ESSAY

Orangutan *Instrumental Gesture-Calls*: Reconciling Acoustic and Gestural Speech Evolution Models

Adriano R. Lameira · Madeleine E. Hardus ·
Serge A. Wich

Call control allows an organism to produce an acoustic signal irrespective of its own underlying emotional state. It is thus a prerequisite to "higher" abilities, such as call imitation, innovation and the use of arbitrary or deceptive calls, and therefore to speech. However, among primates, call control is presumed to be greatly confined to humans (Seyfarth and Cheney 2008). Consequently, there is little agreement about its evolutionary precursors (Christiansen and Kirby 2003). Essentially two major models and lines of evidence have been proposed; speech evolved (1) as an extension of acoustic communication in non-human primates (e.g. Seyfarth et al. 1980; Slocombe and Zuberbühler 2005; Arnold and Zuberbühler 2006; Wich et al. 2009) or (2) from non-human primate gestural communication (e.g. Rizzolatti and Arbib 1998; Corballis 2003; Arbib Michael et al. 2008). These models have been seen as mutually exclusive or as sequential accounts in which calls replace gestures (Brown et al. 1999), however, both face limitations concerning the emergence of call control in our evolutionary lineage. Did call control derive from an

A. R. Lameira (✉)
Behavioural Biology Group, Utrecht University, Utrecht,
The Netherlands
e-mail: a.reiselameira@uu.nl

M. E. Hardus
Institute for Biodiversity and Ecosystem Dynamics,
University of Amsterdam, Amsterdam, The Netherlands

S. A. Wich
Anthropological Institute and Museum, University of Zurich,
Zurich, Switzerland

S. A. Wich
Sumatran Orangutan Conservation Program (PanEco/YEL),
Medan, Indonesia

essentially emotional call use, or from an *essentially* voluntary gesture use, as that of non-human primates? The acoustic model needs to explain how a fundamentally close-ended acoustic system became open-ended (i.e. with limitless number of elements; alike speech). The gestural model needs to clarify the behaviors and respective functional advantages that allowed a shift (or "translation") from an open-ended gestural system to an open-ended acoustic system.

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

Recently we have described (Hardus et al. 2009a) how and why wild orangutans use gestures to functionally alter the acoustic characteristics of a particular sound (sensu Lameira et al. 2010) emitted under disturbing contexts, the kiss squeak (Hardus et al. 2009b). By positioning a hand or holding leaves in front of their lips, wild orangutans lower the maximum frequency (i.e. that of highest dB) but maintain other parameters of the call similar. Evidence suggests that kiss squeaks are under voluntary motor control in orangutans, and when individuals produce these

modified variants of the call, they sound as if their body size is bigger than it actually is, reinforcing this impression on a potential predator and potentially deterring it through functional deception.

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

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

We hypothesise that IGC, dating back to the hominid-pongid split (9–13 m.y.a.; Hobolth et al. 2011) may have provided the direct functional and neural sensory-motor basis towards call control in an early human ancestor essentially lacking this ability, that is, they served as an exaptation for this ability. IGC are remarkable in that they bring into close temporal, motivational, contextual, anatomical and functional association both the gestural and oro-laryngeal systems of motor control in the communication domain. Hand-assisted feeding, for instance, raises the same associations between gestural and oro-laryngeal systems of motor control but in the foraging domain. IGC comprise therefore, obligatorily, the expression of synchronous activations of multiple neural sensory-motor systems in the ape brain. In the ape cerebral cortex, such activations will mainly occur within regions homologous to

the cortical homunculus (that comprises the primary motor cortex, which plays a crucial role in general voluntary motor control) and between the cortical homunculus and other cortical systems involved in the domain of communication, such as those homologous to Broca's and Wernicke's areas (Taglialatela et al. 2011). Such synchronous activations may have provided a neural interface between the brain areas activated, through functional integration and clustering (Tononi et al. 1998a, b) enabling the sharing of abilities which were previously fundamentally restricted or segregated to particular areas. By means of cortical and neural plasticity (Lieberman 2002a), alike for example, use-dependent functional reorganization of sensory cortices (Pantev et al. 1998), this interface would have set the basis for the establishment of enhanced and more resilient short and long distance circuits. Indeed, cortical and neural plasticity is at the basis of hemispheric asymmetries in key areas of the ape and human brain for communicative signaling (Hopkins and Nir 2010; Perani et al. 2011).

