BEHAVIORAL AND BRAIN SCIENCES (2014) **37**, 529–604
doi:10.1017/S0140525X13003099

Brain mechanisms of acoustic communication in humans and nonhuman primates: An evolutionary perspective

Hermann Ackermann
Neurophonetics Group, Centre for Neurology – General Neurology, Hertie Institute for Clinical Brain Research, University of Tuebingen, D-72076 Tuebingen, Germany
hermann.ackermann@uni-tuebingen.de
www.hih-tuebingen.de/neurophonetik

Steffen R. Hage
Neurobiology of Vocal Communication Research Group, Werner Reichardt Centre for Integrative Neuroscience, and Institute for Neurobiology, Department of Biology, University of Tuebingen, D-72076 Tuebingen, Germany
steffen.hage@uni-tuebingen.de
www.vocalcommunication.de

Wolfram Ziegler
Clinical Neuropsychology Research Group, City Hospital Munich-Bogenhausen, D-80992 Munich, and Institute of Phonetics and Speech Processing, Ludwig-Maximilians-University, D-80799 Munich, Germany.
wolfram.ziegler@extern.lrz-muenchen.de
www.ekn.mwn.de

Abstract: Any account of "what is special about the human brain" (Passingham 2008) must specify the neural basis of our unique ability to produce speech and delineate how these remarkable motor capabilities could have emerged in our hominin ancestors. Clinical data suggest that the basal ganglia provide a platform for the integration of primate-general mechanisms of acoustic communication with the faculty of articulate speech in humans. Furthermore, neurobiological and paleoanthropological data point at a two-stage model of the phylogenetic evolution of this crucial prerequisite of spoken language: (i) monosynaptic refinement of the projections of motor cortex to the brainstem nuclei that steer laryngeal muscles, presumably, as part of a "phylogenetic trend" associated with increasing brain size during hominin evolution; (ii) subsequent vocal-laryngeal elaboration of cortico-basal ganglia circuitries, driven by human-specific *FOXP2* mutations. This concept implies vocal continuity of spoken language evolution at the motor level, elucidating the deep entrenchment of articulate speech into a "nonverbal matrix" (Ingold 1994), which is not accounted for by gestural-origin theories. Moreover, it provides a solution to the question for the adaptive value of the "first word" (Bickerton 2009) since even the earliest and most simple verbal utterances must have increased the versatility of vocal displays afforded by the preceding elaboration of monosynaptic corticobulbar tracts, giving rise to enhanced social cooperation and prestige. At the ontogenetic level, the proposed model assumes age-dependent interactions between the basal ganglia and their cortical targets, similar to vocal learning in some songbirds. In this view, the emergence of articulate speech builds on the "renaissance" of an ancient organizational principle and, hence, may represent an example of "evolutionary tinkering" (Jacob 1977).

Keywords: articulate speech; basal ganglia; FOXP2; human evolution; speech acquisition; spoken language; striatum; vocal behavior; vocal learning

1. Introduction: Species-unique (verbal) and primate-general (nonverbal) aspects of human vocal behavior

1.1. Nonhuman primates: Speechlessness in the face of extensive vocal repertoires and elaborate oral-motor capabilities

All attempts to teach great apes spoken language have failed – even in our closest cousins, the chimpanzees (*Pan troglodytes*) and bonobos (*Pan paniscus*) (Hillix 2007; Wallman 1992), despite the fact that these species have "notoriously mobile lips and tongues, surely transcending the human condition" (Tuttle 2007, p. 21). As an example, the cross-fostered chimpanzee infant Viki mastered less than a handful of "words" even after extensive training. These utterances were not organized as speech-like vocal tract activities, but rather as orofacial manoeuvres imposed on a (voiceless) expiratory air stream (Hayes 1951,

 0140-525X/14 $40.00

p. 67; see Cohen 2010). By contrast, Viki was able to skillfully imitate manual and even orofacial movement sequences of her caretakers (Hayes & Hayes 1952) and learned, for example, to blow a whistle (Hayes 1951, pp. 77, 89).

Nonhuman primates are, nevertheless, equipped with rich vocal repertoires, related specifically to ongoing intra-group activities or environmental events (Cheney & Seyfarth 1990; 2007). Yet, their calls seem to be linked to different levels of arousal associated with especially urgent functions, such as escaping predators, surviving in fights, keeping contact with the group, and searching for food resources or mating opportunities (Call & Tomasello 2007; Manser et al. 2002; Seyfarth & Cheney 2003b; Tomasello 2008). Several studies point, indeed, at a more elaborate "cognitive load" to the vocalizations of monkeys and apes in terms of subtle audience effects (Wich & de Vries 2006), conceptual-semantic information (Zuberbühler 2000a; Zuberbühler et al. 1999), proto-syntactical call concatenations (Arnold & Zuberbühler 2006; Ouattara et al. 2009), conditionability (Aitken & Wilson 1979; Hage et al. 2013; Sutton et al. 1973; West & Larson 1995), and the capacity to use distinct calls interchangeably under different conditions (Hage et al. 2013). It remains, however, to be determined whether such communicative skills really represent precursors of higher-order cognitive–linguistic operations. In any case, the motor mechanisms of articulate speech appear to lack significant vocal antecedents within the primate lineage. This limitation of the faculty of acoustic communication is "particularly puzzling because [nonhuman primates] appear to have so many concepts that could, in principle, be articulated" (Cheney & Seyfarth 2005, p. 142). As a consequence, the manual and facial gestures rather than the vocal calls of our primate ancestors have been considered the vantage point of language evolution in our species (e.g., Corballis 2002, p. ix; 2003).

Tracing back to the 1960s, vocal tract morphology has been assumed to preclude production of "the full range of human speech sounds" (Lieberman 2006a; 2006b, p. 289) and, thereby, to constrain imitation of spoken language in nonhuman primates (Lieberman 1968; Lieberman et al. 1969). However, this model cannot account for the inability of nonhuman primates to produce even the most simple verbal utterances. The complete lack of verbal acoustic communication rather suggests more crucial *cerebral* limitations of vocal tract motor control (Boë et al. 2002; Clegg 2012; Fitch 2000a; 2000b). According to a more recent hypothesis, lip smacking – a rhythmic facial expression frequently observed in monkeys – might constitute a precursor of the dynamic organization of speech syllables (Ghazanfar et al. 2012; MacNeilage 1998). As an important evolutionary step, a phonation channel must have been added in order to render lip smacking an audible behavioral pattern (Ghazanfar et al. 2013). Hence, this theory calls for a neurophysiological model of how articulator movements were refined and, finally, integrated with equally refined laryngeal movements to create the complex motor skill underlying the production of speech.

1.2. Dual-pathway models of acoustic communication and the enigma of emotive speech prosody

The calls of nonhuman primates are mediated by a complex network of brainstem components, encompassing a midbrain "trigger structure," located in the periaqueductal gray (PAG) and adjacent tegmentum, and a pontine vocal pattern generator (Gruber-Dujardin 2010; Hage 2010a; 2010b). In addition to various subcortical limbic areas, the medial wall of the frontal lobes, namely, the cingulate vocalization region and adjacent neocortical areas, also projects to the PAG. This region, presumably, controls higher-order motor aspects of vocalization such as operant call conditioning (e.g., Trachy et al. 1981). By contrast, the acoustic implementation of the sound structure of spoken language is bound to a cerebral circuit including the ventrolateral/insular aspects of the language-dominant frontal lobe and the primary sensorimotor cortex, the basal ganglia, and cerebellar structures in either hemisphere (Ackermann & Riecker 2010a; Ackermann & Ziegler 2010; Ackermann et al. 2010). Given the virtually complete speechlessness of nonhuman primates, the behavioral analogues of acoustic mammalian communication might not be sought within the domain of spoken language, but rather in the nonverbal affective vocalizations of our species such as laughing, crying, or moaning (Owren et al. 2011). Against this background, two separate neuroanatomic "channels" with different phylogenetic histories appear to participate in human acoustic communication, supporting nonverbal affective vocalizations and articulate speech, respectively (the "dual-pathway model" of human acoustic communication; see Ackermann 2008; Owren et al. 2011; for an earlier formulation, see Myers 1976).

Human vocal expression of motivational states is not restricted to nonverbal affective displays, but deeply invades articulate speech. Thus, a speaker's arousal-related mood such as anger or joy shape the "tone" of spoken language (emotive/affective speech prosody). Along with nonverbal

Hermann Ackermann is Professor of Neurological Rehabilitation at the Centre for Neurology, Hertie Institute for Clinical Brain Research, University of Tuebingen. His research focuses on the cerebral basis of speech production and speech perception, and he is the author or coauthor of more than 120 publications within the domains of neuropsychology, neurolinguistics, and neurophonetics.

Steffen R. Hage is Head of the Neurobiology of Vocal Communication Research Group at the Werner Reichardt Centre for Integrative Neuroscience, University of Tuebingen. He is the author of more than 20 publications within the area of neuroscience, especially neurophysiology and neuroethology. His major research interests focus on audio-vocal integration as well as vocal-motor control mechanisms in acoustic communication of mammals, as well as cognitive processes involved in vocal behavior of nonhuman primates.

Wolfram Ziegler is Head of the Clinical Neuropsychology Research Group at the City Hospital Munich-Bogenhausen and Professor of Neurophonetics at the Ludwig-Maximilians- University of Munich. He is the author or co-author of more than 150 publications in peer-reviewed journals in the area of speech and language disorders.

affective vocalizations, emotive speech prosody has also be considered a behavioral trait homologous to the calls of nonhuman primates (Heilman et al. 2004; Jürgens 1986; 2002b; Jürgens & von Cramon 1982).[1] Moreover, one's attitude towards a person and one's appraisal of a topic have a significant impact on the "speech melody" of verbal utterances (attitudinal prosody). Often these implicit aspects of acoustic communication – *how* we say something – are more relevant to a listener than propositional content, that is, *what* we say (e.g., Wildgruber et al. 2006). The timber and intonational contour of a speaker's voice, the loudness fluctuations and the rhythmic structure of verbal utterances, including the variation of speaking rate and the local distinctness of articulation, represent the most salient acoustic correlates of affective and attitudinal prosody (Scherer 1986; Scherer et al. 2009; Sidtis & Van Lancker Sidtis 2003). Unlike the propositional content of the speech signal – which ultimately maps onto a digital code of discrete phonetic-linguistic categories – the prosodic modulation of verbal utterances conveys graded/analogue information on a speaker's motivational states and intentional composure (Burling 2005). Most importantly, activity of the same set of vocal tract muscles and a single speech wave simultaneously convey both the propositional and emotional contents of spoken language. Hence, two information sources seated in separate brain networks and creating fundamentally different data structures (analogue versus digital) contribute simultaneously to the formation of the speech signal. Therefore, the two channels must coordinate at some level of the central nervous system. Otherwise these two inputs would distort and corrupt each other. So far, dual-pathway models of human acoustic communication have not specified the functional mechanisms and neuroanatomic pathways that participate in the generation of a speech signal with "intimately intertwined linguistic and expressive cues" (Scherer et al. 2009, p. 446; see also Banse & Scherer 1996, p. 618). This deep entrenchment of articulate speech into a "nonverbal matrix" has been assumed to represent "the weakest point of gestural theories" of language evolution (Ingold 1994, p. 302).

Within the vocal domain, Parkinson's disease (PD) – a paradigmatic dysfunction of dopamine neurotransmission at the level of the striatal component of the basal ganglia – gives predominantly rise to a disruption of prosodic aspects of verbal utterances. Thus, the "addition of prosodic contour" to articulate speech appears to depend on the integrity of the striatum (Darkins et al. 1988; see Van Lancker Sidtis et al. 2006). Against this background, structural reorganization of the basal ganglia during hominin evolution may have been a pivotal prerequisite for the emergence of spoken language, providing a crucial phylogenetic link – at least at the motor level – between the vocalizations of our primate ancestors, on the one hand, and the volitional motor aspects of articulate speech, on the other.[2]

Comparative molecular-genetic data corroborate this suggestion: First, certain mutations of the *FOXP2* gene in humans give rise to developmental verbal dyspraxia. This disorder of spoken language, presumably, reflects impaired sequencing of orofacial movements in the absence of basic deficits of motor execution such as paresis of vocal tract muscles (Fisher et al. 2003; Fisher & Scharff 2009; Vargha-Khadem et al. 2005). Individuals affected with developmental verbal dyspraxia show a reduced volume of the striatum, the extent of which is correlated with the severity of nonverbal oral and speech motor impairments (Watkins et al. 2002b).[3] Second, placement of two hominin-specific *FOXP2* mutations into the mouse genome ("humanized *Foxp2*") gives rise to distinct morphological changes at the cellular level of the cortico-striatal-thalamic circuits in these rodents (Enard 2011). However, verbal dyspraxia subsequent to *FOXP2* mutations is characterized by a fundamentally different profile of speech motor deficits as compared to Parkinsonian dysarthria. The former resembles a communication disorder which, in adults, reflects damage to fronto-opercular cortex (i.e., inferior frontal/lower precentral gyrus) or the anterior insula of the language-dominant hemisphere (Ackermann & Riecker 2010b; Ziegler 2008).

To resolve this dilemma, we propose that ontogenetic speech acquisition depends on close interactions between the basal ganglia and their cortical targets, whereas mature verbal communication requires much less striatal processing capacities. This hypothesis predicts different speech motor deficits in perinatal dysfunctions of the basal ganglia as compared to the acquired dysarthria of PD patients. More specifically, basal ganglia disorders with an onset prior to speech acquisition should severely disrupt articulate speech rather than predominantly compromise the implementation of speech prosody.

1.3. Organization of this target article

The suggestion that structural refinement of cortico-striatal circuits – driven by human-specific mutations of the *FOXP2* gene – represents a pivotal step towards the emergence of spoken language in our hominin ancestors eludes any direct experimental evaluation. Nevertheless, certain inferences on the role of the basal ganglia in speech motor control can be tested against the available clinical and functional-imaging data. As a first step, the neuroanatomical underpinnings of the vocal behavior of nonhuman primates are reviewed in section 2 – as a prerequisite to the subsequent investigation of the hypothesis that in our species this system conveys nonverbal information through affective vocalizations and emotive/attitudinal speech prosody (sect. 3). Based upon clinical and neurobiological data, section 4 then characterizes the differential contribution of the basal ganglia to spoken language at the levels of ontogenetic speech acquisition (sect. 4.2.1) and of mature articulate speech (sect. 4.2.2), and delineates a neurophysiological model of the participation of the striatum in verbal behavior. Finally, these data are put into a paleoanthropological perspective in section 5.

2. Acoustic communication in nonhuman primates: Behavioral variation and cerebral control

2.1. Structural malleability of vocal signals

2.1.1. Ontogenetic emergence of acoustic call morphology. The vocal repertoires of monkeys and apes encompass noise-like and harmonic components (Fig. 1A; De Waal 1988; Goodall 1986; Struhsaker 1967; Winter et al. 1966). Vocal signals of both categories vary considerably across individuals, because age, body size, and stamina influence vocal tract shape and tissue characteristics, for

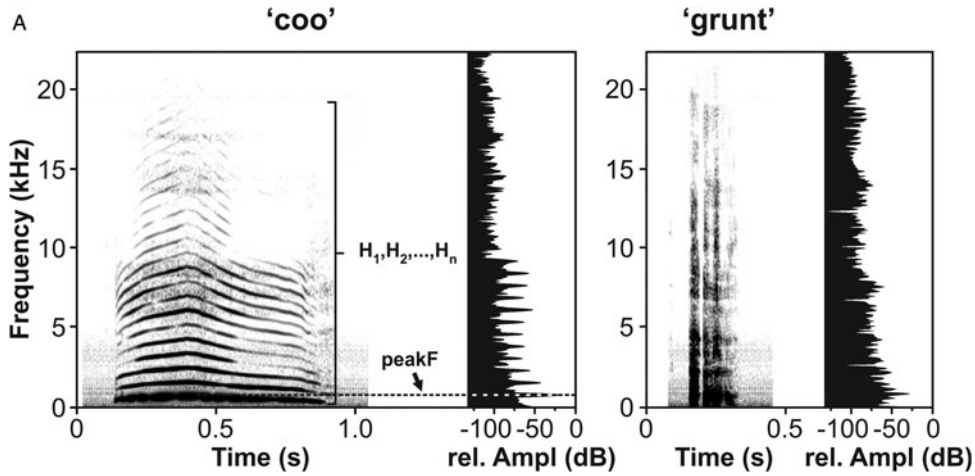

Figure 1A. *Acoustic communication in nonhuman primates: Call structure.*
A. Spectrograms (left-hand section of each panel) and power spectra (right-hand section in each) of two common rhesus monkey vocalizations, that is, a "coo" (left panel) and a "grunt" (right panel). Gray level of the spectrograms codes for spectral energy. Coo calls (left panel) are characterized by a harmonic structure, encompassing a fundamental frequency (F0, the lowest and darkest band) and several harmonics (H_1 to H_n). Measures derived from the F0 contour provide robust criteria for a classification of periodic signals, for example, peak frequency (peakF; Hardus et al. 2009a). Onset F0 seems to be highly predictive for the shape of the intonation contour, indicating the implementation of a "vocal plan" prior to movement initiation (Miller et al. 2009a; 2009b). Grunts (right) represent short and noisy calls whose spectra include more energy in the lower frequency range and a rather flat energy distribution.

example, the distance between the lips and the larynx (Fischer et al. 2002; 2004; Fitch 1997; but see Rendall et al. 2005). However, experiments based on acoustic deprivation of squirrel monkeys (*Saimiri sciureus*) and cross-fostering of macaques and lesser apes revealed that call structure does not appear to depend in any significant manner on species-typical auditory input (Brockelman & Schilling 1984; Geissmann 1984; Hammerschmidt & Fischer 2008; Owren et al. 1992; 1993; Talmage-Riggs et al. 1972; Winter et al. 1973). Thus, ontogenetic modifications of acoustic structure may simply reflect maturation of the vocal apparatus, including "motor-training" effects (Hammerschmidt & Fischer 2008; Pistorio et al. 2006), or the influence of hormones related to social status (Roush & Snowdon 1994; 1999). In contrast, comprehension and usage of acoustic signals show considerably more malleability than acoustic structure both in juvenile and adult animals (Owren et al. 2011).

2.1.2. Spontaneous adult call plasticity: Convergence on and imitation of species-typical variants of vocal behavior. Despite innate acoustic call structures, the vocalizations of nonhuman primates may display some context-related variability in adulthood. For example, two populations of pygmy marmosets (*Cebuella pygmaea*) of a different geographic origin displayed convergent shifts of spectral and durational call parameters (Elowson & Snowdon 1994; see further examples in Snowdon & Elowson 1999 and Rukstalis et al. 2003). Humans may also match their speaking styles inadvertently during conversation ("speech accommodation theory"; Burgoon et al. 2010; see Masataka [2008a; 2008b] for an example). Such accommodation effects could provide a basis for the changes in call morphology during social interactions in nonhuman primates (Fischer 2003; Mitani & Brandt 1994; Mitani & Gros-Louis 1998; Sugiura 1998). Subsequent reinforcement processes may give rise to "regional dialects" of primate species (Snowdon 2008). Rarely, even memory-based imitation capabilities have been observed in great apes: Thus, free-living chimpanzees were found to copy the distinctive intonational and rhythmic pattern of the pant hoots of other subjects – even after the animal providing the acoustic template had disappeared from the troop (Boesch & Boesch-Achermann 2000, pp. 234f). Whatever the precise mechanisms of vocal convergence, these phenomena are indicative of the operation of a neuronal feedback loop between auditory perception and vocalization in nonhuman primates (see Brumm et al. 2004).

A male bonobo infant ("Kanzi") reared in an enriched social environment spontaneously augmented his species-typical repertoire by four "novel" vocalizations (Hopkins & Savage-Rumbaugh 1991). However, these newly acquired signals can be interpreted as scaled variants of a single intonation contour (Fig. 3 in Taglialatela et al. 2003). Since *Pan paniscus* has, to some degree, a graded rather than discrete call system (Bermejo & Omedes 1999; Clay & Zuberbühler 2009), new behavior challenges could give rise to a differentiation of the available "vocal space" – indicating a potential to modulate call structures within the range of innate acoustic constraints rather than the ability to learn new vocal signals. An alternative interpretation is that hitherto un-deployed vocalizations were recruited under those conditions (Lemasson & Hausberger 2004; Lemasson et al. 2005).

2.1.3. Volitional initiation of vocal behavior and modulation of acoustic call structure. It has been a matter of debate for decades, in how far nonhuman primates are capable of volitional call initiation and modulation. A variety of behavioral studies seem to indicate both control over the timing of vocal output and the capacity to "decide" which acoustic signal to emit in a given context. First, at least two species of New World primates (tamarins, marmosets) discontinue acoustic communication during

epochs of increased ambient noise in order to avoid signal interferences and, therefore, to increase call detection probability (Egnor et al. 2007; Roy et al. 2011). In addition, callitrichid monkeys obey "conversational rules" and show response selectivity during vocal exchanges (Miller et al. 2009a; 2009b; but see Rukstalis et al. 2003: independent F0 onset change). Such observations were assumed to indicate some degree of volitional control over call production. As an alternative interpretation, these changes in vocal timing or loudness could simply reflect threshold effects of audio-vocal integration mechanisms. Second, several nonhuman primates produce acoustically different alarm vocalizations in response to distinct predator species, suggesting volitional access to call type (e.g., Seyfarth et al. 1980). Again, variation of motivational states could account for these findings. For example, the approach of an aerial predator could represent a much more threatening event than the presence of a snake. To some extent, even dynamic spectro-temporal features resembling the formant transients of the human acoustic speech signal (see below sect. 4.1.) appear to contribute to the differentiation of predator-specific alarm vocalizations ("leopard calls") in Diana monkeys (*Cercopithecus diana*) (Riede & Zuberbühler 2003a; 2003b; see Lieberman [1968] for earlier data). Yet, computer models insinuate that larynx lowering makes a critical contribution to these changes (Riede et al. 2005; 2006; see critical comments in Lieberman 2006b), thus, eliciting in a receiver the impression of a bigger-than-real body size of the sender (Fitch 2000b; Fitch & Reby 2001). Diana monkeys may have learned this manoeuver as a strategy to mob large predators, a behavior often observed in the wild (Zuberbühler & Jenny 2007).

The question of whether nonhuman primates are able to decouple their vocalizations from accompanying motivational states and to use them in a goal-directed manner has been addressed in several operant-conditioning experiments (Aitken & Wilson 1979; Coudé et al. 2011; Hage et al. 2013; Koda et al. 2007; Sutton et al. 1973; West & Larson 1995). In most of these studies, nonhuman primates learned to utter a vocalization in response to a food reward (e.g., Coudé et al. 2011; Koda et al. 2007). Rather than demonstrating the ability to volitionally vocalize on command, these studies merely confirm, essentially, that nonhuman primates produce adequate, motivationally based behavioral reactions to hedonistic stimuli. A recent study found, however, that rhesus monkeys can be trained to produce different call types in response to arbitrary visual signals and that they are capable to switch between two distinct call types associated with different cues on a trial-to-trial basis (Hage et al. 2013). These observations indicate that the animals are able – within some limits – to volitionally initiate vocalizations and, therefore, are capable to instrumentalize their vocal utterances in order to accomplish behavioral tasks successfully. Likewise, macaque monkeys may acquire control over loudness and duration of coo calls (Hage et al. 2013; Larson et al. 1973; Sutton et al. 1973; 1981; Trachy et al. 1981). A more recent investigation even reported spontaneous differentiation of coo calls in Japanese macaques with respect to peak and offset of the F0 contour during operant tool-use training (Hihara et al. 2003). Such accomplishments may, however, be explained by the adjustment of respiratory functions and do not conclusively imply operant control over spectro-temporal call structure in nonhuman primates (Janik & Slater 1997; 2000).

2.1.4. Observational acquisition of species-atypical sounds. Few instances of species-atypical vocalizations in nonhuman primates have been reported so far. Allegedly, the bonobo Kanzi, mentioned earlier, spontaneously acquired a few vocalizations resembling spoken words (Savage-Rumbaugh et al. 2004). Yet, systematic perceptual data substantiating these claims are not available. As further anecdotal evidence, Wich et al. (2009) reported that a captive-born female orangutan (*Pongo pygmaeus* × *Pongo abelii*) began to produce human-like whistles at an age of about 12 years in the absence of any training. Furthermore, an idiosyncratic pant hoot variant ("Bronx cheer" – resembling a sound called "blowing raspberries") spread throughout a colony of several tens of captive chimpanzees after it had been introduced by a male joining the colony (Hopkins et al. 2007; Marshall et al. 1999; similar sounds have been observed in wild orangutans: Hardus et al. 2009a; 2009b; van Schaik et al. 2003; 2006). Remarkably, these two acoustic displays, "raspberries" and whistles, do not engage laryngeal sound-production mechanisms, but reflect a linguo-labial trill ("raspberries") or arise from oral air-stream resonances (whistles). Thus, the species-atypical acoustic signals in nonhuman primates observed to date spare glottal mechanisms of sound generation. Apparently, laryngeal motor activity cannot be decoupled volitionally from species-typical audiovisual displays (Knight 1999).

2.2. Cerebral control of motor aspects of call production

2.2.1. Brainstem mechanisms (PAG and pontine vocal pattern generator). Since operant conditioning of the calls of nonhuman primates is technically challenging (Pierce 1985), analyses of the neurobiological control mechanisms engaged in phonatory functions relied predominantly on electrical brain stimulation. In squirrel monkeys (*Saimiri sciureus*) – the species studied most extensively so far (Gonzalez-Lima 2010) – vocalizations could be elicited at many cerebral locations, extending from the forebrain to the lower brainstem. This network encompasses a variety of subcortical limbic structures such as the hypothalamus, septum, and amygdala (Fig. 1B; Brown 1915; Jürgens 2002b; Jürgens & Ploog 1970; Smith 1945). In mammals, all components of this highly conserved "communicating brain" (Newman 2003) appear to project to the periaqueductal grey (PAG) of the midbrain and the adjacent mesencephalic tegmentum (Gruber-Dujardin 2010).[4] Based on the integration of input from motivation-controlling regions, sensory structures, motor areas, and arousal-related systems, the PAG seems to gate the vocal dimension of complex multimodal emotional responses such as fear or aggression. The subsequent coordination of cranial nerve nuclei engaged in the innervation of vocal tract muscles depends on a network of brainstem structures, including, particularly, a vocal pattern generator bound to the ventrolateral pons (Hage 2010a; 2010b; Hage & Jürgens 2006).

2.2.2. Mesiofrontal cortex and higher-order aspects of vocal behavior. Electrical stimulation studies revealed that both New and Old World monkeys possess a "cingulate vocalization region" within the anterior cingulate cortex

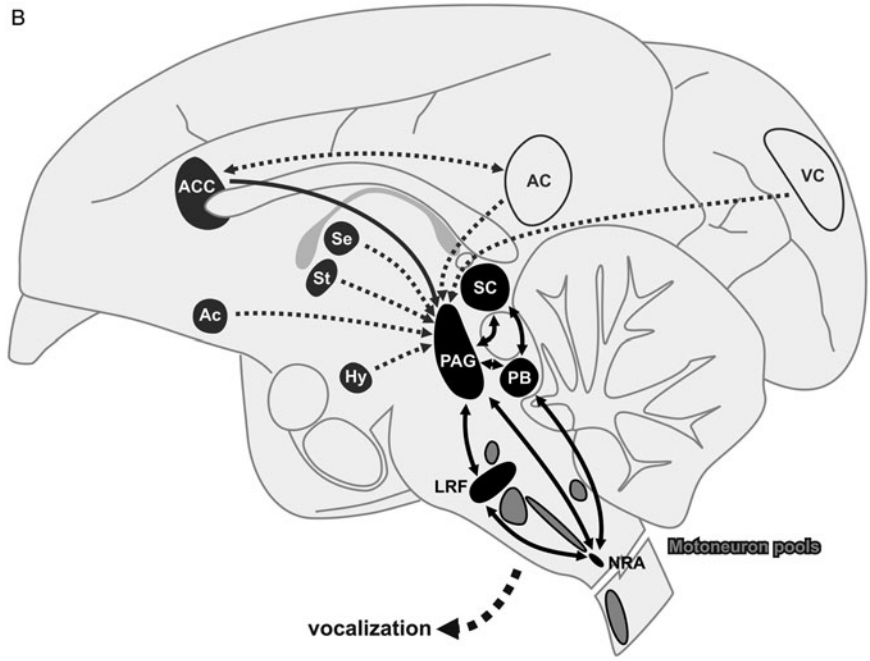

Figure 1B. *Acoustic Communication in nonhuman Primates: Cerebral Organization*.
Cerebral "vocalization network" of the squirrel monkey (as a model of the primate-general "communication brain"). The solid lines represent the "vocal brainstem circuit" of the vocalization network and its modulatory cortical input (ACC), the dotted lines the strong connections of sensory cortical regions (AC, VC) and motivation-controlling limbic structures (Ac, Hy, Se, St) to this circuit.
Key: ACC = Anterior cingulate cortex; AC = Auditory cortex; Ac = Nucleus accumbens; Hy = Hypothalamus; LRF = Lateral reticular formation; NRA = Nucleus retroambigualis; PAG = periaqueductal gray; PB = brachium pontis; SC = superior colliculus; Se = Septum; St = Nucleus stria terminalis; VC = Visual cortex (Unpublished figure. See Jürgens 2002b and Hage 2010a; 2010b for further details).

(ACC), adjacent to the anterior pole of the corpus callosum (Jürgens 2002b; Smith 1945; Vogt & Barbas 1988). Uni- and bilateral ACC ablation in macaques had, however, a minor and inconsistent impact on spontaneously uttered coo calls, but disrupted the vocalizations produced in response to an operant-conditioning task (Sutton et al. 1974; Trachy et al. 1981). Furthermore, damage to preSMA – a cortical area neighboring the ACC in dorsal direction and located rostral to the supplementary motor area (SMA proper) – resulted in significantly prolonged response latencies (Sutton et al. 1985). Comparable lesions in squirrel monkeys diminish the rate of spontaneous isolation peeps, but the acoustic structure of the produced calls remains undistorted (Kirzinger & Jürgens 1982). As a consequence, mesiofrontal cerebral structures appear to predominantly mediate calls driven by an animal's internal motivational milieu.

2.2.3. Ventrolateral frontal lobe and corticobulbar system. Both squirrel and rhesus monkeys possess a neocortical representation of internal and external laryngeal muscles in the ventrolateral part of premotor cortex, bordering areas associated with orofacial structures, namely, tongue, lips, and jaw (Fig. 1 in Hast et al. 1974; Jürgens 1974; Simonyan & Jürgens 2002; 2005). Furthermore, vocalization-selective neuronal activity may arise at the level of the premotor cortex in macaques that are trained to respond with coo calls to food rewards (Coudé et al. 2011). Interestingly, premotor neural firing appears to occur only when the animals produce vocalizations in a specific learned context of food reward, but not under other conditions. Finally, a cytoarchitectonic homologue to Broca's area of our species has been found between the lower branch of the arcuate sulcus and the subcentral dimple just above the Sylvian fissure in Old World monkeys (Gil-da-Costa et al. 2006; Petrides & Pandya 2009; Petrides et al. 2005) and chimpanzees (Sherwood et al. 2003). Nevertheless, even bilateral damage to the ventrolateral aspects of the frontal lobes has no significant impact on the vocal behavior of monkeys (P. G. Aitken 1981; Jürgens et al. 1982; Myers 1976; Sutton et al. 1974). Electrical stimulation of these areas in nonhuman primates also failed to elicit overt acoustic responses, apart from a few instances of "slight grunts" obtained from chimpanzees (Bailey et al. 1950, pp. 334f, 355f). Therefore, spontaneous call production, at least, does not critically depend on the integrity of the cortical larynx representation (Ghazanfar & Rendall 2008; Simonyan & Jürgens 2005). Most likely, however, experimental lesions have not included the full extent or even the bulk of the Broca homologue of nonhuman primates as determined by recent cytoarchitectonic studies (Fig. 4 in Aitken 1981; Fig. 1 in Sutton et al. 1974). The role of this area in the control of vocal behavior in monkeys still remains to be clarified. Nonhuman primates appear endowed with a more elaborate cerebral organization of orofacial musculature as compared to the larynx, which, presumably, provides the basis for their relatively advanced orofacial imitation capabilities (Morecraft et al. 2001). As concerns the basal ganglia and the cerebellum, the lesion and stimulation studies available so far do not provide reliable evidence for a participation of these structures in the control of motor aspects of vocal behavior (Kirzinger 1985; Larson et al. 1978; Robinson 1967).

Prosimians and New World monkeys are endowed solely with polysynaptic corticobulbar projections to lower

brain-stem motoneurons (Sherwood 2005; Sherwood et al. 2005). By contrast, morphological and neurophysiological studies revealed direct connections of the precentral gyrus of Old World monkeys and chimpanzees to the cranial nerve nuclei engaged in the innervation of orofacial muscles (Jürgens & Alipour 2002; Kuypers 1958b; Morecraft et al. 2001) which, together with the aforementioned more elaborate cortical representation of orofacial structures, may contribute to the enhanced facial-expressive capabilities of anthropoid primates (Sherwood et al. 2005). Most importantly, the direct connections between motor cortex and nucleus (nu.) ambiguus appear restricted, even in chimpanzees, to a few fibers targeting its most rostral component (Kuypers 1958b), subserving the innervation of pharyngeal muscles via the ninth cranial nerve (Butler & Hodos 2005). By contrast, humans exhibit considerably more extensive monosynaptic cortical input to the motoneurons engaged in the innervation of the larynx – though still less dense than the projections to the facial and hypoglossal nuclei (Iwatsubo et al. 1990; Kuypers 1958a). In addition, functional imaging data point to a primary motor representation of human internal laryngeal muscles adjacent to the lips of the homunculus and spatially separated from the frontal larynx region of New and Old World monkeys (Brown et al. 2008; 2009; Bouchard et al. 2013). As a consequence, thus, the monosynaptic elaboration of corticobulbar tracts during hominin evolution might have been associated with a refinement of vocal tract motor control at the cortical level ("Kuypers/Jürgens hypothesis"; Fitch et al. 2010).[5]

2.3. Summary: Behavioral and neuroanatomic constraints of acoustic communication in nonhuman primates

The cerebral network controlling acoustic call structure in nonhuman primates centers around midbrain PAG (vocalization trigger) and a pontine vocal pattern generator (coordination of the muscles subserving call production). Furthermore, mesiofrontal cortex (ACC/adjacent preSMA) engages in higher-order aspects of vocal behavior such as conditioned responses. These circuits, apparently, do not allow for a decoupling of vocal fold motor activity from species-typical audio-visual displays (Knight 1999). The resulting inability to combine laryngeal and orofacial gestures into novel movement sequences appears to preclude nonhuman primates from mastering even the simplest speech-like utterances, despite extensive vocal repertoires and a high versatility of their lips and tongue. At best, modification of acoustic call structure is restricted to the "variability space" of innate call inventories, bound to motivational or hedonistic triggers, and confined to intonational, durational, and loudness parameters, that is, signal properties homologous to prosodic aspects of human spoken language.

3. Contributions of the primate-general "limbic communicating brain" to human vocal behavior

The dual-pathway model of human acoustic communication predicts the "limbic communication system" of the brain of nonhuman primates to support the production of affective vocalizations such as laughing, crying, and moaning in our species. In addition, this network might engage in the emotive-prosodic modulation of spoken language. More specifically, ACC and/or PAG could provide a platform for the addition of graded, that is, analogue information on a speaker's motivational states and intentional composure to the speech signal. This suggestion has so far not been thoroughly tested against the available clinical data.

3.1. Brainstem mechanisms of speech production

Ultimately, all cerebral control mechanisms steering vocal tract movements converge on the same set of cranial nerve nuclei. Damage to this final common pathway, therefore, must disrupt both verbal and nonverbal aspects of human acoustic communication. By contrast, clinical observations in patients with bilateral lesions of the frontoparietal operculum and/or the adjacent white matter point at the existence of separate voluntary and emotional motor systems at the supranuclear level (Groswasser et al. 1988; Mao et al. 1989). However, these data do not further specify the course of the "affective-vocal motor system" and, more specifically, the role of the PAG, a major component of the primate-general "limbic communication system" (Lamendella 1977).

According to the dual-pathway model, the cerebral network supporting affective aspects of acoustic communication in our species must include the PAG, but bypass the corticobulbar tracts engaged in articulate speech. Isolated damage to this midbrain structure, thus, should selectively compromise the vocal expression of emotional/motivational states and spare the sound structure of verbal utterances. Yet, lesion data – though still sparse – are at variance with this suggestion. Acquired midbrain lesions restricted to the PAG completely interrupt both channels of acoustic communication, giving rise to the syndrome of akinetic mutism (Esposito et al. 1999). Moreover, comparative electromyographic (EMG) data obtained from cats and humans also indicate that the sound production circuitry of the PAG is recruited not only for nonverbal affective vocalizations, but also during speaking (Davis et al. 1996; Zhang et al. 1994). Likewise, a more recent positron emission tomography (PET) study revealed significant activation of this midbrain component during talking in a voiced as compared to a whispered speaking mode (Schulz et al. 2005).

Conceivably, the PAG contributes to the recruitment of central pattern generators of the brainstem. Besides the control of stereotyped behavioral activities such as breathing, chewing, swallowing, or yawning, these oscillatory mechanisms might, eventually, be entrained by superordinate functional systems as well (Grillner 1991; Grillner & Wallén 2004). During speech production, such brainstem networks could be instrumental in the regulation of highly adaptive sensorimotor operations during the course of verbal utterances. Examples include the control of inspiratory and expiratory muscle activation patterns in response to continuously changing biomechanical forces and the regulation of vocal fold tension following subtle alterations of subglottal pressure (see, e.g., Lund & Kolta 2006). From this perspective, damage to the PAG would interrupt the recruitment of basic adaptive brainstem mechanisms relevant for speech production and, ultimately, cause mutism. However, the crucial assumption of this explanatory model – spoken language engages phylogenetically older, though eventually reorganized, brainstem circuits – remains to be substantiated (Moore 2004; Schulz et al. 2005; Smith 2010).

3.2. Recruitment of mesiofrontal cortex during verbal communication

3.2.1. Anterior cingulate cortex (ACC). There is some evidence that, similar to subhuman primates, the ACC is a mediator of emotional/motivational acoustic expression in humans as well (see sect. 2.2.2). A clinical example is frontal lobe epilepsy, a syndrome characterized by involuntary and stereotyped bursts of laughter ("gelastic seizures"; Wild et al. 2003) that lack any concomitant adequate emotions (Arroyo et al. 1993; Chassagnon et al. 2003; Iannetti et al. 1997; Iwasa et al. 2002). The cingulate gyrus appears to be the most commonly disrupted site based on lesion surveys of gelastic seizure patients (Kovac et al. 2009). This suggestion was further corroborated by a recent case study in which electrical stimulation of the right-hemisphere ACC rostral to the genu of the corpus callosum elicited uncontrollable, but natural-sounding laughter – in the absence of merriment (Sperli et al. 2006). Conceivably, a homologue of the vocalization center of nonhuman primates bound to rostral ACC may underlie stereotyped motor patterns associated with emotional vocalizations in humans.

Does the ACC participate in speaking as well? Based on an early PET study, "two distinct speech-related regions in the human anterior cingulate cortex" were proposed, the more anterior of which was considered to be homologous to the cingulate vocalization center of nonhuman primates (Paus et al. 1996, p. 213). A recent and more focused functional imaging experiment by Loucks et al. (2007) failed to substantiate this claim. However, this investigation was based on rather artificial phonation tasks involving prolonged and repetitive vowel productions which do not allow for an evaluation of the specific role of the ACC in the mediation of *emotional* aspects of speaking. In another study, Schulz et al. (2005) required participants to recount a story in a voiced and a whispered speaking mode and demonstrated enhanced hemodynamic activation during the voiced condition in a region homologous to the cingulate vocalization center, but much larger responses emerged in contiguous neocortical areas of medial prefrontal cortex. It remains unclear, however, how the observed activation differences between voiced and whispered utterances should be interpreted, since both of these phonation modes require specific laryngeal muscle activity. One investigation explicitly aimed at a further elucidation of the role of medial prefrontal cortex in motivational aspects of speech production by analyzing the covariation of induced emotive prosody with blood oxygen level dependent (BOLD) signal changes as measured by functional magnetic resonance imaging (fMRI; Barrett et al. 2004). Affect-related pitch variation was found to be associated with supracallosal rather than pregeniculate hemodynamic activation. However, the observed response modulation may have been related to changes in the induced emotional states rather than pitch control. On the whole, the available functional imaging data do not provide conclusive support for the hypothesis that the prosodic modulation of verbal utterances critically depends on the ACC.

The results of lesion studies are similarly inconclusive. Bilateral ACC damage due to cerebrovascular disorders or tumours has been reported to cause a syndrome of akinetic mutism (Brown 1988; for a review, see Ackermann & Ziegler 1995). Early case studies found the behavioral deficits to extend beyond verbal and nonverbal acoustic communication: Apparently vigilant subjects with normal muscle tone and deep tendon reflexes displayed diminished or abolished spontaneous body movements, delayed or absent reactions to external stimuli, and impaired autonomic functions (e.g., Barris & Schuman 1953). By contrast, bilateral surgical resection of the ACC (cingulectomy), performed most often in patients suffering from medically intractable pain or psychiatric diseases, failed to significantly compromise acoustic communication (Brotis et al. 2009). The complex functional-neuroanatomic architecture of the anterior mesiofrontal cortex hampers, however, any straightforward interpretation of these clinical data. In monkeys, the cingulate sulcus encompasses two or even three distinct "cingulate motor areas" (CMAs), which project to the supplementary motor area (SMA), among other regions (Dum & Strick 2002; Morecraft & van Hoesen 1992; Morecraft et al. 2001). Humans exhibit a similar compartmentalization of the medial wall of the frontal lobes (Fink et al. 1997; Picard & Strick 1996). A closer look at the aforementioned surgical data reveals that bilateral cingulectomy for treatment of psychiatric disorders, as a rule, did not encroach on caudal ACC (Le Beau 1954; Whitty 1955; for a review, see Brotis et al. 2009, p. 276). Thus, tissue removal restricted to rostral ACC components could explain the relatively minor effects of this surgical approach.[6] Conceivably, mesiofrontal akinetic mutism reflects bilateral damage to the caudal CMA and/or its efferent projections, rather than dysfunction of a "cingulate vocalization center" bound to rostral ACC. Instead, the anterior mesiofrontal cortex has been assumed to contribute to reward-dependent selection/inhibition of verbal responses in conflict situations rather than to motor aspects of speaking (Calzavara et al. 2007; Paus 2001). This interpretation is compatible with the fact that psychiatric conditions bound to ACC pathology such as obsessive-compulsive disorder or Tourette syndrome cause, among other things, socially inappropriate vocal behavior (Müller-Vahl et al. 2009; Radua et al. 2010; Seeley 2008).

3.2.2. Supplementary motor area (SMA). Damage to the SMA in the language-dominant hemisphere may give rise to diminished spontaneous speech production, characterized by delayed, brief, and dysfluent, but otherwise well-articulated verbal responses without any central-motor disorders of vocal tract muscles or impairments of other language functions such as speech comprehension or reading aloud ("transcortical motor aphasia"; for a review of the earlier literature, see Jonas 1981; 1987; more recent case studies in Ackermann et al. 1996 and Ziegler et al. 1997).[7] This constellation may arise from initial mutism via an intermediate stage of silent word mouthing (Rubens 1975) or whispered speaking (Jürgens & von Cramon 1982; Masdeu et al. 1978; Watson et al. 1986). Based on these clinical observations, the SMA, apparently, supports the initiation ("starting mechanism") and maintenance of vocal tract activities during speech production (Botez & Barbeau 1971; Jonas 1981). Indeed, movement-related potentials preceding self-paced tongue protrusions and vocalizations were recorded over the SMA (*Bereitschaftspotential*; Ikeda et al. 1992). Calculation of the time course of BOLD signal changes during syllable repetition tasks, preceded by a warning stimulus, revealed an

earlier peak of the SMA response relative to primary sensorimotor cortex (Brendel et al. 2010). These data corroborate the suggestion – based on clinical data – of an engagement of the SMA in the preparation and initiation of verbal utterances, that is, pre-articulatory control processes.

3.3. Summary: Role of the primate-general "limbic communication system" in human vocal behavior

In line with the dual-pathway model of human acoustic communication, the ACC seems to participate in the release of stereotyped motor patterns of affective-vocal displays, even in the absence of an adequate emotional state. Whether this mesiofrontal area also contributes to the control of laryngeal muscles during speech production still remains to be established. An adjacent region, the neocortical SMA, appears, however, to participate in the preparation and initiation of articulate speech. Midbrain PAG also supports spoken language and, presumably, helps to recruit ancient brainstem circuitries which have been reorganized to subserve basic adaptive sensorimotor functions bound to verbal behavior.

4. Contribution of the basal ganglia to spoken language: Vocal-affective expression and acquisition of articulate speech

The basal ganglia represent an ensemble of subcortical gray matter structures of a rather conserved connectional architecture across vertebrate taxa, including the striatum (caudate nucleus and putamen), the external and internal segments of the globus pallidus, the subthalamic nucleus, and the substantia nigra (Butler & Hodos 2005; Nieuwenhuys et al. 2008). Clinical and functional imaging data indicate a significant engagement of the striatum both in ontogenetic speech acquisition and subsequent overlearned speech motor control. We propose, however, a fundamentally different role of the basal ganglia at these two developmental stages: The entrainment of articulatory vocal tract motor patterns during childhood versus the emotive-prosodic modulation of verbal utterances in the adult motor system.

4.1. Facets of the faculty of speaking: The recruitment of the larynx as an articulatory organ

The production of spoken language depends upon "more muscle fibers than any other human mechanical performance" (Kent et al. 2000, p. 273), and the responsible neural control mechanisms must steer all components of this complex action system at a high spatial and temporal accuracy. As a basic constituent, the larynx – a highly efficient sound source – generates harmonic signals whose spectral shape can be modified through movements of the mandible, tongue, and lips (Figs. 2A & 2B). Yet, this physical source-filter principle is not exclusively bound to human speech, but characterizes the vocal behavior of other mammals as well (Fitch 2000a). By contrast to the acoustic communication of nonhuman primates, spoken language depends, however, on a highly articulated larynx whose motor activities must be integrated with the gestures of equally articulated supralaryngeal structures into learned complex vocal tract movement patterns (Fig. 2C). For example, virtually all languages of the world differentiate between voiced and voiceless sounds (e.g., /b/ vs. /p/ or /d/ vs. /t/), a distinction which requires fast and precise laryngeal manoeuvres and a close interaction of the larynx – at a time-scale of tens of milliseconds – with the tongue or lips (Hirose 2010; Munhall & Löfqvist 1992; Weismer 1980). During voiced portions, moreover, the melodic line of the speech signal is modulated in a language-specific meaningful way to implement the intonation patterns inherent to a speaker's native idiom or, in tone languages such as Mandarin, to create different tonal variants of spoken syllables.

Clinical and functional-imaging observations indicate the "motor execution level" of speech production, that is, the adjustment of speed and range of coordinated vocal tract gestures, to depend upon lower primary sensorimotor cortex and its efferent pathways, the cranial nerve nuclei, the thalamus, the cerebellum – and the basal ganglia (Ackermann & Ziegler 2010; Ackermann & Riecker 2010a; Ackermann et al. 2010). More specifically, distributed and overlapping representations of the lips, tongue, jaw, and larynx within the ventral sensorimotor cortex of the dominant hemisphere generate, during speech production, dynamic activation patterns reflecting the gestural organization of spoken syllables (Bouchard et al. 2013). Furthermore, it is assumed that the left anterior peri- and subsylvian cortex houses hierarchically "higher" speech-motor-planning information in the adult brain required to orchestrate the motor execution organs during the production of syllables and words (see Fig. 2C for an illustration; Ziegler 2008; Ziegler et al. 2012). Hence, ontogenetic speech acquisition can be understood as a long-term entrainment of patterned activities of the vocal tract organs and – based upon practice-related plasticity mechanisms – the formation of a speech motor network which subserves this motor skill with ease and precision. In the following sections we argue that the basal ganglia play a key role in this motor-learning process and in the progressive assembly of laryngeal and supralaryngeal gestures into "motor plans" for syllables and words. In the mature system, this "motor knowledge" gets stored within ventrolateral aspects of the left-hemisphere frontal lobe, while the basal ganglia are, by and large, restricted to a fundamentally different role, that is, the mediation of motivational and emotional-affective drive into the speech motor system.

4.2. Developmental shifts in the contribution of the basal ganglia to speech production

4.2.1. The impact of pre- and perinatal striatal dysfunctions on spoken language. Insight into the potential contributions of the basal ganglia to human speech acquisition can be obtained from damage to these nuclei at a prelinguistic age. Distinct mutations of mitochondrial or nuclear DNA may give rise to infantile bilateral striatal necrosis, a constellation largely restricted to this basal ganglia component (Basel-Vanagaite et al. 2006; De Meirleir et al. 1995; Kim et al. 2010; Solano et al. 2003; Thyagarajan et al. 1995). At least two variants, both of them point mutations of the mitochondrial ATPase 6 gene, were associated with impaired speech learning capabilities (De Meirleir et al. 1995: "speech delayed for age"; Thyagarajan et al. 1995, case 1: "no useful language at age 3 years"). As a further clinical paradigm, birth asphyxia may

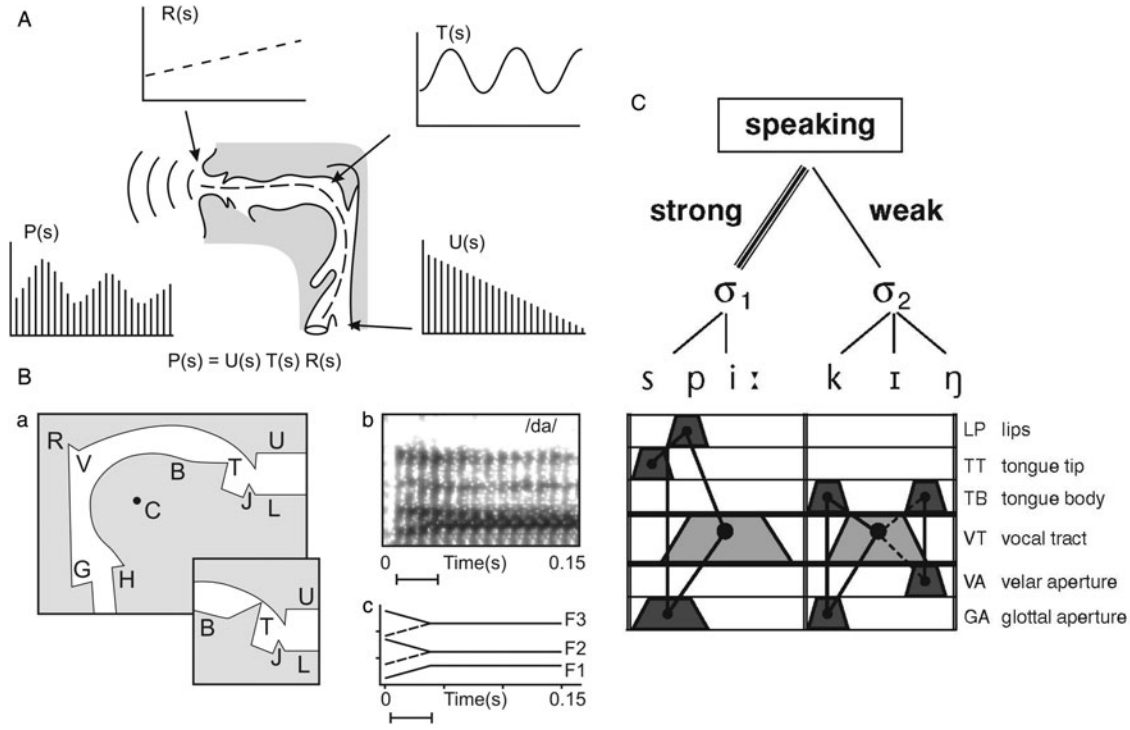

Figure 2. *Vocal tract mechanisms of speech sound production*.
A. Source-filter theory of speech production (Fant 1970). Modulation of expiratory air flow at the levels of the vocal folds and supralaryngeal structures (pharynx, velum, tongue, and lips) gives rise to most speech sounds across human languages (Ladefoged 2005). In case of vowels and voiced consonants, the adducted vocal folds generate a laryngeal source signal with a harmonic spectrum U(s), which is then filtered by the resonance characteristics of the supralaryngeal cavities T(s) and the vocal tract radiation function R (s). As a consequence, these sounds encompass distinct patterns of peaks and troughs (formant structure; P(s)) across their spectral energy distribution.
B. Consonants are produced by constricting the vocal tract at distinct locations (a), for example, through occlusion of the oral cavity at the alveolar ridge of the upper jaw by the tongue tip for /d/, /t/, or /n/ (insert of left panel: T/B=tip/body of the tongue, U/L = upper/lower lips, J = lower jaw with teeth). Such manoeuvres give rise to distinct up- and downward shifts of formants: Right panels show the formant transients of /da/ as a spectrogram (b) and a schematic display (c); dashed lines indicate formant transients of syllable /ba/ (figures adapted from Kent & Read 2002).
C. Schematic display of the gestural architecture of articulate speech, exemplified for the word *speaking*. Consonant articulation is based on distinct movements of lips, tongue, velum, and vocal folds, phase-locked to more global and slower deformations of the vocal tract (VT) associated with vowel production. Articulatory gestures are assorted into syllabic units, and gesture bundles pertaining to strong and weak syllables are rhythmically patterned to form metrical feet. Note that laryngeal activity in terms of glottal opening movements (bottom line) is a crucial part of the gestural patterning of spoken words and must be adjusted to and sequenced with other vocal tract movements in a precise manner (Ziegler 2010).

predominantly impact the basal ganglia and the thalamus (eventually, in addition, the brainstem) under specific conditions such as uterine rupture or umbilical cord prolapse, while the cerebral cortex and the underlying white matter are less affected (Roland et al. 1998). A clinical study found nine children out of a group of 17 subjects with this syndrome completely unable to produce any verbal utterances at the ages of 2 to 9 years (Krägeloh-Mann et al. 2002). Six further patients showed significantly compromised articulatory functions ("dysarthria"). Most importantly, five children had not mastered adequate articulate speech at the ages of 3 to 12 years, though lesions were confined to the putamen and ventro-lateral thalamus, sparing the caudate nucleus and the precentral gyrus.

Data from a severe developmental speech or language disorder of monogenic autosomal-dominant inheritance with full penetrance extending across several generations of a large family provide further evidence of a connection between the basal ganglia and ontogenetic speech acquisition (KE family; Hurst et al. 1990). At first considered a highly selective inability to acquire particular grammatical rules (Gopnik 1990a; for more details, see Taylor 2009), extensive neuropsychological evaluations revealed a broader phenotype of psycholinguistic dysfunctions, including nonverbal aspects of intelligence (Vargha-Khadem & Passingham 1990; Vargha-Khadem et al. 1995; Watkins et al. 2002a). However, the most salient behavioral deficit in the afflicted individuals consists of pronounced abnormalities of speech articulation ("developmental verbal dyspraxia") that render spoken language "of many of the affected members unintelligible to the naive listener" (Vargha-Khadem et al. 1995, p. 930; see also Fee 1995; Shriberg et al. 1997). Furthermore, the speech disorder was found to compromise voluntary control of nonverbal vocal tract movements (Vargha-Khadem et al. 2005). More specifically, the phenotype includes a significant disruption of simultaneous or sequential sets of motor activities to command, in spite of a preserved motility of single vocal tract organs (Alcock et al. 2000a) and uncompromised reproduction of tones and melodies (Alcock et al. 2000b).

A heterozygous point mutation (G-to-A nucleotide transition) of the *FOXP2* gene (located on chromosome 7; coding for a transcription factor) could be detected as the underlying cause of the behavioral disorder (for a review, see Fisher et al. 2003).[8] Volumetric analyses of striatal nuclei revealed bilateral volume reduction in the afflicted family members, the extent of which was correlated with oral-motor impairments (Watkins et al. 2002b). Mice and humans share all but three amino acids in the FOXP2 protein, suggesting a high conservation of the respective gene across mammals (Enard et al. 2002; Zhang et al. 2002). Furthermore, two of the three substitutions must have emerged within our hominin ancestors after separation from the chimpanzee lineage. Since primates lacking the human FOXP2 variant cannot even imitate the simplest speech-like utterances, and since disruption of this gene in humans gives rise to severe articulatory deficits, it appears warranted to assume that the human variant of this gene locus represents a necessary prerequisite for the phylogenetic emergence of articulate speech. Most noteworthy, animal experimentation suggests that the human-specific copy of this gene is related to acoustic communication (Enard et al. 2009) and directly influences the dendritic architecture of the neurons embedded into cortico-basal ganglia–thalamo–cortical circuits (Reimers-Kipping et al. 2011, p. 82).

4.2.2. Motor aprosodia in Parkinson's disease. A loss of midbrain neurons within the substantia nigra pars compacta (SNc) represents the pathophysiological hallmark of Parkinson's disease (PD; idiopathic Parkinsonian syndrome), one of the most common neurodegenerative disorders (Evatt et al. 2002; Wichmann & DeLong 2007). This degenerative process results in a depletion of the neurotransmitter dopamine at the level of the striatum, rendering PD a model of dopaminergic dysfunction of the basal ganglia, characterized within the motor domain by akinesia (bradykinesia, hypokinesia), rigidity, tremor at rest, and postural instability (Jankovic 2008; Marsden 1982). In advanced stages, functionally relevant morphological changes of striatal projection neurons may emerge (Deutch et al. 2007; see Mallet et al. [2006] for other non-dopaminergic PD pathomechanisms). Recent studies suggest that the disease process develops first in extranigral brainstem regions such as the dorsal motor nucleus of the glossopharyngeal and vagal nerves (Braak et al. 2003). These initial lesions affect the autonomic-vegetative nervous system, but do not encroach on gray matter structures engaged in the control of vocal tract movements such as the nu. ambiguus.

A classical tenet of speech pathology assumes that Parkinsonian speech/voice abnormalities reflect specific motor dysfunctions of vocal tract structures, giving rise to slowed and undershooting articulatory movements (brady-/hypokinesia). From this perspective, the perceived speech abnormalities of Parkinson's patients have been lumped together into a syndrome termed "hypokinetic dysarthria" (Duffy 2005). Unlike in other cerebral disorders, systematic auditory-perceptual studies and acoustic measurements identified laryngeal signs such as monotonous pitch, reduced loudness, and breathy/harsh voice quality as the most salient abnormalities in PD (Logemann et al. 1978; Ho et al. 1999a; 1999b; Skodda et al. 2009; 2011).[9] Imprecise articulation appears, by contrast, to be bound to later stages of the disease. In line with these suggestions, attempts to document impaired orofacial movement execution, especially, hypometric ("undershooting") gestures during speech production, yielded inconsistent results (Ackermann et al. 1997a). Moreover, a retrospective study based on a large sample of postmortem-confirmed cases found that PD patients predominantly display "hypophonic/monotonous speech," whereas atypical Parkinsonian disorders (APDs) such as multiple system atrophy or progressive supranuclear palsy result in "imprecise or slurred articulation" (Müller et al. 2001). As a consequence, Müller et al. assume the articulatory deficits of APD to reflect non-dopaminergic dysfunctions of brainstem or cerebellar structures.

Much like early PD, ischemic infarctions restricted to the putamen primarily give rise to hypophonia as the most salient speech motor disorder (Giroud et al. 1997). In its extreme, a more or less complete loss of prosodic modulation of verbal utterances ("expressive or motor aprosodia") has been observed following cerebrovascular damage to the basal ganglia (Cohen et al. 1994; Van Lancker Sidtis et al. 2006).[10] These specific aspects of speech motor disorders in PD or after striatal infarctions suggest a unique role of the basal ganglia in supporting spoken language production in that the resulting dysarthria might primarily reflect a diminished impact of motivational, affective/emotional, and attitudinal states on the execution of speech movements, leading to diminished motor activity at the laryngeal rather than the supralaryngeal level. Similar to other motor domains, thus, the degree of speech deficits in PD appears sensitive to "the emotional state of the patient" (Jankovic 2008), which, among other things, provides a physiological basis for motivation-related approaches to therapeutic regimens such as the Lee Silverman Voice Treatment (LSVT; Ramig et al. 2004; 2007). This general loss of "motor drive" at the level of the speech motor system and the predominant disruption of emotive speech prosody suggest that the intrusion of emotional/affective tone into the volitional motor mechanisms of speaking depends on a dopaminergic striatal "limbic-motor interface" (Mogenson et al. 1980).

4.3. Dual contribution of the striatum to spoken language: A neurophysiological model

4.3.1. Dopamine-dependent interactions between the limbic and motor loops of the basal ganglia during mature speech production. In mammals, nearly all cortical areas as well as several thalamic nuclei send excitatory, glutamatergic afferents to the striatum. This major input structure of the basal ganglia is assumed to segregate into the caudate-putamen complex, the ventral striatum with the nucleus accumbens as its major constituent, and the striatal elements of the olfactory tubercle (e.g., Voorn et al. 2004). Animal experimentation shows these basal ganglia subcomponents to be embedded into a series of parallel reentrant cortico-subcortico-cortical loops (Fig. 3A; Alexander et al. 1990; DeLong & Wichmann 2007; Nakano 2000). Several frontal zones, including primary motor cortex, SMA, and lateral premotor areas, target the putamen, which then projects back via basal ganglia output nuclei and thalamic relay stations to the respective areas of origin (motor circuit). By contrast, cognitive functions relate primarily to connections of prefrontal cortex with the caudate nucleus, and affective

states to limbic components of the basal ganglia (ventral striatum). Functional imaging data obtained in humans are consistent with such an at least tripartite division of the basal ganglia (Postuma & Dagher 2006) and point to a distinct representation of foot, hand, face, and eye movements within the motor circuit (Gerardin et al. 2003). Furthermore, the second basal ganglia output nucleus, the substantia nigra pars reticulata (SNr), projects to several hindbrain "motor centers," for example, PAG, giving rise to several phylogenetically old subcortical basal ganglia–brainstem–thalamic circuits (McHaffie et al. 2005). A brainstem loop traversing the PAG could participate in the recruitment of phylogenetically ancient vocal brainstem mechanisms during speech production (see sect. 3.1; Hikosaka 2007).

The suggestion of parallel cortico-basal ganglia–thalamo–cortical circuits does not necessarily imply strict segregation of information flow. To the contrary, connectional links between these networks are assumed to be a basis for integrative data processing (Joel & Weiner 1994; Nambu 2011; Parent & Hazrati 1995). More specifically, antero- and retrograde fiber tracking techniques reveal a cascade of spiraling striato-nigro-striatal circuits, extending from ventromedial (limbic) via central (cognitive-associative) to dorsolateral (motor) components of the striatum (Fig. 3A; e.g., Haber et al. 2000; for reviews, see Haber 2010a; 2010b). This dopamine-dependent "cascading interconnectivity" provides a platform for a cross-talk between the different basal ganglia loops and may, therefore, allow emotional/motivational states to impact behavioral responses, including the affective-prosodic shaping of the sound structure of verbal utterances.

The massive cortico- and thalamostriatal glutamatergic (excitatory) projections to the basal ganglia input structures target the GABAergic (inhibitory) medium-sized spiny projection neurons (MSN) of the striatum. MSNs comprise roughly 95% of all the striatal cellular elements. Upon leaving the striatum, the axons of these neurons connect via either the "direct pathway" or the "indirect pathway" to the output nuclei of the basal ganglia (Fig. 3B; Albin et al. 1989; for a recent review, see Gerfen & Surmeier 2011; for critical comments, see, e.g., Graybiel 2005; Nambu 2008). In addition, several classes of interneurons and dopaminergic projection neurons impact the MSNs. Dopamine has a modulatory effect on the responsiveness of these cells to glutamatergic input, depending on the receptor subtype involved (David et al. 2005; Surmeier et al. 2010a; 2010b). Against this background, MSNs must be considered the most pivotal computational units of the basal ganglia that are "optimized for integrating multiple

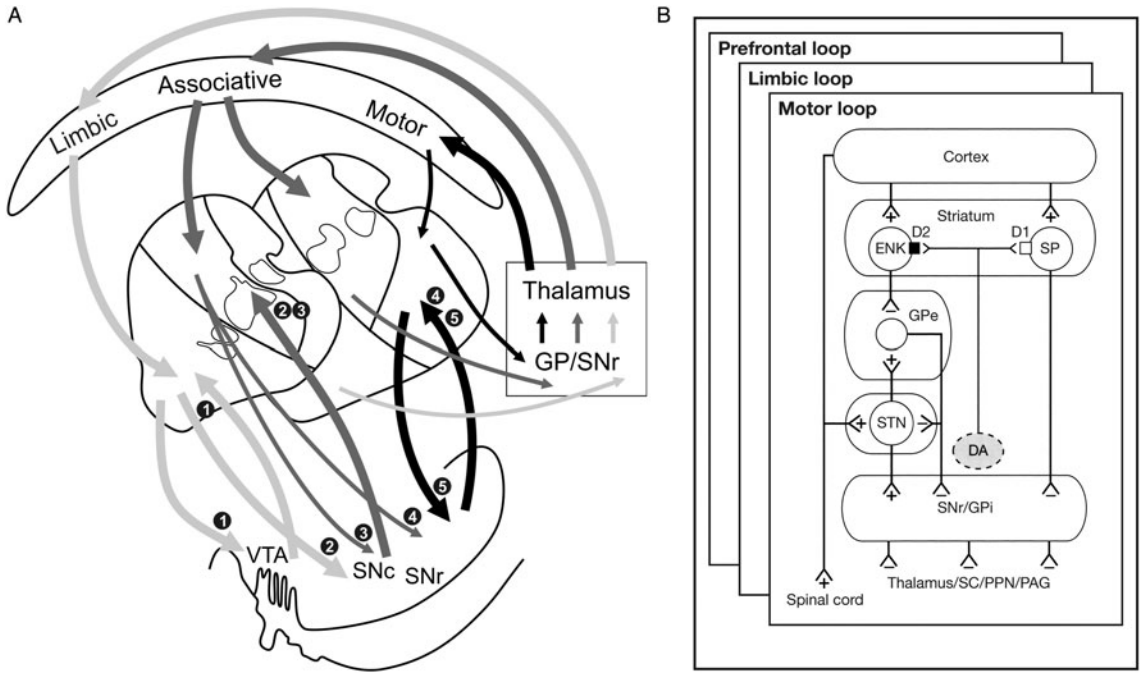

Figure 3. *Structural and functional compartmentalization of the basal ganglia.*
A. Schematic illustration of the – at least – tripartite functional subdivision of the cortico-basal ganglia–thalamo–cortical circuitry. Motor, cognitive/associative, and limbic loops are depicted in different gray shades, and the two cross-sections of the striatum (center) delineate the limbic, cognitive/associative, and motor compartments of the basal ganglia input nuclei. Alternating reciprocal (e.g., 1–1) and non-reciprocal loops (e.g., subsequent trajectory 2) form a spiraling cascade of dopaminergic projections interconnecting these parallel reentrant circuits (modified Fig. 2.3.5. from Haber 2010b).
B. Within the basal ganglia, the motor loop segregates into at least three pathways: a direct (striatum – SNr/GPi), an indirect (striatum – GPe – SNr/GPi), and a hyperdirect (via STN) circuit (based on Fig. 1 in Nambu 2011 and Fig. 25.1 in Walters & Bergstrom 2010). The direct and indirect medium-sized spiny projection neurons of the striatum (MSN) differ in their patterns of receptor and peptide expression (direct pathway: D1-type dopamine receptors, SP = substance P; indirect pathway: D2, ENK = enkephalin) rather than their somatodendritic architecture.
Key: DA = dopamine; GPi/GPe = internal/external segment of globus pallidus; SNr = substantia nigra, pars reticulata; SNc = substantia nigra, pars compacta; VTA = ventral tegmental area; STN = subthalamic nucleus; SC = superior colliculus; PPN = pedunculopontine nucleus; PAG = periaqueductal gray.

distinct inputs" (Kreitzer & Malenka 2008), including dopamine-dependent motivation-related information, conveyed via ventromedial–dorsolateral striatal pathways to those neurons. It is well established that midbrain dopaminergic neurons have a pivotal role within the context of classical/Pavlovian and operant/instrumental conditioning tasks (e.g., Schultz 2006; 2010). More specifically, unexpected benefits in association with a stimulus give rise to stereotypic short-latency/short-duration activity bursts of dopaminergic neurons which inform the brain on novel reward opportunities. Whereas, indeed, such brief responses cannot easily account for the impact of a speaker's mood such as anger or joy upon spoken language, other behavioral challenges, for example, longer-lasting changes in motivational state such as "appetite, hunger, satiation, behavioral excitation, aggression, mood, fatigue, desperation," are assumed to give rise to more prolonged striatal dopamine release (Schultz 2007, p. 207). Moreover, the midbrain dopaminergic system is sensitive to the motivational condition of an animal during instrumental conditioning tasks ("motivation to work for a reward"; Satoh et al. 2003).

