



OPINION ARTICLE

UPDATE From where to what: a neuroanatomically based evolutionary model of the emergence of speech in humans [version 3; referees: 1 approved, 2 approved with reservations]

Oren Poliva

Bangor University, Bangor, UK

v3 First published: 13 Mar 2015, 4:67 (doi: 10.12688/f1000research.6175.1)
 Second version: 21 Jan 2016, 4:67 (doi: 10.12688/f1000research.6175.2)
 Latest published: 20 Sep 2017, 4:67 (doi: 10.12688/f1000research.6175.3)

Abstract

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 a network of limbic and brainstem regions.

Because the ADS of present day humans also performs speech production, 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 his/her 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.

Open Peer Review

Referee Status: ? ✓ ?

	Invited Referees		
	1	2	3
UPDATE			
version 3 published 20 Sep 2017			
REVISED		✓ report	
version 2 published 21 Jan 2016		↑	
version 1 published 13 Mar 2015	? report	? report	? report

- Amy Poremba**, University of Iowa, USA
- Josef Rauschecker**, Georgetown University, USA
- Michael A Arbib**, University of Southern California, USA

Discuss this article

Comments (2)

Corresponding author: Oren Poliva (polivaoren@gmail.com)

Author roles: Poliva O: Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology, Project Administration, Resources, Validation, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

How to cite this article: Poliva O. **From where to what: a neuroanatomically based evolutionary model of the emergence of speech in humans [version 3; referees: 1 approved, 2 approved with reservations]** *F1000Research* 2017, 4:67 (doi: [10.12688/f1000research.6175.3](https://doi.org/10.12688/f1000research.6175.3))

Copyright: © 2017 Poliva O. This is an open access article distributed under the terms of the [Creative Commons Attribution Licence](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Grant information: The author(s) declared that no grants were involved in supporting this work.

First published: 13 Mar 2015, 4:67 (doi: [10.12688/f1000research.6175.1](https://doi.org/10.12688/f1000research.6175.1))

UPDATE Updates from Version 2

In this revised version, I update the paper with a few recent publications and minimal changes in wording.

[See referee reports](#)

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 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. In the present paper, based on findings collected from brain research, genetics and paleoarcheology, I demonstrate that human speech and contact calls use the same brain structures, and consequently argue that human speech emerged from contact call exchange. I then argue that by modifying their contact calls with intonations, infants were capable of signaling their mothers whether they were under high- or low-level of distress. Given the turn taking nature of these calls, and as both mothers and infants were capable of modifying their calls with intonations, the ability to choose the call type eventuated with the first yes-no conversation structure. In this scenario infants were capable of inquiring about the safety of objects in their environment (i.e., with a low-level distress call) and mothers were capable of responding to that question with a high-level distress call to signal danger or a low-level distress call to signal safety. As the use of intonations became more prevalent, conversations became more

complex, and consequently vocal control became more volitional. Speech emerged once individuals acquired sufficient volitional vocal control to invent names for objects in their environment.

2. Models of language processing in the brain and their relation to language evolution

Throughout the 20th century, our knowledge of language processing in the brain was dominated by the Wernicke-Lichtheim-Geschwind model (Geschwind, 1965; Lichtheim, 1885; Wernicke, 1974). This model is primarily based on research conducted on brain-damaged individuals who were reported to possess a variety of language related disorders. In accordance with this model, words are perceived via a specialized word reception center (Wernicke’s area) that is located in the left temporoparietal junction. This region then projects to a word production center (Broca’s area) that is located in the left inferior frontal gyrus. Because almost all language input was thought to funnel via Wernicke’s area and all language output to funnel via Broca’s area, it became extremely difficult to identify the basic properties of each region. This lack of clear definition for the contribution of Wernicke’s and Broca’s regions to human language rendered it extremely difficult to identify their homologues in other primates. (For one attempt, see Aboitiz & García, 1997). With the advent of the MRI and its application for lesion mappings, however, it was shown that this model is based on incorrect correlations between symptoms and lesions and is therefore flawed (Anderson *et al.*, 1999; DeWitt & Rauschecker, 2013; Dronkers *et al.*, 1999; Dronkers, 2000; Dronkers *et al.*, 2004; Mesulam *et al.*, 2015; Poeppel *et al.*, 2012; Vignolo *et al.*, 1986). The refutation of such an influential and dominant model opened the door to new models of language processing in the brain, and as will be presented below, to formulating a novel account of the evolutionary origins of human language from a neuroscientific perspective.

In the last two decades, significant advances occurred in our understanding of the neural processing of sounds in primates. Initially by recording of neural activity in the auditory cortices of monkeys (Bendor & Wang, 2006; Rauschecker *et al.*, 1995) and later elaborated via histological staining (de la Mothe *et al.*, 2006; de la Mothe *et al.*, 2012; Kaas & Hackett, 2000 - review) and fMRI scanning studies (Petkov *et al.*, 2006), 3 auditory fields were identified in the primary auditory cortex, and 9 associative auditory fields were shown to surround them (Figure 1 top left). Anatomical tracing and lesion studies further indicated of a separation between the anterior and posterior auditory fields, with the anterior primary auditory fields (areas R-RT) projecting to the anterior associative auditory fields (areas AL-RTL), and the posterior primary auditory field (area A1) projecting to the posterior associative auditory fields (areas CL-CM; de la Mothe *et al.*, 2006; Morel *et al.*, 1993; Rauschecker & Tian, 2000; Rauschecker *et al.*, 1997). Recently, evidence accumulated that indicates homology between the human and monkey auditory fields. In humans, histological staining studies revealed two separate auditory fields in the primary auditory region of Heschl’s gyrus (Sweet *et al.*, 2005; Wallace *et al.*, 2002), and by mapping the tonotopic organization of the human primary auditory fields with high resolution fMRI and comparing it to the tonotopic organization of the monkey primary auditory fields, homology was established between the human anterior primary auditory field and monkey area R (denoted in humans as area hR) and the human posterior primary auditory field and the monkey area A1 (denoted

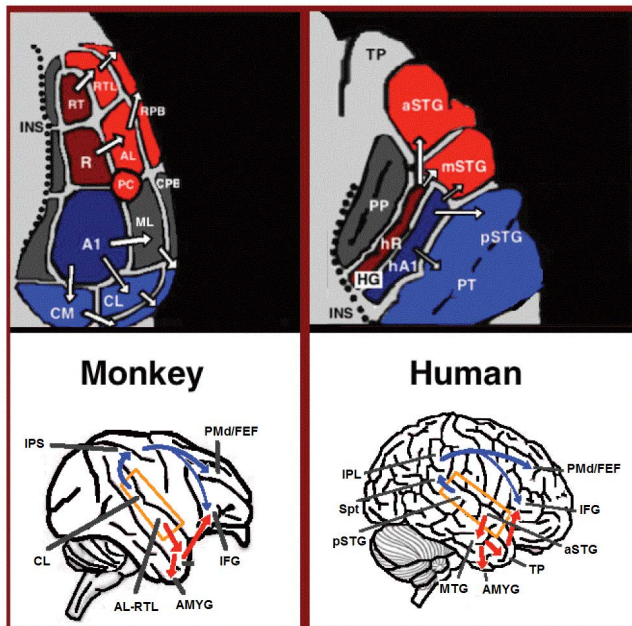


Figure 1. Dual stream connectivity between the auditory cortex and frontal lobe of monkeys and humans. Top: The auditory cortex of the monkey (left) and human (right) is schematically depicted on the supratemporal plane and observed from above (with the parieto-frontal operculi removed). Bottom: The brain of the monkey (left) and human (right) is schematically depicted and displayed from the side. Orange frames mark the region of the auditory cortex, which is displayed in the top sub-figures. Top and Bottom: Blue colors mark regions affiliated with the ADS, and red colors mark regions affiliated with the AVS (dark red and blue regions mark the primary auditory fields). Abbreviations: AMYG-amygdala, HG-Heschl's gyrus, FEF-frontal eye field, IFG-inferior frontal gyrus, INS-insula, IPS-intra parietal sulcus, MTG-middle temporal gyrus, PC-pitch center, PMd-dorsal premotor cortex, PP-planum polare, PT-planum temporale, TP-temporal pole, Spt-sylvian parietal-temporal, pSTG/mSTG/aSTG-posterior/middle/anterior superior temporal gyrus, CL/ML/AL/RTL-caudo-/middle-/antero-/rostromtemporal-lateral belt area, CPB/RPB-caudal/rostral parabelt fields.

in humans as area hA1; [Da Costa et al., 2011](#); [Humphries et al., 2010](#); [Langers & van Dijk, 2012](#); [Striem-Amit et al., 2011](#); [Woods et al., 2010](#)). Intra-cortical recordings from the human auditory cortex further demonstrated similar patterns of connectivity to the auditory cortex of the monkey. Recording from the surface of the auditory cortex (supra-temporal plane) reported that the anterior Heschl's gyrus (area hR) projects primarily to the middle-anterior superior temporal gyrus (mSTG-aSTG) and the posterior Heschl's gyrus (area hA1) projects primarily to the posterior superior temporal gyrus (pSTG) and the planum temporale (area PT; [Figure 1](#) top right; [Gourévitch et al., 2008](#); [Guéguin et al., 2007](#)). Consistent with connections from area hR to the aSTG and hA1 to the pSTG is an fMRI study of a patient with impaired sound recognition (auditory agnosia), who was shown with reduced bilateral activation in areas hR and aSTG but with spared activation in the mSTG-pSTG ([Poliva et al., 2015](#)). This connectivity pattern is also corroborated by a study that recorded activation from the lateral surface of the auditory cortex and reported of simultaneous

non-overlapping activation clusters in the pSTG and mSTG-aSTG while listening to sounds ([Chang et al., 2011](#)).

Downstream to the auditory cortex, anatomical tracing studies in monkeys delineated projections from the anterior associative auditory fields (areas AL-RTL) to ventral prefrontal and premotor cortices in the inferior frontal gyrus (IFG; [Munoz et al., 2009](#); [Romanski et al., 1999](#)) and amygdala ([Kosmal et al., 1997](#)). Cortical recording and functional imaging studies in macaque monkeys further elaborated on this processing stream by showing that acoustic information flows from the anterior auditory cortex to the temporal pole (TP) and then to the IFG ([Perrodin et al., 2011](#); [Petkov et al., 2008](#); [Poremba et al., 2004](#); [Romanski et al., 2005](#); [Russ et al., 2008](#); [Tsunada et al., 2011](#)). This pathway is commonly referred to as the auditory ventral stream (AVS; [Figure 1](#), bottom left-red arrows). In contrast to the anterior auditory fields, tracing studies reported that the posterior auditory fields (areas CL-CM) project primarily to dorsolateral prefrontal and premotor cortices (although some projections do terminate in the IFG; [Cusick et al., 1995](#); [Romanski et al., 1999](#)). Cortical recordings and anatomical tracing studies in monkeys further provided evidence that this processing stream flows from the posterior auditory fields to the frontal lobe via a relay station in the intra-parietal sulcus (IPS; [Cohen et al., 2004](#); [Deacon, 1992](#); [Lewis & Van Essen, 2000](#); [Roberts et al., 2007](#); [Schmahmann et al., 2007](#); [Seltzer & Pandya, 1984](#)). This pathway is commonly referred to as the auditory dorsal stream (ADS; [Figure 1](#), bottom left-blue arrows). Comparing the white matter pathways involved in communication in humans and monkeys with diffusion tensor imaging techniques indicates of similar connections of the AVS and ADS in the two species (Monkey: [Schmahmann et al., 2007](#); Human: [Catani et al., 2004](#); [Frey et al., 2008](#); [Makris et al., 2009](#); [Menjot de Champfleury et al., 2013](#); [Saur et al., 2008](#); [Turken & Dronkers, 2011](#)). In humans, the pSTG was shown to project to the parietal lobe (sylvian parietal-temporal junction-inferior parietal lobule; Spt-IPL), and from there to dorsolateral prefrontal and premotor cortices ([Figure 1](#), bottom right-blue arrows), and the aSTG was shown to project to the anterior temporal lobe (middle temporal gyrus-temporal pole; MTG-TP) and from there to the IFG ([Figure 1](#) bottom right-red arrows).

