FROM WHERE TO WHAT: A NEUROANATOMICALLY BASED EVOLUTIONARY MODEL OF THE EMERGENCE OF SPEECH IN HUMANS

OREN POLIVA

polivaoren@hotmail.com

Psychology department, Bangor University, Bangor, United Kingdom

In the brain of primates, the auditory cortex connects with the frontal lobe via the temporal pole (auditory ventral stream; AVS) and via the inferior parietal lobe (auditory dorsal stream; ADS). The AVS is responsible for sound recognition, and the ADS for sound-localization, voice detection and integration of calls with faces. I propose that the primary role of the ADS in non-human primates is the detection and response to contact calls. These calls are exchanged between tribe members (e.g., mother-offspring) and are used for monitoring location. Detection of contact calls occurs by the ADS identifying a voice, localizing it, and verifying that the corresponding face is out of sight. Once a contact call is detected, the primate produces a contact call in return via descending connections from the frontal lobe to brainstem motor nuclei. Because the ADS of present day humans also performs speech production and repetition, I further propose an evolutionary course for the transition from contact call exchange to an early form of speech. In accordance with this model, structural changes to the ADS endowed early members of the genus Homo with partial vocal control. This development was beneficial as it enabled offspring to modify their contact calls with intonations for signaling high or low levels of distress to their mother. Eventually, individuals were capable of participating in yes-no question-answer conversations. In these conversations the offspring emitted a low-level distress call for inquiring about the safety of objects (e.g., food), and the mother responded with a high- or low-level distress call to signal approval or disapproval of the interaction. Gradually, the ADS and its connections with brainstem motor regions became more robust and vocal control became more volitional. Speech emerged once vocal control was sufficient for inventing novel calls.

1. Introduction

In the past five decades, gorillas, orangutans, chimpanzees and bonobos were shown capable of learning sign language (Blake, 2004; Gibson, 2011). An important cognitive distinction between the language used by humans and the language used by other apes is with the ability to ask questions. This was first noted by (Premack & Premack, 1984) who reported that, although their chimpanzee, Sarah, showed no difficulty answering questions or repeating questions before answering them, she never used the question signs for inquiring about her own environment. Jordania (2006), in his review of the literature, noted that other signing apes did not utilize questions and that their initiation of conversations was limited to commands (e.g., "me more eat") and observational statements (e.g., "bird there"). This absence of a

This paper is distributed under a Creative Commons CC-BY-ND license.

DOI:10.12775/3991-1.094

questioning mind is in direct contrast to human toddlers and children, who are renown for their incessant use of questions. My interpretation of this human-ape distinction is that during human evolution, we transitioned from the display of curiosity toward items that are present in our environment (i.e., observational statements) to curiosity toward items that are absent in our environment (i.e., WH questions). Developing curiosity about out of sight events and objects could thus explain the rapid migration of humans across the globe. Furthermore, this curiosity toward the unknown is the driving force behind scientific exploration and technological development. One could hence argue that it is the ability to ask that separates us from other animals and makes the human species unique.

Although no non-human primate has been reported to ask questions, they were reported to exchange calls for monitoring location (i.e., contact calls). For example, when a mother and her infant are physically separated, each emits in turn a call to signal the other their location. This emission of contact calls could therefore be interpreted as akin in meaning to the question "where are you?". If human communication and contact calls are related, it suggests that the preliminary urge to learn about the unknown is derived from infants and mothers seeking to reunite.



Figure 1. Dual stream connectivity between the auditory cortex and frontal lobe of monkeys and humans. The brain of the monkey (left) and human (right) is schematically depicted and displayed from the side. On the left panel of each species, the auditory cortex is schematically depicted on the supratemporal plane (orange frame in the right panel) and observed from above (with the parieto-frontal operculi removed). Blue colors mark regions affiliated with the ADS, and red colors mark regions affiliated with the AVS.

