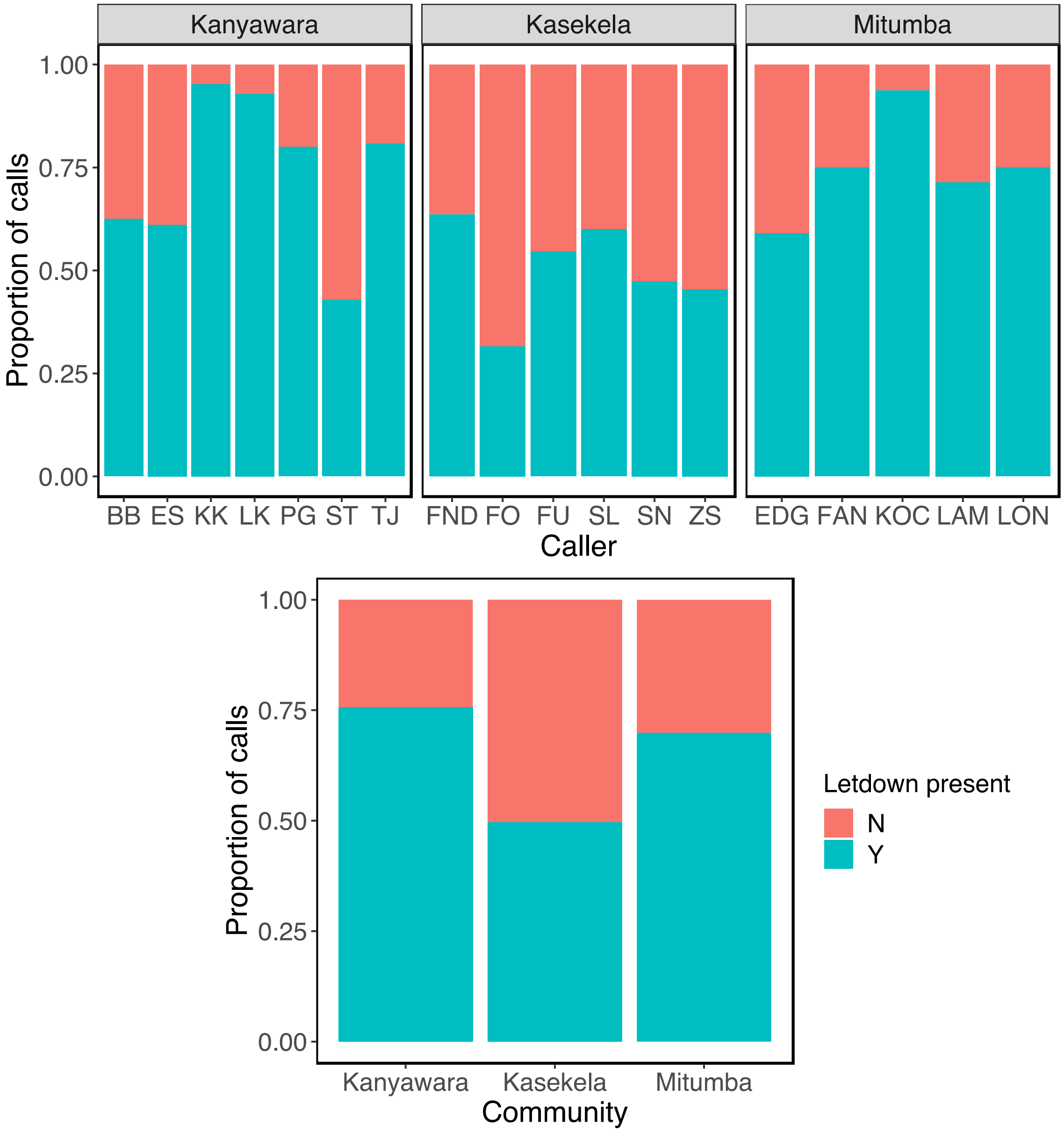


Figure 38: Proportion of calls with letdown present at individual and community levels.

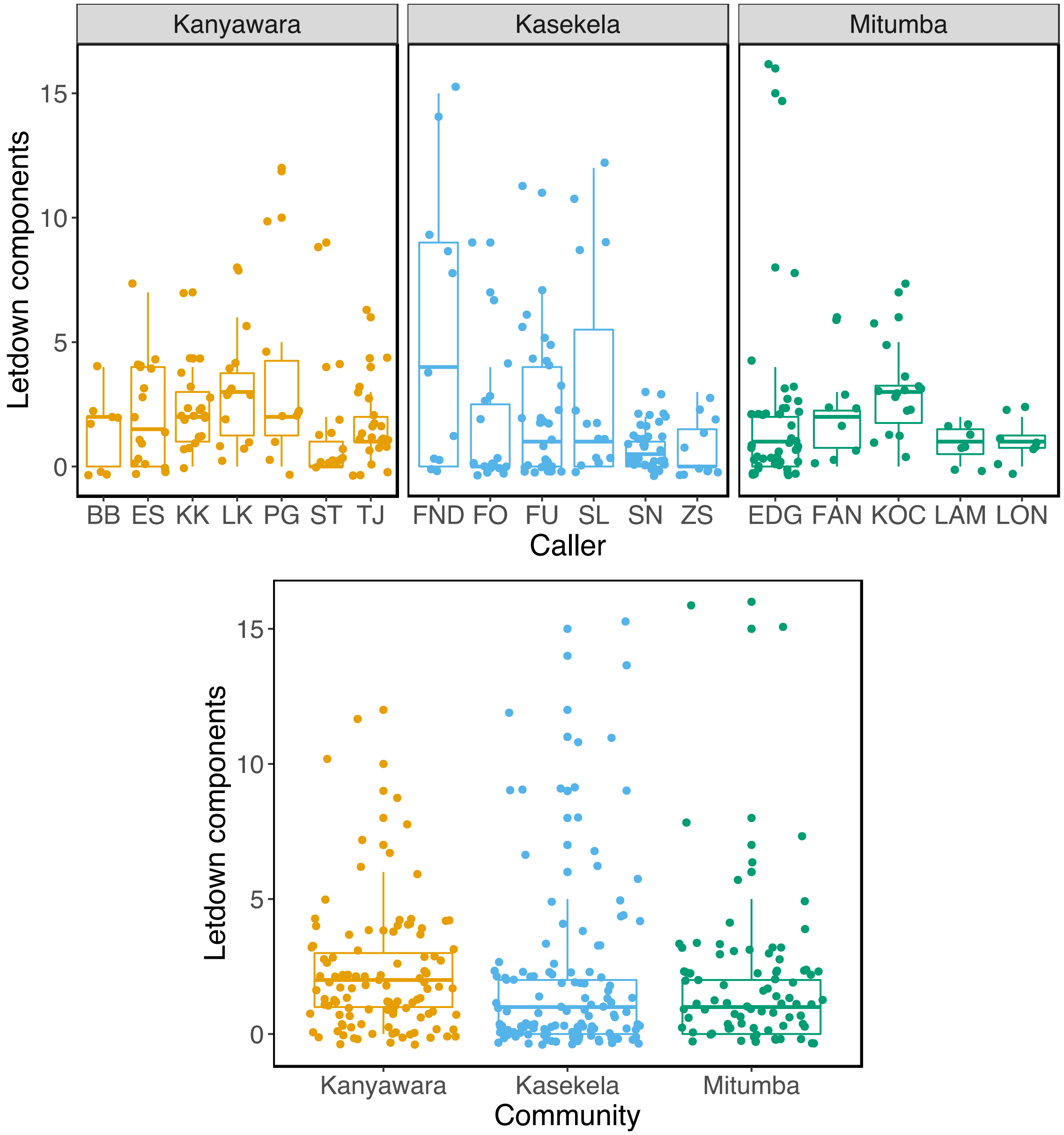


Figure 39: Number of letdown components at individual and community levels.

