University of Zurich^UZH

Small-Scale Spatial Variation in Meerkat Vocalization

GEO 511 Master's Thesis

Author
Antonia Hehli
19-750-017

Supervised by
Dr. Peter Ranacher
Nikola Falk (nikola.falk@ieu.uzh.ch)
Dr. Stuart Watson (stuart.watson@uzh.ch)

Faculty representative
Prof. Dr. Robert Weibel

01.02.2025
Department of Geography, University of Zurich

Small-Scale Spatial Variation in Meerkat Vocalization

GEO511 - Master's Thesis

Author
Antonia Hehli
19-750-017

31st of January, 2025

Supervised by
Dr. Peter Ranacher
Nikola Falk (IEU)
Dr. Stuart Watson (IEA)

Faculty Member
Prof. Dr. Robert Weibel

Department of Geography
University of Zurich

Abstract

Language is central to human life. Yet, its evolutionary origins remain subject to debate. To understand how language evolved, it is necessary to study analogs in other species, where simpler forms of communication provide clues to its development. Vocal learning, the ability to acquire and produce new sounds and words through imitation, which is a key feature of human language, is rare among primates but essential for adapting communication to social contexts. It presents an opportunity to study a critical aspect of language evolution in other species, enabling a comparative approach across various taxa. While vocal convergence or divergence between groups, akin to human accents or dialects, due to spatial drivers have been explored in various animal species, they have not been explored in meerkats, where individual and group identity in vocalization was found in prior studies. This work employed a redundancy analysis to identify the influence of the individual emitting the call and biological, social, and environmental drivers on the variation in meerkat close calls. It further developed a framework for defining the spatial interaction probability between meerkat groups, which presents the basis of a potential exchange of acoustic features due to spatial proximity. This framework was incorporated in a correlation analysis, examining the relationship between spatially and temporally proximal close calls. The results illuminate the potential for spatially driven variation in close calls, especially for frequency-related features. These features showed a positive correlation, suggesting an effect of accommodation, possibly due to vocal learning between groups in spatial proximity. While vocal learning between groups in spatial proximity is a possible explanation for this finding, other explanations, such as genetic relatedness, similar environments, or similar social status of calling individuals are conceivable. The results underscore the need to further investigate the observed differences in frequency-related features in specific, and spatial variation in meerkat close calls in general.

Acknowledgements

This project would not have been possible without the continuous support of many people. Especially, I would like to extend my deepest gratitude to my supervisors who supported me during a time of significant personal transitions.

- Dr. Peter Ranacher: Peter, thank you for being the best supervisor I could have asked for. From the countless hours spent finding the most fitting project in the very beginning to our weekly meetings to discuss the current stage and methodological challenges, your guidance has been invaluable. I also deeply appreciate the (funny) conversations we had about life beyond this thesis. Thank you for believing in me from the very start and for embarking on this little journey to Africa with me!
- Nikola Falk: Thank you, Nikola, for organizing the meerkat part of the project and providing me with very well prepared data. Knowing that a walking meerkat encyclopedia was just two floors below for any questions gave me the confidence to pursue this interdisciplinary topic. Thank you for supporting me in this project, all while finishing your own PhD project.
- Dr. Stuart Watson: Thank you, Stuart, for sharing your extensive expertise in the field of evolutionary biology and in working with animal vocalization data. Your input significantly shaped this thesis.

I would further like to thank Prof. Dr. Robert Weibel for being open to such an interdisciplinary topic for my Master's thesis. I am grateful for your inspiration throughout my studies and all the interesting conversations in passing at my office door.

I would like to express my gratitude to Prof. Dr. Marta Manser for her invaluable feedback on my concept talk, her regular input throughout this thesis, and her help in ensuring its biological plausibility.

Further thanks go to all the people that provided ideas for this thesis. Especially, I would like to thank Dr. Nico Neureiter for his outside perspective which helped me solve one of the methodological challenges. I would also like to thank Lavinia Mihai for sharing her work on Brownian bridges with me.

To my friends: Thank you for your emotional support and your companionship not only during this thesis, but throughout all years at university. You made this experience unforgettable.

Last, but not least, my deepest gratitude goes to my parents for their unwavering support in all my endeavors. Your encouragement and belief in me allowed me to fully focus on my studies, and I could not have done this without you.

Contents

List of Figures

List of Tables

List of Acronyms

ANOVA	analysis of variance
BBMM	Brownian bridge movement model
CC	close call
DBSCAN	density-based spatial clustering of applications with noise
GPS	Global Positioning System
IEU	Department of Evolutionary Biology and Environmental Studies
KL	Kullback-Leibler
KMP	Kalahari Meerkat Project
KRC	Kalahari Research Center
MCP	minimum convex polygon
MDS	multidimensional scaling
OSM	OpenStreetMap
PC	principal component
PCA	principal component analysis
RDA	redundancy analysis
RQ	research question
UD	utilization distribution
UZH	University of Zurich

Chapter 1

Introduction

Language is a cornerstone of human behavior, allowing us to convey complex ideas, emotions, and cultural identities (Pagel, 2017; B. Thompson & Smith, 2015). Despite its centrality to human life, the evolutionary origins of language remain a subject of intense debate (Christiansen & Kirby, 2003; Pagel, 2017; Vernes, Janik, et al., 2021). To understand how language evolved, it is necessary to investigate its precursors and analogs in other species, where simpler forms of communication could provide clues to its development (Hauser et al., 2010).

A particularly striking feature of human language is vocal learning – the ability to acquire and produce new sounds and words through imitation (Ruch et al., 2018). This capacity is rare among primates and represents a critical aspect of language evolution. Beyond merely learning new sounds, humans adapt their vocal behavior through vocal accommodation, modifying accents, dialects, and prosody to align with social partners (Janik & Slater, 2000; Ruch et al., 2018). This dynamic interplay between vocal behavior and social identity underscores the deep connection between vocal learning and culture (Whiten, 2021).

The comparative approach, which compares communication systems across taxa, serves as a powerful tool for probing the evolutionary roots of language (Vernes, Janik, et al., 2021). By studying the vocal behaviors of other species, researchers can identify shared and unique features of communication, shedding light on the building blocks of more advanced systems like human language (Collier et al., 2014; Ruch et al., 2018; Searcy, 2019).

The extent to which non-human mammals exhibit group-specific vocal behavior patterns akin to human accents or dialects remains unclear (Vernes, Kriengwatana, et al., 2021). While some evidence suggests the presence of vocal convergence in species like bats and cetaceans, the adaptive benefits and mechanisms driving such behavior are still poorly understood (Ravignani & Garcia, 2022). Investigating whether similar processes occur in terrestrial mammals could illuminate new dimensions of vocal communication and its evolutionary trajectory (Ravignani & Garcia, 2022).

In this thesis, I investigate the vocal behavior of wild meerkats to further explore the individual and group-specific vocal signatures (Townsend et al., 2010). By examining how biological, social, and spatial drivers shape vocal behavior, this research aims to advance our understanding of the conditions under which vocal convergence and divergence emerge. Such insights not only contribute to the study of meerkat communication but also offer broader implications for the evolution of vocal learning in terrestrial mammals.

Known for their complex social structure and cooperative behaviors, meerkats offer a compelling model for studying vocal communication in social mammals. Their sophisticated vocal repertoire, comprising over 30 distinct call types, plays a vital role in coordinating group activities, maintaining social cohesion, and responding to threats (Clutton-Brock & Manser, 2016; Townsend et al., 2010). Among these, the close call (CC) is particularly common in social-foraging contexts (Townsend et al., 2010). Notably, variation in meerkat CCs between individuals and groups (Townsend et al., 2010) provides an opportunity to investigate the drivers of group-specific vocal behavior.

While many biological and social drivers of such variation have been explored, spatial drivers remain understudied. This represents a significant gap, as studies in other species – such as whales, non-human primates, and songbirds – have demonstrated that spatial factors can influence vocalization patterns (Dawbin & Eyre, 1991; De La Torre & Snowdon, 2009; Hensel et al., 2022; Stafford et al., 2001). By analyzing the potential influence of spatial drivers on meerkat vocal behavior, this study seeks to contribute not only to understanding meerkat vocalization but also to the broader discourse on the evolution of vocal learning and group-specific vocal signatures in terrestrial mammals.

This thesis addresses two primary objectives. First, it examines the influence of the individual emitting the call and of biological, social, and environmental drivers on the acoustic features of meerkat CCs. Second, this thesis investigates the role of a spatial driver in shaping these acoustic features. The spatial driver is quantified using a spatial interaction probability. Brownian bridges transform Global Positioning System (GPS) recordings into continuous trajectories of meerkat groups. The time-continuous trajectories are used to calculate a twofold approach of spatial proximity which is then transformed into a spatial interaction probability. A correlation analysis assesses the similarity between the acoustic features of calls in spatial and temporal proximity. By addressing these objectives, this study aims to provide an enhanced understanding of the factors underlying vocal variation in meerkats, with particular emphasis on the largely unexplored spatial dimension.

Chapter 2

Related Work

This chapter begins by providing background information on the meerkats, on the study area, and on the spatial drivers, which present the main factors to study in this thesis. It reviews recent studies investigating the influence of spatial drivers on variations in animal vocalization and examines research on meerkats (section 2.2.2). Additionally, a literature review of the methodological approaches employed in this study is presented (section 2.2.3). The chapter concludes by identifying research gaps in the existing literature and formulating research questions (RQs) that address these gaps (chapter 2.3 and 2.4).

2.1 Background

Before delving into the literature review, this chapter introduces the study species (section 2.1.1) and the study area (section 2.1.2), providing a foundational understanding of the ecological context. Additionally, it defines the concept of spatial drivers as applied within the scope of this thesis (section 2.1.3).

2.1.1 Meerkats

Meerkats (*Suricata suricatta*) are social mongooses native to the southern part of Africa (figure 2.1). Body sizes reach around 30 centimeters and individuals weigh up to one kilogram (Clutton-Brock & Manser, 2016; van Staaden, 1994). The cooperatively breeding mammals are well-adapted to live in the arid and sparsely vegetated open landscapes of southern Africa. However, these landscapes make them easily detectable by predators (Clutton-Brock & Manser, 2016). Meerkats live in stable groups of three to fifty individuals, with group territories spanning over 2–10 km^2 (Manser et al., 2014). Their territories encompass sleeping burrows and boltholes to hide in case of danger during foraging and are defended against neighboring groups. Groups exhibit strict social hierarchies, characterized by one dominant female and one dominant male, contributing to the birth of most offspring. Additionally, groups contain a number of subordinate individuals some of which are immigrant males; female immigration, however, does not occur (Clutton-Brock & Manser, 2016).

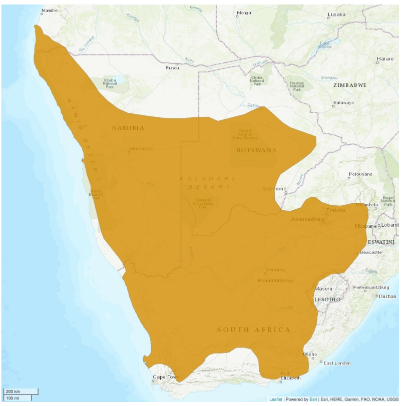


Figure 2.1: Meerkat habitat (Jordan & Do Linh San, 2015).

Close Calls Meerkats have one of the most sophisticated vocal repertoires among mammals, with more than 30 discrete call types (Clutton-Brock & Manser, 2016; Townsend et al., 2010). Vocal communication between group members is vital for meerkats as it enables coordination of collective group actions (Clutton-Brock & Manser, 2016; Engesser & Manser, 2022). Emitted calls depend on behavioral contexts. CCs, for example, are produced in a social foraging context, with individuals searching for food one to ten meters apart from each other (Gall & Manser, 2017). Due to individuals searching for prey in the sand with their head oriented downwards and, therefore, impaired visual communication (figure 2.2), CCs are important to maintain group cohesion and spacing between group members (Gall & Manser, 2017; Townsend et al., 2010, 2011). CCs are short, pulsated calls with a fundamental frequency of 600–1 000 Hz which are produced 5 to 20 times per minute (see figure 2.3). The acoustic signal of CCs can propagate 20 to almost 50 meters without fully degrading (Engesser & Manser, 2022; Garcia Arasco et al., 2022; Townsend et al., 2010). Among individuals and groups, the acoustic signature in CCs varies (Townsend et al., 2010). There is no conclusive evidence on which factor is the main driver for this variation (Townsend et al., 2010). Potential factors are genetic determination and vocal learning (details in section 2.2.1) (Townsend et al., 2010).

Figure 2.2: Meerkats foraging in the Kalahari with their head oriented towards the sand (pictures by Nikola Falk, 2022).

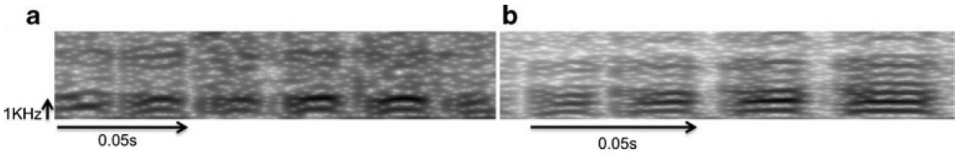


Figure 2.3: Spectrogram of two CCs, in a guarding (a) and foraging (b) context (Townsend et al., 2011, p. 1929).

CCs are particularly suited to study meerkat vocalization due to various reasons. CCs are used for group cohesion during foraging where visibility is impaired (Gall & Manser, 2017; Townsend et al., 2010, 2011). When a CC is heard, it is important who the caller is as the meerkats cannot see each other well, resulting in individual and group signatures (Townsend et al., 2010). Besides the CCs' advantage of data availability, other studies have used them to analyze various influences on meerkat vocalization.

CCs are recorded in a so-called 'focal follow', which is possible when meerkats are habituated to human presence. A researcher follows a meerkat with a directional microphone and a stereo recorder. The calls are recorded while the researchers explains what the meerkat is doing. This way, it can be ensured that the recorded call is assigned to the correct meerkat and the behavioral context to the calls is known.

2.1.2 Study Area

The study area for this Master's thesis project is the Kalahari Research Center (KRC) and its surrounding land located in the Kuruman River Reserve in the northern part of South Africa. The KRC was founded in the Kgalagadi National Park in north-west South Africa and relocated to the present-day Rus en Vrede study site in 1993. The site spans approximately 50 square kilometers. It is located around 30 kilometers west of Van Zylsrus, close to the Botswana border (figure 2.4). The habitat around the reserve is arid, containing the dry Kuruman riverbed, sand dunes, and sparsely vegetated flats. Vegetation, if present, consists of grasses, shrubs, and trees. Rainfall is concentrated on the summer months November until April and can vary strongly depending on the year (Kranstauber et al., 2020). The KRC hosts multiple biological research projects including the Kalahari Meerkat Project (KMP) which is the largest and oldest study at the center, being in operation since 1993. It encompasses between 12 and 15 groups with a total of 250 to 300 meerkats. Meerkats at the KMP are habituated to an observer's presence of one meter and individuals can be identified unambiguously based on characteristic visible marks (KRC, n.d.).

Figure 2.4: Location of the KMP study site in South Africa. Background provided by OSM; map created in QGIS 3.22 Białowieża (QGIS Development Team, 2021).

2.1.3 Spatial Drivers

As mentioned in the introduction (see chapter 1), for this thesis, I am especially interested in spatially driven variation in meerkat CCs. In the following, I define spatial drivers for the scope of this thesis.

Spatial phenomena can be depicted and classified in various ways with spatial data which can be represented in different data formats such as vectors and rasters. Vector data can be further divided into polygon, line, and point data (Heywood et al., 2011). One possibility to classify point data is to separate them into first- and second-order properties. The former focuses on the absolute position of individuals and their interaction with the environment at that exact location. The latter aims to analyze the relative position and interaction between individuals (Yuan et al., 2020). In accordance with these properties, spatial drivers can be defined twofold: environmental drivers and proximity relations.

Environmental drivers can be thought of as first-order properties, as they represent the features on the Earth's surface, in this case the habitat, at the absolute position of an individual or a group. Proximity relations can be thought of as a second-order property. They represent the relative position between individuals or groups. On the one hand, proximity can be expressed by distances such as 'A and B are 500 meters apart'. On the other hand, proximity can be expressed by topology, as in 'A is positioned between B and C'. In terms of acoustic features, proximity may have two consequences. Features between close groups could either be especially alike or especially different from each other (Clink et al., 2017). Acoustic features may have been passed on to neighboring groups, leading to two groups with similar features. This could be an indicator for call accommodation between groups in spatial proximity. Yet, they could also be intentionally produced in a different fashion to clearly separate from groups close by (Clink et al., 2017). Both processes could suggest the possibility of vocal learning in meerkats.

For this work, spatial drivers under consideration are proximity relations, as the land cover, representing the first-order property, is quite consistent across the study area. Analyzing a land cover image of the study area from 2018 (figure 2.5), it becomes evident that low shrubland is predominant (over 95%) across the study area. This land cover class, characterized by a plant canopy cover of heights between 20 centimeters and two meters (M. Thompson, 2019), is mostly consistent across recording locations, and therefore, no influence is expected. This assumption is supported by a study which found no evidence for influence of vegetation density on acoustic structures of CCs (Driscoll et al., 2024). Consequently, the distance to the riverbed plays a minor role in this thesis as an environmental driver, while the primary spatial drivers assessed are proximity relations, i.e., the second-order property, and are termed 'spatial proximity'.

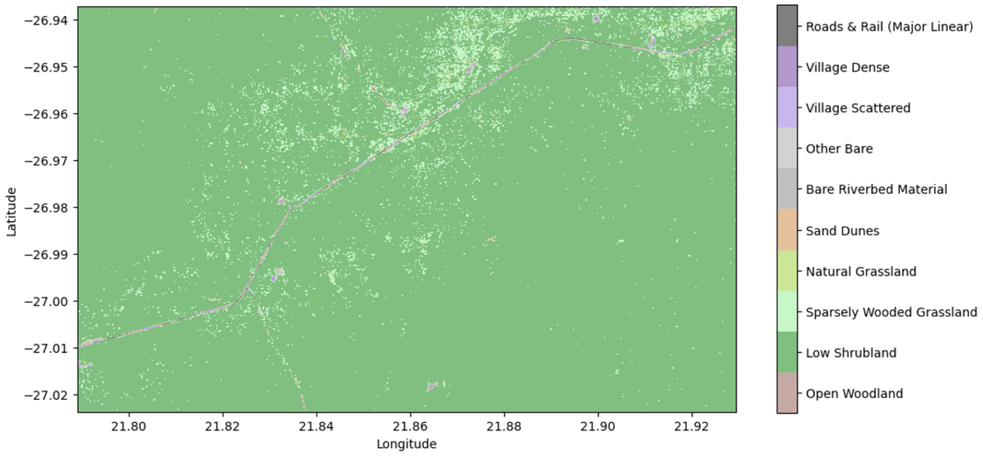


Figure 2.5: Land cover of the study area in South Africa in the year 2018 at 20 meter resolution (M. Thompson, 2019).

2.2 Related Work

This literature review synthesizes existing research on vocal learning, a remarkable feature of human communication, and examines advances in the study of animal vocalization on the same topic. The first part explains the foundations of different forms of vocal learning (section 2.2.1), while the second focuses on animal vocalization, emphasizing the influence of spatial drivers across taxa (section 2.2.2), followed by a detailed examination of research on meerkat CCs and space use. Finally, the review discusses key methodological approaches employed in this thesis, offering an overview of their characteristics (section 2.2.3).

2.2.1 Human Language

Language is a cornerstone of human behavior that allows us to communicate complex ideas, emotions and cultural identities (Pagel, 2017; B. Thompson & Smith, 2015). Despite its importance to humankind, its evolutionary origins remain a subject of scientific debate (Christiansen & Kirby, 2003; Pagel, 2017; Vernes, Janik, et al., 2021). To understand the evolution of language, it is necessary to investigate its precursors as well as analogs and differences to other species (Hauser et al., 2010). Studying species where simpler forms of communications exist may provide clues to the development of communication systems (Hauser et al., 2010).

A striking and fundamental feature of human language is vocal learning – the ability to acquire and produce new sounds and words through imitation (Ruch et al., 2018; Vernes, Janik, et al., 2021). Vocal learning encompasses two main parts which are contextual learning, where signals or calls are not modified, and production learning, where they are (figure 2.6) (Janik & Slater, 2000).

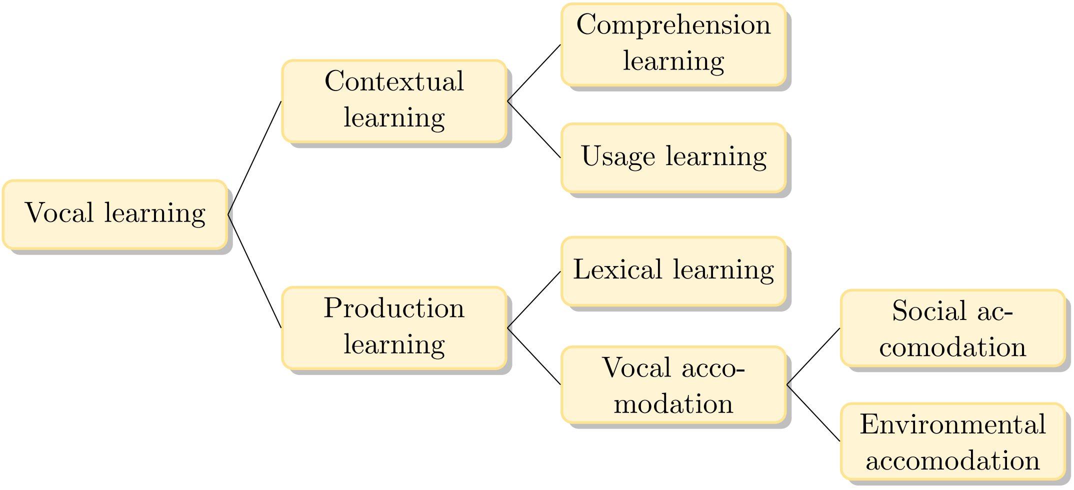


Figure 2.6: Subcategories of vocal learning (Ruch et al., 2018, p. 998).

Contextual learning means that an already existing signal is associated with new context. This is divided into the signaler's and the receiver's perspective. The signaler learns to use a call within a new context (usage learning), while the receiver learns to associate the pre-existing call to a new context (comprehension learning) (Janik & Slater, 2000; Ruch et al., 2018). Production learning can be split into lexical learning, which refers to the acquisition of new call types, and into vocal accommodation, referring to subtle modification of existing call types. Vocal accommodation includes environmental and social accommodation. Environmental accommodation refers to situations where the signal transmission is optimized. For example, people start to speak louder when their communication partner is farther away. Social accommodation is used to express social closeness or distance (Ruch et al., 2018). It is important to note that not every change in vocalization is caused by a form of learning. In fact, it is rather difficult to determine if and in what form learning is responsible for changes (Janik & Slater, 2000). The capacity of vocal learning is rare among primates, yet, it represents a critical aspect of language evolution (Vernes, Janik, et al., 2021). Studying vocal learning in a wide range of animals can help us understand why it arose productively in our own evolution (Vernes, Janik, et al., 2021).

2.2.2 Animal Vocalization

As mentioned above, a comparative approach across human and non-human taxa is beneficial in many lights of language evolution (Collier et al., 2014; Joseph, 2015; Thomason, 2001). The extent to which non-human mammals exhibit group-specific vocal behavior patterns like human accents or dialects is studied intensively across various taxa (Vernes, Kriengwatana, et al., 2021). In the following, there is an overview over recent developments in the field of group-specific variation in vocalization with special attention on spatial drivers of these variations.

2.2.2.1 Influence of Spatial Drivers in Various Taxa

Spatial drivers were found to influence animal vocalizations in different species. In marine mammals, whale songs are often subject to research. In the early 1990s, a study analyzed whale songs at the Australian east and west coast and found variations between these two habitats (Dawbin & Eyre, 1991). A decade later, variation in vocalization patterns between blue whales in the eastern North Pacific and the western Pacific were found (Stafford et al., 2001).