2. Two pathways of auditory processing

In the brains of primates, two pathways connect the auditory cortex with the frontal lobe (Figure 1): In the auditory ventral stream (AVS), the anterior superior temporal gyrus/sulcus connects with the middle temporal gyrus/temporal pole, which further connects with the inferior frontal gyrus (IFG). This pathway is well known for its

role in recognizing speech and sounds in both hemispheres (Dronkers et al., 2004; Hickok et al., 2007; Gow, 2012), and damage to its structures has been associated with the disorders auditory agnosia (Poliva et al., 2015) and semantic dementia (Noppeney et al., 2007). The auditory dorsal stream (ADS) connects the posterior auditory cortex (pSTG) with the Sylvian parietal-temporal junction (Spt) and inferior parietal lobe (IPL)/intra-parietal sulcus (IPS), which also connects with the IFG. As will be detailed below, in non-human primates this pathway is involved in sound localization, speaker identification and integration of calls with faces. In the left hemisphere of the human brain, this pathway is also responsible for speech production and repetition, perception and production of prosody and integration of spoken words with their facial movements. Herein, I argue that the ADS original role was for exchanging contact calls and that it was modification of this pathway that first enabled early members of the species homo (i.e., Hominans) to be able to ask questions, and later develop language.

3. The role of the ADS with the detection of contact calls

Direct evidence for involvement of the ADS in the detection and response to contact calls is demonstrated in a study that sacrificed marmoset monkeys and measured genomic expression of cFos protein. Activation was reported only in the ADS (areas CM-CL and IFG) while hearing and responding to contact calls (phee calls; Miller *et al.*, 2010). Monkeys sacrificed after only hearing contact calls or only emitting them showed neural activity in the same regions but to a much smaller degree (See also Simões *et al.*, 2010 for similar results in a study using the protein Egr-1).

The detection of contact calls likely occurs in 3 stages: 1. Identify the speaker 2. localize the direction of the call 3. Confirm that the face of the speaker is not present in that direction. Consistent with a role of the ADS in the detection of contact calls, the ADS has been implicated with all these functions. Recordings of neural activity from the auditory cortex of monkeys showed selectivity in the posterior, but not anterior auditory cortex, of sound localization (Benson *et al.*, 1981; Miller & Recanzone, 2009; Rauschecker *et al.*, 1995; Tian *et al.*, 2001; Woods *et al.*, 2006). Functional imaging studies in humans also correlated ADS' regions (IPL, IFG) with sound localization (Ahveninen *et al.*, 2006; Barrett & Hall, 2006; Brunetti *et al.*, 2005; De Santis *et al.*, 2007; Hart *et al.*, 2004; Krumbholz *et al.*, 2005; Warren & Griffiths, 2003). A role of the ADS in speaker identification was shown in an intra-cortical recording study of epileptic patients that associated activation in the pSTG with speaker discrimination (Lachaux *et al.*, 2007). Recording from the posterior insula (near pSTG) of monkeys also reported of activity during speaker

discrimination (Remedios *et al.*, 2009). Functional imaging of third trimester human fetuses demonstrated activation in area Spt when the hearing of voices was contrasted to pure tones, and of a sub-region of area Spt that is more selective to maternal voice than unfamiliar female voices (Jardri *et al.*, 2012). The role of the ADS in face-call integration is evident in an fMRI monkey study that reported activation only in the posterior auditory cortex during audio-visual stimuli integration (Kayser *et al.*, 2009). A study that recorded from neurons in A1 also showed that the activation is weaker while hearing contact calls than social calls (grunt call; Ghazanfar *et al.*, 2005). In humans, the neighboring pSTS is responsible integrating speech with lip-movements (McGurk effect; Beauchamp et al., 2010; Nath et al., 2010; Campbell, 2008).

In contrast to most calls, which are produced in limbic regions (Holstege et al., 2015), the IFG appears involved in the emission of contact calls. A study that recorded neural activity from the IFG of macaques reported of neural discharge prior to cued or spontaneous contact call production (coo calls), but not prior to production of vocalizations-like facial movements (i.e., silent vocalizations; Coudé *et al.*, 2011; see also Gemba *et al.*, 1999 for similar results). A tracing study in monkeys also reported of connections between IFG and brainstem motor nuclei (Jürgens & Alipour, 2002), which potentially could endow the animal with partial vocal control. Accordingly, behavioral studies of monkeys reported of ability to modify contact calls with intonations (Biben *et al.*, 1986; Sugiura, 1998) or of learning to use them in order to request objects (Hihara *et al.*, 2003). Anecdotal reports of more generalized volitional vocal control, albeit rudimentary, were also reported in apes (Hayes & Hayes, 1952; Hopkins *et al.*, 2007; Koda *et al.*, 2007; Koda *et al.*, 2015; Taglialatela *et al.*, 2003; Wich *et al.*, 2008).

