

## **Orangutan call communication and the puzzle of speech evolution**

© Adriano Reis e Lameira 2013

Orangutan call communication and the puzzle of speech evolution

PhD dissertation, Utrecht University, the Netherlands, January 2013

ISBN: 978-94-6191-583-2

Cover design & lay-out: Adriano R. Lameira

Cover illustrations: Aloys Zötl, 1831-1887

Printing: Ipskamp Drukkers, Enschede, the Netherlands

Printed on FSC paper

Voor Madeleine

To Izzy

Untuk mawas

Aos meus pais



# **Orangutan call communication and the puzzle of speech evolution**

*Orang-oetan geluidscommunicatie en  
de evolutionaire puzzel van spraak*

(met een samenvatting in het Nederlands)

## **PROEFSCHRIFT**

ter verkrijging van de graad van doctor aan de Universiteit Utrecht  
op gezag van de rector magnificus, prof.dr. G.J. van der Zwaan,  
ingevolge het besluit van het college voor promoties  
in het openbaar te verdedigen  
op woensdag 23 januari 2013 des ochtends te 10.30 uur

door

**Adriano Reis e Lameira**

*geboren op 1 september 1982  
te Lisbon, Portugal*

Promotoren: Prof.dr. B.M. Spruijt

Prof.dr. S.A. Wich

Co-promotor: Dr. E.H.M. Sterck

This thesis was (partly) accomplished with financial support from Fundação para a Ciência e Tecnologia (SFRH/BD/44437/2008, Portugal), Wenner-Gren Foundation for Anthropological Research (USA), Dr. J.L. Dobberke voor Vergelijkende Psychologie (the Netherlands), Lucie Burgers Foundation for Comparative Behaviour Research, Arnhem, the Netherlands, Schure-Beijerinck-Popping Fonds (the Netherlands), Ruggles-Gates Fund for Anthropological Scholarship of the Royal Anthropological Institute of Great Britain and Ireland, Primate Conservation, Inc. (USA) and Fundação Calouste Gulbenkian (Portugal).

# Contents

Chapter I	General introduction	09
Chapter II	Review of geographic variation in terrestrial mammalian acoustic signals: Human speech variation in a comparative perspective	21
Chapter III	Call cultures in orangutan?	45
Chapter IV	Tool use in wild orangutan modifies sound production: A functionally deceptive innovation?	63
Chapter V	Population-specific use of the same tool-assisted alarm sound between two wild orangutan populations ( <i>Pongo pygmaeus wurmbii</i> ) denotes functional arbitrariness	75
Chapter VI	Orangutan whistling and implication for the emergence of an open-ended call repertoire: a replication and extension	89
Chapter VII	Predator guild does not influence orangutan alarm call rates and combinations	111
Chapter VIII	Orangutan <i>instrumental gesture-calls</i> : Reconciling acoustic and gestural speech evolution models	129
Chapter IX	General discussion	137
Chapter X	Summary	147
	Appendices	153
	References	157
	Samenvatting in het Nederlands	177
	Ringkasan bahasa Indonesia	183
	Acknowledgments	189
	Curriculum Vitae	191
	List of publications	193



## Chapter I

# General introduction

### *The evolutionary puzzle of speech*

Language, human's learned communication system, is the most complex communication system known in nature. With the Darwinian theory (Darwin 1859), the study of language opened up to those fascinated by its origins and evolution. Today, the study of language evolution is a multidisciplinary endeavor, mirroring the plurality of its components (Christiansen and Kirby 2003b). One of the pillar components of language evolution is the study of animal communication (aside other components focusing, for example, on language acquisition, language neurological and genetic correlates), in which similarities and differences in characteristics with language are studied. The identification of such differences allows pinpointing features which are unique to language, whereas the identification of similarities allows identifying, either by analogy (i.e. non-primate animal studies) or homology (e.g. nonhuman primate studies), selective pressures that are common in animal and human communication. This task is undertaken by comparative biology and is directly linked to the basic Darwinian premise that language is the result of a gradual evolutionary process stemming from an ancestral form.

Speech, the spoken component of humans' learned communication system, constitutes the acoustic part of language (Fitch 2000). A theory for language evolution in humans must include a valid explanation for the evolution of speech in our ancestors (Christiansen and Kirby 2003b). Allegorically, speech can be understood as a piece of (physical) hardware, together with its respective operations, which has been designed to serve specific (virtual) software – language. As such, speech differs explicitly from language and it may be studied independently, with the advantage of being more accessible to experimental investigation than other components of language (Fitch 2000), just as deconstructing hardware into its composing parts is more straightforward than for software. This is especially true under a comparative approach to speech evolution, where animal calls, the main study subjects of comparative approaches to speech evolution, are readily available to investigation under the natural socio-ecologic settings of the species.

There is some consensus about the distinctive ways speech differs from other call communication systems in the animal kingdom (Fitch 2000), that is, which transformations must have occurred in our evolutionary lineage until speech had evolved as we know it at

present. It is generally accepted that the evolution of human speech from ape call communication required, at least, a modification of the ape vocal tract morphology (Boer 2009) and the acquisition of the capacity to learn new calls (Fitch 2000; Ghazanfar and Rendall 2008). However, speech evolutionary history remains elusive. That is, the exact primate call communication traits that were targeted by positive selection leading to the necessary transformations remain unclear. As such, this topic has always been controversial (Christiansen and Kirby 2003b), leading some researchers to propose that speech precursors are found beyond call communication, for instance, in gestural communication (e.g. Arbib et al. 2008; Corballis 2003). Whether or not specific speech features may have been brought forth initially by behavioral patterns outside call communication, such putative precursors must have been eventually “translated” into call features in speech’s evolutionary path. Therefore, the question remains, which traits and functions of nonhuman primate call communication served as speech evolutionary feedstock. Hence, it is beneficial to the deciphering of the speech puzzle to assess its building blocks within a comparative approach by focusing on homological features of call communication in nonhuman primates closely related to humans, namely great apes. In this study, I investigate how much of speech features can be found in orangutan call communication.

### *A brief history of primate call communication studies*

The first primatological studies focusing on call communication were initiated more than a century ago, in the late nineteenth century (Garner 1892; cf. Radick 2007). This was the first time that nonhuman primate calls were recorded and played back to conspecifics to elicit behavioral responses. By comparing the original observation and the replies from monkeys, Garner hoped to get some clues that would eventually enable him to translate the “monkey language” (Radick 2007), although detailed results remained unpublished. Decades later, primate studies featured dominantly in discussions on language and speech evolution (Savage-Rumbaugh et al. 1978; Savage-Rumbaugh et al. 1985; Gardner and Gardner 1969; Premack 1971; Rumbaugh 1977; Terrace et al. 1979; Patterson and Linden 1981; Furness 1916; Hayes 1951; Hayes and Hayes 1952; Miles 1993; Kellogg and Kellogg 1967). These studies characteristically involved great apes who were reared with humans in the course of language training. Despite this concerted effort, the main objective of teaching an ape how to speak largely failed, since the subjects failed to learn calls or only

learned very few calls (e.g. Furness 1916). Instead, communication between the apes and their caretakers seemed most successful when using gestures, with some ape individuals learning hundreds of symbolic gestures (e.g. Savage-Rumbaugh and Lewin 1994). Based on this apparent lack of results in call communication, together with neurological studies showing little or no voluntarily call control in monkeys (e.g. Jürgens 1979), the general notion that nonhuman primates have an innate call repertoire and cannot learn new calls (Hammerschmidt and Jischer 2008) was implanted, and scientific thought moved away from the view that great ape calls could represent some type of speech precursors. This corroborated the pervasive theoretical idea of the time that language was the product of a saltational, rather than gradual, event (Chomsky 1957) that divides human language and speech from primate call communication. The foundations of language must have rested, hence, in other primate features such as gestures or facial expressions (cf. Slocombe et al. 2011), rather than call communication, and good animal models for speech evolution are better found in other taxa. This notion has remained influential until today, either in studies resuming century-old ideas based on anti-anthropomorphism (i.e. opposition against the search of human features in animals) defending discontinuity between humans and its closest relatives for the evolution of language and speech (e.g. Wynne and Bolhuis 2008; Hemelrijk and Bolhuis 2011), or in studies postulating gradual and continuous evolutionary processes between humans and nonhuman primates but advocating, for instance, a gestural origin for speech (e.g. Arbib et al. 2008; Corballis 2003). This led comparative research to focus on taxa less related to humans. One of the most important insights made by this line of research was that, similar to humans, distantly-related taxa may socially learn new calls and incorporate them into their repertoire. For instance, repertoire ontogeny in song birds and bats can show similarities with speech acquisition (Blake and Fink 1987), where young individuals go through a babbling phase as an opportunity to practice elements of the adult call repertoire (Aronov et al. 2008; Knörnschild et al. 2006; cf. Elowson et al. 1998). In song birds, parrots and aquatic mammals vocal learning may continue in the adult phase (Laland and Janik 2006; Rendell and Whitehead 2001; Noad et al. 2000; Todt 1975), demonstrating that, similarly to humans, the expansion of one's repertoire is not strictly limited to a sensitive phase at young age.

In the second half of the twentieth century, primatology returned to the study of calls produced by animals in the wild (Cheney and Seyfarth 1992; Seyfarth et al. 1980; Zuberbühler et al. 1997; Zuberbühler 2000b), fundamentally reinventing primate playback experiments after almost a century (cf. Garner 1892). These studies were the first to

essentially apply an ethological approach to the examination of monkey calls, where playback stimuli concerned conspecific calls and the subjects' responses took place within animals' natural socio-ecological environment and therefore were embedded with biological meaningful information. This provided a remarkable window into "how monkeys see the world" (Cheney and Seyfarth 1992), revealing the use and understanding of advanced elements, such as the use and understanding of semantic referents to external events in conspecific (Seyfarth et al. 1980; Zuberbühler et al. 1997) and interspecific calls. Nonhuman primates were shown to extract information from calls about objects and/or events in their surroundings and respond to these calls in an equivalent way as if they had encountered instead that same object or event. Hence, as human words, nonhuman calls may carry semantic information to conspecifics, where one call corresponds to one object or event, much in a similar way words name different objects. This body of literature suggests that, unlike some other animal taxa and humans, nonhuman primates did not learn new calls, they were using their own innate calls in sophisticated ways.

Other research during this period based on cross-fostering experiments offered renewed support to the notion that monkey calls were essentially innate and not socially learned (Owren et al. 1993; Marler and Tenaza 1977). Infants of two different monkey species raised in the presence of and by adults of the opposite species only produced the calls of their species of origin and not of their foster species. Nevertheless, this issue remained ambiguous in apes (Geissmann 1984), where some studies suggested the inheritance of stereotyped calls in interbreeding areas in the wild (Brockelman and Schilling 1984), while hybrid calls documented in cross-breeding experiments could not be attributed to the parent or foster species and sources of learning could not be identified.

After the turn of the 20<sup>th</sup> century, the body of literature on primate ethology offered a refined understanding on primate call communication, implying that the traditional view of a fixed call repertoire was too simplistic. The new ideas divided primate call communication in four major components – use, comprehension, production and acquisition – and allowed new research to determine the actual extent of flexibility in primate call communication.

In the domain of call use, nonhuman primates may, for instance, adjust the production of calls in relation to the presence (Wich and Sterck 2003; Slocombe et al. 2010) or composition of their audience (Slocombe and Zuberbühler 2007; Slocombe and Zuberbühler 2005a; Laporte and Zuberbühler 2010; Clay et al. 2011; Slocombe and Zuberbühler 2006; Townsend et al. 2008; Slocombe et al. 2010), where call production is prevented in the absence of (important) social partners. Nonhuman primates may also

regulate the production of calls according to the state of knowledge of their audience (Wich and de Vries 2006; Crockford et al. 2011), where individuals may only call when receivers (show signs to) remain ignorant about events in the environment. Moreover, individuals may accelerate or decelerate the production rate of calls according to external events (Lemasson et al. 2010b), where faster rates are delivered with increased danger, even when danger is experienced by other group members (Papworth et al. 2008). Individuals may also fuse different call-types to form call-combinations and sequences signifying different causing factors (Ouattara et al. 2009b, 2009a; Zuberbühler 2002; Arnold and Zuberbühler 2006, 2008; Clarke et al. 2006; Stephan and Zuberbühler 2008; Schel et al. 2010; Clay and Zuberbühler 2011; Candiotti et al. 2012a; Arnold and Zuberbühler 2012).

In the domain of call comprehension, nonhuman primates may, for instance, recognize the identity of other individuals based on acoustic cues alone (Wich et al. 2002; Cheney and Seyfarth 1980), react more attentively to informed individuals (Lemasson et al. 2010a), associate calls with long-term social memory (Lemasson et al. 2005) and have a causal knowledge of calls heard in their environment (Zuberbühler 2000a). Moreover, old world monkeys (e.g. Zuberbühler et al. 1997; Seyfarth et al. 1980), new world monkeys (Cäsar et al. 2012; Wheeler 2010) and great apes (Slocombe and Zuberbühler 2005b, 2005a) may extract semantic information from calls, that is, perceive calls as though they were exposed to the object or event to which the call refers to. This may occur in agonistic (Slocombe and Zuberbühler 2005a) or foraging contexts (Slocombe and Zuberbühler 2005b), where individuals consistently respond when calls refer either to different types of predators with different hunting techniques or to different types of food with different levels of preference. Besides extracting call meaning from acoustic cues of different call-types, nonhuman primates may also extract call meaning from the sequence and order of different call-types (Ouattara et al. 2009b, 2009a; Zuberbühler 2002; Arnold and Zuberbühler 2006, 2008; Clarke et al. 2006; Stephan and Zuberbühler 2008; Schel et al. 2010; Clay and Zuberbühler 2011; Candiotti et al. 2012a; Arnold and Zuberbühler 2012), where individuals attend to particular combinatorial rules to assess the meaning of the sequences of calls.

In the domain of production, nonhuman primates may, for instance, show control over specific structures of their vocal apparatus and/or involved in call production, such as their vocal tract (Koda et al. 2012), lips (Morrill et al. 2012; Ghazanfar et al. 2012; Wich et al. 2009b), vocal fold action (Lemasson et al. 2011; Candiotti et al. 2012b; Riede et

al. 2005; Taglialatela et al. 2003) and air flow (Wich et al. 2009b), basically allowing individuals to produce variants within certain call types.

This cumulative body of literature demonstrates that primate call communication is plastic in terms of its use, comprehension and production. However, attention has been mainly given to monkeys and lesser apes, and much evidence gathered for these species still remains to be comprehensively assessed in great apes across these three domains. Call acquisition remains the domain where primate behavior in general is considered to be least flexible. There has been some renewed research effort, however, on the capacity of learning new calls across nonhuman primates, where auditory feedback and experience is crucial for proper call production (Wich et al. 2009b). Earlier great ape language projects (see above) were embedded in human-oriented perspectives (Savage-Rumbaugh et al. 1978; Savage-Rumbaugh et al. 1985; Gardner and Gardner 1969; Premack 1971; Rumbaugh 1977; Terrace et al. 1979; Patterson and Linden 1981; Furness 1916; Hayes 1951; Hayes and Hayes 1952; Miles et al. 1996; Kellogg and Kellogg 1967) and aimed to teach human words to nonhuman primates, which may partly explain why these efforts were unproductive according to their original goal. These studies largely ignored great ape call communication as found in the wild or naturalistic conditions and among conspecifics. Moreover, these studies probably just failed because they did not consider the fact that great apes are physically unable to produce all human vowels and consonants (Lieberman et al. 1972; Lieberman 2002a), and did not attend to the calls that great apes are able to produce. Remarkably, evidence of flexibility in call production was overlooked or downplayed on the basis of the expectation that apes should be able to talk fluently, even when an ape actually acquired a call based on learning. Furness (1916) was possibly the first biologist to describe, almost a century ago, the acquisition and production on request of “papa” and “cup” by an orangutan, and later studies described similar capacities with other apes (Hayes 1951). The evidence that these apes could learn some new word-like calls was overshadowed by their lack of learning speech at a similar rate as humans.

### *Great apes: a basic model for speech evolution*

Great apes are our closest relatives in the animal kingdom (Hobolth et al. 2011; Scally et al. 2012; Prufer et al. 2012). Under evolutionary premises, the continuity between humans and other primates implies that every speech aspect evolved directly or indirectly from an ancestral great ape counterpart. For instance, while cetaceans and songbirds are apt vocal

learners in the acoustic domain, this is due to evolutionary convergence and findings on these taxa will remain inconclusive about speech precursors within the primate lineage, their evolutionary timing, sequence, relative importance or interactions. Hence, to address speech evolutionary history, the similarities and differences between great apes and humans have to be examined to identify precursors to human speech.

Distinctively, as humans' closest relatives, primates in general (Riede et al. 2005; Gamba and Giacoma 2006; Gamba and Giacoma 2010), and great apes in particular (Nishimura et al. 2003), present the most similar vocal tract anatomy to humans in the animal kingdom. Human and nonhuman primate vocal tracts are best modeled as a non-uniform vocal tract consisting of multiple tubes (Riede et al. 2005; Gamba and Giacoma 2006). Moreover, vocal folds represent the major source of call production in both groups, and the respective acoustic signals produced may be filtered by articulatory maneuvers performed by similar structures, such as the tongue and lips (Riede et al. 2005; Fitch 2000). Since speech represents a piece of hardware operating under the computational and control commands of language, the anatomy of speech (hardware's structure and parts) is basic in the capacity for speech in humans (Fitch and Reby 2001; Fitch 2000; Riede et al. 2005; Nishimura et al. 2003). Hence, anatomical similarity in oro-laryngeal anatomy between model species and humans is essential for an accurate comparative approach to speech evolutionary history.

### *New comparative tools*

*"No problem can be solved from the same level of consciousness that created it."*

(Albert Einstein)

This thesis is built on the premise that renewed interest in great ape call communication will advance the discipline of speech evolution. I propose that, based on the close relatedness between humans and great apes, translated in anatomical resemblance, primatology can borrow concepts and definitions accumulated by linguistics and semiotics. These disciplines comprise the social sciences dedicated to the study of human speech and human signs, respectively. The first scripts on human speech and signs predate classical antiquity and evolutionary theories by centuries and millennia respectively (Favareau 2010). Thus, they offer mature terminology describing speech features which primatology may use as a comparative tool for the study of speech evolution. For instance, primate

calls may be said to be vowel-like based on their articulatory and acoustic similarity with their human counterparts (e.g. Owren et al. 1997). Such term use is technically correct and does not rely on acoustic similarity alone (as essentially is the case with comparative studies that focus on non-primate taxa). Heuristically, borrowing concepts helps directly differentiating homologies from analogies between speech and great ape call communication, since speech traits that are not found in great apes probably represent derived features. On the other hand, if speech traits, as they are defined in linguistics and semiotics, such as “dialect” or “arbitrariness”, are found in great apes, this may provide an immediate assessment of the trait’s evolutionary age within the human lineage.

As an extension of this terminological use, this study examines the building blocks of speech following a framework outlined by the hands of a linguist, Charles Hockett (1968). I use this system as a reference, because this breakdown of speech has been one of the most influential in the field of animal communication and because more recent, alternative ways to decompose language are subject of considerable debate (Fitch 2010). Hockett’s (1968) system comprises thirteen speech features, divided into those shared with animal calls, and those considered unique to humans. The most advanced features are those which must be accounted for by any valid hypothesis of speech evolution, thus, this system may constitute a logical heuristic basis. The first five, simplest, features are shared within the animal kingdom (“vocal-auditory channel”, “broadcast transmission and directional reception”, “rapid fading”, “interchangeability”, “total feedback”) and do not require further scrutiny. Therefore, I specifically assess in orangutan call communication six of the advanced features considered by Hockett: “traditional transmission”, “prevarication”, “arbitrariness”, “learnability”, “semanticity” and “patterning”. Traditional transmission describes that calls are learned within social groups. Prevarication is the ability to make false statements, to deceive, involving the manipulation of a given shared principle in order to fool other individual. Arbitrariness applies when there is no direct connection between a call and its function. Learnability defines that members of one call system can learn to produce calls from another system. Semanticity concerns calls that can be linked to specific meanings. Patterning describes the ability to combine a finite number of call elements to create new, composed, elements. Two additional advanced features, “displacement” (the ability to refer to things not physically present) and “reflexiveness” (a call system may be used to refer to itself), are not assessed here.

Conceptually, this method represents a paradigm shift from earlier great ape language projects mentioned above. In the latter, the initial focus was to teach apes how to talk. Today, one may examine speech features in the absence of (full-blown) speech

within great ape call repertoires and their use. Supplied with this new comparative method, I revisit and test call abilities of great apes. I focus on orangutans, as they have been argued to represent the great ape species best suited for this endeavor (Furness 1916), and for which recent data on call flexibility are very promising (Wich et al. 2009b; cf. Owren et al. 2010).

### *The study species*

*“On the whole I should say that the orang holds out more promise as a conversationalist than does the chimpanzee; it is more patient, less excitable, and seems to take instruction more kindly.”*

(William Furness, 1916)

Orangutans are the only Asian great apes. They show pronounced dimorphism in body size and secondary sexual characteristics and live in dispersed fission-fusion social communities (Delgado and van Schaik 2000). Today the two recognized species are found on the islands of Sumatra and Borneo (*Pongo abelli* and *P. pygmaeus*, respectively) and are considered critically endangered and endangered, respectively (IUCN 2010). However, in the Pleistocene (2.588 – 0.012 mya ago), Sumatra and Borneo were part of the Sunda landmass, and orangutans also ranged throughout Java and Southern Asia (Delgado and van Schaik 2000). They represent the oldest extant great ape lineage and differentiated from the homininae lineage approximately 9 – 13 mya ago (Hobolth et al. 2011). The orangutan call repertoire was first described comprehensively by MacKinnon (MacKinnon 1974) and Rijksen (Rijksen 1978). Several studies subsequently focused on the conspicuous long calls produced by flanged males as a coordination call within orangutan disperse communities (Delgado 2007; Lameira and Wich 2008; Mitra Setia and van Schaik 2007; Ross and Geissmann 2007; Spillmann et al. 2010). A revised description of the complete orangutan call repertoire recently showed that orangutans produce a larger array of different calls than described before (Hardus et al. 2009a). In addition, geographic variation of the call repertoires (Hardus et al. 2009a), there are strong indications that repertoires may not be similar between populations, suggesting a role for social learning (van Schaik et al. 2003; van Schaik et al. 2006). Although geographic variation in learned behavior, involving tool use and social habits, has been found in both chimpanzees (Whiten et al. 1999) and orangutans (van Schaik et al. 2003; van Schaik et al. 2006), only in

orangutans has cultural variation in calls been proposed. These observations in the wild have been recently substantiated by empirical work in captivity (Dindo et al. 2011), namely demonstrating that orangutans may spontaneously learn new calls that are characteristically human, i.e. whistling, which they voluntarily produce and over which they exert apt control particular features, such as number and duration (Wich et al. 2009b).

### *Aims and overview of this study*

The aim of this research is to extend and broaden the present understanding of speech evolution by investigating orangutan (*Pongo* spp) calls, while adhering to ethological principles but applying terminology borrowed from linguistics and semiotics. In particular, I investigate some of speech's building blocks (cf. Hockett and Altmann 1968) and their putative presence and nature in orangutans.

The first two chapters (2 and 3) derive from a comparative view over Hockett's feature of "traditional transmission", stating that individuals learn their call repertoire from their conspecifics within social groups. This leads to variation within and between groups. In **chapter 2**, I address the identification of different types of call variation between animal groups. I provide definitions for variation in call characteristics between populations of the same species, based on terminology for similar variation in human speech as defined in linguistics. Accents indicate deviations in the production of a particular call while maintaining its basic characteristics, whereas dialects imply the presence/absence of calls in one particular context between two populations. Based on these definitions of accents and dialects, I review the occurrence these two types of call variation in terrestrial mammals and the potential factors determining this variation (i.e. genetic, environmental and/or social factors). This provides a framework allowing a comparative approach to speech variation, together with a better understanding about the importance of social learning in terrestrial mammals.

In **chapter 3**, I specifically investigate the putative presence of dialects in orangutan calls emitted during nest building and by mothers towards their infants as a come-hither call. Variation in the repertoires of several orangutan populations is examined and the possible factors affecting this variation, namely the effects of ecological environment, genetic differentiation and social learning, are disentangled.

The following two chapters consider respectively Hockett's features of "prevarication" (chapter 4), the ability to make false statements, involving the manipulation of a given shared communicative feature in order to deceive other individuals, and "arbitrariness" (chapter 5), a characteristic of a communication system where there is no direct connection between the signal and its function. In **chapter 4**, I examine one of orangutans' alarm call, the kiss-squeaks and its variants – hand kiss-squeaks and kiss-squeaks on leaves. These variants are produced in some populations, but one or two of the modified variants are absent at other populations (Hardus et al. 2009a). Hand and leaf kiss-squeaks may thus represent cultural variants in orangutans (van Schaik et al. 2003; van Schaik et al. 2006; Hardus et al. 2009b), implying that several aspects of its use and production may be determined socially and based on learning. I examine the acoustic and functional differences between these three variants at one particular population, Tuanan (Central Kalimantan).

In **chapter 5**, I replicate the analyses of chapter 4 on kiss-squeaks and its variants in a second population (Cabang Panti, West Kalimantan) belonging to the same subspecies (*Pongo pygmaeus wurmbii*) and living in a partly similar habitat. This allows a direct comparison of the use and function of the three call variants between locations, while controlling for major genetic and environmental biases. When differences are found between sites, then different functions may be attributed by orangutans to the same signal.

**Chapter 6** provides understanding into Hockett's learnability, stating that users of one system of call communication may learn elements from another's system. I revisit a reference study in the discipline (Wich et al. 2009b), demonstrating that great orangutans may expand their innate repertoire by socially learning new calls from humans (i.e. whistling). This reference study offered one of the clearest pieces of evidence of apt flexibility in the domain of call acquisition in great apes. Moreover, it demonstrates that orangutans exert a great extent of voluntary motor control over lip musculature and musculature involved in air flow (i.e. diaphragm, and thoracic and abdominal muscles) during the production of a learned call. I assembled data available over orangutans known to whistle, replicated the previous study with one of the "new" whistling individuals, and I extended the study by comparing the acoustic properties between three whistling orangutans. Because whistling, despite tonal, does not involve vocal fold action, I discuss implications for the evolution of consonants in the human lineage.

**Chapter 7** offers a comparative view over Hockett's semanticity and patterning, respectively stating that some call units have specific meanings and that units may be

combined into larger meaningful units. I investigate whether orangutans are able of semantic and combinatorial acoustic communication. These features are fundamental for speech and are known to be present in primates (e.g. Seyfarth et al. 1980; Zuberbühler 2002), but remain virtually unassessed in great apes. I examine the presence of these features in orangutan alarm calls at two populations (where tigers are extinct or extant, respectively) by analyzing the combinatorial characteristics of their respective alarm calls. Firstly, this allows determining whether specific combinations are predator-specific. Secondly, by comparing populations, this allow determining whether predator guild influences call compositionality in great apes.

Based on the findings of the previous chapters and new data analyses, in **chapter 8**, I propose that some examples of orangutan call communication may have constituted precursors to speech. This evolutionary hypothesis proposes that the two major models of speech evolution, namely via call or gestural evolution, can be reconciled. Contrary to the majority of current hypotheses for speech evolution, the present hypothesis highlights specific behavioral traits which may have been targeted by evolutionary forces due to their proximate adaptive function and which subsequently acted as exaptations for subsequent stages of speech evolution.

**Review of geographic variation in terrestrial mammalian  
acoustic signals: Human speech variation  
in a comparative perspective**

Adriano R. Lameira, Roberto A. Delgado and Serge A. Wich

*Abstract*

Human speech shows an unparalleled richness in geographic variation. However, few attempts have been made to understand this linguistic diversity from an evolutionary and comparative framework. Here, we a) review extensively what is known about geographic variation of acoustic signals in terrestrial mammals, using common terminology adopted from linguistics to define different forms of variation (i.e. accents and dialects), and b) examine which factors may determine this variation (i.e. genetic, environmental and/or social). Heretofore, terminology has been used inconsistently within and across taxa, and geographic variation among terrestrial mammals has never been defined as in human speech. Our results show that accents, phonologically different varieties, occur widely in terrestrial mammals. Conversely, dialects, lexically and phonologically different varieties, have only been documented thus far in great white-lined bats, red deer, chimpanzees and orangutans. Although relatively rare among terrestrial mammals, dialects are thus not unique to humans. This finding also implies that such species possess the capacity for acoustic learning. Within primates, the two great apes showing dialects are those who also show extensive cultures in the wild, suggesting that, in hominoids, intricacy of acoustic geographic variation is potentially associated with cultural complexity; namely, both derive from selection increasingly favoring social learning across varied contexts, including the acoustic domain.

*Published in Journal of Evolutionary Psychology (2010), 8, 309-332*

## *Introduction*

Human acoustic communication shows an unparalleled richness in geographic variation, exemplified in our nearly 7000 extant languages (despite alarming rates of extinction) and a much larger and probably unknown number of dialects and accents (Crystal 2002). In linguistics, *accents* refer to the way in which a speaker (i.e. the acoustic signal's producer) pronounces the signal, and therefore refers to a variety which is only phonetically and/or phonologically different from other varieties (Chambers and Trudgill 1998); *dialects*, on the other hand, refer to varieties which are grammatically (and perhaps lexically) as well as phonologically different from other varieties (Chambers and Trudgill 1998). These forms of geographic variation are the result of human flexibility in innovating, modifying and socially learning acoustic signals (e.g. Cheney and Seyfarth 2005), but genetic (e.g. Cavalli-Sforza et al. 1988) and ecological factors (e.g. Whitfield 2008) are also known to be important in determining some of this variation.

After centuries of research, the evolutionary puzzle of human language (i.e. a cognitive system) and speech (i.e. a form of behavior for which that system is used) remains unsolved, namely due to apparent discontinuities with our closest relatives. For scientists with an interest in the evolution of human language and speech, crucial tasks include a) explaining acoustic geographic variation within an evolutionary framework, b) determining the discontinuities' qualitative and quantitative characteristics, and c) teasing apart perceived from factual discontinuities. Therefore, one important step is to examine whether geographic variation in acoustic communication occurs in other species. It is well known that many bird species show geographic variation in their songs that is similar to the forms of geographic variation in human speech (e.g. Podos and Warren 2007) but for evolutionary comparisons that are more relevant to human evolution, it is essential to examine geographic variation with the same rigor among mammalian species. Studies focusing on taxa less related to humans can reveal how communication systems may evolve analogous traits with language and speech, as well as guide comparative studies focusing on more closely related species. However, ultimately, the latter will be more relevant in interpreting and validating evolutionary scenarios within hominids, as factors such as phylogeny, (neuro)physiology, anatomy, body size, environment, diet, life-history, and social behavior become less confounding. Although there have been numerous studies on geographic variation in mammalian species, and some reviews cover this topic as it relates to aquatic mammals (e.g. Rendell and Whitehead 2001), there has not yet

been a review that examines the available information on terrestrial mammals and compares it in a systematic way by using a common set of definitions.

Here we adopt human linguistic terms and define geographic variation, accents and dialects in acoustic signals of terrestrial mammals (Table 1). It is useful to note that these definitions do not consider the potential factors determining the reported variation. This deviates from some previous studies on geographic variation of acoustic signals in that, traditionally, the definitions of dialect(s) and geographic variation are distinguished based on whether or not gene flow is restricted between populations (Conner 1982a). That is, dialects are defined as “song” differences between neighboring populations of potentially interbreeding individuals, and geographic variation as differences in “song” over long distances and between populations, which normally do not come together (defined for avian species; Nottebohm 1969). However, these traditional definitions pose at least three predicaments. First, these definitions are used inconsistently across taxa (e.g. Conner 1982a) hindering comparative analyses. Second, these definitions are applied across taxa to different types of acoustic signals; for instance, bird studies commonly applied the terms to report variation in songs (i.e. composed calls, see Table 1; explicit in the above definition by Nottebohm 1969) while mammalian studies commonly applied the terms to report variation in single calls (Conner 1982b; see Table 1). Thus, a consistent classification of the forms of acoustic signals is also necessary for informative comparative analyses (Table 1). Third, these definitions do not take account of different forms of acoustic variation. For example, a more recent study in chimpanzees (Mitani et al. 1992) noted that, while applying the traditional definition of dialects in primates and birds, population differences reported for each taxon were not comparable. According to standard definitions adopted from linguistics, the acoustic differences reported by Mitani and colleagues (1992), and in primates in general, are better classified as accents, while those commonly reported in some birds are better classified as dialects. A similar re-assessment as Mitani and colleagues (2009) was given by Elowson and Snowdon (1994). Although traditional definitions convey the notion that vocal learning may play a significant role in the emergence of geographic variation in animal acoustic signals (e.g. Conner 1982a), as it does in humans, it is only by recognizing separately the form of geographic variation of acoustic signals (i.e. accent or dialect) from their respective underlying factor(s) that one may understand the proximate and ultimate mechanisms by which geographic variation in human speech arises. Therefore, the aim of this paper is to provide an extensive review of geographic variation in terrestrial mammalian acoustic

signals, examining which factors determine each form of geographic variation in each species, applying the terms commonly used for geographic variation in human speech.

Table 1. *Glossary*

Term	Definition
Acoustic Signal (type)	The most <i>sensu lato</i> definition of a signal transmitted via pressure waves, irrespective of its acoustic morphology or complexity, produced by one or several structures of the signal producing system of the individual/species.
Call (type)	An acoustic signal consisting of a unit, with a determined morphology and complexity, which is biologically undividable, produced by one or several structures of the signal producing system of the individual/species.
Song/Coda/Composed call (type)	An acoustic signal consisting of a sequence of calls with particular composition and complexity.
Vocalization (type)	A call produced by the vocal folds, in a periodic phonation, biphonation or staccato phonation oscillation regime (see Brown et al., 2003) and thus, with (some) tonal characteristics.
Sound (type)	A call produced without the use of the vocal folds, and thus the product of other structures of the signal producing system of the individual/species (e.g. clicks, smacks, raspberries, whistles), or from vocal folds aperiodic or chaotic phonation (e.g. scream).
Type	A particular category of the referred form of acoustic signal.
Geographic variation	The most <i>sensu lato</i> definition of acoustic differences between groups or populations, usually within the same species (irrespective of underlying factors).
Accents	Geographic variation in the acoustic characteristics of a shared signal between groups or populations of the same species; hence, a difference in the way the same signal is uttered or pronounced (irrespective of the underlying factors). If the acoustic signal is a composed call in particular, differences in the acoustic characteristics of the whole composed call or of one or several of its components may occur.
Dialects	Geographic variation in the presence/absence of a particular acoustic signal and/or in the sequence/composition of calls within a song/coda/composed call between groups or populations of the same species; thus, a functionally analogous difference to lexica and/or grammar differences in humans (irrespective of underlying factors). If the acoustic signal is a composed call in

particular, the whole composed call, one or several of its components may be present/absent or substantially distinct.

Acoustic learning	Traditionally defined as vocal learning. The ability of learning socially the acoustic and/or production features of an acoustic signal.
Vocal learning	The ability of learning socially the acoustic and/or production features of a vocalization.
Sound learning	The ability of learning socially the acoustic and/or production features of a sound.

---

### *Methods*

Published literature in English on acoustic geographic variation in terrestrial mammals was searched using digital search engines (PubMed, Google Scholar and Web of Science) and in the references' sections of the publications found. The keywords used in the digital search engines were: variation *or* population *or* specific *or* dialect *or* accent, *and*, vocalization *or* call *or* sound *or* acoustic *or* signal. Solely geographic variation in the *acoustic* properties of signals was considered. For instance, geographic variation in the production rate of certain acoustic signals was not included. Abstracts or conference presentations were not considered. Studies demonstrating vocal recognition of animals of the same group/population vs. neighboring/stranger animals were not considered either (e.g. Biben and Symmens 1991), unless acoustic differences between the groups/populations were quantified.

In humans, the emergence and maintenance of accents and dialects allows for individuals to have differing acoustic repertoires with varying degrees of mutual intelligibility, whilst being genetically similar overall. Thus, we only considered geographic variation of acoustic signals that occurred *within species* (and subspecies) of terrestrial mammals because this variation will be the most relevant and informative for a comparison with humans and the evolution of speech. Bat studies focusing on cryptic species were not included in this review because these studies suggest that speciation has already occurred between populations (even though populations may still be referred to with the same species name).

For each study, author terminology of the described form of geographic variation and inferences concerning underlying factors were noted. We also classified the described form of geographic variation according to the definitions used in this paper (i.e. Table 1).

Information about the spatial distribution (i.e. adjacent/potentially interbreeding or non-adjacent/non-interbreeding) of the study populations was that of the authors, except in Eiler and Banack (2004), for which we considered the distribution of the populations to be non-adjacent instead of adjacent (since two highways and a lake probably function as dispersal barriers to ground squirrels). When a particular study described several occurrences of acoustic geographic variation, these were mentioned as different reports. Whenever a particular report of geographic variation suggested two possible underlying factors, we considered half count for each factor for statistical purposes (i.e. Table 3).

### *Results*

We identified 82 studies describing 142 occurrences (hereafter reports) of acoustic geographic variation (either present or absent) in 84 species of terrestrial mammals (Table 2 and Table 3).

### *Terminology*

Author terminology on the forms of geographic variation varied. However, no study defined the variation described as accent variation, and no study defined dialects as defined in this review. Other studies used terms specifically defined in the respective publication (21 studies, or 25.6 %), using altogether 9 different terms in this fashion. A considerable number of studies did not define with any particular term the described variation (23 studies, or 28.1 %). Several studies used the traditional definitions (38 studies, or 46.3 %). However, these terms were not used consistently across species or taxa relative to the underlying factors, the form of geographic variation, or the spatial distribution of the populations. For instance, some studies used the term “dialect” based on the adjacent distribution of the populations showing acoustic variation (e.g. Eiler and Banack 2004), others on the fact that variation was determined by social factors (e.g. Esser and Schubert 1998), others still because the variation was described within a subspecies (e.g. de la Torre and Snowdon 2009), and others seemingly just as a general term to define acoustic variation between populations, even when the proposed underlying factors were solely genetic (e.g. Nevo et al. 1987) or environmental (e.g. Perla and Slobodchikoff 2002).

### *Forms of Geographic Variation and Putative Determining Factors*

The taxon with most studies of variation was primates, followed by bats, lagomorphs and rodents (Table 3). However, the taxon with most reports of variation and species showing geographic variation was bats, followed by primates, lagomorphs and rodents (Table 3). There were 15 (10.6 %) reports in which geographic variation was absent, translating to 9 (10.7 %) species, but studies on these species were not replicated. Specifically, all studies reporting absence of variation concerned accents. Geographic variation described concerned mostly accent variation (132 reports, i.e. 93.0 %), with fewer accounts of dialect variation (10 reports, i.e. 7.0 %). Dialects were described in the great white-lined, or sac-winged, bat, the red deer, the chimpanzee and the orangutan.

Among lagomorph and rodent studies that provided suggestions on the underlying factors of geographic variation, genetic factors were the most relevant in explaining the emergence of accents in these taxa (Table 3). Each factor was equally relevant in explaining the emergence of accents in bats (Table 3), and in primates, social factors were the most important (Table 3). Only the great white-lined, or sac-winged, bat showed accent variation in a composed call. All studies that reported dialect variation and provided suggestions for the underlying cause, proposed social factors as the primary determinant (Table 3). Among primates, only 14 (i.e. 40.0%) of the reviewed publications focused on apes. Only the great white-lined, or sac-winged, bat showed dialect variation in a composed call.

## *Discussion*

### *Terminology*

The results indicate that terminology defining geographic variation is used inconsistently across species and taxa of terrestrial mammals, which may lead to the incorrect interpretation of the variation described and makes comparisons complicated. The traditional terms for acoustic geographic variation and dialects were first defined and used in a standardized fashion among avian researchers several decades ago (e.g. Nottebohm 1969). However, the adoption of these terms in terrestrial mammals has been done in an inconsistent way, and their understanding requires a thorough reading of the studies. For instance, definitions by the same authors may differ between publications (e.g. Slobodchikoff and Coast 1980; and Slobodchikoff et al. 1998). On the other hand, because the traditional definitions are based on genetic factors, this seems to have

Table 2. Literature review

Species (common name, <i>scientific name</i> )	Variation type by authors <sup>a</sup>	Variation type <sup>b</sup>	Spatial distribution <sup>c</sup>	Underlying factor type <sup>d</sup>	Reference study <sup>e</sup>
<b>Lagomorphs and Rodents</b>					
Pearson's Tuco-Tuco, <i>Ctenomys pearsoni</i>	D	A	NA	Gen	(Francescoli 2002)
Gunnison's Prairie Dog, <i>Cynomys gunnisoni</i>	D	A	?	Env	(Perla and Slobodchikoff 2002)
	G	A	NA	Gen	(Slobodchikoff et al. 1998)
	G	-	A	-	(Slobodchikoff et al. 1998)
	D	A	B	Env and/or Soc	(Slobodchikoff et al. 1998)
Zuni Gunnison's Prairie Dog, <i>Cynomys gunnisoni zuniensis</i>	D	A	NA	Soc	(Slobodchikoff and Coast 1980) (syllable length and interval length between syllables)
	D	A	NA	Env	(Slobodchikoff and Coast 1980) (call length and number of syllables)
Merriam Chipmunk, <i>Eutamias merriami</i>	G	A	NA	?	(Brand 1976)
Redwood Chipmunk, <i>Eutamias ochrogeny</i>	-	-	NA	-	(Gannon and Lawlor 1989)
Allen's Chipmunk, <i>Eutamias senex</i>	-	-	NA	-	(Gannon and Lawlor 1989)
Siskiyou Chipmunk, <i>Eutamias siskiyou</i>	-	-	NA	-	(Gannon and Lawlor 1989)
Sonoma Chipmunk, <i>Eutamias sonomae</i>	G	A	NA	?	(Brand 1976)
Lodgepole Chipmunk, <i>Eutamias speciosus</i>	G	A	NA	?	(Brand 1976)
Townsend Chipmunk, <i>Eutamias townsendii</i>	G	A	NA	?	(Brand 1976)

	-	-	NA	-	(Gannon and Lawlor 1989)
Long-Tailed Marmot, <i>Marmota caudata</i>	G	A	NA	Gen	(Nicol'skii et al. 1999)
Pika, <i>Ochotona princeps</i>	G	A	B	Gen	(Conner 1982b)
	D	A	NA	Gen	(Somers 1973)
Northern Grasshopper Mouse, <i>Onychomys leucogaster</i>	G	-	NA	-	(Hafner and Hafner 1979)
Southern Grasshopper Mouse, <i>Onychomys torridus</i>	G	-	NA	-	(Hafner and Hafner 1979)
Great Gerbil, <i>Rhombomys opimus</i>	GD	A	A	?	(Randall et al. 2005)
Mole Rat, <i>Spalax ehrenbergi</i>	D	A	NA	Gen	(Nevo et al. 1987)
Belding's ground squirrel, <i>Spermophilus beldingi</i>	-	A	NA	Gen	(McCowan and Hooper 2002)
Golden-Mantled Ground Squirrel, <i>Spermophilus lateralis</i>	G	A	NA	?	(Eiler and Banack 2004)
	G	A	NA	?	(Eiler and Banack 2004) (between <i>S. l. bernardinus</i> and <i>S. l. trepidus</i> )
	D	A	A	?	(Eiler and Banack 2004) (between <i>S. l. chrysodeirus</i> and <i>S. l. trepidus</i> )
Golden-Mantled Ground Squirrel, <i>Spermophilus lateralis bernardinus</i>	D	A	A	Soc	(Eiler and Banack 2004)
Golden-Mantled Ground Squirrel, <i>Spermophilus lateralis chrysodeirus</i>	D	A	A	Soc	(Eiler and Banack 2004)

#### Bats

Pallid Bat, <i>Antrozous pallidus</i>	G	A	NA	?	(Thomas et al. 1987)
Little Free-Tailed Bat, <i>Chaerephon pumilus</i>	G	A	NA	Gen and/or Env	(Aspetsberger et al. 2003)
Long-Tailed Bat, <i>Chalinolobus tuberculatus</i>	G	A	NA	?	(Parsons 1997) (between regions)
	G	-	NA	-	(Parsons 1997) (between populations)
Big Brown Bat, <i>Eptesicus fuscus</i>	-	A	NA	?	(Fenton and Bell 1981)
	-	A	NA	Gen and/or Soc	(Masters et al. 1995)
	G	A	NA	?	(Murray et al. 2001)
	BC	-	NA	-	Rasmuson and Barclay 1992
	G	A	NA	?	(Thomas et al. 1987)
Northern Bat, <i>Eptesicus nilssonii</i>	-	A	NA	Env	(Rydell 1993)
Bicoloured Leaf-Nosed Bat, <i>Hipposideros bicolor</i>	BC	A	NA	?	(Francis and Habersetzer 1998)
Fawn Leaf-Nosed Bat, <i>Hipposideros cervinus</i>	BC	A	NA	?	(Francis and Habersetzer 1998)
	-	A	NA	?	(Heller and Helversen 1989)
Cantor's Roundleaf Bat, <i>Hipposideros galeritus galeritus/insolens</i>	BC	A	NA	Gen	(Francis and Habersetzer 1998)
Noack's Roundleaf Bat, <i>Hipposideros ruber</i>	G	A	B	?	(Guillén et al. 2000)
Silver-Haired Bat, <i>Lasionycteris noctivagans</i>	G	A	NA	?	(Thomas et al. 1987)
Eastern Red Bat, <i>Lasiurus borealis</i>	G	A	NA	?	(Murray et al. 2001)
	G	A	NA	?	(Thomas et al. 1987)

Hoary Bat, <i>Lasiurus cinereus</i>	-	A	NA	Gen	(Barclay et al. 1999) (between Hawai'i and North America)
	-	A	B	Env	(Barclay et al. 1999) (within and between Hawaiian islands)
	G	A	NA	?	(O'Farrell et al. 2000)
	G	A	NA	?	(Thomas et al. 1987)
California Myotis, <i>Myotis californicus</i>	-	A	NA	?	(Fenton and Bell 1981)
	G	A	NA	?	(Thomas et al. 1987)
Western Small-Footed Myotis, <i>Myotis ciliolabrum</i>	G	A	NA	?	(Thomas et al. 1987)
Western Long-Eared Myotis, <i>Myotis evotis</i>	G	A	NA	?	(Thomas et al. 1987)
Gray Bat, <i>Myotis grisescens</i>	G	A	NA	?	(Murray et al. 2001)
Little Brown Bat, <i>Myotis lucifugus</i>	SP	A	A	Soc	(Pearl and Fenton 1996)
	G	A	NA	?	(Murray et al. 2001)
	G	A	NA	?	(Thomas et al. 1987)
Northern Long-Eared Myotis, <i>Myotis septentrionalis</i>	G	A	NA	?	(Murray et al. 2001)
	G	A	NA	?	(Thomas et al. 1987)
Indiana Bat, <i>Myotis sodalis</i>	G	A	NA	?	(Murray et al. 2001)
Long-Legged Myotis, <i>Myotis volans</i>	G	A	NA	?	(Thomas et al. 1987)
Short-Tailed Bat, <i>Mystacina tuberculata</i>	G	A	NA	?	(Parsons 1997)
Evening Bat, <i>Nycticeius humeralis</i>	BC	A	A	Env and/or Soc	(Scherrer and Wilkinson 1993)

Lesser Spear-Nosed Bat, <i>Phyllostomus discolor</i>	-	A	-	Soc	(Esser 1994)
	D	A	NA	Soc	(Esser and Schubert 1998)
Greater Spear-Nosed Bat, <i>Phyllostomus hastatus</i>	GD	A	-	Soc	(Boughman 1997)
	GD	A	-	Soc	(Boughman 1998)
	G	A	NA	Gen and/or Soc	(Boughman and Wilkinson 1998)
Eastern Pipistrelle Bat, <i>Pipistrellus subflavus</i>	G	A	NA	?	(Murray et al. 2001)
Townsend's Big-Eared Bat, <i>Plecotus townsendii</i>	G	A	NA	?	(Thomas et al. 1987)
Orange Leaf-Nosed Bat, <i>Rhinonictis aurantia</i>	SP	A	NA	Gen	(Armstrong and Coles 2007)
Intermediate Horseshoe Bat, <i>Rhinolophus affinis</i>	BC	A	NA	?	(Francis and Habersetzer 1998)
Blasius's Horseshoe Bat, <i>Rhinolophus blassi</i>	-	A	NA	Gen and/or Env	(Heller and Helversen 1989)
Little Japanese Horseshoe Bat, <i>Rhinolophus cornutus pumilus</i>	G	A	B	Gen and/or Env	(Yoshino et al. 2006)
	BC	A	B	Soc	(Yoshino et al. 2008)
Mediterranean Horseshoe Bat, <i>Rhinolophus euryale</i>	-	A	NA	Gen and/or Env	(Heller and Helversen 1989)
	-	A	NA	Env	(Russo et al. 2007)
Large-Eared Horseshoe Bat, <i>Rhinolophus philippinensis</i>	-	A	A	?	(Kingston and Rossiter 2004)
Greater Horseshoe Bat, <i>Rhinolophus ferrumequinum</i>	-	A	NA	Gen and/or Env	(Heller and Helversen 1989)

	-	A	NA	Soc		(Jones and Ransome 1993)
Greater Horseshoe Bat, <i>Rhinolophus ferrumequinum ferrumequinum/nippon</i>	-	A	NA	Gen and/or Env		(Taniguchi 1985)
Lesser Horseshoe Bat, <i>Rhinolophus hipposideros</i>	-	A	NA	Gen and/or Env		(Heller and Helversen 1989)
	-	A	NA	Env		(Russo et al. 2007)
Mehely's Horseshoe Bat, <i>Rhinolophus mehelyi</i>	-	A	NA	Gen and/or Env		(Heller and Helversen 1989)
Formosan Lesser Horseshoe Bat, <i>Rhinolophus monoceros</i>	G	A	B	Gen and/or Soc		(Chen et al. 2009)
Greater White-Lined/Sac-Winged Bat, <i>Saccopteryx bilineata</i>	G	D	B	?		(Davidson and Wilkinson 2002)
	G	-	B	-		(Davidson and Wilkinson 2002) (screech)
	G	A	NA	Gen		(Davidson and Wilkinson 2002) (inverted-V)
	D	A	A	Soc		(Davidson and Wilkinson 2002) (inverted-V)
Brazilian Free-Tailed Bat, <i>Tadarida brasiliensis</i>	G	A	NA	?		(Gillam and McCracken 2007)
Large Forest Bat, <i>Vespadelus darlingtoni</i>	G	A	NA	Gen and/or Env		(Law et al. 2002)
Eastern Forest Bat, <i>Vespadelus pumilus</i>	G	A	NA	Gen and/or Env		(Law et al. 2002)
Southern Forest Bat, <i>Vespadelus regulus</i>	G	A	NA	Gen and/or Env		(Law et al. 2002)
Eastern Cave Bat, <i>Vespadelus troungtoni</i>	G	-	NA	-		(Law et al. 2002)

Little Forest Bat, <i>Vespadelus vulturnus</i>	G	A	NA	Gen and/or Env	(Law et al. 2002)
<b>Other mammalian orders</b>					
Wolf, <i>Canis lupus</i>	G	-	NA	-	(Palacios et al. 2007)
Corsican/European/Scottish Red Deer, <i>Cervus elaphus corsicanus/hippelaphus/scoticus</i>	-	A	NA	?	(Kidjo et al. 2008)
	-	D	NA	?	(Kidjo et al. 2008) (harsh roar)
<b>Primates</b>					
Brown/Northern Brown Howling Monkey, <i>Alouatta fusca clamitans/fusca</i>	-	A	NA	Gen	(Whitehead 1995)
Mantled Howling Monkey, <i>Alouatta palliata</i>	-	A	NA	?	(Whitehead 1995)
Pygmy Marmoset, <i>Cebuella pygmaea</i>	-	A	-	Soc	(Elowson and Snowdon 1994)
Western Pygmy Marmoset, <i>Cebuella pygmaea pygmaea</i>	D	A	NA	?	(de la Torre and Snowdon 2009)
Titi Monkey, <i>Celicebus moloch</i>	-	A	A	?	(Robinson 1979)
Vervet Monkey, <i>Cercopithecus aethiops</i>	-	-	NA	-	(Struhsaker 1970)
Mohol Bushbaby, <i>Galago moholi</i>	SP	-	NA	-	(Anderson et al. 2000)
Agile Gibbon, <i>Hylobates agilis</i>	SP	A	NA	?	(Mitani 1987)
Japanese Macaque, <i>Macaca fuscata</i>	D	A	NA	Gen and/or Soc	(Green 1975)
	D	A	NA	Soc	(Sugiura et al. 2006)
	GD	A	NA	Env and/or	(Tanaka et al. 2006)

				Soc	
Rhesus Macaque, <i>Macaca mulatta</i>	-	A	A	Gen and/or Soc	(Hauser 1992)
Barbary Macaque, <i>Macaca sylvanus</i>	L	A	NA	Gen and/or Soc	(Fischer et al. 1998)
Grey Mouse Lemur, <i>Microcebus murinus</i>	D	A	A	Soc	(Hafen et al. 1998)
	BC	A	NA	Gen and/or Soc	(Zimmermann and Hafen 2001)
Chimpanzee, <i>Pan troglodytes</i>	-	D	NA	Soc	(Hopkins et al. 2007)
East African/Long-Haired Chimpanzee, <i>Pan troglodytes schweinfurthii</i>	BC	D	NA	Soc	(Arcadi 1996)
	BC	A	NA	?	(Mitani and Brandt 1994)
	G	A	NA	Gen and/or Env	(Mitani et al. 1999)
	D	A	NA	Gen and/or Soc	(Mitani et al. 1992)
West African Chimpanzee, <i>Pan troglodytes verus</i>	SP	A	A	Soc	(Crockford et al. 2004) (within Taï National Park)
	SP	-	NA	-	(Crockford et al. 2004) (between Guiroutou and Taï National Park)
	GD	A	NA	Soc	(Marshall et al. 1999)
	GD	D	NA	Soc	(Marshall et al. 1999) (Bronx Cheer)
Olive Baboon, <i>Papio (hamadryas) anubis</i>	-	A	NA	Env	(Ey et al. 2009)

Sumatran Orangutan, <i>Pongo abellii</i>			-	A	NA	Gen and/or Soc	(Ross and Geissmann 2007)
			G	A	NA	?	(Delgado 2007)
			G	A	NA	?	(Delgado et al. 2009)
			G	D	NA	Soc	(van Schaik et al. 2003)
			-	D	NA	Soc	(van Schaik et al. 2006)
Bornean Orangutan, <i>Pongo pygmaeus</i>			-	A	NA	Gen and/or Soc	(Ross and Geissmann 2007)
			G	A	NA	?	(Delgado 2007)
			G	A	NA	?	(Delgado et al. 2009)
			G	D	NA	Soc	(Hardus et al. 2009a)
			G	D	NA	Soc	(van Schaik et al. 2003)
			-	D	NA	Soc	(van Schaik et al. 2006)
North Indian/Himalayan	Langur, <i>Presbytis entellus entellus/schistaceus</i>		-	A	NA	Gen and/or Env	(Vogel 1973)
Nilgiri	Langur, <i>Presbytis johnii</i>		SP	A	NA	Gen	(Hohmann and Vogl 1991)
Thomas	Langur, <i>Presbytis thomasi</i>		G	A	NA	Gen	(Wich et al. 2008)
Spix's/Illiger's/Red-Mantle/Geoffroy's Back Tamarin, <i>Saguinus fuscicollis fuscicollis/illigeri/lagonotus/nigrifrons</i>	Saddle-		SP	A	B	Gen and/or Soc	(Hodun et al. 1981)
Red-Chested Moustached Tamarin, <i>Saguinus labiatus labiatus</i>			D	A	A	Soc	(Maeda and Masataka 1987)

	D	A	A	Soc	(Masataka 1988)
Grey-/Black-Crowned Squirrel Monkey, <i>Saimiri oerstedii citrinellus/oerstedii</i>	SP	A	NA	Gen	(Boinski and Newman 1988)

Notes: <sup>a</sup> (BC) Between/inter- colony/population variation; (D) Dialect variation; (G) Geographic variation; (GD) Group differences/- distinctive calls; (SP) Intra-/Sub-/Population- specific variation; (L) Local variation; (-) Not applicable or type of variation not mentioned by author(s).