Due to their close genetic relation to humans, non-human primates are often subject to vocalization studies. While it had already been demonstrated that certain chimpanzee vocalizations differ between individuals, a study has additionally found variation in vocalization of two chimpanzee populations in different locations in Tanzania (Mitani et al., 1992). Various hypotheses on the cause of this variation were discussed, among them were genetic differences, behavioral contexts, and dialectal reasons. Calls exhibiting dialects would suggest the presence of vocal learning (Mitani et al., 1992).

In Ecuador, a study analyzed calls of five different wild populations of pygmy marmosets. These calls are used to maintain short-range contact and coordinate interaction between group members, comparable to a meerkat's CC. Although partly accounted to habitat acoustics, this study was among the first ones to find evidence of dialects in wild populations of a neotropical primate species, meaning primates living in the tropical regions of the Americas (De La Torre & Snowdon, 2009). Staying in the same family of animals, more recently, a study analyzed three call types of three different captive populations of common marmosets in distinct locations in Europe. They found significant differences between these populations which could neither be attributed to status of the individuals nor to physical and environmental differences (Zürcher & Burkart, 2017).

Besides being the first eusocial mammal that was identified, naked mole-rats are highly cooperative and show a vocal repertoire with at least 17 distinct vocalizations, making them comparable to meerkats (Barker et al., 2021). Analyzing greeting calls from seven naked mole-rat colonies in South Africa and in Germany revealed the call features being predictive of colony identity. Besides this colony identity, naked mole-rats' calls also convey individual identity. Further, they are able to distinguish colony-specific features in vocalizations (Barker et al., 2021).

Among oscine songbirds, patterns of geographical variations – vocal dialects – are very common due to their vocal learning capability. While most of these dialects where described over long distances of hundreds or thousands of kilometers, a recent study investigated variations in local neighborhoods of territorial birds, Savannah sparrows. Their breeding sites might be located within hundreds of meters. The study found microdialects – variation in vocalization spanning over small distances – in male Savannah sparrows:

Individuals living near each other show greater acoustic similarity in their songs than those living farther apart (Hensel et al., 2022).

While vocal dialect analyses are very common for songbirds, vocalizations of parrots were less often studied. A study with focus on Australian palm cockatoos analyzed the number of shared call types between different sites. It was found that the proportion of shared call types between different sites was, in general, low. No combination of sites shared more than half of their calls, with almost one third showing no shared calls. Although there seems to be some spatial influence, no correlation between straight-line distance and call sharing was found (Keighley et al., 2017). Interestingly, the population sharing the greatest number of calls with other populations was located the farthest apart from them. The population site with the fewest shared calls was considerably closer to other sites, indicating that there might be some kind of influence of spatial proximity on call types (Keighley et al., 2017). Further, within and between-population analyses on contact calls were conducted, investigating if call structure changes correlate with geographic distance. Within a population, significant positive relationships between call similarity and straight line geographic distance between sites were found for almost all populations. However, when analyzing between populations, no significant correlation between acoustic and geographic distance between populations was found (Keighley et al., 2017).

The above examples provide, at least partly, evidence for the presence of spatial drivers. Yet, there were also studies unable to find evidence for influence of space on vocalization. Relating to the paper investigating chimpanzee calls in the 1990s discussed above (Mitani et al., 1992), a more recent study analyzed chimpanzee calls at three different sites. Two neighboring sites in Tanzania and one geographically distant site in Uganda were studied. Differences in chimpanzee vocalization of these three populations were argued to represent individual rather than community identity, providing no evidence for dialects and, thus, vocal learning (Desai et al., 2022).

Research on agile gibbons calls in Sumatra and Borneo, which are currently considered as distinct subspecies, focused on evaluating the island effect of geographical isolation on variation in vocalization. No discernible differences were observed between these populations, making it impossible to discern between the two subspecies based on their vocalization. The intraindividual variation was found to be much larger than any interindividual or between-population variations (Heller et al., 2010).

A few years later, a study investigated individuality and small-scale geographic variation in an area of around ten square kilometers among female Bornean gibbons great calls. They found that interindividual variation was greater than intraindividual variation. Additionally, there was no proof for a relationship between call similarity and geographic distance, females of close territories were not more likely to have similar calls than females inhabiting more distant territories (Clink et al., 2017).

Other mammals which were under investigation for spatial variation in their vocalization include bats. Bats have a large vocal repertoire for social communication, so-called social vocalization, alongside them emitting sound pulses for echolocation. Variation in call acoustics of the social vocalization were attributed to differences within a geographic region rather than across geographic regions (Lin et al., 2022).

2.2.2.2 Meerkats

Besides the taxa mentioned above, meerkats are particularly suitable to study animal vocalization due to their complex social structure, sophisticated vocal repertoire as well as their ability to communicate information on external happenings (Clutton-Brock & Manser, 2016; Searcy, 2019; Townsend et al., 2010). Studies on meerkats have extended beyond the scope of vocal communication covering additional topics like space use. Besides introducing studies on meerkat CCs, I address studies on space use that I consider relevant for this work.

Context-Specific Variation Research has explored variations in acoustic features of CCs in different behavioral contexts. For example, a study analyzed the acoustic structures of foraging and guarding CCs. Foraging CCs, used to maintain group cohesion during foraging (see section 2.1.1), and guarding CCs, signaling a predator-free environment after scanning it, differ in their acoustic structures (Townsend et al., 2011). Meerkats can discriminate between these calls, with distinct temporal and spectral features distinguishing guarding from foraging CCs (Townsend et al., 2011). Furthermore, variations in acoustic features have been identified not only across these two contexts, but across six different behavioral contexts (Driscoll et al., 2024). Concluding, CCs encode contextual information regarding many different behavioral contexts (Driscoll et al., 2024) and meerkats are able to attend to the information conveyed in a CC, at least in a guarding and foraging context (Townsend et al., 2010).

Individual Traits Alongside context-specific variation and individuality in meerkat CCs acoustic features (Townsend et al., 2010), it has been shown that group members adjust their behavior based on their social relationship with the calling individual (Reber et al., 2013). This suggests that variations in the acoustic structure of CCs may be influenced by traits such as dominance status, age, or sex (Mausbach et al., 2017). For instance, CCs of dominant females have been found to exhibit the lowest mean fundamental frequencies and the longest mean pulse durations (Mausbach et al., 2017). It is hypothesized that these distinctions are the result of a growth spurt associated with attaining dominance, altering the vocal tract and leading to lower frequencies (Mausbach et al., 2017). Physiological changes associated with increasing age or changing social status could be possible reasons for individuality in the acoustic structure of meerkat CCs.

Group-Specific Variation Beyond individuality in acoustic structure of CCs, researchers are also interested in group-specific information in CCs. Considering that meerkats use CCs to maintain group cohesion during foraging, it is possible that group-specific information is encoded in the call. A study investigated group-specific variation underlying meerkat CCs and if this information can be used by call receivers to discriminate between individuals belonging to their group and the ones belonging to foreign groups (Townsend et al., 2010). Results show that CCs carry information on individual identity (Townsend et al., 2010). Furthermore, when controlling for individual identity, calls encode group-specific acoustic signatures. However, these group-specific signatures cannot be decoded by receivers and have no adaptive function (Townsend et al., 2010). The question arises why CCs encode group-specific features, even if receivers are unable to use the information. Hypotheses are that there is a high level of genetic relatedness within a group, leading to physical similarities between group members and consequently similar calls. Although presented as less likely, it is also possible that vocal learning is responsible for acoustically similar calls along group members. Concluding, meerkat calls encode group-specific signatures. Since receivers cannot make use of the information, there is no conclusive evidence on the causes of group-specific signatures (Townsend et al., 2010).

Neighbor Proximity A recent study investigated the influence of individual neighbor proximity and contextual situation on acoustic structure of CCs (Driscoll et al., 2024). While behavioral context remains a significant factor in acoustic variation, the distance to the nearest individual neighbor of the same group was found to only have limited influence on the acoustic structure. However, the frequency of a CC increases when individuals are in a distance of less than one meter (Driscoll et al., 2024). Moreover, a very small effect of proximity-driven arousal on the acoustic structure was found. In accordance to earlier studies, it was hypothesized that call rates of CCs increase when neighbors are in close proximity (Driscoll et al., 2024; Engesser & Manser, 2022).

Spatiality and Territorial Behavior In addition to their sophisticated vocal communication, meerkats exhibit spatial behaviors that are crucial for their survival. CCs play a key role in coordinating group movements and maintaining group coherence and, thus, safety during foraging, but territorial behavior are also important for their survival. Meerkats are territorial animals that aggressively defend their territory against intruding groups (Clutton-Brock & Manser, 2016; Kranstauber et al., 2020). Research on spatiality explores how meerkats orientate themselves within a territory and how territorial patterns and shifts can be identified and modeled (Bateman et al., 2015; Kranstauber et al., 2020; Manser & Bell, 2004).

Orientation within Territories Meerkat foraging in the open land are frequently threatened by predators (Manser & Bell, 2004). They need to find shelter in boltholes as quickly as possible when hearing an alarm call and therefore, need to have some idea where the closest place to hide is located (Kranstauber et al., 2020; Manser & Bell, 2004). Due to the large number of boltholes in their territories, meerkats are able to find shelter quickly, regardless of the direction they ran to (Manser & Bell, 2004). The study found that, when exposed to alarm calls, in over 80% of the cases, meerkats ran to the closest bolthole available which always was within 60 meters of their present location. They found the hole regardless of whether they passed it during preceding foraging activity. Boltholes created by researchers were mostly ignored, even when they were closer than existing boltholes built by meerkats. These results suggest that adult meerkats have an accurate knowledge of distance and direction to the closest shelter in relation to their own position in the territory at any time (Manser & Bell, 2004). The fact that the meerkats did not always run to the closest bolthole suggests that they have a selective shelter knowledge, but still know a large proportion of their territory. Most likely, meerkats frequently update their present local view of the environment by place recognition or reorientation of specific landmarks. Place recognition refers to the technique of forming ”snapshot views so that the visual image [...] is about the same each time an animal approaches the specified location” (Manser & Bell, 2004, p. 151-152). Reorientation is primarily, but not exclusively, based on geometric features of a meerkat's environment (Manser & Bell, 2004). Summarizing, meerkats likely do not have a complete spatial map of their territory, but rather navigate using momentary and egocentric representation (Manser & Bell, 2004).

Territory Models Research exploring territorial patterns including scent marking, group interactions, and habitat selection was done using mechanistic home range models derived from a spatially biased random-walk process (Bateman et al., 2015). Adding a dynamic extension, territory development and shifts can be assessed. The model that best described meerkat home ranges was found to be independent of group size, but influenced by neighbor groups (Bateman et al., 2015). Groups' sizes fluctuate within more or less consistent territories. A lot of the change in model patterns resulted from altered local interactions like surrounding groups forming or disappearing (Bateman et al., 2015).

Territory Patterns and Shifts A study focusing on space use changes in meerkats aimed to identify different space use patterns and shifts in territories. Investigating territory changes and their causes is appealing as a lot of advantages from a territory are lost when moving (Kranstauber et al., 2020). They found that, generally, home ranges in meerkats are stable and remain mostly unchanged over extended periods of time. Territory shifts can occasionally be observed. However, they were found to occur more often between October and April. Shifts were found to have no relation to environmental conditions: The utilization distribution (UD) in the three months prior to the move did not have a lower average vegetation index than the UD after the move (Kranstauber et al., 2020). Further, shifts were neither found to show relation to dominance or group size changes nor to space use changes of other groups. Rather, in accordance with the above results (Bateman et al., 2015), shifts were found to be related to preceding group interactions and competition (Kranstauber et al., 2020). Causes of such shifts were hypothesized to either be the option to gain better territories, or being forced to move by another group. While such large territory shifts with half or a full diameter of home range size occur suddenly and are not directionally aligned, smaller consecutive shifts within around four months are directionally aligned and rather small (Kranstauber et al., 2020).

2.2.3 Methods

The multitude of studies presented above have used different methodological approaches to investigate drivers of variation in animal vocalization. In this section, I provide related work on the most relevant methods I employed in this thesis, offering insight into their applications as well as their strengths and limitations. These include principal component analysis (PCA), the redundancy analysis (RDA), and Brownian bridges.

2.2.3.1 Principal Component Analysis

PCA is a dimension reduction method, belonging to the class of unconstrained ordination, and was first introduced at the beginning of the 20th century (Borcard et al., 2011; Gelfand et al., 2017; Jolliffe & Cadima, 2016; Quebec Centre for Biodiversity Science, 2023a). The goal of ordination methods is to represent the input data along a reduced number of orthogonal axes, such that they represent the main trends of the data in a decreasing order. PCA rotates the original system of axes and constructs new orthogonal axes that explain most of the variation in the quantitative input data (Borcard et al., 2011; Gelfand et al., 2017). In other words, it creates linear combinations of the original variables that account for as much variation as possible (Quebec Centre for Biodiversity Science, 2023a). The resulting principal components (PCs) give the positions of objects in this new coordinate system. As other unconstrained ordination techniques, PCA is solely descriptive (Borcard et al., 2011). The relationships detected by PCA are linear and the preserved distances between the observations are Euclidean (Borcard et al., 2011). Due to its complexity-reducing characteristics, it has become a commonly employed method in data analysis, is implemented in many statistical softwares, and simple to apply (Borcard et al., 2011; Gelfand et al., 2017; Quebec Centre for Biodiversity Science, 2023a). Further, no prior information about the data is required, making PCA an easily applicable method to get a first overview over the data (Gelfand et al., 2017).

As every other method, PCA comes with drawbacks. The newly defined components change if the units of measurement on one or many variables change (Jolliffe & Cadima, 2016). A standardization of the input data is necessary to amend this problem. PCA on standardized data is called correlation matrix PCA (Jolliffe & Cadima, 2016). Further, PCA is sensitive to outliers and gross errors in the dataset (Jolliffe & Cadima, 2016).

Another disadvantage is that the newly defined components are not trivial to interpret, especially if they are composed by many different input variables. Lastly, rather a challenge than a real disadvantage is the choice of the number of PCs to retain. Approaches can either be of intuitive and descriptive nature, such as setting a proportion of total variance that needs to be explained, or using models to find the optimal amount of components (Jolliffe & Cadima, 2016).

2.2.3.2 Redundancy Analysis

While the above discussed PCA is an unconstrained ordination, the RDA is a canonical ordination which associates two or more data sets in the ordination process itself. The method was developed in 1977 and its goal is to explore relationships between two matrices (Van Den Wollenberg, 1977). One can also formally test statistical hypotheses about the significance of the relationships found (Borcard et al., 2011). The roles of the matrices are not symmetrical, i.e., one is the dependent and one the explanatory matrix, which is why RDA can be classified as an asymmetrical canonical ordination (Borcard et al., 2011).

RDA combines multiple regression with classical ordination, specifically with a PCA. It models multivariate multiple response data, which makes it a direct extension of multiple linear regression (Borcard et al., 2011; Quebec Centre for Biodiversity Science, 2023b). First, it regresses each response variable on the explanatory variables and combines them into a matrix of fitted values. In a next step, a PCA is performed on that matrix. The results are canonical axes that correspond to linear combinations of the explanatory variables in presence of the response variables (Quebec Centre for Biodiversity Science, 2023b).

Partial RDA is the multivariate equivalent to partial linear regression. The effects of the explanatory variables on the response variables are adjusted for the effects of a third set of variables, the covariables (also called covariates) (Borcard et al., 2011; Quebec Centre for Biodiversity Science, 2023c). First, a RDA of the covariables on the response variables is performed and the residuals of this analysis are extracted. Simply speaking, the effect of the covariables on the response variables is removed before performing the RDA with the explanatory variables (Quebec Centre for Biodiversity Science, 2023c).

An advantage of the RDA is that it allows a multivariate response data which is not possible in standard regression (Borcard et al., 2011). Further, RDA allows testing for a hypothesis of absence of a linear relationship between the explanatory and response matrix as opposed to PCA (Borcard et al., 2011). Disadvantages of the RDA are that it assumes a linear relationship between the two matrices and that missing values are not allowed (Borcard et al., 2011).

2.2.3.3 Brownian Bridges

The concept of Brownian bridges originates from Brownian motion, first described by botanist Richard Brown in 1827 (Kahane, 1997). Standard Brownian motion is a Gaussian stochastic process, characterized by independent increments. A Brownian bridge is a Brownian motion confined to the time interval $[0, 1]$ and conditioned to terminate at a value of 0, i.e., at a known end point (Horne et al., 2007). It can be understood as a "continuous-time stochastic model of movement" (Horne et al., 2007, p. 2354). The probability of being in an area is conditioned by three parameters which are the start and end locations, the time elapsed between start and end point, and the speed of movement (Horne et al., 2007). The theory of Brownian motion and bridges has found applications in various fields such as mathematics, physics, geography and ecology (Kahane, 1997).

Brownian bridges were first used to depict animal movement and estimate animal home ranges in the late 1990s (Bullard, 1999; Horne et al., 2007). Since then, Brownian bridges have been widely applied in animal ecology and continuously extended (Horne et al., 2007). For example, the Brownian bridge movement model (BBMM), which is built on a conditional random walk between consecutive pairs of locations with the assumption that these are constantly diffusive without any bias towards an area (Byrne et al., 2014). The BBMM depends on the time elapsed between the first and the second location, on the distance between the locations, and on the Brownian motion variance related to the animal's mobility (Horne et al., 2007). This model has been used to estimate animal home ranges, migration routes, the influence of resource selection on animal movement patterns, and to model disease outbreaks or encounter rates of animals (Horne et al., 2007; Kranstauber et al., 2012). Related to analyzing migration patterns is the risk for birds to collide with hazards such as electrical power lines along their way which was also assessed using the BBMM (Watts et al., 2015).

Another challenge in movement ecology addressed with the help of Brownian bridges are low GPS sampling rates (Buchin et al., 2012). Low sampling rates often occur for movement data collected on wild and free-ranging animals. It is difficult and invasive to equip wild animals with devices, meaning that the observation period is limited by battery-life. Battery-life, in turn, is often anti-proportional to the sampling rate. Due to this constraint, sampling rates are frequently chosen to be lower than where a linear interpolation between two recorded points would make sense (Buchin et al., 2012). Many studies overlook the fact that linear interpolation between relocations with a low sampling rate could lead to high uncertainty. To amend the problem, an efficient framework was developed using the BBMM to interpolate between two relocations with low sampling rates (Buchin et al., 2012).

Advantages of the BBMM include the usage of an animal's movement path rather than individual points, for example, for defining its UD (Kranstauber et al., 2012). Further, it accounts for temporal autocorrelation and high data volumes and is easy to apply to many different movements (Kranstauber et al., 2012). A disadvantage of the BBMM is that homogeneity in movement behavior is assumed since the diffusion coefficient is a constant, which might not represent the reality of animals (Kranstauber et al., 2012).

2.3 Research Gaps

Based on the literature review in the above chapter 2.2, two major research gaps can be identified.

1. There is a lack of evidence on the factors driving individual and group-specific variation in meerkat CCs.

2. There is a lack of evidence on spatial drivers of meerkat vocalization. Space could influence CCs through topography or proximity. An example for topography would be the dry Kuruman riverbed that passes through the study area. Proximity can be understood as the positional relationships between meerkat groups that can either be expressed as distances or topology.

This thesis aims to address these gaps, suggesting a framework to analyze spatial proximity as a possible driver of variation in meerkat vocalization.

2.4 Research Questions

The lack of research surrounding spatial proximity as driver of variation in meerkat vocalization, leads to the following RQs that will be addressed in this work.

> **RQ1**: How do biological, social, and environmental drivers influence the acoustic features of meerkat CCs?

Variation in acoustic structure of meerkat CCs has been evident for over a decade, with potential drivers explored in previous studies (see section 2.2.2.2) (Townsend et al., 2010). Potential drivers assessed in this study include the individual meerkat, group identity, dominance status, sex, age, genetic relatedness, and distance to Kuruman's riverbed. These drivers can be categorized into three groups: biological drivers (age and sex), social drivers (group identity, dominance status, and genetic relatedness), and environmental drivers (distance to the riverbed). The individual meerkat emitting the call does not fit into any of these categories and is a standalone driver. Based on findings of previous studies, I hypothesize that the individual has the greatest influence on the acoustic features of CCs (Townsend et al., 2010). Additionally, I expect group identity to influence acoustic features, as studies have shown group-specific differences after accounting for individual variation (Townsend et al., 2010). I anticipate less influence from the social driver dominance status, followed by the biological driver sex, since it was found that CCs of dominant females differed in fundamental frequency compared to individuals of other status and sex (Mausbach et al., 2017). Differences in sex could result in different body sizes which was found to have some influence on vocalization frequency in different mammalian species (Martin et al., 2017). The second biological driver, age category, is expected to have less influence since mainly differences in calling rates between different meerkat ages were found (Mausbach et al., 2017). Genetic relatedness is hypothesized to have the least influence on the acoustic structure of meerkat CCs among the social drivers. The environmental driver, represented by the distance to the river, is anticipated to have the least influence on the acoustic features of CCs.

> **RQ2**: To what extent do spatially and temporally close meerkats influence each other's CCs? Is there evidence for accommodation, where spatially and temporally close meerkats from different groups have similar CCs, or for divergence, where spatially and temporally close meerkats have different CCs?

I hypothesize that spatially and temporally close meerkats influence each other's CCs, with the potential for both, accommodation and divergence. Accommodation may occur as similar acoustic features diffuse between groups through spatial proximity. This diffusion could result from vocal learning, where groups hearing each other accommodate their calls to neighboring groups. Additionally, genetic relatedness between spatially proximal groups, due to migration or groups splitting, may facilitate acoustic similarity. For example, if a meerkat joins a new group in spatial proximity, the members of the group might accommodate their calls to the new individual. Similarly, if subordinates split from a group to form a new one nearby, shared genetic material could explain call similarities between these groups. Divergence may arise as a mechanism to reinforce group identity, helping to distinguish between members of the same group and those of neighboring groups (Clink et al., 2017). Such divergence could be particularly advantageous in competitive or territorial contexts, where a clear delineation between groups reduces confusion and strengthens group identity. Inklings of such group differentiation were found in the vocalization of palm cockatoos (details in section 2.2.2) (Keighley et al., 2017).

Chapter 3

Data

All data used for my Master's thesis project was provided and preprocessed by the Department of Evolutionary Biology and Environmental Studies (IEU) at the University of Zurich (UZH). I received a dataset with processed meerkat CC recordings (chapter 3.1), access to a meerkat database (chapter 3.2), and group-wise GPS way points (chapter 3.3).

3.1 Close Call Recordings

The main data used for this Master's thesis consists of 9 167 CCs of 155 individuals, belonging to 16 different groups, that were recorded in the years 2006, 2012 to 2020, and 2022. The dataset only includes CCs of two similar behavioral contexts to minimize variation due to context. The number of CCs and their affiliated group varies drastically among the different years (see figure 3.1). The CCs were processed into 26 different, but correlated acoustic features (see table 3.1) using the `spectro_analysis` (formerly `specan`) function of the 'warbleR' package with a window size of 100 ms (Araya-Salas & Smith-Vidaurre, 2017). Further, for each call, there is additional information about the meerkat emitting the call, such as the reference group – the group to which it belonged at the moment of recording (hereafter referred to as 'group') – the group which it was born in, its birth date, and many more. There are also coordinates representing the group's center on the day of recording.

3.2 Meerkat Database

Besides the call recordings and associated variables, I was provided access to the KMP database that has a vast amount of parameters on meerkats available. This database has grown over the past 20 years and is continuously updated by researchers working for the KMP. From the database, I extracted tables with information about births, dominance events, and group composition, such as when an individual was first or last seen in a group.

