

# PHONETISCHE UND KORPUS-LINGUISTISCHE METHODEN BEI DER ANALYSE VOKALER KOMMUNIKATION VON FREILEBENDEN SCHIMPANSEN IM TAÏ NATIONAL FOREST

*Sven Grawunder<sup>1,2</sup>, Natalie T. Uomini<sup>3</sup>, Catherine Crockford<sup>2</sup>*

*<sup>1</sup>Institut für Skandinavistik, Frisistik und allgemeine Sprachwissenschaft, Christian-Albrechts-University, Kiel, <sup>2</sup>Arbeitsgruppe Primatologie, Max-Planck-Institut für Evolutionäre Anthropologie, Leipzig, <sup>3</sup> Abteilung Sprach- und Kulturevolution, Max-Planck-Institut für Menschheitsgeschichte, Jena  
grawunder@eva.mpg.de*

**Kurzfassung:** Über die letzten Jahrzehnte haben sich unsere Kenntnisse über die vokalen Kommunikationssysteme bei Menschenaffen enorm erweitert und aber auch das Verständnis zur Evolution des Kommunikationssystems unserer eigenen Spezies stark gewandelt. Allein nur die Studien zur Kommunikation unter Schimpansen haben gezeigt, dass es eine Reihe von gruppenspezifische und kontextspezifischen Lautäußerungen gibt. Es ist klar geworden, dass die Komplexität der Strukturen und ihrer kommunikativen wie sozialen Funktionen lange unterschätzt wurde und selbst auch die akustischen und artikulatorischen Vorgänge noch nicht vollständig verstanden sind. Die dramatische Abnahme der verbliebenen Populationen ist allerdings wenig hilfreich und wirkt sich ebenso auf die Bedingungen für deren Erforschung und die Anforderung (z.B. an Software) an sie aus. Für diesen Beitrag werden anhand einer aktuellen Studie zu akustischen und artikulatorischen Charakteristika von Vokalisierungen bei freilebenden Schimpansen im Taï National Forest (Elfenbeinküste) grundlegende Probleme analysiert wie sie sich aus Sicht der Erforschung des Sprechens beim Menschen ergeben. Das vorgestellte Projekt fokussiert die individuelle und gruppenspezifische Ausprägung kontextsensitiver Laute unter besonderer Berücksichtigung der beobachtbaren oralen Artikulation. Hierbei möchten wir gerade die Desiderata im methodisch-technischen Bereich (Datenerhebung, Transkription und Annotation, akustische und artikulatorische Parametrisierung von Audio und Video, Datenverwaltung) aufzeigen und versuchen Lösungsvorschläge zu unterbreiten, aber auch Hilfe in der Community suchen.

## 1 Warum studieren wir vokale Kommunikation unter Menschenaffen?

Haben wir es bei den Voaklisationen unter Primaten insbesondere den Menschenaffen mit intentionalen Lautäußerungen in komplexe Gesellschaften zu tun, die hierarchisch geordnet sind und sich dynamisch entwickeln? Oder sind es am Ende nur interagierende Automaten? Warum sollten solche Fragen von Interesse sein? Das Staunen über die Schwierigkeiten die wir beim Nachbau natürlicher menschlicher Kommunikation haben scheint zu wachsen, in dem Maße wie wir über die Details im Spracherwerb, Perception und Produktion erfahren. Wenn wir also etwas über den Ursprung der Vielschichtigkeit der Kommunikation unserer eigenen Primatenart herausfinden wollen, so sollte helfen, auch die unserer nächsten Verwandten etwas genauer zu studieren. Bereits bei den mit dem Menschen nicht so nah verwandten Primatenarten wie z.B. den Pavianen oder Mandrills können Gruppenmitglieder individuelle Identifizierung anhand der Rufe vornehmen. Sie zeigen dabei ein Vokalisierungssystem, das die dynamische Struktur einer

