their article. Our first reflection is on the assumption that independent control over the vocal folds and the upper vocal tract is somehow a given, and our second reflection is on the ability of apes to control vocalization voluntarily.

Following Fitch (2000a), the authors assume that animal vocalizations have a source and a filter. They also appear to assume that source and filter are independent as they are in modern humans, which is not necessarily the case. In many instances, the behavior of the source is in fact strongly coupled to that of the filter (e.g., in woodwind instruments). Source-filter theory was originally formulated in the context of human speech (Fant 1960). However, the fact that independence of source and filter is a good approximation for human speech does not mean it is universally valid.

Fletcher (1993) has investigated the theory of vibrating valves and found that the independence of source and filter depends on the precise shape and configuration of the source. In addition, it depends on the ratio of resonance frequencies of the source and the filter. Titze (2008) has adapted the theory to human-like vocal folds, and found that if the frequency at which the vocal folds vibrate is near the resonance frequencies of the vocal tract, strong coupling can occur. Apparently, modern human vocal folds and vocal tracts avoid strong coupling, but it is an open question whether this was the case in our evolutionary ancestors.

The little that we do know about ape vocal anatomy appears to argue against independence of source and filter. One instance of this is the large air sacs present in all great apes (Hewitt et al. 2002), which lower the resonance frequency of the upper vocal tract considerably (de Boer 2008) and would therefore increase coupling (as found in model experiments by Riede et al. 2008). In addition, chimpanzee vocal folds (the only ones about which we have anatomical data) have so-called vocal lips (Demolin & Delvaux 2006; Kelemen 1969), and thus a very different shape from human vocal folds. Although we do not know the function of these vocal lips, this difference between two closely related species underscores the point that we should not just assume similar behavior of their vocalization systems.

In systems where source and filter cannot behave independently, the set of signals that can be produced is necessarily more limited. This consequence is demonstrated in a modeling study showing that when source and filter are closely coupled, vocalization may be more chaotic, and thus it may be more difficult to time the onset of vocalization precisely (de Boer 2012). Given these observations, it may not just be a lack of neural control that makes precise vocalizations difficult for nonhuman primates. It may also be that the anatomy of their vocal folds and their vocal tracts makes it much harder as well.

Our second point of commentary is to note evidence of at least one case in which a nonhuman primate appears to have some voluntary control over her larynx in the performance of learned, species atypical vocalizations. Koko, a human-reared, female gorilla (Patterson & Linden 1981), has been video-recorded performing numerous instances from a repertoire of play behaviors involving voluntary control over her larynx and supralaryngeal vocal tract in coordination with various gestures and action routines (Perlman et al. 2011). This repertoire includes the production of breathy-voiced sounds and glottal stops in situations that are determined by the particular play routine.

Perlman et al. (2011) describe how Koko exhibits vocal control in her play behavior of "talking" into telephones, when she often directs breathy grunt-like vocalizations into the receiver, which she holds to her mouth (voicing was observed in 42 of 68 exhalations over 11 bouts). That she exercises voluntary control over her larynx in these vocalizations is suggested by the contrast of this behavior to her routine of huffing on the lenses of eyeglasses as if to clean them. As in the real human performance of cleaning eyeglasses, Koko produces, in this case, open-mouthed audible huffs that are distinctly and without exception voiceless (as exhibited in 12 video-recorded bouts involving 25 exhalations). Another dimension of vocal control is demonstrated in her voluntary performance of a mock "cough," which involves a glottal stop, often in coordination with a gesture in which she covers her mouth with an open hand. In several instances, she produces this behavior on command, demonstrating clear voluntary control over the closure of her glottis.

These behaviors appear to be examples of voluntary control over laryngeal motor activity outside of a species-typical audiovisual display, something that Ackermann et al. say has not been attested yet in great apes. Apparently we should not discount the possibility that apes – and by implication our last common ancestor – have more (rudimentary) abilities to control vocalization voluntarily than is often assumed.

Given that (1) control over vocalization is not just limited by neural factors, but also by purely anatomical and physiological ones, and that (2) a gorilla has been shown to have some rudimentary voluntary control over vocalization, we conclude that in the evolution of speech, anatomical and physiological adaptations to the vocal folds and the vocal tract may have been as important as neural adaptations of their control.