3.3 Group Way Points

The GPS way point dataset which is also part of the KMP database was preprocessed and provided by researchers of the IEU at UZH. These way points are manually recorded by following the meerkat groups in two session per day, each lasting for three hours. The GPS way points were recorded per group at its center, meaning that one point represents the location of the group's center at the given time. Alongside the GPS way points, the dataset includes the group and the timestamp of recording. The sampling interval is generally every 15 minutes, but shorter or longer intervals occur as well. The dataset

Table 3.1: Acoustic features as a result of the `spectro_analysis` function (Araya-Salas & Smith-Vidaurre, 2017).

Variable name	**Explanation**	**Unit**
duration	length of the signal	s
meanfreq	mean frequency of the signal which is calculated as the weighted average of the frequency spectrum	kHz
sd	standard deviation of the frequency which is calculated as the weighted standard deviation of the frequency spectrum	kHz
freq.median	median frequency	kHz
freq.Q25	first quartile frequency	kHz
freq.Q75	third quartile frequency	kHz
freq.IQR	interquartile frequency range with frequencies between freq.Q25 and freq.Q75	kHz
time.median	median time at which the time envelope is divided into two time intervals of equal energy	s
time.Q25	first quartile time	s
time.Q75	third quartile time	s
time.IQR	interquartile time range with the time range between time.Q25 and time.Q75	s
skew	skewness, asymmetry of the frequency spectrum	-
kurt	kurtosis, peakedness of the frequency spectrum	-
sp.ent	spectral entropy, energy distribution of the frequency spectrum (pure tone approximately 0; noisy approximately 1)	-
time.ent	time entropy, energy distritbution on the time envelope (amplitude concentrated in a specific time point approximately 0; amplitude equally distributed across time approximately 1)	-
entropy	spectrographic entropy (sp.ent*time.ent)	-
sfm	spectral flatness which is similar to sp.ent	-
meandom	average of dominant frequency measured across the spectrogram	kHz
mindom	minimum of dominant frequency measured across the spectrogram	kHz
maxdom	maximum of dominant frequency measured across the spectrogram	kHz
dfrange	range of dominant frequency measured across the spectrogram	kHz
modindx	modulation index, which is the cumulative absolute difference between adjacent measurements of dominant frequencies divided by the dominant frequency range (signal is not modulated: 1)	-
startdom	dominant frequency measurement at the start of the signal	-
enddom	dominant frequency measurement at the end of the signal	-
dfslope	slope of the change in dominant frequency through time, calculated as $\frac{endom - startdom}{duration}$	kHz/s
meanpeakf	mean peak frequency, frequency with highest energy from the mean frequency spectrum	kHz

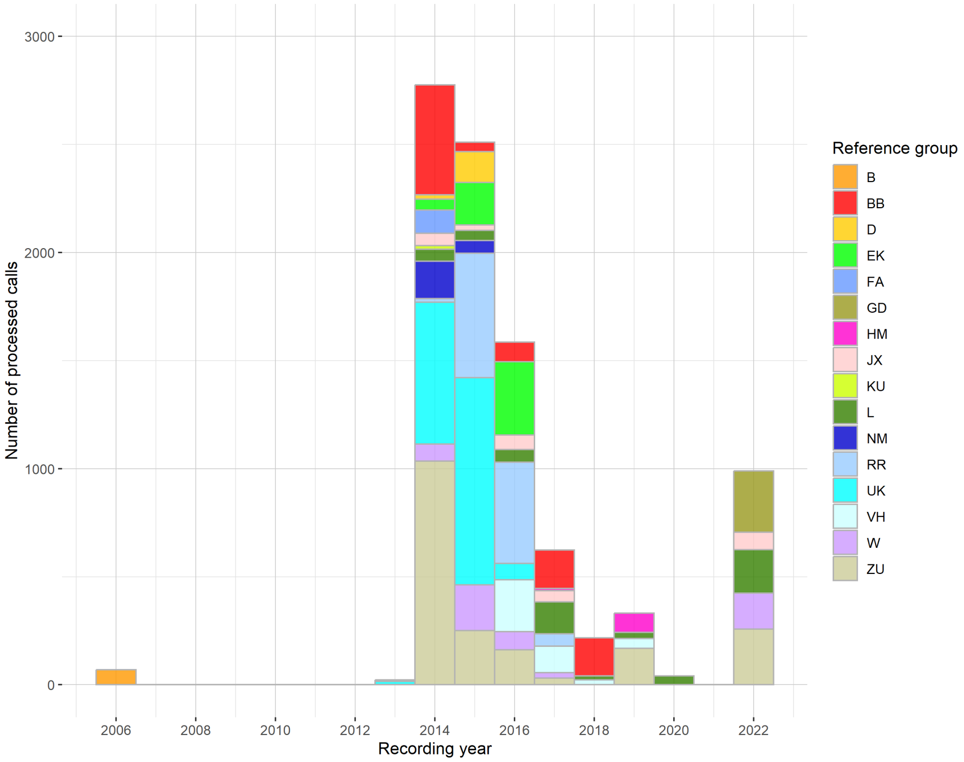


Figure 3.1: Number of preprocessed CCs per year, color coded by reference groups.

contains 567 502 GPS way points of 71 meerkat groups recorded between January, 2001 and March, 2022, with a varying number of recordings per year. 15 080 out of these had no corresponding date and time and were excluded right away, leading to a total of 552 422 GPS way points with all information present, which were provided in WGS84 (EPSG:4326).

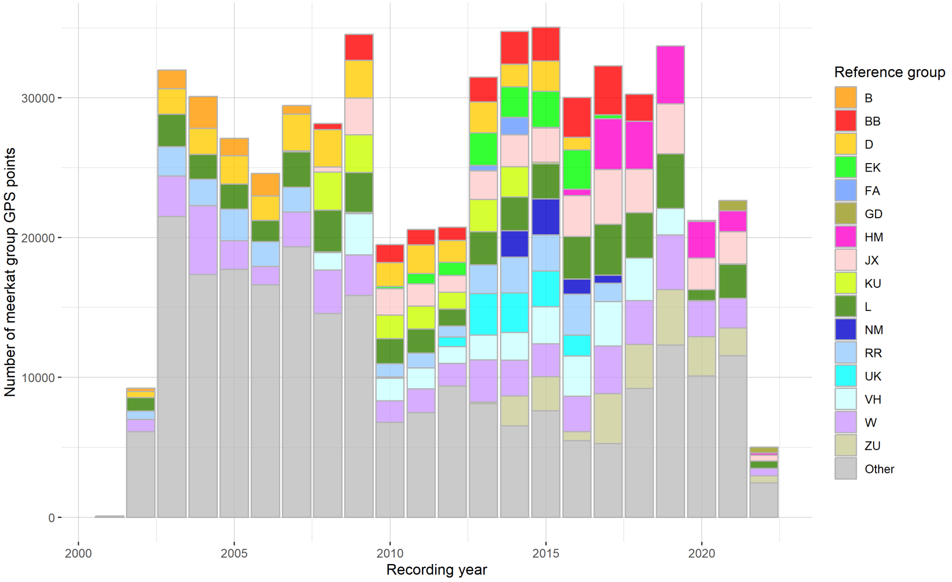


Figure 3.2: Number of recorded GPS points per year. The colors for the groups with CC recordings correspond to those in figure 3.1. Other groups are colored in gray.

Chapter 4

Methodology

The methodology consists of three main parts which are the data preprocessing (chapter 4.1), the preparation and conduction of a RDA to analyze the influence of the individual, and of biological (age and sex), social (group identity, genetic relatedness, dominance), and environmental drivers (distance to the river) (chapter 4.2) as well as the preparation and conduction of a correlation analysis to analyze spatial proximity on CCs (chapter 4.3). Chapter 4.2 is directed towards answering the first RQ and chapter 4.3 is concerned about the second RQ.

To analyze the influence of the individual, and of biological, social, and environmental drivers including group identity, dominance status, sex, age, genetic relatedness, and the distance to the river on their CCs, I employed a RDA. RDA is a multivariate regression method that combines elements of multiple regression and PCA, making it well-suited for analyzing complex ecological data (Borcard et al., 2011). It is particularly effective when dealing with multiple dependent variables, such as the acoustic features of the CCs in this study. Due to the correlation among some acoustic features (see chapter 3.1), I first conducted a PCA on these features to address this problem and reduce dimensionality (section 4.2.1). For additional context on the dependent variables, section 4.2.1.1 provides exploratory statistics on them. The explanatory variables are discussed in section 4.2.2. Finally, the implementation of the RDA is described in section 4.2.3.

To analyze the influence of spatial proximity on CCs, I developed a comparative framework described in chapter 4.3. The goal was to assess the similarity of calls among individuals from different groups that were spatially and temporally proximal. For that, I compared the correlation between spatially and temporally proximal calls and those that were spatially and temporally distant.

4.1 Data Preprocessing

As a first step, I prepared the data by merging various variables from the database, such as dominance events and the beginning and end times of an individual's membership in a certain group, with the corresponding individuals in the CC dataset. The most important step in data preprocessing was determining which years and groups to include in the analysis (section 4.1.1), as the different datasets varied in years and groups they encompassed. I filtered the GPS data (section 4.1.2), since the GPS dataset included data from more years than the CC dataset (see chapter 3.3). All data preprocessing steps were done with R version 4.3.3 (R Core Team, 2024) in RStudio (RStudio Team, 2023).

4.1.1 Observation Years and Groups

Comparing the yearly number of CCs and GPS recordings, it becomes evident that not every group was present every year. Further, not every group present, i.e., recorded by GPS, had CC recordings in the same year. Most processed CCs were recorded in the years 2014 to 2016, which were also the years with the most GPS recordings for the respective groups.

I only included groups and years with a sufficient number of CC and GPS recordings. This was necessary to ensure the analysis was representative of a broad population across different years and to reduce biases from groups or years with only a small amount of data. The group composition should be as diverse as possible with as many different recording dates and as many different individuals recorded per group as possible. For example, one group consisting of only one individual does not represent the group well. I defined the following criteria that needed to be fulfilled by groups and recording years to be part of the analysis:

1. Years need to have a total of at least 1 000 calls which can be from different groups. This ensures a broad range of data within each year and reduces the risk of outliers dominating the data.

2. Groups need to have a total of at least 100 CCs which can be from different years. This minimizes capturing intragroup biases.

3. Groups should have CC recordings in at least two different years, except if they encompass at least ten different individuals with recorded CCs. This ensures diversity within groups.

4. GPS locations of groups with recorded CCs need to be available for at least three months in advance of the respective recording date. This is necessary to analyze possible spatial interactions between groups prior to the call recording.

5. CC recordings should have an appropriate sound quality. This reduces the risk of distortion of the results due to low quality.

The first criterion limited my analysis to the years 2014, 2015, and 2016 as these were the only ones with at least 1 000 processed CCs. Within these three years, 13 groups are present. The second criterion excluded group *Kungfu* (KU), which had only 17 CCs. There were two groups with recordings in only one year, namely *Friksarmy* (FA) and *Van Helsing* (VH). The third criterion excluded group FA, which only had five different individuals recorded as opposed to VH, which had 13. The forth criterion did not lead to further exclusions as all remaining groups adhered to this criterion which resulted in eleven different groups: *Baobab* (BB), *Drie Doring* (D), *Ewoks* (EK), *Jaxx* (JX), *Lazuli* (L), *Nematoads* (NM), *Rascals* (RR), *Uberkatz* (UK), *Van Helsing* (VH), *Whiskers* (W), and *Zulus* (ZU). By filtering the dataset to include only the preprocessed calls from the specified years and groups, I was left with 6 749 of the original 9 167 calls for analysis. The fifth criterion led to a further reduction to 4 958 calls of 99 individuals belonging to the 11 groups mentioned above which were recorded on a total of 91 different dates. The assessment of the sound quality was done by an expert from the IEU. Used for the analysis were, thus, approximately 54% of the initial data. Table 4.1 shows an overview over the number of preprocessed CCs per year and group. Finally, each call was assigned a unique numeric identifier to unambiguously distinguish the calls from one another.

Table 4.1: Number of meerkat CC recordings per year and group for the three observation years chosen.

	BB	**D**	**EK**	**JX**	**L**	**NM**	**RR**	**UK**	**VH**	**W**	**ZU**	**Total**
2014	364	18	31	47	36	69	17	225	0	53	592	1 452
2015	41	87	190	12	36	46	545	786	0	184	237	2 164
2016	69	0	294	51	28	0	394	68	212	76	150	1 342
Total	**474**	**105**	**515**	**110**	**100**	**115**	**956**	**1 079**	**212**	**313**	**979**	**4 958**

4.1.2 Group Way Points

To align the GPS way points with the observation years chosen from the CC dataset, I performed preprocessing steps to filter the GPS way points, as there were more of them than call recordings (see chapter 3.3). Specifically, I filtered the GPS way points to the years 2013 to 2016. I included 2013 because, in a later step, I calculated the interaction probabilities for periods preceding recording date, with the first recording date in January, 2014 (details on the interaction probability in section 4.3.1). The GPS way points were reprojected to a local metric coordinate reference system (UTM zone 34S, EPSG:22234) and checked for outliers outside the study area; none were found. All groups were retained, even those without CC recordings or not belonging to the chosen groups (see section 4.1.1), as other groups present were needed to calculate obstacle interaction probabilities in a later step (see section 4.3.1.2). The filtered GPS way point dataset for the years 2013 to 2016 included 130 896 locations from 32 groups.

4.2 Redundancy Analysis

As a first step, I aimed to analyze the influence of the individual and of biological (age and sex), social (group identity, genetic relatedness, dominance), and an environmental driver (distance to the river) on the variation in the acoustic features of CCs, without including any complex spatial relationships. Given the correlation among the acoustic features (see chapter 3.1), I first conducted a PCA to address this issue and reduce dimensionality (section 4.2.1). Since I had multiple dependent variables, the PCs of the acoustic features of the CCs, I employed a RDA which can handle such cases and is well-suited for analyzing ecological data (Borcard et al., 2011). Alongside the dependent variables, I defined a set of potential explanatory variables for the RDA. These variables, selected based on literature and expert knowledge, are presented in order of their hypothesized importance, as determined by experts:

1. the individual meerkat emitting the call,
2. the group the individual belongs to,
3. the meerkat's dominance status,
4. the meerkat's sex,
5. the meerkat's age category,
6. the coefficient of relatedness r, and
7. the distance to Kuruman's riverbed.

The preparation of the dependent and explanatory variables is described in the following sections. If not stated otherwise, all steps for preparing and conducting the RDA were performed in R version 4.3.3 (R Core Team, 2024) in RStudio (RStudio Team, 2023).

4.2.1 Dependent Variables: Acoustic Features

The preprocessed CC dataset (see section 4.1.1) consisted of 26 acoustic features. Four of these variables – *mindom*, *maxdom*, *startdom*, and *enddom* – were excluded before further processing the acoustic features data (details on these variables in table 3.1). The variables were excluded on the advice of experts as they represent one single point in the acoustic spectrogram. If they were outliers, the parameter space of the respective calls would be greatly influenced. This exclusion resulted in 22 remaining acoustic features, some of which were correlated.

To eliminate correlation among acoustic features, I performed PCA. From the 'stats' package I utilized the `prcomp` to get the PCs (R Core Team, 2024). The data was scaled and centered around 0 during the process of performing PCA. In order to be kept, each PC was required to explain at least 5% of the total variation. This percentage is represented by the dashed red line in figure 4.1. Together, the resulting PCs needed to explain at least 80% of the total variation. For the three years and 11 groups chosen, this resulted in 5 PCs (see figure 4.1) which together explained 82.6% of the total variation.

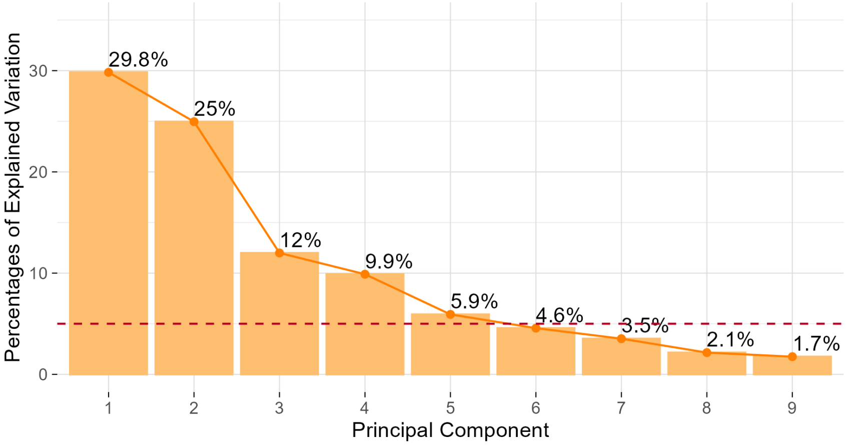


Figure 4.1: Scree plot with the percentage of explained variation for the first 9 PCs of the acoustic features. The red dashed line shows the threshold of 5% explained variation that was set.

4.2.1.1 Exploratory Statistics

Before starting the regression analysis, I checked if the variance of the dependent variables, the five PCs, within different individuals and groups was smaller than the overall variance present in the dataset. Further, I compared the means of the same individuals and groups to see if they are statistically different. If there was no difference, it would not make sense using these in a RDA. I chose five individuals and groups with a large amount of recorded calls for these exploratory analyses (see table 4.2).

I calculated the overall variance and the variance per PC and individual or group to compare them. The variances were calculated using equation 4.1, where X represents an observation, $\bar{x}$ the sample mean, and n the total number of observations in the sample. These calculations were performed using the `var` function from the 'stats' package (R Core Team, 2024).

$$s^2 = \frac{\sum (X - \bar{x})^2}{n - 1} \tag{4.1}$$

Table 4.2: Individuals and groups including their number of calls chosen for the exploratory analysis.

Individual	**#Calls**	**Group**	**#Calls**	**#Individuals**
VZUF014	397	UK	1 079	14
VRRM182	237	ZU	979	16
VUKF018	217	RR	956	12
VEKF017	206	EK	515	6
VRRM181	188	BB	474	19

I used a box plot per PC and individual or group to visually compare the different individuals and groups. In a second step, I used a statistical method to compare the means. To do so, one would commonly use an analysis of variance (ANOVA) (Rogerson, 2020). There are three assumptions underlying the ANOVA. One of them is homoscedasticity – the equality of variances across groups or categories – which is violated by my data. The assumption of normality is further not met by my data which is why I decided to use a Kruskal-Wallis test. A Kruskal-Wallis test is the nonparametric version of the ANOVA, which means that the test does not make any assumption about the distribution of the underlying data (Rogerson, 2020). To perform Kruskal-Wallis tests for the different PCs, I used the `kruskal.test` function from the 'stats' package (R Core Team, 2024) with the following hypotheses:

- H_0: There is no difference between the individuals (groups).
- H_1: At least one individual (group) is different from the others.

In case of significant test results at an α-level of 0.05, I used Dunn's test as a post-hoc analysis since the Kruskal-Wallis test does not make any statement about which individuals or groups differ from others. Dunn's test performs a pairwise comparison between individuals and groups and is an appropriate procedure after a Kruskal-Wallis test (Dinno, 2015). The hypotheses are stated below.

- H_0: There is no difference between the two individuals (groups) compared.
- H_1: There is a difference between the two individuals (groups) compared.

When performing multiple pairwise comparisons, the meaning of α gets redefined. The significance level α is the probability of making a type 1 error, which means falsely rejecting the null hypothesis (Dinno, 2015). For one test, the probability of a type 1 error is 5%. For two tests, however, the probability of not making a type 1 error is 0.95 * 0.95 = 0.9025, i.e., the chance of making a type 1 error increases to almost 10%. This issue can be addressed by applying an adjustment, such as the Bonferroni correction (Dinno, 2015). It modifies the rejection level for each test by dividing α by the total number of tests, thus, requiring a smaller p-value to reject the null hypothesis (Dinno, 2015). The Dunn's test with Bonferroni adjustment was performed using the `dunn_test` function from the 'rstatix' package (Kassambara, 2023).

4.2.2 Explanatory Variables

This section describes the different explanatory variables for the RDA and their procession in the same order as listed in chapter 4.2. The individual and group identity were combined into one section since their procession was the same.

4.2.2.1 Individual and Group Identity

There has been evidence for individual and group signatures in meerkat CCs (Townsend et al., 2010) which is why these two variables were crucial to include in the RDA as possible drivers influencing the acoustic features of CCs.

The individual and the group were included as categorical variables in the CC dataset provided by the IEU. The group, which the meerkat was in when emitting the call, was encoded by the group abbreviations (see section 4.1.1). The individual was identified by a code, for example 'VBBM091'. The first letter, V, was the same for all meerkats included in the data. The second and the third letter (or only the second) stood for the abbreviation of the meerkat's birth group, i.e., the group the meerkat was born in. Due to migration, this group did not necessarily need to be the same as the group the meerkat was in when emitting the call. Following the birth group's code there was the letter for the sex, either F for female or M for male individuals. The last three numbers represented a unique identification of each meerkat.

4.2.2.2 Dominance

Meerkat groups include a dominant pair which monopolizes the breeding within its group (Clutton-Brock & Manser, 2016). Experts and literature suggest to control for dominance as a meerkat's social status might influence its call acoustics (Mausbach et al., 2017; Townsend et al., 2010).

The dominance status of an individual at the recording date of a call was manually defined based on a table with dominance events present in the meerkat database. Manual processing was considered the most efficient, as the events in the table often included a description about the event with additional or more detailed information that cannot be easily extracted automatically (table 4.3). For example, in the second description, it was not only clear that around the end of May, individual VDF173 became dominant, but also that VDF115 disappeared the latest at the end of March, i.e., this individual's dominance surely ended then. Such descriptions were processed fastest in manual work.

Table 4.3: Excerpt of two rows of the dominance event dataset (some columns of the original dataset were excluded).

Comments	**DomDate**	**CallerID**	**GroupRef**	**Sex**
After previous dominant male VBBM080 left group, VBBM091 and VBBM092 competed, on 14/07 M92 seen submitting to M91.	14.07.2015	VBBM091	BB	M
Competed with litter mate F174 since 02/04/2015 after previous dominant VDF115 disappeared (26/03/2015). Eventually VDF174 seen (with large wound from fight) submitting to and receiving sucessful dominance assertions from VDF173.	31.05.2015	VDF173	D	F

The dominance events were filtered to events only including the individuals in question until the end of 2016. For each individual, I defined whether it was dominant at some point and if so, when the dominance period started and ended. Some of the dates might not be accurate, which should not be a big problem as chances were considered low that an individual was recorded on the exact day its dominance started or ended. Unique dates of recorded calls and their respective individuals were extracted from the CC dataset. For each recording date and individual, the individual's dominance status – either dominant (1) or submissive (0) – was defined, resulting in a binary variable.

4.2.2.3 Sex

In accordance to a earlier studies where sex was controlled for (e.g. Hollén and Manser, 2006; Mausbach et al., 2017; Townsend et al., 2010), I included it as an explanatory variable. A meerkat's sex might influence their body size and, therefore, vocal tract morphology (Mausbach et al., 2017).

I extracted a meerkat's sex encoded in its identifier (see section 4.2.2.1) into a separate variable using a regular expression with `str_extract` of the 'stringr' package (Wickham, 2023), resulting in a categorical variable with the two levels F and M.

4.2.2.4 Age Category

Physical similarities such as similar vocal tract morphology could explain part of the variation in CCs (Mausbach et al., 2017; Townsend et al., 2010). Vocal tract morphology may change with meerkats growing up as body size increases which is why age should be included in an analysis (Mausbach et al., 2017). Meerkat ages are typically defined by four categories: Pups are younger than three months, juveniles are between three and six months of age, subadults between six and twelve months, and adults are are older than twelve months (Hollén & Manser, 2006).

I used the same four age categories as a variable specifying meerkat age. I calculated the age at the recording date using the variable indicating the meerkat's date of birth. The age at recording date was then classified into one of the four categories, resulting in a categorical variable with four levels: P for pups, J for juveniles, S for subadults, and A for adults. However, only three categories remained since the data did not include any calls emitted by pups.