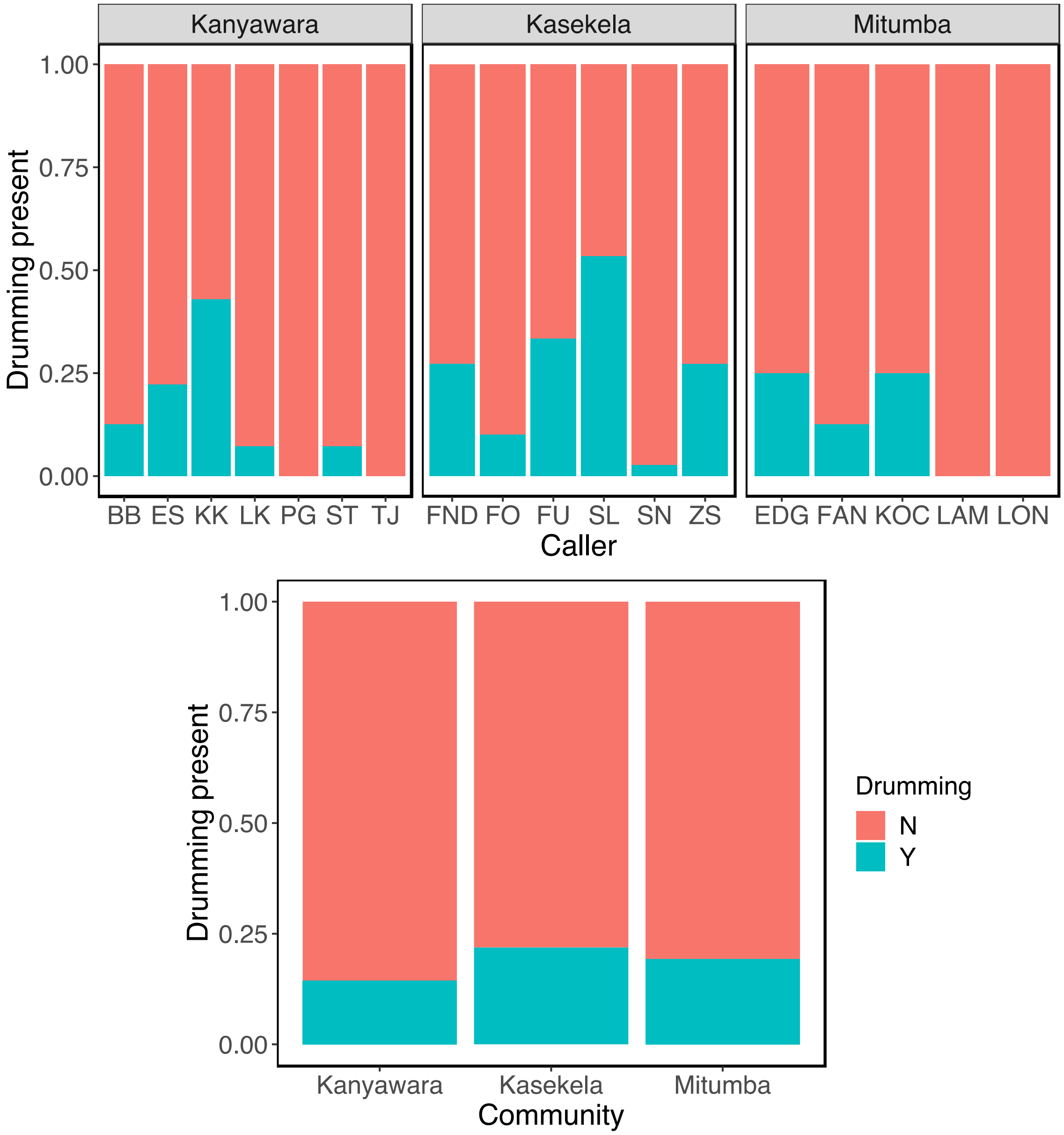


Figure 40: Proportion of calls with drumming present at individual and community levels.

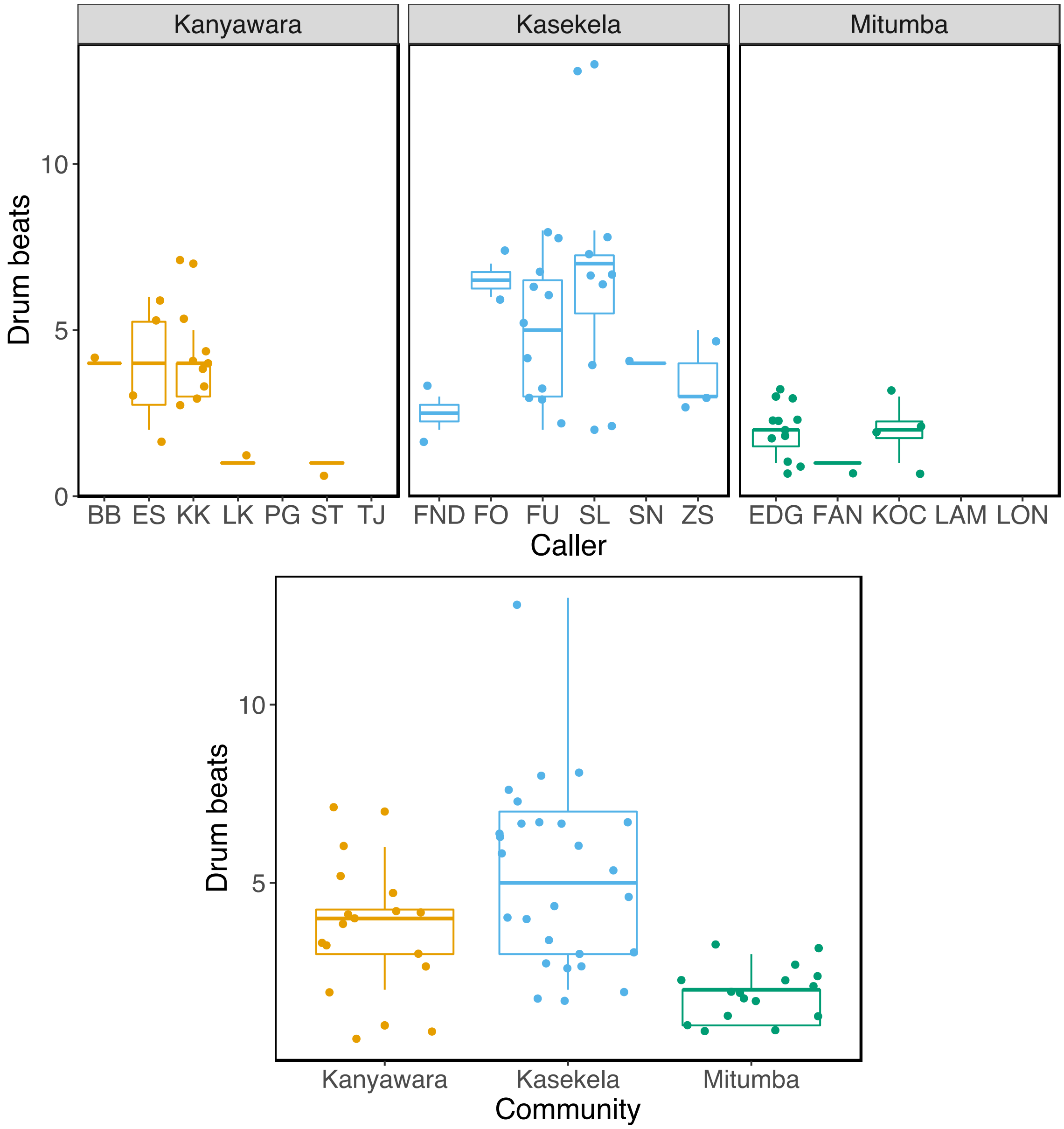


Figure 41: Number of drum beats at individual and community levels.

Table 35: Summary of callers, their ages, the number of pant-hoots calls, percent prevalence of each non-linear phenomenon in the calls (FJ = frequency jumps, SH = subharmonics, BP = biphonation, and CH = deterministic chaos), and percent of calls with at least one NLP.

Community N = 3	Caller N = 23	Number of calls (climax screams) N = 359	Age (years) Range = [10, 56]	% of calls (climax screams) with FJ	% of calls (climax screams) with SH	% of calls (climax screams) with BP	% of calls (climax screams) with CH	% of calls (climax screams) with at least one NLP
Kanyawara	AJ	9 (26)	24	67 (27)	67 (35)	33 (19)	44 (19)	100 (65)
Kanyawara	AJ	7 (15)	37	86 (53)	43 (40)	29 (20)	29 (13)	100 (100)
Kanyawara	BB	4 (4)	31	0 (0)	0 (0)	50 (50)	0 (0)	50 (50)
Kanyawara	BB	8 (16)	45	62 (50)	62 (50)	25 (12)	50 (38)	88 (88)
Kanyawara	BF	9 (23)	23	56 (35)	33 (13)	56 (35)	33 (17)	78 (78)
Kanyawara	BF	6 (14)	32	100 (57)	33 (14)	33 (21)	33 (14)	100 (79)
Kanyawara	KK	1 (3)	13	0 (0)	100 (33)	0 (0)	100 (100)	100 (100)
Kanyawara	KK	9 (15)	26	44 (40)	0 (0)	44 (33)	56 (33)	89 (87)
Kanyawara	LK	6 (16)	16	50 (31)	83 (50)	17 (19)	100 (50)	100 (88)
Kanyawara	LK	10 (19)	29	40 (26)	50 (26)	0 (0)	80 (53)	100 (79)
Kanyawara	MS	6 (13)	10	33 (23)	50 (31)	50 (31)	50 (46)	100 (85)
Kanyawara	MS	21 (37)	19	14 (8)	14 (8)	52 (41)	10 (5)	67 (54)
Kanyawara	ST	3 (5)	34	33 (20)	0 (0)	0 (0)	33 (20)	33 (40)
Kanyawara	ST	5 (11)	43	80 (45)	0 (0)	60 (27)	0 (0)	100 (73)
Kanyawara	ST	11 (24)	56	45 (25)	55 (25)	64 (38)	91 (54)	100 (83)
Kanyawara	SY	15 (65)	24	27 (9)	27 (8)	73 (32)	20 (6)	93 (55)
Kanyawara	SY	8 (46)	34	62 (20)	62 (17)	88 (35)	62 (24)	100 (83)
Kanyawara	TU	3 (6)	29	0 (0)	0 (0)	67 (67)	33 (29)	67 (67)
Kanyawara	TU	6 (7)	38	17 (14)	0 (0)	50 (57)	33 (43)	83 (86)
Kanyawara	TU	1 (4)	51	0 (0)	0 (0)	0 (0)	100 (50)	100 (50)
Kanyawara	YB	2 (4)	25	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)

Kanyawara	YB	6 (17)	38	67 (47)	0 (0)	33 (24)	50 (18)	100 (76)
Kasekela	FND	11 (42)	17	27 (12)	18 (7)	91 (69)	64 (29)	91 (83)
Kasekela	FO	3 (6)	24	67 (50)	0 (0)	0 (0)	0 (0)	67 (50)
Kasekela	FO	17 (36)	28	41 (22)	24 (11)	6 (3)	35 (19)	71 (47)
Kasekela	FD	12 (30)	22	17 (13)	50 (23)	42 (30)	17 (7)	83 (63)
Kasekela	FD	4 (8)	42	0 (0)	50 (38)	50 (38)	50 (25)	75 (75)
Kasekela	FU	1 (2)	16	100 (100)	0 (0)	0 (0)	0 (0)	100 (100)
Kasekela	FU	33 (85)	20	39 (21)	21 (8)	18 (9)	15 (8)	67 (40)
Kasekela	GIM	3 (7)	14	67 (29)	67 (43)	33 (14)	100 (43)	100 (86)
Kasekela	SL	3 (4)	30	100 (75)	33 (25)	0 (0)	0 (0)	100 (100)
Kasekela	SL	12 (23)	34	67 (57)	33 (17)	0 (0)	75 (48)	100 (96)
Kasekela	SN	8 (19)	17	38 (16)	12 (5)	62 (32)	75 (53)	88 (89)
Kasekela	SN	27 (58)	21	89 (59)	19 (10)	59 (36)	33 (26)	93 (83)
Kasekela	ZS	6 (24)	18	83 (29)	50 (29)	0 (0)	33 (12)	83 (54)
Kasekela	ZS	8 (11)	24	50 (55)	50 (36)	0 (0)	38 (27)	62 (73)
Mitumba	EDG	21 (50)	28	67 (56)	24 (22)	57 (36)	29 (14)	86 (74)
Mitumba	FAN	6 (16)	15	83 (75)	83 (75)	17 (12)	67 (38)	100 (94)
Mitumba	KOC	13 (32)	16	77 (38)	31 (12)	8 (3)	54 (31)	92 (66)
Mitumba	LAM	8 (19)	15	25 (16)	0 (0)	62 (53)	0 (0)	75 (58)
Mitumba	LON	7 (15)	16	86 (67)	57 (47)	43 (33)	14 (7)	86 (87)

Table 36: Community-level variation in % of calls with different NLPs.