<sup>b</sup> (A) Accent geographic variation; (D) Dialect geographic variation; (-) Accent variation absent

<sup>c</sup> (NA) Non-adjacent, non-interbreeding populations or groups; (N) Adjacent, potentially interbreeding populations or groups; (B) Both non-adjacent and adjacent populations or groups; (-) Not applicable.

<sup>d</sup> (Gen) Geographic variation influenced by genetic factors; (Env) Geographic variation influenced by environmental factors; (Soc) Geographic variation influenced by social factors; (?) No suggestions given about influencing factors, or geographic variation influenced by genetic and/or environmental and/or social factors; (-) Accent variation absent. Factors presented are those indicated or essentially implied by the author(s).

<sup>e</sup> Comments may be given between brackets to facilitate interpretation.

prompted several studies to avoid the use of these terms when the authors were interested in determining other factor(s) underlying the described variation (e.g. without an a priori notion of the relevance of genetic factors, e.g. Yoshino et al. 2008). Nonetheless, when the traditional definitions were not used, the terms used instead also produce some difficulties. For instance, none of the alternative terms take consideration of different forms of acoustic variation. Particular terms such as population- or colony-specific calls used to refer to (accent) variation may in fact be misleading, seemingly describing calls that are only present in one group (and thus constituting a dialect) contrary to variation in a shared call between groups (e.g. Boughman and Wilkinson 1998; Crockford et al. 2004). Moreover, the use of the term intra-specific variation may also be ambiguous, since variation between sexes and age classes or individuals can also be referred to in this fashion.

This pattern clearly indicates that the study of this topic in terrestrial mammals has brought about a menagerie of terms (or lack of thereof) that hamper comparative analyses. Instead, consistent operational definitions for characterizing the forms of acoustic differences within a species are needed to move the field forward and to provide a clear evolutionary framework for the observed variation. We propose that scientists interested in understanding the evolution of geographic variation in the acoustic signals of terrestrial mammals (including humans) and other taxa should aim to use a terminology that describes different forms of variation independently from the putative factors that underlie their evolution. Here we presented a recommended terminology to be adopted in this field, adapted directly from linguistics, which *ipso facto*, is relevant to evolutionary and comparative studies of human speech. Although our review only included terrestrial mammalian literature, researchers of any animal taxon, such as aquatic mammals or birds, could apply this terminology. In fact, some studies in aquatic mammals have already applied the terms of dialects and accents (e.g. Rendell and Whitehead 2005) but not specifically as defined here. Using a consistent terminology would directly benefit the disciplines of animal acoustics and comparative studies of speech evolution.

At the same time, our findings show that the use of consistent (linguistic) terms across taxa disentangles some perceived discontinuities from factual discontinuities between animal acoustic repertoires and human language and speech, allowing us to better determine whether the nature of factual discontinuities are quantitative or

qualitative. Ultimately, this should also benefit language and speech researchers by offering an integrative basis for general discussions on the evolution of these traits.

#### *Accent Variation and Putative Underlying Factors*

According to our definitions, accents in the acoustic signals of terrestrial mammals were described in numerous species and taxa. Perhaps surprisingly, felids, bears and African large mammals were a few examples of taxa on which acoustic variation remains to be examined. However, we do not exclude the possibility that such variation is present (either in accent or dialects), especially because such species occupy frequently relatively large territories and may range through diverse habitats with different selective pressures, and some have even been shown to be capable of acoustic learning (e.g. elephants, Poole et al. 2005). We did not attempt to formally examine shared traits between species exhibiting accents as only very few published studies reported an absence of such variation between populations (and this could well be the result of a bias against publication of negative results).

An accent may derive from genetic factors when an acoustic signal is linked (directly or via inherited traits such as body size and morphology) to a portion of the gene pool restricted by a barrier between populations, or when its function is selectively strong enough to shape the gene pool of the populations (e.g. a function strongly influenced by sexual selection). An accent may also be determined by environmental factors when a signal's function plays over long distances and thus it is shaped strongly by different habitat acoustics for optimized signal propagation. An accent may also emerge from social factors when individuals have specific fine motor control and flexibility over a signal, and use information relative to its acoustic/production features from different conspecifics in different populations. The reviewed literature, however, suggests that genetic, environmental, and social factors interact differently across taxa in driving the emergence of accents. Altogether, social factors play a larger role than conventionally assumed across all species, being (at least) as important as environmental factors in lagomorphs and rodents, equally important to genetic and social factors in bats, and considerably more important than genetic factors in primates. The percentage of reports suggesting that accent variation is determined by social factors is highest among primates. That habitat characteristics play a minor role in shaping acoustic signals in this taxon has been suggested previously (Schneider et al. 2008). Thus, the learning flexibility necessary to

acquire slight acoustic and/or production features from social partners is possibly a trait that is spread widely among terrestrial mammals showing acoustic geographic variation, and is clearly expressed in humans' close relatives. According to Janik and Slater (1997), this flexibility does not include necessarily acoustic learning (traditionally defined as vocal learning, Table 1) if observed variation relates to the tempo and/or amplitude of signal characteristics.

Table 3. *Summary statistics*

	Number of studies	Number of reports	Number of species	Accent driven by Gen/Env/Soc % <sup>a</sup>	Dialect driven by Gen/Env/Soc % <sup>a</sup>
Lag & Rod <sup>b</sup>	14 (17.1 %)	28 (19.7 %)	17 (20.2 %)	53.9/19.2/26.9 %	-
Bats	31 (37.8 %)	68 (47.9 %)	44 (52.4 %)	35.9/32.8/31.3 %	0/0/100%
Other taxa	2 (2.4 %)	3 (2.1 %)	2 (2.4 %)	-	-
Primates	35 (42.7 %)	43 (30.3 %)	21 (25.0 %)	29.0/12.8/62.9 %	0/0/100%
TOTAL	82 (100%)	137 (100 %)	84 (100%)		

Notes: <sup>a</sup> Proportion of reports with suggestions relatively to each underlying factor.

<sup>b</sup> Lagomorphs and rodents.

#### *Dialect Variation and Putative Underlying Factors*

The proportion of studies describing accents and dialects in the reviewed literature is likely to reflect the natural rarity of the dialects among terrestrial mammals, though biases in research effort may also account for some reported differences. Although this constraint exists, it is still reasonable to assume that dialects are a derived trait in some taxa.

How can different factors bring about the emergence of dialects *within* species? The answers remain challenging with respect to genetic and environmental factors. Environmental factors, for instance, exert selective pressures on a signal towards optimal propagation (e.g. Wiley and Richards 1978) by favoring acoustic properties that are functionally most efficient within their given habitats. However, it is difficult to conceive how such selective pressures could determine the emergence of dialects as it would imply a change in signal function, and in natural circumstances, that different populations held already different habitat preferences.

In relation to genetic factors, to the best of our knowledge, the specific mechanisms by which these could generate dialects within species remain unknown. We presume that mechanisms similar to those generating accents would imply speciation by the time dialects would emerge. Nevertheless, hypothetically, a genetic bottleneck caused by geographic isolation or a mutation could generate multiple changes in the neuro-motor pathways that have influence over the usage of an individual's oro-laryngeal structure, causing the elimination, addition or substantial modification of an acoustic signal. Currently we do not know of studies indicating that genetic phenomena such as bottlenecks and mutations can affect the structure of an acoustic signal in such a way.

On the other hand, social factors imply that species with dialects are capable of acoustic learning (Table 1), since contextual learning alone cannot obliterate, create or generate signals that are substantially distinct in their acoustic morphology from an original morphology (see above). Several studies on aquatic mammals provide evidence that geographic variation in acoustic signals that we would define as dialects may be indeed influenced by acoustic learning (e.g. Morrice et al. 1994b; Noad et al. 2000; Weilgart and Whitehead 1997; Whitehead et al. 1998). Therefore, we infer that dialects can also develop in terrestrial mammals from social factors via acoustic learning; and they can arise (1) rapidly from an innovation subsequently followed by social learning in one population, or (2) gradually from accents (also socially determined *per se*) when the eventual accumulation of transmission "errors" gives rise to substantially distinct signals between populations. According to this seemingly more parsimonious view, dialects were suggested to be determined by social factors in the reviewed literature.

One of the four species showing dialects was the greater white-lined, or sac-winged, bat (Table 2; Davidson and Wilkinson 2002). Although, to the best of our knowledge, there is currently no study that shows acoustic learning in this bat species, this is not an uncommon ability in the taxon (e.g. Boughman 1998). Other remarkable abilities concerning fine motor control over the production of acoustic signals have been shown in this taxon as well, such as harmonic hopping (Kingston and Rossiter 2004). Thus, that dialects have emerged in bats, is perhaps not surprising. But why did dialects emerge solely in the great white-lined bat in particular among other bats? It seems this may have been an artifact of the focus of researchers. Davidson and Wilkinson (2002) were among the few researchers that examined geographic variation in bat acoustic signals other than in echolocation. Because echolocation calls are functionally and ecological strongly

constrained, dialects may remain uncovered in other signals of bat acoustic repertoire until future studies follow a similar approach to Davidson and Wilkinson (2002). Based on the published descriptions of great white-lined bat's dialects (Davidson and Wilkinson 2002) it is uncertain which of the hypothetical evolutionary scenarios described above, relative to social factors, could have resulted in its emergence.

Another species showing dialects was the red deer (Table 2; Kidjo et al. 2008). This was the only study among those reporting the presence of dialects in terrestrial mammals which did not present any suggestions on the potential underlying factors. Because the studied populations of Corsican red deer (showing absence of one call type) derived from very small groups from other insular populations, it is assumed that either genetic or social factors could have determined the emergence of this dialect (see above; e.g. isolation of a pleiotropic gene or absence of relevant tutors, respectively). An interesting aspect of this finding is that the specific call type which is absent in Corsican red deer is related to the use of a descended and mobile larynx during production, an anatomical innovation found in several deer species, but absent in most mammals (Kidjo et al. 2008). Over the last 30 years, the acoustic behavior of red deer has arguably received more attention than that of any other large non-primate terrestrial mammal (Kidjo et al. 2008), and so investigations of geographic variation of acoustic signals in other large non-primate terrestrial mammals is warranted as this suggests that positive results may correlate with research effort.

In addition to the one bat species and one deer species, orangutans and chimpanzees were the only primate species where dialects were described (Table 2 and Table 3), but there was no apparent bias in research effort towards apes within this major taxon. Since environmental and genetic factors are unlikely to determine dialects in these species (see above, cf. Chapter 4; Marshall et al. 1999; van Schaik et al. 2003; van Schaik et al. 2006) and each report has suggested social factors as the main underlying cause, this implies that orangutans and chimpanzees are capable of acoustic learning. Based on the conventional notion that only humans are capable of acoustic learning among primates, it is important to evaluate this assertion carefully. Unequivocal evidence for acoustic learning would be given if animals demonstrate the ability to copy signals (with or without training) that are very different from signals in their natural repertoire (Janik and Slater 1997). The spontaneous acquisition (i.e. without training) of human whistling by an individual orangutan has recently been described (Wich et al. 2009b), and it was verified

that this signal differed significantly in acoustic characteristics from any signal in the repertoire of wild orangutans (comprising more than 30 different signal types, Hardus et al. 2009a). Hence, the acoustic learning hypothesis deserves further study in apes. Based on the published descriptions of great ape dialects, one may speculate that the reported variation was generated by social factors according to the first hypothetical evolutionary scenario described above (i.e. innovation followed by social learning).

Interestingly, extensive primate cultures have been proposed in the species that showed socially determined dialects: orangutans (van Schaik et al. 2003; van Schaik et al. 2006) and chimpanzees (Whiten et al. 1999). Although several primate species show behavioral innovations and rely on social learning to acquire particular behaviors (e.g. Japanese macaques; Kawai 1965), that constitute the basic mechanisms of culture, apes innovate and learn from others more frequently than other primates (after controlling for phylogeny and research effort; Reader and Laland 2002). In apes, cultures entail multiple behavioral variants that include several domains of an individual's life, including subsistence and comfort skills as well as locomotor patterns and communication (e.g. van Schaik et al. 2003; van Schaik et al. 2006; Whiten et al. 1999). Hence, it is reasonable to assume that this pattern of behavioral flexibility in apes reflects a selection pressure that favors a general propensity to use social information across several contexts and functions. If this view is correct, then it may suggest that *among primates* the emergence of multifaceted cultures in hominoids may include, as a natural corollary, some reasonable level of voluntary control over oro-laryngeal structures that qualify as acoustic learning. Alas, the published literature focusing on other taxa of terrestrial mammals (i.e. non-primates) regarding cultural behaviors and, simultaneously, on geographic variation of acoustic signals does not yet allow development of a similar hypothesis for these groups.

### *Geographic Variation in Speech*

The adoption of terminology from linguistics for the varying forms of geographic variation in acoustic signals, together with the review of published literature describing this variation in terrestrial mammals, shows that variation analogous to human accents is present in multiple species. On the other hand, variation analogous to human dialects represents a relatively rare derived trait in terrestrial mammals that is, to some extent, dependent on the ability for acoustic learning. In primates, namely the great apes, the emergence of dialects seems to be related to the evolution of multifaceted cultures (see

above). This positive relationship between increasingly complex cultures (and brain size; see Reader and Laland 2002) and intricacy (i.e. the amount of variant forms) of geographic variation in acoustic signals concords with the fact that humans, who show cultures unmatched in their complexity, are also the ones showing an unparalleled richness in geographic variation in the acoustic repertoire (i.e. speech). Accordingly, geographic languages (i.e. dialects where lexical divergence substantially hampers understanding between varieties; Crystal 2002) possibly emerged in early humans with the evolution of ever-more complex cultures. In spite of this, geographic variation in human speech constitutes only a quantitative discontinuity with our closest relatives. Future studies on the geographic variation of acoustic signals in more primate species (e.g. capuchin monkeys; Perry et al. 2003) and other mammals (e.g. dolphins; Krützen et al. 2005) that show culture will provide insights about whether this connection is valid across mammals or only in those taxa most closely related to humans.

#### *Acknowledgements*

Financial support for this doctoral project was generously provided by: Fundação para a Ciência e Tecnologia (Portugal), Wenner-Gren Foundation for Anthropological Research (USA), Stichting Dobberke (the Netherlands), Lucie Burgers Foundation for Comparative Behaviour Research (the Netherlands), Schure-Beijerinck-Popping Fonds (the Netherlands), Ruggles-Gates Fund for Anthropological Scholarship of the Royal Anthropological Institute of Great Britain and Ireland, Primate Conservation, Inc. (USA) and Fundação Calouste Gulbenkian (Portugal). Special thanks to the librarians of the University of Southern California, who provided access to numerous publications used in this review. AL would like to acknowledge Madeleine E. Hardus for her companionship and support.

Chapter III  
**Call cultures in orangutans?**

Serge A. Wich, Michael Krützen\*, Adriano R. Lameira\*, Alexander Nater, Natasha Arora, Meredith L. Bastian, Elen Meulman, Helen C. Morrogh-Bernard, Sri Suci Utami Atmoko, Joko Pamungkas, Dyah Perwitasari-Farajallah, Madeleine E. Hardus, Maria van Noordwijk and Carel P. van Schaik

*Abstract*

Several studies suggested great ape cultures, arguing that human cumulative culture presumably evolved from such a foundation. These focused on conspicuous behaviours, and showed rich geographic variation, which could not be attributed to known ecological or genetic differences. Although geographic variation within call types (accents) has previously been reported for orangutans and other primate species, we examine geographic variation in the presence/absence of discrete call types (dialects). Because orangutans have been shown to have geographic variation that is not completely explicable by genetic or ecological factors we hypothesized that this will be similar in the call domain and predict that discrete call type variation between populations will be found. We examined long-term behavioural data from five orang-utan populations and collected fecal samples for genetic analyses. We show that there is geographic variation in the presence of discrete types of calls. In exactly the same behavioural context (nest building and infant retrieval), individuals in different wild populations customarily emit either qualitatively different calls or calls in some but not in others. By comparing patterns in call-type and genetic similarity, we suggest that the observed variation is not likely to be explained by genetic or ecological differences. These results are consistent with the potential presence of ‘call cultures’ and suggest that wild orangutans possess the ability to invent arbitrary calls, which spread through social learning. These findings differ substantially from those that have been reported for primates before. First, the results

*\* These authors contributed equally to this work*

*Published in PLoS ONE (2012), 7, e36180*

reported here are on dialect and not on accent. Second, this study presents cases of production learning whereas most primate studies on vocal learning were cases of contextual learning. We conclude with speculating on how these findings might assist in bridging the gap between vocal communication in non-human primates and human speech.

### *Introduction*

Recent studies on various species, especially primates, examined geographic variation in a wide range of behaviours to examine the presence of traditions or cultures (defined as behaviors that are common in at least one site, but are absent in at least one other site, without concomitant genetic or environmental differences among these sites (van Schaik et al. 2003)). Comparisons of different populations of well-studied species such as chimpanzees, orangutans, spider monkeys and capuchin monkeys yielded a large number of behaviours that systematically varied among populations (Whiten et al. 1999; van Schaik et al. 2003; Perry et al. 2003; Santorelli et al. 2011). Application of the method of exclusion (or 'ethnographic method') suggested that individuals acquired many of those variants through socially mediated learning rather than through environmental induction or genetic canalisation because these are excluded by statistical analyses (Whiten et al. 1999; van Schaik et al. 2003; van Schaik et al. 2006; van Schaik et al. 2009). Recent tests that partially control for the effects of environmental and genetic differences among populations support this interpretation for orangutans without directly demonstrating social learning (hence the absence from the definition above) (Krützen et al. 2011). Moreover, the cultural interpretation is consistent with experimental evidence for observational learning in captive great apes (Whiten et al. 2004) and selective visual attention to techniques thought to be cultural among wild immatures (Jaeggi et al. 2010), as well as with experimentally induced diffusion of behavioural alternatives through captive populations of primates (Dindo et al. 2011; Whiten and Mesoudi 2008). Taken together, it has been suggested that these studies indicate that the first hominins had a modest cultural capacity, upon which the much more elaborate cumulative technological and institutional cultures that evolved in the genus *Homo* rest (e.g. (van Schaik et al. 2003)).

These conclusions have been challenged (Laland and Galef 2009). The main point of criticism of the method of exclusion is that the method does not show evidence for social learning, which is essential to claim culture (Laland and Galef 2009). It has been argued that translocation experiments of individuals or populations would unequivocally establish social learning in the wild, but there are ethical and legal obstacles to such experiments in many species, such as great apes. As a result, some argue that the evidence for culture is stronger in fish species than great apes (Laland and Hoppitt 2003). Alternative ways that social learning could potentially be demonstrated could be the introduction of new behaviours by a dispersing individual or when unrelated individuals living in close proximity converge upon the same behavioural variants. In addition, there is considerable debate on the impact of genetic variation on the reported behavioural dissimilarities between sites as witnessed by a recent exchange (Langergraber et al. 2011; Lycett et al. 2011; Langergraber and Vigilant 2011), but see (Krützen et al. 2011). At present, therefore, the evidence for primate cultures rests on plausibility.

Here we aim to advance the discussion on putative great ape culture by extending it to the vocal domain and examining genetic and ecological variation between sites. We hypothesize that like with many of the other behaviours in orangutans, genetic and ecological variation alone cannot easily explain the reported patterns (van Schaik et al. 2006; van Schaik et al. 2003; van Schaik et al. 2009).

Our focus here is not on suggested cultural behavioural variants in great apes that improve subsistence or comfort, or serve as variations on visual and tactile social signals, but rather on qualitatively different calls, so far reported exclusively for orangutans (van Schaik et al. 2003; van Schaik et al. 2006), but present in other animal taxa (Slabbekoorn and Smith 2002; Podos and Warren 2007; Noad et al. 2000; Rendell and Whitehead 2003; Rendell and Whitehead 2001; Schusterman 2008). Here, we focus on different call types made during nest building (nest smacks and raspberries) or by mothers to call infants (throat scrape and harmonic uuh), and will ignore the variation in the production of the so-called kiss-squeak by using hands or leaves in addition to the lips (Hardus et al. 2009b). Thus our study differs from studies that examined geographic variation (or variation among captive groups) in acoustic characteristics of the same call type (e.g. (Delgado 2007; Mitani et al. 1999; Wich et al. 2008; de la Torre and Snowdon 2009; Crockford et al. 2004; Fischer et al. 1998)). These between-population differences represent accents and not dialects, which is the focus of this paper (Lameira et al. 2010).

Such within-call type variation has been attributed to ecological and genetic factors, but also to vocal learning and thus argues for the existence of within-call type vocal learning in nonhuman primates (Crockford et al. 2004). Although evidence for geographic variation in discrete call types (i.e. dialects (Lameira et al. 2010)) that can be attributed to vocal learning has not yet been reported for nonhuman primates, it is common in birds (Slabbekoorn and Smith 2002; Podos and Warren 2007), cetaceans (Noad et al. 2000; Rendell and Whitehead 2003; Rendell and Whitehead 2001) and some non-primate mammal species (Schusterman 2008) . For several of these taxa the possibility of vocal cultures has been investigated (Rendell and Whitehead 2003; Rendell and Whitehead 2001; Noad et al. 2000; Bluff et al. 2010). Here we examine to what extent population specificity in the presence/absence of call types can be explained by genetic and ecological factors. We find that neither of these can sufficiently explain the observed variation and that therefore a cultural explanation remains viable.

### *Methods*

Our analyses focused on four call types (Figure 1). Fieldwork for this study was conducted at five study sites, two on Sumatra and three on Borneo (Figure 2, Table 1). Orangutans at all sites were well habituated to human observers. At each site orangutans were followed from dawn to dusk and behavioural data were collected following a standardized protocol ([www.aim.uzh.ch/orangutannetwork/FieldGuidelines.html](http://www.aim.uzh.ch/orangutannetwork/FieldGuidelines.html)). Researchers active at each site were focused on noting any calls the orangutans made (Hardus et al. 2009a). Many of the researchers worked at multiple sites and were familiar with the calls and behaviours at other sites. It is therefore unlikely that we report false negatives. At each site as many orangutans as possible were followed on a regular basis and the large number of hours and numbers of individuals followed at each site minimise the probability that we report false negatives at some of the sites (Table 1). Faecal samples from orangutans were collected at all sites and stored in ethanol or RNA later. The relevant permits for the observational work and field studies were obtained from the relevant institutes.

### *Call types and recordings*

The nesting calls examined in this study are given by all age-sex classes, except for individuals under 2-3 years of age. Calls of young individuals are made less often and are less loud and therefore we only have recordings of adult males and females. Both sexes are represented in the samples from both sites. The mother-infant calls are made by females with an infant old enough to feed or travel independently from the mother. At all sites such mother-offspring pairs were studied.

Calls were recorded with Marantz Analogue Recorder PMD222 in combination with a Sennheiser microphone ME 64, a Sony Digital Recorder TCD-D100 in combination with a Sony microphone ECM-M907 or a Marantz PMD600 in combination with a Sennheiser microphone ME 66. Calls were digitized in Raven Interactive Sound Analysis Software (2003, Cornell Lab of Ornithology, Ithaca, NY). Because of the brief duration of throat scrapes, non-default spectrogram settings in Raven were used for comparison between harmonic uuh's and throat scrapes in order to increase resolution and measurement accuracy in the time scale (window type = Hanning; spectrogram configuration: time grid spacing = 111 samples/frame overlap = 49.8%; frequency grid spacing = 46.9; window size = 221 samples; 3 dB bandwidth = 312 Hz). Default spectrogram representation in Raven were used for comparison between nest-smacks and raspberries (window type = Hanning; spectrogram configuration: time grid spacing = 256 samples/frame overlap = 50%; frequency grid spacing = 93.8; window size = 512 samples; 3 dB bandwidth = 135 Hz).

Throat scrapes are composed of 1-13 brief glottal pulses. Acoustical measurements were made on the glottal pulses. In total we measured 658 glottal pulses, 44 harmonic uuhs, 89 nest-smacks, and 35 raspberries (for the number of individuals see Table 1). Because the call types varied extensively in their acoustic structure we decided to measure four acoustic parameters that could clearly be measured from each call type and would not be influenced by recording distance so that the call types could be compared statistically. These were: duration (s), max frequency (Hz, the frequency with the maximum power (dB), delta frequency (Hz) and max power position (%). The max power position indicates where in the call the max frequency is occurs. For example, if the max frequency of a call is exactly at half the duration of the call, the max power position will be 50%. We used a discriminant function analysis to determine whether a call could be correctly assigned to its call type based on its acoustic characteristics (cf. Wich et al. 2008). We conducted a separate analysis for nest smacks versus raspberries, and for

throat scrapes versus harmonic uuhs. The analyses were conducted by ARL for a different manuscript (in prep) and therefore could not have been influenced by the aims of the analyses presented here. In addition, a 20% subsample of the calls was analyzed by (MEH) and similar percentages of correct assignments were found when re-running the discriminant function analyses on the subsample.

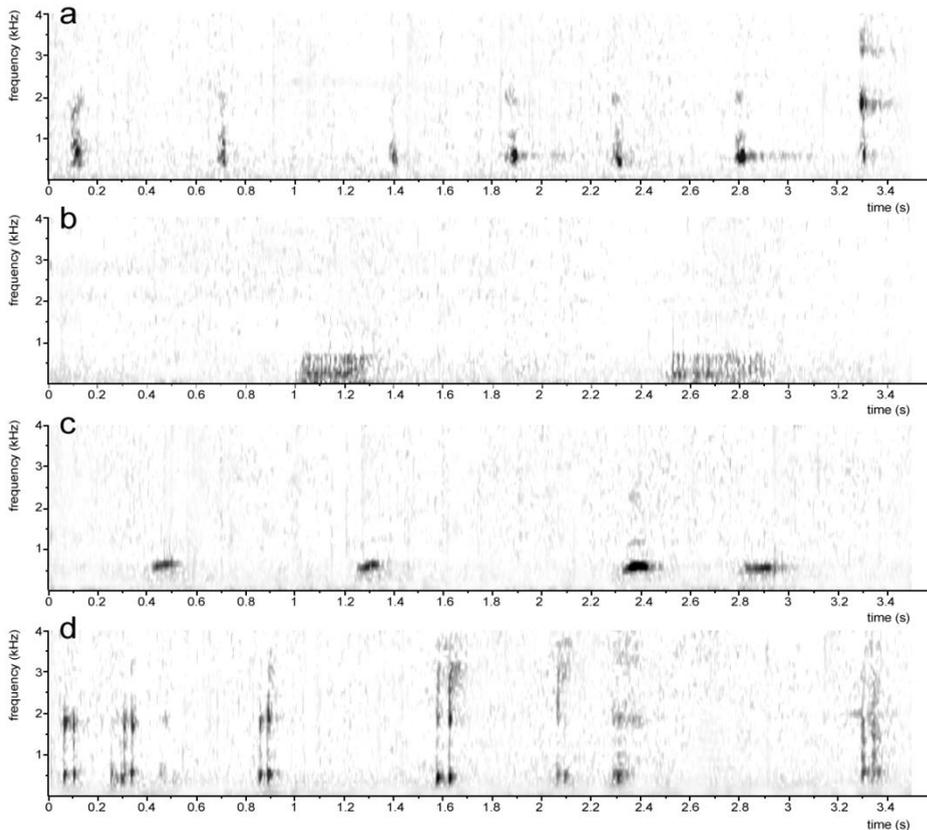


Figure 1. Orangutan call spectrograms. Spectrograms of orangutan calls: a) 'nest smacks'; b) 'raspberries'; c) 'harmonic uuhs'; d) 'throat scrapes'. The nest smack and raspberry are produced by orangutans during nest building. The harmonic uuh and throatscrape are produced by mothers towards infants that are separated from them and functions as a 'come-hither' call because infants return to the mother after these calls.



Figure 2. HVR-I haplotype median-joining network. A median-joining network showing HVR-I haplotypes in the different populations in relation to orang-utan calls: nesting calls and mother-infant calls. The size of each circle corresponds with the number of individuals with this particular haplotype, with the smallest circles representing one individual with this particular haplotype. Black dots indicate mutational steps connecting the sampled haplotypes, and thus represent haplotypes that were not sampled and may or may not exist. Each number on the network indicates a single base-pair mutation. First letter code in blue refers to the kind of nesting call (r = 'raspberry'; s = 'nest smack'; - = no call). Second letter code refers to the mother-infant call (u = 'harmonic uuh'; t = 'throat scrape'; - = no call).

### *Genetic marker systems*

To estimate genetic distance between study sites, we utilised parts of the rapidly evolving hyper-variable segment of the mitochondrial control region (HVRI), which reflects the time since divergence from a common ancestor. Due to the rapid evolution of the HVRI region, this marker may produce homoplasy between the island populations of orangutans, resulting in underestimation of the true genetic distance between populations. Hence, we also calculated genetic distance between sites using 1228 base

pairs of three concatenated mitochondrial genes, which evolve more slowly than the HVRI and at a similar rate as coding nuclear loci.

DNA from 96 wild adult orangutans with known provenance (Table 1) was extracted using the Qiagen stool-kit according to manufacturer's instructions. We obtained sequence information of two mitochondrial segments in order to calculate genetic distances between orang-utan populations. First, we amplified the HVRI region, comprising part of the non-coding control region, using primers DLF 5'-CTGCCCTGTAGTACAAATAAGTA-3' (developed by A.N.) and D5 (Warren et al. 2001), resulting in a 357 base pair product. PCR reactions consisted of 1-40 ng template DNA, 0.25  $\mu$ M of each primer, 0.2 mM dNTPs, 2  $\mu$ g BSA (NEB), 2  $\mu$ l of 10x PCR buffer containing 15 mM  $MgCl_2$ , 0.6 u HotStarTaq DNA polymerase (all Qiagen), and ddH<sub>2</sub>O to a 20  $\mu$ -volume. Hot-start PCR reactions were carried out on a Veriti Thermal Cycler (Applied Biosystems) with the following cycling scheme: initial denaturation for 15 min at 95°C, 35-45 cycles (depending on the starting DNA concentration) of 94 °C for 30 s, 58 °C for 30 s and 72 °C for 60 s, followed by a final extension at 72 °C for 10 min. Second, we amplified a total of 1228 base-pairs (bp) from three mitochondrial genes (NADH dehydrogenase subunit 3, 345 bp; cytochrome b, 496 bp; and 16S rRNA, 387 bp), using primers developed by (Muir et al. 2000; Zhi et al. 1996). Molarities for the PCR reactions were identical to those used in the HVR-I amplifications. Cycling conditions for all three genes were initial denaturation for 15 min at 95°C, followed by cycles of 94 °C for 30 s, 58 °C for 40 s and 72 °C for 40 s. The PCR was finished by a final extension at 72 °C for 10 min.

All PCR products were cycle-sequenced using 1  $\mu$ l of PCR product, 1.75  $\mu$ l 5x sequencing buffer (10 mM  $MgCl_2$ , 400 mM Tris, pH = 9.0), 0.5  $\mu$ l BigDye Terminator v3.1 (Applied Biosystems), 0.4  $\mu$ M sequencing primer and ddH<sub>2</sub>O up to 10  $\mu$ l total volume. The cycling scheme was as follows: initial denaturation at 95 °C for 45 s, 30 cycles of 95 °C for 30 s, 52 °C for 20 s, and 60 °C for 4 min. Sequencing reactions were cleaned up using 75  $\mu$ l of 0.2 mM  $MgSO_4$ , in 70% v/v EtOH. Capillary electrophoresis was performed on a 3730 DNA Analyzer (Applied Biosystems).

Complementary sequences were added to a contig and sequence identity was checked in Lasergene SeqMan Pro v7.1.0 (DNASTAR). Sequences were collapsed into unique haplotypes using Clean Collapse v.1.0.5. For HVR-I, intraspecific gene genealogies were inferred using a median joining network in Network v. 4.5.1 (available from [www.fluxus-technology.com](http://www.fluxus-technology.com)). Genetic distances between pairs of populations were

calculated using the software Mega v. 4.0 (Tamura et al. 2007), employing the Maximum-Composite Likelihood distance with gamma parameters of 0.210 and 0.196 for the HVR-I region and the concatenated mtDNA genes, respectively.

Table 1. Orang-utan site information

Site (start of study)	Coordinates/habitat	No. hrs of focal observation	Individuals observed nesting (# making calls)	Mother-infant pairs observed (making call)	No. of sequenced individuals
Tuanan (B) (2003-)	2° 09' S 114° 26' E/ Peat swamp	> 15,000	21 (21)	8 (8)	20
Sg. Lading (B) (2005-2007)	2° 15' S 114° 22' E/ Peat swamp	> 2,000	6 (0)	4 (0)	24
Sabangau (B) (2003-)	2° 19' S 114° 00' E/ Peat swamp	>3,000	19 (18)	4 (0)	21
Ketambe (S) (1971-)	3° 41' N 97° 39' E/ Dryland	>15,000	20 (0)	6 (5)	16
Suaq (S) (1994-)	3° 04' N 97° 26' E/ Peat swamp/ dryland	>10,000	28 (25)	12 (0)	15

Note: (B) = Borneo, (S) = Sumatra. For the number of individuals making nests, we only included individuals that were followed for more than 10 nights, because after this number of night nests most orang-utans that occur in sites were they make nest calls were found to have made a nest call. At sites where mother-infant calls were heard, they occur once every 7.8 mother-infant follow hours for Ketambe (994 total follow hours) and 42.6 follow hours for Tuanan (5827 total follow hours). At the three sites where these calls were not heard, many more follow hours have been collected (Sabangau: 1709 hrs; Sg. Lading: 2140 hrs; and Suaq: 7665 hrs).

### *Statistical analysis*

The traditional hypothesis is that the call repertoire of primates has a strong genetic basis (e.g. review in (Owren et al. 2010)) and therefore that population differences in the types of calls produced among primates have a genetic basis. If there is a genetic predisposition

for a particular behavioural trait, but there is no information on the genes involved, one would predict that genetic similarity between two individuals is correlated with the similarity in the trait (Falconer and Mackay 1996). Thus, if a genetic signal were present in for example the nesting calls, one would predict that pairs of sites sharing the same state (either presence or absence of the same call type) have a smaller genetic distance on average than pairs of sites with different states (call present in one but absent in another). To evaluate a potential genetic explanation, we applied the following Monte-Carlo procedure. We randomly redistributed the observed behaviours among the five sites a thousand times. For each randomisation, we calculated the genetic differentiation value (GDV), defined as the difference between the averaged genetic distances between the two classes of sites, as above. We carried out this analysis separately for nesting calls and mother-infant calls, using the genetic distances based on HVR-I and the mtDNA genes.

## *Results*

Population comparisons focused on five sites where wild orangutans have been studied extensively (Table 1; Figure 2). Observations revealed at least two behavioural contexts in which orangutans in different populations make very distinct sounds: nest building and infant retrieval by the mother.

All wild orangutans build night nests on a daily basis. During the last stage of nest construction, some produce a call that varies among the five populations compared here. These variations are categorical: the raspberry and nest-smack calls are very different in their acoustic properties (Figures 1a and b). The discriminant function analysis based on the four acoustical measurements explained 100% of the variance between the two calls on the basis of one function. This function had the highest correlations with call parameters duration and maxfreq (-0.756 and -0.705 respectively). For raspberries 96.7% of the cases ( $n = 30$ ) were classified correctly and for nest smacks this was 100% ( $n = 34$ ). Using the leave-one out validation these percentages remained exactly the same. Thus, orangutans in Tuanan routinely produce 'nest smacks' (audio file S1), and those in Suaq and Sabangau 'raspberries' (audio file S2). These calls have high prevalence, *i.e.* are essentially made by all individuals (Table 1), and are made on a daily basis. In contrast, such routine nesting calls are completely absent in Ketambe and Sungai Lading.

The second call concerns the maternal ‘come-hither’ call, made by orang-utan mothers just before retrieving their infant (Figures 1c-d). In three populations, all mothers examined are silent (Figure 2), but in one population, Ketambe on Sumatra, all use one call (‘harmonic uuh’: audio file S3, Figure 1c), whereas in another, Tuanan on Borneo, they use a completely different call (‘throat scrape’: audio file S4, Figure 1d). Similar to the analyses of the nesting calls, the discriminant function analysis based on the four acoustical measurements explained 100% of the variance between the harmonic uuh and the throat scrapes on the basis of one function with which call parameters duration and maxfreq had the highest loadings (-0.932 and 0.175 respectively). For harmonic uuhs 97.7% of the cases (n = 44) were classified correctly and for throat scrapes this was 100% (n = 44). Using the leave-one out validation these percentages remained very similar (95.5 and 100%, respectively). As with the nest-building calls, in populations where these calls are made they are (near)-ubiquitous in their prevalence (Table 1), and are emitted on a regular basis (albeit less than once a day).

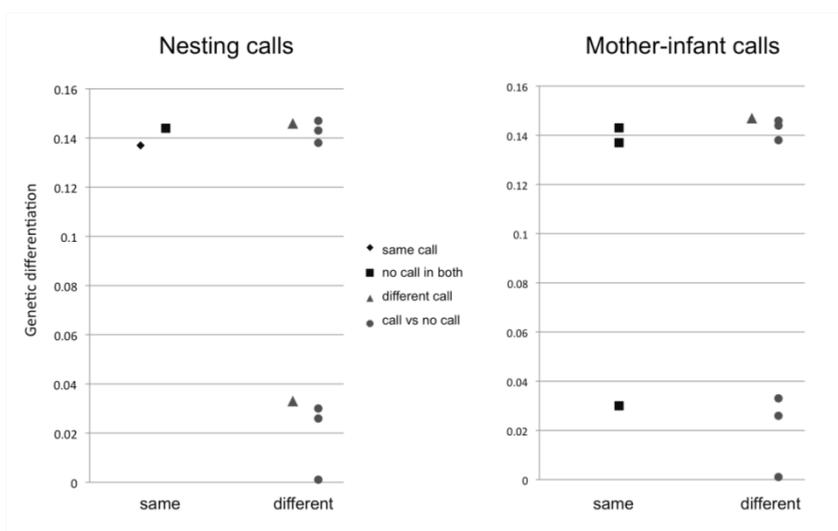


Figure 3. Average genetic distance between pairs of sites. Average genetic distance (maximum composite likelihood distance of HVR-I haplotypes, see Material and Methods) between pairs of sites in five orangutan populations, for two different situations: where nesting calls and mother-infant calls are the same in both sites, and where the two sites are different. If genes play a role in the production of these calls, pairs of sites with the same behavioural state should show smaller average genetic distance than pairs of sites with different behavioural states.

Figure 3 suggests no relationship between average pair-wise genetic distance based on mitochondrial genes and similarities in nesting calls and mother-infant calls in the five orang-utan populations, regardless of whether we used the concatenated mitochondrial genes or the non-coding HVRI region. Indeed, if we separate these points into two classes (high and low genetic differentiation), 4 out of the 5 site-pairs that have the same behavioural state (same call or no call) are in the wrong direction (e.g. high genetic differentiation despite the same behavioural state at a site-pair). A more formal test is reported in Figure 4. The observed value is shown in relation to the cumulative distribution of the randomised genetic differentiation values. There is a trend toward lower genetic similarity between pairs of sites with the same state of the calls, opposite to prediction if the calls were genetically canalized.

Additional evidence against a purely genetic explanation is provided by the median joining network (Figure 2) which shows three of the males sampled in Tuanan as having haplotypes that were genetically closer to those found in Sungai Lading (Figure 2), where no calls are produced. Yet all males in Tuanan observed to date have been found to produce nest calls.

### *Discussion*

This study has shown that orangutans produce population-specific calls that are statistically distinct in their acoustical variables and clearly constitute different call types. Hence, the study expands on an earlier study in orangutans that showed that there was geographic variation in the orang-utan male long-distance call (the long call (Delgado 2007) and other studies showing such within-call type acoustic variation for other primate species (Crockford et al. 2004; Mitani et al. 1999; Wich et al. 2008; de la Torre and Snowdon 2009). Such within-call type geographic variation has also been suggested to be due to vocal learning and not to ecological or genetic factors (Crockford et al. 2004; Marshall et al. 1999).

Here we show that population-specific calls made in the nest-building and infant retrieval contexts are independent of genetic variation across these populations, which implies that the production of these calls is at least not totally genetically canalized during development. These results are therefore in correspondence with a recent study on geographic variation in orangutans that addressed the same question for a large number of behavioural and social variables (Krützen et al. 2011). This study found that genetic dissimilarity between populations for putative cultural behaviours was not correlated with genetic or environmental variation.

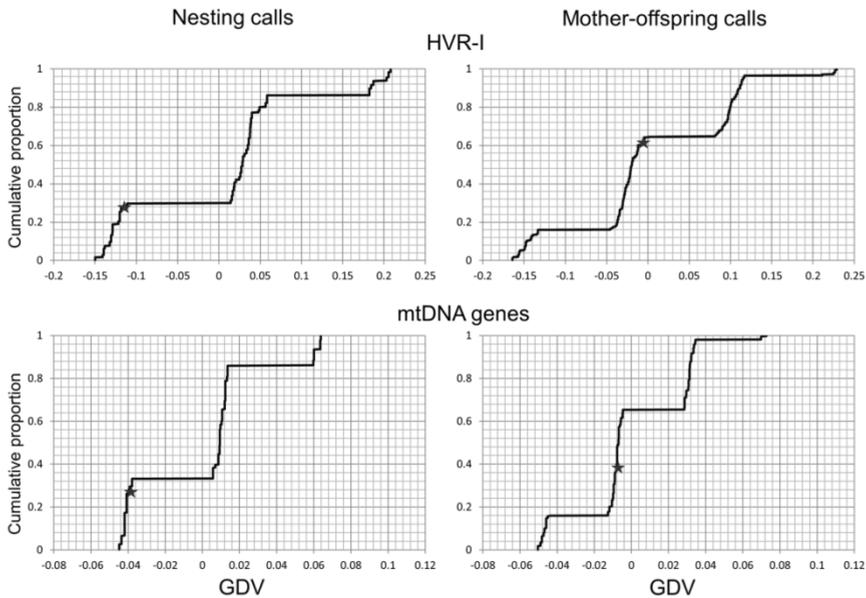


Figure 4. Cumulative distribution of randomised genetic differentiation values (GDV) among populations. GDVs were generated as follows: the observed behavioural states were randomly assigned to each of the 5 sites a thousand times, thereby producing site pairs with the same, but also with different behavioural states compared to those that were originally observed for each randomisation. For each randomisation, we then calculated GDV, defined as the difference between the averaged genetic maximum composite likelihood distance among sites pairs with different behavioural states and the averaged genetic maximum composite likelihood distance among site pairs with the same behavioural state. If genetic similarity played a role in the observed pattern, the observed GDVs are expected to be positive. The star indicates the value actually observed in this study.

The absence of genetic effects is consistent with recent studies using data from autosomal genomes (Locke et al. 2010) and Y chromosome polymorphisms (Nater et al. 2011), which showed a surprisingly recent, not previously documented divergence time of about 400 kya and 168 kya between Sumatran and Bornean orangutans, respectively. Hence, the similarity in patterns between and within islands suggests that these orang-utan calls are not genetically canalised. For instance, the nest raspberries appear in both a Sumatran and a Bornean site, whereas nearby sites have completely different behavioural states.