4.2.2.5 Genetic Relatedness

Genetic relatedness is hypothesized to be a social driver influencing the acoustic features of meerkat CCs (Townsend et al., 2010). The data regarding genetic relations between the recorded meerkats was rather sparse. Genetic relatedness was calculated based on a meerkat's genealogy using the coefficient of relatedness (r) with the assumption of no inbreeding (Widdig, 2017; Wright, 1922). This coefficient is defined as "the probability that two individuals inherited a certain proportion of genes from a common ancestor" (Widdig, 2017, p. 1). The formula for the coefficient r without inbreeding is defined as

$$r = \sum 0.5^L \tag{4.2}$$

with L representing the number of generation links between two individuals. For full siblings, sharing the same mother and the same father, this results in $r = 0.5$. There are two generation links from one sibling to the other sibling hence $L = 2$. Further, there are two common ancestors and thus two pathways with both $L = 2$ to sum up. This results in $r = 2 * (0.5^2) = 2 * 0.25 = 0.5$ (Widdig, 2017) (see figure 4.2 for a graphic explanation). The coefficient of relatedness r for an individual with itself is 1 as there are 0 generation links and, thus, $r = 0.5^0 = 1$. Due to the sparseness of the genealogy data on the meerkats, only three generations, i.e. the meerkat itself, its parents, and its grandparents were considered (see exemplary relationships in table 4.4).

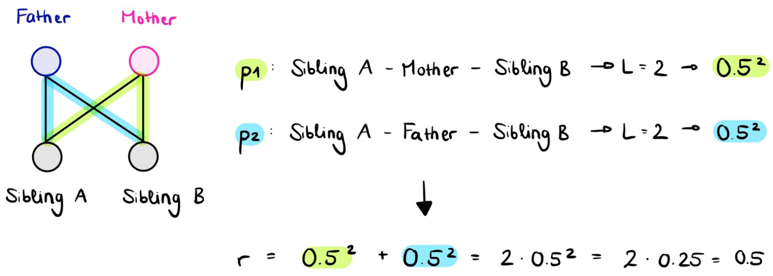


Figure 4.2: Calculation of the coefficient of relatedness for full-siblings.

Table 4.4: Coefficient of relatedness for different relationships between three generations (not complete).

Relationship	**Calculation**	**Coefficient of relatedness**
Full-siblings: both parents shared	$2 * (0.5^2)$	0.5
Half-siblings: only one parent shared, the second is either unknown or different	$1 * (0.5^2)$	0.25
Four grand-parents shared, but no shared parents	$4 * (0.5^4)$	0.25
Three grand-parents shared, but no shared parents	$3 * (0.5^4)$	0.1875
Two grand-parents shared, but no shared parents	$2 * (0.5^4)$	0.125
One grand-parent shared, but no shared parents	$1 * (0.5^4)$	0.0625

Multidimensional Scaling Calculating r for each pair of individuals resulted in a symmetric matrix with 99 rows and columns. To use the information in the RDA, I needed to transform it into columns. For this purpose, I used metric multidimensional scaling (MDS) which is a set of procedures to map distances between objects in a multidimensional space into a space with less dimensions (Živadinović, 2011). If individuals A and B are most similar (or related) in comparison to all other pairs of individuals, MDS positions these individuals such that their distance is smaller than the distance between any other two individuals. A perceptual map is constructed in the chosen number of dimensions (Živadinović, 2011).

An important parameter of MDS is the number of resulting dimensions. A measure of goodness of fit of MDS is the *Stress-1* value, expressed as σ_1, which is a normalized version of the raw *Stress* value σ_r. The smaller the σ_1 value (see equation 4.3), the better the fit (Borg & Groenen, 2005; Kruskal, 1964b).

$$\sigma_1(X) = \sqrt{\frac{\sigma_r(X)}{\sum d_{ij}^2(X)}} = \sqrt{\frac{\sum[f(p_{ij}) - d_{ij}(X)]^2}{\sum d_{ij}^2(X)}} \tag{4.3}$$

$d_{i,j}(X)$ represents the matrix with the original dissimilarity values. $f(p_{i,j})$ is the dissimilarity matrix in the transformed coordinate space (Borg & Groenen, 2005; Kruskal, 1964b). Calculating σ_1 instead of σ_r is advisable as σ_1 is scale-independent. For example, σ_1 is the same if the distances are measured in meters or in kilometers, while σ_r would be 1000-times larger for meters compared to kilometers (Borg & Groenen, 2005).

Since the calculated coefficient of relatedness r was a measure of similarity, the values needed to be transformed to dissimilarity or distance values to use them in MDS. This was done by subtracting r from 1 such that the distance between all individuals with themselves is 0 (equation 4.4).

$$r_{\text{transformed}} = 1 - r \tag{4.4}$$

The MDS was conducted with the transformed values in Python 3.12.3 in PyCharm Community Edition 2024.1.1 (JetBrains, 2024) using the `MDS` function from 'scikit-learn' (Pedregosa et al., 2011). I decided to use 9 dimensions for the genetics as this was the first number of dimensions associated with a σ_1 value below 0.1 (see red dashed line in figure 4.3) which is considered a benchmark for a fair fit (Kruskal, 1964a).

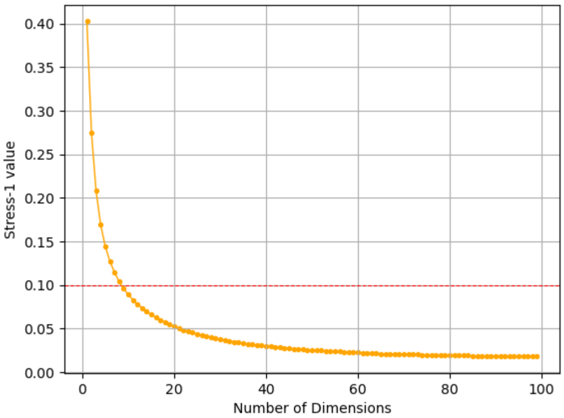


Figure 4.3: *Stress-1* values against the number of dimensions for the matrix of genetic relatedness.

I calculated the pairwise distances in the new coordinate space to compare the distances before and after the transformation. This step ensured that the chosen transformation did not systematically under- or overestimate distances in the newly created space. The histogram of distance changes (figure 4.4) follows a normal distribution which is almost perfectly centered around 0. Although there is a slight tendency for distances to be underestimated in the new coordinate space, I considered the transformation good enough, especially because I wanted to prevent having more dimensions.

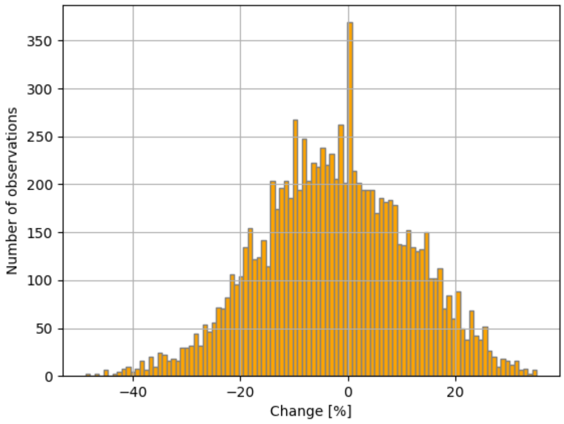


Figure 4.4: Distribution of distance increases and decreases after performing MDS compared to the matrix of genetic relatedness.

4.2.2.6 Distance to Kuruman

The river Kuruman, which is mostly dry, lies within the study area. It is possible that the proximity to the river, representing an environmental driver, influences the acoustic features of a meerkat's CC, for example due to slightly different habitat conditions (Martin et al., 2017).

I calculated the distance to the river using the GPS points in the CC dataset. I extracted the river from OpenStreetMap (OSM) using the QuickOSM plugin in QGIS (QGIS Development Team, 2021) with the keys 'waterway=river' and 'name=Kuruman'. For calculating the distance, I used the `st_distance` function from 'sf' (Pebesma, 2018), resulting in a numerical variable, quantifying the distance in meters.

4.2.3 Conduction of RDA

The input data for the RDA included the dataset with the 4 958 CCs, with the dependent variables being the five PCs of the acoustic features and the possible explanatory variables as listed in chapter 4.2. Explanatory variables can be selected such that the model explains as much of the variation in the response data as possible, e.g., by using a forward selection (Borcard et al., 2011). However, one could select explanatory variables using background knowledge or to gather information, even if the percentage explained by one or more variables is rather low (Heinze et al., 2018; Quebec Centre for Biodiversity Science, 2023b). I used the latter approach to analyze how the individual and biological, social, and environmental drivers influence the variation in the acoustic features of CCs.

I performed the RDA with each possible explanatory variable separately to get a first overview of the influence of each driver. For that, I used the `rda` function of the 'vegan' package (Oksanen et al., 2022). I calculated the adjusted R^2 value, which is the unbiased version of R^2, to assess the explanatory power. The unadjusted R^2 is imprecise because of two factors. Any variable included in the matrix of explanatory variables increases the R^2, be it related to the response matrix or not. Further, due to random correlations, adding explanatory variables increases the amount of explained variation (Borcard et al., 2011). Ezekiel's formula for the adjusted R^2 provides a remedy for the problem as it accounts for the number of objects n and the number of explanatory variables m (see equation 4.5) (Borcard et al., 2011).

$$R^2_{adj} = 1 - \frac{n-1}{n-m-1}(1 - R^2) \tag{4.5}$$

In a second step, I combined different explanatory variables with each other. I observed that combining any drivers (variables 2-7) with the individual (variable 1) did not substantially increase the total variation explained by the model. Its total variation explained was always highly dependent on the variation explained by the individual emitting the call, overshadowing the other variables. I considered a standard RDA not suitable to explore the different biological, social, and environmental drivers influencing the acoustic features of CCs.

A partial RDA (see section 2.2.3.2) and variation partitioning seemed to be more suitable. These methods allow to unveil unique and joint effects of different explanatory variables. Variation partitioning (see example in figure 4.5) combines RDA and partial RDA to divide the variation in the set of dependent variables among two to four sets of explanatory variables X (Quebec Centre for Biodiversity Science, 2023d). Variable X_1 explains the portions a and c (blue and purple) of the total variation in the set of dependent variables

Y. Variable X_2 explains the portions b and c (red and purple). When removing any effects of variable X_2 from the response of X_1, part a (blue) remains. When removing any effects of X_1 from the response of X_2, part b (red) remains. X_1 and X_2 jointly explain portion c (purple) of the total variation. These joint effects cannot be disentangled and clearly assigned to one set of variables. Variables X_1 and X_2 together explain the portions a, b, and c. The remaining variation not explained by either variable is portion d (Borcard et al., 2011; Quebec Centre for Biodiversity Science, 2023d).

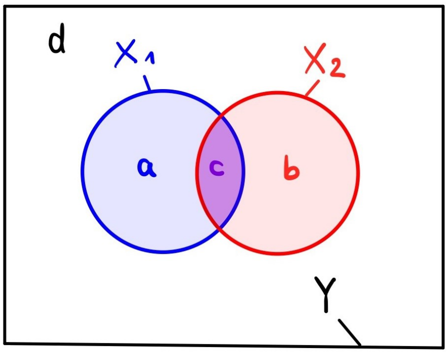


Figure 4.5: Different parts of explained variation that can be uncovered using variation partitioning (Quebec Centre for Biodiversity Science, 2023d).

The starting point for the partial RDA was the outcome of the RDA with a single explanatory variable (for results see table 5.5). Any biological, social, or environmental driver (variables 2-7) with an explanatory power of less than 2% in the standard RDA was excluded from this analysis since it had extremely limited influence on the variation in acoustic features.

4.2.3.1 Interindividual Variation

I removed the variation not explained by the individual (i.e., the intraindividual variation between calls from the same individual) to focus on the influence of the remaining drivers on interindividual variation – the variation that can be explained by the individual emitting the call. The remaining variation was assumed to be either generally inexplicable or due to my lack of other necessary biological variables.

I extracted the variation explained by the individual emitting using the `fitted` function from the 'stats' package (R Core Team, 2024) with the individual as explanatory variable. This resulted in a new set PCs, which I called $\hat{PC}$ with the five components $\hat{PC}_1$, $\hat{PC}_2$, and so on. I performed the RDA for each driver (variables 2-7) separately to see how much of the interindividual variation could be explained by each of these variables. I further performed variation partitioning using the driver with the highest explanatory power, group identity, as explanatory variable. I compared the portion of explained variation between the group identity and every other biological, social, and environmental driver.

4.3 Correlation Analysis: Influence of Spatial Proximity

The second key objective of this thesis was to assess to what extent spatially and temporally close meerkats influence each other's CCs. For that, I compared the correlation between CCs in spatial and temporal proximity, the informed case, versus CCs in spatial and temporal distance, the uninformed case. The uninformed case was included to evaluate whether the results of the informed correlation were greater than a random baseline.

If the informed correlation indicated similarity between spatially and temporally proximal calls, but the uninformed correlation showed the same results, it would not be a meaningful indicator for similarity due to spatial proximity.

I used a single predictor CC in spatial and temporal proximity (or distance) to another CC (the 'dependent call') to correlate it with. I repeated the correlation multiple times for both, informed and uninformed cases, and compared the distributions of correlation coefficients. The informed predictor call was sampled from a set of potential predictor calls in close temporal proximity to the dependent call. The sampling probabilities were drawn from the spatial interaction probability between the dependent call and the predictor call. For the uninformed predictor, calls were sampled from a set of potential predictor calls temporally far apart from the dependent call with the complement of the spatial interaction probability used as sampling probabilities. The distribution of correlation coefficients for the informed and uninformed cases were compared using the Kullback-Leibler (KL) divergence.

The following sections present the preparation of the necessary variables for the correlation analysis. Section 4.3.1 dives into the calculation of the informed spatial interaction probability, while section 4.3.2 focuses on the uninformed interaction probability. Section 4.3.3 describes the comparison of informed and uninformed correlation coefficients. If not stated otherwise, all steps for the calculation of the informed and uninformed predictors as well as the correlation analysis were performed using Python 3.12.3 in PyCharm Community Edition 2024.1.1 (JetBrains, 2024).

4.3.1 Informed Spatial Interaction Probability

The crucial part in defining spatially proximal calls was the definition of a spatial interaction probability between the individuals emitting CCs, which served as the sampling probabilities. I defined spatial interaction probability (hereafter shortened to 'interaction probability') as the likelihood that two meerkats were in close spatial proximity during the 90 days preceding a call recording. The 90-day period was chosen based on a prior study analyzing territory shifts (Kranstauber et al., 2012), and because a single day of high interaction probability between meerkats was deemed insufficient to possibly influence acoustic signatures. The interaction probability reflects the spatial conditions necessary for potential interaction, which is crucial for this thesis, as it determines whether an exchange of acoustic features between groups could have occurred. Since the GPS way points with accurate time stamps were only available per group (see chapter 3.3), the interaction probability was not calculated per pair of calls or individuals, but rather per pair of groups. Each call could be assigned to a group based on the individual emitting the call and its group identity. This meant that all calls emitted in the same group on one date had the same interaction probability with calls from other groups, even if they were emitted by different individuals. The interaction probability could take values between 0 and 1. 0 meant that interaction between this pair of groups was impossible, while 1 meant that two groups showed no spatial restriction for interaction. The values in between represented the likelihood of interaction and exchange of acoustic features between the two groups.

As a basis for the interaction probability, I used two different concepts of spatial proximity. These included a geometry-inspired measure of Euclidean distances and a topology-inspired measure of obstacles between two groups. The Euclidean part of the interaction probability, termed 'Euclidean interaction probability', ensured that two meerkat groups were close enough to reach each other within a reasonable distance, to hear the other's calls, and to potentially adjust their own. The obstacle aspect, called 'obstacle interaction probability' ensured that two groups were immediate neighbors with no third group in between them which would effectively prevent any interaction due to meerkats' territorial behavior (details in section 2.1.1).

I assessed the interaction probability for all groups with processed call recordings during the three observed years. Groups that were solely recorded by GPS without any call recordings were only included as possible obstacles. Their interaction probability with other groups did not matter, however, they could act as obstacles for interaction between two other groups.

Details of the approach to calculating Euclidean distances and the resulting Euclidean interaction probabilities are provided in section 4.3.1.1. Similarly, section 4.3.1.2 outlines the methodology for defining obstacles and computing the obstacle interaction probabilities. To clarify these processes, I introduced a simplified example, which was revisited throughout the explanation of interaction probability.

4.3.1.1 Euclidean Distances

Euclidean distances were included to the calculation of the interaction probability to ensure that two groups were close enough to reach each other and to hear the other's calls. I used multiple timestamps to calculate distances within a 90-day period to reduce the influence of possible outliers.

The simplest way to calculate distances between two groups would be to take two known GPS points, recorded at the same timestamp, and calculate the Euclidean distance between them. However, with the data available, there are multiple problems surfacing, making it impossible to use such an approach. Firstly, although the GPS recordings generally took place during the same three hours in the morning and afternoon, the timestamps of the recording could vary for different groups and distances would be calculated for two different points in time. Secondly, not every group was recorded on every day, making it sometimes impossible to compare two groups on the same day. Lastly, the GPS recordings represent the manually recorded center points of groups, meaning that there is inaccuracy in the GPS data and that group members may spread out further. While this problem could not be eliminated completely – GPS recordings for every single meerkat would be needed – it could be alleviated to parts. Concluding, my method for calculating distances between meerkat groups needed to be capable of continuously estimating locations at times for which an exact GPS recording was missing. The estimation should be based on the recordings at hand and be dependent on the speed of movement of the meerkats.

Brownian bridges (see section 2.2.3.3) are suitable for this type of problem (Horne et al., 2007). They estimate probable locations following a Gaussian distribution for any time t within the interval T, given a start point a, an end b point, and a diffusion coefficient σ_m^2 (see equation 4.6) (Horne et al., 2007):

$$Z(t) \sim \mathcal{N}(\mu(t), \sigma^2(t)I), \quad \text{where} \tag{4.6}$$

$$\mu(t) = a + \frac{t}{T}(b - a), \tag{4.7}$$

$$\sigma^2(t) = \frac{t(T - t)}{T}\sigma_m^2, \quad \text{and} \tag{4.8}$$

$$I = 2 \text{ x } 2 \text{ identity matrix.} \tag{4.9}$$

The diffusion coefficient σ_m^2 greatly influences the extent of variability of the bridges and should be chosen carefully (Krumm, 2021). With an increasing diffusion coefficient, the deviation from the straight-line connection between the two points a and b increases (Horne et al., 2007; Krumm, 2021). In a first step, I defined the diffusion coefficient as 0.15 square meters per second time difference between two consecutive locations. This value was estimated using the help of expert knowledge and the mean speed calculated from the filtered GPS way point data set. Studies reported a movement of several kilometers per day (Clutton-Brock & Manser, 2016) and expert knowledge suggested that meerkats move around 2 kilometers in three hours which equals a speed of 0.185 m/s. Movement in this context did not mean that they covered a straight-line distance of 2 kilometers, but rather them moving around in circles while foraging. The mean speed calculated using the filtered group GPS way points for the years up to 2016 was 0.108 m/s. Calculating the mean between these two led to a value of 0.15 m/s, and therefore to 0.15 m^2/s as a value for the diffusion coefficient.

The meerkats' GPS trajectories were not regularly sampled. In two sessions lasting three hours, locations were generally sampled every 15 minutes. However, there were also instances where the sampling interval was longer or shorter. Between the two daily sessions were a couple of hours for which there was no information on the meerkats' whereabouts. This irregular sampling meant that the diffusion coefficient needed to be adjusted according to the time elapsed between two consecutive sampling points. The diffusion coefficient σ_m^2 per second time difference between two consecutive points was 0.15 which led to the following calculation for any other time difference Δ_t in seconds between two consecutive GPS recordings:

$$\sigma_m^2 = 0.15 * \Delta_t \tag{4.10}$$

During the distance calculation, it became evident that the chosen way of determining the diffusion coefficient was not suitable as it resulted in immense distances of more than 100 kilometers. Such distance values were unrealistic since the study site only encompassed approximately 50 square kilometers (KRC, n.d.). With the approximation of the study area being a square, the maximum distance possible was the diagonal, around 10 kilometers. As an initial measure, I limited the diffusion coefficient to 135, corresponding to the standard 15-minute sampling interval. This adjustment reduced some of the excessively large distances, although a few remained. I scaled the diffusion coefficient further down to reflect the meerkats' tendency to move circularly while foraging. I tested different values between 1.5 and 2, where 2 showed the best results. This value was used as a denominator for the diffusion coefficient (equation 4.11).

$$\sigma_m^2 = \begin{cases} \frac{0.15*\Delta_t}{2}, & \text{if } \Delta_t \leq 900 \\ \frac{0.15*900}{2} = 67.5, & \text{if } \Delta_t > 900 \end{cases} \tag{4.11}$$

With the prerequisites of the Brownian bridges being set, I sampled timestamps and calculated the distances between all pairs of groups for these timestamps. I decided to sample 20 random timestamps per 90-day period (red lines on the left side in figure 4.6). To sample more timestamps at the hours where GPS recordings were done, I included a condition that favored these hours; 5, 6, 7, and 8 for the morning and 14, 15, 16, and 17 for the afternoon session. The favored hours were three times more likely to get sampled compared to other hours of the day.

For each timestamp, I sampled points according to the Brownian bridge probability density at this timestamp using the `multivariate_normal` function from 'SciPy' (Virtanen et al., 2020) with mean and covariance as described in equations 4.7 and 4.8. Points in areas with a higher probability density were more likely to get sampled than points in areas with low probability density. For each pair of groups and timestamp, I sampled one point per group and calculated the distance between the points (see right side in figure 4.6). I repeated the sampling 50 times, leading to a total of 1 000 distance samples per pair of groups and 90-day period. The samples were numbered to be able to unambiguously identify each observation (column 'UID' in table 4.5). Besides the distances, the coordinates used as samples for each group were saved. These were needed for defining the obstacle interaction probability (see section 4.3.1.2).

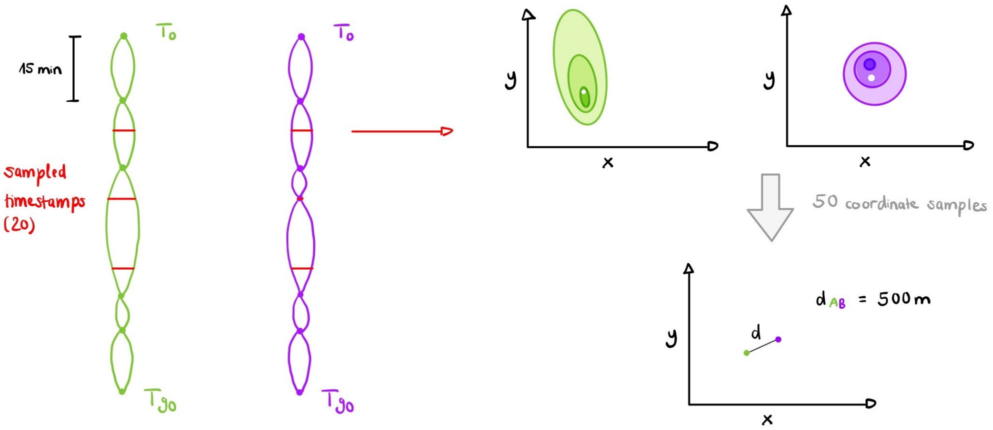


Figure 4.6: Coordinate sampling and distance calculation. The left side shows two trajectories of groups A (green) and B (purple) over time, increasing from top to bottom. The dots represent known GPS locations for known timestamps, the lines connecting the dots symbolize the location uncertainty between recordings. The red lines display sampled timestamps. The top right graph zooms into the probability distribution for these two groups at the first timestamp. The white dots represent two sampled coordinates. On the bottom right graph, the Euclidean distance between the sampled coordinates is calculated.