Community	N calls (screams)	% of calls (screams) with FJ	% of calls (screams) with SH	% of calls (screams) with BP	% of calls (screams) with CH	% of calls (screams) with NL
Kanyawara	156 (390)	44 (25)	33 (17)	45 (28)	42 (23)	87 (72)
Kasekela	148 (355)	52 (30)	28 (14)	31 (22)	36 (21)	81 (66)
Mitumba	55 (132)	67 (49)	33 (26)	40 (27)	33 (18)	87 (73)

Table 37: Parameter estimates from the binomial regression estimating the proportion of scream duration that exhibits NLPs among different communities, their standard deviations (SD), 89% highest posterior density intervals (HPDI), the effective sample size (N_eff), and the Gelman-Rubin convergence diagnostic (Rhat).

Coefficient	**Mean**	**SD**	**HPDI lower (5.50%)**	**HPDI upper (94.50%)**	**N_eff**	**Rhat**
β[Kanyawara]	-0.06	0.36	-0.65	0.52	236.74	1.01
β[Kasekela]	-0.18	0.36	-0.73	0.42	256.2	1
β[Mitumba]	0.12	0.37	-0.49	0.68	239.24	1
$\bar{\alpha}$	-1.07	0.33	-1.6	-0.52	271.72	1
σ	0.83	0.13	0.65	1.07	683.88	1.01
α[AJ]	-1.28	0.36	-1.85	-0.7	236.34	1.01
α[BB]	-0.64	0.36	-1.21	-0.04	237.93	1.01
α[BF]	-0.97	0.36	-1.55	-0.38	237.08	1.01
α[EDG]	-1.65	0.37	-2.21	-1.05	240.58	1
α[FAN]	0.05	0.37	-0.51	0.65	237.92	1
α[FD]	-1.12	0.36	-1.71	-0.56	257.3	1
α[FND]	-0.07	0.36	-0.68	0.48	256.41	1
α[FO]	-2.32	0.36	-2.92	-1.76	256.04	1
α[FU]	-2.5	0.37	-3.11	-1.95	254.56	1
α[GIM]	0.65	0.37	0.04	1.23	258.86	1
α[KK]	-0.75	0.36	-1.33	-0.16	237.61	1.01
α[KOC]	-1.32	0.37	-1.88	-0.72	239.45	1
α[LAM]	-1.4	0.37	-1.97	-0.79	239.36	1
α[LK]	-0.47	0.36	-1.04	0.12	237.48	1.01

α[LON]	-0.98	0.37	-1.54	-0.39	236.73	1
α[MS]	-1.2	0.36	-1.78	-0.62	237.74	1.01
α[SL]	-1.24	0.36	-1.84	-0.68	255.79	1
α[SN]	-0.78	0.36	-1.38	-0.23	256.42	1
α[ST]	-0.62	0.36	-1.19	-0.03	236.44	1.01
α[SY]	-1.6	0.36	-2.17	-1.02	237.39	1.01
α[TU]	-1.23	0.36	-1.8	-0.64	237.68	1.01
α[YB]	-2.34	0.36	-2.91	-1.75	237.28	1.01
α[ZS]	-1.78	0.36	-2.38	-1.22	257.22	1

Table 38: Summary of callers, their ages, the number of pant-hoots, mean percent of climax-scream duration that consists of specific non-linear phenomena (SH = subharmonics, BP = biphonation, and CH = deterministic chaos), and mean percent of climax-scream duration that consists of NLP.

Community	Caller	Number of calls	Age	Mean % of climax screams duration that is SH (Range)	Mean % of climax screams duration that is BP (Range)	Mean % of climax screams duration that is CH (Range)	Mean % of climax screams duration that is NLP (Range)
Kanyawara	AJ	9	24	8 (0,17)	3 (0,10)	8 (0,34)	18 (0,47)
Kanyawara	AJ	7	37	16 (0,51)	5 (0,31)	6 (0,36)	27 (0,51)
Kanyawara	BB	4	31	0 (0,0)	40 (0,100)	0 (0,0)	40 (0,100)
Kanyawara	BB	8	45	13 (0,26)	4 (0,18)	8 (0,31)	25 (0,75)
Kanyawara	BF	9	23	9 (0,42)	13 (0,42)	5 (0,21)	28 (0,64)
Kanyawara	BF	6	32	2 (0,8)	6 (0,20)	8 (0,29)	16 (0,47)

Kanyawara	KK	1	13	5 (5,5)	0 (0,0)	54 (54,54)	58 (58,58)
Kanyawara	KK	9	26	0 (0,0)	8 (0,26)	20 (0,67)	27 (0,67)
Kanyawara	LK	6	16	16 (0,26)	3 (0,18)	32 (6,72)	52 (26,80)
Kanyawara	LK	10	29	15 (0,67)	0 (0,0)	19 (0,45)	34 (0,84)
Kanyawara	MS	6	10	14 (0,40)	10 (0,23)	16 (0,61)	40 (17,86)
Kanyawara	MS	21	19	2 (0,31)	12 (0,55)	1 (0,17)	15 (0,55)
Kanyawara	ST	3	34	0 (0,0)	0 (0,0)	12 (0,36)	12 (0,36)
Kanyawara	ST	5	43	0 (0,0)	21 (0,60)	0 (0,0)	21 (0,60)
Kanyawara	ST	11	56	7 (0,22)	17 (0,34)	26 (0,100)	50 (12,100)
Kanyawara	SY	15	24	2 (0,12)	6 (0,19)	1 (0,13)	10 (0,21)
Kanyawara	SY	8	34	6 (0,11)	12 (0,28)	11 (0,31)	29 (9,50)
Kanyawara	TU	3	29	0 (0,0)	17 (0,43)	4 (0,13)	22 (0,56)
Kanyawara	TU	6	38	0 (0,0)	22 (0,62)	9 (0,32)	31 (0,62)
Kanyawara	TU	1	51	0 (0,0)	0 (0,0)	10 (10,10)	10 (10,10)
Kanyawara	YB	2	25	0 (0,0)	0 (0,0)	0 (0,0)	0 (0,0)
Kanyawara	YB	6	38	0 (0,0)	13 (0,44)	5 (0,23)	18 (0,55)
Kasekela	FND	11	17	2 (0,12)	36 (0,58)	9 (0,23)	47 (0,70)
Kasekela	FO	3	24	0 (0,0)	0 (0,0)	0 (0,0)	0 (0,0)
Kasekela	FO	17	28	3 (0,24)	1 (0,10)	4 (0,18)	8 (0,28)
Kasekela	FD	12	22	8 (0,22)	12 (0,59)	3 (0,17)	22 (0,59)
Kasekela	FD	4	42	8 (0,25)	11 (0,27)	4 (0,8)	23 (0,33)
Kasekela	FU	1	16	0 (0,0)	0 (0,0)	0 (0,0)	0 (0,0)
Kasekela	FU	33	20	2 (0,18)	2 (0,27)	3 (0,43)	7 (0,43)
Kasekela	GIM	3	14	27 (0,49)	3 (0,9)	25 (11,42)	55 (11,84)

Kasekela	SL	3	30	3 (0,10)	0 (0,0)	0 (0,0)	3 (0,10)
Kasekela	SL	12	34	4 (0,17)	0 (0,0)	30 (0,100)	34 (0,100)
Kasekela	SN	8	17	1 (0,4)	14 (0,33)	28 (0,81)	42 (0,85)
Kasekela	SN	27	21	3 (0,47)	10 (0,48)	8 (0,76)	21 (0,76)
Kasekela	ZS	6	18	8 (0,23)	0 (0,0)	3 (0,12)	11 (0,35)
Kasekela	ZS	8	24	12 (0,52)	0 (0,0)	5 (0,20)	16 (0,52)
Mitumba	EDG	21	28	3 (0,27)	11 (0,55)	6 (0,57)	20 (0,57)
Mitumba	FAN	6	15	34 (0,53)	2 (0,12)	18 (0,39)	54 (24,83)
Mitumba	KOC	13	16	5 (0,25)	2 (0,22)	14 (0,36)	20 (0,45)
Mitumba	LAM	8	15	0 (0,0)	22 (0,56)	0 (0,0)	22 (0,56)
Mitumba	LON	7	16	17 (0,47)	11 (0,42)	2 (0,15)	30 (0,60)

Table 39: WAIC values for the four polynomial models, their standard errors (SE), the difference in WAIC from the top model (dWAIC), the standard error of the difference in WAIC (dSE), the number of effective parameters (pWAIC), and model weights.

	WAIC	SE	dWAIC	dSE	pWAIC	weight
4thDegree	147631.3	9789.8	0	NA	9727.2	1
Cubic	148155.6	9823.5	524.2	1675.5	9536.4	0
Quadratic	149845.6	9955.2	2214.2	2871.3	9257.2	0
Linear	157376.5	11348.2	9745.2	6184.7	9281.2	0

Table 40: Parameter estimates from the linear model, their standard deviations (SD), 89% highest posterior density intervals (HPDI), the effective sample size (N_eff), and the Gelman-Rubin convergence diagnostic (Rhat).

