Processing of Graded Signaling Systems

Dissertation

for the award of the degree

"Doctor rerum naturalium"

of the Georg-August Universität Göttingen

within the doctoral program

Sensory and Motor Neuroscience (GGNB)

of the Georg-August University School of Science (GAUSS)

submitted by

Philip Wadewitz

from Bielefeld, Germany

Göttingen 2015

Doctoral Thesis Commitee:

Prof. Dr. Julia Fischer (Advsior, First Referee)
Cognitive Ethology Laboratory
German Primate Center
Kellnerweg 4, 37077 Göttingen

Prof. Dr. Fred Wolf (Second Referee)
Theoretical Neurophysics
Max Plank Institute for Dynamics and Self-Organisation
Fassberg 18, 37077 Göttingen

Prof. Dr. Burkhard Morgenstern
Department of Bioinformatics
Institute for Microbiology and Genetics
Goldschmidtstr. 1, 37077 Göttingen

Further Members
of the Examination Board:

Prof. Dr. Eckhard Heymann
Behavioral Ecology and Sociobiology Unit
German Primate Center
Kellnerweg 4, 37077 Göttingen

Prof. Dr. Alexander Gail
Sensorimotor Group
German Primate Center
Kellnerweg 4, 37077 Göttingen

Prof. Dr. Andreas Stumpner
Department of Cellular Neurobiology
Schwann-Schleiden-Forschungszentrum
Julia-Lermontowa-Weg 3, 37077 Göttingen

Date of Oral Examination: 04.12.2015

Declaration

I herewith declare that I wrote this dissertation independently and that I only used the indicated sources and aids. Places and passages inferred from other factories literally or according to the sense are marked clearly.

Göttingen, 29.09.2015

Philip Wadewitz

Table of Contents

Summary

Vocal repertoires of nonhuman animals and especially of terrestrial mammals are often characterized by their relatively small size of innate vocal types which can show considerable variation in acoustic structure. To understand the proximate and ultimate causes that shape the structure of acoustic communication systems in animals, an objective characterization of the vocal repertoire of a given species is critical, as it provides the foundation for comparative analyses among individuals, populations, and taxa.

The common approach to characterize vocal repertoires is by using unsupervised clustering algorithms to identify call types and to define a repertoire's size. Progress in the field has been hampered by a lack of standard in methodology which can lead to an arbitrary decision about the size of a species' repertoire. To investigate whether this difficulty is based on the used methodology or whether it is intrinsic to the acoustic structure of a given repertoire, the major aim of my dissertation was to investigate and advance the available methods in the field. To do so, I focused on three main aspects of a vocal repertoire analysis: (1) how is the analysis affected by the input parameters, i.e. the acoustic features that are used; (2) how can we quantify the acoustic variation within and between different vocal types; (3) what is the impact of data set composition, i.e. the call recordings that are being used in the analysis.

In the first part of my thesis, I re-analyzed recordings from wild chacma baboons (*Papio ursinus*) to test the impact of the number and type of acoustic features that are included in the analysis. To do this, I constructed data sets with the same 912 call exemplars but with a varying number of acoustic features to describe these calls. To this end, I had three data sets with 9, 38, and 118 acoustic features as well as a data set with 19 factors derived from a principal component analysis. By comparing and validating the resulting classifications of two clustering algorithms, namely k-means and hierarchical Ward's clustering, I could show that the data sets with a higher number of acoustic features lead to better clustering results than data sets with only a few features. I further showed that factors are not suited to cluster the chacma baboon's calls. None of the applied clustering algorithms gave strong support to a specific cluster solution. Since there was substantial acoustic variation within and between the different call types, I applied an approach based on fuzzy logic that we developed to describe the gradation within vocal repertoires and which provides a quantitative description of the gradation within the chacma baboon's repertoire.

To investigate the impact of potential evolutionary forces that shape a species' communication system, comparative studies that quantify the differences in these systems between different species are necessary. In the second part of my thesis, I strove towards such a quantitative comparison by systematically comparing the vocal repertoire of the chacma baboon with the vocal repertoire of the Barbary macaque, *Macaca sylvanus.* I quantified the gradation within and between different call types of both species with an extended version of the fuzzy clustering approach that was used to characterize the chacma baboon's repertoire in the first part of this thesis. The analysis confirmed the findings of previous studies by showing that the repertoire of the Barbary macaque exhibits a significant larger amount of gradation within and between different call types. An important aspect of this method is that it allows the quantification of gradation irrespective of the number of call types by circumventing the problem to settle on one cluster solution when several solutions are largely equivalent.

In the third part of my thesis, I investigated the influence of the data set composition that is used for the analysis of vocal repertoires. Specifically, I was interested in the effects of size- and arousal based differences in the recorded animals and their impact on clustering results. The differences in body size and arousal were simulated with a software-based model that simulates muscle characteristics of the larynx and vocal tract anatomy. With this model I created pseudo repertoires of three distinct baboon call types that varied in subglottic pressure levels (as a proxy of arousal-based differences) and vocal fold and vocal tract characteristics (size-based differences). The preliminary results show that whereas differences in subglottic pressure levels had a minor impact on the characteristics of vocal repertoires and all three call types can be clearly separated from each other, differences in body size can hamper classification and characterization of call types.

In conclusion, I investigated several aspects that have to be taken into account when analyzing vocal repertoires. The composition of the data sets as well as the selection of acoustic features that are used in the analysis can both have a profound effect on the classification outcome and on cluster determination. To overcome the often arbitrary decision about a species repertoire size I developed a method that is useful to describe the gradation within and between different call types over several cluster solutions and therefore circumvents the problem to settle on one specific solution. In addition, the method allows a systematic comparison of different species' vocal repertoires, a prerequisite to investigate potential driving forces in signal evolution.

Zusammenfassung

Vokale Repertoires von nichtmenschlichen Tieren und im Besonderen von terrestrischen Säugetieren sind häufig durch eine relativ kleine Anzahl angeborener Ruftypen gekennzeichnet, welche ein hohes Maß an akustischer Variation aufweisen können. Objektive Beschreibungen von vokalen Repertoires werden benötigt, um die proximaten und ultimaten Faktoren die die Struktur akustischer Kommunikationssysteme beeinflussen zu verstehen, da diese die Grundlage für vergleichende Analysen zwischen Individuen, Population und Taxa bilden.

Üblicherweise werden vokale Repertoires durch nichtüberwachte Clusterverfahren beschrieben, da hiermit Ruftypen identifiziert und die Größe eines Repertoires definiert werden kann. Der Fortschritt in diesem Forschungsfeld wird jedoch durch eine nichtstandardisierte Methodik erschwert, was oft zu einer arbiträren Entscheidung bezüglich der Größe eines vokalen Repertoires führt. Um zu überprüfen ob diese Problematik auf der verwendeten Methodik beruht, oder ob eine klare Einteilung durch die akustische Variation in den analysierten Datensätzen nicht möglich ist, lag der generelle Fokus meiner Dissertation auf der Überprüfung und Weiterentwicklung der vorhanden Klassifizierungsmethoden. Im Speziellen untersuchte ich drei elementare Aspekte der Analyse von vokalen Repertoires: (1) welchen Einfluss hat die Auswahl der akustischen Parameter die genutzt werden um die Rufstruktur zu beschreiben; (2) wie kann die akustische Variation innerhalb und zwischen den verschieden Ruftypen quantifiziert werden; (3) welchen Einfluss hat die Zusammensetzung des Datensatzes, sprich der Rufaufnahmen die in der Analyse verwendet werden.

In dem ersten Teil meiner Arbeit analysierte ich Rufaufnahmen von freilebenden Bärenpavianen (*Papio ursinus*), um die Auswirkungen der Auswahl akustischer Parameter, die in die Analyse einfließen, zu untersuchen. Hierfür erstellte ich Datensätze derselben 912 Rufaufnahmen, variierte jedoch die Anzahl der akustischen Parameter mit der die Rufe beschrieben werden. Insgesamt erstellte ich drei Datensätze mit jeweils 9, 38 und 118 akustischen Parametern, sowie einen Datensatz mit 19 Faktoren, die ich durch eine Hauptkomponentenanalyse gewonnen hatte. Durch den Vergleich und die Validierung der re-

sultierenden Ergebnisse der zwei Clusteralgorithmen (k-means und hierarchisches Ward's Clustering) konnte ich zeigen, dass Datensätze mit einer höheren Anzahl an akustischen Parametern zu besseren Clusterergebnissen führen als Datensätze mit weniger akustischen Parametern. Des Weiteren zeigte ich, dass der Datensatz, der auf Faktoren basiert, nicht geeignet ist, um die Rufe der Bärenpaviane zu klassifizieren. Keiner der angewandten Clusteralgorithmen fand eine eindeutige Lösung bezüglich der Gesamtzahl von Ruftypen. Da der Datensatz jedoch eine substantielle akustische Variation innerhalb und zwischen den verschiedenen Ruftypen aufwies, wendete ich in einem zusätzlichen Analyseschritt ein Clusterverfahren an, welches auf den Prinzipien der Fuzzy-Logik beruht und das von uns entwickelt wurde, um die Variation innerhalb von vokalen Repertoires zu beschreiben. Das Ergebnis dieser Analyse liefert eine quantifizierte Beschreibung dieser Variation innerhalb des vokalen Repertoires des Bärenpavians.

Um den Einfluss von evolutionären Faktoren zu untersuchen, die das Kommunikationssystem einer Art beeinflussen, werden vergleichende Studien zwischen verschiedenen Arten benötigt. Diese Studien müssen in der Lage sein, die strukturellen Unterschiede in diesen Kommunikationssystemen zu quantifizieren. Aus diesem Grund verglich ich im zweiten Teil meiner Arbeit systematisch das vokale Repertoire des Bärenpavians mit dem vokalen Repertoire des Berberaffens, *Macaca sylvanus*. Die Variation innerhalb und zwischen verschiedenen Ruftypen beider Arten wurde mit einer erweiterten Version der Methodik quantifiziert, die bereits zur Beschreibung des Bärenpavian Repertoires im ersten Teil dieser Arbeit entwickelt und genutzt wurde. Die Methodik ermöglichte die Unterschiede in der Variation zwischen den beiden Repertoires zu quantifizieren. Die Analyseergebnisse bestätigten Ergebnisse früherer Studien, welche zeigen konnten, dass Berberaffen ein hohes Maß an akustischer Variation innerhalb und zwischen verschiedenen Ruftypen aufweisen. Ein wichtiger Aspekt dieser Methode ist, dass sie es ermöglicht die Variation innerhalb eines Repertoires ungeachtet der Gesamtzahl der Ruftypen zu quantifizieren.

Im dritten Teil meiner Arbeit untersuchte ich den Einfluss der Zusammenstellung des Datensatzes, der für die Analyse eines vokalen Repertoires genutzt wird. Speziell war ich an dem Einfluss interessiert, den die Körpergröße und der Erregungszustand des

aufgenommenen Tieres auf die Analyseergebnisse spielen. Körpergröße und Erregungszustand wurden durch ein software-basiertes Modell variiert, welches die Eigenschaften von Kehlkopfmuskeln sowie die Anatomie des Vokaltraktes simuliert. Mit Hilfe dieses Models habe ich Pseudo-Repertoires von drei distinkten Ruftypen erstellt, die akustische Variationen aufweisen, die durch Variation des Anpressdrucks in der Lunge (Erregungszustand) sowie Variation in Eigenschaften der Stimmlippen und des Vokaltraktes (Körpergröße) hervorgerufen werden. Die vorläufigen Ergebnisse zeigen, dass während Variation im Erregungszustand einen eher untergeordneten Einfluss auf die Charakteristik eines vokalen Repertoires hat und die drei Ruftypen weiterhin klar voneinander unterschieden werden können, Variationen, die durch Größenunterschiede hervorgerufen werden die Klassifizierung und Charakterisierung von Ruftypen deutlich erschweren kann.

Zusammenfassend habe ich verschiedene analytische Aspekte untersucht, die maßgebliche Auswirkungen auf die Ergebnisse einer vokalen Repertoire Analyse haben können und eine Methode entwickelt, um die akustische Variation innerhalb eines Repertoires zu quantifizieren. Die Zusammenstellung der Datensätze sowie die Auswahl der akustischen Parameter die für die Analyse genutzt werden, können die Bestimmung der Repertoiregröße erheblich erschweren. Um zu vermeiden, dass die Repertoiregröße arbiträr festgelegt wird, kann die von mir entwickelte Methodik angewendet werden, in welcher die akustische Variation eines Repertoires über mehrere mögliche Clusterlösungen beschrieben und die Veränderung der Variation quantifiziert wird. Zusätzlich erlaubt die Methodik einen systematischen Vergleich von vokalen Repertoires verschiedener Arten, welcher eine Grundvoraussetzung darstellt, um die evolutionären Faktoren die die Struktur von Kommunikationssystemen beeinflussen zu untersuchen.

Acknowledgments

I offer my profound thanks to:

Julia Fischer for being a thoughtful and enthusiastic supervisor. For taking me into her academic family and for giving me so much freedom to follow my interests throughout this project. Kurt Hammerschmidt for his excellent methodological and analytical guidance throughout my whole thesis and whose door was always open. Fred Wolf for his interest and advice and for always challenging the status quo. Burkhard Morgenstern for his support as a member of my Thesis Committee and Alexander Gail, Eckhard Heymann, and Andreas Stumpner for being members of my examination board.

The members, past and present, of the Cognitive Ethology Laboratory, with special thanks to Laura Almeling, Matthis Drolet, Rebecca Jürgens, Urs Kalbitzer, Matthias Klapproth, Gisela Kopp, Peter Maciej, and Tabitha Price conversations with whom helped shaping the ideas that went into this thesis and who became friends over the last years.

Demian Battaglia and Annette Witt for their great support on the computational and mathematical aspects of this project and without whom this thesis would not be where it is now.

Ingo Titze and Tobias Riede for giving a stranger the opportunity to come to Salt Lake City and work in their lab. For their hospitality and time to introduce me to the physics of sound production and for their creative input. And Anil Palaparthi for helping me with the modelling.

Mechthild Pohl and Ludwig Ehrenreich for their invaluable help in all bureaucratic and technical questions and for providing me with so many sweets and coffee, even late at night.

My friends in and outside the German Primate Center. Special thanks go to Pascal Marty and the entire volleyball team for great nights out and all kind of sportive activities that helped to stay in shape while spending the majority of the time in front of a computer screen, and to Mariam Lazizi for introducing me to the local music scene and for being a true friend.

Kirsten Spindeldreier for coming to Göttingen even though there are no vineyards around. For the wonderful time we spent together, for her kindness, and for her patience and understanding when things were busy.

And most of all to my entire family, my parents Anke and Dietmar, and my siblings Anna, Miriam, and Benjamin, for their constant encouragement and support of my doing "what's right for me".

The research reported here was funded by the Federal Ministry of Education and Research, and by travel grants from the Göttingen Graduate School for Neurosciences, Biophysics and Molecular Biosciences (GGNB).

1 General Introduction

Human language is strikingly different from communication systems in other species. Whereas human language applies conventional rules about the referential content of words and uses syntactical rules and recursion to generate limitless meaning (Hauser et al. 2002; Fischer 2010), nonhuman animals do not show these key components of human language (or only to a very limited degree). The importance and large interest in language evolution led to a number of studies that explore human language evolution based on hypotheses regarding the evolution of symbolic communication and syntax (Nowak et al. 2000; Komarova et al. 2001; Chater and Manning 2006; Chater et al. 2009). To investigate the evolution of communication at a more general level however, comparative studies of different species' communication systems are necessary.

In order to compare such systems, detailed descriptions of nonhuman animals' vocal repertoires are a prerequisite. Studies to describe vocal repertoires are manifold and investigate several aspects such as the number of calls that a species produces or the acoustic variation within and between different call types. Over the last decades, the upsurge of computer technology has given researchers more sophisticated software-based tools to analyze the fine differences in acoustic structure of calls. However, the available tools require several decisions of the researcher during the analytical process, which often impede objectivity of such studies and hence hinder comparability. Therefore, a major goal of bioacoustics research is to find solutions to overcome these limitations and advance methodology to generate detailed and quantitative descriptions of nonhuman animal communication systems.

In the following sections of this chapter, I will first introduce the basic principles of vocal production in terrestrial mammals and highlight differences in the anatomy of human and primate vocal organs and neural circuits and their implications for speech production. Although the main focus in behavioural bioacoustics research lies on the ultimate evolutionary explanations for the structure of these systems, a basic knowledge of sound production mechanisms is crucial to understand physical factors and potential constraints that can influence the evolution of vocal communication systems. Following from this, I will discuss signal structure and external as well as internal factors that can influence it, before I turn towards signal repertoire design and informational content of signals. I will then summarize the most common analytical tools in bioacoustics research and their applications before I finally outline the overall aim of my thesis and introduce the conducted studies.

1 Sound Production

1.1 Anatomy of the Vocal Organs

The basic mechanisms of sound production in humans and other terrestrial mammals are well explored and show a high level of similarity (Taylor and Reby 2010). Air that is exhaled from the lungs by muscle contraction drives oscillations of the vocal folds which are located in the larynx. Depending on lung capacity and strength of muscle contraction, duration and amplitude of the generated sound can be altered. Since the vocal folds are associated with several laryngeal muscles and cartilages, the fundamental frequency (i.e. pitch) of the produced sound can be changed by lengthening or shortening the vocal folds. If the vocal folds are lengthened, their oscillation rate triggered by the airflow is increased and therefore fundamental frequency is increased. The shorter the vocal folds, the lower their oscillation rate and the lower the fundamental frequency. The generated acoustic energy then passes through the vocal tract where it is filtered before it exits the vocal tract through the nostrils and lips. This filtering process is accomplished by a series of bandpass filters, termed formants. The formants modify the sound that is emitted by allowing only a narrow range of frequencies to pass unhindered. Formants are determined by the length and shape of the vocal tract and are modified during vocalizations by movement of the articulators like lips, tongue, and soft palate (Fitch 2000a).

All terrestrial mammals that have been studied produce sounds in essentially this way, using similar larynges and vocal anatomy. One striking difference in the vocal tract anatomy of humans and most other mammals is the position of the larynx. Whereas in most mammals the larynx is located high enough in the throat to enable simultaneous breathing and swallowing, the lowered human larynx allows the tongue to move both vertically and horizontally within the vocal tract and therefore greatly expands the phonetic repertoire in humans (Lieberman et al. 1969).

1.2 Neural Circuits of Vocal Production

Most terrestrial mammals exhibit a common neurobiological circuitry for volitional vocal control. The analyses of the neurobiological control mechanisms engaged in phonatory functions relied predominantly on brain stimulation studies on squirrel monkeys (Jürgens and Ploog 1970; Gonzales-Lima 2010). Vocal control consists of two hierarchically organized pathways. One of the pathways that controls the readiness to vocalize, centres around the periaqueductal gray (PAG) in the midbrain. The PAG gets input from motivation-controlling regions, sensory structures, motor areas, and arousal-related systems and seems to gate vocalizations in response to emotions such as fear and aggression (Ackermann et al. 2014). After integration of these input signals, the PAG projects into the reticular formation of pons and medulla oblongata, including a vocal pattern generator, which innervate the phonatory motor neurons and finally the vocal tract muscles (Hage and Jürgens 2006). The second pathway that is responsible for the production of innate vocal patterns runs from the motor cortex via the reticular formation to the phonatory motor neurons. Before the final motor commands are generated, two feedback loops provide the motor cortex with pre-processed information from the basal ganglia and the cerebellum (Jürgens 2009). However, the role of basal ganglia and cerebellum in motor aspects of vocal behaviour are still not fully understood (Ackermann et al. 2014).

An additional pathway that directly links regions in the primary motor cortex with the phonatory motor neurons has so far only be found in humans and three distantly related groups of birds (parrots, hummingbirds, and songbirds) (Nottebohm 1972; Janik and Slater 1997). Sparse projections have also recently been identified in mice (Arriaga and Jarvis 2013). This direct link is assumed to enable these species to modify the acoustic structure of produced sounds, including imitation and improvisation, called vocal production learning (Hammerschmidt et al. 2015). Other mammals including bats, cetaceans, seals, and elephants also show vocal learning, however, their brain pathways for learned vocalizations have not yet been studied (see Jarvis 2007 for a review). Notably, nonhuman primates do not have the ability of vocal production learning and the structure of their vocalizations is largely innate.

2 Signal Design

The evolution of signal structure is influenced by a range of ecological factors. These factors can generally be assigned to one of the two opposing components of signal selection, the efficacy and the strategic component. From the signaler's perspective, a signal should influence the receiver in a way that benefits the signaler and at the same time should be energetically cheap. The efficacy component therefore selects for signal structure that provides the optimal trade-off between costs and benefits of the signal (Krebs and Davies 1997). From a receiver's perspective, a signal should be a source of information that benefits the receiver by adjusting its behavior in response. As receivers are under strong selection to only respond to reliable signals, the strategic component of signal structure evolution ensures that signalers pay additional costs that guarantee honest signals (Bradbury and Vehrencamp 2011). In the following subsections I will discuss some of the numerous ecological and biological factors that influence signal structure with respect to these two components of signal selection.

2.1 Signaler Anatomy and Phylogenetic Constraints

A signaler's body size and structure of the vocal apparatus are among the most salient biological factors that influence the acoustic structure of a signal. As I discussed in the previous section, sound production results from a process of three steps, starting with air compression in the lungs, glottal wave generation at the larynx and subsequent filtering in the supralaryngeal vocal tract. Since the variability of signal structure is constrained by the physical properties of these anatomical structures, receivers may be able to use features of the signal to reliably gain information about the physical attributes of the signaler (Fitch and Hauser 1998). This is of particular importance since many terrestrial mammals use acoustic signals in aggressive interactions and mate attraction (Clutton-Brock and Albon 1978) and the outcome of these interactions can depend strongly on physical attributes such as body size, sex, or age (Taylor and Reby 2010).

Generally, reliable cues to physical attributes of the signaler can originate at all three structures, the lungs, larynx, and vocal tract. Since in mammals the lungs occupy most of the thorax, their size is closely related to body size. Acoustic features that are directly linked to body volume (such as signal duration) should therefore be reliable cues for body size. Although there has been no experimental test of this hypothesis to date, MacLarnon and Hewitt showed that primates with air sacs, which are assumed to function as "accessory lungs", have longer signal durations than those without air sacs (MacLarnon and Hewitt 1999).

At the level of the larynx, vocal fold characteristics determine the fundamental frequency (F0) of produced signals. F0 does not seem to be a reliable indicator for body size, since the growth of the vocal folds is not constrained by an individual's body size (Fitch 1997; Riede and Titze 2008). However, several studies have shown that during development, F0 can be correlated with body size (Rendall et al. 2005; Pfefferle and Fischer 2006) and that, among females, F0 can be a reliable indicator of body size even within age classes (Pfefferle and Fischer 2006). Furthermore, in some species F0 has been found to be negatively correlated with reproductive success (Reby and McComb 2003).

At the level of the vocal tract, it has been argued that vocal tract size is constrained by skeletal structures (Fitch 2000,b) and formant dispersion should therefore be a reliable cue of body size (Fitch and Reby 2001). In support of this hypothesis, several studies have found a direct negative correlation of formant dispersion and body size (e.g. in domestic dogs: Riede and Fitch 1999). However, others argue that formant dispersion might not be as reliable as hypothesized since formant dispersion can be altered by lip configuration and jaw movements and, in some species, by lowering down the larynx to the sternum, hence increasing vocal tract length (Pfefferle and Fischer 2006).

Since signal structure is highly influenced by these anatomical features, the variation in signal structure is limited to a small portion of adaptive space that is explorable through genetic recombination and mutation (Fitch and Hauser 1998). Especially in primates, where vocal signal structure is largely innate, vocalizations are expected to represent strong phylogenetic traits. Although studies that systematically compare signal structure and genetic relatedness are rare, existing studies on crested gibbons and leaf monkeys showed a high correlation between signal structure and genetic similarity (Thinh et al. 2011; Meyer et al. 2012). In addition to these phylogenetic traits, several studies have found correlations of signal structure and geographic distance (Geissmann and Nijman 2006), and geographic distance and genetic similarity between populations of the same species (Meyer et al. 2012).

2.2 Physical Properties of the Habitat

As I have discussed in the last section, a signal can be a reliable indicator of several characteristics of the signaler such as its sex, size, fighting ability, or identity. Numerous studies have shown that the distance of signal propagation has profound effects on frequency-dependent features of a signal (e.g. in Maciej et al. 2011). Hence, particularly the structure of long-distance vocal signals should underlie strong environmental selective pressures to minimize propagation losses ("acoustic adaptation hypothesis": Morton 1975).

A strong focus in investigating environmental effects on sound structure is based on the comparison between signal structures of species that inhabit closed and open habitats. Since open habitats provide more variable conditions for sound propagation (Morton 1975) and visual signals of communication can support vocal signals, selection pressure is assumed to be stronger in closed than in open habitats. These environmental-related variations might affect several characteristics of signals (Ey and Fischer 2009), such as signal duration (longer signals in closed habitats increase the chance of detection), signal repetition rate (lower repetition rate in closed habitats avoid reverberation), frequency modulation (lower modulation in closed habitats since transmission is less consistent), or frequency range (lower range in closed habitats since high frequencies experience stronger attenuation). However, whereas some studies support the hypothesis that signals of a species show higher propagation levels under environmental conditions that represent the habitat of the species (e.g. in Japanese macaques: Tanaka et al. 2006) other studies did not find this trend (e.g. in marmosets: Daniel and Blumstein 1998).

Furthermore, environmental-related variations have been hypothesized to impact the structure of a species' entire vocal repertoire. This point is a central part of my work and will be described in more detail in section 1.3.

2.3 Motivational and Affective State of the Signaler

Whereas the physical properties of a habitat have a significant impact on the structure of signals that are used over large distances, structure of signals used in short-distance communication is much less influenced by the selective pressures of habitat characteristics. Nonetheless, selection for effective detection on the structure of short-distance signals exists. According to Morton (1977), the acoustic structure of a signal varies with the signaler's motivational state of fear and aggression. Whereas signals of an aggressive individual are assumed to be characterized by low frequency and broad bandwidth, signals of a fearful individual are characterized by high frequency and narrow bandwidth. Morton assumed that, since larger individuals can produce lower-frequency sounds and larger individuals often win aggressive encounters with smaller individuals, selective pressures act to lower the frequency of vocal threat signals. In contrast, high-frequency calls with narrow bandwidth of fearful animals symbolize small size, indicating appeasement and therefore reducing the likelihood of being attacked during aggressive encounters (Owings and Morton 1998). As aggressive and fearful signalers benefit from the coupling of signal structure and motivational state by making the signals clearly distinguishable, the selective pressures that lead to the divergence of signal structures can be assigned to the efficacy component of signal selection. Evidence for the validity of Morton's motivation-structural code comes from studies on a broad range of vocalizing taxa (e.g. in canids: Brady 1981), or nonhuman primates: Gouzoules and Gouzoules 2000; Fichtel et al. 2001). However, contradicting results have been found and doubts on the general validity of the model across call types have been raised (Cheney and Seyfarth 1990; Hauser 1993).

In humans, preverbal vocalizations like cries and moans that are given in emotional negative situations show higher frequency ranges, higher peak frequencies and longer duration than in emotional positive situations Scheiner et al. (2002). Similarly, in squirrel monkeys calls that are given in aversive situations differ structurally from calls given in pleasant situations (Jürgens and Pratt 1979) by being noisier and having higher peak frequencies. A correlation between aversiveness and peak frequency has also been shown in other species such as pigs and Barbary macaques (Hammerschmidt and Fischer 2008).

3 Signal Repertoires

After I have discussed several factors that can influence signal structure, I will now discuss how all signals used by a given species make up its vocal repertoire, factors that can influence repertoire structure and the potential informational content of a repertoire.

3.1 Signal Repertoire Structure

One of the prominent views on species' repertoires is that signal receivers only gain information from signals if the signaler is sufficiently consistent in emitting a specific signal when a certain condition is true. The mapping between signals and conditions is termed the coding scheme of a species and the sum of all signals makes up a species' vocal repertoire (Bradbury and Vehrencamp 2011).

Signal repertoires can be characterized by the structural variation within and between different signals. If the signals that constitute a signal repertoire are individually distinct and show no structural intermediates, the signal repertoire is considered being discrete. If structural variation occurs and intermediate structures between different signals can be found, the signal repertoire is considered being continuous or graded. For signal repertoires to be discrete, acoustic features of different signals must have separated values so that receivers can easily discriminate between them and assign each to an expected category (see Bradbury and Vehrencamp 2011). In a graded signal repertoire in contrast, signals can vary in one or more features on a continuous scale and therefore the alternatives of signals are potentially infinite in number. However, since variations in signal features have to vary with some minimal resolution to be discriminable by signal receivers, even graded signal repertoires are perceptually finite (categorical response) and many species categorize continuous signals into meaningful categories (categorical perception) (reviewed in Fischer 2006).

For vocal repertoires, several factors have been assumed to have a major impact on the gradation within a species' repertoire structure. Marler suggested that in species that live in habitats that restrict visual access between signaler and receiver and/or show high background noise (such as dense rainforest), discrete repertoires should have evolved to avoid signal misinterpretation (Marler 1975). On the other hand, species that live in open habitats with visual access to each other (such as savannah), graded repertoires should have evolved since the integration of visual signals could be used to avoid misinterpretation. For the same reason, within a species graded repertoire structures have been assumed to occur in close-range signals, whereas signals that are used over long distances should show a more discrete structure (Marler 1967). Marler further assumed that species that live in single-male groups should have evolved discrete signals since single males require loud, unambiguous signals to defend and influence their group (Marler 1976). Another factor that has been assumed to influence vocal repertoire structure is predation (Cheney and Seyfarth 1990; Fischer and Hammerschmidt 2001). In species with predator-specific defense strategies, alarm calls that are easily discriminable by signal receivers should evolve.

In nonhuman primates, graded and discrete vocal repertoires have been described in a number of species (graded: Barbary macaques, *Macaca sylvanus* (Hammerschmidt and Fischer 1998); bonobos, *Pan paniscus* (de Waal 1988); rhesus macaques, *Macaca mulatta* (Rowell and Hinde 1962); and Japanese macaques, *Macaca fuscata* (Green 1975) - discrete: putty-nosed monkeys, *Cercopithecus nictitans* (Arnold and Zuberbühler 2006); blue monkeys, *Cercopithecus mitis* (Papworth et al. 2008), and Diana monkeys, *Cercopithecus* *diana* (Zuberbühler et al. 1997)). As Kennan and colleagues point out however, labelling whole repertories as being either discrete or graded often represents an oversimplification, since gradation can occur within and between call types, and call types may vary to different degrees (Keenan et al. 2013). Whereas between-call-type variation might be dependent on the call's function, within-call-type variation could be linked to an animal's general affective state (Fischer et al. 1995; Manser 2001). Within this general affective state, similar situations can potentially evoke slightly different forms of excitement or fear, which can then relate to dissimilar acoustic structures within call types (Fischer et al. 2001). The importance to differentiate between these different forms of gradation, however, is neglected in most studies on vocal repertoires.

Despite the widespread usage of graded signals, receivers often assign perceived signals to discrete categories, even when signalers emit continuous signals. This phenomenon, called categorical perception, was first described by Liberman et al. (1957) who analyzed the perception of the human spoken phonemes /ba/ and /pa/ (there is no continuous perception although the two phonemes represent an acoustic continuum). Whereas Liberman believed that categorical perception is special to human speech, several studies have shown that categorical perception of continuous signals is a widespread phenomenon across taxa and can be found in insects, rodents, birds, and nonhuman primates (see Fischer 2006 for a review). The widespread presence of categorical perception has led to the question of why receivers would give up potential information that is encoded in continuous signals by lumping received signals into discrete categories (Bradbury and Vehrencamp 2011). Theories why this phenomenon occurs are manifold. In an early work, Ehret hypothesized that the adaptive function of categorical perception is to reliably differentiate discrete call-type-specific features within noisy and variable multidimensional signals that also vary in continuous motivational parameters (Ehret 1987). Another hypothesis is that categorical perception of continuous signals allows groups to communicate within a group-specific communication system and hence fosters group cohesion.

3.2 Vocal Complexity

In vocal repertoires with clear distinct acoustic signals, the number of signals can be a good descriptor for communicative complexity, as it is often the case in repertoires of songbirds. Here, species with a higher number of distinct signals exhibit a more complex communication system than species with fewer signals. As I have discussed in the previous section, vocal repertoires of nonhuman primates and of many other mammalian species can exhibit a substantial level of gradation within and between acoustic signals. Besides the difficulty to verify the number of signals within a graded repertoire, the fine structured variations in signals can also provide an additional dimension of vocal complexity (Freeberg et al. 2012). Another way of accessing the complexity of a vocal repertoire stems from information theory and is based on the measurement of uncertainty (Shannon 1948). The argument goes that the greater the diversity of signals in a vocal repertoire, the greater the uncertainty of a specific signaling event. With the occurrence of a signaling event then, the reduction of uncertainty is higher in repertoires that have a greater diversity of signals. As a consequence, the potential information or complexity that such repertoires possess is higher. Studies on several taxa use repertoire size (e.g. in zebra finch: Boogert et al. 2008; Templeton et al. 2014) or information theory (e.g. in paridae: Krams et al. 2012 or nonhuman primates: Bouchet et al. 2013) to measure vocal complexity. Freeberg and colleagues point out that the actual way in which variation in signals affects the behavior of receivers has to be taken into account, in order to describe all aspects of vocal complexity in a species (Freeberg et al. 2012). Supporting this view Skyrms suggests that in order to measure the information of a signal, it is important to distinguish between the quantity of information in a signal and the informational content of a signal. Whereas the quantity of information can be measured as the extent that the use of a particular signal changes the probability of a specific condition to be true, the informational content lies in the direction the signal affects probabilities, i.e. which condition is more likely to be true (Skyrms 2010).

4 Approaches to Analyze Vocal Repertoires

4.1 Signal Definition

The first step in the analysis of vocal repertoires is to define the signal, i.e. the call unit given by the signaler. Generally, a call unit can be defined on the level of production mechanisms, which focus on how the sounds are generated by the signaler, or by perception mechanisms, which focus on how the sounds are interpreted by the receiver (Kershenbaum et al. 2014). Since the details of acoustic production and perception can be hidden from the researcher however, the acoustic features that can be observed are usually used to define the call unit (Catchpole and Slater 2003). Based on acoustic features, call units are most commonly defined by the presence of silent gaps before and after the unit, which can be identified by the inspection of the time signal or spectrogram of the call (Kershenbaum et al. 2014). Once the call unit has been identified, there are several approaches to extract acoustic features from the sound recording. In species that use less complex call structures Zero-Crossings Analysis (ZCA) is a fast and efficient tool. By counting how many cycles occur in a given time interval, ZCA can be used to identify frequency and frequency modulation. This technique is commonly used to analyze ultrasonic calls in bats (Fenton et al. 2001; Corben 2002) and finds its application in diverse taxa such as anurans (Wilczynski et al. 1995; Huang et al. 2009) or crickets (Bailey et al. 2001). If species use more complex calls and harmonics and amplitude represent important acoustic features of the signal, other methods have to be used. The most common approach to extract acoustic features from acoustically more complex signals is by fast Fourier transforming (FFT) the signal into its frequency-time domain (spectrogram). From this spectrogram, many temporal and spectral features can be extracted that are relevant for acoustic communication using software tools such as Avisoft (Specht 2004), PRAAT (Boersma and Heuven 2001), or Raven (Charif et al. 2006). For an overview of such features, see the method section of Chapter 2. An alternative approach (which is often verified by FFT) is linear predictive coding (LPC). LPC is based on the source-filter model I introduced in Chapter 1.1.1 and is used to measure formant frequencies. From the vocal tract length of the signaler, the number of formants can be determined and subsequently, formant dispersion can be calculated. LPC has its origin in human speech analysis, but is also used in call analysis of primates (Fitch 1997; Rendall et al. 1999; Pfefferle and Fischer 2006) and other mammals (e.g. in dogs: Riede and Fitch 1999 or pigs: Schön et al. 2001). Other techniques, such as wavelet or cepstral analysis, are less common. I will discuss the usage of these alternative techniques in the general discussion of this thesis.

