CrossMark

ORIGINAL PAPER

Learned vocal and breathing behavior in an enculturated gorilla

Marcus Perlman[1] · Nathaniel Clark[2]

Abstract We describe the repertoire of learned vocal and breathing-related behaviors (VBBs) performed by the enculturated gorilla Koko. We examined a large video corpus of Koko and observed 439 VBBs spread across 161 bouts. Our analysis shows that Koko exercises voluntary control over the performance of nine distinctive VBBs, which involve variable coordination of her breathing, larynx, and supralaryngeal articulators like the tongue and lips. Each of these behaviors is performed in the context of particular manual action routines and gestures. Based on these and other findings, we suggest that vocal learning and the ability to exercise volitional control over vocalization, particularly in a multimodal context, might have figured relatively early into the evolution of language, with some rudimentary capacity in place at the time of our last common ancestor with great apes.

✉ Marcus Perlman
  mperlman@wisc.edu

[1] Department of Psychology, University of Wisconsin-Madison, 1202 W. Johnson St., Madison, WI 53707, USA

[2] Department of Psychology, University of California, Santa Cruz, Santa Cruz, CA, USA

Introduction

Examining the vocal abilities of great apes is crucial to understanding the evolution of human language and speech since we diverged from our last common ancestor. Many theories on the origins of language begin with two basic premises concerning the vocal behavior of nonhuman primates, especially the great apes. They assume that (1) apes (and other primates) can exercise only negligible volitional control over the production of sound with their vocal tract, and (2) they are unable to learn novel vocal behaviors beyond their species-typical repertoire (e.g., Arbib et al. 2008; Burling 1993; Call and Tomasello 2007; Corballis 2002; Hauser 1996; Levinson and Holler 2014; Pinker 1994; Pollick and De Waal 2007; Premack 2004). These assumptions are often perpetuated through reference to a few notable reports. For example, Goodall, generalizing from her observations of free-ranging chimpanzees in the Gombe Reserve, came to the conclusion that "the production of sound in the absence of the appropriate emotional state seems to be an almost impossible task for a chimpanzee" (1986: 125). Another prominent example comes from the spoken shortcomings of the human-fostered chimpanzee Viki (Hayes 1951). Despite years of immersion in spoken language, intensive training, and her considerable motivation, Viki was reported to never have succeeded in producing more than four poorly articulated words.

Yet, since these influential studies, accumulating evidence suggests that apes have more capacity for voluntary control and learning of vocal behavior than previously realized[1] (Owren et al. 2011). Here we present a unique case study to this growing body of evidence, describing the repertoire of learned vocal and breathing-related behaviors used by the human-fostered western lowland gorilla (*Gorilla gorilla gorilla*), Koko.

Vocal control and learning in primates

The production of sound through the vocal tract involves the control and coordination of three main components, including the pulmonic system (the diaphragm and lungs), the larynx, and supralaryngeal articulators like the tongue and lips. In technical usage, the term *vocal* refers only to sounds produced by vibration of the vocal folds, while phrases like non-vocal sound production are used for sounds produced through the vocal tract that lack vocal fold vibration (e.g., whistling). Less audible non-vocal behaviors (e.g., blowing out a candle) have received little direct study and are not distinguished by specific terminology. Such distinctions have often been conflated in theorizing about primate vocal and breath control, and thus, in their review of nonhuman primate vocal production, Owren et al. (2011) emphasized that each of these components must be considered in an assessment of vocal ability. Reflecting this point, in this paper, we examine all kinds of learned behaviors that involve control over breathing and the vocal tract, while also making distinctions between the different components of control. The common denominator in each of the behaviors of focus here is the control over egressive airflow in coordination with modification of the laryngeal and/or supralaryngeal portions of the vocal tract, regardless of whether the production of sound is a salient characteristic of the behavior.

Vocal control

Researchers have historically argued that primate vocalizations are entirely involuntary, sub-cortical behaviors, driven by specific emotional and environmental stimuli (Fitch 2006; Lancaster 1968; Myers 1976; Schusterman 2008). A strong and originally influential formulation of this position, put forth by Skinner (1957: 463), described primate vocalizations as "innate responses" that "comprise reflex systems which are difficult, if not impossible, to modify by operant reinforcement." However, a number of

operant conditioning studies on primates run counter to this strong claim, and successful vocal conditioning has been achieved across a wide range of primate taxa, including great apes (Pierce 1985). A typical conditioning study is exemplified by Sutton et al. (1973), who trained rhesus macaques to vocalize in response to a white light trial signal for food rewards. Over the course of training, the monkeys learned to control the intensity and duration of a species-typical *coo* vocalization. Overall, these empirical results make it clear that, given an adequate number of trials and time, primates can at least acquire the ability to voluntarily "gate" production of species typical vocalizations (Fitch 2010).

Further evidence of flexible vocal behavior in primates comes from their tactical use and suppression of calls. For instance, vervet monkeys suppress predator-specific alarm calls when no conspecifics are present (Cheney and Seyfarth 1985). Several examples of this type of flexibility have also been observed in great apes, often referred to as audience effects. One study reported the flexible use of pant grunt vocal greetings by free-ranging female chimpanzees in the Sonso community of Budongo Forest, Uganda (Laporte and Zuberbühler 2010). These vocalizations are typically used by lower-ranking chimpanzees with more dominant individuals. However, females take into account the surrounding audience when producing the greeting, and they are less likely to pant grunt to another individual in the presence of the alpha male and female. Another study examined the probability of calling by wild male chimpanzees in response to the playback of the panthoot calls of an unfamiliar male (Wilson et al. 2001). The probability of a vocal response increased with the number of allied males present, suggesting the possibility that chimpanzees can suppress their vocal behavior when it is tactically advantageous to do so, such as when they lack superior numbers in the presence of a stranger. In another case of tactical suppression, Goodall (1986) described how male chimpanzees at the Gombe Reserve suppress their usual calls as they conduct territorial patrols.

Such findings suggest that apes can exert some degree of voluntary control over their vocal production. However, it is important to distinguish the ability to *suppress* calls in the face of an eliciting stimulus from the ability to voluntary *produce* calls independent of context. Recent work provides some evidence that wild chimpanzees' alarm calls may be voluntarily produced. Experimenters placed a plastic snake along paths commonly travelled by chimpanzees of the Sonso community and then measured the frequency of calls by individuals that encountered the snake (Crockford et al. 2012). While factors like arousal, personal risk, and risk to a high-ranking group-member did affect the number of calls produced, the likelihood of an individual producing any calls at all was only predicted by

---

[1] We limit our review of the literature to primates, but there is of course extensive and highly relevant research on vocal abilities in non-primate species too, including cetaceans, pinnipeds, and birds, among others (e.g., Janik and Slater 1997; Marler 2000; Schusterman 2008).

whether its companions had already seen the snake or had been in hearing range of previous alarm calls. Individuals who had just heard another chimpanzee's alarm almost never produced a call of their own upon seeing the snake, but took care to alter their path to avoid it. Given that these chimpanzees produced calls in less than half of snake encounters, and almost never in response to others' calls, it may be more parsimonious to describe the calls as voluntary productions, rather than describing their absence as voluntary suppression. This conclusion is supported by a similar study that tested the intentionality of chimpanzee alarm calls according to several metrics (Schel et al. 2013). Researchers again observed the responses of chimpanzees to a toy snake, and they noted whether alarm calls were socially directed at particular individuals, whether they were associated with visual monitoring of that recipient, and when the calls ceased. The study found that chimpanzees began alarm calling when another individual neared the snake toy, that the calls were associated with gaze alternation between the snake toy and the recipient, and that they ceased when the recipient withdrew from the snake. These three factors argue strongly for an interpretation of alarm calls as intentionally produced signals, rather than gateable responses to an evoking stimulus.