Landscape-level ecological differences (dryland vs peat swamp forest) could be excluded as potentially explaining the population specificity of these orang-utan calls because several of the differences were found in the same habitat type. All but one of the populations compared live in peat swamp habitats and those on Borneo are in close proximity (but across impassable rivers: Figure 2). They nonetheless vary greatly in either the type or the very presence of nesting calls. In addition, orangutans at Suaq sometimes make nests in dryland forest areas and then still emit the raspberry. Ketambe, where the 'harmonic uuh' is produced, is a dry-land forest, whereas Tuanan, where the 'throat scrape' is produced, is a peat swamp. Thus not all orang-utan populations occurring in peat swamp populations make throat scrapes because these are absent in Suaq, Sg. Lading, and Sabanagau (all peat swamp sites). Similarly, nor do all orang-utan populations that occur in dryland areas produce harmonic uuh calls because part of the Suaq study site is dryland and no harmonic uuh is produced there (Table 1). In addition, even though these are different habitats, the calls are aimed at the infant, which is rarely more than a few meters away in the canopy. This patterning among habitats is therefore consistent with the observation that call propagation properties of different habitats become apparent only at much larger distances than observed here (Wiley and Richards 1978), and that habitat differences have been used in primate studies to explain gradual changes of the same call types, such as subtle frequency changes, but not for the replacement of one call type by another (Mitani et al. 1999; Wich et al. 2008; de la Torre and Snowdon 2009).

The potential role of possible small-scale ecological variation between sites was not addressed and its potential influence on acoustic signals deserves more study. However, for various reasons we think that potential variation of small-scale ecological variation (e.g. leaf density, canopy structure) at best has a limited impact on our results.

First, variation in habitat ecology in general is often considered to have an influence on the acoustic structure of long-distance signals, not of short-distance signals (Marten and Marler 1977; Wiley and Richards 1978). Second, the three sites on Borneo are all in the same peat-swamp forest block and consequently these sites are ecologically very similar (Bastian et al. ; Slik et al. 2011), in which orangutans show many similarities in foraging and nesting behaviour (Wich et al. 2009a). Thus, it is unclear why we find completely different calls types rather than subtle differences within call types. Indeed, studies focusing on bird song that have described variation in bird song have mainly found differences between very different habitats such as open and closed ones within the same species (Slabbekoorn and Smith 2002; Bertelli and Tubaro 2002; Morton 1975), and where within-species dialects have been reported vocal learning has been the predominant explanation, with ecology or genetics given much less prominence (e.g. (Saranathan et al. 2007)). Finally, the influence of potential between-population variation in sound propagation for these calls is probably very limited because excess attenuation differences for the frequency range of the calls studied here are most pronounced below heights of 1m above the ground rather than at the much greater heights (Marten and Marler 1977; Waser and Brown 1986) at which orangutans in these forests build nests and forage (Wich et al. 2009a).

It could also be argued that differences in sociality affect our results, because Sumatran orangutans are more social than those on Borneo (Wich et al. 2009a). The relevance of overall sociality variation on nest calls is probably negligible, however. First, orangutans often nest solitarily (or in the case of mothers with offspring, only with their offspring) and for Tuanan it has been shown that the presence of associates does not affect the production of the nesting calls (Paul 2007). Second, the influence of overall sociality on mother-infant calls is likely to be limited because the three Bornean populations show a similar sociality, but nevertheless vary in the presence or absence of mother-infant calls. The same argument holds for the two Sumatran populations. They also show similar sociality, but a mother-infant call is found in only one of these populations.

The calls had high prevalence where they occurred, basically being made by all relevant individuals (Table 1). Thus, the results presented here strongly suggest that these sounds were invented in each population and subsequently spread through social learning (cf. Hardus et al. 2009a). This interpretation is consistent with evidence in orangutans for

the two critical elements for culture: innovation and social learning. First, fieldworkers have observed that individual apes sometimes produce 'private' calls (i.e. calls that are unique to this individual) in play or nest building (unpublished obs., M. van Noordwijk, M. Bastian, M. Paul), suggesting that the invention of novel calls is not implausible. Second, studies show that captive orangutans and chimpanzees can socially learn to give calls that are not part of the species-specific repertoire, such as a whistle, and subsequently show flexible usage of such calls (Miles et al. 1996; Wich et al. 2009b; Hopkins et al. 2007). Outside the call domain strong indirect evidence for social learning has been found in a number of orangutan field studies (Jaeggi et al. 2010; Jaeggi et al. 2008; Bastian et al.). Taken together, the recent wild studies (van Schaik et al. 2003; van Schaik et al. 2006; Hardus et al. 2009b) and captive studies (Miles et al. 1996; Wich et al. 2009b; Hopkins et al. 2007) on great apes have recently been interpreted to indicate that great apes have some voluntary control over respiration and vocal fold adduction (Owren et al. 2010). It is therefore perhaps no coincidence that the signals recorded in this study are calls that have either no (raspberry and nest smack) or little involvement of the vocal folds (throat scrape and harmonic uuh), with only the latter two showing higher harmonics (even though not clearly depicted for harmonic uuhs in figure 1).

Several types of vocal learning have been described (Janik and Slater 2000) and it is therefore relevant to determine which type of vocal learning could be important for the results presented in this paper. The calls investigated in this study are distinct from other calls types in the orang-utan's call repertoire (Hardus et al. 2009a) and not examples of calls from the existing repertoire used in novel context in some populations. Therefore the results presented here are not an example of contextual learning (Janik and Slater 2000), but are likely to have been innovated in the populations where they are found and thus an example of production learning.

It has been suggested that there is a large gap between the vocal communication of nonhuman primates and human language, making it hard to see how the latter could have evolved from the former (Fitch 2010). However, the presence of these cultural calls in orangutans suggests the gap is perhaps not as wide as often perceived. Orangutans occasionally invent calls with an arbitrary acoustic structure. The spread of these calls can be understood through shared need, so the audience could easily grasp from the context what the function of the calls should be (although the function of the nest calls remains unknown), in a process very similar to the social learning of the functional use of innate

vocalizations in other species (e.g. (Cheney and Seyfarth 1992; Fichtel and Van Schaik 2006)). Thus, the orang-utan findings imply that we are dealing with arbitrary symbols that had acquired a shared meaning — two important elements of language.

#### *Acknowledgements*

We acknowledge the cooperation of the Indonesian Institute of Science (LIPI), the Indonesian Ministry of Research and Technology (RISTEK), the Indonesian Nature Conservation Service (PHKA), Universitas National (UNAS), Indonesian Primate Centre, Bogor Agricultural University.



## **Tool use in wild orangutans modifies sound production: a functionally deceptive innovation?**

Madeleine E. Hardus, Adriano R. Lameira, Carel P. van Schaik and Serge A. Wich

### *Abstract*

Culture has long been assumed to be uniquely human but recent studies, in particular on great apes, have suggested that cultures also occur in non-human primates. The most apparent cultural behaviours in great apes involve tools in the subsistence context where they are clearly functional to obtain valued food. On the other hand, tool-use to modify acoustic communication has been reported only once and its function has not been investigated. Thus, the question whether this is an adaptive behaviour remains opened, even though evidence indicates that it is socially transmitted (i.e. cultural). Here we report on wild orangutans using tools to modulate the maximum frequency of one of their sounds, the kiss squeak, emitted in distress. In this variant, orangutans strip leaves off a twig and hold them to their mouth while producing a kiss squeak. Using leaves as a tool lowers the frequency of the call compared to a kiss squeak without leaves or with only a hand to the mouth. If the lowering of the maximum frequency functions in orangutans as it does in other animals, two predictions follow: 1) kiss squeak frequency is related to body size and 2) the use of leaves will occur in situations of most acute danger. Supporting these predictions, the frequency of kiss squeaks without tools decreased with body size and kiss squeaks with leaves were only emitted by highly distressed individuals. Moreover, we found indications that the

*Published in Proceedings of the Royal Society B: Biological Sciences (2009), 276, 3689-3694*

calls were under volitional control. This finding is significant for at least two reasons. First, although few animal species are known to deceptively lower the maximum frequency of their calls to exaggerate their perceived size to the listener (e.g. vocal tract elongation in male deer) it has never been reported that animals may use tools to achieve this, or that they are primates. Second, it shows that the orangutan culture extends into the communicative domain, thus challenging the traditional assumption that primate calling behaviour is overall purely emotional.

### *Introduction*

Studies of animal cultures have often focused on tool use innovations for subsistence purposes (e.g. Whiten *et al.* 1999; van Schaik & Knott 2001; van Schaik *et al.* 2003, 2006; Krutzen *et al.* 2005; Lefebvre *et al.* 2002), defining tools as objects that are used as an extension of the body and that are held directly in the hand or mouth (Lefebvre *et al.* 2002, see also Beck 1980). However, tool-based innovations in acoustic communication that modify the acoustical properties of a call have been virtually absent, and the only case where such an innovation has been proposed concerns orangutans (van Schaik *et al.* 2003, 2006; Hardus *et al.* 2009). Nevertheless the function of this behaviour is not clear (Peters 2001), raising questions about its adaptiveness and dispersal within populations.

Here we report on tool use and its function in one particular call, the kiss squeak, in wild Bornean orangutans (*Pongo pygmaeus wurmbii*). The kiss squeak has been described as a sharp intake of air through pursed lips causing a kissing sound (Hardus *et al.* 2009; Rijksen 1978). This call is produced by all orang-utan age-sex classes in response to disturbance and/or fear towards potential predators (e.g. snakes, clouded leopards, tigers, humans) or other orangutans upon sight, at times accompanied by display behaviour and most likely to deter the potential predator (Hardus *et al.* 2009; Rijksen 1978). Because orangutans are semi-solitary animals living in fission-fusion societies (Delgado & van Schaik 2000), it often takes hours before conspecifics are attracted to the kiss squeaks (van Noordwijk & van Schaik 2009), so its most likely function is to send a signal to the predator or approaching orang-utan and not to attract other conspecifics.

Kiss squeaks can be given in three different forms: unaided, and with either a hand or leaves positioned in front of the lips (van Schaik *et al.* 2003; Peters 2001), and these forms are given in the same context and occasionally within the same bout. The leaves function as a tool because they are first stripped off from a twig and then held in one hand as a bundle on the mouth while the kiss squeak is produced (see supplemental video file). Leaves used as tools are close at hand and leaf species seem to be chosen indiscriminately. Because the presence or absence of kiss squeaks on leaves in a particular orang-utan population does not appear to be genetically or ecologically determined, this behaviour has been suggested to arise as an innovation and to be culturally transmitted, consistent with its limited geographic distribution and high local prevalence (van Schaik *et al.* 2003, 2006; Hardus *et al.* 2009). Our results indicate that producing kiss squeaks on leaves functions to make the caller sound larger, because it lowers the call's maximum frequency (frequency with the highest power) and because the frequency of unaided kiss squeaks is negatively correlated with body size.

## *Methods*

### *Study Site*

All kiss squeaks were recorded from wild orangutans at research station Tuanan (2°09'06"S; 114°26'26"E), Central Kalimantan, Borneo, Indonesia. This study area is composed of forest on shallow peat with relatively homogenous canopy. Orang-utan kiss squeaks were recorded from January 2003 until August 2005.

### *Data Collection and Data Analysis*

All kiss squeaks produced by a focal orang-utan and/or its associates were recorded opportunistically using a Marantz Analogue Recorder PMD222 in combination with a Sennheiser Microphone ME 64 or a Sony Digital Recorder TCD-D100 in combination with a Sony Microphone ECM-M907. All incidents of recorded kiss squeaks occurred throughout the day. We digitized all the recorded kiss squeaks at 44.1 kHz, using Raven Interactive Sound Analysis Software (2003, Cornell Lab of Ornithology, Ithaca, NY). We transformed all recordings into spectrograms (window type = Hanning; spectrogram configuration: time grid spacing = 256; samples/frame overlap = 50%; frequency grid spacing =86.1;

window size = 512 samples; 3-dB bandwidth = 124 Hz, i.e., narrow-band like spectrogram). A call's maximum frequency (Hz), maximum power (dB) and duration (s) were measured from spectrograms to compare the different kiss squeak forms. Maximum frequency represents the frequency with the highest energy emitted in a call, independently of its location. Maximum power represents the energy of the maximum frequency (i.e. loudness). Duration represents the time period between the beginning and the end of the call. Because kiss squeaks are brief and rather noisy atonal calls (cf. Struhsaker 1967) the number of measurable acoustic parameters was limited. Maximum power was analysed within recording bouts in order to control for distance between recorder and focal, recording volume settings and acoustic environment during recordings. Recordings from adult females were less represented in our sample because these animals were most frequently followed and thus less disturbed by the presence of human observers. Consequently, recordings of kiss squeaks on leaves were biased towards adult males.

### *Results and Discussion*

The maximum acoustic frequency of the three different kiss squeak forms are significantly different from each other within individual orangutans (Figure 1, Friedman test:  $n = 813$  calls emitted in disturbance contexts by 8 adult and 1 adolescent individuals,  $\chi^2 = 1596.0$ ,  $p < 0.001$ ; followed by a multiple comparisons post-hoc test:  $p < 0.001$  between all three forms). Specifically, the maximum frequency decreased progressively from the unaided kiss squeak to kiss squeak with hand, to kiss squeak on leaves (Figure 1). Maximum power did not differ between kiss squeak forms within recording bouts (Paired-samples T-test: unaided/hand,  $n = 62$  call pairs,  $p = 0.848$ ; unaided/leaves,  $n = 14$  call pairs,  $p = 0.357$ ). Duration also did not differ among the three kiss squeak forms (Kruskal Wallis test,  $N = 536$  calls emitted towards observers by 17 adult individuals,  $\chi^2 = 3,813$ ,  $p = 0.149$ ). Hence, these results do not support the suggestion by Peters (2001) that kiss squeak forms with hand and on leaves function to increase the volume, and "somewhat" the frequency of the call. Ambient factors, e.g., branches, leaves, trunks or atmospheric factors, produce rapid attenuation of high acoustic frequencies and a shift to lower frequencies (Lameira & Wich 2008), thus a decrease in a call's frequency generated by positioning some item in

front of the mouth during emission supports this prediction that is well-supported by various studies (e.g. Wiley 1991; Brown 2003).

We hypothesize that lowering the maximum frequency functions to mislead the receiver that the producer has a larger body size than it actually has. Body size enlargement behaviour occurs throughout the animal kingdom in situations in which an individual is disturbed (e.g. fur-bristle in cats, air-swallowing and rising-on-legs in frogs, pilo-erection in chimpanzees), but it is rare in mammalian calling behaviour (e.g. Fitch & Reby 2001; Matrosova *et al.* 2007). In accordance with this hypothesis, unaided kiss squeaks decreased significantly in their maximum frequency with age and thus with increased body size (Kruskal-Wallis Test: KW = 238.7, df = 4,  $p < 0.001$ ; followed by a

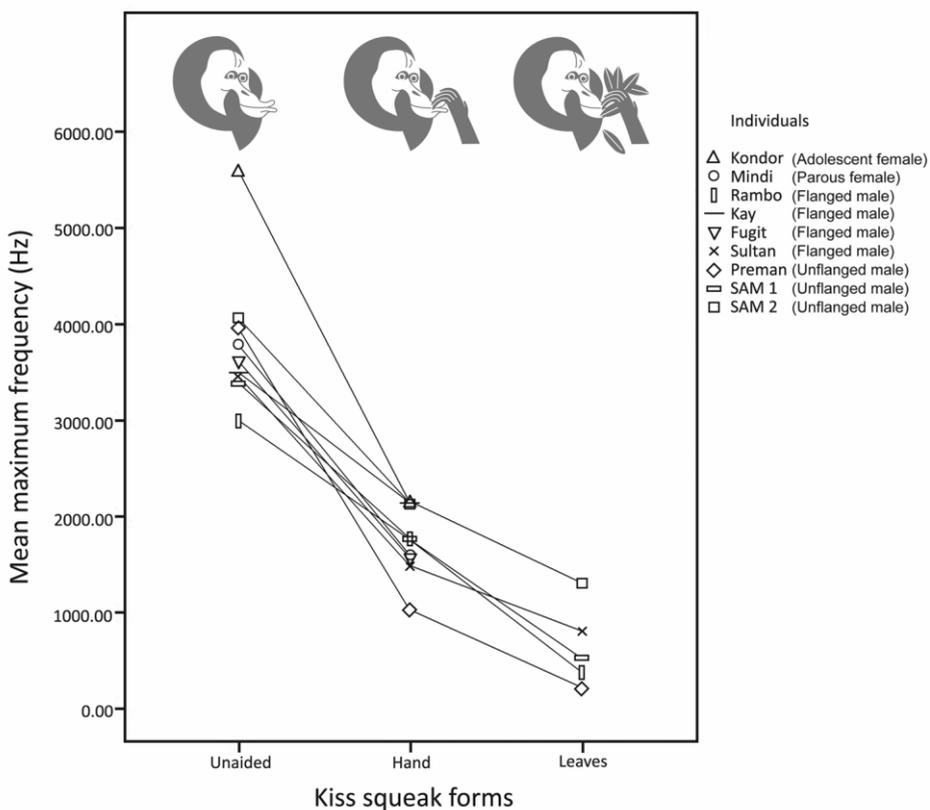


Figure 1. Mean maximum frequency (Hz) of each kiss-squeak form for nine orangutans. Illustrations by A. R. Lameira.

multiple comparisons post-hoc test:  $p < 0.001$  between immatures and all the other size classes and between flanged males and all the other size classes, Figure 2). To our knowledge this is the first evidence of *how* and *why* non-human primates affect their own call variables through the use of hand and tools.

Because dishonestly signalling larger body size might yield the greatest advantage in highly dangerous situations, we expect that the use of kiss squeaks on leaves would increase in such conditions. It is likely that for unhabituated orangutans, encounters (i.e. when parties detect each other) with human observers qualify as dangerous. Unhabituated orangutans were those that could not be followed for several days without showing signs of disturbance, expressed in flight, displays and the emission of disturbance calls. Indeed kiss squeaks on leaves were only produced by unhabituated orangutans ( $n = 5$ ) during encounters with human observers ( $n = 17$  calls, Figure 3). Kiss squeaks with hand towards humans also made up a significantly higher proportion for unhabituated orangutans ( $n = 10$ ) than habituated orangutans ( $n = 12$ ) ( $\chi^2 = 5.46$ ,  $df = 1$ ,  $p = 0.019$ ,  $n = 42$  calls, Figure 3). On the other hand, unaided kiss squeaks were not emitted significantly more common among unhabituated ( $n = 15$ ) than habituated orangutans ( $n = 10$ ) ( $\chi^2 = 0.44$ ,  $df = 1$ ,  $p = 0.5$ ,  $n = 502$  calls, Figure 3). When unhabituated orangutans encountered humans, although it would be beneficial to modify all emitted kiss squeaks, orangutans emitted kiss squeak on leaves/with hand in combination with unaided kiss squeaks (Table 1). It seems that this way orangutans free their hands use for other beneficial purposes during an encounter with a potential predator, such as display, branch-missiles, movement through the canopy or escape. To refrain from emitting unaided kiss squeaks could prove deleterious by decreasing drastically the alarm call rate by the orang-utan facing a predator (see Zuberbühler *et al.* 1999).

This pattern indicates that kiss squeaks with hand and on leaves, i.e. with lower frequencies, were emitted in circumstances assumed to be more dangerous. This functional use is plausible because orangutans are arboreal apes living in dense forests where visibility is usually poor and rarely sufficient to make accurate visual assessments of body size, particularly in disturbing encounters when orangutans use displays and missiles towards potential predators. This way, the visual salience hypothesis (Peters 2001) for the function of the kiss squeak forms also seems unlikely due to poor visibility and because the dropping of leaves is not executed in isolation, but in combination with branch-shaking, breaking and throwing for example. Although one might argue that individuals of

bigger body size classes (e.g. flanged males) would not deploy kiss squeaks with hand and on leaves for the purpose of functional deception, this interpretation ignores that all individuals may benefit from appearing larger in highly disturbing contexts. Thus, it is reasonable to assume that kiss squeak forms function to deceptively convey to the predator a larger body size. Indeed, as expected, orangutans were never observed or reported to emit kiss squeaks with hand and on leaves towards other orangutans, since producer and receiver will generally be familiar with one another and/or with the deceptive technique. At the same time, low sound frequencies travel farther than high frequencies (see Lameira & Wich 2008) and thus the use of hand and leaves during the production of kiss squeaks could be meant for the purpose of conspecifics' recruitment. However this is unlikely since recruitment of conspecifics via the emission of kiss squeak forms, while not uncommon, usually takes much longer than the duration of an encounter (van Noordwijk & van Schaik 2009).

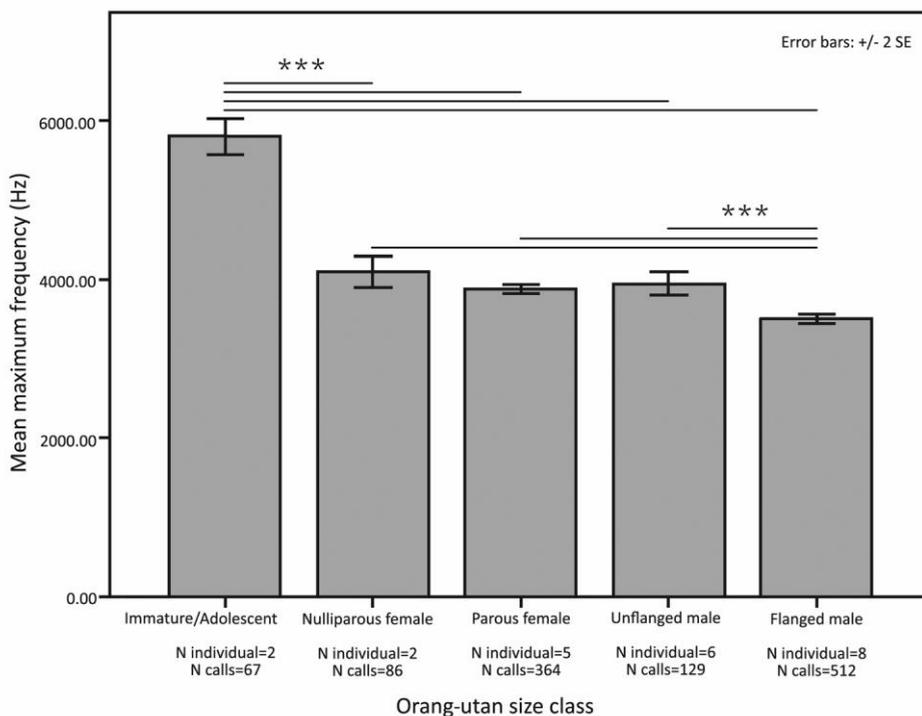


Figure 2. Mean maximum frequency (Hz) per size class of the unaided kiss squeak. Error bars: +2 s.e. \*\*\*p, 0.001.

To confirm that kiss squeaks with hand and on leaves are innovations that spread within populations due to their putative deceptive function, it is important that the behaviour is voluntary and not under strict emotional or motivational control. Although we cannot test this directly, the different kiss-squeak forms were given within the same bout per individual (34 bouts recorded for 11 individuals, Table 1). No patterns were seen per bout that could indicate escalating irritation. Hence, they were used interchangeably within the same context showing no particular relationship to apparent changes in the underlying emotional/ motivational state (see Owren & Rendall 2001). Moreover, immature individuals are known to emit uncoordinated kiss squeaks with hand and on leaves in playful situations (Hardus *et al.* 2009) (not considered in this study), which indicates that practice is important in the acquisition of these techniques. Orangutans in captivity also use kiss squeaks to capture the attention of otherwise inattentive humans (Cartmill & Byrne 2007). Such flexible use of kiss squeaks suggests that the use of hand and leaves during the production of kiss squeaks and its functional use is, to some extent, under volitional control and thus both forms are suitable candidates for innovative call variants. In support of this, Hopkins and colleagues (2007) have reported the novel and functional use of certain calls by captive chimpanzees, implying that these behaviours are voluntary in some measure, and argued that this represents a form of social innovation that parallels those associated with tool use in other contexts (Leavens *et al.* 2005). Moreover, the spontaneous acquisition of human whistling by a captive orang-utan without prior training (Wich *et al.* 2009) demonstrates that orangutans have particular flexibility and control over calls produced by the lips.

Because kiss squeaks with hand are almost certainly a universal behaviour across orang-utan wild populations, while kiss squeaks on leaves are only present at certain sites, where they are highly prevalent (van Schaik *et al.* 2003, 2006; Hardus *et al.* 2009), the results presented here suggest that kiss squeaks on leaves represent a functional innovation that spread locally and became cultural. Degree of habituation also determines whether or not orangutans display this behaviour but highly unhabituated individuals in some survey areas have failed to perform this behaviour. Thus, by further decreasing the kiss squeak's maximum frequency, the use of leaves functions to increase the caller's perceived body size, potentially intimidating a potential predator and possibly improving chances of survival.

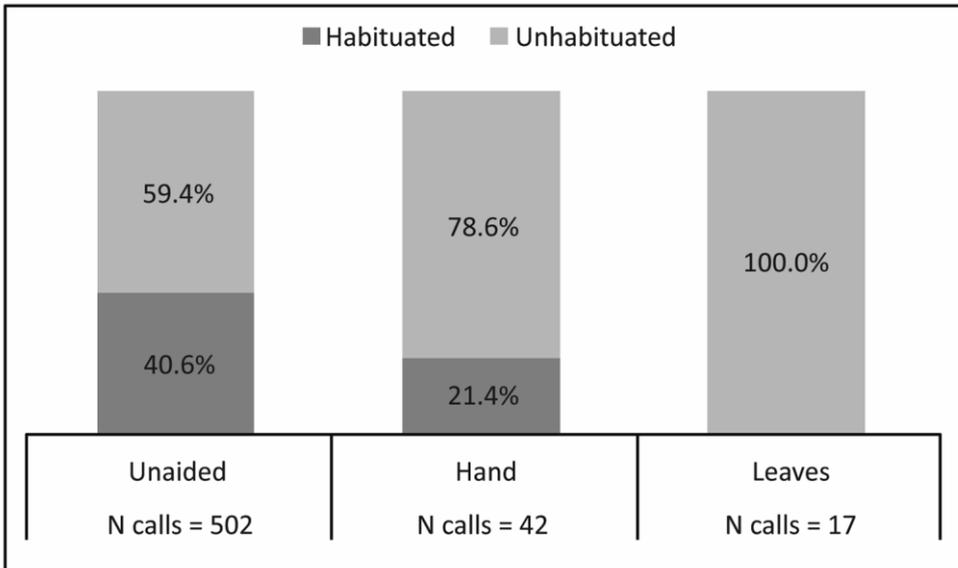


Figure 3. Percentage emitted by habituated and unhabituated individuals for each kiss-squeak form. Light grey, unhabituated; dark grey, habituated.

In the end, the proposed deceptive function of kiss squeak forms can only be evaluated by examining which potential predators elicit the kiss squeaks on leaves and how they respond to them. By following a radio-collared African leopard, Zuberbühler *et al.* (1999) found that after higher alarm call rates by primates the leopard gave up its hiding location and left the hunt significantly faster than would be expected by chance. Schaller (1967) observed similar responses for Asian tigers. Because natural cases of predation on orangutans are rare, playback experiments with potential predators may be needed to elucidate this function.

These results indicate that non-human primate tool-use innovations are more extensive than hitherto appreciated and actually include acoustic communication, where they may play a functional role. This opens the possibility for cultural evolution in non-human primate communication.

Table 1- Kiss squeak bouts comprising two or three forms

Bouts	Individual	<i>Light grey= kiss squeak unaided, dark grey= kiss squeak with hand, white with border= kiss squeak on leaves</i>						
1	Fugit	7	1	2	1	5	1	12
2	Henk	1	4					
3	Jinak	5	1	2				
4	Juni	5	1	4				
5	Kay	2	1	2				
6	Kay	5	1					
7	Kay	8	4	1				
8	Kay	4	2	6				
9	Kondor	9	1	5	1	3		
10	Kondor	2	1					
11	Kondor	2	3	2				
12	Mindi	3	1	2				
13	Mindi	2	1	6				
14	Mindi	6	1	1	1			
15	Mindi	4	6	8				
16	Mindi	2	1	3				
17	Mindi	5	2	5				
18	Mindi	1	4	1				
19	Ucok	2	4	2	3	3	10	
20	Ucok	1	3					
21	SAM	1	2					
22	SAM	2	1	4	1	1		
23	SAM	1	9	5	1			
24	SAM	2	4	4	11	5		
25	SAM	4	2	2	1	1	4	1
26	SAM	15	4	2	1	3		
27	Rambo	1	1	4	1	1	1	3
28	Rambo	2	3					
29	Rambo	4	6					

30	Rambo	2	1	4	4	2	2							
31	Rambo	1	3											
32	Sultan	1	2	5	3	2	7	5	1	6	1	1	1	
33	Sultan	1	1											
34	Sultan	3	2											

---

### *Acknowledgements*

We are grateful to LIPI for authorization to carry out research in Indonesia; BOS for permission to work at Mawas; UNAS for acting as a sponsor; Jo Kolk Foundation, Netherlands Organisation for Scientific Research (NWO), The Leakey Foundation, the National Geographic Society and Utrechts Universiteitsfonds for financial support; Han de Vries and Liesbeth Sterck for (statistical) advice and Tine Geurts and Maria van Noordwijk for the supplemental video data; three anonymous reviewers for suggestions on a previous version of this manuscript. The authors have declared that no competing interests exist.



**Population-specific use of the same tool-assisted alarm sound between two wild orangutan populations (*Pongo pygmaeus wurmbii*) denotes functional arbitrariness**

Adriano R. Lameira, Madeleine E. Hardus, Kim J.J.M. Nouwen, Eva Topelberg, Roberto A. Delgado, Berry M. Spruijt, Elisabeth H.M. Sterck, Cheryl D. Knott and Serge A. Wich

*Abstract*

Arbitrariness is an elementary feature of human language, yet seldom an object of comparative inquiry. While *arbitrary signals* for the same function are relatively frequent between animal populations across taxa, the same signal with *arbitrary functions* is rare and it remains unknown whether it may involve animal call production. To investigate this question, we examined a particular orangutan alarm sound – the kiss-squeak – and two variants – hand and leaf kiss-squeaks. At Tuanan (Central Kalimantan, Indonesia), these variants are correlated with perceived threat, and are hypothesized to increase the conveyed body size of the sender. We examined the use of these variants in the same context in another population of the same sub-species and with partially similar habitat (Cabang Panti, West Kalimantan, Indonesia). Similar analyses at the latter site provided dissimilar results. Hand kiss-squeaks were extremely rare and leaf-use neither enlarged conveyed body size nor related to perceived threat, indicating functional discontinuity between the two sites which implies functional arbitrariness. These results validate the view that these orangutan sound variants are culturally transmitted and reconcile the role of gestures and calls within evolutionary theories based on common ancestry for music and language. Instruments may have been important components of a multimodal language ancestor.

*To be submitted*

## Introduction

Cultural variation in behaviour between animal populations has been suggested across taxa and across behavioural domains, such as the material, foraging, communicative and social domains (Sapolsky 2006). Geographic patterns in behaviour within the communicative domain, for instance, illustrate three major types of cultural variants that may occur across the animal kingdom. First, a signal (with its respective function) may be present in one population but absent in another population (birds: Catchpole and Slater 2003; bats: Davidson and Wilkinson 2002; Morrice et al. 1994a; pinnipeds: Ralls et al. 1985; cetaceans: Ford 1991; Weilgart and Whitehead 1997; non-human primates: Hopkins et al. 2007; Marshall et al. 1999; van Schaik et al. 2003; van Schaik et al. 2006). Second, acoustically different signals with a similar function may be present in different populations (birds: Wright 1996; cetaceans: Noad et al. 2000; non-human primates: van Schaik et al. 2003; Wich et al. 2012). Such examples imply that the signal's *acoustic structure* is arbitrary. In other words, there is no particular relationship between the signal's internal/external determinants (i.e. what actually triggers the signal) and its acoustic structure. The signal's acoustic structure may take in these cases different forms within the limitations of the organism's anatomical structures involved in call production and their respective motor control. Third, the same signal may have different functions in different populations, implying that the signal's *function* is arbitrary. That is, there is no particular relationship between the signal's internal/external determinants and its potential function. However, in contrast to the first two types of cultural variants, there is much less evidence for arbitrary function in animal signals and, as far as we know, this may be restricted to great ape cultural variants (Boesch 2003). Even though arbitrariness is a feature of advanced animal communication systems, pervading human language, the term has been applied in varied ways across the literature (Bonnie et al. 2007; Owren and Rendall 2001; Thornton and Malapert 2009) and it has been relatively little studied. However, the distinction of arbitrariness in relation to a signal or to its function is heuristically relevant, as demonstrated by linguistics' and semiotics' terminology (Saussure et al. 1983); the terms "signifier" and "signified" allow the independent analyses of a signal itself (i.e. its motor action and respective acoustic structure) and its function (i.e. the signal's how, when and why), respectively. Accordingly, this distinction may also offer new understanding about arbitrariness as an object of study in animal species. From

a comparative perspective and considering that human language, and the unparalleled cultural variation therein, is founded on arbitrariness, this approach may allow the identification of possible analogies and/or homologies across animal taxa in signal arbitrariness and provide new clues on language and speech evolution.

A remarkable case of arbitrary function described in animal communication is chimpanzee leaf-clipping (Boesch 2003). Leaf-clipping, the biting of a leaf into pieces to produce a ripping sound without eating the leaf, is used differently by chimpanzees of two different populations. In one population leaf-clipping is functionally used for courtship and in the other for play (Boesch 2003). In a third population leaf-clip is absent. Despite these differences have remained essentially descriptive, they have led some authors to suggest that this behaviour represents a cultural variant (Whiten et al. 1999). This signal represents an object-assisted acoustic gesture, that is, a gesture that produces a sound with the aid of an external object. The question remains: may animal signals involving call production show arbitrary functions as well?

To investigate this possibility we examined an orangutan sound (*sensu* Lameira et al. 2010), the kiss-squeak, a universal (i.e. present at all sites where orangutans have been studied) alarm sound produced by a sharp intake of air through pursed lips (Hardus et al. 2009a). In some, but not all, populations the kiss-squeak is executed (van Schaik et al. 2003) by positioning a hand or a hand with leaves (acting as tools) in front of or against the lips during production (Hardus et al. 2009b). Correspondingly, kiss-squeaks unaided are considered innate (i.e. proper production is not dependent in auditory feedback and/or experience (*sensu* Lameira et al. in press-b)), and hand and leaf kiss-squeaks are suggested to represent cultural variants (van Schaik et al. 2003; van Schaik et al. 2006; van Schaik et al. 2009). A recent description has shown how, when and why orangutans at Tuanan, Central Kalimantan, use these alarm sounds (Hardus et al. 2009b). The positioning of a hand and leaves on their lips progressively lowers the maximum decibel frequency (Hz) of the sound, but does not alter other sound characteristics. Orangutans produce these modified kiss-squeaks more often when confronted with perceived threats. Because the maximum frequency of unaided kiss-squeaks is negatively correlated with body size, orangutans seem to use modified kiss-squeaks to functionally deceive a potential predator by conveying a larger body size, by lowering the sound's frequency with a hand or further lowering the sound's frequency with the use of leaves. Accordingly, at Tuanan, the distinct variants of kiss-squeaks may be said to comprise a graded three-pronged system (i.e.

unaided/hand/leaves) that functionally conveys body size and relates to perceived threat; namely, the more threatening the circumstance, the lower the kiss-squeak's frequency.

Here, we examine potential function and correlates of kiss-squeaks with the same modifiers (hand and leaves) under the same context (towards humans) in a different orangutan population from West Kalimantan – Cabang Panti – of the same sub-species (*Pongo pygmaeus wurmbii*) and with partly overlapping habitat type (i.e. peat swamp). Our fundamental aim is to determine whether modified kiss-squeaks at Cabang Panti function to convey body size and whether they relate to perceived threat, as in the Tuanan population. If these aspects differ between the two populations under the same context, then this sound will likely represent a signal involving call production with arbitrary function.

## *Materials and methods*

### *Study Site*

Kiss-squeaks were recorded from wild orangutans at the Cabang Panti research station (1°13' S, 110°07' E) in the Gunung Palung National Park, West Kalimantan, Indonesia. The study area consists of seven distinct habitat types, including peat swamp (Johnson et al. 2005). This orangutan sub-species also comprises the population at Tuanan research station (2°09' S, 114°26' E), Central Kalimantan, Indonesia, composed of peat swamp forest. Orangutan kiss-squeaks in Cabang Panti were recorded from February to October 2010. The stations as separated by approximately 490 Km.

### *Data Collection and Data Analyses*

At Cabang Panti, twenty-seven identified individuals were followed during a total of 1520.6 hours using focal-animal sampling, comprising 5 age-sex classes (Hardus et al. 2009b), and kiss-squeaks were recorded from 21 of these individuals: immature/adolescents (n = 6; 48/45/18/17/15/6), nulliparous females (n = 2; 98/6), parous females (n = 7; 21/16/12/10/8/6/2), unflanged males (n = 2; 38/11), and flanged males (n = 4; 218/65/33/21). From these individuals, a sub-set of 12 individuals also emitted leaf-squeaks (19/13/11/10/8/4/3/3/2/1/1/1). Kiss-squeaks produced by a focal orangutan and/or its associates were recorded opportunistically throughout the day with

a Marantz Analogue Recorder PMD660 and a ZOOM H4next Handy Recorder with a RØDE NTG-2 directional microphone. The technique (i.e. unaided, hand or hand with leaves) and context (e.g. towards observers) of all kiss-squeaks was noted, using binoculars (magnification power: 10, objective diameter: 25 mm) when necessary. Whenever the technique used could not be observed directly with confidence, these recordings were not considered. This corresponded approximately to 536 recordings (i.e. 30%) out of 1800 recordings. Recordings were transformed into spectrograms according to Hardus et al. (2009b), using Raven (2003, Cornell Lab of Ornithology, NY, USA). To compare the acoustic structure of the different kiss-squeak techniques, the following spectrogram variables were measured: maximum frequency (Hz), maximum power (dB) and duration (s). Maximum frequency is the frequency with the highest energy emitted in a sound. Maximum power is the energy of the maximum frequency (i.e. loudness). Duration represents the time period between the start and the end of the sound. To control for distance to focal, volume setting, and acoustic environment during recordings, maximum power was analysed only within recording bouts (Hardus et al. 2009b). Because kiss-squeaks are brief and noisy *voiceless* calls (i.e. sounds sensu Lameira et al. 2010), the number of measurable acoustic variables is limited. Hardus et al. (2009b) used the exact same variables, allowing a direct and accurate comparison of the sounds between Cabang Panti and Tuanan populations. All aspects of acoustic measures were kept identical to Hardus et al. (2009b) to allow for comparisons. Only kiss-squeaks emitted towards observers were considered, following Hardus et al. (2009b) to avoid any potential contextual biases in the analyses.

Unhabituated orangutans were those with circa 100 observation hours since 2008 (C.D. Knott, unpublished data) *and* that could not be followed during the day without showing signs of disturbance, expressed in flight and displays. Hence, these individuals probably perceived humans as a potential threat. Individuals were considered habituated with more than 100 observation hours since 2008 (C.D. Knott, unpublished data) *and* that could be followed without showing signs of disturbance. These conditions were similar to those met at Tuanan (Hardus et al. 2009b).

Results were qualitatively compared with those generated by data collected during 2510 observation hours according to the same general protocol at Tuanan (Hardus et al. 2009b). At both populations kiss-squeaks were emitted towards other orangutans and other disturbances on the ground (e.g. snakes, sun bears; not considered here), in

addition to humans. Non-parametric statistical tests were conducted using IBM SPSS 19 (2010, SPSS, Inc.), using level of significance at 0.05.

### *Results and Discussion*

At Cabang Panti, kiss-squeaks unaided and kiss-squeaks on leaves were frequently emitted (0.47 and 0.055 kiss-squeaks/hour respectively, i.e. 714 and 84 sounds over 1520.6 follow hours). However, during the same period, only 1 hand kiss-squeak was observed (i.e. < 0.001 kiss-squeaks/hour). Hand kiss-squeaks were thus virtually absent at Cabang Panti while, at Tuanan, hand kiss-squeaks were emitted more often than leaf kiss-squeaks (i.e. 0.036 vs. 0.011 kiss-squeaks/hour, respectively; (Hardus et al. 2009b)). Observed rates of unaided kiss-squeaks were fundamentally the same at Tuanan and Cabang Panti (i.e. 0.47 and 0.46 kiss-squeaks/hour at Cabang Panti and Tuanan, respectively). The observed rate of leaf kiss-squeaks at Cabang Panti was five-fold the respective rate at Tuanan (0.055 and 0.011 kiss-squeaks/hour, respectively). The rarity of hand kiss-squeaks at Cabang Panti did not allow the inclusion of this technique in subsequent analyses. These results show that, in contrast to the three-pronged functional system of kiss-squeaks at Tuanan, the Cabang Panti orangutans only use a two-pronged kiss-squeak system (i.e. unaided/leaf kiss-squeaks).

To assess the functional use of unaided and leaf kiss-squeaks at Cabang Panti, we compared the production rates of habituated and unhabituated orangutans towards humans. At Cabang Panti the two kiss-squeak techniques were not produced at different rates by habituated and unhabituated individuals (Mann-Whitney  $U$  test: unaided and on leaves:  $U = 11.0$ ,  $N_{\text{habituated}} = 5$ ,  $N_{\text{unhabituated}} = 7$ ,  $P = 0.343$ ; Figure 1). This markedly contrasts with Tuanan where only unhabituated individuals, and not habituated, produced leaf kiss-squeaks towards humans (Hardus et al. 2009b) (Figure 2). Thus, the two-pronged kiss-squeak system at Cabang Panti did not relate to perceived threat, as the kiss-squeak system did at Tuanan. Even though unhabituated individuals at Cabang Panti showed signs of disturbance in the presence of human observers (e.g. expressed in flight and displays), this was not translated in their use of leaf kiss-squeaks, and habituated orangutans at Cabang Panti also used leaf kiss-squeaks.

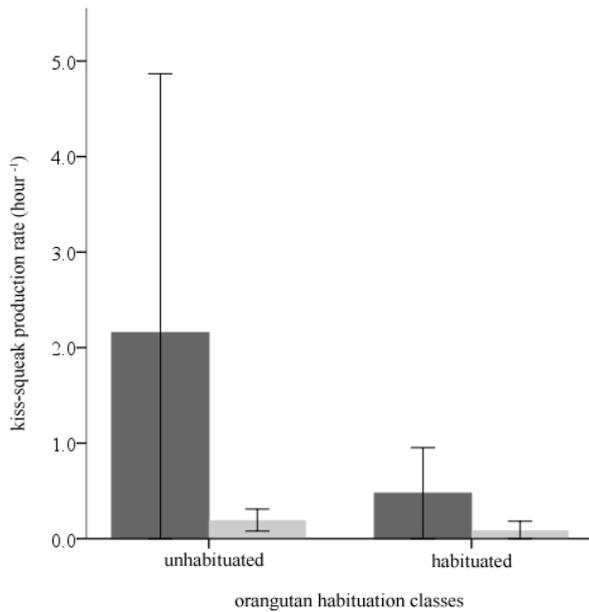


Figure 1. Production rates of unaided kiss-squeaks (dark grey) and leaf kiss-squeaks (bright grey) during total observation hours per individual by (A) habituated ( $N = 5$ ) and unhabituated orangutans ( $N = 7$ ), and by (B) habituated ( $N = 5$ ), semi-habituated ( $N = 3$ ) and unhabituated orangutan ( $N = 4$ ). Error bars: +2 s.e.

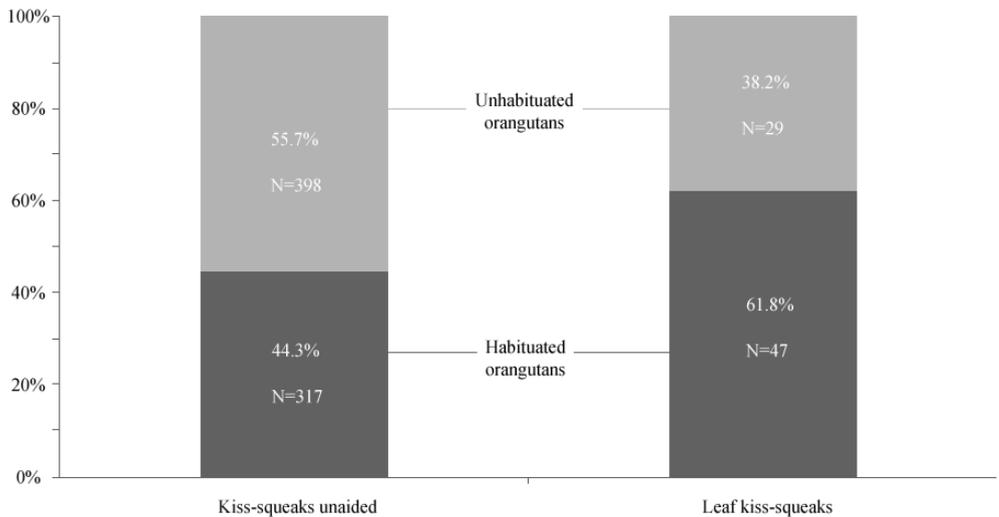


Figure 2. Percentage emitted by habituated and unhabituated individuals for unaided kiss-squeaks and leaf kiss-squeaks.

This difference may have resulted from a lack of relationship between weight/body size (assessed via differences between age-sex classes) and the unaided kiss-squeaks' frequency (Hz) at Cabang Panti, in contrast to Tuanan (Hardus et al. 2009b). This may preclude individuals from functionally simulating body size enlargement through tool use at Cabang Panti. However, similar to Tuanan, a negative relationship between unaided kiss-squeaks' frequency across age-sex classes was found at Cabang Panti (Kruskal-Wallis test:  $H_4 = 270.585$ ,  $P < 0.001$ ; followed by a post hoc test:  $P < 0.001$  between immature/adolescents and all the other classes and between flanged males and all the other classes, Figure 3). That is, the larger body size, the lower kiss-squeaks' maximum frequency. Exact similar results were obtained at Tuanan (Hardus et al. 2009b). This was not unexpected, since this relationship likely reflects a link between the size of source structure (i.e. lips) and the signal's acoustic signature, similar to calls produced at the vocal folds or tract (Fitch 1997).

Number of observed individuals per sex-age class that emitted unaided kiss-squeaks did not differ between populations (Wilcoxon signed-ranks test:  $Z = -0.552$ ,  $N = 5$ ,  $P = 0.581$ ). Number of calls recorded per sex-age class did not differ between populations (Wilcoxon signed-ranks test,  $Z = -0.944$ ,  $N = 5$ ,  $P = 0.345$ ). Accordingly, it is unlikely that differences between datasets affected these results.

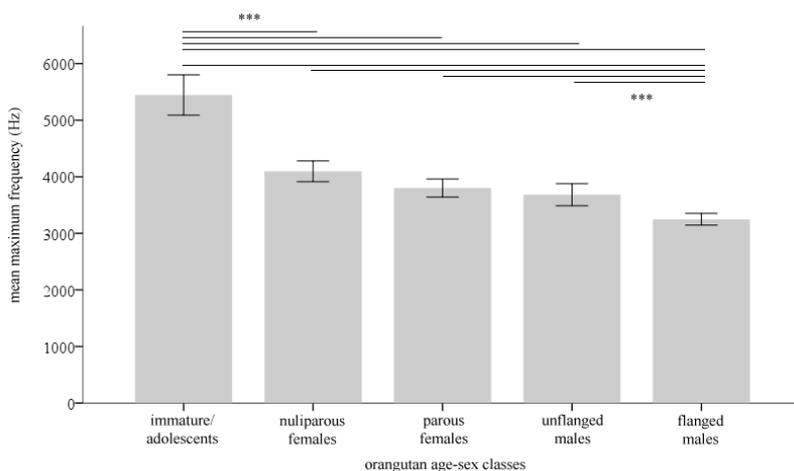


Figure 3. Mean maximum frequency (Hz) per orangutan age-sex class of the kiss-squeak unaided. Immature/adolescent,  $N = 6$ ,  $N_{\text{calls}} = 149$ ; Nulliparous female,  $n = 2$ ,  $N_{\text{calls}} = 104$ ; Parous female,  $N = 7$ ,  $N_{\text{calls}} = 75$ ; Unflanged male,  $N = 2$ ,  $N_{\text{calls}} = 49$ ; Flanged male,  $N = 2$ ,  $N_{\text{calls}} = 337$ . Error bars: +2 s.e. \*\*\*  $p = 0.001$

To investigate the potential acoustic effect of leaf-use on kiss-squeak characteristics at Cabang Panti, the acoustic characteristics of kiss-squeaks unaided and on leaves were analysed. Leaf kiss-squeaks showed a slightly lower maximum frequency than kiss-squeaks unaided produced by the twelve individuals from whom there were recordings of both variants, but this difference did not reach statistical significance (Wilcoxon signed-ranks test:  $T = -1.609$ ,  $N = 12$ ,  $P = 0.108$ ; Figure 4). The difference between the median maximum frequencies of the kiss-squeak unaided and on leaves at Cabang Panti was less than 250Hz (table 1), while at Tuanan this difference surpassed 2000Hz (Hardus et al. 2009b). While positioning an item in front of the mouth during emission can lead to the decrease of a sound's frequency, sufficient proximity or contact between the tool and the lips may be crucial for large manipulations of sound frequency (Lameira et al. in press-a). Thus, at Cabang Panti, the manipulation of leaves was not appropriate to cause a significant acoustic effect and, contrary to Tuanan, did not functionally convey a larger body size.

Table 1. Quantitative acoustic differences between kiss-squeaks unaided and on leaves at Cabang Panti. Median values are presented with 25 and 75 percentiles between brackets.