Coefficient	Mean	SD	HPDI lower (5.50%)	HPDI upper (94.50%)	N_eff	Rhat
$\boldsymbol{\beta}_1$	0.08	0.01	0.07	0.1	831.61	1.01
$\boldsymbol{\alpha}$**[AJ]**	-1.36	0.01	-1.39	-1.34	3985.95	1
$\boldsymbol{\alpha}$**[BB]**	-0.84	0.02	-0.86	-0.81	1313.77	1
$\boldsymbol{\alpha}$**[BF]**	-1.04	0.01	-1.06	-1.02	5271.87	1
$\boldsymbol{\alpha}$**[EDG]**	-1.55	0.02	-1.57	-1.52	4221.29	1
$\boldsymbol{\alpha}$**[FAN]**	0.27	0.02	0.23	0.3	2870.26	1
$\boldsymbol{\alpha}$**[FD]**	-1.31	0.02	-1.33	-1.28	4018.28	1
$\boldsymbol{\alpha}$**[FND]**	-0.17	0.01	-0.2	-0.15	1938.41	1
$\boldsymbol{\alpha}$**[FO]**	-2.51	0.02	-2.54	-2.48	5011.86	1
$\boldsymbol{\alpha}$**[FU]**	-2.63	0.02	-2.66	-2.61	2843.64	1
$\boldsymbol{\alpha}$**[GIM]**	0.58	0.04	0.52	0.65	2616.07	1
$\boldsymbol{\alpha}$**[KK]**	-0.79	0.02	-0.82	-0.76	4179.88	1
$\boldsymbol{\alpha}$**[KOC]**	-1.12	0.02	-1.14	-1.09	2179.79	1
$\boldsymbol{\alpha}$**[LAM]**	-1.19	0.03	-1.23	-1.15	2560.79	1
$\boldsymbol{\alpha}$**[LK]**	-0.51	0.02	-0.53	-0.48	3705.87	1
$\boldsymbol{\alpha}$**[LON]**	-0.78	0.02	-0.82	-0.74	2553	1
$\boldsymbol{\alpha}$**[MS]**	-1.18	0.01	-1.2	-1.16	1936.53	1
$\boldsymbol{\alpha}$**[SL]**	-1.48	0.02	-1.52	-1.45	2701.02	1
$\boldsymbol{\alpha}$**[SN]**	-0.91	0.01	-0.92	-0.89	2702.74	1

α[ST]	-0.89	0.02	-0.92	-0.85	1113.09	1
α[SY]	-1.68	0.01	-1.7	-1.66	3249.8	1
α[TU]	-1.38	0.02	-1.42	-1.35	1870.04	1
α[YB]	-2.49	0.03	-2.53	-2.44	3659.4	1
α[ZS]	-1.91	0.02	-1.94	-1.87	3405.69	1
$\bar{\alpha}$	-1.15	0.16	-1.4	-0.9	3401.86	1
σ	0.82	0.12	0.65	1.04	2600.68	1

Table 41: Parameter estimates from the cubic polynomial model, their standard deviations (SD), 89% highest posterior density intervals (HPDI), the effective sample size (N_eff), and the Gelman-Rubin convergence diagnostic (Rhat).

Coefficient	**Mean**	**SD**	**HPDI lower (5.50%)**	**HPDI upper (94.50%)**	**N_eff**	**Rhat**
β_1	-0.17	0.01	-0.18	-0.15	905.57	1
β_2	0.51	0.01	0.5	0.51	1068.31	1
β_3	-0.09	0	-0.09	-0.09	1841.61	1
α[AJ]	-1.49	0.01	-1.51	-1.47	3428.19	1
α[BB]	-1.5	0.02	-1.54	-1.47	1327.59	1
α[BF]	-1.12	0.01	-1.14	-1.1	3574.82	1
α[EDG]	-1.51	0.02	-1.54	-1.49	3855.23	1
α[FAN]	-0.74	0.02	-0.78	-0.7	1500.8	1
α[FD]	-1.62	0.02	-1.65	-1.59	2542.75	1
α[FND]	-0.86	0.02	-0.89	-0.84	1681.04	1
α[FO]	-2.49	0.02	-2.52	-2.45	3094.32	1

α[FU]	-2.95	0.02	-2.98	-2.92	2565.46	1
α[GIM]	-0.72	0.04	-0.79	-0.66	2001.79	1
α[KK]	-1.04	0.02	-1.07	-1.01	3970.09	1
α[KOC]	-1.98	0.02	-2.02	-1.95	1733.99	1
α[LAM]	-2.12	0.03	-2.16	-2.08	1826.57	1
α[LK]	-0.87	0.02	-0.89	-0.84	2972.24	1
α[LON]	-1.6	0.03	-1.65	-1.56	2250.28	1
α[MS]	-2.17	0.02	-2.2	-2.14	1047.61	1.01
α[SL]	-1.58	0.02	-1.61	-1.54	2449.73	1
α[SN]	-1.3	0.01	-1.32	-1.28	1472.79	1
α[ST]	-1.98	0.03	-2.03	-1.94	1598.99	1
α[SY]	-1.74	0.01	-1.76	-1.72	3620.36	1
α[TU]	-1.83	0.03	-1.88	-1.79	2044.08	1
α[YB]	-2.76	0.03	-2.81	-2.71	2054.15	1
α[ZS]	-2.27	0.02	-2.31	-2.24	2080.18	1
$\bar{\alpha}$	-1.65	0.13	-1.85	-1.43	2611.47	1
σ	0.65	0.1	0.51	0.84	2204.47	1

Table 42: Parameter estimates from the 4th degree polynomial model, their standard deviations (SD), 89% highest posterior density intervals (HPDI), the effective sample size (N_eff), and the Gelman-Rubin convergence diagnostic (Rhat).

Coefficient	**Mean**	**SD**	**HPDI lower (5.50%)**	**HPDI upper (94.50%)**	**N_eff**	**Rhat**
β_1	0.11	0.01	0.09	0.13	740.22	1.01

β_2	0.35	0.01	0.34	0.37	720.49	1
β_3	-0.23	0.01	-0.24	-0.22	1040.82	1
β_4	0.05	0	0.05	0.05	875.81	1
α[AJ]	-1.45	0.01	-1.48	-1.43	1767.69	1
α[BB]	-1.2	0.02	-1.24	-1.17	1093.44	1
α[BF]	-1.1	0.01	-1.12	-1.08	1730.5	1
α[EDG]	-1.57	0.02	-1.59	-1.55	1815.3	1
α[FAN]	-0.51	0.03	-0.55	-0.47	1268	1
α[FD]	-1.46	0.02	-1.49	-1.43	1675.36	1
α[FND]	-0.63	0.02	-0.65	-0.6	824.18	1
α[FO]	-2.53	0.02	-2.57	-2.5	742.8	1
α[FU]	-2.78	0.02	-2.81	-2.75	1537.46	1
α[GIM]	-0.53	0.04	-0.6	-0.47	1824.79	1
α[KK]	-1.02	0.02	-1.05	-0.98	1977.62	1
α[KOC]	-1.74	0.02	-1.78	-1.71	1040.91	1
α[LAM]	-1.88	0.03	-1.93	-1.84	1339.74	1
α[LK]	-0.8	0.02	-0.83	-0.78	1960.79	1
α[LON]	-1.36	0.03	-1.41	-1.32	1101.81	1
α[MS]	-2.09	0.02	-2.12	-2.05	1174.6	1.01
α[SL]	-1.64	0.02	-1.67	-1.6	1525.26	1
α[SN]	-1.12	0.01	-1.14	-1.1	970.3	1
α[ST]	-1.91	0.03	-1.95	-1.87	1450.47	1
α[SY]	-1.74	0.01	-1.76	-1.72	1278.05	1

$\boldsymbol{\alpha}$**[TU]**	-1.69	0.02	-1.73	-1.66	1608.26	1
$\boldsymbol{\alpha}$**[YB]**	-2.71	0.03	-2.76	-2.66	1489.4	1
$\boldsymbol{\alpha}$**[ZS]**	-2.1	0.02	-2.14	-2.07	1528.83	1
$\bar{\boldsymbol{\alpha}}$	-1.53	0.14	-1.75	-1.32	874.57	1
σ	0.67	0.1	0.51	0.86	594.02	1.01