Voluntary calls are also documented in several studies of chimpanzees in captivity that deploy their vocalizations as an attention-getting tactic with human interactants. A set of studies found that captive chimpanzees housed at the Yerkes National Primate Research Center were more likely to use manual gestures to communicate with a visually attentive experimenter, but were more inclined to produce vocalizations and non-vocal sounds when the experimenter was inattentive or facing away (Hopkins et al. 2007; Leavens et al. 2004, 2010). For example, chimpanzees tactically employed vocalizations in the service of gaining the attention of an inattentive human interlocutor with access to food. Considering why this behavior is observed in captive subjects, but not in the wild, Leavens and colleagues reasoned that the production of sound as an attention getter is an effective solution to the adaptive space of captivity, as chimpanzees communicate with humans to obtain food that they cannot physically access.

Altogether these various findings suggest that primates, and especially great apes, can wield at least some degree of control over their vocalizations. A range of primates have been trained to control their vocal behavior under operant conditioning. Chimpanzees in the wild show a tendency to produce vocalizations like alarm calls and greeting grunts only when it is socially advantageous or sensible to do so, suppressing them if not. And in captivity, chimpanzees use vocalizations and non-vocal sounds to get the attention of human partners when their visual attention is lacking. But

do primates also show the ability to *learn* new vocalizations in addition to producing them voluntarily?

Vocal learning

Vocal learning refers both to cases when an animal is able to modify its species typical calls based on experientially guided learning (e.g., hearing the calls of conspecifics), and also to the learning of entirely novel vocal behaviors. One source of evidence for learning in the modification of species typical calls comes from studies of call "dialects," such as the convergence of pant hoots within different populations of chimpanzees (Crockford et al. 2004; Marshall et al. 1999; Mitani and Gros-Louis 1998). These studies found that pant hoots are more similar within populations than between, and rule out factors of heritability and environment. Another study found that captive chimpanzees introduced into a new group showed convergence in the structure of their referential food grunts (Watson et al. 2015). Notably, their calls converged only after they became socially integrated with the new group and established strong affiliative relationships. Together these studies demonstrate that individuals are able to modify some species-typical calls from hearing the calls of others, and further, that call convergence may facilitate social bonding.

Studies of wild orangutans have found region-specific traditions involving novel vocalizations and oral sounds that are thought to be outside of the species' repertoire. Van Schaik et al. (2003, 2006) reported regional variations in the production of *raspberries*, which involve buzzing or sputtering of the lips, and *kiss-squeaks*, which are oral ingressive sounds, not generated by airflow controlled by the lungs. Raspberries are commonly used by orangutans during their nest-building routine, but only by particular populations and not by others. Ruling out ecological factors to explain the regional differences, the authors concluded that it is a novel, socially transmitted behavior, rather than a species typical one. Moreover, even among populations that produce the raspberry, there is regional variation in its precise context of use, with orangutans of one region emitting the sound during the final phase of nest building, and those in a different region emitting the sound just before its start.

Several studies of captive chimpanzees (described above) reported their use of novel vocalizations and non-vocal sounds as voluntarily controlled attention getters with human interactants (Hopkins, et al. 2007; Leavens et al. 2004, 2010). These learned sounds included an unvoiced *raspberry* and also a voiced *extended grunt*, described as a low, loud, guttural sound made with the mouth open (also see Krauss and Fouts 1997, which noted the use of a raspberry as an attention getter by a sign-taught chimpanzee).

With human enculturation, apes show evidence that they are able to learn a more varied repertoire of novel vocalizations. One case is the classic project of Hayes and Hayes with the young female chimpanzee Viki (Hayes 1951). In what may be the most rigorous attempt to teach an ape to speak, the Hayes raised Viki to every extent possible as if she was their human child, beginning at just 3 days of age. The project is often summarized by the conclusion that despite all of her intensive training, Viki never succeeded in pronouncing more than four spoken words: *mama*, *papa*, *cup*, and *up*. According to Pinker (1994: 343), "Viki [was] at a disadvantage … forced to use [her] vocal apparatus, which was not designed for speech and which [she] could not voluntarily control."

However, such a conclusion assesses Viki's vocal abilities against a standard of learning to "speak," rather than considering more general abilities for vocal learning and exercising voluntary control over her vocal apparatus. In these latter respects, Viki's accomplishments are noteworthy (also see informal but comparable findings with orangutans: Furness 1916; Laidler 1980). Originally, through operant conditioning of emotional food barks and anxious *oo oo*s, Viki eventually developed a novel *ahhhh* sound, which she produced frequently across varied contexts, with clear volition. With further conditioning, she learned to coordinate *ahhhh* with the bilabial lip shape until she could articulate the word *mama* at will. Once acquiring the initial skill to vocalize with volitional control, she was able to learn new speech sounds and words more easily by imitation. By copying the model of her parents, Viki learned to produce successive /p/s for a whispered *papa* and added the sounds *tsk* and /k/. She eventually learned to combine /k/ and /p/ into *cup*. Over time, Viki also learned to blow air through her vibrating lips, produce additional sounds transcribed as "blook" and "boo", and blow spit bubbles. Thus although Viki's vocal abilities never amounted to anything resembling the order of complexity involved in human speech, she was nevertheless able to articulate at her volition a variety of sounds with her vocal apparatus. A major limitation to note, however, was Viki's relative lack of laryngeal fluency, as her vowels resembled a hoarse whisper lacking any periodic vibration of the vocal folds.

Another notable case is Kanzi, a bonobo raised from infancy with human enculturation and productive and receptive immersion in symbolic communication (Hopkins and Savage-Rumbaugh 1991; Taglialatela et al. 2003). The reports described Kanzi's use of four novel vocalizations that varied consistently in structure from the typical set of bonobo *peeps*, including qualities of their duration and pitch contour. Each of the four vocalizations was used consistently in distinct contexts and appeared to correspond to semantic usage similar to "banana," "grape," "juice," and "yes."

In two other cases related to enculturation and learning from humans, apes have learned to perform novel behaviors involving breath control in the production of non-vocal sounds. One case is Bonnie, a zoo-housed orangutan who learned to produce, at will, a human whistle without explicit training (Wich et al. 2009). In imitation experiments in which a human experimenter whistled for either a long or a short duration, either once or twice, Bonnie tended to match this pattern. She thus exerted remarkable control over her breathing in coordination with articulating her lips in the proper configuration for a resonating whistle.