4.2 Call Classification Using Unsupervised Clustering

After acoustic features have been extracted from the identified call units using one of the mentioned techniques, call classification is commonly used to separate the calls into discrete types. Traditionally, calls have been categorized by visual inspection of spectrograms (Kroodsma 1974; Marler and Pickert 1984). Although humans are considered to be good at visual categorization (Ripley 1996), this procedure can include bias related to human perceptual processing and therefore lack objectivity (Hopp et al. 1998). Further, this technique is generally not suitable for the categorization of highly graded systems, time consuming, and prone to subjective errors (Burghardt et al. 2012). The upsurge of computational possibilities brought new methodologies that allow standardization across large datasets without the disadvantage of subjective *a priori* classification (Clemins and Johnson 2006). These unsupervised clustering algorithms have proven to be time-saving and more objective (Stowell and Plumbley 2014). Since the notion of a cluster cannot be precisely defined, unsupervised clustering algorithms are manifold and based on different calculations. Two groups of algorithms that are commonly used to categorize vocal repertoires are centroid models like k-means clustering which represent each cluster by a single mean vector, and connectivity models, like hierarchical clustering, that build clusters based on distance connectivity between data samples (Duda et al. 2012).

Unsupervised clustering has been used to categorize vocal repertoires of several species, such as sperm whales (Weilgart and Whitehead 1997), dolphins (McCowan 1995), piglets (Tallet et al. 2013), Barbary macaques (Hammerschmidt and Fischer 1998), and true lemurs (Gamba et al. 2015). Since for unsupervised clustering algorithms the desired classification is unknown, several techniques exist to quantify the stability of the clustering result, as an indicator of clustering quality (Kershenbaum et al. 2014). One common method is to inspect silhouette values which represent the tightness of calls within a cluster and the separation between different clusters in a given repertoire (Rousseeuw 1987). By identifying the cluster solution with the highest silhouette value, the solution that best represents the structure of the dataset can be extracted (e.g. Maciej et al. 2013). Another method that can be used to access cluster quality of by calculating the normalized mutual information (NMI) that compares how well the results of two different clustering schemes match (Fred and Jain 2005).

It is important to keep in mind that these techniques heavily rely on the acoustic features that are used to characterize the structure of the calls and hence the cluster stability gives no evidence for the biological significance of the calculated clusters. To access which features of a signal are perceptually salient, playback experiments are required in which acoustic features are systematically excluded, distorted, or held constant to access their importance for signal receivers (Hauser 1996). Cluster stability is also affected by the composition of the dataset that is analyzed and can change if more calls are included in the analysis (Ben-David et al. 2006). The usage of different unsupervised clustering algorithms, the measurement of clustering quality, and current shortcomings in the analysis of highly graded repertoires are an integral topic of this thesis and will be discussed in detail throughout the next chapters.

5 Aims of this Thesis

Detailed descriptions of vocal repertoires are not only necessary to investigate driving forces in signal evolution (Chapter 1.2), but also needed to determine a repertoire's complexity and to understand consequences for signal processing by signal receivers (Chapter 1.3). In Chapter 1.4 I discussed several approaches to analyze vocal repertoires and highlighted remaining hindrances towards an objective description of vocal repertoires, especially the ones that show a high degree of variation within call structures.

In **Chapter 2** of this thesis, I am investigating several factors that can influence the outcome of a vocal repertoire analysis. The main focus of this chapter is put on the choice of acoustic features that are used in the analysis, differences between alternative unsupervised clustering algorithms that can be applied as well as different approaches of cluster validation. I also present a novel approach based on fuzzy logic that we developed to describe the variation of call structure on a quantitative level. The datasets that have been used for this study come from recordings of chacma baboons, a species which vocal behavior has been intensely studied in the past and therefore served as a good model to access the accuracy of the different approaches. The study was published in PLoS One at the beginning of this year (Wadewitz et al. 2015a).

As a next step, we systematically compare the vocal repertoire of chacma baboons with the vocal repertoire of Barbary macaques in **Chapter 3**. Since we are interested in the differences concerning the level of gradation within vocal repertoires, the comparative approach between the rather discrete repertoire of chacma baboons and the rather graded repertoire of Barbary macaques allows us to evaluate our developed method and to re-examine existing hypotheses about the influencing factors that drive signal evolution. We also present an extension of our approach that circumvents the problem of the determination of the appropriate number of call types. This study was recently submitted to a peer-reviewed journal.

In Chapter 2 we investigated several factors that are dependent on decisions by the researcher during the analytical steps to characterize a vocal repertoire. Another important aspect of a vocal repertoire analysis is the construction of the data set that is used for the analysis and several factors based on the data set composition can have a profound effect on the vocal repertoire analysis. In **Chapter 4**, two of these factors, namely arousal- and size-based differences of the recorded animals, are investigated. To do so, we collaborated with Ingo Titze and colleagues from the National Center for Voice and Speech in Salt Lake City, Utah and Ingo Riede from the Department of Physiology in Glendale, Arizona. We created pseudo vocal repertoires with differing levels of call structure variation by using an elaborated finite element model that simulates muscle characteristics of the larynx and vocal tract anatomy (Chapter 1.1). This model was developed by Titze and colleagues and has been successfully used to model vocalizations of different taxa. The study is currently prepared for submission.

Finally, in **Chapter 5** I summarize the results of my studies and discuss their implications for the ongoing methodological development in bioacoustics research as well as for the general examination of the evolution of vocal communication.

2 | Characterizing Vocal Repertoires - Hard vs. Soft Classification Approaches

Philip Wadewitz[1,2,3], Kurt Hammerschmidt[1], Demian Battaglia[2,3,4], Annette Witt[2,3], Fred Wolf[2,3], Julia Fischer[1,3]

[1] Cognitive Ethology Laboratory, German Primate Center, Göttingen, Germany

[2] Theoretical Neurophysics, Max Plank Institute for Dynamics and Self-Organization, Göttingen, Germany

[3] Bernstein Center for Computational Neuroscience, Göttingen, Germany

[4] Theoretical Neurosciences Group, Institute for Systems Neuroscience, Marseille, France

published in

PLOS ONE 10(4):e0125785 (2015)

1 Abstract

To understand the proximate and ultimate causes that shape acoustic communication in animals, objective characterizations of the vocal repertoire of a given species are critical, as they provide the foundation for comparative analyses among individuals, populations and taxa. Progress in this field has been hampered by a lack of standard in methodology, however. One problem is that researchers may settle on different variables to characterize the calls, which may impact on the classification of calls. More important, there is no agreement how to best characterize the overall structure of the repertoire in terms of the amount of gradation within and between call types. Here, we address these challenges by examining 912 calls recorded from wild chacma baboons (*Papio ursinus*). We extracted 118 acoustic variables from spectrograms, from which we constructed different sets of acoustic features, containing 9, 38, and 118 variables; as well 19 factors derived from principal component analysis. We compared and validated the resulting classifications of k-means and hierarchical clustering. Datasets with a higher number of acoustic features lead to better clustering results than datasets with only a few features. The use of factors in the cluster analysis resulted in an extremely poor resolution of emerging call types. Another important finding is that none of the applied clustering methods gave strong support to a specific cluster solution. Instead, the cluster analysis revealed that within distinct call types, subtypes may exist. Because hard clustering methods are not well suited to capture such gradation within call types, we applied a fuzzy clustering algorithm. We found that this algorithm provides a detailed and quantitative description of the gradation within and between chacma baboon call types. In conclusion, we suggest that fuzzy clustering should be used in future studies to analyze the graded structure of vocal repertoires. Moreover, the use of factor analyses to reduce the number of acoustic variables should be discouraged.

2 Introduction

Objective classifications of animal signals are a prerequisite for addressing a broad array of questions, both at the proximate and ultimate level. Much progress has been made in developing quantitative methods to objectively characterize single acoustic patterns (Boersma and Heuven 2001; Tchernichovski et al. 2000). Less agreement, however, exists on how to objectively characterize the structure of the entirety of a species, that is, its vocal repertoire. Being able to compare the vocal repertoires of different species is crucial to test hypotheses regarding the selective pressures that shape signal repertoires. For instance, the habitat a species lives in was suggested to influence both the spectral characteristics as well as the overall structure of a repertoire (Forrest 1994; Padgham 2004; Waser and Brown 1986). More recently, it was suggested that increased social complexity gives rise to increased vocal complexity (Gustison et al. 2012; McComb and Semple 2005). To rigorously test this assumption, quantitative assessments of vocal complexity are needed. More important, broader comparative or meta-analyses are hampered because studies from different labs often lack consistency in the methods used and in the categorization criteria applied.

Many vocal repertoires are characterized by their graded morphology, meaning that the acoustic structures of vocalizations are not well separated and discrete, but rather form a continuum in the acoustic space (Winter et al. 1966). Such graded systems are assumed to have evolved in species with ready visual access to each other (Marler 1975) and are common in most mammalian vocal systems. Although graded vocal systems are described in a number of nonhuman primates (Arnold and Zuberbühler 2006; Green 1975; Hammerschmidt and Fischer 1998; Marler 1970, 1976; Abbot et al. 2011; Rowell and Hinde 1962; Tomasello and Zuberbühler 2002; de Waal 1988), labelling whole repertories as being either discrete or graded often represents an oversimplification, since gradation can occur within and between call types, and call types may vary to different degrees (Keenan et al. 2013). Whereas between-call-type variation might be dependent on the call's function, within-call-type variation could be linked to an animal's general affective state (Fischer et al. 1995; Manser 2001). Within this general affective state, similar situations can potentially evoke slightly different forms of excitement or fear, which can then relate to dissimilar acoustic structures within call types (Fischer et al. 2001). The importance to differentiate between these different forms of gradation, however, is neglected in most studies on vocal repertoires.

Whereas historically, vocal repertoires were established by human observers via visual categorization of spectrograms (Marler 1976), current approaches largely make use of unsupervised clustering methods (Hammerschmidt and Fischer 1998) that are based on acoustic features extracted from spectrograms. The selection and number of these features may have a potentially critical impact on the subsequent analysis. Thus, the question arises whether a quantitative comparison of repertoires is feasible if repertoires are based on different types and numbers of extracted features. In addition, many studies use factors derived from factor analysis to avoid the use of highly correlating acoustic features (Arnold and Zuberbühler 2006; Bouchet et al. 2012; Gros-Louis et al. 2008). In this study, we use a defined dataset of chacma baboon (*Papio ursinus*) vocalizations to examine how the choice of extracted acoustic features affects clustering results. The structure and function of chacma baboon calls are well known (Fischer et al. 2001; Kitchen et al. 2005; Maciej et al. 2013; Owren et al. 1997), and were partly validated in playback experiments (Fischer et al. 2000, 2001; Rendall et al. 2000). These previous descriptions of call types allowed us to externally validate the structure of the chacma baboon's vocal repertoire.

A second focus of this study was to assess how suited different clustering algorithms are to describe the fine structure of graded vocal systems. In a recent study, Kershenbaum and colleagues tested the performance of different unsupervised clustering-algorithms (k-means, hierarchical clustering, and an adaptive resonance theory neural network) for grouping dolphin signature whistles and compared the results with those of human observers (Kershenbaum et al. 2013). Although all algorithms performed relatively well in the classification of signature whistles, there are some inherent shortcomings that all of them share when constructing vocal repertoires - none of these hard algorithms are able to capture the graded transition of call types that occur in many vocal repertoires. We compared two commonly used non-overlapping models, center-based k-means and hierarchical Ward's clustering, and opposed them to a soft clustering approach, fuzzy c-means clustering (Dunn 1973). Fuzzy set theory has a broad range of applications and has for instance been used in numerical taxonomy (Bezdek 1974) or to cluster ecological data (Equihua 1990). Despite its successful application in these fields, it has not yet been used in vocalization taxonomy. Whereas in k-means and Ward's the existence of a graded separation between call types is not implemented, fuzzy c-means is an algorithm designed to describe systems with not strictly separated categories. We thus expected that fuzzy c-means would be able to describe the graded structure of the chacma baboon's vocal repertoire better than the other methods.

Our overarching goal is to develop recommendations for future analyses of vocal repertoires, with the long-term perspective of creating unified and standardized procedures in the field of bioacoustic research.

3 Methods

3.1 Study Site and Subjects

In this study, we reanalyzed call recordings that were collected during January 1998 and June 1999 in the Moremi Wildlife Reserve in Botswana. A number of comprehensive studies on the social behavior as well as on the vocal communication of this population has been published (see references in Silk et al. 1999).

3.2 Recordings and Call Parameterization

Recordings were taken as part of a number of studies on the monkeys' vocal communication (Fischer et al. 2002). Vocalizations were recorded with a Sony WM TCD-100 DAT recorder and a Sennheiser directional microphone (K6 power module and ME66 recording head with MZW66 pro windscreen) (Fischer et al. 2002). We assembled a data set comprising of 912 calls, which we selected to capture the overall diversity of the chacma baboon's vocalizations. The selected calls were given by 35 adult females and 34 adult males, as well as 5 infant females and 4 infant males (weaning calls). We fast Fourier transformed (FFT) the calls into their frequency-time domain with Avisoft (Avisoft SASLab

Pro, version 5.2.05), using a FFT size of 1024 points, Hamming window and 96.87% overlap. Depending on the frequency range of calls we used a sampling frequency of 5 kHz (grunts) or 20 kHz (all others), resulting in a frequency range of 2.5 or 10 kHz and a frequency resolution of 5 or 20 Hz. The time increment was 6.4 or 1.6 milliseconds. The resulting frequency-time spectra were analyzed with the software LMA 2012 developed by Kurt Hammerschmidt.

To assess the influence of datasets with varying numbers of acoustic features on the clustering results, we constructed 4 different sets for the subsequent analyses, all based on the 912 calls in the analysis. The sets include

a) "sparse set": 9 features, which were used in a previous analysis of the Guinea baboon' vocal repertoire and had proven to be instructive (Maciej et al. 2013)

b) "medium set": 38 features, which are an extension of a) including more detailed features in the frequency- and time domain

c) "full set": 118 features - the maximum amount of features that can be extracted out of the FFT using LMA

d) "factors": 19 features - derived from a factor analysis of the 118 features dataset.

We performed Factor analysis with IBM SPSS Statistics (version 21) using varimax rotation and factors with an Eigenvalue ≥ 1 were selected. Factor loadings, Eigenvalues, and detailed information about all acoustic features used are given in the appendix.

Clustering Schemes

To classify the calls, we performed unsupervised clustering using the above mentioned feature sets. Sets were standardized by z-scoring all of the values and cluster analysis was run within the Matlab environment (Mathworks; version R2011b). We used different clustering methods for comparison, which are described in the following sections in more detail. First, hard algorithms (k-means, Ward's clustering) were used and validated.

Second, a soft classification scheme based on fuzzy set theory (Zadeh 1965) was applied to capture more details of the dataset's underlying structure.

Hard Classification Models and Clustering Validation

Ward's clustering (Ward 1963) is a hierarchical clustering procedure, that is often used to cluster calls and to analyze vocal repertoires (Fuller 2014; Kershenbaum et al. 2013; Laiolo et al. 2000; Shulezhko and Burkanov 2008). The algorithm works by first linking individual calls to their nearest neighbor and then merging the pair of clusters with the minimum between-cluster distance at each time step. This linkage procedure is repeated on these clusters until the top hierarchic level is reached (single-linkage clustering).

In k-means clustering (MacQueen 1967), initial cluster centroids are selected randomly and individual calls are assigned to the cluster whose mean yields the least within-cluster sum of squares (WCSS). In iterative steps the new centroids of the clusters are being calculated and the procedure is repeated until the WCSS cannot longer be improved. Since poor initial cluster centroids can lead to non-optimal solutions by running into local maxima, we executed 100 replications to ensure that the best cluster solution was revealed. K-means clustering has the advantage that initially poorly attributed calls are reassigned by the algorithm and is therefore an often used procedure to classify calls (Hammerschmidt and Fischer 1998; Hammerschmidt and Todt 1995; Kershenbaum et al. 2013; Maciej et al. 2013). However, since in several studies the determination of the optimal number of clusters k showed to be challenging, we here did a further validation of clustering quality.

To assess which of the feature sets give rise to classifications most robust against changes of the clustering method, we measured the Normalized Mutual Information (Dunn 1973) between clusters extracted by two different methods. Normalized mutual information (NMI) is a single metric that measures how well the results of the two different clustering approaches match. If the clusters extracted by Ward and k-means methods are perfectly overlapping, NMI takes a value of 1. If the resulting clusters have little conformity, NMI takes a positive value close to zero. NMI is defined as:

$$NMI = \frac{\sum_{k,c} n_{k,c} log\left[\frac{N \times n_{k,c}}{n_k \times n_c}\right]}{\sqrt{\left(\sum_k n_k log\frac{n_k}{N}\right)\left(\sum_c n_c log\frac{n_c}{N}\right)}} \tag{2.1}$$

where n_c is the number of calls assigned to cluster c by method 1, n_k is the number of calls assigned to cluster k by method 2, $n_{k,c}$ is the number of calls in cluster c and cluster k, and N is the total number of calls.

We also used NMI to compare clustering results with a reference classification. Based on prior studies of the usage, function and meaning of vocalizations, we established six call types, namely male barks (Kitchen et al. 2005); grunts (Owren et al. 1997); weaning calls (Maciej et al. 2013); female barks (Fischer et al. 2001); noisy screams (Maciej et al. 2013); and tonal screams (Maciej et al. 2013). Representative calls are shown in Figure 2.1. Based on acoustic and visual spectrogram evaluation, we assigned each call in the dataset to one of these categories. This procedure provided a defined human expert reference classification.

The quality of a clustering was also validated by the analysis of silhouette values. Silhouette values range from 1 to -1 and represent the tightness of data points within a cluster and the separation between different clusters in a given model (Rousseeuw 1987). Silhouette values are computed as following:

$$S(i) = \frac{b(i) - a(i)}{max[a(i), b(i)]} \tag{2.2}$$

where $a(i)$ denotes the average Euclidean distance between data point i and other data points in the cluster A and $b(i)$ denotes the average Euclidian distance between i and points in the second closest cluster. A silhouette value around zero means that the data point is at similar distance to two clusters. Positive values show that the data point lies closer to one cluster than to the second closest one. Negative values indicate a potential misclassification (even if reassigning a point with a negative silhouette to a different cluster would change as well the cluster means, resulting in a potentially larger number of negative silhouette scores). The overall silhouette width $S(A)$ is defined as the average of the $S(i)$ over the whole dataset and is used as a global measure of the quality of a clustering.

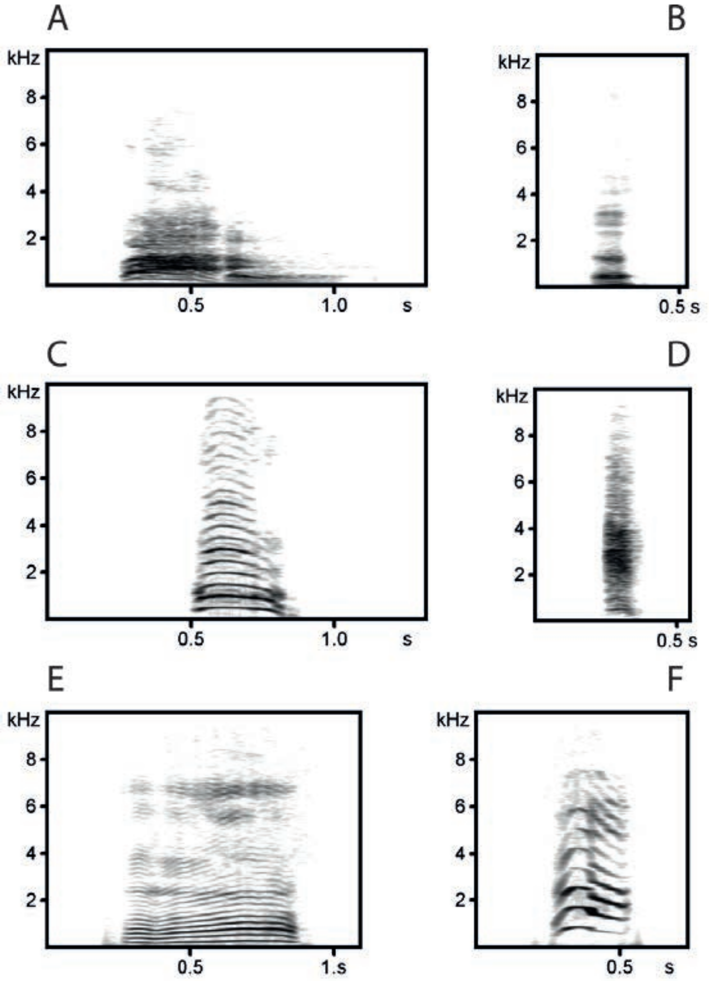

Figure 2.1: Spectrograms of calls in the used dataset. Shown are call types that have been described in the literature. (A) Male bark (Kitchen et al. 2005). (B) Grunt (Owren et al. 1997). (C) Female bark (Fischer et al. 2001). (D) Noisy scream (Maciej et al. 2013). (E) Weaning call (Maciej et al. 2013). (F) Tonal scream (Maciej et al. 2013).

Soft Classification Model: Fuzzy c-means clustering

Fuzzy set theory (Zadeh 1965) extends conventional set theory allowing for the notion of imperfect membership. In this way, it is particularly suited to the classification of data in which the separations between different classes of data-points is gradual rather than sharp (Zadeh 2008). Each call is associated an assigned membership value for each of the clusters, ranging from $m = 1$ (fully displays the properties of the cluster) and $m = 0$ (does not display any of the properties of the cluster). Intermediate membership values

$0 < m_{ia} < 1$ mark calls that do not fully belong to one of the clusters, but can be classified as intermediates between different call types. Membership vectors are normalized in such a way that $\sum_{\alpha=1}^{c} m_{i\alpha} = 1$.

More specifically, we adopted a fuzzy c-means algorithm (Jang and Sun 1997; Xu et al. 2008). To determine the number of clusters that describe the dataset best, two parameters of the algorithm can be adjusted. The first parameter is the maximal number of clusters allowed and the second is the fuzziness parameter μ. If $\mu = 1$, the extracted clusters are very crisp and membership values of data points are either 1 or 0 (in this limit indeed fuzzy c-means converges exactly to k-means). However, by increasing μ, clusters become fuzzier and nearby clusters can eventually merge, unlike in k-means, leading to a smaller number of clusters. We assumed a relatively large possible number of clusters $c = 15$ (larger than the number of reasonably detectable clusters).

Similar to k-means, the fuzzy c-means algorithm builds up clusters by creating randomly selected cluster centroids and a subsequent iterative optimization process. In this aspect both clustering algorithms suffer from the same sensitivity to the initial cluster centroids. Like in k-means, we computed 100 replications to find the optimal cluster solution with fuzzy c-means. In contrast to k-means, where objects do either belong or not belong to a cluster, in fuzzy c-means membership vectors $m_i^{(t)}$ for c clusters are computed at a given iteration t. Cluster centroids are given by vectors $u_\alpha^{(t+i)}(\alpha = 1...c)$ with components $u_{\alpha l}^{(t)}$.

$$\frac{1}{m_{i\alpha}^{(t)}} = \sum_{\lambda=1}^{c} \left(\frac{d_{i\alpha}^{(t)}}{d_{i\lambda}^{(t)}} \right)^{\frac{2}{\mu-1}} \tag{2.3}$$

where $d_{i\lambda}^{(t)}$ is the Euclidean distance between the data-point f_i and the centroid $u_\lambda^{(t)}$ at a given iteration t.

These membership vectors are used in turn to compute a new set of cluster centroids $u(t+1)$ with coordinates:

$$u_{\alpha l}^{(t+1)} = \frac{\sum_{i=1}^{N}(m_{i\alpha}^{(t)})^{\mu} f_{il}}{\sum_{i=1}^{N}(m_{i\alpha}^{(t)})^{\mu}} \tag{2.4}$$

This procedure is designed to minimize a specific cost function (Dunn 1973), namely the sum of the squared distances of the data-points from the different centroids, weighted by the relative fuzzy memberships:

$$J^t = \sum_{i=1}^{N} \sum_{\lambda=1}^{c} (m_{i\lambda}^{(t)})^{\mu} \times (d_{i\lambda}^{(t)})^2 \tag{2.5}$$

Once the fuzziness parameter μ is set and the clusters (i.e. call types) have been computed, the main type α for each call i is the call type with the highest assigned membership component $m_{i\alpha} = m_i^{(1st)}$. By subtracting the second largest membership component $m_{i\alpha} = m_i^{(2nd)}$ from the first, we get the typicality coefficient $d(i) = m_i^{(1st)} - m_i^{(2nd)}$ for each call. The average $\overline{d}$ of all typicality coefficients and their distribution, quantified by the halved mean absolute deviation $\Delta = d - d/2$ were quantified over the entire dataset. Based on the observed distribution of typicality coefficients, calls were then considered as typical if $d > d_{typical} = \overline{d} + \Delta$ and as atypical if $d < d_{typical} = \overline{d} - \Delta$.

4 Results

The hierarchical clustering trees generated by Ward's method show similar classifications of calls for all four sets (Fig. 2.2). However, crucial differences in linkage distances can be found (see y-axes of the four graphs). In the following, the results are exemplified for the full set. Calls are first segregated into two clusters. All calls of cluster 1 ($n = 124$) are characterized by their high frequency distribution over the entire call and are hereafter denoted as "screams". In contrast, all calls of cluster 2 ($n = 788$) are characterized by a substantially lower overall frequency distribution. Cluster 2 was further divided into two second-order branches. Cluster 2.2 ($n = 350$), is characterized by very short and low-frequency calls (grunts). In the next higher order, cluster 2.1 ($n = 438$), splits into cluster 2.1.1 ($n = 97$) and cluster 2.1.2 ($n = 341$). Calls in cluster 2.1.1 are characterized as highly tonal, long and little frequency-modulated (weaning calls), whereas calls in cluster 2.1.2 are shorter and have a higher change in frequency-modulation (barks). On the next level, cluster 1 (screams), is split into cluster 1.1 ($n = 68$) and cluster 1.2 ($n = 56$). Calls between these two sub-clusters differ mainly in their signal to noise ratio (STNR), with calls in cluster 1.1 having a higher average STNR. Further structure was detected by the hierarchical clustering. However we did not analyze it in further detail, due to the instability of these classifications (as revealed by fuzzy c-means, see below). Since the Euclidean distance is defined as the square root of the sum of the squared distances per feature, the less features are included in the analysis, the smaller the average Euclidean distance within a cluster becomes (Fig. 2.2). Although the within-cluster distances are decreasing with decreasing number of acoustic features, the separation of the first three clusters remains rather distinct (see branch structure of dendrograms in Fig. 2.2 A-C). An exception of this pattern is formed by the factorial dataset, which shows a much worse separation of even a small number of call clusters (Fig. 2.2 D).

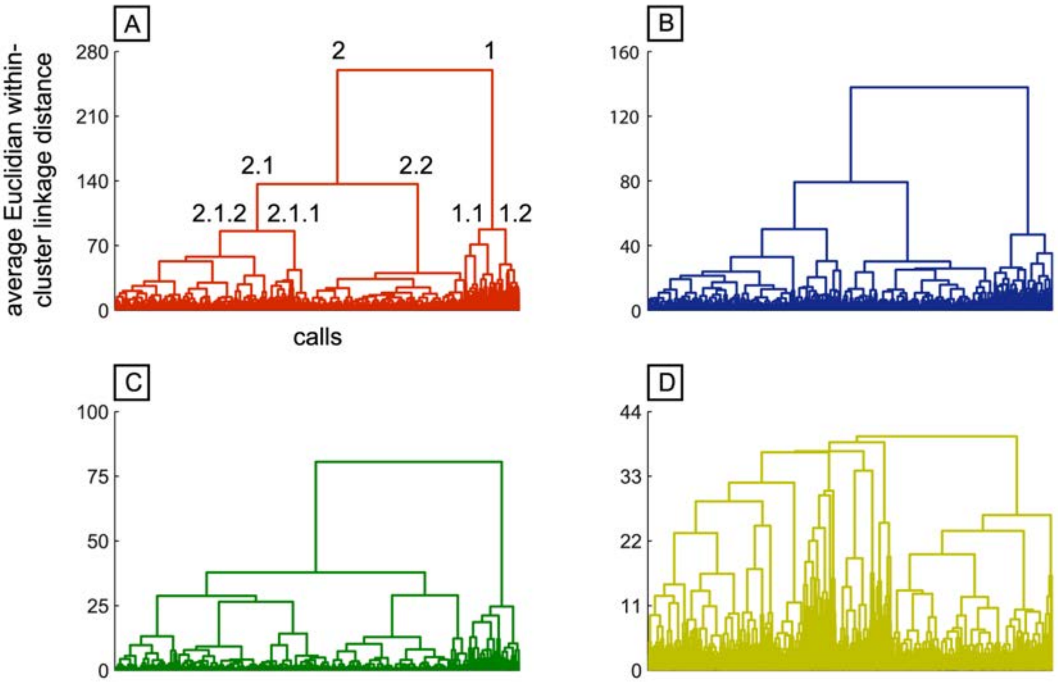

Figure 2.2: Unsupervised Ward's clustering of 912 chacma baboon calls based on different frequency dependent and temporal feature setups. The x-axis represents groups of calls, and the y-axis represents average Euclidian within-cluster linkage distance. (A) Set consisting of 118 features. High-frequency (cluster 1) and low-frequency (cluster 2) were segregated into two first-order clusters. High frequency calls further subdivide into more tonal (cluster 1.1) and relatively noisier (cluster 1.2) calls. Low frequency calls subdivide into short and very low-frequency grunt-calls (cluster 2.2), moderate-frequency and harmonic weaning-calls (cluster 2.1.1), and more noisy, short bark-calls (cluster 2.1.2). (B) Set consisting of 38 features. (C) Set consisting of 9 features. (D) Set consisting of 19 factors determined by factor analysis.

To compare the clustering quality of the four feature sets, we validated the results of k-means clustering. For this purpose we calculated silhouette widths for $k = 2 - 20$ clusters for all four datasets (Fig. 2.3). The general trend for all datasets but the one based on factors was that a 2-cluster solution gained a relatively high value in silhouette widths, followed by a drop and a subsequent stable cluster quality that decreased slowly the more clusters were generated. Silhouette widths for the 9-feature set were generally higher than for the other datasets and silhouette widths for the 19-factor set were generally lower for lower number of cluster solutions.

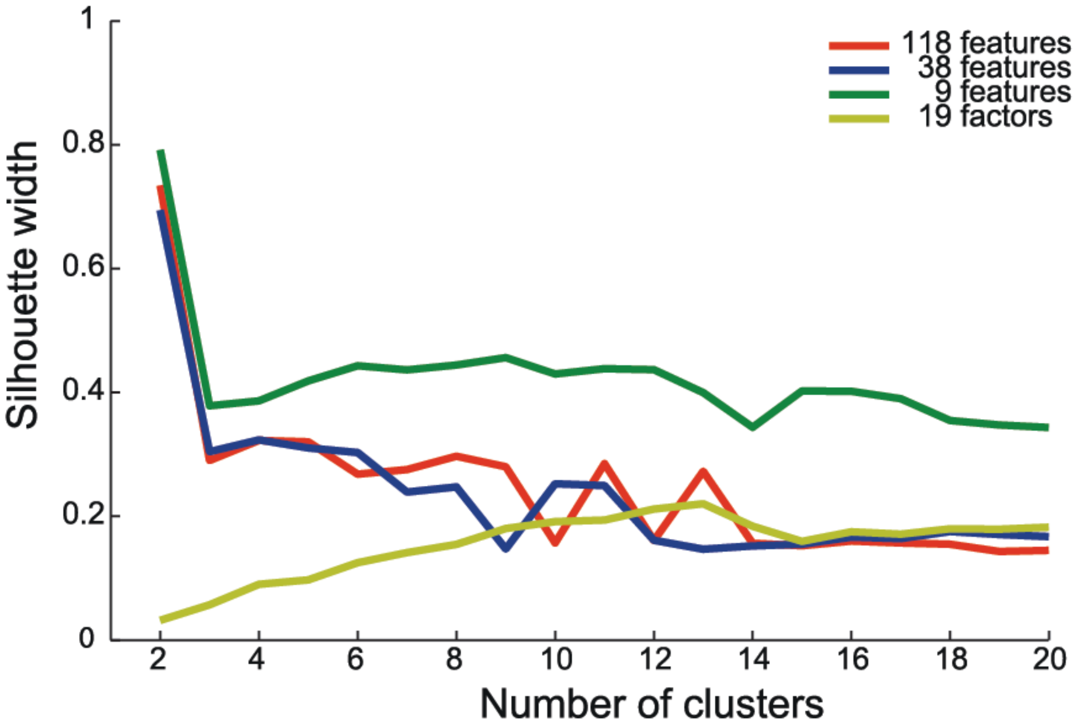

Figure 2.3: **Comparison between the average silhouette width for K-means clustering for k = 2 to 20 clusters for all 4 feature sets.** The 9 feature set (green) shows generally higher silhouette width. For the 2-cluster-solution, all but the set based on factors (yellow) show globally the highest value. Excluding the 2-cluster-solution (not to be retained because of its lack of detail), no solution is markedly superior over all others, although plateau values of average silhouette width are already obtained for cluster numbers as small as k = 5 (apart from factor-based clustering).

We then evaluated the large-scale behavior of the four considered curves. Here we found, for the sets with 38 and 118 features slightly decreasing silhouette widths (for more than two clusters), for the set with 9 features essentially constant values (for more than two clusters) and for the set with 19 factors an increase up to 13 clusters that was followed by saturation. For these reasons, Normalized Mutual Information (NMI) was calculated to further explore the quality of clustering results. If our two unsupervised methods (k-means and Ward's), operating on opposite approaches result in a similar classification, this would be a strong indicator for the robustness of the classification. Classifications extracted by the different methods were overall highly consistent between both algorithms over a wide range of cluster numbers, with peak consistencies for all four datasets nearby $k = 5$ (excluding, as in Fig. 2.3, the too unresolved $k = 2$ clustering).