The dopamine-dependent impact of motivation-related information on MSNs provides a molecular basis for the influence of a speaker's actual mood and actual emotions on the speech control mechanisms bound to the basal ganglia motor loop. Consequently, depletion of striatal dopamine should deprive vocal behavior from the "energetic activation" (Robbins 2010) arising in the various cortical and subcortical limbic structures of the primate brain (see Fig. 1B). The different basic motivational states of our species – shared with other mammals – are bound to distinct cerebral networks (Panksepp 1998; 2010). For example, the "rage/anger" and "fear/anxiety" systems involve the amygdala, which, in turn, targets the ventromedial striatum. On the other hand, the cortico-striatal motor loop is engaged in the control of movement execution, namely, the specification of velocity and range of orofacial and laryngeal muscles. The basal ganglia have an ideal strategic position to translate the various arousal-related mood states (joy or anger) into their respective acoustic signatures by means of a dopaminergic cascade of spiraling striato-nigro-striatal circuits – via adjustments of vocal tract innervation patterns ("psychobiological push effects of vocal affect expression"; Banse & Scherer 1996; Scherer et al. 2009). In addition, spoken language may convey a speaker's attitude towards a person or topic ("attitudinal prosody"; Van Lancker Sidtis et al. 2006). Such higher-order communicative functions of speech prosody involve a more extensive appraisal of the context of a conversation and may exploit learned stylistic (ritualized) acoustic models of vocal-expressive behavior (Scherer 1986; Scherer et al. 2009). Besides subcortical limbic structures and orbitofrontal areas, ACC projects to the ventral striatum in monkeys (Haber et al. 1995; Kunishio & Haber 1994; Öngür & Price 2000). Since these mesiofrontal areas are assumed to operate as a platform of motivational-cognitive interactions subserving response evaluation (see above), the connections of ACC with the striatum, conceivably, engage in the implementation of attitudinal aspects of speech prosody ("sociolinguistic/sociocultural pull factors" as opposed to the "psychobiological push effects" referred to above; Banse & Scherer 1996; Scherer et al. 2009). Thus, both the psychobiological push and the sociocultural pull effects, ultimately, may converge on the ventral striatum, which then, presumably, funnels this information into the basal ganglia motor loops.

4.3.2. Integration of laryngeal and supralaryngeal articulatory gestures into speech motor programs during speech acquisition. The basal ganglia are involved in the development of stimulus-response associations, for example, Pavlovian conditioning (Schultz 2006), and the acquisition of stimulus-driven behavioral routines, such as habit formation (Wickens et al. 2007). Furthermore, striatal circuits are known to engage in motor skill refinement, another variant of procedural (nondeclarative) learning.[11] For example, the basal ganglia input nuclei contribute to the development of "motor tricks" such as the control of a running wheel or the preservation of balance in rodents (Dang et al. 2006; Willuhn & Steiner 2008; Yin et al. 2009). Neuroimaging investigations and clinico-neuropsychological studies suggest that the basal ganglia contribute to motor skill learning in humans as well, though existing data are still ambiguous (e.g., Badgaiyan et al. 2007; Doya 2000; Doyon & Benali 2005; Kawashima et al. 2012; Packard & Knowlton 2002; Wu & Hallett 2005). The clinical observations referred to suggest that bilateral pre-/perinatal damage to the cortico-striatal-thalamic circuits gives rise to severe expressive developmental speech disorders which must be distinguished from the hypokinetic dysarthria syndrome seen in adult-onset basal ganglia disorders. Conceivably, thus, the primary control functions of these nuclei change across different stages of motor skill acquisition. In particular, the basal ganglia may primarily participate in the training phase preceding skill consolidation and automatization: The "engrams" shaping habitual behavior and the "programs" steering skilled movements, thus, may get stored in cortical areas rather than the basal ganglia (for references, see Graybiel 2008; Groenewegen 2003).

Yet, several functional imaging studies of upper-limb movement control failed to document a predominant contribution of the striatum to the early stages of motor sequence learning (Doyon & Benali 2005; Wu et al. 2004) or even revealed enhanced activation of the basal ganglia during overlearned task performance (Ungerleider et al. 2002) and, therefore, do not support this model. As a caveat, these experimental investigations may not provide an appropriate approach to the understanding of the neural basis of speech motor learning. Spoken language represents an outstanding "motor feat" in that its ontogenetic development starts early after or even prior to birth and extends over more than a decade. During this period, the specific movement patterns of an individual's native idiom are exercised more extensively than any other comparable motor sequences. A case similar to articulate speech can at most be made with educated musicians or athletes who have experienced extensive motor practice from early on over many years. In these subject groups, extended motor learning is known to induce structural adaptations of gray and white matter regions related to the level of motor accomplishments (Bengtsson et al. 2005; Gaser & Schlaug 2003). Such investigations into the *mature* neuroanatomic network of highly trained "motor experts" have revealed fronto-cortical and cerebellar regions[12] to be predominantly moulded by the effects of long-term motor learning with little or no evidence for any lasting

changes at the level of the basal ganglia (e.g., Gaser & Schlaug 2003). Against this background, it might be conjectured that the basal ganglia engage primarily in early stages of speech acquisition but do not house the motor representations that ultimately convey the fast, error-resistant, and highly automated vocal tract movement patterns of adult speech. This may explain why pre-/perinatal dysfunctions of the basal ganglia have a disastrous impact on verbal communication and preclude the acquisition of speech motor skills.

How can the contribution of the basal ganglia to the assembly of vocal tract motor patterns during speech acquisition be delineated in neurophysiological terms? One important facet is that the laryngeal muscles should have gained a larger striatal representation in our species as compared to other primates. Humans are endowed with more extensive corticobulbar fiber systems, including monosynaptic connections, engaged in the control of glottal functions (see sect. 2.2.3 above; Iwatsubo et al. 1990; Kuypers 1958a). Furthermore, functional imaging data point to a significant primary-motor representation of human internal laryngeal muscles, spatially separated from the frontal "larynx region" of New and Old World monkeys (Brown et al. 2008; 2009). In contrast to other primates, therefore, a higher number of corticobulbar fibers target the nu. ambiguus. As a consequence, the laryngeal muscles should have a larger striatal representation in our species since the cortico-striatal fiber tracts consist, to a major extent, of axon collaterals of pyramidal tract neurons projecting to the spinal cord and the cranial nerve nuclei, including the nu. ambiguus (Gerfen & Bolam 2010; Reiner 2010). Apart from the nu. accumbens, electrical stimulation of striatal loci in monkeys, in fact, failed to elicit vocalizations. In the latter case, however, the observed vocalizations reflect, most presumably, evoked changes in the animals' internal motivational milieu rather than the excitation of motor pathways (Jürgens & Ploog 1970).

A more extensive striatal representation of laryngeal functions can be expected to enhance the coordination of these activities with the movements of supralaryngeal structures. Briefly, the dorsolateral striatum separates into two morphologically identical compartments of MSNs, which vary, however, in neurochemical markers and input/output connectivity (Graybiel 1990; for recent reviews, see Gerfen 2010; Gerfen & Bolam 2010). While the so-called striosomes (patches) are interconnected with limbic structures, the matrisomes (matrix) participate predominantly in sensorimotor functions. This matrix component creates an intricate pattern of divergent/convergent information flow. For example, primary-motor and somatosensory cortical representations of the same body part are connected with the same matrisomes of the ipsilateral putamen (Flaherty & Graybiel 1993). Conversely, the projections of a single cortical primary-motor or somatosensory area to the basal ganglia appear to "diverge to innervate a set of striatal matrisomes which in turn send outputs that reconverge on small, possibly homologous sites" in pallidal structures further downstream (Flaherty & Graybiel 1994, p. 608). Apparently, such a temporary segregation and subsequent re-integration of cortico-striatal input facilitates "lateral interactions" between striatal modules and, thereby, enhances sensorimotor learning processes.

Similar to other body parts, it must be expected that the extensive larynx-related cortico-striatal fiber tracts of our species feed into a complex divergence/convergence network within the basal ganglia as well. These lateral interactions between matrisomes bound to the various vocal tract structures might provide the structural basis supporting the early stages of ontogenetic speech acquisition. More specifically, a larger striatal representation of laryngeal muscles – split up into a multitude of matrisomes – could provide a platform for the tight integration of vocal fold movements into the gestural architecture of vocal tract motor patterns (Fig. 2C).

4.4. Summary: Basal ganglia mechanisms bound to the integration of primate-general and human-specific aspects of acoustic communication

Dopaminergic dysfunctions of the basal ganglia input nuclei in the adult brain predominantly disrupt the embedding of otherwise well-organized speech motor patterns into an adequate emotive- and attitudinal-prosodic context. Based upon these clinical data, we propose that the striatum adds affective-prosodic modulation to the sound structure of verbal utterances. More specifically, the dopamine-dependent cascading interconnectivity between the various basal ganglia loops allows for a cross-talk between the limbic system and mature speech motor control mechanisms. By contrast, bilateral pre-/perinatal damage to the striato-thalamic components of the basal ganglia motor loops may severely impair speech motor integration mechanisms, resulting in compromised spoken language acquisition or even anarthria. We assume that the striatum critically engages in the initial organization of "motor programs" during speech acquisition, whereas the highly automatized control units of mature speech production, that is, the implicit knowledge of "how syllables and words are pronounced," are stored within anterior left-hemisphere peri-/subsylvian areas.

5. Paleoanthropological perspectives: A two-step phylogenetic/evolutionary scenario of the emergence of articulate speech

In a comparative view, the striatum appears to provide the platform on which a primate-general and, therefore, phylogenetically ancient layer of acoustic communication penetrates the neocortex-based motor system of spoken language production. Given the virtually complete speechlessness of nonhuman primates due to, especially, a limited role of laryngeal/supralaryngeal interactions during call production, structural elaboration of the cortico-basal ganglia–thalamic circuits should have occurred during hominin evolution. Recent molecular-genetic findings provide first specific evidence in support of this notion. More specifically, human-specific *FOXP2* copies may have given rise to an elaboration of somatodendritic morphology of basal ganglia loops engaged in the assemblage of vocal tract movement sequences during early stages of articulate speech acquisition. We propose, however, that the assumed *FOXP2*-driven "vocal-laryngeal elaboration" of the cortico-striatal-thalamic motor loop should have been preceded by a fundamentally different phylogenetic-developmental process, that is, the emergence of monosynaptic corticobulbar tracts engaged in the innervation of the laryngeal muscles.

5.1. Monosynaptic elaboration of the corticobulbar tracts: Enhanced control over tonal and rhythmic characteristics of vocal behavior (Step 1)

In nonhuman primates the larynx functions as an energetically efficient sound source, but shows highly constrained, if any, volitional motor capabilities. Direct projections of the motor cortex to the nu. ambiguus (see sect. 2.2.3) should have endowed this organ in humans with the potential to serve as a more skillful musical organ and an articulator with similar versatility as the lips and the tongue. Presumably, this first evolutionary step toward spoken language emerged independent of the presence of the human-specific *FOXP2* transcription factor. Structural morphometric (Belton et al. 2003; Vargha-Khadem et al. 1998; Watkins et al. 1999; 2002b) and functional imaging studies (Liégeois et al. 2003) in affected KE family members demonstrate abnormalities of all components of the cerebral speech motor control system, except the brainstem targets of the corticobulbar tracts (cranial nerve nuclei, pontine gray) and the SMA (Fig. 4 in Vargha-Khadem et al. 2005).[13] As an alternative to *FOXP2*-dependent neural processes, the increase of monosynaptic elaboration of corticobulbar tracts within the primate order (see sect. 2.2.3) might reflect a "phylogenetic trend" (Jürgens & Alipour 2002) associated with brain volume enlargement. Thus, "evolutionary changes in brain size frequently go hand in hand with major changes in both structural and functional details" (Striedter 2005, p. 12), For example, absolute brain volume predicts – via a nonlinear function – the size of various cerebral components, ranging from the medulla to the forebrain (Finlay & Darlington 1995). The three- to four-fold enlargement of absolute brain size in our species relative to australopithecine forms (Falk 2007), therefore, might have driven this refinement of laryngeal control – concomitant with a reorganization of the respective motor maps at the cortical level (Brown et al. 2008; 2009). Whatever the underlying mechanism, the development of monosynaptic projections of the motor strip to nu. ambiguus should have been associated with an enhanced versatility of laryngeal functions.

From the perspective of the lip-smack hypothesis (Ghazanfar et al. 2012), the elaboration of the corticobulbar tracts might have been a major contribution to turn the visual lip-smacking display into an audible signal (see MacNeilage 1998; 2008). Furthermore, this process should have allowed for a refinement of the rather stereotypic acoustic structure of the vocalizations of our early hominin ancestors (Dissanayake 2009, p. 23; Morley 2012, p. 131), for example, the "discretization" of (innate) glissando-like tonal call segments into "separate tonal steps" (Brandt 2009) or the capacity to match and maintain individual pitches (Bannan 2012, p. 309). Such an elaboration of the "musical characteristics" (Mithen 2006, p. 121) of nonverbal vocalizations, for example, contact calls, must have supported mother–child interactions. In order to impact the attention, arousal, or mood of young infants, caregivers often use non-linguistic materials such as "interjections, calls, and imitative sounds", characterized by "extensive melodic modulations" (Papoušek 2003). Furthermore, monosynaptic corticobulbar projections allow for rapid on/off switching of call segments and, thus, enable synchronization of vocal behavior, first, across individuals (communal chorusing in terms of "wordless vocal exchanges" as a form of "grooming-at-a-distance"; Dunbar 2012) and, second, with other body movements (dance). Such activities support interpersonal emotional bonds ("fellow-feeling") and promote social cohesion/cooperation (Cross 2001; 2003; Cross & Morley 2009). These accomplishments must have emerged after the separation of the hominin lineage since chimpanzees are unable to converge on a regular beat during call production (e.g., Geissmann 2000). More specifically, African apes engage in rhythmical behavior like drumming, but, apparently, lack the capacity of a mutual entrainment of such actions into synchronized group displays (Fitch 2012). Thus, monosynaptic elaboration of the corticobulbar tracts might have provided the phylogenetic basis both for the "communicative musicality" of human infants and for communal "wordless vocal exchanges," preceding both articulate speech and more formal musical activities shaped by culture (Malloch & Trevarthen 2009).[14] As a further indication that these achievements are not bound to the presence of the human-specific FOXP2 transcription factor, reproduction of musical tones and tunes was found largely uncompromised in KE family members with articulatory disorders (Alcock et al. 2000b).

The Kuypers/Jürgens hypothesis (Fitch et al. 2010) assumes that the vocal-behavioral limitations of nonhuman primates are rooted in the absence of direct corticobulbar projections to the brainstem motoneurons engaged in the innervation of laryngeal muscles and housed within the nu. ambiguus. Indeed, this model explains the inability of nonhuman primates to produce sound patterns that impose particularly high demands on the coordination of laryngeal and supralaryngeal activities such as the rapid voiced–voiceless alterations characteristic of articulate speech. Yet, this suggestion cannot account for nonhuman primates' inability to imitate less challenging, fully voiced, speech-like vocalizations such as syllables comprising voiced consonants (see sect. 4.3.2).

5.2. FOXP2-driven vocal elaboration of the basal ganglia motor loop: Enhanced integration of laryngeal and supralaryngeal gestures (Step 2)

As a further prerequisite of spoken language, the vocal folds must serve as an "articulatory organ" that can be "pieced together" with equally versatile orofacial gestures into a tightly integrated meshwork of appropriately timed vocal tract movements. Conceivably, *FOXP2*-driven morphological changes at the level of the basal ganglia in our hominin ancestors provided the physiological basis for these sensorimotor capabilities to emerge as a second phylogenetic step toward articulate speech. More specifically, enhanced "lateral interactions" between striatal representations of vocal tract muscles based on a divergence/convergence architecture of information flow within the basal ganglia (Flaherty & Graybiel 1994) have the potential to support the linkage of vocal tract movements into language-specific syllabic and metrical patterns. This would represent a major step in sensorimotor verbal learning during ontogenetic speech acquisition. The role of the basal ganglia in this process seems to be confined to the phase where the entrainment and automatization of speech motor patterns takes place, while the persistent motor plans evolving during this process get stored within left-hemisphere peri- or subsylvian cortex. In the mature

speech motor system, the contribution of the striatum to speech production appears predominantly restricted to dopamine-dependent, emotive-prosodic shading of the speech signal as a homologue to the vocalizations of nonhuman primates and a vestige of the ancient communication system.

Paleoanthropological data such as endocast traces of Broca's area (Holloway et al. 2004, pp. 15ff) or morphological features of the cranial base (Lieberman 2011) provide only indirect and ambiguous evidence on the evolution of spoken language. "Comparing our behavior and brain with those of other extant primates" (Ghazanfar & Miller 2006, p. R879) still represents the most robust approach to the investigation of the "biological mechanisms underlying the evolution of speech" (Ghazanfar & Rendall 2008, p. R457). Recently, however, molecular-genetic studies have shed light on the phylogeny of verbal communication in the hominin lineage and, more specifically, the contribution of the basal ganglia to the evolution of spoken language. Thus, molecular-genetic analyses found the human form of the FOXP2 protein in 43,000-year-old Neanderthal skeletal remains (Rosas et al. 2006) linked to the same haplotype as in our species (Krause et al. 2007).[15] Since large-scale analyses of the *FOXP2* locus in humans failed to detect any amino acid polymorphisms (Enard et al. 2002), those speech-related mutations must have been the target of strong selection pressures, causing a relatively fast fixation within the human gene pool ("selective sweep"). Assuming modern humans and Neanderthals did not interbreed, positive selection of the relevant *FOXP2* mutation(s) should have occurred in our most recent common ancestor (MRCA). Sequence analyses both of nuclear and mitochondrial DNA "locate" the MRCA to the mid-Middle Pleistocene, around 400,000 to 600,000 years ago (Endicott et al. 2010; Green et al. 2010; Hofreiter 2011; Noonan 2010), and these data are compatible with the fossil record (Weaver et al. 2008). As an alternative scenario, gene flow could explain the presence of the human *FOXP2* variant in Neanderthal bones (Coop et al. 2008). Under these conditions, a later emergence of the respective hominin mutations has been assumed – around 40,000 years ago (see Stringer 2012, pp. 190ff, for a recent discussion of interbreeding between modern humans and archaic populations, i.e., Neanderthals and Denisovans). A more recent molecular-genetic study, finally, points at a positive selective sweep of a regulatory *FOXP2* element – affecting neuronal expression of this gene – within a comparable time domain, that is, during the last 50,000 years (Maricic et al. 2013). In any case, whatever model will prove true, *FOXP2*-driven speech-related modification of cortico-striatal circuits must have emerged in individuals characterized by a cerebral volume similar to that of extant modern humans (Rightmire 2004; 2007).

Assuming a gradual monosynaptic elaboration of corticobulbar projections in parallel with brain size increase across the hominin lineage (see above), the relatively late reorganization of cortico-basal ganglia loops driven by specific *FOXP2* mutations should have occurred on top of a fully developed motoneuronal axis. It is tempting to relate the selective sweep of the hominin *FOXP2* mutations to the evolution of speech and language functions (Enard & Pääbo 2004; Zhang et al. 2002). However, the benefits of full-fledged verbal communication cannot have been the driving force of the emergence of articulate speech. "If the first one or three or five protolanguage signs [such as syllable repetitions or simple words] didn't have a substantial payoff, no one would have bothered to invent any more" (Bickerton 2009, p. 165). The announcement of "displaced" objects such as perished large mammals and the subsequent recruitment of troop members for carcass exploitation has been assumed to provide the necessary "substantial payoff" (Bickerton 2009, pp. 167f). But individuals spending their whole – though often short – lives together in small and intimate troops should have been able to convey such simple messages to a sufficient extent by nonverbal, that is, gestural means (Coward 2010, p. 469).

Rather than semantic-referential functions, the earliest speech-like vocalizations could have served as refined contact calls and, thus, facilitated mother–child interactions (Falk 2004; 2009). Likewise, these vocalizations might have allowed for a vocal elaboration of group activities such as communal dancing or grooming, which consolidate intragroup cohesion and cooperation (Dunbar 1996; Mithen 2006, pp. 208f). In other words, the earliest verbal utterances further expanded and refined the space of versatile vocal displays afforded by the preceding development of monosynaptic corticobulbar projections to the nu. ambiguus. Besides other benefits (see above), these accomplishments should have enhanced a "speaker's" social prestige. Subsequent gradual "conventionalization" (Milo & Quiatt 1994) of speech-like acoustic signals then could have slowly created opportunities for the conveyance of environmental or social information by simply drawing attention to an actual event or situation (Dessalles 2007, p. 360).

6. A look beyond the primate lineage: Birdsong and human speech

In a broader comparative perspective, the emergence of articulate speech appears to have involved the convergent evolution in our species of rather ancient principles of brain wiring, documented already many years ago in songbirds. The avian "song production network" roughly separates into two circuits, that is, the vocal motor pathway (VMP) and the anterior forebrain pathway (AFP; e.g., Bolhuis et al. 2010; Jarvis 2004a; 2004b). Whereas VMP shares essential organizational principles with human corticobulbar tracts such as monosynaptic projections to the cranial nerve centers steering the peripheral vocal apparatus (Wild 2008; see also Ackermann & Ziegler 2013), there are striking similarities between AFP and the cortico-basal ganglia loops of mammals, including our species (Doupe et al. 2005). In zebra finches, area X – a major AFP component that includes both striatal and pallidal elements – shows, for example, specific interdependencies between *FoxP2* level and the accuracy of tutor song imitation (Haesler et al. 2007) or juvenile/adult singing activity (Teramitsu et al. 2010; for an evolutionary perspective on this gene see Scharff & Haesler 2005). Whereas bilateral VMP damage significantly compromises vocal behavior at any stage of an individual's life history, AFP dysfunctions have, by contrast, a more subtle impact upon mature songs, but severely disrupt vocal learning mechanisms (e.g., Brainard & Doupe 2002). Thus, (i) monosynaptic connections between upper and lower motoneurons engaged in the innervation of the sound source and (ii)

cortico-striatal motor loops supporting vocal-laryngeal functions appear to represent common functional-neuroanatomic prerequisites both of spoken language and birdsong (for a review of the parallels between avian and human acoustic communication, see Doupe & Kuhl 1999; Bolhuis & Everaert 2013; Bolhuis et al. 2010). As a consequence, birdsong can serve as an experimental model for the investigation of the neural control of human speech – though, most presumably, syntactic and semantic aspects of verbal utterances elude such an approach (Beckers et al. 2012; Berwick et al. 2011). The hitherto underestimated role of the basal ganglia in spoken language should help to further elucidate the relationship between birdsong and human speech.

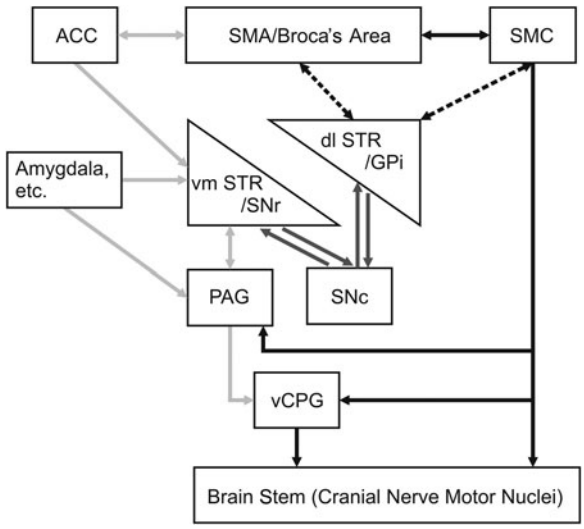

Figure 4. *Cerebral network supporting the integration of primate-general (gray arrows) and human-specific aspects of acoustic communication (black).*
A cascading dopaminergic circuitry (bidirectional arrows) connects the ventromedial-limbic (vm STR) with the dorsolateral-motor components of the striatum (dl STR) and their respective output nuclei, SNr and GPi. We suggest that this circuitry funnels information on a speaker's actual affective/motivational state into the central motor system, thereby modulating spoken language by an emotive-prosodic "tone," a homologue of the vocal behavior of nonhuman primates. Unlike what is postulated by dual-pathway models, the two networks appear to be closely intertwined at the level of the basal ganglia and of midbrain/brainstem structures. In our species, the motor cortex, first, has monosynaptic projections to brainstem nu. ambiguus and, second, the basal ganglia motor loop extends to laryngeal functions – based, probably, on the convergent evolution of a wiring schema already extant in songbirds – whereas nonhuman primates seem to lack such a "vocal elaboration" of subcortical-cortical motor circuitry. The dashed lines indicate that the basal ganglia motor loop, apparently, undergoes a dynamic ontogenetic reorganization during spoken language acquisition in that a left-hemisphere cortical storage site of "motor programs" gradually emerges, bearing the major load of vocal tract control after mature speech production has been established. (This figure does not include the cerebellum, a structure also engaged in speech motor control [see Ackermann 2008], but not relevant for the discussion in this article.)
Key: Amygdala etc. = amygdala and other (allocortical/mesolimbic) structures of the limbic system; ACC = anterior cingulate cortex; SMA = supplementary motor area; SMC = sensorimotor cortex; GPi = internal segment of globus pallidus; SNr/SNc = substantia nigra, pars reticulata/pars compacta; PAG = periaqueductal gray; vCPG = vocal central pattern generator.

7. Conclusions

During recent years, a salient contribution of subcortical structures, including the basal ganglia, to language evolution has been assumed (Lieberman 2000; 2007). More specifically, *FOXP2*-driven modification of neural circuits traversing the basal ganglia must be considered a necessary prerequisite for "the emergence of proficient spoken language" (Vargha-Khadem et al. 2005). However, these suggestions do not account for the developmental dynamics of cortico-striatal interactions and the discrepancies between the sequels of basal ganglia lesions in children and adults. Based upon behavioral–clinical and functional imaging data, in this article we have proposed (1) two successive phylogenetic stages of speech acquisition (monosynaptic refinement of corticobulbar tracts and laryngeal elaboration of cortico-striatal motor circuits), and (2) a functional reorganization of the cortico-striatal motor loops engaged in vocal tract control during ontogenetic speech development (Fig. 4).

It goes without saying that the model outlined here addresses only one out of several building blocks of a comprehensive theory of the evolution of spoken language. Most evidently, our approach still fails to account for the co-evolution of the described linguistic motor skills with the auditory skills underlying speech perception, and, as a consequence, the emergence of the auditory-motor network that underlies the phonological processing capacities of our species. Furthermore, we need to better understand how this elaborate auditory-vocal communication apparatus became overarched by the expanding conceptual-semantic and syntactic capabilities of humans. Thus, language evolution must be considered a multicomponent process, and the specific phylogenetic interactions of emergent speech production with these other traits await further elucidation. Presumably, any such phylogenetic account also needs to integrate, among other things, social and motivational contingencies (e.g., Dunbar 1996), "the desire to use the vocal tract to communicate" (Locke 1993, p. 322f), amodal mimetic capacities (Donald 1999), mirror neuron systems (Arbib 2006), and so-called executive functions (Coolidge & Wynn 2009) as relevant driving forces and prerequisites of spoken language evolution (for a comprehensive overview, see Tallerman & Gibson 2012).

NOTES

1. Though predominantly depending on glottal source characteristics such as the fluctuations of pitch, loudness, and voice quality, vocal-affective prosodic expression may also be associated with changes in speech breathing patterns, alterations of speaking rate, and the degree to which speech sounds are hyper- or hypo-articulated. Thus, motivational factors have, more or less, an impact on all vocal tract subsystems.

Affective-emotive speech prosody, that is, the expression of arousal-related mood states, has been considered as a behavioral trait homologous to the acoustic signals of nonhuman primates in addition to nonverbal affective vocalizations such as laughter ("push-effects" of affective-emotive prosody; see last paragraph in sect. 4.3.1). By contrast, attitudes like doubt or approval cannot unambiguously be expected in nonhuman primates.

Thus, it is questionable whether attitudinal prosody, that is, appraisal-related "pull-effects," can be assumed homologous to the vocal behavior of nonhuman primates.

Besides arousal-related motivational/affective states (e.g., joy) or appraisal-based subjective attitudes (e.g., doubt), speech prosody may also convey linguistic information such as word accent (linguistic prosody) or contribute to the implementation of "speech acts" such as verbal intimidation of another subject (Sidtis & Van Lancker Sidtis 2003; Van Lancker Sidtis et al. 2006). Linguistic and pragmatic prosody are outside the scope of this article.

In addition to a propositional message and affective/attitudinal states, the speech signal also conveys speaker-related ("indexical") information on age, gender, and identity, simply because the size and tissue properties of laryngeal and supralaryngeal structures differ across individuals and change over lifetime (Kreiman & Sidtis 2011).

2. The more recent paleoanthropological literature applies the term *hominin* – rather than *hominid* – to the human *clade* ("family"), that is, the "bush" of all species tracing back to a common ancestor who diverged from the lineage encompassing modern chimpanzees (Lewin & Foley 2004, p. 9).

3. Nucleotide sequences are given in italics, proteins in regular letters; lower- and uppercase serve to distinguish human (*FOXP2*/FOXP2), murine (*Foxp2*/Foxp2), and other, for example, avian (*FoxP2*/FoxP2) variants of the forkhead family of genes (Kaestner et al. 2000).

4. The PAG and the adjacent mesencephalic tegmentum represent a functional-neuroanatomic entity (Holstege 1991). In the subsequent paragraphs, the term "PAG" will always refer to both subcomponents.

5. Monosynaptic projections of (the avian) motor cortex to brainstem nuclei have also been documented in songbirds (for a review see, e.g., Wild 2008), an often neglected prerequisite of vocal learning (see sect. 6).

6. Two cases of a constellation resembling transcortical motor aphasia following ACC infarction have been documented to date (Chang et al. 2007). Diffusion tensor imaging revealed additional disruption of efferent SMA fibers in one patient. Thus, a substantial contribution of premotor mesiofrontal cortex to the observed communication disorders must be considered.

7. Two case studies noted compromised speech prosody after mesiofrontal lesion (Bell et al. 1990; Heilman et al. 2004). In the absence of more detailed neuroanatomic data, such observations are difficult to interpret unambiguously.

8. Further alterations of the *FOXP2* gene – such as a nonsense mutation giving rise to truncated protein products – have been found in association with developmental speech dyspraxia (MacDermot et al. 2005).

9. In contrast to other dysarthria variants, PD subjects show, as a rule, normal speaking rates. A subgroup of patients even displays an accelerated tempo ("hastening phenomenon"; e.g., Duffy 2005). This unique, but rarely studied, phenomenon may reflect a release of oscillatory basal ganglia activity (Ackermann et al. 1997b; Riecker et al. 2006).

10. Tracing back to the late 1970s (Ross & Mesulam 1979), a series of case studies assigned motor aprosodia – disrupted implementation of the "affective tone" of spoken language, concomitant with a preserved "ability to 'feel emotion' inwardly" and an unimpaired comprehension of other subjects' vocal expression of motivational states – to a dysfunction of right-hemisphere fronto-opercular cortex and/or anterior insula (e.g., Ross & Monnot 2008). However, the lesions in these cases appear to have encroached on the basal ganglia, including their connections to mesiofrontal cortex (see Cancelliere & Kertesz 1990).

11. In contrast to habit formation, that is, the incremental emergence of stimulus-driven behavioral routines, motor skill learning is characterized by the incremental refinement of movement execution as reflected in reaction time measurements: "Learning how to ride a bicycle is quite different from having the habit of biking every evening after work" (Graybiel 2008, p. 370).

12. As compared to the upper limbs, the specific contribution of the cerebellum to speech motor learning is less clear. Most noteworthy, the few reported cases of congenital cerebellar hypoplasia/aplasia, apparently, lack any significant disorders of spoken language (Ackermann & Ziegler 1992). Acquired dysfunctions of the cerebellum, nevertheless, compromise speech production, giving rise to, among other things, a slowed speaking rate and imprecise consonant articulation (Ackermann 2008; Duffy 2005).

13. These inferences must be considered with some precautions: We can only conclude that the heterozygous(!) constellations observed so far in the KE family (Bolhuis et al. 2010, p. 753) do not significantly disrupt the corticobulbar pathway – unlike other components of the central motor system.

14. Although contemporary traditional societies of a predominantly hunter-gatherer mode of subsistence "are not necessarily like some form of pre-human and should not be used uncritically as models," the respective ethnographic data, nevertheless, allow limited inferences on the behavioral repertoire of our hominin ancestors (Barnard 2011, p. 15). Thus, extensive communal dancing, often accompanied by rhythmic nonverbal utterances, represents a salient component of many ceremonies associated with important events in the life of an individual (e.g., circumcision rite; Turner 1967, pp. 186ff, 193) or the history of a group (war-/peace-related gatherings; e.g., Rappaport 2000, pp. 173ff). Since the coordination of vocal behavior and body movements may encourage a sense of "unity, harmony, and concord" among a group, social bonding should benefit from a vocal elaboration of ritual forms (Rappaport 1999, pp. 220, 252ff). It must be noted, however, that communal dancing often may include a competitive element aside from social bonding (James 2003, pp. 75f; for examples, see Rappaport 1999, p. 80; 2000, pp. 191ff; Turner 1967, p. 260). Principally, refined musical abilities could have supported to some extent referential communication. Spoken languages may include a broad range of nonverbal signals (Lewis 2009). For example, the Mbendjele people living in the dense equatorial forests of the Congo Basin, a habitat that severely impedes visual orientation, report an encounter with a dangerous animal to other group members by means of meticulous mimicry of the respective auditory scene. These anthropological data support the suggestion that enhanced musicality of nonverbal vocalizations may provide communicative benefits, but do not necessarily imply the notion of a "musical protolanguage" or "musilanguage" (Brown 2000), that is, music-like learned communication systems preceding full-fledged spoken language, a hypothesis tracing back to Charles Darwin (1871).

15. Similar to nonhuman primates, limitations of articulate speech due to vocal tract constraints have been attributed to Neanderthals as well, giving rise to a reduced repertoire of speech sounds (for a critical discussion, see Barney et al. 2012; Clegg 2012).

Open Peer Commentary

The sound of one hand clapping: Overdetermination and the pansensory nature of communication

doi:10.1017/S0140525X13003944

Kenneth John Aitken

Psychology Department, Hillside School, Aberdour, Fife KY3 0RH, Scotland, United Kingdom.

aitken_k@hillsideschool.co.uk drken.aitken@btinternet.com

Abstract: Two substantive issues are relevant to discussions of the evolution of acoustic communication and merit further consideration here. The first is the importance of communicative ontogeny and the impact of the proximal social environment on the early development of communication and language. The second is the emerging evidence for a number of non-linguistic roles of FOXP2 and its orthologs.

Ackermann et al. review evidence that changes to FoxP2 acted as the necessary and specific accelerant for human language development. I will briefly discuss three points relevant to this view: the role of pre-verbal interaction in language acquisition, the range of genes and pathways involved, and lastly the importance of FoxP2 changes in other species.

Communicative ontogeny. The ontogeny of communication stems from a latent genetic potential. This is channelled, constrained, and developed through the neonatal environment (Aitken 2008; Aitken & Trevarthen 1997; Crais et al. 2004; Rowe & Goldin-Meadow 2009). Neonates interact with adults with varied communicative capabilities. Over 100,000 years, newborns have adapted to massive changes in culture and language, while the genetic mechanisms proposed are largely unchanged

Prehistoric behaviour left us no records. We have to look to contemporary ontogenies to observe differences in development. Signing-for-communication by the congenitally deaf infants of signing deaf parents is precocious, while infants with hearing parents and hearing infants with congenitally deaf parents are often slow in signing (Volterra & Erting 1990). Language and social attunement in hearing infants with hearing parents seems little affected by variations in adult gesture (Kirk et al. 2012). Interactional attunement seems critical to infant development (Lundy 2013).

Ontogeny only partially mirrors phylogeny. The communicative environment guides our latent and flexible potential. Our neonatal capacity to cope with, adapt to, and rapidly learn from our social environment is perhaps *the* unique human attribute. We are born with the capacity to develop the language of our parents through their environment. We are socially altricial – our larger cortices enable us to immerse ourselves in and learn through our social environment, to engage our caregivers and to ensure that we are cared for and stimulated. This process is largely articulated well before we develop language (Feldman 2007; Oller et al. 2013).

Foetal brain growth approaches the limits imposed by maternal pelvic size. This is at the cost of the neotenous development and relative vulnerability of most other organ systems. Accelerated early postnatal growth could surely achieve this end to support a cognitive-linguistic system with reduced perinatal risk. The abilities necessary to social survival relies on the intergenerational transmission of adaptability. A dyadic *preverbal* system underpins this process (see Trevarthen & Aitken 2001), but the "second-person neuroscience" required to study its neurobiology is a recent development (Schilbach et al. 2013).

An alternative evolutionary strategy, typified by the social insects, relies on invariance in the social behaviour of its members (see Miller 2010). This can also provide evolutionary success but is less robust in the face of significant environmental change.

FOXP2 – "Human-specific and central to linguistic communication" or a more general key in processing complex information? Forkhead transcription factors are important to a wide range of developmental processes (Carlsson & Mahlapuu 2002; Nudel & Newbury 2013). Interest in FOXP2 came through studying the effects of its mutation on one family pedigree – the KE family (Lai et al. 2001). First referred to as a "developmental verbal dyspraxia" (Hurst et al. 1990), it has also been reported simply as a "dysphasia" (Gopnik 1990b), as a defect in phonology and language-production (Fletcher 1990), and a severe speech disorder affecting all aspects of expressive language (Vargha-Khadem & Passingham 1990). Non-language–related differences have also been reported (Liégeois et al. 2003).

FOXP2 is highly conserved. Only two amino acids differentiate us from orthologs in certain other primates (such as the gorilla and chimpanzee), and three from the orang-utan and mouse. In mice, a defect in *FoxP2* impairs both ultrasonic vocalization (Shu et al. 2005) and motor learning (Groszer et al. 2008), and there are sex differences in gene expression (Bowers et al. 2013). In the human, FOXP transcription differences, including FOXP2, are associated with an increased likelihood of autism spectrum disorders (ASD) (see Bowers & Konopka 2012; Mukamel et al. 2011; Toma et al. 2013).

Is FOXP2 the key "language gene"? A group of genetic factors is reported in association with specific language impairments (SLIs). These include FOXP2, CYP19A1, FOXG1, FOXP1, NRXN1, PCDH11X, PCDH11Y, SETBP1, CNTNAP2, ATP2C2, and CMIP (Deriziotis & Fisher 2013; Marseglia et al. 2012; Newbury et al. 2010; Toma et al. 2013). SLIs are commonly reported in association with ASD (Bowers & Konopka 2012; Chien et al. 2013; Szalontai & Csiszar 2013). In addition, CNTNAP2 KIIA0319/TTRAP/THEM2 mutations have been reported in association with reading disorders (see Newbury et al. 2011; Pinel et al. 2012).

Is the specialised role of FOXP2 confined to human communication? Orthologs of the FOXP2 gene are found across many species, affect vocal communication in many. It is highly conserved and seems likely to have an important function or functions preceding its role in language. The notion of "deep homology" of structural genes in somatotopic development is well known, but its relevance to social behaviour has only recently been suggested (Scharff & Petri 2011). Overly strong parallels to animal models are inappropriate (see Lynch 2009), however knock-in "humanized" *FoxP2* genes in mice have been shown to alter cortico-basal ganglia circuitry (Enard et al. 2009).

FoxP2 is involved in complex non-linguistic systems. It affects birdsong development (Teramitsu et al. 2004), and FoxP2 protein levels alter with the amount of male singing (Miller et al. 2008). In some species of bat, *FoxP2* appears to have evolved in parallel with echolocation (Yin et al. 2008). Here, complex vocalization is used to coordinate flight and prey location (see Metzner & Schuller 2010).

FoxP2 has undergone accelerated evolution in echolocating bats (Li et al. 2007), whales, dolphins (Nery et al. 2013), and humans (Ayub et al. 2013). It has a wider range of functions, across a broader phylogenetic range than was previously appreciated in the brain networks for complex auditory processing. In some species this has served communication, while in others its adaptive function seems more related to complex motor guidance.

Targets in the avian and human brain are now clearer, but their genetic effects and neurochemical cascade are complex (Konopka et al. 2009). To date, some 34 FOXP2 transcription targets have been identified in basal ganglia and inferior frontal cortex alone (Spiteri et al. 2007).

FOXP2 is insufficient to account for the development of human language or its neural and neurochemical substrates. It is a proxy marker for the genetic control of complex biological systems we are only beginning to define or understand.

Comparative analyses of speech and language converge on birds

doi:10.1017/S0140525X13003956

Gabriël J. L. Beckers,[a] Robert C. Berwick,[b] and Johan J. Bolhuis[a]

[a]*Cognitive Neurobiology and Helmholtz Institute, Departments of Psychology and Biology, Utrecht University, 3584 CH Utrecht, The Netherlands;*
[b]*Department of Electrical Engineering and Computer Science and Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139.*
g.j.l.beckers@uu.nl

http://gbeckers.nl
berwick@csail.mit.edu
http://lids.mit.edu/people/faculty/berwick-robert.html
j.j.bolhuis@uu.nl
http://www.bio.uu.nl/behaviour/Bolhuis

Abstract: Unlike nonhuman primates, thousands of bird species have articulatory capabilities that equal or surpass those of humans, and they develop their vocalizations through vocal imitation in a way that is very similar to how human infants learn to speak. An understanding of how speech mechanisms have evolved is therefore unlikely to yield key insights into how the human brain is special.

Ackermann et al.'s efforts to understand the evolution of "brain mechanisms of acoustic communication" focus on neuroanatomical adaptations in nonhuman primates that may have enabled the evolution of articulated speech. Unlike these authors, however, we do not think that an understanding of how articulation evolved in terms of common descent from our primate ancestors must be key to an understanding of "how the human brain is special." Particularly when it comes to speech and language, large-scale patterns of evolutionary convergence provide insights that are at least as important as insights from analyzing recent common descent.

Speech is one possible external interface for human language, and speech-like capabilities per se are not unique to humans or primates but are in fact widespread among species far removed from the primate clade. Ackermann et al. briefly mention songbirds as an experimental model system to study neural control of speech-like behavior, but at least as important is that from a broader comparative view, songbirds also provide important evolutionary insights. Not only do birds have structured, articulated vocalizations, but just like human infants, they acquire these vocalizations through imitation learning, a trait that is rare among mammals and appears to be completely absent in nonhuman primates. In addition, the way in which songbirds learn to sing is very similar to the way that human infants acquire speech. First, in both cases there is a sensitive period during which learning proceeds optimally. Second, developing individuals go through a transitional phase of vocal development, which is called "babbling" in infants and "subsong" in songbirds (Bolhuis & Everaert 2013; Bolhuis et al. 2010). In both species, vocal imitation and learning typically play a large role, though as noted above, in humans, the interface modality can be gestures rather than speech.

Beyond their human-like way of acquiring their vocalizations, many songbird and parrot species also produce highly virtuoso vocalizations, using special adaptations for phonation and articulatory control. Birds have evolved a specialized organ, the syrinx, solely for vocalization, unlike the human larynx. In songbirds, this organ is bipartite, enabling them, for example, to sing with two independent voices at the same time, to use one side for singing and the other side for respiration to avoid running out of breath, or to use one voice for low registers and the other one for high registers. Further, vocal articulation in birds is not restricted to this specialized organ, but also includes fast lingual and oropharyngeal movements that either support voice articulations, or add another layer of complexity on top of it (Beckers 2013; Beckers et al. 2004). In short, there is no question that the vocal capabilities of many species of birds surpass those found in any other clade, including humans.

Vocal virtuosity in birds serves a variety of functions, including the social ones that Ackermann et al. suggest played a role in human speech evolution. Articulatory and vocal imitation capabilities have existed in these large clades for at least 50 million years (Jarvis 2004b), providing ample opportunity for evolutionary tinkering, especially given that birds are very diverse in terms of ecology and behavior. Despite this, none of the many thousands of extant species of vocal learning birds have so far been reported to possess a "special" brain. This comparative result suggests that Ackermann et al. place too much weight on the notion that the evolution of more versatile call-like vocalizations for improved communication in our hominin ancestors played a crucial role in the origin of any traits that may be uniquely human.

Regarding the mechanisms underlying vocal behavior, Ackermann et al. discuss the neural and genetic (*FOXP2*) parallels between humans and nonhuman primates in some detail. Here too, common descent may not be a reliable guiding principle for comparative research, because changes in *FOXP2* are implicated not only in differences between humans and nonhuman primates, but also other mammals (e.g., bats and cetaceans) as well as birds. Songbirds also have a *FOXP2* gene that differs very little from the human variant. Moreover, in zebra finches the *FOXP2* gene is apparently involved in vocalization and vocal learning, as it is in humans (Bolhuis & Everaert 2013; Bolhuis et al. 2010). Additionally, in comparison with humans, songbirds have analogous (and perhaps homologous) brain structures that are involved in vocal production and auditory perception and memory (Bolhuis et al. 2010).

Arguably, an important reason for the uniqueness of the human brain/mind is our capacity for language per se, rather than articulatory competence (Berwick et al. 2013). Given the already strong parallels between humans and songbirds in terms of auditory-vocal imitation learning, and the often remarkable articulatory skills in many avian species, it is reasonable to ask whether songbirds also possess human-like syntactic abilities (Berwick et al. 2011). Recent claims of such syntactic abilities in songbirds (e.g., Abe & Watanabe 2011) have been shown to be based upon flawed experimental methodologies (Beckers et al. 2012). Nevertheless, we argue that the absence of evidence for human-like combinatorial abilities in songbirds does not as of yet constitute evidence of their absence. Should such syntactic capabilities be present in nonhuman animals, songbirds would prove more likely candidates for comparative evolutionary analysis than apes or monkeys. Taken together with the neurocognitive parallels between birdsong and human speech that we have sketched above (see also Berwick et al. 2011; 2013; Bolhuis et al. 2010), this has important consequences for any evolutionary interpretation of speech and language.

Beyond cry and laugh: Toward a multilevel model of language production

doi:10.1017/S0140525X13003968

Marc H. Bornstein[a] and Gianluca Esposito[b,c]

[a]*Child and Family Research, Eunice Kennedy Shriver National Institute of Child Health and Human Development, Rockledge I, Bethesda, MD 20892-7971, USA;* [b]*Department of Psychology and Cognitive Sciences, University of Trento, Trento, 38068, Italy;* [c]*Division of Psychology, Nanyang Technological University, 639798, Singapore.*

Marc_H_Bornstein@nih.gov gianluca.esposito@unitn.it
http://www.cfr.nichd.nih.gov/index.html
http://polorovereto.unitn.it/~esposito/

Abstract: Language production is a multilevel phenomenon, and human capacities to communicate vocally progress from early forms, based on projections of motor cortex to brainstem nuclei, to complex elaborations, mediated by high-order cognition and fostered by socially mediated feedback.

Primates appear to be motorically capable of speaking words insofar as they can articulate sounds and have (in some documented instances) actually articulated "words." For example, rhesus monkeys produce different call types in association with ad hoc visual signals and even switch between call types associated with different signals (Hage & Nieder 2013). So, vocal tract morphology is not the only limitation that accounts for the inability of nonhuman primates to produce even simple verbal utterances.

Contemporary developmental theory and research – language production included – are rooted in systems dynamics of individual-context relations that guide the emergence of behavior and ontogenetic change. Development is associated with dynamic reciprocal relations among structures at multiple levels of organization. Language – *in toto*, comprehension and expression of phonology, morphology, semantics, syntactics, and pragmatics, at least – has such a multilevel organization, extending as it does from the anatomy of the vocal tract through brain-based motor effectors to interpersonal dynamics and on to cultural experience. By focusing on one level of analysis, Ackermann et al.'s hypothesis misses the essential multilevel and developmental nature of vocal production. Bidirectional influences operate across these multiple levels as biological and cognitive systems are nested within individuals, and individuals are nested within complex social and verbal environments. Accordingly, the developmental systems perspective leads away from a singular explanatory focus on organism or on context to how multiple forces, which span from biological pathways to macrolinguistic influences, collaborate in development.

Taking a cue from advances in developmental science, consider two levels above the Ackermann et al.'s focus on vocal tract morphology that may play vital roles in vocal/verbal production. Primates may be lacking in higher-order cognitive-linguistic operations that subserve communicative skills and in social interaction experiences that play key roles in speech development.

Ackermann et al. focus on ontogenetic speech production in interactions between basal ganglia at one end of the spectrum and their cortical targets at the other. Their main argument titularly focuses on the roots and limiting conditions of vocal/verbal production but seems crucially to omit from consideration *comprehension*, which almost by law ontogenetically and cognitively precedes production and therefore places a higher-order limitation on production. The case of human children acquiring language tells us that, outside cry, laugh, and mimicry, "context-restricted" and "context-free" expressions of verbal forms follow comprehension of those forms. Production hardly ever occurs without comprehension as a pre-requisite.

Comprehension *qua* cognition transcends genetic endowment. Ackermann et al. argue that, because primates lacking the (human) *FOXP2* variant cannot even imitate simple speech-like utterances, and because the disruption of this gene in humans gives rise to severe articulatory deficits, it appears warranted to assume that the human variant of this gene locus is pre-requisite to the phylogenetic emergence of articulate speech. From a developmental viewpoint, however, it is well to recall that human babies who are also speechless presumably possess the *FOXP2* gene. Like primates, older infants possess the requisite genetics and neuroanatomy; what they lack, like primates, are cognition and (see below) requisite experience. Here, multilevel development is uncoupled from neuroanatomy and pathology.

Ackermann et al. assert that vocalizations in nonhuman species reflect ontogenetic modifications of acoustic structure rooted in maturation. However, the restriction to maturation again acknowledges only one level of understanding speechlessness in nonhuman primates. Contemporary interactionist models posit that social factors shape human communicative development and early language learning. Communication begins as the product of bidirectional influences between infants and adults. When 9- to 10-month-old English-learning infants experienced a non-native language (Mandarin) through live interactions with adults, television, or audio-only presentations, only those infants who experienced the language through live interactions learned (Kuhl et al. 2003). Similarly, children learn novel verbs during either live interactions or socially contingent video training over video chat, but not during non-contingent video training (Roseberry et al. 2014). Human children's caregivers provide feedback that is vital to infant learning. Furthermore, prospective longitudinal study shows that maternal responsiveness to infants predicts when children achieve various language milestones (Bornstein et al. 1999; Tamis-LeMonda et al. 1996). In human beings, experiences make a telling difference. Recall that 100% of meaningful vocalizations (the lexicon) are learned: Children growing up in Boston learn English-sounding vocal patterns, whereas children growing up in Paris learn French-sounding ones.

These assertions are further supported by understanding what happens when caregivers cannot adequately interpret a child's vocalization and provide adequate feedback. An instructive example occurs when a parent interacts with a child who has a neurological deficit before the deficit is diagnosed, as in the case of children with autism for whom diagnoses are provided after 18–24 months of age. A core deficit of autism occurs in social communication. At least in a subgroup of infants with autism, early vocalizations are atypically produced (Esposito et al. 2013; Sheinkopf et al. 2012), making it challenging for caregivers to interpret (Venuti et al. 2012) and respond to their child in an effective way (Esposito & Venuti 2009).

In summary, Ackermann et al. point to anatomy and neurobiology as rate limiting factors on vocal/verbal production, when it is also the case that language cognition and interactional experience need to be added to neuroanatomical machinery. As language is a multilevel phenomenon, it is good to have one level of the multilevel system better understood, but all levels as well as their interconnectivity need to be analyzed and apprehended. The authors conclude that "birdsong can serve as an experimental model for the investigation of the neural control of human speech" (sect. 6), and this might be the case for the neural control level, but for levels of the complete system above neuroanatomy, including syntactic and semantic aspects of verbal utterances, higher-order cognitions and linguistic experience are requisite. The ultimate goal of the effort here is purportedly to appreciate comprehensively the origins, capacities, and motives of human speech.

The stated aim of Ackermann et al. is to propose phylogenetic stages of speech acquisition which they root in "monosynaptic refinement of corticobulbar tracts and laryngeal elaboration of cortico-striatal motor circuits" (sect. 7, para. 1). This approach leaves untouched virtually all of the higher-order components of mental functioning and social language learning that collaborate in the end state of verbal production.

The evolution of coordinated vocalizations before language

doi:10.1017/S0140525X1300397X

Gregory A. Bryant

Department of Communication, Center for Behavior, Evolution, and Culture, University of California, Los Angeles (UCLA), Los Angeles, CA 90095-1563.
gabryant@ucla.edu
http://gabryant.bol.ucla.edu/

Abstract: Ackermann et al. briefly point out the potential significance of coordinated vocal behavior in the dual pathway model of acoustic communication. Rhythmically entrained and articulated pre-linguistic vocal activity in early hominins might have set the evolutionary stage for later refinements that manifest in modern humans as language-based conversational turn-taking, joint music-making, and other behaviors associated with prosociality.

Ackermann et al. present an excellent overview of the neurocognitive architecture underlying primate vocal production, including a proposal for the evolution of articulated speech in humans. Multiple sources of evidence support the dual pathway model of acoustic communication. The evolution of volitional control over vocalizations might critically involve adaptations for rhythmic entrainment (i.e., a coupling of independent oscillators that have some means of energy transfer between them). Entrained vocal and non-vocal behaviors afford a variety of modern abilities such

as turn-taking in conversation and coordinated music-making, in addition to refinements that lead to the production of speech sounds that interface with the language faculty.

Wilson and Wilson (2005) described an oscillator model of conversational turn-taking where syllable production entrainment allows for efficient interlocutor coordination with minimal gap and overlap in talk. The mechanisms underlying this ability might have been present in the hominin line well before language evolved, and could be closely tied to potential early functions of social signaling including rhythmic musical behavior and dance (Bryant 2013; Hagen & Bryant 2003; Hagen & Hammerstein 2009). Research on error correction mechanisms has revealed several design features of such entrainment mechanisms. Repp (2005) proposed distinct neural systems underlying different kinds of error correction in synchronous tapping. Phase-related adjustments involve dorsal processes controlling action, while ventral perception and planning processes underlie period correction adjustments.

Bispham (2006) and Phillips-Silver et al. (2010) have suggested that behavioral entrainment in humans involves the coupling of perception and action incorporating pre-existing elements of motor control and pulse perception. This coupling is plausibly linked to Ackermann et al.'s first phylogenetic stage including laryngeal elaboration and monosynaptic refinement of corticobulbar tracts. In order to implement proper error correction in improvised contexts of vocal synchrony, volitional control over articulators is necessary. While little comparative work has shown such an ability in nonhuman primates, there is some evidence suggesting control over vocal articulators in gelada baboons, with an ability to control, for example, vocal onset times relative to conspecific vocalizations (Richman 1976). And recently, Perlman et al. (2012) have found that Koko the gorilla exercises breath control in her deliberate play with wind instruments. Other evidence of this sort is certainly forthcoming, and will help us develop an accurate account of the evolutionary precursors to speech production in humans.

Laughter provides a window into the phylogeny of human vocal production as well. Laugh-like vocalizations first appeared prior to the last common ancestor (Davila-Ross et al. 2009), and in humans is likely derived from the breathing patterns exhibited during play activity (Provine 2000). Bryant and Aktipis (2014) found that perceptible proportions of inter-voicing intervals (IVIs) differed systematically between spontaneous and volitional human laughter, and altered versions of the laughs were differentially perceived as being human made, and related to the IVI measures. Specifically, slowed spontaneous laughs were indistinguishable from nonhuman animal calls, while slowed volitional laughs were recognizable as being human produced. These data were interpreted as being evidence for perceptual sensitivity to vocalizations originating from different production machinery – a finding consistent with the dual pathway model presented here by Ackermann et al.

Interestingly, laughter seems to play a role in coordinating conversational timing. Manson et al. (2013) have reported that convergence in speech rate was positively associated with how much interlocutors engaged in co-laughter. While the degree of convergence over a 10-minute conversation predicted cooperative play in an unannounced Prisoner's Dilemma game, the amount of co-laughter did not. The relationship between laughter and speech is not well understood, though evidence suggests that it is integrated to some extent. The placement of laughter in the speech stream follows some linguistic patterns (i.e., a punctuation effect) (Provine 1993), but also manifests itself embedded within words and sentences as well (Bryant 2012). Co-laughter might serve in some capacity to help conversationalists coordinate their talk, and, in early humans, perhaps coordinate other kinds of vocal behavior. Recent work has demonstrated that people can detect in very short co-laughter segments (<2 seconds) whether the co-laughers are acquainted or not (Bryant 2012) suggesting a possible chorusing function.

A surge of recent work is showing how interpersonal synchrony involving entrainment results in cooperative interactions (e.g., Kirschner & Tomasello 2010; Manson et al. 2013; Wiltermuth & Heath 2009), and the effect seems immune to the negative consequences of explicit recognition. That is, when behavior matching is noticed, but does not involve fine temporal coordination, interactants do not respond positively (e.g., Bailenson et al. 2008). Manson et al. (2013) described interpersonal synchrony as a coordination game that does not afford cheating opportunities, unlike mimicry and other behavior matching phenomena where deceptive, manipulative strategies are potentially profitable. Coordinating vocal (and other) behavior provides a means for individuals to assess the fit of others as cooperating partners. Given the extreme cooperative nature of humans relative to other species, mechanisms for such assessment are not surprising, and in fact should be expected.

Taken together, the findings described above point to an important component of human vocal communication that involves the independent and integrated action of emotional vocal production and speech production systems. Selection for articulatory control mechanisms underlying the entrainment of vocal behavior for within- and between-group communicative functions could have set the stage for conversational turn-taking – an ability that incorporated speech. Dual pathway models of acoustic communication should more seriously consider the neurocognitive underpinnings of vocal entrainment abilities and consider these adaptations in the phylogenetic history of human vocal behavior.

Environments organize the verbal brain

doi:10.1017/S0140525X13003981

A. Charles Catania

Department of Psychology, University of Maryland, Baltimore County (UMBC), Baltimore, MD 21250.

catania@umbc.edu

Abstract: FOXP2 expression in the evolution of language derives from its role in allowing vocal articulation that is sensitive to its consequences. The discrete verbal discourse it allows must have evolved recently relative to affective features of vocal behavior such as tone of voice. Because all organ systems must have evolved in the service of behavior, attention is given to ways in which environments may have driven brain organization.

Ackerman et al.'s plausible account of how brain evolution may have led to language would be even more persuasive if it also dealt with how evolutionary environments might have driven brain changes that engendered language. The survival and reproduction of organisms within populations depends on their behavior, so I start from the position that the brain, like all organ systems, evolved in the service of behavior (Catania 2008). For example, brain size may have driven articulatory control, but environments where that articulatory control made a difference must also have driven brain size. Elsewhere I address in more detail these and related issues, including interpretations of learning in terms of selection rather than associations and the distinction between language structure and function (e.g., Catania 1990; 2013a; Catania & Cerutti 1986).

The functional distinction between affective language, as in tone of voice, and substantive language, as in vocal discourse, is illustrated by an account of the different reactions of two audiences to a speech by Ronald Reagan (Sacks 1985). Psychotics without affect responded only to the speech content, whereas aphasics responded only to its affect; only those responsive to both dimensions found the speech persuasive. The affective and the discursive systems involve the same vocal apparatus, so they necessarily evolved in coordination, but Sacks's example demonstrates their separate functionalities. If affective vocal functions are similar to those of other displays usually characterized as emotional, then other functions must have driven evolution of the discursive system: evolution rarely duplicates existing functions.

Emotional behavior substantially predates language, with discursive functions presumably overlaid upon it later. Contented or angry or lustful gorillas do not need new ways to express their emotions.

My candidate for the minimal function of discursive verbal behavior from which all other functions are derived? It is a highly efficient way in which one human can get another to do something (Catania 1995; 2003; 2009). The imperative does not require multiple-word utterances or grammar. Even if nothing else is available, a single utterance functionally equivalent to the command *Stop!* will benefit the members of any hominid group that creates it. Other functions (e.g., communication, narrative) are derivatives of this fundamental one. For example, prestige matters only when some individuals become more important in telling others what to do; cooperation can sometimes be more effectively induced through verbal instructions than by other means. Telling others what to do leads to telling them what to say, and giving information provides expanded ways to tell others how to do things.

The long period during which our ancestors made tools and mastered fire suggests that some form of hominid language has existed for perhaps millions of years, whereas archaeological findings coupled with inferences from linguistic change and human migration implies a source perhaps as recent as 40 to 50 thousand years ago. The longer time makes sense if we include the evolution of affective and single-utterance precursors with simple imperative functions, accompanied by more sophisticated articulations. Significant anatomical developments included bipedal locomotion, freeing respiration from constraints on the rib cage, and elaborations of vocal signals such as laughter (Provine 2000; 2012). But defining language solely in terms of syntactically organized multi-word utterances targets the more recent provenance. The step from single- to multiword utterances with different words having different functions allows for an explosion of language diversity.

Coherent accounts of language evolution must include three concurrent levels of selection (Catania 2001), each entailing different mechanisms by which environments select surviving variants. First, phylogenetic (Darwinian) contingencies must select requisite physiological attributes (e.g., vocal tract structure, neural organization). Second, ontogenetic contingencies (selection of behavior within individual lifetimes) must maintain those language features acquired by individuals, as when native but not non-native speech sounds survive in a child's developing vocalizations. Third, cultural or memetic selection (selection of behavior as it passes among individuals) must perpetuate languages across generations as communities pass them on from one to another.

Ackerman et al. seek an account for the co-evolution of articulatory and perceptual skills. Yet if distributions of both skill levels exist within a population, those at the upper ranges of either skill will be selected. As long as selection operates relative to each population mean, the distributions will change together, just as but more benignly than in the arms races of predators and their prey. For example, mothers with superior acuity along some auditory dimension will sometimes bear offspring with superior differentiation along some articulatory dimension; they and their offspring will both be selected, just as more successful predators are selected as their predation selects prey more successful at escape.

Ackerman et al. cite many relevant studies but provide no taxonomy of relevant processes. We read of reinforcement, goal-directed behavior, instrumental conditioning, learning responses to food rewards, acquisition of stimulus-driven behavioral routines, habit formation, training, and even motor tricks. But these are simply alternative vocabularies for labeling behavior changes that occur because behavior is modified by its consequences (Catania 2013a; Schneider 2012). Consequences are as much involved in stimulus-driven behavioral routines, where responses produce different consequences given different stimuli, as in refinement of skills, where differential consequences shape behavior. Existing taxonomies of behavioral processes could put much of this in good order. They are built not upon associations or conditioning but rather upon the selection of behavior by its consequences (e.g., Catania 2013a; 2013b; Madden 2012; Verplanck 2000). The structure of behavior provides crucial clues for what to look for in the brain.

Nevertheless, Ackerman et al. have made a strong case that FOXP2's expression is prerequisite for the complex vocal articulations of human language. Timing is critical, so these coordinations involve not just tongue, lips, and larynx, but also diaphragm and rib cage. FOXP2's expression may work by allowing motor patterns that are otherwise highly constrained by anatomies and stimuli to be modified by their consequences. Skinner apparently got it right when he wrote: "The human species took a crucial step forward when its vocal musculature came under operant control in the production of speech sounds. Indeed, it is possible that all the distinctive achievements of the species can be traced to that one genetic change" (Skinner 1986, p. 117). In nonverbal organisms, reinforcers can alter only the rate of vocalizations (e.g., cheeps of chicks; Lane 1961); in children, however, their form is sculpted even by such subtle differential consequences as producing sounds more or less resembling those of caregivers (Risley 1977; Vihman 1996). This is as it should be because, as Ackerman et al. so effectively point out, our vocal articulations are perhaps the most sophisticated of human achievements.

Evolution of affective and linguistic disambiguation under social eavesdropping pressures

doi:10.1017/S0140525X13003993

Kevin B. Clark[a,b]

[a]*Research and Development Service, Veterans Affairs Greater Los Angeles Healthcare System, Los Angeles, CA 90073;* [b]*Complex Biological Systems Alliance, North Andover, MA 01845.*

kbclarkphd@yahoo.com kevin.clark@cbsaimtt.com

www.linkedin.com/pub/kevin-clark/58/67/19a

Abstract: Contradicting new dual-pathway models of language evolution, cortico-striatal-thalamic circuitry disambiguate uncertainties in affective prosody and propositional linguistic content of language production and comprehension, predictably setting limits on useful complexity of articulate phonic and/or signed speech. Such limits likely evolved to ensure public information is discriminated by intended communicants and safeguarded against the ecological pressures of social eavesdropping within and across phylogenetic boundaries.

The basal ganglia contribute to acquisition, planning, initiation, and execution of vocal and gestural communication skills in primates, birds, and other animals. Consistent with dual-pathway models of language evolution, Ackermann et al. in the target article now speculate the basal ganglia also integrate and modulate (continuous or analog) affective prosody of vocalizations and gesticulations with little to no influence over (discrete or digital) propositional linguistic content of human phonetic and, presumably, signed speech. The authors cite comparative clinical and basic research findings to support their claim that high-level linguistic processing only occurs in phylogenetically newer brain systems, while omitting the recent small, but credible, neuroimaging literature which contradicts this assertion and implicates human cortico-striatal-thalamic circuitry in disambiguating lexical (Chenery et al. 2008; Copeland 2003), grammatical (Mestres-Missé et al. 2012), and semantic (Ketteler et al. 2008; Marques et al. 2009; Wittforth et al. 2010) uncertainties in perceived language. Failure to assimilate roles of the basal ganglia in both language production and comprehension seriously

weakens the conceptual validity and power of Ackermann et al.'s treatise on selective fitness of advancing animal taxa to evolve increasingly sophisticated dual-pathway communication systems for affective and propositional information exchange.

Evolutionarily older functions of cortico-striatal-thalamic loops to generate and filter variances in affective prosody of non- and/or protolinguistic species-typical/atypical communications, as advocated by Ackermann et al., seem to have eventually and adaptively converged to help perform similar operations on propositional linguistic content, as evidenced in later human language use. Such (lateralized) developments in cortico-striatal-thalamic processing necessarily first enabled language-deficient nonhuman animals to better articulate innate and/or learned primitive communications (e.g., recombinant hierarchical call or song sequences with precise, intricate spectral patterns) and, therefore, to more successfully transmit meanings or labels of both continuously and discretely structured information for receiver understanding (Arnold & Zuberbühler 2006; Berwick et al. 2011; Bolhuis et al. 2010; Doupe et al. 2005; Ouattara et al. 2009; Zuberbühler 2000a; Zuberbühler et al. 1999). Despite lack of direct empirical proof, one can further safely reason that homologous or analogous neuromechanisms for disambiguating communication content arose from ecological forces that continue to drive changes in production, comprehension, and privatization of public vocal and gestural communications ancestral to and descendent from early hominin language innovations.

Capacities of cortico-striatal-thalamic pathways to regulate variability in communication production and comprehension likely coevolved with animal abilities to encrypt and decrypt sensitive public information at risk of corruption or interception from social eavesdroppers. Evolution conserved social eavesdropping across phylogeny, whereby unintended observers breach information security of communicating parties in attempts to gain survival and/or reproductive advantages (Clark 2010; 2013a; 2013b; in press; Dabelsteen 2004; Dall 2005; Danchin et al. 2004; Joint 2006; Peake & McGregor 2004; Seyfarth & Cheney 2010; Stowe et al. 1995). Cortico-striatal-thalamic circuitry, via involvement in automatic and/or volitional processing of affective and propositional content variability, predictably sets limits on useful complexity of naturally communicated information. These constraints determine probabilities that public exchanges may be discriminated by intended observers and safeguarded against social eavesdroppers. When communication complexity processed by phylogenetically or culturally distant unintended observers far subtends upper complexity limits for information processed over superior disambiguation neuromechanisms of intended observers, information content of public messages and replies will remain protected from eavesdropping. Complexity scaling of communication production and comprehension extends along the continuum of signals to protolanguage to language and figures to be an essential evolutionary strategy to secure communications within and across taxonomic boundaries.