Finally, the gorilla Koko, the subject of the current study, has been documented to exert volitional control over her breathing during her play with musical wind instruments (Perlman et al. 2012). Koko commonly plays with instruments like recorders, harmonicas, and party-favor whistles, often without any extrinsic request or reward. While her play would not be considered musical, she exhibits considerable pulmonic fluency stringing together varying numbers of toots of varying durations. Moreover, analysis showed that she uses a faster breathing cycle of inhalations and exhalations during her instrument play compared to her normal breathing. Another study noted Koko's volitional exhalation in a learned imitative eyeglasses-cleaning routine (Perlman and Gibbs 2013). In this routine, Koko huffs through her open mouth onto the lenses of eyeglasses and then wipes them off with a tissue (an act that she appears to pantomime in an event described by Perlman and Gibbs).

Taken together, the research described above shows that ape vocal abilities are not as limited as often assumed. Apes exercise the ability, under some circumstances, to voluntarily control their vocal and non-vocal pulmonic sound production, to modulate species typical vocalizations, and to learn novel vocal and breathing-related behaviors. Moreover, these abilities appear to be enhanced in apes raised within favorable environments, particularly in enculturated settings when such behaviors are more likely to be motivated, as they are modeled by humans and functionally useful with human caregivers. The present study extends research on ape vocal behavior by examining a video corpus of the gorilla Koko's daily interactions with human caregivers and documenting in detail each of her learned vocal and breathing behaviors.

Methods

Subject

The subject of the study is Koko, a female western lowland gorilla (*Gorilla gorilla gorilla*), 37–39 years of age during the reported observations. Koko was born at the San Francisco Zoo in 1971, but became ill at 6 months of age

and was moved from the zoo's gorilla enclosure to be cared for by humans and nursed back to health. Following her recovery, at 1 year of age, she came under the care of Dr. Francine Patterson (FP) and Dr. Ronald Cohn (RC) as the subject of Patterson's doctoral research at Stanford University (Patterson and Linden 1981). From this time until the present day, Koko has been largely immersed in human social interaction and culture. This includes, from the start, many years of intensive instruction in a symbolic communication system with gestures derived from American Sign Language, as well as regular exposure to spoken English. Thus Koko has lived in an environment with profound human socialization and enculturation since the age of 6 months.

Data recording

The data come from video recordings made of regular daily interactions between Koko, FP, and RC, recorded by either FP or RC. In a few instances, other people are present but play a peripheral role in the interaction. The present study focuses on video recordings of Koko from July 2007 through December 2010, which constitutes roughly 71 h of video. The majority of recordings were made with a Canon PowerShot S5 digital camera, but some recordings in 2007 were made with a Canon PowerShot S3.

Coding and analysis

The first author watched the recordings and identified all apparent instances of Koko performing novel vocal and breathing-related behaviors that varied from species-typical in both their manner of production (cf. Salmi et al. 2013) and the context in which they were performed (e.g., directed into a telephone or combined with particular learned manual gestures). In general, the behaviors appeared to be performed intentionally, but they were not necessarily communicative, nor were they necessarily associated with a salient sound. For convenience, hereon we refer to such activities as "vocal and breathing behaviors" (VBBs).

After their initial selection, candidate VBBs were confirmed by a more conservative process. If Koko's production was clearly audible (i.e., evident by listening without viewing), the VBB was counted even when background noise prevented it from being clearly distinguishable in a spectrogram. If the sound was not clearly perceptible in the audio, then visual verification through a spectrogram was required. Instances that could not be confirmed by either audio or spectrogram were omitted from further analysis.

All confirmed VBBs were then coded by the first author according to the procedure described below. To assess reliability, the second author also independently coded a random subset of 50 % of the bouts in the dataset for the

'sounds like' cue to manner and place of articulation, voicing, orofacial features, manual behaviors, initiation, and consequence (see below for explanations of these variables).

VBBs were first classified according to highly distinctive, visible characteristics of articulation and co-occurring gestures and action routines. These characteristics established well-defined categories for each behavior and largely corresponded to previously established labels that were used by FP and RC in interactions with Koko. Some cases matched similar human activities like talking on the telephone or huffing on eyeglasses to clean them.

The behaviors were then coded according to standard articulatory terminology from phonology, including manner (frication or plosive) and place (labial, glottal, nasal, or lingual–labial) of articulation and voicing (voiced or unvoiced). The initial survey of Koko's VBBs established these descriptive categories, and then a two-part procedure was conducted to more rigorously classify each behavior by the articulatory parameters (see Table 1). First, the coder listened to the behaviors, without the visual aid of video. Based on a holistic impression of the sound (the "sounds like" cue), an initial classification was made with respect to four places of articulation (glottal, labial, lingual–labial, and nasal) and two manners of articulation (frication, plosive). Second, specific acoustic and visual cues including orofacial features were coded to independently confirm each classification. When a definitive classification could not be made (e.g., information of the behavior was obscured by noise or viewing angle), then the behavior was classified according to the available information. Instances with conflicting cues were discarded from further analysis. Classifications made without full confirmation are noted in the report of the results.

A definitive classification required the converging evidence from the 'sounds like' cue plus the presence of sufficient additional cues to discriminate the behavior from other possible alternatives. Additional cues included features of the spectrogram and visible features of the face and articulators. Labial and lingual–labial sounds were distinguished from glottal and nasal sounds by visible puffing of the cheeks, created from pressure built up in the mouth. Labial sounds were further distinguished from lingual–labial sounds by the visible rounding of Koko's lips (labial) or by the visible protrusion of her folded tongue (lingual–labial). If cheek puffing was visibly lacking, the sound was determined to be produced either through constriction in the glottis or through the nose. Manually formed nasal constriction was confirmed by the visual observation of Koko manually applying pressure to her nose, either directly with her fingers or with a tissue. As she often produced this sound by moving her fingers around (sometimes applying a tissue) and modulating the pressure on her

Table 1 Prototypical cues for phonetic categories

Articulatory category	Core cues	Supplementary cues
Labial fricative	Sounds like	Frication (sg)
	Cheek puff	
	Pursed/protruded lips	
Lingual–labial fricative	Sounds like	Frication (sg)
	Cheek puff	
	Visible folded tongue	
Glottal fricative	Sounds like	
	No cheek puff or open jaw	
	Smooth frication onset (sg)	
Glottal plosive	Sounds like	
	No cheek puff or open jaw	
	Stop burst (sg)	
Nasal fricative	Sounds like	No cheek puff
	Manual pressure	
	Modulated frication (sg)	
Voicing	Audible pitch	
	Detectable pitch (sg)	

sg cues visible on spectrogram

nose, the second and decisive cue for nasal frication was visible modulation of intensity in the spectrogram that appeared to correspond to the manual manipulation of her nose.

If the features of cheek puffing and manual nose modulation were lacking, then it was determined that the audible turbulence was produced by air passing through the glottis, either through frication or from the release of a plosive. Frication was characterized by a gradual onset and then offset of energy visible in the waveform and spectrogram, whereas a plosive was indicated visibly by an initial burst of energy. Glottal fricatives were further distinguished as being open or closed mouth, and in the case that there was no evidence of cheek puffing or manual nose modulation, an open mouth was considered confirmation of glottal articulation.