	$N_{\text{ind./calls}}$	maximum frequency (Hz)	duration (s)	maximum power (dB)
kiss-squeak unaided	21/714	3273 (2842.4, 4134.4)	0.493 (0.398, 0.618)	99.2 (96.4, 96.4)
kiss-squeak on leaves	12/76	3057.7 (2368.7, 4177.5)	0.562 (0.419, 0.739)	98.2 (96, 102.4)

Moreover, kiss-squeaks unaided and on leaves produced by the twelve individuals from whom there were recordings of both variants did not differ significantly in duration (Wilcoxon signed-ranks test:  $T = -0.863$ ,  $N = 12$ ,  $P = 0.388$ ). Maximum power also did not differ between the two kiss-squeak techniques (Paired  $t$  test:  $t_{26} = 0.249$ ,  $P = 0.805$ ; table 1). Thus, no acoustic differences were found between kiss-squeaks unaided and on leaves in the Cabang Panti population.

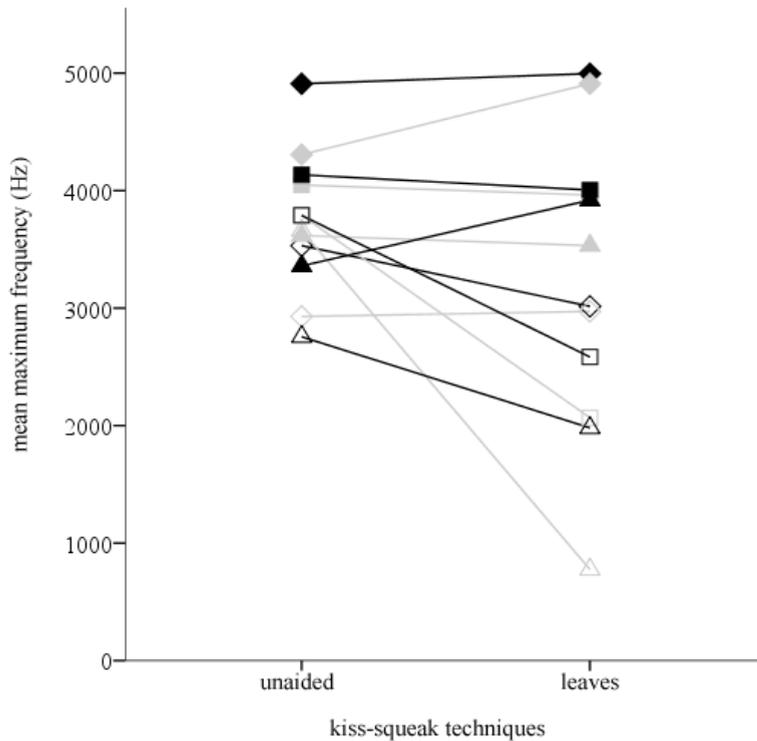


Figure 4. Mean maximum frequency (Hz) of unaided and leaf kiss-squeak for twelve orangutans. Full black diamonds, Aminah (adolescent female, n = 17 unaided vs. 2 leaf kiss-squeaks); full grey diamonds, Asny (parous female, n = 6 vs. 11); blank black diamonds, Beth (parous female, n = 21 vs. 19); blank grey diamonds, Fajar (flanged male, n = 217 vs. 8); full black squares, Walimah (nulliparous female, n = 98 vs. 10); full grey squares, Dewi (nulliparous female, n = 6 vs. 3); blank black squares, Unknown 1 (adolescent male, n = 17 vs. 1); blank grey squares, Codet (flanged male, n = 21 vs. 3); full black triangle, XL (flanged male, n = 33 vs. 4); full grey triangle, Bibi (parous female, n = 12 vs. 1), blank black triangle, Prabu (flanged male, n = 65 vs. 13), blank grey triangle, Indi (parous female, n = 8 vs. 1).

In sum, at Cabang Panti leaf kiss-squeaks towards humans did not function to convey body size enlargement and its use did not relate to perceived threat, despite being produced relatively frequently. This contrasts with their use at Tuanan, where in the same context, leaf-kiss squeaks functionally convey a larger body size than individuals actually have and where they relate to perceived threat (Hardus et al. 2009b). While relatively small individual sample sizes are unavoidable in studies on wild orangutans populations

(who live in dispersed and low-density communities; (Husson et al. 2009)), the results of this study essentially demonstrate qualitative, rather than quantitative, differences between the two populations. In contrast with Tuanan, hand kiss-squeaks were absent at Cabang Panti and habituated individuals produced leaf kiss-squeaks towards observers. Relevantly, these marked differences were not caused by dataset differences.

Effects due to habitat differences between populations are unlikely to have affected the results. First, individuals at Cabang Panti used leaves indiscriminately of plant species, similarly as in Tuanan (Hardus et al. 2009b), simply using any leaves within reach at the moment of disturbance. Second, the possibility of differences in the type of leaves randomly chosen between sites is further unlikely as the sites are partly composed by overlapping habitat types, and thus by (ecologically) similar plant species. Third, compared with factors such as proximity lips-leaves during kiss-squeak emission, leaf type may only bear trivial acoustic effects. Forth, partly similar habitat types are expected to solely impose relaxed selection pressures for acoustically distinct alarm calls. Fifth, the possibility of different influences by habitat acoustics on the maximum frequency of orangutan calls is further unlikely, namely at close- and middle-range (Lameira and Wich 2008). Sixth, production of hand kiss-squeaks does not depend directly on any ecological element. Thus, although we do not exclude the possibility of some effects by cryptic ecological factors, these are most likely artifactual.

Human effects are unlikely to have affected the results. Human observers at both sites collect data according to standardized methods (van Schaik 1999) which are similar across orangutan field sites. Thus, we do not suspect that observers behaved in any different way between sites that could have resulted in orangutans also behaving differently. Nevertheless, we are unaware of any possible means by which observers' behavior could parsimoniously produce the specific results obtained. This similarly applies to potential differences in orangutans' habituation level between sites.

Due to the close genetic relatedness between the two orangutan populations, involving the same sub-species and with a recent mean coalescence date of less than 176ka (Arora et al. 2010), it is unlikely that genetic differentiation is the source of the differences found in this study. To our knowledge, there is no known genetic process which codes and influences differently a particular behaviour when conducted under the same context (and thus probably under the same affective state (*sensu* Owren and Rendall 2001) of the individual) with or without tools and that can be expressed

differently within a particular sub-species. This would seem particularly improbable in the specific case of animal alarm calls which are expected to be under high evolutive inertia due to their importance for individual survivorship. Indeed, the geographic pattern of presence/absence of leaf kiss-squeaks is not sufficiently explained by genetic differentiation between populations nor environmental differences (Krützen et al. 2011). This concurs with evidence demonstrating that genetic differentiation does not sufficiently explain the geographic pattern of other orangutan calls (Wich et al. 2012). On the other hand, orangutans have been empirically shown to be able of tool use observational learning (Dindo et al. 2011). Moreover, proficient use of tools in wild primates is assumed to depend on practice (e.g. de A. Moura and Lee 2004; Inoue-Nakamura and Matsuzawa 1997). Altogether, this study concurs with the literature suggesting that leaf kiss-squeaks may be socially learned within each population and that they represent cultural variants (Hardus et al. 2009b; Krützen et al. 2011; van Schaik et al. 2003). Future work, namely via captive experiments, may assess this in more detail.

This study is, as far as we know, the first to directly demonstrate that an animal signal involving call production does not present a continuous and uniform function across populations. Although, in a broad sense, orangutan leaf kiss-squeaks constitute an alarm sound at both Tuanan and Cabang Panti, this variant's features and correlates at the latter site do not satisfy the conditions in particular for the specific deceptive function proposed at the former site (Hardus et al. 2009b). Similarly to the known example of chimpanzee leaf-clipping which maintains a broad function of eliciting body contact, the specific functions of play or mating vary geographically (Boesch 2003). This demonstrates that the particular function attributed to the signal may be present or absent across communities, implying arbitrariness. Accordingly, although the identification of the specific function of leaf kiss-squeaks at Cabang Panti remains uncertain, it is the absence of Tuanan's function at Cabang Panti that expresses the arbitrary nature of whichever functions may be potentially attributed to this signal. Future experiments using playbacks could investigate the function of leaf kiss-squeaks at Cabang Panti, however, this may prove challenging as the acoustic features and correlates of leaf kiss-squeaks at this site provide few clues. In fact, as leaves serve no apparent *acoustic* function as a tool, this signal could be considered a ghost or superstitious behaviour. This means that the signal may have been maintained within the population but not its function. On the other hand, the function of the tool could have transferred domain to become essentially a visual enhancement to

the kiss-squeak sound. The absence of hand kiss-squeaks could offer support to both these possibilities.

The differences found between these orangutan populations differ from any other types of variation described in animal signals, at least, in three fundamental ways. Leaf kiss-squeak differences are observed in the absence of any contextual variation (cf. Notman and Rendall 2005), they do not constitute one of the species' innate/universal calls, and they involve tool use. Interestingly, the two examples of great ape signals with arbitrary function – chimpanzee leaf-clipping (Boesch 2003) and orangutan leaf kiss-squeaks (this study) – are directly connected with and dependent on tools. This may suggest that the use of a manipulated external element (i.e. tools), executed in synchrony with a communicative signal, may facilitate an operative uncoupling at the neurocognitive level between a signal's internal/external determinants and its attributed function (cf. Lameira et al. in press-a). This finding is relevant to the theory of human language evolution, since gestural and acoustic models of language evolution are commonly seen as mutually exclusive (Lameira et al. in press-a). Our results indicate that, although animal calls, namely great ape alarm calls, are considered affect-based (sensu Owren and Rendall 2001) and innate (sensu Lameira et al. in press-b) due to their importance for the individual's survival, advanced communicative features, such as arbitrariness, may be brought about into the acoustic domain when gestures and calls are synchronous and conjoint (Lameira et al. in press-a). These results are in concordance with “musilanguage” models of human language and speech evolution (Brown 1999) in that they conjure a multimodal common precursor to human language and music, but also suggest that, within this evolutionary model, the role of tools as “instruments” may have been much more relevant than previously assumed.

Evidence of signals with arbitrary function in great apes suggests that primatologists, evolutionary anthropologists, acoustic biologists, and scholars interested in comparative biology may benefit from the use of the linguistic and semiotic terms “signifier” and “signified” to independently refer to a signal and its functional use, respectively. Signals involving call production with arbitrary function may date back to the homininae-ponginae evolutionary split (i.e. 9-13 MYA; (Hobolth et al. 2011)) and may have paved the way towards the use of words with arbitrary meaning in human speech.

### *Acknowledgments*

This study was financially supported by Fundação para a Ciência e Tecnologia (SFRH/BD/44437/2008), Wenner-Gren Foundation for Anthropological Research, Dr. J.L. Dobberke voor Vergelijkende Psychologie, Lucie Burgers Foundation for Comparative Behaviour Research, Schure-Beijerinck-Popping Fonds, Ruggles-Gates Fund for Anthropological Scholarship of the Royal Anthropological Institute of Great Britain and Ireland, and Fundação Calouste Gulbenkian. KN and ET were financially supported by Pongo Foundation. Orangutan fieldwork at Cabang Panti was supported by grants to CDK from the National Science Foundation (0936199), the National Geographic Society, the Leakey Foundation and the US Fish and Wildlife Service. We are grateful to Maria van Noordwijk for her help during the preparation of the manuscript. We thank RISTEK, PHKA and BTNGP for research permission in Indonesia, UNTAN for supporting the project, and the field staff of the Gunung Palung Orangutan Project for their assistance. The authors reveal no conflicting interests.

## **Orangutan whistling and implications for the emergence of an open-ended call repertoire: a replication and extension**

Adriano R. Lameira, Madeleine E. Hardus, Bernd Kowalsky, Han de Vries,  
Berry M. Spruijt, Elisabeth H. M. Sterck, Robert W. Shumaker and Serge A. Wich

### *Abstract*

One of the most apparent discontinuities between non-human primates' call repertoire and human speech concerns the number of calls. The non-human primate repertoire is essentially fixed from birth to a limited number of calls that individuals produce properly without necessary auditory/motor input and without the possibility of expanding the repertoire. By contrast, humans have an open-ended repertoire: they can augment call number through acoustic learning of vocalizations and sounds not previously produced. While learning vocalizations requires voluntary laryngeal control and is largely absent in non-human primates, *sound learning* (not requiring laryngeal control) has hardly been investigated. A recent case of sound learning described the development of human whistling by an orangutan. In the current study, experiments were replicated with another orangutan and variation regarding qualitative and quantitative characteristics was analyzed among three whistling orangutans. The results corroborate the hypothesis that whistling represents a socially learned sound, and demonstrate that orangutans exert voluntary motor control over their lips, tongue, and respiration. Whistling learning conditions were likely based on undirected/unrewarded tutoring. Since orangutans are capable of sound learning, this implies that a portion of the components for human speech control and learning may have been in place before the homininae-ponginae evolutionary split.

*Accepted by Journal of the Acoustic Society of America*

## Introduction

After almost one and a half centuries of theorizing, unanswered questions remain on how human speech could have evolved from the call repertoire of *non-human primates* (hereafter *primates*), which is thought to be representative of the ancestral state (Fitch 2010; Christiansen and Kirby 2003a; Tomasello 2008). One of the main questions revolves around the transition from a primate call repertoire comprised of a limited number of components (i.e. close-ended), that essentially develops reliably without relevant environmental input (hereafter *innate*), to one comprised of a potentially limitless number of components (i.e. open-ended), that is essentially learned, such as human speech. A crucial step for this evolutionary transition to occur is the emergence of acoustic learning, the ability to learn novel *calls* (i.e. vocalizations and/or sounds; sensu Lameira et al. 2010; Hardus et al. 2009a; Owren et al. 2010; Hopkins et al. 2011; Hardus et al. 2009b) from others and to invent novel calls (Ramsey et al. 2007). An individual becomes an effective acoustic learner when auditory and motor experience play a central role in voluntary control over call production and flexible call acoustics (Owren et al. 2010). In other words, at least when some of the calls in an individual's repertoire are derived from reception-first call development, as defined by Owren et al. (2010).

Acoustic learning of novel calls in the animal kingdom, by either individual or social learning, is dependent on three major systems: (1) the respiratory, (2) phonatory and (3) filter systems (Janik and Slater 2000). This ability is not uncommon in animals and is present in several distinct taxa, such as birds and non-primate mammals (Catchpole and Slater 2003; Janik and Slater 1997). Several reference studies have established a solid framework for its investigation (e.g. Fitch and Hauser 2003; Boughman and Moss 2003). However, acoustic learning of novel calls is assumed to be largely absent in primates (Seyfarth and Cheney 2010). This is essentially because, in contrast to the excellent bird and non-primate mammal examples, *social* acoustic learning is virtually absent in primates (Fitch 2010). Nevertheless, recent studies present evidence that primates can exert more flexibility over various aspects of their innate call system than has traditionally been assumed (e.g. Lemasson et al. 2005; Ouattara et al. 2009a, 2009b; Arnold and Zuberbühler 2006; Clarke et al. 2006; Slocombe and Zuberbühler 2007; Slocombe and Zuberbühler 2005b, 2005a; Stephan and Zuberbühler 2008; Hardus et al. 2009b; Wich and de Vries

2006; Seyfarth et al. 1980; Marshall et al. 1999; de la Torre and Snowdon 2009; Brumm et al. 2004).

Investigating Janik and Slater's (2000) perspective offers the possibility to understand which exact structures and respective abilities may or may not be under voluntary motor control in primates. Because these systems differ anatomically and functionally between the animal taxa that are able of acoustic learning, these systems are here described in the following way for primates and humans: (1) diaphragm and abdominal muscles, (2) laryngeal muscles (namely vocal-fold muscles), and (3) tongue, lip and jaw muscles, respectively. To learn a new call, a primate must exert voluntary motor control over at least one of these systems. Tongue and lip control may suffice to learn clicks and smacks (i.e. sounds, *sensu stricto*, see above), and in conjunction with diaphragm and abdominal control may suffice to learn sputters, fizzles and whispers (also sounds), whereas the voluntary control of all three muscle systems is required for learning vocalizations (*sensu stricto*, see above). Accordingly, learning novel sounds does not require laryngeal control but it allows the expansion of an individual's innate call repertoire. Despite its potential importance for theories on speech evolution, sound learning in primates has received little attention, compared to vocal learning (e.g. Fitch and Hauser 2003; Boughman and Moss 2003). For instance, individual differences represent one of the basic levels of acoustic variation in primate vocalizations; this is well described in many primate vocalizations and is commonly thought to be an effect of size and shape differences of the vocal folds and/or tract between different individuals (e.g. Lameira and Wich 2008; Spillmann et al. 2010; Mitani et al. 1996; Rendall et al. 1998; Owren et al. 1997). However, since primate sounds are not necessarily affected by vocal fold/tract physiognomy, is individual variation present in primate sounds? Do primate sounds provide sufficient acoustic range that producers may encode information, for example, about their identity? The answers to these questions can help clarify whether selective pressures are distinct between primate sounds and primate vocalizations, or whether sound learning and vocal learning may have played roles of difference importance during human evolution. Did sound learning represent a transition towards more complex vocal learning in the process of speech evolution, or did they both function together as equal counterparts?

A recent example of sound learning in great apes, namely by an orangutan, involves the spontaneous acquisition (i.e. without training) of a human sound (i.e. via

social learning) – whistling (Wich et al. 2009b). Bonnie, housed in a zoo (see Table 1), most likely learned how to whistle by listening to and observing her caretakers. Empirical data verified that she exerted good control over certain acoustic characteristics (duration and number) of her whistles (Wich et al. 2009b), supported by voluntary motor control of oral and respiratory muscles. Whistling by Bonnie was physiognomically produced similarly to whistling by humans, meaning that she whistled during expiration phases using slightly protruded and constricted lips. Whistling is the result of the periodic vortex shedding of the airstream at the lips' opening and so it differs unequivocally from orangutan vocalizations which are produced at the vocal folds or in the vocal tract (e.g. Hardus et al. 2009a; Lameira and Wich 2008). Moreover, whistles also differ from any other orangutan sounds produced at the lips, such as the raspberry (Hardus et al. 2009a), in that they show very high tonality where raspberries lack any. Orangutan whistles and human whistles are equally classified as “orifice tones” and differ from other whistling-like sounds such as those produced, for example, by blowing at the mouth of a bottle at the right angle (i.e. “edge tone”). Whistling is absent in our database comprising the known call repertoire of wild orangutan populations collected over 6 years, including 5 Indonesian populations and over 5000 recordings (Lameira, Hardus and Wich, personal observation; see Hardus et al. 2009a; Wich et al. 2009b). Thus, whistling by orangutans seems to depend greatly on prior sound learning for proper production and can be classified as a learned sound based on reception-first development (Owren et al. 2010). It is worth mentioning that, for instance, dolphins and sparrows are expert “whistle” mimics (Janik 2000; Soha and Marler 2001), but this terminology is ambiguous: dolphins produce whistles by structures of the nasal apparatus and specialized sacs just below the blowhole (Ridgway 2011) and sparrows produce whistles between the trachea and the lungs by membranes on the left side of the syrinx (Lemon 1973), and thus they are not directly comparable to human and orangutan whistles.

Interestingly, since the time of publication of the study by Wich et al. (2009b), new cases of orangutan whistling have been reported with other captive orangutans (Table 1). The main aim of this article is to replicate the work previously conducted with Bonnie, based on the experience that captive orangutans are motivated to participate in interactions with their caretakers with who they have excellent relationships and occasionally with unfamiliar people (Lameira, Hardus, Wich, Shumaker, personal observation). Moreover, this study lists and presents empirical and acoustic data on the

“new” whistling individuals, and it reassesses the conclusions by Wich and colleagues (2009b) in order to gain insights about sound learning in great apes and how it compares to vocal learning. The following features were experimentally investigated: (1) performance in whistle matching, (2) acoustic individual variation in whistling and (3) physiognomic individual variation in whistling. In the event that replication leads to similar results as that of the previous work (Wich et al. 2009b), the findings and implications are more likely to apply to the whole species and not just to one individual alone.

## *Methods*

### *Focal animals*

Nine orangutans were identified in five different zoos who have demonstrated whistling (Table 1). Three pairs of whistling individuals are/were housed together. These animals are of Sumatran (*Pongo abelii*), Bornean (*Pongo pygmaeus*) or mixed descent. The three individuals with backgrounds in the entertainment industry most likely received training to originally acquire whistling. The six individuals without this background did not receive training to acquire whistling, however, some have received subsequent positive reinforcement for whistling and/or have been trained to improve their whistling. These individuals presently whistle by themselves or use whistling as means to capture the attention of humans (see Hopkins et al. 2007).

### *Performance in whistle matching*

Non-invasive experiments were conducted with Ujian (Table 1), following the methods of Wich et al. (2009b). Experimental results with Bonnie can be found in Wich et al. (2009b). We were unable to conduct tests with Katy, Berani, Feliz, Puan and Tilda. The experiments conducted were based on the concept “do as I do”, where a human demonstrator produced a particular call pattern and the response (if any) of the orangutan was then recorded (Wich et al. 2009b). There was no gestural request for whistling from the orangutan. Three different experimental paradigms were presented to Ujian: (1) pseudo-randomized sequences of single and double whistles (4 sessions), (2) pseudo-randomized sequences of single, double and triple whistles (6 sessions, Fig. 1; Supplementary File) and (3) pseudo-randomized sequences of single (i.e. air outflow) and alternated whistles (i.e. air outflow and inflow alternated; 8 sessions). Whistles presented in the paradigm

Single/Alternating were always sequenced air (1) out- (2) in- (3) outflow. The whistles presented in all paradigms differed structurally from the whistles normally produced by Ujian, which are typically multiple whistles (i.e. > 3) in the air in- and outflow with no apparent order. All whistles produced by the human demonstrator were similar to each other and constant in duration, frequency and overall acoustic structure to the greatest extent possible. The human demonstrator typically waited until 5 seconds after producing the model whistle for the start of Ujian's response. Five seconds were allowed for the completion of Ujian's response for double, triple (Fig. 1; Supplementary File) or alternating whistles. Incorrect matching was considered when there was either no or an incorrect response made. After an incorrect or correct whistle match, a new whistle was presented by the demonstrator without necessarily waiting for the full elapse of the 5 seconds. Whistles by Ujian without a prior whistle model by the demonstrator were not considered.

A total of 26 experimental sessions were conducted with Ujian between the 30<sup>th</sup> of March 2010 and 7<sup>th</sup> of April 2010 at the Heidelberg Zoo, Germany. Sessions lasted approximately 15 minutes (circa 80 whistles; min. 15, max. 115), usually twice a day but never more than three times per day, with at least two hours between sessions.

During the first 6 experimental sessions (30<sup>th</sup> and 31<sup>th</sup> of March) Ujian was for the first time exposed to the experimental paradigm Single/Double and Single/Double/Triple (Fig. 1) and the human demonstrator (MH, i.e. unfamiliar demonstrator). Two experimental sessions (4<sup>th</sup> of April) were dedicated to expose Ujian for the first time to the experimental paradigm Single/Alternating. These training sessions were not considered for analyses. During each experimental session, exclusively the periods during which Ujian was paying close attention to the demonstrator by facing her directly and having his gaze at her were considered. Ujian and the human demonstrator were separated by a mesh at all times. Ujian was not food deprived before the experiments, and was rewarded non-differentially with nuts and/or raisins (i.e. independently of his success during the experiments). He was not separated from his consort (Puan) during or between experiments, who shares the enclosure space permanently with Ujian.

Because of the possible non-independence of Ujian's responses to consecutive whistles by the demonstrator, the following procedure was conducted to obtain a subset of statistically non-independent data points (i.e. trials). For each session, we checked by

means of a chi square cross table test whether the response of Ujian was statistically independent of the whistle made by the human demonstrator 4 trials earlier. For all sessions the p-values of the chi-square tests was larger than 0.10 (range: 0.101 – 0.98), except for 3 sessions which yielded p-values of 0.063, 0.027 and 0.079, indicating a possible non-independence. Similarly, we checked for each session whether the response of Ujian was statistically independent of its own response given 4 trials earlier. Here 4 sessions were found with p-values below 0.10 (0.032, 0.094, 0.091 and 0.079), with 2 of these sessions being the same as found before relatively to the demonstrator's whistles. The p-values of all other sessions ranged between 0.18 and 0.96.

Because of the possible non-independence in 5 sessions (p-values below 0.10) these sessions were not included in our analyses of performance in whistle matching (Results, section A). Moreover, only 4<sup>th</sup> trials (i.e. 4<sup>th</sup>, 8<sup>th</sup>, 12<sup>th</sup>, ...) were used from each of the other sessions in our analyses. Thus, out of the total 1395 trials, 282 statistically non-independent trials were used to examine whether Ujian correctly matched the type of whistle produced by the human demonstrator per paradigm and also for all paradigms together. For this purpose chi-square cross table tests was used.

Cohen's kappa (Martin et al. 2007) was used to assess inter-observer reliability. Twenty percent of all trials of each paradigm were pseudo-randomly chosen and classified by a blind rater (i.e. unaware of the concept "do as I do") familiar with orangutan calls produced in the wild but unfamiliar with orangutan whistles (K.J.J.M. Nouwen).

#### *Production and acoustic measures*

To assess inter-individual variation, whistles were analyzed with the most similar acoustic structure as possible available from three orangutans, Ujian ( $N = 100$ ), Bonnie ( $N = 103$ ) and Katy ( $N = 97$ ). Single whistles were used from experimental paradigms with Ujian (see above) and Bonnie (Wich et al. 2009b). Experimental work with Bonnie was analogous to that conducted and described here with Ujian (Single/Double and Short/Long paradigms presented; Wich et al. 2009b). Whistles from Katy were opportunistically recorded between August 2008 and October 2009 while she whistled "privately" or to gather attention from her caretakers. Two main aspects were compared between individuals; qualitative production characteristics and quantitative acoustic parameters. The following four qualitative production characteristics were visually and aurally inspected during whistling: production style (i.e. with constant or modulated frequency), airflow phase (i.e.

outflow or inflow), airflow duration (i.e. short/momentary or long/sustained) and lip protrusion (i.e. symmetrical/human-like (Shadle 1983) or otherwise). The following eight quantitative acoustic parameters were measured from audio recordings using Praat software (Boersma 2001): duration (s), mean pitch (Hz), pitch standard deviation (Hz), jitter (i.e. frequency perturbations), shimmer (i.e. amplitude perturbations), harmonic-to-noise ratio (i.e. HNR, which is a measure of tonality that does not require the presence of harmonics in a signal as the term may suggest; dB), pitch mean absolute slope (i.e. PMAS; Hz/s) and rise time to maximum amplitude (i.e. RTMA; %). Discriminant analysis was conducted to investigate differences across the acoustic parameters between Ujian, Bonnie and Katy, using all acoustic parameters together followed by the leave-one-out cross-validation procedure (which leaves each whistle, one at a time, out of the dataset after which group membership of this whistle is predicted by the discriminant functions derived from the other cases).

### *Apparatus*

Recordings from Ujian were collected with a Marantz PDM660 recorder via its incorporated microphone, approximately 80 cm away from the human demonstrator and Ujian. Recordings from Bonnie were collected with a Tascam HD-P2 digital stereo recorder with a Sennheiser Microphone ME 66, circa 60 cm away from the human demonstrator and Bonnie. Recordings from Katy were collected with a Marantz PDM661 in combination with a Sennheiser Microphone ME 64, between 2 and 5 meters away. The data were statistically analyzed with IBM SPSS 19 (2010, SPSS, Inc.).

### *Results*

#### *Performance in whistle matching*

The experimental paradigm Single/Double, Single/Double/Triple and Single/Alternating were presented to Ujian 328, 421 and 646 times, respectively. Because of the possible non-independence of Ujian's responses to consecutive trials within a session 282 statistically non-independent trials were selected for statistical testing; paradigm Single/Double: 66; paradigm Single/Double/Triple: 93; paradigm Single/Alternating: 123.

These data points were used for statistical testing, but descriptive counts and percentages are also provided for the full data set within straight brackets.

In paradigm Single/Double, 42 of the 43 [196 of 211] single whistles made by the human demonstrator were followed by a single whistle from Ujian, and 21 of the 23 [95 of 117] double whistles made by the human were followed by a double whistle from Ujian (Chi-square test for 2 x 3 cross table;  $\chi^2 = 63.06$ ,  $df = 2$ ,  $P < 0.0001$ ; row categories: single or double whistle by demonstrator, column categories: incorrect response, single whistle or double whistle by Ujian). [For the full data set: in 15 (7%) of the single and 22 (19%) of the double whistles, Ujian responded incorrectly, that is, he either responded with a single whistle upon hearing a double whistle or vice versa, or he did not whistle at all, or he whistled multiple times.]

In paradigm Single/Double/Triple, 26 of the 34 [136 of 164] single whistles made by the human demonstrator were followed by a single whistle from Ujian, 36 of the 40 [143 of 167] double whistles were followed by a double whistle from Ujian, and 15 of the 19 [73 of 90] triple whistles were answered by Ujian with a triple whistle (Chi-square test for 3 x 4 cross table;  $\chi^2 = 153.40$ ,  $df = 6$ ,  $P < 0.0001$ ). [For the full data set: in 28 (17%) of the single, 24 (14%) of the double and 17 (19%) of the triple whistles presented to Ujian, he responded incorrectly, that is, he either responded with a single whistle upon hearing a double or triple whistle or vice versa, or he did not whistle at all, or he whistled more than 3 times.]

In paradigm Single/Alternating, 59 (78%) of the 76 [in the full data set: 319 (80%) of 397] single whistles made by the human demonstrator were followed by a single whistle from Ujian, and 36 (77%) [in the full data set: 203 (81%) of 249] alternated whistles were followed by an alternated whistle from Ujian (Chi-square test for 2 x 3 cross table;  $\chi^2 = 94.71$ ,  $df = 2$ ,  $P < 0.0001$ ). [For the full data set: On 78 (20%) of the single and 46 (19%) of the alternated whistles presented to Ujian, he responded incorrectly, that is, he either responded with a single whistle upon hearing an alternated whistle or vice versa, or he did not whistle at all, or he whistled multiple times.]

Table 1. List of known whistling orangutans

Focal	Gender	Birth year	Species	Study book ID	Housed at*	Background in entertainment industry	Whistle recordings and analysis	Whistling on cue by caretakers	Remarks
Ujian	M	1973	Sumatran	1930	Heidelberg Zoo, Germany	No	Yes	Yes; gestural request (fore-finger pointed upwards).	Positive reinforcement by caretakers (Lameira, Hardus and Kowalsky, personal observation), housed in the same cage with Puan (see below).
Bonnie <sup>a</sup>	F	1976	hybrid	1545	National Zoo, US	No	Yes	Yes; verbal request.	No whistling reinforcement by caretakers (Lameira and Hardus, personal observation)
Katy	F	1988	hybrid	2248	Indianapolis Zoo, US	Yes	Yes	Yes; verbal request or whistling.	Positive reinforcement by caretakers (Shumaker, personal observation)

Berani	M	1993	Sumatran	2604	Audubon Zoo, USA	No	No. Whistling confirmed by M. Fernandez (pers. comm.)	Yes	Positive reinforcement by caretakers (M. Fernandez, personal communication), housed with Feliz (see above).	whistling by
Buda/Sumi	M	1970 (-1982)	hybrid	1098	n.a. (privately owned)	Yes	No. Whistling confirmed by R. Porec (pers. comm.)	Probably yes	No further info available.	
Chantek	M	1977	hybrid	1617	Atlanta Zoo, USA	No	No. Whistling confirmed by H. L. Miles (1996)	Yes	Enculturated individual (see Miles et al. 1996)	individual

Feliz	F	1988	Sumatran	2157	Audubon Zoo, USA	No	No. Whistling confirmed by M. Fernandez (personal communication).	Yes	Positive reinforcement by caretakers (M. Fernandez, personal communication), housed with Berani (see below).	whistling by
Indah <sup>a</sup>	F	1980 (-2004)	hybrid	1785	n.a.	No	No	Yes; verbal request or whistling.	on	No further info available.
Puan	F	1989	Sumatran	2266	Heidelberg Zoo, Germany	No	No. Whistling confirmed by Kowalsky (pers. comm.)	No.		Housed in the same cage with Ujian (see above)

Tilda	F	≈1967 (wild born)	Bornean	1452	Cologne Zoo, Germany	Yes	No. Whistling confirmed by caretakers and, Lameira and Hardus	?	No reinforcement by caretakers (Lameira and Hardus, personal observation). Whistling as part of stereotypical behavior.	whistling by
-------	---	-------------------------	---------	------	----------------------------	-----	--	---	--	-----------------

\* All individuals housed in group with at least one other conspecific. <sup>a</sup> Wich et al. (2009b)

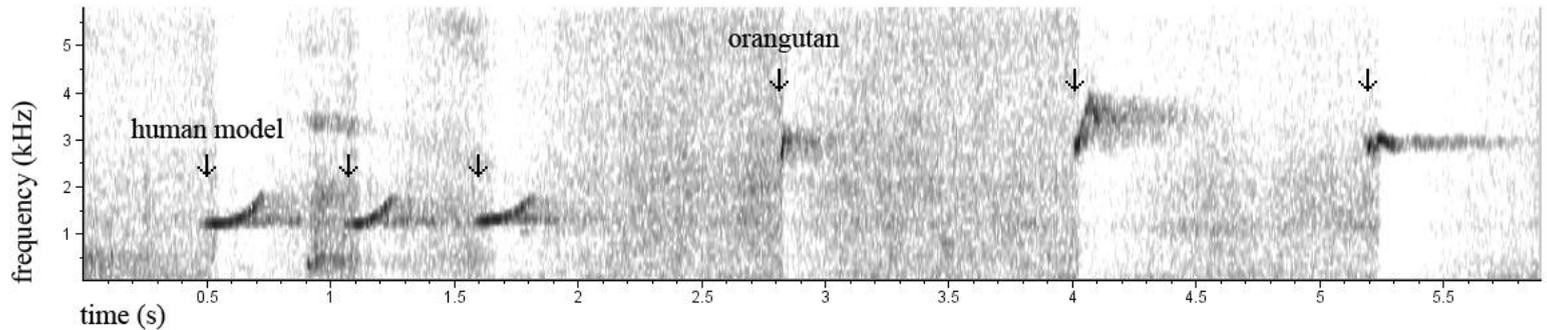


Figure 1. Spectrographic representation of a triple whistle produced by the human demonstrator followed by a triple whistle by Ujian.

Per paradigm, correctly matching whistles were produced 95%, 83% and 77% [in full data set: 89%, 84% and 81%] of the times, respectively. From the total of 282 [1395] single, double, triple or alternated whistles presented by the human demonstrator Ujian replicated 83.3% [83.5%] correctly (Chi-square test for 4 x 5 cross table;  $\chi^2 = 696.29$ ,  $df = 12$ ,  $P < 0.0001$ ).

Cohen's kappa for inter-observer reliability was, between observations of the demonstrator's whistles, for each paradigm, 1.0 ( $n = 66$ ,  $p < 0.001$ ), 1.0 ( $n = 84$ ,  $p < 0.001$ ) and 0.97 ( $n = 130$ ,  $p < 0.001$ ) respectively. Cohen's kappa for inter-observer reliability was, between observations of the orangutan's whistles, for each paradigm, 0.8 ( $n = 66$ ,  $p < 0.001$ ), 0.79 ( $n = 84$ ,  $p < 0.001$ ) and 0.87 ( $n = 130$ ,  $p < 0.001$ ) respectively.

### *Individual Variation*

The qualitative production characteristics of three whistling orangutans were compared (Table 2). Bonnie and Katy exhibited symmetrical lip protrusion during whistling (Fig. 2), however, production style, airflow phase and airflow durations were qualitatively distinct. Ujian presented a unique protrusion of lips with only the lower lip protruded during production (Fig. 2) but his whistles were qualitatively similar in production style and airflow phase to Katy's whistles. Bonnie was the only individual to produce long duration whistles and Ujian was the only individual to produce whistles in the inflow of air.

Table 2. Qualitative whistle production characteristics per individual

Characteristic	Bonnie	Ujian	Katy
<i>Production style</i>	Constant	Modulated	Modulated
<i>Airflow phase</i>	Outflow	Outflow/inflow	Outflow
<i>Airflow duration</i>	Short/long	Short only	Short only
<i>Lip protrusion</i>	Symmetrical	Lower lip protruded	Symmetrical

In addition, the quantitative acoustic parameters of three whistling orangutans were compared and differed considerably for all parameters measured (Table 3, Fig. 3). A discriminant analysis (using all the acoustic parameters) correctly assigned 99.6% of the whistles to the individuals, with the first canonical discriminant function explaining 94.4% of variance and the second function explaining the remainder (Fig. 4). Discriminant

analyses with leave-one-out procedure yielded an overall 97.8% of correct assignments to the individuals. The first discriminant function showed the largest absolute correlation with whistle mean pitch, pitch s.d. and PMAS (Fig. 3). The second discriminant function showed the largest absolute correlation with whistle duration, jitter, shimmer, HTN and RTMA (Fig. 3).



Figure 2. Lip protrusion during whistling; lower lip protruded by Ujian on the left, symmetrical lip protrusion by Bonnie on the right.

### *Discussion*

The results obtained by this study support the hypothesis that orangutan whistling was acquired from humans through sound learning. Results from the replication of the experiments on voluntary control over whistling, previously conducted with Bonnie (Wich et al. 2009b), demonstrate that Ujian exerts at least a similar level of control over whistling production style (i.e. lip control) and airflow phase (i.e. respiration control). Based on the fact that both Bonnie and Ujian produced whistles on cue by caretakers (Table 1), it is reasonable to deduce that at least other orangutans who also whistle on cue or on request (Table 1) exert comparable degrees of control (e.g. Katy, Feliz). Future

experiments will need to verify this. The replication of the experiments and the identification of several whistling individuals (Table 1) verify that Bonnie, and the acquisition of novel sounds, is not an oddity among the population of orangutans living in captivity.

Table 3. Quantitative acoustic whistle parameters per individual. Different superscripts represent significant differences per parameter between individuals following Mann-Whitney U tests, with *nominal* P values adjusted with Bonferroni method (i.e.  $P < 0.0003$ , or  $0.001/3$ ). Significant differences are *nominal* because calls produced by the same individual during the same session/bout are not independent. Median response values are presented with the 25 and 75 percentiles between brackets.

Parameter	Bonnie (N = 103)	Ujian (N = 100)	Katy (N = 97)
<i>Duration (s)</i>	0.356 (0.240, 0.543) <sup>a</sup>	0.126 (0.193, 0.088) <sup>b</sup>	0.079 (0.060, 0.104) <sup>c</sup>
<i>Mean pitch (Hz)</i>	779.2 (752.3, 804.2) <sup>a</sup>	2804.7 (2632.8, 2987.6) <sup>b</sup>	1421.1 (1268.3, 1536.1) <sup>c</sup>
<i>Pitch s.d. (Hz)</i>	17.4 (10.9, 26.4) <sup>a</sup>	139.5 (83.2, 249.7) <sup>b</sup>	35.3 (18.4, 50.2) <sup>c</sup>
<i>Jitter</i>	2.23 (1.39, 3.48) <sup>a</sup>	4.71 (3.29, 6.38) <sup>b</sup>	7.14 (4.84, 8.62) <sup>c</sup>
<i>Shimmer</i>	7.01 (4.07, 13.09) <sup>a</sup>	16.9 (12.5, 20.4) <sup>b</sup>	21.5 (16.9, 23.7) <sup>c</sup>
<i>HNR</i>	18.3 (13.0, 21.8) <sup>a</sup>	10.5 (8.5, 13.7) <sup>b</sup>	6.30 (3.36, 10.5) <sup>c</sup>
<i>PMAS (Hz/s)</i>	1003.4 (714.9, 1403.3) <sup>a</sup>	8958.4 (6122.7, 12985.6) <sup>b</sup>	3794.8 (2815.0, 4952.1) <sup>c</sup>
<i>RTMA (%)</i>	53.1 (32.2, 66.2) <sup>a</sup>	32.6 (18.0, 47.7) <sup>b</sup>	53.0 (27.9, 32.6) <sup>a,b</sup>

Whistling production physiognomy (i.e. lip protrusion) by Ujian, Bonnie and Katy did not predict differences in qualitative production characteristics of whistles between individuals. This suggests that the acquisition mechanisms of whistling by Ujian and Bonnie (who did not receive training to acquire whistling) may have been mostly auditory, meaning that individuals attended, at least more closely, to the acoustic characteristics of whistling production than production physiognomy. This is consistent with the view of several scientists who argue that social learning of novel behaviors may occur without exposure to direct tutoring (Tomasello et al. 1993; Savage-Rumbaugh et al. 1986; Savage-Rumbaugh et al. 1985; Biro et al. 2003), specifically, that this may be also the case in the

sound domain where visual contact is not required for social learning. These results suggest that pioneer language projects implemented in the second half on the 20<sup>th</sup> century (e.g. Shadle 1983; Premack 1971; Rumbaugh 1977; Savage-Rumbaugh et al. 1978; Miles 1993; Gardner and Gardner 1969; Savage-Rumbaugh and Lewin 1994; Patterson and Linden 1981) may have overestimated the lack of acoustic learning in great apes because of their focus on vocal learning, in contrast to sound learning, as described here. Hence, these projects and their results should not be considered as evidence for the absence of acoustic learning in great apes.

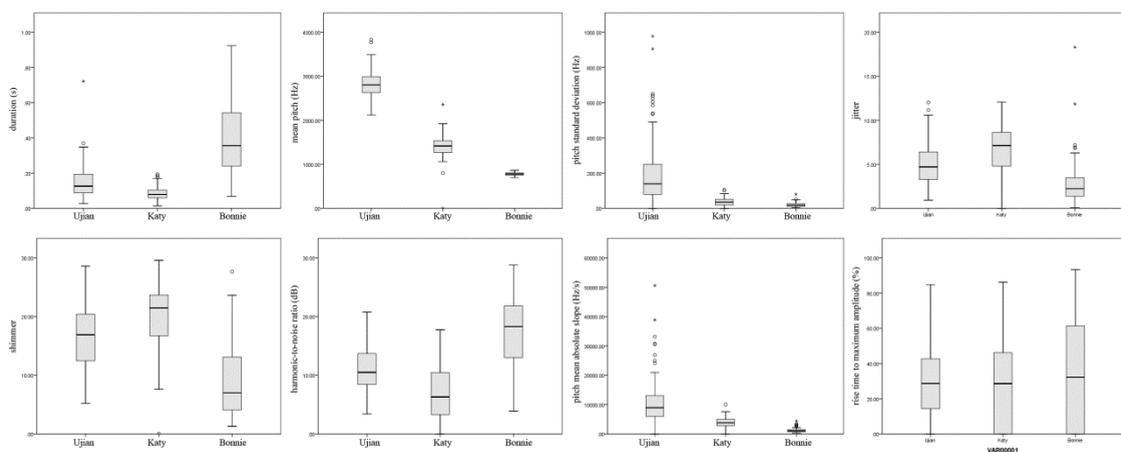


Figure 3. Boxplot graph depicting individual variation per each whistle quantitative acoustic parameter.

All quantitative acoustic parameters that were measured showed significant differences between the three individuals, reflected in high rates of correct assignments by discriminant analyses of whistles to their producer. Although Katy's whistles used for assessing inter-individual variation were not recorded under similar circumstances as Ujian's and Bonnie's, discriminant functions' groups and group centroids demonstrate that this did not influence the analyses; groups did not overlap, and distance between centroids was the longest between Ujian and Bonnie (Fig. 4). Whistling frequency showed significant differences between the three individuals. Even though the acoustic structure of sounds is simpler than that of vocalizations, sounds can nevertheless accommodate sufficient variation to convey information about the identity of the signaler, and perhaps

other aspects of the behavior, such as context. These large differences between individuals were not unexpected since whistle acoustics depend on multiple geometries associated with the production mechanism, such as area of mouth opening, volume velocity of air exhaled, height of mouth opening, depth of the constricted section of the lips, flow velocity through lip constriction, tongue back-/forward position (Shadle 1983), and lip rigidity (Wilson et al. 1971). Indeed, the fact that individuals were able to produce different whistles demonstrates their ability to control a number of these features and that each individual used a particular combination of these features to successfully produce whistles. A similar phenomenon is known to occur in humans – “motor equivalence” – where speakers develop different motor strategies (i.e. use different musculatures) of the larynx to achieve the same voice outcome (Ludlow 2005). Also, in order to change whistle frequencies, a whistler has to increase or decrease the air pressure to raise or drop the whistle frequency respectively (Meyer 2004). Due to the strong link between the respiratory system and affective state (sensu Owren and Rendall 2001) of an individual, whistler motivation, whistling context and/or reward could produce considerable variation in the whistle’s frequency. Altogether, successful whistle production is dependent on the simultaneous control of multiple miniature features of the respiratory and oral systems, which could explain why orangutans (and many humans) do not expand the diversity of whistles they produce in the absence of positive reinforcement or training. Nevertheless, orangutan whistle frequencies are within to the range of human whistles, typically located between 500 and 3000 Hz (Shadle 1983). Future documentation and audio recordings from more orangutans are necessary to understand the exact whistle variation between individuals and its underlying causes.

The record of whistling orangutans catalogued in this article (Table 1) indicates that whistling is apparently more dependent on the presence of relevant tutors (intra or interspecific) and proper learning conditions (the possibility of rewards associated with the acquisition of the behavior) than on aspects of the whistling individuals, such as gender, housing facility or ancestry. For example, Puan and her fraternal twin brother Enche (Studbook ID 2265) were hand raised by BK and remained under his care, housed in the same facility where they were born until they were separated at 11 years of age in 1998. Neither showed whistling behavior at any point during that time. Puan was only heard to whistle after 2005, when she was living in her present facility with Ujian, who is a fluent whistler and receives daily positive reinforcement for the behavior. In addition, the

other 5 full-siblings of Puan (Grisella, Studbook ID 1559; Josephine, Studbook ID 1473; Tigah, Studbook ID 1735; Sandokan, Studbook ID 1844; Tujoh, Studbook ID 1894) and their parents (Beber Nogger, Studbook ID 1333 and Muna, Studbook ID 1334) also received full time care by BK, and are/were non-whistlers. It seems likely that exposure to a whistling conspecific, who obtained benefits from the behavior, induced Puan to whistle. Similar circumstances account for whistling acquisition by Berani, with Feliz functioning as the demonstrator (M. Fernandez, personal communication). These orangutans, in the absence of any training programs, seem to have *originally* learned how to whistle independently of any direct reward by humans. The context and possible functions associated with the acquisition of whistling by orangutans in the entertainment industry remain speculative.

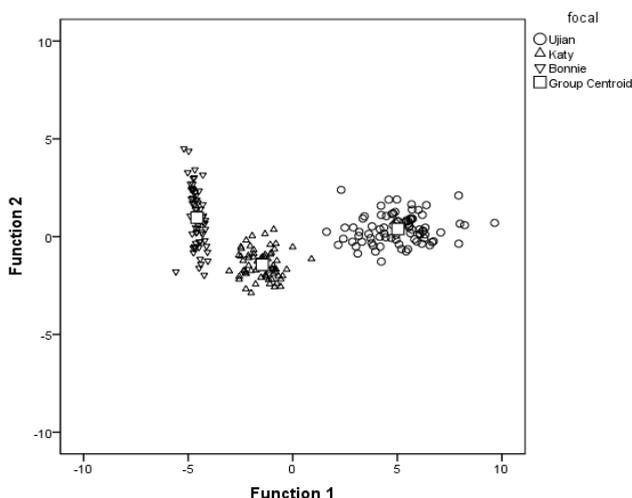


Figure 4. Individual variation in whistling according to discriminant functions based on quantitative acoustic parameters of whistling.

Orangutan whistling is theoretically relevant in relation to theories on speech evolution. To our best knowledge, it reveals, the first great ape example of socially learned reception-first development based on sound copying rather than individually learned reception-first development based on sound invention (e.g. Hopkins et al. 2007). Sound copying is an essential addition to sound invention alone, which does not lead to an open-ended repertoire that could spread in a population. By itself, sound invention

facilitates only individual differences in sound repertoires. The results of this study demonstrate that social learning among orangutans can allow a novel sound to spread within a population. In fact, this exact scenario has been proposed for several sounds among wild orangutan populations (Hardus et al. 2009b; van Schaik et al. 2003; van Schaik et al. 2006). Vocal learning has not been demonstrated among orangutans as it has been in birds and several non-primate mammals. However, orangutans are capable of learning novel sounds. Future comparative studies are needed to further develop the hypotheses related to the adaptive value of acoustic learning during human evolution and the evolutionary pressures favoring its emergence across other taxa. Of course, direct comparisons between taxa must be made cautiously and conservatively.

Social learning of novel sounds by orangutans highlights the evolutionary importance of this specific behavioral process in humans, a neglected topic that deserves further investigation. While the acoustic properties of primate vocalizations closely resemble those of human vowels (Owren et al. 1997), sounds closely resemble those of human consonants, since they are commonly unvoiced (alike non-sonorant consonants). A remarkable example is the use of click consonants in South African languages (Knight et al. 2003). It has been suggested that these sounds existed during the early history of modern humans before they emigrated from Africa. Due to the simpler motor control requirements, it is likely that sound learning preceded vocal learning during the evolution of speech. Sound learning may have offered the advantages of an extended repertoire to our ancestors for the transmission of more (graded) information. This process may have facilitated selective momentum towards the expansion of voluntary control over the vocal folds (c.f. Lameira et al. in press). In fact, it is difficult to imagine vocal learning in place without a degree of skilled control over the air flow exciting the vocal folds as well as control of the oral structures filtering the vocal signal. Sound learning may have represented an evolutionary stepping stone towards vocal learning. At the same time, sound learning also likely played an equally important and parallel role as vocal learning in the maturation of human speech, since human spoken languages are basically composed by vowels and consonants. The theoretical framework presented here for the examination of acoustic learning in primates, and the terminology used for sounds and vocalizations *sensu stricto* (Hardus et al. 2009a; Hardus et al. 2009b; Wich et al. 2009b; Owren et al. 2010; Hopkins et al. 2011) is heuristically relevant for theories on speech evolution.