Gruppe abbildet ([1]), sodass es nicht verwundert, dass beides auch für Menschenaffen zutrifft. Doch haben sich bei selbst genetisch noch weiter entfernten Meerkatzen für mehrere Spezies (Grünmeerkatze, Dianameerkatze) kontextspezifische Alarm-Laute gefunden, d.h. spezifische Rufe die in Abhängigkeit von der Art und Richtung aus der die vermeintliche Gefahr (Greifvogel, Schlange oder Raubkatze) kommt anders lauten (z.B. [2]). Zunächst war die Verhaltensforschung davon ausgegangen, dass es sich bei Alarmrufen eher um Lautäußerungen handelt, die aus den affektiv-emotionalen Zuständen in dieser Situation stammen, doch die Komplexität der Konstellationen in denen diese Rufe (unabhängig von Alarmsituationen) auftauchen, lässt daran zweifeln, sodass inzwischen auch intentionale und volitionale Lautäußerungen diskutiert werden [3]. Doch auch ohne diesen Schritt, gibt es noch eine Menge an Desiderata in der Grundlagenforschung zum generellen Verständnis der vokalen Kommunikation bei höheren Primaten.

Die Verhaltensforschung bei Schimpansen (Goodall [4]), Gorillas (Fossey [5]), Orangutans (Rijksen [6]) oder Bonobos (de Waal.[7]) war zunächst von begrenzten Rufinventaren ausgegangen. Die einfache Frage nach der Größe des Repertoires, d.h. wieviel Rufarten (*calltypes*) es eigentlich gibt, wurde dabei aber im Laufe der Zeit immer schwerer zu beantworten. Die vielmehr beobachteten graduellen Abstufungen, Kombinationen und auch Mischformen ([8, 9, 10]) sind so kaum zu erfassen, so dass ein systematisches, empirisch basiertes Beschreibungssystem auch dieser Vielfalt Rechnung tragen muss. Lediglich bleiben bestimmte Basisbegriffe für Rufarten (*calltypes*) konstant, die in den Beschreibungen der ForscherInnen immer wieder auftauchen. So spricht man bei Schimpansen von *Hoos, Grunts, Barks, Screams, Squeaks* und *Whimpers*). Für die anderen Menschenaffenarten finden sich vergleichbare Terminologien, doch sind diese generell oft nicht kompatibel und oftmals sehr idiosynkratisch gebraucht. Bei Schimpansen sind diese genannten basalen *calltypes* wage mit den menschlichen Stimmregistern zu vergleichen, doch eben nur wage, weil es sich nicht nur um rein phonatorische bzw. stimmliche Unterschiede handelt, sondern auch artikulatorische und sogar situative Aspekte eine Rolle spielen. Es stellt sich also die Aufgabe zumindest für die jeweilige Spezies einheitliche Referenzen zu den Beschreibungskategorien zu finden und den Feldforschern ein Inventar in die Hand zu geben, das es möglich macht, möglichst einfach und reliabel Dokumentationen von Beobachtungen zu erstellen bzw. Annotationen zu Bild- und Tonaufnahmen zu ergänzen. Nichtsdestotrotz handelt es sich (wie auch bei den Methoden zur Analyse menschlichen Kommunikationsverhaltens) um semi-objektive Methoden deren Reproduzierbarkeit und Validität beständig verbessert werden soll.

## 2 Vokalisationen von freilebenden Schimpansen im Tai National Forest

Die Abteilung für Primatologie des MPI betreibt seit über 15 Jahren eine Reihe von Forschungsstationen entlang der verbliebenen Habitate der letzten großen Menschenaffenarten in West- und Zentralafrika, d.h. Uganda, DR Kongo, Guinea, Sierra Leone und Elfenbeinküste. Ziel ist es, über die noch verbliebenen Populationen deren 'kulturelle' Unterschiede zu beschreiben. Das Verständnis von Kultur ist dabei dergestalt erweitert, dass es sich hierbei um tradierte und vom einzelnen Individuum erlernte und in der jeweiligen Gruppe relativ homogen ausgeprägte Verhaltensweisen und -muster handelt, die zum einen eine positive, ein bestimmtes Verlangen erfüllende Handlung darstellen und zum anderen je nach Umgebung für die ganze Gruppe adaptiert werden (s. ausführlicher dazu [11]). Bei den 'Schimpansenkulturen' geht es dabei um solches Verhalten wie das Nüsse knacken mit zwei Steinen (Hammer-Ambos), das Insektenfischen mit einem Blatt, das Ausgraben von Erdbienen mit einem langen Stock, das Früchte waschen ([12]), aber eben auch um (vokales und anderes) kommunikatives Verhalten. Es ist dabei zu betonen, dass diese Art von Forschung nicht ohne Weiteres durch solche an Tieren in Gefangenschaft,