As a final check, since we know from previous studies that the call types of the 5 cluster solution (screams, barks, weaning calls and grunts) are well described calls in baboon vocalizations, NMIs between the 5-cluster partition extracted by k-means or Ward's unsupervised clustering and the human expert-based reference classification were also calculated (Fig. 2.4). The results show, that the classifications generally match well. This confirms that the 5-cluster solution obtained through k-means and Ward's methods are consistent with the results obtained by human expert inspection allowed us endorsing the unsupervised methods as valid alternatives to human inspection when the size of the dataset becomes prohibitively large to be manually parsed. The increase of NMI from the 9-feature set to the 38-feature set is quite large for both clustering algorithms, whereas the 118-feature set only gains a small increase in NMI compared to the 38-feature set. Thus, as a compromise between clustering quality and feature overview, we decided to work with the 38-feature set for the subsequent analysis. We decided against a subsequent usage of the 19-factor set, because factors not only showed the worst separation of clusters (Fig. 2.2), but also because factors are difficult to interpret if feature types are highly mixed (see discussion and appendix).

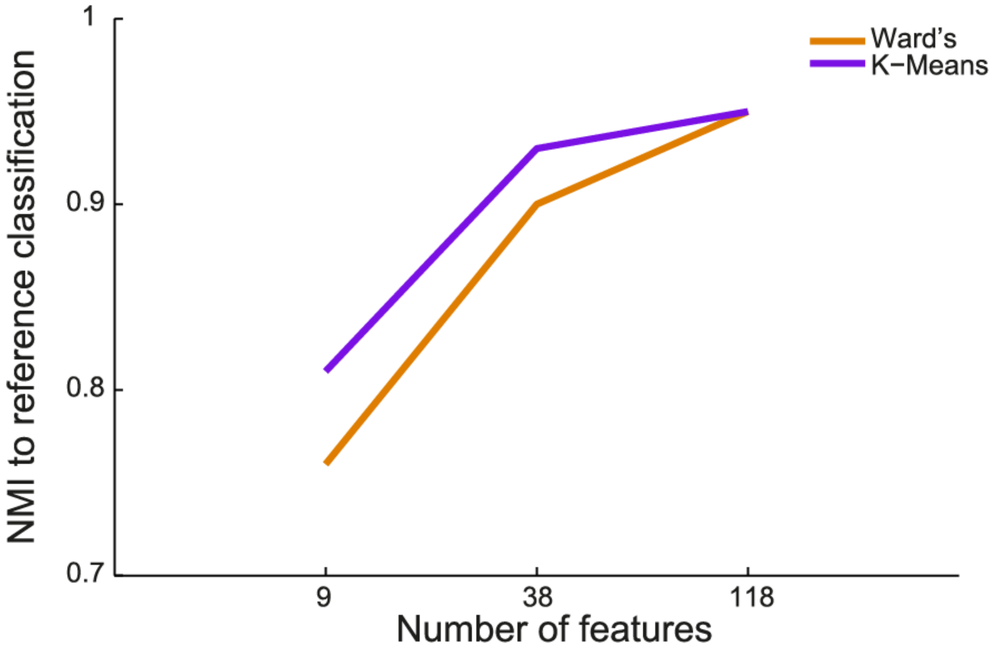

Figure 2.4: Sensitivity of the algorithm performance (normalized mutual information) between the human-made reference classification and K-means (purple), and Ward's (orange) clustering for the three feature sets. NMI values have been calculated for $k = 5$ clusters.

4.1 Fuzzy c-means clustering

To gain better insight into the graded structure of our dataset, we applied fuzzy c-means clustering. This allowed us to determine the best number of clusters in an alternative to silhouette widths for different cluster solutions in the aforementioned algorithms. Hereby, we followed an approach described in (Battaglia et al. 2013), where we made use of the fuzziness parameter μ. By starting with a sufficiently large μ, all calls were grouped indistinctively into one fuzzy class. Decreasing the fuzziness, high-frequency calls ("screams") separated then first at $\mu = 2.38$ (Fig. 2.5 A&C). At $\mu = 1.96$, a second cluster crystallized, consisting of short and low-frequency calls ("grunts") (Fig. 5 A&D). Between $\mu = 1.565$ and 1.515, a third cluster of modulated, short and harsher "bark" calls separated and at $\mu = 1.51$, the high-frequency "scream"-cluster split between calls with a higher and lower signal-to-noise ratio (Fig. 5 A&E). Below $\mu = 1.44$ down to $\mu = 1$ several smaller clusters emerged that were not very stable over μ. Looking at stability (cluster existence over fuzziness parameter μ), the 2-, 3- and 5-cluster solutions are most robust (Fig. 5 A). These results go along with the findings of k-means and Ward's clustering analyses. In Figures 5 C-E membership values for all calls to the existing clusters are shown for selected values of μ. The remaining analyses were performed for the specific classification obtained for $\mu = 1.505$, leading to 5 clusters. The results were very similar to the results of the k-means and Ward's clustering, which provides as a strong indicator that the obtained classification is very robust.

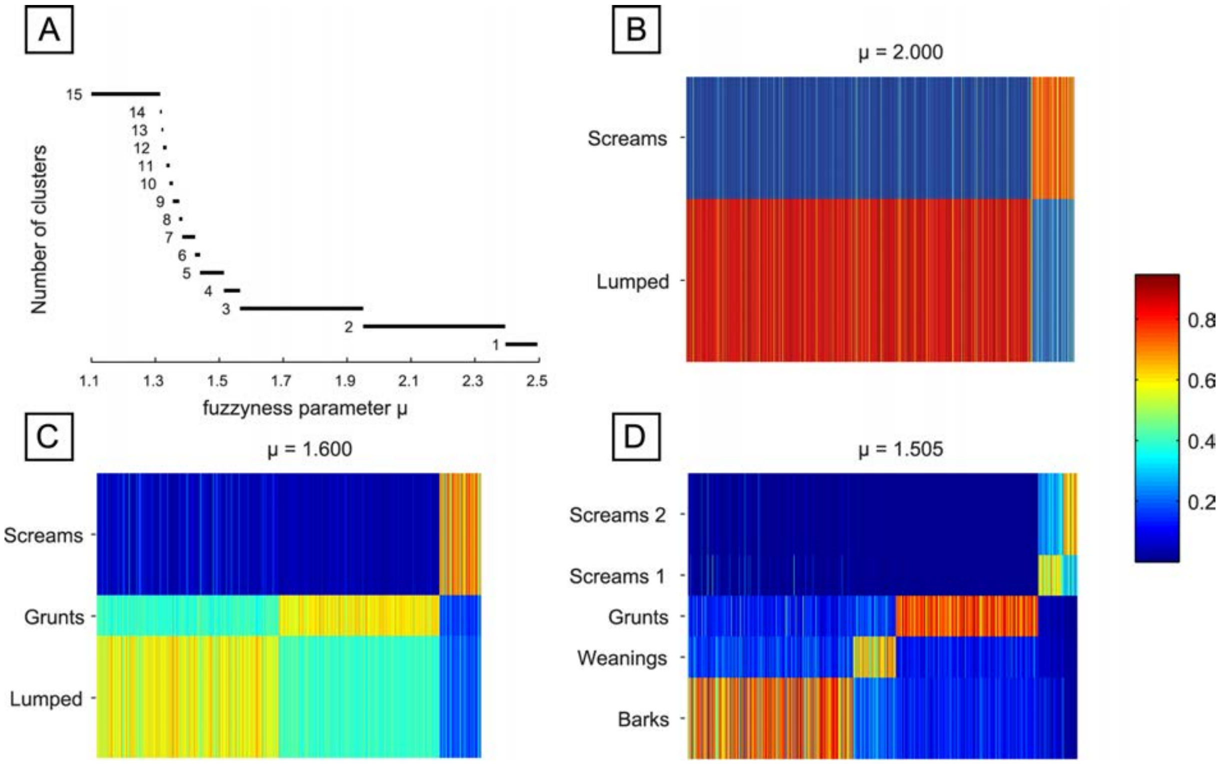

Figure 2.5: Fuzzy partitions with decreasing fuzziness (μ values) are visualized as membership matrices. (A) Number of clusters in dependence on the fuzziness parameter μ. Partitions with more than five clusters exist only over very narrow ranges of μ values (red). (C-D) membership matrices for the identified clusters: Rows correspond to different fuzzy clusters and columns to individual calls. Membership values of single calls to each class are color coded (B). The scream-cluster is the first to emerge (cluster 1, C), followed by grunts (cluster 2, D). The scream-cluster splits into two clusters and the weaning-cluster emerges (cluster 1-2; cluster 4, E).

In Figure 2.6 a 2-dimensional visualization of how calls are scattered in the membership space is presented. Each call is represented by the closest and the second closest cluster. For the five considered calls types we found common boarders between weaning-calls and barks and weaning-calls and grunts. In both cases, highly typical calls can be found along with calls that appear to belong to both clusters. Intermediate calls can also be found between the two scream types and sparsely between the bark and the scream 1 cluster. Calls in the bark- and grunt-clusters share common boarders and typical grunts and barks exist. In contrast to the other pairs, no calls at the very edge to the other cluster can be found and the two clusters remain separated.

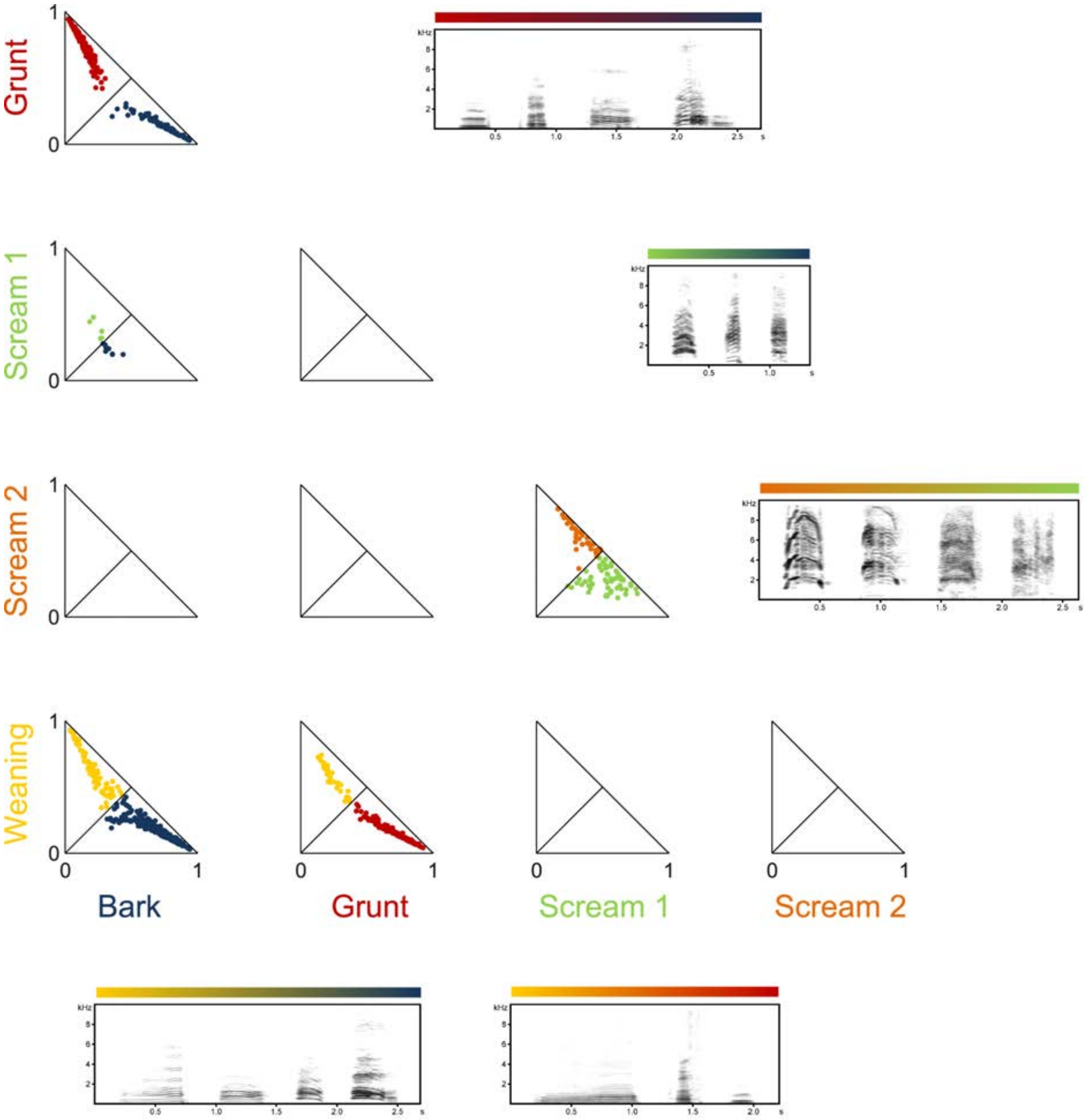

Figure 2.6: Pairwise comparisons of cluster segregations. Two-dimensional projections of memberships of calls belonging to the grunt (red), scream 1 (green), scream 2 (pink), weaning (yellow), and bark (blue) cluster. Every call is represented once (by closest and second closest cluster). Diagonal lines in the panels represent identical memberships. Spectrograms represent transitions from most typical call of cluster A to most typical call of cluster B with hybrids close to the joint cluster borders. Sound examples can be found in the supporting information.

To quantitatively describe the graded structure of our dataset, typicality coefficients for each call were calculated (Fig. 2.7; see Methods). Calls with a typicality larger or smaller than specific thresholds, related to the halved mean absolute deviation of the typicality distribution, were considered as typical or atypical, respectively. According to these criteria, the threshold for atypical calls was calculated at $d_{atypical} = 0.256$ (142 of

912 ≡ 16% of the calls) and for typical calls at $d_{typical} = 0.767$ (120 of 912 ≡ 13% of the calls). However, the distribution of typical and atypical calls was not homogeneous across different clusters. Most grunts and the majority of bark-calls were well-separated from the other call types, as indicated by their large average typicality coefficients (Fig. 2.7). Weaning calls were less detached and the two scream clusters were highly graded towards their shared borders.

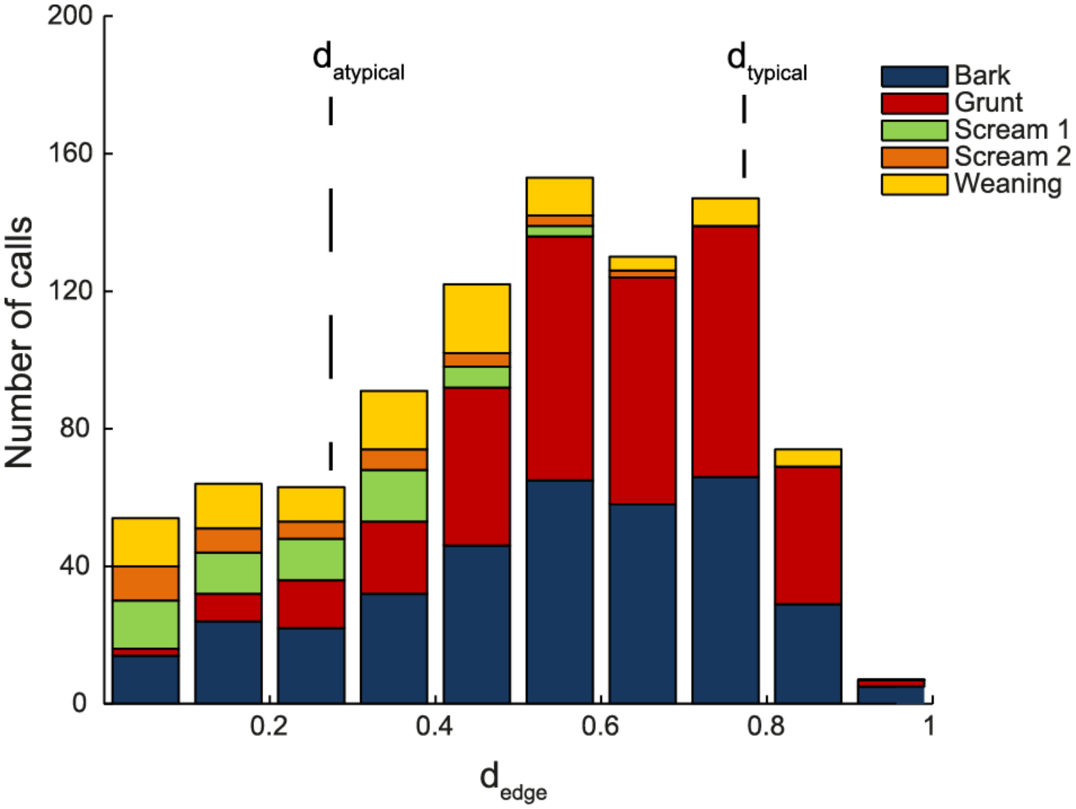

Figure 2.7: Histogram of typicality coefficients. Sections with different colors indicate calls with different main type. Grunts and barks are more distinctly separated from other call types than screams and weaning calls.

5 Discussion

We investigated how different feature setups can affect the clustering quality, and compared the usage of hard and soft clustering methods for the description of a primate vocal repertoire and. Our efforts provided two key results. Firstly, datasets with a higher number of acoustic features led to better clustering results than datasets with only a few features. Secondly, in datasets with considerable gradation within and between clusters, an optimal number of clusters (call types) may not exist, no matter which clustering algorithm is applied. Yet, fuzzy clustering allows one to capture and quantify the extent of variation within and between clusters, providing a potentially fruitful avenue to compare the extent of gradation within and between call types between taxa.

With regard to the number and types of features in the analysis, we found that a low number of features resulted in higher silhouette values. This was not necessarily due to a better separation of the call types, but rather the consequence of a smaller number of acoustic dimensions, and therefore a higher statistical spread of values. For this reason, the usage of absolute silhouette values to compare datasets with varying number of features does not appear to be appropriate. Indeed, when we compared the human-expert reference classification with the cluster solution, we found that the matching success increased with an increasing number of acoustic features. We therefore recommend the usage of a sufficiently large set of features to capture the different acoustic dimensions. Whereas correlated features can cause problems in multivariate statistical hypothesis testing due to colinearity, these restrictions do not apply to clustering procedures. In fact, correlating features can perform badly in classifying call types when taken on their own, but become well performing classifiers when combined. Since every feature has independent measurement noise that can hinder its classification performance, two or more features can share correlating trends but not the stochastic fluctuations around these trends (Guyon and Elisseeff 2003).

We also found that using factors derived from factor analysis resulted in an extremely poor resolution of emerging call types. In addition to the argument above, that correlating features can provide a sort of "error correction" for measurement noise, the weak performance of the factor analysis can be explained by its linear nature, always being based on a matrix decomposition of the covariance matrix. If the established clusters have non-spherical shapes in high-dimensional feature space it might not be possible to properly separate them by hyperplanes orthogonal to the factors. Thus reducing the dimensionality of the data by projecting them to the linear space spanned by only a few factors may conceal non-linear correlations in the data-set, which on the contrary can be exploited for performing clustering by unsupervised algorithms operating on an even smaller number of the original, not factor-reduced features. For these reasons, we generally discourage the use of factors in cluster analysis, and recommend caution when used in acoustic analyses more generally. Factors can be difficult to interpret, especially when highly divergent feature types load onto the same factors (see appendix). In such cases, the usage of selected features, preferably derived from a good understanding of the sound production mechanisms (Fitch and Hauser 1995), is more advisable. If factors are extracted, we recommend inspecting the factors and factor loadings carefully. If parameters load in an interpretable way onto a few factors that explain most of the variance of the dataset, then working with factors may be feasible, but it may also be the case that the construction of apparently meaningful factors results in the loss of crucial variation that would be helpful to distinguish between calls or call types.

A second important insight is that in datasets with a considerable variation an obvious optimal number of call types may not exist. Although the call types in our analysis were easy to distinguish, neither k-means nor Ward's clustering were able to identify an obvious "best solution". Based on the silhouette coefficients, different cluster solutions appeared to be appropriate to partition the dataset. In this aspect, fuzzy c-means clustering did not facilitate the decision on the best cluster solution. The finding that none of the applied approaches gave strong support to a specific cluster solution is somewhat surprising, since the chacma baboon vocal repertoire was previously described as representing a rather discrete system and call types can be easily categorized by human experts. With fuzzy c-means clustering, the 5-cluster solution was the most stable solution for $k > 2$, but differences in cluster stability were relatively small. A 5-cluster solution was also supported by high silhouette values in k-means and the NMI for call classification between k-means and Ward's also had an average peak at the 5-cluster solution. Overall, there appeared to be a trade-off between stability and acuity in our analysis.

When inspecting silhouette values, researchers should be aware that these values are affected by a number of factors. Firstly, with increasing number of features, the dimension of the acoustic space is increased. This leads to higher dispersion within and between clusters and consequentially to smaller silhouette widths. Secondly, although for this reason silhouette widths might be high for low feature sets, these sets may miss some crucial acoustic features to separate between different call types and therefore the clustering does not represent the true structure of the vocal system. Thirdly, within one feature set, silhouette widths indicate which cluster solutions are qualitatively better than others. Nevertheless, if the highest silhouette width commends a low number of clusters, this might be mathematically the best solution, but might not provide sufficient detail to describe a species' vocal repertoire.

Soft clustering allowed us to capture details of the graded nature of vocal repertoires that hard methods did not. Since fuzzy memberships directly represent structural differences of calls, typical and atypical calls within huge datasets can easily be detected and visualized. We propose that the robustness of cluster solutions over the fuzzy parameter in fuzzy c-means clustering (Fig. 2.5 A) should be used in future studies to compare differences in the gradation of vocal repertoires between species on a first level. We further showed that the variation in the level of gradation within and between call types can be visualized and even quantified by calculating typicality scores for each call. Whereas the visualization presents a good overview of the repertoire structure, quantification even allows the systematic comparison of the level of gradation between different species' repertoires.

In sum, although it would be desirable to have completely objective criteria to determine the optimal number of call types, this may not be possible. Therefore, especially in more graded datasets, the researcher's preference to use different features, or to either split or lump data (McKusick 1969), may also come into play. Transparency with regard to these decisions and awareness of their consequences is therefore invaluable.

5.1 Summary

We conclude that the usage of a high number of acoustic features results in better cluster solutions. The use of factors derived from PCA may result in the loss of critical information and may lead to extremely poor solutions. We therefore discourage their usage for the construction of vocal repertoires. We also showed that fuzzy clustering is a powerful tool to describe the graded structure of a species vocal repertoire. It reveals details of the graded nature of vocal repertoires that cannot be captured with classical approaches and allows a quantification of typical and atypical calls. Researchers should be aware of and transparent about the fact that the outcome of their analysis is affected by several decisions and that the choice of the eventual cluster solution eventually depends on researcher preferences and research interests. Therefore, data repositories should be used so that the same methods can be applied to different datasets. This would greatly enhance the possibilities to compare species' vocal repertoires within and across taxa.

A Appendix

Table A.1: Descriptions of all 118 acoustic features that were used in the analyses.

parameter	used in 9- / 38-feature set	description and unit
Duration	9 / 38	Duration [ms]
DFA1 st		Start frequency 1st DFA (distribution of frequency amplitude) [Hz]
DFA1 end		End frequency 1st DFA [Hz]
DFA1 max		Maximum frequency 1st DFA [Hz]
DFA1 min		Minimum frequency 1st DFA [Hz]
DFA1 mean		Mean frequency 1st DFA [Hz]
DFA1 med		Median frequency 1st DFA [Hz]
DFA1 maloc		Location of the maximum frequency 1st DFA [(1/duration)*location]
DFA 2st		Start frequency 2nd DFA (distribution of frequency amplitude) [Hz]
DFA2 end		End frequency 2nd DFA [Hz]
DFA2 max		Maximum frequency 2nd DFA [Hz]
DFA2 min		Minimum frequency 2nd DFA [Hz]
DFA2 mean	9 / 38	Mean frequency 2nd DFA [Hz]
DFA2 med		Median frequency 2nd DFA [Hz]
DFA2 maloc	38	Location of the maximum frequency 2nd DFA [(1/duration)*location]
DFA3 st		Start frequency 3rd DFA (distribution of frequency amplitude) [Hz]
DFA3 end		End frequency 3rd DFA [Hz]
DFA3 max		Maximum frequency 3rd DFA [Hz]
DFA3 min		Minimum frequency 3rd DFA [Hz]
DFA3 mean		Mean frequency 3rd DFA [Hz]
DFA3 med		Median frequency 3rd DFA [Hz]
DFA3 maloc		Location of the maximum frequency 3rd
DFA range	38	DFA3 mean âĂŞ DFA1 mean [Hz]
DFB1 st		start frequency 1st DF (dominant frequency band) [Hz]
DFB1 end		end frequency 1st DF[Hz]
DFB1 max		maximum frequency 1st DF [Hz]
DFB1 min		minimum frequency 1st DF [Hz]
DFB1 mean	9 / 38	mean frequency 1st DF [Hz]
DFB1 med		median frequency 1st DF [Hz]
DFB1 chfre	38	number of changes between original and floating average curve local modulation (LM) 1st DF
DFB1 chmea	9 / 38	mean deviation LM 1st DF [Hz]
DFB1 chmax		maximum deviation LM 1st DF [Hz]
DFB1 pr	38	percent of time segments where a 1st DF could be found [%]
DFB1 maloc	38	location of the maximum frequency 1st DF [(1/duration)*location]
DFB1 miloc	38	location of the minimum frequency 1st DF [(1/duration)*location]
DFB1 trfak	38	factor of linear trend of 1sr DF (global modulation)
DFB1 fretr	38	alternation frequency between 1st DF and linear trend
DFB1 maxtr	38	maximum deviation between 1st DF and linear trend [Hz]
DFB1 mintr		minimum deviation between 1st DF and linear trend [Hz]
DFB2 st		start frequency 2nd DF (dominant frequency band) [Hz]
DFB2 end		end frequency 2nd DF [Hz]
DFB2 max		maximum frequency 2nd DF [Hz]
DFB2 mean	38	mean frequency 2nd DF [Hz]
DFB2 med		median frequency 2nd DF [Hz]
DFB2 pr		percent of time segments where a 2nd DF could be found [%]
DFB3 mean	38	mean frequency 3rd DF [Hz]
DFB3 med		median frequency 3rd DF [Hz]
DFB3 pr		percent of time segments where a 3rd DF could be found [%]
DFB4 pr		percent of time segments where a 4th DF could be found [%]
Diff max		maximum difference between 1st & 2nd DF [Hz]
Diff mean	38	minimum difference between 1st & 2nd DF [Hz]
Diff remax		maximum number of DFâĂŹs
Diff remin		minimum number of DFâĂŹs
Diff req	38	mean number of DFâĂŹs
Ampratio 1		amplitude ratio between 1st & 2nd DF
Ampratio 2		amplitude ratio between 1st & 3rd DF
Ampratio 3		amplitude ratio between 2nd & 3rd DF
F1 mean	38	(global frequency peak) [Hz]
F2 mean	38	[Hz]

Table A.2: Descriptions of all 118 acoustic features that were used in the analyses. continued

parameter	used in 9- / 38-feature set	description and unit
F1 wst		start frequency of 1st P [Hz]
F1 wend		end frequency of 1st P [Hz]
F1 wmax		maximum frequency of 1st P [Hz]
F1 wmin		minimum frequency of 1st P [Hz]
F1 wmean	38	mean frequency of 1st P [Hz]
F1 wmed		median frequency of 1st P [Hz]
FP1 max		maximum frequency 1st P (global frequency peak) [Hz]
FP1 mean		mean frequency 1st P [Hz]
FP1 amax		maximum amplitude 1st P (global frequency peak) [rel. amplitude]
FP1 amean	38	mean amplitude 1st P [rel. amplitude]
F2 pr	38	percent of time segments where a 2nd P could be found [%]
F2 wmean	38	mean frequency of 2nd P [Hz]
F3 pr	38	percent of time segments where a 3rd P could be found [%]
Range max		maximum frequency range [Hz]
Range mean	9	mean frequency range [Hz]
Range min		minimum frequency range [Hz]
PF st		start PF (peak frequency) [Hz]
PF end		end PF [Hz]
PF max		maximum PF [Hz]
PF min		minimum PF [Hz]
PF mean	38	mean PF [Hz]
PF med		median PF [Hz]
PF totmax		frequency of the total maximum amplitude [Hz]
PF totmin		frequency of the total minimum amplitude [Hz]
PF maloc	38	location of the maximum PF [(1/duration)*location]
PF miloc	38	location of the minimum PF [(1/duration)*location]
PF jump	38	maximum difference between successive PFâĂŹs [Hz]
PF trfak	38	factor of linear trend of PF (global modulation)
PF trfre	38	alternation frequency between PF and linear trend
PF trmean	9 / 38	mean deviation between PF and linear trend [Hz]
PF trmax		maximum deviation between PF and linear trend [Hz]
CS mean	9 / 38	mean correlation coefficient of successive time segments
CS maxd		standard deviation correlation coefficient of successive time segments
CS maloc	38	location of maximum correlation coefficient of successive time segments [(1/duration)*location]
F0 mean		mean frequency F0 [Hz]
Noise	9 / 38	percentage of noisy time segments [%]
Disturb		percentage of disturbed time segments [%]
Tonal F0		percentage of tonal time segments and it is possible to estimate the F0 [%]
PF mean		mean PF (peak frequency) [Hz]
PF max		maximum PF [Hz]
PF min		minimum PF [Hz]
Diff mean		mean difference between F0 & PF [Hz]
Diff max		maximum difference between F0 & PF [Hz]
Diff min		minimum difference between F0 & PF [Hz]
Amprat1		amplitude ration between F0 & 1st harmonic
Amprat2		amplitude ration between F0 & 2nd harmonic
Amprat3		amplitude ration between 1st & 3rd harmonic
HNR1 mean		mean harmonic to noise ratio DFA1 (1= no noise)
HNR2 mean	9 / 38	mean harmonic to noise ratio DFA2 (1= no noise)
HNR3 mean		mean harmonic to noise ratio DFA3 (1= no noise)
HNR1 max		max harmonic to noise ratio DFA1 (1= no noise)
HNR2 max		max harmonic to noise ratio DFA2 (1= no noise)
HNR3 max		max harmonic to noise ratio DFA3 (1= no noise)
Shimmer mean	38	mean frequency of vocal fold vibration [Hz]
Shimmer max		max frequency of vocal fold vibration [Hz]
Jitter mean	38	mean amplitude of vocal fold vibration
Jitter max		max amplitude of vocal fold vibration
Range max		maximum frequency range [Hz]
Range min		minimum frequency range [Hz]

Table A.3: Eigenvalues of first 20 factors. Extraction Method: Principal Component Analysis

	Total Variance Explained					
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	47,619	40,355	40,355	45,231	38,332	38,332
2	8,314	7,046	47,401	7,365	6,242	44,573
3	6,569	5,567	52,968	3,765	3,190	47,764
4	5,221	4,425	57,393	3,760	3,186	50,950
5	4,207	3,565	60,958	3,728	3,159	54,109
6	2,854	2,419	63,377	3,491	2,958	57,067
7	2,439	2,067	65,444	3,162	2,679	59,747
8	2,172	1,841	67,284	2,787	2,362	62,109
9	2,016	1,708	68,993	2,747	2,328	64,437
10	1,873	1,588	70,581	2,342	1,985	66,421
11	1,683	1,426	72,007	2,295	1,945	68,367
12	1,561	1,323	73,330	2,277	1,930	70,296
13	1,462	1,239	74,569	2,275	1,928	72,224
14	1,357	1,150	75,718	2,003	1,698	73,922
15	1,293	1,096	76,814	1,765	1,496	75,418
16	1,256	1,065	77,879	1,737	1,472	76,890
17	1,169	,990	78,869	1,603	1,359	78,249
18	1,137	,964	79,833	1,547	1,311	79,560
19	1,031	,874	80,707	1,354	1,148	80,707
20	,977	,828	81,535			

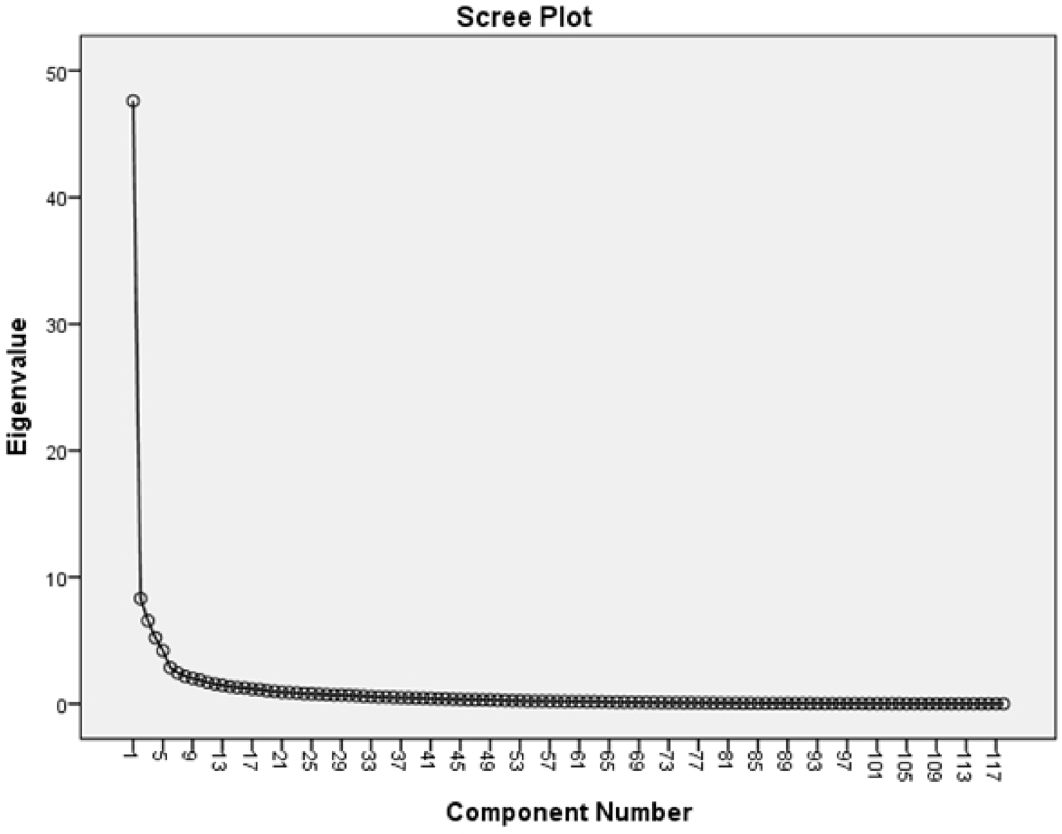

Figure A1: Scree Plot Eigenvalues of 118 factors.