## *Conclusions*

This study demonstrates that orangutans are capable of adding sounds into their innate repertoire through social learning. These results suggest that the last common ancestor of orangutans and humans, dating back to the homininae-ponginae split (i.e. 9 - 13 mya; Hobolth et al. 2011), may already have had a call repertoire that was composed of both innate calls and socially learned sounds, facilitated by a degree of voluntary control over the lips, tongue and respiration. The dawn of an open-ended repertoire that characterizes modern human speech may be much older than previously anticipated.

## *Acknowledgments*

Financial support for this project was generously provided by: Fundação para a Ciência e Tecnologia (Portugal), Wenner-Gren Foundation for Anthropological Research (USA), Dr. J.L. Dobberke voor Vergelijkende Psychologie (the Netherlands), Lucie Burgers Foundation for Comparative Behaviour Research (the Netherlands), Schure-Beijerinck-Popping Fonds (the Netherlands), Ruggles-Gates Fund for Anthropological Scholarship of the Royal Anthropological Institute of Great Britain and Ireland and Fundação Calouste Gulbenkian (Portugal). Special thanks to the curator at Heidelberg Zoo, Sandra Reichler, and the zoological staff, namely Anke Jacob. We would like to acknowledge the zoological staff at the Cologne Zoo, namely Klaus Pyzora and Mike Ebert, and the curator Alex Sliwa. Marsha Fernandez, curator of mammals at the Audubon Zoo, New Orleans, LA, and Robert Porec, curator of mammals at the Honolulu Zoo, Hawaii, provided indispensable reports on whistling orangutans. We are grateful to David Puts and Michael Owren for discussion on the acoustic measurements of orangutan whistles, and to David Dellatore and Michael Krützen for helpful comments. Kim Nouwen kindly provided blind ratings on performance in whistle matching. Authors reveal no conflicts of interest.



## **Predator guild does not influence orangutan alarm call rates and combinations**

Adriano R. Lameira, Han de Vries, Madeleine E. Hardus, Cedric P.A. Hall,  
Tatang Mitra-Setia, Berry M. Spruijt, Arik Kershenbaum, Elisabeth H.M. Sterck,  
Maria van Noordwijk, Carel van Schaik and Serge A. Wich

### *Abstract*

Monkey alarm calls have shown that in the primate clade combinatorial rules in acoustic communication are not exclusive to humans. A recent hypothesis suggests that the number of different call combinations in monkeys increases with increased number of predator species. However, the existence of combinatorial rules in great ape alarm calls remains largely unstudied, despite its obvious relevance to ideas about the evolution of human speech. In this paper we examine the potential use of combinatorial rules in the alarm calls of the only Asian great ape: the orangutan. Alarm calls in orangutans are composed of syllables (with either one or two distinct elements), which in turn are organized into sequences. Tigers and clouded leopards are predators for Sumatran orangutans, but in Borneo tigers are extinct. Thus, orangutans make a suitable great ape model to assess alarm call composition in relation to the size of the predator guild. We exposed orangutans on both islands to a tiger and control model. Response compositionality was analyzed at two levels (i.e. syllable and syllable sequences) between models and populations. Results were corroborated using information theory algorithms. We made specific, directed predictions for the variation expected if orangutans used combinatorial rules. None of these predictions were met, indicating that monkey alarm call combinatorial rules do not have direct homologues in orangutans. If these results are replicated in other great apes, this indicates that predation did not drive selection towards ever more combinatorial rules in the human lineage.

*In press by Behavioural Ecology and Sociobiology*

## *Introduction*

Several non-primate mammal and nonhuman primate species use sophisticated combinatorial capacities in their call systems to influence conspecifics and/or predators (Zuberbühler 2002; Arnold and Zuberbühler 2008; Ouattara et al. 2009a, 2009b; Schel et al. 2010; Candiotti et al. 2012a; Kershenbaum et al. 2012; Clarke et al. 2006; Blumstein 1999; ten Cate and Okanoya 2012). Animals can generate, generalize and categorize call strings or sequences, learn about the co-occurrence of elements within call sequences, and learn combinatorial rules such as attending to phonetic similarity between elements, co-occurrence of elements, duplication of elements, although the animal capacity for learning abstract rules needs substantial more independent research (ten Cate and Okanoya 2012). Within the primate clade, such combinatorial rules have been mainly described in monkeys and lesser apes. Similar primate rules are only known to exist in humans. This is relevant to the debate on the emergence of recursion (the capacity to generate an infinite range of expressions from a finite set of elements), which is suggested to represent a uniquely human component of language (Hauser et al. 2002). Remarkably, however, the call systems of great apes have seldom been studied for the potential occurrence of call combinations (cf. Clay and Zuberbühler 2011; Crockford and Boesch 2005). In non-primate mammals and nonhuman primates, such evidence has mostly been demonstrated in alarm call systems (Zuberbühler 2002; Arnold and Zuberbühler 2008; Ouattara et al. 2009a, 2009b; Schel et al. 2010; Candiotti et al. 2012a; Kershenbaum et al. 2012; Clarke et al. 2006). A relevant question is, thus, whether predation constituted an important ecological effect for the emergence and evolution of combinatorial rules in the primate lineage. Such possibility would be in agreement with several observations indicating that nonhuman primates exert some flexibility over their anti-predator (acoustic) responses, such as variation of alarm calls according to predator guild (Kavanagh 1980; Fichtel and Kappeler 2011; Fichtel and Van Schaik 2006) and adjustment of anti-predator responses to newly introduced predators or local predator guild (Gil-da-Costa et al. 2003; Rainey et al. 2004; Yorzinski and Ziegler 2007; Schel et al. 2009).

A recent study has presented evidence supporting the hypothesis that increased predation pressure (i.e. number of predator species) increases the number of ways monkeys combine alarm call elements, that is, compositionality (Stephan and Zuberbühler 2008). This hypothesis suggests an ecological/evolutionary scenario with predation

operating as selective pressure towards more call combinatorial rules in our early ancestors, and conceivably towards the emergence of acoustic recursion in the hominin lineage. The predation pressure hypothesis (Stephan and Zuberbühler 2008) implies that predator species with different physiognomic/behavioral features elicit different call responses by primate prey; thus a larger number of predator species will elicit a larger number of different call responses by the primate prey. There are potentially two ways to communicate about new predators. The first is to invent or learn novel call types. This, however, requires fine voluntary motor control over call production, which is limited in nonhuman primates (Cheney and Seyfarth 2005; cf. Owren et al. 2010). The second possibility is thus, to assemble the same call types differently into novel sequences generating a larger number of different call responses using combinatorial rules (e.g. Candiotti et al. 2012a; Zuberbühler 2002). Although it is highly likely that hominins experienced some periods of increased predation pressure (e.g. big cats, canids) in the course of evolution (Hart and Sussman 2008), for instance during range expansions or migrations, the generality of the predation pressure hypothesis for increased compositionality in human evolutionary lineage has not yet been assessed in great apes.

Orangutans offer a useful great ape model to examine these two major questions on (1) the possibility of combinatorial rules in great ape alarm calls and (2) the possibility of predation effects on the compositionality of great ape alarm calls. On both islands where orangutans occur, Sumatra and Borneo, alarm calls towards potential predators comprise 4 different single-calls – kiss-squeak (KSQ), grumph (GR), gorkum (GK) and complex call (CXC) – and 3 combined-calls – KSQ+GR, KSQ+GK and KSQ+CXC (Hardus et al. 2009a), thus producing 7 possible syllables. Each of these 7 syllables has been well described and is distinguishable audibly and/or by visual inspection of spectrograms (Hardus et al. 2009a) (see [aim.uzh.ch/orangutannetwork](http://aim.uzh.ch/orangutannetwork)). Orangutan alarm call responses towards potential predators typically last several minutes and can continue for more than an hour, creating sequences of syllables with varying length and with silence gaps between sequences of varying length. Therefore, the orangutan alarm call system allows the analysis of compositionality at two levels: single- and combined calls within syllables and syllables within sequences. Other anti-predator behavioral responses by orangutans include shaking braches and throwing branches at the predator (Hardus et al. 2009b), but here, focus is only given to alarm calls.

The predators preying on orangutans differ between Sumatra and Borneo. While clouded-leopards (*Neofelis diardi*) and pythons (*Python* spp.) constitute potential arboreal predators on both Borneo and Sumatra (Rijksen 1978), tigers (*Panthera tigris sumatrae*), which are largely terrestrial but prey upon orangutans (Rijksen 1978), are only present on Sumatra. On Borneo the presence of tigers has been suggested at two locations for the late Pleistocene and early Holocene (circa 10 – 12 ka (Piper et al. 2007) or 400-480 orangutan generations ago (Wich et al. 2009c)), but there is no evidence that their distribution extended beyond these regions and tigers have become extinct since then. Although the exact timing of this extinction is not known, there are no historical records of tigers on Borneo after this period. Hence, the geographic distribution of orangutan natural predators allows investigating the predation pressure hypothesis on the alarms calls of great apes. In the absence of a predator, a reduction of the number of different combinations is expected on Borneo, as the predator's specific call sequence would become obsolete and there would be no advantages for learning this sequence by individuals in the population. In monkeys, this process is observed over a period of 30 years (Stephan and Zuberbühler 2008) thus one would expect, at least, similar effects in orangutans.

In this study, we exposed female orangutans to a predator model (i.e. tiger model) and to a control model (i.e. white model of the same shape) at two sites, one on each island, Ketambe (Sumatra) and Tuanan (Borneo). Only at Ketambe are orangutans familiar with tigers. We first assessed the possibility of combinatorial rules and the effect of predation at the syllable level (i.e. single- and combined-calls). When orangutans use combinatorial rules at the syllable level, we expect syllable frequencies to vary between models at Ketambe (because one model is familiar and the other unfamiliar), but not at Tuanan (because both models are unfamiliar). When predator presence affects orangutan use of syllables, we expect that syllable frequencies will differ between populations when presented with the tiger model (because the model is familiar in one population but unfamiliar in the other), but not when presented with the white model (because the model is unfamiliar in both populations).

Secondly, we assessed the possibility of combinatorial rules and effects of predation at the level of syllable sequences. When orangutans use combinatorial rules at the level of syllable sequences, we expect that the rate of performing (at least) two- and three-syllable sequences will differ between predator and control model at Ketambe, but

not at Tuanan. When predation affects orangutan syllable sequences, we expect that the rate of (at least) two- and three-syllable sequences will differ between Ketambe and Tuanan towards the tiger model, but not towards the white model. If orangutans use syllable sequences with a larger number of syllables, this will be detectable in the use of two- and three-syllable sequences. Moreover, at each level of compositionality (syllables and syllable sequences), we condense our data and replicate the analyses by categorizing syllable and syllable sequences as (composed by) single- or combined-calls, irrespective of the exact type of call(s). For instance, for KSQ, KSQ+GR and KSQ+GK, we consider instead one single-call and two combined-calls respectively. In addition, we analyzed the data using information theory algorithms (Kershenbaum et al. 2012) to validate the results obtained.



Figure 1. Tiger model in experimental setting.

## *Methods*

### *Site and data collection*

Data were collected at Tuanan (2°09' S; 114°26' E), Central Kalimantan, Borneo, Indonesia, between December 2004 and August 2005 and at Ketambe (3°41' N; 97°39' E), Aceh,

Sumatra, Indonesia, between December 2010 and May 2011, following the same protocol. Six females were tested at Ketambe and five females at Tuanan. These individuals constituted the most habituated adult females resident in the central part of each study site. Orangutans live in dispersed fission-fusion communities, where females spend the majority of their time alone with the exception of their own single offspring (Delgado and van Schaik 2000). Other females could not be experimentally tested due to extremely low encounter rates by human observers. All females tested were parous with dependent offspring, except one female (i.e. nulliparous) at each population. The predator model consisted of a realistic tiger-patterned sheet draped over a human demonstrator on all fours (Figure 1). A similar model, but covered with a white sheet, exposed to the same orangutan females, served as an experimental control, in order to isolate the effect of familiarity with tigers. When a focal female was between 12 and 20 meters height in the forest, feeding, resting or slowly moving, the model moved past in front of the focal. The model halted for approximately 2 minutes when the focal viewed it and then continued moving until out of sight. Model movements relative to the subject orangutan were coordinated by an additional human observer collecting behavioural and acoustic data continuously from the appearance of the model until 30 minutes after disappearance. Experiments were never conducted in the same location in the forest. Each individual was exposed only once to each model, and the two different models were presented to the same individual at least 5 days apart to avoid habituation. Presentation order of the models within population was pseudo-randomized according to the females whose location was known.

The use of a clouded leopard model as control in both populations was not conducted for several reasons. First, the presentation of the clouded leopard model on the ground (to preserve experimental settings consistent across models) could be perceived by orangutans as an encounter presenting little underlying danger, as clouded leopards characteristically hunt through the canopy. Second, the presentation of the model on the canopy would require a different presentation scheme since it would be impossible for a human demonstrator to walk on fours through the canopy. Third, because orangutans are fundamentally arboreal, the presentation of a model on the canopy, instead of on the ground, could be perceived differently by orangutans. For instance, a wild orangutan highly habituated to humans and followed for more than three decades may respond intensively when encountering human in the canopy (Lameira and

Hardus, personal observation). Forth, responses towards a model in the canopy could invoke responses due to proximity to the model, creating biases between models.

### *Data analyses*

For syllable analyses, single- and combined-call syllable rates were considered during the first 10 minutes of the alarm call response towards either model, since in an encounter with a potential predator, this will be the most critical period in prey response. That is, during this period, we measured the number of occurrences of each of the 7 syllables comprising orangutan alarm call repertoire (see above).

For syllable sequence analyses, firstly we identified the optimum duration of silence that would indicate a break between syllable sequences. This was necessary because silence gaps between orangutan syllables have varying durations, and the choice of inter-sequence gap affects the number and length of sequences. For instance, a 5 seconds silence gap between potential syllables sequences yields a smaller number of sequences than a 30 seconds silence gaps for the same data set. We calculated the duration of silence gaps at which the increase of the number of sequences with respect to gap duration decelerated (i.e. when the increase rate became less than 1). We considered this to be the optimum point because of the trade-off between simplicity/speed and versatile/informative communication in alarm calls (cf. Kemp and Regier 2012). For this calculation, we used the first ten minutes of alarm call response towards the tiger model by all six females at Ketambe, as these responses were expected to be the most syllable-dense from all predator model experiments. We determined the number of two- and three-syllable sequences when considering 5, 10, 15, 20, 25 and 30 seconds gaps between syllable sequences for these responses. We fitted a logarithmic curve through the average number of syllable sequences, averaged across the six females. The nonnegative numerator of the second derivative of this logarithmic curve indicates the length of the silence gaps above which the increase rate of the number of sequences becomes less than one. This value is equal to the coefficient of  $\ln(x)$  in the fitted logarithmic curve (Lyons 1995).

Secondly, we identified all two- and three-syllable sequences emitted by all females in both populations towards both models. When orangutans use syllable sequence rules, one expects that these will be frequently used and conventionalised between the individuals of the population. Accordingly, only those two- and three-syllable

sequences that were used by at least two individuals of one population at one type experiment *and* that were produced at least 5 times by one of the two individuals were considered for statistical testing. A syllable comprised by a combined-call (e.g. KSQ+GR) differed unequivocally from a syllable sequence composed by the same calls (e.g. KSQ-GR), in that calls follow each other immediately within a combined-call syllable (i.e. less than 500 milliseconds), while they are separated by more than 3 or 4 seconds within a sequence.

Finally, to examine the presence of combinatorial rules by means of an alternative method, we used a technique taken from information theory to examine the data content of the different call sequences. We generated Markov transition tables separately for each individual, indicating the transition probabilities between common sequences, where each element  $T_{a,b}$  of the transition matrix represents the probability that sequence  $a$  will be followed by sequence  $b$ . We then calculated the entropy of the transition table  $H = -\frac{1}{N^2} \sum_a \sum_b T_{a,b} \ln[T_{a,b}]$  where  $N$  is the number of distinct sequences. Transition table entropy is a measure of the deviation from random sequences (Cover and Thomas 2006) and hence an indication of the extent to which combinatorial rules are being used. Entropy measures have been used to analyse animal acoustic communication sequences in various taxa, including hyraxes (Kershenbaum et al. 2012) and humpback whales (Suzuki et al. 2006). We used two-way ANOVA to test for differences in Markov entropy between populations and between models, using as our null hypothesis that neither population location (Sumatra vs. Borneo) nor model type (tiger vs. white) affects the combinatorial complexity of call sequences.

### *Apparatus*

All calls were recorded during the experiments with a Marantz Recorder PMD-660 with a Rode NTG2 Microphone (at Ketambe), and with a Marantz Recorder PMD222 with a Sennheiser Microphone ME 64 and a Sony Recorder TCD-D100 with a Sony Microphone ECM-M907 (at Tuanan). Statistical tests were conducted using IBM SPSS 19 (2010, SPSS, Inc.), with significance level set at  $p < 0.05$ .

## Results

### Syllable analyses

Total emission rate of syllables did not differ between populations for both models (Mann-Whitney U: tiger model:  $N_{\text{Ketambe}}=6$ ,  $N_{\text{Tuanan}}=5$ , exact  $P=0.20$ ; white model:  $N_{\text{Ketambe}}=6$ ,  $N_{\text{Tuanan}}=5$ , exact  $P=0.18$ ). Median syllable rates (25 and 75% percentiles) were  $6.2 \text{ min}^{-1}$  (2.7, 7.225) and  $8.4 \text{ min}^{-1}$  (5.75, 12.1) at Ketambe and Tuanan respectively towards the tiger model, and  $0 \text{ min}^{-1}$  (0, 3.1) and  $4.3 \text{ min}^{-1}$  (0.2, 8.35) towards the white model.

The emission rate of only one syllable (KSQ) was higher towards the tiger model than the white model at both sites Ketambe and Tuanan (Table 1A and Table S1). None of the syllable rates differed between populations for each model (Table 1B and Table S1). When condensing the data, and considering syllables solely either as single- or combined-call syllables (S or C), the emission rate of single-call syllables differed between models only at Tuanan, with higher emission rates towards the tiger (Table 1A). Moreover, the emission rate of single-call syllables towards the tiger model was higher at Tuanan than at Ketambe (Table 1B). To verify whether these results were solely an effect of high KSQ emission rates, we conducted the analyses once again but excluding KSQs. Emission rates of single-calls differed no longer between models at Tuanan (Mann-Whitney test,  $N = 5$ ,  $Z = -0.904$ ,  $P = 0.366$ ), or between populations towards the tiger model (Mann-Whitney test,  $N_{\text{Ketambe}} = 6$ ,  $N_{\text{Tuanan}} = 5$ ,  $Z = -1.354$ ,  $P = 0.176$ ).

Table 1. Exact P values of statistical comparison per syllable (A) between models within populations and (B) between populations for each models.

		KSQ	GR	GK	CXC	KSQ+ GR	KSQ+GK	KSQ+ CXC	S	C
A	Ketambe	<b>0.041<sup>T</sup></b>	0.655	0.715	0.109	0.102	0.144	0.066	0.225	0.141
	Tuanan	<b>0.043<sup>T</sup></b>	1	0.655	0.317	0.461	0.345	1	<b>0.043<sup>T</sup></b>	0.223
B	Tiger	0.056	0.361	1	1	0.203	0.134	0.104	<b>0.022<sup>Tu</sup></b>	0.853
	White	0.056	0.361	1	1	0.203	0.134	0.104	0.251	0.91

S – single call, C – combined call

Superscripts indicate model or population with significantly higher syllable emission rate:  
A: T = tiger model, W = white model; and B: Ke = Ketambe, Tu = Tuanan

### *Syllable sequences analyses*

The number of two- and three-syllable sequences per minute increased when a longer duration of silence gaps between syllable sequences was permitted (Figure 2). The average number of sequences plotted against the duration of silence gaps fitted a logarithmic curve ( $R^2 = 0.957$ ,  $y = 17.514 * \ln(x) + 0.307$ ; Figure 2). Accordingly, the logarithmic coefficient 17.514 indicates the optimum duration of silence gaps (see Methods). Since we assessed the number of syllables sequences at 5 seconds intervals, we considered the value of 15 seconds for the duration of silence gaps between syllables for subsequent analyses.

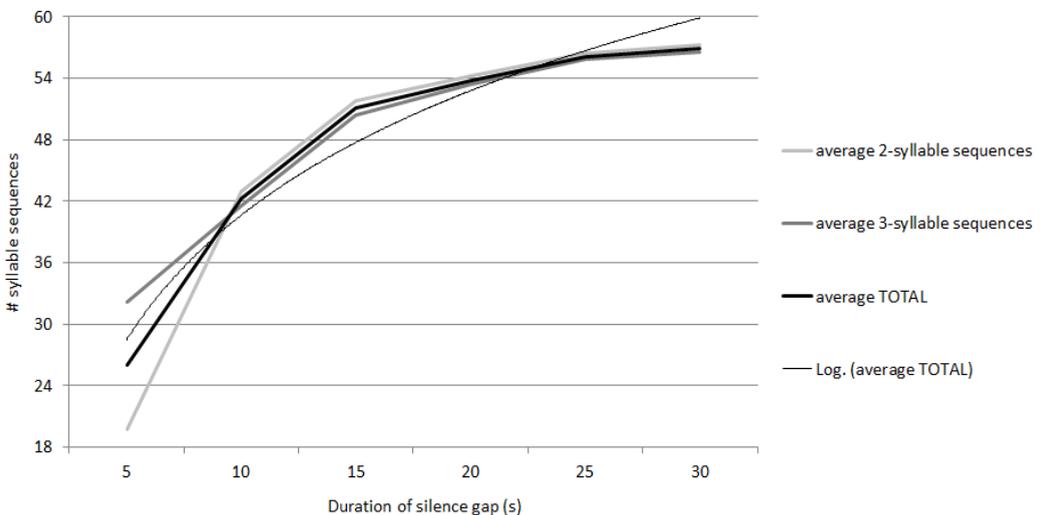


Figure 2. Number of two- and three-syllable sequences vs. duration of silence gaps considered between these sequences.

From a total of 49 (i.e.  $7^2$ ) possible two-syllables sequences, 33 different sequences were recorded at Ketambe towards the tiger model and 11 towards the white model, whereas 11 and 15 different sequences were recorded at Tuanan. The two populations did not differ in the total emission rate of two-syllable sequences to either of

the models (Mann-Whitney U: tiger model:  $N_{\text{Ketambe}}=6$ ,  $N_{\text{Tuanan}}=5$ , exact  $P=0.2$ ; white model:  $N_{\text{Ketambe}}=6$ ,  $N_{\text{Tuanan}}=5$ , exact  $P=0.135$ ). Median two-syllable sequence rates (25 and 75% percentiles) were  $5.5 \text{ min}^{-1}$  (1.725, 6.3) and  $7.4 \text{ min}^{-1}$  (4.65, 11.35) at Ketambe and Tuanan respectively towards the tiger model, and  $0 \text{ min}^{-1}$  (0, 2.725) and  $3.0 \text{ min}^{-1}$  (0.1, 7.5) towards the white model.

We determined whether the use of particular two-syllables sequences seemed habitual. Overall, eight two-syllables sequences were used by at least two individuals in a single population in response to the same model, and used at least five times by at least one of these individuals (Table S2). Within populations, the rate at which these two-syllable rates were given to the two models did not differ significantly (Table 2A). However, two (out of eight) sequences given towards the tiger model (i.e. KSQ-KSQ and KSQ+GR-GR) were given at significantly higher rates, concerning one sequence at each population (Table 2B).

Table 2. Exact P values of statistical comparison of two-syllable sequences (A) between models within populations and (B) between populations for each model.

		R-GR	KSQ+GR- KSQ+GR	KSQ- KSQ	GR- KSQ+GR	KSQ+GR- GR	KSQ+CXC- KSQ+GR	KSQ+GR- KSQ	KSQ- KSQ+GR
A	Ketambe	0.715	0.144	0.18	0.715	0.715	0.18	1	0.593
	Tuanan	1	0.273	0.08	1	0.317	0.854	0.416	0.273
B	Tiger	0.082	0.712	<b>0.015<sup>Tu</sup></b>	0.073	<b>0.034<sup>Ke</sup></b>	0.833	0.109	0.222
	White	0.361	0.29	0.91	1	1	0.104	0.59	0.351

A – Wilcoxon signed ranks test

B – Mann-Whitney test

Superscripts indicate model or population with significantly higher syllable emission rate

When condensing our data, none of the four possible two-syllable sequences (SS, SC, CC and CS) differed between the models at both sites (Table 3A). However, the rate of one two-syllable sequence (i.e. single call-single call) towards the tiger model was higher at Tuanan than at Ketambe (Table 3B).

Based on the same silence criterion, we also distinguished three-syllable sequences. From a total of 343 (i.e.  $7^3$ ) possible three-syllable sequences, 63 and 20 different sequences were recorded at Ketambe towards the tiger- and the white model respectively, and 23 and 27 different sequences were recorded at Tuanan respectively. The rate of three-syllable sequences did not differ between populations for both models (Mann-Whitney U: tiger model:  $N_{\text{Ketambe}}=6$ ,  $N_{\text{Tuanan}}=5$ , exact  $P=0.17$ ; white model:  $N_{\text{Ketambe}}=6$ ,  $N_{\text{Tuanan}}=5$ , exact  $P=0.135$ ). Median three-syllable sequence rates (25 and 75% percentiles) were  $5.35 \text{ min}^{-1}$  (1.5, 6.225) and  $7.3 \text{ min}^{-1}$  (4.9, 11.1) at Ketambe and Tuanan respectively towards the tiger model, and  $0 \text{ min}^{-1}$  (0, 0.27) and  $2.6 \text{ min}^{-1}$  (0.05, 7.35) towards the white model.

Overall, six three-syllable sequences were used by at least two individuals in a single population in response to the same model, and used at least five times by the same individual (Table S3). One three-syllable sequence rate differed between models (i.e. KSQ-KSQ-KSQ) at Tuanan, with higher rates emitted towards the tiger model (Table 4A). Two sequences (out of six) differed between populations (i.e. KSQ-KSQ-KSQ and KSQ+GR-KSQ-KSQ) towards the tiger model, with higher rates emitted at Tuanan (Table 4B).

Table 3. Exact P values of statistical comparison of two-syllable sequences (A) between models within populations and (B) between populations for each models, considering single- and combined-call syllables irrespective of call type(s).

		SS	SC	CC	CS
A	Ketambe	0.345	0.465	0.109	0.465
	Tuanan	0.08	0.336	0.279	0.5
B	Tiger	<b>0.044<sup>Tu</sup></b>	0.926	0.926	1
	White	0.205	0.29	0.09	0.134

S – single call, C – combined call

A – Wilcoxon signed ranks test

B – Mann-Whitney test

When condensing our data pertaining to three-syllable sequences, one of the possible eight (i.e.  $2^3$ ) three-syllable sequences differed between models (i.e. single call-single call) at Tuanan, with higher rates emitted towards the tiger model (Table

5A). The same three-syllable sequence towards the tiger model also differed between populations, with higher rates at Tuanan (Table 5B).

Table 4. Exact P values of statistical comparison of three-syllable sequences between models within populations and between populations for each model.

		KSQ+GR-	KSQ-	KSQ-	KSQ+GR-	KSQ+GR-	KSQ+GR-
		KSQ+GR-	KSQ-	KSQ+GR-	KSQ-KSQ	KSQ-	KSQ+GR-
		KSQ+GR	KSQ	KSQ+GR		KSQ+GR	KSQ
A	Ketambe	0.068	0.180	0.655	0.317	1	1
	Tuanan	0.273	<b>0.043<sup>T</sup></b>	0.414	0.705	0.465	0.593
B	Tiger	0.664	<b>0.015<sup>Tu</sup></b>	0.421	<b>0.037<sup>Tu</sup></b>	0.246	0.146
	White	0.562	0.091	0.351	0.351	0.294	0.139

A – Wilcoxon signed ranks test

B – Mann-Whitney test

### *Information theory analysis*

Markov entropy (and hence combinatorial complexity) was slightly higher at Ketambe (N=7, mean  $1.34 \pm 0.19$  SE) than Tuanan (N=9, mean  $1.11 \pm 0.11$  SE). At Ketambe, animals presented with the tiger model gave slightly more complex calls than when presented with the white model (N=5, mean  $1.48 \pm 0.49$  SE, vs. N=2, mean  $1.00 \pm 0.52$  SE). In contrast, at Tuanan the entropy was lower on presentation of the tiger model than the white model (N=5, mean  $1.00 \pm 0.33$  SE, vs. N=4, mean  $1.25 \pm 0.28$  SE). However, the results of a two-way ANOVA test showed no significant effect for population ( $p=0.601$ ), model ( $p=0.590$ ), or population-model interaction ( $p=0.115$ ). Sample sizes differed between information theory analysis and previous analysis because only individuals who produced alarm calls could be included in the information theory analyses.

## Discussion

The results of this study show few differences between orangutan alarm call responses towards different models and between different populations. Foremost, the only difference between models at both populations concerned the emission rate of kiss-squeaks (KSQs), where individuals of both populations emitted significantly more KSQs per

Table 5. Exact P values of statistical comparison of three-syllable sequences (A) between models within populations and (B) between populations for each model, considering single- and combined-call syllables irrespective of call type(s).

		SSS	SSC	SCS	SCC	CCC	CCS	CSC	CSS
A	Ketambe	0.345	0.715	0.593	0.715	0.068	0.715	1	0.715
	Tuanan	<b>0.043<sup>T</sup></b>	1	0.273	0.465	0.273	0.715	0.498	1
B	Tiger	<b>0.017<sup>Tu</sup></b>	0.925	0.773	0.711	0.644	0.711	0.562	0.848
	White	0.205	0.484	0.562	0.562	0.416	0.134	0.29	0.662

S – single call, C – combined call

<sup>a</sup> Wilcoxon signed ranks test

<sup>b</sup> Mann-Whitney test

Superscripts indicate model or population with significantly higher syllable emission rate

unit of time towards the tiger model than the white model (Table 2). This suggests that both populations perceived a tiger-patterned sheet as being potentially more dangerous than a white sheet. This indicates that, as predicted by the multipredator hypothesis (Blumstein 2006), pattern recognition at Tuanan, where tigers are absent, may have persisted after potentially up to 480 generations due to the presence of other cat predators with a patterned fur, such as the clouded leopard. Nonetheless, it seems unlikely that orangutans in both populations could have mistakenly identified the tiger model as a clouded leopard (which is present in both populations), since the size and hunting techniques differ between the two cats, and their fur pattern is distinct; in Sumatran tiger's fur, the black outline is comprised by narrow and vertical stripes on a largely orange background with pronounced white parts, while in Sunda clouded leopard's fur, the black outline is comprised by round forms, with both dense black and spotted

sections, and with thin orange gaps between black forms, without any pronounced white parts.

At Tuanan, a significantly higher emission rate of single-calls was detected towards the tiger model than the white model (Table 2). However, this was verified to be an effect of the high emission rates of single KSQs. The fact that individuals at Tuanan emitted significantly higher rates of single-call syllables towards the tiger model than individuals at Ketambe seems also to be due to the high emission rates of single KSQs. At the level of syllable sequences, some two- and three-syllable sequences differed between models at Tuanan, and between populations towards the tiger model. However, all these sequences included KSQs as one or two syllable(s). Hence, these results may therefore be an effect of considerably high emission rates of single KSQs. For instance, emission rates of KSQ+GR-GR towards the tiger model were significantly higher at Ketambe than Tuanan, but this sequence was not used significantly more at Ketambe towards the tiger model than the white model, thus this population difference was unlikely meaningful.

Altogether, we did not find evidence to support our predictions related to either the use of combinatorial rules or an effect due to predation guild size on combinatorial rules of orangutan alarm calls. We expected that different syllable or syllable types (i.e. S and C) would be used to indicate different models in the population experiencing tiger predation (i.e. Ketambe), while predicting these differences on syllable use to be absent in the population lacking this predator (i.e. Tuanan). However, this was not found. Syllables were not used differently by individuals of the two populations, nor used differently against the two models. In addition, the analysis of two or three-syllable sequences also did not uncover any patterns. Moreover, the information theory analysis did not reveal any difference in entropy. Therefore, we do not have evidence for the use of combinatorial rules in orangutans to signify predators.

The lack of significant effects may be due to the relatively low statistical power of analyses with small number of females comprising local populations. However, we sampled all possible females at each population and extending the number of individuals would lead to heterogeneous samples. However, the few obtained significant results were opposite to the predictions built at the start of the study. Moreover, information theory analyses corroborated the results. Thus, conclusions can be drawn with relative confidence.

Interestingly, orangutans within populations showed considerable variability in their alarm call responses. These differences suggest that orangutan alarm call system may not be strictly hard-wired and affect-based (*sensu* Owren and Rendall 2001), as commonly assumed. Indeed, orangutans are suggested to socially acquire some aspects of their alarm call system related to the production of KSQs (e.g. use of instrumental tools during KSQ production, van Schaik et al. 2003; van Schaik et al. 2009; Hardus et al. 2009b). This may also be the case with other types of alarm calls. Moreover, these considerable differences within populations exclude the possibility that habitat differences (e.g. canopy height above 12-20 meters available for escape) affected our results in any crucial manner. These intrapopulational differences, in combination with relatively low numbers of individuals in the sample, limited our ability to interpret in detail the obtained results and to detect biologically meaningful differences in orangutan alarm calls towards different predators, specially so at the site where tigers were familiar. In addition to significant differences in the use of KSQ's, other components of orangutan's anti-predator behavioral responses showed differences between the tiger and the white model. For instance, mean and maximum number of branches thrown at the models was one magnitude order higher towards the tiger than the white model (Lameira, Hardus and Wich, unpublished data). While these latter data are preliminary and lay outside the direct scope of this article, they confirm to some extent that orangutans perceived the models differently and suggest that, as it occurs in monkey species (Kavanagh 1980; Fichtel and Kappeler 2011; Fichtel and Van Schaik 2006), alarm call usage may be flexible and dependent on multiple factors.

The results suggest that the combinatorial rules known to organize monkey alarm calls (Zuberbühler 2002; Arnold and Zuberbühler 2008; Ouattara et al. 2009a, 2009b; Schel et al. 2010) do not apply in a straightforward way to orangutan alarm calls. Moreover, the absence of an island difference in orangutans' response suggests no effect of the size of predator guild on alarm call complexity in this genus, suggesting that the selective pressure due to predation on behavioral and alarm call responses is much stronger in monkeys than in orangutans. This may be a result of the relative larger body size of orangutans compared to monkeys, which reduces the number of predator species as, for instance, raptors do not prey on orangutans and an healthy adult may only be successfully hunted by the largest cat species. If this negative proportional rule between prey body size and predators' number is accurate, one may then expect smaller

nonhuman primates (and perhaps smaller non-primate mammals as well) to present a richer repertoire in terms of combinations' number relatively to bigger species (of the same taxon) when their alarm call system is based on call combinations. This may be addressed by future research.

Future work should also investigate whether the effect of predator guild applies to other great apes, which are on average more terrestrial. If it is confirmed, this suggests that predation may have remained a relatively relaxed selective pressure for ground-dwelling hominins, facilitated by the acquisition of effective material culture (e.g. van Schaik et al. 2003), and subsequent fire control (e.g. Wrangham and Books 2010), even though they must have encountered new predator species both before and after migrations out of the region of origin, including new species of felids, canids and ursids (Hart and Sussman 2008).

Overall, our results are consistent with the view that, within the primate clade, the emergence and use of combinatorial rules may be dependent on the number of critical events in the environment about which the individuals communicate, modulated by the species' ability to learn new calls. Thus, combinatorial rules are positively selected when the number of the species' single alarm calls is lower than the number of predators. This is the case in some monkey species, as above mentioned. On the other hand, because great apes may be able of extending their alarm call repertoire with new calls (Lameira et al. in press-b; Wich et al. 2009b; Hopkins et al. 2007; Hardus et al. 2009b; Wich et al. 2012), and because they are preyed upon by fewer species, the threshold for the emergence and use of combinatorial rules is not reached. In other monkey species than Diana monkeys (*Cercopithecus diana*), such as vervets (*Chlorocebus pygerythrus*), selective pressures seem to operate differently, where each predator species is assigned a distinct single alarm call (Cheney and Seyfarth 1992), instead of a combined-call or call sequence. Further studies are needed to examine the conditions that favor the evolution of such a system rather than the combined-call based alarm system of Diana monkeys. Moreover, the work by Stephan and Zuberbühler (2008) should be replicated with other monkey and non-primate mammals species using combined-calls to assess the general validity of the predation pressure hypothesis within such type of alarm systems and the putative indirect role of body size (as a determinant of predators' number). Only such a comparative database on the presence of combinatorial rules in multiple primate and non-primate

mammal species and the conditions favoring such rules, will provide an evolutionary framework for understanding the evolution of speech combinatorial communication.

### *Acknowledgements*

This study was financially supported by Fundação para a Ciência e Tecnologia, Wenner-Gren Foundation for Anthropological Research, Dr. J.L. Dobberke voor Vergelijkende Psychologie, Lucie Burgers Foundation for Comparative Behaviour Research, Schure-Beijerinck-Popping Fonds, Ruggles-Gates Fund for Anthropological Scholarship of the Royal Anthropological Institute of Great Britain and Ireland, Netherlands Organization for Scientific Research (NWO), Primate Conservation, Inc. and Fundação Calouste Gulbenkian. We thank RISTEK, PHKA, TNGI and BPKEL for research permission in Indonesia and UNAS, BOS and SOCP for supporting the project. We also thank Natasha Nieuwenhuis and the field assistants involved in the experimental work.

Chapter VIII

**Orangutan *instrumental gesture-calls*:  
Reconciling acoustic and gestural speech evolution models**

Adriano R. Lameira, Madeleine E. Hardus and Serge A. Wich

*Published in Evolutionary Biology (2012), 39, 415-418*

## Essay

Call control allows an organism to produce an acoustic signal irrespective of its own underlying emotional state. It is thus a prerequisite to “higher” abilities, such as call imitation, innovation and the use of arbitrary or deceptive calls, and therefore to speech. However, among primates, call control is presumed to be greatly confined to humans (Seyfarth and Cheney 2008). Consequently, there is little agreement about its evolutionary precursors (Christiansen and Kirby 2003b). Essentially two major models and lines of evidence have been proposed; speech evolved (1) as an extension of acoustic communication in non-human primates (e.g. Seyfarth et al. 1980; Arnold and Zuberbühler 2006; Slocombe and Zuberbühler 2005b; Wich et al. 2009b) or (2) from non-human primate gestural communication (e.g. Arbib et al. 2008; Corballis 2003; Rizzolatti and Arbib 1998). These models have been seen as mutually exclusive or as sequential accounts in which calls replace gestures (Brown et al. 1999), however, both face limitations concerning the emergence of call control in our evolutionary lineage. Did call control derive from an *essentially* emotional call use, or from an *essentially* voluntary gesture use, as that of non-human primates? The acoustic model needs to explain how a fundamentally close-ended acoustic system became open-ended (i.e. with limitless number of elements; alike speech). The gestural model needs to clarify the behaviors and respective functional advantages that allowed a shift (or “translation”) from an open-ended gestural system to an open-ended acoustic system.

Other important evolutionary models, such as, on syntax (e.g. Scott-Phillips and Kirby 2010), protolanguage (e.g. Mithen 2005), musilanguage (e.g. Brown et al. 1999), linguistic categories (e.g. Puglisi et al. 2008), increased breathing control (e.g. Maclarnon and Hewitt 2004) and iterated learning (e.g. Smith et al. 2003), some of which merge acoustic and gestural models, such as, on Motherese (e.g. Falk 2004) and frame/content (e.g. MacNeilage 1998), commonly begin with a hypothetical organism that is equipped *a priori* with call control, or overlook the behaviors that may have provided the functional advantages towards call control. We propose that recent orangutan (*Pongo pygmaeus wurmbii*) findings answer and reconcile the limitations of these models. Arguments supporting the above mentioned models are compatible with the view presented.

Recently we have described (Hardus et al. 2009b) how and why wild orangutans use gestures to functionally alter the acoustic characteristics of a particular call type

(sensu Lameira et al. 2010) emitted under disturbing contexts, the kiss squeak (Hardus et al. 2009a). By positioning a hand or holding leaves in front of their lips, wild orangutans lower the maximum frequency (i.e. that of highest dB) but maintain other parameters of the call similar. Evidence suggests that kiss squeaks are under voluntary motor control in orangutans, and when individuals produce these modified variants of the call, they sound as if their body size is bigger than it actually is, reinforcing this impression on a potential predator and potentially deterring it through functional deception.

Kiss squeaks with a hand and on leaves represent, to our best knowledge, the only example of *instrumental gesture-calls* (IGC) in non-human primates. They can be defined as gestures that *modify* oro-laryngeal acoustic production, with or without tools, such as finger-assisted whistling or brass-/woodwind-instrument playing. In order to achieve this acoustic modification, some sort of physical contact between hands/tools and lips, and possibly tongue, is critically required. Mere physical proximity is unlikely to modify a call considerably, as for instance, when “loud speaking” through funneled hands. These gestures are importantly distinct from gestures that produce an acoustic signal themselves, with or without tools, and that can be made during call production. Such *acoustic gesture-calls* have been reported in other ape species (Arcadi et al. 1998) and are possibly present in most non-human primate species, such as when making noisy displays during loud calls and/or alarm calling, by slapping the ground or strongly striking branches. Heuristically, gestures may be considered *additive* in acoustic gesture-calls, whereas gestures in IGC may be considered *multiplicative*.

IGC in hominids multiply the number of call-types comprising the acoustic repertoire in an extremely simple way: one call-type used in combination with different gestures produces new call-types. That is, the potential to augment its innate acoustic repertoire can be achieved solely by means of an ability already present— gesture control. It is very likely that our ape/hominid ancestors would have exploited such “new” repertoire when available, as means to transmit more (graded) information, since cognitive abilities in non-human primates have been demonstrated to be richer and more advanced than their acoustic counterparts (Seyfarth and Cheney 2010).

We hypothesise that IGC, dating back to the hominid-pongid split (9 – 13 m.y.a.; Hobolth et al. 2011) may have provided the direct functional and neural sensory-motor basis towards call control in an early human ancestor essentially lacking this ability, that is, they served as an exaptation for this ability. IGC are remarkable in that they bring into

close temporal, motivational, contextual, anatomical and functional association both the gestural and oro-laryngeal systems of motor control in the communication domain. Hand-assisted feeding, for instance, raises the same associations between gestural and oro-laryngeal systems of motor control but in the foraging domain. IGC comprise therefore, obligatorily, the expression of synchronous activations of multiple neural sensory-motor systems in the ape brain. In the ape cerebral cortex, such activations will mainly occur within regions homologous to the cortical homunculus (that comprises the primary motor cortex, which plays a crucial role in general voluntary motor control) and between the cortical homunculus and other cortical systems involved in the domain of communication, such as those homologous to Broca's and Wernicke's areas (Tagliabata et al. 2011). Such synchronous activations may have provided a neural interface between the brain areas activated, through functional integration and clustering (Tononi et al. 1998a; 1998b), enabling the sharing of abilities which were previously fundamentally restricted or segregated to particular areas. By means of cortical and neural plasticity (Lieberman 2002b), alike for example, use-dependent functional reorganization of sensory cortices (Pantev et al. 1998), this interface would have set the basis for the establishment of enhanced and more resilient short and long distance circuits. Indeed, cortical and neural plasticity is at the basis of hemispheric asymmetries in key areas of the ape and human brain for communicative signaling (Hopkins and Nir 2010; Perani et al. 2011).

As the focus of voluntary control, the cortical homunculus would represent the main stage for these circuit modifications. The number of areas activated in this area and their mutual proximity would add up to form a momentary local hotspot of activations sufficient to ignite neighbouring areas over which there was previously little voluntary control. Namely, circuitry between the respiration, hand, face, lips, and tongue (somatotopic) locations would expand to include that of larynx areas. These circuits would not necessarily be required to be established *de novo*, but instead, would only be required to modestly build and expand on previously existing ones. For instance, a rudimentary but functionally relevant interface between hand, respiration and laryngeal locations (and possibly lips and tongue) is already present in the ape brain, in that use of the right hand for gestures is significantly enhanced when the gestures are accompanied by a call (Hopkins and Cantero 2003). At the same time, pathways between the primary motor cortex and nucleus ambiguus (site of the laryngeal motor-neurons in medulla oblongata), which are specifically interpreted as representing a crucial neural step in gaining call

control (Fitch 2005; Brown et al. 2008), are found in apes but not in monkeys (Kuypers 1958), substantiating the view that an rudimentary interface is already present between systems.

In humans, neuroimaging studies support this evolutionary scenario. For instance, the (somatotopic) location of larynx/phonation area (that with control over intrinsic musculature of the larynx, underlying adduction/abduction and tensing/relaxing of the vocal folds) in the cortical homunculus is adjacent to the lips area and the expiratory area (Brown et al. 2008). This means that in humans, phonation, articulation and respiration are neurologically conjunct. Considering that orangutans have been experimentally demonstrated to exert apt voluntary motor control over lips and respiration (Wich et al. 2009b; Lameira et al. in press-b), it is reasonable to view this conjunction as evolutionarily relevant in humans. While laryngeal musculature may operate in complex ways during (online) speech and other functions (Jürgens 2002; Ludlow 2005), the evolutionary genesis of call control theoretically commenced when the first rudimentary neural signal initiating in the primary motor cortex would be transmitted successfully simply to set the larynx into position during air-flow. The view that neural circuitry flexibility could have successfully achieved this in our ancestors is supported by a phenomenon known in human as motor equivalence, where speakers develop different motor strategies, i.e., use different musculatures, of the larynx to achieve the same voice outcome (Ludlow 2005). Accordingly, IGC could potentially explain why the area of representation of the intrinsic laryngeal muscles has seemingly migrated toward the labial area in humans (Brown et al. 2008). In addition, IGC are in concordance with the increasing literature corroborating that gestures and calls/speech are neurally co-processed (Xu et al. 2009; Bernardis and Gentilucci 2006; e.g. Rizzolatti and Arbib 1998).

At the same time, these bimodal behaviors represent cultural variants of orangutan behavior (e.g. van Schaik et al. 2003). Accordingly, enhanced neural connectivity would have also developed across brain systems in areas involved in processing social information, emotional valence and learning, such as the amygdala and the auditory cortex (Remedios et al. 2009). Thus, brain-language (Deacon 1998), biology-culture (Richerson and Boyd 2005) and music-language premises (Brown 1999) are concordant with the IGC hypothesis.

IGC present a parsimonious route to human-like neurophysiology, increased call control and repertoire size in the earliest stages of speech evolution, but one may

question its relevance based on the phylogenetic distance between orangutans and humans. Three clarifications are required. Firstly, comparison between human, chimpanzee and orangutan genomes shows that some regions of the human genome more closely resemble orangutan's (Hobolth et al. 2011). Although this percentage is approximately 1%, a necessarily bigger percentage is equally similar between humans, chimpanzees and orangutans. While broad genetic underpinnings of speech are not well understood beyond FoxP2 gene (e.g. Enard et al. 2002), the relevance of genetic proximity within hominoids remains equivocal. Secondly, speech is a bio-cultural evolutionary phenomenon (Richerson and Boyd 2005), and therefore, theories must encompass some degree of interaction between social and genetic mechanisms in the acquisition and transmission of communication signals. Orangutans and chimpanzees are the only apes to show extensive cultures in the wild (van Schaik et al. 2003; e.g. Whiten et al. 1999), thus, *both* species represent promising models. Thirdly, the description of IGC in orangutans but (so far) not in chimpanzees may constitute a methodological artifact. While cultural variants between populations have been investigated in wild chimpanzees, this record tends to focus on feeding behavior (Watson and Caldwell 2009). Oppositely, researchers have investigated geographical variation in orangutans' complete call repertoire (Hardus et al. 2009a). These conditions may have benefited the description of IGC more readily than in chimpanzees. There are nonetheless anecdotes suggesting that IGC may be part of their repertoire, such as the use of a hand in front of the mouth to muffle a call, as described by Jane Goodall (Deacon 1998).

This essay presents a new view on the earliest stages of speech evolution, based on orangutan IGC. It builds on the concept that enhanced linguistic ability cannot be totally differentiated from enhanced motor activity (Lieberman 2002a), and argues that IGC may have constituted speech exaptations, providing functional advantages in a human ancestor essentially lacking call control but allowing the emergence of the neural and communicative basis for subsequent selection favouring basic abilities for speech. This view provides a new concrete model organism, similar in its abilities of (1) call control, (2) call repertoire size and (3) reliance on social learning as those observed in orangutans for future speech evolution models.

### *Acknowledgements*

ARL was financially supported by Fundação para a Ciência e Tecnologia (SFRH/BD/44437/2008). We thank Carel van Schaik, Asif Ghazanfar and two anonymous reviewers for comments on previous versions of the manuscript.



## Chapter IX

# General discussion

After more than one and a half centuries of theorizing (Darwin 1859), unanswered questions remain on how human speech could have evolved from an ancestral state alike the call repertoire of extant nonhuman primates. This study sought to learn how much of speech characteristics can be found in orangutan call communication, to provide a better understanding of possible speech precursors and to help disentangling perceived from factual differences with human speech. To pursue this, this study used two comparative tools: (1) the use of an ethological framework for the investigation of orangutan call communication and (2) the use of terminology and concepts borrowed from linguistics and semiotics, setting a comparative platform with speech based on articulatory and acoustic homology.