As the focus of voluntary control, the cortical homunculus would represent the main stage for these circuit modifications. The number of areas activated in this area and their mutual proximity would add up to form a momentary local hotspot of activations sufficient to ignite neighbouring areas over which there was previously little voluntary control. Namely, circuitry between the respiration, hand, face, lips, and tongue (somatotopic) locations would expand to include that of larynx areas. These circuits would not necessarily be required to be established de novo, but instead, would only be required to modestly build and expand on previously existing ones. For instance, a rudimentary but functionally relevant interface between hand, respiration and laryngeal locations (and possibly lips and tongue) is already present in the ape brain, in that use of the right hand for gestures is significantly enhanced when the gestures are accompanied by a call (Hopkins and Cantero 2003). At the same time, pathways between the primary motor cortex and nucleus ambiguous (site of the laryngeal motor-neurons in medulla oblongata), which are specifically interpreted as representing a crucial neural step in gaining call control (Fitch 2005; Brown et al. 2008), are found in apes but not in monkeys (Kuypers 1958), substantiating the view that an rudimentary interface is already present between systems.

In humans, neuroimaging studies support this evolutionary scenario. For instance, the (somatotopic) location of larynx/phonation area (that with control over intrinsic musculature of the larynx, underlying adduction/abduction and tensing/relaxing of the vocal folds) in the cortical homunculus is adjacent to the lips area and the expiratory area (Brown et al. 2008). This means that in humans, phonation, articulation and respiration are neurologically conjunct. Considering that orangutans have been experimentally demonstrated to exert apt voluntary motor control

over lips and respiration (Wich et al. 2009; Lameira et al. in review), it is reasonable to view this conjunction as evolutionarily relevant in humans. While laryngeal musculature may operate in complex ways during (online) speech and other functions (Jürgens 2002; Ludlow 2005), the evolutionary genesis of call control theoretically commenced when the first rudimentary neural signal initiating in the primary motor cortex would be transmitted successfully simply to set the larynx into position during airflow. The view that neural circuitry flexibility could have successfully achieved this in our ancestors is supported by a phenomenon known in human as motor equivalence, where speakers develop different motor strategies, i.e., use different musculatures, of the larynx to achieve the same voice outcome (Ludlow 2005). Accordingly, IGC could potentially explain why the area of representation of the intrinsic laryngeal muscles has seemingly migrated toward the labial area in humans (Brown et al. 2008). In addition, IGC are in concordance with the increasing literature corroborating that gestures and calls/speech are neurally co-processed (e.g. Rizzolatti and Arbib 1998; Bernardis and Gentilucci 2006; Xu et al. 2009).

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

IGC present a parsimonious route to human-like neurophysiology, increased call control and repertoire size in the earliest stages of speech evolution, but one may question its relevance based on the phylogenetic distance between orangutans and humans. Three clarifications are required. Firstly, comparison between human, chimpanzee and orangutan genomes shows that some regions of the human genome more closely resemble orangutan's (Hobolth et al. 2011). Although this percentage is approximately 1%, a necessarily bigger percentage is equally similar between humans, chimpanzees and orangutans. While broad genetic underpinnings of speech are not well understood beyond FoxP2 gene (e.g. Enard et al. 2002), the relevance of genetic proximity within hominoids remains equivocal. Secondly, speech is a bio-cultural evolutionary phenomenon (Richerson and Boyd 2005), and therefore, theories must encompass some degree of interaction between social and genetic mechanisms in the acquisition and transmission of communication signals. Orangutans and chimpanzees are the only apes to show extensive cultures in the wild (e.g. Whiten et al. 1999; van Schaik et al. 2003), thus, *both*

species represent promising models. Thirdly, the description of IGC in orangutans but (so far) not in chimpanzees may constitute a methodological artifact. While cultural variants between populations have been investigated in wild chimpanzees, this record tends to focus on feeding behavior (Watson and Caldwell 2009). Oppositely, researchers have investigated geographical variation in orangutans' complete call repertoire (Hardus et al. 2009b). These conditions may have benefited the description of IGC more readily than in chimpanzees. There are nonetheless anecdotes suggesting that IGC may be part of their repertoire, such as the use of a hand in front of the mouth to muffle a call, as described by Jane Goodall (Deacon 1998).

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

References

Arbib Michael, A., Liebal, K., & Pika, S. (2008). Primate vocalization, gesture, and the evolution of human language. *Current Anthropology, 49*(6), 1053–1076.

Arcadi, A., Robert, D., & Boesch, C. (1998). Buttress drumming by wild chimpanzees: Temporal patterning, phrase integration into loud calls, and preliminary evidence for individual distinctiveness. *Primates, 39*(4), 505–518.