Vocal fold vibration was measured first by audible detection and coded with a score of 0 1, or 2, with 0 indicating no audible vocal fold vibration, 1 moderate or unclear vibration, and 2 clearly audible vibration. All cases of 1 and 2 s were then viewed in a spectrogram to determine if the vibration was sufficiently periodic to be detected by the Praat pitch tracker (within the range of 30–45 Hz).

Our preliminary observations determined that VBBs were often performed along with particular manual behaviors or as part of more complex action routines, usually in social interaction with a human caregiver. Thus, in addition to articulatory parameters, we also analyzed the full behavioral context in which each VBB was performed. This included coding the co-performance of particular intransitive manual gestures (i.e., performed without an object), transitive manual gestures (i.e., performed on an object), as well as more complex action routines that characteristically incorporated an object. The particular kinds of objects that were incorporated into the behavior were also noted.

To examine Koko's motivation for performing the behaviors, the VBBs were coded for the context in which they were initiated, and also for their consequences in terms of social response and reward. Because Koko often produced multiple VBBs in quick succession, these factors were considered with respect to bouts, which were defined as successive VBBs occurring within an interval of 8.5 s (see "Results" section for justification of this criterion).

A minimum of 30 s of video preceding the first VBB of a bout was required to determine how it was initiated. Bouts were coded as externally initiated when FP or RC directly asked Koko to perform that, or a similar, behavior, including when it was named or demonstrated without a direct request. Self-initiated behaviors were performed without any request, mention, or demonstration.

Finally, bouts were coded for their social consequence and reinforcement. For a given bout, any consequence taking place from its beginning to 30 s after its completion was counted toward the whole bout. Overlapping bouts were treated as a single extended bout. Possible consequences included reinforcement by food or a desired nonfood object (other), or a verbal response. This latter category included any verbal response to Koko's behavior, from explicit praise to a conversational-style response to simple acknowledgement. In many cases, human caregivers provided multiple forms of reinforcement—these were coded as the most rewarding form (food/other > verbal > none).

Results

Ten VBBs that were identified from the initial audiovisual inspection of video could not be confirmed by independent identification based on audio only or through visual inspection of the spectrogram, and these were excluded from further analysis. The remaining instances comprised a total of 439 individual exhalations. These were each categorized into one of nine main types according to the defining criteria (see Table 2), including *blow/huff with intransitive gesture, blow/huff with transitive gesture, raspberry, cough, blow nose, talk on phone, clean glasses, play instrument*, and *other instrumental blows*. See Supplementary Information for 18 video examples spanning the range of Koko's behaviors. Supplementary Table S1 presents basic information about the type of behavior in each video clip.

Timing of VBBs and bouts

We examined the timing of successive VBBs and how they patterned together into bouts by measuring the intervals of every instance in which a VBB was followed by another in the same video clip. There were 346 such intervals, ranging from <1 to 486 s. 304 of these intervals were between VBBs of the same type. Intervals between the same type (mean = 7.7 s, median = 2.0, sd = 32.3) were much

shorter than those between different types (mean = 59.1 s, median = 22.5, sd = 97.0). Due to the heavy-right-tailed distributions of interval lengths, this difference was tested by Mann–Whitney–Wilcoxon test, which demonstrated it to be highly significant, $U = 1017.00$, $n_{\mathrm{same}} = 304$, $n_{\mathrm{diff}} = 42$, $p < .001$. This result supports our intuition that VBBs of the same type are more likely to appear in a bout together. Therefore, we examined the distribution of same-type intervals and found that 88 % of intervals were 8 s or less, no intervals were between 8 and 10 s, and the remaining 12 % of intervals were >10 s (mean = 50 s, sd = 83). Based on this natural split, we set 8.5 s as the maximum interval between VBBs to consider them part of the same bout. This yielded a total of 161 bouts, of which 154 were composed exclusively of a single type of behavior and 7 of a mix of behaviors. Table 2 describes each behavior type and displays counts of bouts and number of exhalations for each.

Initiation and consequence

The full results of initiation and consequence of VBB bouts are presented in Table 3. The second coding resulted in 92 % agreement on the initiation of bouts ($\kappa = .871$) and 97 % agreement on their consequences ($\kappa = .953$). The majority of bouts (71 %) were self-initiated rather than cued by FP or RC. With respect to consequence, 54 % of

Table 2 Behavior counts and descriptions

Behavior type	Bouts	Exhalations	Mean (E/B)	SD (E/B)	Description
Blow/huff with intransitive gesture	35	74	2.1	1.4	Produces blow or huff-like sound in combination with a manual gesture, not directed at an object
Blow/huff with transitive gesture	20	38	1.9	1.2	Produces blow or huff-like sound in combination with an object-directed manual gesture
Raspberry	20	24	1.2	0.4	Produces sound in combination with folding the tongue and protruding it through the lips
Cough	17	29	1.7	0.7	Produces a burst-like release of air in combination with a manual gesture toward the mouth
Blow nose	13	19	1.5	0.8	Blows air through the nasal passage in combination with manually applied pressure on the nose
Talk on phone	16	68	4.3	3.5	Produces sound into phone-like object, characterized by posture in which object is held to the right ear and mouth with left hand and the fold of the right elbow
Clean glasses	12	23	1.9	1.4	Huff-like behavior performed on the lenses of glasses, with subsequent manual "cleaning" of the frames (usually with a tissue)
Play instrument	29	148	5.1	2.6	Produces sound by blowing into an instrument, most often a recorder or harmonica, but sometimes other objects like bottles and pen caps, with the intention to make a sound
Other instrumental blow	7	16	2.3	1.8	Other instances when Koko blows into things for other instrumental purposes (e.g., blowing at FP through a paper towel tube or attempting to blow up a balloon)
Total	161	439	2.7	2.3	

Total bouts does not match the sum of bouts by behavior type because 7 bouts included 2 or more types. E/B is the number of exhalations per bout

bouts received a verbal response only, 35 % were rewarded with food or some other desired object, and 10 % had no apparent social consequence.

Articulatory features

Play instrument and *other instrumental blows* were excluded from the analysis because their articulation depended on the shape of the instrument. Five VBBs were additionally excluded because of conflicting cues. In total, articulatory features were analyzed for 275 VBBs, which are presented in Table 4. The second coding resulted in 99 % agreement of the 'sounds like' cue to manner and place of articulation ($\kappa = .990$). This very high level of inter-coder reliability reflects the highly distinctive acoustic nature of the sounds as classified by their place and manner of articulation.

In total, Koko produced VBBs with detectable voicing on 61 occasions. The second coding of voicing resulted in 95 % agreement ($\kappa = .866$). Voicing occurred most frequently with *talk on phone* (61.8 %), whereas in contrast, it never occurred with *clean glasses*. Koko also produced voicing more frequently when performing *blow/huff with transitive gesture* (36.8 %) compared to *blow/huff with intransitive gesture* (4.1 %). A Chi-square test showed this to be a reliable difference $\chi^2(1, \ n = 111) = 22.20$, $p < .001$.