On the basis of converging evidence collected from monkeys and humans, it has been established that the AVS is responsible for the extraction of meaning from sounds (see [Appendix A](#) for a review of the literature). Specifically, the anterior auditory cortex is ascribed with the perception of auditory objects, and downstream, the MTG and TP are thought to match the auditory objects with their corresponding audio-visual semantic representations (i.e., the semantic lexicon). This recognition of sounds in the AVS, although critical for intact communication, appears to contribute less to the uniqueness of human language than the ADS. This is demonstrated by the universality of sound recognition, as many mammalian species use it for identifying prey, predators or potential mates. As an example, dogs were reported capable of recognizing spoken words and extract their meaning ([Kaminski et al., 2004](#); [Pillely & Reid, 2011](#)), and with fMRI this sound recognition ability was localized to the TP of the AVS ([Andics et al., 2014](#)). Studies also provided evidence that the sound recognition of non-human apes is equivalent in complexity to ours. Apes trained in human facilities were reported

capable of learning human speech and comprehending its meaning (e.g., the bonobos, Kanzi and Panbanisha, were reported to recognize more than 3000 spoken English words; Blake, 2004; Gibson, 2011). Moreover, a study that compared humans and a chimpanzee in their recognition of acoustically distorted spoken words, reported no differences between chimpanzee and human performance (Heimbauer *et al.*, 2011).

In contrast to the relatively preserved function of the AVS among mammals, converging evidence suggests that the ADS was significantly modified since our *Hominin* ancestors separated from other apes. For instance, a diffusion tensor imaging study that compared the white matter of humans and chimpanzees demonstrated significant strengthening of ADS connectivity, but not AVS connectivity (Rilling *et al.*, 2012). Evidence for restructuring of the ADS during *Hominin* evolution is also demonstrated 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 a dramatic heightening of the IPL and enlargement (though to a lesser degree) of the IFG, whereas the rest of the endocranium remains extremely similar to the endocranium of modern apes (Tobias, 1987). It is also worth reporting that the recently discovered *Australopithecus sediba* (Carlson *et al.*, 2011), which is the closest known relative to the *Australopith* predecessor of *Homo habilis*, is characterized with a very ape-like parietal and frontal lobes (although some modifications of the orbitofrontal surface were noted). These findings also suggest that it was changes to the ADS that initially prompted the brain enlargement that characterized *Hominans* (members of the genus *Homo*; Wood & Richmond, 2000), and separated us from other *Hominins*.

In contrast to the AVS, the ADS was ascribed with a diverse range of seemingly unrelated functions. These functions, which will be detailed throughout this paper, include auditory localization, audiovisual integration, and voice detection in monkeys. In humans, the ADS has been further ascribed with the preparation and production of speech. In the present paper, based on functional differences between the ADS of monkeys and humans, I propose intermediate stages in the development of human speech.

3. The role of the ADS in audiospatial processing

The most established role of the ADS is with audiospatial processing. This is evidenced via studies that recorded neural activity from the auditory cortex of monkeys, and correlated the strongest selectivity to changes in sound location with the posterior auditory fields (areas CM-CL), intermediate selectivity with primary area A1, and very weak selectivity with the anterior auditory fields (Benson *et al.*, 1981; Miller & Recanzone, 2009; Rauschecker *et al.*, 1995; Tian *et al.*, 2001; Woods *et al.*, 2006). In humans, behavioral studies of brain damaged patients (Clarke *et al.*, 2000; Griffiths *et al.*, 1996) and EEG recordings from healthy participants (Anourov *et al.*, 2001) demonstrated that sound localization is processed independently of sound recognition, and thus is likely independent of processing in the AVS. Consistently, a working memory study (Clarke *et al.*, 1998) reported two independent working memory storage spaces, one for acoustic properties and one for locations. Functional imaging studies that contrasted sound discrimination and sound localization reported a correlation between sound discrimination and activation in the mSTG-aSTG, and

correlation between sound localization and activation in the pSTG and PT (Ahveninen *et al.*, 2006; Alain *et al.*, 2001; Barrett & Hall, 2006; De Santis *et al.*, 2007; Viceic *et al.*, 2006; Warren & Griffiths, 2003), with some studies further reporting of activation in the Spt-IPL region and frontal lobe (Hart *et al.*, 2004; Maeder *et al.*, 2001; Warren *et al.*, 2002). Some fMRI studies also reported that the activation in the pSTG and Spt-IPL regions increased when individuals perceived sounds in motion (Baumgart *et al.*, 1999; Krumbholz *et al.*, 2005; Pavani *et al.*, 2002). EEG studies using source-localization also identified the pSTG-Spt region of the ADS as the sound localization processing center (Tata & Ward, 2005a; Tata & Ward, 2005b). A combined fMRI and MEG study corroborated the role of the ADS with audiospatial processing by demonstrating that changes in sound location resulted in activation spreading from Heschl's gyrus posteriorly along the pSTG and terminating in the IPL (Brunetti *et al.*, 2005). In another MEG study, the IPL and frontal lobe were shown active during maintenance of sound locations in working memory (Lutzenberger *et al.*, 2002).

In addition to localizing sounds, the ADS appears also to encode the sound location in memory, and to use this information for guiding eye movements. Evidence for the role of the ADS in encoding sounds into working memory is provided via studies that trained monkeys in a delayed matching to sample task, and reported of activation in areas CM-CL (Gottlieb *et al.*, 1989) and IPS (Linden *et al.*, 1999; Mazzoni *et al.*, 1996) during the delay phase. Influence of this spatial information on eye movements occurs via projections of the ADS into the frontal eye field (FEF; a premotor area that is responsible for guiding eye movements) located in the frontal lobe. This is demonstrated with anatomical tracing studies that reported of connections between areas CM-CL-IPS and the FEF (Cusick *et al.*, 1995; Stricanne *et al.*, 1996), and electro-physiological recordings that reported neural activity in both the IPS (Linden *et al.*, 1999; Mazzoni *et al.*, 1996; Mullette-Gillman *et al.*, 2005; Stricanne *et al.*, 1996) and the FEF (Russo & Bruce, 1994; Vaadia *et al.*, 1986) prior to conducting saccadic eye-movements toward auditory targets.

4. The role of the ADS in the localization of con-specifics

In addition to processing the locations of sounds, evidence suggests that the ADS further integrates sound locations with auditory objects. Demonstrating this integration are electrophysiological recordings from the posterior auditory cortex (Recanzone, 2008; Tian *et al.*, 2001) and IPS (Gifford & Cohen, 2005), as well a PET study (Gil-da-Costa *et al.*, 2006), that reported neurons that are selective to monkey vocalizations. One of these studies (Tian *et al.*, 2001) further reported neurons in this region (CM-CL) that are characterized with dual selectivity for both a vocalization and a sound location. Consistent with the role of the pSTG-PT in the localization of specific auditory objects are also studies that demonstrate a role for this region in the isolation of specific sounds. For example, two functional imaging studies correlated circumscribed pSTG-PT activation with the spreading of sounds into an increasing number of locations (Smith *et al.*, 2010-fMRI; Zatorre *et al.*, 2002-PET). Accordingly, an fMRI study correlated the perception of acoustic cues that are necessary for separating musical sounds (pitch chroma) with pSTG-PT activation (Warren *et al.*, 2003).

When elucidating the role of the primate ADS in the integration of a sound's location with calls, it remains to be determined what kind of information the ADS extracts from the calls. This information could be then used to make inferences about the function of the ADS. Studies from both monkeys and humans suggest that the posterior auditory cortex has a role in the detection of a new speaker. A monkey study that recorded electrophysiological activity from neurons in the posterior insula (near the pSTG) reported neurons that discriminate monkey calls based on the identity of the speaker (Remedios *et al.*, 2009a). Accordingly, human fMRI studies that instructed participants to discriminate voices reported an activation cluster in the pSTG (Andics *et al.*, 2010; Formisano *et al.*, 2008; Warren *et al.*, 2006). A study that recorded activity from the auditory cortex of an epileptic patient further reported that the pSTG, but not aSTG, was selective for the presence of a new speaker (Lachaux *et al.*, 2007-patient 1). The role of this posterior voice area, and the manner in which it differs from voice recognition in the AVS (Andics *et al.*, 2010; Belin & Zatorre, 2003; Nakamura *et al.*, 2001; Perrodin *et al.*, 2011; Petkov *et al.*, 2008), was further shown via electro-stimulation of another epileptic patient (Lachaux *et al.*, 2007-patient 2). This study reported that electro-stimulation of the aSTG resulted in changes in the perceived pitch of voices (including the patient's own voice), whereas electro-stimulation of the pSTG resulted in reports that her voice was "drifting away." This report indicates a role for the pSTG in the integration of sound location with an individual voice. Consistent with this role of the ADS is a study that reported patients, with AVS damage but spared ADS (surgical removal of the anterior STG/MTG), were no longer capable of isolating environmental sounds in the contralesional space, whereas their ability of isolating and discriminating human voices remained intact (Efron & Crandall, 1983). Preliminary evidence from the field of fetal cognition suggests that the ADS is capable of identifying voices in addition to discriminating them. By scanning fetuses of third trimester pregnant mothers with fMRI, the researchers reported of activation in area Spt when the hearing of voices was contrasted to pure tones (Jardri *et al.*, 2012). The researchers also reported that a sub-region of area Spt was more selective to maternal voice than unfamiliar female voices. Based on these findings, I suggest that the ADS has acquired a special role in primates for the localization of conspecifics.

5. The role of the ADS in the detection of contact calls

To summarize, I have argued that the monkey's ADS is equipped with the algorithms required for detecting a voice, isolating the voice from the background cacophony, determining its location, and guiding eye movements for the origin of the call. An example of a behavior that utilizes all these functions is the exchange of contact calls, which are used by extant primates to monitor the location or proximity of conspecific tribe members (Biben *et al.*, 1986; Sugiura, 1998). The utilization of these ADS functions during the exchange of contact calls was demonstrated in studies of squirrel monkeys and vervet monkeys (Biben, 1992; Biben *et al.*, 1989; Cheney & Seyfarth, 1980; Symmes & Biben, 1985). In both species, mothers showed no difficulty in isolating their own infant's call, localizing it, and maintaining this location in their memory while approaching the source of the sound. A similar use of contact calls has been documented in our closest relatives, chimpanzees. The exchange of pant-hoot calls was documented between

chimpanzees that were separated by great distances (Goodall, 1986; Marler & Hobbett, 1975) and was used for re-grouping (Mitani & Nishida, 1993). Because infants respond to their mother's pant-hoot call with their own unique vocalization (staccato call; Matsuzawa, 2006), the contact call exchange appears also to play an important role in the ability of mothers to monitor the location of their infants. It is also worth noting that when a chimpanzee produced a pant-hoot call and heard no call in response, the chimpanzee was reported to carefully scan the forest before emitting a second call (Goodall, 1986). This behavior demonstrates the relationship between the detection of contact calls, the embedding of auditory locations in a map of the environment, and the guidance of the eyes for searching the origin of the call. Further corroborating the involvement of the ADS in the detection of contact calls are intra-cortical recordings from the posterior insula (near area CM-A1) of the macaque, which revealed stronger selectivity for a contact call (coo call) than a social call (threat call; Remedios *et al.*, 2009a). Contrasting this finding is a study that recorded neural activity from the anterior auditory cortex, and reported that the proportion of neurons dedicated to a contact call was similar to the proportions of neurons dedicated to other calls (Perrodin *et al.*, 2011).