Euclidean Interaction Probability For each observation, a single distance sample between one pair of groups in a 90-day period, I used a sigmoid function to define a score between 0 and 1, expressing the interaction probability based on the Euclidean distance. The sigmoid function (equation 4.12) is suitable to represent similarity based on distance because it converges to 0 and to 1, respectively. Observations with value 1 were very close to each other, indicating favorable conditions for an exchange of acoustic features, while observations assigned a 0 were far apart, indicating unfavorable conditions for an exchange.

$$S(x) = \frac{1}{1 + e^{-k(x - x_0)}} \tag{4.12}$$

The parameter k defines the steepness of the curve. A positive k leads to a monotonically increasing function, while a negative k leads to a monotonically decreasing function. The parameter x_0 defines the x-value of the midpoint where $S(x) = 0.5$. I needed a negative steepness, since 1 was supposed to be the value for the shortest distance (smallest x). I first specified where the function should start to converge to 1 and 0, respectively. I chose these values to be 300 and 2 500 meters. The threshold for convergence to 1 was chosen at a distance where I considered it to not make a difference if two groups were even closer. The threshold for convergence to 0 was chosen using expert knowledge, which suggested that meerkats move between 500 and 2 000 meters per day. The function should converge to 0 a little over the maximum distance moved per day such that the maximum distance covered still has an interaction probability value above 0. The parameter x_0 was defined by calculating the midpoint between 300 and 2 500. For defining the steepness, I tried different values to find the most fitting. I aimed for a rather small steepness, as I did

Table 4.5: Example dataset of the Euclidean interaction probability (column 'EIP'), calculated using the sigmoid function (equation 4.13), with column 'UID', representing an observation's unique identifier. The first line corresponds to example groups A and B in figure 4.6.

UID	Date	Group 1	Group 2	Distance [m]	EIP
1	2001-01-01	A	B	500	0.99
2	2001-01-01	A	B	700	0.97
3	2001-01-01	A	C	1250	0.68
4	2001-01-01	A	C	1500	0.38
5	2001-01-01	A	D	1100	0.82
6	2001-01-01	A	D	900	0.92
7	2001-01-01	B	C	2600	0.002
8	2001-01-01	B	C	2300	0.01
9	2001-01-01	B	D	1700	0.18
10	2001-01-01	B	D	1900	0.08
11	2001-01-01	C	D	1450	0.44
12	2001-01-01	C	D	1000	0.85

not want a small change in distance to lead to a big change in the resulting Euclidean interaction probability value. I decided to use $k = -0.005$, which yielded a moderate slope and rather slow convergence to 0 and 1 (see figure 4.7 and equation 4.13). The described procedure led to having a value between 0 and 1 for each distance observation (see example in table 4.5).

$$EIP(x) = \frac{1}{1 + e^{-(-0.005)(x-1400)}} \tag{4.13}$$

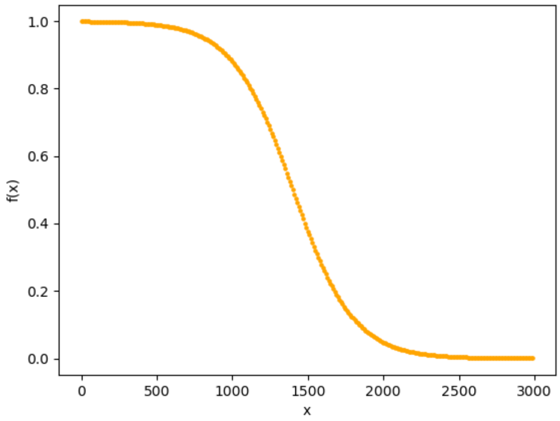


Figure 4.7: Sigmoid function (equation 4.13) for the Euclidean interaction probability, evaluated at distances between 0 and 3 000 meters.

4.3.1.2 Obstacles for Interaction

Meerkats are highly territorial animals (see section 2.1.1), meaning that they do not pass through another group's territory without the intention of aggressive encounter. This effectively prevents acoustic features from being exchanged between groups that are not direct neighbors. While the Euclidean interaction probability gives an idea about the distance between two groups, it does not make any statement about the presence of an obstacle – a third group – between them. Thus, a measure of obstacles between two groups is needed to represent territoriality.

I used the coordinates saved during the Brownian bridge sampling process to connect the same pairs for which distances were calculated, creating a straight line between them using `LineString` from 'shapely' (Gillies et al., 2024). Each line was checked for intersection with any of the other group's territories.

Central to this approach is the definition of territories for the meerkat groups in each of the 90-day periods. The filtered GPS recordings of group centers (see section 4.1.2) needed to be transformed to possible territory polygons. There is an abundance of possibilities how to go about this task, including minimum convex polygon (MCP) and different forms of the kernel density estimation which are turned into UDs and subsequently home ranges (Calenge, 2023). In the following, I introduce the complete habitat area scenario I decided on in discussion with a meerkat expert. According to the expert, the meerkats living in my study area adhere well to territory restrictions and generally avoid foreign territories in their completeness. That led me to choose the most restrictive scenario. Two other scenarios that were presented to the expert as possible territory scenarios are presented in the discussion (chapter 6.2).

Complete Habitat Area Scenario The complete habitat area scenario builds a MCP around all recorded GPS points for each group and 90-day period. Around the MCPs, a buffer is added (figure 4.8). This scenario is the most restrictive as the complete habitat is avoided by a foreign group.

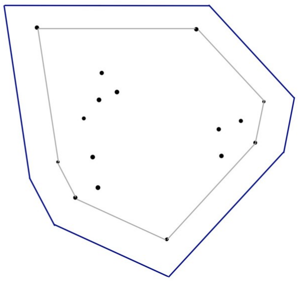


Figure 4.8: Sketch of the complete habitat area scenario. The black dots represent the GPS recordings, the gray polygon the MCP around them, and the blue polygon the outer edges of the defended territory.

This scenario assumes that meerkats leave scent marks everywhere they go to, making all GPS points relevant to the territory definition. The 10-meter buffer around the MCP adds two assumptions. The GPS points which represent the group centers are not the outermost locations of all group's members and meerkats may sometimes move beyond these points, leading to scent marks extending beyond the scope of the MCP of the existing recordings. Further, meerkats need to be able to get from one point to another, meaning they leave scent marks on the paths connecting the recorded GPS points. It is, however, unknown what route they take in between two consecutive relocations and, while they could move along the shortest paths between two consecutive GPS points, they could also roam further and leave scent marks in areas apart from the shortest paths. Due to heavy scent marking in the complete habitat of one group, meerkats from another group notice these and spaciously avoid the marked territory. This scenario is realistic as it is known that meerkats use scent marks as nonviolent cues indicating territory ownership and that foreign scent marks are commonly avoided (Bateman et al., 2015; Jordan et al., 2007).

Obstacle Interaction Probability Every observation – one single line connection between two groups which was checked for an intersection with territories of all other groups at one point in time – was assigned a value of 1 or 0. 1 meant that there was no obstruction with any other group's territory, not restricting interaction between the groups. 0 was assigned if there was obstruction, meaning that the connection crossed through a foreign territory which essentially prevented any possible interaction between the two groups. 0 was assigned independently of the number of foreign groups' territories blocking the connection between two groups (see figure 4.9 and corresponding values in table 4.6).

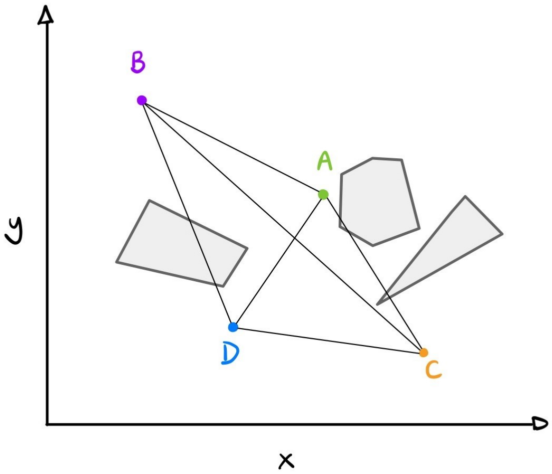


Figure 4.9: Example sketch defining the obstacle interaction probability. Letters A-D refer to one pair of sampled coordinates at one timestamp for each of the groups A-D. The gray polygons represent other groups' territories.

Table 4.6: Example dataset for the obstacle interaction probability (column 'OIP') for groups shown in figure 4.9.

UID	Date	Group 1	Group 2	OIP
1	2001-01-01	A	B	1
3	2001-01-01	A	C	0
5	2001-01-01	A	D	1
7	2001-01-01	B	C	1
9	2001-01-01	B	D	0
11	2001-01-01	C	D	1

4.3.1.3 Interaction Probability Matrices

The two aspects of spatial proximity needed to be combined to get the final interaction probability as matrices. After matching unique observations x_i of the Euclidean and obstacle interaction probability with each other based on the unique identifiers, the two values were multiplied (see equation 4.14 and table 4.7).

$$IP(x_i) = EIP(x_i) * OIP(x_i) \tag{4.14}$$

The mean for each date and pair of groups was calculated from the individual observations to create one interaction probability matrix for each date. Missing group combinations were added with an interaction probability of 0, and groups with itself were assigned an interaction probability of 1 (see table 4.8a). This resulted in one interaction probability

Table 4.7: Example dataset with the Euclidean, the obstacle, and the resulting interaction probability (column 'IP'). For all uneven unique identifiers, the obstacle interaction probability corresponds to table 4.6.

UID	**Date**	**Group 1**	**Group 2**	**Distance [m]**	**EIP**	**OIP**	**IP**
1	2001-01-01	A	B	500	0.99	1	0.99
2	2001-01-01	A	B	700	0.97	1	0.97
3	2001-01-01	A	C	1250	0.68	0	0
4	2001-01-01	A	C	1500	0.38	1	0.38
5	2001-01-01	A	D	1100	0.82	1	0.82
6	2001-01-01	A	D	900	0.92	0	0
7	2001-01-01	B	C	2600	0.002	1	0.002
8	2001-01-01	B	C	2300	0.01	0	0
9	2001-01-01	B	D	1700	0.18	0	0
10	2001-01-01	B	D	1900	0.08	0	0
11	2001-01-01	C	D	1450	0.44	1	0.44
12	2001-01-01	C	D	1000	0.85	1	0.85

matrix for each recording date. Each matrix included the same 11 groups, with values comparable to one another. For example, an interaction probability value of 0.5 meant the same on the first and last date.

I defined a threshold to set the smallest interaction probability values, which served as the sampling probabilities (see section 4.3.3), to 0 to prevent using calls with very small interaction probabilities as predictors. I set the threshold at 0.01, since the interaction probability values were rather small due to the restrictive habitat scenario chosen, All values less than or equal 0.01 were set to 0 (see table 4.8b).

Table 4.8: Unadjusted and final interaction probability matrices for January 1, 2001, using the example data displayed in table 4.7.

(a) Unadjusted interaction probability matrix.

	A	**B**	**C**	**D**
A	1.0	0.98	0.19	0.41
B	0.98	1.0	0.001	0.0
C	0.19	0.001	1.0	0.645
D	0.41	0.0	0.645	1.0

(b) Final interaction probability matrix.

	A	**B**	**C**	**D**
A	1.0	0.98	0.19	0.41
B	0.98	1.0	0.0	0.0
C	0.19	0.0	1.0	0.645
D	0.41	0.0	0.645	1.0

4.3.2 Uninformed Interaction Probability

I defined spatially and temporally uninformed predictor calls to compare the informed with uninformed correlation coefficients. I converted the interaction probability values to an uninformed version to not only use different calls as predictors, but also modify the sampling probabilities in the uninformed analysis.

I created matrices which could be described as the complement of the informed interaction probability matrices to get an uninformed interaction probability for each recording date. The values which were excluded from the informed interaction probability (≤ 0.01) were used to calculated the uninformed interaction probability values. I calculated them by subtracting the informed interaction probability values from 1. All interaction probability values greater than 0.01, as well as all values for groups with themselves, were set to 0 (see equation 4.15 and table 4.9).

$$UIP_{i,j} = \begin{cases} 0, & \text{if } i = j \\ 0, & \text{if } IP_{i,j} > 0.01 \\ 1 - IP_{i.j}, & \text{if } IP_{i,j} \leq 0.01 \text{ and } i \neq j \end{cases} \tag{4.15}$$

Table 4.9: Unadjusted informed interaction probability matrix transformed to an uninformed interaction probability matrix for four groups (unrelated to previous examples).

(a) Informed interaction probability matrix.

	A	**B**	**C**	**D**
A	1	0.57	0.0024	0.35
B	0.57	1	0.0083	0.0097
C	0.0024	0.0083	1	0.72
D	0.35	0.0097	0.72	1

(b) Uninformed interaction probability matrix.

	A	**B**	**C**	**D**
A	0	0	0.9976	0
B	0	0	0.9917	0.9903
C	0.9976	0.9917	0	0
D	0	0.9903	0	0

4.3.3 Conduction of Correlation Analysis

The correlation analysis involved comparing the correlation between a dependent and a predictor call in spatial and temporal proximity versus in distance. I repeated the calculation of the correlation for informed and uninformed cases multiple times to prevent biases from individual sampling and compared the distributions of informed and uninformed correlation coefficients. Using the set of PCs representing interindividual variation ($\hat{PC}$) as dependent calls, I calculated the correlation coefficients. Consequently, before sampling the predictor calls, the dependent CC data needed to be cleaned from duplicate entries on one date, which arose by recording multiple CCs emitted by the same individual on one day.

4.3.3.1 Informed Predictor Calls

All CCs 90 days in advance of the dependent call were defined as potential predictor calls (see figure 4.10). Each of these calls was assigned a sampling probability which corresponded to their interaction probability with neighboring groups (i.e., one column of table 4.8b). The interaction probability for groups with themselves was set to 0, such that the predictor call was never of the same group as the dependent to reduce the influence of group identity. The sampling probabilities for potential predictor calls were normalized to sum to 1 (see example in table 4.10). A sampling with these probabilities was applied for each dependent call. Dependent calls were removed if they had no potential predictor calls, i.e., on the first recording date in January, 2014, or if no calls in spatial proximity were present.

Figure 4.10: Set of potential predictor calls for a call of group A (green with highlighted background).

Table 4.10: Example set of potential predictor calls for a dependent call of group A on January 1, 2001 as shown in figure 4.10, including their interaction probability with group A (table 4.8b) and sampling probability.

Group$_{\text{pred}}$	**Date**	**IP$_{\text{A}}$**	**Sampling probability**
A	2000-10-03	0.0	0.0
A	2000-11-15	0.0	0.0
B	2000-11-17	0.98	0.238
B	2000-12-01	0.98	0.238
B	2000-12-14	0.98	0.238
C	2000-10-30	0.19	0.046
C	2000-11-21	0.19	0.046
C	2000-11-26	0.19	0.046
C	2000-12-30	0.19	0.046
D	2000-12-02	0.41	0.1

4.3.3.2 Uninformed Predictor Calls

The workflow for the uninformed predictor calls was similar to the informed ones, except the time frame of potential predictor calls and the usage of the uninformed interaction probability as sampling probabilities. I used potential predictor calls that were recorded in temporal distance to the dependent call (see figure 4.11) that the uninformed predictor calls are as different from the informed as possible. For calls recorded in 2014, I used a minimum time difference of one year and three months (455 days) after the recording date of the dependent call. For example, for a call recorded on January 3, 2014, the first possible calls to use as predictors would be the ones recorded on April 3, 2015. For calls recorded in 2016, I applied the same principle in reverse, setting the last possible sampling date to 455 days before the recording date of the dependent call. I modified the approach for calls recorded in 2015, since my data was limited between 2014 and 2016 and there were no calls 455 days before or after a recording date in 2015. I employed a sampling strategy that considered calls before and after the dependent call in 2015. Possible sampling dates began at least nine months (274 days) after the recording date of the dependent call or ended at least nine months before it. For instance, the sampling period for a call recorded on November 2, 2015 would include all recording dates up to February 1, 2015 and the ones starting from August 2, 2016.

4.3.3.3 Correlation Comparison

I used Pearson's correlation coefficient, r, as a measure of strength of linear association between the predictor and the dependent call (Rogerson, 2020). The correlation coefficient is quantified between two variables x and y with standard deviations s_x and s_y (equation 4.16).

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{(n-1)s_x s_y} \tag{4.16}$$

The resulting coefficient takes values between -1 and 1. A value of 1 indicates that the points precisely lie along a line with a positive slope, while -1 means the same for a line with a negative slope (Rogerson, 2020). This correlation coefficient is a simple measure to get information on the strength of association between two variables, however, it is important to note that strong correlation does not imply causation (Rogerson, 2020).

The calculation of the correlation coefficient using the `pearsonr` function from the 'SciPy'

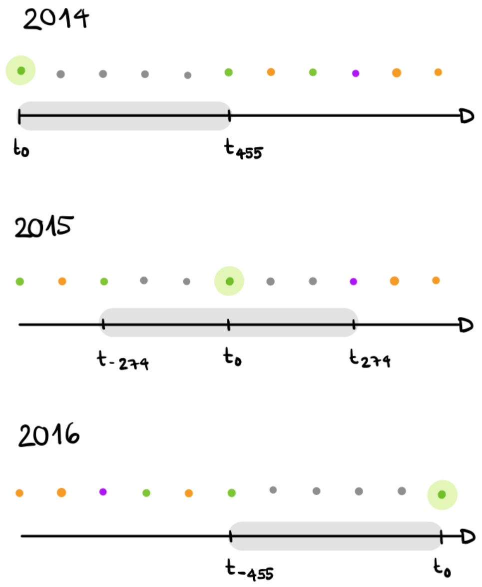


Figure 4.11: Potential uninformed predictor calls for all three years. The green highlighted dot at t_0 represents the dependent call, the gray dots represent calls not eligible as uninformed predictors, while the colored dots are eligible calls. The highlighted part on the timeline represents the time span with calls that cannot be used as predictors for the respective dependent call.

library (Virtanen et al., 2020) was repeated 1 000 times for both cases and each PC. I plotted the distributions of informed and uninformed correlation coefficients in R version 4.3.3 (R Core Team, 2024) in RStudio (RStudio Team, 2023) using the `geom_density` function from the 'ggplot2' package (Wickham, 2016) to make a statement about similarities between calls in spatial proximity. I calculated the KL divergence (Cover & Thomas, 2006), using the `KL` function from the 'philentropy' package (Drost, 2018), for numerical comparisons of the distributions. The KL divergence, also called relative entropy, is an asymmetrical measure of difference between two distributions (Cover & Thomas, 2006). It essentially quantifies how much additional information is needed when one assumes that the data follows a distribution q instead of p (equation 4.17) (Cover & Thomas, 2006).

$$D(p\|q) = \sum_{x \in X} p(x) \log \frac{p(x)}{q(x)} \tag{4.17}$$

Chapter 5

Results

This chapter presents the results of the analyses described in the previous chapter. The first chapter (5.1) focuses on the RDA employed to answer the first RQ. The second chapter (5.2) presents the results of the correlation analysis, aiming to answer the second RQ.

5.1 Redundancy Analysis

A RDA was employed to analyze the influence of the individual, and of biological (age and sex), social (group identity, genetic relatedness, dominance), and environmental drivers (distance to the river) on the acoustic features of CCs, with the aim to answer the first RQ. The following sections provide results on PCA and exploratory statistics applied to the dependent variables (section 5.1.1). Further, the results are presented for the different RDAs performed (sections 5.1.2 and 5.1.3).

5.1.1 Dependent Variables: Acoustic Features

To reduce dimensionality of the set of dependent variables and eliminate correlation among them, I employed PCA, which resulted in a total of five components that each explain at least 5% of the total variation. Together, they explain 82.6% of the total variation in the 22 acoustic features. The first two axes explain 29.84% and 24.96% of the total variation. The loadings of the PCs (see table 5.1) are presented in the following.

PC1 includes high loadings of frequency-related features (*meanfreq*, *freq.IQR*, *sd*), as well as of tonality-related features (*sp.entropy*, *sfm*). Tonality refers to the distribution of energy throughout the spectrum, which is a diagram with frequency on the x-axis and energy on the y-axis (Boersma, 2014). It is a spectrogram analyzed at one specific point. A spectrogram (see figure 2.3) displays frequency versus time (Boersma, 2014). High spectral entropy (*sp.entropy*) means that there is a lot of sound in a tone, such as a dog growling, while low spectral entropy indicates a harmonic, tonal, sound (Araya-Salas & Smith-Vidaurre, 2017), like the voice of an opera singer. PC2 includes high loadings of variables containing information about time, such as *duration*, *time.ent*, or *time.median.* PC3 and PC4 are similar and include information about two acoustic characteristics. First, there is information about the symmetries of the frequency spectrum, whether the frequency distribution is peaked or flat (kurtosis, *kurt*), and, whether the curve is symmetrical or skewed to one side (skewness, *skew*). Second, there is information about the frequency with the highest energy in the frequency spectrum, the dominant frequency (*meandom*, *meanpeakf*). PC5 includes very high loadings of modulation index and dominant frequency range in the spectrogram (*modindx*, *dfrange*). The modulation index provides information,

Table 5.1: Loadings of the acoustic features (see table 3.1) for the PCs. The five highest absolute values for each PC are marked in orange.

Acoustic features	**PC1**	**PC2**	**PC3**	**PC4**	**PC5**
duration	-0.2196	-0.3353	-0.0785	-0.0100	0.0347
meanfreq	-0.3271	0.2113	0.0234	0.0822	-0.0805
sd	-0.3005	0.1117	0.2484	-0.0756	0.0222
freq.median	-0.1529	0.2322	-0.3604	0.0400	-0.0626
freq.Q25	-0.0651	0.1108	-0.2981	0.4472	-0.1351
freq.Q75	-0.2904	0.1915	0.0265	0.1242	-0.1818
freq.IQR	-0.2913	0.1734	0.1098	0.0098	-0.1565
time.median	-0.2186	-0.3313	-0.0840	0.0041	0.0514
time.Q25	-0.1885	-0.3086	-0.1102	0.0090	0.0514
time.Q75	-0.2286	-0.3350	-0.0702	-0.0068	0.0463
time.IQR	-0.2228	-0.3049	-0.0335	-0.0164	0.0342
skew	-0.0296	-0.1105	0.3625	0.4992	-0.1206
kurt	-0.0405	-0.1090	0.3492	0.5001	-0.1336
sp.ent	-0.3490	0.1060	-0.0238	-0.1817	-0.0558
time.ent	0.1781	0.3216	0.0898	-0.0312	-0.0126
entropy	-0.2841	0.2386	0.0113	-0.1983	-0.0627
sfm	-0.3129	0.1402	0.1923	-0.0712	0.0108
meandom	-0.0801	0.1470	-0.4322	0.2513	0.1888
dfrange	-0.1750	0.0290	0.0746	0.0383	0.7121
modindx	-0.0194	-0.2139	-0.2079	-0.1855	-0.5613
dfslope	-0.0101	-0.0069	-0.0094	-0.0035	-0.0233
meanpeakf	-0.0155	0.0616	-0.3743	0.3054	0.0712

whether the dominant frequency stays the same over the total time or varies (Araya-Salas & Smith-Vidaurre, 2017). Since a modulation index of 1 means no modulation (Araya-Salas & Smith-Vidaurre, 2017) and the index has a negative loading, an increase in this component would yield a decrease in modulation index and, thus, an increase in modulation.

5.1.1.1 Exploratory Statistics

Before the RDA, I assessed the variances within and between individuals and groups. I assessed each individual's and group's variance for the five different PCs alongside the variance in the whole dataset (see table 5.2).

The intraindividual variance is consistently smaller than the variance observed in the complete dataset. Examining the intragroup variances, only group RR and ZU exhibit lower variances, compared to the variance of the complete dataset. Group EK demonstrates a higher variance in PC4, compared to the variance of the complete dataset. Further, groups BB and UK show higher variances in four PCs, compared to the variance of the complete dataset. Specifically, group BB has a smaller intragroup variance only in PC3, while group UK shows a lower variance solely in PC4. A box plot for each PC was used to visually compare the individuals and groups (figures 5.1 and 5.2).

I conducted a Kruskal-Wallis test for each PC to statistically compare the observed interindividual and intergroup variances, highlighted by the initial and visual analyses (see table 5.3). The resulting p-values indicate that the null hypothesis of the test can be rejected for all individuals and groups. This finding confirms that at least one individual or group significantly differs from the others.