4. The ADS and the emergence of speech

A role for the ADS in the emergence of human speech is implied in the fossil record. A study that reconstructed the endocranium of early *Hominins* noted that *Homo habilis*, but not any of its *Australopith* ancestors, is characterized by enlargement of the IPL and IFG, whereas the rest of the endocranium remains extremely similar to the endocranium of modern apes (Tobias, 1987). A diffusion tensor imaging study that compared the white matter of humans and chimpanzees also reported of significant strengthening in ADS connectivity, but not AVS connectivity (Rilling *et al.*, 2012). Evidence for a role of the ADS in the transition from mediating contact calls into mediating human speech also includes genetic studies that focused on

mutation to the protein SRPX2 and its regulator protein FOXP2 (Roll et al., 2010). In mice, blockage of SRPX2 or FOXP2 genes resulted in pups not emitting distress calls when separated from their mothers (Shu et al., 2005; Sia et al., 2013). In humans, however, individuals afflicted with a mutated SRPX2 or FOXP2 were reported with speech dyspraxia (Roll et al., 2006; Watkins et al., 2002). A PET imaging study of an individual with a mutated SRPX2 gene correlated this patient's disorder with abnormal activation (hyper-metabolism) along the ADS (pSTG-Spt-IPL; Roll et al., 2006). Similarly, an MRI study that scanned individuals with mutated FOXP2 reported increased grey matter density in the pSTG-Spt and reduced density in the IFG, thus further demonstrating abnormality in ADS' structures (Belton et al., 2003). A role for the ADS in mediating speech production in humans has also been demonstrated in studies that correlated a more severe variant of this disorder, apraxia of speech, with IPL and IFG lesions (Deutsch, 1984; Edmonds & Marquardt, 2004; Hillis et al., 2004; Josephs et al., 2006; Kimura & Watson, 1989; Square et al., 1997). The role of the ADS in speech production is also demonstrated via a series of studies that directly stimulated sub-cortical fibers during surgical operations (Duffau, 2008-review), and reported that interference in the left pSTG and IPL resulted in an increase in speech production errors, and interference in the left IFG resulted in speech arrest (see also Acheson et al., 2011; Stewart et al., 2001 for similar results using magnetic interference in healthy individuals). Damage to Spt-IPL was also correlated with impaired ability to repeat sentences and unfamiliar words (conduction aphasia; Selnes et al., 1985; Axer et al., 2001; Bartha and Benke, 2003; Dronkers et al., 2004; Baldo et al., 2008, 2012; Fridriksson et al., 2010; Buchsbaum et al., 2011). One study even reported that stimulation of the left IPL resulted with patients believing that they spoke, when they didn't, and IFG stimulation resulted with the patients unconsciously moving their lips (Desmurget et al., 2009).

Further support for the transition from contact call exchange to human speech are provided by studies of hemispheric lateralization (Petersen *et al.*, 1978). In one study, Japanese macaques and other old world monkeys were trained to discriminate contact calls of Japanese macaques, which were presented to the right or left ear. Although all the monkeys were capable of completing the task, only the Japanese macaques were noted with right ear advantage, thus indicating left hemispheric processing of contact calls. In a study replicating the same paradigm, Japanese macaques had an impaired ability to discriminate contact calls after suffering unilateral damage to the auditory cortex of the left, but not right, hemisphere (Heffner & Heffner, 1984). This leftward lateralization of contact call detection is

similar to the long established role of the human left hemisphere in the processing human language (Geschwind, 1965).