One may begin to appreciate evolved neurobiological barriers to social eavesdropping by enlisting examples of dual-pathway systems for birdsong and human speech given by Ackermann et al. The cortico-striatal-thalamic circuitry of birds and humans effect complexity scaling through two broad, related domains of complexity – combinatorial and computational complexity – each having particular significance for communication production and comprehension as well as for other aspects of cognition (Clark 2012). Classical combinatorial complexity differentiates levels of comparative language hierarchies and communication repertoires (Changizi 2001; Chomsky 1956; 1966; McNaughton & Papert 1971), where complexity is proportional to number of discrete information elements, length of composite information sequences, and structure of recursive information patterns. Useful complexity under these conditions is defined by strictly ordered inclusive sets of information capable of being both generated and recognized with certain classical computational models, machines, or grammar rules emulating properties of cortico-striatal-thalamic loops. Three fundamental features of all computational complexity classes may be varied – computational resources (e.g., time, space), problem type to be solved (e.g., optimization or decision problem, language production and comprehension), and computational model to be employed (e.g., deterministic Turing Machine, probabilistic Turing Machine, quantum computer) (Clark 2012). Disparities in classical communication complexities between birds and humans reveal dissociations for each computational feature and, consequently, for communication disambiguation involving affective prosodic or propositional information content (Berwick et al. 2011). As disparities narrow and computational features progressively overlap, threats of eavesdropping on public information should escalate for superior communicants, in this case humans.

More instructive scenarios, and ones that help identify flaws in purely classical complexity approaches toward language evolution, concern competing, closely related animals, such as bird or primate subspecies, with very similar communication complexities. Pressures of social eavesdropping rise when quality and/or quantity of niche resources dwindle and acquired public information facilitates selection and acquisition of preferred life necessities shared by conspecifics. Subspecies communication adaptations, including genetically and/or culturally acquired vocal dialects and behavioral modifications (Dabelsteen 2004; Danchin et al. 2004) processed via cortico-striatal-thalamic pathways, increase degrees of freedom for classical information computation, further privatizing public information readily comprehended by conspecifics. However, when disambiguation demands for processing linguistic variations superposed (or nearing maximal entanglement) with affective prosodic variations grow exponentially with information input size, privatization becomes governed by quantum computational models involving the entropic uncertainty principle for indistinguishable communications content (Clark 2012; in press; Nielsen & Chuang 2000). This principle imposes thresholds above which eavesdroppers with inferior, missing, or over-allocated communication disambiguation neuromechanisms cannot definitely and simultaneously decrypt partite affective and linguistic content of public information. However, intended communicants may violate the principle by enhancing public information security through privy subspecies-specific communication and memory specializations (cf. Bennett et al. 1993; Berta et al. 2010).

Physical mechanisms may be as important as brain mechanisms in evolution of speech

doi:10.1017/S0140525X13004007

Bart de Boer[a] and Marcus Perlman[b]

[a]*Artificial Intelligence Lab, Vrije Universiteit Brussel, 1050 Brussels, Belgium;* [b]*Department of Cognitive and Information Sciences, University of California–Merced, Merced, CA 95343.*

bart@ai.vub.ac.be mperlman@ucmerced.edu

http://ai.vub.ac.be/members/bart

Abstract: We present two arguments why physical adaptations for vocalization may be as important as neural adaptations. First, fine control over vocalization is not easy for physical reasons, and modern humans may be exceptional. Second, we present an example of a gorilla that shows rudimentary voluntary control over vocalization, indicating that some neural control is already shared with great apes.

Ackermann et al. propose a model of the evolution of neural adaptations related to the production of spoken language. Although we are convinced of the importance of such adaptations, and although the authors themselves state that "the model outlined here addresses only one out of several building blocks" (sect. 7, para. 2), we would nevertheless like to make two reflections on

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 surpralaryngeal 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
daniellepersyzk2017@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.

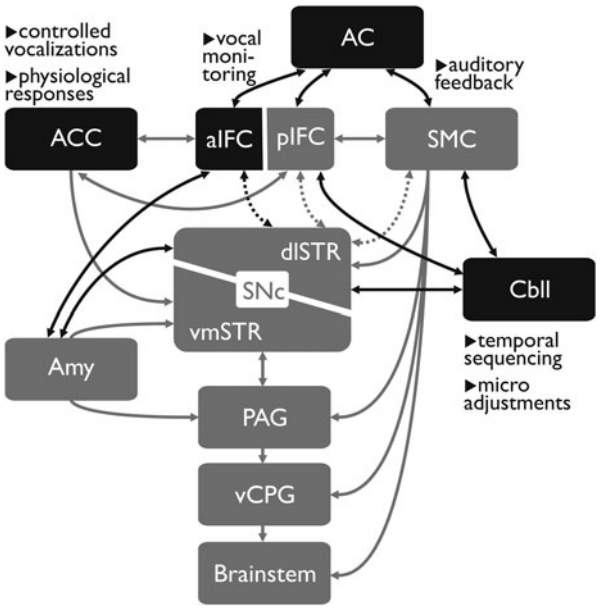

Figure 1 (Frühholz et al.). Suggested extension (black regions and arrows) of Ackermann et al.'s original model (gray regions) beyond the affective (i.e., amygdala) and motor systems. Based on the paradigm of affective vocalizations and emotional speech, we suggest adding the AC and anterior IFC (aIFC), which serve auditory–motor feedback processing and vocal monitoring; the CbII, which serves online micro and macro adjustments of vocal motor output; and the ACC, which appears to be directly involved in controlling vocal output and physiological responses.

Recent neuroimaging evidence also points to two brain structures active during human vocalizations, which are not yet (explicitly) included in the model. As mentioned above, vocalizations strongly depend on auditory feedback for online adjustments and corrections. Accordingly, studies consistently report activity in low- and high-level regions of the auditory cortex (AC) (Aziz-Zadeh et al. 2010; Pichon & Kell 2013), and in the cerebellum (Laukka et al. 2011; Pichon & Kell 2013; Wattendorf et al. 2013). While the AC together with the IFC is thought to serve auditory feedback processing and vocal monitoring, the cerebellum mainly supports online macro- (Pichon & Kell 2013) and micro-adjustments (Wattendorf et al. 2013) of vocal motor behavior.

Concerning the AC feedback-related activity, the online validation of the vocal performance seems critical for vocal expressions. Affective vocalizations for successful social communication depend on a proper vocal production, especially in terms of temporo-dynamic features (Patel et al. 2011). The temporal slow prosodic modulations of emotional speech, in particular, seem to rely on feedback processing in the AC (Aziz-Zadeh et al. 2010; Pichon & Kell 2013). A major part of the slow prosodic modulations is determined by temporal variations of the fundamental frequency, which mainly contribute to the perception of pitch variations. This perceived temporal pitch variations of one's own vocalizations considerably activates the AC, and, surprisingly, also the cerebellum (Pichon & Kell 2013).

Although the cerebellum was a core element in a former model proposed by Ackermann (2008), in the present article Ackermann et al. note that it is not relevant here. However, given the above-mentioned evidence that the cerebellum is related to slow temporal modulations in affective speech (Pichon & Kell 2013), and given the general observation that non-speech (primate-general) and speech-based affective vocalizations (human-specific) considerably activate the cerebellum (Laukka et al. 2011; Wattendorf et al. 2013), we propose that the cerebellum should be an integral part of a neural model of vocal communication. It seems that for emotional vocalizations, the cerebellum supports the online micro adjustment of ongoing motor responses (Wattendorf et al. 2013) and provides a macro temporal event structure (Kotz & Schwartze 2010) for the temporal dynamics embedded in emotional speech. Both are important ingredients for valid affective vocalizations in terms of vocal motor responses (Patel et al. 2011).

Overall, from the perspective of affective vocalizations and emotional speech, neuroimaging evidence supports the neural model of Ackermann et al., but also suggests that the model might be extended to include auditory–motor feedback loops and online adjustment of vocal behavior (Fig. 1). The paradigm of human affective vocalizations thus might be a valid example for a cross-validation of the model proposed by Ackermann et al., because affective vocalizations are an essential ingredient of human communication.

ACKNOWLEDGMENTS

Sascha Frühholz and Didier Grandjean were supported by grants from the Swiss National Science Foundation (105314_146559/1 and 105314_124572/1) and the NCCR in Affective Sciences (51NF40-104897).

Functions of the cortico-basal ganglia circuits for spoken language may extend beyond emotional-affective modulation in adults

doi:10.1017/S0140525X13004032

Takashi Hanakawa[a,b] and Chihiro Hosoda[a,b]

[a]*Department of Advanced Neuroimaging, Integrative Brain Imaging Center, National Center of Neurology and Psychiatry, Kodaira 187-8551, Japan;* [b]*PRESTO, Japan Science and Technology Agency, Kawaguchi, Saitama 332-0012, Japan.*

hanakawa@ncnp.go.jp
http://researchmap.jp/takashihanakawa/
chihiro.hosoda@mail.u-tokyo.ac.jp
http://researchmap.jp/chihiro/

Abstract: We support Ackermann et al.'s proposal that the cortico-basal ganglia circuits may play essential roles in the evolution of spoken language. Here we discuss further evidence indicating that the cortico-basal ganglia circuits may contribute to various aspects of spoken language including planning, learning, and controlling of speech in adulthood.

Ackermann et al. have proposed a two-stage neural control model underlying phylogenetic and ontogenetic evolutions of spoken language. Neural machinery at one stage depends upon the development of monosynaptic projections from the motor cortex to cranial nerve nuclei in the brainstem and the other one involves functions of the cortico-basal ganglia circuits. We appreciate this proposal because we have been interested in the contribution of the cortico-basal ganglia circuits to language and associated abilities in humans. Here we want to extend the authors' view, by arguing for potential roles of the cortico-basal ganglia circuits in various aspects of spoken language in adults.

Accumulating evidence indicates that the basal ganglia participate in speech control in humans. However, the roles of the basal ganglia in language control are still unclear. The functionality of the basal ganglia for spoken language perhaps extends beyond the modulation of laryngeal and orofacial movements. We previously showed basal ganglia activity during a cognitive task involving verbal motor imagery, or "inner speech," in healthy adults (Hanakawa et al. 2002). This basal ganglia activity was accompanied by activity in other speech-related brain regions such as supplementary motor area and frontal opercular regions. Moreover, we reported that performance of this verbal imagery task was impaired in patients with basal ganglia dysfunctions (Parkinson's disease) in comparison with matched control participants (Sawamoto et al. 2002). A neuroimaging experiment

supported that the impaired performance of the verbal imagery task in Parkinson's disease was associated with dysfunctions of the basal ganglia, the caudate nucleus in particular (Sawamoto et al. 2007). Considering that motor imagery is closely related to motor planning (Hanakawa et al. 2008), the contribution of the basal ganglia to spoken language likely involves a planning stage of speech.

Of even more importance is to understand the contribution of the basal ganglia to learning of spoken language. Ackermann et al. propose a fundamentally different role of the basal ganglia at ontogenetic stages: acquisition of articulatory motor patterns during childhood versus emotive-prosodic modulation of verbal utterances during adulthood. We want to modify and extend this view, especially with regard to the contrast between childhood and adulthood stages. The neural underpinnings for the native language development are difficult to study experimentally. Therefore, we want to argue for the role of the basal ganglia in speech acquisition in adults, taking the case of second language (L2) learning as an example.

We recently conducted a cohort study in which Japanese university students were enrolled in a 16-week e-learning program to develop their English vocabulary (Hosoda et al. 2013). Although the training program involved various aspects of vocabulary learning, an emphasis was placed upon the training of pronunciation. The students learned 60 words or idioms in each week. An example sentence for each word and idiom was also presented. The participants were encouraged to dictate each word, idiom, and sentence 10 times in reference to "speech templates" provided by the program. By repeating after the speech templates, the participants were to compare their own utterances and the speech templates, and then try to make corrections to his or her motor programs for pronunciation. Speculatively, this auditory feedback learning should help the trainees achieve adequate spatio-temporal control of laryngeal and orofacial musculature. After 16 weeks, the trainees showed approximately 30% improvement in a test battery of English competence. We performed multidimensional imaging assessment for neuroplastic changes associated with the training. Most notably, probabilistic diffusion tractography demonstrated that connectivity between the inferior frontal gyrus and the caudate nucleus, an input station of the basal ganglia, was enhanced in correlation with the improvement in the trainees' L2 competence. This study has provided the first evidence that the cortico-basal ganglia circuits are involved in language learning in adults. Furthermore, the learning-induced enhancement of the cortico-basal ganglia connectivity was accompanied by enhanced connectivity between the inferior frontal gyrus and superior temporal/supramarginal gyrus (dorsal pathway), but not between the inferior frontal gyrus and middle temporal gyrus (ventral pathway). The dorsal pathway primarily concerns phonological aspects of language control. Hence, the selective involvement of the dorsal pathway indicated that our training program primarily tapped into phonological aspects of L2 vocabulary.

According to the findings in our study (Hosoda et al. 2013), we suggest a possibility that the basal ganglia may contribute to learning of spoken language even in adults. Speech is acquired through experiences of adequate auditory inputs, which is evident in children with hearing loss (Tye-Murray et al. 1995). In addition, we suspect that reinforcement-type learning (Demirezen 1988) subserved by functions of the cortico-basal ganglia circuits may underlie experience-based shaping of spoken language. To improve speech control, it is reasonable for both child and adult learners to rely on information about the success or failure of their speech production. We speculate that the feedback information could be self-generated in adult learners who are enrolled in an e-learning program or be given by family and community members as praise or approval to children. Feedback information indicating successful speech production can be utilized as a positive reinforcer to strengthen the neural circuits a trainee had just activated. The striatum that receives both contextual information from the cortex and reward signals from dopaminergic neurons occupies the best position for reinforcement learning or reward-based temporal difference learning (Doya 2008). It would be extremely interesting to figure out the learning stage at which genetic predispositions such as *FOXP2* play fundamental roles.

Other studies in bilinguals have shown that the caudate nucleus is important for monitoring and controlling of the two languages in use (Crinion et al. 2006; Hernandez et al. 2001; Hosoda et al. 2012).

In conclusion, we generally warrant Ackermann et al.'s proposal that the cortico-basal ganglia circuits may play essential roles in evolutions of spoken language. We, however, consider that the cortico-basal ganglia circuits may contribute to various aspects of spoken language including planning, learning, and controlling of speech in both childhood and adulthood.

Does it talk the talk? On the role of basal ganglia in emotive speech processing

doi:10.1017/S0140525X13004044

Uri Hasson,[a] Daniel A. Llano,[b] Gabriele Miceli,[a] and Anthony Steven Dick[c]

[a]*Center for Mind/Brain Sciences (CIMeC) and Department of Psychology and Cognitive Science, University of Trento, Mattarello (TN), Italy;* [b]*School of Molecular and Cellular Biology, University of Illinois at Urbana-Champaign, Urbana, IL 61801;* [c]*Department of Psychology, Florida International University, Miami, FL 33199.*

uri.hasson@unitn.it d-llano@illinois.edu gabriele.miceli@unitn.it adick@fiu.edu

http://www.hasson.org

http://mcb.illinois.edu/faculty/profile/d-llano/

http://www.unitn.it/en/cimec/11706/gabriele-miceli

http://faculty.fiu.edu/~adick

Abstract: Ackermann et al.'s phylogenetic account of speech argues that the basal ganglia imbue speech with emotive content. However, a body of work on auditory/emotive processing is inconsistent with attributing this function exclusively to these structures. The account further overlooks the possibility that the emotion-integration function may be at least in part mediated by the cortico-ponto-cerebellar system.

Ackermann et al.'s phylogenetic account of speech development hinges, in part, on premises related to the role of basal ganglia (BG) in adult human speech production. It argues that in adults, BG imbue speech with emotive content. While the model targets an important and neglected issue, we argue that it suffers from two structural weaknesses: First, it does not sufficiently consider studies of the role of BG in auditory and emotive processing such as those showing that BG damage does not disrupt emotive processing in speech. Second, the argument also overlooks the possibility that the role attributed to the BG may be at least in part mediated by a different system – the cortico-ponto-cerebellar system. We believe the authors' account would be much strengthened if they address these points, which we detail in turn.

Viability of BG as a speech/emotion synthesizer. A principle incorporated in contemporary models of speech production is that production occurs under one or more levels of feedback, where potential production errors are monitored either after utterance production (sensory feedback) or prior to it (via internal models; e.g., Hickok 2012). Ackermann et al. do not couch their account in an existing speech-production model and leave the issue of feedback underspecified. Nonetheless, if the BG were responsible for imbuing speech with emotive content, they would be expected to have the capacity to monitor and correct for related errors, that is, evaluate that the intended emotive tone/prosody was instantiated. However, BG are a weak candidate for such a function. The authors ignore studies indicating (i) that the auditory response in BG is temporally insufficient to provide feedback (Langers & Melcher 2011) and that it has limited functional connectivity

with areas of the temporal cortex mediating language processing (Choi et al. 2012); (ii) that emotive speech processing is mediated mainly by lateral temporal systems while excluding the BG (Kotz et al. 2013; Wildgruber et al. 2006); and, most importantly, (iii) that individuals with BG infarcts are equally sensitive to emotional speech variations as control populations (Paulmann et al. 2008; 2011). These three points argue against the authors' claim that adding prosody to speech depends on integrity of striatum.

The suggested account relies on two additional premises that are not strongly supported by the literature: The first, that in adults, the BG can afford coding for emotion since adult perisylvian regions code for syllable motor programs, independently of the BG. Empirical support for this point is tenuous at best: Studies using manipulations of syllable frequency have either reported null results (Brendel et al. 2011; Riecker et al. 2008) or documented effects in the anterior insula (Carreiras et al. 2006). The second, that the BG can merge emotional content due to cross talk between cortico-striatal-thalamic circuits. Although there is anatomical evidence for cross-talk across BG circuits in animal models (Haber 2003), the functional significance of these needs to be fleshed out.

On the consideration of alternatives. A BG-oriented account should address questions such as those raised above, and equally importantly argue why the BG is the strongest neurobiological candidate for mediating the function in question. The authors do not make such an argument, which is unfortunate since much of the neurobiological argument made here for BG could be made effectively for other structures, such as the cerebellum.

The involvement of the cerebellum in emotional processing is well established. It is implicated in self-generation of various emotional states (Damasio et al. 2000), with different emotions evoking distinct activity patterns in the structure (Baumann & Mattingley 2012). Damage to the cerebellum affects emotional processing. In animal models, early cerebellar lesions can lead to disrupted emotional processing (Bobee et al. 2000), and in human adults, the Cerebellar Cognitive Affective Syndrome (CCAS; Schmahmann & Sherman 1998) is a recognized clinical entity associated with blunting of affect. CCAS has been attributed to damage to the posterior vermis, which reduces the cerebellar contribution to perisylvian cortical areas via its outflow to the ventral tier thalamic nuclei (Stoodley & Schmahmann 2010).

Arguments used by Ackermann et al. in support of their BG hypothesis could also be applied to the cerebellum. For example, FOXP2 expression is found in the cerebellum as well as the caudate (Lai et al. 2003; Watkins et al. 2002b), and as shown by Ackermann et al. (1992), cerebellar lesions are associated with dysarthia. In addition, activity in the cerebellum, but not BG, discriminates emotive aspects of speech (Kotz et al. 2013). Furthermore, the cerebellum has the capacity for generating an internal forward model of motor-to-auditory predictions of the sort needed to evaluate whether the intended emotive aspect has been communicated (Knolle et al. 2013). While there is no direct examination of this issue for BG, work on motor control suggests that functionally, BG may implement open- rather than closed-loop control of motor actions (Gabrieli et al. 1997).

It is important to point out that these explanations are not mutually exclusive. Cerebellar and BG circuits involved with language converge at the ventral anterior nucleus of the thalamus, which has also been implicated in language, and can serve as a nidus for cortical feedback via cortico-thalamic projections (Crosson 2013). Further, cerebellar outflow can directly influence the BG, and vice versa (Bostan et al. 2013), suggesting that attributing the emotional content of speech to either of these two systems in isolation may not be possible. Given this connectivity, it may be that the cerebellum drives emotion-carrying vocalizations by involving BG, or that the BG trigger emotional behavior that is ultimately modulated by the cerebellum, as would be consistent with a CCAS syndrome. However, data on this issue are lacking.

Summary. Arguing that the BG can imbue speech with emotional content is a significant claim and, as such, requires additional evidence, accompanied by careful consideration of alternative accounts. We hope this commentary will result in more detailed examination of the aforementioned issues.

Differences in auditory timing between human and nonhuman primates

doi:10.1017/S0140525X13004056

Henkjan Honing[a] and Hugo Merchant[b]

[a]*Amsterdam Brain and Cognition, Institute for Logic, Language and Computation, University of Amsterdam, Amsterdam, The Netherlands;* [b]*Department of Cognitive Neuroscience, Instituto de Neurobiología, Universidad Nacional Autónoma de México, Campus Juriquila, Querétaro, México.*

honing@uva.nl hugomerchant@unam.mx
http://www.mcg.uva.nl/hh/
http://132.248.142.13/personal/merchant/members.html

Abstract: The gradual audiomotor evolution hypothesis is proposed as an alternative interpretation to the auditory timing mechanisms discussed in Ackermann et al.'s article. This hypothesis accommodates the fact that the performance of nonhuman primates is comparable to humans in single-interval tasks (such as interval reproduction, categorization, and interception), but shows differences in multiple-interval tasks (such as entrainment, synchronization, and continuation).

Ackermann et al. propose that the monosynaptic elaboration of the corticobulbar tracts, which played a selective role in the origins of speech, might also have provided the phylogenetic basis for "communicative musicality" (sect. 5.1). The term "musicality" is used here to indicate the cognitive and biological mechanisms that underlie the perception and production of music, as opposed to musical activities that are shaped by culture (Honing & Ploeger 2012; Honing et al, in press b). Perceiving a regular pulse – the beat – in music is considered a fundamental component of musicality: It allows humans to dance and make music together. This skill has been referred to as beat perception and synchronization (Patel 2008), beat induction (Honing 2012), or pulse perception and entrainment (Fitch 2013). Furthermore, it is considered a spontaneously developing (Winkler et al. 2009), music-specific (Patel 2008) and species-specific skill (Fitch 2013).

Interestingly, beat perception and synchronization (BPS) has been observed in humans and a selected group of bird species (Hasegawa et al. 2011; Patel et al. 2009b), but appears to show some but not all the behavioral finger prints in nonhuman primates (Honing et al. 2012; Zarco et al. 2009; but see Hattori et al. [2013] for some counter-evidence). This observation is in support of the vocal learning (VL) hypothesis (Patel 2008), which suggests that BPS is a by-product of the VL mechanisms that are shared by several bird and mammal species, including humans, but that are only weakly developed, or missing entirely, in nonhuman primates. Nevertheless it has to be noted that, since no evidence of rhythmic entrainment was found in many vocal learners (including dolphins, seals, and songbirds; Schachner et al. 2009), vocal learning may be necessary, but clearly is not sufficient for BPS. Furthermore, recent evidence for BPS in a non-vocal learner (Cook et al. 2013) weakens vocal learning as a pre-condition for rhythmic entrainment.

The absence of synchronized movements to sound (or music) in certain species is no evidence for the absence of beat perception. With behavioral methods that rely on overt motoric responses (e.g., Hattori et al. 2013; Patel et al. 2009b) it is difficult to distinguish between the contribution of perception and action; more direct, electrophysiological measures such as event-related brain potentials (ERPs) allow testing for neural correlates of beat perception (a pre-condition to rhythmic entrainment). To test this, we measured auditory ERPs in rhesus monkeys (*Macaca mulatta*) using the mismatch negativity (MMN) component as an index of (the violation of) rhythmic expectation (Honing

et al. 2012). Rhythmic expectation was probed by selectively omitting parts of a musical rhythm, randomly inserting gaps at the first position of a musical unit (i.e., the "downbeat"). This oddball paradigm was used previously to probe beat perception in human adults and newborns (Honing et al., in press a; Winkler et al. 2009). The results confirmed the behavioral studies discussed earlier, in that rhesus monkeys are not able to detect the beat in a complex auditory stimulus, although they can detect the start of a rhythmic group (Honing et al. 2012). In fact, a recent paper showed that macaques exhibit changes of gaze and facial expressions when a deviant of a regular rhythmic sequence is presented, supporting the notion that monkeys are sensitive to the structure of simple rhythms (Selezneva et al. 2013).

The question remains of whether more close human relatives, such as the great apes, show a more sophisticated ability for rhythmic entrainment than macaques. While the VL hypothesis predicts that no rhythmic entrainment should be found, a recent study (Hattori et al. 2013) showed that at least one chimpanzee (*Pan troglodytes*), of the three that took part in the experiment, was capable of spontaneously synchronizing her movements with an auditory rhythm. Interestingly, this chimpanzee entrained her tapping behavior to an isochronous 600-msec interval stimuli metronome, but not to other tempos.

Based on these observations, we propose an alternative view: the gradual audiomotor evolution (GAE) hypothesis (Honing et al. 2012; Merchant & Honing 2014), which directly addresses the similarities and differences that are found between human and nonhuman primates (discussed in section 5.1 of the target article). This hypothesis suggests rhythmic entrainment (or beat-based timing) to be gradually developed in primates, peaking in humans but present only with limited properties in other nonhuman primates; while humans share interval-based timing with all nonhuman primates and related species. Thus, the GAE hypothesis accommodates the fact that the performance of rhesus monkeys is comparable to humans in single-interval tasks (such as interval reproduction, categorization, and interception; Mendez et al. 2011; Merchant et al. 2003), but differs substantively in multiple-interval tasks (such as rhythmic entrainment, synchronization, and continuation; Zarco et al. 2009).

Finally, the GAE and VL hypotheses show the following crucial differences. First, the GAE hypothesis does not claim that the neural circuit that is engaged in rhythmic entrainment is deeply linked to vocal perception, production, and learning, even if some overlap between the circuits exists. Second, the GAE hypothesis suggests that rhythmic entrainment could have developed through a gradient of anatomofunctional changes on the interval-based mechanism to generate an additional beat-based mechanism, instead of claiming a categorical jump from non-rhythmic/single-interval to rhythmic entrainment/multiple-interval abilities. Third, since the *cortico-basal ganglia-thalamic* (CBGT) circuit has been involved in beat-based mechanisms in imaging studies (Grahn & Brett 2007; Rao et al. 1997; Teki et al. 2011; Wiener et al. 2010), we suggest that the reverberant flow of audiomotor information that loops across the anterior prefrontal CBGT circuits may be the underpinning of human rhythmic entrainment. Finally, the GAE hypothesis suggests that the integration of sensorimotor information throughout the mCBGT circuit and other brain areas during the perception or execution of single intervals is similar in human and nonhuman primates.

Neanderthals did speak, but *FOXP2* doesn't prove it

doi:10.1017/S0140525X13004068

Sverker Johansson
Dalarna University, Falun, SE-791 88, Sweden.
sja@du.se
http://users.du.se/~sja/

Abstract: Ackermann et al. treat both genetic and paleoanthropological data too superficially to support their conclusions. The case of *FOXP2* and Neanderthals is a prime example, which I will comment on in some detail; the issues are much more complex than they appear in Ackermann et al.

Ackermann et al. provide some interesting speculations about a possible scenario for the evolution of the brain mechanisms of vocal communication and language. But in the areas that I am familiar with, notably Neanderthal language (Johansson 2013), but also the history of the human language capacity in general (Johansson 2005; 2011), their treatment of the evidence is superficial and simplistic (see sect. 5.2), leading to their drawing conclusions that are insufficiently supported.

The authors' Section 5 supposedly provides "paleoanthropological perspectives" on their scenario, but contains little reference to paleoanthropological data. Instead it deals mainly with *FOXP2*, with fossil DNA virtually the only paleo-connection.

When mutations in the gene FOXP2 were found to be associated with specific language impairment (Lai et al. 2001), and it was shown that the gene had changed along the human lineage (Enard et al. 2002), it was heralded as a "language gene." But intensive research has revealed a more complex story, with FOXP2 controlling synaptic plasticity in the basal ganglia (Lieberman 2009) rather than language per se, and playing a role in vocalizations and vocal learning in a wide variety of species, from bats (Li et al. 2007) to songbirds (Haesler et al. 2004). The changes in *FOXP2* in the human lineage quite likely are connected with some aspects of language, but the connection is not nearly as direct as early reports claimed, and as Ackermann et al. apparently assume. While *FOXP2* is clearly relevant at some level when modeling the brain mechanisms of language, Ackermann et al. go far beyond the data when they treat speech evolution as "*FOXP2*-driven" (sect. 5.2).

Likewise, the apparent presence of human *FOXP2* in Neanderthals does not in itself prove that Neanderthals spoke (Benítez-Burraco & Longa 2012). They most likely did speak, but that conclusion rests on a complex web of inferences from diverse sources of evidence, with *FOXP2* just one minor piece of the puzzle (Dediu & Levinson 2013; Johansson 2013; cf. Barceló-Coblijn & Benítez-Burraco 2013).

It is also imprudent to assume that Neanderthals and modern humans did not interbreed (target article, sect. 5.2), and quite improper to invoke Green et al. (2010) in apparent support of this assumption. The jury is still out on the interbreeding issue (Johansson 2013), but evidence favoring interbreeding is accumulating (Green et al. 2010; Dediu & Levinson 2013; Yotova et al. 2011). Ackermann et al. do consider gene flow as an alternative scenario, but here the time frame is off; an emergence of the *FOXP2* mutations 40,000 years ago (sect. 5.2) is not consistent with their presence in all modern human populations, as this postdates our most recent common ancestor (MRCA; Johansson 2011; Macaulay 2005) and is not supported by a proper genetic model either (Diller & Cann 2009).

In their main scenario of no interbreeding, Ackermann et al. have a different time-frame problem; the *FOXP2* change is here constrained to be older than 400,000 years, but the fixation rate is not constrained in this case, nor is there any tight upper time limit (cf. Diller & Cann 2009; 2012), so it is improper to conclude that it must have been "a relatively fast fixation" and thus "strong selection pressures" (target article, sect. 5.2).

Ackermann et al. dismiss the possible contribution of anatomical data from fossils in a single sentence (sect. 5.2, para. 2), and while they are correct that endocasts and cranial bases are not highly informative, other relevant anatomical evidence is available, as reviewed in Johansson (2013) and Dediu & Levinson (2013).

Vocal displays as the selective driver of protolanguage evolution (target article, sect. 5.2; cf. Locke & Bogin 2006) are highly unlikely, as they would drive the evolution of something more resembling birdsong than language (Johansson et al. 2006).

The distinct processing systems for music and language in modern humans likewise do not support such a scenario (Dediu & Levinson 2013).

Ackermann et al. mention briefly many different popular works on language evolution (e.g., Bickerton 2009; Mithen 2005; Falk 2004), but they do not engage with them at any depth, just picking some aspect from each that fits into their own scenario, without integration.

In summary, Ackermann et al. accurately identify brain circuitry issues that need to be addressed in the context of language evolution, and they provide an interesting, if speculative, evolutionary scenario for these circuits. But as soon as they step outside the brain and attempt to engage with other types of evidence, or with possible selective scenarios driving language evolution, their treatment is insufficient.

The forgotten role of consonant-like calls in theories of speech evolution

doi:10.1017/S0140525X1300407X

Adriano R. Lameira[a,b]

[a]*Institute for Biodiversity and Ecosystem Dynamics, University of Amsterdam, 1098 XH Amsterdam, the Netherlands;* [b]*Pongo Foundation, 3421 XN Oudewater, The Netherlands.*

adriano@orangutan.nl

Abstract: Ackermann et al. provide an informative neurological road-map to primate call communication. However, the proposed model for speech evolution inadequately integrates comparative primate evidence. Critically, great ape voiceless calls are explicitly rendered unimportant, leaving the proposed model deprived of behavioral feedstock and proximate selective drivers capable of triggering the neurological transformations described by the authors in the primate brain.

Ackermann et al. compile a manual guide to the neurology of acoustic communication in primates and humans that should be read by any student and scholar interested in one of the oldest questions in evolutionary biology – speech evolution. The authors fail, however, to integrate this important information with critical evidence from comparative primate research, (a recurring pitfall in neurology-based hypotheses for language evolution; Arbib 2005; Seyfarth 2005) and so the proposed evolutionary model falters on central heuristic pillars.

In agreement with the currently dominant view of speech evolution (Fitch et al. 2010; Janik & Slater 1997), Ackermann et al. place a pronounced, but unwarranted importance on vocal learning, underlined primarily by vocal fold control. Because nonhuman primates, including great apes, are assumed to be incapable of vocal learning (Janik & Slater 1997), the authors logically presume that "motor mechanisms of articulate speech appear to lack significant vocal antecedents within the primate lineage" (sect. 1.1, para. 2). Paradoxically, Ackermann et al. argue then for the existence of vocal continuity at the motor level within the primate lineage and pursue an evolutionary model which addresses speech features that primarily relate to nonhuman primate voiced calls, or "vocalizations," and vowels.

Interestingly, Ackermann et al. describe and depict in a clear way that vocal fold control is obligatorily involved *solely* in the production of vowels, while consonants are often voiceless and may be produced via supra-laryngeal articulation *alone* (with or without simultaneous airflow). The authors recognize that "virtually all languages of the world differentiate between voiced and voiceless sounds" (sect. 4.1, para. 1), and the diagrams provided by the authors illustrate well that supra-laryngeal articulation is versatile, multidimensional, multicomponent, and arguably, in some occasions, at least as complex as vocal fold control, both in motor control and acoustics. This fact is well illustrated, for instance, by the size of the consonant repertoire across all the world's spoken languages, which is three-fold larger than that of vowels (Maddieson 1984). Any suitable account of speech evolution must thus account for the evolution of both speech building blocks in our lineage.

Like human consonants, some great apes calls do not obligatorily require the control or action of the vocal folds. Great ape voiceless calls, such as clicks, raspberries, smacks, kiss sounds, and whistles, are underlined by voluntary control and maneuvering of supra-laryngeal articulators (i.e., tongue, lips, and jaw) in apparent homology to the articulatory movements of voiceless consonants (Lameira et al. 2013c). These calls rely on social learning for their acquisition and fine sensory-motor feedback for proper production (Hardus et al. 2009b; Lameira et al. 2013a; 2013b; Marshall et al. 1999; Wich et al. 2009; 2012). Apart from some rare cases across different taxa (e.g., storks, deer, macaques), great apes produce *multiple* voiceless calls. With the exception of humans, it is yet unclear whether any other animal species have explored the acoustic space of their supra-laryngeal vocal tract to an extent similar to great apes. For instance, in some wild orangutan populations, voiceless calls can account for half of the repertoire of an individual who produces more than ten different call types (Hardus et al. 2009b). Unfortunately, Ackermann et al. neglect the importance of the homology in articulation, acoustics, and acquisition between great ape voiceless calls and human voiceless consonants.

Additionally, the authors tentatively suggest that factors like mother–infant interactions, grooming, social prestige, and communal dancing indirectly supported the emergence of vocal learning. Such suggestions find no ground in primate literature. These factors are either shared between most of nonhuman primates or are only known in humans, obscuring possible phylogenetic approaches to relevant primate communicative traits. As described in the target article, any significant and unique role the mentioned factors may have played in the earliest stages of speech evolution remains at least ambiguous and vague. Cooperative breeding, for instance, is also left out, though this is a promising factor capable of prompting a shift in the fundamental way ancestral primate individuals may have communicated with each other (Burkart et al. 2007; 2009a; 2009b; Burkart & van Schaik 2010; Isler & van Schaik 2012; van Schaik & Burkart 2010).

In sum, Ackermann et al. present an evolutionary model inferred virtually from neurology alone, lacking concrete and/or realistic primate behaviors and selective drivers that may have prompted the neural transformations described. Great ape voiceless calls provide one such potent behavioral model and resolve the conflicting notions of motor continuity within the primate lineage. Although further research is needed (Lameira et al. 2013c), evidence suggests that a call repertoire composed of innate vocalizations together with a minority of learned voiceless calls represents a shared feature among all great apes (Lameira et al. 2013c), dating back thus to our ape ancestor. Such an extended repertoire would have offered direct communicative benefits for the transmission of more (detailed) information, disclosing an advanced primate cognition into acoustic communication (Seyfarth & Cheney 2003a; 2008; 2010; Seyfarth et al. 2005) across whatever contexts. Such benefits would have predictively triggered selective pressures towards increased motor control over call production, even though in the absence of vocal fold control. In other words, it is possible that vocal learning did not trigger the emergence of a primate open-ended call repertoire, but represented sequentially a "secondary" evolutionary step (Lameira et al. 2013a). Flexible (e.g., Clay et al. 2011; Koda et al. 2013; Lemasson et al. 2011; Ouattara et al. 2009; Slocombe & Zuberbuhler 2007; Townsend et al. 2008) and intentional (Gruber & Zuberbühler 2013; Schel et al. 2013) use of innate vocalizations by nonhuman primates may have then provided

the basis for the expansion of motor control over the vocal folds sufficient to allow individuals to learn to produce new voiced calls.

Overall, great ape voiceless calls beg for a reconsideration of the premises of the model proposed by Ackermann et al. The homology between great ape voiceless calls and human consonants warrants serious consideration of the former in any historical account of speech evolution. Great ape voiceless calls, for instance, also show fascinating features in that they may be produced simultaneously with "musical" instruments (Hardus et al. 2009b; Lameira et al. 2012), and their cultural transmission within separate populations leads to the emergence of functional arbitrariness in primate acoustic communication (Lameira et al. 2013b). These features are probably based on neurological interactions that are yet to be documented and/or investigated, but that pose intriguing possibilities for our comprehension of speech evolution.

Understanding speech evolution will require integrating evidence collected across multiple levels and disciplines (Christiansen & Kirby 2003). Neurological studies and approaches to the question of speech evolution will be of invaluable importance, but there should be a committed effort to "anchor" neurological data to comparative primate research, mimicking the synergies that likely played out between the primate brain and primate communicative behavior in the course of speech evolution.

ACKNOWLEDGMENT

Adriano R. Lameira was financially support by the Menken Funds of the University of Amsterdam during the preparation of the manuscript.

Early human communication helps in understanding language evolution

doi:10.1017/S0140525X13004081

Daniela Lenti Boero

Department of Social and Human Sciences, University of Valle d'Aosta, 11010 Aosta, Italy.

d.lentiboero@univda.it

http://www.univda.it/lentiboerodaniela

Abstract: Building a theory on extant species, as Ackermann et al. do, is a useful contribution to the field of language evolution. Here, I add another living model that might be of interest: human language ontogeny in the first year of life. A better knowledge of this phase might help in understanding two more topics among the "several building blocks of a comprehensive theory of the evolution of spoken language" indicated in their conclusion by Ackermann et al., that is, the foundation of the coevolution of linguistic motor skills with the auditory skills underlying speech perception, and the possible phylogenetic interactions of protospeech production with referential capabilities.

According to Ackermann et al., human language is a multicomponent process whose evolution must have operated at all life stages (Hogan 1988; Locke & Bogin 2006). In the first year of life, human sounds undergo a radical transformation: the substitution of the cry, an analog signal paralleling the dimension of infant's homeostatic imbalance (Gustafson et al. 2000; Lenti Boero et al. 1998) and similar to mammalian signals by design (Lieberman et al. 1968; 1971), with articulated speech-like sounds and some meaningful words at the end of the first year (de Boysson-Bardie 2001; Lenti Boero & Bottoni 2006; Oller 2000). Thus, in the first few months of life millions of years of language evolution are summarized; and therefore some benchmark might be outlined to make hypotheses about the selective pressures at work.

The how question. A first point is the transformation of the analog signal "cry" into articulated sounds: From the age of 2 months, infants start producing very low intensity protophones which are mostly, but not all, vowel-like, and have more complex melodic contours than cries (Ruzza et al. 2003). In the months to follow sounds are produced with the entire vocal apparatus (mouth, lips, nose, and throat) (de Boysson-Bardie 2001; Oller 2000). This points to the appearance of a better control of sound emission due not only to the maturation of the vocal apparatus, but also to a better nervous motor control of phonation. Vocal tract length in neonates is about 6 to 8 cm (Vorperian & Kent 2007) and reaches 8.5 cm at 18 months, that is, 55% of adult size (Vorperian et al. 2009). Animal and human studies suggest that the nervous motor control of infant cry is similar to that of monkeys: It involves the limbic system that initiates the cry, the midbrain structures that configure the response, and the brainstem that is responsible for the mechanics of the cry. The latter integrates the laryngeal and respiratory activity with the activation of the subroutines for fixed vocal patterns pre-programmed as an answer to external stimuli (Jürgens & Ploog 1988; Lenti Boero 2009, Lester & Boukydis 1992); thus a control for articulated sounds should only come from a rearrangement and a maturation of other centers allowing more motor freedom to lips, mandible, and tongue movements (Davis & MacNeilage 2002).

The why question. Cry is an alarm signal with striking characteristics of loudness and long duration. It can communicate individuality, sex of the caller (Cismaresco & Montagner 1990; Rocca & Lenti Boero 2005), and urgency to a recipient (Lenti Boero et al. 2008). Now, imagine a hominid social group endowed with such communicative tool: A known individual could communicate alarm and urgency to group mates from a distance. This communication might have had a basic referentiality as in other mammalian species (Lenti Boero 1992; Rasa 1986; Seyfarth & Cheney 1980; Zuberbühler 2000b). Why go further? Cries are fixed analog sounds and we know they might be aversive even for mothers (Frodi 1985; Frodi & Senchack 1990; Lenti Boero et al. 2008; Levitzky & Cooper 2000), while articulated sounds are considered music-like and very pleasant to the care giver (Papoušek & Papoušek 1981). Newborns having a capacity for music-like sounds might have been preferentially selected by parents (Locke 2006), as a pilot experiment suggests (Lenti Boero & Bottoni 2009). Those same infants might have been selected when adults (Hogan 1988) because they were able to use frequency modulated sounds in courtship in a kind of hominids' ancestral serenade, enabling them to communicate felt emotions (Banse & Scherer 1996).

Auditory-motor coevolution. All communication devices, human language included, imply the coevolution of both receiver and emitter, which is evident in the specialized adult language brain areas: Wernicke's and Broca's. During early development we know that infant perception of surrounding sounds, including language, is much more advanced than motor competence: Infants are capable of auditory streaming at 2–5 days old (Winkler et al. 2003), and they discriminate vowel and phonetic sounds from the first month (Clarkson & Berg 1983; Eimas et al. 1971; Mehler et al. 1988; Teinonen et al. 2009), sharing this capacity with many animal species: rhesus macaques, dogs, chinchilla, quails, and parrots (Adams et al. 1987; Bottoni et al. 2009; Dewson 1964; Kluender et al. 1987; Kuhl & Miller 1975; Miller 1977; Morse & Snowdon 1975; Pepperberg 2007). On the melodic and musical side newborn infants recognize musical melodies heard before birth (Kisilevsky et al. 2004). In addition, event-related brain potential (ERP) and magnetoencephalography MEG studies show that newborns can form expectation of a musical pitch and that infants detect substitution of musical notes (Tervaniemi & Huotilainen 2003). Eventually, infants shape their cries' melodic contours upon their native language (Mampe et al. 2009), thus showing a foundation for auditory-motor connection and imitation. Infants' sense of hearing is "encyclopaedic," because it is open to all linguistic and music sounds. This capacity is lost from 5 to 6 months (de Boysson-Bardie 2001), when infants attach their attention to motherese (Oller 2000), a nonexistent feature at the dawn of language.

Thus, sound imitation by means of protophones might have been concentrated on surrounding sounds, especially those uttered by predator or prey animals, to convey information about their presence and denote them in the acoustic channel. Though many theories point to enhanced sociality (Dunbar 1993), the possibility to refer to an object is still a core point for language evolution (Lenti Boero & Bottoni 2006) and might have been a key factor for the selection for articulated sounds emission.

ACKNOWLEDGMENTS

The research going into the preparation of this commentary was supported in 1995, from 2001 to 2003, and from 2005 to 2007 by grants from MURST (Ministero dell'Università e della Ricerca Scientifica e Tecnologica), and by funding from the University of Valle d'Aosta from 2009 to 2013.

Why we can talk, debate, and change our minds: Neural circuits, basal ganglia operations, and transcriptional factors

doi:10.1017/S0140525X13004093

Philip Lieberman

Department of Cognitive, Linguistic and Psychological Sciences, Brown University, Providence, RI 02912.

Philip.Lieberman@gmail.com

www.Brown.edu

Abstract: Ackermann et al. disregard attested knowledge concerning aphasia, Parkinson disease, cortical-to-striatal circuits, basal ganglia, laryngeal phonation, and other matters. Their dual-pathway model cannot account for "what is special about the human brain." Their human cortical-to-laryngeal neural circuit does not exist. Basal ganglia operations, enhanced by mutations on FOXP2, confer human motor-control, linguistic, and cognitive capabilities.

It has been clear for decades that aphasia never occurs without subcortical damage, and can occur absent insult to the cortex (Naeser et al. 1982; Stuss & Benson 1986). The speech production deficits of Parkinson disease and focal lesions to the basal ganglia are qualitatively similar to ones occurring in aphasia (Blumstein 1995; Blumstein et al. 1980; Lieberman et al. 1990; 1992; Pickett et al. 1998; Usui et al. 2004) and are not limited to aberrant laryngeal phonation. Motor control is slow and imprecise, thus degrading speech, walking, and other internally guided motor tasks (Harrington & Haaland 1991; Marsden & Obeso 1994). A suite of cognitive deficits occurs (Flowers & Robertson 1985; Lange et al. 1992), including impairment of cognitive inflexibility and comprehending distinctions in meaning conveyed by syntax (Grossman et al. 1991; Lieberman et al. 1990; 1992; Natsopoulos et al. 1993). Similar, less pronounced, motor and cognitive deficits occur when hypoxic insult degrades the metabolically active basal ganglia (Lieberman et al. 1994; 2005).

These behavioral deficits derive from insult to a network of segregated cortical-to-basal neural circuits linking areas of motor cortex and prefrontal cortex. Marsden and Obeso (1994), taking into account a comprehensive range of studies, concluded that the basal ganglia act as a neural "switch" in circuits linking them to the motor cortex, activating and linking submovements in internally guided acts such as walking or talking. When circumstances suggest a different motor response, the basal ganglia switch to a different sequence. The basal ganglia perform similar operations during cognitive tasks in circuits that include areas of the prefrontal cortex. fMRI studies confirm their supposition. For example, the ventrolateral prefrontal cortex and the caudate nucleus of the basal ganglia are active when a subject is planning to change how he or she is sorting images on the basis of their shapes, to sorting them by color; or selecting words that rhyme and, instead, shifting to selecting words that have similar meanings. A cortical-to-basal ganglia circuit that includes the putamen and posterior prefrontal cortex is active during the execution of a sorting set-shift. The dorsolateral prefrontal cortex is involved whenever subjects make any decision, apparently monitoring whether the subjects' responses were consistent with the chosen sorting criterion (Monchi et al. 2001; Simard et al. 2011). Other neuroimaging studies, reviewed in Lieberman (2000; 2002; 2006b; 2012; 2013), show that the prefrontal cortex and the basal ganglia are active when subjects have to understand the meaning of a sentence, recall words from memory, subtract numbers, and cognitive tasks. All primates, including humans, appear to have similar cortical-to-basal ganglia circuits (Lehericy et al. 2004).

Ackermann et al. instead place great weight on a hypothetical direct cortical-to-laryngeal neural circuit that bypasses the basal ganglia, accepting a premise advanced in Fitch (2010). The circuit does not exist, being based on flawed attempts to adapt a lethal tracer technique to study humans. The Nauta and Gygax (1954) technique necessitates destroying discrete neural structures in an animal's brain. After some weeks the animal is sacrificed and its brain is impregnated with a silver solution that delineates neuronal structure. Microscopic examination of sectioned brain tissue can then reveal damage to downstream neurons in circuits to the neural structure that was destroyed. Using this technique, Kuypers (1958a) and Iwatsubo et al. (1990) claimed that changes to spinal cord neurons that enervate the larynx revealed a direct cortical-laryngeal circuit. However, the deceased patients studied had massive brain damage that included the basal ganglia and pathways to it. Similar changes to brainstem neurons occurred in patients who had died from non-neurological disease processes (Terao et al. 1997). Jürgens (2002b) concludes his review article on the neural bases of motor control by noting that "motor coordination of learned vocal patterns comes from the motor cortex and basal ganglia" (p. 251). Moreover, in itself, enhanced laryngeal control of phonation would not have yielded the encoding of segmental phonemes that is a unique property of human speech (Liberman et al. 1967).

Ackermann et al. claim that basal ganglia circuits are devoted to learning "digital" linguistic contrasts in the first years of life, then shift to learning emotional prosody. However, no data are presented to support this claim, and developmental studies show that this is not the case. For example, prosodic patterns signaling intent are apparent in the first year of life in infants in a Catalan-speaking environment (Esteve-Gibert & Prieto 2013). Both lexical tones and prosodic patterns emerge in the early years of life for Mandarin-learning infants (Chen & Kent 2009).

As my publications have pointed out, transcriptional factors such as FOXP2 may hold the key to why the human brain enables us to talk, continually create new forms of art, and possess language (Lieberman 2006b; 2009; 2013). The basal ganglia, which initially played a role in motor control, appear to have been modified in the course of evolution. The version of FOXP2 that differs with respect to two amino acids from chimpanzees enhances synaptic plasticity in basal ganglia neurons and in the substantia nigra. It also increases dendritic connectivity. A third mutation on FOXP2 (on interon 8, close to the amino acid substitutions) appears to enhance transcription. This uniquely human mutation occurred when modern humans first appeared in Africa (Maricic et al. 2013). It resulted in a "selective sweep." Selective sweeps on genetic mutations, such as those that confer adult lactose tolerance (Tishkoff et al. 2007), occur when a mutation enhances the survival of progeny. One of the tenets of neurophysiology is that synaptic plasticity is the key to learning anything. Virtually all human knowledge is transmitted through the medium of language, and FOXP2 appears to have played a role in the evolution of human language by enhancing basal ganglia synaptic plasticity and connectivity.

It is puzzling that Ackermann et al., disputing my views on the physiology of speech production, included a direct quotation from

page 289 of my 2006 book, *Toward an Evolutionary Biology of Language*. In pages 131 to 245 of this book I discuss, in detail, the issues noted above and other points raised by Ackermann et al.

En route to disentangle the impact and neurobiological substrates of early vocalizations: Learning from Rett syndrome

doi:10.1017/S0140525X1300410X

Peter B. Marschik,[a] Walter E. Kaufmann,[b] Sven Bölte,[c] Jeff Sigafoos,[d] and Christa Einspieler[a]

[a]*Institute of Physiology, Research Unit iDN – Interdisciplinary Developmental Neuroscience, Medical University of Graz Austria, 8010 Graz, Austria;* [b]*Department of Neurology, Boston Children's Hospital and Harvard Medical School, Boston, MA 02115;* [c]*Center of Neurodevelopmental Disorders (KIND), Department of Women's and Children's Health, Karolinska Institutet, Astrid Lindgren Children's Hospital, Solna 171 76 Stockholm, Sweden;* [d]*School of Educational Psychology, Victoria University of Wellington, PO Box 600, Wellington 6012, New Zealand.*

peter.marschik@medunigraz.at
walter.kaufmann@childrens.harvard.edu
sven.bolte@ki.se jeff.sigafoos@vuw.ac.nz
christa.einspieler@medunigraz.at
www.medunigraz.at/physiologie/pbmarschik
http://www.iddrc.org/childrens-hospital-boston/index.php/investigators/details/walter_e._kaufmann_md
www.ki.se/kind
http://www.victoria.ac.nz/education/about/staff/ed-psy-ped-staff/jeff-sigafoos
www.medunigraz.at/physiologie/ceinspieler

Abstract: Research on acoustic communication and its underlying neurobiological substrates has led to new insights about the functioning of central pattern generators (CPGs). CPG-related atypicalities may point to brainstem irregularities rather than cortical malfunctions for early vocalizations/babbling. The "vocal pattern generator," together with other CPGs, seems to have great potential in disentangling neurodevelopmental disorders and potentially predict neurological development.

Acoustic communication has become the focus of intensive research aiming to assess, delineate, and interpret the integrity of neural functions within different theoretical frameworks. For example, an increasing number of studies have aimed to document early difficulties in this domain and their potential implications with participants of various developmental disorders, such as autism spectrum disorders or Rett syndrome (RTT). In this research a series of peculiarities have been reported. Publications on delay in acquisition of milestones are increasingly complemented by documentation of qualitative deviances, even at the earliest stages of speech-language acquisition such as cooing and babbling vocalizations (e.g., Marschik et al. 2012; 2013; Paul et al. 2011).

Our species-unique ability to frame our world with words and the required neurobiological underpinnings that enable it have fascinated researchers studying phylogenetic and ontogenetic perspectives of language and communication. Contemplations about the origin, evolution, and development of verbal communicative abilities have led to many assumptions, speculations, theories, and attempts to deliver the one and only plausible explanation. In their article, Ackermann et al. postulate an ontogenetic model that assumes age-dependent interactions between basal ganglia and their cortical targets. We discuss the plausibility of and support for this assumption by reviewing recent findings on early vocalizations in infants with neurodevelopmental disorders, more specifically RTT.

From an ontogenetic perspective, early vocalizations – both nonhuman (e.g., as shown for pygmy marmosets) and human – actively promote the proximity and attention of caregivers, and therefore represent an advantage for the babbling infant (Elowson et al. 1998). But what can we say at this point about the neurobiological substrates of these vocalizations? And, more pragmatically, what does it tell us with regard to progressive neurodevelopmental conditions, such as RTT? Prosodic features of spoken language were reported to be dependent on the integrity of the basal ganglia, especially the striatum (Darkins et al. 1988; Van Lancker Sidtis et al. 2006). From an evolutionary perspective Ackermann et al. argue that a "structural reorganization of the basal ganglia during hominin evolution may have been a pivotal prerequisite for the emergence of spoken language" (sect. 1.2, para. 3). A great body of clinical evidence for the involvement of the basal ganglia in speech-language functions stems from patients with basal ganglia dysfunctions, such as Parkinson's disease or Tourette syndrome. The focus has been on the substantia nigra pars reticulata that exerts inhibitory control of the midbrain periaqueductal gray matter (PAG), a major relay of the descending motor system across vertebrates, and its role in converting emotional and cognitive commands into vocalization (Kittelberger & Bass 2013; Menuet et al. 2011).

The PAG does not directly control the coordinated activity of respiratory movements, and laryngeal and orofacial muscle groups, but rather projects to the closely related brainstem central pattern generators (CPGs; Hikosaka 2007). CPGs are neuronal circuits that can produce rhythmic motor patterns in the absence of oscillatory input. Some CPGs operate continuously (e.g., respiratory movements), whereas others are activated to perform specific behavioral tasks (e.g., locomotion). To provide motor output flexibility, supraspinal projections activate, inhibit, and, most of all, modulate the CPG-activity, as does sensory feedback (Einspieler & Marschik 2012; Grillner et al. 1995). CPGs for vocalization have been studied to a great extent not only in amphibians and avians, but also in mammals such as cats (CPGs located in the nucleus retroambiguus; Zhang et al. 1995) or squirrel monkeys (CPGs in the parvocellular reticular formation around the nucleus ambiguus; Hage & Jürgens 2006). Barlow et al. (2009) have suggested the same mechanism for early human vocalizations/babbling. A rudimentary understanding of the CPG-circuitry for respiration and mouth movements suggests multiple loci in the brainstem, with a significant role for integration among subsystems and the PAG (Barlow & Estep 2006).

The above-mentioned tight interconnection of CPGs (Barlow et al. 2009) becomes functionally evident when observing individuals with RTT, a neurodevelopmental disorder mainly arising from mutations in the X-linked *MECP2* gene (Neul et al. 2010). We have speculated that the interconnectivity of CPGs is pictured in RTT by the apparent evolution of early atypical vocalizations, with inspiratory-modulated sound patterns, into oro-motor dyspraxia and breathing irregularities later in childhood (Marschik et al. 2012). We propose the neuropathology of RTT, a condition with well-documented early atypical vocalizations in both humans and animal models (De Filippis et al. 2010; Marschik et al. 2012, 2013), as a model for elucidating abnormalities and their mechanisms involving the CPGs.

In terms of neurobiological substrates, studies of knock-out mouse models of RTT have revealed reduced striatal dopamine release after stimulation that coincided with motor abnormalities (Gantz et al. 2011). Whether such a nigro-striatal pathway involvement could also be associated with abnormal ultrasonic vocalizations, as demonstrated in the *Mecp2*-308 mouse model (De Filippis et al. 2010), remains open. Of relevance to the vocalization-generating circuitry is the demonstration of decreased PAG volume and length in yet another RTT mouse model (*Mecp2B*; Belichenko et al. 2008). Ultimately, the ontogeny of MeCP2 expression in the human brain (Kaufmann et al. 2005) supports an early involvement of brainstem monoaminergic nuclei and related brain regions in the pathogenesis of multiple neurologic deficits, including language.

In conclusion, developmental delays and atypicalities in verbal behaviors and other neurologic functions in RTT support CPG

and, consequently, brainstem involvement. Future human and animal model studies are needed to further elucidate developing brain–behavior interfaces, disentangle specific traits, and help detect affected children at an earlier age. The "vocal pattern generator" together with other CPGs seems to have great potential in disentangling neurodevelopmental disorders and potentially predict neurological development.

Speech as a breakthrough signaling resource in the cognitive evolution of biological complex adaptive systems

doi:10.1017/S0140525X13004111

Tobias A. Mattei

Department of Neurosurgery – Brain & Spine Center / InvisionHealth, Buffalo, NY 14221.

tobiasmattei@yahoo.com

http://www.invisionhealth.com/providers/neurosurgery/tobias-mattei-md/

Abstract: In self-adapting dynamical systems, a significant improvement in the signaling flow among agents constitutes one of the most powerful triggering events for the emergence of new complex behaviors. Ackermann and colleagues' comprehensive phylogenetic analysis of the brain structures involved in acoustic communication provides further evidence of the essential role which speech, as a breakthrough signaling resource, has played in the evolutionary development of human cognition viewed from the standpoint of complex adaptive system analysis.

In the target article, Ackermann et al. contend that speech has emerged as a major evolutionary advantage in hominin ancestors as a result of a refinement in the projections from the motor cortex to the brainstem nuclei responsible for the control of laryngeal muscles as well as the further development of vocalization-specific cortico-basal ganglia circuitries driven by certain mutations in the FOXP2 gene which were unique to humans.

Complex adaptive system (CAS) analysis has emerged as a powerful research approach that has been successfully used to study the basic mechanisms underlying the evolution of dynamical systems composed of multiple agents interacting through complex and interdependent networks. As a broad and general theoretical tool, CAS analysis has been employed in a variety of research fields in both biological and social sciences in order to unveil the common general principles responsible for the evolution of apparently unrelated complex systems, such as global macroeconomics (Gintis 2006), the stock market (Mauboussin 2002), geopolitical organizations (Braman 2004), the cyberspace (Phister 2010), natural ecosystems (Levin 1998), the immune system (Grilo et al. 2002), the human brain (Gomez Portillo & Gleiser 2009), and intracellular signaling networks (Schwab & Pienta 1997).

Recently it has been suggested that, taking into account the dynamic nature of grammar and semantics' evolution throughout the centuries, language should be considered a typical example of a complex adaptive system (Ellis 2009). More importantly, the emergence of the cognitive apparatus responsible for the processing of acoustic communication can be regarded as a unique breakthrough within biological complex adaptive systems, as it fostered the development of new signaling networks not only among different individuals, but also within the subsystems operating inside each specific agent (Pinker 2010). By enabling dynamical inter-individual interactions through fast and instantaneous feedback loops, the emergence of speech granted the biological systems harboring such a new cognitive resource an enormous evolutionary advantage not only from the individual standpoint, but also from the perspective of further development of the whole species through social collaboration. Ultimately the combination of such new signaling networks built upon oral communication and language gave rise to more complex and general social elements, which, in turn, began to play a central role in the very interactions among such individuals through new types of information exchange pathways such as advertisement, mass communication vehicles, and, more recently, social media (Fitch 2006).

By exploring the underlying brain structures involved in the phylogenetic emergence of speech, the target article demonstrates how the use of words, as standardized vocal utterances filled with specific meaning (or, as Ackermann et al. call them, phonetic-linguistic categories), represents a unique cognitive resource of the human species. In a simple manner, speech can be understood as a voluntary pattern of vocalization common to a social community which has definite connotations and which is carefully manipulated at each individual communication event for the transferring of specific information, intentions, and abstract ideas. As discussed by the authors, despite the fact that close primates may demonstrate elaborated oral-motor capabilities and possess an extensive vocal repertoire, they fail in producing a pattern of vocal communication that resembles speech. As properly pointed by Ackermann et al., nonhuman primate oral communication would be much more similar to other nonverbal affective forms of human vocal expressions (such as laughing, crying, or moaning) than to any type of organized and standardized pattern of vocalization that might possibly deserve the status of language. Based on such considerations, Ackermann et al. argue that a unique state of development of the neurophysiological networks responsible for coupling intentional planning and the refined coordination between the several motor elements involved in phonation (such as the tongue, laryngeal, jaw and facial muscles), ultimately enabled the human race to cross the critical edge that separates the crude vocalization patterns observed in other primates from human speech and language. In this sense, it could be said that, from the phylogenetic standpoint of brain evolution, the observed advancements in the neuroanatomical areas responsible for acoustic communication mentioned in the target article (such as the cortico-brainstem connections and the FOXP2 gene-induced new cortico-basal projections) fostered the further development of the primary cortical areas related to language emission (Broca's area, which is localized in the left inferior frontal gyrus of the dominant hemisphere – Brodmann's areas 44 and 45) and language comprehension (Wernicke's area, which is localized in the posterior section of the superior temporal gyrus – Brodmann's area 22) as well as of the white matter connection tracts and the accessory heteromodal association areas involved in the generation and processing of different speech features such as prosody, melody, rhythm, pitch, and syntax (Rauschecker 2012).

Such refinement in the brain networks responsible for the production and processing of speech in conjunction with advances in other brain regions which enabled the emergence of more complex non-pictographic forms of written language (i.e. systems which provided symbolic representations for the phonemes and words that became established in the oral culture throughout early human history) were also a decisive factor for the development of other higher cognitive functions which ended up achieving a uniquely sophisticated status in humans, such as semantic memory, abstraction, future anticipation and planning, and mathematical reasoning (Aboitiz et al. 2006).

In summary, the specific pattern observed in the evolutionary development of speech highlighted in the target article (which involves one breakthrough change leading to the percolation of the whole system and the emergence of new unpredictable attributes) represents a typical feature of complex adaptive systems. In fact, in such types of self-adapting and dynamical systems, it has already been demonstrated that a significant improvement in the signaling flow among agents (such as that proportioned by the development of speech and language) constitutes one of the most powerful triggering events for the emergence of new complex behaviors, very often leading to a complete reformulation

of the boundaries and hierarchical structures within the system (Holland 2012). As an academic masterpiece on the issue, Ackermann and colleagues' comprehensive phylogenetic analysis of the brain structures responsible for speech in humans and nonhumans primates provides further evidence of the essential role that speech and language, as breakthrough signaling resources, have played in the evolutionary development of human cognition viewed from the standpoint of complex adaptive system analysis.

Voluntary and involuntary processes affect the production of verbal and non-verbal signals by the human voice

doi:10.1017/S0140525X13004123

Carolyn McGettigan[a] and Sophie Kerttu Scott[b]

[a]*Department of Psychology, Royal Holloway, University of London, Egham TW20 0EX, United Kingdom;* [b]*Institute of Cognitive Neuroscience, University College London, London WC1N 3AR, United Kingdom.*

Carolyn.McGettigan@rhul.ac.uk sophie.scott@ucl.ac.uk
www.carolynmcgettigan.com
https://sites.google.com/site/speechskscott/

Abstract: We argue that a comprehensive model of human vocal behaviour must address both voluntary and involuntary aspects of articulate speech and non-verbal vocalizations. Within this, plasticity of vocal output should be acknowledged and explained as part of the mature speech production system.

In their account of the neural systems supporting vocal expression in humans, Ackermann et al. suggest that emotional and "attitudinal" aspects of prosody might influence the execution of speech via cross-talk between basal ganglia loops processing emotion, motivation, and speech motor programmes. It is problematic to claim one system for spoken language, plus one or more others for paralinguistic or non-linguistic signals that are then added together to make a finished product of fluent, emotionally inflected speech. The division between lateral motor cortex and other systems in the production of human vocal signals is not a simple one, and might be better characterized by the degree of *voluntary control* over the vocal tract rather than according to the type of signals generated. For example, patients who have sustained lateral cortical injuries disrupting the voluntary production of speech can still produce spontaneous and natural-sounding laughter and crying, and *swearing* (Van Lancker & Cummings 1999). Thus, articulate speech – swear words – can be produced involuntarily. Similarly, non-verbal emotional vocalizations can be produced under voluntary control – social laughter is typically timed to occur at the end of linguistic phrases, during both speaking and signing (Provine & Emmorey 2006). Recent work using functional MRI to explore the neural underpinnings of laughter showed a considerable involvement of lateral sensorimotor systems in the production of laughter under varying amounts of voluntary control (Wattendorf et al. 2013).

In everyday spoken language, voluntary modulation of the way we speak plays an essential role in the intentional expression of mood, intentions and aspirations. Hawkins and Smith (2001) illustrate this with the English phrase, "I do not know," the pragmatic sense of which can vary dramatically depending on how the words are articulated (compare the casual manner of "I dunno" with the suggestion of irritation in "I… do… not… know!"). We recently investigated the neural correlates of voluntary modulations of spoken language by asking participants in an MRI scanner to perform spoken impressions of accents and impersonations of familiar individuals (McGettigan et al. 2013). The peak activations associated with deliberate changes to speaking style (compared with speaking in a "normal voice") were found in the left anterior insula and inferior frontal gyrus. These areas are classically associated with the production of spoken language (Blank et al. 2002; Dronkers 1996), yet in this case the linguistic content of the

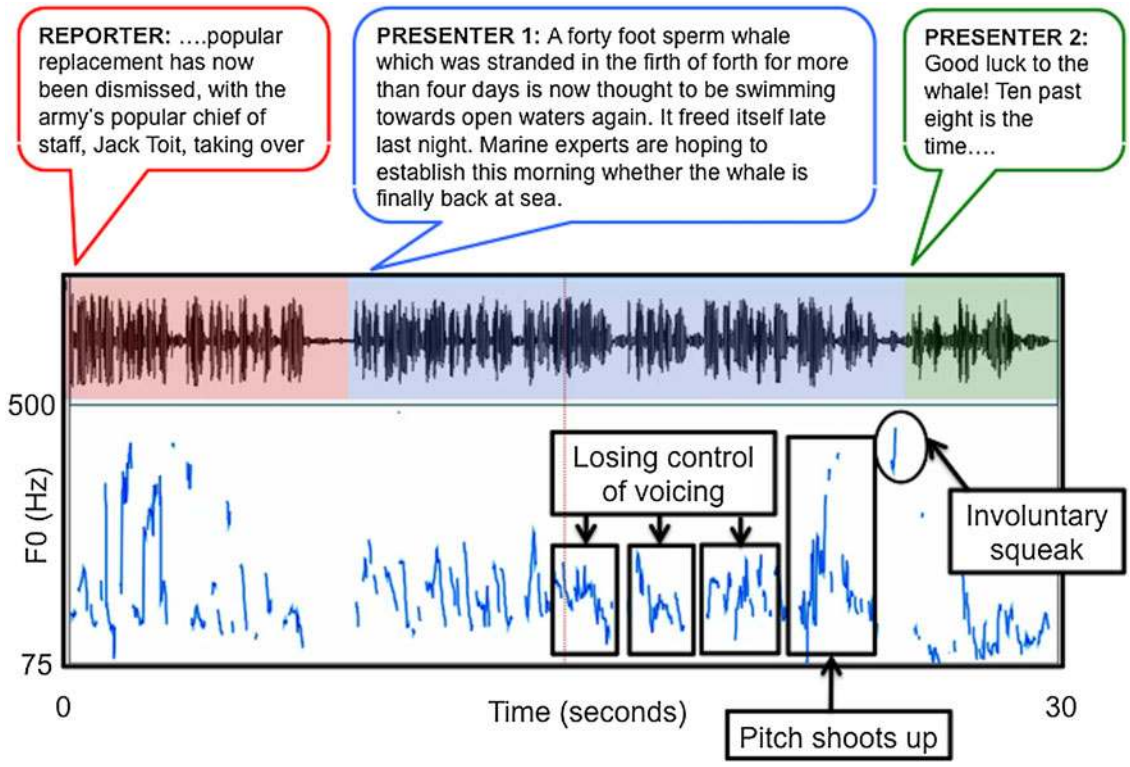

Figure 1 (McGettigan & Scott). On live radio, Presenter 1 is amused by the Reporter's pronunciation of "Jack Toit." Although she manages to deliver her script, the pitch (F0) of her voice rises sharply as her emotional state constricts the vocal tract and renders her less able to control the source of the vocal signal (Ruch & Ekman 2001).

utterances was kept constant across the different conditions of the experiment. It is difficult to assert that these voluntary aspects of speech production should, or *could*, be added to speech separately from the "digital" information bound up in the phonemes, syllables, and words of a language. Our recent results suggest that this kind of flexibility is an integral part of the planning and control of speech and voluntary vocal behaviour.