Perceiving a contact call can be viewed as a three-step process. The individual is required to detect a voice, integrate it with its location and verify that no face is visible in that location (Figure 2). In the previous paragraphs, I provided evidence for the involvement of the ADS in the first two stages (voice detection and localization). Evidence for the role of the ADS in the integration of faces with their appropriate calls is provided by a study that recorded activity from the monkey auditory cortex (areas A1 and ML; Ghazanfar *et al.*, 2005). The monkeys were presented with pictures of a monkey producing a call in parallel to hearing the appropriate call, or only saw the face or heard the call in isolation. Consistent with the prediction from the present model that visual inspection of faces inhibits processing of contact calls, the face-call integration was much more enhanced for the social call (grunt call) than for the contact call (coo call). Associating this integration of faces with calls with processing in the ADS is consistent with a monkey fMRI study that correlated audio-visual integration with activation in the posterior, but not in the anterior, auditory fields (Kayser *et al.*, 2009).

6. The role of the ADS in the response to contact calls

Hitherto, I have argued that the ADS is responsible for the perception of contact calls. However, as the perception of a contact call leads to producing a contact call in return, it is also desirable to suggest a pathway through which the ADS mediates vocal production.

Cumulative evidence suggests that most vocalizations in non-human primates are prepared and produced in a network of limbic and brainstem regions, and do not appear to be controlled by the ADS. For instance, studies that damaged the temporoparietal and/or the IFG regions of monkeys reported that such lesions had no effect on spontaneous vocal production (Aitken, 1981; Sutton *et al.*, 1974). This conclusion is also consistent with comprehensive electro-stimulation mappings of the monkey's brain, which reported no spontaneous vocal production during stimulation of the temporal, occipital, parietal, or frontal lobes (Jürgens & Ploog, 1970; Robinson, 1967). These studies, however, reported emission of

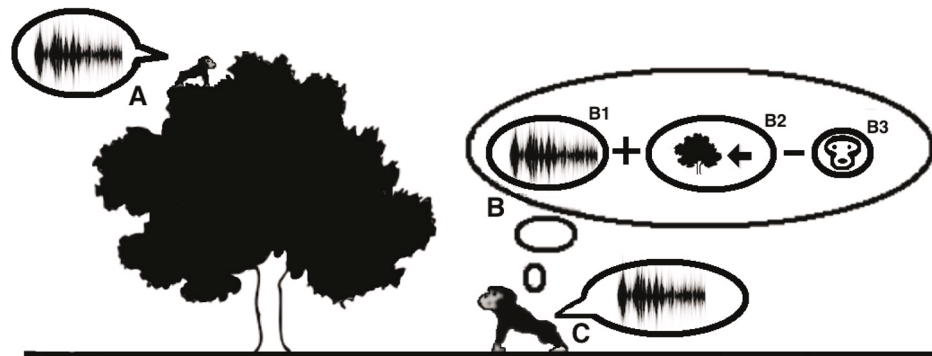


Figure 2. Discrete stages in contact call exchange. In accordance with the model, the original function of the ADS is for the localization of and the response to contact calls that are exchanged between mothers and their infants. When an infant emits a contact call (A), the mother identifies her offspring's voice (B1) localizes the call (B2) and maintains this information in visual working memory. Then, if the corresponding face is absent in that location (B3), the mother emits a contact call in return (C).

vocalizations after stimulating limbic and brainstem regions (amygdala, anterior cingulate cortex, basal forebrain, hypothalamus, mid-brain periaqueductal gray). Moreover, based on a study that correlated chemical activation in the mid-brain periaqueductal gray with vocal production, it was inferred that all the limbic regions project to central pattern generators in the periaqueductal gray, which orchestrates the vocal production (Zhang *et al.*, 1994). In a series of tracing studies and electrophysiological recordings, it was also shown that the periaqueductal gray projects to pre-motor brainstem areas (Hage & Jürgens, 2006; Hannig & Jürgens, 2006), which in turn project to brainstem motor nuclei (Holstege, 1989; Holstege *et al.*, 1997; Lüthe *et al.*, 2000; Vanderhorst *et al.*, 2000; Vanderhorst *et al.*, 2001). The brainstem motor nuclei then directly stimulate the individual muscles of the vocal apparatus. Because documented calls of non-human primates (including chimpanzees) were shown with very little plasticity (Arcadi, 2000) and were observed only in highly emotional situations (Goodall, 1986), these limbic-brainstem generated calls are likely more akin to human laughter, sobbing, and screaming than to human speech.

Although most monkey vocalizations can be produced without cortical control, some calls, such as alarm calls and contact calls are context dependent and are thus likely under cortical influence (Biben, 1992; Seyfarth *et al.*, 1980). Furthermore, several studies demonstrated that contact calls are directly regulated by the ADS. For instance, 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). Consistently, a study that sacrificed marmoset monkeys immediately after responding to contact calls (phee calls) measured highest neural activity (genomic expression of cFos protein) in the posterior auditory fields (CM-CL), and IFG (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). Anatomical tracing studies (Jürgens & Alipour, 2002; Roberts *et al.*, 2007) demonstrated direct

connections from the IFG of monkeys to limbic and brainstem regions, thus providing a possible route for controlling the contact call response. The former study (Jürgens & Alipour, 2002), however, further reported of a second direct connection from the IFG to a brainstem motor nucleus (hypoglossal nucleus) which controls tongue movements. Although the role of this pathway is not yet known, its anatomical connectivity implies that it is capable of bypassing the limbic-brainstem vocal network, and provides some volitional control over the vocal apparatus. This conclusion is further consistent with behavioral studies of monkeys that reported partial volitional control in the contact call response. For instance, a study that followed macaque mothers and babies reported that the macaque mothers were capable, to a limited extent, of modifying their contact calls to acoustically match those of their infants (Masataka, 2009). Squirrel monkeys and macaque monkeys were also reported to modify the frequencies of their contact calls, which resulted with the caller and responder emitting slightly different calls (Biben *et al.*, 1986; Sugiura, 1998). In one study, macaque monkeys were even observed to spontaneously modify the vocal properties of their contact call for requesting different objects from the experimenter (Hihara *et al.*, 2003). Anecdotal reports of more generalized volitional vocal control, albeit rudimentary, in apes (Hayes & Hayes, 1952; Hopkins *et al.*, 2007; Kalan *et al.*, 2015; Koda *et al.*, 2007; Koda *et al.*, 2012; Lameira *et al.*, 2015; Laporte & Zuberbühler, 2010; Perlman & Clark, 2015; Tagliatalata *et al.*, 2003; Wich *et al.*, 2008) suggest that the direct connections between the IFG and the brainstem motor nuclei were strengthened prior to our divergence from our apian relatives.

7. From contact calls to speech

In the previous sections I provided evidence that the ADS of non-human primates is responsible for the detection and response to contact calls. In the present section I present converging evidence that in humans the ADS performs speech production, and argue that human speech emerged from the exchange of contact calls.

Evidence for a role of the ADS in the transition from mediating contact calls into mediating human speech 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). 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).

8. Prosodic speech and the emergence of conversations

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 in section 5) 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 context of the present model, such evolutionary course implies that throughout *Hominan*

evolution, the ADS gained increased control over the vocal apparatus, possibly by strengthening the direct connections of the IFG with the brainstem motor nuclei. Consistent with this view, many studies demonstrated a role for the ADS in the perception and production of intonations. For instance, an fMRI study that instructed participants to rehearse speech, reported that perception of prosodic speech, when contrasted with flattened speech, results in a stronger activation of the PT-pSTG of both hemispheres (Meyer *et al.*, 2004). In congruence, an fMRI study that compared the perception of hummed speech to natural speech didn't identify any brain area that is specific to humming, and thus concluded that humming is processed in the speech network (Ischebeck *et al.*, 2004). fMRI studies that instructed participants to analyze the rhythm of speech also reported of ADS activation (Spt, IPL, IFG; Geiser *et al.*, 2008; Gelfand & Bookheimer, 2003). An fMRI study that compared speech perception and production to the perception and production of humming noises, reported in both conditions that the overlapping activation area for perception and production (i.e., the area responsible for sensory-motor conversion) was located in area Spt of the ADS (Hickok *et al.*, 2003). Supporting evidence for the role of the ADS in the production of prosody are also studies reporting that patients diagnosed with apraxia of speech are additionally diagnosed with expressive dysprosody (Odell *et al.*, 1991; Odell & Shriberg, 2001; Shriberg *et al.*, 2006 - FOXP2 affected individuals). Finally, the evolutionary account proposed here from vocal exchange of calls to a prosodic-based language is similar to the recent development of whistling languages, since these languages were documented to evolve from exchanging simple calls used to report speakers' locations into a complex semantic system based on intonations (Meyer, 2008).

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 ask questions likely co-emerged with the ability to modify calls with prosodic intonations, 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 (Figure 3). 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

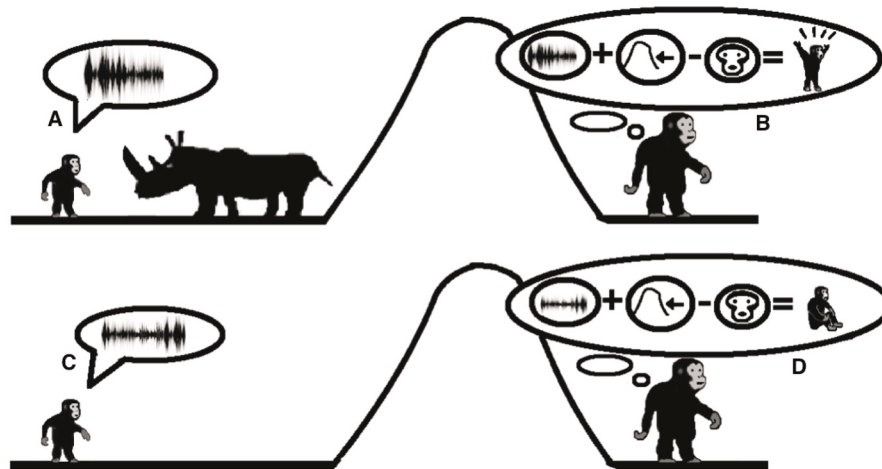


Figure 3. The use of prosody to signal levels of distress. In accordance with the model, *early* Hominans became capable of modifying their contact calls with intonations (prosody). This modification could have originated for the purpose of expressing different levels of distress. In this figure, we see a *Homo habilis* child using prosody to modify the contact call to express a high level of distress (A) or a low level of distress (C). The child's mother then registers the call (by integrating his prosodic intonation and voice, location, and the absence of his face) to recognize whether her child requires immediate (B) or non-immediate (D) attention.

the ability to ask 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 (Figure 4). Eventually, the infant emitted the question call and waited for an appropriate answer from their mother before proceeding with their intended action. This conversation structure could be the precursor to present-day yes/no questions.

The proto-conversations described so far are very limited in their content as the meaning of each call is dependent on context. In order for speech to become more versatile, early *Hominans* needed a method for acquiring vocabulary. A possible route for the acquisition of words is that the prevalence of using intonations gradually resulted with increase in volitional control over the vocal apparatus.