Table 4 also shows the orofacial features associated with each behavior. The second coding of these features resulted in 82 % agreement ($\kappa = .771$). Orofacial features were largely constrained by supralaryngeal features of articulation, but we point to one noteworthy pattern in which Koko differentially opened her mouth according to the context of behavior. When cleaning glasses, she without exception huffed with a wide-open mouth, yet when huffing into the phone, her mouth was open in only 19 of 34 (55.9 %) visible exhalations. This distribution significantly differed from an independent distribution of mouth shapes across behavior types, $p < .001$ by Fisher's Exact Test.

Coordination with gestures and action routines

Six VBBs were excluded from the analysis because they were directed at an object being held for Koko by FP. Table 5 shows the breakdown of each VBB articulation type by specific gestures and action routines. The second coding resulted in 95 % agreement on these manual behaviors ($\kappa = .936$). Almost all VBBs (96 %) were produced in coordination with some kind of manual behavior or routine, including intransitive gestures (27 %), transitive gestures (33 %), and more complex routines with objects (25 %). Glottal fricatives were performed almost exclusively with objects (96 %, excluding instances when FP held the object), either directing the VBB at a whole object (e.g., a doll, a baseball card) or the tip of an object (e.g., a pen, the antenna tip of a walkie talkie). Glottal fricatives directed toward object tips typically involved longer objects and may have originated from play with a microphone. Nasal fricatives were also performed most frequently on objects (68 %), specifically tissues.

The wide array of objects incorporated into Koko's VBBs is cataloged in Table 6. Simple transitive gestures tended to incorporate a mix of object types. Routines, on the other hand, tended to incorporate the same type of object. For example, the talk on phone routine was used almost exclusively with telephones and walkie talkies, with just a single exception in which a harmonica was played

Table 3 Initiation and consequence of behaviors

Behavior type	Initiation			Consequence		
	External	Self	Undetermined	None	Verbal only	Verbal + food/other
Blow/huff with intransitive gesture	3	20	4	0 (0)	18 (4)	9
Blow/huff with transitive gesture	5	5	4	0 (0)	9 (1)	5
Raspberry	5	8	4	0 (0)	3 (2)	14
Cough	7	5	1	0 (0)	5 (0)	8
Blow nose	1	4	1	1 (0)	3 (1)	2
Talk on phone	2	8	3	2 (0)	11 (3)	0
Clean glasses	1	9	2	6 (1)	4 (0)	2
Play instrument	3	4	11	2 (0)	14 (5)	2
Other instrumental blow	0	3	2	2 (0)	1 (0)	2
Total	27	66	32	13 (1)	68 (16)	44

Initiation and consequence were determined by observing the preceding and following 30 s from each bout, respectively. In some cases, another bout occurred before the end of this 30 s window. These cases were treated as extended bouts. The "undetermined" column for initiation refers to cases where a bout began too early in a clip to determine its initiation. Numbers in parentheses for consequence refer to cases in which the clip ended before the end of the 30 s window. Some of these cases may have had more rewarding consequences than reported here

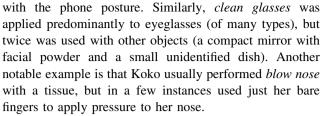**Table 4** Articulatory features of vocal and breathing-related behaviors

Behavior	Count	Supralaryngeal articulation						Voicing			Orofacial features				
		Labial fricative	Lingual–labial fricative	Glottal plosive	Glottal fricative	Nasal Frication	Undetermined	None	Intermediate	Full	Open mouth	Neutral/closed mouth	Rounded lips	Protruded tongue	Undetermined
Blow/huff with intransitive gesture	74	67	0	0	4	0	3	71	3	0	2	6	49	0	17
Blow/huff with transitive gesture	38	13	0	0	24	0	1	24	9	5	18	0	14	0	6
Raspberry	24	0	22	0	1	0	1	22	2	0	0	0	0	23	1
Cough	29	0	0	29	0	0	0	29	0	0	11	3	0	0	15
Blow nose	19	0	0	0	0	19	0	19	0	0	1	2	0	0	16
Talk on phone	68	0	0	0	68	0	0	26	21	21	19	15	0	0	34
Clean glasses	23	0	0	0	23	0	0	23	0	0	23	0	0	0	0

The numbers refer to counts of individual exhalations. Nasal frication was achieved with manual pressure on the nostrils

Table 5 Coordination of manual behavior and VBBs

Articulation type	No manual behavior	Intransitive gestures						Transitive gestures and action routines						
		Open hand	Perpendicular hands	Overlapping hands	Rub nose	Miscellaneous gestures	Total	Hold object to mouth	At object	At object tip	Phone routine posture	Glasses cleaning routine	Tissue to nose	Total
Voiced glottal fricative	3	1	0	0	0	1	2	0	0	12	42	0	0	54
Voiceless glottal fricative	4	1	0	0	0	1	2	0	0	6	26	23	0	55
Glottal plosive	0	26	0	0	0	0	26	0	0	0	0	0	3	3
Labial fricative	0	47	13	5	0	2	67	0	12	1	0	0	0	13
Lingual–labial fricative	10	7	0	0	0	5	12	0	0	0	0	0	0	0
Nasal fricative	0	0	0	0	6	0	6	0	0	0	0	0	13	13
Play instrument	0	0	0	0	0	0	0	144	0	0	2	0	0	146
Other instrumental blow	0	0	0	0	0	0	0	16	0	0	0	0	0	16

with the phone posture. Similarly, *clean glasses* was applied predominantly to eyeglasses (of many types), but twice was used with other objects (a compact mirror with facial powder and a small unidentified dish). Another notable example is that Koko usually performed *blow nose* with a tissue, but in a few instances used just her bare fingers to apply pressure to her nose.

In comparison with glottal and nasal fricatives, glottal plosives, labial fricatives, and lingual–labial fricatives were performed more frequently with intransitive gestures. The most common gesture was *open hand,* in which Koko brought her flat palm to her mouth in synchrony with exhalation. She also sometimes used other gestures like *perpendicular hands*, bringing her two palms together and her thumbs to her mouth, and *overlapping hands*, placing one hand on top of the other and bringing her bottom palm to her mouth.

Finally, to summarize these results, we constructed profiles of each VBB type based on the analyses of temporal structure, social context, articulatory characteristics, and associated manual behaviors described above.

VBB profiles

Blow/huff with intransitive gesture

Koko tended to articulate blow/huffs with intransitive gestures as voiceless labial fricatives (i.e., a "blow"). The behavior was associated with the visible orofacial feature of rounded lips as part of producing a labial fricative. They were commonly produced in bouts with multiple exhalations and often combined with an open hand gesture, but sometimes with other intransitive gestures. The behaviors were most often initiated by Koko and resulted in a verbal response and sometimes food.

Blow/huff with transitive gesture

Koko produced both glottal and labial fricatives with transitive gestures, often with rounded lips for both. Glottal fricatives were often voiced and characteristically directed toward the tips of objects. Labial fricatives were not voiced and directed toward the center of the object. Both were commonly produced in bouts of multiple exhalations. Koko performed these VBBs most often in response to external motivation and often elicited a verbal response and sometimes received food as a result.