Not all vocal modulations can be added to speech in a controlled manner. Ackermann and colleagues argue that linguistic and emotional prosodic information, which they see as digital and analogue, respectively, are coordinated in the basal ganglia, as "Otherwise these two inputs would distort and corrupt each other" (target article, sect. 1.2, para. 2). It is reductive to draw boundaries between linguistic and paralinguistic aspects of vocal behaviour, particularly when considering the role of linguistic prosody in disambiguation (e.g., the contrast between a question and a statement). Furthermore, it is certainly the case that emotional states do corrupt articulate speech, as is shown when a person tries to produce speech during a fit of laughter, when overcome with grief, or when feeling extremely nervous – here, the *voluntary control* of vocalization is compromised, and articulate speech is taken over by the physiological effects of emotion on the functions of the vocal tract; see our Figure 1 (cf. Levenson 2003).

Ackermann et al. claim that the basal ganglia might be essential for the acquisition of articulate speech during early childhood, while the behaviours of the mature speech production system are controlled by perisylvian cortical structures. There is evidence that the plasticity of vocal learning reduces in adolescence and adulthood, for example, the marked persistence of first-language pronunciation in adult learners of a second language (Flege et al. 1999a; 1999b). However, speech can change in adulthood – one study showed that vowels in the speech of Queen Elizabeth II have, over several decades, gradually moved closer to the standard British English spoken by her subjects (Harrington et al. 2000). Similarly, there is extensive evidence for the recovery of speech in the adult system after stroke (Blank et al. 2003). It is difficult to estimate the extent to which these gradual changes in speech come about under conscious voluntary control. We continue to learn new information at all levels of the linguistic hierarchy throughout the lifespan, and the extent to which an individual changes their speech, voluntarily or not, can vary over both long and short timescales. With reference to the authors' proposal, we therefore pose the question: How do relearned and remapped behaviours in the adult speech production system fit within a model where the contributions of the basal ganglia end after childhood language acquisition?.

We are encouraged by an approach to modelling human vocal behaviour that incorporates its social, emotional, and linguistic aspects. However, we urge caution in attempts to divide the speech signal into distinct types of information served by specific underlying functional subsystems. We argue that vocal behaviour is better characterized in terms of voluntary versus involuntary control of a complex motor act, regardless of its informational content. Further, given the evidence that vocal behaviour remains plastic and flexible into adulthood, we question the extent to which this plasticity need be mechanistically distinct from childhood language acquisition.

Why vocal production of atypical sounds in apes and its cerebral correlates have a lot to say about the origin of language

doi:10.1017/S0140525X13004135

Adrien Meguerditchian,[a] Jared P. Taglialatela,[b,c] David A. Leavens,[d] and William D. Hopkins[c,e]

[a]*Laboratory of Cognitive Psychology, UMR7290, Aix-Marseille University – CNRS, Brain and Language Research Institute,13331 Marseille, France;* [b]*Department of Biology and Physics, Kennesaw State University, Kennesaw, GA 30144;* [c]*Division of Developmental and Cognitive Neuroscience, Yerkes National Primate Research Center, Atlanta, GA 30329;* [d]*School of Psychology, University of Sussex, Falmer BN1 9QH, England, United Kingdom;* [e]*Neuroscience Institute and Language Research Center, Georgia State University, Atlanta, GA 30302.*

adrien.meguerditchian@univ-amu.fr
http://gsite.univ-provence.fr/gsite/document.php?pagendx=12531&project=lpc
jtaglial@kennesaw.edu
http://science.kennesaw.edu/~jtaglial/Taglialatela_-_Ape_Communication_Lab/Home.html
davidl@sussex.ac.uk
http://www.sussex.ac.uk/profiles/114996
whopkin@emory.edu
http://neuroscience.gsu.edu/profile/william-hopkins/

Abstract: Ackermann et al. mention the "acquisition of species-atypical sounds" in apes without any discussion. In our commentary, we demonstrate that these atypical sounds in chimpanzees not only include laryngeal sounds, but also have a major significance regarding the origins of language, if we consider looking at their context of use, their social properties, their relations with gestures, their lateralization, and their neurofunctional correlates as well.

Whether apes are able to voluntarily and intentionally control their vocal production remains a topic of intense debate (e.g., Hopkins et al. 2011). In a brief paragraph in their target article (sect. 2.1.4.), Ackermann et al. mention the "observational acquisition of species-atypical sounds" in apes and acknowledge that chimpanzees are able to produce voluntary sounds using the modulation of the air through the lips ("blowing raspberries" or "kiss"). However, the authors also claimed that apes are not able to "engage laryngeal sound-production mechanisms" that can be "decoupled volitionally from species-typical audiovisual displays." In fact, this latter claim is not accurate.

Hopkins et al. (2007) have indeed described the use of two atypical novel "learned" sounds produced by several chimpanzees among the captive groups from the Yerkes Primate Research Center: Some chimpanzees are not only able to produce non-voiced "raspberries" or "kiss" sounds (involving only the lips with the air of the mouth) but also "extended grunts," which clearly engage the vocal tract and laryngeal sound-production mechanisms. Hopkins and colleagues showed that the production of these atypical sounds and vocalizations is often produced with pointing gestures and is used exclusively in the presence of both a human and an out-of-reach food in order to beg for food, while typical species-specific "food calls" were more frequent in the presence of food alone (Hopkins et al. 2007). Such atypical productions were interpreted as signals used intentionally to capture the attention of the human. Indeed, great apes have been shown to use those acoustic signals – vocal and lips sounds, cage banging or clapping gestures – especially when the recipient is not attentive, whereas visual pointing gestures are preferentially used when the recipient is attentive (e.g., Leavens et al. 2004; 2010; see also in orangutans: Cartmill & Byrne 2007; for a review of the literature, see Hopkins et al. 2011). In other words, the multimodal flexibility of communicative signaling (sounds, vocalizations, and gestures) is a manifestation of the ability of the great apes to adjust the modality of the signal to the attentional state of the recipient, and such an intentional property might be thus a special feature of social cognition that is needed in language processing.

In addition, given the inter-individual variability among chimpanzees concerning the ability to produce or not those novel sounds, it has been interpreted that, as for human speech but in contrast to species-typical vocalizations, those atypical vocal and lip sounds might be socially learned. In fact, it has been reported that chimpanzees raised by biological mothers who were able to produce those sounds, were more likely to also be able to do so than chimpanzees raised by humans in a nursery (Taglialatela et al. 2012). Moreover, among the chimpanzees that were not able to produce these atypical vocalizations, a recent study not

only showed that (i) it was possible to explicitly train them to do so using operant conditioning, but also (ii) that those subjects would further use these novel vocalizations in a communicative context for getting the attention of a human (Russell et al. 2013).

Finally, the investigation of lateralization of those atypical sounds and its functional cerebral correlates show some continuity with the language system. Indeed, most of the language functions involve a left-hemispheric dominance (Knecht et al. 2000). Interestingly, it turns out that these chimpanzee auditory signals, when produced simultaneously with food-begging pointing gestures, induce a stronger right-hand preference than when the gesture is produced alone (Hopkins & Cantero 2003), indicating that the left hemisphere may be more activated when producing both gestures and these atypical vocal and lip sounds simultaneously. Moreover, measures of orofacial asymmetries for vocal production in chimpanzees have showed that species-typical vocalizations – such as food barks or pant-hoot – elicited a left-sided orofacial asymmetry (i.e., right-hemispheric dominance), whereas atypical attention-getting sounds elicited an asymmetry toward the right side of the mouth, indicating that, as for right-handedness for communicative clapping gestures (Meguerditchian et al. 2012), a left-hemispheric dominance might be involved for producing those acoustical signals (Losin et al. 2008). More impressively, brain imaging studies (PET [positive emission tomography]) conducted in three captive individuals have found that communicative signaling for begging food from a human by using either gestures, atypical attention-getting sounds, or both of these modalities simultaneously, activated a homologous region of Broca's area (IFG) predominantly in the left hemisphere (Taglialatela et al. 2008), a pattern of activation which is enhanced in subjects who used both gestural and vocal signals simultaneously (Taglialatela et al. 2011).

These collective findings support the idea that the atypical orofacial and vocal sounds in chimpanzees are a good illustration of the potential existence of a multimodal intentional system that integrates gestures, orofacial, and atypical vocal sounds into the same lateralized system. This multimodal communicative system not only shares some features of social cognition and social learning with human language, but also seems to be ultimately related to brain specialization for language (Meguerditchian et al. 2011). This theory is consistent with the evidence that in humans, a single integrated communication system in the left cerebral hemisphere might be in charge of both vocal and gestural linguistic communication (e.g., Gentilucci & Dalla Volta 2008). For all of these reasons, and their implications for the precursors of human language and its brain specialization, we believe that Ackermann et al. should better consider these voluntary laryngeal sound-production mechanisms in chimpanzees and the related multimodal communicative system, in their theoretical model.

ACKNOWLEDGMENT

This research was supported by a grant from Agence National de la Recherche ANR-12-PDOC-0014-01 (LangPrimate).

Speech, vocal production learning, and the comparative method

doi:10.1017/S0140525X13004147

Bjorn Merker

Fjälkestadsv. 410-82, SE-29194, Kristianstad, Sweden.

gyr694c@tninet.se

Abstract: The faith that "comparative analysis of the behaviour of modern primates, in conjunction with an accurate phylogenetic tree of relatedness, has the power to chart the early history of human cognitive evolution" (Byrne 2000 p. 543) runs afoul of the fact that no other primate besides humans is capable of vocal production learning. This basic enabling adaptation for articulate speech bears crucially on the reconstruction of language origins.

In their target article Ackermann et al. make a valiant attempt to assemble a comprehensive account of the origin and neural organization of human speech on the basis of arguments confined by and large to comparative primatology. The nature of their topic is ill-suited to such an approach, because at its core lies a behavioral adaptation and corresponding neural mechanism which we share with some species of cetaceans, pinnipeds, and birds, but not with any nonhuman primate. For such a situation, the comparative method offers analogy instead of homology as guiding concept (e.g., the elucidation of body form in cetaceans is better served by turning to distant fishes rather than to far closer relatives among extant mammals).

The capacity in question is the ability to learn to reproduce, by voice, patterns of sound first received by ear. This capacity is of singular biological uselessness except in special cases, one of which happens to be us humans, because every word and phrase we know how to pronounce has become ours by such means. Technically, the capacity is known as vocal production learning (Janik & Slater 1997; 2000), and though the concept does occur in the target article, it is more by way of an afterthought than as a principal pivot of analysis.

Putting vocal production learning at center stage removes the mystery of the "speechlessness" of even our closest primate relatives rightly emphasized by Ackermann and colleagues. Lacking the vocal learning mechanism (Janik & Slater 1997), they naturally cannot do that which inherently is dependent upon it, namely, learn to pronounce words and phrases of rather arbitrary phonemic composition. That vocal learning is, in fact, the crux of the matter is demonstrated by the ease with which numerous species of parrots and other mimics among the birds do what no chimpanzee has ever done: acquire a substantial repertoire of human words and phrases pronounced with a fidelity that fools the human ear (Nottebohm 1976).

The diction of bird mimics tells us that the entire pronunciatory part of the speech equation is a matter of being a vocal learner. Step 1 on the path to speech is accordingly to come into possession of the capacity for vocal learning. *This* first step, moreover, provides a plausible evolutionary context for the first step invoked by Ackermann et al., namely, the addition of direct (monosynaptic) cortical efference to lower brainstem motor nuclei controlling larynx, pharynx, tongue, and lips.

The species distribution of such direct connections (to which can be added direct cortical innervation of the nucleus retroambiguus for respiratory control) suggests that they evolve specifically for cerebral fine control of respiration and vocalization and not (as the target article assumes) as a general concomitant of brain expansion (Arriaga & Jarvis 2013; Fitch et al. 2010; Iwatsubo et al. 1990; Jürgens 2002a; Kuypers 1958a; 1958b; Merker 2009; Okanoya & Merker 2007; Okanoya et al. 2007; Wild 1993; 1997).

As suggested in a previous BBS commentary (Merker 2009), it is even conceivable that the "simple" addition, in ancestral *Homo*, of a direct primary motor cortex efference to those medullary motor nuclei sufficed to recruit the already present cerebral territories centered on Wernicke's and Broca's areas (see Fig. 12.4 of Falk [2007] for putative homologs in *Pan* and *Macaca;* see also Neubert et al. 2014) to the practice-based acquisition of complex vocal output matching auditory models, thus making our ancestors vocal learners.

The most common use of vocal production learning in nature is as a means to impress potential mates and rivals by mastery of a complex song tradition (for the evolutionary logic, see Merker [2012] and review by Spencer & MacDougall-Shackleton [2011]). Humans are a singing species (von Humboldt 1836/1971), so the default assumption would be that the vocal learning capacity of our ancestors was exercised for similar purposes. If so, they were maintaining traditions of intergenerationally transmitted and culturally learned vocal lore (song) long before that lore became verbal by being semanticized.

Such a situation brings the "learner bottleneck" principle of iterated (transgenerational) learning into play (Kirby 2002). In the non-human examples the narrow focus of song displays on potential mates and rivals limits the operation of that principle to effecting progressive refinement of the formal properties of songstrings (their purely formal syntax; see Kirby et al. [2008] and references therein). Should, however, circumstances spread singing to the full range of daily and seasonal activities, the same principle would ensure a gradual and progressive differentiation of song repertoire by behavioral context, amounting to an implicit assortative contextual semanticization of the songstrings repertoire as a whole (Merker 2012; Merker & Okanoya 2007).

Therein lies the point of departure for a gapless path to human language, details of which are presented in a recent publication of mine (Merker 2012). At the point at which that path is about to arrive at fully instrumental language, it reaches an impasse. Attempts to use semanticized songstrings in a "displacement" mode of reference (Hockett 1960) would undermine the very contextual basis underpinning the semantic meaning of strings. To overcome that hurdle, I have postulated a naturally (as opposed to sexually) selected enhancement (expansion) of the cerebral storage capacity hosting the vocal learning mechanism, a capacity increase driven by individual self-interest in reaping the benefits of instrumental uses of songstring semantics.

Such an expanded storage capacity would allow string (and gestural) markers for communicative intent to be appended to the songstring repertoire by learning, launching it upon fully instrumental language use. Perhaps the FOXP2 enhancement of cortico-basal ganglia function in the human line provided the required extra storage capacity. As far as is known, it affects relevant neurons at the microscopic and functional level, promoting lengthening of dendrites (potentially increasing synapse numbers) as well as affecting synaptic plasticity by enhancing long term depression (Reimers-Kipping et al. 2011). Since such changes are general for the basal ganglia as a whole (along with associated thalamic and cortical domains), they fit better with a general expansion of storage capacity than with a remodeling of lateral interactions among circuits serving components of articulate speech assumed in the target article.

Phonation takes precedence over articulation in development as well as evolution of language

doi:10.1017/S0140525X13004159

D. Kimbrough Oller[a,b,c]

[a]*School of Communication Sciences and Disorders, The University of Memphis, Memphis, TN 38105;* [b]*Institute for Intelligent Systems, The University of Memphis, Memphis, TN 38152;* [c]*Konrad Lorenz Institute for Evolution and Cognition Research, A-3400 Klosterneuburg, Austria.*

koller@memphis.edu

http://umwa.memphis.edu/fcv/viewprofile.php?uuid=koller

Abstract: Early human vocal development is characterized first by emerging control of phonation and later by prosodic and supraglottal articulation. The target article has missed the opportunity to use these facts in the characterization of evolution in language-specific brain mechanisms. Phonation appears to be the initial human-specific brain change for language, and it was presumably a key target of selection in early hominin evolution.

The Ackermann et al. target article offers intriguing suggestions about a dual-pathway approach to evolution of vocal capabilities and language. But the approach could be enhanced with regard to its behavioral assumptions by taking into account key information on vocal development. Without this information, the article misses the opportunity to augment some of its most interesting claims with evidence that could help situate the model in a more evolutionary-developmental (evo-devo) frame (Bertossa 2011).

In particular, the article does not take note of the evidence that early vocal development of humans (especially in the first 3 months) shows emergent *phonatory* control, rather than emergent articulatory control (Buder et al. 2008; Koopmans-van Beinum & van der Stelt 1986; Oller 1980). Phonatory control does not necessarily imply "prosodic" control (the focus of much of the target article) in the sense that the term *prosody* is used in literature about mature languages. Prosody is normally meant to denote the capacity to integrate suprasegmental variations across multisyllabic strings, and the human infant in the first months produces no such syllabic strings. Instead, categories of infant vocalization that are recognized in the first few months by parents and laboratory staff cross-culturally include "protophones" (Oller 2000), such as vowel-like "vocants" (sounds produced in the mid-pitch range of the individual infant, with "normal" phonation, the kind that typically occurs in speech), squeals (high-pitched sounds for the infant in question, often in falsetto), and growls (low-pitched or raucous sounds often in creaky voice). These typical protophones occur as phonatory events with little or no articulatory modulation, and to the limited extent that supraglottal modulation occurs, it appears to be disorganized and unpredictable at this stage. On the other hand, the protophones are easily recognized as distinguishable categories because they tend to occur in clusters of the same types (a series of squeals, for example, followed later by a series of vocants) even as early as 3 months (Kwon et al. 2007). The high rate of production of the protophones, along with the fact that they occur both in solitary and social circumstances (Locke 1993; Stark 1980; Yale et al. 1999), suggests endogenous motivation in the infant to explore and seemingly to practice vocalization, as well as to use vocalization to serve social functions. In addition, all the protophones are used by infants in expression of positive, neutral, and negative affective states (as indicated by facial affect), and these expressions are predictably related to responses of caregivers ranging from encouraging interaction in response to positive expressions, to changing the situation (or talking about the need for it) in response to negative expressions (Oller et al. 2013). All these properties of very early vocal development (spontaneous production, the ability to repeat sounds in clusters, vocal social interaction, and the ability to use sounds to express differing emotional states on differing occasions) in the human infant at 3 months are based on phonatory control, and all of them are foundational for language, since every aspect of human language requires flexible control of phonation.

Phonatory control takes naturally logical precedence over supraglottal control in the sequence of development, and that logical precedence is reflected in the facts of development. Phonatory categories appear in development *before* systematic supraglottal articulated categories such as "canonical" syllables, wherein well-formed syllables (heard as, for example, "dada" or "baba") are produced through coordination of phonation and systematically repeatable supraglottal articulations (Koopmans-van Beinum & van der Stelt 1986; Oller 1980; Stark 1980). In 40 years of longitudinal research in human infant vocalization, I have never witnessed any infant developing systematic and communicative supraglottal movements and using them in practice-like play or social communication, prior to developing phonatory categories as described above.

These facts offer an opportunity to Ackermann et al. to more thoroughly incorporate developmental patterns into their expectations regarding the "monosynaptic refinement of the projections of motor cortex to the brainstem nuclei that steer laryngeal muscles." Such refinement, it would appear, must begin to be manifest in the human infant's brain by 3 months or earlier, although by that age, the behavioral data do not suggest "prosodic" control, but something simpler – control over gross differences in phonatory pattern and pitch. The developmental picture appears

to be sufficiently clear to help illuminate differences between human and nonhuman primate brain organization at maturity, by suggesting notable cross-species similarity of phonatory and articulatory control capabilities across nonhuman primates and humans early in life with much larger differences developing as time passes.

In the near future, it may also be possible for the modeling of Ackermann et al. to be enhanced by *quantified direct comparisons* among vocal capabilities of humans and nonhuman primates. On the one hand, studies in human infants are rapidly tying down facts regarding vocal rate (volubility) for protophones across ages and across social and non-social circumstances (Franklin et al. 2014; Goldstein et al. 2009; Nathani et al. 2001) as well as substantially improving our understanding of vocal types and their flexibility in human infants (Griebel & Oller 2008; Scheiner et al. 2006; Stark et al. 1993). Similarly, considerable progress is being made on the description of both the amount of vocalization that occurs across age in nonhuman primates and the degree to which these vocalizations are used in differing contexts, the latter representing an attempt to characterize the degree to which the social functions of nonhuman calls may indeed show flexibility (Crockford & Boesch 2003; Laporte & Zuberbühler 2010). With the recent development of a facial affect coding system for chimpanzees (Parr et al. 2008) modeled on the Ekman scheme for human affect (Ekman & Friesen 1978), quantitative comparison of functional flexibility in vocalization across humans and chimpanzees should soon be reached. Such improvements in our quantitative understanding of development, amplified by cross-species comparisons, should fundamentally enhance the modeling of the evolution of language and the brain mechanisms that underlie it.

ACKNOWLEDGMENTS

This work was supported by the National Institutes of Deafness and Other Communication Disorders (DC011027, D. K. Oller), by the Konrad Lorenz Institute for Evolution and Cognition Research, and by the Plough Foundation.

The basal ganglia within a cognitive system in birds and mammals

doi:10.1017/S0140525X13004160

Christopher I. Petkov[a,b] and Erich D. Jarvis[c]

[a]*Institute of Neuroscience, Newcastle University, Medical School, Newcastle upon Tyne NE2 4HH, United Kingdom;* [b]*Centre for Behaviour and Evolution, Newcastle University, Newcastle upon Tyne NE2 4HH, United Kingdom;* [c]*Howard Hughes Medical Institute and Department of Neurobiology, Duke University, Durham, NC 27710.*

chris.petkov@ncl.ac.uk jarvis@neuro.duke.edu

http://www.staff.ncl.ac.uk/chris.petkov

http://www.jarvislab.net/

Abstract: The primate basal ganglia are fundamental to Ackermann et al.'s proposal. However, primates and rodents are models for human cognitive functions involving basal ganglia circuits, and links between striatal function and vocal communication come from songbirds. We suggest that the proposal is better integrated in cognitive and/or motor theories on spoken language origins and with more analogous nonhuman animal models.

In the target article, Ackermann et al. present an interesting twist on the well-weathered hypothesis of a direct cortico-bulbar tract as a key step in the evolution of spoken language in humans, or song in vocal-learning birds. The authors seek to generate a new hypothesis that the basal ganglia, in particular, are functionally reorganized during human evolution for spoken language and also change in function during ontogeny with the learning of speech. Curiously, however, the basal ganglia, after supporting a language-learning role during child development, are proposed to revert to a seemingly more evolutionarily conserved functional role of supporting "emotive-prosodic" modulation in adult humans. This illustrates how the proposal flexes to encompass most data and risks being empirically untestable. Especially unclear is what similarities or differences are hypothesized to exist between humans and different animal models, where presumably homologous or analogous neurobiological mechanisms can be clarified.

Although we have little doubt that the basal ganglia were an evolutionary substrate for spoken language, one among many others, the current proposal requires considerable strengthening. We make two key suggestions. First, the hypothesis needs to be grounded in, or its key tenets distinguished from, certain cognitive and/or motor theories. Such theories have proposed that specific improvements occurred in vocal-learning systems or motor pathways of humans and some birds, including cortico-striatal-thalamic circuits (Arriaga & Jarvis 2013; Feenders et al. 2008; Fitch et al. 2010; Fitch & Jarvis 2012; Petkov & Jarvis 2012; Wild 1997). Second, we propose that the key tenets of the proposal, if clarified, can be comparatively tested in studies between, for instance, human and nonhuman primates, and songbirds and vocal non-learning birds, and any of these species and rodents (see our Figure 1). Such comparative analyses have already been used in the past to test for the hypothesized differences in the cortico-striatal system between some of these species, and can still be used to comparatively test additional aspects of the current proposal.

One issue is whether and which basal ganglia–dependent differences exist between humans and other nonhuman primates or mammals. There is little direct comparative evidence in the primate literature to suggest that the cortico-striatal-thalamic system is strikingly different in humans relative to nonhuman primates. In fact, as Ackermann et al. note, nonhuman primates and rodents are used as cellular model systems for human basal ganglia–related cognitive function on motor and procedural learning, habit forming, reward and decision-making, and sensory-motor timing relationships (Matell & Meck 2004; Schultz et al. 2000). Presumably, the proposal is that the basal ganglia, as part of a cognitive system, increased in capacity in humans to support language learning (Friederici 2011; Petkov & Jarvis 2012; Petkov & Wilson 2012). In this regard, it is possibly interesting that Artificial Grammar learning tasks, which were developed in the infant learning literature and that tap into rule-based procedural learning, appear to show differences between different species of monkeys (Wilson et al. 2013) and between monkeys and humans (Fitch & Hauser 2004). These observations were predicted by cognitive theories on spoken language origins (Arriaga & Jarvis 2013; Petkov & Jarvis 2012).

Thus, the proposal lacks the strength of the specificity of the direct cortico-bulbar hypothesis, and at the same time suffers from the limitation of overemphasis on a region vital for cognition, whose function is lost without the context of the cortico-striatal-thalamic circuits that are formed in the brains of birds and mammals. As a historical example, the direct cortico-bulbar hypothesis is now seen to be grounded in motor theories of spoken language origins (Petkov & Jarvis 2012). It is very specific that a monosynaptic change allowed learned sensory patterns to be vocally produced. But its strength in specificity was also its Achilles heel, leaving unanswered how humans and other mammals differ in their neurobiological substrates for learned auditory patterns, and which are linked to vocal motor output (via the nucleus ambiguus). Cognitive theories and the current proposal aim to address this shortcoming. Moreover, even the tenet of a presence versus absence of a direct cortico-bulbar tract is being challenged by recent data: Mice appear to have a sparse but still present direct cortico-bulbar projection to the nucleus ambiguus and greater vocal-production-plasticity capabilities than had been thought (Arriaga & Jarvis 2013; Arriaga et al. 2012), features that had been thought to be unique to humans and vocal-learning birds.

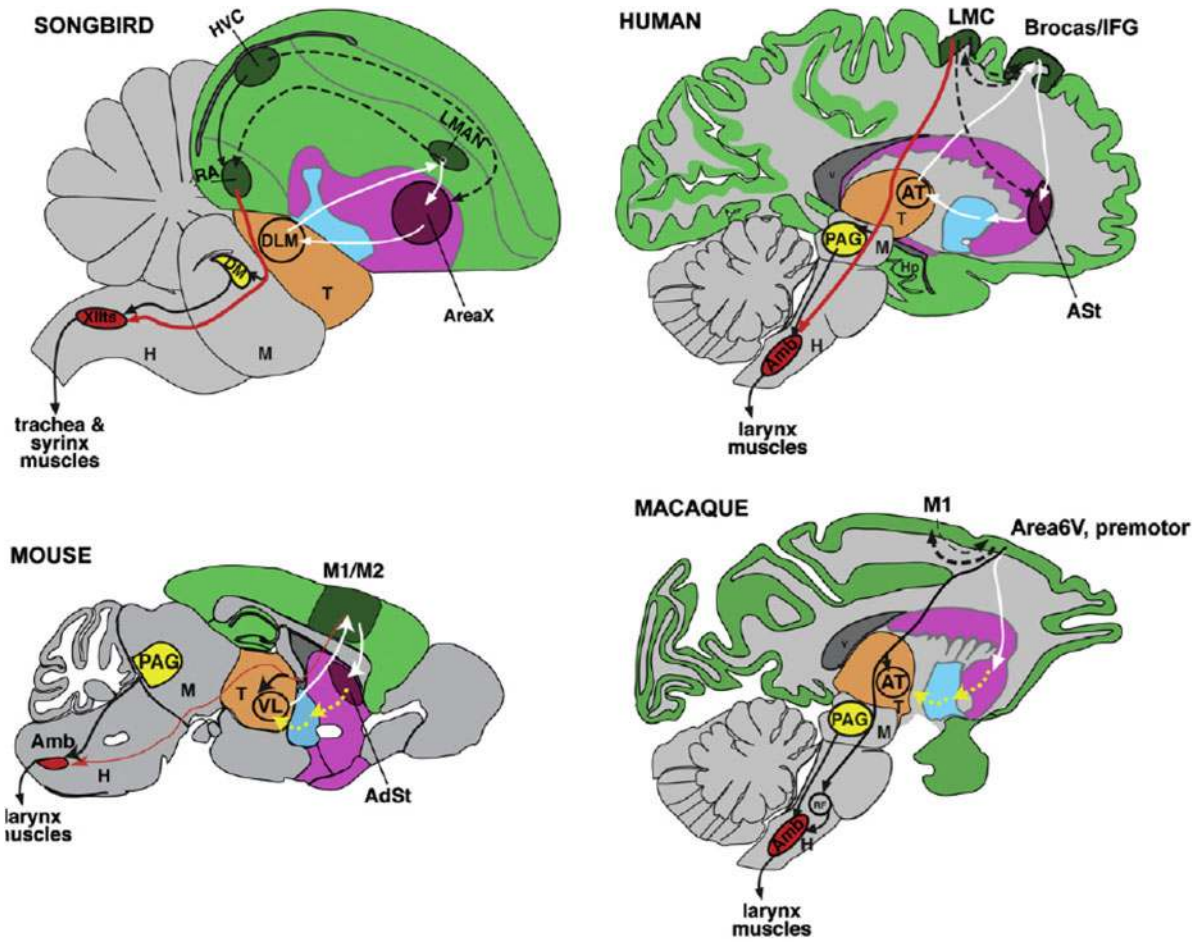

Figure 1 (Petkov & Jarvis). Summary diagrams of vocal systems in songbirds, humans, monkeys, and mice. Modified from Arriaga and Jarvis (2013). Cortico-striatal-thalamic loops are schematized from data in humans and songbirds. Yellow dashed lines in macaque monkeys and mice show proposed cortico-striatal-thalamic connections for vocalization that need to be tested.

Notably, the more precise link that the authors are pursuing with regard to the origins of spoken language and basal ganglia function, already has an evolutionary counterpart in vocal-learning and vocal-non-learning birds. The avian striatal vocal nucleus (called Area X in songbirds) sits within a cortico-striatal-thalamic loop, which is important for song learning (Jarvis 2004b; 2006; Jarvis et al. 2000), including covert-skill song learning (Charlesworth et al. 2012). Moreover, Feenders et al. (2008), by comparing the anterior-forebrain pathway in vocal-learning birds to this pathway in vocal-non-learning birds, found evidence to develop a motor theory of vocal-learning origin.

This theory proposes that the anterior-forebrain song pathway (including Area X) independently arose multiple times in vocal-learning birds from a set of regions that in vocal-non-learning birds control non-vocal motor actions. The discrete striatal Area X that sits within the cortico-striatal-thalamic vocal-learning loop (Fig. 1) is not present in vocal-non-learning birds. Motor striatal regions outside of Area X, or the comparable forebrain regions in vocal-non-learning birds, are more diffuse and relate to these animals' non-vocal motor learning abilities. Thus, considerable insights on the cortico-striatal-thalamic system have already been provided by avian models. These are only briefly alluded to but not meaningfully used to inform the current proposal.

In summary, Ackermann et al.'s proposal is an interesting review of the literature with an emphasis on the basal ganglia as an evolutionary substrate for spoken language. However, we found it heavy on conjecture and light on empirical hypotheses, which, as we have suggested, can be strengthened by (1) taking a broader evolutionary perspective that allows integrating data from birds and mammals, and (2) delineating more carefully how the current proposal can be integrated within or distinguished from other theories on spoken language origins.

The sensorimotor and social sides of the architecture of speech

doi:10.1017/S0140525X13004172

Giovanni Pezzulo,[a] Laura Barca,[a] and Alessando D'Ausilio[b]

[a]*Institute of Cognitive Sciences and Technologies, National Research Council, 00185 Rome, Italy;* [b]*Robotics, Brain and Cognitive Sciences Department, Italian Institute of Technology, 16163 Genova, Italy.*

giovanni.pezzulo@istc.cnr.it laura.barca@istc.cnr.it
alessandro.dausilio@iit.it
https://sites.google.com/site/giovannipezzulo/
https://sites.google.com/site/laurabarcahomepage/
http://www.iit.it/people/robotics-brain-and-cognitive-sciences-mirror-neurons-and-interaction-lab/researcher/alessandro-dausilio.html

Abstract: Speech is a complex skill to master. In addition to sophisticated phono-articulatory abilities, speech acquisition requires neuronal systems configured for vocal learning, with adaptable sensorimotor maps that couple heard speech sounds with motor programs for speech production; imitation and self-imitation mechanisms that can train the sensorimotor maps to reproduce heard speech sounds; and a "pedagogical" learning environment that supports tutor learning.

Besides sophisticated phono-articulatory abilities, the architecture of speech has key computational, neuronal, and social prerequisites that can shed light on its phylogenetic and ontogenetic origins.

As a first important requirement, the architecture of speech has to be configured for vocal learning, with adaptable sensorimotor circuits that couple heard speech sounds with motor programs for speech production. From a computational perspective, mastering speech in naturalistic environments plagued by uncertainty and noise is hard; this fact has long motivated control-theoretic views of speech emphasizing error-correction mechanisms and internal modeling (Guenther & Perkell 2004; Moore 2007).

Computational considerations also suggest that speech processing (and learning, see below) might benefit from a close interaction of perception and production systems. For example, production systems might support perceptual processes by predicting and "synthesizing" auditory candidates (as in *analysis by synthesis*), while perceptual systems might support the self-monitoring and error-correction of vocal production by affording an advance auditory analysis of the produced speech sounds. Neurobiological experiments support this idea by showing that the neuronal mechanisms for speech production and perception are not segregated in the brain; for example, specific motor circuits are recruited for the analysis of speech sound features (D'Ausilio et al. 2012). An organic proposal on the architecture of speech can be formulated within the framework of *generative systems*, in which perception and action systems share computational (and neuronal) resources and are both guided by a common prediction-error minimization process (Dindo et al. 2011; Friston 2010; Kiebel et al. 2008; Pezzulo 2012a; 2013; Yildiz et al. 2013).

A second important requirement is a learning method powerful enough to train the aforementioned sensorimotor architecture to perceive and (re)produce sounds and speech. This problem has been studied particularly in songbirds that, while not speaking, have sophisticated vocal learning abilities. Most theories assume that songbird learning is a staged process (Brainard & Doupe 2002). An initial period of auditory learning is needed to tune sensory maps to represent sensory "prototypes" of heard speech sounds (e.g., memorize learned song patterns heard by conspecifics). These prototypes are then used as "reference signals" for imitation learning; by learning to reproduce the stored template, an animal can acquire equivalent vocal sound production skills. In control-theoretic terms, this process uses (auditory and articulatory) feedback error-correction mechanisms to produce a sound (sing or speech) that closely matches the stored template (Guenther & Perkell 2004). During the learning process, internal (inverse and forward) models are trained, too, that successively afford skilled sing or speech processing.

To speed up learning, learners benefit from using self-imitation, too. Covert rather than overt singing (or speaking) might reproduce frequently heard speech sounds in the same way they are encoded in their sensory maps (note that generative architectures afford this form of learning quite naturally; Hinton 2007). Using both overt and covert processes, animals (including humans) might reproduce their stored prototypes with high fidelity, including the local *accents* of their communities.

The brain architecture supporting the aforementioned learning processes is incompletely known. Indeed, speech is a computationally challenging skill as it requires sensorimotor circuits to be sensitive enough to discriminate subtle changes in speech sounds, and accurate enough to afford extremely precise control (e.g., of the timing of speech). The brain could finesse these problems by recruiting cortico-subcortical loops (especially those involving the basal ganglia and the cerebellum) especially during learning. The role of these loops is seldom recognized in "cortico-centric" theories of motor skills (including speech), but the evidence indicates that they could play an important role in skill learning and mastery (Ackermann 2008; Caligiore et al. 2013). For example, vocal learning in the swamp sparrow might involve a loop between forebrain neurons that establish auditory-vocal correspondences and striatal structures important for song learning (Prather et al. 2008).

The high-fidelity reproduction of sounds could be key to cultural transmission and the evolutionary value of singing in songbirds (Merker 2012). However, human communities have richer social structures than other animals, which might have favored an open-ended instrumental use of vocal production besides ritualized display. The importance of this skill might have led to a greater investment of parental time in teaching and, we propose, to advanced forms of "tutor learning" (Canevari et al. 2013). Of note, a so-called pedagogical learning environment (Csibra & Gergely 2011) might have afforded specialized teaching strategies that could be uniquely human and that greatly improve on imitation and self-teaching learning methods. One example is "motherese": Mothers modify their speech when speaking to young children in order to simplify their auditory processing and learning (see Pezzulo et al. 2013). This example suggests that social and interactive aspects of the learning environment are important prerequisites – or at least a useful scaffold – for speech acquisition and cultural transmission.

In sum, speech processing requires a sophisticated neuro-computational architecture in which physiologic, motoric, sensory, and social aspects mutually constrain each other and plausibly co-evolve. In addition to studying genetic determinants, it is important to recognize that speech could have found a suitable "neuronal niche" (Dehaene & Cohen 2007) in existing brain structures (cortical and subcortical) supporting skilled action. For example, speech could have re-used "generative" dynamics of such structures for imitation and self-imitation, and re-deployed existing computational resources for combinatorial processing (Chersi et al. 2014; Fadiga et al. 2009).

In parallel, speech could have found a suitable "socio-cultural niche": It could have been incubated within the sophisticated interactive and social dynamics of our species. The social context in which human speech is acquired is extremely rich, and human speech learning operates on top of the sophisticated interactive, joint action, mutual emulation, and pedagogical abilities, most of which are unique or at least much more developed in our species (Pickering & Garrod 2013; Sebanz et al. 2006). The demands of sophisticated social interactions might have contributed to transform vocalization from an initially quite limited sensorimotor feat to a powerful, open-ended instrumental tool that permits conveying rich communicative intentions and forming extremely varied cultures (Pezzulo 2012b). In turn, we should not neglect how the intertwined sensorimotor and social sides of speech had a transformative impact on the destiny of our species.

Vocal learning, prosody, and basal ganglia: Don't underestimate their complexity[1]

doi:10.1017/S0140525X13004184

Andrea Ravignani,[a] Mauricio Martins,[a,b] and W. Tecumseh Fitch[a]

[a]*Department of Cognitive Biology, University of Vienna, A-1090 Vienna, Austria;* [b]*Language Research Laboratory, Lisbon Faculty of Medicine, 1649-028 Lisbon, Portugal.*

andrea.ravignani@univie.ac.at mauricio.martins@univie.ac.at
tecumseh.fitch@univie.ac.at
http://homepage.univie.ac.at/andrea.ravignani/
www.researchgate.net/profile/Mauricio_Martins4/
http://homepage.univie.ac.at/tecumseh.fitch/

Abstract: Ackermann et al.'s arguments in the target article need sharpening and rethinking at both mechanistic and evolutionary levels. First, the authors' evolutionary arguments are inconsistent with recent evidence concerning nonhuman animal rhythmic abilities. Second, prosodic intonation conveys much more complex linguistic information

than mere emotional expression. Finally, human adults' basal ganglia have a considerably wider role in speech modulation than Ackermann et al. surmise.

While Ackermann et al.'s theory is interesting, seems plausible, and may initially appear tempting, it is based on incomplete readings of several literatures. First, it is unclear why some of their arguments should only apply to the specific instances of rhythmic and prosodic control the authors discuss or why they *fail* to apply in other animal species. Their model assumes that enhancement of in-group cooperation and cohesion was the main driving force for the evolution of speech via the intermediate step where vocal control and rhythm production would serve as chorusing and bonding tools. A key assumption is that speech would produce rhythmic abilities as an evolutionary by-product. This scenario is in line with some empirical observations (for reviews, see Fitch 2012; Geissmann 2000) and previous theoretical frameworks for the *origins of music* (Hagen & Bryant 2003; Hagen & Hammerstein 2009; Merker 2000; Merker et al. 2009). However, when applied to *language*, Ackermann et al.'s evolutionary model does not withstand cross-species validation: Many nonhuman animals exhibit rhythmic behaviors while lacking speech. Before primate rhythmic abilities can be compared with humans' at all, more evidence regarding flexibility in vocalizations' temporal patterning (Fedurek et al. 2013) and motor synchronization (Hattori et al. 2013) is needed in apes (cf. (Ravignani et al. 2013).

Evidence from non-primate species also seems to undermine Ackermann et al.'s model. Two bird species, both vocal learners, have been shown to entrain to steady pulses (Hasegawa et al. 2011; Patel et al. 2009a), supporting Ackermann et al.'s model and Patel's hypothesis, whereby auditory-motor entrainment skills would be evolutionary by-products of vocal learning abilities (Patel 2006). However, recent evidence suggests that vocal learning and rhythmic abilities might be dissociated. Sea lions, unlike seals, show no evidence of vocal learning (Janik & Slater 1997) but nonetheless can reliably synchronize their movements to a range of musical stimuli at different tempi (Cook et al. 2013). Humans and sea lions are both rhythmically skilled, but only humans evolved vocal learning and speech. Therefore, sea lions constitute outliers inconsistent with the prediction of Ackermann et al.'s model. This species evolved cognitive rhythmic abilities, without evolving speech. Invoking additional evolutionary forces and physiological mechanisms thus appears necessary: How can Ackermann et al.'s model be modified to avoid incorrectly predicting vocal learning in rhythmic-skilled species?

Second, Ackermann et al.'s model assumes that prosodic modulation of speech conveys mainly simple motivational-emotional information, and thus, that prosody and complex speech production had separate evolutionary histories. But evidence showing a tight connection between prosody and complex linguistic functions argues against this "double pathway" theory. Prosodic contour is influenced by syntactic constituent structure, semantic relations, phonological rhythm, pragmatic considerations, as well as by the length, complexity, and predictability of linguistic material (Wagner & Watson 2010). Furthermore, prosodic cues are used in childhood during acquisition of words (Christophe et al. 2008) and grammatical constructions (Männel et al. 2013), and in adulthood for syntactic processing (Christophe et al. 2008; Kjelgaard & Speer 1999; Langus et al. 2012; Wagner 2010) and word recognition (Cutler et al. 1997).

Contra Ackermann et al., such complex linguistic modulation of prosody seems to be a prerequisite for the acquisition and use of language, and this process is likely to be influenced by cognitive mechanisms specially modified in the human lineage. Comparative research on syntax precursors favors this hypothesis: The ability to assemble sequences of sounds into *hierarchical* patterns might be either human-specific, or very poorly developed in other species (Conway & Christiansen 2001; ten Cate & Okanoya 2012). Hence, developmental and comparative evidence point to a more complex cognitive integration of prosody and speech than allowed by the dual-pathway proposal of Ackermann et al. The challenge for Ackermann et al.'s theory is, therefore, to account for the modulation of prosody by human-specific cognitive functions (e.g., syntax), which are clearly not evolutionary homologues of primate emotional vocalizations controlled by the anterior cingulate cortex.

Finally, Ackermann et al. propose an ontogenetic pathway in which: (1) basal ganglia (BG) are important to generate integrated templates of orofacial and laryngeal movements during childhood, but (2) in adulthood can be retrieved from cortical areas because these motor templates become well-trained. Later in ontogeny, BG would mostly subserve the modulation of emotional prosody, and not the coordination of speech production. These claims are not supported by currently available empirical data. For instance, Ackermann et al. cite Parkinson's Disease (PD) data to support their claims that, in adults, BG lesions only impair emotional prosody. In fact, PD patients with normal cognitive functioning are more impaired in semantic fluency tasks than in phonetic fluency (Henry & Crawford 2004). Additionally, contra Ackermann et al., BG subserve complex syntactic and semantic processing in adults, with empirical findings consistent across PD (Dominey & Inui 2009; Henry & Crawford 2004; Lewis et al. 1998), BG lesion (Kotz et al. 2003; Teichmann et al. 2008; Ullman et al. 1997), and neuroimaging research (Friederici & Kotz 2003). These data suggest that in adults the BG support multiple functions relevant to spoken language, not just simple emotional prosodic modulation.

Furthermore, contrary to the developmental pathway proposed by Ackermann et al., the acquisition of novel syntactic structures in adults depends on the medial temporal cortex, and the retrieval of syntactic templates *after* thorough learning mostly recruits the BG and perisylvian structures (Ullman 2004). This evidence shows that, contra Ackermann et al., BG are active in the retrieval of over-learnt procedures. Ackermann et al. therefore need to propose alternative explanations to reconcile child and adult data concerning the function of BG.

In conclusion, to make their model robust, Ackermann et al. must modify and refine their evolutionary and mechanistic explanations, and clarify which assumptions are necessary, and which are sufficient, for their explanatory framework to hold. Is their model robust enough to stand up to the clear, strong relationship between prosody and complex linguistic functions? How can Ackermann et al.'s model account for the complex functions of BG in adulthood? If in-group cohesion had to be achieved, why was precise vocal control specifically selected for, rather than general non-vocal rhythmic abilities? These and other questions need to be addressed if Ackermann et al.'s model is to become convincing.

NOTE

1. Andrea Ravignani and Mauricio Martins contributed equally to this commentary as joint first authors.

ACKNOWLEDGMENTS

This work was supported by Fundação para a Ciência e Tecnologia grant SFRH/BD/64206/2009 (to Mauricio Martins) and European Research Council Advanced Grant 230604 SOMACCA (to Andrea Ravignani and W. Tecumseh Fitch).

Perceptual elements in brain mechanisms of acoustic communication in humans and nonhuman primates

doi:10.1017/S0140525X13004196

David H. Reser[a] and Marcello Rosa[a,b]

[a]*Department of Physiology, Monash University, Melbourne, VIC3800, Australia;* [b]*Australian Research Council Centre of Excellence for Integrative Brain Function, Monash University Node, Melbourne, VIC 3800.*

David.Reser@monash.edu
http://www.med.monash.edu.au/physiology/staff/reser.html
Marcello.Rosa@monash.edu
http://www.med.monash.edu.au/physiology/staff/rosa.html

Abstract: Ackermann et al. outline a model for elaboration of subcortical motor outputs as a driving force for the development of the apparently unique behaviour of language in humans. They emphasize circuits in the striatum and midbrain, and acknowledge, but do not explore, the importance of the auditory perceptual pathway for evolution of verbal communication. We suggest that understanding the evolution of language will also require understanding of vocalization perception, especially in the auditory cortex.

In all primate species examined so far the auditory cortex of consists of a "core" region, comprising three areas; a complex of surrounding "belt" areas, which are thought to provide more complex types of processing; and a poorly characterised "parabelt" region, which is also likely to consist of multiple areas. Although human homologues of the core, belt, and parabelt regions have been tentatively identified, the system remains best characterized in nonhuman primates. The primary auditory cortex (A1) remains the best understood component of this organization, although recent progress has also illuminated some of the functions of other core regions, including the rostral auditory area (R), the rostrotemporal auditory area (RT), and some belt areas (Bendor & Wang 2008; Petkov et al. 2006; Rajan et al. 2013).

It has been shown that trained animals can resolve temporal features of natural and synthetic human speech (Kuhl & Miller 1975), and that populations of neurons in A1 of awake, untrained monkeys exhibit responses that are consistent with categorical perceptual boundaries. In particular, the voice onset time (VOT) parameter, which distinguishes pairs of spectrally similar phonemes in many languages, elicits a characteristic pattern of excitatory and inhibitory activity in A1 of awake monkeys, which is consistent with activity recorded in the auditory cortex of human subjects undergoing intracortical preoperative epilepsy monitoring (Steinschneider et al. 2005; 2013). Although the categorical nature of the VOT parameter has come under scrutiny (Toscano et al. 2010), it remains clear that the temporal features of human speech can be modelled across species – in short, the basic apparatus employed for processing of speech sound parameters is phylogenetically conserved. Therefore, it is feasible to address questions about the interaction between vocalization production and reception in animal models, and (carefully) extrapolate those results to the process of human speech. A particularly ripe area for investigation of the interactions between speech production and perception is in the realm of affective nonverbal content. It has been suggested that monkey vocalizations are akin to the nonverbal and automatic features of human vocalizations, such as laughter (Ross et al. 2010) and infant–mother interaction vocalizations (Whitham et al. 2007), and some monkey vocalizations have rhythmic similarities to human speech (Bergman 2013).

A key feature of the vocalization production model proposed by Ackermann et al. is developmental change in the role of striatal and cortico-striatal circuits in vocal skill learning. They suggest that in early life, the cortico-striatal circuits are critical for development of motor expertise, which is essential for normal speech production. Although some evidence suggests that there is limited developmental modification of monkey vocalizations (Owren et al. 1993; Winter et al. 1973), monkeys do exhibit maturational improvement of control over call features, a form of vocal "skill" (reviewed in Fedurek & Slocombe 2011), and marmoset calls, in particular, undergo maturational change during progress toward adult communication (Pistorio et al. 2006). The motor programs that result from expert learning of speech in the Ackermann et al. model are ascribed to para- and subsylvian cortical areas, though it is unclear which areas in particular are implicated. This developmental trajectory leads to several testable hypotheses regarding the functional anatomy of auditory cortex, and of the rostral and ventral auditory and auditory associated areas of the temporal lobe in particular.

Anatomical tracing data have demonstrated that the rostral, most core area, RT, together with the belt area, RTL, have consistent monosynaptic connections with limbic areas in the rostral prefrontal, anterior cingulate, and temporal pole cortex, as well as subcortical limbic structures such as the lateral amygdala and ventral striatum (Reser et al. 2009). Moreover, evidence from electrophysiological and imaging studies in monkeys indicates that areas in the anterior and lateral temporal lobe cortex exhibit response selectivity to sounds of increasing complexity (Kikuchi et al. 2010; Kusmierek et al. 2012), including vocalizations, in preference to environmental sounds (Perrodin et al. 2011). Selective responses to speech, and, to a lesser extent, nonverbal vocalizations including laughter and baby cries, have been obtained in recordings from likely homologous areas in the human anterior superior temporal lobe (Chan et al. 2014). It is reasonable to suspect that the anatomical and functional circuits formed by these auditory areas undergo developmental modification in parallel with the vocal output circuits described by Ackermann et al., given that we and other primates are expert listeners to conspecific vocalizations, in addition to being expert producers. This proposition could be tested in longitudinal studies of sub-adult animals (e.g., via implanted electrode arrays), and by tract tracing experiments involving rostral auditory areas at different developmental stages.

Another issue with the proposed model involves where and how the learned motor programs would be stored and encoded in "para- and subsylvian" cortical areas, and how this information could be accessed by the subcortical centers controlling laryngeal and pharyngeal movements. A notable feature of the connectional anatomy of primate auditory cortex is a paucity of projections to motor cortex. While in macaques parts of the parabelt and adjacent polysensory cortex send connections to putative homologues of Broca's area, which may be classified as a premotor area, there are few or no monosynaptic projections to the cingulate or subcortical output areas which feature prominently in Ackermann et al.'s proposal. Phonological and motor aspects of speech should be considered jointly, rather than as disparate perceptual and productive components (Ziegler et al. 2012). Cortical microstimulation, as well as polysynaptic tract tracing using modified viruses make it feasible to map the connections from the auditory receptive areas to vocalization output pathways. We believe that further studies of descending cortical modulatory areas in the anterior cingulate will likely help understand the early development and evolution of language.

Vocal communication is multi-sensorimotor coordination within and between individuals

doi:10.1017/S0140525X13004202

Daniel Y. Takahashi[a,b] and Asif A. Ghazanfar[a,b,c]

[a]*Princeton Neuroscience Institute, Princeton University, Princeton, NJ 08544;* [b]*Department of Psychology, Princeton University, Princeton, NJ 08544;* [c]*Department of Ecology and Evolutionary Biology, Princeton University, Princeton, NJ 08544.*

takahashiyd@gmail.com asifg@princeton.edu
www.princeton.edu/~asifg

Abstract: Speech is an exquisitely coordinated interaction among effectors both within and between individuals. No account of human communication evolution that ignores its foundational multisensory characteristics and cooperative nature will be satisfactory. Here, we describe two additional capacities – rhythmic audiovisual speech and cooperative communication – and suggest that they may utilize the very same or similar circuits as those proposed for vocal learning.

Both speech and nonhuman primate vocalizations are produced by the coordinated movements of the lungs, larynx (vocal folds), and the supralaryngeal vocal tract (Ghazanfar & Rendall 2008). During vocal production, the shape of the vocal tract can be changed by moving the various effectors of the face (including the lips, jaw, and tongue) into different positions. The different shapes, along with changes in vocal fold tension and respiratory power, are what give rise to different sounding vocalizations. Different vocalizations (including different speech sounds) are produced in part by making different facial expressions. Thus vocalizations are inherently "multisensory" (Ghazanfar 2013).

Given the inextricable link between vocal output and facial expressions, it is perhaps not surprising that nonhuman primates, like humans, readily recognize the correspondence between the visual and auditory components of vocal signals (Ghazanfar & Logothetis 2003; Ghazanfar et al. 2007; Habbershon et al. 2013; Jordan et al. 2005; Sliwa et al. 2011) and use facial motion to more accurately and more quickly detect vocalizations (Chandrasekaran et al. 2011). However, one striking dissimilarity between monkey vocalizations and human speech is that the latter has a unique bi-sensory *rhythmicity*, in that both the acoustic output and the movements of the mouth share a 3–8 Hz rhythmicity and are tightly correlated (Chandrasekaran et al. 2009; Greenberg et al. 2003). According to one hypothesis, this bimodal speech rhythm evolved through the linking of rhythmic facial expressions to vocal output in ancestral primates to produce the first babbling-like speech output (Ghazanfar & Poeppel 2014; MacNeilage 1998). Lip-smacking, a rhythmic facial expression commonly produced by many primate species, may have been one such ancestral expression. It is used during affiliative and often face-to-face interactions (Ferrari et al. 2009; Van Hooff 1962); it exhibits a 3–8 Hz rhythmicity like speech (Ghazanfar et al. 2010); and the coordination of effectors during its production (Ghazanfar et al. 2012) and its developmental trajectory are similar to speech (Morrill et al. 2012).

Very little is known about the neural mechanisms underlying the production of rhythmic communication signals in human and nonhuman primates. The mandibular movements shared by lip-smacking, vocalizations, and speech all require the coordination of muscles controlling the jaw, face, tongue, and respiration, and their foundational rhythms are likely produced by homologous central pattern generators in the brainstem (Lund & Kolta 2006). These circuits are modulated by feedback from peripheral sensory receptors. The neocortex may be an additional source influencing orofacial movements and their rhythmicity. Indeed, lip-smacking and speech production are both modulated by the neocortex, in accord with social context and communication goals (Bohland & Guenther 2006; Caruana et al. 2011). Thus, one hypothesis for the similarities between lip-smacking and visual speech (i.e., the orofacial component of speech production) is that they are a reflection of the development of neocortical circuits influencing brainstem central pattern generators.

One important neocortical node likely to be involved in this circuit is the insula, a structure that has been a target for selection in the primate lineage (Bauernfiend et al. 2013). The human insula is involved in, among other socio-emotional behaviors, speech production (Ackermann & Riecker 2004; Bohland & Guenther 2006; Dronkers 1996). Consistent with an evolutionary link between lip-smacking and speech, the insula also plays a role in generating monkey lip-smacking (Caruana et al. 2011). It is conceivable that for both monkey lip-smacking and human speech, the development and coordination of effectors related to their shared orofacial rhythm are due to the socially guided development of the insula. However, a neural substrate is needed to link the production of lip-smack-like facial expressions to concomitant vocal output (the laryngeal source) in order to generate that first babbling-like vocal output. This link to laryngeal control remains a mystery. One scenario is the evolution of insular cortical control over the brainstem's nucleus ambiguus. The fact that gelada baboons produce lip-smacks concurrently with vocal output, generating a babbling-like sound (Bergman 2013), is evidence that a coordination between lip-smacking and vocal output may be easy to evolve.

Human vocal communication is also a coordinated and cooperative exchange of signals *between* individuals (Hasson et al. 2012). Foundational to all cooperative verbal communicative acts is a more general one: taking turns to speak. Given the universality of turn-taking (Stivers et al. 2009), it is natural to ask how it evolved. Recently, we tested whether marmoset monkeys communicate cooperatively like humans (Takahashi et al. 2013). Among the traits marmosets share with humans are a cooperative breeding strategy and volubility. Cooperative care behaviors scaffold prosocial motivational and cognitive processes not typically seen in other primate species (Burkart et al. 2009a). We capitalized on the fact that marmosets are not only prosocial, but are also highly vocal and readily exchange vocalizations with conspecifics. We observed that they exhibit cooperative vocal communication, taking turns in extended sequences of call exchanges (Takahashi et al. 2013), using conversation rules that are strikingly similar to human rules (Stivers et al. 2009). Such exchanges did not depend upon pair-bonding or kinship with conspecifics and are more sophisticated than simple call-and-responses exhibited by other species. Moreover, our data show that turn-taking in marmosets shares with humans the characteristics of coupled oscillators with self-monitoring as a necessary component (Takahashi et al. 2013) – an example of convergent evolution.

The lack of evidence for such turn-taking (vocal or otherwise) in apes suggests that human cooperative vocal communication could have evolved in a manner very different than what the gestural-origins hypotheses predict (Rizzolatti & Arbib 1998; Tomasello 2008). In this alternative scenario, existing vocal repertoires could begin to be used in a cooperative, turn-taking manner when prosocial behaviors in general emerged. Although the physiological basis of cooperative breeding is unknown (Fernandez-Duque et al. 2009), the "prosociality" that comes with it certainly would require modifications to the organization of social and motivational neuroanatomical circuitry. This must have been an essential step in the evolution of both human and marmoset cooperative vocal communication – one that may, like vocal production learning, also include changes to the cortical-basal ganglia loops as well as changes to socially related motivational circuitry in the hypothalamus and amygdala (Syal & Finlay 2011). These neuroanatomical changes would link vocalizations and response contingency to reward centers during development. Importantly, given the small encephalization quotient of marmosets, such changes may not require an enlarged brain.

Speech prosody, reward, and the corticobulbar system: An integrative perspective

doi:10.1017/S0140525X13004214

Carmelo M. Vicario

School of Psychology, Bangor University, Bangor, Gwynedd LL57 2AS, United Kingdom.

carmelo.vicario@uniroma1.it

c.m.vicario@bangor.ac.uk

http://www.bangor.ac.uk/psychology/people/profiles/carmelo_vicario.php.en

Abstract: Speech prosody is essential for verbal communication. In this commentary I provide an integrative overview, arguing that speech prosody is subserved by the same anatomical and neurochemical mechanisms involved in the processing of reward/affective outcomes.

Speech prosody can be conceptually intended as the affective dimension of verbal communication. The recognition of speech

prosody during social interactions has an adaptive function since it provides information about the speaker's intention and its emotional states, allowing an appropriate response in different situations (Frith 2009).

Current models of brain organization for prosody propose lateralized representation based on featural (timing vs. pitch) or functional (affective vs. linguistic) characteristics of prosodic material (see Sidtis et al. 2003). However, the role of subcortical structures in prosody is being increasingly described.

From the arguments provided by Ackermann et al. in the target article, one could argue that the corticobulbar tract is a key structure linking both the affective related dimensions of verbal communication and the reward system. For example, the authors discuss the pivotal role of dopamine in speech prosody. Surprisingly, no mention was made about the role of serotonin, another key monoamine of the reward system (Vicario 2013b), whose involvement at corticobulbar level has been reported in several articles (e.g., Raul 2003).

Here, I expand upon these issues by providing arguments in support of the suggestion that, in humans, speech prosody might have evolved from the basic mechanisms implied in reward/affective-related functions. In particular, I propose an integrative overview which argues that speech prosody is subserved, at least in part, by the same anatomical and neurochemical mechanisms involved in the processing of reward/affective outcomes.

Evidence is provided by the case of patients with Parkinson's disease (PD), a clinical condition characterized by a dysfunctional neurotransmission of dopaminergic and serotoninergic neural circuits (Bédard et al. 2011). PD patients are affected by a disruption of the prosodic aspect of verbal utterance (Van Lancker Sidtis et al. 2006), but these patients are known also for their swallowing disorder (Potulska et al. 2003). Interestingly, De Letter et al. (2007) have shown that Levodopa induces modifications of prosody in advanced PD. Moreover, it was recently reported that the deep brain stimulation of the subthalamic nucleus (STN), which modulates the activity of both dopaminergic (Lhommée et al. 2012) and serotoninergic (Creed et al. 2012) neurons, improves tongue force (Skodda 2012) and emotional speech (Brück et al. 2011) in PD.

Studies on animal models provide further support to this link between reward, prosody, and the corticobulbar tract. For example, Nuckolls et al. (2012) have shown that nigrostriatal dopaminergic depletion affects tongue force output. This phenomenon is probably mediated by the effect of dopamine depletion on several structures of the neural pathway connecting the tongue muscle with midbrain, for example, the nucleus tractus solitarius (NTS) (Granata & Woodruff 1982), a key area of the corticobulbar system involved in the regulation of reward-related behaviors such as food intake and swallowing. Interestingly, studies have shown that agonists and antagonists of dopamine spontaneously activate neurons in the NTS (Granata & Woodruff 1982).

Tongue muscle control might also involve serotoninergic mechanisms. For example, there is evidence that clozapine, a serotoninergic antagonist, affects lick frequency in rats (Das & Fowler 1995). Moreover, it has been documented that NTS is an important site of action for serotonin (Lam et al. 2009).

The study of animal models has also shown an involvement of both these two monoamines in reward-oriented communication. For example, the research by Huang and Hessler (2008) on male songbirds reports a relationship between direct rewarding communications (such as singing used for courtship) and dopaminergic mechanisms. These authors speculate that the ventral tegmentum (VT), a dopaminergic area which mediates rewarding and motivated behaviors (Ghanbarian & Motamedi 2013), might modulate songbirds output to the higher stereotypy typical of courtship. Interestingly, a direct connectivity between VT and NTS has recently been shown (Alhadeff et al. 2012), which suggests some influence of VT at corticobulbar level. Salvante et al. (2010) have also reported that serotonin may modulate communication in birds since it influences the effect of extrinsic social factors on their singing effort.

Finally, the work by Alipour et al. (2002) on new world monkeys, *Saguinus fuscicollis*, provides the anatomical rationale to the link between corticobulbar pathway, reward, and prosody. In fact, this study documents the existence of a direct connectivity between the motor cortical tongue area and several subcortical regions – such as the anterior cingulate cortex, the insula, the ventral putamen, the caudate nucleus, and the amygdale – involved in the processing of affective and rewarding outcomes.

A further support to the current discussion is provided by neurogenetic investigation. Ackermann et al. discuss the role of the *FOXP2* gene on verbal communication. For example, a mutation in this gene has been associated with apraxia of speech (Laffin et al. 2012; see also Vicario 2013a). Moreover, Shriberg et al. (2006) have shown that these alterations may extend to prosody. The impact of *FOXP2* on the activity of striatum (French et al. 2012) suggests the dopaminergic nature (Gale et al. 2013) of the effect played by this gene on verbal communication. However, it is known that *FOXP2* may modulate the serotoninergic activity of corticobulbar structures involved in the regulation of rewarding signals. For example, the neurons of the parabrachial nucleus (PB) of rats constitutively express the *FOXP2* transcription factor (Miller et al. 2011). PB is another key structure of the reward system; consisting of taste-responsive neurons (Simon et al. 2006), it receives an obligatory relay from the NTS, in rodents (Tokita et al. 2004).

The research described above offers an overview about the relationship between prosody, reward, and the corticobulbar system. In particular, it shows that speech prosody and reward processing may share similar neuroanatomical and neurochemical mechanisms. However, the evidence that both serotoninergic and dopaminergic mechanisms are involved in prosody, though corroborating the initial hypothesis of this commentary, poses the problem of understanding how these monoamines work. A possible insight is provided by two recent articles suggesting a distinct role for dopamine (Fiorillo 2013) and serotonin (Vicario 2013c), respectively, in the processing of reward and aversiveness. By extending this argument to speech prosody, one could speculate that dopamine subserves reward-oriented (e.g., approach) communication, while serotonin subserves punishment-oriented (e.g., threat) communication.

ACKNOWLEDGMENT

I thank Abdul Malik Ibraim and the copy editorial office for their help in checking the English used in this commentary.

Modification of spectral features by nonhuman primates

doi:10.1017/S0140525X13004226

Daniel J. Weiss,[a] Cara F. Hotchkin,[b] and Susan E. Parks[c]

[a]*Department of Psychology and Program in Linguistics, Pennsylvania State University, University Park, PA 16802;* [b]*Naval Facilities Engineering Command, Atlantic, Norfolk, VA 23508;* [c]*Department of Biology, Syracuse University, Syracuse, NY 13244.*

djw21@psu.edu
http://weisslab.weebly.com/
hotchkin.cara@gmail.com sparks@syr.edu
http://parkslab.syr.edu/

Abstract: Ackermann et al. discuss the lack of evidence for vocal control in nonhuman primates. We suggest that nonhuman primates may be capable of achieving greater vocal control than previously supposed. In support of this assertion, we discuss new evidence that nonhuman primates are capable of modifying spectral features in their vocalizations.

In discussing the modulation of acoustic call structure, Ackermann et al. question the extent to which nonhuman primates exert operant control over the spectro-temporal features of their

calls. This concern echoes a long-standing notion that nonhuman primates are capable of controlling acoustic parameters that can be modulated by changes in exhalation, such as loudness and duration, but lack control over spectral features that may require more nuanced control over the vocal apparatus (Janik & Slater 1997).

In this connection, we would like to briefly review our work on noise-induced vocal modifications in nonhuman primates. When vocalizing in noisy environments, humans and several other species (e.g., whales, bats, etc.; Hage et al. 2013; Parks et al. 2011; see review in Hotchkin & Parks 2013) involuntarily raise the amplitude of their vocalizations (i.e., the Lombard effect; Lombard 1911; Pick et al. 1989). Associated changes in vocalization duration have also been documented in humans and nonhumans (e.g., Garnier et al. 2010). Some species also modify spectral features by shifting energy to higher harmonics (reviewed in Hotchkin & Parks 2013). However, consistent with Ackermann et al.'s concerns, previous studies with nonhuman primates have suggested that they may be restricted to manipulating the amplitude and temporal aspects of their calls in response to noise in the transmission environment (Egnor & Hauser 2006; Sinnott et al. 1975). To the best of our knowledge, no nonhuman primate has previously demonstrated the ability to modify spectral features of their calls in response to noise.

We have recently discovered this ability in cotton-top tamarins (*Saguinus oedipus*), a small arboreal New World species known to have an extensive vocal repertoire (Cleveland & Snowdon 1982). A previous study with this species found that they modify the amplitude and duration of their calls in response to noise (Egnor & Hauser 2006), and, using a different method, we have determined they are also capable of modifying spectral components of their calls. Our findings suggest the possibility for greater vocal control in nonhuman primates than previously supposed (e.g., Janik & Slater 1997).

Unlike prior studies with nonhuman primates (Brumm et al. 2004; Egnor & Hauser 2006; Sinnott et al. 1975), our study used playbacks of both broad- and narrow-band white noise at a range of amplitudes to investigate vocal control of two call types, chirps and combination long calls (CLCs; Hotchkin 2012; Hotchkin et al. 2013). We measured a variety of acoustic parameters including spectral tilt (a relative measure of the distribution of energy between the high- and low-frequency components of the call), which is an acoustic parameter that has been studied almost exclusively in human responses to noise (e.g., Lu & Cooke 2009). We found that individuals modified the structure of both call types in response to changes in noise amplitude and bandwidth. In CLCs, whose frequencies were overlapped by the noise, peak fundamental frequency and spectral tilt changed in response to increased noise amplitude and bandwidth (see our Fig. 1). While the noise stimuli did not overlap with the chirp frequencies, this call type also showed changes to frequency content during noise. Noise with frequency components at or slightly below vocalization frequency may result in masking, due to a phenomenon known as the upward spread of masking which has been observed in both humans and nonhuman mammals (Egan & Hake 1950; Nachtigall et al. 2004). Increases in chirp frequency may provide release from masking by low-frequency noise, thereby improving the detectability even when noise frequencies do not overlap the vocalization. In chirps, the peak and maximum components of the fundamental frequency increased as a result of noise level, with no changes to spectral tilt. Other vocal modifications observed included the Lombard effect (i.e., an increase in amplitude) and longer chirp duration.