Eventually, vocal control was sufficient for inventing novel calls. Offspring, which so far communicated vocally with their parents for signaling interest in interacting with objects, began mimicking their parents' vocal response. Eventually, by practicing mimicry, the offspring learned the names of objects and enhanced their vocabulary. Transitioning to children demonstrating curiosity for the names of objects could have also prompted the curiosity towards the unknown that characterizes our species. This period of mimicry in language development could be the reason present day babies constantly mimic their parents' vocalizations. In depth discussion about the role of vocal mimicry in language development and its relation with the ADS is beyond the scope of the present paper. However, an evolutionary account of the emergence of language from mimicry based conversations and its relation with the ADS and AVS is discussed in detail in a follow up paper, titled 'From Mimicry to Language: A Neuroanatomically Based Evolutionary Model of the Emergence of Vocal Language' (Poliva, 2016).

9. Comparisons of the 'From Where to What' model to previous language evolution models

Following in the footsteps of Dean Falk and Nobuo Masataka, the present model argues that human speech emerged from the exchange of contact calls via a transitory prosodic phase. Since the principle of natural selection was first acknowledged by the scientific community however, several other accounts of language evolution were proposed. Here, I'll present two schools of thought, and discuss their validity in the context of the present model.

The earliest model for language evolution was proposed by Charles Darwin. In his book, *The Descent of Man* (1871), Darwin equated speech exchange to bird song, and proposed that the perception and production of songs during mating rituals were the precursor to human language (singing ape hypothesis). Similar accounts suggesting music to participate in the evolutionary development

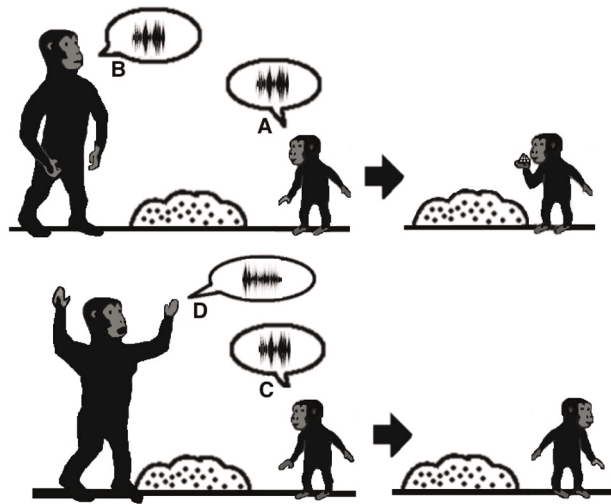


Figure 4. Prosody and the emergence of question-answer conversations. In accordance with the model, the modification of contact calls with intonations for reporting distress levels eventually transitioned into question-answer conversations about items in their environment. In this figure, a child is using low-level distress call (A,C) to ask permission to eat an unfamiliar food (berries). The mother can then respond with a high-level distress call (D) that signals danger or a low-level distress (B) that signals safety.

of speech were also proposed by more recent researchers (Jordania, 2006; Masataka, 2009; Mithen, 2006). However, so far the idea of music as precursor to language has not taken hold in the scientific community due to lack of substantiating evidence. In Appendix A, I cite evidence that the perception of melodies occurs in the aSTG of the AVS. Given the mounting evidence indicating that speech is processed primarily in the ADS, we would expect that precursors to speech would also be processed in the same pathway (although, see the review by Stewart *et al.*, 2006 who suggests roles also for other auditory fields in music perception). Since I hypothesize that singing-like calls were utilized for communication prior to complex vocal language, the idea of music perception and production isn't too different from the present model. However, arguing that music served as precursor to speech is different than arguing that music and speech emerged from a common proto-function. Investigating whether music served as a precursor to vocal language is problematic since such a model implies that music perception is a unique human trait. Therefore, in order to resolve the conundrum of music evolution and its level of contribution to the emergence of vocal language, future studies should first attempt to determine whether non-human primates can perceive music. (See Remedios *et al.*, 2009b for preliminary findings).

A more recent school of thought argues that language with complex semantics and grammar was first communicated via the exchange of gestures and only recently became vocal (Gestural language model; Arbib, 2008; Corballis, 2010; Donald, 2005; Gentilucci & Corballis, 2006; Hewes, 1973; Studdert-Kennedy, 2005). In accordance with this model, speech could have served for increasing communication distance and enabling communication under low visibility conditions (e.g., night, caves). This model is primarily based on the natural use of gestural communication between non-human primates, the ability of apes to learn sign language, and the natural development of sign languages in deaf communities. This model also

received increased popularity since the discovery of mirror neurons, as these neurons are interpreted by proponents of the model as evidence of a mechanism dedicated to the imitation of gestures. From a neuroanatomical perspective it is plausible that vocal communication emerged from gestures. For instance, an fMRI study correlated hearing animal calls with bilateral activation in the mSTG-aSTG, whereas hearing manual tool sounds (e.g., hammer, saw) correlated with activation in the pSTG and IPL of the hemisphere contralateral to the dominant hand (Lewis *et al.*, 2006). This recognition of tool sounds in the ADS instead of AVS is surprising because it could suggest that the teaching of tool use, which required gestures, was associated with speech production. This view is also supported by a study that reported of an area that is co-selective to the detection of hands and manual tools (i.e., area responsible for the perception of tool usage), which is located near the pSTG (Bracci *et al.*, 2012), and not in the area most often responsible for visual object recognition, the inferior temporal lobe. Finally, it is interesting to note that damage to the ADS (areas Spt, IPL and IFG) in the left hemisphere were strongly associated with errors gesturing tool use (Manuel *et al.*, 2013). Based on these findings I find the hypothesis that speech and gestures co-evolved compelling. However, given that my model delineates a course for the development of proto-conversations from calls that are used by extant primates, it is incongruent with the argument that a gestural language with complex grammar and semantics preceded vocal language.

10. 'From Where to What'- Future Research

In the present paper, I delineate a course for the early development of language by proposing four hypotheses: 1. In non-human primates, the ADS is responsible for perceiving and responding to contact calls; 2. Mother-offspring vocal exchange was the predominant force that guided the emergence of speech in the ADS; 3. Speech emerged from modifying calls with intonations for signaling a low-level and high-level of distress, and these calls are the

precursor to our use of intonations for converting words into questions and commands, respectively. 4. Asking questions is a unique human characteristic and the primary driving force behind our species' cognitive success. Cumulative and converging evidence for the veracity of each of these hypotheses was provided throughout the paper. However, as the veracity of a model can only be measured by its ability to predict experimental results, I will present here outlines for 4 potential studies that can test each of these hypotheses.

In accordance with the first hypothesis, the ADS of non-human primates is responsible for the detection and vocal response to contact calls. A possible way of testing this hypothesis is by inducing bilateral lesions to the temporo-parietal junction of a monkey and then measuring whether the monkey no longer responds vocally to contact calls or responds less than before the lesion induction.

In accordance with the second hypothesis, mother infant interaction was the guiding force that endowed the ADS with its role in speech. This hypothesis is primarily based on the finding that a sub-region of area Spt in human fetuses was shown selective to the voice of their mothers (Jardri *et al.*, 2012). Future studies should further explore whether this region remains active in the brain of infants and toddlers and whether mothers also possess a region in the ADS that is selective to the voice of their children.

In accordance with the third hypothesis, the ADS originally served for discriminating calls that signal different levels of distress by analyzing their intonations. At present day, this development is reflected in our ability to modify intonations for converting spoken words into questions and commands. A way of testing this hypothesis is by using fMRI to compare the brain regions active when participants discriminate spoken words into questions and commands, with the brain regions active when they discriminate these words based on their emotional content (e.g., scared and happy). I predict that the former will activate the ADS whereas the latter the AVS.

In accordance with the fourth hypothesis, the unique human mind is the result of our ability to ask questions. To test whether this hypothesis is true, when teaching apes sign language, more effort should be allocated in training them to ask questions.

Competing interests

No competing interests were disclosed.

Grant information

The author(s) declared that no grants were involved in supporting this work.

Acknowledgments

First, I would like to thank my advisor and mentor, Robert Rafal for his advice, comments and support when writing this paper. I would also like to thank Ben Crossey, Iva Ivanova, Cait Jenkins, Ruth Fishman and Catherine Le Pape for their help with reviewing this paper; and to the editors of American Journal Experts, Journal Prep and NPG language editing for their participation in the editing, proofreading and reviewing of this paper at its different stages.

Appendix A: The auditory ventral stream and its role in sound recognition

Accumulative converging evidence indicates that the AVS is involved in recognizing auditory objects. At the level of the primary auditory cortex, recordings from monkeys showed higher percentage of neurons selective for learned melodic sequences in area R than area A1 (Yin *et al.*, 2008), and a study in humans demonstrated more selectivity for heard syllables in the anterior Heschl's gyrus (area hR) than posterior Heschl's gyrus (area hA1; Steinschneider *et al.*, 2005). In downstream associative auditory fields, studies from both monkeys and humans reported that the border between the anterior and posterior auditory fields (Figure 1-area PC in the monkey and mSTG in the human) processes pitch attributes that are necessary for the recognition of auditory objects (Bendor & Wang, 2006). The anterior auditory fields of monkeys were also demonstrated with selectivity for con-specific vocalizations with intra-cortical recordings (Perrodin *et al.*, 2011; Rauschecker *et al.*, 1995; Russ *et al.*, 2008) and functional imaging (Joly *et al.*, 2012; Petkov *et al.*, 2008; Poremba *et al.*, 2004). One fMRI monkey study further demonstrated a role of the aSTG in the recognition of individual voices (Petkov *et al.*, 2008). The role of the human mSTG-aSTG in sound recognition was demonstrated via functional imaging studies that correlated activity in this region with isolation of auditory objects from background noise (Scheich *et al.*, 1998; Zatorre *et al.*, 2004) and with the recognition of spoken words (Binder *et al.*, 2004; Davis & Johnsrude, 2003; Liebenthal *et al.*, 2005; Narain *et al.*, 2003; Obleser *et al.*, 2006; Obleser *et al.*, 2007; Scott *et al.*, 2000), voices (Belin & Zatorre, 2003), melodies (Benson *et al.*, 2001; Leaver & Rauschecker, 2010), environmental sounds (Lewis *et al.*, 2006; Maeder *et al.*, 2001; Viceic *et al.*, 2006), and non-speech communicative sounds (Shultz *et al.*, 2012). A Meta-analysis of fMRI studies (DeWitt & Rauschecker, 2012) further demonstrated functional dissociation between the left mSTG and aSTG, with the former processing short speech units (phonemes) and the latter processing longer units (e.g., words, environmental sounds). A study that recorded neural activity directly from the left pSTG and aSTG reported that the aSTG, but not pSTG, was more active when the patient listened to speech in her native language than unfamiliar foreign language (Lachaux *et al.*, 2007-patient 1). Consistently, electro stimulation to the aSTG of this patient resulted in impaired speech perception (Lachaux *et al.*, 2007-patient 1; see also Matsumoto *et al.*, 2011; Roux *et al.*, 2015 for similar results). Intra-cortical recordings from the right and left aSTG further demonstrated that speech is processed laterally to music (Lachaux *et al.*, 2007-patient 2). An fMRI study of a patient with impaired sound recognition (auditory agnosia) due to brainstem damage was also shown with reduced activation in areas hR and aSTG of both hemispheres when hearing spoken words and environmental sounds (Poliva *et al.*, 2015). Recordings from the anterior auditory cortex of monkeys while maintaining learned sounds in working memory (Tsunada *et al.*, 2011), and the debilitating effect of induced lesions to this region on working memory recall (Fritz *et al.*, 2005; Stepien *et al.*, 1960; Strominger *et al.*, 1980), further implicate the AVS in maintaining the perceived auditory objects in working memory. In humans, area mSTG-aSTG was also reported active during rehearsal of heard syllables with MEG (Kaiser *et al.*, 2003) and fMRI (Buchsbbaum *et al.*, 2005). The latter study further demonstrated that working

memory in the AVS is for the acoustic properties of spoken words and that it is independent to working memory in the ADS, which mediates inner speech. Working memory studies in monkeys also suggest that in monkeys, in contrast to humans, the AVS is the dominant working memory store (Scott *et al.*, 2012).

