

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

Hermann Ackermann (a1), Steffen R. Hage (a2) and Wolfram Ziegler (a3)

<https://doi.org/10.1017/S0140525X1400003X>

Published online: 17 December 2014

In response to commentaries on: Brain mechanisms of acoustic communication in humans and nonhuman primates: An evolutionary perspective

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. 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. 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. 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 cortico-striatal 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, Sigafos, & 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:

1. (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.
2. (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.
3. (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

- Ackermann, H. & Brendel, B. (in press) Cerebellum. In: *The neurobiology of language*, ed. Hickok, G. & Small, S. L. Elsevier.
- Ackermann, H., Konczak, J. & Hertrich, J. (1997b) The temporal control of repetitive articulatory movements in Parkinson's disease. *Brain and Language* 56:312–19.
- 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.
- Ackermann, H. & Ziegler, W. (1994) Acoustic analysis of vocal instability in cerebellar dysfunctions. *Annals of Otology, Rhinology and Laryngology* 103:98–104.
- 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.
- 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.
- Brendel, B. & Ziegler, W. (2008) Effectiveness of metrical pacing in the treatment of apraxia of speech. *Aphasiology* 22:77–102.
- Brumm, H. & Slabbekoorn, H. (2005) Acoustic communication in noise. *Advances in the Study of Behavior* 35:151–209.
- Brumm, H. & Zollinger, S. A. (2011) The evolution of the Lombard effect: 100 years of psychoacoustic research. *Behaviour* 148:1173–98.
- Cardona, J. F., Gershanik, O., Gelormini-Lezama, C., Houck, A. L., Cardona, S., Kargieman, L., Trujillo, N., Arévalo, A., Amoroso, L., Manes, F. & Ibáñez, A. (2013) Action-verb processing in Parkinson's disease: New pathways for motor-language learning. *Brain Structure and Function* 218:1355–73.
- 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
- Cummins, F. (2009) Rhythm as entrainment: The case of synchronous speech. *Journal of Phonetics* 37:16–28.
- Ellgring, H. & Scherer, K. R. (1996) Vocal indicators of mood change in depression. *Journal of Nonverbal Behavior* 20:83–110.
- Fagan, B. (2010) *Cro-Magnon: How the Ice Age gave birth to the first modern humans*. Bloomsbury.
- 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.
- 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. Arbib, M. A., pp. 215–49. Cambridge University Press.

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.

Hage, S. R. & Nieder, A. (2013) Single neurons in monkey prefrontal cortex encode volitional initiation of vocalizations. *Nature Communications* 4:2409.

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.

Hurford, J. R. (2007) *The origins of meaning: Language in the light of evolution*. Oxford University Press.

Jankovic, J. (2008) Parkinson's disease: Clinical features and diagnosis. *Journal of Neurology, Neurosurgery, and Psychiatry* 79:368–76.

Johansson, S. (2005) *Origins of language: Constraints on hypotheses*. John Benjamins.

Kirzinger, A. (1985) Cerebellar lesion effects on vocalization of the squirrel monkey. *Behavioural Brain Research* 16:177–81.

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.

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.

MacNeilage, P. F. (1998) The frame/content theory of evolution of speech production. *Behavioral and Brain Sciences* 21(4):499–511.

MacNeilage, P. F. (2008) *The origin of speech*. Oxford University Press.

MacNeilage, P. F. & Davis, B. L. (2001) Motor mechanisms in speech ontogeny: Phylogenetic, neurobiological and linguistic implications. *Current Opinion in Neurobiology* 11:696–700.

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.

Mithen, S. J. (2006) *The singing Neanderthals: The origins of music, language, mind and body*. Harvard University Press. (Original work published in 2005).

Peelle, J. E. & Davis, M. H. (2012) Neural oscillations carry speech rhythm through to comprehension. *Frontiers in Psychology* 3:320.

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.

Rothermich, K., Schmidt-Kassow, M. & Kotz, S. A. (2012) Rhythm's gonna get you: Regular meter facilitates semantic sentence processing. *Neuropsychologia* 50:232–44.

Schmahmann, J. D. & Sherman, J. C. (1998) The cerebellar cognitive affective syndrome. *Brain* 121 (Pt. 4):561–79.

Sidtis, J. J. & Van Lancker Sidtis, D. (2003) A neurobehavioral approach to dysprosody. *Seminars in Speech and Language* 24:93–105.

Sterelny, K. (2012) *The evolved apprentice: How evolution made humans unique*. MIT Press.

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.

Ullman, M. T. (2001) A neurocognitive perspective on language: The declarative/procedural model. *Nature Reviews Neuroscience* 2:717–26.

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.

Wilson, M. & Wilson, T. P. (2005) An oscillator model of the timing of turn-taking. *Psychonomic Bulletin and Review* 12:957–68.

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.