seien es Zoos oder Reservaten, ersetzt werden können, weil sich hier vehemente Unterschiede zu den freilebenden Gruppen gezeigt haben: Freilebende Schimpansen (aber auch andere Menschenaffenarten) vokalisieren weitaus mehr und möglicherweise auch differenzierter als solche in Gefangenschaft. Nichtsdestotrotz sind natürlich auch experimentelle Verhaltensstudien mit Tieren in Gefangenschaft wichtige Ergänzungen oder auch Vorstudien zu solchen in der Wildnis. Gerade die Studien zu freilebenden Schimpansen haben gezeigt, dass es eine Fülle von kontextspezifischen Lauten bzw. Rufen ([13, 14]) gibt, aber auch einen weitaus großen Teil von funktional flexiblem vokalem Verhalten, für dessen Verständnis also eine pragmatische Analyse notwendig ist um es in seinem semantischen Gehalt bewerten zu können. Dabei sind die Rufe zum Teil sehr gruppenspezifisch (z.B. [15]) und passen sich an, wenn Individuen in eine andere Gruppe wechseln [16].

Für das Projekt zur 'Phonetik' der Schimpansenlauten - um das es hier hauptsächlich geht - liegt es also auf der Hand, bestehende Erfahrungen aus den Bereichen der verhaltenorientierten Biologie und der Phonetik/Linguistik zusammen zu bringen. Die dem Projekt vorliegenden Daten sind Videoaufnahmen aus dem Taï National Forest (Elfenbeinküste), die über die letzten sieben Jahre von unterschiedlichen Feldforschern (hauptsächlich unter Leitung von Dr. Crockford) erstellt wurden. Sie sind unter dem Methodenparadigma der Fokus-Tier-Methode (*focal observation, focal(-animal)-sampling*) erhoben worden, bei der meist über längere Zeit Aktionen und Interaktionen eines Tieres (Fokus-Tier) erfasst werden. Ein solche Erfassung kann eine, mehrere oder alle Verhaltensweisen eines Individuums umfassen und kann, muss aber nicht kontinuierlich erfolgen (<http://www.spektrum.de/lexikon/biologie/fokus-tier-methode/25473>). Diese Anforderung charakterisiert auch die Aufnahmen, bei denen die Feldforscher üblicherweise auch während der Aufnahme das Geschehen soweit wie nötig kommentieren (Akteure, Handlung, etc.). Hauptsächlich handelt es sich um zwei verschiedene Schimpansengruppen mit derzeit 27 erfassten Tieren unterschiedlichen Alters, Rangs und Geschlechts.

### 3 Akustisch-artikulatorische Analyse von Vokalisierungen

Wir verallgemeinern an dieser Stelle zu 'Vokalisierungen' und 'akustisch-artikulatorischer' Analyse, weil wir nicht *per se* von phonetischen Eigenschaften sprechen wollen, um die Analogie zum menschlichen Sprachlaut nicht zu überstrapazieren. Dennoch besteht genau hier das Ziel vieler Arbeiten die sich mit dem vokalen Verhalten von Primaten beschäftigen: die kontextspezifischen, erlernten Details der vokalen Kommunikationen zu analysieren und dabei möglichst viel über die Gruppendynamik und die individuelle Variation von Kontext zu Kontext zu kontrollieren. Es steht zum Beispiel der Verdacht, dass in Nah- und Kontaktkommunikation weitaus größere Freiheitsgrade der artikulatorischen Variation genutzt werden können als das in Distanzrufen der Fall ist, weil bei letzteren Anforderungen der akustischen Effizienz bzw. Effektivität im Vordergrund stehen. Eine solche passiert multimodal unter Ausnutzung der speziesspezifischen Charakteristika (Körpergröße, Geschlechtsdimorphismus, Vokaltraktmorphologie) aber auch der Gegebenheiten in der Umwelt (Trommeln auf Wurzeln, Gesten, Proxemik). Fokussieren wir auf vokale Kommunikation und Artikulation, so spielt insbesondere bei den Schimpansen die Modulation der Lippen eine zentrale Rolle. Umgekehrt erscheint die Artikulation der Zunge im Vergleich zum Menschen eingeschränkt, was im Widerspruch zu Modellen **möglicher** Artikulation und mithin Lautbildung führt. Fitch et al. [17] argumentieren, dass bereits Rhesusmakaken theoretische alle primären Vokalqualitäten erzeugen können. Allerdings werden diese bisher nicht beobachtet. Dabei stellt sich schlicht zum einen die Aufgabe, die beobachtbaren Indizien und akustischen Korrelate in einen zeitlichen Zusammenhang zu bringen. Auf der anderen Seite sind jedoch noch grundlegendere Daten etwa zur Steuerung

