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).