Table 5.2: Comparison of overall variances with intraindividual and intragroup variances. Values marked in orange symbolize a greater variance within the individual or group, compared to the variance of all calls.

	Overall	**VEKF017**	**VUKF018**	**VZUF014**	**VRRM181**	**VRRM181**
PC1	6.56	3.65	3.46	1.74	3.46	2.16
PC2	5.49	3.96	3.65	1.80	3.72	1.85
PC3	2.64	2.13	1.77	1.35	1.88	1.80
PC4	2.18	1.37	1.82	0.74	1.25	0.98
PC5	1.30	1.40	1.15	0.44	0.75	0.69
	Overall	**BB**	**EK**	**RR**	**UK**	**ZU**
PC1	6.56	11.56	4.62	3.57	7.29	3.10
PC2	5.49	7.11	3.39	4.24	7.13	4.20
PC3	2.64	2.25	2.43	1.90	2.76	1.69
PC4	2.18	2.37	3.28	1.96	1.94	1.85
PC5	1.30	1.93	1.28	0.95	1.56	0.83

Table 5.3: Kruskal-Wallis test p-values per PC for the five individuals and groups.

	PC1	**PC2**	**PC3**	**PC4**	**PC5**
Individual	$< 2 * 10^{-16}$	$< 2 * 10^{-16}$	$2.613 * 10^{-15}$	$< 2 * 10^{-16}$	$2.287 * 10^{-4}$
Group	$< 2 * 10^{-16}$	$< 2 * 10^{-16}$	$< 2 * 10^{-16}$	$1.924 * 10^{-14}$	$1.232 * 10^{-13}$

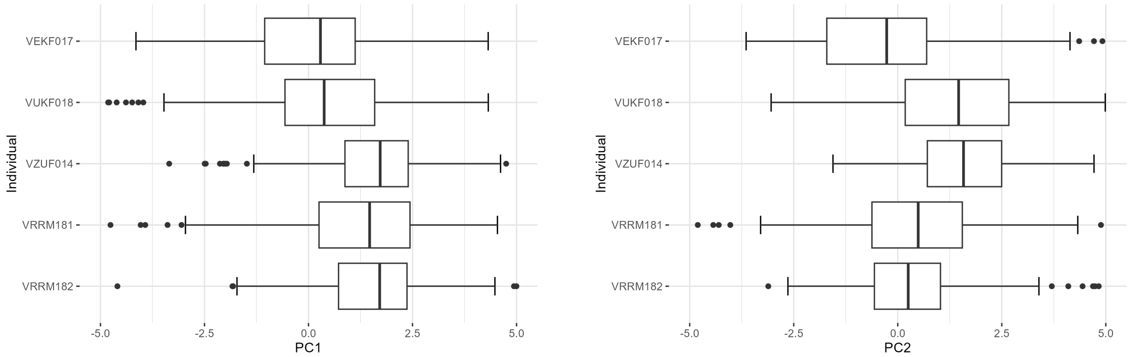
A Dunn's test with Bonferroni adjustment (Dinno, 2015) was performed to identify pairwise differences between individuals and groups. The percentages of significant pairwise comparisons for the individuals and groups, respectively, across each PC were summarized (table 5.4). The detailed set of p-values of the Dunn's test is provided in the appendix (see table A.1). For both groups and, to a lesser extent, individuals, the number of pairs with significant differences decreases as the component number increases, with PC5 showing the fewest pairs with significant differences among groups and individuals. Pairs of groups are more often significantly different than pairs of individuals, with the exception of PC4.

Table 5.4: Percentages of pairs of individuals and groups with significant p-values in the Dunn's test at α-levels of 0.05 and 0.1, split by PC.

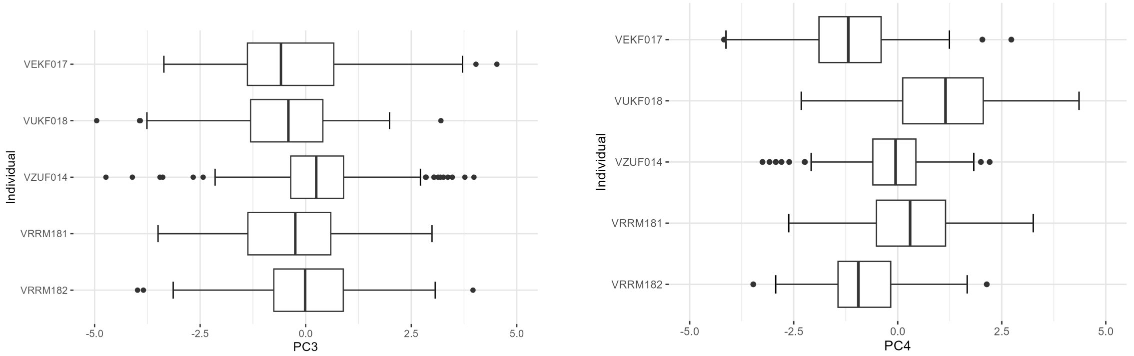
	% Significant ($\alpha = 0.05$)	**% Significant ($\alpha = 0.1$)**
	Individuals	
PC1	60	60
PC2	70	70
PC3	60	60
PC4	90	90
PC5	10	20
	Groups	
PC1	90	90
PC2	70	70
PC3	70	70
PC4	70	80
PC5	50	60

5.1.2 Biological, Social, and Environmental Drivers

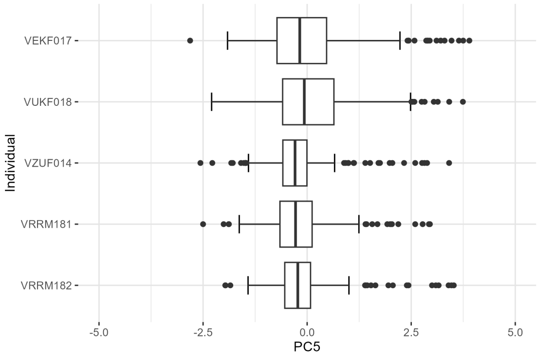
In the first RDA, each driver (listed in chapter 4.2) was used separately as a variable explaining the variation in the set of five PCs. The explanatory power was assessed using



(a) PC1. (b) PC2.



(c) PC3. (d) PC4.



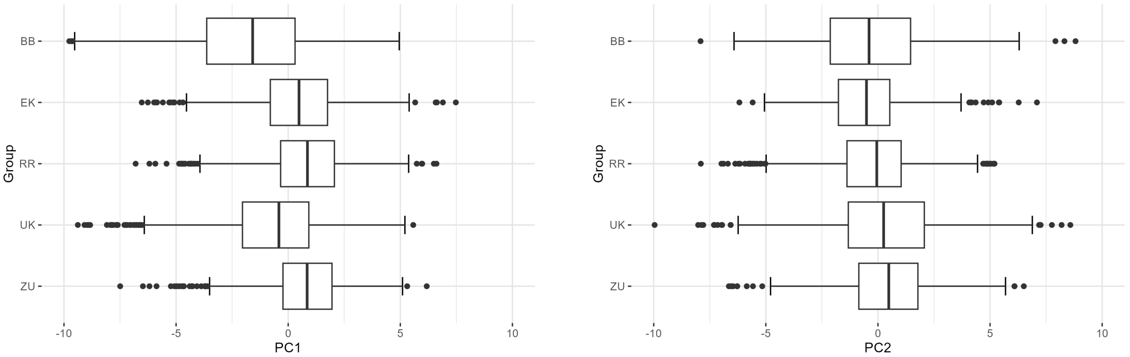
(e) PC5.

Figure 5.1: Interindividual variation in the different PCs for five individuals.

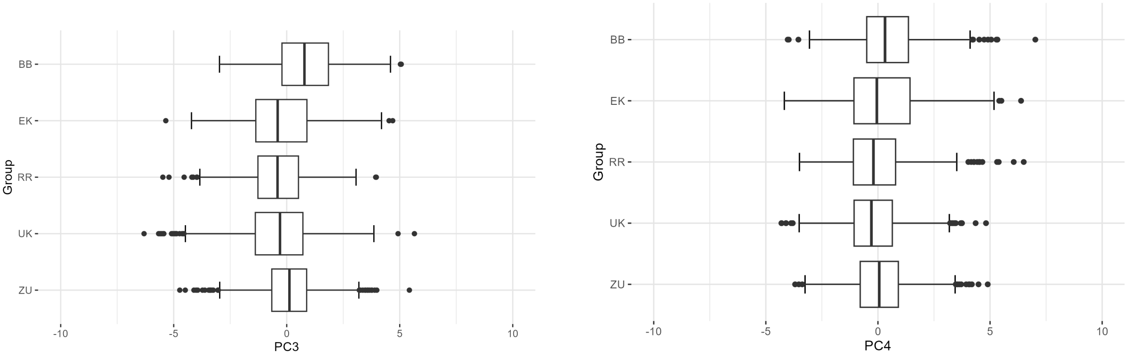
the adjusted R^2 value (table 5.5).

When using the individual as a single explanatory variable in a RDA, it explains almost a third of the variation in the PCs ($R^2_{adj} = 28.86\%$). The social driver group identity explains almost 10% of the variation in the PCs. Another social driver, dominance, as well as the two biological drivers, sex, and age category, have very low explanatory power of less than 2%. According to the rule set in chapter 4.2, these variables were excluded from the partial RDA due to their low explanatory power. They were assumed to be inherent to the individual and reflected well enough in the variable defining the individual.

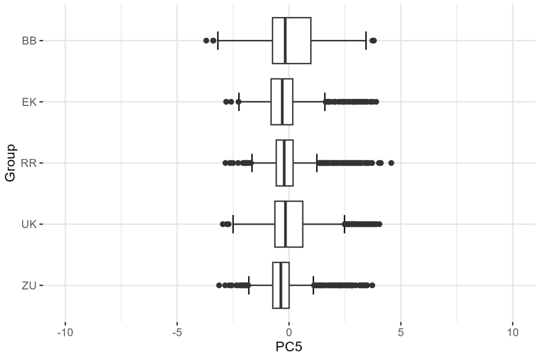
The adjusted R^2 did not substantially increase when adding other explanatory variables to the individual. For example, when adding the group to the individual, the variation



(a) PC1. (b) PC2.



(c) PC3. (d) PC4.



(e) PC5.

Figure 5.2: Intergroup variation in the different PCs for five groups.

Table 5.5: Adjusted R^2 values for all explanatory variables analyzed separately, with the PCs as dependent variables.

Explanatory variable	$\mathbf{R^2_{adj}}$ [%]
Individual	28.86
Group	9.60
Dominance	1.19
Sex	0.65
Age Category	1.70
Genetics	4.10
Distance to Kuruman	2.51

explained increased by less than half a percent to $R^2_{adj} = 29.31\%$. The effect of group identity on the acoustic signal seemed to be entangled with the individual's signal. A partial RDA and variation partitioning were employed with the individual as explanatory variable alongside different covariables to unveil unique and joint effects (table 5.6).

Table 5.6: Partial RDA and variation partitioning results using the individual as explanatory variable alongside different covariables. Refer to table 5.5 to get fractions of the individual variables, i.e., $[a + c]_{adj}$ or $[b + c]_{adj}$. Some percentages are inaccurate due to rounding.

Explanatory	**Covariable**	$\mathbf{a_{adj}}$ [%]	$\mathbf{b_{adj}}$ [%]	$\mathbf{c_{adj}}$ [%]	$\mathbf{[a + b + c]_{adj}}$ [%]
Individual	Group	19.70	0.45	9.15	29.31
Individual	Genetics	24.76	0.0	4.1	28.86
Individual	Distance to Kuruman	26.4	0.06	2.45	28.92

The proportion of variation explained by the covariables with any effects of the individual removed is much lower, compared to their explanatory power in the standard RDA. Among the three covariables, the social driver group identity explains the most variation with any effects of the individual removed. However, its explanatory power still lies below 1%. The percentage explained jointly by the individual and the group (c_{adj}) is rather high with around 9%. The explanatory power of genetic relatedness decreases to 0, when removing any effects of the individual, yet, there are some joint effects. For the distance to the river, the unique and joint explanatory powers are low. The explanatory power when adding the covariables to the individual ($[a + b + c]_{adj}$) is almost the same for all three combinations, with the combination of individual and group identity yielding slightly higher values than the other two.

5.1.3 Interindividual Variation

The interindividual variation was extracted to analyze it in more detail. Each explanatory variable was used separately in the RDA with $\hat{PC}$ as dependent variable (table 5.7).

Table 5.7: RDA regressing the dependent variables, representing interindividual variation ($\hat{PC}$), on each explanatory variable separately.

Explanatory variable	$\mathbf{R^2_{adj}}$ [%]
Group	30.75
Dominance	4.18
Sex	2.18
Age Category	2.73
Genetics	13.95
Distance to Kuruman	7.65

The social drivers group identity and genetic relatedness have the most influence on the interindividual variation. The environmental driver, distance to the river, also shows some influence by explaining almost 8% of the total interindividual variation. Similarly to the results of the first RDA (see table 5.5), the biological drivers sex and age category, and the social driver dominance have low influence. Among these, the dominance status has the largest influence, with a value of slightly above 4%. The biological drivers (sex and age category) have low influence, between 2% and 3%. Because group identity explained almost a third of the variation in $\hat{PC}$, I used it as the variable to compare all others with in a partial RDA and variation partitioning (table 5.8).

Table 5.8: Partial RDA and variation partitioning results on $\hat{PC}$ using the group as an explanatory variable with different covariables. Refer to table 5.7 to get the fractions of the variables by themselves ($[a+c]_{adj}$ or $[b+c]_{adj}$). Some percentages are inaccurate due to rounding.

Explanatory	**Covariable**	$\mathbf{a_{adj}}$ **[%]**	$\mathbf{b_{adj}}$ **[%]**	$\mathbf{c_{adj}}$ **[%]**	$\mathbf{[a+b+c]_{adj}}$ **[%]**
Group	Dominance	29.43	2.87	1.31	33.62
Group	Sex	31.61	3.04	-0.86	33.79
Group	Age Category	30.32	1.98	0.63	32.93
Group	Genetics	26.56	9.76	4.19	40.51
Group	Distance to Kuruman	23.54	0.44	7.21	31.18

The joint effects of group and dominance, as well as those of group and age category are minimal. The negative value for the joint effects between group and sex suggests that the response matrix is more effectively explained by a combination of these variables than by either variable independently (Quebec Centre for Biodiversity Science, 2023d). When removing the effects of the group, genetic relatedness accounts for approximately 10% of the variation in the data. The two social drivers group identity and genetic relatedness also share a considerable amount of joint effects. The joint effects between group identity and the distance to the river are even more influential. When removing the influence of the group, the explanatory power of the distance to the river is reduced to almost 0. Using group identity alongside genetic relatedness as explanatory variable for the interindividual variation, the RDA explains the largest amount of interindividual variation, with a little over 40%. Other variable combinations do not offer substantially greater explanatory power compared to using group identity as a sole explanatory variable.

5.2 Correlation Analysis: Influence of Spatial Proximity

This chapter presents the results of the correlation analysis to assess to what extent spatially and temporally close meerkats influence each other's CCs. For that, I compared the distributions of informed (blue) and uninformed (pink) correlation coefficients by plotting them (figure 5.3) and quantified the difference between the distributions using the KL divergence (table 5.9).

The plot shows that the distributions of correlation coefficients for the first and fourth PC are similar for informed and uninformed predictors. The distributions of the second, third, and fifth component show small to substantial differences. For PC1, the informed distribution is more peaked and narrower, compared to the uninformed distribution, while both consist of negative correlation coefficients. PC2 shows similar shapes of distribution for both distributions, with the informed one tending more to the negative coefficients. PC3 shows great differences between the two distributions in regard to their values ranges albeit they are similar in shape. The informed coefficients are mostly positive, while the uninformed ones are mostly negative. The distributions of PC4 are very similar and centered around 0. PC5 also shows differences. The majority of informed coefficients are negative and the distribution is more peaked, compared to the uninformed one, which has

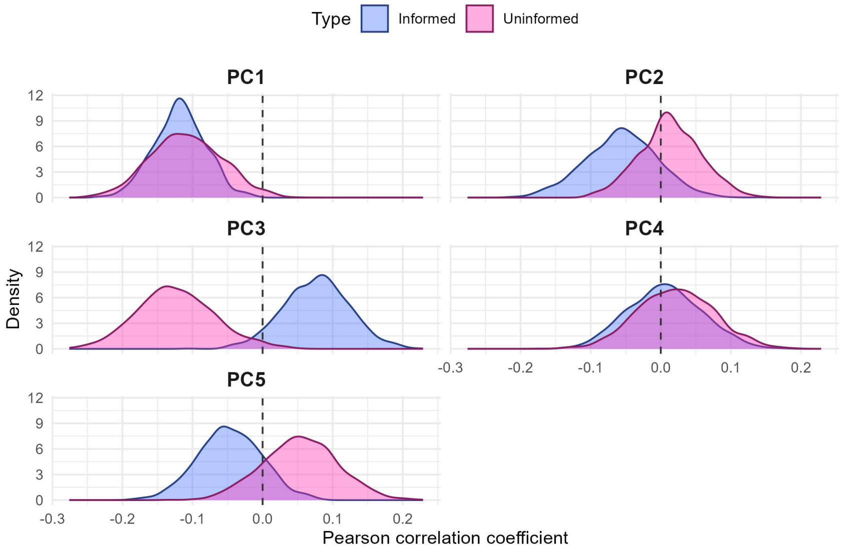


Figure 5.3: Distributions of informed and uninformed correlation coefficients.

a majority of positive correlation coefficients. The KL divergences served as numerical comparisons of the distributions (table 5.9).

Table 5.9: Kullback-Leibler divergence between probability distribution of informed and uninformed correlation coefficients.

Component	**KL divergence**
PC1	0.13
PC2	3.49
PC3	15.04
PC4	0.07
PC5	4.53

The KL divergence for PC4 is the lowest, with a value of 0.07, indicating almost no differences. For PC1, it yielded similar results, with a value below 1. The distribution differences for PC2 and PC5 are similar, with PC5 showing a slightly higher divergence of 4.53. For PC3, for which the KL divergence is more than three times greater than for PC5, is by far the highest, with a value of over 15.

Chapter 6

Discussion

This chapter discusses the results, aiming to answer the two RQs posed (chapters 6.1 and 6.3). Additionally, it explores two other territory scenarios for calculating the obstacle interaction probability that were under consideration (chapter 6.2). The chapter concludes by highlighting limitations of the study (chapter 6.4).

6.1 Redundancy Analysis

The first part of the discussion focuses on the results for the first RQ which is answered at the end of this chapter (section 6.1.3). These include the results of the exploratory statistics on the five PCs and RDA results.

6.1.1 Exploratory Statistics on Dependent Variables

The exploratory analysis aimed to assess my data's alignment with earlier studies on meerkat CCs. As expected from previous findings (Townsend et al., 2010), intraindividual variance was smaller than overall variance, supporting the presence of individuality in acoustic signature. However, patterns of variation among individuals and groups also revealed unexpected trends, particularly for specific PCs, which are discussed in the following paragraphs.

Boxplots (figure 5.1) revealed substantial interindividual variation, with notable differences across PCs. While most PCs (1 to 3, and 5) followed the expected trend of reduced interindividual differences with decreasing explanatory power, PC4 deviated significantly. This component exhibited large interindividual differences, even between full siblings VRRM181 and VRRM182, who showed similar values for other PCs. This suggests that PC4, associated with symmetry and dominant frequency-related features, strongly drives interindividual variation. The sibling pair VRRM181 and VRRM182 underscored the genetic influence on acoustic features. Their indistinguishable PC1, PC2, and PC5 values (see table A.1) suggest that frequency, tonality, modulation and time-related variables may run within a genetic line. Their difference in PC3 and very clear difference in PC4 highlight the role of symmetry and dominant frequency-related features in shaping individual variation.

Pairwise statistical comparisons betwen individuals supported the observed patterns. The Kruskal-Wallis test indicated significant differences across individuals for all PCs, while Dunn's test revealed that the proportion of significantly different individual pairs decreased with higher PCs numbers. Coinciding with the finding from the boxplots, PC4 showed the highest percentage of significant differences (90%), outperforming PC1, which explained the most variation in the data but only had 60% significant pairwise differences between individuals.

Variance patterns among groups revealed several interesting trends. Groups BB and UK exhibited higher intragroup variance than others. I expected these groups to have a larger amount of calls which would facilitate the occurrence of outlying calls which would increase the intragroup variance. While group UK aligns with expectations due to its high number of calls (1 079), variance in group BB was unexpectedly high despite having fewer calls (474). In contrast, groups RR and ZU had the lowest variances despite recording only slightly fewer calls than UK. Group variance was also compared with the number of individuals in each group. I expected that groups with many calls recorded per individual showed lower intragroup variance than groups with a low number of calls per individual due to individuality in CCs. Group BB, with the highest number of individuals (19) and the lowest number of calls, showed high variance as expected. Group ZU, with a high amount of calls and fewer individuals, leading to around 60 calls per individual, fits my expectation with lower intragroup variance. Group UK, with the highest number of calls per individual (around 80 per individual), still showed bigger variance than overall in four out of five PCs, suggesting that call number alone does not explain the observed patterns of intragroup variances.

Boxplots indicated that intergroup differences decreased with increasing PCs, as expected. PC1 and PC2 displayed the most significant intergroup variation, whereas higher PCs showed minimal differences. These findings suggest that group-level signatures are less pronounced than individual signatures in this dataset, a tendency consistent with earlier studies that found group signatures after accounting for individual variation (Townsend et al., 2010).

6.1.2 Biological, Social, and Environmental Drivers

The RDA aimed to analyze the drivers influencing variation in CCs, including the individual, and biological (age and sex), social (group identity, dominance, and genetic relatedness), and an environmental driver (distance to the river). Consistent with previous research (Townsend et al., 2010), individual variation was found to play the most significant role in explaining differences in CCs. When extracting only interindividual variation, the social drivers group identity and genetics were also found to shape meerkat CCs. The results of the various analyses are discussed in detail below.

The results of the first RDA, where each driver was analyzed separately, align with previous findings of individuality in the acoustic structure of meerkat CCs (Townsend et al., 2010) and with the results of the exploratory statistics. The individual accounts for the majority of the variation in the data, far exceeding the explanatory power of other variables. The social driver group identity explains the second-largest proportion of variation, though its contribution is only one-third that of the individual. The remaining drivers each explain less than 5% of the variation. The partial RDA suggests that the social drivers group identity, genetic relatedness, and the environmental driver can serve as proxies for the individual, as much or all of the variation of these variables are caused by joint effects with the individual. An earlier study found group-specific variation when accounting for individual differences (Townsend et al., 2010). The results provide an initial indication that similar effects may be present in my data. The follow-up analysis explores this in more detail.

When performing a RDA using only the interindividual variation, I gained insight into the drivers of this variation. Group identity and genetic relatedness appear to be the most significant drivers, consistent with existing literature (Townsend et al., 2010). Interestingly, dominance, sex, and age category only explain only a small portion of the interindividual variation. This is surprising given that previous studies suggested differences in the acoustic features of CCs based on social status and sex (Mausbach et al., 2017). For instance, differences in vocal tract anatomy between males and females could lead to variations in call production, as could age-related changes. Younger meerkats, being smaller, may have a smaller vocal tract that influences their calls, while acquiring dominance and the accompanying increase in body size might also explain acoustic differences between dominant and subordinate individuals (Mausbach et al., 2017). However, my results indicate that these drivers have a minor influence on interindividual variation. One possible explanation is the analysis setup. Over the three-year study period, some individuals were recorded in multiple age categories (e.g., as subadults and adults), making it harder to explain the variation solely by age since the intraindividual variation possibly showing differences was removed. Similarly, dominance status may have changed over time for individuals recorded repeatedly, complicating explaining the variation only using dominance status. For sex, however, the low explanatory power is surprising, as sex does not change over time. The RDA results using the PCs, representing interindividual variation, align with previous findings of the importance of the social drivers group identity and genetic relatedness (Townsend et al., 2010). The discrepancies for dominance status, age category, and sex may partly reflect the differences in analysis setup and in aims of the analysis.