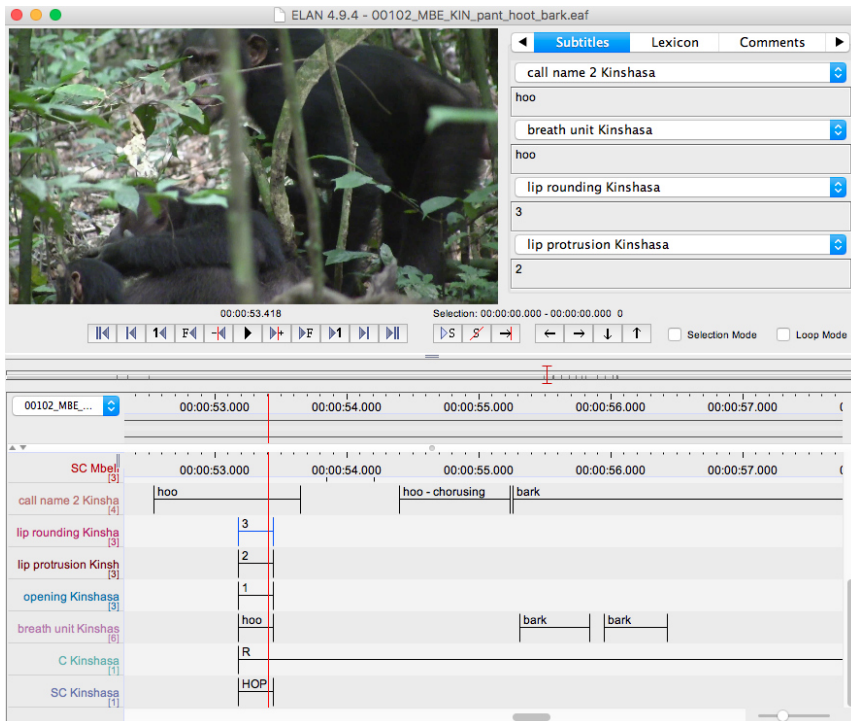
der Zunge und anderer Artikulatoren nötig um die entsprechenden Modelle zu ergänzen und zu verfeinern ([18, 19]). Rein auf akustischen Analysen basierend sind bei Schimpansen- und Bonobolauten bisher eine Reihe von häufigen nichtlinearen stimmlichen Phänomenen (*subharmonics, octave jumps, biphonation*) beschrieben worden (s. z.B. [20]), die so im menschlichen Repertoire nur als Extreme auftauchen. Zudem hat aber der Menschenaffen-Vokaltrakt doch ein paar Besonderheiten, deren Eigenschaften bisher nicht geklärt sind, so etwa die Funktion der Kehlsäcke ([21, 22]) oder die Modulation der ingressiven Laute oder der velopharyngealen/nasalen Passagen bzw. deren Auswirkungen.