The application of the first tool implied avoiding a priori (mis)conceptions about orangutan capacities in the call domain, as occurred in part in previous great ape language projects (Savage-Rumbaugh et al. 1978; Savage-Rumbaugh et al. 1985; Gardner and Gardner 1969; Premack 1971; Rumbaugh 1977; Terrace et al. 1979; Patterson and Linden 1981; Furness 1916; Hayes 1951; Hayes and Hayes 1952; Miles 1993), such as the notion that great apes should be able to perform fluent speech when properly thought. The application of this first tool entailed, thus, primarily the appreciation of the calls that orangutans in fact produce, their use and function, and the boundaries of orangutan capacity for call learning and flexible use.

The application of the second tool allowed the use of this ethological framework while simultaneously maintaining a comparative standpoint. For instance, orangutan calls were categorized as vocalizations and sounds when either involving or not involving patterned vocal-fold action, and opened- and closed-mouth phases during production, respectively. These calls may thus be classified as vowel- and consonant-like respectively, since human vowel and consonants are likewise voiced and unvoiced (excluding voiced consonants), and produced on opened- and closed-mouth phases during production. The use of linguistic and semiotic terminology was also applied to variation in the production of the same call between communities (i.e. accents) and variation in the call repertoire between communities (i.e. dialects). This loan of terms was based on the similar patterns

in variation as found in humans and, in great apes, expressed articulatory and acoustic resemblance between the species call production due to the close relatedness with humans (Hobolth et al. 2011; Prufer et al. 2012; Scally et al. 2012)

In addition, the second comparative tool provided a comparative framework to the investigation of orangutan call communication for the presence of speech building blocks, as defined by the linguist Charles Hockett (Hockett and Altmann 1968). Hockett described thirteen essential features of speech, five shared with animal calls, and eight unique to human speech, of which the presence in great apes had not yet been determined. This study examined the presence of six of these eight features in orangutan call communication: “traditional transmission” (chapter 2 and 3), “prevarication” (chapter 4), “arbitrariness” (chapter 5), “learnability” (chapter 6), “semanticity” (chapter 7) and “patterning” (chapter 7).

### *Traditional transmission*

Traditional transmission maintains that speech is learned within social groups. When the feature of traditional transmission is found in animal call communication, it follows that learning biases within and between animal communities will give rise to acoustic variation that is not attributable to genetic or environmental physical variation. While the ability of learning new calls has been demonstrated across taxa, from song birds (Aronov et al. 2008; Boughman and Moss 2003; Wilbrecht and Nottebohm 2003) to bats (Knörnschild et al. 2006; Boughman and Moss 2003; Boughman 1998), from aquatic mammals (Foote et al. 2006; Janik and Slater 1997; Noad et al. 2000) to elephants (Poole et al. 2005), the common and traditional assumption is that nonhuman primates are unable of learning new calls, that is, that they exert no flexibility in the domain of call acquisition and their calls are genetically hard-wired. To address this question, this study firstly verified the presence of variation between communities as defined in humans (i.e. accents and dialects) across terrestrial mammals (Chapter 2). Even though this group of animals has motivated less interest on this topic than other animal groups, they share human’s environment and have not been imposed major selectively pressures arising from aquatic (e.g. aquatic mammals) or aerial (e.g. birds) life-styles. Hence, this study provided the first review focused on these animal taxa illustrating the presence or prevalence of each of these forms of acoustic variation between communities, along with

putative underlying factors (i.e. genetic, environmental and/or social). This review showed that genetic, environmental and social factors affect variation in the same call differently per taxon of terrestrial mammals. In nonhuman primates, social factors represent the major driver of acoustic variation in accents between populations. Although accents are relatively common across terrestrial mammals, dialects are extremely rare, with only four registered occurrences, all of which have been offered a socially-driven explanation, and half of which have been described in great apes, including orangutans. Thus, this review detected potential traces of traditional learning for call production across terrestrial mammals taxa, usually concerning accents, but also concerning dialects and especially in nonhuman primate species.

To resume investigating the potential presence of traditional learning in orangutans, an assessment was made of the presence of dialects between orangutan communities (Chapter 3). For this purpose a method was designed and used to disentangle genetic, ecological and social factors potentially underlying call repertoire geographic differences. To control for differences in the physical environment between orangutan sites, we examined presence/absence variation of two close-range calls, that is, calls that only suffer minor physical degradation between producer and receiver (cf. Lameira and Wich 2008). Variation in population-specific calls emitted in nest-building and infant retrieval contexts did not correlate with genetic differentiation between orangutan communities. Moreover, ecological variation between sites could also not explain the presence/absence pattern found for these calls between communities. This demonstrated that production of these calls was not fully genetically canalized, and that each call likely was invented and subsequently socially transmitted within each population, generating local call traditions. These results represent a major advance in the study of great ape calls, implying that traditional learning is present in great ape call communication and that capacity to learn at least some new calls is shared between humans and human's closest relatives.

### *Prevarication*

Prevarication expresses the speech ability to make false statements, to deceive, involving the manipulation of a given shared principle in order to fool other individuals. Within the animal call domain, this requires that call production is not a direct manifestation of the individual's affective state. However, the use of deceptive call production is rare in

mammals (Fitch and Reby 2001; Matrosova et al. 2007), and in the cases that have been shown concern the use of exceptionally low calls produced to increase the perceived body size of the producer. Nevertheless, to our knowledge, this was never examined in nonhuman primate call communication. We examined this possibility at one orangutan population in the use of one of orangutan's alarm calls, the kiss squeak (KSQ), and its two variants – hand and leaf KSQ (Chapter 4). The function of these variants remained hitherto unknown. At this population, orangutans cumulatively lowered the loudest frequency (Hz) of KSQ by positioning a hand or further positioning leaves in front of their lips during KSQ production. If this effect was indeed used with a deceptive function then, KSQ frequency should relate to body size, and the use of low/modified KSQs should relate to perceived threat, where a deceptive function would be most advantageous. Indeed, KSQs' loudest frequency was negatively correlated to body size, where the largest adult male morph (i.e. flanged males) produced the lowest frequencies across orangutan sex-age classes. Moreover, higher percentages of orangutans unhabituated to the presence of humans produced low/modified KSQ variants towards humans, where the lowest variants (i.e. leaf KSQ) were solely emitted by unhabituated individuals. The use of orangutan KSQs satisfies thus the predictions for deceptive use of calls, constituting the first indication for this kind of call use in primates and in nonhuman animals by intermediary of tool-use (i.e. leaves). At the same time, KSQ variants were produced without any particular order when they were produced within the same bouts, confirming that no escalating disturbance underlined the production of KSQ variants. This evidence for the deceptive use of orangutan call differs from previous non-primate mammal examples, since the latter were underlined by anatomical innovations (e.g. descendent larynx in deer (Fitch and Reby 2001)), while the former was underlined by manipulative and tool innovations (KSQ variants comprise the cultural repertoire of wild orangutans (Krützen et al. 2011)). Because skilled animal tool-use requires practice for proper performance (Osvath and Osvath 2008; von Bayern et al. 2009; de A. Moura and Lee 2004; Inoue-Nakamura and Matsuzawa 1997), the deceptive use of KSQs by orangutans strongly likely represents a local tradition, and so, provides further evidence for the presence of traditional transmission (see above) in orangutan call communication, besides prevarication.

### *Arbitrariness*

Arbitrariness defines the absence of a direct connection between a call and its function or meaning. Despite being commonly used in linguistics, this concept was theoretically further developed by semiotics, which operationalized the concept by defining separately the communication signal per se, the “signifier”, and its potential function, the “significant” (Saussure et al. 1983). This allowed separate analyses between these two elements. For instance, according to these concepts, there are several examples in animal call communication for acoustically different calls (signifiers) with a similar function (significant) between conspecific populations (Wright 1996; Noad et al. 2000; van Schaik et al. 2009). However, it remains unknown if different functions (significants) can be attributed to the same call (signifier) within the animal kingdom. The only available example relates to the gestural domain in chimpanzees (Boesch 2003). To explore the possibility of arbitrariness of the same orangutan signal (Chapter 5), this study replicated the assessment on the use and function of KQs and its variants (as in Chapter 4) in a second population. These populations were of the same sub-species and lived in partly similar habitat types. This allowed controlling for major differences in KSQ use and function that could result from genetic differentiation between populations or major habitat differences between sites. At the second population the KSQ and only one, instead of two, KSQ variant were used in the population. The use of this variant (signifier) did not convey increased body size and did not relate to perceived threat, indicating that its function (significant) is distinct from that at the former population, therefore providing evidence for arbitrariness of the second type: the same call used for two different functions. To our knowledge, this constitutes the first demonstration of an animal signal involving call production with arbitrary functions. Moreover, it shows that the functions attributed to a particular signal involving call production are likely learned within populations. The use of semiotic terminology is adequate in the examination of the different facets of arbitrariness in animal communication.

### *Learnability*

Learnability defines the capacities of members of one call system to learn to produce calls from another system. Similarly to traditional transmission (see above), learnability depends on the apt control over call production allowing the acquisition of new calls into

one's repertoire. The main difference with traditional transmission resides on the fact that, when applied to animals, learnability mainly refers to the capacity to learn new calls from a non-conspecific, that is, a tutor from another species. The comprehension of calls from another species has been demonstrated in nonhuman primates (Zuberbühler 2000b). The production of another species calls is, however, restricted to a few species of aquatic mammals (Janik and Slater 1997) and birds (Dobkin 1979), which may imitate human utterances (spontaneously or after training) or other species' calls. A recent study demonstrated for the first time a case of spontaneous acquisition of a human sound by an orangutan – whistling (Wich et al. 2009b). This reference study constitutes thus a significant step in the progress of the study of primate call communication and warrants further investigation (Chapter 6). These new data show that several orangutan individuals are able to whistle and that this is not restricted to a particular orangutan species or hybrids. The motor control over whistling production is precise and voluntary and individuals may control the number and duration of whistles. However, when comparing several orangutans, each individual produces whistles on its own particular way, either positioning their lips differently and/or whistling in the air in-flow or out-flow. This refutes the view that whistling could represent an innate call, as one would accordingly expect no differences in whistling technique. Together with evidence for traditional transmission, prevarication, arbitrariness (see above), learnability shows that orangutans are able to acquire calls, a capacity that was fundamentally denied at the start of this study.

### *Semanticity and patterning*

Semanticity defines the property of calls that can be linked to specific meanings. Patterning describes the ability to combine a finite number of call elements to create new, composed, elements. Both features have been shown in nonhuman primates (semanticity: (Seyfarth et al. 1980; Slocombe and Zuberbühler 2005b, 2006), patterning (Arnold and Zuberbühler 2006, 2008; Clarke et al. 2006; Candiotti et al. 2012a; Crockford and Boesch 2005)). Nevertheless, only for monkeys and lesser apes there is evidence that call-combinations or sequences (patterning) may refer to objects or events in the individuals' environment (semanticity). These cases have only been demonstrated in predation contexts, but great ape call repertoires in alarm contexts remain unexamined. This examination is, however, necessary to determine the potential evolutionary

continuity of monkey and lesser ape evidence with human speech, which fully relies on “patterned semanticity”. At the same time, a recent hypothesis based on alarm call data from a monkey species has proposed that predation pressure (expressed by the number of a species’ predators) increases the number of different call combinations used by monkeys, possibly operating as a selection pressure towards ever more (complex) call-combinations within the human lineage. Hence, it is relevant to ascertain the possibility of semantic communication in great ape (combined) alarm calls in relation to differences in predator guild composition (Chapter 7). In contrast to evidence from monkey and lesser ape species, orangutans do not organize alarm calls in specific ways according to different predators. In addition, populations differing in predator guild composition do not differ in the way they organize alarm calls. Together, this indicates that alarm call-combinations or sequences with semantic content in monkey and lesser ape do not have a direct orangutan counterpart. This suggests that predation may have only represented a relaxed selective pressure on call communication after the evolutionary split with lesser ape ancestors. Similar to monkey and lesser ape species for which evidence on semantic call-combinations is available, orangutans are arboreal. They differ, however, in their much larger body size, which reduces the number of species capable of preying on orangutans, reducing in turn the need to combine alarm calls in new ways conveying semantic information about different predators. Assuming that human ancestors maintained an ape-like body size, that their environment presented up to 40% of woody cover in the past 6 million years (Cerling et al. 2011), and that their locomotor repertoire included a substantial amount of climbing (up to *Australopithecus afarensis*) (Green and Alemseged 2012), it is reasonable to assume that predation pressure probably did not pose significant selection pressure on human speech evolution, in particular on semanticity and patterning.

#### *New hypothesis for speech evolution*

As presented above, orangutan behavioural flexibility extends over call use, production and acquisition. Call comprehension remained unassessed. Studies focusing on call comprehension typically rely on playback experimental methods to examine how calls are perceived by conspecifics, but at present, there is not sufficient data on orangutan playback experiments to incorporate call comprehension. The present findings challenge

the traditional assumption that primate calls are inflexible in the domain of control and acquisition. Orangutans can, for instance, control and manipulate call production via tool use, generating new variants, and learn new calls into their repertoire, generating local-specific call repertoires. Moreover, orangutan call communication shows evidence of multiple speech building blocks. Nevertheless, the extent of great ape call use, control and acquisition seems quantitatively much less extensive than humans. Speech comprises multiple learned elements and combinatorial rules that confer limitlessness to humans' repertoire. As such, this open-endedness of speech has been proposed to be uniquely human (Hauser et al. 2002), and possible evolutionary precursors have been suggested outside call communication, namely in the gestural domain, to circumvent this difference in call repertoire between great apes and humans. The gestural and acoustical hypotheses are usually assumed to be mutually exclusive. Both lines of thought face, however, limitations in explaining the emergence of an ever expanding communicative repertoire in the human lineage. Basically, this is due to the fact that none of the hypotheses is able to present a concrete great ape trait or behaviour which, due to its proximate adaptation advantages, could have been targeted by positive selective pressures leading to speech (Chapter 8). I present the new hypothesis that orangutan tool-modified calls, as described previously in this study (Chapter 4 and 5) constitute a promising speech precursor. Orangutan tool-modified calls represent instrumental gesture-calls, signals composed by a gestural and a call component where the former modifies the latter. This multimodality is expressed by synchronous activations of multiple neural sensory-motor systems in the ape brain, creating a momentaneous hotspot able of igniting neighboring areas, such as those responsible of voluntary call control. That is, via gestural control, which is developed in nonhuman primates, individuals may bring increased control aptitude over call production, whilst present (see above), is less developed. Thus, the combination of a call with modifying gestures represents a relative easy and straightforward way to increase call repertoire in the absence of developed call control, allowing the selective pressures favoring larger repertoires to take action. This hypothesis is compatible with multiple other hypotheses presented for language and speech evolution, and it reconciles the gestural and acoustic hypothesis for cumulative increase in control over call production and repertoire size during speech evolution.

### *Conclusion*

The use of linguistic and semiotic terminology is beneficial for primatological (and probably also for broader) comparative studies focusing on speech evolution. The borrowing of terms provides a well-defined comparative approach to speech, while no a priori human-like capacities are necessarily attributed to nonhuman subjects, preventing anthropocentricity. When applying this approach to examine orangutan call communication, several building blocks of speech are not recognized as factual, but instead only constituted perceived, discontinuities between human speech and nonhuman primate call communication. That is, speech basic ingredients, namely “traditional transmission” (chapter 2 and 3), “prevarication” (chapter 4), “arbitrariness” (chapter 5) and “learnability” (chapter 6), but so far not “semanticity” (chapter 7) and “patterning” (chapter 7), are found in great ape call communication. Renewed effort in the study of great ape calls will remain valuable for the reconstruction of speech evolution and the selective forces acting upon each of these building blocks. The identification of shared features between the two communication systems of apes and humans allows speech precursors to be subsequently pursued, providing ever more pieces to the puzzle of speech evolution.



## Chapter X

# Summary

Speech is a human hallmark, yet its evolution is little understood. It remains largely unknown which features of call communication in our closest relatives – the great apes – may have constituted the starting point of human speech evolution. Over the last four decades, scientific thought has moved away from the notion that call communication in great apes is valuable to answer this puzzle. As a result, other animal models have received much focus and research effort, and it has been suggested that in nonhuman primates gestural communication represents a better candidate as speech precursor. Although heuristically legitimate, these notions bypass the fact that speech operates fundamentally in the acoustic modality. Hence, any valid speech precursor, found in non-primate taxa or non-acoustic modalities, must have been firstly present and functionally relevant in human ancestors and, when present outside the acoustic domain, it must have expanded or transferred into the call communication. Therefore, the main focus of this study was to investigate the extent to which speech building blocks can be found in orangutan call communication.

Orangutans represent the oldest, and therefore humans' most distant, great ape lineage, but old and recent studies indicate they represent a promising model species in this unraveling of speech evolution. While behavioral flexibility in call use, comprehension and production may be found across nonhuman primates, the rare cases of flexibility in call acquisition predominantly concern orangutans. That is, orangutans show some evidence for the capacity of learning and incorporating new calls into their repertoire. Therefore, I investigated orangutan call communication in both natural and naturalistic conditions. Instead of seeking to teach human words, I take as a starting point the calls individuals actually produce. From these calls, I identified speech features potentially present in orangutans in the absence of full-blown speech. Moreover, I borrowed terminology from linguistics and semiotics (social sciences dedicated to the study of human speech and signs, respectively) to establish a direct comparative approach based on structural and behavioral homology between great apes and humans. Accordingly, I examined speech building blocks traditionally presumed to represent uniquely human features, since any valid hypothesis for speech evolution must explain the emergence of

these features. These advanced communication features are: “traditional transmission”, “prevarication”, “arbitrariness”, “learnability”, “semanticity” and “patterning”.

The first speech feature addressed is “traditional transmission” (chapter 2 and 3), stating that individuals learn their call repertoire from their conspecifics within social groups. This leads to variation between groups and two different potential types of variation: “accents” when the same call-type varies acoustically between populations, or “dialects” when some call-types are site-specific and absent in other sites (chapter 2). Accents commonly occur in terrestrial mammals, either due to genetic, environmental and/or social factors. These factors interact differently per taxon that shows accents, but social factors represent the most important factor in the emergence of accents in nonhuman primates. In contrast, dialects are extremely rare among terrestrial mammals and encompass 4 reported cases, two of which concern great apes, namely chimpanzees and orangutans. A further examination of the presence of dialects in orangutans is thus of interest (chapter 3). The presence and absence patterns of population-specific orangutan calls produced during nest-building or mother-infant contact calls could not be satisfactorily explained by genetic or ecological variation between sites, supporting the view that these calls were most likely invented and subsequently spread via social learning within the populations where they are present. The observation that two populations may have distinct calls for the same function also denotes that there is no particular relationship between the type of calls orangutans produce to this function, implying that these calls are arbitrary.

Subsequently, I investigate “prevarication”, the ability to make false statements. This involves the manipulation of a particular shared communicative feature, in order to deceive other individuals (chapter 4). Orangutans at the Tuanan population (Central Kalimantan, Indonesia) produce three variants of one of their alarm sounds: unaided, hand and leaf kiss-squeaks. Consistently with a general rule in bioacoustics, the frequencies (Hz) of unaided kiss-squeaks are negatively proportional with body size, where larger individuals produce lower kiss-squeaks. The use and positioning of a hand or leaves in front of the lips during kiss-squeak production reduced the frequency of kiss-squeaks (but no other acoustic measure of kiss-squeaks), potentially signaling a larger body size. Unhabituated orangutans produced the same proportions of unaided kiss-squeaks towards humans as habituated individuals. However, they proportionally emitted more hand kiss-squeaks than habituated individuals and were the only ones to produce

leaf-squeaks. That is, the production of modified kiss-squeak variants was correlated with perceived threat by orangutans, and these calls conveyed larger body size than their actual size.

I then addressed “arbitrariness”, a characteristic of a communication system where there is no direct connection between the signal and its function. Since orangutans can produce arbitrary calls for the same function (see chapter 3), I was interested in investigating whether orangutans can also produce the same call types for arbitrary different functions (chapter 5), providing a complete approach to the examination of arbitrariness. For this purpose I analyzed the use and potential function of unaided, hand and leaf kiss-squeaks at another population of orangutans (Cabang Panti, West Kalimantan, Indonesia), replicating the analyses previously made at Tuanan (chapter 4). These two populations belong to the same sub-species and live in partly similar habitats. At Cabang Panti, unaided kiss-squeaks’ frequencies presented the same negative proportion in relation to body size, potentially allowing body-size manipulations by the individuals via the use of hands or leaves. Nonetheless, hand kiss-squeaks were extremely rare. More importantly, the regularly emitted leaf-use did not result in a lower frequency, and therefore did not convey a larger body size. Moreover, leaf kiss squeaks were not related to perceived threat, since they were not used more often by unhabituated individuals. This difference between populations in the use of the same call implies that the functions attributed to leaf kiss-squeaks are arbitrary.

Subsequently, I focused on “learnability”, stating that users of one system of call communication may learn elements from another species’ system. This ability is traditionally assumed to be absent in nonhuman primates, however, recently an orangutan has been reported to have acquired a human sound – whistling (Wich et al. 2009b). This case represents one of the clearest examples of interspecific learning by a nonhuman primate. To further investigate orangutan whistling, I replicated and extended the previous reference study (chapter 6). Currently, 10 whistling orangutans are known in Europe and USA, indicating that orangutan whistling is not an oddity. One particular orangutan, Ujian, correctly matches single, double and triple whistles produced by a human experimenter, confirming that whistling is not an affective and reflexive display but that, contrarily to most nonhuman primate calls, is under voluntary control. Whistling recordings from 3 different orangutans were compared qualitatively and quantitatively. Each individual showed specific whistling styles resulting in acoustically distinct whistles.

This can be explained when each individual was subject to different learning opportunities, indicating that orangutan whistling is experience-dependent. Altogether, these results support the view that whistling in orangutan is acquired through learning.

I then addressed “semanticity” and “patterning”, respectively stating that some call units have specific meanings and that units may be combined into larger meaningful units. While these features have been demonstrated in several monkey species, they remain unassessed in great apes. Such data are required to verify whether monkey semantic and patterned calls persisted in the human evolutionary lineage. I investigated specifically the emission rates of different orangutan alarm calls towards a predator (tiger) model and a control (white) model in two populations, one where tigers are extinct and one where tigers hunt for orangutans (chapter 7). There were no differences in call rates between models at both populations, indicating that calls were not predator-specific. Moreover, there were no differences in call rates between populations, showing that the size of predator guild did not affect the use of alarm calls by orangutans. Therefore, monkey semantic and patterned calls in alarm contexts do not appear to have direct counterparts in great apes.

Altogether, from the six speech features examined in this study (traditional transmission, prevarication, arbitrariness, learnability, semanticity and patterning), four were found in orangutans. Semanticity and patterning are not present in the alarm call system of orangutans, suggesting that these speech features should be investigated in great ape calls used in other contexts, such as foraging. Within orangutan call repertoire from wild populations, kiss-squeaks represent promising calls for future studies interested in addressing new aspects of behavioral flexibility in the call domain of great apes. In captivity, calls are solely produced by particular individuals may provide in the future new insight into the mechanisms underlying call learning in great apes.

Based on the findings of the previous chapters, I propose that some examples of orangutan call communication may indicate features of precursors of speech (chapter 8), namely orangutan hand and leaf kiss-squeaks. These calls represent unique multimodal signals, involving the action of hands and the manipulation of objects (i.e. leaves) synchronously with call production. As such, these calls bring together the gestural and call domain of great apes, demonstrating that theoretic lines focusing on each domain are reconcilable. This presents a new viewpoint over putative speech precursors, which is fully compatible with data supporting either gesture- or call-based hypotheses. In addition,

orangutan hand and leaf kiss-squeaks present a means by which a human ape-like ancestor with a limited capacity to learn new calls could have expanded its communicative repertoire involving call production. The “simple” combination of a call with a modifying gesture produces a new call variant, triggering positive selection towards an increased repertoire by which individuals could transmit more or more diverse information. As such, multimodal signals in our ancestors involving synchronous gestures and calls could have provided the basis for the emergence of selective pressures towards subsequent steps in speech evolution. By further studying the features of great ape call communication, combined with our understanding of gestural communication, we can deduce the likely starting points of human speech and language evolution in our hominin ancestors.

Renewed effort in the study of great ape call use, production and acquisition will remain valuable for the reconstruction of speech evolution. The identification of shared features between human and great ape communication systems allows the identification of putative speech precursors. Such features may then be further examined with the potential of providing ever more pieces to the puzzle of speech evolution. The features explored in this study in orangutans may be studied in other great apes as well, and linguistic and semiotic concepts may further provide important cues for the search and investigation of yet unexamined shared features between human speech and great ape call communication.



## Appendices

Table S1. Emission rates ( $\text{min}^{-1}$ ) of each syllable emitted per individual at both populations.

		KSQ		GR		GK		CXC		KSQ+GR		KSQ+GK		KSQ+CXC		S		C	
Ketambe (Sumatra)	Yet	0.1	0	0	0	1.4	0	0.1	0	0.1	0	3.7	0	0.7	0	1.6	0	4.5	0
	Kely	3.6	0.1	0	0.2	0	0	0	0	0	0	0	0	0	0	3.6	0.3	0	0
	Puji	0.1	0	0	0	1.1	0	0.6	0	0	0	3.9	0	0.6	0	1.8	0	4.5	0
	Elisa	1.3	0.9	0	0	0.2	3.1	0	0	0.3	0.2	5.3	7.3	0.1	0	1.5	4	5.7	7.5
	Sina	0.4	0	0.6	0	1.7	0	0.9	0	1.4	0	2.2	0	0.1	0	3.6	0	3.7	0
	Chris	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Tuanan (Borneo)	Jinak	4.8	1.4	0	0	0	0.1	0.2	0	0	0.6	1.7	3.9	1.7	0	5	1.5	3.4	4.5
	Juni	1.6	0.2	0	0	0.2	0	0	0	0.2	0	3.8	0.2	0.1	0	1.8	0.2	4.1	0.2
	Mindi	3.8	3.1	0	0	0	0	0	0	0	0.1	10	6.5	0	1	3.8	3.1	10	7.6
	Kery	10.4	3.2	0	0	0	0	0	0	0	0.1	0	0.2	0	0.8	10.4	3.2	0	1.1
	Sumi	4.3	0	0	0	0	0	0	0	0	0	1.3	0	0	0	4.3	0	1.3	0

S – single call, C – combined call

Grey columns – tiger model

White columns – white model

Table S2. Emission rates ( $\text{min}^{-1}$ ) of the eight two-syllable sequences used by at least two individuals in a single population in response to the same model, and used at least five times by the same individual.

		GR-		KSQ+GR-		KSQ-		GR-		KSQ+GR-		KSQ+CXC		KSQ+GR-		KSQ-	
		GR		KSQ+GR		KSQ		KSQ+GR		GR		-KSQ+GR		KSQ		KSQ+GR	
Ketambe (Sumatra)	Yet	0.8	0	2.4	0	0	0	0.3	0	0.3	0	0.4	0	0.1	0	0.1	0
	Kely	0	0	0	0	2.3	0	0	0	0	0	0	0	0	0	0	0
	Puji	0.1	0	2.6	0	0	0	0.6	0	0.7	0	0.2	0	0	0	0	0
	Elisa	0	1.3	4.4	5	0.5	0.3	0.2	1.6	0.1	1.5	0	0	0.2	0.4	0.2	0.4
	Sina	0.3	0	0.7	0	0	0	0.9	0	0.5	0	0	0	0.1	0	0.3	0
	Chris	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Tuanan (Borneo)	Jinak	0	0	1	3	3.9	0.6	0	0.1	0	0.1	0.6	0	0.1	0.5	0.1	0.4
	Juni	0	0	3.1	0.1	0.7	0	0.1	0	0	0	0.1	0	0.3	0.1	0.2	0
	Mindi	0	0	7.8	4.4	1.4	1.6	0	0	0	0	0	0.8	2.1	1	2	0.9
	Kery	0	0	0	0	9.4	1.9	0	0	0	0	0	0.1	0	0.2	0	0
	Sumi	0	0	0.5	0	2.8	0	0	0	0	0	0	0	0.7	0	0.6	0

Grey columns – tiger model

White columns – white model

Table S3. Emission rates ( $\text{min}^{-1}$ ) of the six three-syllable sequences used by at least two individuals in a single population in response to the same model, and used at least five times by the same individual.

		KSQ+GR-		KSQ-		KSQ-		KSQ+GR-		KSQ+GR-		KSQ+GR-	
		KSQ+GR-		KSQ-		KSQ+GR-		KSQ-		KSQ-		KSQ+GR-	
		KSQ+GR		KSQ		KSQ+GR		KSQ		KSQ+GR		KSQ	
Ketambe (Sumatra)	Yet	1.5	0	0	0	0	0	0	0	0.1	0	0.1	0
	Kely	0	0	1.4	0	0	0	0	0	0	0	0	0
	Puji	1.7	0	0	0	0	0	0	0	0	0	0	0
	Elisa	3.7	3.4	0.3	0.2	0.2	0.3	0	0.1	0.1	0.3	0.2	0.3
	Sina	0.2	0	0	0	0.2	0	0	0	0.1	0	0	0
	Chris	0	0	0	0	0	0	0	0	0	0	0	0
Tuanan (Borneo)	Jinak	0.6	2.4	3.1	0.3	0.1	0.3	0.1	0	0	0.4	0.1	0.4
	Juni	2.8	0	0.4	0	0	0	0	0	0.2	0	0.1	0.1
	Mindi	6.2	3	0.7	0.6	1.4	0.4	0.5	0.6	1.6	0.4	1.5	0.7
	Kery	0	0	8.5	1.2	0	0	0	0.1	0	0	0	0
	Sumi	0.2	0	2	0	0.2	0	0.4	0	0.3	0	0.2	0

Grey columns – tiger model

White columns – white model



## References

- Anderson, M. J., Ambrose, L., Bearder, S. K., Dixon, A. F., & Pullen, S. (2000). Intraspecific Variation in the Vocalizations and Hand Pad Morphology of Southern Lesser Bush Babies (*Galago moholi*): A Comparison with *G. senegalensis*. *International Journal of Primatology*, *21*(3), 537-555, doi:10.1023/a:1005400205038.
- Arbib, Michael A., Liebal, K., & Pika, S. (2008). Primate Vocalization, Gesture, and the Evolution of Human Language. *Current Anthropology*, *49*(6), 1053-1076, doi:10.1086/593015.
- Arcadi, A., Robert, D., & Boesch, C. (1998). Buttress drumming by wild chimpanzees: Temporal patterning, phrase integration into loud calls, and preliminary evidence for individual distinctiveness. *Primates*, *39*(4), 505-518, doi:10.1007/bf02557572.
- Arcadi, A. C. (1996). Phrase structure of wild chimpanzee pant hoots: Patterns of production and interpopulation variability. *American Journal of Primatology*, *39*(3), 159-178, doi:10.1002/(sici)1098-2345(1996)39:3<159::aid-ajp2>3.0.co;2-y.
- Armstrong, K. N., & Coles, R. B. (2007). ECHOLOCATION CALL FREQUENCY DIFFERENCES BETWEEN GEOGRAPHIC ISOLATES OF RHINONICTERIS AURANTIA (CHIROPTERA: HIPPOSIDERIDAE): IMPLICATIONS OF NASAL CHAMBER SIZE. *Journal of Mammalogy*, *88*(1), 94-104, doi:10.1644/06-mamm-a-115r1.1.
- Arnold, K., & Zuberbühler, K. (2006). Semantic combinations in primate calls. *Nature*, *441*(7091), 303-303, doi:10.1038/441303a.
- Arnold, K., & Zuberbühler, K. (2008). Meaningful call combinations in a non-human primate. *Current Biology*, *18*(5), R202-R203.
- Arnold, K., & Zuberbühler, K. (2012). Call combinations in monkeys: Compositional or idiomatic expressions? *Brain and Language*, *120*(3), 303-309, doi:10.1016/j.bandl.2011.10.001.
- Aronov, D., Andalman, A. S., & Fee, M. S. (2008). A Specialized Forebrain Circuit for Vocal Babbling in the Juvenile Songbird. *Science*, *320*(5876), 630-634, doi:10.1126/science.1155140.
- Arora, N., Nater, A., van Schaik, C. P., Willems, E. P., van Noordwijk, M. A., Goossens, B., et al. (2010). Effects of Pleistocene glaciations and rivers on the population structure of Bornean orangutans (*Pongo pygmaeus*). *Proceedings of the National Academy of Sciences*, *107*(50), 21376-21381, doi:10.1073/pnas.1010169107.
- Aspetsberger, F., Brandsen, D., & Jacobs, D. S. (2003). Geographic variation in the morphology, echolocation and diet of the little free-tailed bat, *Chaerephon pumilus* (Molossididae). *African Zoology*, *38*(2), 245-254.
- Barclay, R. M. R., Fullard, J. H., & Jacobs, D. S. (1999). Variation in the echolocation calls of the hoary bat (*Lasiurus cinereus*): influence of body size, habitat structure, and geographic location. *Canadian Journal of Zoology*, *77*(4), 530-534, doi:10.1139/z99-008.
- Bastian, M. L., Zweifel, N., Vogel, E. R., Wich, S. A., & van Schaik, C. P. (2010). Diet traditions in wild orangutans. *American Journal of Physical Anthropology*, *143*(2), 175-187, doi:10.1002/ajpa.21304.
- Bernardis, P., & Gentilucci, M. (2006). Speech and gesture share the same communication system. [doi: DOI: 10.1016/j.neuropsychologia.2005.05.007]. *Neuropsychologia*, *44*(2), 178-190.
- Bertelli, S., & Tubaro, P. L. (2002). Body mass and habitat correlates of song structure in a primitive group of birds. *Biological Journal of the Linnean Society*, *77*, 423-430.
- Biben, M., & Symmens, D. (1991). Playback Studies of Affiliative Vocalizing in Captive Squirrel Monkeys: Familiarity as a Cue to Response. *Behaviour*, *117*(1/2), 1-19.

- Biro, D., Inoue-Nakamura, N., Tonooka, R., Yamakoshi, G., Sousa, C., & Matsuzawa, T. (2003). Cultural innovation and transmission of tool use in wild chimpanzees: evidence from field experiments. *Animal Cognition*, *6*(4), 213-223, doi:10.1007/s10071-003-0183-x.
- Blake, J., & Fink, R. (1987). Sound-meaning correspondences in babbling. *Journal of Child Language*, *14*(02), 229-253, doi:doi:10.1017/S0305000900012903.
- Bluff, L. A., Kacelnik, A., & Rutz, C. (2010). Vocal culture in New Caledonian crows *Corvus moneduloides*. *Biological Journal of the Linnean Society*, *101*, 767-776.
- Blumstein, D. T. (1999). Alarm calling in three species of marmots. *136*, 731-757.
- Blumstein, D. T. (2006). The multipredator hypothesis and the evolutionary persistence of antipredator behavior. *Ethology*, *112*(3), 209-217, doi:10.1111/j.1439-0310.2006.01209.x.
- Boer, B. d. (2009). Acoustic analysis of primate air sacs and their effect on vocalization. *Journal of the Acoustical Society of America*, *126*(6), 3329-3343.
- Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glott International*, *5*(9/10), 341-345, doi:citeulike-article-id:4016976.
- Boesch, C. (2003). Is culture a golden barrier between human and chimpanzee? *Evolutionary Anthropology*, *12*(2), 82-91, doi:10.1002/evan.10106.
- Boinski, S., & Newman, J. D. (1988). Preliminary observations on squirrel monkey (*Saimiri oerstedii*) vocalizations in Costa Rica. *American Journal of Primatology*, *14*(4), 329-343, doi:10.1002/ajp.1350140403.
- Bonnie, K. E., Horner, V., Whiten, A., & de Waal, F. B. M. (2007). Spread of arbitrary conventions among chimpanzees: a controlled experiment. *Proceedings of the Royal Society B: Biological Sciences*, *274*(1608), 367-372, doi:10.1098/rspb.2006.3733.
- Boughman, J., & Moss, C. (2003). Social Sounds: Vocal Learning and Development of Mammal and Bird Calls. In A. Simmons, R. Fay, & A. Popper (Eds.), *Acoustic Communication* (Vol. 16, pp. 138-224, Springer Handbook of Auditory Research). New York: Springer.
- Boughman, J. W. (1997). Greater spear-nosed bats give group-distinctive calls. *Behavioral Ecology and Sociobiology*, *40*(1), 61-70, doi:10.1007/s002650050316.
- Boughman, J. W. (1998). Vocal learning by greater spear-nosed bats. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *265*(1392), 227-233, doi:10.1098/rspb.1998.0286.
- Boughman, J. W., & Wilkinson, G. S. (1998). Greater spear-nosed bats discriminate group mates by vocalizations. *Animal Behaviour*, *55*(6), 1717-1732, doi:10.1006/anbe.1997.0721.
- Brand, L. R. (1976). The vocal repertoire of chipmunks (Genus *Eutamias*) in California. *Animal Behaviour*, *24*(2), 319-335, doi:10.1016/s0003-3472(76)80040-1.
- Brockelman, W., & Schilling, D. (1984). Inheritance of stereotyped gibbon calls. *Nature*, *312*, 634-636, doi:10.1038/312634a0.
- Brown, S. (1999). The "musilanguage" model of music evolution. In N. L. Wallin, B. Merker, & S. Brown (Eds.), *The Origins of Music* (pp. 271-300). Cambridge, MA: MIT Press.
- Brown, S., Merker, B., & Wallin, N. L. (1999). An introduction to evolutionary musicology. In N. L. Wallin, B. Merker, & S. Brown (Eds.), *The Origins of Music* (pp. 3-24). Cambridge, MA: MIT Press.
- Brown, S., Ngan, E., & Liotti, M. (2008). A Larynx Area in the Human Motor Cortex. *Cerebral Cortex*, *18*(4), 837-845, doi:10.1093/cercor/bhm131.
- Brumm, H., Voss, K., Köllmer, I., & Todt, D. (2004). Acoustic communication in noise: regulation of call characteristics in a New World monkey. *Journal of Experimental Biology*, *207*(3), 443-448, doi:10.1242/jeb.00768.
- Candiotti, A., Zuberbühler, K., & Lemasson, A. (2012a). Context-related call combinations in female Diana monkeys. *Animal Cognition*, *15*(3), 327-339, doi:10.1007/s10071-011-0456-8.

- Candiotti, A., Zuberbühler, K., & Lemasson, A. (2012b). Convergence and divergence in Diana monkey vocalizations. *Biology Letters*, doi:10.1098/rsbl.2011.1182.
- Cäsar, C., Byrne, R. W., Hoppitt, W., Young, R. J., & Zuberbühler, K. (2012). Evidence for semantic communication in titi monkey alarm calls. *Animal Behaviour*, *84*(2), 405-411, doi:10.1016/j.anbehav.2012.05.010.
- Catchpole, C. K., & Slater, P. J. B. (2003). *Bird Song: Biological Themes and Variations*. Cambridge, New York: Cambridge University Press.
- Cavalli-Sforza, L. L., Piazza, A., Menozzi, P., & Mountain, J. (1988). Reconstruction of human evolution: bringing together genetic, archaeological, and linguistic data. *Proceedings of the National Academy of Sciences*, *85*(16), 6002-6006.
- Cerling, T. E., Wynn, J. G., Andanje, S. A., Bird, M. I., Korir, D. K., Levin, N. E., et al. (2011). Woody cover and hominin environments in the past 6[thinsp]million years. [10.1038/nature10306]. *Nature*, *476*(7358), 51-56, doi:http://www.nature.com/nature/journal/v476/n7358/abs/nature10306.html#supplementary-information.
- Chambers, J. K., & Trudgill, P. (1998). *Dialectology*: Cambridge University Press.
- Chen, S.-F., Jones, G., & Rossiter, S. J. (2009). Determinants of echolocation call frequency variation in the Formosan lesser horseshoe bat (*Rhinolophus monoceros*). *Proceedings of the Royal Society B: Biological Sciences*, *276*(1674), 3901-3909, doi:10.1098/rspb.2009.1185.
- Cheney, D. L., & Seyfarth, R. M. (1980). Vocal recognition in free-ranging vervet monkeys. *Animal Behaviour*, *28*(2), 362-367, doi:10.1016/s0003-3472(80)80044-3.
- Cheney, D. L., & Seyfarth, R. M. (1992). *How monkeys see the world: Inside the mind of another species*. London: University of Chicago Press.
- Cheney, D. L., & Seyfarth, R. M. (2005). Constraints and preadaptations in the earliest stages of language evolution. [doi: 10.1515/tlir.2005.22.2-4.135]. *The Linguistic Review*, *22*(2-4), 135-159, doi:10.1515/tlir.2005.22.2-4.135.
- Chomsky, N. (1957). *Syntactic Structures*: Mouton.
- Christiansen, M., & Kirby, S. (2003a). *Language evolution*. Oxford, New York: Oxford University Press.
- Christiansen, M. H., & Kirby, S. (2003b). Language evolution: consensus and controversies. [10.1016/S1364-6613(03)00136-0]. *Trends in Cognitive Sciences*, *7*(7), 300-307, doi:10.1016/S1364-6613(03)00136-0.
- Clarke, E., Reichard, U. H., & Zuberbühler, K. (2006). The syntax and meaning of wild gibbon songs. *PLoS ONE*, *1*, e73, doi:10.1371/journal.pone.0000073.
- Clay, Z., Pika, S., Gruber, T., & Zuberbühler, K. (2011). Female bonobos use copulation calls as social signals. *Biology Letters*, *7*(4), 513-516, doi:10.1098/rsbl.2010.1227.
- Clay, Z., & Zuberbühler, K. (2011). Bonobos extract meaning from call sequences. *PLoS ONE*, *6*(4), e18786.
- Conner, D. A. (1982a). Dialects versus geographic variation in mammalian vocalizations. *Animal Behaviour*, *30*(1), 297-298, doi:10.1016/s0003-3472(82)80269-8.
- Conner, D. A. (1982b). Geographic variation in short calls of pikas (*Ochotona princeps*). *Journal of Mammalogy*, *63*(1), 48-52.
- Corballis, M. C. (2003). From mouth to hand: Gesture, speech, and the evolution of right-handedness. *Behavioral and Brain Sciences*, *26*(02), 199-208, doi:10.1017/S0140525X03000062.
- Cover, T. M., & Thomas, J. A. (2006). *Elements of information theory*. Hoboken, New Jersey: Wiley-Interscience.

- Crockford, C., & Boesch, C. (2005). Call combinations in wild chimpanzees. *Behaviour*, *142*(4), 397-421, doi:10.1163/1568539054012047.
- Crockford, C., Herbinger, I., Vigilant, L., & Boesch, C. (2004). Wild Chimpanzees Produce Group-Specific Calls: a Case for Vocal Learning? *Ethology*, *110*(3), 221-243.
- Crockford, C., Wittig, Roman M., Mundry, R., & Zuberbühler, K. (2011). Wild Chimpanzees Inform Ignorant Group Members of Danger. *Current biology : CB*.
- Crystal, D. (2002). *Language death*: Cambridge University Press.
- Darwin, C. (1859). *On the origin of the species by means of natural selection: or, The preservation of favoured races in the struggle for life*: John Murray.
- Davidson, S. M., & Wilkinson, G. S. (2002). Geographic and individual variation in vocalizations by male Saccopteryx B«bilineata (Chiroptera: Emballonuridae). *Journal of Mammalogy*, *83*(2), 526-535, doi:10.1644/1545-1542(2002)083<0526:gaiviv>2.0.co;2.
- de A. Moura, A. C., & Lee, P. C. (2004). Capuchin Stone Tool Use in Caatinga Dry Forest. *Science*, *306*(5703), 1909, doi:10.1126/science.1102558.
- de la Torre, S., & Snowdon, C. T. (2009). Dialects in pygmy marmosets? Population variation in call structure. *American Journal of Primatology*, *71*(4), 333-342, doi:10.1002/ajp.20657.
- Deacon, T. (1998). *The symbolic species: the co-evolution of language and the brain*: W.W. Norton.
- Delgado, R. A. (2007). Geographic variation in the long calls of male orangutans (*Pongo* spp.). *Ethology*, *113*(5), 487-498, doi:10.1111/j.1439-0310.2007.01345.x.
- Delgado, R. A., Lameira, A., Davila Ross, M., Husson, S., Morrogh-Bernard, H., & Wich, S. (2009). Geographical variation in orangutan long calls. In S. Wich, T. Mitra Setia, S. S. Utami Atmoko, & C. P. Van Schaik (Eds.), *Orangutans: Geographic Variation in Behavioral Biology and Conservation* (pp. 215-224). New York: Oxford University Press.
- Delgado, R. A., & van Schaik, C. P. (2000). The behavioral ecology and conservation of the orangutan (*Pongo pygmaeus*): A tale of two islands. *Evolutionary Anthropology*, *9*(5), 201-218, doi:10.1002/1520-6505(2000)9:5<201::aid-evan2>3.0.co;2-y.
- Dellatore, D., Waitt, C., & Foitova, I. (2009). Two cases of mother–infant cannibalism in orangutans. *Primates*, *50*(3), 277-281, doi:10.1007/s10329-009-0142-5.
- Dindo, M., Stoinski, T., & Whiten, A. (2011). Observational learning in orangutan cultural transmission chains. *Biology Letters*, *7*(2), 181-183, doi:10.1098/rsbl.2010.0637.
- Dobkin, D. S. (1979). Functional and Evolutionary Relationships of Vocal Copying Phenomena in Birds. *Zeitschrift für Tierpsychologie*, *50*(4), 348-363, doi:10.1111/j.1439-0310.1979.tb01037.x.
- Eiler, K. C., & Banack, S. A. (2004). Variability in the Alarm Call of Golden-mantled Ground Squirrels (*Spermophilus lateralis* and *S. saturatus*). *Journal of Mammalogy*, *85*(1), 43-50, doi:10.1644/1545-1542(2004)085<0043:vitaco>2.0.co;2.
- Elowson, A. M., & Snowdon, C. T. (1994). Pygmy marmosets, *Cebuella pygmaea*, modify vocal structure in response to changed social environment. *Animal Behaviour*, *47*(6), 1267-1277, doi:10.1006/anbe.1994.1175.
- Elowson, A. M., Snowdon, C. T., & Lazaro-Perea, C. (1998). 'Babbling' and social context in infant monkeys: parallels to human infants. *Trends in Cognitive Sciences*, *2*(1), 31-37, doi:10.1016/s1364-6613(97)01115-7.
- Enard, W., Przeworski, M., Fisher, S. E., Lai, C. S. L., Wiebe, V., Kitano, T., et al. (2002). Molecular evolution of FOXP2, a gene involved in speech and language. [10.1038/nature01025]. *Nature*, *418*(6900), 869-872, doi:http://www.nature.com/nature/journal/v418/n6900/suppinfo/nature01025\_S1.html.
- Esser, K.-H. (1994). Audio-vocal learning in a non-human mammal: the lesser spear-nosed bat *Phyllostomus discolor*. *Neuroreport*, *5*(14), 1718-1720.

- Esser, K.-H., & Schubert, J. (1998). Vocal Dialects in the Lesser Spear-Nosed Bat *Phyllostomus discolor*. *Naturwissenschaften*, *85*(7), 347-349, doi:10.1007/s001140050513.
- Ey, E., Rahn, C., Hammerschmidt, K., & Fischer, J. (2009). Wild Female Olive Baboons Adapt their Grunt Vocalizations to Environmental Conditions. *Ethology*, *115*(5), 493-503, doi:10.1111/j.1439-0310.2009.01638.x.
- Falconer, D. S., & Mackay, T. F. C. (1996). *Introduction to quantitative genetics*: Benjamin Cummings.
- Falk, D. (2004). Prelinguistic evolution in early hominins: Whence motherese? *Behavioral and Brain Sciences*, *27*(04), 491-503, doi:10.1017/S0140525X04000111.
- Favareau, D. (2010). Introduction: An Evolutionary History of Biosemiotics. In M. Barbieri (Ed.), *Essential Readings in Biosemiotics: Anthology and Commentary* (pp. 1-68). Dordrecht, the Netherlands: Springer.
- Fenton, M. B., & Bell, G. P. (1981). Recognition of Species of Insectivorous Bats by Their Echolocation Calls. *Journal of Mammalogy*, *62*(2), 233-243.
- Fichtel, C., & Kappeler, P. (2011). Variation in the meaning of alarm calls in Verreaux's and Coquerel's sifakas (*Propithecus verreauxi*, *P. coquereli*). *International Journal of Primatology*, *32*(2), 346-361, doi:10.1007/s10764-010-9472-9.
- Fichtel, C., & Van Schaik, C. P. (2006). Semantic differences in sifaka (*Propithecus verreauxi*) alarm calls: A reflection of genetic or cultural variants? *Ethology*, *112*(9), 839-849, doi:10.1111/j.1439-0310.2006.01239.x.
- Fischer, J., Hammerschmidt, K., & Todt, D. (1998). Local variation in Barbary macaque shrill barks. *Animal Behaviour*, *56*(3), 623-629, doi:10.1006/anbe.1998.0796.
- Fitch, W. (2010). *The Evolution of Language*. Cambridge: Cambridge University Press.
- Fitch, W. T. (1997). Vocal tract length and formant frequency dispersion correlate with body size in rhesus macaques. *Journal of the Acoustical Society of America*, *102*(2), 1213-1222, doi:10.1121/1.421048.
- Fitch, W. T. (2000). The evolution of speech: a comparative review. [doi: DOI: 10.1016/S1364-6613(00)01494-7]. *Trends in Cognitive Sciences*, *4*(7), 258-267.
- Fitch, W. T. (2005). Protomusic and protolanguage as alternatives to protosign. *Behavioral and Brain Sciences*, *28*(02), 132-133, doi:10.1017/S0140525X05290039.
- Fitch, W. T., & Hauser, M. D. (2003). Unpacking "honesty": vertebrate vocal production and the evolution of acoustic signals. In A. M. Simmons, R. R. Fay, & A. N. Popper (Eds.), *Acoustic communication* (pp. 65-137). New York: Springer.
- Fitch, W. T., & Reby, D. (2001). The descended larynx is not uniquely human. *Proceedings of the Royal Society of London Series B: Biological Sciences*, *268*(1477), 1669-1675, doi:10.1098/rspb.2001.1704.
- Foote, A. D., Griffin, R. M., Howitt, D., Larsson, L., Miller, P. J. O., & Rus Hoelzel, A. (2006). Killer whales are capable of vocal learning. *Biology Letters*, *2*(4), 509-512, doi:10.1098/rsbl.2006.0525.
- Ford, J. K. B. (1991). Vocal traditions among resident killer whales (*Orcinus orca*) in coastal waters of British Columbia. *Canadian Journal of Zoology*, *69*(6), 1454-1483, doi:10.1139/z91-206.
- Francescoli, G. (2002). Geographic variation in vocal signals of *Ctenomys pearsoni*. *Acta Theriologica*, *47*(1), 35-44, doi:10.1007/bf03193564.
- Francis, C. M., & Habersetzer, J. (1998). Inter-and intra-specific variation in echolocation call frequency and morphology of horseshoe bats (*Rhinolophus* and *Hipposideros*). In T. H. Kunz, & P. A. Racey (Eds.), *Bat Biology and Conservation* (pp. 169-179). Washington and London: Smithsonian Institution Press.
- Furness, W. H. (1916). Observations on the Mentality of Chimpanzees and Orang-Utans. *Proceedings of the American Philosophical Society*, *55*(3), 281-290.