Table A.4: Rotated Component Matrix. Extraction Method: Principal Component Analysis; Rotation Method: Varimax with Kaiser Normalization; Rotation converged in 21 iterations

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
Duration										,781									
DAF1 1st	,910																		
DFA1 end	,916																		
DFA1 max	,966																		
DFA1 min	,907																		
DFA1 mean	,973																		
DFA1med	,968																		
DFA1 maloc														,471				,346	
DFA2 st	,853																		
DFA2 end	,849																		
DFA2 max	,878																		
DFA2 min	,931																		
DFA2 mean	,979																		
DFA2 med	,973																		
DFA2 maloc														,761					
DFA3 st	,724						,430												
DFA3 end	,748						,355												
DFA3 max	,731				-,347		,389												
DFA3 min	,933																		
DFA3 mean	,933																		
DFA3 med	,941																		
DFA3 maloc														,764					
DFB1 st	,589					,475									-,305				
DFB1 end	,651														,332				
DFB1 max	,783					,550													
DFB1 min	,686																		
DFB1 mean	,947																		
DFB1 med	,928																		
DFB1 chfre	,614	-,402																	
DFB1 chmea	,709					,413													
DFB1 chmax	,513					,746													
DFB1 pr					,818														
DFB1 maloc															,456				
DFB1 miloc		-,448													-,382				
DFB1 trfak															,793				
DFB1 fretr																			,729
DFB1 mtr	,736					,415													
DFB1 maxtr	,578					,725													
DFB2 st	,642					,431													
DFB2 end	,696					,356													
DFB2 max	,815					,390													
DFB2 mean	,954																		
DFB2 med	,944																		
DFB2 pr		,343			,722														
DFB3 mean	,942																		
DFB3 med	,913																		
DFB3 pr		,618			,386			,309											
DFB4 pr		,765																	
Diff max	,639					,534													
Diff mean	,807																		
Diff remax	,594	,650																	
Diff remin		,702			,342														
Diff req		,862																	
Ampratio1								-,903											
Ampratio2								-,789					-,379						
Ampratio3													-,761						
F1 mean	,950																		
F2 mean	,789																		

Table A.5: Rotated Component Matrix. continued

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
F1 wst	,357								,490										
F1 wend	,401								,499			,309							
F1 wmax	,837								,361										
F1 wmin		-,339			,406				,364										
F1 wmean	,672								,609										
F1 wmed	,566								,639										
FP1 max	,925																		
FP1 mean	,944																		
FP1 amax		-,695																	
FP1 amean		-,724																	
F2 pr		,710																	
F2 wmean	,777																		
F3 pr	,443	,427														,329			
Range mean	,900																		
Range max	,756				-,302														
Range min	,507								,407										
PF st	,685											,367							
PF end	,601											,613							
PF max	,931																		
PF min	,499											,702							
PF mean	,942																		
PF med	,911																		
PF totmax	,874																		
PF totmin	,683											,452							
PF maloc							,428												
PF miloc																		,731	
PF jump	,848																		
PF trfak																		,636	
PF trfre		,327								-,506									
PF trmean	,784											-,342							
PF trmax	,849																		
CS mean					,751														
CS maxd					-,638		,343												
CS maloc														,455				-,305	
F0 mean	,704		,317																
Noise	,316						-,681												
Disturb																,593			
Tonal F0	-,343						,734												
PF mean	,694			,506															
PF max	,591			,522															
PF min	,498			,536								,333							
Diff mean	,515			,766															
Diff max	,485			,680															
Diff min	,311			,770															
Amprat1		-,361						,595			,443								
Amprat2								,544					,565						
Amprat3													,865						
HNR1 mean		-,696																	
HNR2 mean	,329		,596																
HNR3 mean			,757																
HNR1 max	-,341	-,320	,511	,342						,305									
HNR2 max			,815																
HNR3 max			,883																
Shimmer mean		-,321									,753								
Shimmer max		-,381									,745								
Jitter mean																	,840		
Jitter max																	,752		
Range mean	,637		,301	,395															
Range max	,445		,354	,361			,315												
Range min	,493	,379		,437												,343			

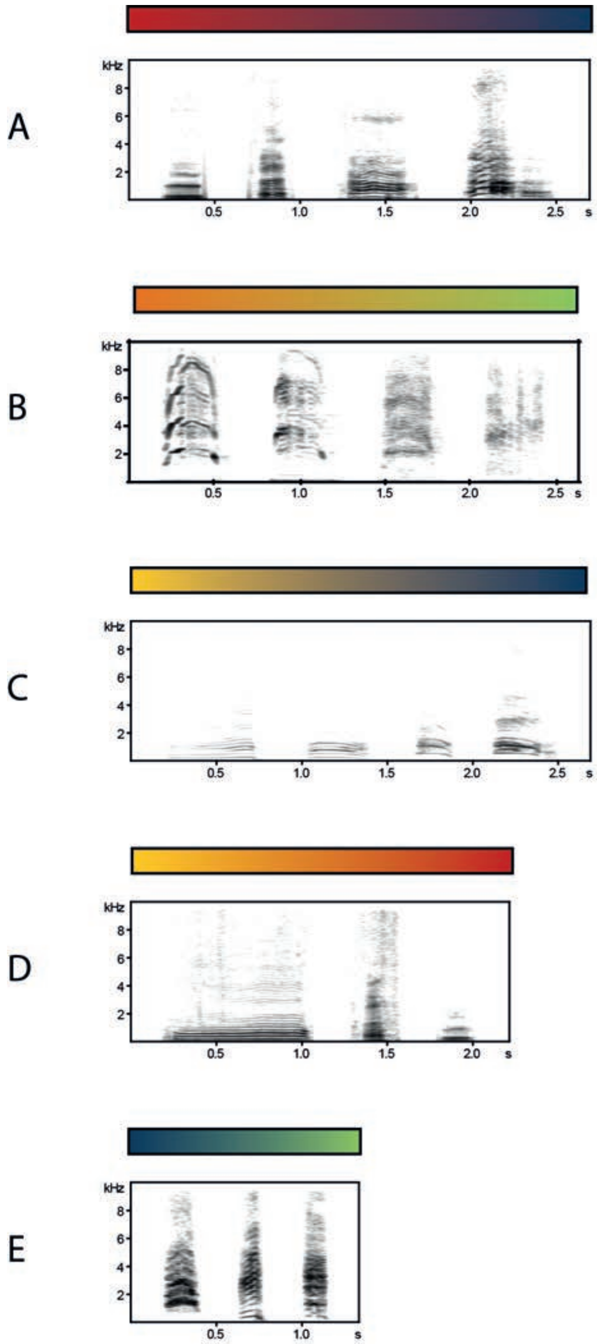

Figure A2: **Call exemplars of typical and hybrid calls.** (A) Grunt to bark. (B) Tonal scream to noisy scream. (C) Weaning call to bark. (D) Weaning call to grunt. (E) Bark to noisy scream. Colors represent the color code for call types in Figure 2.6 and 2.7.

3 | Quantifying and Comparing the Level of Gradation between Vocal Repertoires

Philip Wadewitz[1,2,3], Kurt Hammerschmidt[1], Demian Battaglia[3,4], Fred Wolf[2,3], Julia Fischer[1,3]

[1] Cognitive Ethology Laboratory, German Primate Center, Göttingen, Germany

[2] Theoretical Neurophysics, Max Plank Institute for Dynamics and Self-Organization, Göttingen, Germany

[3] Bernstein Center for Computational Neuroscience, Göttingen, Germany

[4] Aix-Marseille University, Institute for Systems Neuroscience, INSERM UMR 1106, Marseille, France

submitted

1 Abstract

A core problem in biology is to sort tokens such as haplotypes, ecological communities, or behavioral patterns that vary in multi-dimensional trait-spaces into discrete types or categories. To establish such categories, cluster analyses are frequently used although it is typically difficult to identify a unique decomposition of biological data into clusters. As a consequence, the exact number of categories remains empirically under-constrained, and the certainty with which categories can be established is hard to explicitly represent and assess. Here we develop a method to quantitatively compare the degree to which high-dimensional data sets can be partitioned into distinct clusters. Using the vocal repertoires of two nonhuman primate species as an example, we show that the distribution of typicality coefficients (DTC) enables a systematic comparison of the structure of different data sets. It allows moving on from frequently contentious statements about specific numbers of types or categories that can be identified to a quantitative assessment of the overall differentiation within a complex data set. This method may thus be useful in a wide range of biological disciplines.

2 Introduction

Categorization of objects that show a high degree of variation has already preoccupied Darwin and Hooker (Endersby 2009) and remains indispensable in many fields of modern biology. One of the core problems in such endeavors is the need to either split or lump objects that are to be categorized (McKusick 1969). Whereas splitting creates new categories by focusing on often subtle differences of samples, lumping results in broader categories, emphasizing similarities over differences. The identification of species is a typical case where different preferences for lumping or splitting appear to complicate an objective assessment of biodiversity (Isaac et al. 2004). At the same time, progress in feature detection and data technology raise the hope that the acquisition of large and high-dimensional data sets can facilitate an objective categorization of biological objects (Gomez-Marin et al. 2014; Vogelstein et al. 2014).

Unsupervised clustering is considered to be one of the best candidate tools to detect and cluster objects that share similar features (Larrañaga et al. 2006). If the data is intrinsically organized in discrete categorical classes it should in principle be possible to obtain the number and definition of these classes by an objective data driven approach. A variety of unsupervised clustering algorithms have thus been used across biological fields for exploratory data mining and statistical data analysis (Jain 2010). Most of the available algorithms are indeed based on the assumption that the computed classes reflect discrete categories (MacQueen 1967). Numerous biological systems, however, exhibit a substantial level of gradation. This is expected in particular if sample differences are produced by continuous differentiation over evolutionary (or ontogenetic) time scales (Handley et al. 2007). Hard clustering algorithms by construction are not well suited for characterizing such graded systems. In population dynamics, neighboring populations may show gradual transition zones (Evanno et al. 2005). In neurobiology, cortical neurons can display a large structural and physiological diversity which can either form a continuum or discrete cell types (Armañanzas and Ascoli 2015; Battaglia et al. 2013; Markram et al. 2004). In community ecology, plant species may show continuous dispersal patterns along climate gradients (Gauch and Whittaker 1972; Whittaker 1953). Nevertheless, unsupervised clustering is still commonly used to classify graded systems, evoking the misleading impression that objects in these systems would belong to distinct classes (Jain 2010). In addition, in such cases there is typically not one, but several equally good (or poor) cluster solutions that may result in a rather idiosyncratic choice which cluster solution to accept. Thus, across a wide range of biological problems, it would be desirable to develop a methodology that does not presume the existence of discrete categories but extracts from the structure of the data a set of indicators that characterize the degree of gradation within a system.

Here we develop a method to critically probe and quantitatively compare the degree to which a high-dimensional data set can be partitioned into distinct clusters. Our approach is based on Fuzzy set theory, in which every item of a data set is assessed with regard to its memberships of all generated clusters (Zadeh 1965). In contrast to other clustering algorithms, fuzzy clustering (Jang and Sun 1997) does not rely on strict boundaries between types as a basic assumption. In short, our approach begins by examining the quality of different cluster solutions by assessing their stability over a parameter μ that controls the level of fuzziness of the emerged clusters. The fuzzy clustering yields typicality coefficients, which quantify to which degree an item shares its membership not only with its main cluster, but also with other cluster(s) (Zadeh 1965). The critical next step that we are proposing here is the quantitative analysis of the distribution of typicality coefficients (DTC) for the obtained solutions, which provide an indicator of the level of gradation within a high-dimensional data set. Whereas discrete data sets are characterized by a majority of typical items resulting in a DTC with a prominent peak at high typicality coefficients and an overall left-skewed shape, graded data sets contain a substantial amount of items with low typicality coefficients and DTCs can range from none-skewed to even highly right-skewed, with a peak near zero. Importantly, these measures are largely independent from the number of clusters in the chosen solution.

To examine the utility of the proposed method, we compared the structure of the vocal repertoire of chacma baboons (*Papio ursinus*) to the structure of Barbary macaques (*Macaca sylvanus*). The vocal behavior of both species has been intensely studied and therefore the two repertoires serve as good model data sets to evaluate the accuracy of our approach (Fischer and Hammerschmidt 2002; Hammerschmidt and Fischer 1998;

Wadewitz et al. 2015a).

From an evolutionary perspective, the comparison of the structure of different vocal repertoires is relevant because of an interest in the driving forces in signal evolution, and consequences for processing of graded signals by signal recipients (Bailey 1994; Marler 1977; Rowell and Hinde 1962; Wadewitz et al. 2015a).

3 Results

3.1 Fuzzy c-means as a tool to compare data sets with different level of gradation

We adopted a variant of the fuzzy c-means algorithm (Battaglia et al. 2013; Jang and Sun 1997) to systematically compare our two reference primate vocalization data sets. However, to first illustrate how the approach works, we created two artificial "toy" data sets both including three distinct classes, but with highly discrete (Fig. 3.1 A) and highly graded levels of separation, respectively (Fig. 3.1 E). For small values of the fuzziness parameter μ, our clustering algorithm identified three distinct clusters for both data sets. However, whereas in the discrete data set the retrieved solution robustly continued to yield a partition into three clusters throughout the entire range of μ (Fig. 3.1 B), the graded data set dropped from three to one cluster solutions with increasing μ (Fig. 3.1 F). While all three clusters are extremely well separated in the discrete data set (Fig. 3.1 C), clusters of the graded data set show typical and intermediate objects between all three clusters (Fig. 3.1 G). The visualized segregation of clusters can be quantified by object co-memberships (how well is an object separated from the corresponding cluster) and residual co-memberships (how much of an object's membership is captured by the two corresponding clusters). The overall gradation of objects within a data set can be quantified by calculating typicality coefficients (TCs), which show constantly high values for the discrete data set even when μ is set to relatively high values (Fig. 3.1 D). In the graded data set, typicality coefficients drop significantly with increasing fuzziness (Fig. 3.1 H). These different evolution patterns of the DTCs with increasing μ provide a strong qualitative signal of the different level of gradedness of the considered toy examples, which can be identified even when analyzing data sets from real world applications.

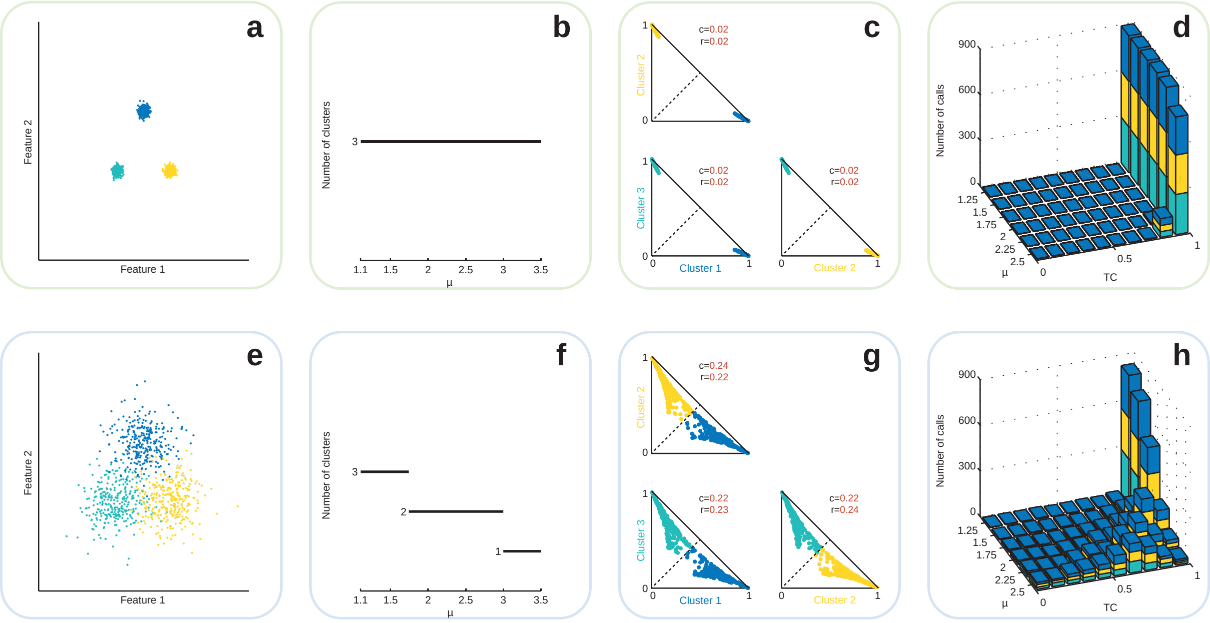

Figure 3.1: Modelled data sets to illustrate all performed calculations. To illustrate our approach, two-dimensional data sets were created that vary substantially in their level of gradation. In a first step, objects in both data sets are clustered by using the fuzzy c-means algorithm at a very low level of fuzziness (see Online Methods). By stepwise incrementing μ, nearby clusters eventually merge which leads to an overall smaller amount of cluster. This way the stability of cluster solutions can be evaluated over μ. Whereas in our discrete sample the three-cluster solution remains stable over the entire range of μ, clusters in the graded sample eventually merge to one all-embracing cluster. Once the most robust cluster solution and its corresponding μ value (here $\mu = 2.5$) are identified, cluster segregation can be visualized by pairwise comparison of memberships. If clusters are well separated, objects in the projections do not share borders (dashed diagonals). If clusters are graded, a continuum between objects of both clusters can be observed. Typicality coefficients of the cluster solutions allow the quantification of the level of gradation within our two data sets. Highly typical objects have a typicality coefficient close to 1; highly atypical objects have a typicality coefficient close to 0. For very low values of μ differences in typicality coefficients between discrete and graded data sets are not apparent since for low μ fuzzy c-means operates like any hard clustering algorithm (no intermediate forms possible). By investigating the evolution of typicality coefficients over μ, differences in the two data sets become conspicuous. Whereas in the discrete data set objects remain highly typical over the entire range of μ, in the graded data set typicality scores of objects drop significantly with increasing μ.

3.2 Determining the number of call types in the analyzed vocal repertoires

Figure 3.2 shows the number of call types in dependence of the fuzziness parameter μ for the two real data sets of chacma baboon (A) and Barbary macaque (B) calls. For the chacma baboon data set, the two-, three- and five-cluster solutions are most robust, before several smaller clusters emerge that are not very stable over μ (Fig. 3.2 A). In the Barbary macaque data set, solutions for $k > 5$ are equally unstable and lower cluster solutions are even less structured (Fig. 3.2 B). In the Barbary macaque data set, the first split into two clusters appears at a much lower value of μ (2.02 vs. 2.40) and successive cluster splits appear systematically at lower μ values than in the Barbary macaque data set (Fig. 3.2 C). This continuous pattern indicates that fuzzy c-means finds less defined and, therefore, fewer clusters in the Barbary macaque data set for any fixed resolution (i.e. level of fuzziness). In general, no clear preferable solution for $k > 3$ can be found in either of the analyzed data sets. In the subsequent analysis the graded structure of the repertoire is compared between the two species. For this purpose, cluster solutions with $k = 2 - 10$ were established with μ values at the lower end of the cluster stability range.

3.3 Describing the Gradation of Vocal Repertoires

As an example, Figure 3.3 visualizes the segregation of call types for both data sets in a 2-dimensional space. In both sets, the most stable cluster solutions with $k > 3$ have been chosen (Fig. 3.2 A&B - dashed lines). The position of each call between two select call types can be quantified by calculating each call's co-membership and residual co-membership. For chacma baboon call types, common borders can be found between weaning calls and barks and weaning calls and grunts (Fig. 3.3 A). In both cases, highly typical calls (low co-membership) can be found along with calls that appear to belong to both call types (high co-membership). Intermediate calls can also be found between tonal and noisy screams. Although typical grunts and barks exist, calls in these clusters share common borders. In contrast to the other pairs, no calls at the very edge to the other

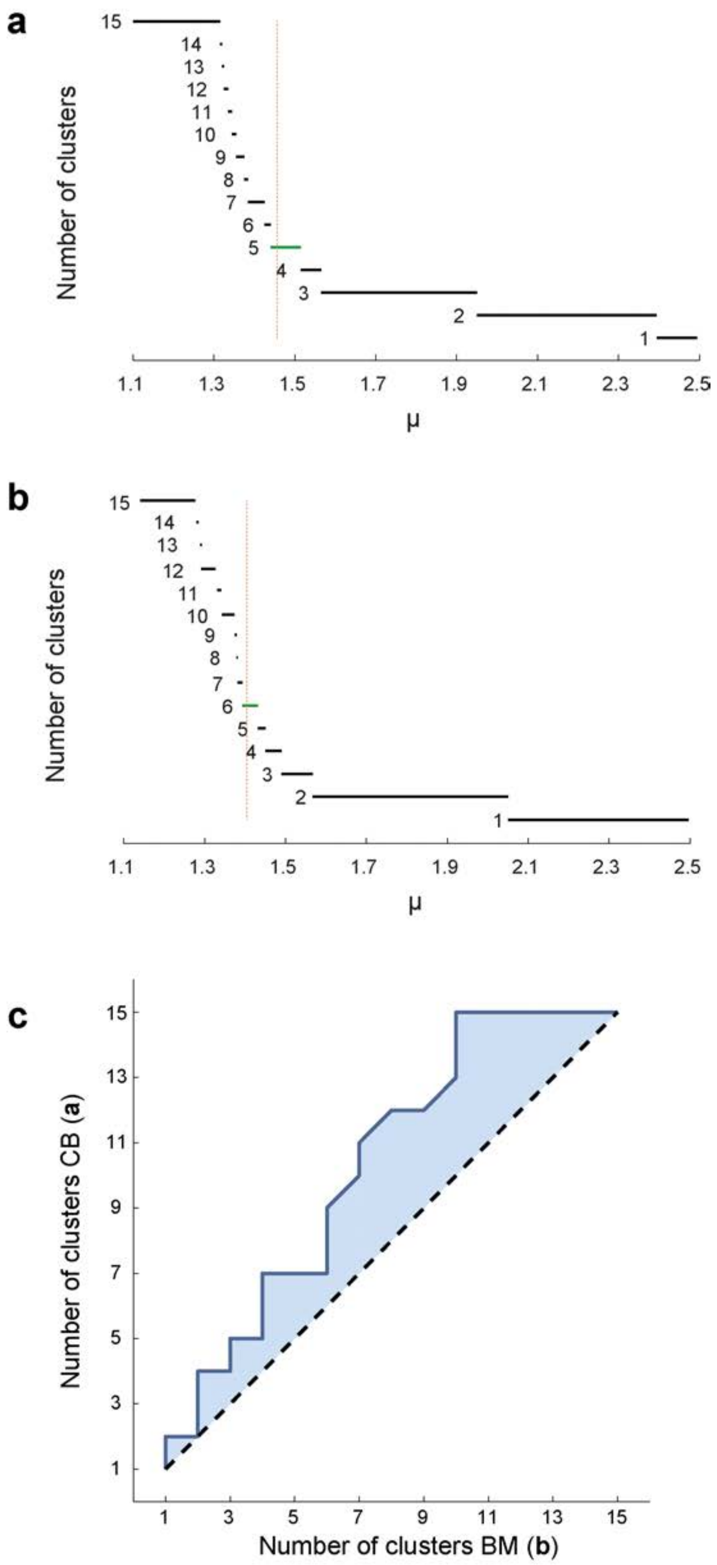

Figure 3.2: **Number of call types in dependence on the fuzziness parameter μ.** For low values of μ the initial 15 clusters remain separated. With an increase in μ, nearby clusters eventually merge and form larger clusters leading to a smaller total number of clusters. (A) In the chacma baboon repertoire, partitions with more than five call types exist only over very narrow ranges of μ values. (B) In the Barbary macaque repertoire, no clear superior cluster solution for $k > 2$ can be identified. Dashed lines indicate values for μ that have been used in the example of two-dimensional pairwise projections (Fig 3.4). (C) Combined plot of A and B. Dashed line represents the same number of call types for identical μ values in A and B. Since the parametric curve lies constantly above the dashed line, call types in the Barbary macaque data set are less well defined than call types in the chacma baboon data set for any given level of resolution.

cluster can be found and the two clusters remain separated. Calls between the grunt, bark, and weaning call clusters share most of their membership between the two corresponding clusters, which results in low residual co-memberships. Calls between the other cluster pairs share their membership between several clusters (high residual co-membership) and are therefore even more hybrid.

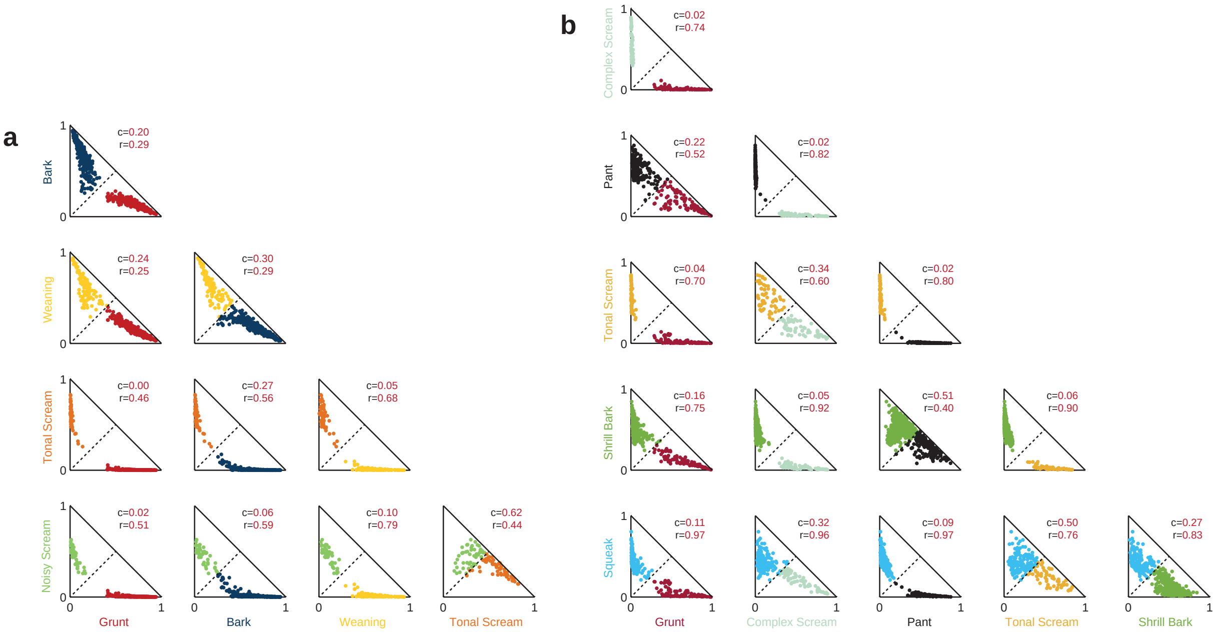

Figure 3.3: Pairwise comparisons of cluster segregations. Exemplary, two-dimensional projections of memberships of calls belonging to the most stable-cluster solutions for $k > 3$ are shown for chacma baboons (A) and Barbary macaques (B). Every call is represented by its closest and its corresponding second membership. Dashed lines in the panels represent identical memberships. For each cluster pair, means of co-memberships and residual co-memberships are shown. In chacma baboons, common boarders can be found between weaning calls and barks and weaning calls and grunts and between tonal and noisy screams. Calls in the bark- and grunt-clusters share common borders and typical grunts and barks exist. In contrast to the other pairs, no calls at the very edge to the other cluster can be found and the two clusters remain separated. Whereas calls between grunts, barks, and weaning calls share most of the membership between the two corresponding clusters (calls close to the solid diagonal; low residual co-membership), calls in the other cluster pairs are even more hybrid, lying between several clusters (high residual co-membership). In Barbary macaques, common borders can be found between almost all established call types. With some exceptions, highly typical calls can only be found in grunts. The mean residual co-memberships are systematically higher in the macaque cluster pairs, indicating that calls generally share their membership with more clusters and therefore exhibit a higher level of gradation.

In the Barbary macaque data set, common borders can be found between almost all established call types (Fig. 3.3 B) and with some exceptions, highly typical calls can only be found in grunts. The mean residual co-memberships are systematically higher in the Barbary macaque cluster pairs, indicating that calls generally share their membership with more than just one or two clusters and therefore exhibit a higher level of gradation compared to the chacma baboons. Average co-memberships and residual co-memberships of all pairwise clusters including their 95% confidence intervals can be found in the appendix.

To quantify the overall typicality of calls in both data sets and for all cluster solutions, typicality coefficients have been computed by subtracting the membership value of the second closest from the membership value of the closest call type. Results show that highly typical calls (TC close to 1) can be found along with calls that appear to belong to two or more call types and therefore have a highly atypical character (TC close to 0). Between these extreme cases a continuum of calls with varying degree of typicality coefficients is present (Fig. 3.4).

In the chacma baboon data set, typicality plots for all cluster solutions show a left-skewed distribution with an obvious mode at a typicality coefficient of > 0.5, meaning that the majority of calls are typical, i.e. well related to a main call prototype. The left-skewed distribution is quantified by the slope of the regression line for all typicality coefficient values within a cluster solution and is highest for the 2-cluster solution, drops significantly for the 3- to 5-cluster solution and increases again for the 6- and 7-cluster solution (see appendix). A further split into more clusters does not increase the slope and the overall typicality of the repertoire remains stable. In general, the grunt and bark clusters are very stable over the different solutions, with the scream and weaning calls forming clusters with mostly atypical calls (i.e. calls between cluster centers). For the Barbary macaque data set, a left-skewed distribution of typicality coefficients can only be found for the 2-cluster solution. All solutions with $k > 2$ show a significant decrease in regression line slope which becomes stronger with increasing number of clusters until a highly right-skewed distribution is reached (i.e. most of the calls have a highly atypical character, i.e. cannot be easily described as related to a unique call prototype).

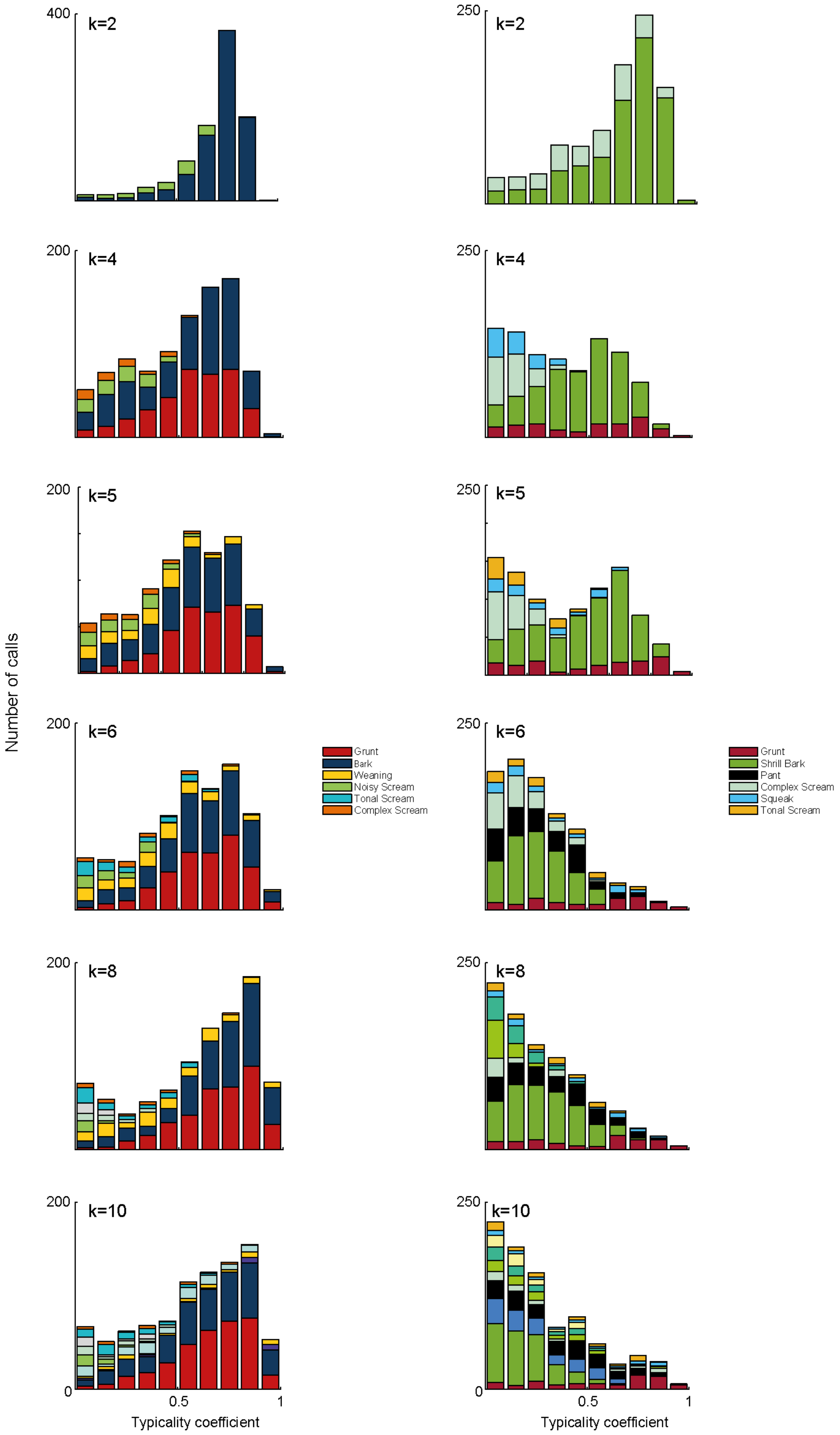

Figure 3.4: Histograms of typicality coefficients. Quantification of the overall typicality of calls in both data sets and for all cluster solutions by typicality coefficients. Typicality coefficients are calculated by subtracting the membership value of the second closest from the membership value of the closest call type. Repertoires of chacma baboons (left) and Barbary macaques (right). Shown are histograms for k = 2, 4, 6, 8 & 10 (top to bottom). Sections with different colors indicate calls of different main type. Names and color codes of call types are denoted explicitly for $k = 6$. Values of μ and regression line slope for $k = 2 - 10$ stated in table 1. In both species, repertoires with a very low number of call types consist of mainly typical calls (left-skewed distribution). Repertoires in the chacma baboon remain typical with increasing number of call types, especially grunt and bark types were very stable over the different solutions. The repertoires of the Barbary macaques show a significant decrease in typicality coefficients for all solutions with $k > 2$ until a highly right-skewed distribution is reached (i.e. most of the calls have a highly atypical character).

To quantify the distribution of typicality coefficients over k, we used different metrics (Fig. 3.5). Mean values of chacma baboon typicality coefficients are higher for all solutions with $k = 2 - 10$ in comparison to the Barbary macaque calls with no overlapping 95% confidence intervals. These differences even increase with an increase of clusters and remain relatively stable for $k \geq 6$. The finding that the DTC of the Barbary macaque repertoire decreases with the number of clusters is supported by the two other measured metrics, the mode and Kelley's measure of skewness.