The focus of the authors is primarily on volitional control and modification of vocalizations in nonhuman primates, and they could therefore dismiss this finding because responses to noise are thought to reflect involuntary processes, as noted above. In fact, we do not argue that the flexibility exhibited in the Lombard effect and conditioning studies necessarily set the stage for the evolution of vocal flexibility as it is manifest in humans (as noted by Owren et al. 2011). However, in order to refine the dual stage hypothesis, it is important to fully describe the range of flexible features available in nonhuman primate vocal communication. Research with cotton-top tamarins alone has demonstrated they are capable of producing long-term changes to the acoustic structure of their calls (i.e., vocal convergence; see Weiss et al. 2001); perceiving changes to the spectral features of the harmonics in their CLCs (Weiss & Hauser 2002); modifying the timing of their calls (Egnor et al. 2007);

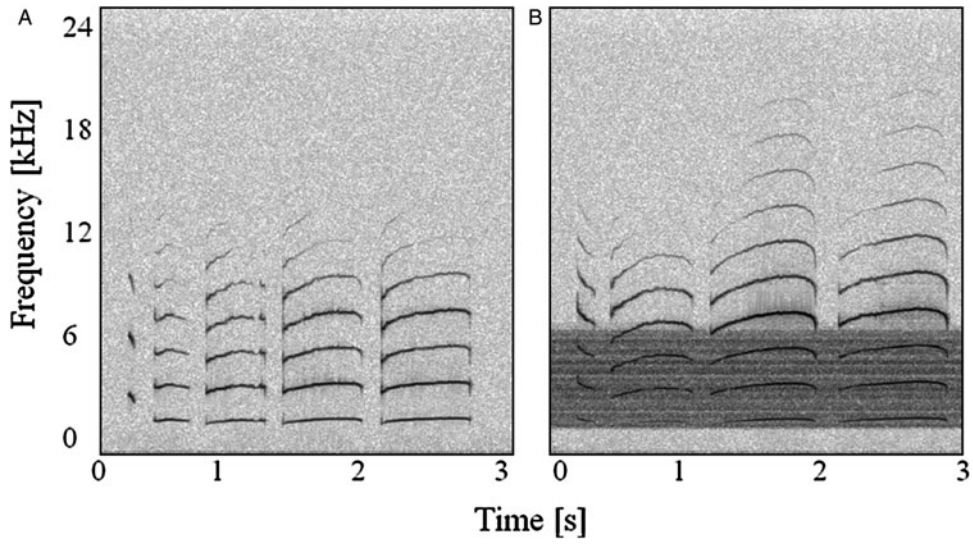

Figure 1 (Weiss et al.). Representative CLCs produced during (a) control and (b) treatment trials demonstrating changes to spectral tilt. All whistles from (a) have strong fundamental frequencies and maximum energy in the 2nd harmonic, while in (b) the first whistle has a very faint fundamental frequency at approximately 2 kHz, and peak frequencies for all whistles occur in the 4th harmonics. Reduced energy in the fundamental frequency is also apparent in the second and third whistles. Spectrogram parameters: 1024 point Hamming window, 75% overlap, 11.7 Hz frequency resolution.

and also altering the loudness, duration, and spectral components of their calls when transmitting signals in noisy environments (Egnor & Hauser 2006; Egnor et al. 2006; Hotchkin 2012). Further, it has been suggested that multiple levels of vocal control may be active during Lombard vocalizations and could involve a complex array of neural structures extending beyond brainstem reflexes (Eliades & Wang 2012). Thus, the main implication of our findings for the target article is to indicate that nonhuman primates may possess greater vocal control than has previously been supposed. Developing a more complete understanding of the ways in which nonhuman primates are capable of manipulating their vocalizations, and the supporting neural networks, may ultimately help Ackermann et al. further refine their theory.

Contribution of the basal ganglia to spoken language: Is speech production like the other motor skills?

doi:10.1017/S0140525X13004238

Alexandre Zenon[a] and Etienne Olivier[a,b]

[a]*Institute of Neuroscience, School of Medicine, Université catholique de Louvain, UCL, B-1200 Brussels, Belgium;* [b]*FondazioneIstituto Italiano di Tecnologia (IIT), 16163 Genova, Italy.*

alexandre.zenon@uclouvain.be etienne.olivier@uclouvain.be

www.coactionslab.com

Abstract: Two of the roles assigned to the basal ganglia in spoken language parallel very well their contribution to motor behaviour: (1) their role in sequence processing, resulting in syntax deficits, and (2) their role in movement "vigor," leading to "hypokinetic dysarthria" or "hypophonia." This is an additional example of how the motor system has served the emergence of high-level cognitive functions, such as language.

Besides the well-known contribution of the basal ganglia (BG) to motor behaviour, the numerous language deficits reported in patients with Parkinson's disease (PD), or with other BG lesions, suggest they participate in language production, as Ackermann et al. discuss in the target article. However, most inferences about the contribution of BG to motor control based on deficits observed in patients have proven to be flawed, and, despite decades of investigation, the actual role of BG remains debated. For example, recent studies in humans and monkeys have shown that a lesion of the internal part of the globus pallidus – one of the main BG outputs – leads to rather subtle motor deficits, mostly unrelated to PD motor symptoms (Desmurget & Turner 2008; 2010; Obeso et al. 2009). This suggests that many of the symptoms resulting from BG lesions, including language deficits, are likely to result from the perturbation, by noisy BG signals, of a larger network of cortical and subcortical structures.

Nonetheless, two motor functions have emerged recently as being distinctively imputable to BG. First, it appears almost undisputable that BG contribute to motor sequence learning (Desmurget & Turner 2010; Obeso et al. 2009; Turner & Desmurget 2010). This is in sharp contrast to the idea that BG are involved in the storage and execution of overlearned movements, or "habits" (Graybiel 2008). Interestingly, a recent study has shown that BG are involved in motor chunking (Wymbs et al. 2012), as already suggested by the finding that this process is dopamine-dependent (Boyd et al. 2009; Tremblay et al. 2010; 2009). Chunking is a key mechanism in sequence learning, and it comprises two distinct processes occurring at different stages: a first operation, called "segmentation," consists of parsing sequences into shorter clusters (Clerget et al. 2012; Sakai et al. 2003; Verwey 2001; Verwey & Eikelboom 2003), and this is followed by "concatenation," which consists of assembling these short chunks into longer clusters (Sakai et al. 2003; Verwey 1996). Whereas BG would be involved in the concatenation process, the initial segmentation would rely on a left fronto-parietal network, including Broca's area (Wymbs et al. 2012). The second function currently assigned to BG is the control of movement "vigor" according to motivational factors (Kurniawan et al. 2010; Mazzoni et al. 2007; Turner & Desmurget 2010). Indeed, a loss of dopaminergic neurones in the *substantia nigra* affects the reward and decision-making processes (Mimura et al. 2006). Bigger rewards typically lead to higher efforts (Berridge 2004; Schmidt et al. 2008; Takikawa et al. 2002) and, when selecting an action among multiple candidates, the choice critically depends on a comparison of the cost/benefit ratio for each option (Niv et al. 2006). The motivational effect that an anticipated reward has on the execution of an action is known as the "incentive salience," which is then translated into movement "vigor" to optimize the balance between the effort invested and the outcome value. An impairment of this process would lead to an inability to adjust the level of effort invested in movements, explaining the occurrence of bradykinesia in PD patients (Mazzoni et al. 2007).

It is remarkable that two of the main roles assigned to BG in spoken language by Ackermann and colleagues match very well with the contribution of BG to motor behaviour. First, following perinatal BG lesions or in patients with inherited language disorders, one major spoken language impairment is "a significant disruption of simultaneous or sequential sets of motor activities to command, in spite of a preserved motility of single vocal tract organs" (target article, sect. 4.2.1, para. 2) (Alcock et al. 2000a; Watkins et al. 2002a), which is consistent with the role of BG in processing structured sequences. Intriguingly, these patients also present a deficit in grammatical rules acquisition (Alcock et al. 2000a; Gopnik 1990a), as also reported in PD patients (Chan et al. 2013; Pell & Monetta 2008). It is noteworthy that the other brain region critically involved in syntax and sequence processing is Broca's area (Clerget et al. 2009; 2011; 2013; Fadiga et al. 2009; Tettamanti & Weniger 2006), which is tightly interconnected with BG (Ullman 2006). It is remarkable that the two brain regions involved in chunking (Wymbs et al. 2012) are also those in which a lesion yields a syntax deficit suggesting that the chunking process might provide the basis for hierarchical processing and represent the common denominator of BG contribution to motor and language production (Kotz et al. 2009).

Second, the "hypokinetic dysarthria" or "hypophonia" reported in PD, and viewed by Ackermann et al. as a consequence of "a diminished impact of motivational, affective/emotional, and attitudinal states on the execution of speech movements, leading to sparse motor activity" and interpreted as a "general loss of 'motor drive' at the level of the speech motor system" (sect. 4.2.2, para. 3) fits quite well with the view that BG play a central role in controlling the movement "vigor" according to motivational factors (Turner & Desmurget 2010). According to this view, bradykinesia and hypophonia would not arise from a mere impairment of the motor command execution but from a loss of influence of the motivational drive on the motor output. Ackermann and colleagues go one step further in proposing that, in the case of language, this impairment would not only result in a reduced amplitude of the motor output but also in a decreased motivational and emotional modulation of speech. This could be paralleled, in the context of motor behaviour, to the relative lack of spontaneous facial expression exhibited by PD patients despite their preserved ability to produce posed facial expression (Smith et al. 1996). Even though there are currently very little data supporting this hypothesis, this is an interesting idea that deserves further investigation.

The similitude between the putative functions of BG to motor behaviour and language provides an additional example of how the motor system has served the emergence of high-level cognitive functions, by minimal transformation, from ancestral structures already present in nonhuman primates (Andres et al. 2008; Dehaene & Cohen 2007; Olivier et al. 2007).

Authors' Response

Phylogenetic reorganization of the basal ganglia: A necessary, but not the only, bridge over a primate Rubicon of acoustic communication

doi:10.1017/S0140525X1400003X

Hermann Ackermann,[a] Steffen R. Hage,[b] and Wolfram Ziegler[c]

[a]*Neurophonetics Group, Centre for Neurology – General Neurology, Hertie Institute for Clinical Brain Research, University of Tuebingen, D-72076 Tuebingen, Germany;* [b]*Neurobiology of Vocal Communication Research Group, Werner Reichardt Centre for Integrative Neuroscience, and Institute for Neurobiology, Department of Biology, University of Tuebingen, D-72076 Tuebingen, Germany;* [c]*Clinical Neuropsychology Research Group, Municipal Hospital Munich-Bogenhausen, D-80992 Munich, and Institute of Phonetics and Speech Processing, Ludwig-Maximilians-University, D-80799 Munich, Germany.*

hermann.ackermann@uni-tuebingen.de
steffen.hage@uni-tuebingen.de
wolfram.ziegler@extern.lrz-muenchen.de
http://www.hih-tuebingen.de/neurophonetik/
http://www.vocalcommunication.de
http://www.ekn.mwn.de

Abstract: In this response to commentaries, we revisit the two main arguments of our target article. Based on data drawn from a variety of research areas – vocal behavior in nonhuman primates, speech physiology and pathology, neurobiology of basal ganglia functions, motor skill learning, paleoanthropological concepts – the target article, first, suggests a two-stage model of the evolution of the crucial motor prerequisites of spoken language within the hominin lineage: (1) monosynaptic refinement of the projections of motor cortex to brainstem nuclei steering laryngeal muscles, and (2) subsequent "vocal-laryngeal elaboration" of cortico-basal ganglia circuits, driven by human-specific *FOXP2* mutations. Second, as concerns the ontogenetic development of verbal communication, age-dependent interactions between the basal ganglia and their cortical targets are assumed to contribute to the time course of the acquisition of articulate speech. Whereas such a phylogenetic reorganization of cortico-striatal circuits must be considered a necessary prerequisite for ontogenetic speech acquisition, the 30 commentaries – addressing the whole range of data sources referred to – point at several further aspects of acoustic communication which have to be added to or integrated with the presented model. For example, the relationships between vocal tract movement sequencing – the focus of the target article – and rhythmical structures of movement organization, the connections between speech motor control and the central-auditory and central-visual systems, the impact of social factors upon the development of vocal behavior (in nonhuman primates and in our species), and the interactions of ontogenetic speech acquisition – based upon *FOXP2*-driven structural changes at the level of the basal ganglia – with preceding subvocal stages of acoustic communication as well as higher-order (cognitive) dimensions of phonological development. Most importantly, thus, several promising future research directions unfold from these contributions – accessible to clinical studies and functional imaging in our species as well as experimental investigations in nonhuman primates.

R1. Introduction

The 30 commentaries have elaborated upon all aspects of the target article, extending from vocal behavior in nonhuman primates to speech physiology and pathology, the neurobiology of basal ganglia functions, as well as motor skill learning and paleoanthropological concepts. In particular, the following issues have been addressed: (i) the capacities of nonhuman primates to control vocal behavior and to produce species-atypical calls; (ii) the constraints of vocal tract anatomy on vocalizations; (iii) the scope of birdsong as a model of – at least some aspects of – human spoken language; (iv) the relationship of the *FOXP2* gene to motor functions – or, more specifically – vocal behavior across mammalian and avian taxa; (v) the contribution of corticobulbar tracts and brainstem central pattern generators – besides and beyond the basal ganglia – to acoustic human communication; (vi) the rhythmic organization and oscillatory underpinnings of behavior; (vii) the impact of auditory and audiovisual information as well as social factors on speech acquisition; (viii) the interactions of motor speech learning with preceding subverbal stages of acoustic communication; (ix) the contribution of cortico-striatal circuitry to "speech learning" in adulthood; (x) the broad range of cognitive basal ganglia functions beyond vocal-emotional expression and motor aspects of language; and, finally, (xi) paleoanthropological aspects of the target article such as the benefits of the initial articulatory efforts of our species and the speaking capabilities of Neanderthals.

We gratefully appreciate all the contributions which have helped us to further specify our argument and have broadened our view on primate acoustic communication – in extant nonhuman cousins, extinct relatives from the genus *Homo*, and in our own species. In this response, we have organized the various commentaries into four broad subject areas: (a) nonhuman primate vocal behavior (and birdsong), which we discuss in section R2; (b) contributions of the basal ganglia to mature spoken language production/affective-vocal behavior (sect. R3); (c) role of the basal ganglia in ontogenetic speech acquisition (sect. R4); and (d) paleoanthropological perspectives of articulate speech acquisition (sect. R5). In the concluding section, R6, we summarize some of the main points/key questions likely to be entailed in further investigations of the phylogenetic reorganization of the basal ganglia.

R2. Nonhuman primate vocal behavior: An underestimated or an inadequate vantage point for models of spoken language evolution?

R2.1. Volitional control of vocal behavior in nonhuman primates

Based upon a review of the behavioral organization and the neuroanatomic underpinnings of acoustic communication in nonhuman primates, we proposed in the target article that these species lack the capacity "to combine laryngeal and orofacial gestures into novel movement sequences" (sect. 2.3), rendering them virtually unable to generate even the simplest speech-like vocal emissions, that is, acoustic events in the form of one or more syllable-shaped signal pulses. Several commentaries suggest that we might have underestimated the versatility of vocal functions in our primate relatives:

1. For example, commentators **de Boer & Perlman** report that Koko, a human-reared female gorilla, learned to display some species-atypical vocalizations ("breathy

grunt-like vocalizations" and "mock 'coughs'"), indicating at least rudimentary voluntary laryngeal control. Comparable observations of species-atypical acoustic events ("extended grunts") in captive chimpanzees, often as a component of multimodal and intentional display scenarios, are mentioned in the commentary by **Meguerditchian, Taglialatela, Leavens, & Hopkins** (**Meguerditchian et al.**)

2. Recent experiments by **Weiss, Hotchkin, & Parks** (**Weiss et al.**) found modification of the spectral structure ("spectral tilt") of the vocalizations of cotton-top tamarins under specific conditions such as a noisy environment.

3. Finally, **Lameira** points at an eventually salient role of the voiceless calls of great apes in speech evolution, which are "underlined by voluntary control and maneuvering of supra-laryngeal articulators (...) in apparent homology to the articulatory movements of voiceless consonants."

We readily admit the existence of – though highly limited – volitional control over some aspects of vocal behavior in nonhuman primates. In fact, recent studies by one of us (Hage, and colleagues) show that rhesus monkeys are capable of volitionally initiating vocal output, that is, able to switch between two distinct call types from trial to trial in response to different visual cues in an operant conditioning task (Hage & Nieder 2013; Hage et al. 2013). Furthermore, single-cell recordings identified neurons in the monkey homologue of human Broca's area – located within the ventrolateral prefrontal cortex – that specifically predict such volitionally triggered calls, suggesting a crucial engagement of the monkey homologue of human Broca's area in vocal initiation processes, a putative precursor for speech control in the primate lineage (Hage & Nieder 2013).

However, such preadaptations of human vocal tract motor control in our nonhuman relatives do not pose a threat to our model. To the contrary, a complete absence of any precursors would raise the question of how the suggested *FOXP2*-driven reorganization of cortico-striatal circuits could have gained a foothold in the primate "communicating brain" in the first place. At the laryngeal level, nevertheless, learned species-atypical sounds are restricted to breathy-voiced (**de Boer & Perlman**) or extended grunts (**Meguerditchian et al.**). These vocalizations, therefore, lack a property which we consider essential to the communicative efficiency and the generative potential of the sound structures prevailing in all spoken languages, that is, the syllabic patterning of vocal tract movement sequences. This specific compositional principle requires the control of the laryngeal sound source to become part of a meshwork of phonetic gestures which are organized – on the basis of precisely defined phase-relationships – as syllable-shaped gestural scores (e.g., Goldstein et al. 2006; see Figure 2C of the target article).

Besides changes in spectral call features, the experiments by **Weiss et al.** – referred to in their commentary – gave rise to an increase in vocal amplitude in response to noise (Lombard effect). Under these conditions, modifications of call amplitude and spectral structure, conceivably, are rooted in a common cerebral mechanism and, thus, may represent components of a multifaceted vocal response pattern. Most probably, the Lombard effect – and its associated acoustic sequels – reflects involuntary changes of several call parameters such as amplitude, duration, repetition rate, and spectral composition in response to masking ambient noise rather than volitionally controlled modification of vocal output (e.g., Brumm & Slabbekoorn 2005; Brumm & Zollinger 2011). Recently, Hage and colleagues reported such vocal shifts to show an extremely brief delay and to emerge at a latency of less than a hundred milliseconds after noise onset (Hage et al. 2013). Taking into account that single neurons in the periaqueductal gray (PAG) change their vocalization-related firing rates already around 400 msec prior to call onset (Larson & Kistler 1984), these results indicate that the Lombard effect – and at least one of its acoustic correlates – might be controlled by a neuronal network located within the brainstem rather than by superordinate higher-order brain structures. Furthermore, modifications of the spectral features of vocal output such as those reported by Weiss et al. might be caused by alterations of an animal's motivational state under different noise conditions. A study in squirrel monkeys has, for example, found an increase in aversion to be correlated with an upward shift of the maximal energy of the power spectrum of some call types (Fichtel et al. 2001). Taken together, changes in call structures do not necessarily point at specific volitional control capabilities, but may be mediated by lower-level brainstem mechanisms.

R2.2. Auditory-motor interactions in nonhuman (and human) primates

Reser & Rosa call attention to the tight relationship between perception and production of species-typical vocal behavior in nonhuman primates. Most importantly, "the basic apparatus employed for processing of speech sound parameters is phylogenetically conserved" and, thus, available to our cousins as well. As a hint towards tight connections between the auditory and the motor domains of human vocal behavior, specific motor circuits have been found to be recruited during the analysis of speech sound features, as described in the commentary by **Pezzulo, Barca, & D'Ausilio** (**Pezzulo et al.**). Besides frontal cortex, subcortical structures may contribute to these encoding processes as well (Ackermann & Brendel, in press).

More specifically, speech acquisition represents a variant of "vocal production learning," that is, the capacity "to reproduce by voice patterns of sound *first received by ear*," as **Merker** writes (italics ours), and, therefore, must be expected to involve tight auditory-motor interconnections. However, the target article focuses on the *motor* side of vocal *production* learning, and herein rests, in our view, a major obstacle for speech acquisition in nonhuman primates (see also sect. R4 here). Nevertheless, as alluded to by **Reser & Rosa**, studies of the connections between central-auditory and central-motor systems in nonhuman primates, including limbic structures, should provide further opportunities for an elucidation of language evolution. As a highly intriguing aspect of the perception-production links within the domain of musicality, **Honing & Merchant** discuss the differential sensitivity to rhythm and beat in nonhuman primates as a basis for the proposed gradual audiomotor evolution hypothesis (see also the commentary by **Ravignani, Martins, & Fitch** [**Ravignani et al.**]).

R2.3. Rhythmical entrainment and interlocutor coordination as speech precursors?

Takahashi & Ghazanfar and **Bryant** have contributed two elucidating commentaries which suggest a precursor role of rhythmical facial activities and rhythmically entrained vocal and non-vocal behaviors in nonhuman primates for the rhythmical organization of verbal utterances, on the one hand, and for the coordination of interlocutors in human conversation, on the other. This notion conforms to recent phonetic accounts of speaking as a quasi-rhythmically entrained motor activity (e.g., Cummins 2009), interlinked with rhythmical principles engaged in the organization of auditory speech perception (Rothermich et al. 2012). Thus, Peelle and Davis (2012) consider slow oscillatory activity of cortical neuronal assemblies as a physiological basis for the processing of quasi-rhythmical structures in speech comprehension, and Wilson and Wilson (2005) provided an oscillator model of the turn-taking behavior of speakers during conversation. Hence, the rhythmical entrainment approach embarks on a close interlacing of vocal tract motor mechanism with auditory-perceptual processes in speech, and relates it to the cooperative nature of linguistic interactions. Allusions to the rhythmicity of spontaneous and posed laughter and to the role of laughter "in coordinating conversational timing," as highlighted by Bryant, point at a deeply entrenched rhythmical basis of verbal utterances. Besides brainstem centers and the insula (see comments by Takahashi & Ghazanfar), most importantly, clinical and functional imaging studies in humans suggest the rhythmical organization of verbal vocal behavior to be associated with the basal ganglia (e.g., Ackermann et al. 1997b; Konczak et al. 1997; Riecker et al. 2002). Furthermore, rhythmical entrainment processes during speech production may serve as a target of therapeutic intervention techniques in speech-disordered patients (e.g., Brendel & Ziegler 2008). So far, nevertheless, "very little is known about the neural mechanisms underlying the production of rhythmic communication signals in human and nonhuman primates" (as Takahashi & Ghazanfar point out in their commentary), and this issue, surely, deserves further investigations.

The commentary by **Takahashi & Ghazanfar** draws attention to, among other things, experimental work on lip-smacking in nonhuman primates, an emotional social signal whose frequency largely corresponds to the syllabic rhythm of human speech. It is an intriguing idea – and a valuable expansion of the frame/content concept developed by MacNeilage and Davis (2001; see also MacNeilage 1998; 2008) – that the superimposition of a voice signal onto the lip-smacking cycle in gelada baboons has rendered this social signal audible and may, thereby, have paved the way for the evolution of speech as a rhythmical oral-facial-laryngeal activity within auditory-visual displays. From the perspective of the model developed in our target article, however, the notion of two parallel layers of lip-smacking and vocalization behavior still lacks an important ingredient: The phonatory mechanisms generating the voice signal during speaking involve a precisely timed and smooth interaction of laryngeal gestures with the movements of supralaryngeal movements as sketched in Figure 2C of the target article. Considering the "inextricable link between vocal output and facial expressions" mentioned by Takahashi & Ghazanfar, comparative investigations of the neural bases of vocal behavior and non-vocal facial expression are definitely warranted. As noted in the commentary by **Meguerditchian et al.**, the vocal behavior of chimpanzees is associated, depending upon communicative content, with differential orofacial motor asymmetries.

R2.4. Commonalities between birdsong and human spoken language: A more adequate vantage point for scenarios of spoken language evolution?

Apart from a brief final paragraph related to birdsong, the target article focuses on precursors of spoken language within the primate clade, trying to "delineate how these remarkable motor capabilities [underlying speech production] could have emerged in our hominin ancestors" (target article, Abstract). Four commentaries plead for a broader perspective, including, especially, avian vocal behavior (**Beckers, Berwick, & Bolhuis** [**Beckers et al.**], **Merker**, **Petkov & Jarvis**, **Pezzulo et al.**). Beckers et al. even raise the concern that – with respect to speech and language – "common descent may not be a reliable guiding principle for comparative research" and, most importantly, that this approach may miss the unique aspects of language per se, "given the already strong parallels between humans and songbirds in terms of auditory-vocal imitation learning, and the often remarkable articulatory skills in many avian species" (see also the first paragraph of the commentary by Merker for a similar argument). It goes without saying that a broader perspective would have provided a more elucidating scenario, and might have helped to define the major constraints acting upon speech evolution mechanisms and to narrow down research questions in primate studies. But all the commonalities between human verbal communication and the acoustic behavior of non-primate mammals or songbirds cannot dispense us from the challenge of clarifying – in sufficient detail – how highly vocal, but speechless primates have ultimately acquired the unique motor capabilities that enable us to gossip in well-articulated utterances. As a matter of fact, "there is little direct comparative evidence in the primate literature to suggest that the cortico-striatal-thalamic system is strikingly different in humans relative to nonhuman primates" (Petkov & Jarvis). In our proposal, the differences are restricted to the vocal domain and involve a – within the primate lineage – human-specific vocal elaboration of otherwise primate-general cortico-striatal circuits, allowing for the sequencing of laryngeal and supralaryngeal gestures according to auditory templates (see comments of **Zenon & Olivier** for a discussion of sequencing as a basic basal ganglia function, see also **Lieberman**'s commentary).

R3. The basal ganglia in mature speech production and affective-vocal behavior: A major player or a negligible factor?

Based upon behavioral and neurobiological data obtained from nonhuman primates and from our species, we have argued for a crucial role of the basal ganglia during mature speech production in terms of the implementation of emotive prosody, that is, the "affective tone" of verbal utterances. A series of recent functional imaging studies,

indeed, provides further evidence for an engagement of the basal ganglia in affective-vocal behavior, as highlighted in the commentary by **Frühholz, Sander, & Grandjean** (**Frühholz et al.**). However, we are by no means suggesting that basal ganglia functions are restricted to "just simple emotional prosodic modulation" – a critical objection brought forward by **Ravignani et al.** By contrast, we fully acknowledge that "the basal ganglia support multiple functions relevant to spoken language" and that, more specifically, these subcortical structures must be expected to engage in "complex syntactic and semantic processing in adults" (see fifth paragraph of the commentary by Ravignani et al.). Against the background of several parallel but interacting basal ganglia loops, including limbic, motor, and cognitive components (see, e.g., Fig. 3 of the target article), multiple contributions of the basal ganglia to speech and language are not only conceivable, but must even be expected. Thus, we agree that syntactic (Teichmann et al. 2005; Ullman 2001) and semantic (Cardona et al. 2013) processes may hinge upon corticostriatal circuits (see also our response to **Lieberman** in subsequent paragraphs).

Furthermore, the target article by no means "assumes that prosodic modulation of speech conveys mainly simple motivational-emotional information" – a concern raised by **Ravignani et al.** (see Note 1 of the target article). We excluded linguistic prosody from our review because the modulation of prosody by human-specific cognitive functions (e.g., syntax) is, most presumably, a component of the left-hemisphere language system and must be strictly separated – both at the functional and the neuroanatomic level – from emotive prosody (see, e.g., Sidtis & Van Lancker Sidtis 2003). As a consequence, we fully support the suggestion that linguistic prosody is related to "human-specific cognitive functions," which – in contrast to emotional tone – "are clearly not evolutionary homologues of primate emotional vocalizations" (Ravignani et al.).

The first part of **Lieberman**'s comments also raises a strong argument for a broad variety of motor, cognitive, and behavioral functions of the basal ganglia, based upon "a network of segregated cortical-to-basal neural circuits linking areas of motor cortex and prefrontal cortex." The common basic operation across these domains seems to be the task-dependent "switching" between motor and cognitive responses or movements during "internally guided acts." Section 4.3.1. of the target article pays full credit to this firmly established model. Nevertheless, more recent work shows that interconnections between these loops are also of considerable importance (see Fig. 3 of the target article), especially in order to better understand the striatal interface of emotional/motivational and motor functions as well as the psychomotor aspects of striatal disorders (see, e.g., Jankovic 2008).

While we support the main thrust of **Lieberman**'s argument, we have some concerns over the clinical data referred to, that is, the contention that the "speech production deficits of Parkinson's disease and focal lesions to the basal ganglia are qualitatively similar to ones occurring in aphasia." As regards speech motor impairments in a narrow sense, there is definitely no similarity between Parkinson's dysarthria, on the one hand, and speech apraxia or phonological impairments after left anterior cortical lesions, on the other. We acknowledge that disorders of the basal ganglia have been observed to give rise – though rather infrequently – to mostly transitory syndromes of an aphasia (but not compromised speech), and the concept of "subcortical aphasia" has been widely acknowledged. Nevertheless, any interpretation of these findings in terms of the relevant functional-neuroanatomic substratum must take into account alternative interpretations. First, left-hemispheric subcortical lesions may give rise to diaschisis effects within the overlying fronto-temporal cortex, that is, hypometabolism – and subsequent dysfunction – of the perisylvian "language zones" (Weiller et al. 1993). Second, more advanced stages of Parkinson's disease and so-called atypical Parkinsonian syndromes may be associated with damage to cortical areas affecting, eventually, language functions.

A further critical comment put forward by **Ravignani et al.** also relates to the role of the basal ganglia in higher-order language processing. Based on experiments probing the learning of novel syntactic structures in adults, they claim that these subcortical nuclei engage in the retrieval – rather than the acquisition – of overlearned procedures, implicitly suggesting that a similar relationship should hold for motor speech processes as well. Yet, the short-term encoding of artificial syntactic structures under experimental conditions in adulthood and their subsequent retrieval are not necessarily the same thing as the long-term acquisition of speech motor routines during infancy and childhood, and their retrieval in adults need not depend on the same cerebral network. These suggestions could explain why the findings of novel syntax learning experiments are not compatible with the clinical data obtained from speech-disordered infants and adults cited in our target article (sect. 4.3.2.), which demonstrate that the engagement of the basal ganglia declines – though it does not necessarily cease – across the time course of speech acquisition.

Commentators **Hasson, Llano, Miceli, & Dick** (**Hasson et al.**) raise principal concerns over the "viability of BG [basal ganglia] as a speech/emotion synthesizer," since these subcortical structures lack "the capacity to monitor and correct for related errors, that is, evaluate that the intended emotive tone/prosody was instantiated." They argue that: (i) The basal ganglia cannot provide the necessary fast auditory feedback; (ii) processing of emotive prosody is mainly bound to lateral-temporal systems of the cortex; and (iii) basal ganglia dysfunctions fail to compromise the perception of "emotional speech variations." Parenthetically, it is indeed the case that patients with Parkinson's disease – at least in more advanced stages – may show impaired emotion recognition (see, e.g., Breitenstein et al. 1998). More importantly, however, the basic premise of the argument is – in our view – unwarranted. We by no means want to curtail the relevance of (auditory) feedback within the domain of (speech) motor control, but why must the basal ganglia – in order to implement emotive prosody – be embedded into a "fast" feedback loop? Rather, as suggested by **Frühholz et al.**, the "temporal slow prosodic modulations of emotional speech … seem to rely on feedback processing in the AC [auditory cortex]." But whatever the role of auditory feedback within the area of vocal-emotional processing, the suggestions of Hasson et al. are at variance with a solid tradition of clinical neurology. All Parkinsonian symptoms are, for example, "dependent on the emotional state of the patient" (Jankovic 2008). Based upon, among other things, such

observations, it is widely acknowledged that the basal ganglia operate as a dopamine-dependent interface between the limbic system and various motor areas (see, e.g., Mogenson et al. 1980, referred to in sect. 4.2.2. of the target article). Vocal-affective expression represents just one aspect of this broader spectrum of psychomotor basal ganglia functions (the second part of the commentary by **Zenon & Olivier** provides a lucent account of these relationships). The projections from the limbic to the motor basal ganglia loop can be considered the neurobiological substratum of psychomotor interactions, and this circuitry represents – contrary to the claims by Hasson et al. – a relatively well-established functional-neuroanatomic model at this time, extending from the level of systems physiology to the level of molecular biology (see sect. 4.3. of the target article).

Besides the structures depicted in Figure 4 of the target article, which centers around the basal ganglia, further cortical and subcortical structures engage in speech motor control or, more generally, contribute to verbal communication – such as the anterior cingulate cortex (briefly referred to in the last part of sect. 4.3.1. of the target article), rostral parts of the inferior frontal gyrus, auditory cortical areas, and the cerebellum (see **Frühholz et al.**'s commentary). Whereas these regions do not play a significant role in our argument, we, nevertheless, highly appreciate Frühholz et al.'s Figure 1, which incorporates the afore-mentioned structures into Figure 4 of the target article. Interestingly, both Hasson et al.'s and Frühholz et al.'s commentaries proffer the cerebellum – rather than the basal ganglia – as the region most likely to "imbue speech with emotive content" (Hasson et al.'s phrase for the role these authors see us attributing to the BG). A significant contribution of the "small brain" to speech motor control is beyond any dispute (Ackermann 2008), though, in parentheses, the cerebellum does not appear to pertain to the cerebral network underlying acoustic communication in nonhuman primates (e.g., Kirzinger 1985). However, cerebellar disorders do not give rise to a constellation of motor aprosodia, that is, a monotonous and hypophonic voice lacking affective deflections as in Parkinson's disease (for reviews, see Ackermann & Brendel, in press; Ackermann et al. 2007). Instead, the syndrome of ataxic dysarthria is predominantly characterized by articulatory deficits with irregular distortions of consonants and vowels. The cerebellar cognitive affective syndrome – referred to by Hasson et al. – has been reported, admittedly, to comprise abnormalities of speech prosody in terms of a high-pitched voice of a "whining, childish and hypophonic quality," emerging, especially, in bilateral or generalized disease processes (Schmahmann & Sherman 1998, p. 564; eight patients out of a total of 20 subjects with cerebellar pathology). Most presumably, these perceived voice abnormalities reflect impaired lower-level, that is, reflex-mediated control of pitch stability in a subgroup of cerebellar patients as documented, for example, by Ackermann and Ziegler (1994) – rather than a compromised ability to "imbue speech with emotive content."

Vicario points out that the target article does not pay any attention to the role of serotonin, that is, "another key monoamine of the reward system" besides the neurotransmitter dopamine. We highly appreciate this observation. Apart from Parkinson's disease, major depression may also give rise to a monotonous/hypophonic voice lacking affective deflection (e.g., Alpert et al. 2001; Cohn et al. 2009; Ellgring & Scherer 1996), and this clinical constellation is assumed to be associated with an imbalance of serotonergic (and noradrenergic) neurotransmission – a still central, though not sufficient pathophysiological model (Massart et al. 2012). Vicario speculates that "dopamine subserves reward-oriented (e.g., approach) communication, while serotonin subserves punishment-oriented (e.g., threat) communication." Conceivably, thus, both dopamine and serotonin depletion might converge upon "motor aprosodia" as a default vocal constellation. In contrast to the dopamine, unfortunately, the neurobiological bases of serotonin effects are still by far less elaborated. Any attempt towards an integration of both neurotransmitter systems into a common functional-neuroanatomic framework of the control of vocal behavior remains, thus, premature at the moment.

R4. Basal ganglia and ontogenetic speech acquisition: A so far neglected role of cortico-striatal circuits?

Besides adult speech production (see sect. R3.) and phylogenetic language evolution (see sects. R5.1. and R5.2.), the target article proposes a crucial role of the basal ganglia in the ontogenetic development of verbal communication. Several commentaries correctly point at the multilevel and multifaceted organization of an individual's speech development and, correctly, complain that the target article misses one or another aspect of this more complex picture: For example, (i) "the impact of the proximal social environment" (**Aitken**) on the ontogenetic emergence of communicative capacities (Aitken and **Bornstein & Esposito**); (ii) the influence of auditory-perceptual abilities already available to newborns and young infants (auditory streaming, speech sound discrimination, melody processing) upon vocal imitation capacities (**Lenti Boero**; see also **Reser & Rosa** for the domain of nonhuman primates); (iii) the role of comprehension "which almost by law ontogenetically and cognitively precedes production" during speech development (Bornstein & Esposito); (iv) the – highly intriguing – influence of listening to the vocalizations of nonhuman primates on cognitive core-capacities such as concept formation in infants during the first months of life (**Ferguson, Perszyk, & Waxman** [**Ferguson et al.**]); (v) the "possibility to refer to an object" (Lenti Boero); (v) and the obvious fact that speech motor plasticity does not – or at least must not – end after childhood (**McGettigan & Scott**).

At the end of the target article (sect. 7, "Conclusions"), we have briefly mentioned the importance of auditory-motor networks and the social environment within the context of phylogenetic language evolution. We readily acknowledge that these functional interconnections also hold for ontogenetic speech development. However, the target article focuses on a distinct, but crucial, motor aspect of the acquisition of articulate speech, that is, the concatenation of vocal tract movements into coarticulated syllabic sequences; and a more exhaustive account would have been beyond the scope of the review. Nevertheless, two commentaries touch upon the motor level of ontogenetic speech development. Whereas the target article focuses on the emergence of increasingly overlearned sequences of consonant-vowel syllables, the commentaries by **Oller**

and **Lenti Boero** further specify the preverbal vocalizations of infants.

Oller points out that "phonatory events" ("protophones") lacking significant supralaryngeal, that is, articulatory, modification characterize the early stages of human vocal development, especially, the first 3 months of life. These observations indicate the maturation of the laryngeal apparatus to precede the maturation of the cortico-striatal circuits bound to language production. At least in this regard, ontogeny, thus, appears to recapitulate phylogeny. Furthermore, **Lenti Boero** highlights the "radical transformation" of human vocal behavior during the first year of life, that is, "the substitution of the cry, an analog signal . . . with articulated speech-like sounds." Whereas infant cries, most presumably, depend upon a primate-general cerebral network, it is, in our view, the cortico-striatal circuitry which then steps in as a prerequisite of speech motor learning.

Our focus on the contribution of cortico-striatal circuits to speech acquisition in childhood by no means excludes a persisting engagement of the basal ganglia in speech motor plasticity mechanisms at a more advanced age. Indeed, as illustrated by **McGettigan & Scott** in their comment, adaptive adjustments of speaking extend well into adulthood and even senescence – in response to a variety of internal and external conditions such as alterations of peripheral-anatomic structures during aging or ambient dialectal influences causing gradual sound changes in adults. We are not aware of any data supporting the implication of the basal ganglia in such extended speech motor adaptation mechanisms, but a recent functional imaging found cortico-striatal circuits to be engaged in second language vocabulary learning (Hosoda et al. 2013; see commentary by **Hanakawa & Hosoda**). Since the experimental design of this study emphasized pronunciation training, the task must, apparently, have challenged the motor aspects of speech production. Though adult second language learning cannot be equated with the adaptive mechanisms influencing adult speech, these data point at least at the possibility of a significant contribution of the basal ganglia to a continuing process of modulation of motor speech mechanisms across adulthood – based, presumably, upon dopaminergic reward signals associated with successful articulatory performance (see also the comments by **Vicario**, and further discussion below). Hence, our proposal does not assume two distinct computational subsystems of the basal ganglia supporting immature and mature speech motor control, respectively. We rather aimed at presenting a model in which these subcortical nuclei assume two roles, that is, (i) a system supporting speech motor learning mechanisms, and (ii) a pivot between motivational-emotive and volitional mechanisms during speaking, with a gradual decrease of the importance of the former component during the maturation of speech motor control.

Any attempt towards a more comprehensive neurobiological model of human speech production, integrating phylogenetically older (vocal-emotional displays, including affective prosody) and more recent components (construction of syllables and wordforms), must address the contribution of the various central pattern generators of the brainstem to spoken language (see sect. 3.1. and Fig. 4 of the target article). Admittedly, however, the respective discussion of the target article has a still highly preliminary character – because (adult) speech pathology lacks adequate clinical model systems. **Marschik, Kaufmann, Bölte, Sigafoos, & Einspieler** (**Marschik et al.**) point at a further approach to the analysis of the operation of the central pattern generators within the speech domain, that is, neurodevelopmental disorders such as Rett syndrome, a highly promising future research area.

R5. Paleoanthropological perspectives of articulate speech acquisition: How did peripheral and cerebral adaptations interact, and does a focus on functional anatomy miss the crucial parts of the story?

R5.1. Corticobulbar-laryngeal and striatal contributions to spoken language evolution: Who takes the lead?

The introductory section of the target article suggests the "inability of nonhuman primates to produce even the most simple verbal utterances" to be due to "more crucial" cerebral limitations of motor control rather than vocal tract anatomy (sect. 1.1, para. 3). Deliberately, this formulation ("more crucial") does not exclude additional phylogenetic adaptations of the human speech apparatus at a peripheral level, including the shape of the vocal folds – as suggested by **de Boer & Perlman**. These authors hint at a larger source-filter coupling in apes as compared to human vocal tract anatomy – an observation that seems to reinforce our notion of the human larynx as an independent and coordinate player within the orchestra of speech organs (see Fig. 2C of the target article). Obviously, the strongly coupled source-filter system of apes does not allow for the same versatility of acoustic pattern generation as the (relatively) uncoupled human system. As a consequence, the control of the more independent source and filter mechanisms of the human vocal apparatus – specifically, the coordination of laryngeal and supralaryngeal gestures – must involve the regulation of a greater number of degrees-of-freedom and, therefore, should require enhanced neural control mechanisms. Against this background, the "vocal elaboration" of the cortico-striatal circuitry described in our model nicely meshes with the peripheral vocal tract modifications that may have occurred within the hominin lineage – in line with the comments by de Boer & Perlman.

Lieberman strongly rejects the assumption of a major contribution of monosynaptic corticobulbar connections to the phylogenetic development of articulate speech: He writes, "in itself, enhanced laryngeal control of phonation would not have yielded the encoding of segmental phonemes that is a unique property of human speech." In stark contrast, **Merker** deemphasizes the role of the basal ganglia and puts the corticobulbar connections to the front of the stage: He suggests "it is even conceivable that the 'simple' addition, in ancestral *Homo*, of a direct primary motor cortex efference to . . . medullary motor nuclei sufficed to recruit the already present cerebral territories centered on Wernicke's and Broca's areas (...) to the practice-based acquisition of complex vocal output" in terms of articulate speech. In this perspective, the role of "FOXP2 enhancement of cortico-basal ganglia function in the human line" is restricted to the provision of "extra storage capacity" (Merker). As convincingly argued for by Lieberman in his commentary (and relevant books), enhanced, *FOXP2*-driven "basal ganglia synaptic plasticity and connectivity" represents a necessary prerequisite for

vocal learning, including speech acquisition. In accordance with the commentary of Merker, we assume, however, that enhanced cerebral control of the larynx via monosynaptic corticobulbar connections represents a necessary prerequisite of speech production as well, providing, for instance, the basis for the generation of fast, ballistic laryngeal gestures such as those engaged in the production of unvoiced stop consonants (two-stage model of the phylogenetic development of articulate speech; see target article, Abstract).

R5.2. FOXP2-driven striatal reorganization during spoken language evolution

The (second part of the) commentary by **Aitken** provides a concise review of the multiple linguistic/nonlinguistic targets of *FOXP2* (and its nonhuman cognates) across a variety of species as well as the linguistic/nonlinguistic dysfunctions following disruption of this gene locus. It concludes: "FOXP2 is insufficient to account for the development of human language or its neural and neurochemical substrates. It is a proxy marker for the genetic control of complex biological systems we are only beginning to define or understand." Similarly, **Johansson** curtails the contribution of this gene to phylogenetic language development: "The changes in FOXP2 in the human lineage quite likely are connected with some aspects of language, but the connection is not nearly as direct as early reports claimed, and as Ackermann et al. apparently assume."

We fully agree with these statements, which deny an – exclusive and/or exhaustive – contribution of *FOXP2* to the evolution of the human language system. Our model proposes only a significant – and necessary – contribution of *FOXP2* to the phylogenetic emergence of motor aspects (!) of spoken language (we leave open the question of an engagement in higher-order cognitive dimensions of acoustic communication, see our response to **Lieberman** above). Against this background, we really – in the words of Johansson (2005, p. 27) – "begin to define or understand the genetic control of the complex biological system" of spoken language at the motor level since a plausible account of the underlying neurophysiological mechanisms and molecular-biological substrata can be envisaged in terms of enhanced "basal ganglia synaptic plasticity and connectivity" (Lieberman).

Admittedly, "the apparent presence of human *FOXP2* in Neanderthals does not in itself prove that Neanderthals spoke" (an argument put forward by **Johansson**) in terms of mastering the syntactic, semantic, and pragmatic level of a full-fledged language system, and the target article does not make such a claim. Yet, there is no reason to assume that Neanderthals were "quiet people" who "lacked completely articulate speech" (Fagan 2010, Ch. 4). We think that Neanderthals – even if they did not attain higher-order linguistic capabilities – had the functional-anatomic prerequisites to enrich their "Hmmmmm" vocalizations (Mithen 2006) by syllabic articulatory gestures – giving rise, presumably, to more salient vocal displays (some kind of elaborated "babbling"). The target article leaves open the question of the origin of the human *FOXP2* variant in Neanderthals and does not – cannot – rule out the still controversial topic of interbreeding between these two hominin species. However, this issue is not a crucial aspect of our argument, which rests upon the notion that at least the functionally relevant human *FOXP2* mutation arose in a large brain with monosynaptic corticobulbar connections to the distal cranial nerve nuclei at its disposal. Any modifications of the proposed scenario that shift these events into a more recent time window do not compromise our suggestions.

Two commentaries raise concerns over the paleoanthropological scenario put forward in the target article, linking the emergence of articulate speech to a preceding elaboration of nonverbal vocal displays. **Ravignani et al.** challenge the – alleged – assumption of our model that "enhancement of in-group cooperation and cohesion was the main driving force for the evolution of speech" (their words). And **Johansson** claims: "Vocal displays as the selective driver of protolanguage evolution (...) are highly unlikely, as they would drive the evolution of something more resembling birdsong than language." First, *FOXP2*-driven striatal reorganization in humans does not give rise to "something more resembling birdsong than language" since it took place within a human brain, endowed with a highly differentiated conceptual system even, most presumably, prior to the emergence of language (see, e.g., Hurford 2007). And, furthermore, this development played out in a more elaborate social environment as compared to other species (see commentaries by **Catania** and **Pezzulo et al.**).

In our view, second, preverbal vocal displays – whether or not within the context of coordinated group activities – served as a preadaptation for speech acquisition rather than a "selective driver of protolanguage evolution." More specifically, vocal displays enriched by sequences of syllable-sized articulatory gestures (resembling elaborated "babbling" instead of "Hmmmmm"; see above) could have supported and promoted the initial stages of the phylogenetic trajectory towards spoken language – at a point in time when the benefits of a full-fledged spoken language were not yet available, even not imaginable. Most importantly, this model aims at an answer to the quest for the adaptive benefits of a "first word" as raised by Bickerton (2009; see second last paragraph of sect. 5.2. in the target article). The commentaries by **Catania** as well as **Pezzulo et al.** provide lucid and valuable ideas relevant for a further specification of the forces which "might have contributed to transform vocalization from an initially quite limited sensorimotor feat to a powerful, open-ended instrumental tool that permits conveying rich communicative intentions" (Pezzulo et al.). For example, the more sophisticated interactions at the disposal of our species, such as joint attention (Pezzulo et al.) and/or environmental contingencies in the social context of how "one human can get another to do something" (Catania), should have paved the way towards a verbal code of acoustic communication – after a *FOXP2*-driven vocal reorganization of cortico-striatal circuits provided the sensorimotor prerequisites of spoken language.

R5.3. Extensions of the proposed model of phylogenetic articulate speech development

The new "dual-pathway model" of language evolution presented in the target article is vividly rejected by **Clark** because it omits "the recent small, but credible, neuroimaging literature which contradicts this assertion and implicates human cortico-striatal-thalamic circuitry in disambiguating lexical (…), grammatical (…), and semantic (…)

uncertainties in perceived language." Most presumably, the task of disambiguation of verbal utterances rather hinges predominantly on cortical areas (see, e.g., Wittforth et al. 2010). In any case, there is ample clinical and experimental evidence for multiple contributions of the basal ganglia to language perception and production, and the model of multiple cortico-striatal loops (see above) allows these subcortical nuclei to subserve both motor-limbic and cognitive aspects of spoken language. More specifically, elementary basal ganglia operations such as the generation and filtering of signal variances – as assumed by Clark in his commentary (second paragraph) – may be recruited within different domains of behavior (see also the comments by **Zenon & Olivier** and **Lieberman**). Interestingly, these comments put the suggestion of a contribution of cortico-striatal circuits to the disambiguation of vocal behavior/verbal information into an evolutionary context: The basal ganglia are assumed to set "limits on useful complexity of naturally communicated information" (Clark) in terms of a trade-off between the (desired) signal recognition by intended observers and (unwanted) social eavesdropping. Although Clark does not further specify the mechanisms of the assumed cortico-striatal "complexity scaling of communication," assumed to extend "along the continuum of signals to protolanguage to language," these considerations, nevertheless, touch upon a significant problem of language evolution: Whereas a speaker should take measures to safeguard the signal against social eavesdroppers, a listener must ascertain signal honesty. Increased voluntary control over vocal behavior and the "low costs" of verbal utterances facilitate deception and raise the question of how trust as a prerequisite of human cooperation can emerge and be maintained (e.g., Sterelny 2012, Ch. 5). Rather than the basal ganglia, enhanced mind-reading capabilities and memory storage capacities – associated with neocortical areas – must be considered the relevant tools for the evaluation of the reliability of a signal's content.

The contribution by **Mattei** adds an interesting novel aspect to the evolutionary scenario of the target article, which further strengthens – in our view – the suggested proposal: This commentary puts the paleoanthropological inferences of the target article into the perspective of complex adaptive system (CAS) analysis and highlights that the phylogenetic processes driving the emergence of speech production within the hominin lineage – "refinement in the projections from the motor cortex to the brainstem nuclei . . . as well as the further development of vocalization-specific cortico-basal ganglia circuitries" – can be considered a "breakthrough change" of signaling resources triggering the "percolation of the whole system and the emergence of new unpredictable features" (Mattei). As a consequence, relatively small reorganizational processes within the motor system may have supported "the emergence of high-level cognitive functions . . . from ancestral structures already present in nonhuman primates" (as **Zenon & Olivier** observe).

R6. Summary/conclusions

The target article focuses upon the – often neglected – motor aspects of spoken language evolution and emphasizes the crucial role of a vocal elaboration of cortico-striatal circuits within the hominin lineage – driven, most presumably, by a human-specific variant of the *FOXP2* gene. As a consequence, the control of the laryngeal sound source could have become part of a meshwork of phonetic gestures that are molded – via precisely defined phase-relationships – into syllable-shaped motor patterns. Such a phylogenetic reorganization of the basal ganglia must be considered necessary, but does not represent an already sufficient prerequisite for ontogenetic speech acquisition in our species – as demonstrated by the highly appreciated comments to the target article. Furthermore, the various commentaries point at a series of research questions which deserve further consideration and which are accessible to clinical/experimental investigations in our species as well as, at least partially, nonhuman primates. For example:

(a) *Basal ganglia*: Given a multitude of distinct cortico-striatal circuits, a "variegated" engagement of the basal ganglia in human communication must be taken into account, including, among other things, the modulation of higher-order aspects of speech production – bound, presumably, to the operation of the so-called "cognitive loop" – and the integration of vocal and non-vocal (facial, gestural) aspects of emotional expression. Against the background of well-established analogies between the human or mammalian basal ganglia and the avian "song brain," the interactions of the cortico-striatal circuits with the central-auditory system both during ontogenetic speech acquisition and mature speech production must be addressed in more detail. Finally, the conceivable interactions between the neurotransmitter serotonin and the "striatal messenger" dopamine during vocal-emotional expression await further elucidation.

(b) *Speech motor control mechanism*: The relationship between vocal tract movement sequencing – the focus of the target article – and the rhythmic structure of verbal utterances as well as other domains of behavior must be further addressed in a comparative-biological perspective. For example, the influential frame/content model of speech development (MacNeilage 2008) points at the supplementary motor area (SMA) as a crucial component of the cortical network of spoken language production, a mesiofrontal structure tightly interconnected with the basal ganglia.

(c) *Ontogenetic speech acquisition*: The suggested model of a pivotal role of the basal ganglia during ontogenetic speech/language development must be further substantiated. As an important research perspective within the clinical domain, the articulatory/phonatory deficits due to specific cerebral disorders such as Rett syndrome or isolated damage to the putamen must be further characterized, based upon hypothesis-driven fine-grained perceptual and acoustic evaluation procedures. Furthermore, the notion of a pivotal contribution of the basal ganglia to the ontogenetic acquisition of speech motor skills must be embedded into a broader framework, including the preceding subverbal stages of vocal behavior and higher-order aspects of phonological development.

Unfortunately, the most interesting aspect of spoken language, that is, its emergence in the first place, eludes so far a more direct examination, although molecular-genetic data begin to shed some light on this issue. As exemplified by the commentaries on the target article, this light does not yet unravel a brightly illuminated and, thus, unambiguous scenario. Nevertheless, the *FOXP2*-story nicely fits into

the context of our current understanding of speech motor control mechanisms and primate vocal behavior. Ultimately, we hope that the suggestions of the target article on phylogenetic and ontogenetic speech acquisition, centered around the basal ganglia, will help to pave the way towards a better understanding of the "end-point" of these developmental trajectories, that is, the cortical organization of mature speech production in relation to, for example, the hemispheric lateralization effects of communicative behavior in our closest cousins.

References

[The letters "a" and "r" before author's initials stand for target article and response references, respectively]

Abe, K. & Watanabe, D. (2011) Songbirds possess the spontaneous ability to discriminate syntactic rules. *Nature Neuroscience* 14:1067–74. [GJLB]

Aboitiz, F., García, R. R., Bosman, C. & Brunetti, E. (2006) Cortical memory mechanisms and language origins. *Brain and Language* 98:40–56. [TAM]

Ackermann, H. (2008) Cerebellar contributions to speech production and speech perception: Psycholinguistic and neurobiological perspectives. *Trends in Neurosciences* 31(6):265–72. doi: 10.1016/j.tins.2008.02.011. [aHA, SF, GP]

Ackermann, H. & Brendel, B. (in press) Cerebellum. In: *The neurobiology of language*, ed. G. Hickok & S. L. Small. Elsevier. [rHA]

Ackermann, H., Hertrich, I., Daum, I., Scharf, G. & Spieker, S. (1997a) Kinematic analysis of articulatory movements in central motor disorders. *Movement Disorders* 12:1019–27. [aHA]

Ackermann, H., Hertrich, I. & Ziegler, W. (2010) Dysarthria. In: *The handbook of language and speech disorders*, ed. J. S. Damico, N. Müller & M. J. Ball, pp. 362–90. Wiley-Blackwell. [aHA]

Ackermann, H., Hertrich, I., Ziegler, W., Bitzer, M. & Bien, S. (1996) Acquired dysfluencies following infarction of the left mesiofrontal cortex. *Aphasiology* 10:409–17. [aHA]

Ackermann, H., Konczak, J. & Hertrich, J. (1997b) The temporal control of repetitive articulatory movements in Parkinson's disease. *Brain and Language* 56:312–19. [arHA]

Ackermann, H., Mathiak, K. & Riecker, A. (2007) The contribution of the cerebellum to speech production and speech perception: Clinical and functional imaging data. *Cerebellum* 6:202–13. [rHA]

Ackermann, H. & Riecker, A. (2004) The contribution of the insula to motor aspects of speech production: A review and a hypothesis. *Brain and Language* 89:320–28. [DYT]

Ackermann, H. & Riecker, A. (2010a) Cerebral control of motor aspects of speech production: Neurophysiological and functional imaging data. In: *Speech motor control: New developments in basic and applied research*, ed. B. Maassen & P. van Lieshout, pp. 117–34. Oxford University Press. [aHA]

Ackermann, H. & Riecker, A. (2010b) The contribution(s) of the insula to speech production: A review of the clinical and functional imaging literature. *Brain Structure and Function* 214:419–33. [aHA]

Ackermann, H., Vogel, M., Petersen, D. & Poremba, M. (1992) Speech deficits in ischaemic cerebellar lesions. *The Journal of Neuroscience* 239(4):223–27. [UH]

Ackermann, H. & Ziegler, W. (1992) Cerebellar dysarthria: A review. *Fortschritte der Neurologie und Psychiatrie* 60:28–40. (German). [aHA]

Ackermann, H. & Ziegler, W. (1994) Acoustic analysis of vocal instability in cerebellar dysfunctions. *Annals of Otology, Rhinology and Laryngology* 103:98–104. [rHA]

Ackermann, H. & Ziegler, W. (1995) Akinetic mutism: A review of the literature. *Fortschritte der Neurologie und Psychiatrie* 63:59–67. (German). [aHA]

Ackermann, H. & Ziegler, W. (2010) Brain mechanisms underlying speech motor control. In: *The handbook of phonetic sciences*, 2nd edition, ed. W. J. Hardcastle, J. Laver & F. E. Gibbon, pp. 202–50. Wiley-Blackwell. [aHA]

Ackermann, H. & Ziegler, W. (2013) A "birdsong perspective" on human speech production. In: *Birdsong, speech, and language: Exploring the evolution of mind and brain*, ed. J. J. Bolhuis & M. Everaert, pp. 331–52. MIT Press. [aHA]

Adams, C. L., Molfese, D. L. & Betz, J. C. (1987) Electrophysiological correlates of categorical speech perception for voicing contrasts in dogs. *Developmental Neuropsychology* 3(3–4):175–89. [DLB]

Aitken, K. J. (2008) Intersubjectivity, affective neuroscience, and the neurobiology of autistic spectrum disorders: A systematic review. *Keio Journal of Medicine* 57 (1):15–36. [KJA]

Aitken, K. J. & Trevarthen, C. (1997) Self/other organization in human psychological development. *Development and Psychopathology* 9(4):653–77. [KJA]

Aitken, P. G. (1981) Cortical control of conditioned and spontaneous vocal behavior in rhesus monkeys. *Brain and Language* 13:171–84. [aHA]

Aitken, P. G. & Wilson, W. A., Jr. (1979) Discriminative vocal conditioning in rhesus monkeys: Evidence for volitional control? *Brain and Language* 8:227–40. [aHA]

Albin, R. L., Young, A. B. & Penney, J. B. (1989) The functional anatomy of basal ganglia disorders. *Trends in Neurosciences* 12:366–75. [aHA]

Alcock, K. J., Passingham, R. E., Watkins, K. E. & Vargha-Khadem, F. (2000a) Oral dyspraxia in inherited speech and language impairment and acquired dysphasia. *Brain and Language* 75(1):17–33. doi: 10.1006/brln.2000.2322. [aHA, AZ]

Alcock, K. J., Passingham, R. E., Watkins, K. E. & Vargha-Khadem, F. (2000b) Pitch and timing abilities in inherited speech and language impairment. *Brain and Language* 75:34–46. [aHA]

Alexander, G. E., Crutcher, M. D. & DeLong, M. R. (1990) Basal ganglia-thalamocortical circuits: Parallel substrates for motor, oculomotor, "prefrontal" and "limbic" functions. In: *The prefrontal cortex: Its structure, function and pathology*, ed. H. B. M. Uylings, C. G. van Eden, J. P. C. de Bruin, M. A. Corner & M. G. P. Feenstra, pp. 119–46. Elsevier. (Elsevier Book Series on Neuroscience: *Progress in Brain Research, vol. 85*). [aHA]

Alhadeff, A. L., Rupprecht, L. E. & Hayes, M. R. (2012) GLP-1 neurons in the nucleus of the solitary tract project directly to the ventral tegmental area and nucleus accumbens to control for food intake. *Endocrinology* 153:647–58. [CMV]

Alipour, M., Chen, Y. & Jürgens, U. (2002) Anterograde projections of the motor-cortical tongue area in the saddle-back tamarin (*Saguinus fuscicollis*). *Brain and Behavioral Evolution* 60:101–16. [CMV]

Alpert, M., Pouget, E. R. & Silva, R. R. (2001) Reflections of depression in acoustic measures of the patient's speech. *Journal of Affective Disorders* 66:59–69. [rHA]

Andres, M., Olivier, E. & Badets, A. (2008) Actions, words, and numbers: A motor contribution to semantic processing? *Current Directions in Psychological Science* 17(5):313–17. doi: 10.1111/j.1467-8721.2008.00597.x. [AZ]

Arbib, M. A. (2005) From monkey-like action recognition to human language: An evolutionary framework for neurolinguistics. *Behavioral and Brain Sciences* 28:105–24. [ARL]

Arbib, M. A. (2006) The Mirror System Hypothesis on the linkage of action and language. In: *Action to language via the mirror neuron system*, ed. M. A. Arbib, pp. 3–47. Cambridge University Press. [aHA]

Arnold, K. & Zuberbühler, K. (2006) Semantic combinations in primate calls. *Nature* 441(7091):303. [aHA, KBC]

Arriaga, G. & Jarvis, E. D. (2013) Mouse vocal communication system: Are ultrasounds learned or innate? *Brain and Language* 124(1):96–116. doi: 10.1016/j.bandl.2012.10.002. [BM, CIP]

Arriaga, G., Zhou, E. P. & Jarvis, E. D. (2012) Of mice, birds, and men: The mouse ultrasonic song system has some features similar to humans and song-learning birds. *PLOS ONE* 7(10):e46610. doi: 10.1371/journal.pone.0046610. [CIP]

Arroyo, S., Lesser, R. P., Gordon, B., Uematsu, S., Hart, J., Schwerdt, P., Andreasson, K. & Fisher, R. S. (1993) Mirth, laughter, and gelastic seizures. *Brain* 116:757–80. [aHA]

Ayub, Q., Yngvadottir, B., Chen, Y., Xue, Y., Hu, M., Vernes, S. C., Fisher, S. E. & Tyler-Smith, C. (2013) FOXP2 Targets show evidence of positive selection in European populations. *American Journal of Human Genetics* 92:696–706. [KJA]

Aziz-Zadeh, L., Sheng, T. & Gheytanchi, A. (2010) Common premotor regions for the perception and production of prosody and correlations with empathy and prosodic ability. *PLoS ONE* 5(1):1–8. [SF]

Badgaiyan, R. D., Fischman, A. J. & Alpert, N. M. (2007) Striatal dopamine release in sequential learning. *NeuroImage* 38:549–56. [aHA]

Bailenson, J. N., Yee, N., Patel, K. & Beall, A. C. (2008) Detecting digital chameleons. *Computers in Human Behavior* 24:66–87. [GAB]

Bailey, P., von Bonin, G. & McCulloch, W. S. (1950) *The isocortex of the chimpanzee*. University of Illinois Press. [aHA]

Bannan, N. (2012) Harmony and its role in human evolution. In: *Music, language, and human evolution*, ed. N. Bannan, pp. 288–339. Oxford University Press. [aHA]

Banse, R. & Scherer, K. R. (1996) Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology* 70(3):614–36. [aHA, DLB]

Barceló-Coblijn, L. & Benítez-Burraco, A. (2013) Disentangling the Neanderthal net: A comment on Johansson (2013). *Biolinguistics* 7:199–216. [SJ]

Barlow, S. M. & Estep, M. (2006) Central pattern generation and the motor infrastructure for suck, respiration, and speech. *Journal of Communication Disorders* 39:366–80. [PBM]

Barlow, S. M., Lund, J. P., Estep, M. & Kolta, A. (2009) Central pattern generators for orofacial movements and speech. In: *Handbook of mammalian vocalization*, ed. S. M. Brudzynski, pp. 351–70. Academic Press. [PBM]

Barnard, A. (2011) *Social anthropology and human origins*. Cambridge University Press. [aHA]

Barrett, J., Pike, G. B. & Paus, T. (2004) The role of the anterior cingulate cortex in pitch variation during sad affect. *European Journal of Neuroscience* 19:458–64. [aHA]

Barney, A., Martelli, S., Serrurier, A. & Steele, J. (2012) Articulatory capacity of Neanderthals, a very recent and human-like fossil hominin. *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences* 367:88–102. [aHA]

Barris, R. W. & Schuman, H. R. (1953) Bilateral anterior cingulate gyrus lesions: Syndrome of the anterior cingulate gyri. *Neurology* 3:44–52. [aHA]

Basel-Vanagaite, L., Muncher, L., Straussberg, R., Pasmanik-Chor, M., Yahav, M., Rainshtein, L., Walsh, C. A., Magal, N., Taub, E., Drasinover, V., Shalev, H., Attia, R., Rechavi, G., Simon, A. J. & Shohat, M. (2006) Mutated *nup62* causes autosomal recessive infantile bilateral striatal necrosis. *Annals of Neurology* 60:214–22. [aHA]

Bauernfiend, A. L., de Sousa, A. A., Avasthi, T., Dobson, S. D., Raghanti, M. A., Lewandowski, A. H., Zilles, K., Semendeferi, K., Allman, J. M., Craig, A. D., Hof, P. R. & Sherwood, C. C. (2013) A volumetric comparison of the insular cortex and its subregions in primates. *Journal of Human Evolution* 64:263–79. [DYT]

Baumann, O. & Mattingley, J. B. (2012) Functional topography of primary emotion processing in the human cerebellum. *NeuroImage* 61(4):805–11. doi: 10.1016/j.neuroimage.2012.03.044. [UH]

Beckers, G. J. L. (2013) Peripheral mechanisms of vocalization in birds: A comparison with human speech. In: *Birdsong, speech, and language. Exploring the evolution of mind and brain*, ed. J. J. Bolhuis & M. Everaert, pp. 399–422. MIT Press. [GJLB]

Beckers, G. J. L., Bolhuis, J. J., Okanoya, K. & Berwick, R. C. (2012) Birdsong neurolinguistics: Songbird context-free grammar claim is premature. *NeuroReport* 23:139–45. [aHA, GJLB]

Beckers, G. J. L., Nelson, B. S. & Suthers, R. A. (2004) Vocal tract filtering by lingual articulation in a parrot. *Current Biology* 14:1592–97. [GJLB]

Bédard, C., Wallman, M. J., Pourcher, E., Gould, P. V., Parent, A. & Parent, M. (2011) Serotonin and dopamine striatal innervation in Parkinson's disease and Huntington's chorea. *Parkinsonism Related Disorders* 17:593–98. [CMV]

Belichenko, N. P., Belichenko, P. V., Li, H. H., Mobley, W. C. & Francke, U. (2008) Comparative study of brain morphology in *Mecp2* mutant mouse models of Rett syndrome. *Journal of Comparative Neurology* 508:184–95. [PBM]

Bell, W. L., Davis, D. L., Morgan-Fisher, A. & Ross, E. D. (1990) Acquired aprosodia in children. *Journal of Child Neurology* 5:19–26. [aHA]

Belton, E., Salmond, C. H., Watkins, K. E., Vargha-Khadem, F. & Gadian, D. G. (2003) Bilateral brain abnormalities associated with dominantly inherited verbal and orofacial dyspraxia. *Human Brain Mapping* 18:194–200. [aHA]

Bendor, D. & Wang, X. (2008) Neural response properties of primary, rostral, and rostrotemporal core fields in the auditory cortex of marmoset monkeys. *Journal of Neurophysiology* 100:888–906. [DHR]

Bengtsson, S. L., Nagy, Z., Skare, S., Forsman, L., Forssberg, H. & Ullén, F. (2005) Extensive piano practicing has regionally specific effects on white matter development. *Nature Neuroscience* 8:1148–50. [aHA]

Benítez-Burraco, A. & Longa, V. M. 2012. On the inference "Neanderthals had FOXP2=they had complex language." In: *Evolution of language. Proceedings of the 9th International Conference (Evolang9)*, ed. T. Scott-Phillips, M. Tamariz, E. Cartmill & J. R. Hurford, pp 50–57. World Scientific. [SJ]

Bennett, C. H., Brassard, G., Crépeau, C., Jozsa, R., Peres, A. & Wootters, W. (1993) Teleporting an unknown quantum state via dual classical and EPR channels. *Physical Review Letters* 70:1895–99. [KBC]

Bergman, T. J. (2013) Speech-like vocalized lip-smacking in geladas. *Current Biology* 23(7):R268–69. [DHR, DYT]

Bermejo, M. & Omedes, A. (1999) Preliminary vocal repertoire and vocal communication of wild bonobos (*Pan paniscus*) at Lilungu (Democratic Republic of Congo). *Folia Primatologica* 70:328–57. [aHA]

Berridge, K. C. (2004) Motivation concepts in behavioral neuroscience. *Physiology and Behavior* 81(2):179–209. doi: 10.1016/j.physbeh.2004.02.004. [AZ]

Berta, M., Christandl, M., Colbeck, R., Renes, J. M. & Renner, R. (2010) The uncertainty principle in the presence of quantum memory. *Nature Physics* 6:659–62. [KBC]

Bertossa, R. C., ed. (2011) Theme issue: Evolutionary developmental biology (evo-devo) and behaviour. *Philosophical Transactions of the Royal Society B: Biological Sciences* 366(1574):2056–180. doi: 10.1098/rstb.2011.0035. [DKO]

Berwick, R. C., Friederici, A. D., Chomsky, N. & Bolhuis, J. J. (2013) Evolution, brain, and the nature of language. *Trends in Cognitive Sciences* 17(2):89–98. [GJLB, BF]

Berwick, R. C., Okanoya, K., Beckers, G. J. L. & Bolhuis, J. J. (2011) Songs to syntax: The linguistics of birdsong. *Trends in Cognitive Sciences* 15(3):113–21. [aHA, GJLB, KBC]

Bickerton, D. (2009) *Adam's tongue: How humans made language, how language made humans*. Hill & Wang. [aHA, SJ]

Bispham, J. (2006) Rhythm in music: What is it? Who has it? And why? *Music Perception* 24:125–34. [GAB]

