Supplementary Materials for

Evolution of vocal learning and spoken language

Erich D. Jarvis

Email: ejarvis@rockefeller.edu

Published 4 October 2019, Science 366, 50 (2019)
DOI: 10.1126/science.aax0287

This PDF file includes:

- Supplementary Text
- Figs. S1 to S8
- References
Supplementary Text

Box 1: Comparative summary of forebrain vocal communication circuits across species

The brain summaries in Figures 3 and 4 represent a synthesis of many studies from many laboratories, and yet are not exhaustive. Here, I highlight additional details, including what differs from the past literature. This summary is meant to be help to serve as a guide for future studies. The human brain circuit diagram (Fig. 3B) is revised from my previous 2004 synthesis (4), where instead of considering Broca’s area in BA44 and BA45 as part of a larger strip of prefrontal language cortex between it and the aSMA, this strip is now just limited to Broca’s in the inferior frontal gyrus. Subsequent literature has not supported an entire strip of cortex as spoken-language related. Between Broca’s area and the primary LMC in BA4, I added a premotor LMC in BA6. Prior hypothesis from the Simoyan and other labs had indicated that non-human primates have an old LMC in BA6 premotor cortex, that they called Area 6v (BA6 ventral), which shifted in position in humans to a new LMC in BA4 primary motor cortex, simultaneously acquiring a direct projection (while maintaining the indirect) to ambiguus (Am) vocal motor neurons (18, 41). Here instead, I propose that LMC did not shift position in humans, but that premotor LMC and primary LMC exist together. In addition, in humans a dLMC was discovered separate from a vLMC, based on imaging, gene expression, and electrophysiological studies (10, 15, 16). Because a rudimentary vLMC maybe present in mice and non-human primates, the dLMC has been proposed to be a duplication of the vLMC (17). The aSMA in this model is still considered to be involved in spoken-language. It has been proposed to be connected with the LMC (dorsal and ventral not specified), and many studies show that it is active in speech task, including initiation of speech, but not needed for non-human primate vocalizations (41).

For humans and songbirds, often times comparisons are made between an entire brain region of birds (like HVC) and an entire brain region of human (like Broca’s area) (42, 43). This approach, however, goes against our current understanding of the two competing hypothesis of cortex-pallial homologies between songbirds and humans: 1) different subdivisions of the avian pallium have cell types homologous to different layers of the mammalian cortex; and/or different avian subdivisions have cell types homologous to those within the mammalian amygdala and claustrum (44). These hypotheses are cell-type based, not necessarily regional based.

In the context of the first cell type homology hypothesis, this would mean that avian arcopallium in which songbird RA resides is similar to layer 5 of mammalian cortex, and RA thereby similar to LMC layer 5 neurons, a conclusion supported by gene expression, connectivity, and functional evidence (4, 15, 45). Songbird HVC neurons in the nidopallium that project to RA would be similar to layer 2 neurons in human dLMC or vLMC that are presumed to connect to layer 5 neurons in the same column. NIf also would be considered similar to layers 2, of which human cortical region remains to be determined; another part of LMC or preLMC would be the closest candidate. Av in the mesopallium would be considered similar to layer 3 (or 6), again of LMC. Songbird MAN and songbird MO (the later barely developed in songbirds) would be considered parallel to neurons of layers 2 and 3 respectively of premotor LMC and/or Broca’s area, due to these neurons having similar intratelencephalic connectivity as layers 2 and 3 neurons of mammals, higher level functions in vocal learning beyond production, and gene expression parallels with these layers of the cortex. However, specialized gene expression in HVC, MAN, MO, NIf, and Av have not been thoroughly invested yet relative to its direct surrounding brain regions and the regions directly surrounding speech areas in humans. In contrast, in the striatum, the surrounding gene expression profiles have been assessed, and strong parallels were found
between songbird Area X and the human aSt at the junction between the caudate and putamen (15), whose coordinates overlap with a region involved in vocal learning in humans (46).

Not included in these Figures are the differences between the vocal learning systems of the three vocal learning bird lineages. Unlike songbirds, the parrot and hummingbird analogs of HVC and RA and of parrot NIf and Av analogs are physically separate from the auditory regions, and more anterior in the brain, similar in position to human dLMC and vLMC. In songbirds, all four of these song nuclei are surrounded by both auditory and non-vocal motor regions, whereas in parrots and hummingbirds they are surrounded only by motor regions (28). It is not known if the additional parrot shell song system has specialized gene expression convergence with the human speech regions, but the parrot HVC and RA shell analogs are good candidates for human dLMC and the parrot MO and MAN shell analogs are good candidates for premotor LMC and/or Broca’s area.