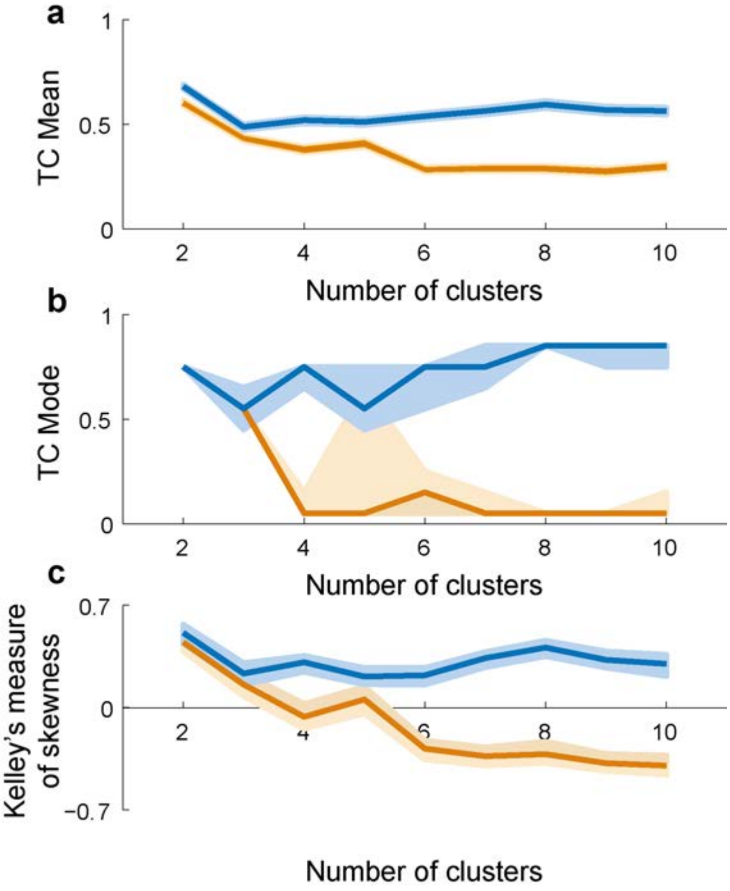

Figure 3.5: Quantification of typicality coefficients over all cluster solutions. Mean (A), Mode (B), and Kelley's measure of skewness (C) including 95% confidence intervals have been calculated. Whereas in the chacma baboon repertoire typicality remains stable over different solutions, the Barbary macaque typicality is decreasing with increasing number of call types. The slight increase in mean, skewness, and modes' confidence interval for the 5-cluster solution can be explained by the bimodal distribution of the typicality coefficients, which represents a transition stage between low-cluster solutions with mainly typical calls and high-cluster solutions with mainly atypical calls. The metrics support the findings illustrated in Fig. 3.4 that differences in typicality coefficients between repertoires of different species might only become conspicuous with higher values of k. For larger k, all three metrics eventually stabilize and give a clear image of the gradation in the two given repertoires.

The results show that for very low cluster solutions, differences in typicality coefficients might not be obvious between repertoires with varying level of gradation. However, by examining the change of the metrics over k, values eventually stabilize and give a clear image of the gradation in a given repertoire. Overall, Figures 3.2-3.5 highlight that the Barbary macaque repertoire shows a significantly higher degree of gradation independently of the (in principle arbitrary) decision on how many call types the repertoires consist of.

4 Discussion

The data-driven comparison of the vocal repertoires of Barbary macaques and chacma baboons presented above revealed clear differences in the level of gradation, confirming previous studies (Fischer and Hammerschmidt 2001; Fischer et al. 2001; Hammerschmidt and Fischer 1998), with Barbary macaques having a more graded system than chacma baboons. The assessment of the distribution of typicality coefficients (DTC) allowed us to describe the differences in the degree of gradation in a quantitative manner. Importantly, above a certain number of clusters, the derived metrics were robustly insensitive to the number of clusters obtained. Thus, the strength of our approach may lie precisely in circumventing the problem of settling on one specific cluster solution when multiple solutions are largely equivalent. Furthermore, the output of the analysis provides an intuitively accessible depiction of the structure of the data set. We suggest that the DTC method is not only promising for characterizing and comparing vocal repertoires (Bouchet et al. 2013; Freeberg et al. 2012; Krams et al. 2012), but should also be a useful tool for similar research questions, such as assessing variations of allele frequencies between and within populations (Evanno et al. 2005) or quantifying vegetation distribution patterns (Collins et al. 1993).

The DTC method allows for a quantification of the differences in the degree of gradation, largely irrespective of the number of clusters under investigation. Thus, our proposed method overcomes the limitation that in graded repertoires a unique solution can rarely be found and it is left to the investigator to settle on some rather arbitrary number of clusters. Differently from other clustering methods that automatically determine the "best" number of clusters (Blatt et al. 1996; Frey and Dueck 2007), our method does not force the convergence to a crisp classification, but preserves and exploits the extracted information about the whole spectrum of diversity present in the data set, beyond simply partitioning it into some number of classes. The analysis of the discrete artificial data set furthermore revealed that a highly robust discrete clustering solution can be found if it exists (Battaglia et al. 2013).

The fact that the most stable cluster solutions (see Fig. 3.2 A&B, k=2, 3) are not necessarily the best solutions to describe the biological structure of the data set is a major problem for characterization or comparison of different assemblages. In past analyses of vocal repertoires a two cluster solution often turned out to be the most stable solution (Hammerschmidt and Fischer 1998; Wadewitz et al. 2015a). This does not mean that a two call category solution is really appropriate to describe the communicative diversity of these repertoires (Fischer and Hammerschmidt 2001, 2002). It remains to be seen whether such stable but inappropriately low numbers of clusters also emerge in data sets of entirely different provenance.

For a comprehensive assessment of difference in overall structure between different data sets, we suggest inspecting the DTC over a range of possible cluster solutions. As we demonstrated in this study, the gradation of the two vocal repertoires does not differ significantly for solutions with a low number of clusters and the differences only become apparent if the level of resolution (i.e. number of clusters) is increased.

4.1 Comparing Vocal Repertoires

In the Barbary macaque data set, clusters split systematically at lower values of μ. This finding indicates that for any given resolution (i.e. level of fuzziness) the established Barbary macaque call types exhibited a higher level of gradation compared to the established chacma baboon call types. For instance, in the Barbary macaque data set at the fuzziness level of $\mu = 2.02$, the two previously existing call types merged to one all-embracing call type, whereas the two previously existing call types in the chacma baboon repertoire still remained separated. This indicates that already at the level of two call types, the chacma baboon repertoire showed a higher separation than the two Barbary macaque call types. Calls do not vary only between different call types, but can also exhibit considerable within-call type variation (Marler 1976; Vehrencamp 2000).

With our approach we were able to capture both, between- and within-call type variation over multiple cluster solutions as well as in detail for specific solutions. While the majority of calls in the chacma baboon repertoire shared their membership between two call types, the calls in the Barbary macaque repertoire had common borders with several call types. For the chacma baboons, all solutions showed a right-skewed distribution of typicality scores (i.e. the repertoires had a rather discrete structure), while the Barbary macaque repertoires dropped significantly in typicality coefficients for all repertoires with $k > 2$, resulting in a highly left-skewed distribution. In other words, many calls in the repertoires had a very atypical character, lying between the centers of established call types.

4.2 Implications for Acoustic Communication

Describing the gradation of vocal repertoires at a quantitative level is a prerequisite to evaluate hypotheses on the possible selective factors that drive the structure of communication systems. The results of our approach allow to re-examine some classic hypotheses, for instance, that graded repertoires should occur in species with higher visual access towards each other (Marler 1976, 1977). Moreover, the DTC method in principle provides a way to quantitatively test the hypothesis that social complexity (McComb and Semple 2005) drives vocal complexity (Bouchet et al. 2013; Freeberg et al. 2012; Krams et al. 2012). More specifically, it allows us to reconsider the notion of vocal complexity. Previous studies used the number of call types in the repertoire as a proxy (Freeberg et al. 2012), but as discussed above, this number is difficult to determine with certainty in graded repertoires, which are typical in mammalian species. Instead, we propose to use the mean typicality coefficient. It is also worth considering a more information-based notion of complexity. According to such an information-based notion, the level of gradation between different call types might be a better indicator for vocal complexity than the sheer number of call types. As expected by information theory, complex systems are neither completely ordered, nor completely disordered, but rather stand in between these two extremes (Crutchfield 2011; Tononi et al. 1998). In the case of vocal complexity, the information that a completely discrete repertoire exhibits can only lie in the existing call types itself. In contrast in a system with no differentiation, the information may be high, but difficult to distinguish from noise, and thus altogether useless for communication. Complex vocal repertoires would therefore show both, typical calls that serve as a "reference map" that guides interpretation (decoding) and atypical calls that convey additional information via variation in the acoustical structure. The metrics that we extracted from the distribution of typicality coefficients revealed that the repertoires of the two analyzed species exhibit both, typical and atypical calls and that this distribution changes significantly with an increasing number of call types in the more graded repertoire of Barbary macaques.

4.3 General Implications

The usage of typicality coefficients to describe structured data sets and the investigation of the distribution of these typicality coefficients over several possible cluster solutions allows for an objective description of the level of differentiation in a given data set. Representations in fuzzy membership space (Fig. 3.3) provide a strong visual hook on the structure of a data set, and their interpretation in terms of similarity to concrete data prototypes appears more direct and natural than for other powerful but abstract dimensional reduction approaches, such as deep learning (Hinton and Salakhutdinov 2006). Beyond the application in bioacoustics research, we believe that the DTC method may be helpful in other scientific domains where graded data sets need to be quantified, compared, and visualized. In neurobiology, the large diversity of morphological, synaptic, electrophysiological, and molecular properties of inhibitory interneurons can be quantified by the use of typicality coefficients (Battaglia et al. 2013). In population genetics, edge and core populations show a different level of diversity in haplotypes (Eckert et al. 2008) and the DTC method could be used to describe these differences in detail. In vegetation classification, different models exist that make assumptions of whether plant communities form discrete units or whether they are structured continuously (Collins et al. 1993). Typicality coefficients may be used to quantify the distribution of these vegetation patches. In sum, we suggest that this approach has a good potential to facilitate the comparison of complex structured data sets in a novel and productive fashion. More specifically, it provides an alternative to settling on rather arbitrary numbers of types or categories, and focus on the degree of differentiation within a given data set instead.

4.4 Acknowledgements

We thank the Office of the President and the Department of Wildlife and National Parks of the Republic of Botswana to conduct research in the Moremi Wildlife Reserve to get recordings of chacma baboons that we have used in this study. The calls of Barbary macaques have been recorded at "La Forêt des Singes, Rocamadour" which would have not been possible without the kind permission of Ellen Merz and Gilbert de Turckheim.

5 Methods

5.1 Data sets and Recordings

Chacma baboons

In a previous study (Wadewitz et al. 2015a), we re-analyzed call recordings that were collected during January 1998 and June 1999 in the Moremi Wildlife Reserve in Botswana. A number of comprehensive studies on the social behavior of this population has been published (Silk et al. 1999) and recordings were taken as part of an array of studies on the monkeys' vocal communication (Fischer et al. 2002). Vocalizations were recorded with a Sony WM TCD-100 DAT recorder and a Sennheiser directional microphone (K6 power module and ME66 recording head with MZW66 pro windscreen) (Fischer et al. 2002). We assembled a data set comprising of 912 calls, which we selected to capture the overall diversity of the chacma baboon's vocalizations. The selected calls were given by

45 females (41 adults, 4 infants) and 28 males (24 adults, 4 infants). For an overview of the call types included in the data set, spectrograms of typical calls are shown in Figure 3.6 A.

Barbary macaques

For the Barbary macaque data set, we used recordings that were taken in an outdoor enclosure at Rocamadour, France, between 1987 and 1993. The enclosure is a visitor park where monkeys range freely while visitors are restricted to a path. Vocalizations were recorded at a distance of 1-10 m with a Marantz cp430 or a Sony WM DC6 cassette recorder and a Sennheiser directional microphone (KN3 power module and ME80 or ME88 recording head with Sennheiser windscreen) and transferred to a DAT tape (SONY TCD-D3) for storage. We assembled a data set of 934 calls to match the sample of the chacma baboon repertoire. The selected calls were given by 41 females (34 adults, 7 infants) and 33 males (27 adults, 6 infants). Spectrograms of calls in the data set that have been described in previous studies are shown in Figure 6 B.

5.2 Call Parameterization

Calls of both data sets were fast Fourier transformed (FFT) into their frequency-time domain with Avisoft (Avisoft SASLab Pro, version 5.2.05), using a FFT size of 1024 points, Hamming window and 96.87% overlap. Depending on the frequency structure of calls we used a sampling frequency of 5 kHz (low frequency grunts of chacma baboons) or 20 kHz (all others), resulting in a frequency range of 2.5/10 kHz, a frequency resolution of 5/20 Hz and a time increment of 1.6/6.4 ms. The resulting frequency-time spectra were analysed with the software LMA 2012 developed by Kurt Hammerschmidt. For all acoustic analysis we chose a set of 38 acoustic features that broadly describe the temporal- and spectral characteristics of the vocalizations as well as the call tonality and the spectral modulation of the calls (see appendix). This set of acoustic features has proven to sufficiently describe the call morphology of the different call types in chacma baboons (Wadewitz et al. 2015b). Since Barbary macaque vocalizations have fairly similar temporal- and spectral

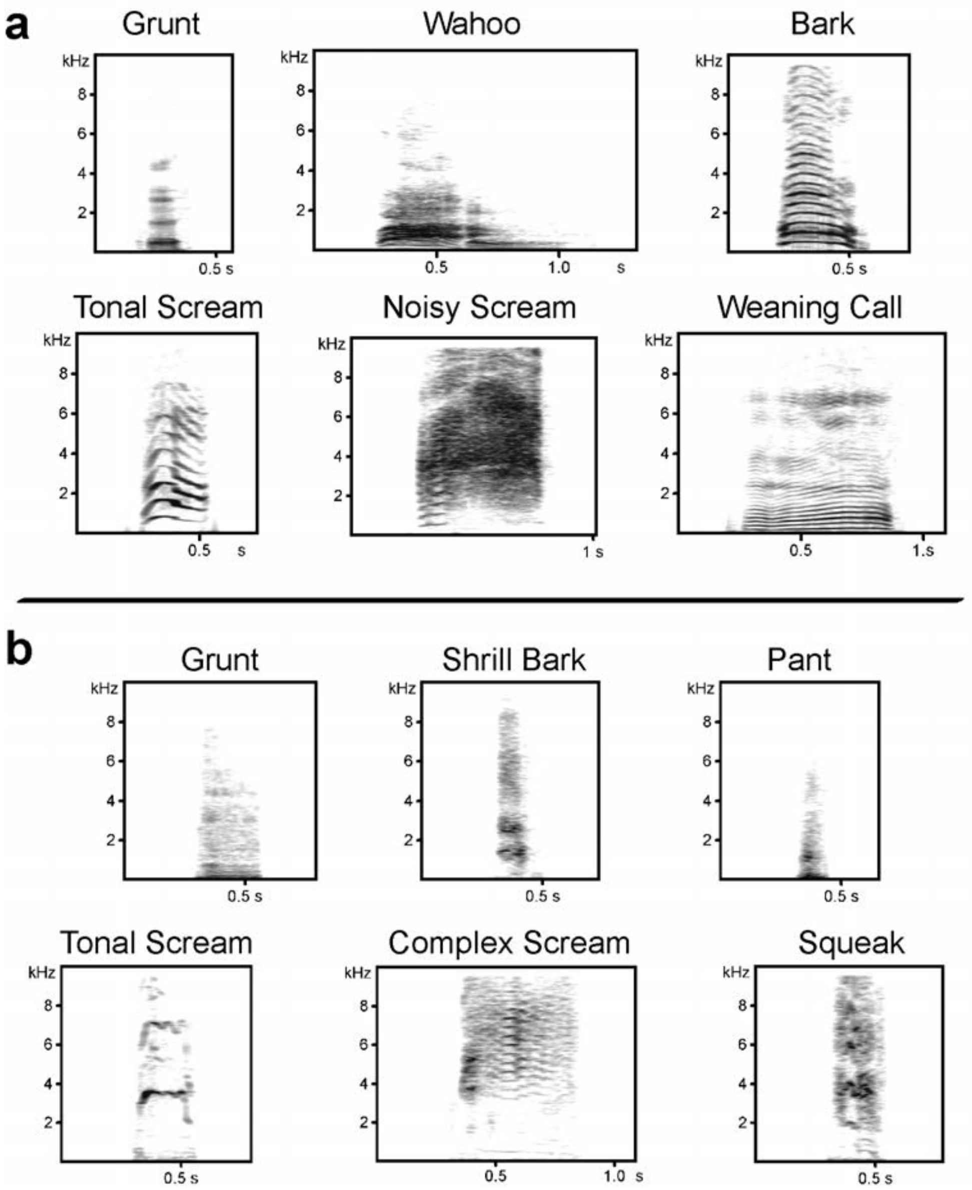

Figure 3.6: Spectrograms of calls in the used data sets. Shown are call types given by the two investigated species. (A) The chacma baboon repertoire consists of several call types that have been well described in the literature and are mostly referred to as being rather discrete. Next to more noisy calls like wahoos and the majority of screams, chacma baboons also possess several tonal call types such as barks, grunts, weaning calls and some tonal screams. (B) The vocal repertoire of Barbary macaques consists of calls that seem to be much noisier and exhibit a high degree of variation. Prior studies have shown that the categorization of Barbary macaque calls proves to be difficult. The repertoire consists of a variety of different scream types as well as lower frequency calls such as grunts and pants.

characteristics (Hammerschmidt and Fischer 1998) the same acoustic features have been used for the analysis.

5.3 Fuzzy c-means clustering

Fuzzy set theory (Zadeh 1965) extends conventional set theory allowing for the notion of imperfect membership. In this way, it is particularly suited to the clustering of data in which the separations between different classes of data-points is gradual rather than sharp (Zadeh 2008). Each call is associated an assigned membership value for each of the clusters, ranging from $m = 1$ (fully displays the properties of the cluster) and $m = 0$ (does not display any of the properties of the cluster). Intermediate membership values $0 < m_{ia} < 1$ mark calls that do not fully belong to one of the clusters, but can be classified as intermediates between different call types. Membership vectors are normalized in such a way that $\sum_{\alpha=1}^{c} m_{i\alpha} = 1$.

More specifically, we adopted a fuzzy c-means algorithm (Jang and Sun 1997; Xu et al. 2008). To determine the number of clusters that describe the dataset best, two parameters of the algorithm can be adjusted. The first parameter is the maximal number of clusters allowed and the second is the fuzziness parameter μ. If $\mu = 1$, the extracted clusters are very crisp and membership values of data points are either 1 or 0 (in this limit indeed fuzzy c-means converges exactly to k-means). However, by increasing μ, clusters become fuzzier and nearby clusters can eventually merge, unlike in k-means, leading to a smaller number of clusters. We assumed a relatively large possible number of clusters $c = 15$ (larger than the number of reasonably detectable clusters) and increased μ until all calls were grouped indistinctively into one fuzzy cluster ($\mu = 2.5$).

Similar to k-means, the fuzzy c-means algorithm builds up clusters by an iterative optimization process. In contrast to k-means, where objects do either belong or not belong to a cluster, in fuzzy c-means membership vectors $m_i^{(t)}$ for c clusters are computed at a given iteration t. Cluster centroids are given by vectors $u_\alpha^{(t+i)}(\alpha = 1...c)$ with components $u_{\alpha l}^{(t)}$.

$$\frac{1}{m_{i\alpha}^{(t)}} = \sum_{\lambda=1}^{c} \left(\frac{d_{i\alpha}^{(t)}}{d_{i\lambda}^{(t)}} \right)^{\frac{2}{\mu-1}} \tag{3.1}$$

where $d_{i\lambda}^{(t)}$ is the Euclidean distance between the data-point f_i and the centroid $u_{\lambda}^{(t)}$ at a given iteration t.

These membership vectors are used in turn to compute a new set of cluster centroids $u(t+1)$ with coordinates:

$$u_{\alpha l}^{(t+1)} = \frac{\sum_{i=1}^{N} (m_{i\alpha}^{(t)})^{\mu} f_{il}}{\sum_{i=1}^{N} (m_{i\alpha}^{(t)})^{\mu}} \tag{3.2}$$

This procedure is designed to minimize a specific cost function (Dunn 1973), namely the sum of the squared distances of the data-points from the different centroids, weighted by the relative fuzzy memberships:

$$J^t = \sum_{i=1}^{N} \sum_{\lambda=1}^{c} (m_{i\lambda}^{(t)})^{\mu} \times (d_{i\lambda}^{(t)})^2 \tag{3.3}$$

In practice we randomly initialized a collection of 15 cluster centroids $u_{\alpha}^{(0)}$ in the feature space, by selecting 15 arbitrary data-points f_i. Initial membership vectors $m_i^{(0)}$ were then computed and the procedure was iterated until the positions of the 15 centroids converged to a fixed point (with a prescribed tolerance) or until a fixed maximal number of iterations was reached. As an extension to the original fuzzy c-means strategy (Jang and Sun 1997), the final set of centroids was then inspected to identify potential coalescences and drop redundant centroids. Following our previous studies (Battaglia et al. 2013; Wadewitz et al. 2015a), whenever the Euclidean distances between different centroids was smaller than a tolerance threshold (set to $\epsilon = 0.01$), the associated fuzzy classes were merged, and the membership vectors of data-points correspondingly shrunk to a length $c^* < c$, by adding up memberships of the merged classes. Thus, given a data set and a maximum number of

15 allowed clusters, the effective number c^* of clusters in the final fuzzy partition depended on μ.

5.4 Call Type Memberships

Once the fuzziness parameter μ is set and the clusters (i.e. call types) have been computed, the main type α for each call i is the call type with the highest assigned membership component $m_{i\alpha} = m_i^{1st}$. In a pairwise cluster comparison, the co-membership of every call between its main type α and corresponding cluster β is calculated by:

$$c(i) = \frac{min(m_{i\alpha} - m_{i\beta})}{max(m_{i\alpha} - m_{i\beta})} \tag{3.4}$$

Its residual co-membership is calculated by:

$$r(i) = \frac{1 - m_{i\alpha} - m_{i\beta}}{max(m_{i\alpha} - m_{i\beta})} \tag{3.5}$$

The interval of both indices lies between 0 and 1. C-values close to 0 indicate a high separation of the call from its corresponding cluster. R-values close to 0 indicate that the membership of the call is shared by several clusters. By subtracting the second largest membership component m_i^{2nd} from the first, we get the typicality coefficient for each call, which represents the overall typicality of a call:

$$TC(i) = m_i^{(1st)} - m_i^{(2nd)} \tag{3.6}$$

The typicality coefficient is bounded in the interval $0 \leq TC \leq 1$. A typicality coefficient of 1 indicates that the call lies right in the center of a call type and therefore highly typical, a typicality coefficient of 0 indicates that the call lies exactly between the centers of two call types and has an intermediate morphological structure (i.e. is a highly atypical call).

In this way, the typicality coefficient can be used to quantify how discrete or graded call types are within a repertoire. Note that here we phrase our notion of typicality simply in terms of the fuzzy memberships themselves unlike in other available definitions (Lesot et al. 2006).

To see how typicality coefficients are affected by the decision on one of the possible cluster solutions, typicality coefficients have been calculated for $k = 2 - 10$. To compare the two data sets, regression lines and their slopes for each cluster solution have been calculated (estimator: ordinary least squares).

5.5 Quantification of Typicality Coefficients

Different metrics have been used to describe the developing of typicality coefficients over k. Besides the arithmetic mean including 95% confidence intervals, and the mode (value that occurs most often in the data set), we calculated Kelley's measure of skewness as an additional percentile based measure:

$$s_k = \frac{P_{10}+P_{90}-2\times P_{50}}{P_{10}-P_{90}} \tag{3.7}$$

where P_{10},P_{50}, and P_{90} are the 10^{th}, 50^{th}, and 90^{th} percentile of the ordered list of typicality coefficients for a given k. Positive values for s_k indicate negative skewness (more typical calls), whereas negative values indicate a positive skewness (more atypical calls).

A Appendix

Table A.1: Descriptions of all 38 acoustic features used in the analysis.

Acoustic Feature	Description and unit
(01) Duration	Duration [ms]
(02) DFA range	Frequency Range (DFA3-DFA1) [Hz]
(03) DFA2 mean	2nd quartile of frequency amplitudes distribution [Hz]
(04) DFA2 maloc	Location of the maximum frequency 2nd DFA [Hz]
(05) DFB1 mean	1st dominant frequency band [Hz]
(06) DFB1 chfre	Number of changes between original and floating average curve local modulation 1st DF
(07) DFB1 chmean	Deviation local modulation 1st DFB [Hz]
(08) DFB1 pr	Time segments where a 1st DF could be found [%]
(09) DFB1 maloc	Location of the maximum frequency 1st DFB [Hz]
(10) DFB1 miloc	Location of the minimum frequency 1st DFB [(1/duration)*location]
(11) DFB1 trfak	Factor of linear trend of 1st DFB
(12) DFB1 fretr	Alternation frequency between 1st DF and linear trend
(13) DFB1 mtr	Max deviation between 1st DFB and linear trend [Hz]
(14) DFB2 mean	2nd dominant frequency band [Hz]
(15) DFB3 mean	3rd dominant frequency band [Hz]
(16) Diff mean	Minimum difference between 1st & 2nd DFB [Hz]
(17) Diff req	Mean number of DFâĂŹs
(18) F1 mean	1st global frequency peak [Hz]
(19) F2 mean	2nd global frequency peak [Hz]
(20) F1 wmean	Mean frequency 1st peak [Hz]
(21) FP1 mean	Mean frequency 1st peak [Hz]
(22) FP1 amean	Mean amplitude 1st P [rel. amplitude]
(23) F2 pr	Time segments where a 2nd P could be found [%]
(24) F2 wmean	Mean frequency 2nd peak [Hz]
(25) F3 pr	Time segments where a 3rd P could be found [%]
(26) PF mean	Peak frequency [Hz]
(27) PF maloc	Location of the maximum PF [(1/duration)*location]
(28) PF miloc	Location of the minimum PF [(1/duration)*location]
(29) PF jump	Maximum difference between successive PFâĂŹs [Hz]
(30) PF trfak	Factor of linear trend of PF
(31) PF trfre	Alternation frequency between PF and linear trend
(32) PF trmean	Deviation between PF and linear trend [Hz]
(33) CS mean	Correlation coefficient of successive time segments
(34) CS maloc	Location of maximum correlation coefficient of successive time segments [(1/duration)*location]
(35) Noise	Noisiness [%]
(36) Hnr2	Mean signal to noise ratio (1=no noise) [%]
(37) Shimmer mean	Mean frequency of vocal fold vibration [Hz]
(38) Jitter mean	Mean amplitude of vocal fold vibration [rel. amplitude]

Table A.2: Average co-memberships and residual co-memberships including 95% confidence intervals of chacma baboon calls.

Cluster1	Cluster2	Co-Memb	95% CI	Residual	95% CI
Grunt	Bark	0.2012	0.2012 0.2019	0.2858	0.2848 0.2862
Grunt	Weaning	0.235	0.2342 0.2351	0.2464	0.2457 0.2468
Grunt	Tonal Scream	0.0099	0.0099 0.0101	0.4556	0.4543 0.4563
Grunt	Noisy Scream	0.0236	0.0234 0.0238	0.5057	0.5057 0.5080
Bark	Weaning	0.3005	0.2997 0.3010	0.2903	0.2888 0.2905
Bark	Tonal Scream	0.0272	0.0271 0.0276	0.5635	0.5623 0.5649
Bark	Noisy Scream	0.0636	0.0635 0.0645	0.5907	0.5905 0.5931
Weaning	Tonal Scream	0.0511	0.0504 0.0513	0.6771	0.6747 0.6784
Weaning	Noisy Scream	0.1029	0.1016 0.1033	0.7887	0.7868 0.7911
Tonal Scream	Noisy Scream	0.6214	0.6201 0.6228	0.4381	0.4330 0.4394

Table A.3: Average co-memberships and residual co-memberships including 95% confidence intervals of Barbary macaque calls.

Cluster1	Cluster2	Co-Mem	95% CI	Residual	95% CI
Grunt	Complex Scream	0.0244	0.0243 0.0247	0.7434	0.7409 0.7460
Grunt	Pant	0.222	0.2216 0.2232	0.5219	0.5217 0.5244
Grunt	Tonal Scream	0.0357	0.0353 0.0359	0.6973	0.6943 0.6990
Grunt	Shrill Bark	0.1451	0.1449 0.1458	0.7527	0.7513 0.7535
Grunt	Squeak	0.1147	0.1137 0.1149	0.972	0.9699 0.9745
Complex Scream	Pant	0.0192	0.0192 0.0195	0.8223	0.8223 0.8258
Complex Scream	Tonal Scream	0.3361	0.3359 0.3384	0.5954	0.5937 0.5983
Complex Scream	Shrill Bark	0.0507	0.0505 0.0509	0.9207	0.9194 0.9218
Complex Scream	Squeak	0.3226	0.3207 0.3228	0.9608	0.9606 0.9655
Pant	Tonal Scream	0.0196	0.0193 0.0197	0.7996	0.7977 0.8009
Pant	Shrill Bark	0.5143	0.5138 0.5148	0.3975	0.3961 0.3980
Pant	Squeak	0.0941	0.0935 0.0944	0.9687	0.9669 0.9701
Tonal Scream	Shrill Bark	0.0636	0.0635 0.0639	0.8956	0.8939 0.8962
Tonal Scream	Squeak	0.4989	0.4977 0.4999	0.7632	0.7627 0.7673
Shrill Bark	Squeak	0.2659	0.2646 0.2659	0.8301	0.8288 0.8309

Table A.4: **Values of μ and slope of regression line for $k = 2 - 10$.**

k	**chacma baboons**		**Barbary macaques**	
	μ	**slope**	μ	**slope**
2	1.98	19.7	1.56	11.6
3	1.54	-1.5	1.46	-7.5
4	1.46	2	1.44	-13.7
5	1.42	1.3	1.4	-11.4
6	1.4	4.5	1.36	-24.4
7	1.38	7	1.34	-23.9
8	1.36	10.5	1.332	-23.7
9	1.344	7.6	1.324	-25.3
10	1.34	7.2	1.3	-22.8

4 | The effect of subglottic pressure- and size-dependent variations in animal vocalization on the ability to retrieve existing vocal types

Philip Wadewitz[1,2], Kurt Hammerschmidt[1], Tobias Riede[3], Anil Palaparthi[4], Ingo Titze[4], Julia Fischer[1,2]

[1] Cognitive Ethology Laboratory, German Primate Center, Göttingen, Germany

[2] Bernstein Center for Computational Neuroscience, Göttingen, Germany

[3] Department of Physiology, Midwestern University, Glendale, AZ 85308, USA

[4] National Center for Voice and Speech, University of Utah, Salt Lake City, UT 84111, USA

in preparation for submission

1 Abstract

Objective characterizations of vocal repertoires are a prerequisite to understand the proximate and ultimate causes that shape acoustic communication in animals, as they provide the foundation for comparative analyses among individuals, populations and taxa. To achieve progress in this field it is important to standardize the methodological approach. In two former studies we evaluate the influence of different cluster procedures and the effect of acoustic feature selection on the result of on acoustic analysis. In this study we focused on the effect of call selection, how the inclusion of arousal- and size-dependent variation influences the ability to retrieve given vocal types. We create different repertoires using a larynx finite element model based on the estimated anatomy of a baboon larynx and varied arousal and size of possible callers. Our results showed that whereas moderate variation in arousal has only a minor impact on the ability to retrieve the given vocal types, differences in body size made the recognition of the existing three vocal types impossible. Higher variation of arousal strengthens the negative effect of size. We suggest limiting size variation to improve vocal type recognition.

2 Introduction

In acoustically communicating species, the analysis of vocal repertoires is a widespread approach to investigate how communicative systems are shaped by possible selective factors like environmental conditions or the social structure of a species. Despite the large interest in an objective methodology to analyze vocal repertoires, some major hindrances still remain. One of these hindrances is the determination of the ultimate number of call types within a vocal repertoire. Since many vocal repertoires show a considerable amount of variation within and between call types, commonly used clustering techniques fail in providing satisfying results. With an approach that we currently proposed that is based on typicality coefficients we were able to describe and compare the graded structure of the vocal repertoires of two selected nonhuman primates in detail (Wadewitz et al. 2015b). In addition to individual decisions that can influence the result of a constructed repertoire (e.g. the selection of acoustic features or the choice of one supposedly optimal number of call types), another important aspect concerns the size of the analyzed dataset and the selection of recordings. Since differences in physiology and arousal have been shown to influence the structure of calls (Fischer et al. 2002; Ey et al. 2007; Bouchet et al. 2010), the composition of the dataset regarding factors like sex or age of the individuals will presumably have a compound effect on the shape of the constructed vocal repertoire. In this study, we aim to investigate the effect of these individual and arousal based variations in animal calls by creating controlled sets of vocal repertoires that cover arousal- and size-dependent differences. To create these repertoires we used a larynx finite element model based on the estimated anatomy of a baboon larynx. The model simulates oscillating vocal folds positioned within a laryngeal cartilaginous framework, applies intrinsic laryngeal muscle activations (Alipour and Titze 1999) and includes a wave propagation model of the vocal tract (Story and Titze 1995). The created sets are subsequently quantified using a novel approach that we proposed to compare vocal repertoires in a recent study (Wadewitz et al. 2015b).

3 Methods

3.1 General Design of the Finite Element Model

To create controlled sets of pseudo repertoires that cover natural individual variation and arousal based variation in animal calls, we used a finite element (FE) model of vibrating vocal fold tissue which is continuously developed by Titze and colleagues and has been successfully used to model vocalizations of cervids (Titze and Riede 2010). The model is based on the combination of physical modeling of tissue and air movement with physiologic modeling that progresses first from muscle activation to muscle mechanics, then to cartilage and soft tissue posturing, then to self-sustained oscillation of tissue, then to glottal airflow, and finally to wave propagation in the vocal tract. A full description of the soft tissue simulation and its physical properties can be found in (Titze and Riede 2010). A schematic overview of the working model is given in Figure 4.1. The model is superior over other existing vocal fold models by its simulation of vocal fold posturing with realistic biomechanics and muscle activation.

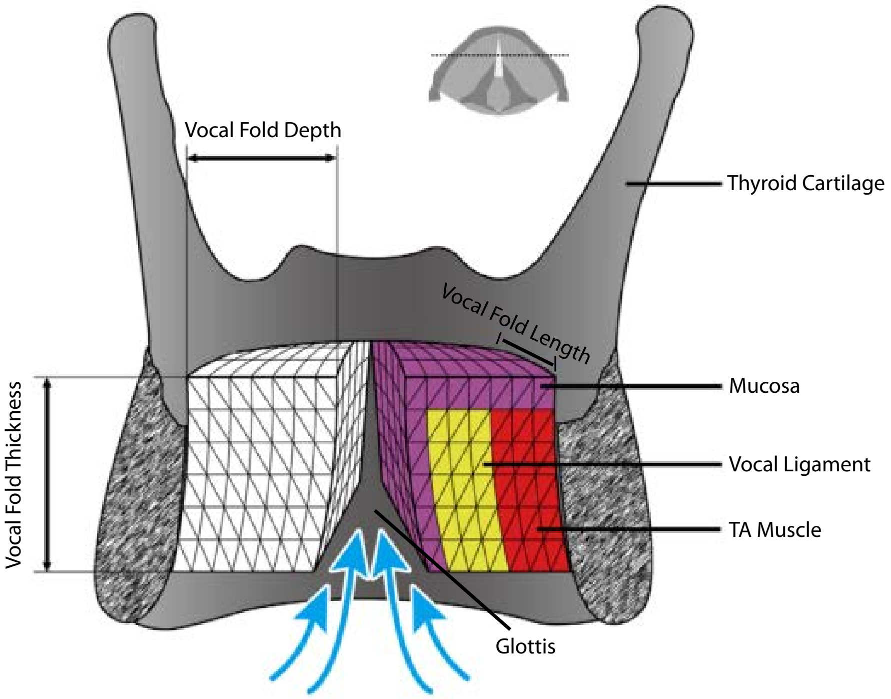

Figure 4.1: 3-dimensional FE model of the vocal folds. Frontal section through the thyroid cartilage and the vibrating portion of the vocal folds. Mucosa, ligament, and TA muscle are shown in color for the left vocal fold. *Taken and modified from (Titze and Riede 2010)*

.