5. From vocal control to the first conversation

A possible route for the transition from contact call exchange to proto-speech was proposed by Dean Falk (2004). She argued that due to bipedal locomotion and the loss of hair in early Hominins, mothers were not capable of carrying their infants while foraging. As a result, the mothers maintained contact with their infant through a vocal exchange of calls that resembles contemporary "motherese" (the unique set of intonations that caregivers use when addressing infants). As previously suggested by another researcher (Masataka, 2009), such intermediate prosodic phase in the development of speech is consistent with evidence presented above that monkeys, to a limited extent, are capable of modifying their contact calls with intonations, and that apes are endowed with slightly more versatile vocal control. In the opening paragraph of this paper, I described the inability of apes to ask questions, and proposed that the ability to ask questions emerged from contact calls. Because the ability to modify calls with prosodic intonations likely further developed as we began conversing with questions, I expand Falk's and Masataka's views regarding the prosodic origins of vocal language, and propose that the transition from contact calls to prosodic intonations could have emerged as a means of enabling infants to express different levels of distress. In such a scenario, the modification of a call with intonations designed to express a high level of distress is akin in meaning to the sentence "mommy, come here now!". Hence, the modification of calls with intonations could have served as a precursor for the development of prosody in contemporary vocal commands. On the other hand, the use of intonations for expressing a low-level of distress is akin in meaning to the sentence "mommy, where are you?". Therefore, this use of prosody for asking the first question could have served as the precursor for pragmatically converting calls into questions by using prosody as well. This transition could be related to the ability of present-day infants of using intonations for changing the pragmatic utilization of a word from a statement to a command/demand ("MOMMY!") or a question ("mommy?"). This view is consistent with a longitudinal developmental study of toddlers, which reported of the toddlers utilizing prosodic intonations in their speech prior to construction of sentences (Snow, 1994). A study of speech perception in adults also demonstrated that our ability to discriminate questions sentences from statement sentences is dependent on analysis of prosodic intonations (Srinivasan et al., 2003). Evidence of the relationship between asking questions and processing in the ADS is demonstrated in a diffused tensor imaging and fMRI study (Sammler *et al.*, 2015), which reported the participation of both the ADS and AVS in the discrimination of mono-syllabic words into questions or statements. The researchers further showed that this discrimination was impaired while interference was induced with TMS in the pre-motor cortex of the ADS. Supporting the role of the ADS in the discrimination of questions and statements is the finding that patients with phonological dementia, who are known to suffer from degeneration along the ADS (Gorno-Tempini *et al.*, 2008; Rohrer *et al.*, 2010), were impaired in distinguishing whether a spoken word was a question or a statement (Rohrer *et al.*, 2012).

A possible route for the transition from emitting low-level distress calls to asking questions is by individuals starting to utilize the former to signal interest about objects in their environment. Given that both contact call exchange and contemporary speech are characterized with turn taking, early *Hominans* could have responded to the low-level distress calls with either high- or low-level distress calls. For example, when an infant expressed a low-level distress call prior to eating berries, his/her mother could have responded with a high-level distress call that indicated the food is dangerous or a low-level distress call that indicated the food is safe. Eventually, the infant emitted the question call and waited for an appropriate answer from their mother before proceeding with their intended action. This vocal exchange was most likely the precursor to yes-no question-answer conversations.

References

- Acheson, D. J., Hamidi, M., Binder, J. R., & Postle, B. R. (2011). A common neural substrate for language production and verbal working memory. *Journal of Cognitive Neuroscience* 23, 1358–1367. doi:10.1162/jocn.2010.21519.
- Ahveninen, J., Jaaskelainen, I. P., Raij, T., Bonmassar, G., Devore, S., Hämäläinen, M., et al. (2006). Task-modulated "what" and 'where' pathways in human auditory cortex. *Proc. Natl. Acad. Sci. U.S.A.* 103, 14608–14613.
- Axer, H., Keyserlingk, von, A. G., Berks, G., & Keyserlingk, von, D. G. (2001). Supra- and Infrasylvian Conduction Aphasia. *Brain and Language* 76, 317– 331. doi:10.1006/brln.2000.2425.
- Baldo, J. V., Katseff, S., & Dronkers, N. F. (2012). Brain regions underlying repetition and auditory-verbal short-term memory deficits in aphasia: Evidence from voxel-based lesion symptom mapping. *Aphasiology* 26, 338–354. doi: 10.1080/02687038.2011.602391.
- Baldo, J. V., Klostermann, E., & Dronkers, N. F. (2008). It's either a cook or a baker: Patients with conduction aphasia get the gist but lose the trace. *Brain and Language* 105, 134–140. doi:10.1016/j.bandl.2007.12.007.
- Barrett, D. J. K., & Hall, D. A. (2006). Response preferences for "what" and 'where'

in human non-primary auditory cortex. *NeuroImage* 32, 968–977. doi:10.1016/j.neuroimage.2006.03.050.