Blank, S. C., Bird, H., Turkheimer, F. & Wise, R. J. (2003) Speech production after stroke: The role of the right pars opercularis. *Annals of Neurology* 54(3):310–20. doi: 10.1002/ana.10656. [CM]

Blank, S. C., Scott, S. K., Murphy, K., Warburton, E. & Wise, R. J. (2002) Speech production: Wernicke, Broca and beyond. *Brain* 125(Pt. 8):1829–38. [CM]

Blumstein, S. E. (1995) The neurobiology of language. In: *Speech, language and communication*, ed. J. Miller & P. D. Eimas, pp. 339–370. Academic Press. [PL]

Blumstein, S. E., Cooper, W. E., Goodglass, H., Statlender, S. & Gottlieb J. E. (1980) Production deficits in aphasia: A voice-onset time analysis. *Brain and Language* 9:153–70. [PL]

Bobee, S., Mariette, E., Tremblay-Leveau, H. & Caston, J. (2000) Effects of early midline cerebellar lesion on cognitive and emotional functions in the rat. *Behavioural Brain Research* 112(1–2):107–17. [UH]

Boë, L. J., Heim, J. L., Honda, K. & Maeda, S. (2002) The potential Neandertal vowel space was as large as that of modern humans. *Journal of Phonetics* 30:465–84. [aHA]

Boesch, C. & Boesch-Achermann, H. (2000) *The chimpanzees of the Taï Forest: Behavioural ecology and evolution*. Oxford University Press. [aHA]

Bohland, J. W. & Guenther, F. H. (2006) An fMRI investigation of syllable sequence production. *NeuroImage* 2:821–41. [DYT]

Bolhuis, J. J. & Everaert, M., eds. (2013) *Birdsong, speech and language. Exploring the evolution of mind and brain*. MIT Press. [aHA, GJLB]

Bolhuis, J. J., Okanoya, K. & Skarff, C. (2010) Twitter evolution: Converging mechanisms in birdsong and human speech. *Nature Reviews Neuroscience* 11 (11):747–59. [aHA, GJLB, KBC]

Bornstein, M. H., Tamis-LeMonda, C. S., & Haynes, O. M. (1999) First words in the second year: Continuity, stability, and models of concurrent and predictive correspondence in vocabulary and verbal responsiveness across age and context. *Infant Behavior and Development* 22:65–85. [MHB]

Bostan, A. C., Dum, R. P. & Strick, P. L. (2013) Cerebellar networks with the cerebral cortex and basal ganglia. *Trends in Cognitive Sciences* 17(5):241–54. doi: 10.1016/j.tics.2013.03.003. [UH]

Botez, M. I. & Barbeau, A. (1971) Role of subcortical structures, and particularly of the thalamus, in the mechanisms of speech and language: A review. *International Journal of Neurology* 8:300–20. [aHA]

Bottoni, L., Masin, S. & Lenti Boero, D. (2009) Vowel-like sound structure in an African Grey Parrot (*Psittacus erithacus*) vocal production. *The Open Behavioral Science Journal* 3:1–16. [DLB]

Bouchard, K. E., Mesgarani, N., Johnson, K. & Chang, E. F. (2013) Functional organization of human sensorimotor cortex for speech articulation. *Nature* 495:327–32. doi:10.1038/nature11911. [aHA]

Bowers, J. M. & Konopka, G. (2012) The role of the FOXP family of transcription factors in ASD. *Disease Markers* 33(2012):251–60. doi: 10.3233/DMA-2012-0919. [KJA]

Bowers, J. M., Perez-Pouchoulen, M., Edwards, N. S. & McCarthy, M. M. (2013) Foxp2 mediates sex differences in ultrasonic vocalization by rat pups and directs order of maternal retrieval. *Journal of Neuroscience* 33(8):3276–83. [KJA]

Boyd, L. A., Edwards, J. D., Siengsukon, C. S., Vidoni, E. D., Wessel, B. D. & Linsdell, M. A. (2009) Motor sequence chunking is impaired by basal ganglia stroke. *Neurobiology of Learning and Memory* 92(1):35–44. doi: 10.1016/j.nlm.2009.02.009. [AZ]

Braak, H., Del Tredici, K., Rüb, U., de Vos, R. A. I., Jansen Steur, E. N. H. & Braak, E. (2003) Staging of brain pathology related to sporadic Parkinson's disease. *Neurobiology of Aging* 24:197–211. [aHA]

Brainard, M. S. & Doupe, A. J. (2002) What songbirds teach us about learning. *Nature* 417:351–58. [aHA, GP]

Braman, S. (2004) The emergent global information policy regime. In: *The emergent global information policy regime,* ed. S. Braman, pp. 12–37. Palgrave Macmillan. [TAM]

Brandt, P. A. (2009) Music and how we became human – a view from cognitive semiotics: Exploring imaginative hypotheses. In: *Communicative musicality: Exploring the basis of human companionship*, ed. S. Malloch & C. Trevarthen, pp. 31–44. Oxford University Press. [aHA]

Breitenstein, C., Daum, I. & Ackermann, H. (1998) Emotional processing following cortical and subcortical brain damage: Contribution of the fronto-striatal circuitry. *Behavioural Neurology* 11:29–42. [rHA]

Brendel, B., Erb, M., Riecker, A., Grodd, W., Ackermann, H. & Ziegler, W. (2011) Do we have a "mental syllabary" in the brain? An fMRI study. *Motor Control* 15(1):34–51. [UH]

Brendel, B., Hertrich, I., Erb, M., Lindner, A., Riecker, A., Grodd, W. & Ackermann, H. (2010) The contribution of mesiofrontal cortex to the preparation and execution of repetitive syllable productions: An fMRI study. *NeuroImage* 50:1219–30. [aHA]

Brendel, B. & Ziegler, W. (2008) Effectiveness of metrical pacing in the treatment of apraxia of speech. *Aphasiology* 22:77–102. [rHA]

Brockelman, W. Y. & Schilling, D. (1984) Inheritance of stereotyped gibbon calls. *Nature* 312:634–36. [aHA]

Brotis, A. G., Kapsalaki, E. Z., Paterakis, K., Smith, J. R. & Fountas, K. N. (2009) Historic evolution of open cingulectomy and stereotactic cingulotomy in the management of medically intractable psychiatric disorders, pain and drug addiction. *Stereotactic and Functional Neurosurgery* 87:271–91. [aHA]

Brown, J. W. (1988) Cingulate gyrus and supplementary motor correlates of vocalization in man. In: *The physiological control of mammalian vocalization*, ed. J. D. Newman, pp. 227–43. Plenum Press. [aHA]

Brown, S. (2000) The "musilanguage" model of music evolution. In: *The origins of music*, ed. N. L. Wallin, B. Merker & S. Brown, pp. 271–300. MIT Press. [aHA]

Brown, S., Ngan, E. & Liotti, M. (2008) A larynx area in the human motor cortex. *Cerebral Cortex* 18:837–45. [aHA]

Brown, S., Laird, A. R., Pfordresher, P. Q., Thelen, S. M., Turkeltaub, P. & Liotti, M. (2009) The somatotopy of speech: Phonation and articulation in the human motor cortex. *Brain and Cognition* 70:31–41. [aHA]

Brown, T. G. (1915) Note on the physiology of the basal ganglia and mid-brain of the anthropoid ape, especially in reference to the act of laughter. *Journal of Physiology* 49:195–207. [aHA]

Brück, C., Wildgruber, D., Kreifelts, B., Krüger, R. & Wächter, T. (2011) Effects of subthalamic nucleus stimulation on emotional prosody comprehension in Parkinson's disease. *PLOS ONE* 6:e19140. [CMV]

Brumm, H. & Slabbekoorn, H. (2005) Acoustic communication in noise. *Advances in the Study of Behavior* 35:151–209. [rHA]

Brumm, H., Voss, K., Köllmer, I. & Todt, D. (2004) Acoustic communication in noise: Regulation of call characteristics in a New World monkey. *Journal of Experimental Biology* 207(3):443–48. [aHA, DJW]

Brumm, H. & Zollinger, S. A. (2011) The evolution of the Lombard effect: 100 years of psychoacoustic research. *Behaviour* 148:1173–98. [rHA]

Bryant, G. A. (2012) Shared laughter in conversation as coalition signaling. Paper presented at the XXI Biennial International Conference on Human Ethology, Vienna, Austria, August 13, 2012. [GAB]

Bryant, G. A. (2013) Animal signals and emotion in music: Coordinating affect across groups. *Frontiers in Psychology* 4(Article 990):1–13. [GAB]

Bryant, G. A. & Aktipis, A. (2014) The animal nature of spontaneous human laughter. *Evolution and Human Behavior* 35(4):327–35. [GAB]

Buder, E. H., Chorna, L., Oller, D. K. & Robinson, R. (2008) Vibratory regime classification of infant phonation. *Journal of Voice* 22:553–64. [DKO]

Burgoon, J. K., Floyd, K. & Guerrero, L. K. (2010) Nonverbal communication theories of interaction adaptation. In: *The handbook of communication science*, 2nd edition, ed. C. R. Berger, M. E. Roloff & D. R. Roskos-Ewoldsen, pp. 93–108. Sage. [aHA]

Burkart, J. M., Fehr, E., Efferson, C. & van Schaik, C. P. (2007) Other-regarding preferences in a non-human primate: Common marmosets provision food altruistically. *Proceedings of the National Academy of Sciences USA* 104:19762–66. [ARL]

Burkart, J. M., Hrdy, S. B. & van Schaik, C. P. (2009a) Cooperative breeding and human cognitive evolution. *Evolutionary Anthropology* 18:175–86. [ARL, DYT]

Burkart, J. M., Strasser, A. & Foglia M. (2009b) Trade-offs between social learning and individual innovativeness in common marmosets, *Callithrix jacchus*. *Animal Behaviour* 77:1291–301. [ARL]

Burkart, J. & van Schaik, C. (2010) Cognitive consequences of cooperative breeding in primates? *Animal Cognition* 13:1–19. [ARL]

Burling, R. (2005) *The talking ape: How language evolved*. Oxford University Press. [aHA]

Butler, A. B. & Hodos, W. (2005) *Comparative vertebrate neuroanatomy: Evolution and adaptation*, 2nd edition. Wiley. [aHA]

Byrne, R. W. (2000) Evolution of primate cognition. *Cognitive Science* 24:543–70. [BM]

Caligiore, D., Pezzulo, G., Miall, R. C. & Baldassarre, G. (2013) The contribution of brain sub-cortical loops in the expression and acquisition of action understanding abilities. *Neuroscience and Biobehavioral Reviews* 37(10):2504–15. [GP]

Call, J. & Tomasello, M., eds. (2007) *The gestural communication of apes and monkeys*. Erlbaum. [aHA]

Calzavara, R., Mailly, P. & Haber, S. N. (2007) Relationship between the corticostriatal terminals from areas 9 and 46, and those from area 8A, dorsal and rostral premotor cortex and area 24c: An anatomical substrate for cognition to action. *European Journal of Neuroscience* 26:2005–24. [aHA]

Cancelliere, A. E. B. & Kertesz, A. (1990) Lesion localization in acquired deficits of emotional expression and comprehension. *Brain and Cognition* 13:133–47. [aHA]

Canevari, C., Badino, L., D'Ausilio, A., Fadiga, L. & Metta, G. (2013) Modeling speech imitation and ecological learning of auditory-motor maps. *Frontiers in Psychology* 4:364. [GP]

Cardona, J. F., Gershanik, O., Gelormini-Lezama, C., Houck, A. L., Cardona, S., Kargieman, L., Trujillo, N., Arévalo, A., Amoruso, L., Manes, F. & Ibánez, A. (2013) Action-verb processing in Parkinson's disease: New pathways for motor-language learning. *Brain Structure and Function* 218:1355–73. [rHA]

Carlsson, P. & Mahlapuu, M. (2002) Forkhead transcription factors: Key players in development and metabolism. *Developmental Biology* 250(1):1–23. [KJA]

Carreiras, M., Mechelli, A. & Price, C. J. (2006) Effect of word and syllable frequency on activation during lexical decision and reading aloud. *Human Brain Mapping* 27(12):963–72. doi: 10.1002/hbm.20236. [UH]

Cartmill, E. A. & Byrne, R. W. (2007) Orangutans modify their gestural signalling according to their audience's comprehension. *Current Biology* 17:1345–48. [AM]

Caruana, F., Jezzini, A., Sbriscia-Fioretti, B., Rizzolatti, G. & Gallese, V. (2011) Emotional and social behaviors elicited by electrical stimulation of the insula in the macaque monkey. *Current Biology* 21:195–99. [DYT]

Catania, A. C. (1990) What good is five percent of a language competence? *Behavioral and Brain Sciences* 13:729–31. [ACC]

Catania, A. C. (1995) Single words, multiple words, and the functions of language. *Behavioral and Brain Sciences* 18:184–85. [ACC]

Catania, A. C. (2001) Three varieties of selection and their implications for the origins of language. In: *Language evolution: Biological, linguistic and philosophical perspectives*, ed. G. Györi, pp. 55–71. Peter Lang. [ACC]

Catania, A. C. (2003) Why behavior should matter to linguists. *Behavioral and Brain Sciences* 26:670–72. [ACC]

Catania, A. C. (2008) Brain and behavior: Which way does the shaping go? *Behavioral and Brain Sciences* 31:516–17. [ACC]

Catania, A. C. (2009) Language evolution: Two tracks are not enough. *Behavioral and Brain Sciences* 32:451–52. [ACC]

Catania, A. C. (2013a) *Learning*, 5th edition. Sloan. [ACC]

Catania, A. C. (2013b) A natural science of behavior. *Review of General Psychology* 17:133–39. [ACC]

Catania, A. C. & Cerutti, D. (1986) Some nonverbal properties of verbal behavior. In: *Analysis and integration of behavioral units*, ed. T. Thompson & M. D. Zeiler, pp. 185–211. Erlbaum. [ACC]

Chan, A. M., Dykstra, A. R., Jayaram, V., Leonard, M. K., Travis, K. E., Gygi, B., Baker, J. M., Eskandar, E., Hochberg, L.R., Halgren, E. & Cash, S. S. (2014) Speech-specific tuning of neurons in human superior temporal gyrus. *Cerebral Cortex* 24(10):2679–93. [DHR]

Chan, S.-H., Ryan, L. & Bever, T. G. (2013) Role of the striatum in language: Syntactic and conceptual sequencing. *Brain and Language* 125(3):283–94. doi: 10.1016/j.bandl.2011.11.005. [AZ]

Chandrasekaran, C., Lemus, L., Trubanova, A., Gondan, M. & Ghazanfar, A. A. (2011) Monkeys and humans share a common computation for face/voice integration. *PLOS Computational Biology* 7(9):e1002165. [DYT]

Chandrasekaran, C., Trubanova, A., Stillittano, S., Caplier, A. & Ghazanfar, A. A. (2009) The natural statistics of audiovisual speech. *PLOS Computational Biology* 5:e1000436. [DYT]

Chang, C.-C., Lee, Y. C., Lui, C.-C. & Lai, S.-L. (2007) Right anterior cingulate cortex infarction and transient speech aspontaneity. *Archives of Neurology* 64:442–46. [aHA]

Changizi, M. A. (2001) Universal scaling laws for hierarchical complexity in languages, organisms, behaviors and other combinatorial systems. *Journal of Theoretical Biology* 211(3):277–95. [KBC]

Charlesworth, J. D., Warren, T. L. & Brainard, M. S. (2012) Covert skill learning in a cortical-basal ganglia circuit. *Nature* 486(7402):251–55. doi: 10.1038/nature11078. [CIP]

Chassagnon, S., Minotti, L., Kremer, S., Verceuil, L., Hoffmann, D., Benabid, A. L. & Kahane, P. (2003) Restricted frontomesial epileptogenic focus generating dyskinetic behavior and laughter. *Epilepsia* 44:859–63. [aHA]

Chen, L.-M. & Kent, R. D. (2009) Development of prosodic patterns in Mandarin-learning infants. *Journal of Child Language* 36(1):73–84. [PL]

Chenery, H. J., Angwin, A. J. & Copeland, D. A. (2008) The basal circuits, dopamine, and ambiguous word processing: A neurobiological account of priming studies in Parkinson's disease. *Journal of International Neuropsychology Society* 14 (3):351–64. [KBC]

Cheney, D. L. & Seyfarth, R. M. (1990) *How monkeys see the world: Inside the mind of another species*. University of Chicago Press. [aHA]

Cheney, D. L. & Seyfarth, R. M. (2005) Constraints and preadaptations in the earliest stages of language evolution. *The Linguistic Review* 22:135–59. [aHA]

Cheney, D. L. & Seyfarth, R. M. (2007) *Baboon metaphysics: The evolution of a social mind*. University of Chicago Press. [aHA]

Chersi, F., Ferro, M., Pezzulo, G. & Pirrelli, V. (2014) Topological self-organization and prediction learning can support both action and lexical chains in the brain. *Topics in Cognitive Science* 6(3):476–91. [GP]

Chien, W.-H., Gau, S. S.-F., Chen, C.-H., Tsai, W.-C., Wu, Y.-Y., Chen, P.-H., Shang, C.-Y. & Chen, C.-H. (2013) Increased gene expression of FOXP1 in patients with autism spectrum disorders. *Molecular Autism* 4:23. Available at: http://www.molecularautism.com/content/4/1/23 [KJA]

Choi, E. Y., Yeo, B. T. & Buckner, R. L. (2012) The organization of the human striatum estimated by intrinsic functional connectivity. *Journal of Neurophysiology* 108(8):2242–63. doi: 10.1152/jn.00270.2012. [UH]

Chomsky, N. (1956) Three models for the description of language. *IRE Transactions on Information Theory* 2:113–24. [KBC]

Chomsky, N. (1965) *Aspects of the theory of syntax*. MIT Press. [BF]

Chomsky, N. (1966) *Cartesian linguistics: A chapter in the history of rationalist thought*. Harper & Row. [KBC]

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

Christophe, A., Millotte, S., Bernal, S. & Lidz, J. (2008) Bootstrapping lexical and syntactic acquisition. *Language and Speech* 51(1–2):61–75. doi: 10.1177/00238309080510010501. [AR]

Cismaresco, A. S. & Montagner, H. (1990) Mother's discrimination of their neonates' cry in relation to cry acoustics: The first week of life. *Early Child Development and Care* 65:3–13. [DLB]

Clark, K. B. (2010) On classical and quantum error-correction in ciliate mate selection. *Communicative & Integrative Biology* 3(4):374–78. [KBC]

Clark, K. B. (2012) A statistical mechanics definition of insight. In: *Computational intelligence*, ed. A. G. Floares, pp. 139–62. Nova Science. [KBC]

Clark, K. B. (2013a) Ciliates learn to diagnose and correct classical error syndromes in mating strategies. *Frontiers in Microbiology* 4:229. [KBC]

Clark, K. B. (2013b) The mating judgments of microbes. In: *Social learning theory: Phylogenetic considerations across animal, plant, and microbial taxa*, ed. K. B. Clark, pp. 173–200. Nova Science. [KBC]

Clark, K. B. (in press) Entropic uncertainty of ciliate behavioral signals limits eavesdropping by mating rivals and predators. *Frontiers in Microbiology*. [KBC]

Clarkson, M. G. & Berg, W. K. (1983) Cardiac orienting and vowel discrimination in newborns: Crucial stimulus parameters. *Child Development* 54:162–71. [DLB]

Clay, Z., Pika, S., Gruber, T. & Zuberbühler, K (2011) Female bonobos use copulation calls as social signals. *Biology Letters* 7:513–16. [ARL]

Clay, Z. & Zuberbühler, K. (2009) Food-associated calling sequences in bonobos. *Animal Behaviour* 77:1387–96. [aHA]

Clegg, M. (2012) The evolution of the human vocal tract: Specialized for speech? In: *Music, language, and human evolution*, ed. N. Bannan, pp. 58–80. Oxford University Press. [aHA]

Clerget, E., Andres, M. & Olivier, E. (2013) Deficit in complex sequence processing after a virtual lesion of left BA45. *PLOS ONE* 8(6):e63722. doi: 10.1371/journal.pone.0063722. [AZ]

Clerget, E., Badets, A., Duque, J. & Olivier, E. (2011) Role of Broca's area in motor sequence programming: A cTBS study. *NeuroReport* 22(18):965–69. doi: 10.1097/WNR.0b013e32834d87cd. [AZ]

Clerget, E., Poncin, W., Fadiga, L. & Olivier, E. (2012) Role of Broca's area in implicit motor skill learning: Evidence from continuous theta-burst magnetic stimulation. *Journal of Cognitive Neuroscience* 24(1):80–92. doi: 10.1162/jocn_a_00108. [AZ]

Clerget, E., Winderickx, A., Fadiga, L. & Olivier, E. (2009) Role of Broca's area in encoding sequential human actions: A virtual lesion study. *NeuroReport* 20 (16):1496–99. doi: 10.1097/WNR.0b013e3283329be8. [AZ]

Cleveland, J. & Snowdon, C. T. (1982) The complex vocal repertoire of the adult cotton-top tamarin (*Saguinus oedipus oedipus*). *Zeitschrift für Tierpsychologie* 58(3):231–70. [DJW]

Cohen, J. (2010) *Almost chimpanzee: Searching for what makes us human, in rainforests, labs, sanctuaries, and zoos*. Henry Holt. [aHA]

Cohen, M. J., Riccio, C. A. & Flannery, A. M. (1994) Expressive aprosodia following stroke to the right basal ganglia: A case report. *Neuropsychology* 8:242–45. [aHA]

Cohn, J. F., Kruez, T. S., Matthews, I., Yang, Y., Nguyen, M. H., Padilla, M. T., Zhou, F. & De la Torre, F. (2009) Detecting depression from facial actions and vocal prosody. In: *Proceedings of the Third International Conference on Affective Computing and Intelligent Interaction (ACII-09), 10-12 September 2009, Amsterdam, The Netherlands*, pp. 1–7. IEEE Xplore Digital Library. [rHA]

Conway, C. M. & Christiansen, M. H. (2001) Sequential learning in non-human primates. *Trends in Cognitive Sciences* 5(12):539–46. [AR]

Cook, P., Rouse, A., Wilson, M. & Reichmuth, C. (2013) A California sea lion (*Zalophus californianus*) can keep the beat: Motor entrainment to rhythmic auditory stimuli in a non-vocal mimic. *Journal of Comparative Psychology* 127 (4):412–27. doi: 10.1037/a0032345. [HH, AR]

Coolidge, F. L. & Wynn, T. (2009) *The rise of Homo sapiens: The evolution of modern thinking*. Wiley-Blackwell. [aHA]

Coop, G., Bullaughey, K., Luca, F. & Przeworski, M. (2008) The timing of selection at the human *FOXP2* gene. *Molecular Biology and Evolution* 25:1257–59. [aHA]

Copeland, D. (2003) The basal ganglia and semantic engagement: Potential insights from semantic priming in individuals with subcortical vascular lesions, Parkinson's disease, and cortical lesions. *Journal of International Neuropsychology Society* 9(7):1041–52. [KBC]

Corballis, M. C. (2002) *From hand to mouth: The origins of language*. Princeton University Press. [aHA]

Corballis, M. C. (2003) From mouth to hand: Gesture, speech, and the evolution of right-handedness. *Behavioral and Brain Sciences* 26:199–260. [aHA]

Coudé, G., Ferrari, P. F., Rodà, F., Maranesi, M., Borelli, E., Veroni, V., Monti, F., Rozzi, S. & Fogassi, L. (2011) Neurons controlling voluntary vocalization in the macaque ventral premotor cortex. *PLoS ONE* 6:e26822. [aHA]

Coward, F. (2010) Small worlds, material culture and ancient Near Eastern social networks. In: *Social brain, distributed mind*, ed. R. Dunbar, C. Gamble & J. Gowlett, pp. 449–79. Oxford University Press. (*Proceedings of the British Academy, vol. 158*). [aHA]

Crais, E., Douglas, D. & Campbell, C. (2004) The intersection of the development of gestures and intentionality. *Journal of Speech, Language and Hearing Research* 47:678–94. [KJA]

Creed, M. C., Hamani, C., Bridgman, A., Fletcher, P. J. & Nobrega, J. N. (2012) Contribution of decreased serotonin release to the antidyskinetic effects of deep brain stimulation in a rodent model of tardive dyskinesia: Comparison of the subthalamic and entopeduncular nuclei. *Journal of Neuroscience* 32:9574–81. [CMV]

Crinion, J., Turner, R., Grogan, A., Hanakawa, T., Noppeney, U., Devlin, J. T., Aso, T., Urayama, S., Fukuyama, H., Stockton, K., Usui, K., Green, D. W. & Price, C. J. (2006) Language control in the bilingual brain. *Science* 312:1537–40. [TH]

Crockford, C. & Boesch, C. (2003) Context-specific calls in wild chimpanzees, *Pan troglodytes verus*: Analysis of barks. *Animal Behaviour* 66:115–25. doi: 10.1006/anbe.2003.2166. [DKO]

Cross, I. (2001) Music, mind and evolution. *Psychology of Music* 29:95–102. [aHA]

Cross, I. (2003) Music, cognition, culture, and evolution. In: *The cognitive neuroscience of music*, ed. I. Peretz & R. Zatorre, pp. 42–56. Oxford University Press. [aHA]

Cross, I. & Morley, I. (2009) The evolution of music: Theories, definitions and the nature of the evidence. In: *Communicative musicality: Exploring the basis of human companionship*, ed. S. Malloch & C. Trevarthen, pp. 61–81. Oxford University Press. [aHA]

Crosson, B. (2013) Thalamic mechanisms in language: A reconsideration based on recent findings and concepts. *Brain and Language* 126(1):73–88. doi: 10.1016/j.bandl.2012.06.011. [UH]

Csibra, G. & Gergely, G. (2011) Natural pedagogy as evolutionary adaptation. *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences* 366:1149–57. [GP]

Cummins, F. (2009) Rhythm as entrainment: The case of synchronous speech. *Journal of Phonetics* 37:16–28. [rHA]

Cutler, A., Oahan, D. & van Donselaar, W. (1997) Prosody in the comprehension of spoken language: A literature review. *Language and Speech* 40(2):141–201. [AR]

Dabelsteen, T. (2004) Strategies that facilitate or counter eavesdropping on vocal interactions in songbirds. *Anais de Academia Brasileirade Ciências* 76 (2):274–78. [KBC]

Dall, S. R. X. (2005) Information and its use by animals in evolutionary ecology. *Trends in Ecological Evolution* 20(4):187–93. [KBC]

Damasio, A. R., Grabowski, T. J., Bechara, A., Damasio, H., Ponto, L. L., Parvizi, J. & Hichwa, R. D. (2000) Subcortical and cortical brain activity during the feeling of self-generated emotions. *Nature Neuroscience* 3(10):1049–56. doi: 10.1038/79871. [UH]

Danchin, E., Giraldeau, L. A., Valone, T. J. & Wagner, R. H. (2004) Public information: From nosy neighbours to cultural evolution. *Science* 305 (5683):487–91. [KBC]

Dang, M. T., Yokoi, F., Yin, H. H., Lovinger, D. M., Wang, Y. & Li, Y. (2006) Disrupted motor learning and long-term synaptic plasticity in mice lacking *NMDAR1* in the striatum. *Proceedings of the National Academy of Sciences USA* 103:15254–59. [aHA]

Darkins, A. W., Fromkin, V. A. & Benson, D. F. (1988) A characterization of the prosodic loss in Parkinson's disease. *Brain and Language* 34:315–27. [aHA, PBM]

Darwin, C. (1871) *The descent of man, and selection in relation to sex*. John Murray [2nd edition 1879 by John Murray, reprint 2004 by Penguin Books]. [aHA]

Das, S. & Fowler, S. C. (1995) Acute and subchronic effects of clozapine on licking in rats: Tolerance to disruptive effects on number of licks, but no tolerance to rhythm slowing. *Psychopharmacology (Berlin)* 120:249–55. [CMV]

D'Ausilio, A., Craighero, L. & Fadiga, L. (2012) The contribution of the frontal lobe to the perception of speech. *Journal of Neurolinguistics* 25:328–35. [GP]

David, H. N., Ansseau, M. & Abraini, J. H. (2005) Dopamine–glutamate reciprocal modulation of release and motor responses in the rat caudate-putamen and nucleus accumbens of "intact" animals. *Brain Research. Brain Research Reviews* 50:336–60. [aHA]

Davila-Ross, M., Owren, M. & Zimmermann, E. (2009) Reconstructing the evolution of laughter in great apes and humans. *Current Biology* 19:1106–11. [GAB]

Davis, B. L. & MacNeilage, P. F. (2002) The internal structure of the syllable. In: *The evolution of language out of pre-language*, ed. T. Givòn & B. F. Malle, pp. 135–54. John Benjamins. [DLB]

Davis, P. J., Zhang, S. P., Winkworth, A. & Bandler, R. (1996) Neural control of vocalization: Respiratory and emotional influences. *Journal of Voice* 10:23–38. [aHA]

de Boer, B. (2008) The acoustic role of supralaryngeal air sacs. *Journal of the Acoustical Society of America* 123(5 Pt 2):3732–33. [BdB]

de Boer, B. (2012) Air sacs and vocal fold vibrations: Implications for evolution of speech. *Theoria et Historia Scientiarum* 9:13–28. [BdB]

de Boysson-Bardie, B. (2001) *How language comes to children*. MIT Press. [DLB]

Dediu, D. & Levinson, S. C. (2013) On the antiquity of language: The reinterpretation of Neandertal linguistic capacities and its consequences. *Frontiers in Psychology* 4(397):1–17. doi: 10.3389/fpsyg.2013.00397. [SJ]

De Filippis, B., Ricceri, L. & Laviola, G. (2010) Early postnatal behavioral changes in the *Mecp2*-308 truncation mouse model of Rett syndrome. *Genes, Brain and Behavior* 9:213–23. [PBM]

Dehaene, S. & Cohen, L. (2007) Cultural recycling of cortical maps. *Neuron* 56 (2):384–98. doi: 10.1016/j.neuron.2007.10.004. [GP, AZ]

De Letter, M., Santens, P., Estercam, I., Van Maele, G., De Bodt, M., Boon, P. & Van Borsel, J. (2007) Levodopa-induced modifications of prosody and comprehensibility in advanced Parkinson's disease as perceived by professional listeners. *Clinical Linguistic Phonology* 21:783–91. [CMV]

DeLong, M. R. & Wichmann, T. (2007) Circuits and circuit disorders of the basal ganglia. *Archives of Neurology* 64:20–24. [aHA]

De Meirleir, L., Seneca, S., Lissens, W., Schoentjes, E. & Desprechins, B. (1995) Bilateral striatal necrosis with a novel point mutation in the mitochondrial ATPase 6 gene. *Pediatric Neurology* 13:242–46. [aHA]

Demirezen, M. (1988) Behaviorist theory and language learning. *Hacettepe Vniversitesi Eðitim Fakültesi Dergisi* 3:135–40. [TH]

Demolin, D. & Delvaux, V. (2006) A comparison of the articulatory parameters involved in the production of sounds of bonobos and modern humans. In: *The evolution of language: Proceedings of the 6th International Conference (Evolang6)*, ed. A. Cangelosi, A. D. M. Smith & K. Smith. pp. 67–74. World Scientific. [BdB]

Deriziotis, P. & Fisher, S. E. (2013) Neurogenomics of speech and language disorders: The road ahead. *Genome Biology* 14:204. Available at: http://genomebiology.com/2013/14/4/204 [KJA]

Desmurget, M. & Turner, R. S. (2008) Testing basal ganglia motor functions through reversible inactivations in the posterior internal globus pallidus. *Journal of Neurophysiology* 99(3):1057–76. doi: 10.1152/jn.01010.2007. [AZ]

Desmurget, M. & Turner, R. S. (2010) Motor sequences and the basal ganglia: Kinematics, not habits. *Journal of Neuroscience* 30(22):7685–90. doi: 10.1523/JNEUROSCI.0163-10.2010. [AZ]

Dessalles, J.-L. (2007) *Why we talk: The evolutionary origins of language*. Oxford University Press. [aHA]

Deutch, A. Y., Colbran, R. J. & Winder, D. J. (2007) Striatal plasticity and medium spiny neuron dendritic remodeling in Parkinsonism. *Parkinsonism and Related Disorders* 13 (Suppl. 3):S251–58. [aHA]

De Waal, F. B. M. (1988) The communicative repertoire of captive bonobos compared to that of chimpanzees. *Behaviour* 106:183–251. [aHA]

Dewson, J. H. (1964) Speech sound discrimination by cats. *Science* 144:555–56. [DLB]

Diller, K. C. & Cann, R. L. (2009) Evidence against a genetic-based revolution in language 50,000 years ago. In: *The cradle of language*, ed. R. Botha & C. Knight, pp. 135–49. Oxford University Press. [SJ]

Diller, K. C. & Cann, R. L. (2012) Genetic influences on language evolution: An evaluation of the evidence. In: *The Oxford Handbook of Language Evolution,* ed. M. Tallerman & K. R. Gibson, pp. 168–75. Oxford University Press. [SJ]

Dindo, H., Zambuto, D. & Pezzulo, G. (2011) Motor simulation via coupled internal models using sequential Monte Carlo. In: *Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence (IJCAI), Barcelona, Catalonia, Spain, 16–22 July 2011*, ed. Toby Walsh, pp. 2113–19. AAAI Press/International Joint Conferences on Artificial Intelligence. [GP]

Dissanayake, E. (2009) Root, leaf, blossom, or bole: Concerning the origin and adaptive function of music. In: *Communicative musicality: Exploring the basis of human companionship*, ed. S. Malloch & C. Trevarthen, pp. 17–30. Oxford University Press. [aHA]

Dominey, P. F. & Inui, T. (2009) Cortico-striatal function in sentence comprehension: Insights from neurophysiology and modeling. *Cortex* 45(8):1012–18. doi: 10.1016/j.cortex.2009.03.007. [AR]

Donald, M. (1999) Preconditions for the evolution of protolanguages. In: *The descent of mind: Psychological perspectives on hominid evolution,* ed. M. C. Corballis & S. E. G. Lea, pp. 138–54. Oxford University Press. [aHA]

Doupe, A. J. & Kuhl, P. K. (1999) Birdsong and human speech: Common themes and mechanisms. *Annual Review of Neuroscience* 22:567–631. [aHA]

Doupe, A. J., Perkel, D. J., Reiner, A. & Stern, E. A. (2005) Birdbrains could teach basal ganglia research a new song. *Trends in Neurosciences* 28(7):353–63. [aHA, KBC]

Doya, K. (2000) Complementary roles of basal ganglia and cerebellum in learning and motor control. *Current Opinion in Neurobiology* 10:732–39. [aHA]

Doya, K. (2008) Modulators of decision making. *Nature Neuroscience* 11:410–16. [TH]

Doyon, J. & Benali, H. (2005) Reorganization and plasticity in the adult brain during learning of motor skills. *Current Opinion in Neurobiology* 15:161–67. [aHA]

Dronkers, N. F. (1996) A new brain region for coordinating speech articulation. *Nature* 384(6605):159–61. doi: 10.1038/384159a0. [CM, DYT]

Duffy, J. R. (2005) *Motor speech disorders: Substrates, differential diagnosis, and management*, 2nd edition. Elsevier Mosby. [aHA]

Dum, R. P. & Strick, P. L. (2002) Motor areas in the frontal lobe of the primate. *Physiology and Behavior* 77:677–82. [aHA]

Dunbar, R. I. M. (1993) Coevolution of neocortical size, group size and language in humans. *Behavioral and Brain Sciences* 16:681–94. [DLB]

Dunbar, R. I. M. (1996) *Grooming, gossip, and the evolution of language*. Harvard University Press. [aHA]

Dunbar, R. I. M. (2012) On the evolutionary function of song and dance. In: *Music, language, and human evolution*, ed. N. Bannan, pp. 201–14. Oxford University Press. [aHA]

Egan, J. P. & Hake, H. W. (1950) On the masking pattern of a simple auditory stimulus. *Journal of the Acoustical Society of America* 22:622–30. [DJW]

Egnor, S. E. R. & Hauser, M. D. (2006) Noise-induced vocal modulation in cotton-top tamarins (*Saguinus oedipus*). *American Journal of Primatology* 68 (12):1183–90. [DJW]

Egnor, S. E. R., Iguina, C. G. & Hauser, M. D. (2006) Perturbation of auditory feedback causes systematic perturbation in vocal structure in adult cotton-top tamarins. *Journal of Experimental Biology* 209(18):3652–63. [DJW]

Egnor, S. E. R., Wickelgren, J. G. & Hauser, M. D. (2007) Tracking silence: Adjusting vocal production to avoid acoustic interference. *Journal of Comparative Physiology A: Neuroethology, Sensory, Neural, and Behavioral Physiology* 193 (4):477–83. [aHA, DJW]

Eimas, P. D., Siqueland, E. R., Jusczyk, P. & Vigorito, J. (1971) Speech perception in infants. *Science* 171:303–306. [DLB]

Einspieler, C. & Marschik, P. B. (2012) Central Pattern Generators and their significance for the foetal motor function. *Klinische Neurophysiologie* 43:16–21. [PBM]

Ekman, P. & Friesen, W. (1978) *The facial action coding system*. Consulting Psychologists Press. [DKO]

Eliades, S. J. & Wang, X. (2012) Neural correlates of the Lombard effect in primate auditory cortex. *The Journal of Neuroscience* 32(31):10737–48. [DJW]

Ellgring, H. & Scherer, K. R. (1996) Vocal indicators of mood change in depression. *Journal of Nonverbal Behavior* 20:83–110. [rHA]

Ellis, N. C. (2009) Language is a complex adaptive system: Position paper. *Language Learning* 59:1–26. [TAM]

Elowson, A. M. & Snowdon, C. T. (1994) Pygmy marmosets, *Cebuella pygmaea*, modify vocal structure in response to changed social environment. *Animal Behaviour* 47:1267–77. [aHA]

Elowson, A. M., Snowdon, C. T. & Lazaro-Perea, C. (1998) "Babbling" and social context in infant monkeys: Parallels to human infants. *Trends in Cognitive Sciences* 2:31–37. [PBM]

Enard, W. (2011) FOXP2 and the role of cortico-basal ganglia circuits in speech and language evolution. *Current Opinion in Neurobiology* 21:415–24. [aHA]

Enard, W., Gehre, S., Hammerschmidt, K., Hölter, S. M., Blass, T., Somel, M., Brückner, M. K., Schreiweis, C., Winter, C., Sohr, R., Becker, L., Wiebe, V., Nickel, B., Giger, T., Müller, U., Groszer, M., Adler, T., Aguilar, A., Bolle, I., Calzada-Wack, J., Dalke, C., Ehrhardt, N., Favor, J., Fuchs, H., Gailus-Durner, V., Hans, W., Hölzlwimmer, G., Javaheri, A., Kalaydjiev, S., Kallnik, M., Kling, E., Kunder, S., Mossbrugger, I., Naton, B., Racz, I., Rathkolb, B., Rozman, J., Schrewe, A., Busch, D. H., Graw, J., Ivandic, B., Klingenspor, M., Klopstock, T., Ollert, M., Quintanilla-Martinez, L., Schulz, H., Wolf, E., Wurst, W., Zimmer, A., Fisher, S. E., Morgenstern, R., Arendt, T., de Angelis, M. H., Fischer, J., Schwarz, J. & Pääbo, S. (2009) A humanized version of Foxp2 affects cortico-basal ganglia circuits in mice. *Cell* 137:961–71. [aHA, KJA]

Enard, W. & Pääbo, S. (2004) Comparative primate genomics. *Annual Review of Genomics and Human Genetics* 5:351–78. [aHA]

Enard, W., Przeworski, M., Fisher, S. E., Lai, C. S., Wiebe, V., Kitano, T., Monaco, A. P. & Pääbo, S. (2002) Molecular evolution of FOXP2, a gene involved in speech and language. *Nature* 418(6900):869–72. [aHA, SJ]

Endicott, P., Ho, S. Y. W. & Stringer, C. (2010) Using genetic evidence to evaluate four palaeoanthropological hypotheses for the timing of Neanderthal and modern human origins. *Journal of Human Evolution* 59:87–95. [aHA]

Esposito, A., Demeurisse, G., Alberti, B. & Fabbro, F. (1999) Complete mutism after midbrain periaqueductal gray lesion. *NeuroReport* 10:681–85. [aHA]

Esposito, G., del Carmen Rostagno, M., Venuti, P., Haltigan, J. D. & Messinger, D. S. (2013) Brief Report: Atypical expression of distress during the separation phase of the strange situation procedure in infant siblings at high risk for ASD. *Journal of Autism and Developmental Disorders* 44(4):975–80. [MHB]

Esposito, G. & Venuti, P. (2009) Comparative analysis of crying in children with autism, developmental delays, and typical development. *Focus on Autism and Other Developmental Disabilities* 24(4):240–47. [MHB]

Esteve-Gibert, N. & Prieto, P. (2013) Prosody signals the emergence of intentional communication in the first year of life: Evidence from Catalan-babbling infants. *Journal of Child Language* 40(5):919–44. [PL]

Evatt, M. L., DeLong, M. R. & Vitek, J. L. (2002) Parkinson's disease. In: *Diseases of the nervous system, vol. 1*, 3rd edition, ed. A. K. Asbury, G. M. McKhann, W. I. McDonald & P. J. Goadsby, pp. 477–89. Cambridge University Press. [aHA]

Fadiga, L., Craighero, L. & D'Ausilio, A. (2009) Broca's area in language, action, and music. *Annals of the New York Academy of Sciences* 1169(1):448–58. doi: 10.1111/j.1749-6632.2009.04582.x. [GP, AZ]

Fagan, B. (2010) *Cro-Magnon: How the Ice Age gave birth to the first modern humans*. Bloomsbury. [rHA]

Falk, D. (2004) Prelinguistic evolution in early hominins: Whence motherese? *Behavioral and Brain Sciences* 27(4):491–503. [aHA, SJ]

Falk, D. (2007) Evolution of the primate brain. In: *Handbook of palaeoanthropology, vol. 2: Primate evolution and human origins*, ed. W. Henke & I. Tattersall, pp. 1133–62. Springer-Verlag. [aHA, BM]

Falk, D. (2009) *Finding our tongues: Mothers, infants and the origins of language*. Basic Books. [aHA]

Fant, G. (1960) *Acoustic theory of speech production*. Mouton. [BdB]

Fant, G. (1970) *Acoustic theory of speech production – with calculations based on X-ray studies of Russian articulations*, 2nd edition. Mouton. [aHA]

Fedurek, P., Schel, A. M. & Slocombe, K. E. (2013) The acoustic structure of chimpanzee pant-hooting facilitates chorusing. *Behavioral Ecology and Sociobiology* 67(11):1781–89. [AR]

Fedurek, P. & Slocombe, K. E. (2011) Primate vocal communication: A useful tool for understanding human speech and language evolution? *Human Biology* 83:153–73. [DHR]

Fee, E. J. (1995) The phonological system of a specifically language-impaired population. *Clinical Linguistics and Phonetics* 9:189–209. [aHA]

Feenders, G., Liedvogel, M., Rivas, M., Zapka, M., Horita, H., Hara, E. & Jarvis, E. D. (2008) Molecular mapping of movement-associated areas in the avian brain: A motor theory for vocal learning origin. *PLOS ONE* 3(3):e1768. doi: 10.1371/journal.pone.0001768. [CIP]

Feldman, R. (2007) Parent–infant synchrony: Biological foundations and developmental outcomes. *Current Directions in Psychological Science* 16(6):340–45. [KJA]

Ferguson, B. & Waxman, S. R. (2013) Communication and categorization: New insights into the relation between speech, labels and concepts for infants. In: *Proceedings of the 35th Annual Conference of the Cognitive Science Society*, ed. M. Knauff, M. Pauen, N. Sebanz & I. Wachsmuth, pp. 2267–72. Cognitive Science Society. [BF]

Fernandez-Duque, E., Valeggia, C. R. & Mendoza, S. P. (2009) The biology of paternal care in human and nonhuman primates. *Annual Review of Anthropology* 38:115–30. [DYT]

Ferrari, P. F., Paukner, A., Ionica, C. & Suomi, S. J. (2009) Reciprocal face-to-face communication between rhesus macaque mothers and their newborn infants. *Current Biology* 19:1768–72. [DYT]

Ferry, A. L., Hespos, S. J. & Waxman, S. R. (2010) Categorization in 3- and 4-month-old infants: An advantage of words over tones. *Child Development* 81 (2):472–79. [BF]

Ferry, A. L., Hespos, S. J. & Waxman, S. R. (2013) Nonhuman primate vocalizations support categorization in very young human infants. *Proceedings of the National Academy of Sciences USA* 110(38):15231–35. [BF]

Fichtel, C., Hammerschmidt, K. & Jürgens, U. (2001) On the vocal expression of emotion: A multi-parametric analysis of different states of aversion in the squirrel monkey. *Behaviour* 138:97–116. [rHA]

Fink, G. R., Frackowiak, R. S. J., Pietrzyk, U. & Passingham, R. E. (1997) Multiple nonprimary motor areas in the human cortex. *Journal of Neurophysiology* 77:2164–74. [aHA]

Finlay, B. L. & Darlington, R. B. (1995) Linked regularities in the development and evolution of mammalian brains. *Science* 268:1578–84. [aHA]

Fiorillo, C. D. (2013) Two dimensions of value: Dopamine neurons represent reward but not aversiveness. *Science* 341:546–49. [CMV]

Fischer, J. (2003) Developmental modifications in the vocal behavior of non-human primates. In: *Primate audition: Ethology and neurobiology*, ed. A. A. Ghazanfar, pp. 109–25. CRC Press. [aHA]

Fischer, J., Hammerschmidt, K., Cheney, D. L. & Seyfarth, R. M. (2002) Acoustic features of male baboon loud calls: Influences of context, age, and individuality. *Journal of the Acoustical Society of America* 111:1465–74. [aHA]

Fischer, J., Kitchen, D. M., Seyfarth, R. M. & Cheney, D. L. (2004) Baboon loud calls advertise male quality: Acoustic features and their relation to rank, age, and exhaustion. *Behavioral Ecology and Sociobiology* 56:140–48. [aHA]

Fisher, S. E., Lai, C. S. L. & Monaco, A. P. (2003) Deciphering the genetic basis of speech and language disorders. *Annual Review of Neuroscience* 26:57–80. [aHA]

Fisher, S. E. & Scharff, C. (2009) FOXP2 as a molecular window into speech and language. *Trends in Genetics* 25:166–77. [aHA]

Fitch, W. T. (1997) Vocal tract length and formant frequency dispersion correlate with body size in rhesus macaques. *Journal of the Acoustical Society of America* 102:1213–22. [aHA]

Fitch, W. T. (2000a) The evolution of speech: A comparative review. *Trends in Cognitive Sciences* 4(7):258–67. [aHA, BdB]

Fitch, W. T. (2000b) The phonetic potential of nonhuman vocal tracts: Comparative cineradiographic observations of vocalizing animals. *Phonetica* 57:205–18. [aHA]

Fitch, W. T. (2006) The biology and evolution of music: A comparative perspective. *Cognition* 100:173–215. [TAM]

Fitch, W. T. (2010) *The evolution of language*. Cambridge University Press. [PL]

Fitch, W. T. (2011) Speech perception: A language-trained chimpanzee weighs in. *Current Biology* 21(14):R543–46. [BF]

Fitch, W. T. (2012) The biology and evolution of rhythm: Unraveling a paradox. In: *Language and music as cognitive systems*, ed. P. Rebuschat, M. Rohrmeier, J. A. Hawkins & I. Cross, pp. 73–95. Oxford University Press. [aHA, AR]

Fitch, W. T. (2013) Rhythmic syntax and rhythmic cognition in humans and animals: A computational and comparative perspective. *Frontiers in Systems Neuroscience* **7**:68. doi: 10.3389/fnsys.2013.00068. [HH]

Fitch, W. T. & Hauser, M. D. (2004) Computational constraints on syntactic processing in a nonhuman primate. *Science* 303(5656):377–80. doi: 10.1126/science.1089401. [CIP]

Fitch, W. T., Huber, L. & Bugnyar, T. (2010) Social cognition and the evolution of language: Constructing cognitive phylogenies. *Neuron* 65(6):795–814. doi: 10.1016/j.neuron.2010.03.011. [aHA, ARL, BM, CIP]

Fitch, W. T. & Jarvis, E. D. (2012) Birdsong and other animal models for human speech, song, and vocal learning. In: *Language, music and the brain*, ed. M. Arbib, pp. 499–540. MIT Press. [CIP]

Fitch, W. T. & Reby, D. (2001) The descended larynx is not uniquely human. *Proceedings of the Royal Society, B: Biological Sciences* 268:1669–75. [aHA]

Flaherty, A. W. & Graybiel, A. M. (1993) Two input systems for body representations in the primate striatal matrix: Experimental evidence in the squirrel monkey. *Journal of Neuroscience* 13:1120–37. [aHA]

Flaherty, A. W. & Graybiel, A. M. (1994) Input-output organization of the sensorimotor striatum in the squirrel monkey. *Journal of Neuroscience* 14:599–610. [aHA]

Flege, J. E., MacKay, I. R. & Meador, D. (1999a) Native Italian speakers' perception and production of English vowels. *Journal of the Acoustical Society of America* 106(5):2973–87. [CM]

Flege, J. E., Yeni-Komshian, G. H. & Liu, S. (1999b) Age constraints on Second-Language Acquisition. *Journal of Memory and Language* 41:78–104. [CM]

Fletcher, N. H. (1993) Autonomous vibration of simple pressure-controlled valves in gas flows. *Journal of the Acoustical Society of America* 93(4, Pt. 1):2172–80. [BdB]

Fletcher, P. (1990) Speech and language defects. *Nature* 346(6281):226. [KJA]

Flowers, K. A. & Robertson, C. (1985) The effects of Parkinson's disease on the ability to maintain a mental set. *Journal of Neurology, Neurosurgery, and Psychiatry* 48(6):517–29. [PL]

Franklin, B. S., Warlaumont, A. S., Messinger, D. S., Bene, E. R., Iyer, S. N., Lee, C.-C., Lambert, B. & Oller, D. K. (2014) Effects of parental interaction on infant vocalization rate, variability and vocal type. *Language Learning and Development* 10(3):279–96. doi.org/10.1080/15475441.2013.849176. [DKO]

French, C. A., Jin, X., Campbell, T. G., Gerfen, E., Groszer, M., Fisher, S. E. & Costa, R. M. (2012) An aetiological Foxp2 mutation causes aberrant striatal activity and alters plasticity during skill learning. *Molecular Psychiatry* 17:1077–85. [CMV]

Friederici, A. D. (2011) The brain basis of language processing: From structure to function. *Physiological Reviews* 91(4):1357–92. doi: 10.1152/physrev.00006.2011. [CIP]

Friederici, A. D. & Kotz, S. A. (2003) The brain basis of syntactic processes: Functional imaging and lesion studies. *NeuroImage* 20:S8–S17. doi: 10.1016/S1053-8119(03)00522-6. [AR]

Friston, K. (2010) The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience* 11:127–38. [GP]

Frith, C. (2009) Role of facial expressions in social interactions. *Philosophical Transactions of the Royal Society B: Biological Sciences* 364:3453–58. [CMV]

Frodi, A. (1985) When empathy fails: Aversive infant crying and child abuse. In: *Infant crying: Theoretical and research perspective*, ed. B. M. Lester & C. F. Z. Boukydis, pp. 263–78. Plenum Press. [DLB]

Frodi A. & Senchack, M. (1990) Verbal and behavioral responsiveness to the cries of atypical infants. *Child Development* 61:76–84. [DLB]

Frühholz, S. & Grandjean, D. (2013) Processing of emotional vocalizations in bilateral inferior frontal cortex. *Neuroscience and Biobehavioral Reviews*. 37 (10):2847–55. [SF]

Fulkerson, A. L. & Waxman, S. R. (2007) Words (but not tones) facilitate object categorization: Evidence from 6- and 12-month-olds. *Cognition* 105(1):218–28. [BF]

Gabrieli, J. D., Stebbins, G. T., Singh, J., Willingham, D. B. & Goetz, C. G. (1997) Intact mirror-tracing and impaired rotary-pursuit skill learning in patients with Huntington's disease: Evidence for dissociable memory systems in skill learning. *Neuropsychology* 11(2):272–81. [UH]

Gale, J. T., Lee, K. H., Amirnovin, R., Roberts, D. W., Williams, Z. M., Blaha, C. D. & Eskandar, E. N. (2013) Electrical stimulation-evoked dopamine release in the primate striatum. *Stereotactic Functions Neurosurgery* 91(6):355–63. doi: 10.1159/000351523. [CMV]

Gantz, S. C., Ford, C. P., Neve, K. A. & Williams, J. T. (2011) Loss of Mecp2 in substantia nigra dopamine neurons compromises the nigrostriatal pathway. *The Journal of Neuroscience* 31:12629–37. [PBM]

Garnier, M., Henrich, N. & Dubois, D. (2010) Influence of sound immersion and communicative interaction on the Lombard effect. *Journal of Speech, Language, and Hearing Research* 53:588–608. [DJW]

Gaser, C. & Schlaug, G. (2003) Brain structures differ between musicians and nonmusicians. *Journal of Neuroscience* 23:9240–45. [aHA]

Geissmann, T. (1984) Inheritance of song parameters in the gibbon song, analysed in 2 hybrid gibbons (*Hylobates pileatus* X *H. lar*). *Folia Primatologica* 42:216–35. [aHA]

Geissmann, T. (2000) Gibbon songs and human music from an evolutionary perspective. In: *The origins of music*, ed. N. L. Wallin, B. Merker & S. Brown, pp. 103–23. MIT Press. [aHA, AR]

Gentilucci, M. & Dalla Volta, R. (2008) Spoken language and arm gestures are controlled by the same motor control system. *The Quarterly Journal of Experimental Psychology* 61:944–57. [AM]

Gerardin, E., Lehéricy, S., Pochon, J.-B., Tézenas du Montcel, S., Mangin, J.-F., Poupon, F., Agid, Y., Le Bihan, D. & Marsault, C. (2003) Foot, hand, face and eye representation in the human striatum. *Cerebral Cortex* 13:162–69. [aHA]

Gerfen, C. R. (2010) Functional neuroanatomy of dopamine in the striatum. In: *Dopamine handbook*, ed. L. L. Iversen, S. D. Iversen, S. B. Dunnett & A. Björklund, pp. 11–21. Oxford University Press. [aHA]

Gerfen, C. R. & Bolam, J. P. (2010) The neuroanatomical organization of the basal ganglia. In: *Handbook of basal ganglia structure and function*, ed. H. Steiner & K. Y. Tseng, pp. 3–28. Elsevier. [aHA]

Gerfen, C. R. & Surmeier, D. J. (2011) Modulation of striatal projection systems by dopamine. *Annual Review of Neuroscience* 34:441–66. [aHA]

Ghanbarian, E. & Motamedi, F. (2013) Ventral tegmental area inactivation suppresses the expression of CA1 long term potentiation in anesthetized rat. *PLOS ONE* 8:e58844. [CMV]

Ghazanfar, A. A. (2013) Multisensory vocal communication in primates and the evolution of rhythmic speech. *Behavioral Ecology and Sociobiology* 67(9):1441–48. [DYT]

Ghazanfar, A. A., Chandrasekaran, C. & Morrill, R. J. (2010) Dynamic, rhythmic facial expressions and the superior temporal sulcus of macaque monkeys: Implications for the evolution of audiovisual speech. *European Journal of Neuroscience* 31:1807–17. [DYT]

Ghazanfar, A. A. & Logothetis, N. K. (2003) Facial expressions linked to monkey calls. *Nature* 423(6943):937–38. [DYT]

Ghazanfar, A. A. & Miller, C. T. (2006) Language evolution: Loquacious monkey brains? *Current Biology* 16:R879–81. [aHA]

Ghazanfar, A. A., Morill, R. J. & Kayser, C. (2013) Monkeys are perceptually tuned to facial expressions that exhibit a theta-like speech rhythm. *Proceedings of the National Academy of Sciences USA* 110:1959–63. [aHA]

Ghazanfar, A. A. & Poeppel, D. (2014) The neurophysiology and evolution of the speech rhythm. In: *The cognitive neurosciences V* (5th edition), ed. M. S. Gazzaniga & G. R. Mangun, pp. 629–38. MIT Press. [DYT]

Ghazanfar, A. A. & Rendall, D. (2008) Evolution of human vocal production. *Current Biology* 18(11):R457–60. [aHA, DYT]

Ghazanfar, A. A., Takahashi, D. Y., Mathur, N. & Fitch, W. T. (2012) Cineradiography of monkey lipsmacking reveals the putative origins of speech dynamics. *Current Biology* 22:1176–82. [aHA, DYT]

Ghazanfar, A. A., Turesson, H. K., Maier, J. X., van Dinther, R., Patterson, R. D. & Logothetis, N. K. (2007) Vocal tract resonances as indexical cues in rhesus monkeys. *Current Biology* 17:425–30. [DYT]

Gil-da-Costa, R., Martin, A., Lopes, M. A., Muñoz, M., Fritz, J. B. & Braun, A. R. (2006) Species-specific calls activate homologs of Broca's and Wernicke's areas in the macaque. *Nature Neuroscience* 9:1064–70. [aHA]

Gintis, H. (2006) The economy as a complex adaptive system. (Online paper). Available at: http://www.umass.edu/preferen/Class%20Material/Readings%20in%20Market%20Dynamics/Complexity%20Economics.pdf [TAM]

Giroud, M., Lemesle, M., Madinier, G., Billiar, T. & Dumas, R. (1997) Unilateral lenticular infarcts: Radiological and clinical syndromes, aetiology, and prognosis. *Journal of Neurology, Neurosurgery, and Psychiatry* 63:611–15. [aHA]

Goldin-Meadow, S. & Mylander, C. (1983) Gestural communication in deaf children: Noneffect of parental input on language development. *Science, New Series* 221(4608):372–74. [BF]

Goldstein, L. M., Byrd, D. & Saltzman, E. (2006) The role of vocal tract gestural action units in understanding the evolution of phonology. In: *Action to language via the mirror neuron system*, ed. M. A. Arbib, pp. 215–49. Cambridge University Press. [rHA]

Goldstein, M. H., Schwade, J. A. & Bornstein, M. H. (2009) The value of vocalizing: Five-month-old infants associate their own noncry vocalizations with responses from adults. *Child Development* 80(3):636–44. [DKO]

Gomez Portillo, I. J. & Gleiser, P. M. (2009) An adaptive complex network model for brain functional networks. *PLoS ONE* 4(9):e6863. [TAM]

Gonzalez-Lima, F. (2010) Responses of limbic, midbrain and brainstem structures to electrically-induced vocalizations. In: *Handbook of mammalian vocalization: An integrative neuroscience approach*, ed. S. M. Brudzynski, pp. 293–301. Elsevier. [aHA]

Goodall, J. (1986) *The chimpanzees of Gombe: Patterns of behavior*. Belknap Press/Harvard University Press. [aHA]

Gopnik, M. (1990a) Feature-blind grammar and dysphasia. *Nature* 344(6268):715. doi: 10.1038/344715a0. [aHA, AZ]

Gopnik, M. (1990b) Genetic basis of grammar defect. *Nature* 347(6288):26. [KJA]

Grahn, J. A. & Brett, M. (2007) Rhythm and beat perception in motor areas of the brain. *Journal of Cognitive Neuroscience* 19(5):893–906. doi: 10.1162/jocn.2007.19.5.893. [HH]

Granata, A. R. & Woodruff, G. N. (1982) Dopaminergic mechanisms in the nucleus tractus solitarius and effects on blood pressure. *Brain Research Bulletin* 8:483–88. [CMV]

Graybiel, A. M. (1990) Neurotransmitters and neuromodulators in the basal ganglia. *Trends in Neurosciences* 13:244–54. [aHA]

Graybiel, A. M. (2005) The basal ganglia: Learning new tricks and loving it. *Current Opinion in Neurobiology* 15:638–44. [aHA]

Graybiel, A. M. (2008) Habits, rituals, and the evaluative brain. *Annual Review of Neuroscience* 31:359–87. doi: 10.1146/annurev.neuro.29.051605.112851. [aHA, AZ]

Green, R. E., Krause, J. Briggs, A. W., Maricic, T., Stenzel, U., Kircher, M., Patterson, N., Li, H., Zhai, W., Fritz, M. H.-Y., Hansen, N. F., Durand, E. Y., Malaspinas, A.-S., Jensen, J. D., Marques-Bonet, T., Alkan, C., Prüfer, K., Meyer, M., Burbano, H. A., Good, J. M., Schultz, R., Aximu-Petri, A., Butthof, A., Höber, B., Höffner, B., Siegemund, M., Weihmann, A., Nusbaum, C., Lander, E. S., Russ, C., Novod, N., Affourtit, J., Egholm, M., Verna, C., Rudan, P., Brajkovic, D., Kucan, E., Gusic, I., Doronichev, V. B., Golovanova, L. V., Lalueza-Fox, C. , de la Rasilla, M., Fortea, J., Rosas, A., Schmitz, R. W., Johnson, P. L. F., Eichler, E. E., Falush, D., Birney, E., Mullikin, J. C., Slatkin, M., Nielsen, R., Kelso, J., Lachmann, M., Reich, D. & Pääbo, S. (2010) A draft sequence of the Neandertal genome. *Science* 328(5979):710–22. [aHA, SJ]

Greenberg, S., Carvey, H., Hitchcock, L. & Chang, S. (2003) Temporal properties of spontaneous speech – a syllable-centric perspective. *Journal of Phonetics* 31 (3–4):465–85. [DYT]

Griebel, U. & Oller, D. K. (2008) Evolutionary forces favoring contextual flexibility. In: *Evolution of communicative flexibility: Complexity, creativity and adaptability in human and animal communication*, ed. D. K. Oller & U. Griebel, pp. 9–40. MIT Press. [DKO]

Grillner, S. (1991) Recombination of motor pattern generators. *Current Biology* 1:231–33. [aHA]

Grillner, S., Deliagina, T., Ekeberg, O., el Manira, A., Hill, R. H., Lansner, A., Orlovsky, G. N. & Wallén, P. (1995) Neural networks that co-ordinate locomotion and body orientation in lamprey. *Trends in Neurosciences* 18:270–79. [PBM]

Grillner, S. & Wallén, P. (2004) Innate versus learned movements – a false dichotomy? In: *Brain mechanisms for the integration of posture and movement*, ed. S. Mori, D. G. Stuart & M. Wiesendanger, pp. 3–12. (*Progress in Brain Research, vol. 143*). Elsevier. [aHA]

Grilo, A., Caetano, A. & Rosa, A. (2002) Immune system simulation through a complex adaptive system model. In: *Soft Computing and Industry*, ed. R. Roy, M. Köppen, S. Ovaska, T. Furuhashi, & F. Hoffmann, pp. 675–98. Springer. [TAM]

Groenewegen, H. J. (2003) The basal ganglia and motor control. *Neural Plasticity* 10:107–20. [aHA]

Grossman, M., Carvell, S., Gollomp S., Stern, M. B., Vernon, G. & Hurtig, H. I. (1991) Sentence comprehension and praxis deficits in Parkinson's disease. *Neurology* 41(10):1620–26. [PL]

Groswasser, Z., Korn, C., Groswasser-Reider, I. & Solzi, P. (1988) Mutism associated with buccofacial apraxia and bihemispheric lesions. *Brain and Language* 34:157–68. [aHA]

Groszer, M., Keays, D. A., Deacon, R. M. J., de Bono, J. P., Prasad-Mulcare, S., Gaub, S., Baum, M. G., French, C. A., Nicod, J., Coventry, J. A., Enard, W., Fray, M., Brown, S. D. M., Nolan, P. M., Pääbo, S., Channon, K. M., Costa, R. M., Ellers, J., Ehret, G., Rawlins, J. N. P. & Fisher, S. E. (2008) Impaired synaptic plasticity and motor learning in mice carrying a point mutation implicated in human speech deficits. *Current Biology* 18:354–62. [KJA]

Gruber, T. & Zuberbühler, K. (2013) Vocal recruitment for joint travel in wild chimpanzees. *PLOS ONE* 8:e76073. [ARL]

Gruber-Dujardin, E. (2010) Role of the periaqueductal gray in expressing vocalization. In: *Handbook of mammalian vocalization: An integrative neuroscience approach*, ed. S. M. Brudzynski, pp. 313–27. Elsevier. [aHA]

Guenther, F. H. & Perkell, J. S. (2004) A neural model of speech production and its application to studies of the role of auditory feedback in speech. In: *Speech motor control in normal and disordered speech*, ed. B. Maassen, R. Kent, H. Peters, P. Van Lieshout & W. Hulstijn, pp. 29–49. Oxford University Press. [GP]

Gustafson, G. E., Wood, R. M. & Green, J. A. (2000) Can we hear the causes of infants' crying? In: *Crying as a sign, a symptom, and a signal: Clinical, emotional and developmental aspects of infant and toddler crying*, ed. R. G. Barr, B. Hopkins & J. A. Green, pp. 8–22. Mac Keith Press/Wiley. [DLB]

Habbershon, H. M., Ahmed, S. Z. & Cohen, Y. E. (2013) Rhesus macaques recognize unique multimodal face-voice relations of familiar individuals and not of unfamiliar ones. *Brain, Behavior, and Evolution* 81:219–25. [DYT]

Haber, S. N. (2003) The primate basal ganglia: Parallel and integrative networks. *Journal of Chemical Neuroanatomy* 26(4):317–30. [UH]

Haber, S. N. (2010a) Integrative networks across basal ganglia circuits. In: *Handbook of basal ganglia structure and function*, ed. H. Steiner & K. Y. Tseng, pp. 409–27. Elsevier. [aHA]

Haber, S. N. (2010b) Convergence of limbic, cognitive, and motor cortico-striatal circuits with dopamine pathways in primate brain. In: *Dopamine handbook*, ed. L. L. Iversen, S. D. Iversen, S. B. Dunnett & A. Björklund, pp. 38–48. Oxford University Press. [aHA]

Haber, S. N., Fudge, J. L. & McFarland, N. R. (2000) Striatonigrostriatal pathways in primates form an ascending spiral from the shell to the dorsolateral striatum. *Journal of Neuroscience* 20:2369–82. [aHA]

Haber, S. N., Kunishio, K., Mizobuchi, M. & Lynd-Balta, E. (1995) The orbital and medial prefrontal circuit through the primate basal ganglia. *Journal of Neuroscience* 15:4851–67. [aHA]

Haesler, S., Rochefort, C., Georgi, B., Licznerski, P., Osten, P. & Scharff, C. (2007) Incomplete and inaccurate vocal imitation after knockdown of FoxP2 in songbird basal ganglia nucleus area X. *PLoS Biology* 5:2885–97. [aHA]

Haesler, S., Wada, K., Nshdejan, A., Morrisey, E. E., Lints, T., Jarvis, E. D. & Scharff, C. (2004) FoxP2 expression in avian vocal learners and non-learners. *The Journal of Neuroscience* 24(13):3164–75. [SJ]

Hage, S. R. (2010a) Localization of the central pattern generator for vocalization. In: *Handbook of mammalian vocalization: An integrative neuroscience approach*, ed. S. M. Brudzynski, pp. 329–37. Elsevier. [aHA]

Hage, S. R. (2010b) Neuronal networks involved in the generation of vocalization. In: *Handbook of mammalian vocalization: An integrative neuroscience approach*, ed. S. M. Brudzynski, pp. 339–49. Elsevier. [aHA]

Hage, S. R., Gavrilov, N. & Nieder, A. (2013) Cognitive control of distinct vocalizations in rhesus monkeys. *Journal of Cognitive Neuroscience* 25:1692–701. doi:0.1162/jocn_a_00428. [aHA]

Hage, S. R., Jiang, T., Berquist, S. W., Feng, J. & Metzner, W. (2013) Ambient noise induces independent shifts in call frequency and amplitude within the Lombard effect in echolocating bats. *Proceedings of the National Academy of Sciences USA* 110(10):4063–68. [DJW, rHA]

Hage, S. R. & Jürgens, U. (2006) On the role of the pontine brainstem in vocal pattern generation: A telemetric single-unit recording study in the squirrel monkey. *Journal of Neuroscience* 26:7105–15. [aHA, PBM]

Hage, S. R. & Nieder, A. (2013) Single neurons in monkey prefrontal cortex encode volitional initiation of vocalizations. *Nature Communications* 4:2409. [rHA, MHB]

Hagen, E. H. & Bryant, G. A. (2003) Music and dance as a coalition signaling system. *Human Nature* 14(1):21–51. [GAB, AR]

Hagen, E. H. & Hammerstein, P. (2009) Did Neanderthals and other early humans sing? Seeking the biological roots of music in the territorial advertisements of primates, lions, hyenas, and wolves. *Musicae Scientiae* 13(Suppl. 2):291–320. [GAB, AR]

Hammerschmidt, K. & Fischer, J. (2008) Constraints in primate vocal production. In: *Evolution of communicative flexibility: Complexity, creativity, and adaptability in human and animal communication*, ed. D. K. Oller & U. Griebel, pp. 93–119. MIT Press. [aHA]

Hanakawa, T., Dimyan, M. A. & Hallett, M. (2008) Motor planning, imagery, and execution in the distributed motor network: A time-course study with functional MRI. *Cerebral Cortex* 18:2775–88. [TH]

Hanakawa, T., Honda, M., Sawamoto, N., Okada, T., Yonekura, Y., Fukuyama, H. & Shibasaki, H. (2002) The role of rostral Brodmann area 6 in mental-operation tasks: An integrative neuroimaging approach. *Cerebral Cortex* 12:1157–70. [TH]

Hardus, M. E., Lameira, A. R., Singleton, I., Morrogh-Bernard, H. C., Knott, C. D., Ancrenaz, M., Utami Atmoko, S. S. & Wich, S. A. (2009a) A description of the orangutan's vocal and sound repertoire, with a focus on geographic variation. In: *Orangutans: Geographic variation in behavioral ecology and conservation*, ed. S. A. Wich, S. S. Utami Atmoko, T. M. Setia & C. P. van Schaik, pp. 49–64. Oxford University Press. [aHA]

Hardus, M. E., Lameira, A. R., van Schaik, C. P. & Wich, S. A. (2009b) Tool use in wild orang-utans modifies sound production: A functionally deceptive innovation? *Proceedings of the Royal Society B: Biological Sciences* 276:3689–94. [aHA, ARL]

Harrington, D. L. & Haaland, K. Y. (1991) Sequencing in Parkinson's disease: Abnormalities programming and controlling movement. *Brain* 114:99–115. [PL]

Harrington, J., Palethorpe, S. & Watson, C. I. (2000) Does the Queen speak the Queen's English? Elizabeth II's traditional pronunciation has been influenced by modern trends. *Nature* 408(6815):927–28. doi: 10.1038/35050160. [CM]

Hasegawa, A., Okanoya, K., Hasegawa, T. & Seki, Y. (2011) Rhythmic synchronization tapping to an audio–visual metronome in budgerigars. *Scientific Reports* 1(Article 120):1–8. (Online journal). doi:10.1038/srep00120. [HH, AR]

Hasson, U., Ghazanfar, A. A., Galantucci, B., Garrod, S. & Keysers, C. (2012) Brain-to-brain coupling: A mechanism for creating and sharing a social world. *Trends in Cognitive Sciences* 16(2):114–21. [DYT]

Hast, M. H., Fischer, J. M., Wetzel, A. B. & Thompson, V. E. (1974) Cortical motor representation of the laryngeal muscles in *Macaca mulatta*. *Brain Research* 73:229–40. [aHA]

Hattori, Y., Tomonaga, M. & Matsuzawa, T. (2013) Spontaneous synchronized tapping to an auditory rhythm in a chimpanzee. *Scientific Reports* 3(Article 1566):1–6. (Online journal). doi: 10.1038/srep01566. [HH, AR]

Hauser, M. D., Chomsky, N. & Fitch, W. T. (2002) The faculty of language: What is it, who has it, and how did it evolve? *Science* 298(5598):1569. [BF]

Hawkins, S. & Smith, R. (2001) Polysp: A polysystemic, phonetically-rich approach to speech understanding. *Italian Journal of Linguistics* 13:99–188. [CM]

Hayes, C. (1951) *The ape in our house*. Harper & Brothers. [aHA]

Hayes, K. J. & Hayes, C. (1952) Imitation in a home-raised chimpanzee. *Journal of Comparative and Physiological Psychology* 45:450–59. [aHA]

Heilman, K. M., Leon, S. A. & Rosenbek, J. C. (2004) Affective aprosodia from a medial frontal stroke. *Brain and Language* 89:411–16. [aHA]

Henry, J. D. & Crawford, J. R. (2004) Verbal fluency deficits in Parkinson's disease: A meta-analysis. *Journal of the International Neuropsychological Society* 10:608–22. [AR]

Hernandez, A. E., Dapretto, M., Mazziotta, J. & Bookheimer, S. (2001) Language switching and language representation in Spanish-English bilinguals: An fMRI study. *NeuroImage* 14:510–20. [TH]

Hewitt, G. P., MacLarnon, A. & Jones, K. E. (2002) The functions of laryngeal air sacs in primates: A new hypothesis. *Folia Primatologica* 73:70–94. [BdB]

Hickok, G. (2012) Computational neuroanatomy of speech production. *Nature Reviews. Neuroscience* 13(2):135–45. doi: 10.1038/nrn3158. [UH]

Hihara, S., Yamada, H., Iriki, A. & Okanoya, K. (2003) Spontaneous vocal differentiation of coo-calls for tools and food in Japanese monkeys. *Neuroscience Research* 45:383–89. [aHA]

Hikosaka, O. (2007) GABAergic output of the basal ganglia. In: *GABA and the basal ganglia: From molecules to systems*, ed. J. M. Tepper, E. D. Abercrombie & J. P. Bolam, pp. 209–26. (*Progress in Brain Research, vol. 160*). Elsevier. [aHA, PBM]

Hillix, W. A. (2007) The past, present, and possible futures of animal language research. In: *Primate perspectives on behavior and cognition*, ed. D. A. Washburn, pp. 223–34. American Psychological Association. [aHA]

Hinton, G. E. (2007) Learning multiple layers of representation. *Trends in Cognitive Sciences* 11:428–34. [GP]

Hirose, H. (2010) Investigating the physiology of laryngeal structures. In: *The handbook of phonetic sciences*, 2nd edition, ed. W. J. Hardcastle, J. Laver & F. E. Gibbon, pp. 130–52. Wiley-Blackwell. [aHA]

Ho, A. K., Bradshaw, J. L., Iansek, R. & Alfredson, R. (1999a) Speech volume regulation in Parkinson's disease: Effects of implicit cues and explicit instructions. *Neuropsychologia* 37:1453–60. [aHA]

Ho, A. K., Iansek, R. & Bradshaw, J. L. (1999b) Regulation of Parkinsonian speech volume: The effect of interlocuter distance. *Journal of Neurology, Neurosurgery, and Psychiatry* 67:199–202. [aHA]

Hockett, C. F. (1960) The origin of speech. *Scientific American* 203:89–96. [BM]

Hofreiter, M. (2011) Drafting human ancestry: What does the Neanderthal genome tell us about hominid evolution? Commentary on Green et al. (2010). *Human Biology* 83:1–11. [aHA]

Hogan, J. A. (1988) Cause and function in the development of behavior systems. In: *Handbook of Behavioral Neurobiology, vol. 9: Developmental psychobiology and behavioral ecology*, ed. E. M. Blass, pp. 63–106. Plenum. [DLB]

Holland, J. H. (2012) *Signals and boundaries: Building blocks for complex adaptive systems*. MIT Press. [TAM]

Holloway, R. L., Broadfield, D. C. & Yuan, M. S. (2004) *The human fossil record: Vol. III. Brain endocasts – the paleoneurological evidence*. Wiley. [aHA]

Holstege, G. (1991) Descending pathways from the periaqueductal gray and adjacent areas. In: *Midbrain periaqueductal gray matter: Functional, anatomical, and neurochemical organization*, ed. A. Depaulis & R. Bandler, pp. 239–65. Plenum Press. (NATO ASI Series, *Life Sciences, vol. 213*). [aHA]

Honing, H. (2012) Without it no music: Beat induction as a fundamental musical trait. *Annals of the New York Academy of Sciences* 1252(1):85–91. doi: 10.1111/j.1749-6632.2011.06402.x. [HH]

Honing, H., Bouwer, F. & Háden, G. P. (in press) Perceiving temporal regularity in music: The role of auditory event-related potentials (ERPs) in probing beat perception. In: *Neurophysiology of temporal processing*, ed. H. Merchant & V. de Lafuente. Springer. [HH]

Honing, H., Merchant, H., Háden, G. P., Prado, L. & Bartolo, R. (2012) Rhesus monkeys (*Macaca mulatta*) detect rhythmic groups in music, but not the beat. *PLoS ONE* 7(12):1–10. doi: 10.1371/journal.pone.0051369. [HH]

Honing, H. & Ploeger, A. (2012) Cognition and the evolution of music: Pitfalls and prospects. *Topics in Cognitive Science* 4:513–24. doi: 10.1111/j.1756-8765.2012.01210.x. [HH]

Honing, H., ten Cate, C., Peretz, I., & Trehub, S. (in press) Without it no music: Cognition, biology, and evolution of musicality. *Philosophical Transactions B*.