Very young infants' responses to human and nonhuman primate vocalizations

doi:10.1017/S0140525X13004019

Brock Ferguson, Danielle R. Perszyk, and Sandra R. Waxman
Psychology Department, Northwestern University, Evanston, IL 60208.
brock@u.northwestern.edu
danielleperszyk2017@u.northwestern.edu
s-waxman@northwestern.edu
http://www.psychology.northwestern.edu/people/faculty/core/profiles/sandra-waxman.html

Abstract: Recent evidence from very young human infants' responses to human and nonhuman primate vocalizations offers new insights – and brings new questions – to the forefront for those who seek to integrate primate-general and human-specific mechanisms of acoustic communication with theories of language acquisition.

In their target article, Ackermann et al. contribute to a long-standing debate concerning the extent to which the uniquely human propensity for language is the product of species-unique cognitive mechanisms (e.g., Hauser et al. 2002; Penn et al. 2008; Pinker & Bloom 1990). Their comprehensive analysis of neurological and behavioral evidence strengthens the proposal for evolutionary continuity in the mechanisms underlying acoustic communication in human and nonhuman primates. Our goal in this commentary is to amplify their proposal by highlighting recent behavioral evidence from human infants between 3 and 6 months of age. This evidence, which documents how infants respond to vocalizations of humans and nonhuman primates, bears on Ackermann et al.'s formidable challenge to consider the evidence of evolved acoustic communication architecture within the broader faculties of human language.

Recent studies have documented that even in infants too young to speak, listening to human speech supports core cognitive processes, including the formation of object categories (Ferry et al. 2010; Fulkerson & Waxman 2007; Waxman & Gelman 2009). Perhaps more surprisingly, this precocious link between human language and cognition is initially broad enough to include the vocalizations of nonhuman primates. For 3- and 4-month-olds, nonhuman primate vocalizations (from a blue-eyed Madagascar lemur) also promote object categorization, mirroring exactly the effects of human speech. However, by 6 months, lemur vocalizations no longer have this language-like effect: Instead, the link to categorization is tuned specifically to human language (Ferry et al. 2013). These findings reveal that a link between language and object categories, evident as early as 3 months in human infants, derives from a broader template that initially encompasses vocalizations of human and nonhuman primates, and is rapidly tuned specifically to human vocalizations (see also Vouloumanos et al. 2010).

This striking ontogenetic evidence has strong implications for theories of language acquisition. It also offers insights into Ackermann et al.'s proposal for integrating primate-general and human-specific mechanisms of acoustic communication. We focus here on three. First, the evidence from human infants is consistent with the Ackermann et al.'s proposal that, broadly speaking, the faculties that give rise to human language may be related to those predating *Homo sapiens* (see also Fitch 2011; Stoeger et al. 2012). What remains to be seen is how precisely the relations between homologous neural structures can be specified. For example, one promising investigation might be to ascertain whether infants' responses to human and nonhuman primate vocalizations engage the neural mechanisms described in the target article.

Second, the evidence from human infants converges with Ackermann et al.'s claim that human language acquisition may be built upon mechanisms that are specialized for acoustic communication. One must, however, consider the necessity of these acoustically-based mechanisms in human language acquisition. Although most humans acquire language in the aural-oral modality, our linguistic capacities are distinctly amodal. The signature of human language is not its perceptual form, but rather its ability to enable its users to express an infinite number of ideas using a discrete number of meaningful elements (Chomsky 1965). Thus, a complete account of the evolution of human language will be one that considers not only the acoustic-spoken modality but also the visual-manual modality in which deaf infants naturally acquire language. One question is whether, given the evidence for evolved neural hardware underpinning acoustic communication, infants acquiring spoken language might have some advantage. Evidence from infants acquiring sign language casts doubt on this possibility (e.g., Goldin-Meadow & Mylander 1983; Newport & Meier 1985; Petitto & Marentette 1991). More recent evidence from our lab underscores infants' flexibility in identifying language-like signals beyond human speech. If a novel signal (consisting of pure sine-wave tone sequences) is embedded within a social communicative exchange, infants endow the signal with communicative status and its effects mirror those of human speech in a subsequent categorization task (Ferguson & Waxman 2013).

Finally, evidence from infants can mutually constrain and inform developing theories of language evolution, acquisition, and usage. For example, we have recently discovered that unlike nonhuman primate vocalizations, zebra finch birdsong does not promote object categorization in human infants at any age (Perszyk & Waxman 2013). This outcome is consistent with claims that, although birdsong shares some structural features with human language, it lacks the links to meaning that characterize human language and, to a much lesser extent, nonhuman primate vocalizations (e.g., Berwick et al. 2013).