- Bartha, L., & Benke, T. (2003). Acute conduction aphasia: An analysis of 20 cases. *Brain and Language* 85, 93–108. doi:10.1016/S0093-934X(02)00502-3.
- Beauchamp, M. S., Nath, A. R., and Pasalar, S. (2010). fMRI-Guided Transcranial Magnetic Stimulation Reveals That the Superior Temporal Sulcus Is a Cortical Locus of the McGurk Effect. Journal of Neuroscience 30, 2414–2417. doi: 10.1523/JNEUROSCI.4865-09.2010.
- Belton, E., Salmond, C. H., Watkins, K. E., Vargha-Khadem, F., & Gadian, D. G. (2003). Bilateral brain abnormalities associated with dominantly inherited verbal and orofacial dyspraxia. *Hum. Brain Mapp.* 18, 194–200. doi:10.1002/ hbm.10093.
- Benson, D. A., Hienz, R. D., & Goldstein, M. H. (1981). Single-unit activity in the auditory cortex of monkeys actively localizing sound sources: spatial tuning and behavioral dependency. *Brain Res.* 219, 249–267.
- Biben, M., Symmes, D., & Masataka, N. (1986). Temporal and structural analysis of affiliative vocal exchanges in squirrel monkeys (Saimiri sciureus). *Behaviour*, 259–273.
- Blake, J. (2004). Gestural communication in the great apes. In The Evolution of Thought: Evolutionary Origins of Great Ape Intelligence. Cambridge University Press.
- Brunetti, M., Belardinelli, P., Caulo, M., Del Gratta, C., Penna, Della, S., Ferretti, A., et al. (2005). Human brain activation during passive listening to sounds from different locations: An fMRI and MEG study. *Hum. Brain Mapp.* 26, 251–261. doi:10.1002/hbm.20164.
- Buchsbaum, B. R., Baldo, J., Okada, K., Berman, K. F., Dronkers, N., D'Esposito, M., et al. (2011). Conduction aphasia, sensory-motor integration, and phonological short-term memory - An aggregate analysis of lesion and fMRI data. *Brain and Language* 119, 119–128. doi:10.1016/j.bandl.2010.12.001.
- Campbell, R. (2008). The processing of audio-visual speech: empirical and neural bases. Philosophical Transactions of the Royal Society B: Biological Sciences 363, 1001–1010. doi:10.1093/cercor/13.10.1034.
- Coudé, G., Ferrari, P. F., Rodà, F., Maranesi, M., Borelli, E., Veroni, V., et al. (2011). Neurons Controlling Voluntary Vocalization in the Macaque Ventral Premotor Cortex. *PLoS ONE* 6, e26822. doi:10.1371/journal.pone.0026822.s004.
- De Santis, L., Clarke, S., & Murray, M. M. (2006). Automatic and Intrinsic Auditory "What" and "Where" Processing in Humans Revealed by Electrical Neuroimaging. *Cerebral Cortex* 17, 9–17. doi:10.1093/cercor/bhj119.
- Desmurget, M., & Sirigu, A. (2009). A parietal-premotor network for movement intention and motor awareness. *Trends Cogn. Sci. (Regul. Ed.)* 13, 411–419. doi:10.1016/j.tics.2009.08.001.
- Deutsch, S. E. (1984). Prediction of site of lesion from speech apraxic error patterns. In *Apraxia of speech: Physiology, acoustics, linguistics, management.* 113–134.