Raspberry

Koko's version of a "raspberry" (as it is called by FP and RC) is articulated as a lingual–labial fricative. She combined raspberries with a manual intransitive gesture about half of the time. However, they were the only VBB that

Table 6 Objects incorporated with VBB articulation types

Articulation	Objects used
Voiced glottal fricative	Phone (25), walkie talkie (22), gorilla doll (4), pen (4), none (3), bottle (1)
Voiceless glottal fricative	Glasses (22), phone (15), walkie talkie (14), pen (3), cable (3), none (2), paper towel tube (1)
Glottal plosive	None (26)
Play instrument	Recorder (118), harmonica (21), whistle (9)
Other instrumental blow	Paper towel tube (9), balloon (3), bottle (3), pen cap (1)
Labial fricative	None (67), doll (6), card (4), balloon (1), makeup compact (1), paper towel tube (1)
Lingual–labial fricative	None (22)
Nasal fricative	Tissue (13), none (6)

Number in parentheses is the number of uses. Excludes cases when FP held objects for Koko

Koko commonly performed without any kind of coordinated manual behavior, and they also contained on average the fewest number of exhalations per bout—characteristics that might result from its relatively difficult articulation. Koko's production of raspberries most frequently resulted in food (nuts in particular) compared to other behaviors, and in practice, it was often interpreted as a vocal signal used by Koko to request a nut.

Cough

Koko articulated coughs as a voiceless glottal plosive, typically in combination with an open hand gesture (i.e., like someone coughing and covering their mouth). They were usually performed with an open mouth, but on a few occasions with a closed or neutral mouth position. They were often produced in response to external motivation, such as requests by FP and RC to perform the behavior.

Blow nose

Koko performed blow nose by blowing air through her nostrils while pressing on her nose, usually with a tissue, but sometimes just by applying pressure directly with her fingers. The behavior appears similar to the human act of nose blowing. Indeed, while Koko sometimes performed the behavior on request, she also appeared to blow her nose as a pragmatic means to clear congestion from her nasal cavity.

Talk on phone

Koko performed talk on phone with a characteristic posture in which she tucked a phone or other suitable object in the crook of her elbow and held it to her ear and face. Koko exclusively produced glottal fricatives into the phone, often with voicing. Bouts of the behavior typically were performed with a large number of exhalations, second only to play instrument. These could be produced with an open mouth or with a closed/neutral mouth. Koko's performance of talk on phone often resulted in a verbal response by FP and RC, but never in food.

Clean glasses

Koko performed clean glasses by huffing on the lenses of eyeglasses and wiping them manually with a tissue. In an interesting contrast to the talk on phone routine, she produced clean glasses as voiceless glottal fricatives, articulated with a wide-open mouth. Koko usually performed clean glasses with eyeglasses but in a few instances with other objects having flat surfaces like a compact mirror. Koko's performance of clean glasses often resulted in no social consequence, indicating that she sometimes may have performed the routine for her own amusement.

Play instrument and other instrumental blows

Koko's play with instruments differed from other VBBs in that sound production involved an external source, most often an instrument such as a recorder or harmonica, but sometimes other sources like a glass bottle or paper towel roll. The behavior was more variable in form compared to the other object-directed routines as it was largely dictated by the constraints imposed by the different instrument shapes. Koko's play with instruments showed the highest average number of exhalations per bout, sometimes up to 10 or more. Her instrument play usually resulted in a verbal response from FP and RC, sometimes food, and on a few occasions, she performed the activity without any social attention.

Discussion

We described the repertoire of novel vocal and breathing behaviors (VBBs) performed by the enculturated gorilla Koko, as documented in video recordings from 2007 to 2010 when she was 37–39 years of age. Aside from the few

reports of efforts to teach chimpanzees and orangutans to speak (Furness 1916; Kellogg and Kellogg 1933; Hayes 1951; Laidler 1980), this is one of the first studies to examine the influence of intensive human rearing in the development of an ape's vocal abilities (see also Hopkins and Savage-Rumbaugh 1991; Perlman et al. 2012; Taglialatela et al. 2003). It is the first detailed description of learned vocal and breathing-related behavior in a gorilla and documents one of the most extensive repertoires recorded for any nonhuman primate.

Drawing on approximately 71 h of footage of Koko, we identified 161 bouts of VBBs comprising 439 individual exhalations. Based on salient characteristics, the VBBs were classified as one of nine distinctive behaviors or routines: *blow/huff with transitive gesture*, *blow/huff with intransitive gesture*, *raspberry*, *cough*, *blow nose*, *talk on phone*, *clean glasses*, *play instrument*, and *other instrumental blows*. The vast majority of VBBs were produced in characteristic combination with manual behaviors, including intransitive or transitive gestures (i.e., performed on objects) and also more complex routines with objects. To our knowledge, nothing similar to these behaviors has ever been reported as part of the species-typical repertoire of gorillas or other apes.

Koko appears to perform these behaviors with a large degree of volitional control. She initiated their production in the majority of instances and flexibly varied the number of exhalations she produced for any given bout. Koko's voluntary control over her vocal apparatus is also suggested by the sheer variety of behavior, both with respect to the different types of VBBs in her repertoire, and also their combination with various gestures, routines, and objects. Her production of such a large number of combinations indicates a degree of flexibility that seems unlikely if the behaviors were involuntary reflexes to environmental stimuli.

An issue of particular interest is Koko's control over vocalization, since previous research has suggested that apes have less volitional control over their larynx compared to their breath and supralaryngeal articulators (Fitch 2010). While the majority of Koko's VBBs were unvoiced, she does appear to wield some voluntary control over her larynx. For example, she often produced voiced huffs in her phone routine compared to voiceless huffs with eyeglasses.[2] Skeptics may contend that the voicing we observed might instead be due to the vibration of some other part of the vocal tract anatomy, perhaps velar tissue as in snoring. While we cannot completely rule out this

possibility, we point to Koko's production of voiceless huffs and glottal stops as further evidence of voluntary laryngeal control. She is able to open her glottis wide, or alternatively, close it shut.

Articulatory features and multimodal combination

We found that Koko's vocal and breathing-related behaviors could be aptly described in terms of articulatory features used in the phonological description of human speech, including place (lingual–labial, labial, glottal, nasal) and manner of articulation (frication, plosive) and voicing (unvoiced, voiced). Furthermore, our results show how these features serve in functional contrasts as they vary consistently with respect to particular co-occurring manual movements and routines. Koko's use of voiceless huffs with eyeglasses compared to voiced huffs with the phone is one such example. Her use of glottal frication with objects and labial frication with intransitive gestures is another.

Lieberman (1969: 173) once remarked that, "If apes did communicate by means of cries that were differentiated by phonologic feature contrasts that were a subset of phonologic features available to man, we would see a link between human language and nonhuman primate behavior." Indeed, Koko's repertoire reveals the emergence of articulatory contrasts along a subset of phonological features that are found in human languages. While she does not string these features together in any way that resembles speech, Koko demonstrates the potential to use a rudimentary system of contrastive sounds that she produces with her vocal tract. The question of whether and how these behaviors serve specific communicative functions is a direction for future research. As we have described, they are typically performed within a social, interactive context, and often appear to be expressive. However, our impression is that it is unlikely that Koko's production of articulatory contrasts—either with her vocal tract or with combined gestures—convey anything like the kinds of semantic distinctions made by symbolic languages.