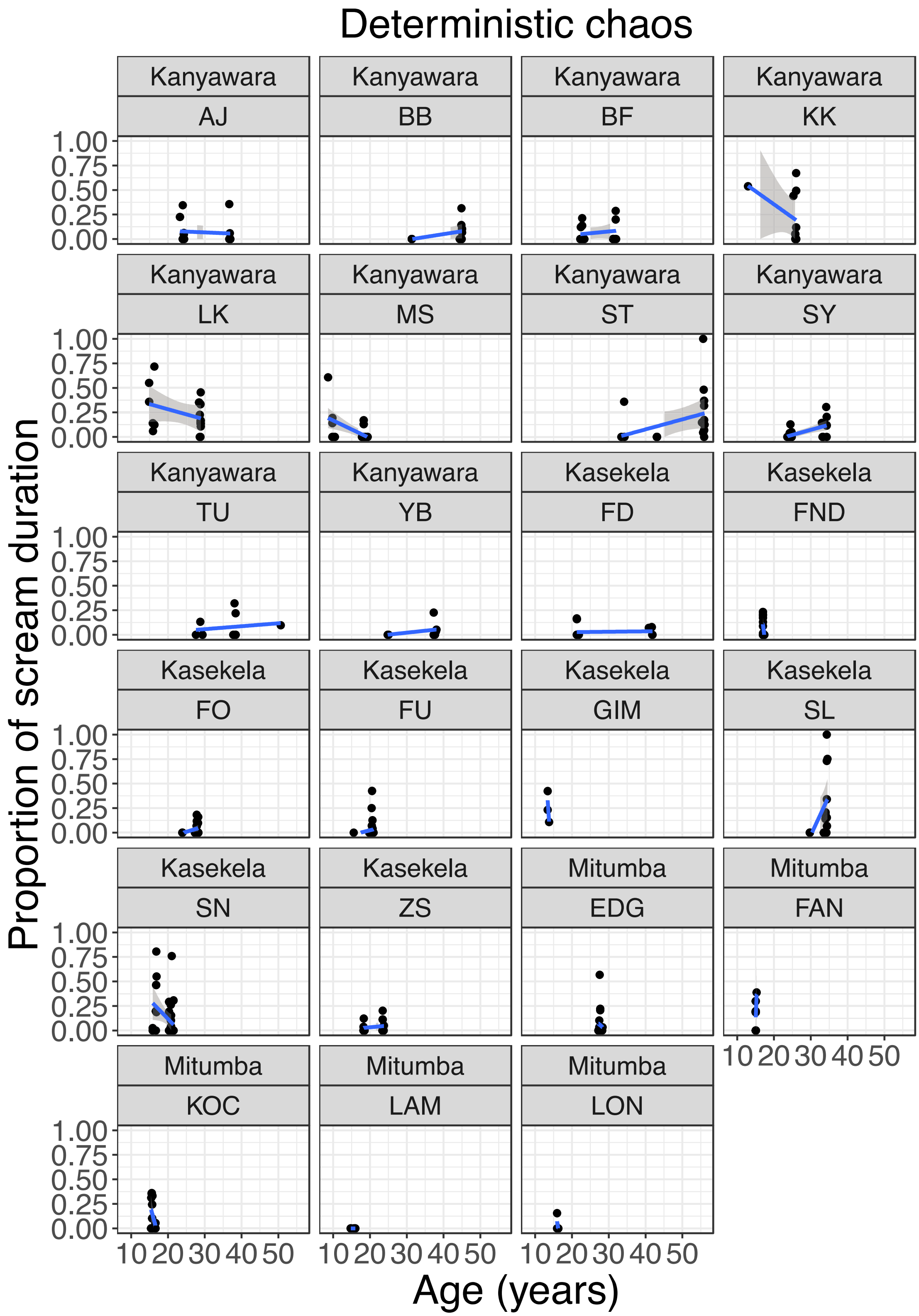


Figure 42: Within-individual longitudinal variation in proportion of total climax-scream duration that consists of deterministic chaos.

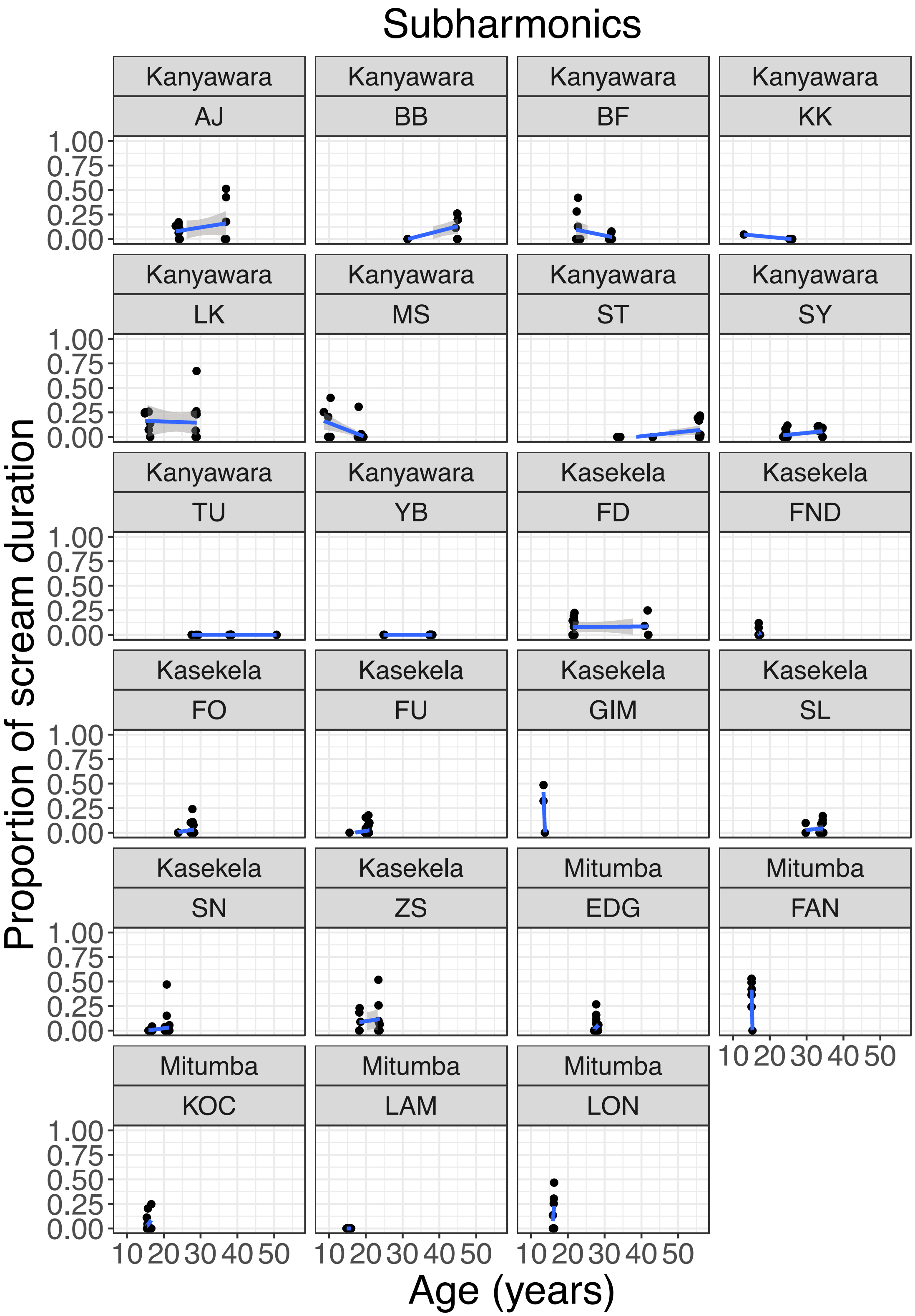


Figure 43: Within-individual longitudinal variation in proportion of total climax-scream duration that consists of subharmonics.

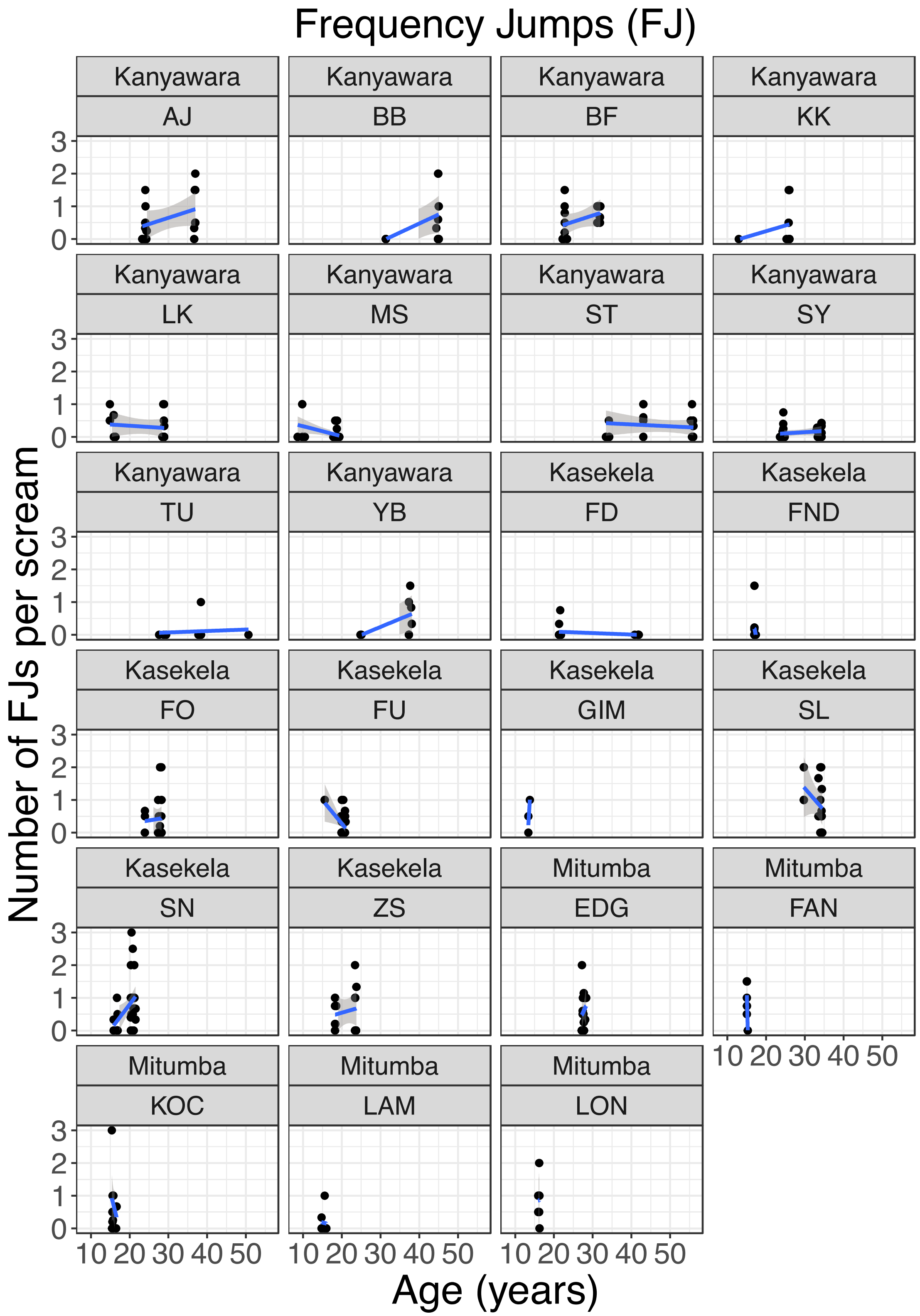


Figure 44: Within-individual longitudinal variation in the number of frequency jumps per climax-scream.