Ackerman et al.'s target article invites researchers across disciplines to engage in the larger enterprise of uncovering the origins of human language. Within this enterprise, the biggest leaps will be made by those who integrate seemingly disparate neurological, behavioral, and developmental evidence to unearth the evolutionary continuities and discontinuities in both modality-specific (e.g., vocalizations) and modality-independent capacities that provide humans alone with the capacity to acquire language.

ACKNOWLEDGMENTS

Portions of this research was supported by a SSHRC Doctoral Fellowship to Brock Ferguson, an NSF Graduate Research Fellowship to Danielle R. Perszyk, and a National Science Foundation grant to Sandra R. Waxman (BCS-0950376).

Functional neuroimaging of human vocalizations and affective speech

doi:10.1017/S0140525X13004020

Sascha Frühholz,[a,b] David Sander,[a,b] and Didier Grandjean[a,b]

[a]*Swiss Center for Affective Sciences, University of Geneva, 1211 Geneva, Switzerland;* [b]*Department of Psychology, University of Geneva, 1205 Geneva, Switzerland.*

Sascha.fruehholz@unige.ch David.sander@unige.ch
Didier.grandjean@unige.ch http://www.affective-sciences.org/user/286
http://cms.unige.ch/fapse/EmotionLab/Home.html
http://cms.unige.ch/fapse/neuroemo/

Abstract: Neuroimaging studies have verified the important integrative role of the basal ganglia during affective vocalizations. They, however, also point to additional regions supporting vocal monitoring, auditory–motor feedback processing, and online adjustments of vocal motor responses. For the case of affective vocalizations, we suggest partly extending the model to fully consider the link between primate-general and human-specific neural components.

Ackermann et al. provide a remarkable neural model of human vocalizations linking affective and motor brain systems underlying vocal communication. Recent neuroimaging studies on human affective vocalizations provide additional insights on this close link between the affective and motor component. Although human communication is mostly non-affective, the case of affective expressions provides an ideal paradigm to test the validity of the affective-motor model of human communication proposed by Ackermann et al.

Recent neuroimaging studies have specified the neural mechanisms underlying affective vocalizations (Aziz-Zadeh et al. 2010; Laukka et al. 2011; Wattendorf et al. 2013). These studies confirm the central role of the basal ganglia (BG) in vocalizations (Laukka et al. 2011; Pichon & Kell 2013), as proposed by Ackermann et al., and show the close connection between the ventromedial and dorsolateral striatum during emotional speech (Pichon & Kell 2013). They also support the notion of a close connection of the BG to the cortico-subcortical vocalization network (Laukka et al. 2011; Pichon & Kell 2013) as well as to the limbic system, which adds the emotional component of speech (Laukka et al. 2011; Péron et al. 2013; Wattendorf et al. 2013).

Although these studies support several of the main assumptions by Ackermann et al., they, first, also provide conflicting evidence for the suggested roles of some brain regions, and, second, suggest additional areas to be included in the neural network of vocalizations. Concerning the first point, Ackermann et al. propose, for example, that the anterior cingulate cortex (ACC) has no central role for prosodic vocal modulations, and that the inferior frontal cortex (IFC) is only involved in speech output behavior. Recent studies, however, indicate that the ACC plays a central role in the regulation of vocal behavior (Wattendorf et al. 2013), probably supporting the interaction between cognitive, physiological, and emotional-motivational states (Laukka et al. 2011) and serving as an auditory–motor interface between the perception and production of vocalizations (Aziz-Zadeh et al. 2010); see our Figure 1. Furthermore, the portion of the inferior frontal cortex (IFC) that lies rostral to the premotor cortex and Broca's area seems also to be involved in processing vocalizations, especially in the recognition and the generation of emotional intonated speech (Aziz-Zadeh et al. 2010; Frühholz & Grandjean 2013). Similar to the ACC, the IFC might thus act as an auditory–motor interface linking the perception and the production of emotional speech. This interface seems critical, because auditory–motor feedback loops are important for online adjustments of vocal behavior based on the forward and backward mapping of performance predictions (Rauschecker & Scott 2009). This is closely related to the second point.