In humans, downstream to the aSTG, the MTG and TP are thought to constitute the semantic lexicon, which is a long-term memory repository of audio-visual representations that are interconnected on the basis of semantic relationships. (See also the reviews by Hickok & Poeppel, 2007 and Gow, 2012, discussing this topic). The primary evidence for this role of the MTG-TP is that patients with damage to this region (e.g., patients with semantic dementia or herpes simplex virus encephalitis) are reported with an impaired ability to describe visual and auditory objects and a tendency to commit semantic errors when naming objects (i.e., semantic paraphasia; Noppeney *et al.*, 2007; Patterson *et al.*, 2007). Semantic paraphasias were also expressed by aphasic patients with left MTG-TP damage (Dronkers *et al.*, 2004; Schwartz *et al.*, 2009) and were shown to occur in non-aphasic patients after electro-stimulation to this region (Hamberger *et al.*, 2007; Roux *et al.*, 2015) or the underlying white matter pathway (Duffau, 2008). Two meta-analyses of the fMRI literature also reported that the anterior MTG and TP were consistently active during semantic analysis of speech and text (Binder *et al.*, 2009; Vigneau *et al.*, 2006); and an intra-cortical recording study correlated neural

discharge in the MTG with the comprehension of intelligible sentences (Creutzfeldt *et al.*, 1989).

In contradiction to the Wernicke-Lichtheim-Geschwind model that implicates sound recognition to occur solely in the left hemisphere, studies that examined the properties of the right or left hemisphere in isolation via unilateral hemispheric anesthesia (i.e., the WADA procedure; Hickok *et al.*, 2008) or intra-cortical recordings from each hemisphere (Creutzfeldt *et al.*, 1989) provided evidence that sound recognition is processed bilaterally. Moreover, a study that instructed patients with disconnected hemispheres (i.e., split-brain patients) to match spoken words to written words presented to the right or left hemifields, reported vocabulary in the right hemisphere that almost matches in size with the left hemisphere (Zaidel, 1976). (The right hemisphere vocabulary was equivalent to the vocabulary of a healthy 11-years old child). This bilateral recognition of sounds is also consistent with the finding that unilateral lesion to the auditory cortex rarely results in deficit to auditory comprehension (i.e., auditory agnosia), whereas a second lesion to the remaining hemisphere (which could occur years later) does (Poeppel, 2001; Ulrich, 1978). Finally, as mentioned earlier, an fMRI scan of an auditory agnosia patient demonstrated bilateral reduced activation in the anterior auditory cortices (Poliva *et al.*, 2015), and bilateral electro-stimulation to these regions in both hemispheres resulted with impaired speech recognition (Lachaux *et al.*, 2007-patient 2).

References

- Abowitz F, García VR: **The evolutionary origin of the language areas in the human brain. A neuroanatomical perspective.** *Brain Res Brain Res Rev.* 1997; **25**(3): 381–396.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Acheson DJ, Hamidi M, Binder JR, *et al.*: **A common neural substrate for language production and verbal working memory.** *J Cogn Neurosci.* 2011; **23**(6): 1358–1367.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Ahveninen J, Jääskeläinen IP, Raij T, *et al.*: **Task-modulated “what” and “where” pathways in human auditory cortex.** *Proc Natl Acad Sci U S A.* 2006; **103**(39): 14608–14613.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Aitken PG: **Cortical control of conditioned and spontaneous vocal behavior in rhesus monkeys.** *Brain Lang.* 1981; **13**(1): 171–184.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Alain C, Arnott SR, Hevenor S, *et al.*: **“What” and “where” in the human auditory system.** *Proc Natl Acad Sci U S A.* 2001; **98**(21): 12301–12306.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Anderson JM, Gilmore R, Roper S, *et al.*: **Conduction aphasia and the arcuate fasciculus: A reexamination of the Wernicke-Geschwind model.** *Brain Lang.* 1999; **70**(1): 1–12.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Andics A, Gácsi M, Faragó T, *et al.*: **Voice-sensitive regions in the dog and human brain are revealed by comparative fMRI.** *Curr Biol.* 2014; **24**(5): 574–578.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Andics A, McQueen JM, Petersson KM, *et al.*: **Neural mechanisms for voice recognition.** *Neuroimage.* 2010; **52**(4): 1528–1540.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Anourova I, Nikouline VV, Ilmoniemi RJ, *et al.*: **Evidence for dissociation of spatial and nonspatial auditory information processing.** *Neuroimage.* 2001; **14**(6): 1268–1277.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Arbib MA: **From grasp to language: embodied concepts and the challenge of abstraction.** *J Physiol Paris.* 2008; **102**(1–3): 4–20.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Arcadi AC: **Vocal responsiveness in male wild chimpanzees: implications for the evolution of language.** *J Hum Evol.* 2000; **39**(2): 205–223.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Barrett DJ, Hall DA: **Response preferences for “what” and “where” in human non-primary auditory cortex.** *Neuroimage.* 2006; **32**(2): 968–977.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Baumgart F, Gaschler-Markefski B, Woldorff MG, *et al.*: **A movement-sensitive area in auditory cortex.** *Nature.* 1999; **400**(6746): 724–726.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Belin P, Zatorre RJ: **Adaptation to speaker’s voice in right anterior temporal lobe.** *Neuroreport.* 2003; **14**(16): 2105–2109.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Belton E, Salmond CH, Watkins KE, *et al.*: **Bilateral brain abnormalities associated with dominantly inherited verbal and orofacial dyspraxia.** *Hum Brain Mapp.* 2003; **18**(3): 194–200.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Bendor D, Wang X: **Cortical representations of pitch in monkeys and humans.** *Curr Opin Neurobiol.* 2006; **16**(4): 391–399.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Benson DA, Hienz RD, Goldstein MH Jr: **Single-unit activity in the auditory cortex of monkeys actively localizing sound sources: spatial tuning and**

- behavioral dependency. *Brain Res.* 1981; **219**(2): 249–267.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Benson RR, Whalen DH, Richardson M, *et al.*: **Parametrically dissociating speech and nonspeech perception in the brain using fMRI.** *Brain Lang.* 2001; **78**(3): 364–396.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Biben M, Symmes D, Bernhards D: **Contour variables in vocal communication between squirrel monkey mothers and infants.** *Dev Psychobiol.* 1989; **22**(6): 617–631.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Biben M, Symmes D, Masataka N: **Temporal and structural analysis of affiliative vocal exchanges in squirrel monkeys (*Saimiri sciureus*).** *Behaviour.* 1986; **98**(1): 259–273.
[Publisher Full Text](#)
- Biben M: **Allomaternal vocal behavior in squirrel monkeys.** *Dev Psychobiol.* 1992; **25**(2): 79–92.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Binder JR, Desai RH, Graves WW, *et al.*: **Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies.** *Cereb Cortex.* 2009; **19**(12): 2767–2796.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Binder JR, Liebenthal E, Possing ET, *et al.*: **Neural correlates of sensory and decision processes in auditory object identification.** *Nat Neurosci.* 2004; **7**(3): 295–301.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Blake J: **Gestural communication in the great apes.** In *The Evolution of Thought: Evolutionary Origins of Great Ape Intelligence.* Cambridge University Press. 2004; 61–75.
[Publisher Full Text](#)
- Bracci S, Cavina-Pratesi C, Ietswaart M, *et al.*: **Closely overlapping responses to tools and hands in left lateral occipitotemporal cortex.** *J Neurophysiol.* 2012; **107**(5): 1443–1456.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Brunetti M, Belardinelli P, Caulo M, *et al.*: **Human brain activation during passive listening to sounds from different locations: an fMRI and MEG study.** *Hum Brain Mapp.* 2005; **26**(4): 251–261.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Buchsbaum BR, Olsen RK, Koch P, *et al.*: **Human dorsal and ventral auditory streams subserve rehearsal-based and echoic processes during verbal working memory.** *Neuron.* 2005; **48**(4): 687–697.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Carlson KJ, Stout D, Jashashvili T, *et al.*: **The endocast of MH1, *Australopithecus sediba*.** *Science.* 2011; **333**(6048): 1402–1407.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Catani M, Jones DK, Fyfe DH: **Perisylvian language networks of the human brain.** *Ann Neurol.* 2004; **57**(1): 8–16.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Chang EF, Edwards E, Nagarajan SS, *et al.*: **Cortical spatio-temporal dynamics underlying phonological target detection in humans.** *J Cogn Neurosci.* 2011; **23**(6): 1437–1446.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Cheney DL, Seyfarth RM: **Vocal recognition in free-ranging vervet monkeys.** *Anim Behav.* 1980; **28**(2): 362–367.
[Publisher Full Text](#)
- Clarke S, Adriani M, Bellmann A: **Distinct short-term memory systems for sound content and sound localization.** *Neuroreport.* 1998; **9**(15): 3433–3437.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Clarke S, Bellmann A, Meuli RA, *et al.*: **Auditory agnosia and auditory spatial deficits following left hemispheric lesions: evidence for distinct processing pathways.** *Neuropsychologia.* 2000; **38**(6): 797–807.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Cohen YE, Russ BE, Gifford GW 3rd, *et al.*: **Selectivity for the spatial and nonspatial attributes of auditory stimuli in the ventrolateral prefrontal cortex.** *J Neurosci.* 2004; **24**(50): 11307–11316.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Corballis MC: **Mirror neurons and the evolution of language.** *Brain Lang.* 2010; **112**(1): 25–35.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Coudé G, Ferrari PF, Rodà F, *et al.*: **Neurons controlling voluntary vocalization in the macaque ventral premotor cortex.** *PLoS One.* 2011; **6**(11): e26822.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Creutzfeldt O, Ojemann G, Lettich E: **Neuronal activity in the human lateral temporal lobe. I. Responses to speech.** *Exp Brain Res.* 1989; **77**(3): 451–475.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Cusick CG, Seltzer B, Cola M, *et al.*: **Chemoarchitectonics and corticocortical terminations within the superior temporal sulcus of the rhesus monkey: evidence for subdivisions of superior temporal polysensory cortex.** *J Comp Neurol.* 1995; **360**(3): 513–535.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Da Costa S, van der Zwaag W, Marques JP, *et al.*: **Human primary auditory cortex follows the shape of Heschl's gyrus.** *J Neurosci.* 2011; **31**(40): 14067–14075.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Darwin C: **The Descent of Man and Selection in Relation to Sex.** Appleton. 1871.
[Publisher Full Text](#)
- Davis MH, Johnsrude IS: **Hierarchical processing in spoken language comprehension.** *J Neurosci.* 2003; **23**(8): 3423–3431.
[PubMed Abstract](#)
- de la Mothe LA, Blumell S, Kajikawa Y, *et al.*: **Cortical connections of the auditory cortex in marmoset monkeys: Core and medial belt regions.** *J Comp Neurol.* 2006; **496**(1): 27–71.
[PubMed Abstract](#) | [Publisher Full Text](#)
- de la Mothe LA, Blumell S, Kajikawa Y, *et al.*: **Cortical connections of auditory cortex in marmoset monkeys: lateral belt and parabelt regions.** *Anat Rec (Hoboken).* 2012; **295**(5): 800–821.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- De Santis L, Clarke S, Murray MM: **Automatic and intrinsic auditory “what” and “where” processing in humans revealed by electrical neuroimaging.** *Cereb Cortex.* 2007; **17**(1): 9–17.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Deacon TW: **Cortical connections of the inferior arcuate sulcus cortex in the macaque brain.** *Brain Res.* 1992; **573**(1): 8–26.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Desmurget M, Reilly KT, Richard N, *et al.*: **Movement intention after parietal cortex stimulation in humans.** *Science.* 2009; **324**(5928): 811–813.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Deutsch SE: **Prediction of site of lesion from speech apraxic error patterns. In apraxia of speech: physiology, acoustics, linguistics, management.** College Hill Pr. 1984; 113–134.
- DeWitt I, Rauschecker JP: **Phoneme and word recognition in the auditory ventral stream.** *Proc Natl Acad Sci U S A.* 2012; **109**(8): E505–14.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- DeWitt I, Rauschecker JP: **Wernicke's area revisited: parallel streams and word processing.** *Brain Lang.* 2013; **127**(2): 181–191.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Donald M: **Imitation and Mimesis.** In *Perspectives on Imitation: Mechanisms of Imitation in animals by Hurley and Chater.* MIT Press. 2005; 284–300.
[Reference Source](#)
- Dronkers NF, Redfern BB, Knight RT: **The neural architecture of language disorders.** In M. S. Gazzaniga (Ed.), *The Cognitive Neurosciences.* Cambridge MA MIT Press. 1999; 949–958.
[Reference Source](#)
- Dronkers NF, Wilkins DP, Van Valin RD Jr, *et al.*: **Lesion analysis of the brain areas involved in language comprehension.** *Cognition.* 2004; **92**(1–2): 145–177.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Dronkers NF: **The pursuit of brain-language relationships.** *Brain Lang.* 2000; **71**(1): 59–61.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Duffau H: **The anatomo-functional connectivity of language revisited. New insights provided by electrostimulation and tractography.** *Neuropsychologia.* 2008; **46**(4): 927–934.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Edmonds L, Marquardt T: **Syllable use in apraxia of speech: Preliminary findings.** *Aphasiology.* 2004; **18**(12): 1121–1134.
[Publisher Full Text](#)
- Efron R, Crandall PH: **Central auditory processing. II. Effects of anterior temporal lobectomy.** *Brain Lang.* 1983; **19**(2): 237–253.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Falk D: **Prelinguistic evolution in early hominins: whence motherese?** *Behav Brain Sci.* 2004; **27**(4): 491–503.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Formisano E, De Martino F, Bonte M, *et al.*: **“Who” is saying “what”? Brain-based decoding of human voice and speech.** *Science.* 2008; **322**(5903): 970–973.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Frey S, Campbell JS, Pike GB, *et al.*: **Dissociating the human language pathways with high angular resolution diffusion fiber tractography.** *J Neurosci.* 2008; **28**(45): 11435–11444.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Fritz J, Mishkin M, Saunders RC: **In search of an auditory engram.** *Proc Natl Acad Sci U S A.* 2005; **102**(26): 9359–9364.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Geiser E, Zaehle T, Jancke L, *et al.*: **The neural correlate of speech rhythm as evidenced by metrical speech processing.** *J Cogn Neurosci.* 2008; **20**(3): 541–552.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Gelfand JR, Bookheimer SY: **Dissociating neural mechanisms of temporal sequencing and processing phonemes.** *Neuron.* 2003; **38**(5): 831–842.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Gemba H, Kyuhou S, Matsuzaki R, *et al.*: **Cortical field potentials associated with audio-initiated vocalization in monkeys.** *Neurosci Lett.* 1999; **272**(1): 49–52.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Gentilucci M, Corballis MC: **From manual gesture to speech: a gradual transition.** *Neurosci Biobehav Rev.* 2006; **30**(7): 949–960.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Geschwind N: **Disconnexion syndromes in animals and man. I.** *Brain.* 1965;