3.2 Laryngeal Structures in the Model

Inside the larynx, interactions between the laryngeal muscles and cartilages determine the shape of the vocal folds and are therefore crucial for phonation. The four critical cartilages that can be found in the model are the thyroid, cricoid, and two arytenoid cartilages. The intrinsic laryngeal muscles that activities can be simulated in the model are the cricothyroid muscle (CT), the thryoarytenoid muscle (TA), the interarytenoid muscle (IA), the lateral crico-arytenoid muscle (LCA), and the posterior cricothyroid muscle (PCA). A detailed description of the laryngeal cartilage and muscle structures is given in Fig. 4.2. The activities of each mentioned muscle could range from 0 to 1 (0-100%) and the effects of adduction/elongation of them are described in Table 4.1. In our approach, vocal fold dimensions were set to 10mm length, 5mm thickness, and 4.5mm depth. Since we did not have exact measures on baboon vocal fold characteristics, we based our estimates on measurements from other species (Riede et al. 2005; Pfefferle and

Fischer 2006).

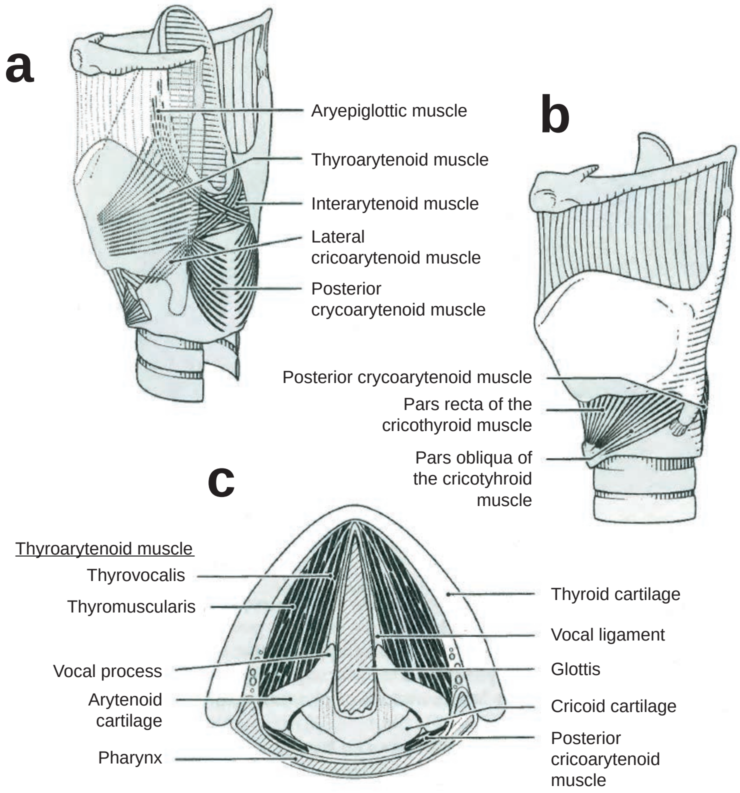

Figure 4.2: Intrinsic muscles and cartilages of the larynx. (A) posterior-lateral view, (B) anterior-lateral view, and (C) superior view. *Taken and modified from (Titze 1994)*

.

Table 4.1: Description of laryngeal muscles

Muscle	**Function**	**Effect**
Cricothyroid muscle **CT**	Tilts back the upper border of the cricoid cartilage lamina -> Tension and lengthening of vocal folds	Increases voice pitch
Lateral cricoarytenoid muscles **LCA**	Adduct and internally rotate the arytenoid cartilages -> Adduction of vocal folds	Closes glottis
Thyroarytenoid muscle **TA**	Draws the arytenoid cartilages forward toward the thyroid -> Relaxation and shortening of the vocal folds	Lowers voice pitch
Interarytenoid muscles **IA**	Adducts the arytenoid cartilages -> Adduction of vocal folds	Closes glottis
posterior cricoarytenoid muscle **PCA**	Abducts and externally rotates the arytenoid cartilages -> Abduction of vocal folds	Opens glottis; Responsible for breathing

3.3 Vocal Tract Design

In mammals the vocal tract consists of the laryngeal cavity, the pharynx, the oral cavity, and the nasal cavity and has a filtering function of the sound that is produced at the larynx. Filtering is generated by resonances of the vocal tract that allow certain frequencies (formants) to pass and radiate from the mouth better than others. Which frequencies are amplified and which suppressed is highly dependent on the shape of the vocal tract (Story et al. 1996; Riede and Titze 2008).

In our approach we worked with a simple model of the vocal tract since we did not have precise measures of nonhuman primate vocal tract geometry. The supraglottal tract was modeled with 50 tubelets of equal length for a total length of 19.9*cm* (Fig. 4.3 B). The vocal tract length was based on existing measurements of hamadryas baboons (Pfefferle and Fischer 2006). The shape of the oral cavity was designed to simulate a simplified version of the vocal tract geometry of a human producing an /a/-vowel with jaws separated and lips open (Fig. 4.3 A).

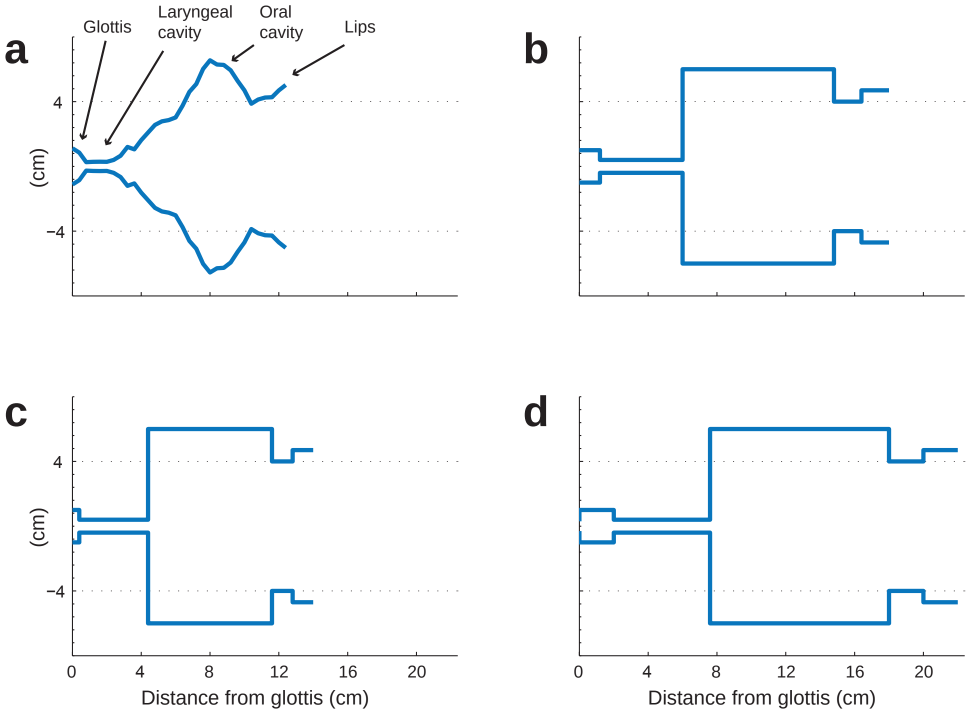

Figure 4.3: Schematic drawing of the vocal tract (VT) simulation. (A) Measurement of human VT when producing an /a/ vowel and description of the different sections. (B) VT simulation for a medium sized animal. (C) VT simulation for a small sized animal. (D) VT simulation for a large sized animal.

3.4 Created Datasets

Datasets with different level of gradation

To create controlled repertoires with variation of call structure within and between call types we aimed to simulate three known call types given by chacma baboons (Wadewitz et al. 2015a) by adjusting laryngeal muscle activities and subglottic pressure of the model. To identify the settings of these features that match our desired call templates, we first created a muscle activation plot (MAP). The MAP plots CT activity against TA activity and indicates the relationship between these two muscle activities, subglottic pressure, and the fundamental frequency. Fundamental frequency lines indicate muscle activity settings that produce specific fundamental frequencies (Fig. 4.4). To calculate these lines, calls for all possible combinations of TA and CT activities in steps of 10% were simulated for 750ms duration. Each simulation created a wav-file which was visually inspected and the fundamental frequency was calculated using the software PRAAT (Boersma and Heuven 2001). Based on these calculations, the fundamental frequency lines were calculated via linear interpolation. Additionally, for every simulation, signal to noise ratio of the simulated call was measured in PRAAT and color coded in the MAP. Highlighted in the MAP are settings to create the three simulated call prototypes. In figure 4.5 the original templates and their corresponding simulations are shown.

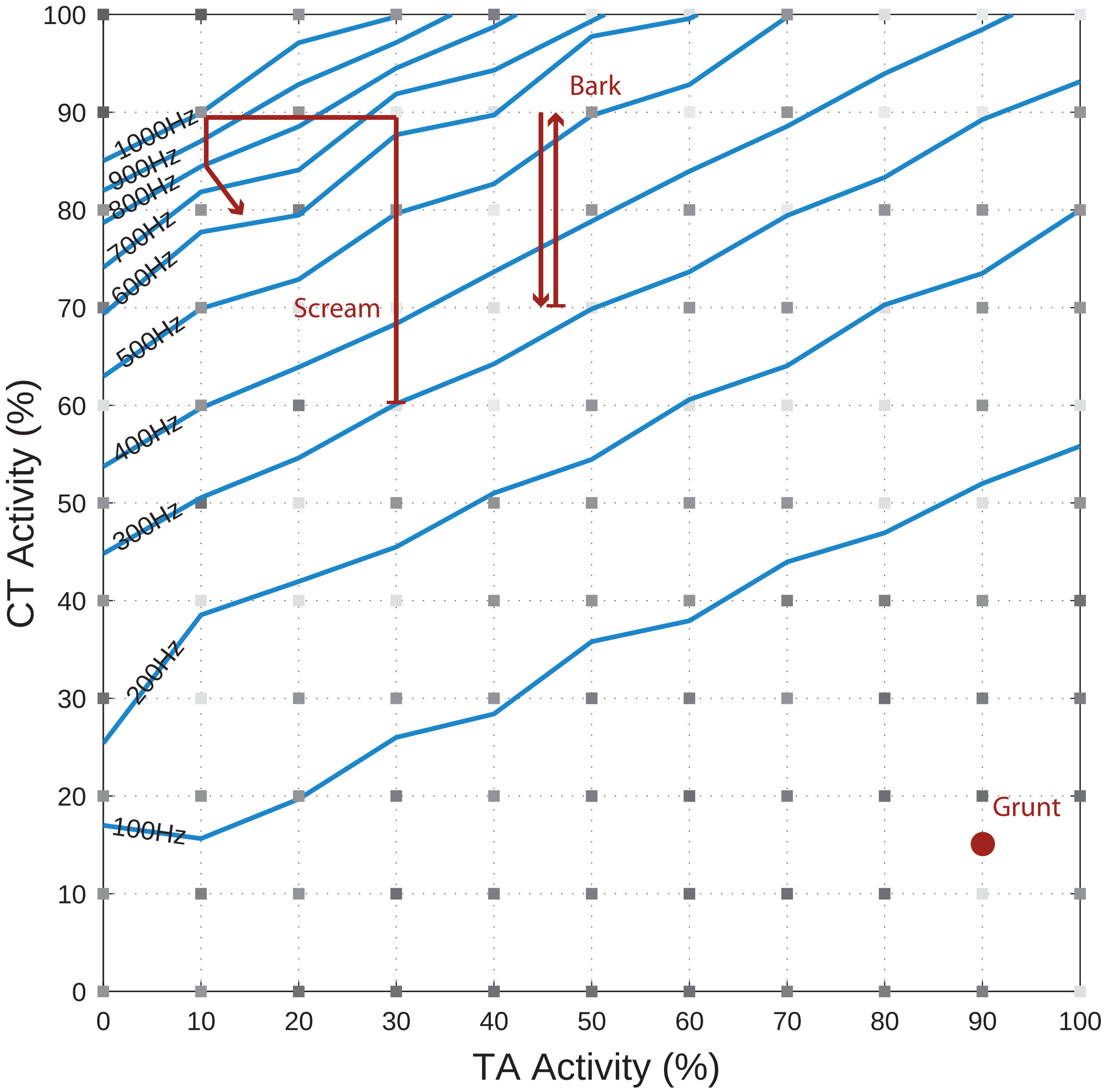

Figure 4.4: Muscle activation plot (MAP). Fundamental frequency lines for vocal fold oscillation based on simulations with the finite element model. Gray scaled squares indicate signal-to-noise ratio of the simulated sounds (Darker squares indicate lower signal-to-noise ratios). Settings of muscle activity for the four simulated calls are indicated in the plot.

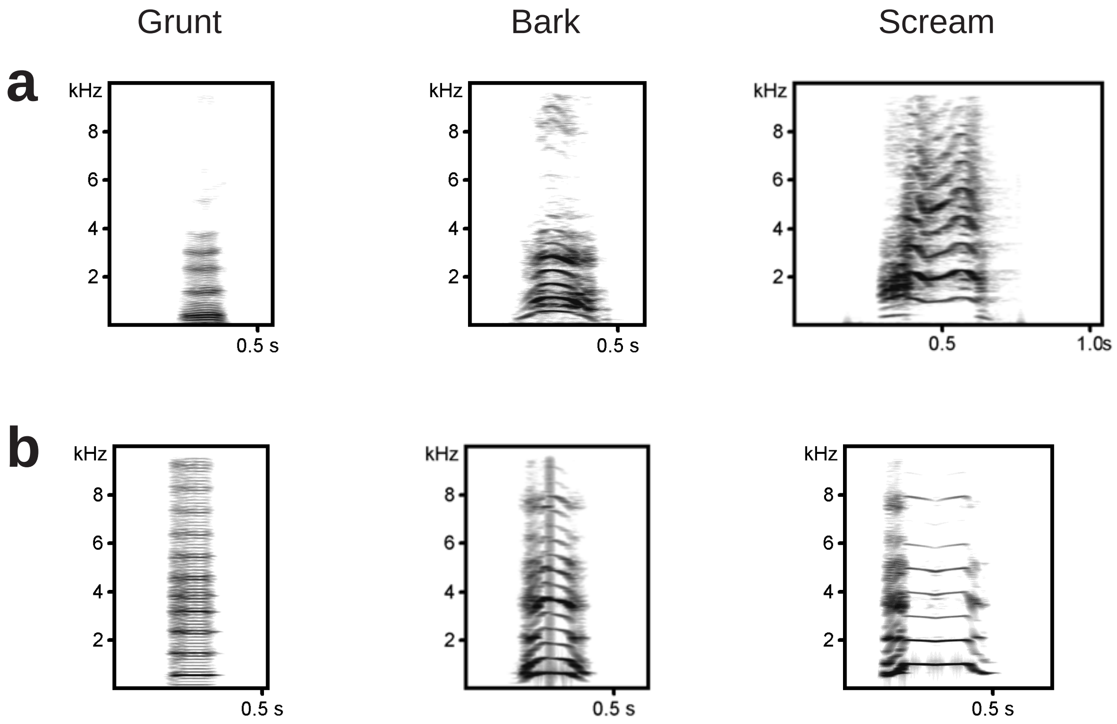

Figure 4.5: Spectrograms of recorded call templates (A) and their simulated calls (B). Call simulations were created with the laryngeal finite element model and vocal fold and vocal tract settings for a medium sized individual.

Modelling Arousal- and Size-dependent Differences in Call Structure

To simulate arousal based differences in call structure, 23 simulations of each call type have been constructed with different subglottic pressures. For all call types subglottic pressures ranged from $0.78kPa$ to $2.94kPa$. Subglottic pressure data during phonation are available for a few mammals (in vivo measurements: human: $0.3 - 6kPa$ (Bouhuys et al. 1968); bat: $0.5 - 7kPa$ (Fattu and Suthers 1981); horses: $0.5 - 8kPa$ (Rakesh et al. 2008); excised larynx experiments in various species: $0.3 - 5kPa$ (e.g. in squirrel monkeys: Brown et al. 2003). We therefore considered our simulated range of subglottic pressure level as a realistic estimate when low to high effort is exerted. The change of subglottic pressure level is meant to simulate a reduction in individual arousal leading to a smaller motivation to build up subglottic pressure (Stoeger et al. 2011).

To simulate size-dependent differences in call structure, the same sets have been created with different sizes of vocal tract and vocal fold characteristics. We aimed to simulate a smaller individual as well as a larger individual. Settings for the different simulations are summarized in Table 4.2 and differences in vocal fold lengths are additionally visualized in Figure 4.3 B&C. An overview of all created datasets can be found in Table 4.3.

Table 4.2: Vocal fold (VF) and vocal tract (VT) characteristics in the model

Individual	VF length	VF thickness	VF depth	VT length
Small	7mm	3.5mm	3.15mm	15.89cm
Medium	10mm	5.0mm	4.50mm	19.87cm
Large	13mm	6.5mm	5.85mm	23.34cm

Table 4.3: Overview of created datasets

Individual	Variation	Call Types	Rep Size
Small	Arousal	Grunt	23
		Bark	23
		Scream	23
			69
Medium	Arousal	Grunt	23
		Bark	23
		Scream	23
			69
Large	Arousal	Grunt	23
		Bark	23
		Scream	23
			69
All	Arousal + Size	Grunt	69
		Bark	69
		Scream	69
			207

3.5 Sound Analysis

Call Parameterization

Every simulation created a wav file of the simulated call. This call was subsequently fast Fourier transformed (FFT) into its frequency-time domain with Avisoft (Avisoft SASLab Pro, version 5.2.05), using a FFT size of 1024 points, Hamming window and 96.87% overlap. We used a sampling frequency of 20 kHz, resulting in a frequency range of 10 kHz, a frequency resolution of 20 Hz and a time increment of 6.4ms. The resulting frequency-time spectra were analysed with the software LMA 2012 developed by Kurt Hammerschmidt. For all acoustic analysis we chose a set of 118 acoustic features that describe the temporal- and spectral characteristics of the vocalizations as well as the call tonality and the spectral modulation of the calls (see chapter 2 - Table A.1). The acoustic features have proven to sufficiently describe the call morphology of the different call types in chacma baboons (Wadewitz et al. 2015a).

Determining the Number of Call Types

Since we simulated three call types, we expected the three cluster solution to be superior over other possible solutions. To assess our expectations we used a fast validation method that is based on k-means clustering (MacQueen 1967) and the analysis of silhouette values (Rousseeuw 1987). For a description of the implementation of k-means clustering and silhouette validation see (Wadewitz et al. 2015a). With this method the clustering quality of different cluster solutions with $k = 2 - 10$ has been validated for every dataset.

Quantification of Call Structure Variation

To quantify the variation of call structure in the simulated datasets, we applied a method that is based on fuzzy c-means clustering and which has been successfully used for the description of call variation in the acoustic repertoires of chacma baboons and Barbary macaques. The detailed descriptions of our approach can be found in the Methods section of (Wadewitz et al. 2015b). In fuzzy c-means clustering, every call is associated an assigned membership value for each of the clusters, ranging from $m = 1$ (fully displays the properties of the cluster) and $m = 0$ (does not display any of the properties of the cluster). Intermediate membership values $0 < m_{ia} < 1$ mark calls that do not fully belong to one of the clusters, but can be classified as intermediates between different call types. Membership values are affected by a parameter (μ) of the algorithm that defines the fuzziness of the system. Since we knew the optimal cluster solution of our created datasets from prior cluster quality validation, we set μ to 2.0 for all datasets. This allowed us to standardize the analyses of the different datasets and make results best comparable.

For each call, typicality coefficients have been calculated by subtracting the second largest membership component from the first:

$$TC(i) = m_i^{(1st)} - m_i^{(2nd)} \tag{4.1}$$

The typicality coefficient is bounded in the interval $0 \leq TC \leq 1$. A typicality coefficient of 1 indicates that the call lies right in the center of a call type and therefore highly typical, a typicality coefficient of 0 indicates that the call lies exactly between the centers of two call types and has an intermediate morphological structure (i.e. is a highly atypical call). In this way, the typicality coefficient can be used to quantify how discrete or graded call types are within a repertoire.

4 Results

In the three moderate datasets, the three-cluster solutions show the highest validity and all simulated calls fall into their pre-assigned categories. Higher numbers of clusters do not lead to higher validity but on the contrary silhouette values drop significantly. In the fourth dataset that includes all calls from the other three datasets ('All'), no cluster solution is significantly more valid than the others. In contrary to the moderate datasets, in the extended datasets with a higher variation in subglottic pressure levels (Fig. 4.6 B) the three-cluster solution is not superior over the other solutions. In addition, the dataset including size variation ('All') has generally lower silhouette values, which means that the

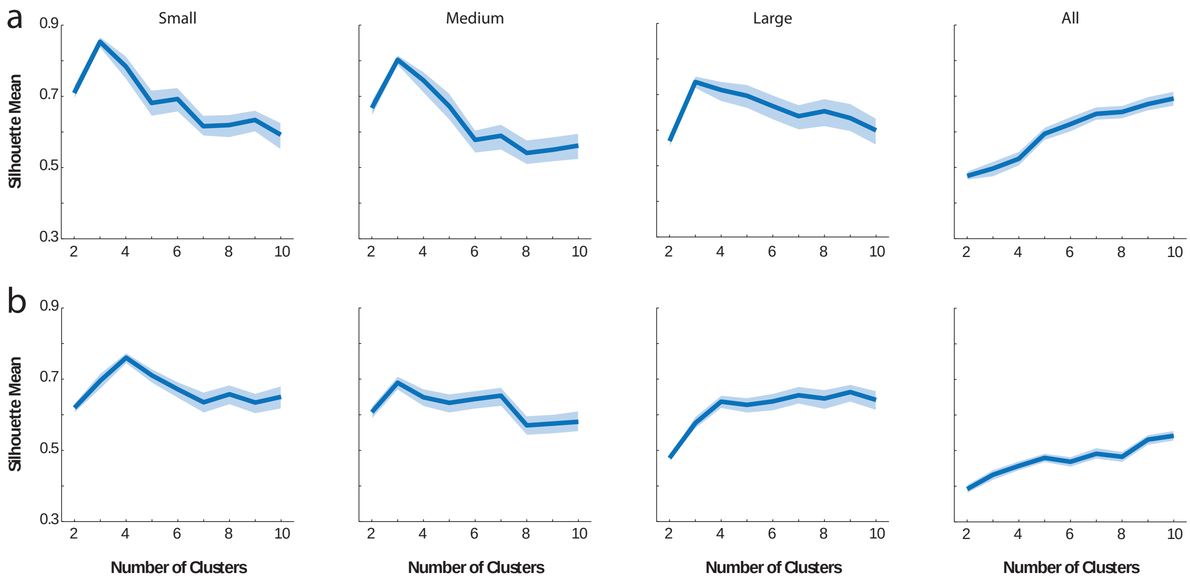

Figure 4.6: **Comparison between the average silhouette width for K-means clustering for k=2-10 cluster for all four datasets.** All moderate data sets without size-variation show a similar curve of average silhouette widths with a three-cluster solution that is superior over all other created solutions. For the extended data sets with higher variation in subglottic pressure levels the three-cluster solution is not superior over the others and an optimal solution is hard to determine. Data sets with variation in body size show generally lower silhouette values, i.e. the established clusters are less well separated. Shaded areas represent 95% confidence intervals.

clusters are less well separated than in the other three datasets. To be able to compare the quantitative variation in call structure between all datasets, subsequent analysis has been carried out for the three-cluster solution.

In Figure 4.7, two acoustic features, namely fundamental frequency (F0) and peak frequency (PF), have been chosen to visualize their distribution over the three call types when subglottic pressure is changed. The change in subglottic pressure does not affect F0 in a significant way and the three call types remain separated (Fig. 4.7 A-C). However, if all data sets are being combined, F0s of the small individual's barks are hardly discriminable from the F0s of large individual's screams (Fig. 4.7 D). Concerning the peak frequency, grunts and barks are not discriminable in any of the data sets, whereas the screams remain separated (Fig. 4.7 E-G). If all data sets are being combined, PFs of the large individual's screams are not discriminable from the PFs of smaller individual's grunts and barks (Fig. 4.7 H).

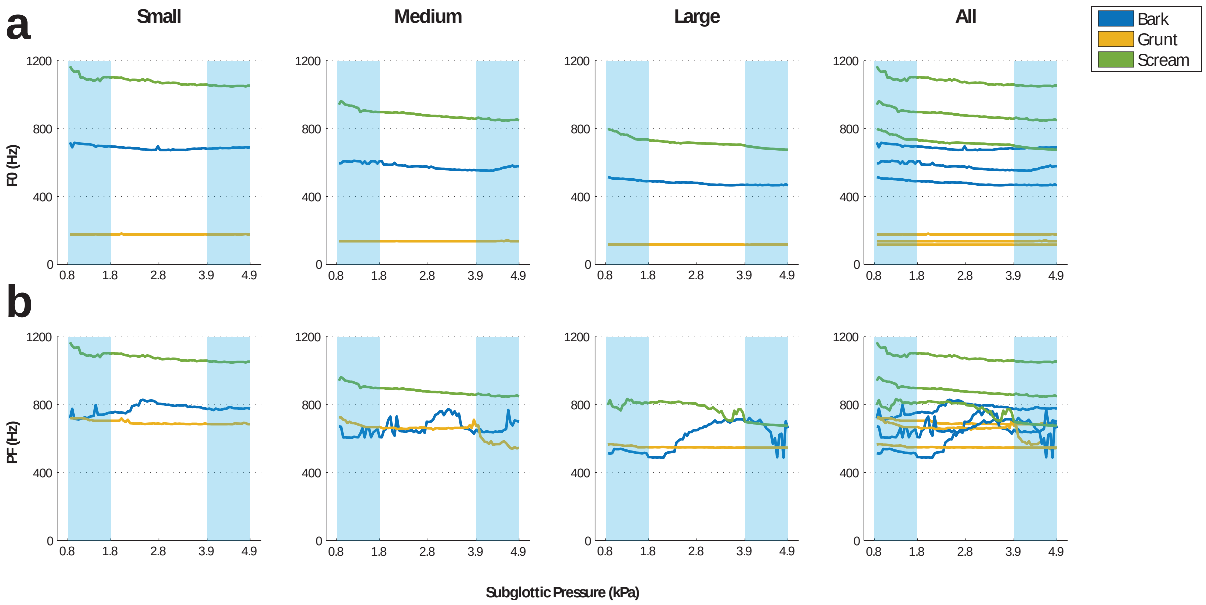

Figure 4.7: Distribution of fundamental frequency (F0) and peak frequency (PF) over changes in subglottic pressure levels and body size. Whereas changes in subglottic pressure level do not affect the discriminability of the three call type in the data sets without body size variation, with changing body size F0 auf small individual's barks and large individual's screams overlap (a). Considering PF, the grunt and bark cluster are not discriminable whereas the scream cluster remains separated for data sets without size variation (b). Taking body size into account, PFs of large individual's screams overlap with PFs of smaller individual's grunts and barks. Shaded areas represent the calls incorporated in the extended data sets.

Taking all acoustic features into account and calculating membership values for all calls, the three call types are segregated in the acoustic space in a rather discrete fashion if the data sets do not include individuals of different sizes (Fig. 4.8 A). A higher variation of subglottic pressure levels leads to higher acoustic variation, but does not change the general pattern of rather discrete call types (Fig. 4.8 B). In the fourth dataset that incorporates calls from different sized individuals, a larger amount of variation within and between call types can be found (Fig. 4.8 'All'). Especially between the bark and scream call types, there is a considerable amount of gradation and some calls that we modelled as "scream" types with the settings of a large individual were "misclassified" into the bark cluster.

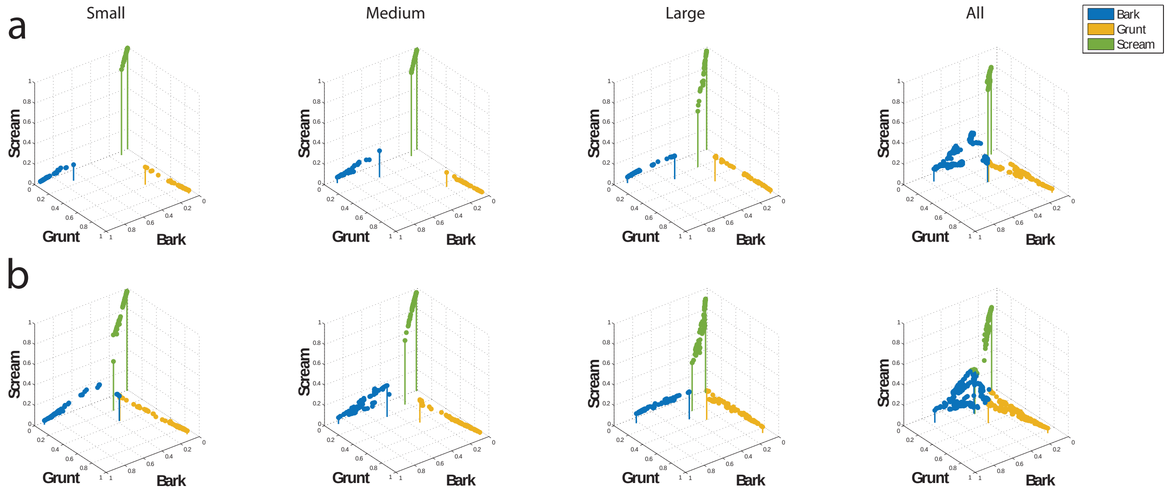

Figure 4.8: Tetrahedral comparison of cluster segregations. Three-dimensional representation linking all three simulated call types. Individual calls are depicted as dots in the 3D space of memberships. The X, Y, and Z axes correspond to membership of the bark, grunt, and scream clusters, respectively. Spectrograms represent transitions from most typical calls to most atypical calls of the three clusters. Sound examples can be found in the supporting information. Whereas in the data sets without size-variation call types remain relatively separated, data sets with size-variation show a considerable amount of gradation between the bark and scream call types. This pattern holds for the extended data sets which show a generally higher variation in acoustic structure.

The level of gradation that is visualized in Fig. 4.8 has been quantified by the calculation of typicality coefficients and the distribution of typicality coefficients (DTC) is shown in Fig. 4.9. In all data sets, screams show higher typicality coefficients than grunts and barks, indicating a better separation of this call type. Although this pattern remains throughout all of the created datasets, typicality coefficients are generally lower in the dataset with size variation (Fig. 4.9 'All'). This reduction in typicality coefficients confirms the visual impression from the three dimensional representation in Fig. 4.8, where the bark cluster shows much more variation in call structure.

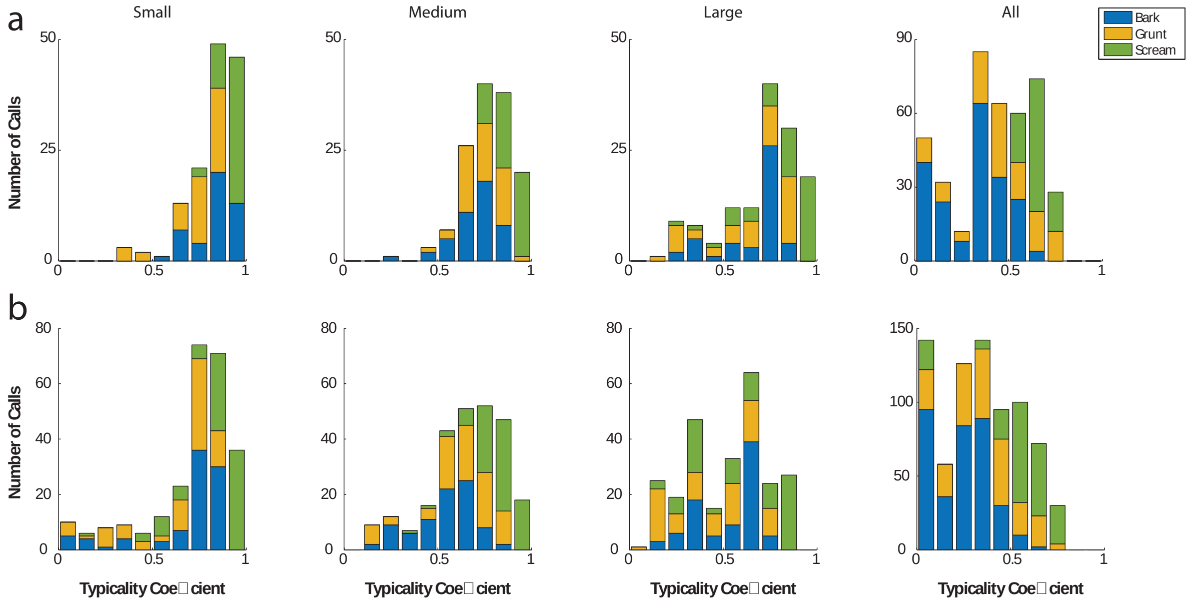

Figure 4.9: Distribution of typicality coefficients. Quantification of the overall typicality of calls in all four datasets by typicality coefficients. Typicality coefficients are calculated by subtracting the membership value of the second closest from the membership value of the closest call type. Sections with different colors indicate calls with different main type. In all data sets, screams show higher typicality coefficients than grunts and barks, indicating a better separation of this call type. In the data sets with size-variation typicality coefficients are generally lower.

5 Discussion

In this study we showed that the composition of a species' vocal repertoire can have a profound influence on the analysis of its acoustic structure. Whereas moderate variation in subglottic pressure has only a minor impact on the ability to retrieve the given vocal types, differences in body size made the recognition of the existing three vocal types impossible. Higher variation of arousal strengthens the negative effect of size.

The application of the FE model to simulate calls of a nonhuman primate allows us to imitate physical and physiological mechanics of sound production. Despite the general accuracy of the model, some limitations have to be kept in mind. First, the muscle parameters of the three simulated call types are not based on measurements of a vocalizing animal, but are chosen to simulate our call templates based on the inspection of spectrograms, measurements of fundamental frequencies and muscle activation plots including iso-fundamental frequency contours. Generally, other muscle activation patterns are possible that could lead to a similar acoustic output. Another limiting factor concerns the laryngeal anatomy as well as precise measurements of the vocal tract of baboons. These characteristics have been estimated by existing literature on Hamadryas baboons (Pfefferle and Fischer 2006), Diana monkeys (Riede et al. 2005) and humans (Titze 1994). Because the flexibility of vocal tract geometry of nonhuman primates is highly debated amongst bio-acousticians and linguists (Lieberman et al. 1969), we kept the shape of the vocal tract simple and constant in all simulations. A third limitation is that subglottic pressure range is estimated based on available data of other mammals. To ensure that our pressure levels could actually be applied by an animal with the size of our simulated baboons, we took a rather conservative approach with moderate lower and upper pressure level boundaries (1.8kPa - 3.9kPa).