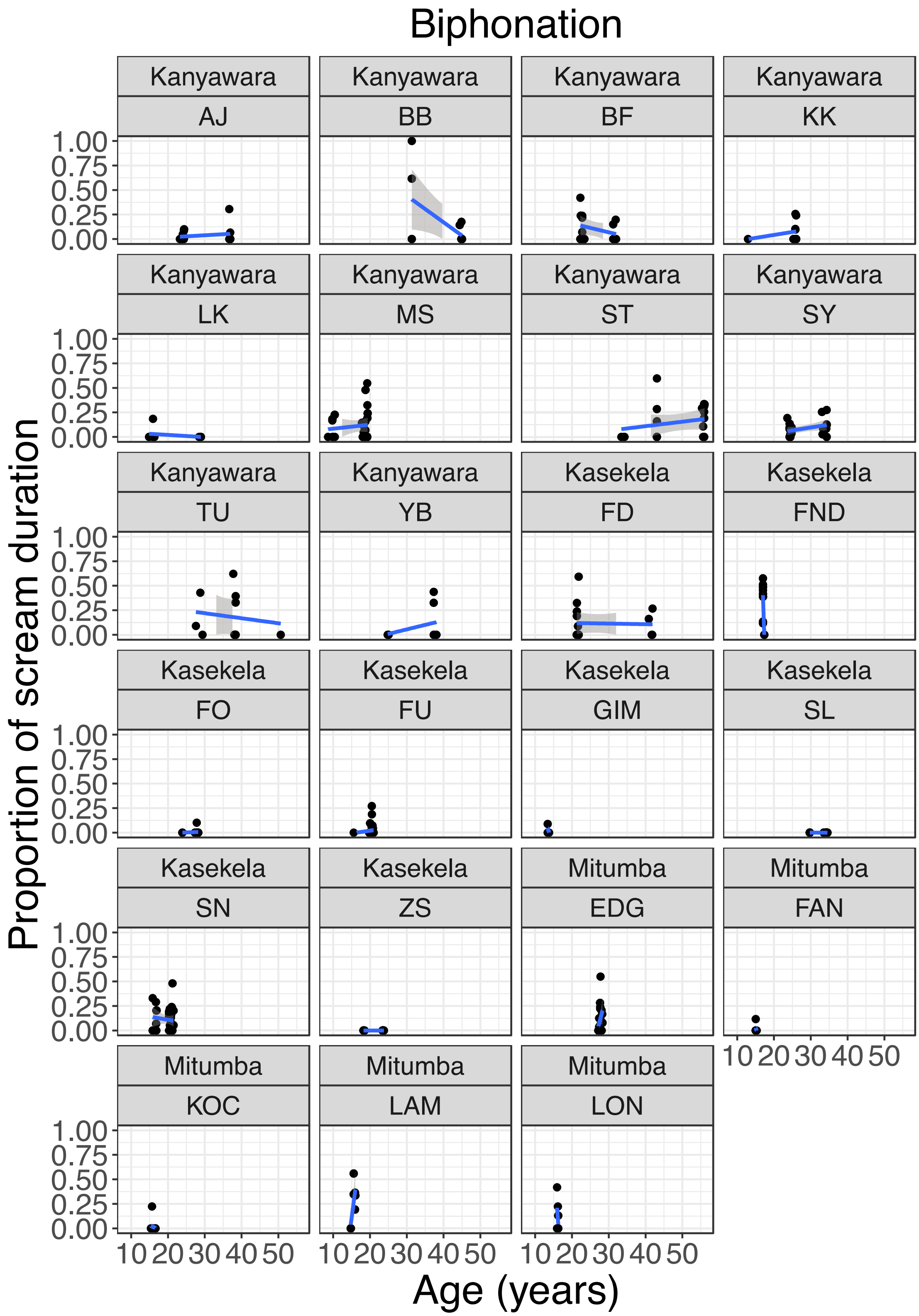


Figure 45: Within-individual longitudinal variation in proportion of total climax-scream duration that consists of biphonation.

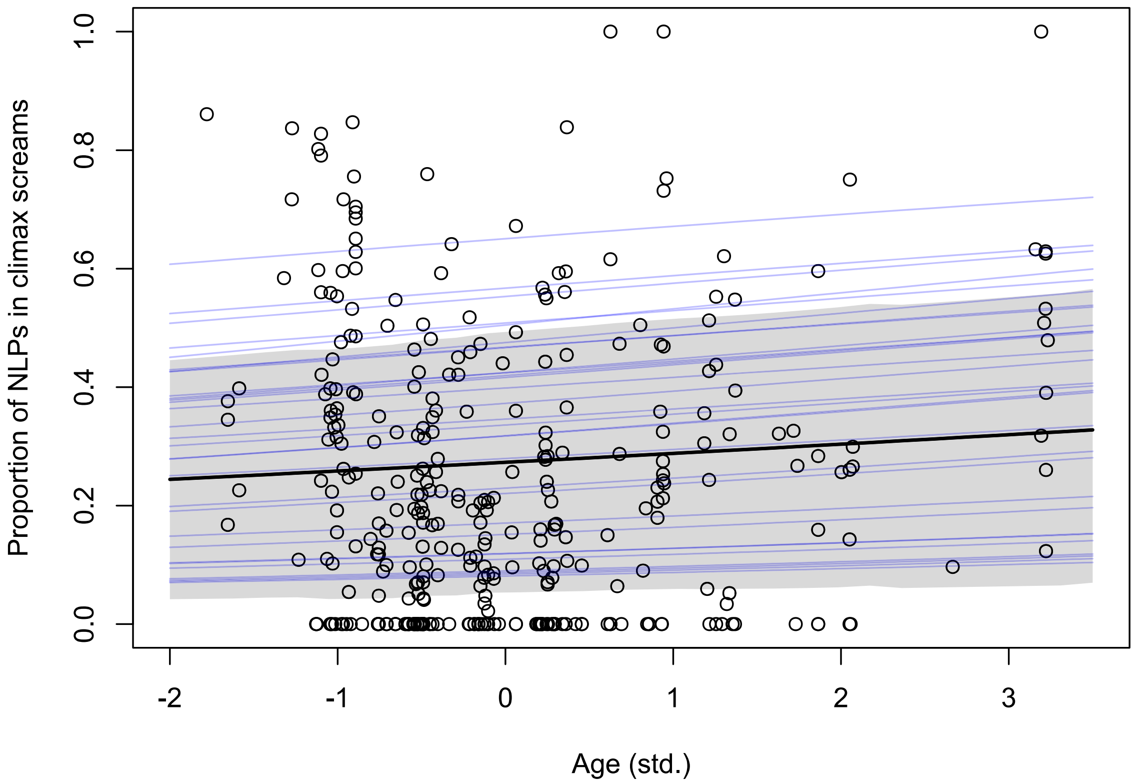


Figure 46: Posterior predictions simulated from the estimated joint posterior distribution from the linear model fitting the relationship of the proportion of NLPs in climax-screams to age (standardized). Dots represent raw data, the solid black line represents the predicted average chimpanzee, blue lines represent 30 random chimpanzees simulated from the posterior, and the gray shade represents the 89% HPDI of the population of chimpanzees.

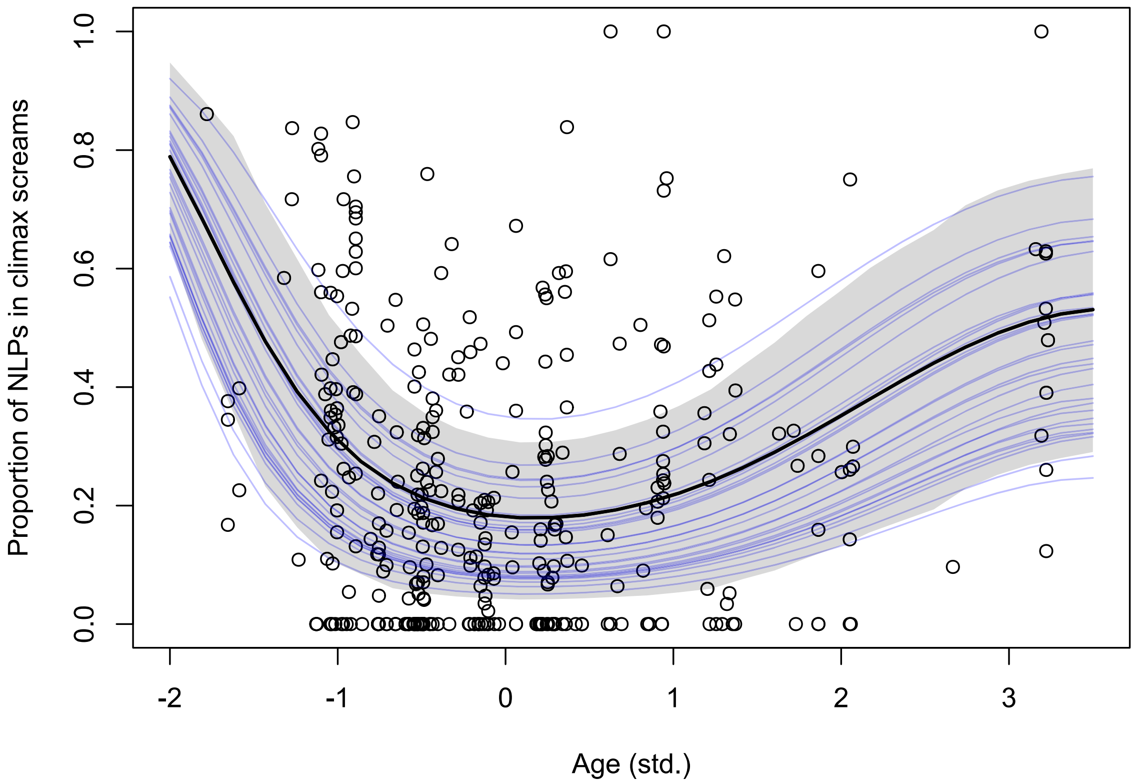


Figure 47: Posterior predictions simulated from the estimated joint posterior distribution from the cubic model fitting the relationship of the proportion of NLPs in climax-screams to age (standardized). Dots represent raw data, the solid black line represents the predicted average chimpanzee, blue lines represent 30 random chimpanzees simulated from the posterior, and the gray shade represents the 89% HPDI of the population of chimpanzees.