Arnold, K., & Zuberbühler, K. (2006). Semantic combinations in primate calls. *Nature, 441*(7091), 303.

Bernardis, P., & Gentilucci, M. (2006). Speech and gesture share the same communication system. *Neuropsychologia, 44*(2), 178–190.

Brown, S. (1999). The "musilanguage" model of music evolution. In N. L. Wallin, B. Merker, & S. Brown (Eds.), *The origins of music* (pp. 271–300). Cambridge, MA: MIT Press.

Brown, S., Merker, B., & Wallin, N. L. (1999). An introduction to evolutionary musicology. In N. L. Wallin, B. Merker, & S. Brown (Eds.), *The origins of music* (pp. 3–24). Cambridge, MA: MIT Press.

Brown, S., Ngan, E., & Liotti, M. (2008). A larynx area in the human motor cortex. *Cerebral Cortex, 18*(4), 837–845.

Christiansen, M. H., & Kirby, S. (2003). Language evolution: Consensus and controversies. *Trends in Cognitive Sciences, 7*(7), 300–307.

Corballis, M. C. (2003). From mouth to hand: Gesture, speech, and the evolution of right-handedness. *Behavioral and Brain Sciences, 26*(02), 199–208.

Deacon, T. (1998). *The symbolic species: The co-evolution of language and the brain*. NY: Norton.

Enard, W., Przeworski, M., Fisher, S. E., Lai, C. S. L., Wiebe, V., Kitano, T., et al. (2002). Molecular evolution of FOXP2, a gene involved in speech and language. *Nature, 418*(6900), 869–872.

Falk, D. (2004). Prelinguistic evolution in early hominins: Whence motherese? *Behavioral and Brain Sciences, 27*(04), 491–503.

Fitch, W. T. (2005). Protomusic and protolanguage as alternatives to protosign. *Behavioral and Brain Sciences, 28*(02), 132–133.

Hardus, M. E., Lameira, A. R., van Schaik, C. P., & Wich, S. A. (2009a). Tool use in wild orang-utans modifies sound production: A functionally deceptive innovation? *Proceedings of the Royal Society B: Biological Sciences, 276*(1673), 3689–3694.

Hardus, M. E., Lameira, A. R., Singleton, I., Morrough-Bernard, H. C., Knott, C. D., Ancrenaz, M., et al. (2009b). A description of the orangutan's vocal and sound repertoire, with a focus on geographic variation. In S. Wich, T. Mitra Setia, S. S. Utami, & C. P. Schaik (Eds.), *Orangutans* (pp. 49–60). New York: Oxford University Press.

Hobolth, A., Dutheil, J. Y., Hawks, J., Schierup, M. H., & Mailund, T. (2011). Incomplete lineage sorting patterns among human, chimpanzee, and orangutan suggest recent orangutan speciation and widespread selection. *Genome Research, 21*, 349–356.

Hopkins, W. D., & Cantero, M. (2003). From hand to mouth in the evolution of language: The influence of vocal behavior on lateralized hand use in manual gestures by chimpanzees (Pan troglodytes). *Developmental Science, 6*(1), 55–61.

Hopkins, W. D., & Nir, T. M. (2010). Planum temporale surface area and grey matter asymmetries in chimpanzees (Pan troglodytes): The effect of handedness and comparison with findings in humans. *Behavioural Brain Research, 208*(2), 436–443.

Jürgens, U. (2002). Neural pathways underlying vocal control. *Neuroscience and Biobehavioral Reviews, 26*(2), 235–258.

Kuypers, M. G. J. M. (1958). Some projections from the peri-central cortex to the pons and lower brain stem in monkeys and chimpanzee. *Journal of Comparative Neurology, 110*, 211–255.

Lameira, A., Delgado, R., & Wich, S. (2010). Review of geographic variation in terrestrial mammalian acoustic signals: Human speech variation in a comparative perspective. *Journal of Evolutionary Psychology, 8*(4), 309–332.

Lameira, A. R., Hardus, M. E., Kowalsky, B., de Vries, H., Spruijt, B. M., Sterck, E. H. M., et al. in review. Orangutan whistling and implications for the emergence of an open-ended call repertoire: A replication and extension. *Journal of the Acoustical Society of America*.

Lieberman, P. (2002a). On the nature and evolution of the neural bases of human language. *American Journal of Physical Anthropology, 119*(S35), 36–62.

Lieberman, P. (2002b). *Human language and our reptilian brain*. Cambridge, Massachusetts and London, England: Harvard University Press.