Keeping these limitations in mind, we were able to simulate all three call templates and to analyze the differences in call structure that were induced by differences in body size and subglottic pressure level. Our analysis shows that the applied clustering algorithm reveals a three-cluster solution as the optimal solution to describe the structure of the data sets with moderate subglottic pressure variation and no body size variation. The determination of the optimal cluster solution becomes however problematic if variation in subglottic pressure levels is extended towards the boundaries of physical call production. Differences in subglottic pressure levels do therefore influence the general determination of the number of call types in our datasets if taken to extreme values, whereas data sets with moderate pressure variation are not affected. We assume that the difficulty to separate call types with an extreme variation in subglottic pressure levels is due to chaotic oscillation patterns of the vocal folds if the subglottal airstream exceeds a certain threshold. If a data set contains calls that do not only differ in the applied subglottic pressure (i.e. state of arousal of the signaling animal), but also contains calls of animals with strongly varying body size, the clustering quality drops significantly and an optimal solution regarding the number of call types in the dataset becomes harder to determine (Fig. 4.6 'All').

Interestingly, the constructed data sets without body size differences already show some degree of gradation within and between their call types. Although most calls show a highly typical structure and only a minority of calls shows lower typicality coefficients, the differences between typical and atypical calls is not strikingly obvious when inspecting their spectrograms. A possible explanation could be that these calls differ in some hidden acoustic features that cannot easily been assessed by eye in the spectrograms. Despite these atypical calls, all three call types remain separated in all of the data sets. In the datasets with size variation, the differences in body size of the simulated calls leads to a much higher variation in call structure which results into lower typicality coefficients in all three call types. A striking result is that especially the calls in the bark cluster show considerably lower typicality. A reason for this effect might lie in the acoustic structure of the bark type. Both, fundamental frequency as well as frequency modulation show medium characteristics that are located between the non-modulated and lower-pitched grunts and the stronger-modulated higher-pitched screams (see also Fig. 4-7). Through the variation in body size, grunts of smaller individuals might approach call characteristics of barks given by large individuals and similarly screams of larger individuals might approach call characteristics of barks given by small individuals. The intermediate character of the bark call type might therefore lead to reduced typicality in their calls, whereas in the scream and grunt cluster, gradation towards a different call type can only occur in one direction (towards the bark type). This effect is demonstrated and even increased by the misclassification of scream types (given by a large individual) into the bark cluster.

In summary, our approach shows that whereas moderate differences in the state of arousal have a minor impact on the characteristics of vocal repertoires, differences in body size can hamper classification and characterization of call types. Although it would be desirable to have precise measurements of the modelled species' anatomy and especially its muscle activation patterns while vocalizing, we assume that the shown effects are widely applicable. The degree of these effects is, as discussed in this manuscript, dependent on several factors and not least at the general level of gradation within a species' repertoire. Researchers should be aware of these effects and should construct their data sets based on the underlying research question. To determine the number of call types and investigate their acoustic structure, our results show that it is beneficial to only incorporate recordings from animals of one age class (i.e. minimizing the variation in body size). To investigate a communication system from the perspective of signal receivers, all age classes should be taken into account.

5 General Discussion

The quantitative characterization of a species' vocal repertoire is not only necessary to investigate potential driving forces in signal evolution but is also important to understand consequences for signal processing by signal receivers. For these reasons, the analysis of vocal repertoires has a broad significance in animal communication studies. Analytical tools that are currently used to characterize vocal repertoires, however, have some shortcomings and often lack objectivity.

In my dissertation I evaluated some of the common methods used for the analysis of vocal repertoires, investigated the major factors that hinder objectivity and developed a method that allows a quantitative assessment of call structure variation within and between vocal repertoires. In the following chapter, I will first summarize the most important findings of the different studies imbedded in this thesis, highlight their contribution to the current way of how vocal repertoires are analyzed and discuss their implications for bioacoustics research in general. Finally, I will give an outlook and make suggestions for future research.

1 Common Ways to Analyze Vocal Repertoires

As I have introduced in Chapter 1.4.2, several methods to cluster acoustic data exist. In Chapter 2, two of the most common clustering algorithms have been used to analyze the vocal repertoire of chacma baboons, namely k-means (MacQueen 1967) and hierarchical Ward's clustering (Ward 1963). Although based on different metrics (in k-means random cluster centers are improved to minimize within-cluster distance whereas in Ward's method nearby clusters are linked in a bottom-up fashion), both procedures resulted in very similar categorization of calls which matched our reference classification (gained by visual inspection of spectrograms) to a high degree. Therefore, our results confirm the findings of several studies that these algorithms are useful to cluster calls in a given repertoire into call types that consist of similarly structured calls (e.g. Ward's: Laiolo et al. 2000; Shulezhko and Burkanov 2008; Kershenbaum et al. 2013; Fuller 2014; K-means: Hammerschmidt and Todt 1995; Hammerschmidt and Fischer 1998; Maciej et al. 2013).

We also investigated clustering approaches that have been developed to overcome some shortcomings of the more traditional clustering algorithms like k-means. Affinity propagation (AP) circumvents the problem of running into a local maximum that is globally not the best cluster solutions (Frey and Dueck 2007) and super-paramagnetic clustering (SPC) is based on the physical properties of a ferromagnetic model that outperforms other approaches in sophisticated data structures (Blatt et al. 1996). Both algorithms have been successfully applied to cluster large amounts of data (for instance on genetic data: Getz et al. 2000 (AP); Leone et al. 2007 (SPC)) and recently Gamba and colleagues used AP to analyze and compare vocal repertoires of *Eulemur* (Gamba et al. 2015). To test the strength of both approaches for vocal repertoire analysis we applied both algorithms on our chacma baboon data set (data not shown). Both approaches resulted in similar results than k-means and Ward's clustering. One of the reasons why we already gained high reference matching with k-means is due to the high number of repetitions of the iterative optimization procedure of the algorithm. The basic idea of this procedure is to find some reasonable initial partition of the data and to then move the samples from one cluster to another if such a move improves the clustering result. In this way, the algorithm reveals the locally optimal cluster solution. If the procedure is only run once, the argument of Frey & Dueck that clustering results with k-means can be inferior by finding the local but not global optimization is valid, since the starting points of the procedure might have been badly chosen. If, however, the optimization procedure is repeated often enough, using newly assigned initial cluster centers for every repetition, the globally optimal solution can be found (Duda et al. 2012). Although from a computational perspective this procedure might be less efficient, the relatively small data sets that are used to characterize vocal repertoires allow the usage of this procedure. Ward's clustering on the other hand hierarchically groups nearby clusters in a bottom-up fashion starting with one sample in every cluster. In this way, the approach avoids the problem of initial cluster selection and can also reveal subclusters in the data (Duda et al. 2012).

Several factors in the usage of these and similar unsupervised clustering algorithms can have a profound influence on the outcome of repertoire analysis, such as feature selection and determination of the number of call types. In the following sections of this chapter, I will discuss our findings in regards to these factors which can hinder objectivity in the analysis of vocal repertoires and make comparative studies difficult.

2 Acoustic Feature Selection

The decision of the acoustic features that are used in the analysis of vocal repertoires is one of the most crucial steps in the analysis of vocal repertoires and is often a matter of opinion among researchers. Whereas some prefer to use a small set of selected features that are thought to describe the structure of the analyzed calls in the most important aspects (e.g. Bastos et al. 2015), others prefer to use a large set of features to take all structural dimensions into account (e.g. Hammerschmidt and Fischer 1998). Others again use a small amount of factors derived from feature reduction methods that explain a large amount of structural variation (e.g. Gros-Louis et al. 2008). To test the influence of feature choice in a systematic way, we created four data sets of the recorded chacma baboon calls that varied in the number of acoustic features that were taken into account and compared the clustering results with our human-expert reference classification. Interestingly, the usage of a very low number of features resulted in a rather poor matching success with the reference, whereas a high number of clusters resulted in a high matching success. Our results therefore suggest that sufficiently large sets of acoustic features should be incorporated to capture all acoustic dimensions of the calls. This result may be surprising since correlated features can cause problems in multivariate statistical hypothesis testing and are therefore generally avoided. However, these restrictions do not apply in clustering procedures and in fact correlating features can perform well in classifying call types when combined. This effect can be explained by the fact that every measured feature shows some amount of measurement noise, which can be cancelled out if features are taken into account that share correlating trends but differ in measurement noise (Guyon and Elisseeff 2003). The assumed problem that a high number of correlating features can make it difficult to find appropriate cluster centers could not be confirmed in our analysis.

Another reason why some researchers prefer the usage of a small set of acoustic features is due to the results of validation methods that are often used to access the quality of cluster solutions. One of these validation methods is to inspect silhouette values of cluster solutions which describe how tight data points lie within a cluster and how separated the different clusters in a given data set are (Rousseeuw 1987). Data sets with a lower number of acoustic features intrinsically have a lower statistical spread of values since every additional acoustic feature adds variation to the data. Silhouette values in sets with lower number of acoustic features will therefore be higher. If the quality of clustering is evaluated by silhouette values, this results in the misleading impression that a lower number of acoustic features is leading to a better separation of clusters. I have shown this effect by calculating silhouette values from k-means clustering results based on different combinations of 13 acoustic features and comparing them to the matching success with our reference classification (Fig. 5.1). As the Figure indicates, data sets with a lower number of acoustic features (light gray) show higher silhouette coefficients, but are less similar to the human-expert reference. With an increase in acoustic feature (dark grey), the similarity to the reference (i.e. the cluster quality) increases, whereas the silhouette coefficient decreases.

2.1 Factor Analysis

A common way to circumvent the assumed problem of correlation between acoustic features in the analysis of vocal repertoires is the usage of factor analysis (e.g. Fuller 2014; Gros-Louis 2006; Templeton et al. 2014). However, as our data set that was based on factors resulted in a poor resolution of emerging call types, we generally discourage the usage of factors in cluster analysis. The weak performance of the factor analysis can be explained by its linear nature, always being based on a matrix decomposition of the co-

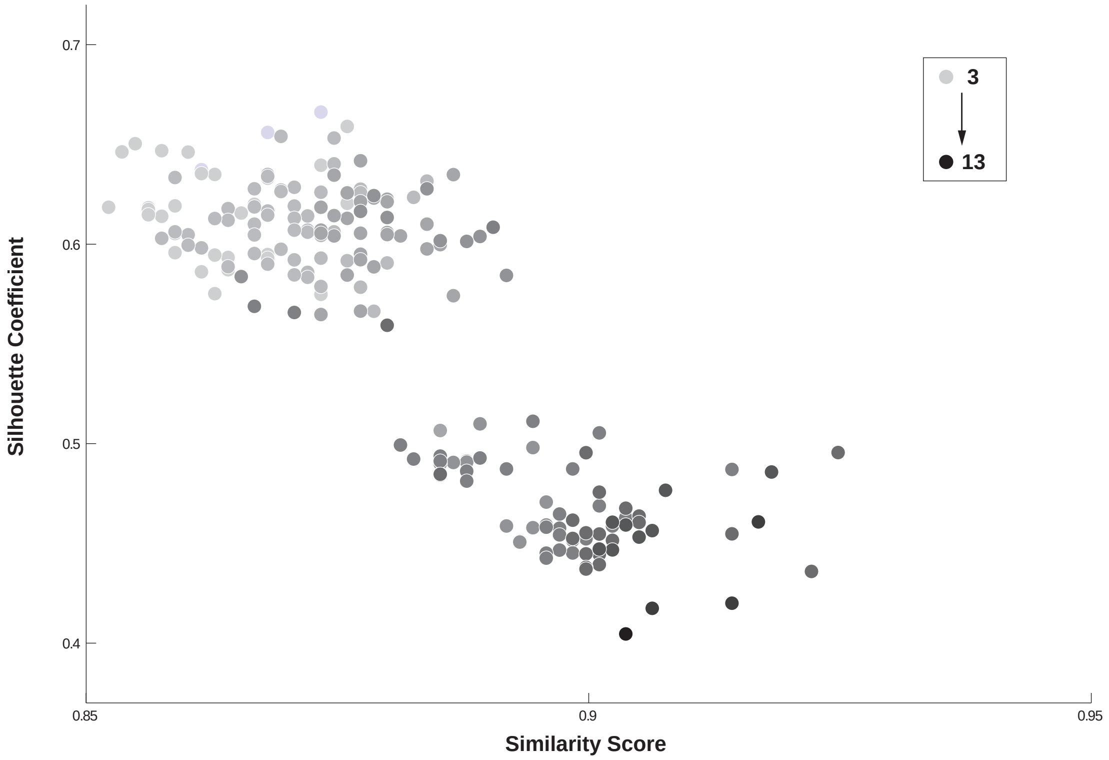

Figure 5.1: Top 200 feature combinations that achieved the highest score based on silhouette values and similarity to reference classification. Gray scale represents the number of features used in the k-means clustering. Data sets with more features gained generally lower silhouette values, but were more similar to reference classification than data sets with fewer features.

variance matrix. If the established clusters have non-spherical shapes in high dimensional feature space it might not be possible to properly separate them by hyperplanes orthogonal to the factors. Thus reducing the dimensionality of the data by projecting them to the linear space spanned by only a few factors may conceal non-linear correlations in the data set.

Although this limitation does not necessarily have to be the case in every kind of data set, there are additional reasons why factors have to be used with caution. Factors can be difficult to interpret, especially when highly divergent features load onto the same factors. If the acoustic features load in an interpretable way onto a few factors working with factors may be feasible. However, the construction of apparently meaningful factors may also result in the loss of crucial variation that would be helpful to distinguish between calls or call types.

2.2 Fast Fourier Transfrom vs. Wavelet Transform

To extract acoustic features from sound recordings, signals are usually transformed into their time-frequency spectrum by using Fast Fourier transformation (FFT). In this way, the signal is represented as the sum of a series of sines and cosines. Since the original representation of the FFT has only frequency resolution but no time resolution (i.e. all frequency components of a signal can be determined, but there is no temporal information of when the frequencies occur), the signal is cut into small sections (windows) which are analyzed separately and from which frequency and temporal features of the signal can be extracted. Depending on the size of the windows that are analyzed, either frequency resolution or time resolution is increased, while the other resolution is decreased (Heisenberg's uncertainty principle). However, by overlapping the windows, this limitation can be significantly decreased.

The wavelet transform was developed to overcome the FFT's initial shortcomings concerning the frequency- and time resolution by using a scalable modulated window that is shifted along the signal. For every position, the spectrum is calculated. By stretching or compressing the window, a collection of time-frequency representations is gained which all differ in resolutions. Since this stretch/compression (scaling) of the wavelet is directly correlated with its frequency, the scaling coefficients can be seen as the frequency components of the signal (for comprehensible summaries of wavelet transforms see Valens 1999 and Torrence and Compo 1998).

To test whether the analysis of vocal repertoires could gain from wavelet analysis, we extracted scaling features from the chacma baboon recordings and used them in cluster analysis. In Figure 5.2 representations of the same call in spectrogram and wavelet representation are shown.

As Figure 5.2 shows, the main difference between the FFT and the wavelet representation of the signal is that whereas in the FFT representation the frequency and time resolution are constant throughout the signal, these resolutions change in the wavelet transform. Here, low frequencies (high scale) are characterized by a good frequency but poor temporal resolution whereas high frequencies (low scale) are characterized by good

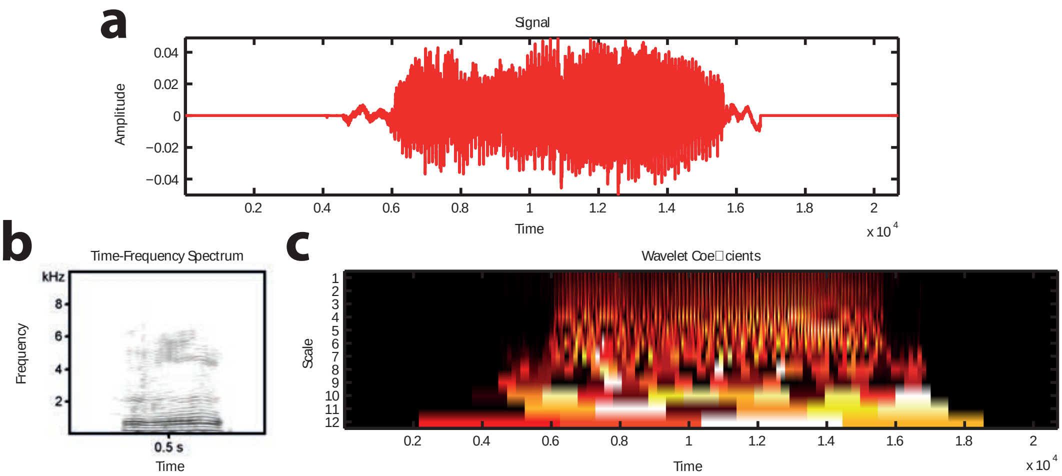

Figure 5.2: Different representations of a chacma baboon weaning call. (A) Time signal; (B) Spectrogram after Fast Fourier transformation with a window overlap of 97%; (C) Wavelet power spectrum after Discrete Wavelet transformation.

temporal but poor frequency resolution (note that the y-axis is logarithmic).

Using the wavelet coefficients (instead of extracted acoustic features of the FFT) to cluster the calls, the matching success with our reference classification is extremely poor (Tables 5.1 & 5.2). A possible reason for this poor matching success might be that although the high frequency resolution in the low frequency areas allows a precise determination of some crucial features such as fundamental or dominant frequency, the low temporal resolution in the low frequencies of the calls does not cover the frequency modulations that are potentially important to discriminate between call variants appropriately. The results of our analyses confirm the general view that although wavelet analysis might be helpful in specific bioacoustics applications such as the recognition of discontinuities and sharp spikes in bird sounds (Selin et al. 2007) or the detection of sperm whale clicks (Lopatka et al. 2005), it is rather unsuited to analyze vocalizations of nonhuman primates.

Table 5.1: Cross-tabulation of reference classification and k-means cluster solution for k=4. The cluster solution is based on 123 acoustic features that have been extracted from *Fast Fourier transforms* of the recorded calls. Similarity score: 0.95.

label	bark	grunt	scream	wean.
bark	**334**	0	24	2
grunt	0	**339**	0	0
scream	0	0	**106**	0
wean.	6	11	4	**86**

Table 5.2: Cross-tabulation of reference classification and k-means cluster solution for k=4. The cluster solution is based on 12 wavelet coefficients that have been extracted from *Discrete Wavelet transforms* of the recorded calls. Similarity score: 0.73.

label	bark	grunt	scream	wean.
bark	**286**	67	51	13
grunt	1	**239**	0	19
scream	1	0	**83**	0
wean.	52	44	0	**56**

3 Determination of the Number of Clusters

Determining the optimal number of call types in a given repertoire can be challenging or even impossible, especially if there is a high level of variation in the acoustic structure within and between different call types. Several cluster validation methods are available to determine the optimal number of call types. By calculating the reduction in variance of calls within different call types, cluster solutions that partition a data set best can be revealed (Hammerschmidt and Fischer 1998). Silhouette values, which are based on similar calculations, can in principle be used as a reliable indicator of the quality of a specific cluster solution. However, results are often less obvious than assumed (e.g. Maciej et al. 2013). In the analysis of the chacma baboon repertoire, different cluster solutions appeared to be appropriate to partition the data set. As discussed in Chapter 5.2, silhouette values are affected by the number of acoustic features that are taken into account and therefore the comparison of cluster quality should not be compared between data sets if the number of acoustic features differs in a significant way. Another approach that can be used to determine the number of call types is based on neutral mutual information

(NMI) between different clustering methods (e.g. Kershenbaum et al. 2013). We successfully used NMI between the classification results of k-means and Ward's clustering which resulted in similar cluster stabilities that we gained from silhouette analysis. However, the same limitations that I described for the silhouette analysis applies for NMI and researchers should be aware of them when characterizing vocal repertoires.

A critical aspect that concerns all validation methods is that often a low number of clusters can lead to a high cluster quality. It is important to mention that although these solutions can be mathematically superior, they might not provide sufficient detail to describe a species' vocal repertoire. An additional problem which hinders objectivity in the determination of the number of call types irrespective of the used validation method is the researcher's preference to either split or lump data (McKusick 1969).

In the following section I will discuss the results of our developed approach to describe vocal repertoires in detail. An integral part of this approach is the determination of the optimal number of call types, which discussion I will therefore shift into the next section.

4 Fuzzy Clustering to Describe Vocal Repertoires

One of the main deficits that the hard clustering algorithms I discussed in the previous sections possess is the initial assumption that computed call types reflect discrete categories. Since vocal repertoires of most terrestrial mammals exhibit a substantial level of gradation within and between call types (e.g. pigs: Tallet et al. 2013; giant otters: Leuchtenberger et al. 2015; mice: Scattoni et al. 2008; nonhuman primates: Rowell and Hinde 1962; Green 1975), we developed a method that is based on fuzzy logic and that allows us to capture details of the graded structure of vocal repertoires.

To examine the utility of our approach, we quantified the graded structure of the chacma baboon vocal repertoires in Chapter 2 and systematically compared it to the vocal repertoire of the Barbary macaque in Chapter 3. We chose these two species because the vocal behavior of both species has been intensely studied and therefore they served as good models to evaluate the accuracy of our approach (Hammerschmidt and Fischer 1998; Fischer et al. 2001). Our results confirmed the findings of previous studies that Barbary macaques show a higher level of gradation within and between different call types.

The first step in our approach is the determination of the number of clusters, which often proves to be difficult as I discussed before. To tackle this problem we tested the stability of different cluster solutions over a parameter that determines the fuzziness of the algorithm. Surprisingly, except from solutions with very low number of clusters, in neither of the two vocal repertoires one superior cluster solution could be identified. This result leads to the question of whether the methodology is not suited to find the "true" best cluster solution, or whether several cluster solutions are simply largely equivalent. Since this technique has been applied to identify the best cluster solution among morphological variation in interneurons (Battaglia et al. 2013), I assume the latter is the case. Support for this assumption comes from our results in Chapter 2, were we showed that other cluster validation methods like NMI or silhouette values could not identify one superior solution in the chacma baboon data set either. Although our approach did not find one superior cluster solution, the direct comparison of cluster stabilities of the chacma baboon and the Barbary macaque repertoire still revealed interesting insights. Since in the Barbary macaque data set, clusters split systematically at lower fuzziness values, the established Barbary macaque call types exhibit a higher level of gradation for any given resolution. This is already a first indication that the call types in the Barbary macaque repertoire are less well separated than in the chacma baboon repertoire without the detailed analysis that follows after this first step.

Although the decision on how many call types can be found in a given repertoire can be difficult, in Chapter 2 we focused on one specific cluster solution and described how we can quantify each call's typicality in the vocal repertoire of chacma baboons. With this first study we could already measure the variation in call structure between and within call types and reveal details of the level of gradation that cannot be captured with other clustering approaches. One aspect of this detailed analysis is that call types differ in their level of gradation. Whereas for instance barks and weaning calls have shared boundaries showing intermediate call structures, variation in grunt structure is less pronounced, resulting in a stronger separation of this call type with overall higher typicality coefficients.

In Chapter 3 we extended our approach to circumvent the problem of settling on one specific cluster solution when multiple solutions are largely equivalent. In order to do this, we assessed the distribution of typicality coefficients over a range of possible cluster solutions. Our results showed that the gradation of the two vocal repertoires does not differ significantly for solutions with a low number of clusters. When increasing the number of clusters however, the differences between the two repertoires become apparent and remain stable above a certain number of clusters. We showed that the majority of calls in the chacma baboon repertoire shared their membership value between two call types and all solutions showed a left-skewed distribution (i.e. the repertoires had a rather discrete structure). In contrast, calls in the Barbary macaque repertoire had common borders with several call types and typicality coefficients dropped significantly with an increase in the number of clusters, resulting in a highly right-skewed distribution. With our approach we were able to show that the Barbary macaque repertoire exhibits significantly more variation within and between different call types in comparison to the chacma baboon repertoire which confirms previous studies on the vocal behavior of chacma baboons (Owren et al. 1997; Fischer et al. 2001a) and Barbary macaques (Hammerschmidt and Fischer 1998).

5 Repertoire Composition

In addition to the analytical factors such as feature selection and determination of the number of call types that I discussed in the previous sections, the composition of a species' vocal repertoire is another factor that can have a profound influence on repertoire structure. A change in the level of acoustic variation within a given repertoire can be caused by several factors such as variations in sex and age of the recorded animals (e.g. Ey et al. 2007) which seem to correspond to variations in body size, according to the mechanisms of sound production (Fitch and Hauser 1995). Other potential factors are the animal's arousal (e.g. Stoeger et al. 2011; Townsend and Manser 2011) or different recording conditions leading to signal fragmentation (Maciej et al. 2011). Whereas the latter can be avoided or at least diminished by trying to keep recording conditions constant and dismiss bad quality recordings, differences in signalers body size or state of arousal might be harder to access. In Chapter 4 of this thesis, I simulated differences in body size and arousal and evaluated the influence of these factors on the constructed vocal repertoire. Simulation of arousal has been done by changing subglottic pressure level, which has been shown to be affected by arousal (Stoeger et al. 2011). However, it is important to mention that this is only one (even though crucial) factor that can be affected by arousal and influence signal structure. The results of our analysis show that by changing the subglottic pressure level only, the three call types that have been simulated remain separated and the three cluster solution is superior over all other solutions. One limitation of the study in this respect is that subglottic pressure range was estimated based on available data of other mammals. To ensure that our pressure levels could actually be applied by an animal with the size of our simulated baboons, we took a rather conservative approach with a moderate upper pressure level boundary.

Interestingly, when a data set contains calls that do not only differ in the applied subglottic pressure level, but also contains calls of animals with strongly varying body size, the clustering quality drops significantly and an optimal solution regarding the number of call types in the data set becomes harder to determine. In this constructed repertoire, especially the calls in the bark cluster show considerably lower typicality. I assume that this effect can be explained by the intermediate character of the bark cluster considering frequency-related acoustic features. In this regard, taking calls of animals into account that vary in body size, screams (characterized by higher-frequency components) given by larger individuals can converge bark characteristics given by smaller individuals and grunts (characterized by lower-frequency components) given by smaller individuals can converge bark characteristics given by larger individuals. Despite the general accuracy of the used model, some limitations have to be kept in mind when interpreting the results. A limitation concerning the simulated body size variation is that laryngeal anatomy as well as vocal tract size has been estimated by existing literature on Diana monkeys (Riede et al. 2005) and humans (Titze 1994) and does not represent true characteristics of the baboon's morphology.

Although the study is still in preparation for submission and it is planned to produce larger data sets and do additional analyses, the results indicate the strong influence that the data set composition can have on the resulting repertoire. Whereas differences in epiglottal pressure seem to have a minor impact on the characteristics of vocal repertoires, differences in body size can hamper classification and characterization of call types and researchers should be aware of these effects when preparing data sets to analyze a species' vocal repertoire.

6 General Implications

In addition to the detailed descriptions of the chacma baboon and Barbary macaque vocal repertoires, the results of my dissertation have some general implications on bioacoustics research and potentially on other scientific domains. The importance of feature selection, identification of the number of clusters and composition of data sets which I discussed in detail in the previous sections should be taken into account in future repertoire analysis and interpretation. Our approach based on fuzzy clustering (DTC) has the potential to evaluate possible selective factors that drive the structure of signals and entire vocal repertoires.

Within a repertoire, several studies have shown that structural variation in calls can differ between different call types. In Campbell's monkeys, for instance, contact calls are rather graded, whereas alarm calls are more discrete (Outarra et al. 2009; Lemasson and Hausberger 2011). In chimpanzees, screams show a high level of variability, whereas copulation calls are much more discrete (Slocombe et al. 2009; Townsend and Zuberbuhler 2009). Other examples include gray mouse lemurs (Leliveld et al. 2011) and baboons (Fischer et al. 2001; Rendall et al. 2009). The differences in the variability between different call types are assumed to be related to the call function (Bouchet et al. 2012). The results of our analyses confirm the findings that within a repertoire call types can differ in their variability. In the chacma baboon repertoire, for instance, grunts show a much more discrete structure than screams, which is preserved over all constructed cluster solutions. The level of detail that our method provides can be used to quantify these differences in variability between different call types that other methods cannot.

Between vocal repertoires of different species, the quantification of gradation can be used to re-examine some classic hypotheses, for instance, that graded repertoires should occur in species with higher visual access towards each other, since additional sensory modalities can be used by the signaler and receiver (Marler 1976, 1977). Following this hypothesis, Barbary macaques with a higher level of gradation within their repertoire should have better visual conditions than chacma baboons. The characteristics of the habitats of both species are strikingly different with Barbary macaques living more arboreal, inhabiting mainly cedar, fir, and oak forests in the Atlas Mountain range of Morroco and Algeria (Fa 1984) and chacma baboons inhabiting mainly desert, savannah grassland and woodland in southern Africa (Cowlishaw 2013). However, in both species individuals have generally good visual access towards each other. Therefore, our finding that the vocal repertoire of Barbary macaques exhibits a higher level of gradation does not support this hypothesis. One important aspect that has to be taken into account is the historic distribution of a species that might differ from the current distribution but might have shaped the communication system. However, since chacma baboons are endemic to southern Africa and never inhabitated closed habitats, and Barbary macaques are endemic to the extreme North Africa, this possibility can be ruled out.

Detailed descriptions of vocal repertoires are needed to determine a repertoire's complexity. To evaluate the theory that vocal complexity (Freeberg et al. 2012; Krams et al. 2012; Bouchet et al. 2013) is driven by social complexity (McComb and Semple 2005), an explicit definition of vocal complexity is necessary. As I have introduced in Chapter 1, two definitions of vocal complexity exist. The first and simpler one looks at the total number of call types within a species' repertoire (Freeberg et al. 2012), which can be complicated to determine if repertoires show a higher level of gradation. The second definition of vocal complexity is based on information theory. According to this concept, complex systems are neither completely discrete, nor completely graded, but rather lie in between these two extremes (Tononi et al. 1998; Crutchfield 2011). Whereas in discrete systems the information only lies in the call types itself, totally graded systems without structure would be useless for communication since potential information could not be distinguished from noise. Following this definition, the quantification of typicality within repertoires that our method provides could be a useful indicator for vocal complexity.

Beyond the application in bioacoustics research, our developed method can be potentially useful in other scientific domains where graded data sets need to be quantified, compared, and visualized. In neurobiology, Battaglia and colleagues successfully quantified the large diversity of morphological and molecular properties of inhibitory interneurons using typicality coefficients (Battaglia et al. 2013). In population genetics, edge and core populations show a different level of diversity in haplotypes (Eckert et al. 2008) and our method could be used to describe these differences in detail. In vegetation classification, different models exist that make assumptions of whether plant communities form discrete units or whether they are structured continuously (Collins et al. 1993) and typicality coefficients may be used to quantify the distribution of these vegetation patches.

7 Outlook

In my dissertation, I investigated different factors that can hinder objectivity in the analysis of vocal repertoires and provide a method that allows quantifying and comparing the level of gradation within species' repertoires. Although I already discussed several potential areas of application in the previous sections of this chapter, I would like to conclude with some thoughts on additional future research perspectives.

As I have discussed, feature selection is a crucial step in the analysis of vocal repertoires and I recommend taking a large number of features into account. If a smaller set of features is desired, one has to make sure that features that might be important for the discrimination of call variants are not ignored. Concerning the analytical process, a systematic study could be designed in a way that data sets with a large number of features serve as a reference classification. From these data sets features could be iteratively excluded measured by the smallest loss in reference matching until a threshold is reached that accounts for a good trade-off between number of features and loss in information. An important aspect that has to be kept in mind, however, is that the features that are important to discriminate call variants on the analytical level are not necessarily the features that receivers within a species' communication system use to discriminate between call variants.

To investigate this question and additionally validate the accuracy of our developed approach, playback experiments could be conducted using the "habituation-dishabituation" paradigm (Fischer 1998). Since several species show categorical perception (Fischer 2006), call variants that grade from one into another call type could be used to detect the change points in the acoustic structures that are perceived as different categories and lead to a change in response behavior. To identify the call variants that might be in the relevant acoustic area for categorical perception, membership values of constructed vocal repertoires could be used. If the acoustic features that have been chosen in the analysis accurately reflect the acoustic features that are used by receivers to discriminate between call variants, the relevant intermediate calls that lie between two call types could be detected at the category boundaries (Figures 2.6 & 3.3, dashed lines).

Although this might be an enormous endeavor, systematic changes of specific acoustic features of call variants could be modelled to measure differences in response behavior and evaluate the importance of these features for call discrimination. It is important to mention however that when using the "habituation-dishabituation" paradigm, one does not measure the acoustic difference that the receiver is able to distinguish, but the acoustic difference that is meaningful to the receiver (Nelson and Marler 1989).

Another interesting aspect that could be investigated using our approach is whether the level of gradation within a repertoire and specifically within different call types changes with ontogeny. Several studies on alarm call development (for a review see Hollén and Radford 2009), but also on other call types (e.g. Hauser 1989) have found changes in call structure with ontogeny that are beyond changes in body size. Quantifying these changes using typicality coefficients could allow re-examining existing hypotheses concerning the function of these changes.

6 References

Abbot P, Abe J, Alcock J (2011) Inclusive fitness theory and eusociality. *Nature* 471.

Ackermann H, Hage SR, Ziegler W (2014) Brain mechanisms of acoustic communication in humans and nonhuman primates : An evolutionary perspective. *Behavioral and Brain Sciences* 37:529–604.

Alipour F, Titze I (1999) Active and passive characteristics of the canine cricothyroid muscles. *Journal of Voice* 13:1–10.

Armañanzas R, Ascoli GA (2015) Towards the automatic classification of neurons. *Trends in Neurosciences* 38:307–318.

Arnold K, Zuberbühler K (2006) The alarm-calling system of adult male putty-nosed monkeys, *Cercopithecus nictitans martini*. *Animal Behaviour* 72:643–653.

Arriaga G, Jarvis ED (2013) Mouse vocal communication system: are ultrasounds learned or innate? *Brain and Language* 124:1–44.

Bailey K (1994) Numerical Taxonomy and Cluster Analysis In *Typologies and Taxonomies*, p. 34.

Bailey WJ, Bennet-Clark HC, Fletcher NH (2001) Acoustics of a small Australian burrowing cricket: the control of low-frequency pure-tone songs. *The Journal of Experimental Biology* 204:2827–2841.

Battaglia D, Karagiannis A, Gallopin T, Gutch HW, Cauli B (2013) Beyond the frontiers of neuronal types. *Frontiers in Neural Circuits* 7:13.

Ben-David S, v. Luxburg U, Pal D (2006) A Sober Look on Clustering Stability In *Learning Theory*, pp. 5–19. Springer Berlin Heidelberg.

Bezdek JC (1974) Numerical Taxonomy with Fuzzy Sets. *Journal of Mathematical Biology* 1:57–71.

Blatt M, Wiseman S, Domany E (1996) Superparamagnetic clustering of data. *Physical Review Letters* 76:3251–3254.

Boersma BP, Heuven VV (2001) Speak and unSpeak with PRAAT. *Glot International* 5:341–347.

Boogert NJ, Giraldeau LA, Lefebvre L (2008) Song complexity correlates with learning ability in zebra finch males. *Animal Behaviour* 76:1735–1741.