- Gamba, M., & Giacoma, C. (2006). Vocal Tract Modeling in a Prosimian Primate: The Black and White Ruffed Lemur. *Acta Acustica united with Acustica*, 92(5), 749-755.
- Gamba, M. M., & Giacoma, C. C. (2010). Key issues in the study of primate acoustic signals, an update. *J Anthropol Sci*, 88, 215-220.
- Gannon, W. L., & Lawlor, T. E. (1989). Variation of the Chip Vocalization of Three Species of Townsend Chipmunks (Genus *eutamias*). *Journal of Mammalogy*, 70(4), 740-753.
- Gardner, R. A., & Gardner, B. T. (1969). Teaching Sign Language to a Chimpanzee. *Science*, 165(3894), 664-672, doi:10.1126/science.165.3894.664.
- Garner, R. L. (1892). *The speech of monkeys*: Heinemann.
- Geissmann, T. (1984). Inheritance of Song Parameters in the Gibbon Song, Analysed in 2 Hybrid Gibbons (*Hylobates pileatus* X *H. lar*). *Folia Primatologica*, 42(3-4), 216-235.
- Ghazanfar, A. A., & Rendall, D. (2008). Evolution of human vocal production. *Current Biology*, 18(11), R457-R460, doi:10.1016/j.cub.2008.03.030.
- Ghazanfar, Asif A., Takahashi, Daniel Y., Mathur, N., & Fitch, W. T. (2012). Cineradiography of Monkey Lip-Smacking Reveals Putative Precursors of Speech Dynamics. *Current biology : CB*.
- Gil-da-Costa, R., Palleroni, A., Hauser, M. D., Touchton, J., & Kelley, J. P. (2003). Rapid acquisition of an alarm response by a neotropical primate to a newly introduced avian predator. *Proceedings of the Royal Society B: Biological Sciences*, 270(1515), 605-610, doi:10.1098/rspb.2002.2281.
- Gillam, E. H., & McCracken, G. F. (2007). Variability in the echolocation of *Tadarida brasiliensis*: effects of geography and local acoustic environment. *Animal Behaviour*, 74(2), 277-286, doi:10.1016/j.anbehav.2006.12.006.
- Green, D. J., & Alemseged, Z. (2012). *Australopithecus afarensis* Scapular Ontogeny, Function, and the Role of Climbing in Human Evolution. *Science*, 338(6106), 514-517, doi:10.1126/science.1227123.
- Green, S. (1975). Dialects in Japanese Monkeys: Vocal Learning and Cultural Transmission of Locale-specific Vocal Behavior? *Zeitschrift für Tierpsychologie*, 38(3), 304-314, doi:10.1111/j.1439-0310.1975.tb02006.x.
- Guillén, Juste, B., & Ibáñez (2000). Variation in the frequency of the echolocation calls of *Hipposideros ruber* in the Gulf of Guinea: an exploration of the adaptive meaning of the constant frequency value in rhinolophoid CF bats. *Journal of Evolutionary Biology*, 13(1), 70-80, doi:10.1046/j.1420-9101.2000.00155.x.
- Hafen, T., Neveu, H., Rimpler, Y., Wilden, I., & Zimmermann, E. (1998). Acoustically Dimorphic Advertisement Calls Separate Morphologically and Genetically Homogenous Populations of the Grey Mouse Lemur (*Microcebus murinus*). *Folia Primatologica*, 69(Suppl. 1), 342-356.
- Hafner, M. S., & Hafner, D. J. (1979). Vocalizations of Grasshopper Mice (Genus *Onychomys*). *Journal of Mammalogy*, 60(1), 85-94.
- Hammerschmidt, K., & Jischer, J. (2008). Constraints in primate vocal production. In G. U., & O. K (Eds.), *The evolution of communicative creativity: From fixed signals to contextual flexibility* (pp. 93-119). Cambridge: The MIT Press.
- Hardus, M. E., Lameira, A. R., Singleton, I., Morrough-Bernard, H. C., Knott, C. D., Ancrenaz, M., et al. (2009a). A description of the orangutan's vocal and sound repertoire, with a focus on geographic variation. In S. Wich, T. Mitra Setia, S.S. Utami, & C. P. Schaik (Eds.), *Orangutans* (pp. 49-60). New York: Oxford University Press.

- Hardus, M. E., Lameira, A. R., van Schaik, C. P., & Wich, S. A. (2009b). Tool use in wild orang-utans modifies sound production: a functionally deceptive innovation? *Proceedings of the Royal Society B: Biological Sciences*, *276*(1673), 3689-3694, doi:10.1098/rspb.2009.1027.
- Hart, D., & Sussman, R. W. (2008). *Man the hunted: Primates, predators, and human evolution*. Boulder, CO: Westview Press.
- Hauser, M. D. (1992). Articulatory and social factors influence the acoustic structure of rhesus monkey vocalizations: A learned mode of production? *The Journal of the Acoustical Society of America*, *91*(4), 2175-2179.
- Hauser, M. D., Chomsky, N., & Fitch, W. T. (2002). The faculty of language: What is it, who has it, and how did it evolve? *Science*, *298*(5598), 1569-1579, doi:10.1126/science.298.5598.1569.
- Hayes, C. (1951). *The ape in our house*: Harper.
- Hayes, K. J., & Hayes, C. (1952). Imitation in a home-raised chimpanzee. *Journal of Comparative and Physiological Psychology*, *45*(5), 450-459, doi:10.1037/h0053609.
- Heller, K.-G., & Helversen, O. (1989). Resource partitioning of sonar frequency bands in rhinolophoid bats. *Oecologia*, *80*(2), 178-186, doi:10.1007/bf00380148.
- Hemelrijk, C. K., & Bolhuis, J. J. (2011). A minimalist approach to comparative psychology. *Trends in Cognitive Sciences*, *15*(5), 185-186.
- Hobolth, A., Duthel, J. Y., Hawks, J., Schierup, M. H., & Mailund, T. (2011). Incomplete lineage sorting patterns among human, chimpanzee, and orangutan suggest recent orangutan speciation and widespread selection. *Genome Research*, (in press), doi:10.1101/gr.114751.110.
- Hockett, C. F., & Altmann, S. (1968). A note on design features. In T. A. Sebeok (Ed.), *Animal communication: techniques of study and results of research* (pp. 61-72). Bloomington: Indiana University Press.
- Hodun, A., Snowdon, C. T., & Soini, P. (1981). Subspecific Variation in the Long Calls of the Tamarin, *Saguinus fuscicollis*. *Zeitschrift für Tierpsychologie*, *57*(2), 97-110, doi:10.1111/j.1439-0310.1981.tb01315.x.
- Hohmann, G., & Vogl, L. (1991). Loud calls of male Nilgiri langurs (*Presbytis johnii*): Age-, individual-, and population-specific differences. *International Journal of Primatology*, *12*(5), 503-524, doi:10.1007/bf02547636.
- Hopkins, W., Tagliatalata, J. P., & Leavens, D. (2011). Do chimpanzees have voluntary control of their facial expressions and vocalizations? In A. Vilain, C. Abry, J. Schwartz, & J. Vauclair (Eds.), *Primate Communication and Human Language: Vocalisation, Gestures, Imitation and Deixis in Humans and Non-Humans* (pp. 71-90). Amsterdam, Philadelphia: John Benjamins Publishing Company.
- Hopkins, W. D., & Cantero, M. (2003). From hand to mouth in the evolution of language: the influence of vocal behavior on lateralized hand use in manual gestures by chimpanzees (*Pan troglodytes*). *Developmental Science*, *6*(1), 55-61, doi:10.1111/1467-7687.00254.
- Hopkins, W. D., & Nir, T. M. (2010). Planum temporale surface area and grey matter asymmetries in chimpanzees (*Pan troglodytes*): The effect of handedness and comparison with findings in humans. *Behavioural Brain Research*, *208*(2), 436-443, doi:10.1016/j.bbr.2009.12.012.
- Hopkins, W. D., Tagliatalata, J. P., & Leavens, D. A. (2007). Chimpanzees differentially produce novel vocalizations to capture the attention of a human. *Animal Behaviour*, *73*(2), 281-286, doi:10.1016/j.anbehav.2006.08.004.
- Husson, S., Wich, S., Marshall, A. J., Dennis, R., Ancrenaz, M., Brasseur, R., et al. (2009). Orangutan distribution, density, abundance and impacts of disturbance. In S. Wich, S. S. Utami Atmoko, T. M. Setia, & C. P. Van Schaik (Eds.), *Orangutans: Geographic Variation in Behavioral Ecology and Conservation* (pp. 77-96): Oxford University Press.

- Inoue-Nakamura, N., & Matsuzawa, T. (1997). Development of stone tool use by wild chimpanzees (*Pan troglodytes*). *Journal of Comparative Psychology*, *111*(2), 159-173, doi:10.1037/0735-7036.111.2.159.
- Jaeggi, A. V., Dunkel, L. P., Van Noordwijk, M. A., Wich, S. A., Sura, A. A. L., & Van Schaik, C. P. (2010). Social learning of diet and foraging skills by wild immature Bornean orangutans: implications for culture. *American Journal of Primatology*, *72*(1), 62-71, doi:10.1002/ajp.20752.
- Jaeggi, A. V., van Noordwijk, M. A., & van Schaik, C. P. (2008). Begging for information: mother-offspring food sharing among wild Bornean orangutans. *American Journal of Primatology*, *70*(6), 533-541, doi:10.1002/ajp.20525.
- Janik, V., & Slater, P. B. (1997). Vocal Learning in Mammals. In J. S. R. C. T. S. Peter J.B. Slater, & M. Manfred (Eds.), *Advances in the Study of Behavior* (Vol. Volume 26, pp. 59-99). San Diego, London: Academic Press.
- Janik, V. M. (2000). Whistle matching in wild bottlenose dolphins (*Tursiops truncatus*). *Science*, *289*(5483), 1355-1357.
- Janik, V. M., & Slater, P. J. B. (2000). The different roles of social learning in vocal communication. *Animal Behaviour*, *60*(1), 1-11, doi:10.1006/anbe.2000.1410.
- Johnson, A. E., Knott, C. D., Pamungkas, B., Pasaribu, M., & Marshall, A. J. (2005). A survey of the orangutan (*Pongo pygmaeus wurmbii*) population in and around Gunung Palung National Park, West Kalimantan, Indonesia based on nest counts. *Biological Conservation*, *121*(4), 495-507, doi:10.1016/j.biocon.2004.06.002.
- Jones, G., & Ransome, R. D. (1993). Echolocation Calls of Bats are Influenced by Maternal Effects and Change over a Lifetime. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *252*(1334), 125-128, doi:10.1098/rspb.1993.0055.
- Jürgens, U. (1979). Neural control of vocalization in nonhuman primates. In H. D. Steklis, & M. J. Raleigh (Eds.), *Neurobiology of social communication in primates* (pp. 87-92). New York: Addison-Wesley.
- Jürgens, U. (2002). Neural pathways underlying vocal control. [doi: DOI: 10.1016/S0149-7634(01)00068-9]. *Neuroscience & Biobehavioral Reviews*, *26*(2), 235-258.
- Kavanagh, M. (1980). Invasion of the forest by an african savannah monkey: Behavioural adaptations. *Behaviour*, *73*(3-4), 238-260, doi:10.1163/156853980x00258.
- Kawai, M. (1965). Newly-acquired pre-cultural behavior of the natural troop of Japanese monkeys on Koshima islet. *Primates*, *6*(1), 1-30, doi:10.1007/bf01794457.
- Kellogg, W. N., & Kellogg, L. A. (1967). *The ape and the child: a study of environmental influence upon early behavior*: Hafner Pub. Co.
- Kemp, C., & Regier, T. (2012). Kinship categories across languages reflect general communicative principles. *Science*, *336*(6084), 1049-1054, doi:10.1126/science.1218811.
- Kershenbaum, A., Ilany, A., Blaustein, L., & Geffen, E. (2012). Syntactic structure and geographical dialects in the songs of male rock hyraxes. *Proceedings of the Royal Society B: Biological Sciences*, doi:10.1098/rspb.2012.0322.
- Kidjo, N., Cargnelutti, B., Charlton, B. D., Wilson, C., & Reby, D. (2008). VOCAL BEHAVIOUR IN THE ENDANGERED CORSICAN DEER: DESCRIPTION AND PHYLOGENETIC IMPLICATIONS. *Bioacoustics*, *18*(2), 159-181, doi:10.1080/09524622.2008.9753598.
- Kingston, T., & Rossiter, S. (2004). Harmonic hopping in Wallacea's bats. *Nature*, *429*, 654-657, doi:citeulike-article-id:1149835.
- Knight, A., Underhill, P. A., Mortensen, H. M., Zhivotovsky, L. A., Lin, A. A., Henn, B. M., et al. (2003). African Y Chromosome and mtDNA Divergence Provides Insight into the History of Click Languages. *Current Biology*, *13*(6), 464-473.

- Knörnschild, M., Behr, O., & von Helversen, O. (2006). Babbling behavior in the sac-winged bat (<i>Saccopteryx bilineata</i>). *Naturwissenschaften*, 93(9), 451-454, doi:10.1007/s00114-006-0127-9.
- Koda, H., Nishimura, T., Tokuda, I. T., Oyakawa, C., Nihonmatsu, T., & Masataka, N. (2012). Soprano singing in gibbons. *American Journal of Physical Anthropology*, n/a-n/a, doi:10.1002/ajpa.22124.
- Krützen, M., Mann, J., Heithaus, M. R., Connor, R. C., Bejder, L., & Sherwin, W. B. (2005). Cultural transmission of tool use in bottlenose dolphins. *Proceedings of the National Academy of Sciences of the United States of America*, 102(25), 8939-8943, doi:10.1073/pnas.0500232102.
- Krützen, M., Willems, Erik P., & van Schaik, Carel P. (2011). Culture and Geographic Variation in Orangutan Behavior. *Current biology : CB*, 21(21), 1808-1812.
- Kuypers, M. G. J. M. (1958). Some projections from the peri-central cortex to the pons and lower brain stem in monkeys and chimpanzee. *Journal of Comparative Neurology*(110), 211-255.
- Laland, K. N., & Galef, B. G. (2009). *The Question of Animal Culture*: Harvard University Press.
- Laland, K. N., & Hoppitt, W. (2003). Do animals have culture? *Evolutionary Anthropology: Issues, News, and Reviews*, 12(3), 150-159, doi:10.1002/evan.10111.
- Laland, K. N., & Janik, V. M. (2006). The animal cultures debate. [doi: DOI: 10.1016/j.tree.2006.06.005]. *Trends in Ecology & Evolution*, 21(10), 542-547.
- Lameira, A., Delgado, R., & Wich, S. (2010). Review of geographic variation in terrestrial mammalian acoustic signals: Human speech variation in a comparative perspective. [10.1556/JEP.8.2010.4.2]. *Journal of Evolutionary Psychology*, 8(4), 309-332.
- Lameira, A., Hardus, M., & Wich, S. (in press-a). Orangutan Instrumental Gesture-Calls: Reconciling Acoustic and Gestural Speech Evolution Models. *Evolutionary Biology*, 1-4, doi:10.1007/s11692-011-9151-6.
- Lameira, A., & Wich, S. (2008). Orangutan Long Call Degradation and Individuality Over Distance: A Playback Approach. *International Journal of Primatology*, 29(3), 615-625, doi:10.1007/s10764-008-9253-x.
- Lameira, A. R., Hardus, M. E., Kowalsky, B., de Vries, H., Spruijt, B. M., Sterck, E. H. M., et al. (in press-b). Orangutan whistling and implications for the emergence of an open-ended call repertoire: a replication and extension. *Journal of the Acoustical Society of America*.
- Langergraber, K. E., Boesch, C., Inoue, E., Inoue-Murayama, M., Mitani, J. C., Nishida, T., et al. (2011). Genetic and 'cultural' similarity in wild chimpanzees. *Proceedings of the Royal Society B: Biological Sciences*, 278(1704), 408-416, doi:10.1098/rspb.2010.1112.
- Langergraber, K. E., & Vigilant, L. (2011). Genetic differences cannot be excluded from generating behavioural differences among chimpanzee groups. *Proceedings of the Royal Society B: Biological Sciences*, 278(1715), 2094-2095, doi:10.1098/rspb.2011.0391.
- Laporte, M. N. C., & Zuberbühler, K. (2010). Vocal greeting behaviour in wild chimpanzee females. *Animal Behaviour*, 80(3), 467-473, doi:10.1016/j.anbehav.2010.06.005.
- Law, B. S., Reinhold, L., & Pennay, M. (2002). Geographic variation in the echolocation calls of *Vespadelus* spp. (Vespertilionidae) from New South Wales and Queensland, Australia. *Acta Chiropterologica*, 4(2), 201-215.
- Lemasson, A., Gandon, E., & Hausberger, M. (2010a). Attention to elders' voice in non-human primates. *Biology Letters*, 6(3), 325-328, doi:10.1098/rsbl.2009.0875.
- Lemasson, A., Hausberger, M., & Zuberbühler, K. (2005). Socially meaningful vocal plasticity in adult Campbell's monkeys (*Cercopithecus campbelli*). *Journal of Comparative Psychology*, 119(2), 220-229.

- Lemasson, A., Ouattara, K., Bouchet, H., & Zuberbühler, K. (2010b). Speed of call delivery is related to context and caller identity in Campbell's monkey males. *Naturwissenschaften*, 97(11), 1023-1027, doi:10.1007/s00114-010-0715-6.
- Lemasson, A., Ouattara, K., Petic, J. E., & Zuberbühler, K. (2011). Social learning of vocal structure in a nonhuman primate? *BMC Evolutionary Biology*(11), 362, doi:10.1186/1471-2148-11-362.
- Lemon, R. E. (1973). Nervous control of the syrinx in White-throated sparrows (*Zonotrichia albicollis*). *Journal of Zoology*, 171(1), 131-140, doi:10.1111/j.1469-7998.1973.tb07520.x.
- Lieberman, P. (2002a). *Human Language and Our Reptilian Brain: The Subcortical Bases of Speech, Syntax, and Thought*: Harvard University Press.
- Lieberman, P. (2002b). On the nature and evolution of the neural bases of human language. *American Journal of Physical Anthropology*, 119(S35), 36-62, doi:10.1002/ajpa.10171.
- Lieberman, P., Crelin, E. S., & Klatt, D. H. (1972). Phonetic Ability and Related Anatomy of the Newborn and Adult Human, Neanderthal Man, and the Chimpanzee. *American Anthropologist*, 74(3), 287-307, doi:10.1525/aa.1972.74.3.02a00020.
- Locke, D. P., Hillier, L. W., Warren, W. C., Worley, K. C., Nazareth, L. V., Muzny, D. M., et al. (2010). Comparative and demographic analysis of orang-utan genomes. *Nature*, 469(7331), 529-533.
- Ludlow, C. L. (2005). Central nervous system control of the laryngeal muscles in humans. [doi: DOI: 10.1016/j.resp.2005.04.015]. *Respiratory Physiology & Neurobiology*, 147(2-3), 205-222.
- Lycett, S. J., Collard, M., & McGrew, W. C. (2011). Correlations between genetic and behavioural dissimilarities in wild chimpanzees (*Pan troglodytes*) do not undermine the case for culture. *Proceedings of the Royal Society B: Biological Sciences*, 278(1715), 2091-2093, doi:10.1098/rspb.2010.2599.
- Lyons, L. (1995). *All you wanted to know about mathematics but were afraid to ask: Mathematics for science students* (Vol. 1-2). Cambridge: Cambridge University Press.
- MacKinnon, J. (1974). The behaviour and ecology of wild orang-utans (*Pongo pygmaeus*). [doi: DOI: 10.1016/S0003-3472(74)80054-0]. *Animal Behaviour*, 22(1), 3-74.
- Maclarnon, A., & Hewitt, G. (2004). Increased breathing control: Another factor in the evolution of human language. *Evolutionary Anthropology: Issues, News, and Reviews*, 13(5), 181-197, doi:10.1002/evan.20032.
- MacNeilage, P. F. (1998). The frame/content theory of evolution of speech production. *Behavioral and Brain Sciences*, 21(04), 499-511, doi:10.1017/S0140525X98001265.
- Maeda, T., & Masataka, N. (1987). Locale-specific Vocal Behaviour of the Tamarin (*Saguinus l. labiatus*). *Ethology*, 75(1), 25-30, doi:10.1111/j.1439-0310.1987.tb00639.x.
- Marler, P., & Tenaza, R. (1977). Signalling behavior of apes with special reference to vocalization. In T. A. Sebeok (Ed.), *How animals communicate* (pp. 965-1033). Bloomington: Indiana University Press.
- Marshall, A. J., Wrangham, R. W., & Arcadi, A. C. (1999). Does learning affect the structure of vocalizations in chimpanzees? [doi: DOI: 10.1006/anbe.1999.1219]. *Animal Behaviour*, 58(4), 825-830.
- Marten, K., & Marler, P. (1977). Sound transmission and its significance for animal vocalization. *Behavioral Ecology and Sociobiology*, 2(3), 271-290, doi:10.1007/bf00299740.
- Martin, P. R., Martin, P., & Bateson, P. P. G. (2007). *Measuring behaviour: an introductory guide*. Cambridge, New York: Cambridge University Press.
- Masataka, N. (1988). The response of red-chested moustached tamarins to long calls from their natal and alien populations. *Animal Behaviour*, 36(1), 55-61, doi:10.1016/s0003-3472(88)80249-5.

- Masters, W. M., Raver, K. A. S., & Kazial, K. A. (1995). Sonar signals of big brown bats, *Eptesicus fuscus*, contain information about individual identity, age and family affiliation. *Animal Behaviour*, *50*(5), 1243-1260, doi:10.1016/0003-3472(95)80041-7.
- Matrosova, V., Volodin, I., Volodina, E., & Babitsky, A. (2007). Pups crying bass: vocal adaptation for avoidance of age-dependent predation risk in ground squirrels? *Behavioral Ecology and Sociobiology*, *62*(2), 181-191, doi:10.1007/s00265-007-0452-9.
- McCowan, B., & Hooper, S. L. (2002). Individual acoustic variation in Belding's ground squirrel alarm chirps in the High Sierra Nevada. *The Journal of the Acoustical Society of America*, *111*(3), 1157-1160.
- Meyer, J. (2004). Bioacoustics of human whistled languages: an alternative approach to the cognitive processes of language. *Anais da Academia Brasileira de Ciências*, *76*, 406-412.
- Miles, H. L. W. (1993). Language and the Orang-utan: The Old 'Person' of the Forest. In P. Cavalieri, & P. Singer (Eds.), *The Great ape project: equality beyond humanity* (pp. 42-57). New York: St. Martin's Griffin.
- Miles, H. L. W., Mitchell, R. W., & Harper, S. E. (1996). Simon says: The development of imitation in an encultured orangutan. In A. Russon, K. Bard, & S. T. Parker (Eds.), *Reaching into thought: The minds of the great apes* (pp. 278-299). Cambridge: Cambridge University Press.
- Mitani, J., Gros-Louis, J., & Macedonia, J. (1996). Selection for acoustic individuality within the vocal repertoire of wild chimpanzees. *International Journal of Primatology*, *17*(4), 569-583, doi:10.1007/bf02735192.
- Mitani, J. C. (1987). Species discrimination of male song in gibbons. *American Journal of Primatology*, *13*(4), 413-423, doi:10.1002/ajp.1350130406.
- Mitani, J. C., & Brandt, K. L. (1994). Social Factors Influence the Acoustic Variability in the Long-distance Calls of Male Chimpanzees. *Ethology*, *96*(3), 233-252, doi:10.1111/j.1439-0310.1994.tb01012.x.
- Mitani, J. C., Hasegawa, T., Gros-Louis, J., Marler, P., & Byrne, R. (1992). Dialects in wild chimpanzees? *American Journal of Primatology*, *27*(4), 233-243, doi:10.1002/ajp.1350270402.
- Mitani, J. C., Hunley, K. L., & Murdoch, M. E. (1999). Geographic variation in the calls of wild chimpanzees: A reassessment. *American Journal of Primatology*, *47*(2), 133-151, doi:10.1002/(sici)1098-2345(1999)47:2<133::aid-ajp4>3.0.co;2-i.
- Mithen, S. J. (2005). *The singing Neanderthals: the origins of music, language, mind and body*, (Vol. 16, Cambridge Archaeological Journal, Vol. 01). London: Cambridge Journals Online.
- Mitra Setia, T., & van Schaik, C. P. (2007). The response of adult orang-utans to flanged male long calls: inferences about their function. *Folia Primatologica*, *78*(4), 215-226, doi:10.1159/000102317.
- Morrice, M. G., Burton, H. R., & Green, K. (1994a). Microgeographic variation and songs in the underwater vocalisation repertoire of the Weddell seal (&#x26;Leptonychotes weddellii&#x26;) from the Vestfold Hills, Antarctica. *Polar Biology*, *14*(7), 441-446, doi:10.1007/bf00239046.
- Morrice, M. G., Burton, H. R., & Green, K. (1994b). Microgeographic variation and songs in the underwater vocalisation repertoire of the Weddell seal (*Leptonychotes weddellii*) from the Vestfold Hills, Antarctica. *Polar Biology*, *14*(7), 441-446, doi:10.1007/bf00239046.
- Morrill, R. J., Paukner, A., Ferrari, P. F., & Ghazanfar, A. A. (2012). Monkey lipsmacking develops like the human speech rhythm. *Developmental Science*, no-no, doi:10.1111/j.1467-7687.2012.01149.x.

- Morton, E. S. (1975). Ecological sources of selection on avian sounds. *American Naturalist*, *109*, 17-34.
- Muir, C. C., Galdikas, B. M. F., & Beckenbach, A. T. (2000). mtDNA sequence diversity of orangutans from the islands of Borneo and Sumatra. *J. Mol. Evol.*, *51*, 471-480.
- Murray, K. L., Britzke, E. R., & Robbins, L. W. (2001). VARIATION IN SEARCH-PHASE CALLS OF BATS. *Journal of Mammalogy*, *82*(3), 728-737, doi:10.1644/1545-1542(2001)082<0728:vispco>2.0.co;2.
- Nater, A., Nietlisbach, P., Arora, N., van Schaik, C. P., van Noordwijk, M. A., Willems, E. P., et al. (2011). Sex-Biased Dispersal and Volcanic Activities Shaped Phylogeographic Patterns of Extant Orangutans (genus: Pongo). *Molecular Biology and Evolution*, *28*(8), 2275-2288, doi:10.1093/molbev/msr042.
- Nevo, E., Heth, G., Beiles, A., & Frankenberg, E. (1987). Geographic dialects in blind mole rats: role of vocal communication in active speciation. *Proceedings of the National Academy of Sciences*, *84*(10), 3312-3315.
- Nikol'skii, A. A., Kotlyakov, V. M., & Blumstein, D. T. (1999). Glaciation as a factor of geographic variation in the long-tailed marmot. *Doklady Biological Sciences*, *368*, 509-513.
- Nishimura, T., Mikami, A., Suzuki, J., & Matsuzawa, T. (2003). Descent of the larynx in chimpanzee infants. *Proceedings of the National Academy of Sciences*, *100*(12), 6930-6933, doi:10.1073/pnas.1231107100.
- Noad, M. J., Cato, D. H., Bryden, M. M., Jenner, M. N., & Jenner, K. C. S. (2000). Cultural revolution in whale songs. [10.1038/35046199]. *Nature*, *408*(6812), 537-537.
- Notman, H., & Rendall, D. (2005). Contextual variation in chimpanzee pant hoots and its implications for referential communication. *Animal Behaviour*, *70*(1), 177-190, doi:10.1016/j.anbehav.2004.08.024
- Nottebohm, F. (1969). The song of the chingolo, *Zonotrichia capensis*, in Argentina: description and evaluation of a system of dialects. *The Condor*, *71*(3), 299-315.
- O'Farrell, M. J., Corben, C., & Gannon, W. L. (2000). Geographic variation in the echolocation calls of the hoary bat (*Lasiurus cinereus*). *Acta Chiropterologica*, *2*(2), 185-196.
- Osvath, M., & Osvath, H. (2008). Chimpanzee (*Pan troglodytes*) and orangutan (*Pongo abelii*) forethought: self-control and pre-experience in the face of future tool use. *Animal Cognition*, *11*(4), 661-674, doi:10.1007/s10071-008-0157-0.
- Ouattara, K., Lemasson, A., & Zuberbühler, K. (2009a). Campbell's monkeys concatenate vocalizations into context-specific call sequences. *Proceedings of the National Academy of Sciences*, *106*(51), 22026-22031, doi:10.1073/pnas.0908118106.
- Ouattara, K., Lemasson, A., & Zuberbühler, K. (2009b). Campbell's monkeys use affixation to alter call meaning. *PLoS ONE*, *4*(11), e7808.
- Owren, M. J., Amoss, R. T., & Rendall, D. (2010). Two organizing principles of vocal production: Implications for nonhuman and human primates. *American Journal of Primatology*, *73*(6), 530-544, doi:10.1002/ajp.20913.
- Owren, M. J., Dieter, J. A., Seyfarth, R. M., & Cheney, D. L. (1993). Vocalizations of rhesus (*Macaca mulatta*) and Japanese (*M. fuscata*) macaques cross-fostered between species show evidence of only limited modification. *Developmental Psychobiology*, *26*(7), 389-406, doi:10.1002/dev.420260703.
- Owren, M. J., & Rendall, D. (2001). Sound on the rebound: Bringing form and function back to the forefront in understanding nonhuman primate vocal signaling. *Evolutionary Anthropology*, *10*(2), 58-71, doi:10.1002/evan.1014.

- Owren, M. J., Seyfarth, R. M., & Cheney, D. L. (1997). The acoustic features of vowel-like grunt calls in chacma baboons (*Papio cyncephalus ursinus*): Implications for production processes and functions. *Journal of the Acoustical Society of America*, *101*(5), 2951-2963.
- Palacios, V., Font, E., & Márquez, R. (2007). Iberian Wolf Howls: Acoustic Structure, Individual Variation, and a Comparison with North American Populations. *Journal of Mammalogy*, *88*(3), 606-613, doi:10.1644/06-mamm-a-151r1.1.
- Pantev, C., Oostenveld, R., Engelien, A., Ross, B., Roberts, L. E., & Hoke, M. (1998). Increased auditory cortical representation in musicians. *Nature*, *392*(6678), 811-814, doi:10.1038/33918.
- Papworth, S., Böse, A.-S., Barker, J., Schel, A. M., & Zuberbühler, K. (2008). Male blue monkeys alarm call in response to danger experienced by others. *Biology Letters*, *4*(5), 472-475, doi:10.1098/rsbl.2008.0299.
- Parsons, S. (1997). Search-phase echolocation calls of the New Zealand lesser short-tailed bat (*Mystacina tuberculata*) and long-tailed bat (*Chalinolobus tuberculatus*). *Canadian Journal of Zoology*, *75*(9), 1487-1494, doi:10.1139/z97-772.
- Patterson, F., & Linden, E. (1981). *The education of Koko*. Ann Arbor: Holt, Rinehart, and Winston.
- Paul, M. (2007). *Nesting sounds in wild Bornean orang-utans (Pongo pygmaeus): Cultural and functional aspects*. University of Zurich, Zurich.
- Pearl, D. L., & Fenton, M. B. (1996). Can echolocation calls provide information about group identity in the little brown bat (*Myotis lucifugus*)? *Canadian Journal of Zoology*, *74*(12), 2184-2192, doi:10.1139/z96-247.
- Perani, D., Saccuman, M. C., Scifo, P., Awander, A., Spada, D., Baldoli, C., et al. (2011). Neural language networks at birth. *Proceedings of the National Academy of Sciences*, *108*(38), 16056-16061, doi:10.1073/pnas.1102991108.
- Perla, B. S., & Slobodchikoff, C. N. (2002). Habitat structure and alarm call dialects in Gunnison's prairie dog (*Cynomys gunnisoni*). *Behavioral Ecology*, *13*(6), 844-850, doi:10.1093/beheco/13.6.844.
- Perry, S., Baker, M., Fedigan, L., Gros-Louis, J., Jack, K., MacKinnon, K. C., et al. (2003). Social conventions in wild white-faced capuchin monkeys: evidence for traditions in a Neotropical primate. *Current Anthropology*, *44*(2), 241-268.
- Piper, P. J., Cranbrook, E. O., & Rabett, R. J. (2007). Confirmation of the presence of the tiger *Panthera tigris* (L.) in late Pleistocene and Holocene Borneo. *Malayan Nature Journal*, *59*, 257-265.
- Podos, J., & Warren, P. S. (2007). The Evolution of Geographic Variation in Birdsong. In T. J. R. M. N. K. E. W.-E. C. B. H. Jane Brockmann, & M. John (Eds.), *Advances in the Study of Behavior* (Vol. Volume 37, pp. 403-458): Academic Press.
- Poole, J. H., Tyack, P. L., Stoeger-Horwath, A. S., & Watwood, S. (2005). Elephants are capable of vocal learning. [10.1038/434455a]. *Nature*, *434*(7032), 455-456, doi:http://www.nature.com/nature/journal/v434/n7032/supinfo/434455a\_S1.html.
- Premack, D. (1971). Language in Chimpanzee? *Science*, *172*(3985), 808-822, doi:10.1126/science.172.3985.808.
- Prufer, K., Munch, K., Hellmann, I., Akagi, K., Miller, J. R., Walenz, B., et al. (2012). The bonobo genome compared with the chimpanzee and human genomes. [10.1038/nature11128]. *Nature*, *advance online publication*, doi:http://www.nature.com/nature/journal/vaop/ncurrent/abs/nature11128.html#supplementary-information.

- Puglisi, A., Baronchelli, A., & Loreto, V. (2008). Cultural route to the emergence of linguistic categories. *Proceedings of the National Academy of Sciences*, 105(23), 7936-7940, doi:10.1073/pnas.0802485105.
- Radick, G. (2007). *The Simian Tongue: The Long Debate About Animal Language*: University of Chicago Press.
- Rainey, H. J., Zuberbühler, K., & Slater, P. J. B. (2004). Hornbills can distinguish between primate alarm calls. *Proceedings of the Royal Society B: Biological Sciences*, 271(1540), 755-759, doi:10.1098/rspb.2003.2619.
- Ralls, K., Fiorelli, P., & Gish, S. (1985). Vocalizations and vocal mimicry in captive harbor seals, *Phoca vitulina*. *Canadian Journal of Zoology*, 63(5), 1050-1056, doi:10.1139/z85-157.
- Ramsey, G., Bastian, M. L., & van Schaik, C. (2007). Animal innovation defined and operationalized. *Behavioral and Brain Sciences*, 30(04), 393-407, doi:doi:10.1017/S0140525X07002373.
- Randall, J. A., McCowan, B., Collins, K. C., Hooper, S. L., & Rogovin, K. (2005). Alarm signals of the great gerbil: Acoustic variation by predator context, sex, age, individual, and family group. *The Journal of the Acoustical Society of America*, 118(4), 2706-2714.
- Reader, S. M., & Laland, K. N. (2002). Social intelligence, innovation, and enhanced brain size in primates. *Proceedings of the National Academy of Sciences*, 99(7), 4436-4441, doi:10.1073/pnas.062041299.
- Remedios, R., Logothetis, N. K., & Kayser, C. (2009). Monkey drumming reveals common networks for perceiving vocal and nonvocal communication sounds. *Proceedings of the National Academy of Sciences*, 106(42), 18010-18015, doi:10.1073/pnas.0909756106.
- Rendall, D., Owren, M. J., & Rodman, P. S. (1998). The role of vocal tract filtering in identitycueing in rhesus monkey (*Macaca mulatta*) vocalizations. *Journal of the Acoustical Society of America*, 103(1), 602-614.
- Rendell, L., & Whitehead, H. (2001). Culture in whales and dolphins. *Behavioral and Brain Sciences*, 24(02), 309-324, doi:doi:10.1017/S0140525X0100396X.
- Rendell, L., & Whitehead, H. (2005). Spatial and temporal variation in sperm whale coda vocalizations: stable usage and local dialects. *Animal Behaviour*, 70(1), 191-198, doi:10.1016/j.anbehav.2005.03.001.
- Rendell, L. E., & Whitehead, H. (2003). Vocal clans in sperm whales (*Physeter macrocephalus*). *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 270(1512), 225-231, doi:10.1098/rspb.2002.2239.
- Richerson, P., & Boyd, R. (2005). *Not by genes alone: how culture transformed human evolution*: University of Chicago Press.
- Ridgway, S. H. (2011). Neural time and movement time in choice of whistle or pulse burst responses to different auditory stimuli by dolphins. *Journal of the Acoustical Society of America*, 129(2), 1073-1080, doi:10.1121/1.3523431.
- Riede, T., Bronson, E., Hatzikirou, H., & Zuberbühler, K. (2005). Vocal production mechanisms in a non-human primate: morphological data and a model. [doi: DOI: 10.1016/j.jhevol.2004.10.002]. *Journal of Human Evolution*, 48(1), 85-96.
- Rijksen, H. D. (1978). *A field study on sumatran orangutans (*Pongo pygmaeus abelii*, Lesson 1827): Ecology, behavior, and conservation*. Wageningen, the Netherlands: H. Veenman and Zonen.
- Rizzolatti, G., & Arbib, M. A. (1998). Language within our grasp. *Trends in Neurosciences*, 21(5), 188-194, doi:10.1016/S0166-2236(98)01260-0.
- Robinson, J. (1979). Vocal regulation of use of space by groups of titi monkeys *Callicebus moloch*. *Behavioral Ecology and Sociobiology*, 5(1), 1-15, doi:10.1007/bf00302691.

- Ross, M. D., & Geissmann, T. (2007). Call diversity of wild male orangutans: a phylogenetic approach. *American Journal of Primatology*, 69(3), 305-324, doi:10.1002/ajp.20356.
- Rumbaugh, D. M. (1977). *Language learning by a chimpanzee: the Lana project*. San Diego, London: Academic Press.
- Russo, D., Mucedda, M., Bello, M., Biscardi, S., Pidinchedda, E., & Jones, G. (2007). Divergent echolocation call frequencies in insular rhinolophids (Chiroptera): a case of character displacement? *Journal of Biogeography*, 34(12), 2129-2138, doi:10.1111/j.1365-2699.2007.01762.x.
- Rydell, J. (1993). Variation in the Sonar of an Aerial-hawking Bat (*Eptesicus nilssonii*). *Ethology*, 93(4), 275-284, doi:10.1111/j.1439-0310.1993.tb01209.x.
- Santorelli, C. J., Schaffner, C. M., Campbell, C. J., Notman, H., Pavelka, M. S., Weghorst, J. A., et al. (2011). Traditions in Spider Monkeys Are Biased towards the Social Domain. *PLoS ONE*, 6(2), e16863.
- Sapolsky, R. (2006). Social cultures among nonhuman primates. *Current Anthropology*, 47(4), 641-656, doi:citeulike-article-id:397003.
- Saranathan, V., Hamilton, D., Powell, G. V. N., Kroodsma, D. E., & Prum, R. O. (2007). Genetic evidence supports song learning in the three-wattled bellbird *Procnias tricarunculata* (Cotingidae). *Molecular Ecology*, 16, 3689-3702.
- Saussure, F., Bally, C., Sechehaye, A., & Riedlinger, A. (1983). *Course in general linguistics*: Open Court.
- Savage-Rumbaugh, E., Rumbaugh, D., & Boysen, S. (1978). Symbolic communication between two chimpanzees (*Pan troglodytes*). *Science*, 201(4356), 641-644, doi:10.1126/science.675251.
- Savage-Rumbaugh, E. S., & Lewin, R. (1994). *Kanzi: the ape at the brink of the human mind*. New York: John Wiley & Sons, Inc.
- Savage-Rumbaugh, E. S., Sevcik, R. A., Rumbaugh, D. M., & Rubert, E. (1985). The Capacity of Animals to Acquire Language: Do Species Differences have Anything to Say to Us? *Philosophical Transactions of the Royal Society B: Biological Sciences*, 308(1135), 177-185, doi:10.1098/rstb.1985.0018.
- Savage-Rumbaugh, S., McDonald, K., Sevcik, R. A., Hopkins, W. D., & Rubert, E. (1986). Spontaneous Symbol Acquisition and Communicative Use By Pygmy Chimpanzees (*Pan paniscus*). [doi: DOI: 10.1037/0096-3445.115.3.211]. *Journal of Experimental Psychology*, 115(3), 211-235.
- Scally, A., Dutheil, J. Y., Hillier, L. W., Jordan, G. E., Goodhead, I., Herrero, J., et al. (2012). Insights into hominid evolution from the gorilla genome sequence. [10.1038/nature10842]. *Nature*, 483(7388), 169-175, doi:http://www.nature.com/nature/journal/v483/n7388/abs/nature10842.html#supplementary-information.
- Schel, A. M., Candiotti, A., & Zuberbühler, K. (2010). Predator-detering alarm call sequences in Guereza colobus monkeys are meaningful to conspecifics. *Animal Behaviour*, 80(5), 799-808, doi:10.1016/j.anbehav.2010.07.012.
- Schel, A. M., Zuberb, & hler, K. (2009). Responses to leopards are independent of experience in Guereza colobus monkeys. *Behaviour*, 146(12), 1709-1737, doi:10.1163/000579509x12483520922007.
- Scherrer, J. A., & Wilkinson, G. S. (1993). Evening bat isolation calls provide evidence for heritable signatures. *Animal Behaviour*, 46(5), 847-860, doi:10.1006/anbe.1993.1270.
- Schneider, C., Hodges, K., Fischer, J., & Hammerschmidt, K. (2008). Acoustic Niches of Siberut Primates. *International Journal of Primatology*, 29(3), 601-613, doi:10.1007/s10764-007-9181-1.

- Schusterman, R. J. (2008). Vocal learning in mammals with special emphasis on pinnipeds. In D. K. Oller, & U. Griebel (Eds.), *The evolution of communicative flexibility: Complexity, creativity, and adaptability in human and animal communication* (pp. 41-70). Cambridge, MA: MIT Press.
- Scott-Phillips, T. C., & Kirby, S. (2010). Language evolution in the laboratory. *Trends in Cognitive Sciences*, 14(9), 411-417, doi:doi:10.1016/j.tics.2010.06.006.
- Seyfarth, R., & Cheney, D. (2008). Primate social knowledge and the origins of language. [10.1007/s11299-007-0038-2]. *Mind & Society*, 7(1), 129-142, doi:10.1007/s11299-007-0038-2.
- Seyfarth, R. M., & Cheney, D. L. (2010). Production, usage, and comprehension in animal vocalizations. [doi: DOI: 10.1016/j.bandl.2009.10.003]. *Brain and Language*, 115(1), 92-100.
- Seyfarth, R. M., Cheney, D. L., & Marler, P. (1980). Monkey responses to three different alarm calls - evidence of predator classification and semantic communication. *Science*, 210(4471), 801-803, doi:10.1126/science.7433999.
- Shadle, C. H. (1983). Experiments on the acoustics of whistling. *The Physics Teacher*, 21(3), 148-154.
- Slabbekoorn, H., & Smith, T. B. (2002). Bird song, ecology and speciation. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 357(1420), 493-503, doi:10.1098/rstb.2001.1056.
- Slik, J. W. F., Aiba, S.-I., Bastian, M., Brearley, F. Q., Cannon, C. H., Eichhorn, K. A. O., et al. (2011). Soils on exposed Sunda Shelf shaped biogeographic patterns in the equatorial forests of Southeast Asia. *Proceedings of the National Academy of Sciences*, 108(30), 12343-12347, doi:10.1073/pnas.1103353108.
- Slobodchikoff, C. N., Ackers, S. H., & van Ert, M. (1998). Geographic Variation in Alarm Calls of Gunnison's Prairie Dogs. *Journal of Mammalogy*, 79(4), 1265-1272.
- Slobodchikoff, C. N., & Coast, R. (1980). Dialects in the alarm calls of prairie dogs. *Behavioral Ecology and Sociobiology*, 7(1), 49-53, doi:10.1007/bf00302518.
- Slocombe, K., Kaller, T., Turman, L., Townsend, S., Papworth, S., Squibbs, P., et al. (2010). Production of food-associated calls in wild male chimpanzees is dependent on the composition of the audience. *Behavioral Ecology and Sociobiology*, 64(12), 1959-1966, doi:10.1007/s00265-010-1006-0.
- Slocombe, K. E., Waller, B. M., & Liebal, K. (2011). The language void: the need for multimodality in primate communication research. [doi: DOI: 10.1016/j.anbehav.2011.02.002]. *Animal Behaviour*, 81(5), 919-924.
- Slocombe, K. E., & Zuberbühler, K. (2007). Chimpanzees modify recruitment screams as a function of audience composition. *Proceedings of the National Academy of Sciences of the United States of America*, 104(43), 17228-17233, doi:10.1073/pnas.0706741104.
- Slocombe, K. E., & Zuberbühler, K. (2005a). Agonistic Screams in Wild Chimpanzees (*Pan troglodytes schweinfurthii*) Vary as a Function of Social Role. [doi: DOI: 10.1037/0735-7036.119.1.67]. *Journal of Comparative Psychology*, 119(1), 67-77.
- Slocombe, K. E., & Zuberbühler, K. (2005b). Functionally Referential Communication in a Chimpanzee. *Current Biology*, 15(19), 1779-1784, doi:10.1016/j.cub.2005.08.068.
- Slocombe, K. E., & Zuberbühler, K. (2006). Food-associated calls in chimpanzees: responses to food types or food preferences? *Animal Behaviour*, 72(5), 989-999, doi:10.1016/j.anbehav.2006.01.030.
- Smith, K., Kirby, S., & Brighton, H. (2003). Iterated Learning: A Framework for the Emergence of Language. *Artificial Life*, 9(4), 371-386, doi:10.1162/10645460322694825.

- Soha, J. A., & Marler, P. (2001). Vocal Syntax Development in the White-Crowned Sparrow (*Zonotrichia leucophrys*). *Journal of Comparative Psychology*, *115*(2), 172-180.
- Somers, P. (1973). Dialects in southern Rocky Mountain pikas, *Ochotona princeps* (Lagomorpha). *Animal Behaviour*, *21*(1), 124-137, doi:10.1016/s0003-3472(73)80050-8.
- Spillmann, B., Dunkel, L. P., Van Noordwijk, M. A., Amda, R. N. A., Lameira, A. R., Wich, S. A., et al. (2010). Acoustic Properties of Long Calls Given by Flanged Male Orang-Utans (*Pongo pygmaeus wurmbii*) Reflect Both Individual Identity and Context. *Ethology*, *116*(5), 385-395, doi:10.1111/j.1439-0310.2010.01744.x.
- Stephan, C., & Zuberbühler, K. (2008). Predation increases acoustic complexity in primate alarm calls. *Biology Letters*, *4*, 641-644, doi:10.1098/rsbl.2008.0488.
- Struhsaker, T. (1970). Phylogenetic implications of some vocalizations of Cercopithecus monkeys. In J. R. Napier, & P. H. Napier (Eds.), *Old World Monkeys: Evolution, Systematics, and Behavior* (pp. 365-444). New York: Academic Press.
- Sugiura, H., Tanaka, T., & Masataka, N. (2006). Sound transmission in the habitats of Japanese macaques and its possible effect on population differences in coo calls. *Behaviour*, *143*(8), 993-1012, doi:10.1163/156853906778623617.
- Suzuki, R., Buck, J. R., & Tyack, P. L. (2006). Information entropy of humpback whale songs. *Journal of the Acoustical Society of America*, *119*(3), 1849-1866.
- Tagliatalata, J. P., Russell, J. L., Schaeffer, J. A., & Hopkins, W. D. (2011). Chimpanzee Vocal Signaling Points to a Multimodal Origin of Human Language. *PLoS ONE*, *6*(4), e18852.
- Tagliatalata, J. P., Savage-Rumbaugh, S., & Baker, L. A. (2003). Vocal Production by a Language-Competent *Pan paniscus*. *International Journal of Primatology*, *24*(1), 1-17, doi:10.1023/a:1021487710547.
- Tamura, K., Dudley, J., Nei, M., & Kumar, S. (2007). MEGA4: Molecular evolutionary genetics analysis (MEGA) software version 4. *Molecular Biology & Evolution*, *24*, 1596-1599.
- Tanaka, T., Sugiura, H., & Masataka, N. (2006). Cross-Sectional and Longitudinal Studies of the Development of Group Differences in Acoustic Features of Coo Calls in Two Groups of Japanese Macaques. *Ethology*, *112*(1), 7-21, doi:10.1111/j.1439-0310.2006.01103.x.
- Taniguchi, I. (1985). Echolocation sounds and hearing of the greater Japanese horseshoe bat (*Rhinolophus ferrumequinum nippon*). *Journal of Comparative Physiology A*, *156*(2), 185-188, doi:10.1007/bf00610860.
- ten Cate, C., & Okanoya, K. (2012). Revisiting the syntactic abilities of non-human animals: natural vocalizations and artificial grammar learning. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1598), 1984-1994, doi:10.1098/rstb.2012.0055.
- Terrace, H., Petitto, L., Sanders, R., & Bever, T. (1979). Can an ape create a sentence? *Science*, *206*(4421), 891-902, doi:10.1126/science.504995.
- Thomas, D. W., Bell, G. P., & Fenton, M. B. (1987). Variation in Echolocation Call Frequencies Recorded from North American Vespertilionid Bats: A Cautionary Note. *Journal of Mammalogy*, *68*(4), 842-847.
- Thornton, A., & Malapert, A. (2009). The rise and fall of an arbitrary tradition: an experiment with wild meerkats. *Proceedings of the Royal Society B: Biological Sciences*, *276*(1660), 1269-1276, doi:10.1098/rspb.2008.1794.
- Todt, D. (1975). Social Learning of Vocal Patterns and Modes of their Application in Grey Parrots (*Psittacus erithacus*)<sub>1,2,3</sub>. *Zeitschrift für Tierpsychologie*, *39*(1-5), 178-188, doi:10.1111/j.1439-0310.1975.tb00907.x.
- Tomasello, M. (2008). *Origins of human communication*. Cambridge, MA: MIT Press.