#### 4 Phonetische und korpus-linguistische Methoden bei den Erforschung vokaler Kommunikation von Menschenaffen

Ziel des Projektes ist es also die vorhandenen heterogenen Daten der *focal observation* soweit aufzubereiten, dass eine systematische, replizierbare akustische und artikulatorische Analyse. Als Basiseinheit wurde für die Annotation die *breath unit (BU)* gewählt, weil diese noch in aller Regel einem *calltype* zugewiesen werden kann und akustisch-auditiv gut abgegrenzt erscheint, etwa durch hörbare Atmung bzw. Luftstromwechsel. Die BU wäre in diesem Sinne also mit einer Art Silbe zu vergleichen. Mehrere BUs können einen komplex zusammengesetzten *call* bilden, wie einen *pant hoot* der z.B. zu beobachten ist, wenn ein ranghöheres Tier der Gruppe begrüßt wird [23]. *Pant hoots* bestehen aus egressiven und ingressiven *hoos*, gefolgt von einer *built-up* Phase und einem *scream* als Klimax. Die eingebetteten BUs können sich dabei natürlich von einzeln auftretenden unterscheiden, wobei die Frequenz der Wiederholung oder auch Intensität der einzelnen Rufe ebenfalls eine Rolle spielen [20]. D.h. ein *call* kann mehrere BUs enthalten und steht selbst in einem zeitlich weitergefassten *context*. Solch ein *context* kann im engeren Sinn sozialer Art (Grüßen, Betteln, Aggressionsanzeige etc.) oder anderer Art (Alarm, Jagd, Begegnung mit andere Gruppe etc.) sein.

Als geeignete Videotranskriptionssoftware wurde das vom MPI für psycholinguistik in Nijmegen entwickelte ELAN (<https://tla.mpi.nl/tools/tla-tools/elan/>) gewählt u.a. weil es eine weitaus frame-genaue Analyse erlaubt, eine unbegrenzte Anzahl von Annotations-ebenen zulässt und auch einen Export in andere Formate (z.B. PRAAT TextGrids) ermöglicht. Allerdings ist eine manuelle bzw. halbautomatische Adjustierung der Vorsegmentierung der BUs für eine artefaktfreie akustische Analyse unerlässlich, da die videoframe-basierte akustische Signalalignierung sich als zu ungenau erwiesen hat. Ausgehend von einer dreidimensionalen Parametrisierung der Lippenartikulation (*jaw opening, lip protrusion (= Lippenstülpung), lip rounding*) wurden für sichtbare Artikulationen mit nichtüberlappenden BUs (d.h. solche in den nur ein Tier vokalisiert). Die semi-objektiven Parameter finden auf einer maximal 4-stufigen Messskala (z.B. für *lip rounding*: 0 = *retracted*; 1 = *unrounded*; 2 = *little-rounded*; 3 = *very-rounded*) ihren Eintrag. Wie in Abbildung 1 zu sehen ist, bedeutet dies, dass auch zum Teil verdeckte oder nicht frontal zur Kamera befindliche Gesichter zu beurteilen. Die initiale Zuordnung der derzeit ca. 1300 BUs zu den *calltypes* konnte bis auf eine kleine Anzahl von ca. 2% bei einem *recheck* bestätigt werden. Bei einem Reliabilitätscheck mit 10% der uns zur Verfügung stehenden Datenpunkte bei den Ratings zur Lippenartikulation wurde ca. 80% Übereinstimmung gemessen.

Die akustische Analyse versucht dazu die essenziellen Dimensionen der Rufe mit so wenig Parametern wie möglich zu erfassen; so in Form von Energiekonzentration auf der Frequenzebene sowie Signal-Rausch-Abstand (präziser Harmonics-to-Noise Ratio). Die Messungen sind jeweils auf die gesamte *breath unit* ausgerichtet und werden als Durchschnittswerte erfasst. Als Ergänzung dienen dazu in der Zeitebene die Dauern der BUs sowie die Anstiege (*slopes*) der Grundfrequenz und der Hüllkurve. Damit sollen zunächst die stabilsten Eigenschaften der *call-*



**Abbildung 1** – Videoannotation der Lippenartikulation mit Hilfe von ELAN: multilayer annotation von *call (name)*, *lip protrusion*, *lip rounding*, (*jaw*) *opening*, *breath unit*, *social context (SC)*, (*other non-social*) *context*. Das Gesicht des Schimpansenweibchens (Kinshasa) ist teilweise verdeckt. Ein anderes Weibchen (Mbeli) ist in der Nähe und ruft ebenfalls. Der situative Kontext ist (*R*)esting und der soziale Kontext *HOP* = *HearOtherParty* (Videoaufnahme: Liran Samuni)