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

Hopkins, W. D. & Savage-Rumbaugh, E. S. (1991) Vocal communication as a function of differential rearing experiences in *Pan paniscus*: A preliminary report. *International Journal of Primatology* 12:559–83. [aHA]

Hopkins, W. D., Taglialatela, J. P. & Leavens, D. A. (2007) Chimpanzees differentially produce novel vocalizations to capture the attention of a human. *Animal Behaviour* 73(2):281–86. [aHA, AM]

Hopkins, W. D., Taglialatela, J. P. & Leavens, D. A. (2011) Do chimpanzees have voluntary control of their facial expressions and vocalizations? In: *Primate communication and human language: Vocalisation, gestures, imitation and deixis in humans and non-humans*, ed. A. Vilain, C. Abry, J. C. Schwartz & J. Vaucair, pp. 71–88. [*Advances in Interaction Studies, vol. 1*]. John Benjamins. [AM]

Hosoda, C., Hanakawa, T., Nariai, T., Ohno, K. & Honda, M. (2012) Neural mechanisms of language switch. *Journal of Neurolinguistics* 25:44–61. [TH]

Hosoda, C., Tanaka, K., Nariai, T., Honda, M. & Hanakawa, T. (2013) Dynamic neural network reorganization associated with second language vocabulary acquisition: A multimodal imaging study. *The Journal of Neuroscience* 33:13663–72. [TH, rHA]

Hotchkin, C. F. (2012) *Vocal noise compensation in nonhuman mammals: Modification types and usage patterns.* Ph.D. thesis, Department of Ecology, The Pennsylvania State University. [DJW]

Hotchkin, C. F. & Parks, S. (2013) The Lombard effect and other noise-induced vocal modifications: Insight from mammalian communication systems. *Biological Reviews* 88(4):809–24. [DJW]

Hotchkin, C. F., Parks, S. E. & Weiss, D. J. (2013) Vocal modifications in primates: Effects of noise and behavioral context on vocalization structure. In: *Proceedings of Meetings on Acoustics*, Vol. 19, No. 1, Article 010061. Acoustical Society of America. [DJW]

Huang, Y. C. & Hessler, N. A. (2008) Social modulation during songbird courtship potentiates midbrain dopaminergic neurons. *PLOS ONE* 3:e3281. [CMV]

Hurford, J. R. (2007) *The origins of meaning: Language in the light of evolution*. Oxford University Press. [rHA]

Hurst, J., Baraitser, M., Auger, E., Graham, F. & Norell, S. (1990) An extended family with a dominantly inherited speech disorder. *Developmental Medicine and Child Neurology* 32:347–55. [aHA, KJA]

Iannetti, P., Spalice, A., Raucci, U., Atzei, G. & Cipriani, C. (1997) Gelastic epilepsy: Video-EEG, MRI and SPECT characteristics. *Brain and Development* 19:418–21. [aHA]

Ikeda, A., Lüders, H. O., Burgess, R. C. & Shibasaki, H. (1992) Movement-related potentials recorded from supplementary motor area and primary motor area. *Brain* 115:1017–43. [aHA]

Ingold, T. (1994) Tool-using, toolmaking, and the evolution of language. In: *Hominid culture in primate perspective*, ed. D. Quiatt & J. Itani, pp. 279–314. University Press of Colorado. [aHA]

Isler, K. & van Schaik, C. P. (2012) How our ancestors broke through the gray ceiling: Comparative evidence for cooperative breeding in early homo. *Current Anthropology* 53:S453–65. [ARL]

Iwasa, H., Shibata, T., Mine, S., Koseki, K., Yasuda, K., Kasagi, Y., Okada, M., Yabe, H., Kaneko, S. & Nakajima, Y. (2002) Different patterns of dipole source localization in gelastic seizure with or without a sense of mirth. *Neuroscience Research* 43:23–29. [aHA]

Iwatsubo, T., Kuzuhara, S., Kanemitsu, A., Shimada, H. & Toyokura, Y. (1990) Corticofugal projections to the motor nuclei of the brainstem and spinal cord in humans. *Neurology* 40(2):309–12. [aHA, PL, BM]

Jacob, F. (1977) Evolution and tinkering. *Science* 196:1161–66. [aHA]

James, W. (2003) *The ceremonial animal: A new portrait of anthropology*. Oxford University Press. [aHA]

Janik, V. & Slater, P. J. B. (1997) Vocal learning in mammals. In: *Advances in the Study of Behavior, vol. 26*, ed. P. J. B. Slater, J. S. Rosenblatt, C. T. Snowdon, & M. Milinski, pp. 59–99. Academic Press. [aHA, ARL, BM, AR, DJW]

Janik, V. M. & Slater, P. J. B. (2000) The different roles of social learning in vocal communication. *Animal Behaviour* 60:1–11. [aHA, BM]

Jankovic, J. (2008) Parkinson's disease: Clinical features and diagnosis. *Journal of Neurology, Neurosurgery, and Psychiatry* 79:368–76. [arHA]

Jarvis, E. D. (2004a) Brains and birdsong. In: *Nature's music: The science of birdsong*, ed. P. Marler & H. Slabbekoorn, pp. 226–71. Elsevier. [aHA]

Jarvis, E. D. (2004b) Learned birdsong and the neurobiology of human language. In: *Behavioral neurobiology of birdsong*, ed. H. P. Zeigler, P. Marler, pp. 749–77. (*Annals of the New York Academy of Sciences, vol. 1016*). New York Academy of Sciences. [aHA, GJLB, CIP]

Jarvis, E. D. (2006) Selection for and against vocal learning in birds and mammals. *Ornithological Science* 5:5–14. [CIP]

Jarvis, E. D., Ribeiro, S., da Silva, M. L., Ventura, D., Vielliard, J. & Mello, C. V. (2000) Behaviourally driven gene expression reveals song nuclei in hummingbird brain. *Nature* 406(6796):628–32. doi: 10.1038/35020570. [CIP]

Joel, D. & Weiner, I. (1994) The organization of the basal ganglia-thalamocortical circuits: Open interconnected rather than closed segregated. *Neuroscience* 63:363–79. [aHA]

Johansson, S. (2005) *Origins of language: Constraints on hypotheses*. John Benjamins. [SJ, rHA]

Johansson, S. (2011) Constraining the time when language evolved. *Linguistic and Philosophical Investigations* 10:45–59. [SJ]

Johansson, S. (2013) The talking Neanderthals: What do fossils, genetics, and archeology say? *Biolinguistics* 7:35–74. [SJ]

Johansson, S., Zlatev, J. & Gärdenfors, P. (2006) Why don't chimps talk and humans sing like canaries? Comment on Locke & Bogin's "Language and life history". *Behavioral and Brain Sciences* 29(3):287–88. [SJ]

Joint, I. (2006) Bacterial conversations: Talking, listening and eavesdropping. A NERC Discussion Meeting held at the Royal Society on 7 December 2005. *Journal of the Royal Society Interface* 3(8):459–63. [KBC]

Jonas, S. (1981) The supplementary motor region and speech emission. *Journal of Communication Disorders* 14:349–73. [aHA]

Jonas, S. (1987) The supplementary motor region and speech. In: *The frontal lobes revisited*, ed. E. Perecman, pp. 241–50. Erlbaum. [aHA]

Jordan, K. E., Brannon, E. M., Logothetis, N. K. & Ghazanfar, A. A. (2005) Monkeys match the number of voices they hear with the number of faces they see. *Current Biology* 15:1034–38. [DYT]

Jürgens, U. (1974) On the elicitability of vocalization from the cortical larynx area. *Brain Research* 81:564–66. [aHA]

Jürgens, U. (1986) The squirrel monkey as an experimental model in the study of cerebral organization of emotional vocal utterances. *European Archives of Psychiatry and Neurological Sciences* 236:40–43. [aHA]

Jürgens, U. (2002a) A study of the central control of vocalization using the squirrel monkey. *Medical Engineering and Physics* 7–8:473–77. [BM]

Jürgens, U. (2002b) Neural pathways underlying vocal control. *Neuroscience and Biobehavioral Reviews* 26:235–58. [aHA, PL]

Jürgens, U. & Alipour, M. (2002) A comparative study on the cortico-hypoglossal connections in primates, using biotin dextranamine. *Neuroscience Letters* 328:245–48. [aHA]

Jürgens, U., Kirzinger, A. & von Cramon, D. (1982) The effects of deep-reaching lesions in the cortical face area on phonation: A combined case report and experimental monkey study. *Cortex* 18:125–39. [aHA]

Jürgens, U. & Ploog, D. (1970) Cerebral representation of vocalization in the squirrel monkey. *Experimental Brain Research* 10:532–54. [aHA]

Jürgens, U. & Ploog, D. (1988) On the motor control of monkey calls. In: *The physiological control of mammalian vocalization*, ed. J. D. Newman, pp. 7–19. Plenum Press. [DLB]

Jürgens, U. & von Cramon, D. (1982) On the role of the anterior cingulate cortex in phonation: A case report. *Brain and Language* 15:234–48. [aHA]

Kaestner, K. H., Knöchel, W. & Martínez, D. E. (2000) Unified nomenclature for the winged helix/forkhead transcription factors. *Genes and Development* 14:142–46. [aHA]

Kaufmann, W. E., Johnston, M. V. & Blue, M. E. (2005) MeCP2 expression and function during brain development: Implications for Rett syndrome's pathogenesis and clinical evolution. *Brain and Development* 27:S77–S87. [PBM]

Kawashima, S., Ueki, Y., Kato, T., Matsukawa, N., Mima, T., Hallett, M., Ito, K. & Ojika, K. (2012) Changes in striatal dopamine release associated with human motor-skill acquisition. *PLOS ONE* 7:e31728. [aHA]

Kelemen, G. (1969) Anatomy of the larynx and the anatomical basis of vocal performance. In: *The chimpanzee*, ed. G. H. Bourne, pp. 165–86. Karger. [BdB]

Kent, R. D., Kent, J. F., Weismer, G. & Duffy, J. R. (2000) What dysarthrias can tell us about the neural control of speech. *Journal of Phonetics* 28:273–302. [aHA]

Kent, R. D. & Read, C. (2002) *The acoustic analysis of speech*, 2nd edition. Singular/Thomson Learning. [aHA]

Ketteler, D., Kastrau, F., Vohn, R. & Huber, W. (2008) The subcortical role of language processing. High level linguistic features such as ambiguity-resolution and the human brain: An fMRI study. *NeuroImage* 39(4):2002–2009. [KBC]

Kiebel, S. J., Daunizeau, J. & Friston, K. J. (2008) A hierarchy of time-scales and the brain. *PLOS Computational Biology* 4:e1000209. [GP]

Kikuchi, Y., Horwitz, B. & Mishkin, M. (2010) Hierarchical auditory processing directed rostrally along the monkey's supratemporal plane. *Journal of Neuroscience* 30:13021–30. [DHR]

Kim, I.-S., Ki, C.-S. & Park, K.-J. (2010) Pediatric-onset dystonia associated with bilateral striatal necrosis and G14459A mutation in a Korean family: A case report. *Journal of Korean Medical Science* 25:180–84. [aHA]

Kirby, S. (2002) Learning, bottlenecks and the evolution of recursive syntax. In: *Linguistic evolution through language acquisition: Formal and computational models*, ed. T. Briscoe, pp. 173–204. Cambridge University Press. [BM]

Kirby, S., Cornish, H. & Smith, K. (2008) Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language. *Proceedings of the National Academy of Sciences USA* 105:10681–86. doi: 10.1073#pnas.0707835105. [BM]

Kirk, E., Howlett, N., Pine, K. J. & Fletcher, B. (C). (2012) To sign or not to sign? The impact of encouraging infants to gesture on infant language and maternal mind-mindedness. *Child Development* 84(2):574–90. [KJA]

Kirschner, S. & Tomasello, M. (2010) Joint music making promotes prosocial behavior in 4- year-old children. *Evolution and Human Behavior* 31:354–64. [GAB]

Kirzinger, A. (1985) Cerebellar lesion effects on vocalization of the squirrel monkey. *Behavioural Brain Research* 16:177–81. [arHA]

Kirzinger, A. & Jürgens, U. (1982) Cortical lesion effects and vocalization in the squirrel monkey. *Brain Research* 233:299–315. [aHA]

Kisilevsky, B. S., Hains, S. M. J., Jacquet, A.-Y., Granier-Deferre, D. & Lecanuet, J. P. (2004) Maturation of fetal responses to music. *Developmental Science* 7 (5):550–59. [DLB]

Kittelberger, J. M. & Bass, A. H. (2013) Vocal-motor and auditory connectivity of the midbrain periaqueductal gray in a teleost fish. *Journal of Comparative Neurology* 521:791–812. [PBM]

Kjelgaard, M. M. & Speer, S. R. (1999) Prosodic facilitation and interference in the resolution of temporary syntactic closure ambiguity. *Journal of Memory and Language* 40:153–94. [AR]

Kluender, K. R., Diehl, R. L. & Killeen, P. R. (1987) Japanese quail can learn phonetic categories. *Science* 237:1195–97. [DLB]

Knecht, S., Dräger, B., Deppe, M., Bobe, L., Lohmann, H., Flöel, A., Ringelstein, E. B. & Henningsen, H. (2000) Handedness and hemispheric language dominance in healthy humans. *Brain* 123:2512–18. [AM]

Knight, C. (1999) Sex and language as pretend-play. In: *The evolution of culture: An interdisciplinary view*, ed. R. Dunbar, C. Knight & C. Power, pp. 228–47. Edinburgh University Press. [aHA]

Knolle, F., Schroger, E. & Kotz, S. A. (2013) Cerebellar contribution to the prediction of self-initiated sounds. *Cortex* 49(9):2449–61. doi: 10.1016/j.cortex.2012.12.012. [UH]

Koda, H., Lemasson, A., Oyakawa, C., Rizaldi, P. J. & Masataka, N. (2013) Possible role of mother-daughter vocal interactions on the development of species-specific song in gibbons. *PLOS ONE* 8:e71432. [ARL]

Koda, H., Oyakawa, C., Kato, A. & Masataka, N. (2007) Experimental evidence for the volitional control of vocal production in an immature gibbon. *Behaviour* 144:681–92. [aHA]

Konczak, J., Ackermann, H., Hertrich, I., Spieker, S. & Dichgans, J. (1997) Control of repetitive lip and finger movements in Parkinson's disease: Influence of external timing signals and simultaneous execution on motor performance. *Movement Disorders* 12:665–76. [rHA]

Konopka, G., Bomar, J. M., Winden, K., Coppola, G., Jonsson, Z. O., Gao, F., Peng, S., Preuss, T. M., Wohlschlegel, J. A. & Geschwind, D. H. (2009) Human-specific transcriptional regulation of CNS development genes by FOXP2. *Nature* 462(7270):213–17. [KJA]

Koopmans-van Beinum, F. J. & van der Stelt, J. M. (1986) Early stages in the development of speech movements. In: *Precursors of early speech*, ed. B. Lindblom & R. Zetterstrom, pp. 37–50. Stockton Press. [DKO]

Kotz, S. A., Frisch, S., Cramon, S. Y. & Friederici, A. D. (2003) Syntactic language processing: ERP lesion data on the role of the basal ganglia. *Journal of the International Neuropsychological Society* 9:1053–60. [AR]

Kotz, S. A., Kalberlah, C., Bahlmann, J., Friederici, A. D. & Haynes, J. D. (2013) Predicting vocal emotion expressions from the human brain. *Human Brain Mapping* 34(8):1971–81. doi: 10.1002/hbm.22041. [UH]

Kotz, S. A. & Schwartze, M. (2010) Cortical speech processing unplugged: A timely subcortico-cortical framework. *Trends in Cognitive Sciences* 14(9):392–99. [SF]

Kotz, S. A., Schwartze, M. & Schmidt-Kassow, M. (2009) Non-motor basal ganglia functions: A review and proposal for a model of sensory predictability in auditory language perception. *Cortex* 45(8):982–90. doi: 10.1016/j.cortex.2009.02.010. [AZ]

Kovac, S., Deppe, M., Mohammadi, S., Schiffbauer, H., Schwindt, W., Möddel, G., Dogan, M. & Evers, S. (2009) Gelastic seizures: A case of lateral frontal lobe epilepsy and review of the literature. *Epilepsy and Behavior* 15:249–53. [aHA]

Krägeloh-Mann, I., Helber, A., Mader, I., Staudt, M., Wolff, M., Groenendaal, F. & DeVries, L. (2002) Bilateral lesions of thalamus and basal ganglia: Origin and outcome. *Developmental Medicine and Child Neurology* 44:477–84. [aHA]

Krause, J., Lalueza-Fox, C., Orlando, L., Enard, W., Green, R. E., Burbano, H. A., Hublin, J. J., Hänni, C., Fortea, J., De la Rasilla, M., Bertranpetit, J., Rosas, A. & Pääbo, S. (2007) The derived *FOXP2* variant of modern humans was shared with Neandertals. *Current Biology* 17:1–5. [aHA]

Kreiman, J. & Sidtis, D. (2011) *Foundations of voice studies: An interdisciplinary approach to voice production and perception*. Wiley-Blackwell. [aHA]

Kreitzer, A. C. & Malenka, R. C. (2008) Striatal plasticity and basal ganglia circuit function. *Neuron* 60:543–54. [aHA]

Kuhl, P. K. & Miller, J. D. (1975) Speech perception by chinchilla: Voiced voiceless distinction in alveolar plosive consonants. *Science* 190:69–72. [DLB, DHR]

Kuhl, P. K., Tsao, F. M. & Liu, H. M. (2003) Foreign-language experience in infancy: Effects of short-term exposure and social interaction on phonetic learning. *Proceedings of the National Academy of Sciences USA* 100(15):9096–101. [MHB]

Kunishio, K. & Haber, S. N. (1994) Primate cingulostriatal projection: Limbic striatal versus sensorimotor striatal input. *Journal of Comparative Neurology* 350:337–56. [aHA]

Kurniawan, I. T., Seymour, B., Talmi, D., Yoshida, W., Chater, N. & Dolan, R. J. (2010) Choosing to make an effort: The role of striatum in signaling physical effort of a chosen action. *Journal of Neurophysiology* 104(1):313–21. doi: 10.1152/jn.00027.2010. [AZ]

Kusmierek, P., Ortiz, M. & Rauschecker, J. P. (2012) Sound-identity processing in early areas of the auditory ventral stream in the macaque. *Journal of Neurophysiology* 107:1123–41. [DHR]

Kuypers, H. G. J. M. (1958a) Corticobulbar connection to the pons and lower brainstem in man. *Brain* 81:364–88. [aHA, PL, BM]

Kuypers, H. G. J. M. (1958b) Some projections from the peri-central cortex to the pons and lower brain stem in monkey and chimpanzee. *Journal of Comparative Neurology* 110:221–55. [aHA, BM]

Kwon, K., Oller, D. K. & Buder, E. H. (2007) Evidence of systematic repetition in infant vocalizations. Paper presented at the Annual Meeting of the American Speech-Language-Hearing Association, Boston, MA, November 2007. [DKO]

Ladefoged, P. (2005) *Vowels and consonants: An introduction to the sounds of languages*, 2nd edition. Blackwell. [aHA]

Laffin, J. J., Raca, G., Jackson, C. A., Strand, E. A., Jakielski, K. J. & Shriberg, L. D. (2012) Novel candidate genes and regions for childhood apraxia of speech identified by array comparative genomic hybridization. *Genetic Medicine* 14:928–36. [CMV]

Lai, C. S. L., Fisher, S. E., Hurst, J. A., Vargha-Khadem, F. & Monaco, A. P. (2001) A forkhead-domain gene is mutated in a severe speech and language disorder. *Nature* 413(6855):519–23. [KJA, SJ]

Lai, C. S., Gerrelli, D., Monaco, A. P., Fisher, S. E. & Copp, A. J. (2003) FOXP2 expression during brain development coincides with adult sites of pathology in a severe speech and language disorder. *Brain* 126(Pt 11):2455–62. doi: 10.1093/brain/awg247. [UH]

Lam, D. D., Zhou, L., Vegge, A., Xiu, P. Y., Christensen, B. T., Osundiji, M. A., Yueh, C. Y., Evans, M. L. & Heisler, L. K. (2009) Distribution and neurochemical characterization of neurons within the nucleus of the solitary tract responsive to serotonin agonist-induced hypophagia. *Behavioral Brain Research* 196:139–43. [CMV]

Lameira, A. R., Hardus, M. E., Kowalsky, B., de Vries, H., Spruijt, B. M., Sterck, E. H. M., Shumaker, R. W. & Wich, S. A. (2013a) Orangutan (*Pongo spp.*) whistling and implications for the emergence of an open-ended call repertoire: A replication and extension. *Journal of the Acoustical Society of America* 134:1–11. [ARL]

Lameira, A. R., Hardus, M. E., Nouwen, K. J. J. M., Topelberg, E., Delgado, R. A., Spruijt, B. M., Sterck, E. H. M., Knott, C. D. & Wich, S. A. (2013b) Population-specific use of the same tool-assisted alarm call between two wild orangutan populations (*Pongopygmaeus wurmbii*) indicates functional arbitrariness. *PLOS ONE* 8:e69749. [ARL]

Lameira, A., Hardus, M., & Wich, S. (2012) Orangutan instrumental gesture-calls: Reconciling acoustic and gestural speech evolution models. *Evolutionary Biology* 39:415–18. [ARL]

Lameira, A. R., Maddieson, I., & Zuberbühler, K. (2013c) Primate feedstock for the evolution of consonants. *Trends in Cognitive Science* 18:60–62. [ARL]

Lamendella, J. T. (1977) The limbic system in human communication. In: *Studies in neurolinguistics, vol. 3*, ed. H. Whitaker & H. A. Whitaker, pp. 157–222. Academic Press (*Perspectives in Neurolinguistics and Psycholinguistics Series*). [aHA]

Lane, H. (1961) Operant control of vocalizing in the chicken. *Journal of the Experimental Analysis of Behavior* 4:171–77. [ACC]

Lange, K. W., Robbins, T. W., Marsden, C. D., James, M., Owen, A. & Paul, G. M. (1992) L-dopa withdrawal in Parkinson's disease selectively impairs cognitive performance in tests sensitive to frontal lobe dysfunction. *Psychopharmacology* 107:394–404. [PL]

Langers, D. R. & Melcher, J. R. (2011) Hearing without listening: Functional connectivity reveals the engagement of multiple nonauditory networks during basic sound processing. *Brain Connect* 1(3):233–44. doi: 10.1089/brain.2011.0023. [UH]

Langus, A., Marchetto, E., Bion, R. A. H. & Nespor, M. (2012) Can prosody be used to discover hierarchical structure in continuous speech? *Journal of Memory and Language* 66(1):285–306. doi: 10.1016/j.jml.2011.09.004. [AR]

Laporte, M. N. C. & Zuberbühler, K. (2010) Vocal greeting behaviour in wild chimpanzee females. *Animal Behaviour* 80:467–73. [DKO]

Larson, C. R. & Kistler, M. K. (1984) Periaqueductal gray neuronal activity associated with laryngeal EMG and vocalization in the awake monkey. *Neuroscience Letters* 46:261–66. [rHA]

Larson, C. R., Sutton, D. & Lindeman, R. C. (1978) Cerebellar regulation of phonation in rhesus monkey (Macaca mulatta). *Experimental Brain Research* 33:1–18. [aHA]

Larson, C. R., Sutton, D., Taylor, E. M. & Lindeman, R. (1973) Sound spectral properties of conditioned vocalization in monkeys. *Phonetica* 27:100–10. [aHA]

Laukka, P., Åhs, F., Furmark, T. & Fredrikson, M. (2011) Neurofunctional correlates of expressed vocal affect in social phobia. *Cognitive, Affective, and Behavioral Neuroscience* 11(3):413–25. [SF]

Leavens, D. A., Hostetter, A. B., Wesley, M. J. & Hopkins, W. D. (2004) Tactical use of unimodal and bimodal communication by chimpanzees, *Pan troglodytes*. *Animal Behaviour* 67:467–76. [AM]

Leavens, D. A., Russell, J. L. & Hopkins, W. D. (2010) Multimodal communication by captive chimpanzees (*Pan troglodytes*). *Animal Cognition* 13:33–40. [AM]

Le Beau, J. (1954) Anterior cingulectomy in man. *Journal of Neurosurgery* 11:268–76. [aHA]

Lehericy, S., Ducros, M., Van de Moortele, P-F., Francois, C. L., Thivard, L., Poupon, C., Swindale, N., Ugurbil, K. & Kim, D. S. (2004) Diffusion tensor fiber tracking shows distinct corticostriatal circuits in humans. *Annals of Neurology* 55(4):522–29. [PL]

Lemasson, A. & Hausberger, M. (2004) Patterns of vocal sharing and social dynamics in a captive group of Campbell's monkeys (*Cercopithecus campbelli campbelli*). *Journal of Comparative Psychology* 118:347–59. [aHA]

Lemasson, A., Hausberger, M. & Zuberbühler, K. (2005) Socially meaningful vocal plasticity in adult Campbell's monkeys (*Cercopithecus campbelli*). *Journal of Comparative Psychology* 119:220–29. [aHA]

Lemasson, A., Ouattara, K., Petic, J. E. & Zuberbühler, K. (2011) Social learning of vocal structure in a nonhuman primate? *BMC Evolutionary Biology* 11:362. doi:10.1186/1471-2148-11-362. [ARL]

Lenti Boero, D. (1992) Alarm calls in marmots: Evidence for semantic communication. *Ethology, Ecology, Evolution* 4(2):125–38. [DLB]

Lenti Boero, D. (2009) Neurofunctional spectrographic analysis of the cry of brain injured asphyxiated infants: A physioacoustic and clinical study. In: *Models and analysis of vocal emissions for biomedical applications*, ed. C. Manfredi, pp. 3–6, Università di Firenze-Firenze University Press. [DLB]

Lenti Boero, D. & Bottoni, L. (2006) From crying to words: Unique or multilevel selective pressures? *Behavioral and Brain Sciences* 29(3):292–93. [DLB]

Lenti Boero, D. & Bottoni, L. (2009) Contrasting early cry and early babbling: Results from a pilot study. In: *Book of Abstracts*, pp. 25–27. University of Torun. Available at: http://www.protolang.umk.pl/2009/public/files/book_of_abstracts.pdf [DLB]

Lenti Boero, D., Miraglia, S. Ortalda, F. Nuti, G. Bottoni, L. & Lenti, C. (2008) Biomusicological approach in infant cry listening. In: *Musical Development and Learning: Proceedings of the Second European Conference on Developmental Psychology of Music, Roehampton University, England, 10–12 September, 2008*, ed. A. Daubney, E. Longhi, A. Lamont & D. Hargreaves, pp. 162–65. GK Publishing, Hull. (ISBN: 978-0-9553329-1-3). [DLB]

Lenti Boero, D., Volpe, C., Marcello, A., Bianchi, C. & Lenti, C. (1998) Newborns crying in different situational contexts: Discrete or graded signals? *Perceptual and Motor Skills* 86:1123–40. [DLB]

Lester, B. M. & Boukydis, C. F. Z. (1992) No language but a cry. In: *Nonverbal vocal communication. Comparative and developmental approaches*, ed. H. Papoušek, U. Jurgens & M. Papoušek, pp. 145–73. Cambridge University Press. [DLB]

Levenson, R. W. (2003) Blood, sweat, and fears: The autonomic architecture of emotion. *Annals of the New York Academy of Sciences* 1000:348–66. [CM]

Levin, S. A. (1998) Ecosystems and the biosphere as complex adaptive systems. *Ecosystems* 1:431–36. [TAM]

Levitzky, S. & Cooper, R. (2000) Infant colic syndrome maternal fantasies of aggression and infanticide. *Clinical Pediatrics* 39(7):395–400. [DLB]

Lewin, R. & Foley, R. A. (2004) *Principles of human evolution*, 2nd edition. Blackwell. [aHA]

Lewis, F. M., Lapointe, L. L., Murdoch, B. E. & Chenery, H. J. (1998) Language impairment in Parkinson's disease. *Aphasiology* 12(3):193–206. doi: 10.1080/02687039808249446. [AR]

Lewis, J. (2009) As well as words: Congo Pygmy hunting, mimicry, and play. In: *The cradle of language*, ed. R. Botha & C. Knight, pp. 236–56. Oxford University Press. [aHA]

Lhommée, E., Klinger, H., Thobois, S., Schmitt, E., Ardouin, C., Bichon, A., Kistner, A., Fraix, V., Xie, J., Aya, K. M., Chabardès, S., Seigneuret, E., Benabid, A. L., Mertens, P., Polo, G., Carnicella, S., Quesada, J. L., Bosson, J. L., Broussolle, E. Pollak, P. & Krack, P. (2012) Subthalamic stimulation in Parkinson's disease: Restoring the balance of motivated behaviours. *Brain* 135:1463–77. [CMV]

Li, G., Wang, J., Rossiter, S. J., Jones, G. & Zhang, S. (2007) Accelerated FoxP2 evolution in echolocating bats. *PLoS ONE* 2(9):e900. doi: 10.1371/journal.pone.0000900. [KJA, SJ]

Liberman, A. M., Cooper, F. S., Shankweiler, D. P. & Studdert-Kennedy, M. (1967) Perception of the speech code. *Psychological Review* 74:431–61. [PL]

Lieberman, D. E. (2011) *The evolution of the human head*. Harvard University Press. [aHA]

Lieberman, P. (1968) Primate vocalizations and human linguistic ability. *Journal of the Acoustical Society of America* 44:1574–84. [aHA]

Lieberman, P. (2000) *Human language and our reptilian brain: The subcortical bases of speech, syntax, and thought*. Harvard University Press. [aHA, PL]

Lieberman, P. (2002) On the nature and evolution of the neural bases of human language. *Yearbook of Physical Anthropology* 45:36–62. [PL]

Lieberman, P. (2006a) Limits on tongue deformation: Diana monkey formants and the impossible vocal tract shapes proposed by Riede et al. (2005). *Journal of Human Evolution* 50:219–21. [aHA]

Lieberman, P. (2006b) *Toward an evolutionary biology of language*. Harvard University Press. [aHA, PL]

Lieberman, P. (2007) The evolution of human speech: Its anatomical and neural bases. *Current Anthropology* 48:39–66. [aHA]

Lieberman, P. (2009) FOXP2 and human cognition. *Cell* 137:800–802. [SJ, PL]

Lieberman, P. (2012) Vocal tract anatomy and the neural bases of talking. *Journal of Phonetics* 40:608–22. [PL]

Lieberman, P. (2013) *The unpredictable species: What makes humans unique*. Princeton University Press. [PL]

Lieberman, P., Friedman, J. & Feldman, L. S. (1990) Syntactic deficits in Parkinson's disease. *Journal of Nervous and Mental Disease* 178:360–65. [PL]

Lieberman, P., Harris, K. S. & Wolff, P. (1968) Newborn infant cry in relation to nonhuman primate vocalizations. *Journal of Acoustic Society of America A* 44:365. [DLB]

Lieberman, P., Harris, K. S., Wolff, P. & Russel, L. H. (1971) Newborn infant cry and nonhuman primate vocalization. *Journal of Speech and Hearing Research* 14:718–27. [DLB]

Lieberman, P., Kako, E. T., Friedman, J., Tajchman, G., Feldman, L. S. & Jiminez, E. B. (1992) Speech production, syntax comprehension, and cognitive deficits in Parkinson's disease. *Brain and Language* 43:169–89. [PL]

Lieberman, P., Kanki, B. G., Protopapas, A., Reed, E. & Youngs, J. W. (1994) Cognitive defects at altitude. *Nature* 372:325. [PL]

Lieberman, P., Klatt, D. H. & Wilson, W. H. (1969) Vocal tract limitations on the vowel repertoires of rhesus monkey and other nonhuman primates. *Science* 164:1185–87. [aHA]

Lieberman, P., Morey, A., Hochstadt, J., Larson, M. & Mather, S. (2005) Mount Everest: A space-analog for speech monitoring of cognitive deficits and stress. *Aviation, Space and Environmental Medicine* 76:198–207. [PL]

Liégeois, F., Baldeweg, T., Connelly, A., Gadian, D. G., Mishkin, M. & Vargha-Khadem, F. (2003) Language fMRI abnormalities associated with FOXP2 gene mutation. *Nature Neuroscience* 6:1230–37. [aHA, KJA]

Locke, J. L. (1993) *The child's path to spoken language*, First edition. Harvard University Press. [aHA, DKO]

Locke, J. L. (2006) Parental selection of vocal behavior. Crying, cooing, babbling, and the evolution of language. *Human Nature* 17(2):155–68. [DLB]

Locke, J. L. & Bogin, B. (2006) Language and life history: A new perspective on the development and evolution of human language. *Behavioral and Brain Sciences* 29(3):301–11. [SJ, DLB, SJ]

Logemann, J. A., Fisher, H. B., Boshes, B. & Blonsky, E. R. (1978) Frequency and cooccurrence of vocal tract dysfunctions in the speech of a large sample of Parkinson patients. *Journal of Speech and Hearing Disorders* 43:47–57. [aHA]

Lombard, E. (1911) Le signe de l'elevation de la voix [The sign of the elevation of the voice]. *Annales des Maladies de l'Oreille et du Larynx* 37:101–19. [DJW]

Losin, E., Russell, J. L., Freeman, H., Meguerditchian, A. & Hopkins, W. D. (2008) Left hemisphere specialization for oro-facial movements of learned vocal signals by captive chimpanzees. *PLOS ONE* 3(6):e2529. [AM]

Loucks, T. M. J., Poletto, C. J., Simonyan, K., Reynolds, C. L. & Ludlow, C. L. (2007) Human brain activation during phonation and exhalation: Common volitional control for two upper airway functions. *NeuroImage* 36:131–43. [aHA]

Lu, Y. & Cooke, M. (2009) The contribution of changes in F0 and spectral tilt to increased intelligibility of speech produced in noise. *Speech Communication* 51 (12):1253–62. [DJW]

Lund, J. P. & Kolta, A. (2006) Brainstem circuits that control mastication: Do they have anything to say during speech? *Journal of Communication Disorders* 39:381–90. [aHA, DYT]

Lundy, B. L. (2013) Paternal and maternal mind-mindedness and preschoolers' theory of mind: The mediating role of interactional attunement. *Social Development* 22:58–74. [KJA]

Lynch, V. J. (2009) Use with caution: Developmental systems divergence and potential pitfalls of animal models. *Yale Journal of Biology and Medicine* 82:53–66. [KJA]

Macaulay, V. (2005) Single, rapid coastal settlement of Asia revealed by analysis of complete mitochondrial genomes. *Science* 308:1034–36. [SJ]

MacDermot, K. D., Bonora, E., Sykes, N., Coupe, A. M., Lai, C. S. L., Vernes, S. C., Vargha-Khadem, F., McKenzie, F., Smith, R. L., Monaco, A. P. & Fisher, S. E. (2005) Identification of FOXP2 truncation as a novel cause of developmental speech and language deficits. *American Journal of Human Genetics* 76:1074–80. [aHA]

MacNeilage, P. F. (1998) The frame/content theory of evolution of speech production. *Behavioral and Brain Sciences* 21(4):499–511. [arHA, DYT]

MacNeilage, P. F. (2008) *The origin of speech*. Oxford University Press. [arHA]

MacNeilage, P. F. & Davis, B. L. (2001) Motor mechanisms in speech ontogeny: Phylogenetic, neurobiological and linguistic implications. *Current Opinion in Neurobiology* 11:696–700. [rHA]

Madden, G. J., ed. (2012) *APA handbook of behavior analysis*. American Psychological Association. [ACC]

Maddieson, I. (1984) *Patterns of sounds*. Cambridge University Press. [ARL]

Mallet, N., Ballion, B., Le Moine, C. & Gonon, F. (2006) Cortical inputs and GABA interneurons imbalance projection neurons in the striatum of parkinsonian rats. *Journal of Neuroscience* 26:3875–84. [aHA]

Malloch, S. & Trevarthen, C., eds. (2009) *Communicative musicality: Exploring the basis of human companionship*. Oxford University Press. [aHA]

Mampe, B., Friederici, A. D., Christophe, A. & Wermke, K. (2009) Newborns' cry melody is shaped by their native language. *Current Biology* 19:1994–97. [DLB]

Männel, C., Schipke, C. S. & Friederici, A. D. (2013) The role of pause as a prosodic boundary marker: Language ERP studies in German 3- and 6-year-olds. *Developmental Cognitive Neuroscience* 5:86–94. doi: 10.1016/j.dcn.2013.01.003. [AR]

Manser, M. B., Seyfarth, R. M. & Cheney, D. L. (2002) Suricate alarm calls signal predator class and urgency. *Trends in Cognitive Sciences* 6:55–57. [aHA]

Manson, J. E., Bryant, G. A., Gervais, M. & Kline, M. (2013) Convergence of speech rate in conversation predicts cooperation. *Evolution and Human Behavior* 34 (6):419–26. [GAB]

Mao, C. C., Coull, B. M., Golper, L. A. & Rau, M. T. (1989) Anterior operculum syndrome. *Neurology* 39:1169–72. [aHA]

Maricic, T., Günther, V., Georgiev, O., Gehre, S., Curlin, M., Schreiweis, C., Naumann, R., Burbano, H. A., Meyer, M., Lalueza-Fox, C., de la Rasilla, M., Rosas, A., Gajovic, S., Kelso, J., Enard, W., Schaffner, W. & Pääbo, S. (2013) A recent evolutionary change affects a regulatory element in the human FOXP2 gene. *Molecular Biology and Evolution* 30(4):844–52. doi: 10.1093/molbev/mss271. [aHA, PL]

Marques, J. F., Canessa, N. & Cappa, S. (2009) Neural differences in the processing of true and false sentences: Insights into the nature of "truth" in language comprehension. *Cortex* 45(6):759–68. [KBC]

Marschik, P. B., Kaufmann, W. E., Sigafoos, J., Wolin, T., Zhang, D., Bartl-Pokorny, K. D., Pini, G., Zappella, M., Tager-Flusberg, H., Einspieler, C. & Johnston, M. V. (2013) Changing the perspective on early development of Rett syndrome. *Research in Developmental Disabilities* 34:1236–39. [PBM]

Marschik, P. B., Pini, G., Bartl-Pokorny, K. D., Duckworth, M., Gugatschka, M., Vollmann, R., Zappella, M. & Einspieler, C. (2012) Early speech-language development in females with Rett syndrome: Focusing on the preserved speech variant. *Developmental Medicine and Child Neurology* 54:451–56. [PBM]

Marsden, C. D. (1982) The mysterious motor function of the basal ganglia: The Robert Wartenberg Lecture. *Neurology* 32:514–39. [aHA]

Marsden, C. D. & Obeso, J. A. (1994) The functions of the basal ganglia and the paradox of sterotaxic surgery in Parkinson's disease. *Brain* 117:877–97. [PL]

Marseglia, G., Scordo, M. R., Pescucci, C., Nannetti, G., Biagini, E., Scandurra, V., Gerundino, F., Magi, A., Benelli, M. & Torricelli, F. (2012) 372 kb Microdeletion in 18q12.3 causing SETBP1 haploinsufficiency associated with mild mental retardation and expressive speech impairment. *European Journal of Medical Genetics* 55:216–21. [KJA]

Marshall, A. J., Wrangham, R. W. & Arcadi, A. C. (1999) Does learning affect the structure of vocalizations in chimpanzees? *Animal Behaviour* 58:825–30. [aAH, ARL]

Masataka, N. (2008a) The gestural theory of and the vocal theory of language origins are not incompatible with one another. In: *The origins of language: Unraveling evolutionary forces*, ed. N. Masataka, pp. 1–10. Springer. [aHA]

Masataka, N. (2008b) Implication of the human musical faculty for evolution of language. In: *The origins of language: Unraveling evolutionary forces*, ed. N. Masataka, pp. 133–51. Springer. [aHA]

Masdeu, J. C., Schoene, W. C. & Funkenstein, H. (1978) Aphasia following infarction of the left supplementary motor area: A clinicopathologic study. *Neurology* 28:1220–23. [aHA]

Massart, R., Mongeau, R. & Lanfumey, L. (2012) Beyond the monoaminergic hypothesis: Neuroplasticity and epigenetic changes in a transgenic mouse model of depression. *Philosophical Transactions of the Royal Society B* 367:2485–94. [rHA]

Matell, M. S. & Meck, W. H. (2004) Cortico-striatal circuits and interval timing: Coincidence detection of oscillatory processes. *Brain Research: Cognitive Brain Research* 21(2):139–70. doi: 10.1016/j.cogbrainres.2004.06.012. [CIP]

Mauboussin, M. J. (2002) Revisiting market efficiency: The stock market as a complex adaptive system. *Journal of Applied Corporate Finance* 14:47–55. [TAM]

Mazzoni, P., Hristova, A. & Krakauer, J. W. (2007) Why don't we move faster? Parkinson's disease, movement vigor, and implicit motivation. *The Journal of Neuroscience* 27(27):7105–16. doi: 10.1523/JNEUROSCI.0264-07.2007. [AZ]

McGettigan, C., Eisner, F., Agnew, Z. K., Manly, T., Wisbey, D. & Scott, S. K. (2013) T'ain't what you say, it's the way that you say it – Left insula and inferior frontal cortex work in interaction with superior temporal regions to control the performance of vocal impersonations. *Journal of Cognitive Neuroscience* 25 (11):1875–86. doi:10.1162/jocn_a_00427. [CM]

McHaffie, J. G., Stanford, T. R., Stein, B. E., Coizet, V. & Redgrave, P. (2005) Subcortical loops through the basal ganglia. *Trends in Neurosciences* 28:401–407. [aHA]

McNaughton, R. & Papert, S. (1971) *Counter-free Automata*. MIT Press. [KBC]

Meguerditchian, A., Cochet, H. & Vauclair, J. (2011) From gesture to language: ontogenetic and phylogenetic perspectives on gestural communication and its cerebral lateralization. In: *Primate communication and human language: Vocalisation, gestures, imitation and deixis in humans and non-humans*, ed. A. Vilain, J. L. Schwartz, C. Abry & J. Vauclair, pp. 89–118. [*Advances in Interaction Studies, vol. 1*]. John Benjamins. [AM]

Meguerditchian, A., Gardner, M. J., Schapiro, S. J. & Hopkins, W. D. (2012) The sound of one hand clapping: Handedness and perisylvian neural correlates of a communicative gesture in chimpanzees. *Proceedings of the Royal Society B: Biological Sciences* 279:1959–66. [AM]

Mehler, J., Juskzyc, P., Lamberz, G., Halsted, N., Bertoncini, J. & Amiel-Tison, C. (1988) A precursor of language acquisition in young infants. *Cognition* 29:143–78. [DLB]

Mendez, J. C., Prado, L., Mendoza, G. & Merchant, H. (2011) Temporal and spatial categorization in human and non-human primates. *Frontiers in Integrative Neuroscience* 5(50):1–10. doi: 10.3389/fnint.2011.00050. [HH]

Menuet, C., Cazals, Y., Gestreau, C., Borghgraef, P., Gielis, L., Dutschmann, M., Van Leuven, F. & Hilaire, G. (2011) Age-related impairment of ultrasonic vocalization in Tau.P301 L mice: Possible implication for progressive language disorders. *PLOS ONE* 6:e25770. [PBM]

Merchant, H., Battaglia-Mayer, A. & Georgopoulos, A. P. (2003) Interception of real and apparent circularly moving targets: Psychophysics in human subjects and monkeys. *Experimental Brain Research* 152:106–12. [HH]

Merchant, H. & Honing, H. (2014) Are non-human primates capable of rhythmic entrainment? Evidence for the gradual audiomotor evolution hypothesis. *Frontiers in Auditory Cognitive Neuroscience* 7:274 doi: 10.3389/fnins.2013.00274. [HH]

Merker, B. (2000) Synchronous chorusing and the origins of music. *Musicae Scientiae* 3(Suppl. 1):59–73. [AR]

Merker, B. (2009) Returning language to culture by way of biology. Commentary on Evans & Levinson (2009). *Behavioral and Brain Sciences* 32:460–61. [BM]

Merker, B. (2012) The vocal learning constellation: Imitation, ritual culture, encephalization. In: *Music, language and human evolution*, ed. N. Bannan, pp. 215–60. Oxford University Press. [BM, GP]

Merker, B., Madison, G. & Eckerdal, P. (2009) On the role and origin of isochrony in human rhythmic entrainment. *Cortex* 45(1):4–17. [AR]

Merker, B. & Okanoya, K. (2007) The natural history of human language: Bridging the gaps without magic. In: *Emergence of communication and language*, ed. C. Lyon, L. Nehaniv & A. Cangelosi, pp. 403–20. Springer-Verlag. [BM]

Mestres-Missé, A., Turner, R. & Friederici, A. D. (2012) An anterior–posterior gradient of cognitive control within the dorsomedial striatum. *NeuroImage* 62 (1):41–47. [KBC]

Metzner, W. & Schuller, G. (2010) Vocal control in echolocating bats. In: *Handbook of mammalian vocalization: An integrative neuroscience approach*, ed. S. M. Brudzynski, Ch. 9.4, pp. 403–15. Academic Press/Elsevier. [KJA]

Miller, C. T., Beck, K., Meade, B. & Wang, X. (2009a) Antiphonal call timing in marmosets is behaviorally significant: Interactive playback experiments. *Journal of Comparative Physiology, A: Neuroethology, Sensory, Neural, and Behavioral Physiology* 195:783–89. [aHA]

Miller, C. T., Eliades, S. J. & Wang, X. (2009b) Motor planning for vocal production in common marmosets. *Animal Behaviour* 78:1195–203. [aHA]

Miller, J. D. (1977) Perception of speech sounds in animals: evidence for speech processing by mammalian auditory mechanisms. In: *Dahlem Workshop on Recognition of Complex Acoustic Signals. Life Sciences Report, vol. 5,* ed. T. Bullock, pp. 49–58. Abakon. [DLB]

Miller, J. E., Spiteri, E., Condro, M. C., Dosumu-Johnson, R. T., Geschwind, D. H. & White, S. A. (2008) Birdsong decreases protein levels of FoxP2, a molecule required for human speech. *Journal of Neurophysiology* 100:2015–25. [KJA]

Miller, P. (2010) *The smart swarm.* Avery Books. [KJA]

Miller, R. L., Stein, M. K. & Loewy, A. D. (2011) Serotonergic inputs to FoxP2 neurons of the pre-locus coeruleus and parabrachial nuclei that project to the ventral tegmental area. *Neuroscience* 193:229–40. [CMV]

Milo, R. G. & Quiatt, D. (1994) Language in the middle and late stone ages: Glottogenesis in anatomically modern homo sapiens. In: *Hominid culture in primate perspective,* ed. D. Quiatt & J. Itani, pp. 321–39. University Press of Colorado. [aHA]

Mimura, M., Oeda, R. & Kawamura, M. (2006) Impaired decision-making in Parkinson's disease. *Parkinsonism and Related Disorders* 12(3):169–75. doi: 10.1016/j.parkreldis.2005.12.003. [AZ]

Mitani, J. C. & Brandt, K. L. (1994) Social factors influence acoustic variability in the long-distance calls of male chimpanzees. *Ethology* 96:233–52. [aHA]

Mitani, J. C. & Gros-Louis, J. (1998) Chorusing and call convergence in chimpanzees: Tests of three hypotheses. *Behaviour* 135:1041–64. [aHA]

Mithen, S. J. (2006) *The singing Neanderthals: The origins of music, language, mind and body*. Harvard University Press. (Original work published in 2005). [arHA, SJ]

Mogenson, G. J., Jones, D. L. & Yim, C. Y. (1980) From motivation to action: Functional interface between the limbic system and the motor system. *Progress in Neurobiology* 14:69–97. [aHA]

Monchi, O., Petrides, M., Petre, V., Worsley, K. & Dagher, A. (2001) Wisconsin card sorting revisited: Distinct neural circuits participating in different stages of the task identified by event-related functional magnetic resonance imaging. *Journal of Neuroscience* 21(19):7733–41. [PL]

Moore, C. A. (2004) Physiologic development of speech production. In: *Speech motor control in normal and disordered speech*, ed. B. Maassen, R. D. Kent, H. F. M. Peters, P. H. H. M. van Lieshout & W. Hulstijn, pp. 191–209. Oxford University Press. [aHA]

Moore, R. K. (2007) PRESENCE: A human-inspired architecture for speech-based human-machine interaction. *IEEE Transactions on Computers* 56:1176–88. [GP]

Morecraft, R. J. & van Hoesen, G. W. (1992) Cingulate input to the primary and supplementary motor cortices in the Rhesus monkey: Evidence for somatotopy in areas 24c and 23c. *Journal of Comparative Neurology* 322:471–89. [aHA]

Morecraft, R. J., Louie, J. L., Herrick, J. L. & Stilwell-Morecraft, K. S. (2001) Cortical innervation of the facial nucleus in the non-human primate: A new interpretation of the effects of stroke and related subtotal brain trauma on the muscles of facial expression. *Brain* 124:176–208. [aHA]

Morley, I. (2012) Hominin physiological evolution and the emergence of musical capacities. In: *Music, language, and human evolution*, ed. N. Bannan, pp. 109–41. Oxford University Press. [aHA]

Morrill, R. J., Paukner, A., Ferrari, P. F. & Ghazanfar, A. A. (2012) Monkey lipsmacking develops like the human speech rhythm. *Developmental Science* 15:557–68. [DYT]

Morse, P. A. & Snowdon, C. T. (1975) An investigation of categorical speech discrimination by rhesus monkeys. *Perception and Psychophysics* 17:9–16. [DLB]

Mukamel, Z., Konopka, G., Wexler, E., Osborn, G. E., Dong, H., Bergman, M. Y., Levitt, P. & Geschwind, D. H. (2011) Regulation of MET by FOXP2, genes implicated in higher cognitive dysfunction and autism risk. *Journal of Neuroscience* 31(32):11437–42. [KJA]

Müller, J., Wenning, G. K., Verny, M., McKee, A., Chaudhuri, K. R., Jellinger, K., Poewe, W. & Litvan, I. (2001) Progression of dysarthria and dysphagia in postmortem-confirmed Parkinsonian disorders. *Archives of Neurology* 58:259–64. [aHA]

Müller-Vahl, K. R., Kaufmann, J., Grosskreutz, J., Dengler, R., Emrich, H. M. & Peschel, T. (2009). Prefrontal and anterior cingulate cortex abnormalities in Tourette syndrome: Evidence from voxel-based morphometry and magnetization transfer imaging. *BMC Neuroscience* 10:47. Available at: www.biomedcentral.com/1471–2202/10/47 [aHA]

Munhall, K. & Löfqvist, A. (1992) Gestural aggregation in speech: Laryngeal gestures. *Journal of Phonetics* 20:111–26. [aHA]

Myers, R. E. (1976) Comparative neurology of vocalization and speech: Proof of a dichotomy. In: *Origins and evolution of language and speech*, ed. S. R. Harnad, H. D. Steklis & J. Lancaster, pp. 745–57. (*Annals of the New York Academy of Sciences, vol. 280*). New York Academy of Sciences. [aHA]

Nachtigall, P., Supin, A. Y., Pawloski, J. & Au, W. W. L. (2004) Temporary threshold shifts after noise exposure in the bottlenose dolphin (*Tursiops truncatus*) measured using evoked auditory potentials. *Marine Mammal Science* 20:673–87. [DJW]

Naeser, M. A., Alexander, M. P., Helm-Estabrooks, N., Levine, H. L., Laughlin, S. A. & Geschwind, N. (1982) Aphasia with predominantly subcortical lesion sites: Description of three capsular/putaminal aphasia syndromes. *Archives of Neurology* 39(1):2–14. [PL]

Nakano, K. (2000) Neural circuits and topographic organization of the basal ganglia and related regions. *Brain and Development* 22:S5–16. [aHA]

Nambu, A. (2008) Seven problems on the basal ganglia. *Current Opinion in Neurobiology* 18:595–604. [aHA]

Nambu, A. (2011) Somatotopic organization of the primate basal ganglia. *Frontiers in Neuroanatomy* 5:26. [aHA]

Nathani, S., Neal, A. R., Olds, H., Brill, J. & Oller, D. K. (2001) Canonical babbling and volubility in infants with moderate to severe hearing impairment. Paper presented at the International Child Phonology Conference, Boston, MA, April 2001. [DKO]

Natsopoulos, D., Grouios, G., Bostantzopoulou, S., Mentenopoulos, G., Katsarou, Z. & Logothetis, J. (1993) Algorithmic and heuristic strategies in comprehension of complement clauses by patients with Parkinson's disease. *Neuropsychologia* 31(9):951–64. [PL]

Nauta, W. J. H. & Gygax, P. A. (1954) Silver impregnation of degenerating axons in the central nervous system: A modified technic. *Stain Technology* 29:91–93. [PL]

Nery, M. F., González, D. J. & Opazo, J. C. (2013) How to make a dolphin: Molecular signature of positive selection in Cetacean Genome. *PLoS ONE* 8(6): e65491. doi: 10.1371/journal.pone.0065491. [KJA]

Neubert, F.-X., Mars, R. B., Thomas, A. G., Sallet, J. & Rushworth, M. F. S. (2014) Comparison of human ventral frontal cortex areas for cognitive control and language with areas in monkey frontal cortex. *Neuron* 81:1–14. doi:10.1016/j.neuron.2013.11.012. [BM]

Neul, J. L., Kaufmann, W. E., Glaze, D. G., Christodolou, J., Clarke, A. J., Bahi-Buisson, N., Leonard, H., Bailey, M. E., Schanen, N. C., Zappella, M., Renieri, A., Huppke, P., Percy, A. K. & RettSearch Consortium. (2010) Rett syndrome: Revised diagnostic criteria and nomenclature. *Annals of Neurology* 68:944–50. [PBM]

Newbury, D. F., Fisher, S. E. & Monaco, A. P. (2010) Recent advances in the genetics of language impairment. *Genome Medicine* 2:6. [KJA]

Newbury, D. F., Paracchini, S., Scerri, T. S., Winchester, L., Addis, L., Richardson, A. J., Walter, J., Stein, J. F., Talcott, J. B. & Monaco, A. P. (2011) Investigation of dyslexia and SLI risk variants in reading- and language-impaired subjects. *Behavioral Genetics* 41:90–104. [KJA]

Newman, J. D. (2003) Vocal communication and the triune brain. *Physiology and Behavior* 79:495–502. [aHA]

Newport, E. L. & Meier, R. P. (1985) The acquisition of American Sign Language. In: *The crosslinguistic study of language acquisition*, ed. D. Slobin, pp. 881–38. Erlbaum. [BF]

Nielsen, M. A. & Chuang, I. L. (2000) *Quantum computation and quantum information*. Cambridge University Press. [KBC]

Nieuwenhuys, R., Voogd, J. & van Huijzen, C. (2008) *The human central nervous system*, 4th edition. Springer. [aHA]

Niv, Y., Joel, D. & Dayan, P. (2006) A normative perspective on motivation. *Trends in Cognitive Sciences* 10(8):375–81. doi: 10.1016/j.tics.2006.06.010. [AZ]

Noonan, J. P. (2010) Neanderthal genomics and the evolution of modern humans. *Genome Research* 20:547–53. [aHA]

Nottebohm, F. (1976) Discussion paper. Vocal tract and brain: A search for evolutionary bottlenecks. In: *Origins and evolution of language and speech*, ed. S. R. Harnad, H. D. Steklis & J. Lancaster, pp. 643–49. *Annals of the New York Academy of Sciences, vol. 280*. New York Academy of Sciences. [BM]

Nuckolls, A. L., Worley, C., Leto, C., Zhang, H., Morris, J. K. & Stanford, J. A. (2012) Tongue force and tongue motility are differently affected by unilateral vs. bilateral nigrostriatal dopamine depletion in rats. *Behavioral Brain Research* 234:343–48. [CMV]

Nudel, R. & Newbury, D. F. (2013) FOXP2 advanced review. *WIREs Cognitive Science* 4(5):547–60. doi: 10.1002/wcs.1247. [KJA]

Obeso, J. A., Jahanshahi, M., Alvarez, L., Macias, R., Pedroso, I., Wilkinson, L., Pavon, N., Day, B., Pinto, S., Rodríguez-Oroz, M. C., Tejeiro, J., Artieda, J., Talelli, P., Swayne, O., Rodríguez, R., Bhatia, K., Rodriguez-Dias, M., Lopez,

G., Guridi, J. & Rothwell, J. C. (2009) What can man do without basal ganglia motor output? The effect of combined unilateral subthalamotomy and pallidotomy in a patient with Parkinson's disease. *Experimental Neurology* 220 (2):283–92. doi: 10.1016/j.expneurol.2009.08.030. [AZ]

Okanoya, K., Hihara, S., Tokimoto, N., Tobari, Y. & Iriki, A. (2007) Complex vocal behaviour and cortical-medullar projection. *Lecture Notes in Computer Science* 3609:362–67. [BM]

Okanoya, K. & Merker, B. (2007) Neural substrates for string-context mutual segmentation: A path to human language. In: *Emergence of communication and language*, ed. C. Lyon, L. Nehaniv & A. Cangelosi, pp. 421–34. Springer-Verlag. [BM]

Olivier, E., Davare, M., Andres, M. & Fadiga, L. (2007) Precision grasping in humans: From motor control to cognition. *Current Opinion in Neurobiology* 17 (6):644–48. doi: 10.1016/j.conb.2008.01.008. [AZ]

Oller, D. K. (1980) The emergence of the sounds of speech in infancy. In: *Child phonology, vol. 1: Production*, ed. G. Yeni-Komshian, J. Kavanagh & C. Ferguson, pp. 93–112. Academic Press. [DKO]

Oller, D. K. (2000) *The emergence of the speech capacity*. Erlbaum. [DLB, DKO]

Oller, D. K., Buder, E. H., Ramsdell, H. L., Chorna, L., Warlaumont, A. S. & Bakeman, R. (2013) Functional flexibility of infant vocalization and the emergence of language. In: *Proceedings of the National Academy of Sciences USA* 110(16):6318–23. doi: 10.1073/pnas.1300337110. [KJA, DKO]

Öngür, D. & Price, J. L. (2000) The organization of networks within the orbital and medial prefrontal cortex of rats, monkeys and humans. *Cerebral Cortex* 10:206–19. [aHA]

Ouattara, K., Lemasson, A. & Zuberbühler, K. (2009) Campbell's monkeys concatenate vocalizations into context-specific call sequences. *Proceedings of the National Academy of Sciences USA* 106(51): 22026–31. [aHA, KBC, ARL]

Owren, M. J., Amoss, R. T. & Rendall, D. (2011) Two organizing principles of vocal production: Implications for nonhuman and human primates. *American Journal of Primatology* 73(6):530–44. [aHA, DJW]

Owren, M. J., Dieter, J. A., Seyfarth, R. M. & Cheney, D. L. (1992) "Food" calls produced by adult female rhesus (*Macaca mulatta*) and Japanese (*M. fuscata*) macaques, their normally-raised offspring, and offspring cross-fostered between species. *Behaviour* 120:218–31. [aHA]

Owren, M. J., Dieter, J. A., Seyfarth, R. M. & Cheney, D. L. (1993) Vocalizations of rhesus (*Macaca mulatta*) and Japanese (*M. fuscata*) macaques cross-fostered between species show evidence of only limited modification. *Developmental Psychobiology* 26:389–406. [aHA, DHR]

Packard, M. G. & Knowlton, B. J. (2002) Learning and memory functions of the basal ganglia. *Annual Review of Neuroscience* 25:563–93. [aHA]

Panksepp, J. (1998) *Affective neuroscience: The foundations of human and animal emotions*. Oxford University Press. [aHA]

Panksepp, J. (2010) Emotional causes and consequences of social-affective vocalization. In: *Handbook of mammalian vocalization: An integrative neuroscience approach*, ed. S. M. Brudzynski, pp. 201–208. Elsevier. [aHA]

Papoušek, M. (2003) Intuitive parenting: A hidden source of musical stimulation in infancy. In: *Musical beginnings: Origins and development of musical competence*, ed. I. Deliège & J. Sloboda, pp. 88–112. Oxford University Press. [aHA]

Papoušek, M. & Papoušek, H. (1981) Musical elements in the infant's vocalization: Their significance for communication, cognition, and creativity. In: *Advances in infancy research, vol. 1*, ed. L. P. Lipsitt, pp. 163–224. Ablex. [DLB]

Parent, A. & Hazrati, L. N. (1995) Functional anatomy of the basal ganglia: I. The cortico-basal ganglia-thalamo-cortical loop. *Brain Research Brain Research Reviews* 20:91–127. [aHA]

