A neurobiological perspective on the evolutionary origins of multimodal communication

Elizabeth Qing Zhang [1], *Edward Ruoyang Shi* [2], *Michael Pleyer* [3]

[1] School of Linguistic Sciences and Arts, Jiangsu Normal University

[2] Department of Translation and Language Sciences, University Pompeu Fabra

[3] Center for Language Evolution Studies, Nicolaus Copernicus University in Toruń

zqelizabeth@gmail.com, edwardshiruoyangend@gmail.com, pleyer@umk.pl

Recently, the multimodal origin view of language evolution has received increasingly strong support (e.g. Zlatev et al., 2020; Vigliocco et al., 2014; Levinson & Holler, 2014). In particular, robust evidence comes from experimental and observational research on the role of multimodality in human interaction, experimental semiotics, and research on multimodal communication in our primate cousins. Research on multimodal language has shown that multimodal visual and vocal cues contribute to utterance construction (Perniss, 2018; Vigliocco et al., 2014). Recent experimental semiotics studies also provide evidence on the advantage of multimodality in bootstrapping a communication system over the vocal modality and gestural modalities on their own (Macuch Silva et al., 2020). Lastly, research on communication in non-human primates has demonstrated that their communication is inherently multimodal, combining multiple modalities in interaction (Liebal et al., 2014; Fröhlich & van Schaik, 2018).

Although the multimodal origins of communication have gained increasing attention, few studies have addressed the question outside primates to explore the deep evolutionary roots and origins of multimodality. In the current work, from a comparative perspective, we show that multimodality of communication seems to date back to invertebrates, suggesting an evolutionary continuity. From a neurobiological perspective, we propose that FoxP2 could be an important factor in the evolution of multimodal animal communication. In humans, the evolution of the connections between the basal ganglia and hippocampus could have provided humans with the ability to integrate vocalizations and gestures. The human version of FOXP2 with two specific mutations could have been instrumental in this change.

Multimodal communication has been detected in various species. Combinations of different modalities such as visual, acoustic, olfactory and tactile channels have been reported for fruit flies (Ewing, 1983), fish (Tavolga, 1956), and birds (Dalziell & Peters, 2013). Despite songbirds being famous for their vocal communications, closer observations have revealed that movements of other body parts are also concurrent with songs. For example, courtship displays integrate songs with beak movements, head motions and hops (William, 2001). Furthermore, song type repertoire has been shown to be coordinated temporally with a dance-like movement repertoire (Dalziell & Peters, 2013). Nonhuman primates also show simultaneous production of communication signals in vocal, facial and manual modalities (Liebal et al. 2004; Micheletta et al., 2013; Genty et al. 2014). In chimpanzees, for example, approximately 50 percent of vocalizations are produced in conjunction with another communicative modality (Taglialatela et al., 2015). These data indicate that multimodal communication has a phylogenetic origin dating back to invertebrates.

In terms of the brain, the basal ganglia and hippocampus are both conserved subcortical structures that can be found in all vertebrates. Homologous brain structures were also proposed in invertebrates. The functions of the basal ganglia are mainly about motor control and action selection, and also cognitive functions such as procedural memory and learning. The functions of the hippocampus include navigation, episodic memory and declarative memory and learning. Studies on vocal production learning animals, especially birds, have shown that the basal ganglia are crucial in song learning (Jarvis, 2019). Comparison between humans and songbirds revealed that analogous (perhaps homologous) cortico-basal ganglia-thalamo-cortical circuits are important for the learning process of vocalization (Pfenning et al., 2014). Studies on human amnesia patients have suggested that the hippocampus is important for gesture production (Hilliard, Cook & Duff, 2017). Moreover, an interesting hypothesis on the hippocampal contribution to gestures based on the spatial cognition function of the hippocampus in evolution also indicates a possible involvement of the hippocampus in gesture (Levinson, 2023). Therefore, the connection between the basal ganglia and hippocampus could be underlying multimodal communication across species.