*types* erfasst werden und mit den Kontext-/Situationsanalysen und den artikulatorischen Analysen verknüpft werden. So lassen sich auch wo es möglich ist LPC-basierte Formantschätzungen in den vokoiden (oder wenn man so will vokalartigen) Anteilen der BUs vornehmen. Hierbei handelt es sich um quasi speziesspezifische Probleme der F0- und F1-Messungen, weil z.B. in einer Reihe von *calls* Grundfrequenzen über der natürlichen ersten Vokaltraktresonanz auftreten. Um Oversampling bzw. Pseudoreplikation zu vermeiden, wird pro *call* nur eine BU ausgewählt bzw. die Messungen aller BUs vom gleichen *calltype* in einem *call* zusammengefasst.

Die bisherigen Ergebnisse, basierend auf Hauptkomponentenanalyse und Diskriminantenanalyse, bestätigen die Basisrufkategorien (*calltypes*: *hoo*, *grunt*, *bark*, *scream*) als Hauptattraktoren in dem von uns aufgestellten Parameterraum ([24]). Ebenfalls scheinen die Ergebnisse der artikulatorischen Parameter zu bekräftigen, dass die bisherigen Modelle der Vokaltraktakustik (vgl. [19]) bestätigt bleiben.

## 5 Methodische-technische Desiderata

Die Hürden bei der Datengewinnung sind neben denen der hohen Temperaturen und hohen Luftfeuchtigkeit manchmal auch die Bedingungen unter denen die Feldforschung stattfindet, weil die Gebiete zum Teil an Zonen angrenzen in denen bürgerkriegsähnliche Zustände herrschen. Andere Einschränkungen stellen die selbst auferlegten ethischen Richtlinien dar (<http://www.eva.mpg.de/primat/ethical-guidelines.html>), die zum Schutz der bedrohten

Arten beitragen sollen. So soll direkter Kontakt mit den Tieren vermieden werden und eine Mindestabstand (von 7 m) eingehalten werden um die Übertragung von Krankheiten auf die Tiere zu verhindern. Damit verbieten sich jegliche invasiven und halb-invasiven Methoden an lebenden Tieren. In diesem Zusammenhang stehen auch die erhöhten Anforderung an die Qualität der Aufnahmen bzw. Robustheit der Aufnahmetechnik, die nicht nur eine Richtcharakteristik bieten muss sondern auch die räumliche Zuordnung der akustischen Vorgängen ermöglichen soll. Umgekehrt steht in der Rohdatenverarbeitung die Dedektierung und Filterung der Hintergrundgeräusche im Vordergrund. Diese sind zum einen von der Tageszeit abhängig in verschiedenen Frequenzbändern auftretende ‘Insektenpegel’. Desweiteren gleichzeitige Rufe der Gruppe also Mehrstimmigkeit die aufgefunden werden muss. Daneben stehen die Kommentare der FeldforscherInnen, die in ihrer Häufigkeit kein Problem darstellen aber dennoch herausgefiltert werden müssen. Zusätzliche adhoc-Notizen (z.B. in Form von simplen *tags*) sollten synchron mitaufgezeichnet werden können.

Fragen zu den Möglichkeiten des Arten- und Naturschutzes sind ebenso Teil der Forschung am MPI EVA, wobei der Einfluss des Menschen auf die Lebensräume der Menschenaffen (Jagd, Wilderei, Abholzung) untersucht wird, um somit auch ein genaueres Bild von den Auswirkungen bestimmter Gegenmaßnahmen u.a. erworben werden kann. Sonst vielfach übliche Ausstattung einzelner Tiere mit Funktransmittern ist aus oben genannten Gründen nicht möglich. Eine alternative Möglichkeit bietet der Einsatz von *passive acoustic monitoring (PAM)* Systemen in einem Areal [25]. Ideal wäre dabei extrem robuste (u. kostengünstige) Kamera-Fallen mit in diese Art von Monitoring einbinden zu können und dieses Material gezielt für sog. *Animal Biometry* nutzen zu können. Während die Videoaufnahme aus solchen *arrays* oft qualitativ nicht hochwertig genug sind um einzelne Tier zu identifizieren, steht dies bei den Fokus-Tieren und ihren Interaktanten an vorderster Stelle. Würde nun die Fokus-Tier-Methode soweit erweitert werden, dass längere Zeiträume zusammenhängend aufgezeichnet würden, setzte dies höhere Anforderung in der Vorsortierung des Materials. Eine effektive Gesichtsdetektion (zur Vorsegmentierung von Videosequenzen) wäre dabei sehr wünschenswert um die sehr zeitaufwändige manuelle Arbeit der Identifizierung und vor allem der Annotation zu erleichtern. Einer zuverlässigen Gesichtererkennung sind durch die dunkle bis schwarze Pigmentierung der meisten Menschenaffenarten und kleine Sclerafläche wahrscheinlich z.Z. noch Grenzen gesetzt. Neben den Anforderungen an eine in diesem Sinne erweiterbare Videotranskriptionssoftware steht die Identifizierung einzelner Individuen über das akustische Signal als nächste Herausforderung.