- Tomasello, M., Savage-Rumbaugh, S., & Kruger, A. C. (1993). Imitative Learning of Actions on Objects by Children, Chimpanzees, and Enculturated Chimpanzees. *Child Development*, 64(6), 1688-1705, doi:10.1111/j.1467-8624.1993.tb04207.x.
- Tononi, G., Edelman, G. M., & Sporns, O. (1998a). Complexity and coherency: integrating information in the brain. [10.1016/S1364-6613(98)01259-5]. *Trends in Cognitive Sciences*, 2(12), 474-484, doi:10.1016/S1364-6613(98)01259-5
- Tononi, G., McIntosh, A. R., Russell, D. P., & Edelman, G. M. (1998b). Functional clustering: Identifying strongly interactive brain regions in neuroimaging data. *Neuroimage*, 7(2), 133-149, doi:10.1006/nimg.1997.0313.
- Townsend, S. W., Deschner, T., & Zuberbühler, K. (2008). Female Chimpanzees Use Copulation Calls Flexibly to Prevent Social Competition. *PLoS ONE*, 3(6), e2431, doi:10.1371/journal.pone.0002431.
- van Schaik, C. (1999). The socioecology of fission-fusion sociality in Orangutans. *Primates*, 40(1), 69-86, doi:10.1007/bf02557703.
- van Schaik, C. P., Ancrenaz, M., Borgen, G., Galdikas, B., Knott, C. D., Singleton, I., et al. (2003). Orangutan cultures and the evolution of material culture. *Science*, 299(5603), 102-105, doi:10.1126/science.1078004.
- van Schaik, C. P., Ancrenaz, M., Djojomoro, R., Knott, C. D., Morrogh-Bernard, H., Nuzuar, et al. (2009). Orangutan cultures revisited. In S. A. Wich, S. S. U. Atmoko, T. Mitra Setia, & C. P. Van Schaik (Eds.), *Orangutans: geographic variation in behavioral ecology and conservation* (pp. 299-310). New York: Oxford University Press.
- van Schaik, C. P., van Noordwijk, M. A., & Wich, S. A. (2006). Innovation in wild Bornean orangutans (*Pongo pygmaeus wurmbii*). *Behaviour*, 143, 839-876, doi:10.1163/156853906778017944.
- Vogel, C. (1973). Acoustical communication among free-ranging common Indian langurs (*Presbytis entellus*) in two different habitats of North India. *American Journal of Physical Anthropology*, 38(2), 469-479, doi:10.1002/ajpa.1330380250.
- von Bayern, A. M. P., Heathcote, R. J. P., Rutz, C., & Kacelnik, A. (2009). The Role of Experience in Problem Solving and Innovative Tool Use in Crows. *Current biology : CB*, 19(22), 1965-1968.
- Warren, K. S., Verschoor, E. J., Langenhuijzen, S., Heriyanto, Swan, R. A., Vigilant, L., et al. (2001). Speciation and intraspecific variation of Bornean orangutans, *Pongo pygmaeus pygmaeus*. *Molecular Biology and Evolution*, 18(4), 472-480.
- Waser, P. M., & Brown, C. H. (1986). Habitat acoustics and primate communication. *American Journal of Primatology*, 10(2), 135-154, doi:10.1002/ajp.1350100205
- Watson, C., & Caldwell, C. (2009). Understanding Behavioral Traditions in Primates: Are Current Experimental Approaches Too Focused on Food? *International Journal of Primatology*, 30(1), 143-167, doi:10.1007/s10764-009-9334-5.
- Weilgart, L., & Whitehead, H. (1997). Group-specific dialects and geographical variation in coda repertoire in South Pacific sperm whales. *Behavioral Ecology and Sociobiology*, 40(5), 277-285, doi:10.1007/s002650050343.
- Wheeler, B. (2010). Production and perception of situationally variable alarm calls in wild tufted capuchin monkeys (<i>Cebus apella nigritus</i>). *Behavioral Ecology and Sociobiology*, 64(6), 989-1000, doi:10.1007/s00265-010-0914-3.
- Whitehead, H., Dillon, M., Dufault, S., Weilgart, L., & Wright, J. (1998). Non-geographically based population structure of South Pacific sperm whales: dialects, fluke-markings and genetics. *Journal of Animal Ecology*, 67(2), 253-262, doi:10.1046/j.1365-2656.1998.00187.x.

- Whitehead, J. (1995). Vox alouattinae: A preliminary survey of the acoustic characteristics of long-distance calls of howling monkeys. *International Journal of Primatology*, 16(2), 121-144, doi:10.1007/bf02700156.
- Whiten, A., Goodall, J., McGrew, W. C., Nishida, T., Reynolds, V., Sugiyama, Y., et al. (1999). Cultures in chimpanzees. [10.1038/21415]. *Nature*, 399(6737), 682-685, doi:http://www.nature.com/nature/journal/v399/n6737/supinfo/399682a0\_S1.html.
- Whiten, A., Horner, V., Litchfield, C., & Marshall-Pescini, S. (2004). How do apes ape? *Animal Learning & Behavior*, 32(1), 36-52, doi:10.3758/bf03196005.
- Whiten, A., & Mesoudi, A. (2008). Establishing an experimental science of culture: animal social diffusion experiments. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 363(1509), 3477-3488, doi:10.1098/rstb.2008.0134.
- Whitfield, J. (2008). Across the Curious Parallel of Language and Species Evolution. *PLoS Biol*, 6(7), e186.
- Wich, S., Assink, P. R., Becher, F., & Sterck, E. H. M. (2002). Playbacks of Loud Calls to Wild Thomas Langurs (Primates; *Presbytis thomasi*): The Effect of Familiarity. *Behaviour*, 139(1), 79-87.
- Wich, S., Atmoko, S., Setia, T., & Schaik, C. (2009a). *Orangutans: geographic variation in behavioral ecology and conservation*: Oxford University Press.
- Wich, S., Swartz, K., Hardus, M., Lameira, A., Stromberg, E., & Shumaker, R. (2009b). A case of spontaneous acquisition of a human sound by an orangutan. *Primates*, 50(1), 56-64, doi:10.1007/s10329-008-0117-y.
- Wich, S. A., & de Vries, H. (2006). Male monkeys remember which group members have given alarm calls. *Proceedings of the Royal Society B: Biological Sciences*, 273(1587), 735-740, doi:10.1098/rspb.2005.3320.
- Wich, S. A., De Vries, H., Ancrenaz, M., Perkins, L. W., S. R., Suzuki, A., et al. (2009c). Orangutan life history variation. In S. A. Wich, S. S. U. Atmoko, T. M. Setia, & C. P. v. Schaik (Eds.), *Orangutans: Geographic variation in behavioral ecology and conservation* (pp. 65-76). Oxford, New York: Oxford University Press.
- Wich, S. A., Krützen, M., Lameira, A. R., Nater, A., Arora, N., Bastian, M. L., et al. (2012). Call cultures in orang-utans? *PLoS ONE*, 7(5), e36180, doi:10.1371/journal.pone.0036180.
- Wich, S. A., Schel, A. M., & de Vries, H. (2008). Geographic variation in Thomas langur (*Presbytis thomasi*) loud calls. *American Journal of Primatology*, 70, 566-574, doi:10.1002/ajp.20527.
- Wich, S. A., & Sterck, E. H. M. (2003). Possible audience effect in thomas langurs (primates; *presbytis thomasi*): An experimental study on male loud calls in response to a tiger model. *American Journal of Primatology*, 60(4), 155-159, doi:10.1002/ajp.10102.
- Wilbrecht, L., & Nottebohm, F. (2003). Vocal learning in birds and humans. *Mental Retardation and Developmental Disabilities Research Reviews*, 9(3), 135-148, doi:10.1002/mrdd.10073.
- Wiley, R. H., & Richards, D. (1978). Physical constraints on acoustic communication in the atmosphere: Implications for the evolution of animal vocalizations. *Behavioral Ecology and Sociobiology*, 3(1), 69-94, doi:10.1007/bf00300047.
- Wilson, T. A., Beavers, G. S., DeCoster, M. A., Holger, D. K., & Regenfuss, M. D. (1971). Experiments on the Fluid Mechanics of Whistling. *Journal of the Acoustical Society of America*, 50(1B), 366-372.
- Wrangham, R., & Books, B. (2010). *Catching fire: How cooking made us human*. London: Profile.
- Wright, T. F. (1996). Regional Dialects in the Contact Call of a Parrot. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 263(1372), 867-872, doi:10.1098/rspb.1996.0128.

- Wynne, C. D. L., & Bolhuis, J. J. (2008). Minding the gap: Why there is still no theory in comparative psychology. *Behavioral and Brain Sciences*, 31(02), 152-153, doi:doi:10.1017/S0140525X08003786.
- Xu, J., Gannon, P. J., Emmorey, K., Smith, J. F., & Braun, A. R. (2009). Symbolic gestures and spoken language are processed by a common neural system. *Proceedings of the National Academy of Sciences*, 106(49), 20664-20669, doi:10.1073/pnas.0909197106.
- Yorzinski, J. L., & Ziegler, T. (2007). Do naïve primates recognize the vocalizations of felid predators? *Ethology*, 113(12), 1219-1227, doi:10.1111/j.1439-0310.2007.01435.x.
- Yoshino, H., Armstrong, K. N., Izawa, M., Yokoyama, J. U. N., & Kawata, M. (2008). Genetic and acoustic population structuring in the Okinawa least horseshoe bat: are intercolony acoustic differences maintained by vertical maternal transmission? *Molecular Ecology*, 17(23), 4978-4991, doi:10.1111/j.1365-294X.2008.03975.x.
- Yoshino, H., Matsumura, S., Kinjo, K., Tamura, H., Ota, H., & Izawa, M. (2006). Geographical Variation in Echolocation Call and Body Size of the Okinawan Least Horseshoe Bat, *Rhinolophus pumilus* (Mammalia: Rhinolophidae), on Okinawa-jima Island, Ryukyu Archipelago, Japan. *Zoological Science*, 23(8), 661-667, doi:10.2108/zsj.23.661.
- Zhi, L., Karesh, W. B., Janzowski, D. N., Frazier-Taylor, H., Sajuthi, D., Gombek, F., et al. (1996). Genomic differentiation among natural populations of orang-utan (*Pongo pygmaeus*). *Current Biology*, 6, 1326-1336.
- Zimmermann, E., & Hafen, T. G. (2001). Colony specificity in a social call of mouse lemurs (*Microcebus* spp.). *American Journal of Primatology*, 54(3), 129-141, doi:10.1002/ajp.1018.
- Zuberbühler, K. (2000a). Causal cognition in a non-human primate: field playback experiments with Diana monkeys. *Cognition*, 76(3), 195-207, doi:10.1016/s0010-0277(00)00079-2.
- Zuberbühler, K. (2000b). Interspecies semantic communication in two forest primates. *Proceedings of the Royal Society of London Series B: Biological Sciences*, 267(1444), 713-718, doi:10.1098/rspb.2000.1061.
- Zuberbühler, K. (2002). A syntactic rule in forest monkey communication. *Animal Behaviour*, 63(2), 293-299, doi:10.1006/anbe.2001.1914.
- Zuberbühler, K., Noë, R., & Seyfarth, R. M. (1997). Diana monkey long-distance calls: messages for conspecifics and predators. *Animal Behaviour*, 53(3), 589-604, doi:10.1006/anbe.1996.0334

## Samenvatting

Spraak is een kenmerk van de mens. Echter, men weet nog weinig over de evolutie hiervan. Het blijft grotendeels onbekend welke kenmerken van geluidcommunicatie van onze naastverwanten – de mensapen- het basismateriaal zou kunnen vertegenwoordigen van spraak. De laatste vier decennia is het idee dat geluidscommunicatie in mensapen waardevol is in het ontrafelen van deze puzzel in de wetenschap afgenomen. Dit heeft ertoe geleid dat andere dierlijke modellen veel (onderzoeks) aandacht hebben gekregen, en er wordt gesuggereerd dat in primaten gebarentaal een betere kandidaat is als voorganger van spraak. Alhoewel heuristisch legitiem, deze theorie gaat voorbij aan het feit dat spraak fundamenteel opereert in de akoestische modaliteit. Daarom moet elke gegronde voorganger van spraak, die gevonden worden in non-primate taxa of niet-akoestische modaliteiten, én eerst aanwezig zijn én functioneel relevant zijn in menselijke voorouders. Wanneer deze aanwezig zijn buiten het akoestische domein, dan moet het verspreid zijn of getransformeerd zijn in geluidscommunicatie. Het voornaamste doel van deze studie is dan ook het onderzoeken in wat voor mate kenmerken van spraak gevonden kunnen worden in orang-oetan geluidscommunicatie.

Orang-oetans vertegenwoordigen de oudste afstamming van mensapen. Ondanks de ideeën in de laatste vier decennia in de wetenschap, zijn er recente studies die laten zien dat orang-oetans een veelbelovend modelsoort zijn voor het ontrafelen van spraakevolutie. Hoewel flexibiliteit van gedrag in het gebruik van geluiden, begrip en productie gevonden kunnen worden in verschillende primaten, zijn de zeldzame voorbeelden van flexibiliteit in het *verwerven* van geluid voornamelijk een aangelegenheid in orang-oetans. Dit wil zeggen, orang-oetans laten enig bewijs zien voor de capaciteit van het leren en het opnemen van nieuwe geluiden in hun repertoire. Zodoende, onderzoek ik orang-oetan geluidscommunicatie in natuurlijke omstandigheden. In dit onderzoek neem ik als beginpunt de geluiden die de individuen feitelijk produceren in plaats van het streven om menselijke woorden aan te leren. Van deze geluiden identificeer ik kenmerken van spraak die mogelijk aanwezig zijn in orang-oetans in de afwezigheid van volledige spraak. Tevens gebruik ik vakterminologie van taalkunde en semiologie (sociale wetenschappen toegewijd aan de studie van, respectievelijk menselijke spraak en

symbolen) om een directe vergelijkende benadering te maken, gebaseerd op structurele en gedrags-, homologie tussen mensapen en mensen. Daarom bestudeer ik kenmerken van spraak die traditioneel beschouwd werden als uniek menselijk, aangezien elke gegronde hypothese voor spraak evolutie het ontstaan van deze kenmerken passend zou moeten verklaren. Deze kenmerken van vergevorderde communicatie zijn: “traditionele transmissie”, “dubbelzinnigheid”, “willekeurigheid”, “leerbaarheid”, “semantiek”, en “samenstelling”. Het eerste kenmerk dat wordt behandeld is “traditionele transmissie” (hoofdstuk 2 en 3), welke verklaart dat individuen hun geluidsrepertoire leren van gelijksoortigen binnen sociale groepen. Dit leidt tot variatie tussen groepen en twee verschillende potentiële variatie typen: “accenten”, wanneer hetzelfde geluidstype akoestisch varieert tussen populaties, of “dialecten”, wanneer sommige geluidstypen gebiedsspecifiek zijn en afwezig zijn in andere gebieden (hoofdstuk 2). Accenten komen veelal voor in landzoogdieren door genetische, omgevings-, en/of sociale factoren. Deze factoren reageren verschillend op elkaar per taxon waar accenten voorkomen, echter sociale factoren geven de meest belangrijke factor weer in het ontstaan van accenten in niet menselijke primaten. Tegengesteld zijn dialecten, die zeer zelden voorkomen in landzoogdieren en maar 4 keer gerapporteerd zijn, waarvan twee keer in mensapen; in chimpanzees en in orang-oetans. Een nadere analyse van de aanwezigheid van dialecten in orang-oetans is daarom van belang (hoofdstuk 3). De aan-, en afwezigheid van populatie specifieke orang-oetan geluiden die worden geproduceerd tijdens het maken van een nest of het contact tussen moeder-kind kunnen niet volledig verklaard worden door genetische of ecologische variatie tussen gebieden. Dit ondersteunt de opvatting dat deze geluiden meest waarschijnlijk zijn ontstaan en vervolgens verspreid via sociaal leren binnen de populatie waar ze aanwezig zijn. De observatie dat twee populaties verschillende geluiden kunnen hebben met dezelfde functie geeft ook aan dat er geen specifieke relatie is tussen het type geluiden die orang-oetans produceren voor deze functie, wat aangeeft dat deze geluiden arbitrair zijn.

Vervolgens onderzoek ik “dubbelzinnigheid”, het vermogen om foute beweringen te maken. Dit gaat over de manipulatie van een specifiek gedeeld communicatief kenmerk, om andere individuen te misleiden (hoofdstuk 4). Orang-oetans in de Tuanan populatie (Centraal Kalimantan, Indonesie), produceren drie varianten van één van hun alarmgeluiden: kiss squeaks zonder hulpmiddel (unaided), met de hand en op bladeren. Consistent met de algemene regel in de bio-akoestiek, zijn de frequenties (Hz) van

unaided kiss-squeaks negatief proportioneel met de lichaamsgrootte; grotere individuen produceren een lagere frequentie kiss-squeaks dan kleinere individuen. Het gebruik en de positie van de hand of bladeren voor hun lippen tijdens de productie van de kiss-squeak verminderd de frequentie van de kiss-squeak (maar geen andere akoestische maat van de kiss-squeaks), mogelijk tot het signaleren van een grotere lichaamsgrootte. Niet gehabituëerde orang-oetans produceren dezelfde proportie van unaided kiss-squeaks tegen mensen als gehabituëerde individuen. Echter, ze produceerden meer kiss-squeaks met hand dan gehabituëerde individuen en ze waren de enige die kiss-squeaks op bladeren produceerden. In andere woorden, de productie van gemodificeerde kiss-squeak varianten was gecorreleerd met het waargenomen gevaar van de orang-oetans, en deze geluiden omvatten een grotere lichaamsgrootte dan hun eigenlijke grootte.

Verder onderzoek ik “willekeurigheid”, een kenmerk van een communicatie systeem waar geen directe connectie is tussen het signaal en de functie. Sinds orang-oetans willekeurige geluiden kunnen produceren voor dezelfde functie (zie hoofdstuk 3), werd ik geïnteresseerd in het onderzoeken of orang-oetans ook dezelfde type geluiden kunnen produceren voor willekeurige verschillende functies (hoofdstuk 5). Om te voorzien in een complete benadering in het onderzoek naar willekeurigheid, analyseerde ik het gebruik en de potentiële functie van de kiss-squeaks unaided, met hand en op bladeren in een andere orang-oetan populatie (Cabang Panti, West Kalimantan, Indonesië), de analyse herhalend die voorheen uitgevoerd was in de Tuanan populatie (hoofdstuk 4). Deze twee populaties behoren tot dezelfde sub-soorten en leven gedeeltelijk in eenzelfde habitat. In Cabang Panti vertonen de frequenties van unaided kiss-squeaks dezelfde negatieve relatie met lichaamsgrootte, wat ook de mogelijkheid geeft tot manipulatie van lichaamsgrootte door de individuen wanneer ze handen of bladeren gebruiken. Desalniettemin waren kiss-squeaks met de hand buitengewoon ongebruikelijk. Nog belangrijker was dat de regelmatig geproduceerde kiss-squeak op bladeren niet resulteerde in een lagere frequentie, en daarom geen grotere lichaamsgrootte overbracht. Bovendien waren kiss-squeaks op bladeren niet gerelateerd aan waargenomen bedreigingen, omdat ze niet meer gebruikt werden door niet gehabituëerde individuen. Deze verschillen tussen de populaties in het gebruik van hetzelfde geluid impliceert dat de functies die worden toegeschreven aan de kiss-squeak op bladeren willekeurig zijn.

Vervolgens focus ik op “leerbaarheid”, wat bepaald dat gebruikers van een systeem van geluidscommunicatie elementen kunnen leren van het systeem van een

andere soort. Traditioneel, wordt dit vermogen verondersteld afwezig te zijn in niet menselijke primaten. Echter recentelijk is er gerapporteerd dat een orang-oetan een menselijk geluid heeft aangeleerd – fluiten (Wich et al. 2009b). Deze kwestie geeft één van de duidelijkste voorbeelden weer van interspecifiek leren door een niet menselijke primate. Om het fluiten in orang-oetans verder te onderzoeken, herhaal ik en breid ik de voorgaande referentie studie uit (hoofdstuk 6). Tot op heden zijn er 10 orang-oetans bekend die kunnen fluiten in Europa en in de USA, wat laat zien dat het geen curiositeit is. Een specifieke orang-oetan, Ujian, paart correct enkele, dubbele en driedubbele fluitjes die door een menselijke onderzoeker wordt geproduceerd. Dit bevestigt dat fluiten geen gemoedsvertoon en geen reflex is, maar, dat tegengesteld tot de meeste niet menselijke primaten geluiden, het onder vrijwillige controle gegeven wordt. Verder zijn fluitopnamens van 3 verschillende orang-oetans kwalitatief en kwantitatief vergeleken. Elk individu liet een specifieke fluitstijl zien, wat resulteerde in akoestisch verschillend fluiten. Dit kan verklaard worden doordat elk individu onderworpen was aan verschillende leermogelijkheden, wat laat zien dat het fluiten in orang-oetans ervaringsafhankelijk is. Samengevat ondersteunen deze resultaten het beeld dat fluiten in orang-oetans verworven is door leren.

Ik behandel daarna respectievelijk “semantiek” en “samenstelling”, welke inhouden dat sommige onderdelen van geluid specifieke betekenissen hebben en dat onderdelen gecombineerd kunnen worden in grotere betekenisvolle delen. Ondanks dat deze kenmerken gedemonstreerd zijn in verschillende aapsoorten, zijn ze in mensapen nog niet onderzocht. Dit soort data zijn nodig om te verifiëren of aap semantiek en samengestelde geluiden standhouden in de evolutie van de mens. Ik onderzoek specifiek de mate van emissie van verschillende orang-oetan alarmgeluiden tegen een predator (tijger) model en een controle (wit) model in twee populaties, een waarvan tijgers uitgestorven zijn en een waar tijgers op orang-oetans jagen (hoofdstuk 7). Er zijn geen verschillen gevonden in de hoeveelheid geluiden per tijdseenheid tussen de modellen in beide populaties, wat weergeeft dat de geluiden niet predator specifiek waren. Eveneens waren er geen verschillen in hoeveelheid geluiden per tijdseenheid tussen de populaties, wat laat zien dat de hoeveelheid predatoren het gebruik van alarmgeluiden in orang-oetans niet beïnvloed. Op grond hiervan lijken aap semantiek en samengestelde geluiden in de context van alarm niet homolog te zijn in mensapen.

Samengevat, van de zes kenmerken van spraak die onderzocht zijn in deze studie (traditionele transmissie, dubbelzinnigheid, willekeurigheid, leerbaarheid, semantiek, en samenstelling), zijn er vier gevonden in orang-oetans. Semantiek en samenstelling zijn niet aanwezig in het alarmgeluid systeem van orang-oetans, wat suggereert dat deze kenmerken van spraak onderzocht zouden moeten worden in geluiden die gemaakt worden in een andere context, zoals tijdens het foerageren. Binnen het orang-oetan geluidsrepertoire van wilde populaties zijn kiss-squeaks veelbelovende geluiden voor toekomstige studies die geïnteresseerd zijn in het behandelen van nieuwe aspecten van gedragsflexibiliteit in het geluidsdomein van mensapen. Geluiden die slechts gemaakt worden door bepaalde individuen in gevangenschap kunnen in de toekomst een nieuw inzicht verschaffen in het onderliggende mechanisme van het leren van geluiden in mensapen.

Gebaseerd op de bevindingen van voorgaande hoofdstukken stel ik voor dat sommige voorbeelden van orang-oetan geluidscommunicatie kunnen wijzen op kenmerken van de voorganger van spraak (hoofdstuk 8), namelijk de kiss-squeaks met hand en op bladeren. Deze geluiden geven unieke multimodale signalen weer, waarvan de handen en manipulatie van objecten (bladeren) synchroon betrokken worden bij het maken van het geluid. Op deze wijze brengen deze geluiden het nonverbale (gebarentaal) domein en het geluidsdomein van mensapen samen, wat laat zien dat theoretische opvattingen die gefocused zijn op één van de domeinen verenigbaar zijn. Dit laat een nieuw gezichtspunt zien van vermoedelijke spraak voorgangers, welke volledig verenigbaar is met data die of op gebarentaal of op geluiden gebaseerde hypothesen ondersteunen. Bovendien vertonen kiss-squeaks met hand en op bladeren een manier hoe een menselijke aapachtige voorouder, met een beperkte capaciteit om nieuwe geluiden te leren, zijn communicatieve repertoire, betreft de productie van geluid, vergroot zou kunnen hebben. De “simpele” combinatie van een geluid met een veranderend gebaar produceert een nieuwe variant van het geluid, wat een positieve selectie stimuleert voor een vergroot repertoire. Hierdoor kunnen individuen meer en meer diverse informatie overbrengen. Op deze wijze kunnen multimodale signalen in onze voorouders, betreft synchrone gebaren en geluiden, gezorgd hebben voor het ontstaan van verdere stappen in de evolutie van spraak. Door het verder onderzoeken naar kenmerken van geluidscommunicatie in mensapen, gecombineerd met onze

verstandhouding van gebarencommunicatie, kunnen we mogelijke startpunten afleiden van menselijke spraak en de evolutie van taal in onze hominin voorouders.

Hernieuwde inspanning in de studie van mensapen in het gebruik van geluiden, productie en verwerving zullen waardevol blijven voor de reconstructie van de evolutie van spraak. De identificatie van gedeelde kenmerken tussen de communicatiesystemen van mensen en mensapen staan de identificatie van vermeende voorgangers van spraak toe. Zulke kenmerken kunnen dan verder onderzocht worden voor het verkrijgen van meer puzzelstukken van de spraakevolutie. De kenmerken die onderzocht zijn in deze studie in orang-oetans kunnen ook in andere mensapen bestudeerd worden. Concepten van de taalkunde en semiologie kunnen verder voor belangrijke aanwijzingen zorgen voor het zoeken van en onderzoeken naar nog ononderzochte gedeelde kenmerken tussen menselijke spraak en de geluidscommunicatie in mensapen.

## Ringkasan

Berbicara/bersuara adalah merupakan ciri dari manusia, walaupun evolusinya masih belum terlalu dipahami. Sampai saat ini masih belum diketahui fitur komunikasi yang digunakan oleh kerabat terdekat kita – kera besar – yang mungkin merupakan titik awal evolusi suara manusia. Selama empat dekade terakhir, pemikiran ilmiah telah berpindah dari gagasan bahwa panggilan komunikasi di kera besar berharga untuk menjawab teka-teki ini. Akibatnya, hewan model lainnya menerima lebih banyak usaha dan upaya penelitian, dan telah diusulkan bahwa komunikasi gestural primata bukan manusia merupakan kandidat yang lebih baik sebagai *precursor*/pendahulu dalam hal berbicara. Meskipun sah secara heuristik, pengertian ini mengabaikan fakta bahwa berbicara beroperasi *fundamental* dalam modalitas akustik. Oleh karena itu, setiap pendahulu dalam hal berbicara yang ditemukan dalam non-primata modalitas taksa atau non-akustik, akan menunjukkan dan berfungsi relevan pada leluhur manusia dan, ketika hadir di luar domain akustik, itu harus diperluas atau ditransfer ke dalam komunikasi suara. Oleh karena itu, fokus utama dari penelitian ini adalah untuk mengetahui sejauh mana blok dalam hal berbicara/bersuara yang dapat ditemukan di orangutan yang dinamakan komunikasi.

Orangutan adalah yang tertua, dan merupakan garis keturunan tertua dalam kera besar, tetapi studi terbaru mengindikasikan bahwa mereka mewakili spesies model yang menjanjikan dalam mengungkap evolusi dalam hal berbicara. Sementara fleksibilitas perilaku dalam penggunaan panggilan, pemahaman dan produksi dapat ditemukan di primata bukan manusia, kasus-kasus langka orangutan fleksibilitas dalam akuisisi suara didominasi orangutan. Artinya, orangutan menunjukkan beberapa bukti kemampuan belajar dan menggabungkan panggilan baru ke dalam repertoar mereka. Oleh karena itu, saya menyelidiki komunikasi panggilan orangutan dalam dua kondisi: alam (hutan) dan naturalistik (diluar hutan). Daripada mencoba untuk mengajarkan bahasa manusia, saya memulai dengan suara yang dihasilkan oleh masing-masing individu. Dari suara yang telah dihasilkan, saya mengidentifikasi fitur suara orangutan yang hampir mendekati suara manusia. Selanjutnya, saya menggunakan pendekatan terminologi dari *linguistic* dan *semiotic* (ilmu sosial yang digunakan untuk mempelajari bahasa manusia dan isyarat, masing-masing) untuk membangun pendekatan perbandingan langsung berdasarkan struktur dan perilaku homologi antara kera besar dan manusia. Maka dari itu, saya

memeriksa blok suara yang dianggap mewakili keunikan fitur manusia, karena setiap hipotesis untuk evolusi berbicara/bersuara harus menjelaskan munculnya fitur ini. Fitur-fitur yang dimaksud adalah: *traditional transmission*, *prevarication*, *arbitrariness*, *learnability*, *semanticity*, dan *patterning*.

Hal pertama tentang berbicara yang saya bicarakan adalah *traditional transmission* (Bab 2 dan 3), yang menyatakan bahwa individu belajar bagaimana untuk bersuara dari individu lain dalam kelompok sosial yang sama. Hal tersebut yang menyebabkan variasi antara kelompok dan variasi yang berbeda tersebut adalah: *accents* adalah jenis yang sama dari suara, tapi kedengarannya berbeda dalam populasi, atau *dialects* yang merupakan suara yang hanya dapat didengar di lokasi tertentu, dan tidak dapat didengar di lokasi lain (Bab 2). *Accent* pada umumnya terjadi pada mamalia darat, baik karena faktor genetik, lingkungan dan atau sosial. Faktor-faktor ini berinteraksi secara berbeda per takson yang menunjukkan aksen, namun faktor sosial merupakan faktor yang paling penting dalam munculnya *accent* pada primata bukan manusia. Sebaliknya, *dialect* sangat langka di antara mamalia darat, hanya ditemukan 4 kasus, yang dua di antaranya terjadi kera besar, yaitu pada simpanse dan orangutan. Pembahasan lebih lanjut dari kehadiran dialek di orangutan ini sehingga (Bab 3). Ada tidaknya suara orangutan yang dilakukan selama membangun sarang atau saat berinteraksi antara ibu dan anak tidak dapat menjelaskan perbedaan genetik atau ekologis yang jelas antar daerah, yang menunjukkan bahwa suara tersebut diciptakan dan kemudian tersebar melalui hubungan sosial dalam suatu populasi. Pengamatan bahwa dua populasi mungkin memiliki suara yang berbeda dengan fungsi yang sama menandakan bahwa tidak ada hubungan tertentu antara jenis suara yang dihasilkan orangutan untuk fungsi tertentu, menyiratkan bahwa hal tersebut tidak mempunyai dasar.

Selanjutnya, saya menyelidiki *prevarication*, yaitu kemampuan untuk membuat pernyataan palsu. Hal tersebut melibatkan manipulasi tertentu dalam fitur komunikasi bersama, dengan tujuan untuk membohongi individu lain. Orangutan pada populasi Tuanan (Kalimantan Tengah, Indonesia) menghasilkan tiga varian dalam satu suara tanda bahaya mereka: tanpa alat-bantu, tangan, dan daun *kiss-squeak*. Secara konsisten sesuai dengan aturan umum di bioakustika, frekuensi dari tanpa alat-bantu *kiss-squeak* tidak tergantung dari ukuran badan, dimana individu yang lebih besar akan menghasilkan *kiss-squeak* yang lebih rendah. Penggunaan dan posisi tangan atau daun di depan bibir selama *kiss-squeak* mengurangi frekuensi *kiss-squeak* (tetapi tidak terjadi perubahan lain dalam

*kiss-squeak*), untuk menandakan ukuran badan yang lebih besar. Orangutan liar yang belum mengenal manusia menghasilkan unaided *kiss-squeak* kepada manusia sama seperti orangutan lain yang telah mengenal manusia. Namun orangutan yang tidak terbiasa dengan manusia melakukan lebih banyak *kiss-squeak* dengan tangan pada manusia, dibandingkan dengan orangutan yang sudah mengenal manusia, dan hanya orangutan yang belum mengenal manusia menggunakan daun untuk menghasilkan suara. Artinya, *kiss-squeak* yang dihasilkan telah dimodifikasi berkorelasi dengan ancaman yang dirasakan oleh orangutan, dan suara tersebut menyampaikan ukuran tubuh yang lebih besar dari ukurannya yang sebenarnya.

Saya kemudian menjelaskan *arbitrariness*, karakteristik dari sistem komunikasi di mana tidak ada hubungan langsung antara sinyal dan fungsinya. Karena orangutan dapat membuat suara abritrasi dengan fungsi yang sama (lihat Bab 3), saya ingin melihat apakah orangutan juga dapat membuat jenis suara yang sama untuk fungsi abritrasi yang berbeda (Bab 5), yang akan memberikan pemahaman lengkap untuk prinsip dalam komunikasi orangutan. Untuk keperluan tersebut saya telah melakukan analisis yang sama pada penggunaan dan fungsi dari tanpa-alat-bantu, tangan, dan daun *kiss-squeak*, dalam populasi orangutan yang berbeda (Cabang Panti, Kalimantan Barat, Indonesia) sebagai perbandingan dengan populasi di Tuanan (Bab 4). Kedua populasi tersebut berasal dari sub-spesies yang sama dan tinggal di terpisah di habitat yang sama. Di Cabang Panti, *kiss-squeak* tanpa alat bantu menunjukkan proporsi yang negatif dalam hubungannya dengan ukuran badan, memungkinkan potensi manipulasi ukuran badan oleh individu dengan menggunakan tangan atau daun. Meskipun demikian, *kiss-squeak* dengan tangan sangat langka. Dan yang lebih penting, penggunaan daun dalam panggilan tidak menghasilkan frekuensi akustik yang lebih rendah pada populasi ini, yang karena itu tidak menyampaikan ukuran tubuh yang lebih besar. Selain itu, *kiss-squeak* dengan menggunakan daun yang tidak berhubungan dengan ancaman yang dirasakan, karena tidak sering digunakan oleh individu yang belum pernah mengenal manusia. Perbedaan antara populasi dalam penggunaan suara menunjukkan bahwa fungsi *kiss-squeak* dikaitkan dengan abritrasi.

Setelah itu, saya fokus pada *learnability*, yaitu bahwa pengguna dari salah satu sistem suara dapat mempelajari elemen dari sistem species yang lain. Kemampuan untuk belajar dari spesies lain, yang sampai saat ini di asumsikan tidak terdapat pada primata selain manusia, tetapi dengan laporan terbaru yang memaparkan bagaimana orangutan

belajar bersiul dari manusia, yang kemudian merubah pendapat tersebut. Untuk mempelajari lebih dalam tentang orangutan yang dapat bersiul, saya mereplikasi dan meneruskan referensi penelitian sebelumnya (Bab 6). Saat ini, terdapat 10 orangutan yang dapat bersiul yang berada di Eropa dan Amerika Serikat, menunjukkan bahwa orangutan bersiul bukanlah suatu keanehan. Satu orangutan tertentu, 'Ujian', cocok dengan keadaan tersebut, dua dan tiga siul dihasilkan oleh eksperimen manusia, mengkonfirmasi bahwa bersiul bukanlah tindakan afektif dan *reflexive* tetapi, sebaliknya untuk suara primata bukan manusia hal tersebut berada di bawah kontrol kesadaran. Rekaman bersiul dari 3 orangutan yang berbeda dibandingkan secara kualitatif dan kuantitatif. Setiap individu menunjukkan gaya bersiul spesifik sehingga menghasilkan akustik yang berbeda. Hal ini dapat dijelaskan bahwa setiap individu mendapat kesempatan belajar yang berbeda. Secara keseluruhan, hasil ini mendukung pandangan bahwa bersiul pada orangutan diperoleh melalui pembelajaran.

Saya kemudian menjelaskan *semanticity* dan *patterning*, masing-masing menyatakan bahwa beberapa unit suara memiliki makna tertentu dan bahwa unit dapat digabungkan menjadi unit yang lebih besar. Fitur ini telah di demonstrasikan terhadap beberapa spesies monyet, tetapi belum dilakukan pada kera besar. Data tersebut diperlukan untuk memverifikasi apakah suara monyet *semantic* dan *pattern* muncul dalam silsilah evolusi manusia. Saya menyelidiki tingkat emisi suara tanda bahaya orangutan yang berbeda dibandingkan dengan predator model (harimau) dan kontrol (putih) dalam dua populasi yang berbeda, satu di mana harimau punah dan satu di mana harimau memburu orangutan (Bab 7). Tidak ada perbedaan frekuensi suara antara model pada kedua populasi, menunjukkan bahwa suara tidak spesifik untuk predator. Selain itu, tidak ada perbedaan frekuensi suara diantara populasi, menunjukkan bahwa ukuran predator tidak mempengaruhi penggunaan suara tanda bahaya pada orangutan. Oleh karena itu, suara tanda bahaya pada monyet *semantic* dan *patterned* tampaknya tidak memiliki hubungan langsung dengan kera besar.

Secara keseluruhan, dari enam fitur suara yang diperiksa dalam penelitian ini (*traditional transmission, prevarication, arbitrariness, learnability, semanticity* dan *patterning*), empat ditemukan pada orangutan. *Semanticity* dan *patterning* yang tidak muncul dalam sistem suara tanda bahaya orangutan, menunjukkan harus diselidiki lebih lanjut bahwa kedua fitur tersebut dalam kera besar dapat digunakan dalam konteks lain, seperti mencari makan. Dalam repertoar suara orangutan dari populasi liar, *kiss-squeak*

berpotensi besar untuk menjadi bahan penelitian yang menarik dalam menangani aspek-aspek baru fleksibilitas perilaku dalam domain suara kera besar. Dalam penangkaran, suara yang dihasilkan oleh individu tertentu dapat memberikan wawasan baru masa depan dalam mekanisme suara yang mendasari pembelajaran pada kera besar.

Berdasarkan temuan dari bab-bab sebelumnya, saya mengusulkan bahwa beberapa contoh komunikasi suara orangutan dapat menunjukkan fitur prekursor berbicara (Bab 8), yaitu orangutan yang menghasilkan suara dengan menggunakan tangan dan daun. Suara ini merupakan sinyal *multimodal* yang unik, yang melibatkan aksi tangan dan manipulasi obyek (yaitu daun) serentak dengan menghasilkan suara. Dengan demikian, suara tersebut merupakan gabungan dari domain gestural dan suara dari kera besar, menunjukkan bahwa fokus teori pada masing-masing domain dapat dipertemukan. Hal ini menunjukkan sudut pandang baru untuk prekursor pembicaraan manusia, yang sepenuhnya kompatibel dengan data pendukung baik hipotesis gerakan-atau panggilan berbasis. Selain itu, *kiss-squeak* yang dihasilkan dengan tangan dan daun menunjukkan bahwa nenek moyang kita juga memiliki kapasitas untuk mempelajari suara baru dan dapat memperluas kemampuan untuk bersuara. Kombinasi yang sederhana suara dengan modifikasi gerakan menghasilkan varian panggilan baru, menghasilkan seleksi positif dalam meningkatkan repertoar sehingga individu dapat mengirimkan informasi yang lebih banyak dan beragam. Dengan demikian, sinyal *multimodal* di nenek moyang kita yang melibatkan gerakan sinkron dan panggilan bisa memberikan dasar munculnya tekanan selektif terhadap langkah-langkah berikutnya dalam evolusi berbicara. Dengan mempelajari fitur berkomunikasi kera besar, dikombinasikan dengan pemahaman kita tentang komunikasi gestural, kita dapat menyimpulkan titik awal suara manusia dan evolusi bahasa nenek moyang hominin.

Memperbaharui upaya penelitian penggunaan suara pada kera besar, dalam menghasilkan dan akuisisi berharga untuk rekonstruksi evolusi berbicara. Identifikasi fitur komunikasi diantara manusia dan kera besar memungkinkan identifikasi prekursor berbicara putatif. Fitur tersebut kemudian diperiksa lebih lanjut untuk memberikan informasi tentang evolusi berbicara. Fitur yang dieksplorasi dalam penelitian juga dapat dilakukan untuk kera besar lain, dan konsep *linguistic* dan *semiotic* lebih lanjut dapat memberikan informasi penting bagi pencarian dan penyelidikan fitur yang belum teruji antara suara manusia dan kera besar.



## Acknowledgments

This research project would not have happened without authorization from the Indonesian Ministry of Research and Tecnology (RISTEK, in particular Ibu Yeti), the Indonesian Ministry of Forestry (PHKA), the National Park of the Gunung Leuser (TNGL), the Leuser Ecosystem Management Board in Aceh (BPKEL), and the Gunung Palung National Park (BTNGP). Logistic support was provided by the Sumatran Orangutan Conservation Programme (SOCP, in particular Ian Singleton and Desanti), and by Universitas Nasional in Jakarta (UNAS, in particular Suci Utami-Atmoko, Tatang Mitra Setia, Fitri, Achi and Dewi) and Universitas Tanjungpura (UNTAN, in particular Peter Widmann, Hari Prayogo, Yuliati Indrayani) who acted as sponsor and counterpart.

Crucial financial support was provided by Fundação para a Ciência e Tecnologia (Portugal, SFRH/BD/44437/2008), the Wenner-Gren Foundation (USA), Lucie Burgers Foundation for Comparative Behaviour Research, Arnhem, the Netherlands, Stichting Dr. J.L. Dobberke voor Vergelijkende Psychologie (the Netherlands), Schure-Beijerinck-Popping Fonds (the Netherlands), Ruggles-Gates Fund for Anthropological Scholarship of the Royal Anthropological Institute of Great Britain and Ireland, Primate Conservation, Inc. (USA) and Fundação Calouste Gulbenkian (Portugal).

Field work was helped by Ike Nayasilana, Misdi, Sumurudin, Basarudin, Mdin, Roma, Pak Usman, Pak Matplin and Arwin, Cheryl Knott, Meredith Bastian, and all the field staff at Sampan Getek and Gunung Palung. Good nutrition in the field was assured by Atun and Ahmed. During field work, I supervised and was helped by Kim Nouwen, Cedric Hall, Eva Topelberg, Sanne Gasteren, and Wiebe Rinsma. Life in camp was pleasantly shared with Madeleine Hardus, Adam van Casteren, Julia Myatt, Kirsten Tuson, Cedric Girard-Buttoz and Anna Marzec.

Thank you Paul, Gerold, Sabrina and Nadine for your wonderful visit to Indonesia.

Life in Medan was enriched by the company of Madeleine Hardus, David Dellatore, Ian Singleton, Serge Wich, Gabriella Fredriksson, Graham Usher, Ellen Meulman, Gail Campbell-Smith, Miran, Andrea Permana, Edit and Simone Sauren.

Thank you Gail and Gabriella for allowing us in your field sites when permission to enter Aceh was not provided. Cheryl, it was a pleasure to collaborate with you. I am grateful to all co-authors of this work.

All work done in captivity was conducted under fantastic hosting by Robert Shumaker, Lisa Goodwin, JP, Bern Kowalsky, Sandra Reichler, Anke Jacob, Klaus Pyzora, Mike Ebert, and Alex Sliwa.

Liesbeth, Han, Berry and Marusha, thank you for your conversations and discussions, and for not letting me feel lost in the department.

Serge, thank you for believing in the person, more than in the project.

Outstanding summary translations were made by Madeleine and Pris (Dutch translation) and by Dave and Diana (Indonesian translation).

Special thanks to those who offered shelter and support to Madeleine and me between field seasons: Marga and Peter, Pris and René, Marc and Trudi, Dan and Rhonda, Serge and Tine, Pai e Mãe.

And still, all this would have been impossible without the omnipresence of my soul mate, my best friend, lover, wife and mother of my child. Thank you, Madeleine, meu anjo!

## Curriculum Vitae

Adriano Reis e Lameira was born in Lisbon on the 1<sup>st</sup> of September 1982. After finishing his secondary school at the Liceu Camões in Lisbon in 2000, he started his Biology study at the Faculty of Sciences of the University of Lisbon. He finished his Masters in 2004, and in the same year, seeking more adventure, he joined a group of students and researchers to study the behavior of wild orangutans in the peat swamps of Indonesian Borneo, under the supervision of Serge Wich and Carel van Schaik. This work led him to co-found the Pongo Foundation for orangutan research and conservation in 2006. To pursue the study of orangutans, this time in Sumatra, he returned to Indonesia in 2007. A year later his actual PhD project started, of which the present thesis is the outcome.





## List of publications

### Book Chapters

1. Hardus M., **Lameira A.**, Singleton I., Knott C., Morrogh-Bernard H., Ancrenaz M., Utami S., Wich S. (2009) The Orangutan Vocal and Sound Repertoire: with a focus on Geographical variation. *In* Orangutans. Wich S., Utami S., Setia T., van Schaik C., eds. Oxford University Press, pp. 49-64.
2. Delgado R., **Lameira A.**, Davila Ross M., Husson S., Morrogh-Bernard H., Wich S. (2009) Geographic variation in Orangutan Long Calls. *In* Orangutans. Wich S., Utami S., Setia T., van Schaik C., eds. Oxford University Press, pp. 215-224.

### Articles

1. **Lameira A.**, Hardus M., Hall C., Spruijt B., Sterck E., de Vries H., van Noordwijk M., van Schaik C., Wich s. (in press). Predator guild does not influence orangutan alarm call rates and combinations. *Behavioural Biology and Socioecology*. doi: 10.1007/s00265-012-1471-8
2. **Lameira A.**, Hardus M., Kowalsky B., de Vries H., Spruijt B., Sterck E., Shumaker R., Wich S. (accepted). Orangutan whistling and implications for the emergence of an open-ended call repertoire: a replication and extension. *Journal of the Acoustical Society of America*
3. Hardus M., de Vries H., Dellatore D., **Lameira A.**, Menken S., Wich S. (in press). Socioecological correlates of inter-individual variation in orangutan diets at Ketambe, Sumatra. *Behavioural Biology and Socioecology*. doi: 10.1007/s00265-012-1463-8
4. **Lameira A.**, Hardus M., Wich S. (2012). Orangutan Instrumental Gesture-Calls: Reconciling Acoustic and Gestural Speech Evolution Models. *Evolutionary Biology*, 39, 415-418. doi: 10.1007/s11692-011-9151-6
5. Wich S., Krützen M. , **Lameira A.** , Nater A., Arora N., Bastian M., Meulman E., Morrogh-Bernard H., Atmoko, S., Pamungkas J., Perwitasari-Farajallah D., Hardus M., van

Noordwijk M., van Schaik C. (2012) Call cultures in orang-utans? PLoS ONE, 7, e36180. doi:10.1371/journal.pone.0036180

☺ These authors contributed equally to this work.

6. Hardus M., **Lameira A.**, Zulfa A., Atmoko S., de Vries H., Wich S. (2012). Behavioral, Ecological, and Evolutionary Aspects of Meat-Eating by Sumatran Orangutans (*Pongo abelii*). *International Journal of Primatology*, 33, 287-304. doi: 10.1007/s10764-011-9574-z

7. Hardus M., **Lameira A.**, Menken S., Wich S. (2012). Effects of logging on orangutan behavior. *Biological Conservation*, 146:177-187. doi: 10.1016/j.biocon.2011.12.014

8. **Lameira A.**, Delgado R., Wich S. (2010). Review of geographic variation in terrestrial mammalian acoustic signals: Human speech variation in a comparative perspective. *Journal of Evolutionary Psychology*, 8:309-332. doi: 10.1556/JEP.8.2010.4.2

9. Spillmann B., Dunkel L., van Noordwijk M., Amda R., **Lameira A.**, Wich S., van Schaik C. (2010). Acoustic properties of long calls given by flanged male orang-utans (*Pongo pygmaeus wurmbii*) reflect both individual identity and context. *Ethology*, 116:385-395. doi: 10.1111/j.1439-0310.2010.01744.x

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

11. Hardus M., **Lameira A.**, van Schaik C., Wich S. (2009). Tool use modifies sound production in wild orangutans: a functionally deceptive innovation? *Proceedings of the Royal Society B*, 276:3689-3694. doi: 10.1098/rspb.2009.1027

12. **Lameira A.**, Wich S. (2008). Orangutan long call degradation and individuality over distance: a playback approach. *International Journal of Primatology*, 29:615-625. doi: 10.1007/s10764-008-9253