One possible contribution to the integration of speech and gesture could be attributed to the human version of FOXP2. In evolution, FoxP2 is a conserved transcription factor among vertebrates. In drosophila, the homolog of FoxP2, FoxP has been shown to be responsible for pulse-song structure and sex-specific walking and flight (Lawton et al., 2014). In mice, when the FoxP2 is knocked-out, infant mice will produce abnormal ultrasonic vocalizations (Shu et al., 2005). The heterozygous mutations of Foxp2 impair their sensorimotor association learning (Kurt, Fisher & Ehret, 2012). Zbtb20, a repressor of FoxP2, was reported to bind to and repress cortical layer marker genes (including Foxp2) in the developing hippocampus (Nielsen et al., 2014). In birds, the knockdown of Foxp2 in Area X in juvenile zebra finches affects the accuracy and completeness of the produced song (Haesler et al., 2007).

Human FOXP2 incorporated two fixed amino acid changes in a broadly defined transcription suppression domain (Zhang et al., 2002). These two amino acid changes (N325S, T303N) occurred in the short timescale of the evolutionary split from the lineage of chimpanzees (Enard et al., 2002). When injected with a humanized version of FOXP2, mice showed a reduced dopamine level, increased dendritic length and long-term synaptic depression (Enard et al., 2009), suggesting the role of human FOXP2 in altering the basal ganglia function. Moreover, humanized version of FOXP2 in mice also show accelerated transition between declarative learning to procedural learning (Schreiweis et al., 2014), the neural basis of which are the hippocampus and basal ganglia respectively, suggesting a key role of FOXP2 in better connecting the basal ganglia and hippocampus.

Index Terms: multimodal origin of communication, comparative perspective, basal ganglia, hippocampus, FOXP2

References

Dalziell, A. H., Peters, R. A., Cockburn, A., Dorland, A. D., Maisey, A. C., & Magrath, R. D. (2013). Dance choreography is coordinated with song repertoire in a complex avian display. *Current Biology*, 23(12), 1132–1135.

Enard, W., Przeworski, M., Fisher, S. E., Lai, C. S. L., Wiebe, V., Kitano, T., … Pääbo, S. (2002). Molecular evolution of FOXP2, a gene involved in speech and language. *Nature*, 418(6900), 869–872.

Enard, W., Gehre, S., Hammerschmidt, K., Hölter, S. M., Blass, T., Somel, M., … Pääbo, S. (2009). A Humanized Version of Foxp2 Affects Cortico-Basal Ganglia Circuits in Mice. *Cell*, 137(5), 961–971.

Ewing, A. W. (1983). Functional Aspects of Drosophila Courtship. Biological Reviews, 58(2), 275–292.

Fröhlich, M., & van Schaik, C. P. (2018). The function of primate multimodal communication. Animal Cognition, 21, 619-629.

Genty, E., Clay, Z., Hobaiter, C., & Zuberbühler, K. (2014). Multi-modal use of a socially directed call in bonobos. *PloS one*, 9(1), e84738.

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(12), 2885–2897.

Hilliard, C., Cook, S. W., & Duff, M. C. (2016). Hippocampal declarative memory supports gesture production: Evidence from amnesia. *Cortex*, 85, 25–36.

Jarvis, E. D. (2019). Evolution of vocal learning and spoken language. *Science*, 366(6461), 50–54.

Kurt, S., Fisher, S. E., & Ehret, G. (2012). Foxp2 mutations impair auditory-motor association learning. *PloS one*, 7(3), e33130.

Lawton, K. J., Wassmer, T. L., & Deitcher, D. L. (2014). Conserved role of Drosophila melanogaster FoxP in motor coordination and courtship song. *Behavioural Brain Research*, 268, 213–221.

Levinson, S. C. (2023). Gesture, spatial cognition and the evolution of language. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 378(1875), 20210481.