- Dronkers, N. F., Wilkins, D. P., Van Valin, R. D., redfern, B. B., & Jaeger, J. J. (2004). Lesion analysis of the brain areas involved in language comprehension. *Cognition* 92, 145–177. doi:10.1016/j.cognition.2003.11.002.
- Duffau, H. (2008). The anatomo-functional connectivity of language revisited. *Neuropsychologia* 46, 927–934. doi:10.1016/j.neuropsychologia.2007.10.025.
- Edmonds, L., & Marquardt, T. (2004). Syllable use in apraxia of speech: Preliminary findings. *Aphasiology* 18, 1121–1134. doi:10.1080/02687030444000561.
- Falk, D. (2004). Prelinguistic evolution in early hominins: whence motherese? *Behav. Brain Sci.* 27, 491–503– discussion 503–83.
- Fridriksson, J., Kjartansson, O., Morgan, P. S., Hjaltason, H., Magnusdottir, S., Bonilha, L., et al. (2010). Impaired Speech Repetition and Left Parietal Lobe Damage. *Journal of Neuroscience* 30, 11057–11061. doi:10.1523/JNEUROSCI. 1120-10.2010.
- Gemba, H., Kyuhou, S., Matsuzaki, R., & Amino, Y. (1999). Cortical field potentials associated with audio-initiated vocalization in monkeys. *Neuroscience Letters* 272, 49–52.
- Geschwind, N. (1965). Disconnexion syndromes in animals and man. I. *Brain* 88, 237–294.
- Ghazanfar, A. A. (2005). Multisensory Integration of Dynamic Faces and Voices in Rhesus Monkey Auditory Cortex. *Journal of Neuroscience* 25, 5004–5012. doi: 10.1523/JNEUROSCI.0799-05.2005.
- Gibson, K. R. (2011). Language or protolanguage? A review of the ape language literature. In The Oxford Handbook of Language Evolution. Oxford University Press, USA.
- Gorno Tempini, M. L., Brambati, S. M., Ginex, V., Ogar, J., Dronkers, N. F., Marcone, A., et al. (2008). The logopenic/phonological variant of primary progressive aphasia. *Neurology* 71, 1227–1234. doi:10.1212/01.wnl. 0000320506.79811.da.
- Gow, D. W., Jr (2012). The cortical organization of lexical knowledge: A dual lexicon model of spoken language processing. *Brain and Language* 121, 273– 288. doi:10.1016/j.bandl.2012.03.005.
- Hart, H. C., Palmer, A. R., & Hall, D. A. (2004). Different areas of human nonprimary auditory cortex are activated by sounds with spatial and nonspatial properties. *Hum. Brain Mapp.* 21, 178–190. doi:10.1002/hbm.10156.
- Hayes, C., & Hayes, K. J. (1952). Imitation in a home-raised chimpanzee. J Comp Physiol Psychol 45, 450–459.
- Heffner, H. E., & Heffner, R. S. (1984). Temporal lobe lesions and perception of species-specific vocalizations by macaques. *Science* 226, 75–76.
- Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience* 8, 393–402. doi:10.1038/nrn2113.
- Hihara, S., Yamada, H., Iriki, A., & Okanoya, K. (2003). Spontaneous vocal differentiation of coo-calls for tools and food in Japanese monkeys. *Neuroscience Research* 45, 383–389. doi:10.1016/S0168-0102(03)00011-7.

- Hillis, A. E., Work, M., Barker, P. B., Jacobs, M. A., Breese, E. L., & Maurer, K. (2004). Re-examining the brain regions crucial for orchestrating speech articulation. *Brain* 127, 1479–1487. doi:10.1093/brain/awh172.
- Holstege, G., and Subramanian, H. H. (2015). Two different motor systems are needed to generate human speech. J. Comp. Neurol. 524, 1558–1577. doi: 10.1016/0304-3940(92)90692-Z.
- Hopkins, W. D., Taglialatela, J. P., & Leavens, D. A. (2007). Chimpanzees differentially produce novel vocalizations to capture the attention of a human. *Animal Behaviour* 73, 281–286. doi:10.1016/j.anbehav.2006.08.004.
- Jardri, R., Houfflin-Debarge, V., Delion, P., Pruvo, J.-P., Thomas, P., & Pins, D. (2012). Assessing fetal response to maternal speech using a noninvasive functional brain imaging technique. *International Journal of Developmental Neuroscience* 30, 159–161. doi:10.1016/j.ijdevneu.2011.11.002.
- Jordania, J. (2006). Who Asked the First Question? *The Origins of Human Choral Singing, Intelligence, Language and Speech*, p. 334–338.
- Josephs, K. A. (2006). Clinicopathological and imaging correlates of progressive aphasia and apraxia of speech. *Brain* 129, 1385–1398. doi:10.1093/brain/ awl078.
- Kayser, C., Petkov, C. I., & Logothetis, N. K. (2009). Multisensory interactions in primate auditory cortex: fMRI and electrophysiology. *Hear. Res.* 258, 80–88. doi:10.1016/j.heares.2009.02.011.
- Kimura, D., & Watson, N. (1989). The relation between oral movement control and speech. *Brain and Language* 37, 565–590.
- Koda, H., Nishimura, T., Tokuda, I. T., Oyakawa, C., Nihonmatsu, T., and Masataka, N. (2012). Soprano singing in gibbons. *Am. J. Phys. Anthropol.* 149, 347–355. doi:10.1002/ajpa.22124.
- Koda, H., Oyakawa, C., Kato, A., & Masataka, N. (2007). Experimental evidence for the volitional control of vocal production in an immature gibbon. *Behaviour*, 681–692.
- Krumbholz, K., Schönwiesner, M., Rübsamen, R., Zilles, K., Fink, G. R., and Cramon, von, D. Y. (2005). Hierarchical processing of sound location and motion in the human brainstem and planum temporale. *European Journal of Neuroscience* 21, 230–238. doi:10.1111/j.1460-9568.2004.03836.x.
- Lachaux, J.-P., Jerbi, K., Bertrand, O., Minotti, L., Hoffmann, D., Schoendorff, B., et al. (2007). A blueprint for real-time functional mapping via human intracranial recordings. *PLoS ONE* 2, e1094. doi:10.1371/journal.pone.0001094.
- Lameira, A. R., Hardus, M. E., Bartlett, A. M., Shumaker, R. W., Wich, S. A., & Menken, S. B. J. (2015). Speech-Like Rhythm in a Voiced and Voiceless Orangutan Call. *PLoS ONE* 10, e116136. doi:10.1371/journal.pone. 0116136.s003.
- Laporte, M. N. C., & Zuberbühler, K. (2010). Vocal greeting behaviour in wild chimpanzee females. *Animal Behaviour* 80, 467–473. doi:10.1016/j.anbehav. 2010.06.005.