In summary, many insights have been gained from studying the evolution and mechanisms of vocal communication in song learning birds, humans, and other species. In this article, I have highlighted hypotheses that I believe have been investigated most intensively. There are other areas that I did not cover, because of space limitations and because they have not yet been investigated as intensively. These include what environmental (e.g. diversity, domestication) and/or behavioral (e.g. sexual, social, predatory, tool use, manual gestures) variables have selected for or against vocal learning and spoken-language (47). When studying these other areas, I believe that the same principles presented here apply.
Supplementary Figures

The supplementary figures S1-S7 include those from the published literature with additional annotations for this report, to help guide understanding of hypotheses presented here and some of their evidence.

![Vocal learning continuum hypothesis](attachment:Vocal_learning_continuum_hypothesis.png)

**Fig. S1.**

**Vocal learning continuum hypothesis.** Diagram of hypothesized step-wise continuous ability of vocal learning among vertebrates (right y axis), from simple to more complex forms (x axis). As vocal learning complexity increases, there are decreasing number of species with the ability (left y axis). (A to H) Proposed example species at each step is given. The continuum ranges from lizards that do not vocalize and have no vocal learning, to non-human primates with limited vocal learning, to songbirds with advanced vocal learning, to parrots and humans with high vocal learning (2). Technically the, vocal learning is considered sounds learned using the vocal organ, larynx or syrinx. However, other oral facial musculature could be used to produce learned sounds. Modulating the movements of the oral cavity with little voluntary control of the larynx could be the mechanism used by many pet dogs and cats trained to try to say sounds like “I love you”, “hello”, and more. These ways of producing sounds, as well as other component traits of spoken language (Fig. 1) could be in continuums among species. Hypothesis based on (2, 6, 48) and diagram updated from (2).
Fig. S2.

**General brain differences in vocal learning species.** (A) Birds have 2-3 times higher neuron density compared to mammals of similar brain size. (B) Some vocal learning birds (songbirds, green; and parrots, red) have 1-2 times higher density of neurons in their cortical region (pallium) than do vocal non-learning birds (black), non-human primates (blue), which in turn is higher than other mammals (rodents, purple; artiodactyls, brown). For all species, as brain mass increases (x axis), neuron density decreases (y axis). (C) The extra copies of *SRGAP2* (*SRGAP2B* and *SRGAP2C*) only found in human genomes inhibits the original copy of *SRGAP2A* found in all species, resulting in human cortical dendritic spines remaining longer and at a higher density into adulthood. (A) from [https://www.sciencespacerobots.com/study-finds-bird-brains-are-more-neuron-dense-than-62020161](https://www.sciencespacerobots.com/study-finds-bird-brains-are-more-neuron-dense-than-62020161) based on (49), (B) from (49), and C from (50) based on (32, 51).
Fig. S3.

**Vocal organ function and morphology between vocal learners and vocal non-learners.** (A) Hypothesis on larynx descent (yellow) away from the tongue (red) and loss of airsac (blue) as being responsible for greater vocalization diversity of sounds in human relative to chimpanzee. However, when chimpanzees vocalize, they open the vocal tract with decent of the larynx. (B) Experimentally determined space of vowel-like sounds, capable of being produced by both baboon and human larynx. (C) Examples of simple and complex syrinx muscle organization in vocal non-learning (top) and vocal learning (bottom) bird species. (A) from (52), (B) from (53), and (C) from (54).
Species differences in density of direct projection from motor cortex to brainstem vocal motor neurons. (A) Many fibers (white) with direct projections to songbird XIIts vocal motor and adjacent respiratory RAm premotor neurons from a tracer injection into motor cortical RA song nucleus (coronal section; left is lateral, top is dorsal). (B) Many fibers (black) with direct projections innervating a human Am vocal motor neuron (arrow pointing to cell body) from a patient with damage to M1 including LMC (coronal section). (C) Two fibers (black, small arrows) with direct projections onto a mouse Am vocal motor neuron (brown, large arrow) from a tracer injection into LMC-M1 (coronal section). (D) Mouse LMC-M1 layer 5 neurons (white) labeled from a transsynaptic tracer injected into laryngeal muscles (coronal section). (E) Putative mouse vocal communication pathway according to the continuum hypothesis view, with a weak direct projection (red arrow) from LMC-M1. Abbreviations same as Fig. 3 and color-coding the same as Fig. 4., except yellow arrows are predicted. (A) from (55), (B) from (55, 56), and (C-E) from (48).
Fig. S5.
**Auditory-vocal premotor cortical hypothesis in mammals.** (A-C) Illustration showing interpretation of results that indicate that the arcuate fasciculus connection (or dorsal stream) between A2 (BA22 and adjacent regions) with the vocal premotor cortex and Broca’s (BA44 and BA45) increases in strength (thicker green arrows) the closer a species is related to humans: monkey (macaque, C), to great ape (chimpanzee, B), to hominid (human, A) (6). A more ventral stream does not change as dramatically across primate species, and if anything decreases in humans. Findings determined from diffusion-weighted magnetic resonance imaging (DTI) imaging. (D-F) Tracers injected into mouse LMC-M2/M1 (D) reveal layer 3 (III) neurons in A2 (E) that project to LMC-M2/M1 similar to the dorsal stream in primates (compare F versus A-C). Abbreviations same as Figure 3; additional ones are: PrCS, precentral sulcus; CS, central sulcus; IFS, inferior frontal sulcus; STS, superior temporal sulcus; IPS, intraparietal sulcus; CC, corpus colosum. (A-C) modified from (57), (D) and (E) modified from (48), and (F) brain model from the website listed.
Fig. S6.