Our findings further show that Koko's VBBs should not be considered as unimodal behaviors that she performs with her vocal tract alone. Instead they seem to be understood most accurately as integrated components of multimodal behaviors. Koko regularly produces VBBs in coordination with intransitive and transitive manual gestures, and within more complicated behavioral routines that incorporate characteristic types of objects. Her flexible use of VBBs within these multimodal complexes reveals a system of behavior that could potentially support combinatorial expansion in a communication system. The rotation of different objects in her transitive gestures further expands this potential.

---

[2] Her production of voicing specifically when performing *talk on phone* and when vocalizing into an object tip, often referred to as her *microphone*, hints at the possibility that Koko is doing her "imitation" of human speech.

However, in addition to these multimodal behaviors, we note a previous report of Koko that described two VBBs not recorded in our video corpus (Perlman et al. 2012). One of these—*the blow test*—is a gentle ritual Koko performs when greeting visitors through the mesh of her enclosure. She leans forward toward her visitor and blows gently toward their face as an invitation to blow back so she can smell their breath. Koko performs the other VBB—called *you blew it*—when she is especially agitated at someone. She fills her lungs with air and blows forcefully at the transgressor (through mesh). Notably, neither of these behaviors involves a distinctive manual component. Thus Koko may tend to incorporate VBBs into multimodal complexes of behavior, and yet, her voluntary control over her breathing and vocal tract does exhibit some independence from manual behavior. This independence is also reflected in Koko's productions of some raspberries without any accompanying manual behavior.

Implications for language evolution

Scholars of language evolution have often assumed that apes are severely limited in their ability to voluntarily control their vocalizing and breathing and that they are unable to learn new behaviors beyond their species-typical repertoire. These assumptions are shared by theorists who otherwise disagree fundamentally about the phylogenetic history (and indeed the very nature) of human language. Those inclined toward a vocal origin of language (considered primarily in the form of speech) often dismiss the relevance of great ape vocal behavior because they believe it to be entirely involuntarily and reflexive. They posit evolutionary scenarios in which the human language capacity evolved essentially de novo in the *Homo* lineage, distinctly after our divergence from the great apes (e.g., Bickerton 1990; Pinker and Bloom 1990). Alternatively, theorists positing a gestural origin of language often emphasize apes' supposed lack of flexible learning and control in the vocal modality as evidence against a vocal origin theory. In comparison, they argue that apes' flexible learning and control over their manual gestures constitutes a major piece of evidence for the argument that symbolic communication arose first in the form of gesture (Arbib 2012; Call and Tomasello 2007; Corballis 2002; Hewes 1973).

The present study adds to accumulating findings that invalidate these strong claims about negligible vocal abilities in apes. Koko, for one, has acquired an impressive repertoire of vocal and breathing behaviors and, in performing them, exhibits considerable dexterity and voluntary control in coordinating her exhalation with movements of her glottis and supralaryngeal articulators. Moreover, as we have detailed, Koko produces these behaviors as components of multimodal complexes of behavior. These findings fit with the view that human language use is a profoundly multimodal activity that combines speech, gesture, facial expression, and other bodily postures within a tightly integrated and synchronized system of expressive movements (Birdwhistell 1970; Kendon 2004; McNeill 1992). As evidence suggests that multiple modalities function together interdependently in modern language, a plausible hypothesis for its evolution is that multiple modalities also evolved together interdependently (Kendon 2009; McNeill 2012). Some ape gesture researchers are similarly coming to the conclusion that the evolution of language is best understood from a multimodal perspective (Hopkins et al. 2007; Leavens 2003; Taglialatela et al. 2011; Waller et al. 2013). For example, Leavens (2003) proposed that, "Because visual and vocal communication seem to be functionally linked in extant apes, language may have been multimodal from its inception" (p. 233).

Although Koko's ability to perform novel vocal and breathing-related behaviors are in line with similar abilities documented in other apes, her combination of these behaviors with gestures appears somewhat unique. Thus caution is warranted in generalizing from the multimodality of Koko's behavior. She obviously acquired these patterns of behavior from her unusual life experience of immersive interaction with humans in an environment enriched with communication with conventional gestures—including extensive teaching of gestures derived from American Sign Language—as well as with speech and other human vocal behavior. What Koko shows, however, is that gorillas have the potential for coordinating learned vocal and breathing behavior with gesture, and likely they share this capacity with chimpanzees, bonobos, and orangutans.

We point to one example in orangutans that illustrates how this potential might be adaptive for apes who are not, like Koko, raised amongst humans interested in teaching them to talk. Above we described how orangutans in certain wild populations produce a culturally transmitted kiss squeak (Van Schaik et al. 2003). Further study has found that some individuals produce the sound while holding a stripped leaf to their mouth, which functions to decrease the maximum frequency of the sound (Hardus et al. 2009). These modified calls tend to be produced by smaller individuals in high distress, suggesting that they produce this culturally learned variant to sound bigger and ward off predators. This is only a single case, but it shows how the ape ability to learn and control behaviors combining sound production through the vocal tract with manual actions can be adaptive in certain contexts. Decades of field studies have surely only scratched the surface of culturally transmitted VBBs that have cycled in and out of ape populations across their millions of years of history.

Conclusion

Koko—a 43-year-old gorilla immersed in human interaction since the age of 6 months—offers a unique case study for scholars interested in the evolution of language and speech. Like Viki decades before, Koko has not learned to speak like a human, but this does not leave her bound by a reflexive, involuntary vocal system. Koko, Viki, Kanzi, Bonnie, and increasingly many other apes show us that they are able to exercise volitional control over their vocalization and sound production and can even learn to produce new vocal and breathing-related behaviors. The size of Koko's repertoire may be unusual, but it presumably developed from an entirely ordinary capacity of gorillas and other apes that was fostered under extraordinary environmental circumstances.

References

Arbib MA (2012) How the brain got language: the mirror system hypothesis. Oxford University Press, Oxford

Arbib MA, Liebal K, Pika S (2008) Primate vocalization, gesture and the evolution of human language. Curr Anthropol 49:1053–1076

Bickerton D (1990) Language and species. University of Chicago Press, Chicago

Birdwhistell R (1970) Kinesics and context: essays on body motion communication. University of Pennsylvania Press, Philadelphia

Burling R (1993) Primate calls, human language, and nonverbal communication. Curr Anthropol 34:25–53

Call J, Tomasello M (2007) The gestural communication of apes and monkeys. Taylor and Francis, New York

Cheney DL, Seyfarth RM (1985) Vervet monkey alarm calls: manipulation through shared information? Behaviour 94:150–166

Corballis M (2002) From hand to mouth: the origins of language. Princeton University Press, Princeton

Crockford C, Herbinger I, Vigilant L, Boesch C (2004) Wild chimpanzees produce group-specific calls: a case for vocal learning? Ethology 110:221–243

Crockford C, Wittig RM, Mundry R, Zuberbühler K (2012) Wild chimpanzees inform ignorant group members of danger. Curr Biol 22:142–146

Fitch WT (2006) Production of vocalizations in mammals. In: Brown K (ed) Encyclopedia of language and linguistics. Oxford University Press, Oxford, pp 115–121