- Masataka, N. (2009). The origins of language and the evolution of music: A comparative perspective. *Physics of Life Reviews* 6, 11–22. doi:10.1016/j.plrev. 2008.08.003.
- Miller, C. T., Mandel, K., & Wang, X. (2010). The communicative content of the common marmoset phee call during antiphonal calling. *Am. J. Primatol.* 72, 974–980. doi:10.1002/ajp.20854.
- Miller, L. M., & Recanzone, G. H. (2009). Populations of auditory cortical neurons can accurately encode acoustic space across stimulus intensity. *Proc. Natl. Acad. Sci. U.S.A.* 106, 5931–5935.
- Nath, A. R., and Beauchamp, M. S. (2012). A neural basis for interindividual differences in the McGurk effect, a multisensory speech illusion. NeuroImage 59, 781–787. doi:10.1016/j.neuroimage.2011.07.024.
- Noppeney, U., Josephs, O., Hocking, J., Price, C. J., & Friston, K. J. (2007). The Effect of Prior Visual Information on Recognition of Speech and Sounds. *Cerebral Cortex* 18, 598–609. doi:10.1093/cercor/bhm091.
- Perlman, M., & Clark, N. (2015). Learned vocal and breathing behavior in an enculturated gorilla. Anim Cogn 18, 1165-1179. doi:10.1007/ s10071-015-0889-6.
- Petersen, M. R., Beecher, M. D., Zoloth, S. R., Moody, D. B., & Stebbins, W. C. (1978). Neural lateralization of species-specific vocalizations by Japanese macaques (Macaca fuscata). *Science* 202, 324–327.
- Poliva, O., Bestelmeyer, P. E. G., Hall, M., Bultitude, J. H., Koller, K., & Rafal, R. D. (2015). Functional Mapping of the Human Auditory Cortex: fMRI Investigation of a Patient with Auditory Agnosia from Trauma to the Inferior Colliculus. *Cogn Behav Neurol* 28, 160–180. doi:10.1097/WNN. 000000000000072.
- Premack, D., & Premack, A. J. (1984). The Mind of an Ape. W. W. Norton.
- Rauschecker, J. P., Tian, B., & Hauser, M. (1995). Processing of complex sounds in the macaque nonprimary auditory cortex. *Science*.
- Remedios, R., Logothetis, N. K., & Kayser, C. (2009). An Auditory Region in the Primate Insular Cortex Responding Preferentially to Vocal Communication Sounds. *Journal of Neuroscience* 29, 1034–1045. doi:10.1523/JNEUROSCI. 4089-08.2009.
- Rilling, J. K., Glasser, M. F., Jbabdi, S., Andersson, J., and Preuss, T. M. (2011). Continuity, divergence, and the evolution of brain language pathways. Frontiers in evolutionary neuroscience 3. doi:10.3389/fnevo.2012.00011/abstract.
- Rohrer, J. D., Ridgway, G. R., Crutch, S. J., Hailstone, J., Goll, J. C., Clarkson, M. J., et al. (2010). Progressive logopenic/phonological aphasia: Erosion of the language network. *NeuroImage* 49, 984–993. doi:10.1016/j.neuroimage. 2009.08.002.
- Rohrer, J. D., Sauter, D., Scott, S., Rossor, M. N., & Warren, J. D. (2012). Receptive prosody in nonfluent primary progressive aphasias. *Cortex* 48, 308–316. doi: 10.1016/j.cortex.2010.09.004.