## Literatur

- [1] FISCHER, J., K. HAMMERSCHMIDT, D. L. CHENEY, und R. M. SEYFARTH: *Acoustic features of male baboon loud calls: Influences of context, age, and individuality*. *The Journal of the Acoustical Society of America*, 111(3), S. 1465–1474, 2002.
- [2] SEYFARTH, R. M., D. L. CHENEY, und P. MARLER: *Vervet monkey alarm calls: semantic communication in a free-ranging primate*. *Animal Behaviour*, 28(4), S. 1070–1094, 1980.
- [3] TOWNSEND, S. W., S. E. KOSKI, R. W. BYRNE, K. E. SLOCOMBE, B. BICKEL, M. BOECKLE, I. BRAGA GONCALVES, J. M. BURKART, T. FLOWER, F. GAUNET, H. J. GLOCK, T. GRUBER, D. A. W. A. M. JANSEN, K. LIEBAL, A. LINKE, Á. MIKLÓSI, R. MOORE, C. P. VAN SCHAIK, S. STOLL, A. VAIL, B. M. WALLER, M. WILD, K. ZUBERBÜHLER, und M. B. MANSER: *Exorcising grice’s ghost: an empirical approach to studying intentional communication in animals*. *Biological Reviews*, S. 000–000, 2016. doi:10.1111/brv.12289. URL <http://dx.doi.org/10.1111/brv.12289>.