Ludlow, C. L. (2005). Central nervous system control of the laryngeal muscles in humans. *Respiratory Physiology & Neurobiology, 147*(2–3), 205–222.

Maclarnon, A., & Hewitt, G. (2004). Increased breathing control: Another factor in the evolution of human language. *Evolutionary Anthropology: Issues, News and Reviews, 13*(5), 181–197.

MacNeilage, P. F. (1998). The frame/content theory of evolution of speech production. *Behavioral and Brain Sciences, 21*(04), 499–511.

Mithen, S. J. (2005). *The singing Neanderthals: The origins of music, language, mind and body* (pp. 97–112). London: Cambridge Journals Online.

Pantev, C., Oostenveld, R., Engelien, A., Ross, B., Roberts, L. E., & Hoke, M. (1998). Increased auditory cortical representation in musicians. *Nature, 392*(6678), 811–814.

Perani, D., Saccuman, M. C., Scifo, P., Awander, A., Spada, D., Baldoli, C., et al. (2011). Neural language networks at birth. *Proceedings of the National Academy of Sciences, 108*(38), 16056–16061.

Puglisi, A., Baronchelli, A., & Loreto, V. (2008). Cultural route to the emergence of linguistic categories. *Proceedings of the National Academy of Sciences, 105*(23), 7936–7940.

Remedios, R., Logothetis, N. K., & Kayser, C. (2009). Monkey drumming reveals common networks for perceiving vocal and nonvocal communication sounds. *Proceedings of the National Academy of Sciences, 106*(42), 18010–18015.

Richerson, P., & Boyd, R. (2005). *Not by genes alone: How culture transformed human evolution*. Chicago: University of Chicago Press.

Rizzolatti, G., & Arbib, M. A. (1998). Language within our grasp. *Trends in Neurosciences, 21*(5), 188–194.

Scott-Phillips, T. C., & Kirby, S. (2010). Language evolution in the laboratory. *Trends in Cognitive Sciences, 14*(9), 411–417.

Seyfarth, R., & Cheney, D. (2008). Primate social knowledge and the origins of language. *Mind & Society, 7*(1), 129–142.

Seyfarth, R. M., & Cheney, D. L. (2010). Production, usage, and comprehension in animal vocalizations. *Brain and Language, 115*(1), 92–100.

Seyfarth, R. M., Cheney, D. L., & Marler, P. (1980). Monkey responses to three different alarm calls—evidence of predator classification and semantic communication. *Science, 210*(4471), 801–803.

Slocombe, K. E., & Zuberbühler, K. (2005). Functionally referential communication in a chimpanzee. *Current Biology, 15*(19), 1779–1784.

Smith, K., Kirby, S., & Brighton, H. (2003). Iterated learning: A framework for the emergence of language. *Artificial Life, 9*(4), 371–386.

Taglialatela, J. P., Russell, J. L., Schaeffer, J. A., & Hopkins, W. D. (2011). Chimpanzee vocal signaling points to a multimodal origin of human language. *PLoS ONE, 6*(4), e18852.

Tononi, G., McIntosh, A. R., Russell, D. P., & Edelman, G. M. (1998a). Functional clustering: Identifying strongly interactive brain regions in neuroimaging data. *NeuroImage, 7*(2), 133–149.

Tononi, G., Edelman, G. M., & Sporns, O. (1998b). Complexity and coherency: Integrating information in the brain. *Trends in Cognitive Sciences, 2*(12), 474–484.

van Schaik, C. P., Ancrenaz, M., Borgen, G., Galdikas, B., Knott, C. D., Singleton, I., et al. (2003). Orangutan cultures and the evolution of material culture. *Science, 299*(5603), 102–105.

Watson, C., & Caldwell, C. (2009). Understanding behavioral traditions in primates: Are current experimental approaches too focused on food? *International Journal of Primatology, 30*(1), 143–167.

Whiten, A., Goodall, J., McGrew, W. C., Nishida, T., Reynolds, V., Sugiyama, Y., et al. (1999). Cultures in chimpanzees. *Nature, 399*(6737), 682–685.

Wich, S., Swartz, K., Hardus, M., Lameira, A., Stromberg, E., & Shumaker, R. (2009). A case of spontaneous acquisition of a human sound by an orangutan. *Primates, 50*(1), 56–64.

Xu, J., Gannon, P. J., Emmorey, K., Smith, J. F., & Braun, A. R. (2009). Symbolic gestures and spoken language are processed by a common neural system. *Proceedings of the National Academy of Sciences, 106*(49), 20664–20669.