88(2): 237–294.

[PubMed Abstract](#) | [Publisher Full Text](#)

Ghazanfar AA, Maier JX, Hoffman KL, *et al.*: **Multisensory integration of dynamic faces and voices in rhesus monkey auditory cortex.** *J Neurosci.* 2005; 25(20): 5004–5012.

[PubMed Abstract](#) | [Publisher Full Text](#)

Gibson KR: **Language or protolanguage? A review of the ape language literature.** In *The Oxford Handbook of Language Evolution.* Oxford University Press, USA. 2011; 46–58.

[Publisher Full Text](#)

Gifford GW 3rd, Cohen YE: **Spatial and non-spatial auditory processing in the lateral intraparietal area.** *Exp Brain Res.* 2005; 162(4): 509–512.

[PubMed Abstract](#) | [Publisher Full Text](#)

Gil-da-Costa R, Martin A, Lopes MA, *et al.*: **Species-specific calls activate homologs of Broca's and Wernicke's areas in the macaque.** *Nat Neurosci.* 2006; 9(8): 1064–1070.

[PubMed Abstract](#) | [Publisher Full Text](#)

Goodall J: **The chimpanzees of Gombe: patterns of behavior.** Belknap Press, 1986.

[Reference Source](#)

Gorno-Tempini ML, Brambati SM, Ginex V, *et al.*: **The logopenic/phonological variant of primary progressive aphasia.** *Neurology.* 2008; 71(16): 1227–1234.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Gottlieb Y, Vaadia E, Abeles M: **Single unit activity in the auditory cortex of a monkey performing a short term memory task.** *Exp Brain Res.* 1989; 74(1): 139–148.

[PubMed Abstract](#) | [Publisher Full Text](#)

Gourévitch B, Le Bouquin Jeannès R, Faucon G, *et al.*: **Temporal envelope processing in the human auditory cortex: response and interconnections of auditory cortical areas.** *Hear Res.* 2008; 237(1–2): 1–18.

[PubMed Abstract](#) | [Publisher Full Text](#)

Gow DW Jr: **The cortical organization of lexical knowledge: a dual lexicon model of spoken language processing.** *Brain Lang.* 2012; 121(3): 273–288.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Griffiths TD, Rees A, Witton C, *et al.*: **Evidence for a sound movement area in the human cerebral cortex.** *Nature.* 1996; 383(6599): 425–427.

[PubMed Abstract](#) | [Publisher Full Text](#)

Guéguin M, Le Bouquin-Jeannès R, Faucon G, *et al.*: **Evidence of functional connectivity between auditory cortical areas revealed by amplitude modulation sound processing.** *Cereb Cortex.* 2007; 17(2): 304–313.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Hage SR, Jürgens U: **Localization of a vocal pattern generator in the pontine brainstem of the squirrel monkey.** *Eur J Neurosci.* 2006; 23(3): 840–844.

[PubMed Abstract](#) | [Publisher Full Text](#)

Hamberger MJ, McClelland S 3rd, McKhann GM 2nd, *et al.*: **Distribution of auditory and visual naming sites in nonlesional temporal lobe epilepsy patients and patients with space-occupying temporal lobe lesions.** *Epilepsia.* 2007; 48(3): 531–538.

[PubMed Abstract](#) | [Publisher Full Text](#)

Hannig S, Jürgens U: **Projections of the ventrolateral pontine vocalization area in the squirrel monkey.** *Exp Brain Res.* 2006; 169(1): 92–105.

[PubMed Abstract](#) | [Publisher Full Text](#)

Hart HC, Palmer AR, Hall DA: **Different areas of human non-primary auditory cortex are activated by sounds with spatial and nonspatial properties.** *Hum Brain Mapp.* 2004; 21(3): 178–190.

[PubMed Abstract](#) | [Publisher Full Text](#)

Hayes KJ, Hayes C: **Imitation in a home-raised chimpanzee.** *J Comp Physiol Psychol.* 1952; 45(5): 450–459.

[PubMed Abstract](#) | [Publisher Full Text](#)

Heffner HE, Heffner RS: **Temporal lobe lesions and perception of species-specific vocalizations by macaques.** *Science.* 1984; 226(4670): 75–76.

[PubMed Abstract](#) | [Publisher Full Text](#)

Heimbauer LA, Beran MJ, Owren MJ: **A chimpanzee recognizes synthetic speech with significantly reduced acoustic cues to phonetic content.** *Curr Biol.* 2011; 21(14): 1210–1214.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Hewes GW: **Primate communication and the gestural origin of language.** *Curr Anthropol.* 1973; 14(1/2): 5–24.

[Reference Source](#)

Hickok G, Buchsbaum B, Humphries C, *et al.*: **Auditory-motor interaction revealed by fMRI: speech, music, and working memory in area Spt.** *J Cogn Neurosci.* 2003; 15(5): 673–682.

[PubMed Abstract](#)

Hickok G, Okada K, Barr W, *et al.*: **Bilateral capacity for speech sound processing in auditory comprehension: evidence from Wada procedures.** *Brain Lang.* 2008; 107(3): 179–184.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Hickok G, Poeppel D: **The cortical organization of speech processing.** *Nat Rev Neurosci.* 2007; 8(5): 393–402.

[PubMed Abstract](#) | [Publisher Full Text](#)

Hihara S, Yamada H, Iriki A, *et al.*: **Spontaneous vocal differentiation of coo-calls for tools and food in Japanese monkeys.** *Neurosci Res.* 2003; 45(4): 383–389.

[PubMed Abstract](#) | [Publisher Full Text](#)

Hillis AE, Work M, Barker PB, *et al.*: **Re-examining the brain regions crucial for orchestrating speech articulation.** *Brain.* 2004; 127(Pt 7): 1479–1487.

[PubMed Abstract](#) | [Publisher Full Text](#)

Holstege G, Kerstens L, Moes MC, *et al.*: **Evidence for a periaqueductal gray-nucleus retroambiguus-spinal cord pathway in the rat.** *Neuroscience.* 1997; 80(2): 587–598.

[PubMed Abstract](#) | [Publisher Full Text](#)

Holstege G: **Anatomical study of the final common pathway for vocalization in the cat.** *J Comp Neurol.* 1989; 284(2): 242–252.

[PubMed Abstract](#) | [Publisher Full Text](#)

Hopkins WD, Tagliatela JP, Leavens DA: **Chimpanzees Differentially Produce Novel Vocalizations to Capture the Attention of a Human.** *Anim Behav.* 2007; 73(2): 281–286.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Humphries C, Liebenthal E, Binder JR: **Tonotopic organization of human auditory cortex.** *Neuroimage.* 2010; 50(3): 1202–1211.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Ischebeck A, Indefrey P, Usui N, *et al.*: **Reading in a regular orthography: an fMRI study investigating the role of visual familiarity.** *J Cogn Neurosci.* 2004; 16(5): 727–741.

[PubMed Abstract](#) | [Publisher Full Text](#)

Jardri R, Houfflin-Debarge V, Delion P, *et al.*: **Assessing fetal response to maternal speech using a noninvasive functional brain imaging technique.** *Int J Dev Neurosci.* 2012; 30(2): 159–161.

[PubMed Abstract](#) | [Publisher Full Text](#)

Joly O, Pallier C, Ramus F, *et al.*: **Processing of vocalizations in humans and monkeys: a comparative fMRI study.** *Neuroimage.* 2012; 62(3): 1376–1389.

[PubMed Abstract](#) | [Publisher Full Text](#)

Jordania J: **Who Asked the First Question? The Origins of Human Choral Singing, Intelligence, Language and Speech.** Tbilisi: Logos, 2006; 334–338.

[Reference Source](#)

Josephs KA, Duffy JR, Strand EA, *et al.*: **Clinicopathological and imaging correlates of progressive aphasia and apraxia of speech.** *Brain.* 2006; 129(Pt 6): 1385–1398.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Jürgens U, Alipour M: **A comparative study on the cortico-hypoglossal connections in primates, using biotin dextranamine.** *Neurosci Lett.* 2002; 328(3): 245–248.