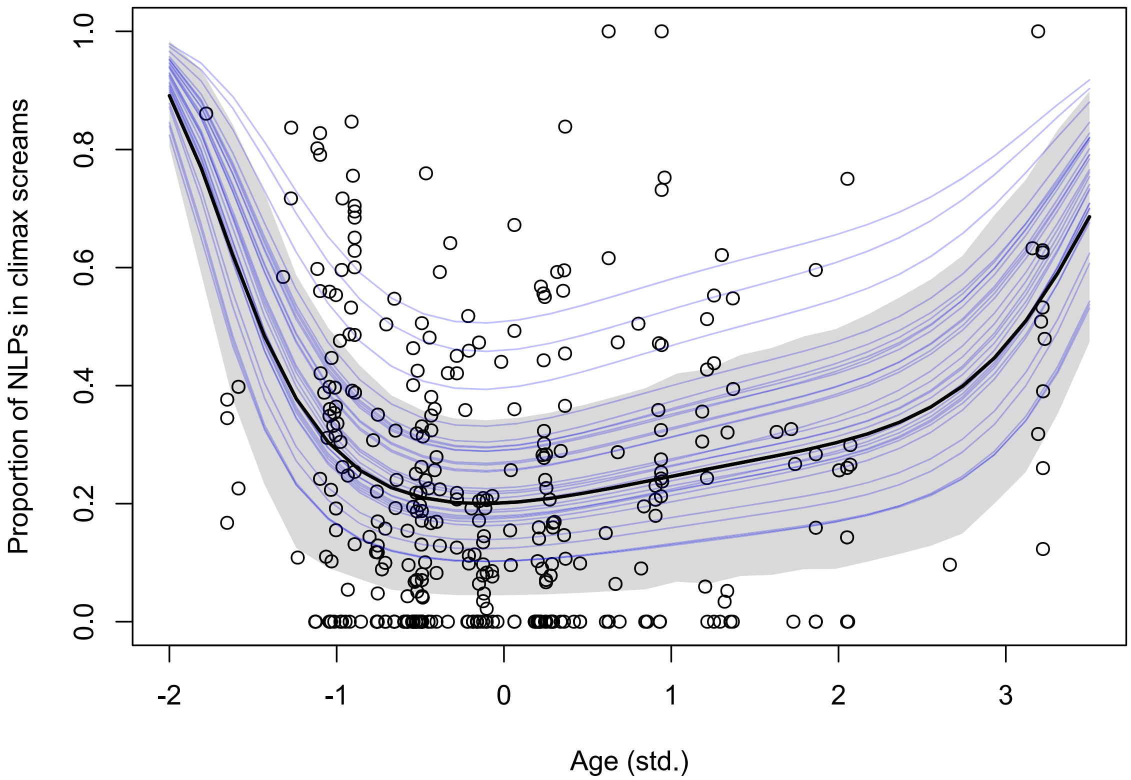


Figure 48: Posterior predictions simulated from the estimated joint posterior distribution from the 4th degree polynomial model fitting the relationship of the proportion of NLPs in climax-screams to age (standardized). Dots represent raw data, the solid black line represents the predicted average chimpanzee, blue lines represent 30 random chimpanzees simulated from the posterior, and the gray shade represents the 89% HPDI of the population of chimpanzees.

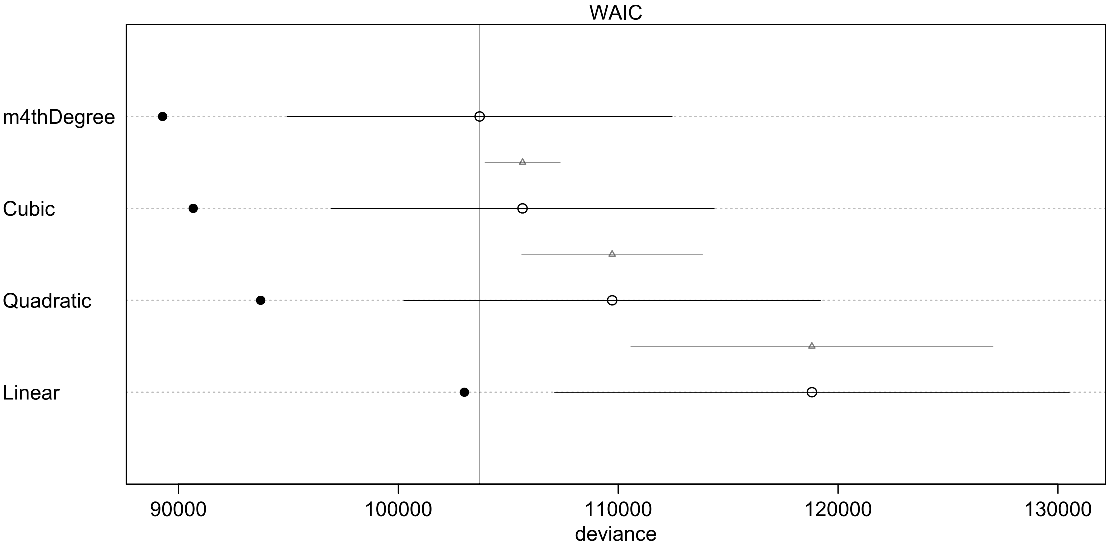


Figure 49: Comparison of four polynomial regression models assessing the relationship of age and proportion of deterministic chaos in climax-screams. The filled black dots represent in-sample deviance. The open dots represent the WAIC values, which are a theoretical estimate of the out-of-sample deviance. The black line segments represent the standard error of the WAIC values. The gray line segments represent the standard error of the difference in the WAIC values. The 4th degree polynomial was the best model.

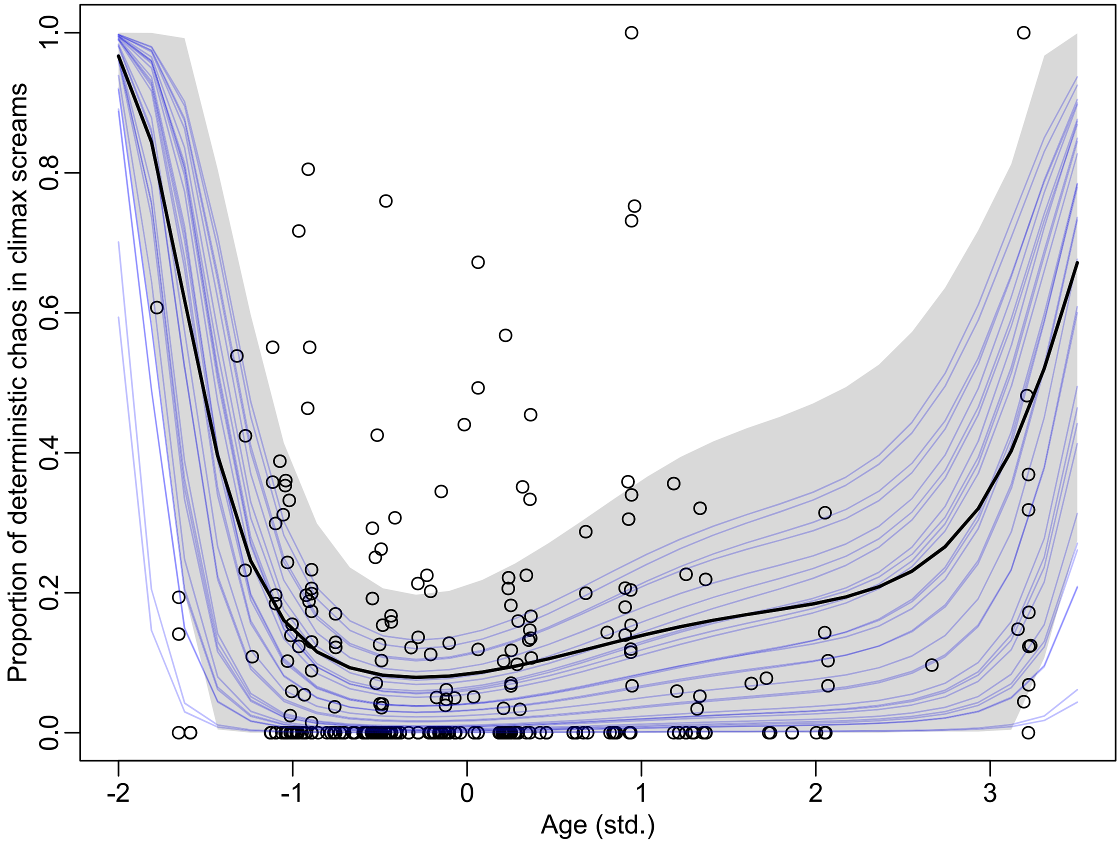


Figure 50: Posterior predictions simulated from the estimated joint posterior distribution from the 4th degree polynomial model fitting the relationship of the proportion of deterministic chaos in climax-screams to age (standardized). Dots represent raw data, the solid black line represents the predicted average chimpanzee, blue lines represent 30 random chimpanzees simulated from the posterior, and the gray shade represents the 89% HPDI of the population of chimpanzees.

