Published in final edited form as:

Behav Brain Sci. 2014 December ; 37(6): 572–604. doi:10.1017/S0140525X13004202.

Vocal communication is multi-sensorimotor coordination within and between individuals

Daniel Y. Takahashi^{a,b} and Asif A. Ghazanfar^{a,b,c}

Daniel Y. Takahashi: takahashi.d@gmail.com; Asif A. Ghazanfar: asifg@princeton.edu

^aPrinceton Neuroscience Institute, Princeton University, Princeton, NJ 08544

^bDepartment of Psychology, Princeton University, Princeton, NJ 08544

^cDepartment of Ecology and Evolutionary Biology, Princeton University, Princeton, NJ 08544

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.