Fitch WT (2010) The evolution of language. Cambridge University Press, Cambridge

Furness WH (1916) Observations on the mentality of chimpanzees and orang-utans. Proc Am Philos Soc 55:281–290

Goodall J (1986) The chimpanzees of Gombe: patterns of behavior. Harvard University Press, Cambridge

Hardus ME, Lameira AR, Van Schaik CP, Wich SA (2009) Tool use in wild orang-utans modifies sound production: a functionally deceptive innovation? Philos Trans R Soc B 276:3689–3694

Hauser MD (1996) The evolution of communication. MIT Press, Cambridge

Hayes C (1951) The ape in our house. Harper, New York

Hewes GW (1973) Primate communication and the gestural origins of language. Curr Anthropol 14:5–24

Hopkins WD, Savage-Rumbaugh S (1991) Vocal communication as a function of differential rearing experiences in *Pan paniscus*: a preliminary report. Int J Primatol 12:559–583

Hopkins WD, Taglialatela JP, Leavens DA (2007) Chimpanzees differentially produce novel vocalizations to capture the attention of a human. Anim Behav 73:281–286

Janik VM, Slater PJB (1997) Vocal learning in mammals. Adv Study Behav 26:59–99

Kellogg WN, Kellogg LA (1933) The ape and the child: a study of environmental influence upon early behavior. Whittelsey House, Oxford

Kendon A (2004) Gesture: visible action as utterance. Cambridge University Press, Cambridge

Kendon A (2009) Language's matrix. Gesture 9:352–372

Krauss MA, Fouts RS (1997) Chimpanzee (*Pan troglodytes*) pointing: hand shapes, accuracy, and the role of eye gaze. J Comp Psychol 11:330–336

Laidler K (1980) The talking ape. Stein and Day, New York

Lancaster JB (1968) Primate communication systems and the emergence of human language. In: Jay PC (ed) Primates: studies in adaptation and variability. Holt, Rinehart & Winston, New York, pp 439–457

Laporte MNC, Zuberbühler K (2010) Vocal greeting behavior in wild chimpanzee females. Anim Behav 80:467–473

Leavens DA (2003) Integration of visual and vocal communication: evidence for Miocene origins. Behav Brain Sci 26:232–233

Leavens DA, Hostetter AB, Wesley MJ, Hopkins WD (2004) Tactical use of unimodal and bimodal communication by chimpanzees, *Pan troglodytes*. Anim Behav 80:467–476

Leavens DA, Russell JL, Hopkins WD (2010) Multimodal communication by captive chimpanzees (*Pan troglodytes*). Anim Cogn 13:33–40

Levinson SC, Holler J (2014) The origin of human multi-modal communication. Philos Trans R Soc B 369:20130302. doi:10.1098/rstb.2013.0302

Lieberman P (1969) On the acoustic analysis of primate vocalizations. Behav Res Methods Instrum 1:169–174

Marler P (2000) Origins of music and speech: insights from animals. In: Wallin NL, Merker B, Brown S (eds) The origins of music. MIT Press, Cambridge, pp 31–48

Marshall AJ, Wrangham RW, Arcadi AC (1999) Does learning affect the structure of vocalizations in chimpanzees? Anim Behav 58:825–830

McNeill D (1992) Hand and mind. University of Chicago Press, Chicago

McNeill D (2012) How language began: gesture and speech in human evolution. Cambridge University Press, Cambridge

Mitani JC, Gros-Louis J (1998) Chorusing and call convergence in chimpanzees: tests of three hypotheses. Behaviour 135:1041–1064

Myers RE (1976) Comparative neurology of vocalization and speech: proof of a dichotomy. Ann N Y Acad Sci 280:745–757

Owren MJ, Amoss RT, Rendall D (2011) Two organizing principles of vocal production: implications for nonhuman and human primates. Am J Primatol 73:530–544

Patterson FG, Linden E (1981) The education of Koko. Holt, Rinehart &Winston, New York

Perlman MS, Gibbs RW Jr (2013) Pantomimic gestures reveal the sensorimotor imagery of a human-fostered gorilla. J Ment Imaging 37:73–96

Perlman MS, Patterson FG, Cohn RH (2012) The human-fostered gorilla Koko shows breath control in play with wind instruments. Biolinguistics 6:433–444

Pierce JD (1985) A review of attempts to condition operantly alloprimate vocalizations. Primates 26:202–213

Pinker S (1994) The language instinct. William Morrow & Co., New York

Pinker S, Bloom P (1990) Natural language and natural selection. Behav Brain Sci 13:707–727

Pollick AS, de Waal FBM (2007) Ape gestures and language evolution. Proc Natl Acad Sci 104:8184–8189

Premack D (2004) Is language the key to human intelligence? Science 303:318–320

Salmi R, Hammerschmidt K, Doran-Sheehy DM (2013) Western gorilla vocal repertoire and contextual use of vocalization. Ethology 119:831–847

Schel AM, Townsend SW, Machanda Z, Zuberbühler K, Slocombe K (2013) Chimpanzee alarm call production meets key criteria for intentionality. PLoS One 8:e76674. doi:10.1371/journal.pone.0076674

Schusterman RJ (2008) Vocal learning in mammals with a special emphasis on pinnipeds. In: Oller DK, Gribel U (eds) The evolution of communicative flexibility: complexity, creativity, and adaptability in human and animal communication. MIT Press, Cambridge, pp 41–70

Skinner BF (1957) Verbal behavior. Appleton-Century-Crofts, New York

Sutton D, Larson C, Taylor EM, Lindeman RC (1973) Vocalization in rhesus monkeys: conditionability. Brain Res 52:225–231

Taglialatela JP, Savage-Rumbaugh S, Baker LA (2003) Vocal production by a language-competent *Pan paniscus*. Int J Primatol 24:1–17

Taglialatela JP, Russell JL, Schaeffer JA, Hopkins WD (2011) Chimpanzee vocal signaling points to a multimodal origin of human language. PLoS One 6:e18852. doi:10.1371/journal.pone.0018852

van Schaik CP, Ancrenaz M, Borgen G, Galdikas B, Knott CD, Singleton I, Suzuki A, Utami SS, Merrill M (2003) Orangutan cultures and the evolution of material culture. Science 3:102–105

van Schaik CP, van Noordwijk MA, Wich SA (2006) Innovation in wild Bornean orangutans (*Pongo pygmaeus wurmbii*). Behaviour 143:839–876

Waller B, Liebal K, Burrows A, Slocombe K (2013) How can a multimodal approach to primate communication help us understand the evolution of communication? Evol Psychol 11:538–549

Watson SK, Townsend SW, Schel AM, Wilke C, Wallace EK, Cheng L, West V, Slocombe KE (2015) Vocal learning in the functionally referential food grunts of chimpanzees. Curr Biol 25:495–499

Wich SA, Swartz KB, Hardus ME, Lameira AR, Stromberg E, Shumaker RW (2009) A case of spontaneous acquisition of a human sound by an orangutan. Primates 50:56–64

Wilson ML, Hauser MD, Wrangham RW (2001) Does participation in intergroup conflict depend on numerical assessment, range location, or rank for wild chimpanzees? Anim Behav 61:1203–1216