- [4] GOODALL, J.: *The chimpanzees of Gombe: patterns of behavior*. The Belknap Press of Harvard University Press, Cambridge, 1986.
- [5] FOSSEY, D.: *Gorillas in the Mist*. Houghton Mifflin Harcourt, 2000.
- [6] RIJKSEN, H. D.: *A field study on Sumatran orangutans (Pongo pygmaeus abelli Lesson 1872): ecology, behaviour and conservation*. Veenman & Zonen, Wageningen, The Netherlands, 1978.
- [7] DE WAAL, F. B.: *The Communicative Repertoire of Captive Bonobos (Pan Paniscus), Compared To That of Chimpanzees*. *Behaviour*, 106(3), S. 183–251, 1988. doi:<https://doi.org/10.1163/156853988X00269>. URL <http://booksandjournals.brillonline.com/content/journals/10.1163/156853988x00269>.
- [8] CROCKFORD, C. und C. BOESCH: *Call combinations in wild chimpanzees*. *Behaviour*, S. 397–421, 2005.
- [9] HEDWIG, D., K. HAMMERSCHMIDT, R. MUNDRY, M. M. ROBBINS, und C. BOESCH: *Acoustic structure and variation in mountain and western gorilla close calls: a syntactic approach*. *Behaviour*, 151(8), S. 1091–1120, 2014.
- [10] SCHAMBERG, I., D. L. CHENEY, Z. CLAY, G. HOHMANN, und R. M. SEYFARTH: *Call combinations, vocal exchanges and interparty movement in wild bonobos*. *Animal Behaviour*, 122, S. 109–116, 2016. doi:<http://dx.doi.org/10.1016/j.anbehav.2016.10.003>. URL <http://www.sciencedirect.com/science/article/pii/S0003347216302421>.
- [11] BOESCH, C.: *Wild Cultures - A comparison between chimpanzee and human culture*. Cambridge University Press, 2012.
- [12] WHITEN, A., J. GOODALL, W. C. MCGREW, T. NISHIDA, V. REYNOLDS, Y. SUGIYAMA, C. E. G. TUTIN, R. W. WRANGHAM, und C. BOESCH: *Cultures in chimpanzees*. *Nature*, 399(6737), S. 682–685, 1999. URL <http://dx.doi.org/10.1038/21415>.
- [13] CROCKFORD, C., R. M. WITTIG, R. MUNDRY, und K. ZUBERBÜHLER: *Wild chimpanzees inform ignorant group members of danger*. *Current Biology*, 22(2), S. 142–146, 2012.
- [14] CROCKFORD, C. und C. BOESCH: *Context-specific calls in wild chimpanzees, Pan troglodytes verus: analysis of barks*. *Animal Behaviour*, 66(115-125), 2003.
- [15] CROCKFORD, C., I. HERBINGER, L. VIGILANT, und C. BOESCH: *Wild chimpanzees produce group-specific calls: a case for vocal learning?* *Ethology*, 110(3), S. 221–243, 2004.
- [16] WATSON, S. K., S. W. TOWNSEND, A. M. SCHEL, C. WILKE, E. K. WALLACE, L. CHENG, V. WEST, und K. E. SLOCOMBE: *Vocal learning in the functionally referential food grunts of chimpanzees*. *Current Biology*, 25(4), S. 495–499, 2015. doi:[10.1016/j.cub.2014.12.032](https://doi.org/10.1016/j.cub.2014.12.032). URL <http://dx.doi.org/10.1016/j.cub.2014.12.032>.
- [17] FITCH, W. T., B. DE BOER, N. MATHUR, und A. A. GHAZANFAR: *Monkey vocal tracts are speech-ready*. *Science Advances*, 2(12), S. e1600723, 2016.
- [18] HOLSTEGE, G. und H. H. SUBRAMANIAN: *Two different motor systems are needed to generate human speech*. *Journal of Comparative Neurology*, 2015.

- [19] BOË, L.-J., F. BERTHOMMIER, T. LEGOU, G. CAPTIER, C. KEMP, T. R. SAWALLIS, Y. BECKER, A. REY, und J. FAGOT: *Evidence of a vocalic proto-system in the baboon (*papio papio*) suggests pre-hominin speech precursors*. *PLOS ONE*, 12(1), S. e0169321–, 2017. URL <http://dx.doi.org/10.1371/journal.pone.0169321>.
- [20] RIEDE, T., A. C. ARCADI, und M. J. OWREN: *Nonlinear acoustics in the pant hoots of common chimpanzees (*Pan troglodytes*): Vocalizing at the edge*. *The Journal of the Acoustical Society of America*, 121(3), S. 1758–1767, 2007.
- [21] HEWITT, G., A. MACLARNON, und K. E. JONES: *The functions of laryngeal air sacs in primates: A new hypothesis*. *Folia Primatologica*, 73(2-3), S. 70–94, 2002. URL <http://www.karger.com/DOI/10.1159/000064786>.
- [22] DE BOER, B.: *Loss of air sacs improved hominin speech abilities*. *Journal of Human Evolution*, 62(1), S. 1–6, 2012. doi:<http://dx.doi.org/10.1016/j.jhevol.2011.07.007>. URL <http://www.sciencedirect.com/science/article/pii/S0047248411002004>.
- [23] NOTMAN, H. und D. RENDALL: *Contextual variation in chimpanzee pant hoots and its implications for referential communication*. *Animal Behaviour*, 70(1), S. 177–190, 2005.
- [24] GRAWUNDER, S., N. UOMINI, und C. CROCKFORD: *Acoustic characteristics of chimpanzee vocalizations based on field data from two Tai groups (East/South)*, 2017 in Vorb.
- [25] KALAN, A., A. K. PIEL, R. MUNDRY, R. M. WITTIG, C. BOESCH, und H. S. KÜHL: *Passive acoustic monitoring reveals group ranging and territory use: A case study of wild chimpanzees (*Pan troglodytes*)*. *Frontiers in Zoology*, 13, 34, 2016. doi:10.1186/s12983-016-0167-8.