Parks, S. E., Johnson, M., Nowacek, D. P. & Tyack, P. L. (2011) Individual right whales call louder in increased environmental noise. *Biology Letters* 7:33–35. [DJW]

Parr, L., Waller, B. M. & Heintz, M. (2008) Facial expression categorization by chimpanzees using standardized stimuli. *Emotion* 8(2):216–31. doi: 10.1037/1528-3542.8.2.216. [DKO]

Passingham, R. (2008) *What is special about the human brain*? Oxford University Press. [aHA]

Patel, A. D. (2006) Musical rhythm, linguistic rhythm, and human evolution. *Music Perception: An Interdisciplinary Journal* 24(1):99–104. [AR]

Patel, A. D. (2008) *Music, language, and the brain*. Oxford University Press. [HH]

Patel, A. D., Iversen, J. R., Bregman, M. R. & Schulz, I. (2009a) Experimental evidence for synchronization to a musical beat in a nonhuman animal. *Current Biology* 19(10):827–30. [AR]

Patel, A. D., Iversen, J. R., Bregman, M. R. & Schulz, I. (2009b) Studying synchronization to a musical beat in nonhuman animals. *Annals of the New York Academy of Sciences* 1169:459–69. doi: 10.1111/j.1749-6632.2009.04581.x. [HH]

Patel, S., Scherer, K. R., Bjorkner, E. & Sundberg, J. (2011) Mapping emotions into acoustic space: The role of voice production. *Biological Psychology* 87 (1):93–98. [SF]

Patterson, F. G. & Linden, E. (1981) *The education of Koko*. Holt, Rinehart, & Winston. [BdB]

Paul, R., Fuerst, Y., Ramsay, G., Chawarska, K. & Klin, A. (2011) Out of the mouths of babes: Vocal production in infant siblings of children with ASD. *Journal of Child Psychology and Psychiatry* 52:588–98. [PBM]

Paulmann, S., Ott, D. V. & Kotz, S. A. (2011) Emotional speech perception unfolding in time: The role of the basal ganglia. *PLOS ONE* 6(3):e17694. doi: 10.1371/journal.pone.0017694. [UH]

Paulmann, S., Pell, M. D. & Kotz, S. A. (2008) Functional contributions of the basal ganglia to emotional prosody: Evidence from ERPs. *Brain Research* 1217:171–78. doi: 10.1016/j.brainres.2008.04.032. [UH]

Paus, T. (2001) Primate anterior cingulate cortex: Where motor control, drive and cognition interface. *Nature Reviews Neuroscience* 2:417–24. [aHA]

Paus, T., Tomaiuolo, F., Otaky, N., MacDonald, D., Petrides, M., Atlas, J., Morris, R. & Evans, A. C. (1996) Human cingulate and paracingulate sulci: Pattern, variability, asymmetry, and probabilistic map. *Cerebral Cortex* 6:207–14. [aHA]

Peake, T. M. & McGregor, P. K. (2004) Information and aggression in fishes. *Learning and Behavior* 32(1):114–21. [KBC]

Peelle, J. E. & Davis, M. H. (2012) Neural oscillations carry speech rhythm through to comprehension. *Frontiers in Psychology* 3:320. [rHA]

Pell, M. D. & Monetta, L. (2008) How Parkinson's disease affects non-verbal communication and language processing. *Language and Linguistics Compass* 2 (5):739–59. doi: 10.1111/j.1749-818X.2008.00074.x. [AZ]

Penn, D. C., Holyoak, K. J. & Povinelli, D. J. (2008) Darwin's mistake: Explaining the discontinuity between human and nonhuman minds. *Behavioral and Brain Sciences* 31(2):109–78. [BF]

Pepperberg, I. M. (2007) Grey parrots do not always "parrot": The roles of imitation and phonological awareness in the creation of new labels from existing vocalizations. *Language Science* 29:1–13. [DLB]

Perlman, M., Patterson, F. G. & Cohn, R. H. (2011) The incorporation of learned breathing-related behavior into the multimodal gestures and rituals of an enculturated gorilla. Paper presented at the New College Oxford Conference on Embodied Language, Oxford, United Kingdom, September 2011. [BdB]

Perlman, M., Patterson, F. G. & Cohn, R. H. (2012) The human-fostered gorilla Koko shows breath control in play with wind instruments. *Biolinguistics* 6:433–44. [GAB]

Péron, J., Frühholz, S., Verin, M. & Grandjean, D. (2013) Subthalamic nucleus: A key structure for emotional component synchronization in humans. *Neuroscience and Biobehavioral Reviews* 37(3):358–73. [SF]

Perrodin, C., Kayser, C., Logothetis, N. K. & Petkov, C. I. (2011) Voice cells in the primate temporal lobe. *Current Biology* 21:1408–15. [DHR]

Perszyk, D. R. & Waxman, S. R. (2013) The role of experience in linking sounds and meaning in language acquisition. Paper Presented at the 38th Annual Boston University Conference on Language Development, Boston, MA, November 3, 2013. [BF]

Petitto, L. A. & Marentette, P. F. (1991) Babbling in the manual mode: Evidence for the ontogeny of language. *Science* 251(5000):1493–96. [BF]

Petkov, C. I. & Jarvis, E. D. (2012) Birds, primates, and spoken language origins: Behavioral phenotypes and neurobiological substrates. *Frontiers in Evolutionary Neuroscience* 4:12. doi: 10.3389/fnevo.2012.00012. [CIP]

Petkov, C. I., Kayser, C., Augath, M. & Logothetis, N. K. (2006) Functional imaging reveals numerous fields in the monkey auditory cortex. *PLOS Biology* 4:e215. [DHR]

Petkov, C. I. & Wilson, B. (2012) On the pursuit of the brain network for proto-syntactic learning in non-human primates: Conceptual issues and neurobiological hypotheses. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* 367(1598):2077–88. doi: 10.1098/rstb.2012.0073. [CIP]

Petrides, M., Cadoret, G. & Mackey, S. (2005) Orofacial somatomotor responses in the macaque monkey homologue of Broca's area. *Nature* 435:1235–38. [aHA]

Petrides, M. & Pandya, D. N. (2009) Distinct parietal and temporal pathways to the homologues of Broca's area in the monkey. *PLoS Biology* 7:e1000170. [aHA]

Pezzulo, G. (2012a) An Active Inference view of cognitive control. *Frontiers in Psychology* 3:478. doi: 10.3389/fpsyg.2012.00478. [GP]

Pezzulo, G. (2012b) The "Interaction Engine": A common pragmatic competence across linguistic and non-linguistic interactions. *IEEE Transactions on Autonomous Mental Development* 4:105–23. [GP]

Pezzulo, G. (2013) Studying mirror mechanisms within generative and predictive architectures for joint action. *Cortex* 49:2968–69. [GP]

Pezzulo, G., Donnarumma, F. & Dindo, H. (2013) Human sensorimotor communication: A theory of signaling in online social interactions. *PLOS ONE* 8: e79876. [GP]

Phillips-Silver, J., Aktipis, A. & Bryant, G. A. (2010) The ecology of entrainment: Foundations of coordinated rhythmic movement. *Music Perception* 28(1):3–14. [GAB]

Phister, P. W., Jr. (2010) Cyberspace: The ultimate complex adaptive system. *International C2 Journal* 4:1–30. [TAM]

Picard, N. & Strick, P. L. (1996) Motor areas of the medial wall: A review of their location and functional activation. *Cerebral Cortex* 6:342–53. [aHA]

Pichon, S. & Kell, C. A. (2013) Affective and sensorimotor components of emotional prosody generation. *The Journal of Neuroscience* 33(4):1640–50. [SF]

Pick, H. L., Jr., Siegel, G. M., Fox, P. W., Garber, S. R. & Kearney, J. K. (1989) Inhibiting the Lombard effect. *Journal of the Acoustical Society of America* 85:94. [DJW]

Pickering, M. J. & Garrod, S. (2013) An integrated theory of language production and comprehension. *Behavioral and Brain Sciences* 36(4):329–47. [GP]

Pickett, E. R., Kuniholm, E., Protopapas, A., Friedman, J. & Lieberman, P. (1998) Selective speech motor, syntax and cognitive deficits associated with bilateral damage to the putamen and the head of the caudate nucleus: A case study. *Neuropsychologia* 36:173–88. [PL]

Pierce, J. D., Jr. (1985) A review of attempts to condition operantly alloprimate vocalizations. *Primates* 26:202–13. [aHA]

Pinel, P., Fauchereau, F., Moreno, A., Barbot, A., Lathrop, M., Zelenika, D., Le Bihan, D., Poline, J.-B., Bourgeron, T. & Dehaene, S. (2012) Genetic variants of *FOXP2* and *KIAA0319/TTRAP/THEM2* locus are associated with altered brain activation in distinct language-related regions. *Journal of Neuroscience* 32 (3):817–25. [KJA]

Pinker, S. (2010) Colloquium paper: The cognitive niche: Coevolution of intelligence, sociality, and language. *Proceedings of the National Academy of Sciences USA* 107:8993–99. doi: 10.1073/pnas.0914630107. [TAM]

Pinker, S. & Bloom, P. (1990) Natural language and natural selection. *Behavioral and Brain Sciences* 13:707–84. [BF]

Pistorio, A. L., Vintch, B. & Wang, X. (2006) Acoustic analysis of vocal development in a New World primate, the common marmoset (*Callithrix jacchus*). *Journal of the Acoustical Society of America* 120:1655–70. [aHA, DHR]

Postuma, R. B. & Dagher, A. (2006) Basal ganglia functional connectivity based on a meta-analysis of 126 positron emission tomography and functional magnetic resonance imaging publications. *Cerebral Cortex* 16:1508–21. [aHA]

Potulska, A., Friedman, A., Królicki, L. & Spychala, A. (2003) Swallowing disorders in Parkinson's disease. *Parkinsonism Related Disorders* 9:349–453. [CMV]

Prather, J. F., Peters, S., Nowicki, S. & Mooney, R. (2008) Precise auditory–vocal mirroring in neurons for learned vocal communication. *Nature* 451:305–10. [GP]

Provine, R. R. (1993) Laughter punctuates speech: Linguistic, social and gender contexts of laughter. *Ethology* 95:291–98. [GAB]

Provine, R. R. (2000) *Laughter: A scientific investigation*. Viking/Penguin Press. [GAB, ACC]

Provine, R. R. (2012) *Curious behavior: Yawning, laughing, hiccupping, and beyond*. Belknap/Harvard University Press. [ACC]

Provine, R. R. & Emmorey, K. (2006) Laughter among deaf signers. *Journal of Deaf Studies and Deaf Education* 11(4):403–409. doi: 10.1093/deafed/enl008. [CM]

Radua, J., van den Heuvel, O. A., Surguladze, S. & Mataix-Cols, D. (2010) Meta-analytical comparison of voxel-based morphometry studies in obsessive-compulsive disorder vs other anxiety disorders. *Archives of General Psychiatry* 67:701–11. [aHA]

Rajan, R., Dubaj, V., Reser, D. H. & Rosa, M. G. (2013) Auditory cortex of the marmoset monkey – complex responses to tones and vocalizations under opiate anaesthesia in core and belt areas. *European Journal of Neuroscience* 37:924–41. [DHR]

Ramig, L. O., Fox, C. & Sapir, S. (2004) Parkinson's disease: Speech and voice disorders and their treatment with the Lee Silverman Voice Treatment. *Seminars in Speech and Language* 25:169–80. [aHA]

Ramig, L. O., Fox, C. & Sapir, S. (2007) Speech disorders in Parkinson's disease and the effects of pharmacological, surgical and speech treatment with emphasis on Lee Silverman Voice Treatment (LSVT®). In: *Parkinson's disease and related disorders, Part 1*, ed. W. C. Koller & E. Melamed, pp. 385–99. (*Handbook of Clinical Neurology, vol. 83*, 3rd series). Elsevier Press. [aHA]

Rao, S. M., Harrington, D. L., Haaland, K. Y., Bobholz, J. A., Cox, R. W. & Binder J. R. (1997) Distributed neural systems underlying the timing of movements. *Journal of Neuroscience* 17:5528–35. [HH]

Rappaport, R. A. (1999) *Ritual and religion in the making of humanity*. Cambridge University Press. [aHA]

Rappaport, R. A. (2000) *Pigs for the ancestors: Ritual in the ecology of a New Guinea people*, 2nd edition. Waveland Press. [aHA]

Rasa, O. A. E. (1986) Coordinated vigilance in dwarf mongoose family groups: The "watchman's song" hypothesis and the costs of guarding. *Ethology* 71(4):340–44. [DLB]

Raul, L. (2003) Serotonin2 receptors in the nucleus tractus solitarius: Characterization and role in the baroreceptor reflex arc. *Cell Molecular Neurobiology* 23:709–26. [CMV]

Rauschecker, J. P. (2012) Ventral and dorsal streams in the evolution of speech and language. *Frontiers in Evolutionary Neuroscience* 4:7. doi: 10.3389/fnevo.2012.00007. [TAM]

Rauschecker, J. P. & Scott, S. K. (2009) Maps and streams in the auditory cortex: Nonhuman primates illuminate human speech processing. *Nature Neuroscience* 12(6):718–24. [SF]

Ravignani, A., Olivera, V. M., Gingras, B., Hofer, R., Hernández, C. R., Sonnweber, R.-S. & Fitch, W. T. (2013) Primate drum kit: A system for studying acoustic pattern production by non-human primates using acceleration and strain sensors. *Sensors* 13(8):9790–820. [AR]

Reimers-Kipping, S., Hevers, W., Pääbo, S. & Enard W. (2011) Humanized Foxp2 specifically affects cortico-basal ganglia circuits. *Neuroscience* 175:75–84. doi: 10.1016/j.neuroscience.2010.11.042. [aHA, BM]

Reiner, A. (2010) Organization of corticostriatal projection neuron types. In: *Handbook of basal ganglia structure and function*, ed. H. Steiner & K. Y. Tseng, pp. 323–39. Elsevier. [aHA]

Rendall, D., Kollias, S., Ney, C. & Lloyd, P. (2005) Pitch (*F0*) and formant profiles of human vowels and vowel-like baboon grunts: The role of vocalizer body size and voice-acoustic allometry. *Journal of the Acoustical Society of America* 117:944–55. [aHA]

Repp, B. H. (2005) Sensorimotor synchronization: A review of the tapping literature. *Psychonomic Bulletin and Review* 12:969–92. [GAB]

Reser, D. H., Burman, K. J., Richardson, K. E., Spitzer, M. W. & Rosa, M. G. (2009) Connections of the marmoset rostrotemporal auditory area: Express pathways for analysis of affective content in hearing. *European Journal of Neuroscience* 30:578–92. [DHR]

Richman, B. (1976) Some vocal distinctive features used by gelada monkeys. *Journal of the Acoustical Society of America* 60:718–24. [GAB]

Riecker, A., Brendel, B., Ziegler, W., Erb, M. & Ackermann, H. (2008) The influence of syllable onset complexity and syllable frequency on speech motor control. *Brain and Language* 107(2):102–13. doi: 10.1016/j.bandl.2008.01.008. [UH]

Riecker, A., Kassubek, J., Gröschel, K., Grodd, W. & Ackermann, H. (2006) The cerebral control of speech tempo: Opposite relationship between speaking rate and BOLD signal changes at striatal and cerebellar structures. *NeuroImage* 29:46–53. [aHA]

Riecker, A., Wildgruber, D., Dogil, G., Grodd, W. & Ackermann, H. (2002) Hemispheric lateralization effects of rhythm implementation during syllable repetitions: An fMRI study. *NeuroImage* 16:169–76. [rHA]

Riede, T., Bronson, E., Hatzikirou, H. & Zuberbühler, K. (2005) Vocal production mechanisms in a non-human primate: Morphological data and a model. *Journal of Human Evolution* 48:85–96. [aHA]

Riede, T., Bronson, E., Hatzikirou, H. & Zuberbühler, K. (2006) Multiple discontinuities in nonhuman vocal tracts: A response to Lieberman (2006). *Journal of Human Evolution* 50:222–25. [aHA]

Riede, T., Tokuda, I. T., Munger, J. B. & Thomson, S. L. (2008) Mammalian laryngseal air sacs add variability to the vocal tract impedance: Physical and computational modeling. *Journal of the Acoustical Society of America* 124(1):634–47. [BdB]

Riede, T. & Zuberbühler, K. (2003a) Pulse register phonation in Diana monkey alarm calls. *Journal of the Acoustical Society of America* 113:2919–26. [aHA]

Riede, T. & Zuberbühler, K. (2003b) The relationship between acoustic structure and semantic information in Diana monkey alarm vocalization. *Journal of the Acoustical Society of America* 114:1132–42. [aHA]

Rightmire, G. P. (2004) Brain size and encephalization in early to mid-Pleistocene *Homo*. *American Journal of Physical Anthropology* 124:109–23. [aHA]

Rightmire, G. P. (2007) Later middle Pleistocene *Homo*. In: *Handbook of paleoanthropology, vol. 3: Phylogeny of hominids*, ed. W. Henke & I. Tattersall, pp. 1695–715. Springer. [aHA]

Risley, T. R. (1977) The development and maintenance of language: An operant model. In: *New developments in behavioral research*, ed. B. C. Etzel, J. M. LeBlanc & D. M. Baer, pp. 81–101. Erlbaum. [ACC]

Rizzolatti, G. & Arbib, M. A. (1998) Language within our grasp. *Trends in Neurosciences* 21:188–94. [DYT]

Robbins, T. W. (2010) From behavior to cognition: Functions of mesostriatal, mesolimbic, and mesocortical dopamine systems. In: *Dopamine handbook*, ed. L. L. Iversen, S. D. Iversen, S. B. Dunnett & A. Björklund, pp. 203–14. Oxford University Press. [aHA]

Robinson, B. W. (1967) Vocalization evoked from forebrain in *Macaca mulatta*. *Physiology and Behavior* 2:345–54. [aHA]

Rocca, F. & Lenti Boero, D. (2005) Sex differences in human infant cry: A comparative view. In: *Abstracts of the XXIX International Ethological Conference*, ed. R. Sàndor, p. 131. Késult a Codex Print Nyondàbam, Budapest. [DLB]

Roland, E. H., Poskitt, K., Rodriguez, E., Lupton, B. A. & Hill, A. (1998) Perinatal hypoxic-ischemic thalamic injury: Clinical features and neuroimagery. *Annals of Neurology* 44:161–66. [aHA]

Rosas, A., Martínez-Maza, C., Bastir, M., García-Tabernero, A., Lalueza-Fox, C., Huguet, R., Ortiz, J. E., Julià, R., Soler, V., de Torres, T., Martínez, E., Canaveras, J. C., Sánchez-Moral, S., Cuezva, S., Lario, J., Santamaría, D., de la Rasilla, M. & Fortea, J. (2006) Paleobiology and comparative morphology of a late Neandertal sample from El Sidrón, Asturias, Spain. *Proceedings of the National Academy of Sciences USA* 103:19266–71. [aHA]

Roseberry, S., Hirsh-Pasek, K. & Golinkoff, R. M. (2014) Skype me! Socially contingent interactions help toddlers learn language. *Child Development* 85 (3):956–70. [MHB]

Ross, E. D. & Mesulam, M.-M. (1979) Dominant language functions of the right hemisphere? *Archives of Neurology* 36:144–48. [aHA]

Ross, E. D. & Monnot, M. (2008) Neurology of affective prosody and its functional-anatomic organization in right hemisphere. *Brain and Language* 104:51–74. [aHA]

Ross, M. D., Owren, M. J. & Zimmermann, E. (2010) The evolution of laughter in great apes and humans. *Communicative and Integrative Biology* 3:191–94. [DHR]

Rothermich, K., Schmidt-Kassow, M. & Kotz, S. A. (2012) Rhythm's gonna get you: Regular meter facilitates semantic sentence processing. *Neuropsychologia* 50:232–44. [rHA]

Roush, R. S. & Snowdon, C. T. (1994) Ontogeny of food-associated calls in cotton-top tamarins. *Animal Behaviour* 47:263–73. [aHA]

Roush, R. S. & Snowdon, C. T. (1999) The effects of social status on food-associated calling behaviour in captive cotton-top tamarins. *Animal Behaviour* 58:1299–305. [aHA]

Rowe, M. & Goldin-Meadow, S. (2009) Early gesture selectively predicts later language learning. *Developmental Science* 12(1):182–87. [KJA]

Roy, S., Miller, C. T., Gottsch, D. & Wang, X. (2011) Vocal control by the common marmoset in the presence of interfering noise. *Journal of Experimental Biology* 214:3619–29. [aHA]

Rubens, A. B. (1975) Aphasia with infarction in the territory of the anterior cerebral artery. *Cortex* 11:239–50. [aHA]

Ruch, W. & Ekman, P. (2001) The expressive pattern of laughter. In: *Emotion, qualia, and consciousness*, ed. A. W. Kaszniak, pp. 426–43. World Scientific. [CM]

Rukstalis, M., Fite, J. E. & French, J. A. (2003) Social change affects vocal structure in a callitrichid primate (*Callithrix kuhlii*). *Ethology* 109:327–40. [aHA]

Russell, J. L., McIntyre, J. M., Hopkins, W. D. & Taglialatela, J. P. (2013) Vocal learning of a communicative signal in captive chimpanzees, *Pan troglodytes*. *Brain and Language* 127(3):520–25. [AM]

Ruzza, B., Rocca, F., Lenti Boero, D. & Lenti, C. (2003) Investigating the musical qualities of early infant sounds. In: *The neurosciences and music*, ed. G. Avanzini, C. Faienza, D. Minciacchi, L. Lopez & L. Majno, pp. 527–29. New York Academy of Sciences. [DLB]

Sacks, O. (1985) The president's speech. In: *The man who mistook his wife for a hat and other clinical tales*, by O. Sacks, pp. 80–85. Simon & Schuster. [ACC]

Sakai, K., Kitaguchi, K. & Hikosaka, O. (2003) Chunking during human visuomotor sequence learning. *Experimental Brain Research* 152(2):229–42. doi: 10.1007/s00221-003-1548-8. [AZ]

Salvante, K. G., Racke, D. M., Campbell, C. R. & Sockman, K. W. (2010) Plasticity in singing effort and its relationship with monoamine metabolism in the songbird telencephalon. *Developmental Neurobiology* 70:41–57. [CMV]

Satoh, T., Nakai, S., Sato, T. & Kimura, M. (2003) Correlated coding of motivation and outcome of decision by dopamine neurons. *Journal of Neuroscience* 23:9913–23. [aHA]

Savage-Rumbaugh, S., Fields, W. M. & Spircu, T. (2004) The emergence of knapping and vocal expression embedded in a Pan/Homo culture. *Biology and Philosophy* 19:541–75. [aHA]

Sawamoto, N., Honda, M., Hanakawa, T., Aso, T., Inoue, M., Toyoda, H., Ishizu, K., Fukuyama, H. & Shibasaki, H. (2007) Cognitive slowing in Parkinson disease is accompanied by hypofunctioning of the striatum. *Neurology* 68:1062–68. [TH]

Sawamoto, N., Honda, M., Hanakawa, T., Fukuyama, H. & Shibasaki, H. (2002) Cognitive slowing in Parkinson's disease: A behavioral evaluation independent of motor slowing. *The Journal of Neuroscience* 22:5198–203. [TH]

Schachner, A., Brady, T. F., Pepperberg, I. M. & Hauser, M. D. (2009) Spontaneous motor entrainment to music in multiple vocal mimicking species. *Current Biology* 19(10):831–36. doi: 10.1016/j.cub.2009.03.061. [HH]

Scharff, C. & Haesler, S. (2005) An evolutionary perspective on FoxP2: Strictly for the birds? *Current Opinion in Neurobiology* 15:694–703. [aHA]

Scharff, C. & Petri, J. (2011) Evo-devo, deep homology and FoxP2: Implications for the evolution of speech and language. *Philosophical Transactions of the Royal Society, Series B* 366:2124–40. [KJA]

Scheiner, E., Hammerschmidt, K., Jürgens, U. & Zwirner, P. (2006) Vocal expression of emotions in normally hearing and hearing-impaired infants. *Journal of Voice* 20(4):585–604. [DKO]

Schel, A. M., Townsend, S. W., Machanda, Z., Zuberbühler, K. & Slocombe, K. E. (2013) Chimpanzee alarm call production meets key criteria for intentionality. *PLOS ONE* 8:e76674. [ARL]

Scherer, K. R. (1986) Vocal affect expression: A review and a model for future research. *Psychological Bulletin* 99:143–65. [aHA]

Scherer, K. R., Johnstone, T. & Klasmeyer, G. (2009) Vocal expression of emotion. In: *Handbook of affective sciences*, ed. R. J. Davidson, K. R. Scherer & H. Hill Goldsmith, pp. 433–56. Oxford University Press. [aHA]

Schilbach, L., Timmermans, B., Reddy, V., Costall, A., Bente, G., Schlicht, T. & Vogeley, K. (2013) Toward a second-person neuroscience. *Behavioral and Brain Sciences* 36(4):393–462. [KJA]

Schmahmann, J. D. & Sherman, J. C. (1998) The cerebellar cognitive affective syndrome. *Brain* 121 (Pt. 4):561–79. [UH, rHA]

Schmidt, L., Forgeot d'Arc, B., Lafargue, G., Galanaud, D., Czernecki, V., Grabli, D., Schübach, M., Hartmann, A., Lévy, R., Dubois, B. & Pessiglione, M. (2008) Disconnecting force from money: Effects of basal ganglia damage on incentive motivation. *Brain: A Journal of Neurology* 131(Pt. 5):1303–10. doi: 10.1093/brain/awn045. [AZ]

Schneider, S. M. (2012) *The science of consequences*. Prometheus. [ACC]

Schultz, W. (2006) Behavioral theories and the neurophysiology of reward. *Annual Review of Psychology* 57:87–115. [aHA]

Schultz, W. (2007) Behavioral dopamine signals. *Trends in Neurosciences* 30:203–10. [aHA]

Schultz, W. (2010) Dopamine signals for reward value and risk: Basic and recent data. *Behavioral and Brain Functions* 6:24. [aHA]

Schultz, W., Tremblay, L. & Hollerman, J. R. (2000) Reward processing in primate orbitofrontal cortex and basal ganglia. *Cerebral Cortex* 10(3):272–84. [CIP]

Schulz, G. M., Varga, M., Jeffires, K., Ludlow, C. L. & Braun, A. R. (2005) Functional neuroanatomy of human vocalization: An $H_2^{15}O$ PET study. *Cerebral Cortex* 15:1835–47. [aHA]

Schwab, E. D. & Pienta, K. J. (1997) Modeling signal transduction in normal and cancer cells using complex adaptive systems. *Medical Hypotheses* 48:111–23. [TAM]

Sebanz, N., Bekkering, H. & Knoblich, G. (2006) Joint action: Bodies and minds moving together. *Trends in Cognitive Sciences* 10:70–76. [GP]

Seeley, W. W. (2008) Selective functional, regional, and neuronal vulnerability in frontotemporal dementia. *Current Opinion in Neurology* 21:701–707. [aHA]

Selezneva, E., Deike, S., Knyazeva, S., Scheich, H., Brechmann, A. & Brosch, M. (2013) Rhythm sensitivity in macaque monkeys. *Frontiers in Systems Neuroscience* 7:49. doi: 10.3389/fnsys.2013.00049. [HH]

Seyfarth, R. M. (2005) Continuities in vocal communication argue against a gestural origin of language. *Behavioral and Brain Sciences* 28:144–45. [ARL]

Seyfarth, R. M. & Cheney, D. L. (1980) Monkey responses to three different alarm calls: Evidence of predator classification and semantic communication. *Science* 210:801–803. [DLB]

Seyfarth, R. M. & Cheney, D. L (2003a) Meaning and emotion in animal vocalizations. *Annals of the New York Academy of Sciences* 1000:32–55. [ARL]

Seyfarth, R. M. & Cheney, D. L. (2003b) Signalers and receivers in animal communication. *Annual Review of Psychology* 54:145–73. [aHA]

Seyfarth, R. M. & Cheney, D. L. (2008) Primate social knowledge and the origins of language. *Mind and Society* 7:129–42. [ARL]

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

Seyfarth, R. M., Cheney, D. L. & Bergman, T. J. (2005) Primate social cognition and the origins of language. *Trends in Cognitive Sciences* 9:264–66. [ARL]

Seyfarth, R. M., Cheney, D. L. & Marler, P. (1980) Vervet monkey alarm calls: Semantic communication in a free-ranging primate. *Animal Behaviour* 28:1070–94. [aHA]

Sheinkopf, S. J., Iverson, J. M., Rinaldi, M. L. & Lester, B. M. (2012) Atypical cry acoustics in 6-month-old infants at risk for autism spectrum disorder. *Autism Research* 5(5):331–39. [MHB]

Sherwood, C. C. (2005) Comparative anatomy of the facial motor nucleus in mammals, with an analysis of neuron numbers in primates. *The Anatomical Record, Part A: Discoveries in Molecular, Cellular, and Evolutionary Biology* 287(1):1067–79. [aHA]

Sherwood, C. C., Broadfield, D. C., Holloway, R. L., Gannon, P. J. & Hof, P. R. (2003) Variability of Broca's area homologue in African great apes: Implications for language evolution. *The Anatomical Record, Part A: Discoveries in Molecular, Cellular, and Evolutionary Biology* 271:276–85. [aHA]

Sherwood, C. C., Hof, P. R., Holloway, R. L., Semendeferi, K., Gannon, P. J., Frahm, H. D. & Zilles, K. (2005) Evolution of the brainstem orofacial motor system in primates: A comparative study of trigeminal, facial, and hypoglossal nuclei. *Journal of Human Evolution* 48:45–84. [aHA]

Shriberg, L. D., Aram, D. M. & Kwiatkowski, J. (1997) Developmental apraxia of speech: I. Descriptive and theoretical perspectives. *Journal of Speech, Language, and Hearing Research* 40:273–85. [aHA]

Shriberg, L. D., Ballard, K. J., Tomblin, J. B., Duffy, J. R., Odell, K. H. & Williams, C. A. (2006) Speech, prosody, and voice characteristics of a mother and daughter with a 7;13 translocation affecting FOXP2. *Journal of Speech Language Hearing Research* 49:500–25. [CMV]

Shu, W. G., Cho, J. Y., Jiang, Y. H., Zhang, M. H., Weisz, D., Elder, G. A., Schmeidler, J., De Gasperi, R., Gama Sosa, M. A., Rabidou, D., Santucci, A. C., Perl, D., Morrisey, E. & Buxbaum, J. D. (2005) Altered ultrasonic vocalization in mice with a disruption in the *FoxP2* gene. *Proceedings of the National Academy of Science USA* 102:9643–48. [KJA]

Sidtis, J. J. & Van Lancker Sidtis, D. (2003) A neurobehavioral approach to dysprosody. *Seminars in Speech and Language* 24:93–105. [arHA, CMV]

Simard, F., Joanette, Y., Petrides, M. Jubault, T., Madjar, C. & Monchi, O. (2011) Fronto-striatal contributions to lexical set-shifting. *Cerebral Cortex* 21(5):1084–93. [PL]

Simon, S. A., de Araujo, I. E., Gutierrez, R. & Nicolelis, M. A. (2006) The neural mechanisms of gustation: A distributed processing code. *Nature Reviews Neuroscience* 7:890–901. [CMV]

Simonyan, K. & Jürgens, U. (2002) Cortico-cortical projections of the motorcortical larynx area in the rhesus monkey. *Brain Research* 949:23–31. [aHA]

Simonyan, K. & Jürgens, U. (2005) Afferent subcortical connections into the motor cortical larynx area in the rhesus monkey. *Neuroscience* 130:119–31. [aHA]

Sinnott, J. M., Stebbins, W. C. & Moody, D. B. (1975) Regulation of voice amplitude by the monkey. *Journal of the Acoustical Society of America* 58:412–14. [DJW]

Skinner, B. F. (1986) The evolution of verbal behavior. *Journal of the Experimental Analysis of Behavior* 45:115–22. [ACC]

Skodda, S. (2012) Effect of deep brain stimulation on speech performance in Parkinson's Disease. *Parkinson's Disease* 2012(Article No. 850596):1–10. Available at: http://dx.doi.org/10.1155/2012/850596 [CMV]

Skodda, S., Grönheit, W. & Schlegel, U. (2011) Intonation and speech rate in Parkinson's disease: General and dynamic aspects and responsiveness to levodopa admission. *Journal of Voice* 25:199–205. [aHA]

Skodda, S., Rinsche, H. & Schlegel, U. (2009) Progression of dysprosody in Parkinson's disease over time – a longitudinal study. *Movement Disorders* 24:716–22. [aHA]

Sliwa, J., Duhamel, J. R., Pascalis, O. & Wirth, S. (2011) Spontaneous voice-face identity matching by rhesus monkeys for familiar conspecifics and humans. *Proceedings of the National Academy of Sciences USA* 108:1735–40. [DYT]

Slocombe, K. E. & Zuberbuhler, K. (2007) Chimpanzees modify recruitment screams as a function of audience composition. *Proceedings of the National Academy of Sciences USA* 104:17228–33. [ARL]

Smith, A. (2010) Development of neural control of orofacial movements for speech. In: *The handbook of phonetic sciences*, 2nd edition, ed. W. J. Hardcastle, J. Laver & F. E. Gibbon, pp. 251–96. Wiley-Blackwell. [aHA]

Smith, M. C., Smith, M. K. & Ellgring, H. (1996) Spontaneous and posed facial expression in Parkinson's disease. *Journal of the International Neuropsychological Society* 2(5):383–91. [AZ]

Smith, W. K. (1945) The functional significance of the rostral cingulate cortex as revealed by its responses to electrical excitation. *Journal of Neurophysiology* 8:241–55. [aHA]

Snowdon, C. T. (2008) Contextually flexible communication in nonhuman primates. In: *Evolution of communicative flexibility: Complexity, creativity, and adaptability in human and animal communication*, ed. D. K. Oller & U. Griebel, pp. 71–91. MIT Press. [aHA]

Snowdon, C. T. & Elowson, A. M. (1999) Pygmy marmosets modify call structure when paired. *Ethology* 105:893–908. [aHA]

Solano, A., Roig, M., Vives-Bauza, C., Hernandez-Peña, J., Garcia-Arumi, E., Playan, A., Lopez-Perez, M. J., Andreu, A. L. & Montoya, J. (2003) Bilateral striatal necrosis associated with a novel mutation in the mitochondrial ND6 gene. *Annals of Neurology* 54:527–30. [aHA]

Spencer, K. A. & MacDougall-Shackleton, S. A. (2011) Indicators of development as sexually selected traits: The developmental stress hypothesis in context. *Behavioral Ecology* 22:1–9. doi: 10.1093/beheco/arq068. [BM]

Sperli, F., Spinelli, L., Pollo, C. & Seeck, M. (2006) Contralateral smile and laughter, but no mirth, induced by electrical stimulation of the cingulate cortex. *Epilepsia* 47:440–43. [aHA]

Spiteri, E., Konopka, G., Coppola, G., Bomar, J., Oldham, M., Ou, J., Vernes, S. C., Fisher, S. E., Ren, B. & Geschwind, D. H. (2007) Identification of the transcriptional targets of *FOXP2*, a gene linked to speech and language, in developing human brain. *American Journal of Human Genetics* 8:1144–57. [KJA]

Stark, R. E. (1980) Stages of speech development in the first year of life. In: *Child phonology, vol. 1: Production*, ed. G. Yeni-Komshian, J. Kavanagh & C. Ferguson, pp. 73–90. Academic Press. [DKO]

Stark, R. E., Bernstein, L. E. & Demorest, M. E. (1993) Vocal communication in the first 18 months of life. *Journal of Speech and Hearing Research* 36:548–58. [DKO]

Steinschneider, M., Nourski, K. V. & Fishman, Y. I. (2013) Representation of speech in human auditory cortex: Is it special? *Hearing Research* 305:57–73. [DHR]

Steinschneider, M., Volkov, I. O., Fishman, Y. I., Oya, H., Arezzo, J. C. & Howard, M. A., 3rd. (2005) Intracortical responses in human and monkey primary auditory cortex support a temporal processing mechanism for encoding of the voice onset time phonetic parameter. *Cerebral Cortex* 15(2):170–86. [DHR]

Sterelny, K. (2012) *The evolved apprentice: How evolution made humans unique.* MIT Press. [rHA]

Stivers, T., Enfield, N. J., Brown, P., Englert, C., Hayashi, M., Heinemann, T., Hoymann, G., Rossano, F., de Ruiter, J. P., Yoon, K. E. & Levinson, S. C. (2009) Universals and cultural variation in turn-taking in conversation. *Proceedings of the National Academy of Sciences USA* 106:10587–92. [DYT]

Stoeger, A. S., Mietchen, D., Oh, S., de Silva, S. & Herbst, C. T. (2012) An Asian elephant imitates human speech. *Current Biology* 22(22):2144–48. [BF]

Stoodley, C. J. & Schmahmann, J. D. (2010) Evidence for topographic organization in the cerebellum of motor control versus cognitive and affective processing. *Cortex* 46(7):831–44. doi: 10.1016/j.cortex.2009.11.008. [UH]

Stowe, M. K., Turlings, T. C., Loughrin, J. H., Lewis, W. J. & Tumlinson, J. H. (1995) The chemistry of eavesdropping, alarm, and deceit. *Proceedings of the National Academy of Science USA* 92(1):23–28. [KBC]

Striedter, G. F. (2005) *Principles of brain evolution*. Sinauer. [aHA]

Stringer, C. (2012) *Lone survivors: How we came to be the only humans on earth*. Times Books, Henry Holt. [aHA]

Struhsaker, T. T. (1967) Auditory communication among vervet monkeys (*Cercopithecus aethiops*). In: *Social communication among primates*, ed. S. A. Altmann, pp. 281–324. University of Chicago Press. [aHA]

Stuss, D. T., Benson, D. F. (1986) *The frontal lobes*. Raven. [PL]

Sugiura, H. (1998) Matching of acoustic features during the vocal exchange of coo calls by Japanese macaques. *Animal Behaviour* 55:673–87. [aHA]

Surmeier, D. J., Day, M., Gertler, T., Chan, S. & Shen, W. (2010a) D1 and D2 dopamine receptor modulation of glutamatergic signaling in striatal medium spiny neurons. In: *Handbook of basal ganglia structure and function*, ed. H. Steiner, K. Y. Tseng, pp. 113–32. Elsevier. [aHA]

Surmeier, D. J., Day, M., Gertler, T.S, Chan, C. S. & Shen, W. (2010b) Dopaminergic modulation of striatal glutamatergic signaling in health and Parkinson's disease. In: *Dopamine handbook*, ed. L. L. Iversen, S. D. Iversen, S. B. Dunnett & A. Björklund, pp. 349–67. Oxford University Press. [aHA]

Sutton, D., Larson, C. & Lindeman, R. C. (1974) Neocortical and limbic lesion effects on primate phonation. *Brain Research* 71:61–75. [aHA]

Sutton, D., Larson, C., Taylor, E. M. & Lindeman, R. C. (1973) Vocalization in rhesus monkeys: Conditionability. *Brain Research* 52:225–31. [aHA]

Sutton, D., Trachy, R. E. & Lindeman, R. C. (1981) Vocal and nonvocal discriminative performance in monkeys. *Brain and Language* 14:93–105. [aHA]

Sutton, D., Trachy, R. E. & Lindeman, R. C. (1985) Discriminative phonation in macaques: Effects of anterior mesial cortex damage. *Experimental Brain Research* 59:410–13. [aHA]

Syal, S. & Finlay, B. L. (2011) Thinking outside the cortex: Social motivation in the evolution and development of language. *Developmental Science* 14:417–30. [DYT]

Szalontai, A. & Csiszar, K. (2013) Genetic insights into the functional elements of language. *Human Genetics* 132:959–86. [KJA]

Taglialatela, J. P., Reamer, L., Schapiro, S. J. & Hopkins, W. D. (2012) Social learning of a communicative signal in captive chimpanzees. *Biology Letters* 8:498–50. [AM]

Taglialatela, J. P., Russell, J. L., Schaeffer, J. A. & Hopkins, W. D. (2008) Communicative signaling activates Broca's homolog in chimpanzees. *Current Biology* 18(5):343–48. [AM]

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:e18852. [AM]

Taglialatela, J. P., Savage-Rumbaugh, S. & Baker, L. A. (2003) Vocal production by a language-competent *Pan paniscus*. *International Journal of Primatology* 24:1–17. [aHA]

Takahashi, D. Y., Narayanan, D. Z. & Ghazanfar, A. A. (2013) Coupled oscillator dynamics of vocal turn-taking in monkeys. *Current Biology* 23:2162–68. [DYT]

Takikawa, Y., Kawagoe, R., Itoh, H., Nakahara, H. & Hikosaka, O. (2002) Modulation of saccadic eye movements by predicted reward outcome. *Experimental Brain Research* 142(2):284–91. doi: 10.1007/s00221-001-0928-1. [AZ]

Tallerman, M. & Gibson, K. R. (2012) *The Oxford handbook of language evolution*. Oxford University Press. (*Oxford Handbooks in Linguistics* Series). [aHA]

Talmage-Riggs, G., Winter, P., Ploog, D. & Mayer, W. (1972) Effect of deafening on the vocal behavior of the squirrel monkey (*Saimiri sciureus*). *Folia Primatologica* 17:404–20. [aHA]

Tamis-LeMonda, C. S., Bornstein, M. H., Baumwell, L. & Damast, A. M. (1996) Responsive parenting in the second year: Specific influences on children's language and play. *Early Development and Parenting* 5:173–83. [MHB]

Taylor, J. (2009) *Not a chimp: The hunt to find the genes that make us human*. Oxford University Press. [aHA]

Teichmann, M., Dupoux, E., Kouider, S., Brugières, P., Boissé, M. F., Baudic, S., Cesaro, P., Peschanski, M. & Bachoud-Lévi, A. C. (2005) The role of the striatum in rule application: The model of Huntington's disease at early stage. *Brain* 128:1155–67. [rHA]

Teichmann, M., Gaura, V., Demonet, J. F., Supiot, F., Delliaux, M., Verny, C., Renou, P., Remy, P. & Bachoud-Levi, A. C. (2008) Language processing within the striatum: Evidence from a PET correlation study in Huntington's disease. *Brain* 131(4):1046–56. doi: 10.1093/brain/awn036. [AR]

Teinonen, T., Fellman, V., Näätänen, R., Alku, P. & Huotilainen, M. (2009) Statistical language learning in neonates revealed by event-related brain potentials. *BMC Neuroscience* 10:21. doi: 10.1186/1471-2202-10-21. [DLB]

Teki, S., Grube, M., Kumar, S. & Griffiths, T. D. (2011) Distinct neural substrates of duration-based and beat-based auditory timing. *The Journal of Neuroscience* 31 (10):3805–12. doi: 10.1523/jneurosci.5561-10.2011. [HH]

ten Cate, C. & Okanoya, K. (2012) Revisiting the syntactic abilities of non-human animals: Natural vocalizations and artificial grammar learning. *Philosophical Transactions of the Royal Society, B: Biological Sciences* 367(1598):1984–94. [AR]

Teramitsu, I., Kudo, L. C., London, S. E., Geschwind, D. H. & White, S. A. (2004) Parallel *FoxP1* and *FoxP2* expression in human and songbird brain predicts functional interaction. *Journal of Neuroscience* 24:3152–63. [KJA]

Teramitsu, I., Poopatanapong, A., Torrisi, S. & White, S. A. (2010) Striatal *FoxP2* is actively regulated during songbird sensorimotor learning. *PLoS ONE* 5: e8548. [aHA]

Terao, S., Li, M., Hashizume, Y., Osano, Y, Mitsuma, T. & Sobue, G. (1997) Upper motor neuron lesions in stroke patients do not induce anterograde transneuronal degeneration in spinal anterior horn cells. *Stroke* 28(12):2553–56. [PL]

Tervaniemi, M. & Huotilainen, M. (2003) The promises of change-related brain potentials in cognitive neuro science of music. In: *The neurosciences and music,* ed. G. Avanzini, C. Faienza, D. Minciacchi, L. Lopez & L. Majno, pp. 29–39. New York Academy of Sciences. [DLB]

Tettamanti, M. & Weniger, D. (2006) Broca's area: A supramodal hierarchical processor? *Cortex* 42(4):491–94. [AZ]

Thyagarajan, D., Shanske, S., Vazquez-Memije, M., DeVivo, D. & DiMauro, S. (1995) A novel mitochondrial ATPase 6 point mutation in familial bilateral striatal necrosis. *Annals of Neurology* 38:468–72. [aHA]

Tishkoff, S. A., Reed, F., Ranciaro, A., Voight, B. F., Babbitt, C. C., Silverman, J. S., Powell, K., Mortensen, H. M., Hirbo, J. B., Osman, M., Ibrahim, M., Omar, S. A., Lema, G., Nyambo, T. B., Ghori, J., Bumpstead, S., Pritchard, J. K., Wray, G. A. & Deloukas, P. (2007) Convergent adaptation of human lactase persistence in Africa and Europe. *Nature Genetics* 39(1):31–40. [PL]

Titze, I. R. (2008) Nonlinear source–filter coupling in phonation: Theory. *Journal of the Acoustical Society of America* 123(5):2733–49. [BdB]

Tokita, K., Karadi, Z., Shimura, T. & Yamamoto, T. (2004) Centrifugal inputs modulate taste aversion learning associated parabrachial neuronal activities. *Journal of Neurophysiology* 92:265–79. [CMV]

Toma, C., Hervásk, A., Torrico, B., Balmañnak, N., Salgado, M., Maristany, M., Vilella, E., Martínez-Leal, R., Planelles, M. I., Cuscóc, I., del Campo, M., Pérez-Jurado, L. A., Caballero-Andaluz, R., de Diego-Otero, Y., Pérez-Costillas, L., Ramos-Quiroga, J. A., Ribasés, M., Bayés, M. & Cormand, B. (2013) Analysis of two language-related genes in autism: A case–control association study of FOXP2 and CNTNAP2. *Psychiatric Genetics* 23(2):82–85. doi: 10.1097/YPG.0b013e32835d6fc6. [KJA]

Tomasello, M. (2008) *Origins of human communication.* MIT Press. [aHA, DYT]

Toscano, J. C., McMurray, B., Dennhardt, J. & Luck, S. J. (2010) Continuous perception and graded categorization: Electrophysiological evidence for a linear relationship between the acoustic signal and perceptual encoding of speech. *Psychological Science* 21:1532–40. [DHR]

Townsend, S. W., Deschner, T. & Zuberbühler, K. (2008) Female chimpanzees use copulation calls flexibly to prevent social competition. *PLOS ONE* 3:e2431. [ARL]

Trachy, R. E., Sutton, D. & Lindeman, R. C. (1981) Primate phonation: Anterior cingulate lesion effects on response rate and acoustical structure. *American Journal of Primatology* 1:43–55. [aHA]

Tremblay, P.-L., Bedard, M.-A., Langlois, D., Blanchet, P. J., Lemay, M. & Parent, M. (2010) Movement chunking during sequence learning is a dopamine-dependant process: A study conducted in Parkinson's disease. *Experimental Brain Research.* 205(3):375–85. doi: 10.1007/s00221-010-2372-6. [AZ]

Tremblay, P.-L., Bedard, M.-A., Levesque, M., Chebli, M., Parent, M., Courtemanche, R. & Blanchet, P. J. (2009) Motor sequence learning in primate: Role of the D2 receptor in movement chunking during consolidation. *Behavioural Brain Research* 198(1):231–39. doi: 10.1016/j.bbr.2008.11.002. [AZ]

Trevarthen, C. & Aitken, K. J. (2001) Infant intersubjectivity: Research, theory, and clinical applications. (Annual Research Review.) *Journal of Child Psychology and Psychiatry* 42(1):3–48. [KJA]

Turner, R. S. & Desmurget, M. (2010) Basal ganglia contributions to motor control: A vigorous tutor. *Current Opinion in Neurobiology* 20(6):704–16. doi: 10.1016/j.conb.2010.08.022. [AZ]

Turner, V. (1967) *The forest of symbols: Aspects of Ndembu ritual*. Cornell University Press. [aHA]

Tuttle, R. H. (2007) Apes, intelligent science, and conservation. In: *Primate perspectives on behavior and cognition*, ed. D. A. Washburn, pp. 17–28. American Psychological Association. [aHA]

Tye-Murray, N., Spencer, L. & Woodworth, G. G. (1995) Acquisition of speech by children who have prolonged cochlear implant experience. *Journal of Speech and Hearing Research* 38(2):327–37. [TH]

Ullman, M. T. (2001) A neurocognitive perspective on language: The declarative/procedural model. *Nature Reviews Neuroscience* 2:717–26. [rHA]

Ullman, M. T. (2004) Contributions of memory circuits to language: The declarative/procedural model. *Cognition* 92(1–2):231–70. [AR]

Ullman, M. T. (2006) Is Broca's area part of a basal ganglia thalamocortical circuit? *Cortex* 42(4):480–85. doi: 10.1016/S0010-9452(08)70382-4. [AZ]

Ullman, M. T., Corkin, S., Coppola, M., Hickok, G., Growdon, J. H., Koroshetz, W. J. & Pinker, S. (1997) A neural dissociation within language: Evidence that the mental dictionary is part of declarative memory, and that grammatical rules are processed by the procedural system. *Journal of Cognitive Neuroscience* 9(2):266–76. [AR]

Ungerleider, L. G., Doyon, J. & Karni, A. (2002) Imaging brain plasticity during motor skill learning. *Neurobiology of Learning and Memory* 78:553–64. [aHA]

Usui, C., Inoue, Y., Kimura, M., Kirino, E., Nagaoka, S., Abe, M., Nagata, T. & Arai, H. (2004) Irreversible subcortical dementia following high altitude illness. *High Altitude Medicine and Biology* 5(1):77–81. [PL]

Van Hooff, J. A. R. A. M. (1962) Facial expressions of higher primates. *Symposium of the Zoological Society, London* 8:97–125. [DYT]

Van Lancker, D. & Cummings, J. L. (1999) Expletives: Neurolinguistic and neurobehavioral perspectives on swearing. *Brain Research Reviews* 31:83–104. [CM]

Van Lancker Sidtis, D., Pachana, N., Cummings, J. L. & Sidtis, J. J. (2006) Dysprosodic speech following basal ganglia insult: Toward a conceptual framework for the study of the cerebral representation of prosody. *Brain and Language* 97:135–53. [aHA, PBM, CMV]

van Schaik, C. P., Ancrenaz, M., Borgen, G., Galdikas, B., Knott, C. D., Singleton, I., Suzuki, A., Utami, S. S. & Merrill, M. (2003) Orangutan cultures and the evolution of material culture. *Science* 299:102–105. [aHA]

van Schaik, C. P. & Burkart, J. (2010) Mind the gap: Cooperative breeding and the evolution of our unique features. In: *Mind gap*, ed. P. M. Kappeler & J. Silk, pp. 477–96. Springer. [ARL]

van Schaik, C. P., van Noordwijk, M. A., & Wich, S. A. (2006) Innovation in wild Bornean orangutans (*Pongo pygmaeus wurmbii*). *Behaviour* 143:839–76. [aHA]

Vargha-Khadem, F., Gadian, D. G., Copp, A. & Mishkin, M. (2005) *FOXP2* and the neuroanatomy of speech and language. *Nature Reviews Neuroscience* 6:131–38. [aHA]

Vargha-Khadem, F. & Passingham, R. (1990) Speech and language defects. *Nature* 346(6281):226. [aHA, KJA]

Vargha-Khadem, F., Watkins, K. E., Alcock, K. J., Fletcher, P. & Passingham, R.E. (1995) Praxic and nonverbal cognitive deficits in a large family with a genetically transmitted speech and language disorder. *Proceedings of the National Academy of Sciences USA* 92:930–33. [aHA]

Vargha-Khadem, F., Watkins, K. E., Price, C. J., Ashburner, J., Alcock, K. J., Connelly, A., Frackowiak, R. S. J., Friston, K. J., Pembrey, M. E., Mishkin, M., Gadian, D. G. & Passingham, R. E. (1998) Neural basis of an inherited speech and language disorder. *Proceedings of the National Academy of Sciences USA* 95:12695–700. [aHA]

Venuti, P., Caria, A., Esposito, G., De Pisapia, N., Bornstein, M. H. & de Falco, S. (2012) Differential brain responses to cries of infants with autistic disorder and typical development: An fMRI study. *Research in Developmental Disabilities* 33 (6):2255–64. [MHB]

Verplanck, W. S. (2000) *Glossary/thesaurus of behavioral terms* (CD-ROM). CMS Software. [ACC]

Verwey, W. B. (1996) Buffer loading and chunking in sequential keypressing. *Journal of Experimental Psychology: Human Perception and Performance* 22(3):544–62. [AZ]

Verwey, W. B. (2001) Concatenating familiar movement sequences: The versatile cognitive processor. *Acta Psychologica* 106:69–95. [AZ]

Verwey, W. B. & Eikelboom, T. (2003) Evidence for lasting sequence segmentation in the discrete sequence-production task. *Journal of Motor Behavior* 35(2):171–81. doi: 10.1080/00222890309602131. [AZ]

Vicario, C. M. (2013a) FOXP2 gene and language development: The molecular substrate of the gestural-origin theory of speech? *Frontiers in Behavioral Neuroscience* 7:99. [CMV]

Vicario, C. M. (2013b) Inborn mechanisms of food preference and avoidance: The role of polymorphisms in neuromodulatory systems. *Frontiers in Molecular Neuroscience* 6:16. [CMV]

Vicario, C. M. (2013c) Uncovering the neurochemistry of reward and aversiveness. *Frontiers in Molecular Neuroscience*.6:41. [CMV]

Vihman, M. M. (1996) *Phonological development: The origins of language in the child*. Blackwell. [ACC]

Vogt, B. A. & Barbas, H. (1988) Structure and connections of the cingulate vocalization region in the rhesus monkey. In: *The physiological control of mammalian vocalization*, ed. J. D. Newman, pp. 203–25. Plenum Press. [aHA]

Volterra, V. & Erting C. J., eds. (1990) From gesture to language in hearing and deaf children. *Springer Series in Language and Communication* 27:97–106. [KJA]

von Humboldt, W. (1836/1971) *Über die Verschiedenheit des Menschlichen Sprachbaues und ihren Einfluss auf die geistige Entwicklung des Menschengeschlechts* [*Linguistic variability and intellectual development*], trans. G. C. Buck & F. Raven. Royal Academy of Sciences/University of Miami Press. (Original work published in 1836; English translation from the German by Buck & Raven in 1971). [BM]

Voorn, P., Vanderschuren, L. J. M. J., Groenewegen, H. J., Robbins, T. W. & Pennartz, C. M. A. (2004) Putting a spin on the dorsal–ventral divide of the striatum. *Trends in Neurosciences* 27:468–74. [aHA]

Vorperian, H. K. & Kent, R. D. (2007) Vowel acoustic space development in children: A synthesis of acoustic and anatomic data. *Journal of Speech Language and Hearing Research* 50(6):1510–45. [DLB]

Vorperian, H. K., Wang, S., Chung, M. K., Schimek, E. M., Durtschi, R. B., Kent, R. D., Ziegert, A. J. & Gentry, L. R. (2009) Anatomic development of the oral and pharyngeal portions of the vocal tract: An imaging study. *Journal of Acoustic Society of America* 125:1666–78. [DLB]

Vouloumanos, A., Hauser, M. D., Werker, J. F. & Martin, A. (2010) The tuning of human neonates' preference for speech. *Child Development* 81(2):517–27. [BF]

Wagner, M. (2010) Prosody and recursion in coordinate structures and beyond. *Natural Language and Linguistic Theory* 28(1):183–237. doi: 10.1007/s11049-009-9086-0. [AR]

Wagner, M. & Watson, D. G. (2010) Experimental and theoretical advances in prosody: A review. *Language and Cognitive Processes* 25(7–9):905–45. doi: 10.1080/01690961003589492. [AR]

Wallman, J. (1992) *Aping language*. Cambridge University Press. [aHA]

Walters, J. R. & Bergstrom, D. A. (2010) Synchronous activity in basal ganglia circuits. In: *Handbook of basal ganglia structure and function*, ed. H. Steiner & K. Y. Tseng, pp. 429–43. Elsevier. [aHA]

Watkins, K. E., Dronkers, N. F. & Vargha-Khadem, F. (2002a) Behavioural analysis of an inherited speech and language disorder: Comparison with acquired aphasia. *Brain: A Journal of Neurology* 125(Pt. 3):452–64. [aHA, AZ]

Watkins, K. E., Gadian, D. G. & Vargha-Khadem, F. (1999) Functional and structural brain abnormalities associated with a genetic disorder of speech and language. *American Journal of Human Genetics* 65:1215–21. [aHA]

Watkins, K. E., Vargha-Khadem, F., Ashburner, J., Passingham, R. E., Connelly, A., Friston, K. J., Frackowiak, R. S., Mishkin, M. & Gadian, D. G. (2002b) MRI analysis of an inherited speech and language disorder: Structural brain abnormalities. *Brain* 125 (Pt. 3):465–78. [aHA, UH]

Watson, R. T., Fleet, W. S., Gonzalez-Rothi, L. & Heilman, K. M. (1986) Apraxia and the supplementary motor area. *Archives of Neurology* 43:787–92. [aHA]

Wattendorf, E., Westermann, B., Fiedler, K., Kaza, E., Lotze, M. & Celio, M. R. (2013) Exploration of the neural correlates of ticklish laughter by functional magnetic resonance imaging. *Cerebral Cortex* 23(6):1280–89. doi: 10.1093/cercor/bhs094. [SF, CM]

Waxman, S. R. & Gelman, S. A. (2009) Early word-learning entails reference, not merely associations. *Trends in Cognitive Sciences* 13(6):258–63. [BF]

Weaver, T. D., Roseman, C. C. & Stringer, C. B. (2008) Close correspondence between quantitative- and molecular-genetic divergence times for Neandertals and modern humans. *Proceedings of the National Academy of Sciences USA* 105:4645–49. [aHA]

Weiller, C., Willmes, K., Reiche, W., Thron, A., Isensee, C., Buell, U. & Ringelstein, E. B. (1993) The case of aphasia or neglect after striatocapsular infarction. *Brain* 116:1509–25. [rHA]

Weismer, G. (1980) Control of the voicing distinction for intervocalic stops and fricatives: Some data and theoretical considerations. *Journal of Phonetics* 8:427–38. [aHA]

Weiss, D. J. & Hauser, M. D. (2002) Perception of harmonics in the combination long call of cotton-top tamarins (*Saguinus oedipus*). *Animal Behaviour* 64:415–26. [DJW]

Weiss, D. J., Garibaldi, B. T. & Hauser, M. D. (2001) The production and perception of long calls by cotton-top tamarins (*Saguinus oedipus*): Acoustic analyses and playback experiments. *Journal of Comparative Psychology* 15(3):258–71. [DJW]

West, R. A. & Larson, C. R. (1995) Neurons of the anterior mesial cortex related to faciovocal activity in the awake monkey. *Journal of Neurophysiology* 74:1856–69. [aHA]

Whitham, J., Gerald, M. & Maestripieri, D. (2007) Intended receivers and functional significance of grunt and gurney vocalizations in free-ranging rhesus macaques. *Ethology* 113:862–74. [DHR]

Whitty, C. W. M. (1955) Effects of anterior cingulectomy in man. *Proceedings of the Royal Society of Medicine* 48:463–69. [aHA]

Wich, S. A. & de Vries, H. (2006) Male monkeys remember which group members have given alarm calls. *Proceedings of the Royal Society of London, Series B: Biological Sciences* 273:735–40. [aHA]

Wich, S. A., Krützen, M., Lameira, A. R., Nater, A., Arora, N., Bastian, M. L., Meulman, E., Morrogh-Bernard, H. C., Atmoko, S. S. U., Pamungkas, J., Perwitasari-Farajallah, D., Hardus, M. E., van Noordwijk, M. & van Schaik, C. P. (2012) Call cultures in orang-utans? *PLOS ONE* 7:e36180. [ARL]

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

Wichmann, T. & DeLong, M. R. (2007) Epidemiology of Parkinson's disease. In: *Parkinson's disease and related disorders, Part 1*, ed. W. C. Koller & E. Melamed, pp. 3–18. (*Handbook of Clinical Neurology, vol. 83*, 3rd series). Elsevier Press. [aHA]

Wickens, J. R., Horvitz, J. C., Costa, R. M. & Killcross, S. (2007) Dopaminergic mechanisms in actions and habits. *Journal of Neuroscience* 27:8181–83. [aHA]

Wiener, M., Turkeltaub, P. & Coslett, H. H. (2010) The image of time: A voxel-wise meta-analysis. *NeuroImage* 49:1728–40. [HH]

Wild, B., Rodden, F. A., Grodd, W. & Ruch, W. (2003) Neural correlates of laughter and humour. *Brain* 126:2121–38. [aHA]

Wild, J. M. (1993) Descending projections of the songbird nucleus robustus, archistriatalis. *Journal of Comparative Neurology* 338:225–41. [BM]

Wild, J. M. (1997) Neural pathways for the control of birdsong production. *Journal of Neurobiology* 33:653–70. [BM, CIP]

Wild, J. M. (2008) Birdsong: Anatomical foundations and central mechanisms of sensorimotor integration, In: *Neuroscience of birdsong*, ed. H. P. Zeigler & P. Marler, pp. 136–51. Cambridge University Press. [aHA]

Wildgruber, D., Ackermann, H., Kreifelts, B. & Ethofer, T. (2006) Cerebral processing of linguistic and emotional prosody: fMRI studies. In: *Understanding emotions*, ed. S. Anders, G. Ende, M. Junghofer, J. Kissler & D. Wildgruber, pp. 249–68. (Series: *Progress in Brain Research, vol. 156*). Elsevier. [aHA, UH]

Willuhn, I. & Steiner, H. (2008) Motor-skill learning in a novel running-wheel task is dependent on D1 dopamine receptors in the striatum. *Neuroscience* 153:249–58. [aHA]

Wilson, B., Slater, H., Kikuchi, Y., Milne, A. E., Marslen-Wilson, W., Smith, K. & Petkov, C. I. (2013) Auditory artificial-grammar learning in macaque and marmoset monkeys. *Journal of Neuroscience* 33(48):18825–35. Open Access publication. PMC3841451. [CIP]

Wilson, M. & Wilson, T. P. (2005) An oscillator model of the timing of turn-taking. *Psychonomic Bulletin and Review* 12:957–68. [GAB, rHA]

Wiltermuth, S. S. & Heath, C. (2009) Synchrony and cooperation. *Psychological Science* 20:1–5. [GAB]

Winkler, I., Háden, G. P., Ladinig, O., Sziller, I. & Honing, H. (2009) Newborn infants detect the beat in music. *Proceedings of the National Academy of Sciences USA* 106(7):2468–71. doi: 10.1073/pnas.0809035106. [HH]

Winkler, I., Kushnerenko, E., Horvath, J., Ceponiené, R., Fellman, V., Huotilainen, M., Näätänen, R. & Sussman, E. (2003) Newborn infants can organize the auditory world. *Proceedings of the National Academy of Sciences USA* 100 (2):11812–15. [DLB]

Winter, P., Handley, P., Ploog, D. & Schott, D. (1973) Ontogeny of squirrel monkey calls under normal conditions and under acoustic isolation. *Behaviour* 47:230–39. [aHA, DHR]

Winter, P., Ploog, D. & Latta, J. (1966) Vocal repertoire of the squirrel monkey (*Saimiri sciureus*), its analysis and significance. *Experimental Brain Research* 1:359–84. [aHA]

Wittforth, M., Schröder, C., Schardt, D. M., Dengler, R., Heinze, H. J. & Kotz, S. A. (2010) On emotional conflict: Interference resolution of happy and angry prosody reveals valence-specific effects. *Cerebral Cortex* 20(2):383–92. [KBC, rHA]

Wu, T. & Hallett, M. (2005) A functional MRI study of automatic movements in patients with Parkinson's disease. *Brain* 128:2250–59. [aHA]

Wu, T., Kansaku, K. & Hallett, M. (2004) How self-initiated memorized movements become automatic: A functional MRI study. *Journal of Neurophysiology* 91:1690–98. [aHA]

Wymbs, N. F., Bassett, D. S., Mucha, P. J., Porter, M. A. & Grafton, S. T. (2012) Differential recruitment of the sensorimotor putamen and frontoparietal cortex during motor chunking in humans. *Neuron* 74(5):936–46. doi: 10.1016/j.neuron.2012.03.038. [AZ]

Yale, M. E., Messinger, D. S., Cobo-Lewis, A. B., Oller, D. K. & Eilers, R. E. (1999) An event-based analysis of the coordination of early infant vocalizations and facial actions. *Developmental Psychology* 35(2):505–13. [DKO]

Yildiz, I. B., von Kriegstein, K. & Kiebel, S. J. (2013) From birdsong to human speech recognition: Bayesian inference on a hierarchy of nonlinear dynamical systems. *PLOS Computational Biology* 9:e1003219. [GP]

Yin, H. H., Mulcare, S. P., Hilário, M. R. F., Clouse, E., Holloway, T., Davis, M. I., Hansson, A. C., Lovinger, D. M. & Costa, R. M. (2009) Dynamic reorganization of striatal circuits during the acquisition and consolidation of a skill. *Nature Neuroscience* 12:333–41. [aHA]

Yin, J., Ma, J., Zhang, S. & Metzner, W. (2008) FoxP2 expression in the brain of echolocating and non-echolocating bats and its possible role in vocalization. Paper presented at the 38th Annual Meeting of the Society for Neuroscience, Washington, DC, 2008. Program No. 796.13. [KJA]

Yotova, V., Lefebvre, J.-F., Moreau, C., Gbeha, E., Hovhannesyan, K., Bourgeois, S., Bédarida, S., Azevedo, L., Amorim, A., Sarkisian, T., Avogbe, P., Chabi, N.,

Dicko, M. H., Kou'Santa Amouzou, E. S., Sanni, A., Roberts-Thomson, J., Boettcher, B., Scott, R. J. & Labuda, D. (2011) An X-linked haplotype of Neandertal origin is present among all non-African populations. *Molecular Biology and Evolution* 28:1957–62. [SJ]

Zarco, W., Merchant, H., Prado, L. & Mendez, J. C. (2009) Subsecond timing in primates: Comparison of interval production between human subjects and rhesus monkeys. *Journal of Neurophysiology* 102(6):3191–202. doi: 10.1152/jn.00066.2009. [HH]

Zhang, J., Webb, D. M. & Podlaha, O. (2002) Accelerated protein evolution and origins of human-specific features: FOXP2 as an example. *Genetics* 162:1825–35. [aHA]

Zhang, S. P., Bandler, R. & Davis, P. J. (1995) Brain stem integration of vocalization: Role of the nucleus retroambigualis. *Journal of Neurophysiology* 74:2500–12. [PBM]

Zhang, S. P., Davis, P. J., Bandler, R. & Carrive, P. (1994) Brain stem integration of vocalization: Role of the midbrain periaqueductal gray. *Journal of Neurophysiology* 72:1337–56. [aHA]

Ziegler, W. (2008) Apraxia of speech. In: *Neuropsychology and behavioral neurology*, ed. G. Goldenberg & B. L. Miller, pp. 269–85. (*Handbook of clinical neurology, vol. 88*, 3rd series). Elsevier Press. [aHA]

Ziegler, W. (2010) Apraxic failure and the hierarchical structure of speech motor plans: A nonlinear probabilistic model. In: *Assessment of motor speech disorders*, ed. A. Lowit & R. D. Kent, pp. 305–23. Plural Publishing. [aHA]

Ziegler, W., Aichert, I. & Staiger, A. (2012) Apraxia of speech: Concepts and controversies. *Journal of Speech, Language, and Hearing Research* 55:S1485–501. [aHA, DHR]

Ziegler, W., Kilian, B. & Deger, K. (1997) The role of the left mesial frontal cortex in fluent speech: Evidence from a case of left supplementary motor area hemorrhage. *Neuropsychologia* 35:1197–208. [aHA]

Zuberbühler, K. (2000a) Causal cognition in a nonhuman primate: Field playback experiments with Diana monkeys. *Cognition* 76(3):195–207. [aHA, KBC]

Zuberbühler, K. (2000b) Referential labelling in Diana monkeys. *Animal Behaviour* 59(5):917–27. [DLB]

Zuberbühler, K., Cheney, D. L. & Seyfarth, R. M. (1999) Conceptual semantics in a nonhuman primate. *Journal of Comparative Psychology* 113:33–42. [aHA, KBC]

Zuberbühler, K. & Jenny, D. (2007) Interaction between leopard and monkeys. In: *Monkeys of the Taï Forest: An African primate community*, ed. W. S. McGraw, K. Zuberbühler, & R. Noe, pp. 133–54. Cambridge University Press. [aHA]