[PubMed Abstract](#) | [Publisher Full Text](#)

Jürgens U, Ploog D: **Cerebral representation of vocalization in the squirrel monkey.** *Exp Brain Res.* 1970; 10(5): 532–554.

[PubMed Abstract](#) | [Publisher Full Text](#)

Kaas JH, Hackett TA: **Subdivisions of auditory cortex and processing streams in primates.** *Proc Natl Acad Sci U S A.* 2000; 97(22): 11793–11799.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Kaiser J, Ripper B, Birbaumer N, *et al.*: **Dynamics of gamma-band activity in human magnetoencephalogram during auditory pattern working memory.** *Neuroimage.* 2003; 20(2): 816–827.

[PubMed Abstract](#) | [Publisher Full Text](#)

Kalan AK, Mundry R, Boesch C: **Wild chimpanzees modify food call structure with respect to tree size for a particular fruit species.** *Anim Behav.* 2015; 101: 1–9.

[Publisher Full Text](#)

Kaminski J, Call J, Fischer J: **Word learning in a domestic dog: evidence for “fast mapping”.** *Science.* 2004; 304(5677): 1682–1683.

[PubMed Abstract](#) | [Publisher Full Text](#)

Kayser C, Petkov CI, Logothetis NK: **Multisensory interactions in primate auditory cortex: fMRI and electrophysiology.** *Hear Res.* 2009; 258(1–2): 80–88.

[PubMed Abstract](#) | [Publisher Full Text](#)

Kimura D, Watson N: **The relation between oral movement control and speech.** *Brain Lang.* 1989; 37(4): 565–590.

[PubMed Abstract](#)

Koda H, Nishimura T, Tokuda IT, *et al.*: **Soprano singing in gibbons.** *Am J Phys Anthropol.* 2012; 149(3): 347–355.

[PubMed Abstract](#) | [Publisher Full Text](#)

Koda H, Oyakawa C, Kato A, *et al.*: **Experimental evidence for the volitional control of vocal production in an immature gibbon.** *Behaviour.* 2007; 144(6): 681–692.

[Publisher Full Text](#)

Kosmal A, Malinowska M, Kowalska DM: **Thalamic and amygdaloid connections of the auditory association cortex of the superior temporal gyrus in rhesus monkey (*Macaca mulatta*).** *Acta Neurobiol Exp (Wars).* 1997; 57(3): 165–188.

[PubMed Abstract](#)

Krumbholz K, Schönwiesner M, Rübsamen R, *et al.*: **Hierarchical processing of sound location and motion in the human brainstem and planum temporale.** *Eur J Neurosci.* 2005; 21(1): 230–238.

[PubMed Abstract](#) | [Publisher Full Text](#)

Lachaux JP, Jerbi K, Bertrand O, *et al.*: **A blueprint for real-time functional mapping via human intracranial recordings.** *PLoS One.* 2007; 2(10): e1094.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Lameira AR, Hardus ME, Bartlett AM, *et al.*: **Speech-like rhythm in a voiced and voiceless orangutan call.** *PLoS One.* 2015; 10(1): e116136.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

- Langers DRM, van Dijk P: **Mapping the tonotopic organization in human auditory cortex with minimally salient acoustic stimulation.** *Cereb Cortex.* 2012; **22**(9): 2024–2038.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Laporte MN, Zuberbühler K: **Vocal greeting behaviour in wild chimpanzee females.** *Anim Behav.* 2010; **80**(3): 467–73.
[Publisher Full Text](#)
- Leaver AM, Rauschecker JP: **Cortical representation of natural complex sounds: effects of acoustic features and auditory object category.** *J Neurosci.* 2010; **30**(22): 7604–7612.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Lewis JW, Phinney RE, Breczynski-Lewis JA, et al.: **Lefties get it “right” when hearing tool sounds.** *J Cogn Neurosci.* 2006; **18**(8): 1314–1330.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Lewis JW, Van Essen DC: **Corticocortical connections of visual, sensorimotor, and multimodal processing areas in the parietal lobe of the macaque monkey.** *J Comp Neurol.* 2000; **428**(1): 112–137.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Lichtheim L: **On aphasia.** *Brain.* 1885; **7**: 433–484.
[Publisher Full Text](#)
- Liebenthal E, Binder JR, Spitzer SM, et al.: **Neural substrates of phonemic perception.** *Cereb Cortex.* 2005; **15**(10): 1621–1631.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Linden JF, Grunewald A, Andersen RA: **Responses to auditory stimuli in macaque lateral intraparietal area. II. Behavioral modulation.** *J Neurophysiol.* 1999; **82**(1): 343–358.
[PubMed Abstract](#)
- Lüthe L, Häusler U, Jürgens U: **Neuronal activity in the medulla oblongata during vocalization. A single-unit recording study in the squirrel monkey.** *Behav Brain Res.* 2000; **116**(2): 197–210.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Lutzenberger W, Ripper B, Busse L, et al.: **Dynamics of gamma-band activity during an audiospatial working memory task in humans.** *J Neurosci.* 2002; **22**(13): 5630–5638.
[PubMed Abstract](#)
- Maeder PP, Meuli RA, Adriani M, et al.: **Distinct pathways involved in sound recognition and localization: a human fMRI study.** *Neuroimage.* 2001; **14**(4): 802–816.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Makris N, Papadimitriou GM, Kaiser JR, et al.: **Delineation of the middle longitudinal fascicle in humans: a quantitative, in vivo, DT-MRI study.** *Cereb Cortex.* 2009; **19**(4): 777–785.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Manuel AL, Radman N, Mesot D, et al.: **Inter- and intrahemispheric dissociations in ideomotor apraxia: a large-scale lesion-symptom mapping study in subacute brain-damaged patients.** *Cereb Cortex.* 2013; **23**(12): 2781–9.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Marler P, Hobbett L: **Individuality in a long-range vocalization of wild chimpanzees.** *Z Tierpsychol.* 1975; **38**(1): 37–109.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Masataka N: **The origins of language and the evolution of music: A comparative perspective.** *Phys Life Rev.* 2009; **6**(1): 11–22.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Matsumoto R, Imamura H, Inouchi M, et al.: **Left anterior temporal cortex actively engages in speech perception: A direct cortical stimulation study.** *Neuropsychologia.* 2011; **49**(5): 1350–1354.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Matsuzawa T: **Evolutionary Origins of the Human Mother-Infant Relationship.** In *Cognitive development in chimpanzees.* Tokyo: Springer-Verlag. 2006; 127–141.
[Publisher Full Text](#)
- Mazzoni P, Bracewell RM, Barash S, et al.: **Spatially tuned auditory responses in area LIP of macaques performing delayed memory saccades to acoustic targets.** *J Neurophysiol.* 1996; **75**(3): 1233–1241.
[PubMed Abstract](#)
- Menjot de Champfleury N, Lima Maldonado I, Moritz-Gasser S, et al.: **Middle longitudinal fasciculus delineation within language pathways: a diffusion tensor imaging study in human.** *Eur J Radiol.* 2013; **82**(1): 151–157.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Mesulam MM, Thompson CK, Weintraub S, et al.: **The Wernicke conundrum and the anatomy of language comprehension in primary progressive aphasia.** *Brain.* 2015; **138**(Pt 8): 2423–37.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Meyer J: **Typology and acoustic strategies of whistled languages: Phonetic comparison and perceptual cues of whistled vowels.** *J Int Phon Assoc.* 2008; **38**(01): 69–94.
[Publisher Full Text](#)
- Meyer M, Steinhauer K, Alter K, et al.: **Brain activity varies with modulation of dynamic pitch variance in sentence melody.** *Brain Lang.* 2004; **89**(2): 277–289.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Miller CT, Dimauro A, Pistorio A, et al.: **Vocalization Induced CFos Expression in Marmoset Cortex.** *Front Integr Neurosci.* 2010; **4**: 128.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Miller LM, Recanzone GH: **Populations of auditory cortical neurons can accurately encode acoustic space across stimulus intensity.** *Proc Natl Acad Sci U S A.* 2009; **106**(14): 5931–5935.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Mitani JC, Nishida T: **Contexts and social correlates of long-distance calling by male chimpanzees.** *Anim Behav.* 1993; **45**(4): 735–746.
[Publisher Full Text](#)
- Mithen S: **The Singing Neanderthals: the Origins of Music, Language, Mind and Body.** Harvard University Press. 2006.
[Reference Source](#)
- Morel A, Garraghty PE, Kaas JH: **Tonotopic organization, architectonic fields, and connections of auditory cortex in macaque monkeys.** *J Comp Neurol.* 1993; **335**(3): 437–459.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Mullette-Gillman OA, Cohen YE, Groh JM: **Eye-centered, head-centered, and complex coding of visual and auditory targets in the intraparietal sulcus.** *J Neurophysiol.* 2005; **94**(4): 2331–52.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Muñoz M, Mishkin M, Saunders RC: **Resection of the medial temporal lobe disconnects the rostral superior temporal gyrus from some of its projection targets in the frontal lobe and thalamus.** *Cereb Cortex.* 2009; **19**(9): 2114–2130.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Nakamura K, Kawashima R, Sugiura M, et al.: **Neural substrates for recognition of familiar voices: a PET study.** *Neuropsychologia.* 2001; **39**(10): 1047–1054.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Narain C, Scott SK, Wise RJ, et al.: **Defining a left-lateralized response specific to intelligible speech using fMRI.** *Cereb Cortex.* 2003; **13**(12): 1362–1368.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Noppeney U, Patterson K, Tyler LK, et al.: **Temporal lobe lesions and semantic impairment: a comparison of herpes simplex virus encephalitis and semantic dementia.** *Brain.* 2007; **130**(pt 4): 1138–1147.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Obleser J, Boecker H, Drzezga A, et al.: **Vowel sound extraction in anterior superior temporal cortex.** *Hum Brain Mapp.* 2006; **27**(7): 562–571.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Obleser J, Zimmermann J, Van Meter J, et al.: **Multiple stages of auditory speech perception reflected in event-related fMRI.** *Cereb Cortex.* 2007; **17**(10): 2251–2257.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Odell K, McNeil MR, Rosenbek JC, et al.: **Perceptual characteristics of vowel and prosody production in apraxic, aphasic, and dysarthric speakers.** *J Speech Hear Res.* 1991; **34**(1): 67–80.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Odell K, Shriberg DL: **Prosody-voice characteristics of children and adults with apraxia of speech.** *Clin Linguist Phon.* 2001; **15**(4): 275–307.
[Publisher Full Text](#)
- Patterson K, Nestor PJ, Rogers TT: **Where do you know what you know? The representation of semantic knowledge in the human brain.** *Nat Rev Neurosci.* 2007; **8**(12): 976–987.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Pavani F, Macaluso E, Warren JD, et al.: **A common cortical substrate activated by horizontal and vertical sound movement in the human brain.** *Curr Biol.* 2002; **12**(18): 1584–1590.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Perلمان M, Clark N: **Learned vocal and breathing behavior in an enculturated gorilla.** *Anim Cogn.* 2015; **18**(5): 1165–79.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Perrodin C, Kayser C, Logothetis NK, et al.: **Voice cells in the primate temporal lobe.** *Curr Biol.* 2011; **21**(16): 1408–1415.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Petersen MR, Beecher MD, Zoloth SR, et al.: **Neural lateralization of species-specific vocalizations by Japanese macaques (*Macaca fuscata*).** *Science.* 1978; **202**(4365): 324–327.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Petkov CI, Kayser C, Augath M, et al.: **Functional imaging reveals numerous fields in the monkey auditory cortex.** *PLoS Biol.* 2006; **4**(7): e215.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Petkov CI, Kayser C, Steudel T, et al.: **A voice region in the monkey brain.** *Nat Neurosci.* 2008; **11**(3): 367–374.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Pilley JW, Reid AK: **Border collie comprehends object names as verbal referents.** *Behav Processes.* 2011; **86**(2): 184–195.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Poeppel D: **Pure word deafness and the bilateral processing of the speech code.** *Cogn Sci.* 2001; **25**(5): 679–693.
[Publisher Full Text](#)
- Poeppel D, Emmorey K, Hickok G, et al.: **Towards a new neurobiology of language.** *J Neurosci.* 2012; **32**(41): 14125–14131.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Poliva O: **From Mimicry to Language: A Neuroanatomically Based Evolutionary Model of the Emergence of Vocal Language.** *Front Neurosci.* 2016; **10**: 307.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