The final partial RDA, using PCs representing interindividual variation as dependent and group as an explanatory variable, provided a more detailed understanding of the above results. The joint effects of group and dominance as a predictors are negligible, which is expected given that each group typically has a single dominant pair. Together, these two variables explain about one-third of the total interindividual variation. Similarly, the joint contribution of group and sex is non-existent, as groups consist of both male and female individuals. The combination of group and sex has the same explanatory power as group and dominance, with both dominance and sex adding roughly equal amounts to the explained variation by group. The combination of group and age category shows similar results, with only a small portion attributed to joint effects. This may reflect a tendency for groups to have more individuals sharing the same age, for instance, an overall younger or older group. However, the explanatory power of group and age combined is slightly lower than that of group combined with dominance or sex. The combination of group and genetic relatedness reveals effects that are explained both individually by each variable and joint effects. The joint effects likely arise from the high genetic relatedness within groups, as expected. However, migration may also lead to genetic relatedness between groups, which could explain the unique variation attributed to genetic relatedness when removing effects of the group. This combination of variables explains just over 40% of the interindividual variation, the highest value observed in the analysis. Lastly, the combination of group and distance to the river Kuruman, while having the lowest overall explanatory power, offers an interesting finding. When the effects of distance to the river are removed, the portion of variation uniquely explained by the group drops to around 23%. This suggests that groups tend to remain in consistent locations over time, making the distance to the river a proxy for group. The low unique explanatory power of distance is likely caused by its calculation being group-based, as GPS data were only available at the group level. In conclusion, the partial RDA reveals that group identity and genetic relatedness play a key role in explaining interindividual variation. Genetic relatedness is present within as well as between groups. Other drivers like dominance, sex, age, and distance to the river contribute less, though the latter highlights the spatial stability of groups over time.

6.1.3 Research Question 1

The first RQ, in the box below, can be answered under consideration of the above discussed results.

> **RQ1**: How do biological, social, and environmental drivers influence the acoustic features of meerkat CCs?

The individual most prominently influences acoustic features of meerkat CCs. When considering only interindividual variation, the social driver group identity is the main influence on variation. Genetic relatedness also drives interindividual variation. Genetic relatedness showed partly joint effects with group identity, indicating genetic relatedness between groups. However, genetic relatedness is also present between groups. The environmental driver, distance to the river, can be used as a proxy for group and indicates location stability of groups over time. The remaining social driver dominance and both biological drivers (sex and age) were less influential for the variation in meerkat CCs.

6.2 Interaction Probability: Territory Scenarios

Conceptualizing meerkat territories was an important task in defining the obstacle interaction probability (see section 4.3.1.2). Besides the scenario I decided on, I created two other scenarios that are supported by literature for the discussion with the meerkat expert. While the complete habitat area scenario I used is simple and rather restrictive, the other two presented below are more complex and liberal.

6.2.1 Scent-Highway Scenario

The scent-highway scenario first clusters the recorded GPS points using the density-based spatial clustering of applications with noise (DBSCAN) algorithm. It forms clusters of dense points and removes any noise points defined by points in areas of low density (Hahsler et al., 2019). In a second step, multiple MCPs are formed around the clusters – areas of high density, so called core areas – found. These core areas are combined to a complete graph using shortest path connections between them. A buffer around the paths is added to create a polygon representing highways as connections between the core areas (see figure 6.1). This scenario is of medium restrictiveness, with less area as restricted area than in the complete habitat area scenario.

This model assumes that the recorded GPS points cluster around resource hot spots or frequently used boltholes critical to the meerkats' survival. The scenario assumes that the meerkats spend most of their time in these areas and that they rarely move outside. Scent marking occurs more intensely in these highly frequented areas and along the frequently traveled paths between them. While meerkats may move beyond the outermost GPS points recorded, these additional locations are assumed to not be relevant for the defended territory as groups are assumed to stay close to the recorded centers. Due to heavy scent marking in the core areas, foreign groups avoid passing through these areas. Shortest paths are used to connect two hot spot areas, assuming that meerkats prefer to stay in the resource-rich areas and transfer from one to another as efficiently as possible. Along these paths, however, it is assumed that there are some refuge locations in case danger is approaching when transferring between core areas. Adding a buffer around the shortest path line supports the assumption that a meerkat group does not move in a line one after

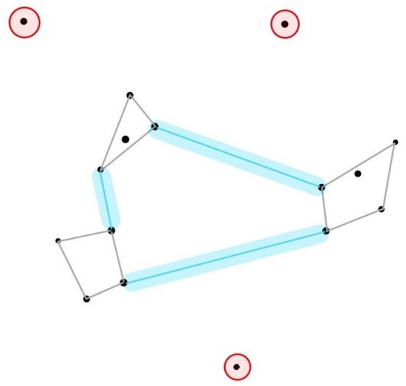


Figure 6.1: Sketch of the scent-highway scenario. The black dots represent the GPS recordings and the red circles the noise points which were removed. The gray lines are MCPs around core areas, and the light blue lines the highways including a buffer.

another, but rather in a flock. While scent-highways are assumed to be avoided by foreign groups whenever possible, they are crossed occasionally. It has been shown in other species using scent marks that centrally clustered scent marks are preferred over scent marks along the boundary, especially if territories are fairly large (Jordan et al., 2007). For meerkats, scent marks were found to be denser in the core area of the territory than in the border regions (Jordan et al., 2007). Due to relatively large territory sizes, it is likely not possible for meerkats to maintain a well-marked perimeter against intruders and it is more feasible to only keep the core areas well-marked (Jordan et al., 2007). Further, meerkats prefer to stay close to a refuge and scent marks are selectively positioned close to these refuges (Jordan et al., 2007).

6.2.2 Core Area Scenario

The core area scenario clusters the recorded GPS points using the DBSCAN clustering algorithm as in the scent-highway scenario. The algorithm forms clusters of dense points and removes the noise points. A concave hull is used to form a territory around each cluster of GPS points. This scenario is the least restrictive, since it only considers core areas as restricted zones.

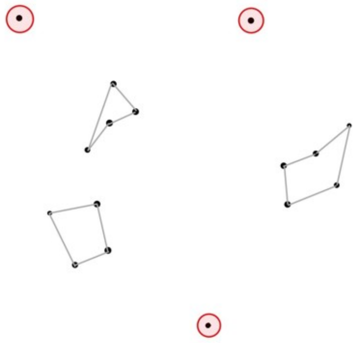


Figure 6.2: Sketch of the core area scenario. The black dots represent the GPS recordings, the red circles the noise points that were removed and gray lines the concave hulls around clustered recordings.

This model assumes that the recorded GPS points are an indication of where meerkats move around. They are, however, not assumed to consider the complete area as their territory and, consequently, might not use scent markings in the complete area. Meerkats are assumed to rather mark and defend only their core areas which they heavily frequent. Remote points at the edges are not assumed to be critical to the territory definition and do not need to be included since they exhibit only limited scent marking. Further, the scenario assumes that while a group may spread out from the recorded center point, it is not relevant to the defended area as a group generally is compact while foraging and moving around (Driscoll et al., 2024). The few meters deviation from the center point do not significantly affect the marked territory and can be ignored. Foreign groups are assumed to avoid the core areas as they are heavily scent marked. This scenario is supported by the finding that well-maintained scent marks along the border are costly for meerkats which leads them to concentrate the marks around important and highly-frequented areas of their territory (Jordan et al., 2007).

6.3 Correlation Analysis: Influence of Spatial Proximity

The influence of spatial proximity on the acoustic features of CCs was assessed by comparing correlation coefficients between two calls in spatial and temporal proximity with those of two calls in spatial and temporal distance. The comparison of the distributions of informed and uninformed correlation coefficients yielded markedly different results across PCs. For the discussion, I focus on the components that showed notable differences. As PC1 and PC4 exhibited minimal to no differences, they will not be discussed further. Considerable differences were observed in PC2 and PC5, while PC3 showed substantial differences. These components are highlighted in the following discussion (sections 6.3.1 and 6.3.2). Finally, I address the second RQ (section 6.3.3).

6.3.1 Principal Components 2 and 5

The informed correlation coefficients for the second component tend to fall more within the negative value range compared to the uninformed coefficients and showed a KL divergence of 3.49. This component has high loadings of time-related features (e.g., *duration*, *time.median*), most of which have negative loadings, except for the feature specifying time entropy (*time.ent*). An increase in PC2 corresponds to a decrease in call duration, meaning the call shortens.

Since the informed correlation coefficients tend to be negative, an increase in PC2 of a call would result in a decrease of PC2 of a call in spatial proximity. Given the negative loadings, this implies that a decrease in duration would lead to an increase in duration, and vice versa. For example, an increase in duration of a call would lead to a decrease in duration of a call in spatial proximity.

Call length has previously been associated with seasonality, with longer calls observed during the non-reproductive season compared to the reproductive season (Mausbach et al., 2017). It is possible that the 90-day period used to define temporal proximity sometimes spanned the transition between reproductive and non-reproductive seasons, leading to variations in call length between the start and end of the observed period. This could explain the trend for temporally proximal calls to differ in length.

In conclusion, divergence in acoustic features, such as call length, may arise between groups in spatial proximity. However, factors like the inclusion of calls from different seasons could also drive this variation. Considering that the KL divergence is only 3.49, it is likely that additional drivers beyond spatial proximity contribute to the observed differences.

The distributions of informed and uninformed correlation coefficients for PC5 show similar patterns to those observed for the second component, with a KL divergence of 4.53. The informed distribution is more peaked and tends toward the negative value range, while the uninformed distribution is flatter and leans toward the positive range.

PC5 is characterized by very high loadings for the dominant frequency range (*dfrange*) and the modulation index (*modindx*). The dominant frequency range is positively loaded, while the modulation index is negatively loaded. Consequently, an increase in PC5 corresponds to an increase in the dominant frequency range but a decrease in the modulation index, i.e., an increase in modulation.

Relating these findings to the distribution of correlation coefficients, an increase in the dominant frequency range of a call leads to a decrease in the dominant frequency range of a temporally and spatially proximal call. Similarly, an increase in modulation of a call results in a decrease in modulation of a proximal call. In other words, strongly modulated calls tend to be followed by less modulated calls within nearby groups, and vice versa.

As with PC2, these results suggest the possibility of divergence between groups in close proximity via dominant frequency range and call modulation. However, it is important to note that this component explains only 5% of the variation in acoustic features, indicating that its influence may be relatively minor.

The analyzed PC2 and PC5 reveal negative relationships – or divergence – between two calls in spatial and temporal proximity. Beyond the explanations already discussed, this presumed divergence might be influenced by genetic relatedness. While the effects of genetic relatedness within a group were excluded by avoiding correlations between calls from the same group, the effects of genetic relatedness between groups – accounting for nearly 10% of the interindividual variation (see table 5.8) – were not controlled for. Migration could lead to an individual relocating to a spatially distant group, resulting in proximity to groups with less shared genetic material and distance from its original group with higher genetic similarity. Given the considerable contribution of genetic relatedness to interindividual variation, this may partly explain the reduced similarity between calls of spatially proximal groups. This hypothesis is supported by the findings of the pairwise comparisons between two full siblings, which showed that they were indistinguishable in PC2 and PC5, leading to the assumption that these features are more strongly influenced by genetic relatedness than others (see section 6.1.1).

6.3.2 Principal Component 3

The third component shows the largest difference between the informed and uninformed distributions, with a KL divergence of approximately 15. The informed correlation coefficients are almost exclusively in the positive value range, while the uninformed coefficients are largely negative and less peaked. PC3 includes high negative loadings for frequency-related features (*meandom*, *meanpeakf*, *freq.median*) and high positive loadings for symmetry-related features (*skew*, *kurt*). An increase in PC3 generally reflects a decrease in mean dominant and median frequency. Since the informed correlations are positive, increases (or decreases) in the mean dominant and median frequency of one call correspond to similar changes in spatially proximal calls, suggesting a pattern of accommodation in PC3 between such calls.

While symmetry-related features are present in both PC3 and PC4, the latter shows almost no difference between informed and uninformed distributions. Symmetriy-related features have the highest loadings in PC4 (around 0.5) but are less influential in PC3 (around 0.35). Instead, frequency-related features are more dominant in PC3, with notable differences in specific loadings. For example, median frequency (*freq.median*) has a high negative loading in PC3 but has almost no influence on PC4. Similarly, the absolute loading of the mean dominant frequency (*meandom*) is higher in PC3 (-0.43) than in PC4 (0.25). These differences suggest that median and mean dominant frequencies are primarily responsible for the observed divergence between informed and uninformed correlation coefficients, indicating that these features are more influenced by spatially and temporally proximal calls than other frequency-related features.

This pattern can be related to previous findings associating lower fundamental frequencies, which are similar to dominant frequencies (Telgarsky, 2013), with dominant female individuals (Mausbach et al., 2017). CCs with lower median and lower mean dominant frequency tend to influence calls in spatial proximity to also have lower median and mean dominant frequencies. If the lower frequencies are linked to dominant females, this could suggest that calls of dominant females are more likely to be followed by similar calls from proximal dominant individuals, whereas calls from subordinate individuals are followed by calls from subordinates.

Another possible explanation for this similarity between spatially proximal groups is genetic relatedness. While genetic effects within groups were controlled for, genetic similarity between proximal groups might contribute to these patterns. For example, newly emerging groups formed by subordinate individuals that split from their original group and inhabited an area close to their old group, might increase genetic similarity between groups in spatial proximity. Genetic similarity accounts for a portion of interindividual variation and could influence the observed accommodation in call features.

Lastly, external factors, i.e., the environment groups live in, may also play a role. Spatially proximal groups might inhabit similar environments, leading to greater similarity in call frequencies. For example, groups near the riverbed versus those in sand dunes might experience subtle environmental variations that influence call frequencies. While environmental differences in the KMP study area were presumed to be minor, such habitat-related factors could still contribute to variation in CCs, as suggested by prior research on the influence of environment on mammalian vocalization (Martin et al., 2017).

In conclusion, while the observed pattern of accommodation in PC3 may result from various factors, spatial proximity could be a contributing factor. If accommodation occurs through spatial proximity, this would indicate that meerkats have some capacity for vocal learning, specifically, production learning (see section 2.2.1).

6.3.3 Research Question 2

Under consideration of the above discussed results, I answer the second RQ, which is stated below.

RQ2: To what extent do spatially and temporally close meerkats influence each other's CCs? Is there evidence for accommodation, where spatially and temporally close meerkats from different groups have similar CCs, or for divergence, where spatially and temporally close meerkats have different CCs?

Correlation coefficients between CCs in spatial and temporal proximity are generally low, ranging from -0.3 to 0.3, suggesting a limited influence on CCs between spatially and temporally proximal meerkats. However, notable differences emerge when comparing correlations between calls from spatially and temporally proximal meerkats with those from distant individuals, particularly in PC2 (KL divergence: 3.49), PC3 (15.04), and PC5 (4.53). The direction of these relationships varies. For PC2 and PC5, the negative correlations suggest divergence in time- and modulation-related acoustic features, whereas for PC3, the positive correlation indicates accommodation in frequency-related features among spatially proximal meerkats.

These findings suggest that while spatial proximity may influence certain acoustic features, alternative explanations, such as genetic relatedness between groups, similar environmental conditions (Martin et al., 2017), or the social status of calling individuals (Mausbach et al., 2017), cannot be ruled out. Although overall correlations are modest, incorporating spatial proximity into the analysis provides additional insights into the interindividual variation of meerkat CCs.

6.4 Limitations

Although this thesis provides insights into the influence of spatial proximity on meerkat CCs, certain limitations must be acknowledged. These include challenges related to data availability, methodological choices, and parameter selection, which are elaborated in this chapter.

Level of Aggregation A notable limitation of this analysis arises from the disparity in accuracy between the CC and GPS data. While each CC is assigned to an individual meerkat belonging to a specific group, the available GPS data are recorded at the group level, representing the central location of the group at a given time. As a result, all individuals in the same group share a single GPS point within a sampling interval, obscuring the precise spatial locations of individual meerkats and introducing inaccuracies into the analysis. The interaction probability could only be calculated at the group level, not for individual, whereas calls were assigned to individuals. Consequently, all calls from a group recorded on the same day share the same interaction probability with other groups. This averaging could lead to either an underestimation or an overestimation of the influence of spatial proximity on acoustic features, as the specific locations of individuals are not captured.

Choice of Territory Scenario The decision to use the complete habitat area scenario, including a buffer, to calculate obstacle interaction probabilities is biologically realistic but conservative in terms of data. This method assumes that the entire habitat area, including the buffer, is inaccessible to other groups. As a result, many group combinations had no connectivity, leading to a high frequency of zero values for the obstacle interaction probabilities. When multiplied by the Euclidean interaction probabilities, many non-zero values were reduced to zero. For each group combination on a given date, the final interaction probability – calculated as the mean value of all observations of interaction probabilities for that date – was often very small. This affected the sampling probabilities, leading to some pairs of groups having very similar interaction probability values.

Parameter Selection The study's outcomes were influenced by several parameters which were determined based on estimations and informed assumptions. Modifying these parameters could yield different results. One critical parameter was the 90-day period used to evaluate interaction probabilities and select calls for the informed analyses. Extending this period might have increased interaction probability values but also increased the number of potential calls for the sampling, potentially leading to greater averaging. The parameters of the sigmoid function used to transform distances into values between 0 and 1 played an important role. These parameters controlled the curve's steepness and defined its convergence points. While the selected values were biologically motivated, alternative parameter sets could potentially better reflect ecological realities and alleviate the issue of small interaction probability values. Additionally, the choice of a ten-meter buffer around the MCP for the complete habitat area scenario was an informed estimate. Adjusting this buffer could shift the distribution of obstacle interaction probability values, increasing the proportion of values equaling 1 or 0, thereby altering the analysis outcomes. Finally, the temporal thresholds for the uninformed correlation analysis were selected within the scope of the data. However, alternative thresholds might have resulted in different outcomes, as these choices were not rooted in a biological rationale.

Method Selection and Interpretation The choice of methods was primarily driven by the constraints of the available data. While these methods allowed for meaningful analyses, they also posed challenges in terms of result interpretation. The use of PCA to reduce the dimensionality and eliminate correlation among acoustic features was essential given the volume of the CC dataset. However, the resulting PCs represent linear combinations of all acoustic features, which makes it difficult to attribute them to a specific feature. While some general interpretations can be made, the precise meaning of the PCs often remains unclear, especially for those without a background in acoustics or evolutionary biology. For the correlation analysis, it could have been helpful to analyze each feature separately, but this would require running many more correlation analyses, which would increase computational demands. The use of RDA added further complexity to the interpretation. This method involves multiple steps, including regressions followed by PCA, which complicates the understanding of the results. Additionally, the complexity of the dependent variables in the RDA made the interpretation even more intricate. Finally, at a preliminary level, the correlation analysis served as a useful tool for exploring potential relationships between spatially and temporally proximal CCs. While it helped identifying patterns, this method cannot determine the slope of the relationship, as would be possible with regression analysis. Additionally, it is crucial to emphasize that correlation does not imply causation (Rogerson, 2020). An observed correlation between two variables does not necessarily indicate a direct or causal relationship; it could simply be coincidental or influenced by external factors. Thus, caution is needed to avoid drawing premature conclusions.

Chapter 7

Conclusion

7.1 Contributions

Spatial Variation in Meerkat Vocalization This thesis introduced a framework to analyze spatial variation in meerkat vocalization, addressing a gap in existing research. This approach for including the spatial dimension into the analysis of vocal behavior in meerkats presents a new perspective to meerkat vocalization studies which have traditionally focused on non-spatial drivers.

Redundancy Analysis This thesis explored how the individual and biological (age, sex), social (group identity, dominance, and genetic relatedness), and environmental drivers (distance to the river) influence variation in meerkat vocalization. The study examined which of these drivers are most influential and how the results align with previous research on variation in meerkat vocalization.

Territory Scenarios This thesis conceptualized meerkat territories by devising three different scenarios based on the available GPS recordings of group centers, using geospatial methods such as MCPs. These scenarios varied in restrictiveness and complexity, offering a range of possible territorial representations.

Spatial Interaction Probability This thesis developed a novel method to estimate the probability of spatial interactions between meerkat groups using Brownian bridges derived from GPS recordings, which model time-continuous trajectories of meerkat movement. Two spatial concepts were incorporated in calculating the interaction probability. The first used a geometry-inspired Euclidean distance calculation based on the Brownian bridge trajectories, while the second included a topology-inspired measure of obstacle between two points of the Brownian bridge trajectories.

Correlation Analysis This thesis employed a correlation analysis to assess if there is evidence for influence of spatial proximity on meerkat vocalization. Correlations were assessed between the acoustic features of two calls in close spatial and temporal proximity, with sampling probabilities derived from the spatial interaction probability. This correlation was repeated multiple times with different call combinations in each run. The distribution of resulting coefficients was compared to a random baseline, calculated using calls in spatial and temporal distance.

7.2 Findings

This work explored the individual and biological, social, and environmental drivers influencing variation in meerkat vocalization, uncovering differences in acoustic features of CCs across individuals, groups and shared genetics. Through various exploratory statistical analyses, I found that CCs vary among individual meerkats and among groups. These analyses emphasized the critical role of individual identity and group identity, as found in earlier studies (e.g., Townsend et al., 2010), in shaping vocalization characteristics. The RDA further supported these findings, revealing that the individual along with the social driver group identity are the primary drivers of CCs variations. Additionally, the social driver genetic relatedness was found to influence the variation in meerkat CCs within and between groups. Overall, these results align with earlier research.

Beyond insights into biological, social and environmental drivers, which were not novel, a key finding of this research is the potential influence of spatial proximity on variation in meerkat CCs. The results indicate vocal accommodation of calls in spatial and temporal proximity in frequency-related features and a divergence in time- and modulation-related features. These findings indicate that spatial proximity relations might contribute to variation in vocalization. This finding is not only novel, but also highlights the need to further investigate spatial proximity and the potential for vocal learning in meerkats.

Overall, the findings of this study confirm the significance of biological, social, and environmental drivers, as well as spatial proximity, in shaping meerkat vocalization, especially the influences of the individual meerkat, the group identity, the genetic relatedness and the spatial proximity between meerkats. By considering multiple drivers influencing vocalization, including spatial proximity, this study provides a comprehensive view of the elements shaping meerkat vocal behavior. It also opens new avenues for research into the relationship between spatial proximity and vocal communication in social animals.

7.3 Outlook

This work provided an initial approach to including spatial proximity in the analysis of variation in meerkat CCs, establishing a baseline for future research aimed at exploring spatial variation in meerkat vocalization or the potential for vocal learning in meerkats. My work could be extended as explained in the following.

To address limitations in the analysis, future work could focus on two key areas. First, to overcome the challenge of identifying features influenced by spatial proximity when using PCs, similar analyses could be conducted on features where spatial proximity may play a role. This would not only help verify the results but also offer deeper insights into features influenced by spatial proximity. Second, to minimize the influence of residual genetic relatedness in the input data used for the correlation analysis, future studies could focus on eliminating this artifact from the call recordings prior to the analysis. This would provide a clearer understanding of spatial proximity as a driver of variation in CCs.

Future work could follow a similar workflow using data that has a better match of aggregation. Specifically, GPS data on an individual or, ideally, call level, could be used to calculate interaction probabilities. These interaction probabilities would then reflect the spatially favored interactions between individual or calls, rather than groups, refining subsequent analyses. Essentially, the GPS aggregation level should match the level of acoustic features. For example, if the acoustic features include interindividual variation, individual-level GPS recordings could be used. If one wishes to analyze each call separately, call-level GPS recordings could be more appropriate.