One view of external and internal concepts influenced by the language module hypothesis. Internalization functions of a language (i.e. language module) are proposed to generate high-level syntax rules (blue) and cognitive computations (orange), and then through externalization send instructions to and receive feedback from sensory and motor circuits (red) for perception and production of spoken and other forms of language in humans. As specified in (58) “Externalization is the mapping from internal linguistic representations to their ordered output form, either spoken or manually gestured. Internalization is the computations that construct mental syntactic and conceptual-intentional representations internal to the mind/brain.” The boxes represent different brain regions or circuits, but the specific circuits are not specified. Figure from (58) with permission.
Fig. S7.
Convergent gene regulation, function, and brain regions between songbirds and humans for song and spoken-language. (A) Example gene, SLIT1, with specialized downregulation in the RA song nucleus analog across the three independently evolved vocal learning bird species, not found in vocal non-learner species. White, SLIT1 mRNA expression detected by in-situ hybridization. Red, cresyl violate stain of all cells. Sections are coronal (left is lateral, top is dorsal). (B) Heatmap of SLIT1 mRNA expression in human cortex, with white labelled samples showing down regulation of SLIT1 relative to the surrounding cortex. Right hemisphere is shown, although left is similar. (C) Heatmap of in-vivo electrophysiological activity showing the regions in M1 cortex with the highest activity during production of speech sounds that use the larynx, in
human patients. The two regions of highest laryngeal speech activation (dorsal LMC and ventral LMC) correspond to the two regions of convergent gene expression specializations between humans and the RA analog of song-learning birds. These findings were used to identify and support a newly defined separation of dorsal LMC and ventral LMC in humans (10, 15-17). (D) Hypothesis of how SLIT1 down-regulation in layer 5 human LMC and songbird RA could result in formation of a direct projection to vocal motor neurons. As SLIT1 binding to its receptor ROBO1 can be repulsive, presence of SLIT1 in the projecting neurons may prevent the direct projection from forming (left); absent or reduced SLIT1 may allow the direction projection to form (right). Abbreviations same as Figure 3; additional names are: VA, vocal nucleus of the arcopallium; AAC, central nucleus of the anterior arcopallium. (A) and (B) from (15), (C) from (10), and (D) from (59), used with permission.
**Fig. S8.**

**Vocal learning continuum hypothesis explained in anatomical stages.** Written is a speculative trajectory to explain all the finding of differences between species. Arrows point to evolutionary milestone changes, or steps. Thus far no one has observed a species at stage 1b, nor a brain pathway to duplicate. This scheme sets up a series of sub-hypotheses that can be tested. Description of Lombard effect can be found at (60)

1. **Duplicate forebrain motor learning pathway to become a vocal learning pathway caused by changes in brain development genes (e.g. mouse and non-human primates)**

2. **Move and isolate new vocal motor learning pathway out of motor learning pathway using neural migration genes**

3. **Enhance vocal motor pathway with modification of neural connectivity, protection, and plasticity genes (e.g. songbirds and hummingbirds)**

4. **Duplicate vocal motor learning pathway one or more times in more advanced vocal learners (e.g. parrots and humans)**
References and Notes