Bouchet H, Blois-heulin C, Lemasson A (2013) Social complexity parallels vocal complexity : a comparison of three non-human primate species. *Frontiers in Psychology* 4:1–15.

Bouchet H, Blois-Heulin C, Pellier AS, Zuberbühler K, Lemasson A (2012) Acoustic variability and individual distinctiveness in the vocal repertoire of red-capped mangabeys (*Cercocebus torquatus*). *Journal of Comparative Physiology* 126:45–56.

Bouchet H, Pellier AS, Blois-Heulin C, Lemasson A (2010) Sex differences in the vocal repertoire of adult red-capped mangabeys (*Cercocebus torquatus*): A multi-level acoustic analysis. *American Journal of Primatology* 72:360–75.

Bouhuys A, Mead J, Proctor D, Stevens K (1968) Pressure-Flow Events During Singing. *Annals of the New York Academy of Sciences* 155:165–176.

Bradbury JW, Vehrencamp SL (2011) *Principles of Animal Communication.* Sinauer Associations, Sunderland.

Brady CA (1981) The vocal repertoires of the bush dog (*Speothos venaticus*), crab-eating fox (*Cerdocyon thous*), and maned wolf (*Chrysocyon brachyurus*). *Animal Behaviour* 29:649–669.

Brown CH, Alipour F, Berry Da, Montequin D (2003) Laryngeal biomechanics and vocal communication in the squirrel monkey (*Saimiri boliviensis*). *The Journal of the Acoustical Society of America* 113:2114–2126.

Burghardt GM, Bartmess-Levasseur JN, Browning Sa, Morrison KE, Stec CL, Zachau CE, Freeberg TM (2012) Perspectives - Minimizing Observer Bias in Behavioral Studies: A Review and Recommendations. *Ethology* 118:511–517.

Catchpole CK, Slater PJB (2003) *Bird Song: Biological Themes and Variations.* Cambridge University Press, Cambridge.

Charif R, Ponirakis D, Krein T (2006) Raven Lite 1.0 User's Guide.

Chater N, Manning CD (2006) Probabilistic models of language processing and acquisition. *Trends in Cognitive Sciences* 10:335–344.

Chater N, Reali F, Christiansen MH (2009) Restrictions on biological adaptation in language evolution. *Proceedings of the National Academy of Sciences of the United States of America* 106:1015–20.

Cheney DL, Seyfarth RM (1990) *How Monkeys See the World: Inside the Mind of Another Species* University of Chicago Press, Chicago.

Clemins PJ, Johnson MT (2006) Generalized perceptual linear prediction features for animal vocalization analysis. *The Journal of the Acoustical Society of America* 120:527–534.

Clutton-Brock TH, Albon SD (1978) The roaring of red deer and the evolution of honest advertisement. *Behaviour* 3-4:145–170.

Collins S, Glenn SM, Roberts DW (1993) The hierarchical continuum concept. *Journal of Vegetation Science* 4:149–156.

Corben C (2002) Zero-Crossings Analysis for Bat Identification: An Overview Technical report, Austin.

Cowlishaw G (2013) *Mammals of Africa - Volume 2 - Primates* Bloomsbury Publishing, London, New Delhi, New York and Sydney.

Crutchfield JP (2011) Between order and chaos. *Nature Physics* 8:17–24.

Daniel J, Blumstein D (1998) A test of the acoustic adaptation hypothesis in four species of marmots. *Animal Behaviour* 56:1517–1528.

de Waal FBM (1988) The Communicative Repertoire of Captive Bonobos (*Pan Paniscus*), Compared to That of Chimpanzees. *Behaviour* 106:183–251.

Duda R, Hart P, Stork D (2012) *Pattern Classification* John Wiley & Sons, second edi edition.

Dunn JC (1973) A Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact Well-Separated Clusters. *Journal of Cybernetics* 3:32–57.

Eckert CG, Samis KE, Lougheed SC (2008) Genetic variation across species' geographical ranges: The central-marginal hypothesis and beyond. *Molecular Ecology* 17:1170–1188.

Ehret G (1987) Categorical Perception of Speech Signals Facts and Hypotheses from Animal Studies In Harnad S, editor, *Categorical perception: The groundwork of Cognition.* Cambridge University Press, New York, NY.

Endersby J (2009) Lumpers and Splitters: Darwin, Hooker, and the Search for Order. *Science (New York, N.Y.)* 326:1496–1499.

Equihua M (1990) Fuzzy Clustering of Ecological Data. *Journal of Ecology* 78:519–534.

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE : a simulation study. *Molecular Ecology* 14:2611–2620.

Ey E, Fischer J (2009) The Acoustic Adaptation Hypothesis - A Review of the Evidence From Birds, Anurans and Mammals. *Bioacoustics* 19:21–48.

Ey E, Hammerschmidt K, Seyfarth RM, Fischer J (2007) Age- and Sex-Related Variations in Clear Calls of *Papio ursinus*. *International Journal of Primatology* 28:947–960.

Fa JE (1984) Habitat distribution and habitat preference in Barbary macaques (*Macaca sylvanus*). *International Journal of Primatology* 5:273–286.

Fattu J, Suthers R (1981) Subglottic pressure and the control of phonation by the echolocating bat, *Eptesicus*. *Journal of Comparative Physiology* 143:465–475.

Fenton MB, Bouchard S, Vonhof MJ, Zigouris J (2001) Time-Expansion and Zero-Crossing Period Meter Systems Present Significantly Different Views of Echolocation Calls of Bats. *Journal of Mammalogy* 82:721.

Fichtel C, Hammerschmidt K, Jürgens U (2001) On the vocal expression of emotion. A multi-parametric analysis of different states of aversion in the squirrel monkey. *Behaviour* 138:97–116.

Fischer J (1998) Barbary macaques categorize shrill barks into two call types. *Animal Behaviour* 55:799–807.

Fischer J, Cheney DL, Seyfarth RM (2000) Development of infant baboons' responses to graded bark variants. *Proceedings of the Royal Society B: Biological Sciences* 267:2317–21.

Fischer J (2006) Categorical Perception in Animals In Brown K, editor, *Encyclopedia of Language & Linguistics*, number 1, pp. 248–251. Elsevier, Oxford.

Fischer J (2010) Nothing to talk about. On the linguistic abilities of nonhuman primates (and some other animal species) In Frey U, Störmer C, Willführ K, editors, *Homo Novus - A Human Without Illusions.*

Fischer J, Hammerschmidt K (2001) Functional referents and acoustic similarity revisited: the case of Barbary macaque alarm calls. *Animal Cognition* 4:29–35.

Fischer J, Hammerschmidt K (2002) An overview of the Barbary macaque, *Macaca sylvanus*, vocal repertoire. *Folia Primatologica* 73:32–45.

Fischer J, Hammerschmidt K, Cheney DL, Seyfarth RM (2001) Acoustic Features of Female Chacma Baboon Barks. *Ethology* 107:33–54.

Fischer J, Hammerschmidt K, Cheney DL, Seyfarth RM (2002) Acoustic features of male baboon loud calls: Influences of context, age, and individuality. *The Journal of the Acoustical Society of America* 111:1465–1474.

Fischer J, Hammerschmidt K, Todt D (1995) Factors Affecting Acoustic Variation in Barbary-macaque (Macaca sylvanus) Disturbance Calls. *Ethology* 101:51–66.

Fischer J, Metz M, Cheney DL, Seyfarth RM (2001) Baboon responses to graded bark variants. *Animal Behaviour* 61:925–931.

Fitch T (1997) Vocal tract length and formant frequency dispersion correlate with body size in rhesus macaques. *The Journal of the Acoustical Society of America* 102:1213–22.

Fitch T (2000a) The evolution of speech: A comparative review. *Trends in Cognitive Sciences* 4:258–267.

Fitch T (2000b) The phonetic potential of nonhuman vocal tracts: Comparative cineradiographic observations of vocalizing animals. *Phonetica* 57:205–218.

Fitch T, Reby D (2001) The descended larynx is not uniquely human. *Proceedings of the Royal Society B: Biological Sciences* 268:1669–1675.

Fitch WT (2000) Skull dimensions in relation to body size in nonhuman mammals The causal bases for acoustic allometry. *Zoology* 103:40–58.

Fitch WT, Hauser MD (1995) Vocal Production in Nonhuman Primates: Acoustics, Physiology and Functional Constraints on "Honest" Advertisment. *American Journal of Primatology* 37:191–219.

Fitch WT, Hauser MD (1998) Unpacking " Honesty ": Vertebrate Vocal Production and the Evolution of Acoustic Signals In Simmons A, Fay RR, Popper N, editors, *Acoustic Communication*, pp. 65–137. Springer New York.

Forrest TG (1994) From Sender to Receiver: Propagation and Environmental Effects on Acoustic Signals. *American Zoologist* 34:644–654.

Fred NL, Jain K (2005) Combining multiple clusterungs using evidence accumulation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 27:835–850.

Freeberg TM, Dunbar RIM, Ord TJ (2012) Social complexity as a proximate and ultimate factor in communicative complexity. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* 367:1785–801.

Frey BJ, Dueck D (2007) Clustering by passing messages between data points. *Science (New York, N.Y.)* 315:972–6.

Fuller JL (2014) The vocal repertoire of adult male blue monkeys (*Cercopithecus mitis stulmanni*): a quantitative analysis of acoustic structure. *American Journal of Primatology* 76:203–16.

Gamba M, Friard O, Riondato I, Righini R, Colombo C, Miaretsoa L, Torti V, Nadhurou B, Giacoma C (2015) Comparative Analysis of the Vocal Repertoire of Eulemur: A Dynamic Time Warping Approach. *International Journal of Primatology* 36:894–910.

Gauch HG, Whittaker RH (1972) Coenocline simulation. *Ecology* 53:446–451.

Geissmann T, Nijman V (2006) Calling in Wild Silvery Gibbons (Hylobates moloch) in Java (Indonesia): Behavior, Phylogeny, and Conservation. *American Journal of Primatology* 68:1–19.

Getz G, Levine E, Domany E (2000) Coupled two-way clustering analysis of gene microarray data. *Proceedings of the National Academy of Sciences of the United States of America* 97:12079–12084.

Gomez-Marin A, Paton JJ, Kampff AR, Costa RM, Mainen ZF (2014) Big behavioral data: psychology, ethology and the foundations of neuroscience. *Nature Neuroscience* 17:1455–1462.

Gonzales-Lima F (2010) Responses of limbic, midbrain and brainstem structures to electrically-induced vocalizations In Brudzynski, editor, *Handbook of Mammalian Vocalization - An Integrative Neuroscience Approach*, pp. 293–301. Elsevier.

Gouzoules H, Gouzoules S (2000) Agonistic screams differ among four species of macaques: the significance of motivation-structural rules. *Animal Behaviour* 59:501–512.

Green S (1975) Communication by a graded vocal system in Japanese monkeys In Rosenblum, editor, *Primate Behaviour*, pp. 1–102. New York: Academic Press.

Gros-Louis J (2006) Acoustic Analysis and Contextual Description of Food-Associated Calls in White-Faced Capuchin Monkeys (*Cebus capucinus*). *International Journal of Primatology* 27:273–294.

Gros-Louis JJ, Perry SE, Fichtel C, Wikberg E, Gilkenson H, Wofsy S, Fuentes A (2008) Vocal Repertoire of *Cebus capucinus*: Acoustic Structure, Context, and Usage. *International Journal of Primatology* 29:641–670.

Gustison ML, le Roux A, Bergman TJ (2012) Derived vocalizations of geladas (*Theropithecus gelada*) and the evolution of vocal complexity in primates. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* 367:1847–59.

Guyon I, Elisseeff A (2003) An Introduction to Variable and Feature Selection. *Journal of Machine Learning Research* 3:1157–1182.

Hage SR, Jürgens U (2006) On the role of the pontine brainstem in vocal pattern generation: a telemetric single-unit recording study in the squirrel monkey. *The Journal of Neuroscience* 26:7105–7115.

Hammerschmidt K, Fischer J (1998) The Vocal Repertoire of Barbary Macaques: A Quantitative Analysis of a Graded Signal System. *Ethology* 104:203–216.

Hammerschmidt K, Fischer J (2008) Constraints in Primate Vocal Production In Oller DK, Griebel U, editors, *Evolution of communicative flexibility: complexity, creativity, and adaptability in human and animal communication*, pp. 93–119. MIT Press, Cambridge.

Hammerschmidt K, Todt D (1995) Individual Differences in Cocalisations of Young Barbary Macaques: A Multi-Parametric Analysis to Identify Critical Cues in Acoustic Signalling. *Behaviour* 132:381–399.

Hammerschmidt K, Whelan G, Eichele G, Fischer J (2015) Mice lacking the cerebral cortex develop normal song: Insights into the foundations of vocal learning. *Scientific Reports* 5:8808.

Handley LJL, Manica A, Goudet J, Balloux F (2007) Going the distance : human population genetics in a clinal world. *Trends in Genetics* 23:432–439.

Hauser MD (1989) Ontogenetic changes in the comprehension and production of vervet monkey (*Cercopithecus aethiops*) vocalizations. *Journal of Comparative Psychology* 103:149–158.

Hauser MD (1993) The evolution of nonhuman primate vocalizations: effects of phylogeny, body weight, and social context. *The American Naturalist* 142:528–42.

Hauser MD (1996) *The Evolution of Communication* MIT Press.

Hauser MD, Chomsky N, Fitch WT (2002) The faculty of language: what is it, who has it, and how did it evolve? *Science (New York, N.Y.)* 298:1569–79.

Hinton, Salakhutdinov (2006) Reducing the Dimensionality of Data with Neural Networks. *Science* 313:504–507.

Hollén LI, Radford AN (2009) The development of alarm call behaviour in mammals and birds. *Animal Behaviour* 78:791–800.

Hopp L, Owren MJ, Evans C (1998) *Animal Acoustic Communication* Springer Verlag.

Huang CJ, Yang YJ, Yang DX, Chen YJ (2009) Frog classification using machine learning techniques. *Expert Systems with Applications* 36:3737–3743.

Isaac NJB, Mallet J, Mace GM (2004) Taxonomic inflation: its influence on macroecology and conservation. *Trends in Ecology & Evolution* 19:464–469.

Jain AK (2010) Data clustering : 50 years beyond K-means. *Pattern Recognition Letters* 31:651–666.

Jang JSR, Sun CT (1997) *Neuro-Fuzzy and Soft Computing - A Computational Approach to Learning and Machine Intelligence*, Vol. 42 Upper Saddle River, NJ: Prentice Hall.

Janik VM, Slater PJB (1997) Vocal learning in mammals. *Advances in the Study of Behavior* 26:59–99.

Jarvis ED (2007) Neural systems for vocal learning in birds and humans: A synopsis. *Journal of Ornithology* 148:S35–S44.

Jürgens U, Ploog D (1970) Cerebral representations of vocalizations in the squirrel monkey. *Experimental Brain Research* 10:532–554.

Jürgens U, Pratt R (1979) Role of the periaqueductal grey in vocal expression of emotion. *Brain Research* 167:367–378.

Jürgens U (2009) The neural control of vocalization in mammals: A review. *Journal of Voice* 23:1–10.

Keenan S, Lemasson A, Zuberbühler K (2013) Graded or discrete? A quantitative analysis of Campbell's monkey alarm calls. *Animal Behaviour* 85:109–118.

Kershenbaum A, Blumstein DT, Roch MA, Backus G, Bee MA, Bohn K, Cao Y, Carter G, Cäsar C, Coen M, Deruiter SL, Doyle L, Edelman S, Ferrer-i cancho R, Freeberg TM, Garland EC, Gustison M, Harley HE, Huetz C, Hughes M, Bruno JH, Ilany A, Jin DZ, Manser MB, Mccowan B, Iii EM, Narins PM, Piel A, Rice M, Salmi R, Sasahara K, Sayigh L, Shiu Y, Taylor C, Vallejo EE (2014) Acoustic sequences in non-human animals : a tutorial review and prospectus. *Biological Reviews* pp. 1–42.

Kershenbaum A, Sayigh LS, Janik VM (2013) The encoding of individual identity in dolphin signature whistles: how much information is needed? *PLoS ONE* 8:e77671.

Kitchen DM, Cheney DL, Seyfarth RM (2005) Male chacma baboons (Papio hamadryas ursinus) discriminate loud call contests between rivals of different relative ranks. *Animal Cognition* 8:1–6.

Komarova NL, Niyogi P, Nowak Ma (2001) The evolutionary dynamics of grammar acquisition. *Journal of Theoretical Biology* 209:43–59.

Krams I, Krama T, Freeberg TM, Kullberg C, Lucas JR (2012) Linking social complexity and vocal complexity: A parid perspective. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* 367:1879–91.

Krebs JR, Davies NB (1997) *Behavioural Ecology: An Evolutionary Approach* Blackwill Scientific Publications, Oxford.

Kroodsma DE (1974) Song Learning, Dialects, and Dispersal in the Bewick's Wren. *Zeitschrift für Tierpsychologie* 35:353–380.

Laiolo P, Palestrini C, Rolando A (2000) A study of Choughs' vocal repertoire: variability related to individuals, sexes and ages. *Journal of Ornithology* 141:168–179.

Larrañaga P, Calvo B, Santana R, Bielza C, Galdiano J, Inza I, Lozano Ja, Armañanzas R, Santafé G, Pérez A, Robles V (2006) Machine learning in bioinformatics. *Briefings in Bioinformatics* 7:86–112.

Leliveld LMC, Scheumann M, Zimmermann E (2011) Acoustic correlates of individuality in the vocal repertoire of a nocturnal primate (*Microcebus murinus*). *The Journal of the Acoustical Society of America* 129:2278–2288.

Lemasson a, Hausberger M (2011) Acoustic variability and social significance of calls in female Campbell's monkeys (*Cercopithecus campbelli campbelli*). *The Journal of the Acoustical Society of America* 129:3341–52.

Leone M, Weigt S, Weigt M (2007) Clustering by soft-constraint affinity propagation: Applications to gene-expression data. *Bioinformatics* 23:2708–2715.

Lesot MJ, Mouillet L, Bouchon-Meunier B (2006) Fuzzy prototypes based on typicality degrees. *Advances in Soft Computing* 33:125–138.

Leuchtenberger C, Duplaix N, Magnusson WE (2015) Vocal repertoire of the social giant otter. *Journal of the Acoustic Society of America* 136.

Liberman aM, Harris KS, Hoffman HS, Griffith BC (1957) The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology* 54:358–368.

Lieberman P, Klatt D, Wilson W (1969) Vocal tract limiations on the vowel repertoires of rhesus monkey and other nonhuman primates. *Science* 164:1185–1187.

Lopatka M, Adam O, Laplanche C, Zarzycki J, Motsch JF (2005) An Attractive Alternative for Sperm Whale Click Detection Using the Wavelet Transform in Comparison to the Fourier Spectrogram. *Aquatic Mammals* 31:463–467.

Maciej P, Fischer J, Hammerschmidt K (2011) Transmission characteristics of primate vocalizations: implications for acoustic analyses. *PLoS ONE* 6:e23015.

Maciej P, Ndao I, Hammerschmidt K, Fischer J (2013) Vocal communication in a complex multi-level society: constrained acoustic structure and flexible call usage in Guinea baboons. *Frontiers in Zoology* 10:58.

MacLarnon A, Hewitt G (1999) The Evolution of Human Speech: The Role of Enhanced Breathing Control. *American Journal of Physical Anthropology* 109:341–363.

MacQueen J (1967) Some methods of classification and analysis of multi- variate observations In *Proceedings of the fifth Berkeley symposium in mathematical statistics and probability*, Vol. 1, pp. 281–297.

Manser MB (2001) The acoustic structure of suricates' alarm calls varies with predator type and the level of response urgency. *Proceedings of the Royal Society B: Biological Sciences* 268:2315–2324.

Markram H, Toledo-Rodriguez M, Wang Y, Gupta A, Silberberg G, Wu C (2004) Interneurons of the neocortical inhibitory system. *Nature Reviews. Neuroscience* 5:793–807.

Marler P (1967) Animal Communication Signals. *Science* 157:769–774.

Marler P (1970) Vocalizations of East African monkeys: I. Red colobus. *Folia Primatologia* 13:81–91.

Marler P (1975) On the origin of speech from animal sounds In Kavanagh J, Cutting J, editors, *The Role of Speech in Language*, pp. 11–37. Cambridge, Massachusetts: MIT Press.

Marler P (1976) Social organization, communications and graded signals: the chimpanzee and the gorilla In Bateson, Hinde, editors, *Growing Points in Ethology*, pp. 239–280. Cambridge: Cambridge University Press.

Marler P (1977) The structure of animal communication sounds In Bullock TH, editor, *Recognition of Complex Acoustic Signals*, pp. 17–35. Springer Verlag, Berlin.

Marler P, Pickert R (1984) Species-universal microstructure in the learned song of the swamp sparrow (*Melospiza georgiana*). *Animal Behaviour* 32:673–689.

McComb K, Semple S (2005) Coevolution of vocal communication and sociality in primates. *Biology Letters* 1:381–5.

McCowan B (1995) A New Quantitative Technique for Categorizing Whistles Using Simulated Signals and Whistles from Captive Bottlenose Dolphins (*Delphinidae, Tursops truncatus*). *Ethology* pp. 177–193.

McKusick Va (1969) On lumpers and splitters, or the nosology of genetic disease. *Perspectives in Biology and Medicine* 12:298–312.

Meyer D, Hodges JK, Rinaldi D, Wijaya A, Roos C, Hammerschmidt K (2012) Acoustic structure of male loud-calls support molecular phylogeny of Sumatran and Javanese leaf monkeys (genus *Presbytis*). *BMC Evolutionary Biology* 12:16.

Morton ES (1975) Ecological sources of selection on avian sounds. *The American Naturalist* 108:17–34.

Morton ES (1977) On the Occurrence and Significance of Motivation-Structural Rules in Some Bird and Mammal Sounds. *The American Naturalist* 111:855.

Nelson DA, Marler P (1989) Categorical Perception of a Natural Stimulus: Birdsong. *Science* 244:976–978.

Nottebohm F (1972) The origins of vocal learning. *The American Naturalist* pp. 116–140.

Nowak Ma, Plotkin JB, Jansen Va (2000) The evolution of syntactic communication. *Nature* 404:495–8.

Outarra K, Lemasson A, Zuberbühler K (2009) Campbell's Monkeys Use Affixation to Alter Call Meaning. *PLoS ONE* 4.

Owings DH, Morton ES (1998) *Animal Vocal Communication - A New Approach* Cambridge University Press.

Owren MJ, Seyfarth RM, Cheney DL (1997) The acoustic features of vowel-like grunt calls in chacma baboons (*Papio cyncephalus ursinus*): implications for production processes and functions. *The Journal of the Acoustical Society of America* 101:2951–63.

Padgham M (2004) Reverberation and frequency attenuation in forests - implications for acoustic communication in animals. *The Journal of the Acoustical Society of America* 115:402.

Papworth S, Böse AS, Barker J, Schel AM, Zuberbühler K (2008) Male blue monkeys alarm call in response to danger experienced by others. *Biology Letters* 4:472–5.

Pfefferle D, Fischer J (2006) Sounds and size: identification of acoustic variables that reflect body size in hamadryas baboons, *Papio hamadryas*. *Animal Behaviour* 72:43–51.

Rakesh V, Datta AK, Ducharme NG, Pease AP (2008) Simulation of Turbulent Aiflow Using a CT Based Upper Airway Model of a Racehorse. *Journal of Biomechanical Engineering* 130.

Reby DD, McComb DK (2003) Anatomical constraints generate honesty: acoustic cues to age and weight in the roars of red deer stags. *Animal Behaviour* 65:519–530.

Rendall D, Seyfarth RM, Cheney DL, Owren MJ (1999) The meaning and function of grunt variants in Baboons. *Animal Behaviour* 57:583–592.

Rendall D, Cheney DL, Seyfarth RM (2000) Proximate factors mediating "contact" calls in adult female baboons (*Papio cynocephalus ursinus*) and their infants. *Journal of Comparative Psychology* 114:36–46.

Rendall D, Kollias S, Ney C, Lloyd P (2005) Pitch (F0) and formant profiles of human vowels and vowel-like baboon grunts: the role of vocalizer body size and voice-acoustic allometry. *The Journal of the Acoustical Society of America* 117:944–955.

Rendall D, Notman H, Owren MJ (2009) Asymmetries in the individual distinctiveness and maternal recognition of infant contact calls and distress screams in baboons. *The Journal of the Acoustical Society of America* 125:1792–1805.

Riede T, Bronson E, Hatzikirou H, Zuberbühler K (2005) Vocal production mechanisms in a non-human primate: morphological data and a model. *Journal of Human Evolution* 48:85–96.

Riede T, Fitch T (1999) Vocal Tract Length and Acoustics of Vocalization in the Domestic Dog (*Canis Familiaris*). *The Journal of Experimental Biology* 202:2859–2867.

Riede T, Titze IR (2008) Vocal fold elasticity of the Rocky Mountain elk (*Cervus elaphus nelsoni*) - producing high fundamental frequency vocalization with a very long vocal fold. *The Journal of Experimental Biology* 211:2144–2154.

Ripley BD (1996) *Pattern Recognition and Neural Networks* Cambridge University Press.

Rousseeuw PJ (1987) Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics* 20:53–65.

Rowell TE, Hinde RA (1962) Vocal communication by the rhesus monkey (*Macaca mulatta*). *Proceedings of the Zoological Society of London* 138:279–294.

Scattoni ML, Gandhy SU, Ricceri L, Crawley JN (2008) Unusual repertoire of vocalizations in the BTBR T+tf/J mouse model of autism. *PLoS ONE* 3:48–52.

Scheiner E, Hammerschmidt K, Jürgens U, Zwirner P (2002) Acoustic analyses of developmental changes and emotional expression in the preverbal vocalizations of infants. *Journal of Voice* 16:509–29.

Schön PC, Puppe B, Manteuffel G (2001) Linear prediction coding analysis and self-organizing feature map as tools to classify stress calls of domestic pigs (*Sus scrofa*). *The Journal of the Acoustical Society of America* 110:1425–1431.

Selin A, Turunen J, Tanttu JT (2007) Wavelets in recognition of bird sounds. *Journal on Advances in Signal Processing* pp. 141–150.

Shannon CE (1948) A Mathematical Theory of Communication. *The Bell System Technical Journal* 27:379–423.

Shulezhko TS, Burkanov VN (2008) Stereotyped acoustic signals of the killer whale *Orcinus orca* (*Cetacea: Delphinidae*) from the Northwestern Pacific. *Russian Journal of Marine Biology* 34:118–125.

Silk JB, Seyfarth RM, Cheney DL (1999) The structure of social relationships among female savanna baboons in Moremi Reserve, Botswana. *Behaviour* 136:679–703.

Skyrms B (2010) *Signals, Evolution, Learning and Information* Oxford University Press, Oxford.

Slocombe KE, Townsend SW, Zuberbühler K (2009) Wild chimpanzees (*Pan troglodytes schweinfurthii*) distinguish between different scream types: Evidence from a playback study. *Animal Cognition* 12:441–449.

Specht R (2004) Avisoft -SASLab Pro. Sound Analysis and Synthesis Laboratory.

Stoeger AS, Charlton BD, Kratochvil H, Fitch WT (2011) Vocal cues indicate level of arousal in infant African elephant roars. *The Journal of the Acoustical Society of America* 130:1700.

Story BH, Titze IR, Hoffman Ea (1996) Vocal tract area functions from magnetic resonance imaging. *The Journal of the Acoustical Society of America* 100:537–554.

Story BH, Titze IR (1995) Voice simulation with a body-cover model of the vocal. *Journal of the Acoustic Society of America* 97:1249–1260.

Stowell D, Plumbley MD (2014) Automatic large-scale classification of bird sounds is strongly improved by unsupervised feature learning. *PeerJ* 2.

Tallet C, Linhart P, Policht R, Hammerschmidt K, Šimeček P, Kratinova P, Špinka M (2013) Encoding of situations in the vocal repertoire of piglets (*Sus scrofa*): a comparison of discrete and graded classifications. *PLoS ONE* 8.

Tanaka T, Masataka N, Sugiura H (2006) Sound transmission in the habitats of Japanese macaques and its possible effect on population differences in coo calls. *Behaviour* 143:993–1012.

Taylor M, Reby D (2010) The contribution of source-filter theory to mammal vocal communication research. *Journal of Zoology* 280:221–236.

Tchernichovski O, Nottebohm F, Ho C, Pesaran B, Mitra P (2000) A procedure for an automated measurement of song similarity. *Animal Behaviour* 59:1167–1176.

Templeton CN, Laland KN, Boogert NJ (2014) Does song complexity correlate with problem-solving performance in flocks of zebra finches? *Animal Behaviour* 92:63–71.

Thinh VN, Hallam C, Roos C, Hammerschmidt K (2011) Concordance between vocal and genetic diversity in crested gibbons. *BMC Evolutionary Biology* 11:36.

Titze IR (1994) *Principles of Voice Production* Prentice Hall, Upper Saddle River, NJ 07458.

Titze IR, Riede T (2010) A Cervid Vocal Fold Model Suggests Greater Glottal Efficiency in Calling at High Frequencies. *PLoS Computational Biology* 6.

Tomasello M, Zuberbühler K (2002) Primate Vocal and Gestural Communication In Allen C, Bekoff M, Burghardt C, editors, *The Cognitive Animal*, pp. 293–299. MIT Press, Cambridge.

Tononi G, Edelman GM, Sporns O (1998) Complexity and coherency : integrating information in the brain. *Trends in Cognitive Sciences* 2:474–484.

Torrence C, Compo GP (1998) A Practical Guide to Wavelet Analysis. *Bulletin of the American Meteorological Society* 79:61–78.

Townsend SW, Manser MB (2011) The function of nonlinear phenomena in meerkat alarm calls. *Biology Letters* 7:47–49.

Townsend SW, Zuberbuhler K (2009) Audience effects in chimpanzee copulation calls. *Communicative & Integrative Biology* 2:282–284.

Valens C (1999) A Really Friendly Guite to Wavelets Technical report.

Vehrencamp S (2000) Handicap, index, and conventional signal elements of bird song In *Animal Signals: Signalling and Signal Design in Animal Communication*, pp. 277–300.

Vogelstein JT, Park Y, Ohyama T, Kerr RA, Truman JW, Priebe CE, Zlatic M (2014) Discovery of Brainwide Neural-Behavioral Maps via Multiscale Unsupervised Structure Learning. *Science* 344:386–392.

Wadewitz P, Hammerschmidt K, Battaglia D, Witt A, Wolf F, Fischer J (2015a) Characterizing Vocal Repertoires - Hard vs . Soft Classification Approaches. *PLoS ONE* 10.

Wadewitz P, Hammerschmidt K, Battaglia D, Wolf F, Fischer J (2015b) A quantitative comparison of graded vocal repertoires. *submitted* .

Ward JH (1963) Hierarchical Grouping to Optimize an Objective Function. *Journal of the American Statistical Association* 58:236–244.

Waser PM, Brown CH (1986) Habitat acoustics and primate communication. *American Journal of Primatology* 10:135–154.

Weilgart L, Whitehead H (1997) Group-specific dialiects and geographical variation in coda repertoire in South Pacific sperm whales. *Behavioral Ecology and Sociobiology* 40:277–285.

Whittaker RH (1953) A Consideration of the Climax Theory: The Climax as a Population and Pattern. *Ecological Monographs* 23:41–78.

Wilczynski W, Rand Sa, Ryan MJ (1995) The processing of spectral cues by the call analysis system of the túngara frog,*Physalaemus pustulosus*. *Animal Behaviour* 49:911–929.

Winter P, Ploog D, Latta J (1966) Vocal repertoire of the squirrel monkey (*Saimiri sciureus*), its analysis and significance. *Experimental Brain Research* 1:359–84.

Xu D, Keller JM, Popescu M, Bondugula R (2008) *Applications of fuzzy logic in bioinformatics* London: Imperial College Press.

Zadeh LA (1965) Fuzzy Sets. *Information and Control* 8:338–353.

Zadeh LA (2008) Is there a need for fuzzy logic? *Information Sciences* 178:2751–2779.

Zuberbühler K, Noë R, Seyfarth RM (1997) Diana monkey long-distance calls: messages for conspecifics and predators. *Animal Behaviour* pp. 589–604.

Philip Wadewitz

German Citizen: Born Bielefeld, North Rhine-Westphalia (31.10.1983)

Education and Qualification

2011 - 2015
PhD, German Primate Center, University of Göttingen
Dissertation: "Processing of graded signaling systems."

2008 - 2010
Master of Science, Biology, Behavioral Sciences, University of Zurich, Switzerland
Thesis: "Olfactory discrimination of predators and conspecifics in meerkats (*Suricata suricatta*)"

2005 - 2008
Bachelor of Science, Life Sciences, University of Münster, Germany
Thesis: "Colony housing during pregnancy and lactation in guinea pigs: effects on the female offspring endocrine profile and behavior"

Professional & Teaching Experience

2014
Cognitive Ethology Laboratory, University of Göttingen, Germany
Graduate assistant. Supervision of students during the biannual field course on social behavior and communication in Rocamadour, France

2010-2011
International Otter Survival Fund, Isle of Skye, Scotland
Volunteer. Tracking wild otters and fecal sample analysis

2009 - 2010
Department of Ecology, University of Zurich, Switzerland
Graduate assistant. Lab work assistance for a project on Sepsis fly development

2008 - 2009
Department of Animal Behavior, University of Zurich, Switzerland
Graduate assistant. Collection of life-history data for a long-term project on house mice

Publications

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

Price T, Wadewitz P, Cheney D, Seyfarth R, Hammerschmidt K, Fischer J (2015) Vervets revisited: A quantitative analysis of alarm call structure and context specificity. Scientific Reports.

Conference Papers and Presentations

Sep 2015
Wadewitz P, Hammerschmidt K, Fischer J. Advances in the quantitative analysis of graded vocal repertoires. Talk presented at the 25th International Bioacoustics Congress, Murnau, Germany.

Aug 2015
Wadewitz P, Hammerschmidt K, Fischer J. Characterizing vocal repertoires - Shortcoming of current methods and future perspectives. Talk presented at 6th European Federation for Primatology Meeting, Rome, Italy.

Aug 2014
Wadewitz P, Hammerschmidt K, Fischer J. Towards an objective classification of vocal repertoires - analyzing the graded nature of primate calls. Talk presented at 25th International Primatological Society Congress, Hanoi, Vietnam.

Sep 2013
Wadewitz P, Hammerschmidt K, Battaglia D, Witt A, Wolf F, Fischer J. Towards an objective classification of vocal repertoires. Talk presented at 24th International Bioacoustics Congress, Pirénopolis, Brazil.

Dec 2012
Wadewitz P, Hammerschmidt K, Battaglia D, Witt A, Wolf F, Fischer J. Acoustic analysis of the chacma baboon's (*Papio ursinus*) vocal repertoire. Poster presented at ASAB Winter Meeting 2012, London, UK.

Nov 2011
Wadewitz P, Hammerschmidt K, Fischer J. Processing of graded signaling systems. Talk presented at Winter School 2011, Cognition and Communication, Vienna, Austria.