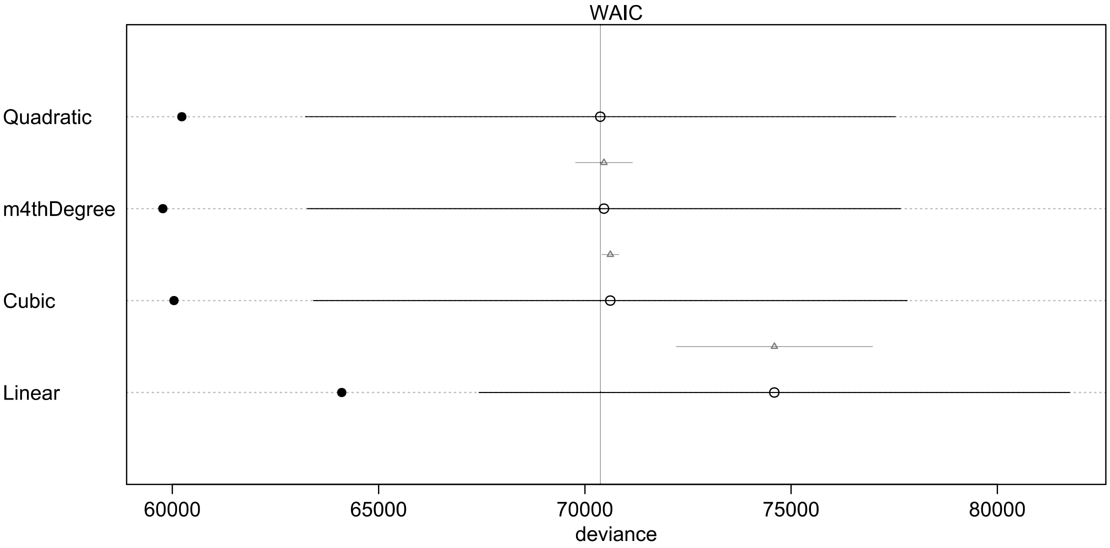


Figure 51: Comparison of four polynomial regression models assessing the relationship of age and proportion of subharmonics in climax-screams. The filled black dots represent in-sample deviance. The open dots represent the WAIC values, which are a theoretical estimate of the out-of-sample deviance. The black line segments represent the standard error of the WAIC values. The gray line segments represent the standard error of the difference in the WAIC values. The quadratic model was the best model, but equivalent to the cubic and 4th degree polynomial models.

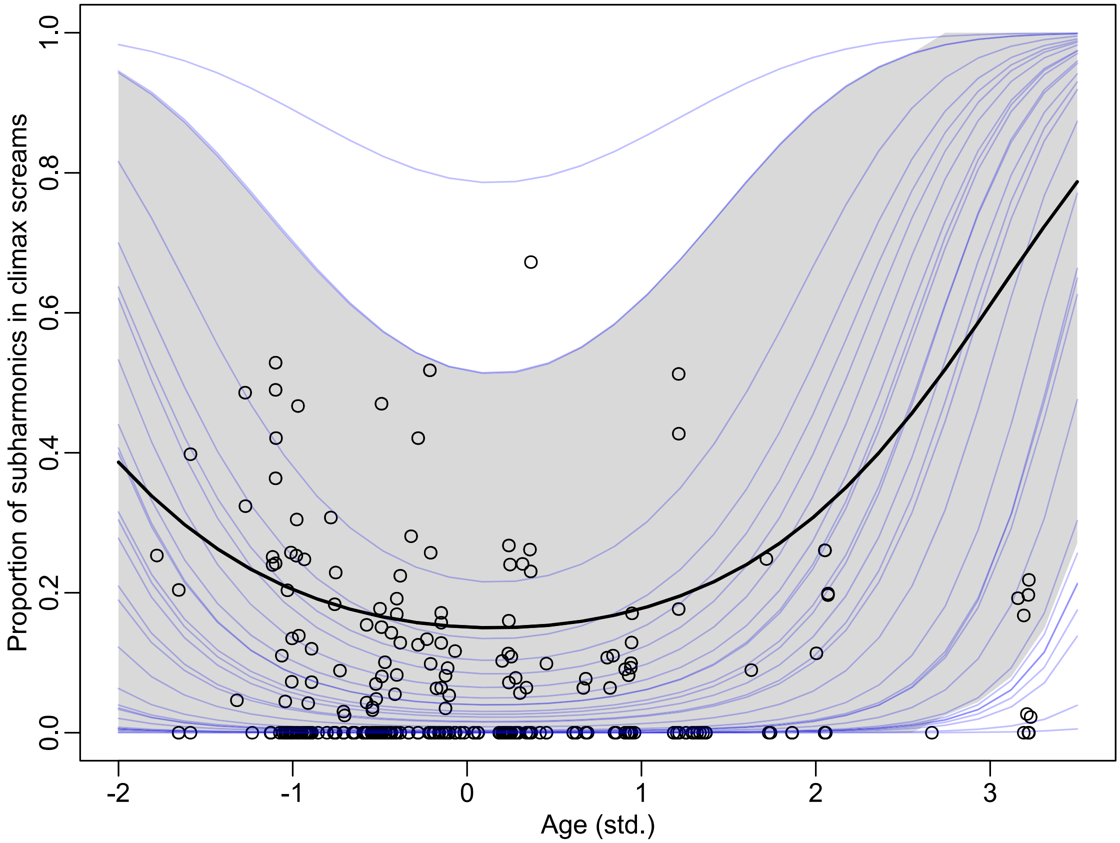


Figure 52: Posterior predictions simulated from the estimated joint posterior distribution from the quadratic polynomial model fitting the relationship of the proportion of subharmonics in climax-screams to age (standardized). Dots represent raw data, the solid black line represents the predicted average chimpanzee, blue lines represent 30 random chimpanzees simulated from the posterior, and the gray shade represents the 89% HPDI of the population of chimpanzees.

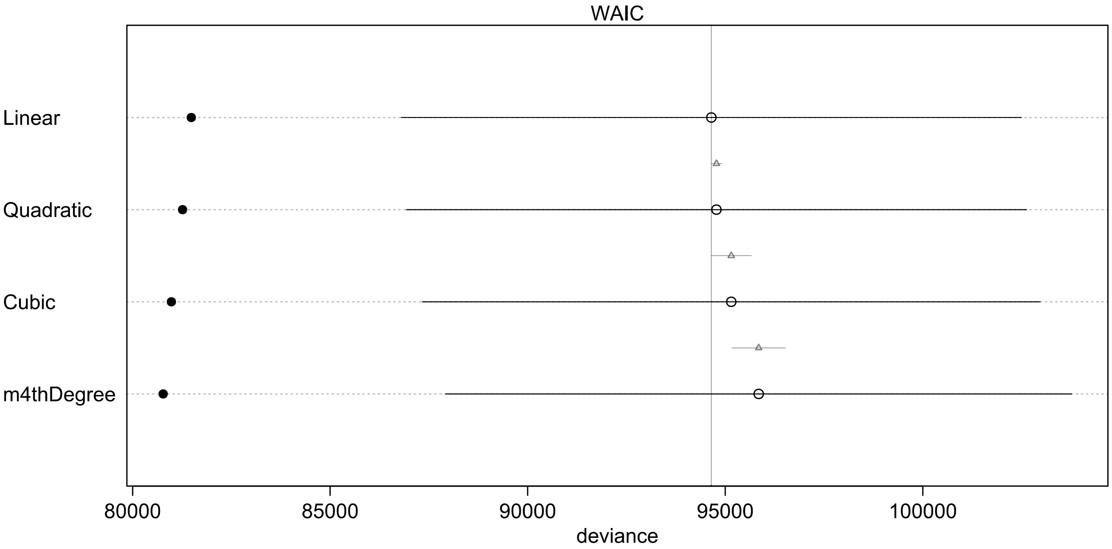


Figure 53: Comparison of four polynomial regression models assessing the relationship of age and proportion of biphonation in climax-screams. The filled black dots represent in-sample deviance. The open dots represent the WAIC values, which are a theoretical estimate of the out-of-sample deviance. The black line segments represent the standard error of the WAIC values. The gray line segments represent the standard error of the difference in the WAIC values. The linear model was the best model but equivalent to the quadratic and cubic model.

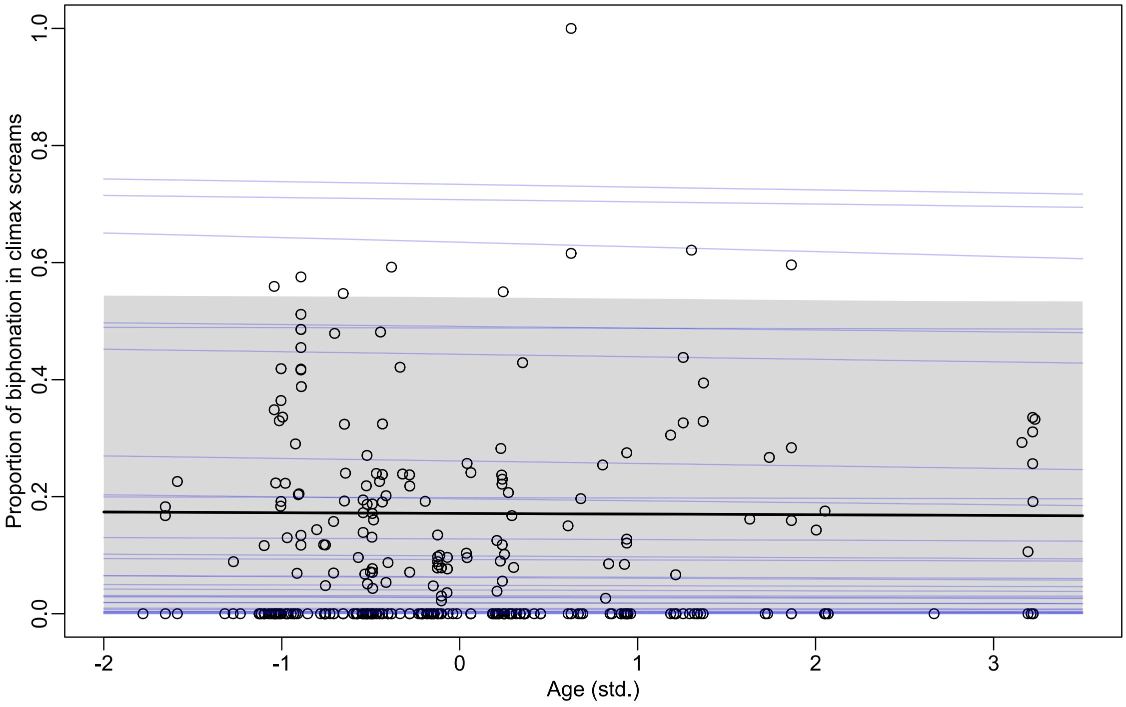


Figure 54: Posterior predictions simulated from the estimated joint posterior distribution from the linear model fitting the relationship of the proportion of biphonation in climax-screams to age (standardized). Dots represent raw data, the solid black line represents the predicted average chimpanzee, blue lines represent 30 random chimpanzees simulated from the posterior, and the gray shade represents the 89% HPDI of the population of chimpanzees.