- Roll, P., Rudolf, G., Pereira, S., Royer, B., Scheffer, I. E., Massacrier, A., et al. (2006). SRPX2 mutations in disorders of language cortex and cognition. *Human Molecular Genetics* 15, 1195–1207. doi:10.1093/hmg/ddl035.
- Roll, P., Vernes, S. C., Bruneau, N., Cillario, J., Ponsole-Lenfant, M., Massacrier, A., et al. (2010). Molecular networks implicated in speech-related disorders: FOXP2 regulates the SRPX2/uPAR complex. *Human Molecular Genetics* 19, 4848–4860. doi:10.1093/hmg/ddq415.
- Sammler, D., Grosbras, M.-H., Anwander, A., Bestelmeyer, P. E. G., and Belin, P. (2015). Dorsal and Ventral Pathways for Prosody. CURBIO 25, 3079–3085. doi:10.1016/j.cub.2015.10.009.
- Selnes, O. A., Knopman, D. S., Niccumm, N., & Rubens, A. B. (1985). The critical role Wernicke's area in sentence repetition. *Ann. Neurol.* 17, 549–557.
- Shu, W., Cho, J. Y., Jiang, Y., Zhang, M., Weisz, D., Elder, G. A., et al. (2005). Altered ultrasonic vocalization in mice with a disruption in the Foxp2 gene. *Proc. Natl. Acad. Sci. U.S.A.* 102, 9643–9648. doi:10.1073/pnas.0503739102.
- Sia, G. M., Clem, R. L., & Huganir, R. L. (2013). The Human Language-Associated Gene SRPX2 Regulates Synapse Formation and Vocalization in Mice. *Science*. doi:10.1126/science.1245079.
- Simões, C. S. (2010). Activation of frontal neocortical areas by vocal production in marmosets. *Front.Integr.Neurosci.* 4. doi:10.3389/fnint.2010.00123.
- Snow, D. (1994). Phrase-final syllable lengthening and intonation in early child speech. J. Speech Hear Res.; 37(4): 831–840.
- Square, P. A., Roy, E. A., & Martin, R. E. (1997). Apraxia of speech: Another form of praxis disruption. In *Apraxia: The neuropsychology of action*. 173–206.
- Srinivasan, R.J., Massaro, D.W. (2003). Perceiving prosody from the face and voice: distinguishing statements from echoic questions in English. *Lang Speech*, 46(Pt 1): 1–22.
- Stewart, L., Walsh, V., Frith, U., & Rothwell, J. C. (2001). TMS Produces Two Dissociable Types of Speech Disruption. *NeuroImage* 13, 472–478. doi: 10.1006/nimg.2000.0701.
- Sugiura, H. (1998). Matching of acoustic features during the vocal exchange of coo calls by Japanese macaques. *Animal Behaviour* 55, 673–687.
- Taglialatela, J. P., Savage-Rumbaugh, S., & Baker, L. A. (2003). Vocal production by a language-competent Pan paniscus. *Int J Primatol* 24, 1–17.
- Tian, B., Reser, D., Durham, A., Kustov, A., & Rauschecker, J. P. (2001). Functional Specialization in Rhesus Monkey Auditory Cortex. *Science* 292, 290–293. doi: 10.1126/science.1058911.
- Tobias, P. V. (1987). The brain of Homo habilis: A new level of organization in cerebral evolution. Journal of Human Evolution 16, 741–761.
- Warren, J. D., & Griffiths, T. D. (2003). Distinct mechanisms for processing spatial sequences and pitch sequences in the human auditory brain. *Journal of Neuroscience* 23, 5799–5804.
- Wich, S. A., Swartz, K. B., Hardus, M. E., Lameira, A. R., Stromberg, E., &

Shumaker, R. W. (2008). A case of spontaneous acquisition of a human sound by an orangutan. *Primates* 50, 56–64. doi:10.1007/s10329-008-0117-y.

Woods, T. M., Lopez, S. E., Long, J. H., Rahman, J. E., & Recanzone, G. H. (2006). Effects of Stimulus Azimuth and Intensity on the Single-Neuron Activity in the Auditory Cortex of the Alert Macaque Monkey. J. Neurophysiol. 96, 3323– 3337. doi:10.1152/jn.00392.2006.