Levinson, S. C., & Holler, J. (2014). The origin of human multi-modal communication. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369(1651), 20130302. https://doi.org/http://dx.doi.org/10.1098/rstb.2013.0302

Liebal, K., Pika, S., & Tomasello, M. (2004). Social communication in siamangs (Symphalangus syndactylus): Use of gestures and facial expressions. *Primates*, 45(1), 41–57.

Liebal, K., Waller, B. M., Burrows, A. M., & Slocombe, K. E. (2014). *Primate Communication: A Multimodal Approach*. Cambridge University Press.

Macuch Silva, V., Holler, J., Ozyurek, A., & Roberts, S. G. (2020). Multimodality and the origin of a novel communication system in face-to-face interaction. *Royal Society Open Science*, 7(1), 182056.

Micheletta, J., Engelhardt, A., Matthews, L., Agil, M., & Waller, B. M. (2013). Multicomponent and multimodal lip smacking in crested macaques (Macaca nigra). *American Journal of Primatology*, 75(7), 763–773.

Nielsen, J. V., Thomassen, M., Møllgård, K., Noraberg, J., & Jensen, N. A. (2014). Zbtb20 defines a hippocampal neuronal identity through direct repression of genes that control projection neuron development in the isocortex. *Cerebral Cortex*, 24(5), 1216-1229.

Nölle, J. & Galantucci, B. (2022). Experimental semiotics: past, present and future. In A. M. García & A. Ibánez (Eds.), *Routledge handbook of semiotics and the brain*. Routledge

Pfenning, A. R., Hara, E., Whitney, O., Rivas, M. V., Wang, R., Roulhac, P. L., Howard, J. T., Wirthlin, M., Lovell, P. V., Ganapathy, G., Mouncastle, J., Moseley, M. A., Thompson, J. W., Soderblom, E. J., Iriki, A., Kato, M., Gilbert, M. T. P., Zhang, G., Bakken, T., … Jarvis, E. D. (2014). Convergent transcriptional specializations in the brains of humans and song-learning birds. *Science*, 346(6215), 1256846–1256846. https://doi.org/10.1126/science.1256846

Pika, S., & Bugnyar, T. (2011). The use of referential gestures in ravens (Corvus corax) in the wild. *Nature Communications*, 2, 560.

Schreiweis, C., Bornschein, U., Burguière, E., Kerimoglu, C., Schreiter, S., Dannemann, M., … Graybiel, A. M. (2014). Humanized Foxp2 accelerates learning by enhancing transitions from declarative to procedural performance. *Proceedings of the National Academy of Sciences*, 111(39), 14253–14258.

Shu, W., Cho, J. Y., Jiang, Y., Zhang, M., Weisz, D., Elder, G. A., … Buxbaum, J. D. (2005). Altered ultrasonic vocalization in mice with a disruption in the Foxp2 gene. *Proceedings of the National Academy of Sciences of the United States of America*, 102(27), 9643–8.

Taglialatela, J. P., Russell, J. L., Pope, S. M., Morton, T., Bogart, S., Reamer, L. A., … Hopkins, W. D. (2015). Multimodal communication in chimpanzees. *American Journal of Primatology*, 77(11), 1143–1148.

Tavolga, W. N. (1956). Visual, chemical and sound stimuli as cues in the sex discriminatory behavior of the gobiid fish Bathygobius soporator. *Zoologica*, 41(2), 49-64.

Vigliocco, G., Perniss, P., & Vinson, D. (2014). Language as a multimodal phenomenon: implications for language learning, processing and evolution. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369, 20130292. https://doi.org/10.1098/rstb.2013.0292

Williams, H. (2001). Choreography of song, dance and beak movements in the zebra finch (Taeniopygia guttata). *The Journal of Experimental Biology*, 204(Pt 20), 3497–506.

Zhang, J., Webb, D. M., & Podlaha, O. (2002). Accelerated protein evolution and origins of human-specific features: FOXP2 as an example. *Genetics*, 162(4), 1825–1835.

Zlatev, J., Żywiczyński, P., & Wacewicz, S. (2020). Pantomime as the original human-specific communicative system. *Journal of Language Evolution*, 5(2), 156-174.