Another advancement for future research would be to use a method more informative than a correlation analysis. While correlation could provide an initial overview and suggest the potential relevance of spatial proximity on variation in meerkat CCs, it does not necessarily indicate causation. A method like regression would offer more detailed insights into the influence and potential causation of the variables involved. Future research could aim to integrate the interaction probability between calls, individuals, or groups in a more comprehensive analysis to gain a deeper understanding of spatial drivers in meerkat vocalization.

Finally, future studies could apply this approach to other taxa with territorial behavior. The framework for defining spatial interaction probability is adaptable, allowing for the inclusion of various territory scenarios. Researchers can adjust the parameters of the sigmoid function or apply alternative functions, providing a flexible foundation for studying spatial variation in vocalization of other taxa.

Bibliography

Araya-Salas, M., & Smith-Vidaurre, G. (2017). warbleR: An r package to streamline analysis of animal acoustic signals (N. Golding, Ed.). *Methods in Ecology and Evolution*, *8*(2), 184–191. https://doi.org/10.1111/2041-210X.12624

Barker, A. J., Veviurko, G., Bennett, N. C., Hart, D. W., Mograby, L., & Lewin, G. R. (2021). Cultural transmission of vocal dialect in the naked mole-rat. *Science*, *371*(6528), 503–507. https://doi.org/10.1126/science.abc6588

Bateman, A. W., Lewis, M. A., Gall, G., Manser, M. B., & Clutton-Brock, T. H. (2015). Territoriality and home-range dynamics in meerkats, suricata suricatta: A mechanistic modelling approach (L. Borger, Ed.). *Journal of Animal Ecology*, *84*(1), 260–271. https://doi.org/10.1111/1365-2656.12267

Boersma, P. (2014). Acoustic analysis. In R. J. Podesva & D. Sharma (Eds.), *Research methods in linguistics* (1st ed., pp. 375–396). Cambridge University Press. https://doi.org/10.1017/CBO9781139013734.020

Borcard, D., Gillet, F., & Legendre, P. (2011). *Numerical ecology with r* [OCLC: ocn690089213]. Springer.

Borg, I., & Groenen, P. J. F. (2005). *Modern multidimensional scaling* (2nd ed.). Springer New York. https://doi.org/10.1007/0-387-28981-X

Buchin, K., Sijben, S., Arseneau, T. J. M., & Willems, E. P. (2012). Detecting movement patterns using brownian bridges. *Proceedings of the 20th International Conference on Advances in Geographic Information Systems*, 119–128. https://doi.org/10.1145/2424321.2424338

Bullard, F. (1999). *Estimating the home range of an animal: A brownian bridge approach* [Doctoral dissertation, University of North Carolina].

Byrne, M. E., Clint McCoy, J., Hinton, J. W., Chamberlain, M. J., & Collier, B. A. (2014). Using dynamic brownian bridge movement modelling to measure temporal patterns of habitat selection. *Journal of Animal Ecology*, *83*, 1234–1243. https://doi.org/10.1111/1365-2656.12205

Calenge, C. (2023). adehabitatHR: Home range estimation [R package version 0.4.21]. https://CRAN.R-project.org/package=adehabitatHR

Christiansen, M. H., & Kirby, S. (2003, July 24). Language evolution: The hardest problem in science? In M. H. Christiansen & S. Kirby (Eds.), *Language evolution* (pp. 1–15). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199244843.003.0001

Clink, D. J., Bernard, H., Crofoot, M. C., & Marshall, A. J. (2017). Investigating individual vocal signatures and small-scale patterns of geographic variation in female bornean gibbon (hylobates muelleri) great calls. *International Journal of Primatology*, *38*(4), 656–671. https://doi.org/10.1007/s10764-017-9972-y

Clutton-Brock, T., & Manser, M. B. (2016). Meerkats: Cooperative breeding in the kalahari. In W. D. Koenig & J. L. Dickinson (Eds.), *Cooperative breeding in vertebrates: Studies of ecology, evolution, and behavior* (pp. 294–317). Cambridge University Press.

Collier, K., Bickel, B., Van Schaik, C. P., Manser, M. B., & Townsend, S. W. (2014). Language evolution: Syntax before phonology? *Proceedings of the Royal Society B: Biological Sciences*, *281*(1788), 20140263. https://doi.org/10.1098/rspb.2014.0263

Cover, T. M., & Thomas, J. A. (2006). *Elements of information theory* (2nd). Wiley-Interscience.

Dawbin, W., H., & Eyre, E. J. (1991). Humpback whale songs along the coast of western australia and some comparison with east coast songs. *Memoirs of the Queensland Museum*, *30*(2), 249–254.

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

Desai, N. P., Fedurek, P., Slocombe, K. E., & Wilson, M. L. (2022). Chimpanzee pant-hoots encode individual information more reliably than group differences. *American Journal of Primatology*, *84*(11), 1–23. https://doi.org/10.1002/ajp.23430

Dinno, A. (2015). Nonparametric pairwise multiple comparisons in independent groups using dunn's test. *The Stata Journal*, *15*(1), 292–300.

Driscoll, I., Briefer, E. F., & Manser, M. B. (2024). The role of neighbour proximity and context on meerkat close call acoustic structure. *Animal Behaviour*, *212*, 113–126. https://doi.org/10.1016/j.anbehav.2024.03.021

Drost, H.-G. (2018). Philentropy: Information theory and distance quantification with r. *Journal of Open Source Software*, *3*(26), 1–4. https://doi.org/10.21105/joss.00765

Engesser, S., & Manser, M. B. (2022). Collective close calling mediates group cohesion in foraging meerkats via spatially determined differences in call rates. *Animal Behaviour*, *185*, 73–82. https://doi.org/10.1016/j.anbehav.2021.12.014

Gall, G. E. C., & Manser, M. B. (2017). Group cohesion in foraging meerkats: Follow the moving 'vocal hot spot'. *Royal Society Open Science*, *4*(4), 170004. https://doi.org/10.1098/rsos.170004

Garcia Arasco, A., Manser, M., Watson, S. K., Kyabulima, S., Radford, A. N., Cant, M. A., & Garcia, M. (2022). Testing the acoustic adaptation hypothesis with vocalizations from three mongoose species. *Animal Behaviour*, *187*, 71–95. https://doi.org/10.1016/j.anbehav.2022.02.016

Gelfand, A. E., Fuentes, M., Hoeting, J. A., & Smith, R. L. (Eds.). (2017). *Handbook of environmental and ecological statistics*. Chapman; Hall/CRC. https://doi.org/10.1201/9781315152509

Gillies, S., van der Wel, C., Van den Bossche, J., Taves, M. W., Arnott, J., Ward, B. C., et al. (2024). *Shapely* (Version 2.0.5). https://doi.org/10.5281/ZENODO.5597138

Hahsler, M., Piekenbrock, M., & Doran, D. (2019). Dbscan: Fast density-based clustering with r. *Journal of Statistical Software*, *91*(1). https://doi.org/10.18637/jss.v091.i01

Hauser, M. D., Chomsky, N., & Fitch, W. T. (2010). The faculty of language: What is it, who has it, and how did it evolve? In R. K. Larson, V. Déprez, & H. Yamakido (Eds.), *The evolution of human language* (1st ed., pp. 14–42). Cambridge University Press. https://doi.org/10.1017/CBO9780511817755.002

Heinze, G., Wallisch, C., & Dunkler, D. (2018). Variable selection – a review and recommendations for the practicing statistician. *Biometrical Journal*, *60*(3), 431–449. https://doi.org/10.1002/bimj.201700067

Heller, R., Sander, A., Wang, C., Usman, F., & Dabelsteen, T. (2010). Macrogeographical variability in the great call of *Hylobates agilis*: Assessing the applicability of vocal analysis in studies of fine-scale taxonomy of gibbons. *American Journal of Primatology*, *72*(2), 142–151. https://doi.org/10.1002/ajp.20762

Hensel, A. L., Dobney, S. L., Doucet, S. M., Ryan Norris, D., Newman, A. E., Williams, H., & Mennill, D. J. (2022). Microgeographical variation in birdsong: Savannah sparrows exhibit microdialects in an island population. *Animal Behaviour*, *188*, 119–131. https://doi.org/10.1016/j.anbehav.2022.04.008

Heywood, D. I., Cornelius, S., & Carver, S. (2011). *An introduction to geographical information systems* (4th ed.). Pearson Prentice Hall.

Hollén, L. I., & Manser, M. B. (2006). Ontogeny of alarm call responses in meerkats, suricata suricatta: The roles of age, sex and nearby conspecifics. *Animal Behaviour*, *72*(6), 1345–1353. https://doi.org/10.1016/j.anbehav.2006.03.020

Horne, J. S., Garton, E. O., Krone, S. M., & Lewis, J. S. (2007). Analyzing animal movements using brownian bridges. *Ecology*, *88*(9), 2354–2363. https://doi.org/10.1890/06-0957.1

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

JetBrains. (2024). *PyCharm* (Version Community Edition 2024.1.1). https://www.jetbrains.com/pycharm/

Jolliffe, I. T., & Cadima, J. (2016). Principal component analysis: A review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *374*(2065), 1–16. https://doi.org/10.1098/rsta.2015.0202

Jordan, N. R., & Do Linh San, E. (2015, February 28). Suricata suricatta. the IUCN red list of threatened species in 2015 [Institution: International Union for Conservation of Nature]. https://doi.org/10.2305/IUCN.UK.2015-4.RLTS.T41624A45209377.en

Jordan, N. R., Cherry, M. I., & Manser, M. B. (2007). Latrine distribution and patterns of use by wild meerkats: Implications for territory and mate defence. *Animal Behaviour*, *73*(4), 613–622. https://doi.org/10.1016/j.anbehav.2006.06.010

Joseph, B. D. (2015). Language contact. In *International encyclopedia of the social & behavioral sciences* (2nd ed., pp. 300–306, Vol. 13). Elsevier. https://doi.org/10.1016/B978-0-08-097086-8.53016-5

Kahane, J.-P. (1997). A century of interplay between taylor series, fourier series and brownian motion. *Bulletin of the London Mathematical Society*, *29*(3), 257–279. https://doi.org/10.1112/S0024609396002913

Kassambara, A. (2023). *Rstatix: Pipe-friendly framework for basic statistical tests* [R package version 0.7.2]. https://CRAN.R-project.org/package=rstatix

Keighley, M. V., Langmore, N. E., Zdenek, C. N., & Heinsohn, R. (2017). Geographic variation in the vocalizations of australian palm cockatoos (*Probosciger aterrimus*). *Bioacoustics*, *26*(1), 91–108. https://doi.org/10.1080/09524622.2016.1201778

Kranstauber, B., Gall, G. E. C., Vink, T., Clutton-Brock, T., & Manser, M. B. (2020). Long-term movements and home-range changes: Rapid territory shifts in meerkats (J. Chapman, Ed.). *Journal of Animal Ecology*, *89*(3), 772–783. https://doi.org/10.1111/1365-2656.13129

Kranstauber, B., Kays, R., LaPoint, S. D., Wikelski, M., & Safi, K. (2012). A dynamic brownian bridge movement model to estimate utilization distributions for heterogeneous animal movement. *Journal of Animal Ecology*, *81*, 738–746. https://doi.org/10.1111/j.1365-2656.2012.01955.x

KRC. (n.d.). *Meerkat research* [Kalahari research centre and the kalahari meerkat project]. Retrieved January 1, 2024, from https://kalahariresearchcentre.org/

Krumm, J. (2021). Brownian bridge interpolation for human mobility? *Proceedings of the 29th International Conference on Advances in Geographic Information Systems*, 175–183. https://doi.org/10.1145/3474717.3483942

Kruskal, J. B. (1964a). Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika*, *29*(1), 1–27. https://doi.org/10.1007/BF02289565

Kruskal, J. B. (1964b). Nonmetric multidimensional scaling: A numerical method. *Psychometrika*, *29*(2), 115–129. https://doi.org/10.1007/BF02289694

Lin, A., Feng, J., & Kanwal, J. S. (2022). Geographic variation in social vocalizations of the great himalayan leaf-nosed bat, hipposideros armiger: Acoustic overflow across population boundaries. *Frontiers in Ecology and Evolution*, *10*, 948324. https://doi.org/10.3389/fevo.2022.948324

Manser, M. B., & Bell, M. B. (2004). Spatial representation of shelter locations in meerkats, suricata suricatta. *Animal Behaviour*, *68*(1), 151–157. https://doi.org/10.1016/j.anbehav.2003.10.017

Manser, M. B., Jansen, D. A., Graw, B., Hollén, L. I., Bousquet, C. A., Furrer, R. D., & Le Roux, A. (2014). Vocal complexity in meerkats and other mongoose species. In *Advances in the study of behavior* (pp. 281–310, Vol. 46). Elsevier. https://doi.org/10.1016/B978-0-12-800286-5.00006-7

Martin, K., Tucker, M. A., & Rogers, T. L. (2017). Does size matter? examining the drivers of mammalian vocalizations. *Evolution*, *71*(2), 249–260. https://doi.org/10.1111/evo.13128

Mausbach, J., Braga Goncalves, I., Heistermann, M., Ganswindt, A., & Manser, M. B. (2017). Meerkat close calling patterns are linked to sex, social category, season and wind, but not fecal glucocorticoid metabolite concentrations (E. Zimmermann, Ed.). *PLOS ONE*, *12*(5), 1–19. https://doi.org/10.1371/journal.pone.0175371

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

Oksanen, J., Simpson, G. L., Blanchet, F. G., Kindt, R., Legendre, P., Minchin, P. R., O'Hara, R. B., Solymos, P., Stevens, M. H. H., Szoecs, E., Wagner, H., Barbour, M., Bedward, M., Bolker, B., Borcard, D., Carvalho, G., Chirico, M., Caceres, M. D., Durand, S., . . . Weedon, J. (2022). Vegan: Community ecology package [R package version 2.6-4]. https://CRAN.R-project.org/package=vegan

Pagel, M. (2017). Q&a: What is human language, when did it evolve and why should we care? *BMC Biology*, *15*(1), 64. https://doi.org/10.1186/s12915-017-0405-3

Pebesma, E. (2018). Simple features for r: Standardized support for spatial vector data. *The R Journal*, *10*(1), 439–446. https://doi.org/10.32614/RJ-2018-009

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, *12*, 2825–2830.

QGIS Development Team. (2021). *QGIS Geographic Information System* (Version 3.22 Białowieża). QGIS Association. https://www.qgis.org

Quebec Centre for Biodiversity Science. (2023a). *Chapter 13 principal component analysis* [Workshop 9: Multivariate analyses in r]. Retrieved August 10, 2024, from https://r.qcbs.ca/workshop09/book-en/principal-component-analysis.html

Quebec Centre for Biodiversity Science. (2023b). *Chapter 6 redundancy analysis* [Workshop 10: Advanced multivariate analyses in r]. Retrieved February 23, 2024, from https://r.qcbs.ca/workshop10/book-en/redundancy-analysis.html

Quebec Centre for Biodiversity Science. (2023c). *Chapter 7 partial redundancy analysis* [Workshop 10: Advanced multivariate analyses in r]. Retrieved February 23, 2024, from https://r.qcbs.ca/workshop10/book-en/partial-redundancy-analysis.html

Quebec Centre for Biodiversity Science. (2023d). *Chapter 8 variation partitioning* [Workshop 10: Advanced multivariate analyses in r]. Retrieved February 23, 2024, from https://r.qcbs.ca/workshop10/book-en/variation-partitioning.html

R Core Team. (2024). *R: A language and environment for statistical computing* (Version 4.3.3). Vienna, Austria, R Foundation for Statistical Computing. https://www.R-project.org/

Ravignani, A., & Garcia, M. (2022). A cross-species framework to identify vocal learning abilities in mammals. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *377*(1841), 1–12. https://doi.org/10.1098/rstb.2020.0394

Reber, S. A., Townsend, S. W., & Manser, M. B. (2013). Social monitoring via close calls in meerkats. *Proceedings of the Royal Society B: Biological Sciences*, *280*(1765), 1–7. https://doi.org/10.1098/rspb.2013.1013

Rogerson, P. A. (2020). *Statistical methods for geography: A student's guide* (Fifth edition). SAGE.

RStudio Team. (2023). *RStudio: Integrated development for r.* (Version 2023.06.1.524). Boston, RStudio PBC. http://www.rstudio.com/

Ruch, H., Zürcher, Y., & Burkart, J. M. (2018). The function and mechanism of vocal accommodation in humans and other primates. *Biological Reviews*, *93*(2), 996–1013. https://doi.org/10.1111/brv.12382

Searcy, W. A. (2019). Animal communication, cognition, and the evolution of language. *Animal Behaviour*, *151*, 203–205. https://doi.org/10.1016/j.anbehav.2019.03.001

Stafford, K. M., Nieukirk, S. L., & Cox, C. G. (2001). Geographic and seasonal variation of blue whale calls in the north pacific. *J. Cetacean Res. Manage.*, *3*(1), 65–76. https://doi.org/10.47536/jcrm.v3i1.902

Telgarsky, R. (2013). Dominant frequency extraction. https://doi.org/10.48550/arXiv.1306.0103

Thomason, S. (2001). Language contact. In *International encyclopedia of the social & behavioral sciences* (pp. 8325–8329). Elsevier. https://doi.org/10.1016/B0-08-043076-7/03032-1

Thompson, B., & Smith, K. (2015). Evolution and language: Cultural transmission. In *International encyclopedia of the social & behavioral sciences* (pp. 357–363). Elsevier. https://doi.org/10.1016/B978-0-08-097086-8.81067-3

Thompson, M. (2019). *South african national land-cover 2018 report & accuracy assessment.* Department of Environmental Affairs & Department of Rural Development and Land Reform. Republic of South Africa. Retrieved February 16, 2024, from https://egis.environment.gov.za/sa_national_land_cover_datasets

Townsend, S. W., Hollén, L. I., & Manser, M. B. (2010). Meerkat close calls encode group-specific signatures, but receivers fail to discriminate. *Animal Behaviour*, *80*(1), 133–138. https://doi.org/10.1016/j.anbehav.2010.04.010

Townsend, S. W., Zöttl, M., & Manser, M. B. (2011). All clear? meerkats attend to contextual information in close calls to coordinate vigilance. *Behavioral Ecology and Sociobiology*, *65*(10), 1927–1934. https://doi.org/10.1007/s00265-011-1202-6

Van Den Wollenberg, A. L. (1977). Redundancy analysis an alternative for canonical correlation analysis. *Psychometrika*, *42*(2), 207–219. https://doi.org/10.1007/BF02294050

van Staaden, M. J. (1994). Suricata suricatta. *Mammalian Species*, (483), 1–8. https://doi.org/10.2307/3504085

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

Vernes, S. C., Kriengwatana, B. P., Beeck, V. C., Fischer, J., Tyack, P. L., Ten Cate, C., & Janik, V. M. (2021). The multi-dimensional nature of vocal learning. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *376*(1836), 1–15. https://doi.org/10.1098/rstb.2020.0236

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., Van Der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., … Vázquez-Baeza, Y. (2020). SciPy 1.0: Fundamental algorithms for scientific computing in python. *Nature Methods*, *17*(3), 261–272. https://doi.org/10.1038/s41592-019-0686-2

Watts, B. D., Mojica, E. K., & Paxton, B. J. (2015). Using brownian bridges to assess potential interactions between bald eagles and electrical hazards within the upper chesapeake bay: Brownian bridges and eagle hazards. *The Journal of Wildlife Management*, *79*(3), 435–445. https://doi.org/10.1002/jwmg.853

Whiten, A. (2021). The burgeoning reach of animal culture. *Science*, *372*, 1–7. https://doi.org/10.1126/science.abe6514

Wickham, H. (2016). *Ggplot2: Elegant graphics for data analysis* (Second edition). Springer international publishing.

Wickham, H. (2023). Stringr: Simple, consistent wrappers for common string operations [R package version 1.5.1]. https://CRAN.R-project.org/package=stringr

Widdig, A. (2017). Coefficient of relatedness. In J. Vonk & T. Shackelford (Eds.), *Encyclopedia of animal cognition and behavior* (pp. 1–3). Springer International Publishing. https://doi.org/10.1007/978-3-319-47829-6_700-1

Wright, S. (1922). Coefficients of inbreeding and relationship. *The American Naturalist*, *56*(645), 330–338. https://doi.org/10.1086/279872

Yuan, Y., Qiang, Y., Bin Asad, K. M., & Chow, E. (2020). Point pattern analysis (J. P. Wilson, Ed.). *Geographic Information Science & Technology Body of Knowledge*, *2020*. https://doi.org/10.22224/gistbok/2020.1.13

Živadinović, N. K. (2011). Multidimensional scaling: An introduction. In M. Lovric (Ed.), *International encyclopedia of statistical science* (pp. 878–879). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-04898-2_386

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 A

Detailed Results of Dunn's Test

Table A.1: Dunn's test p-values for five PCs for all pairs of individuals and groups. The five PCs are ordered from top to bottom per block of individual or group, i.e., PC1 being the top row. Dark blue indicates significant differences for $\alpha = 0.05$ and light blue for $\alpha = 0.1$. Red indicates non-significant differences.

		VUKF018	**VZUF014**	**VRRM181**	**VRRM182**	
		0.96	$< 2 * 10^{-16}$	$1.60 * 10^{-11}$	$< 2 * 10^{-16}$	**VEKF017**
		$< 2 * 10^{-16}$	$< 2 * 10^{-16}$	0.03	0.3	
		1.0	$1.75 * 10^{-9}$	1.0	$1.08 * 10^{-3}$	
		$< 2 * 10^{-16}$	$< 2 * 10^{-16}$	$< 2 * 10^{-16}$	0.11	
		1.0	0.14	1.0	1.0	
EK	$< 2 * 10^{-16}$		$< 2 * 10^{-16}$	$3.27 * 10^{-7}$	$6.95 * 10^{-13}$	**VUKF018**
	0.21		1.0	$2.33 * 10^{-11}$	$3.08 * 10^{-16}$	
	$< 2 * 10^{-16}$		$9.65 * 10^{-11}$	1.0	$2.53 * 10^{-4}$	
	$2.86 * 10^{-3}$		$< 2 * 10^{-16}$	$1.32 * 10^{-5}$	$< 2 * 10^{-16}$	
	$8.52 * 10^{-3}$		$1.27 * 10^{-4}$	0.1	0.16	
RR	$< 2 * 10^{-16}$	0.01		0.31	1.0	**VZUF014**
	1.0	$6.95 * 10^{-3}$		$< 2 * 10^{-16}$	$< 2 * 10^{-16}$	
	$< 2 * 10^{-16}$	1.0		$5.43 * 10^{-7}$	0.29	
	$1.12 * 10^{-10}$	0.06		0.02	$4.60 * 10^{-14}$	
	1.0	0.09		1.0	0.84	
UK	$7.50 * 10^{-9}$	$< 2 * 10^{-16}$	$< 2 * 10^{-16}$		1.0	**VRRM181**
	$2.83 * 10^{-4}$	$1.78 * 10^{-11}$	$1.51 * 10^{-4}$		1.0	
	$< 2 * 10^{-16}$	1.0	1.0		0.02	
	$7.84 * 10^{-13}$	$7.31 * 10^{-3}$	1.0		$< 2 * 10^{-16}$	
	1.0	$8.46 * 10^{-5}$	0.28		1.0	
ZU	$< 2 * 10^{-16}$	0.01	1.0	$< 2 * 10^{-16}$		
	$3.22 * 10^{-7}$	$5.19 * 10^{-16}$	$2.55 * 10^{-8}$	0.75		
	$7.93 * 10^{-13}$	$6.15 * 10^{-6}$	$3.00 * 10^{-14}$	$3.35 * 10^{-13}$		
	$8.23 * 10^{-4}$	1.0	$3.95 * 10^{-3}$	$1.44 * 10^{-4}$		
	$1.20 * 10^{-6}$	1.0	$7.50 * 10^{-6}$	$2.79 * 10^{-12}$		
	BB	**EK**	**RR**	**UK**		

Personal Declaration

I hereby declare that the submitted thesis is the result of my own, independent work. All external sources are explicitly acknowledged in the thesis.

Wolfhausen, 31.01.2025

A.Hehli

Antonia Hehli