- Poliva O, Bestelmeyer PE, Hall M, *et al.*: **Functional Mapping of the Human Auditory Cortex: fMRI Investigation of a Patient with Auditory Agnosia from Trauma to the Inferior Colliculus.** *Cogn Behav Neurol.* 2015; **28**(3): 160–80.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Poremba A, Malloy M, Saunders RC, *et al.*: **Species-specific calls evoke asymmetric activity in the monkey's temporal poles.** *Nature.* 2004; **427**(6973): 448–451.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Premack D, Premack AJ: **The Mind of an Ape.** W. W. Norton. 1984.
[Reference Source](#)
- Rauschecker JP, Tian B: **Mechanisms and streams for processing of “what” and “where” in auditory cortex.** *Proc Natl Acad Sci U S A.* 2000; **97**(22): 11800–11806.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Rauschecker JP, Tian B, Hauser M: **Processing of complex sounds in the macaque nonprimary auditory cortex.** *Science.* 1995; **268**(5207): 111–114.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Rauschecker JP, Tian B, Pons T, *et al.*: **Serial and parallel processing in rhesus monkey auditory cortex.** *J Comp Neurol.* 1997; **382**(1): 89–103.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Recanzone GH: **Representation of con-specific vocalizations in the core and belt areas of the auditory cortex in the alert macaque monkey.** *J Neurosci.* 2008; **28**(49): 13184–13193.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Remedios R, Logothetis NK, Kayser C: **An auditory region in the primate insular cortex responding preferentially to vocal communication sounds.** *J Neurosci.* 2009a; **29**(4): 1034–1045.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Remedios R, Logothetis NK, Kayser C: **Monkey drumming reveals common networks for perceiving vocal and nonvocal communication sounds.** *Proc Natl Acad Sci U S A.* 2009b; **106**(42): 18010–18015.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Rilling JK, Glasser MF, Jbabdi S, *et al.*: **Continuity, divergence, and the evolution of brain language pathways.** *Front Evol Neurosci.* 2012; **3**: 11.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Roberts AC, Tomic DL, Parkinson CH, *et al.*: **Forebrain connectivity of the prefrontal cortex in the marmoset monkey (*Callithrix jacchus*): an anterograde and retrograde tract-tracing study.** *J Comp Neurol.* 2007; **502**(1): 86–112.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Robinson BW: **Vocalization evoked from forebrain in *Macaca mulatta*.** *Physiol Behav.* 1967; **2**(4): 345–354.
[Publisher Full Text](#)
- Rohrer JD, Ridgway GR, Crutch SJ, *et al.*: **Progressive logopenic/phonological aphasia: erosion of the language network.** *Neuroimage.* 2010; **49**(1): 984–993.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Rohrer JD, Sauter D, Scott S, *et al.*: **Receptive prosody in nonfluent primary progressive aphasias.** *Cortex.* 2012; **48**(3): 308–316.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Roll P, Rudolf G, Pereira S, *et al.*: **SRPX2 mutations in disorders of language cortex and cognition.** *Hum Mol Genet.* 2006; **15**(7): 1195–1207.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Roll P, Vernes SC, Bruneau N, *et al.*: **Molecular networks implicated in speech-related disorders: FOXP2 regulates the SRPX2/uPAR complex.** *Hum Mol Genet.* 2010; **19**(24): 4848–4860.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Romanski LM, Averbeck BB, Diltz M: **Neural representation of vocalizations in the primate ventrolateral prefrontal cortex.** *J Neurophysiol.* 2005; **93**(2): 734–747.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Romanski LM, Bates JF, Goldman-Rakic PS: **Auditory belt and parabelt projections to the prefrontal cortex in the rhesus monkey.** *J Comp Neurol.* 1999; **403**(2): 141–157.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Roux FE, Mishkin K, Durand JB, *et al.*: **Electrostimulation mapping of comprehension of auditory and visual words.** *Cortex.* 2015; **71**: 398–408.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Russ BE, Ackelson AL, Baker AE, *et al.*: **Coding of auditory-stimulus identity in the auditory non-spatial processing stream.** *J Neurophysiol.* 2008; **99**(1): 87–95.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Russo GS, Bruce CJ: **Frontal eye field activity preceding aurally guided saccades.** *J Neurophysiol.* 1994; **71**(3): 1250–1253.
[PubMed Abstract](#)
- Sammler D, Grosbras MH, Anwander A, *et al.*: **Dorsal and Ventral Pathways for Prosody.** *Curr Biol.* 2015; **25**(23): 3079–3085.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Saur D, Kreher BW, Schnell S, *et al.*: **Ventral and dorsal pathways for language.** *Proc Natl Acad Sci U S A.* 2008; **105**(46): 18035–18040.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Scheich H, Baumgart F, Gaschler-Markefski B, *et al.*: **Functional magnetic resonance imaging of a human auditory cortex area involved in foreground-background decomposition.** *Eur J Neurosci.* 1998; **10**(2): 803–809.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Schmahmann JD, Pandya DN, Wang R, *et al.*: **Association fibre pathways of the brain: parallel observations from diffusion spectrum imaging and autoradiography.** *Brain.* 2007; **130**(Pt 3): 630–653.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Schwartz MF, Kimberg DY, Walker GM, *et al.*: **Anterior temporal involvement in semantic word retrieval: voxel-based lesion-symptom mapping evidence from aphasia.** *Brain.* 2009; **132**(Pt 12): 3411–3427.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Scott SK, Blank CC, Rosen S, *et al.*: **Identification of a pathway for intelligible speech in the left temporal lobe.** *Brain.* 2000; **123**(Pt 12): 2400–2406.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Scott BH, Mishkin M, Yin P: **Monkeys have a limited form of short-term memory in audition.** *Proc Natl Acad Sci U S A.* 2012; **109**(30): 12237–41.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Seltzer B, Pandya DN: **Further observations on parieto-temporal connections in the rhesus monkey.** *Exp Brain Res.* 1984; **55**(2): 301–312.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Seyfarth RM, Cheney DL, Marler P: **Monkey responses to three different alarm calls: evidence of predator classification and semantic communication.** *Science.* 1980; **210**(4471): 801–3.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Shriberg LD, Ballard KJ, Tomblin JB, *et al.*: **Speech, prosody, and voice characteristics of a mother and daughter with a 7;13 translocation affecting FOXP2.** *J Speech Lang Hear Res.* 2006; **49**(3): 500–525.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Shu W, Cho JY, Jiang Y, *et al.*: **Altered ultrasonic vocalization in mice with a disruption in the *Foxp2* gene.** *Proc Natl Acad Sci U S A.* 2005; **102**(27): 9643–9648.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Shultz S, Vouloumanos A, Pelphrey K: **The superior temporal sulcus differentiates communicative and noncommunicative auditory signals.** *J Cogn Neurosci.* 2012; **24**(5): 1224–1232.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Sia GM, Clem RL, Haganir RL: **The human language-associated gene SRPX2 regulates synapse formation and vocalization in mice.** *Science.* 2013; **342**(6161): 987–991.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Simões CS, Vianey PV, de Moura MM, *et al.*: **Activation of frontal neocortical areas by vocal production in marmosets.** *Front Integr Neurosci.* 2010; **4**: pii: 123.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Smith KR, Hsieh IH, Saberi K, *et al.*: **Auditory spatial and object processing in the human planum temporale: no evidence for selectivity.** *J Cogn Neurosci.* 2010; **22**(4): 632–639.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Snow D: **Phrase-final syllable lengthening and intonation in early child speech.** *J Speech Hear Res.* 1994; **37**(4): 831–840.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Square PA, Roy EA, Martin RE: **Apraxia of speech: Another form of praxis disruption.** In *Apraxia: The neuropsychology of action.* Psychology Press, 1997; 173–206.
[Reference Source](#)
- Srinivasan RJ, Massaro DW: **Perceiving prosody from the face and voice: distinguishing statements from echoic questions in English.** *Lang Speech.* 2003; **46**(Pt 1): 1–22.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Steinschneider M, Volkov IO, Fishman YI, *et al.*: **Intracortical responses in human and monkey primary auditory cortex support a temporal processing mechanism for encoding of the voice onset time phonetic parameter.** *Cereb Cortex.* 2005; **15**(2): 170–186.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Stepien LS, Cordeau JP, Rasmussen T: **The effect of temporal lobe and hippocampal lesions on auditory and visual recent memory in monkeys.** *Brain.* 1960; **83**(3): 470–489.
[Publisher Full Text](#)
- Stewart L, von Kriegstein K, Warren JD, *et al.*: **Music and the brain: disorders of musical listening.** *Brain.* 2006; **129**(Pt 10): 2533–2553.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Stewart L, Walsh V, Frith U, *et al.*: **TMS produces two dissociable types of speech disruption.** *Neuroimage.* 2001; **13**(3): 472–478.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Striccanne B, Andersen RA, Mazzoni P: **Eye-centered, head-centered, and intermediate coding of remembered sound locations in area LIP.** *J Neurophysiol.* 1996; **76**(3): 2071–2076.
[PubMed Abstract](#)
- Striemi-Amit E, Hertz U, Amedi A: **Extensive cochleotopic mapping of human auditory cortical fields obtained with phase-encoding fMRI.** *PLoS One.* 2011; **6**(3): e17832.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Strominger NL, Oesterreich RE, Neff WD: **Sequential auditory and visual discriminations after temporal lobe ablation in monkeys.** *Physiol Behav.* 1980; **24**(6): 1149–1156.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Studdert-Kennedy M: **How did language go discrete?** *Language Origins: Perspectives on Evolution.* 2005; 48–67.
[Reference Source](#)

- Sugiura H: **Matching of acoustic features during the vocal exchange of coo calls by Japanese macaques.** *Anim Behav.* 1998; **55**(3): 673–687.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Sutton D, Larson C, Lindeman RC: **Neocortical and limbic lesion effects on primate phonation.** *Brain Res.* 1974; **71**(1): 61–75.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Sweet RA, Dorph-Petersen KA, Lewis DA: **Mapping auditory core, lateral belt, and parabelt cortices in the human superior temporal gyrus.** *J Comp Neurol.* 2005; **491**(3): 270–289.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Symmes D, Biben M: **Maternal recognition of individual infant squirrel monkeys from isolation call playbacks.** *Am J Primatol.* 1985; **9**(1): 39–46.
[Publisher Full Text](#)
- Tagliabue JP, Savage-Rumbaugh S, Baker LA: **Vocal production by a language-competent *Pan paniscus*.** *Int J Primatol.* 2003; **24**(1): 1–17.
[Publisher Full Text](#)
- Tata MS, Ward LM: **Early phase of spatial mismatch negativity is localized to a posterior “where” auditory pathway.** *Exp Brain Res.* 2005a; **167**(3): 481–486.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Tata MS, Ward LM: **Spatial attention modulates activity in a posterior “where” auditory pathway.** *Neuropsychologia.* 2005b; **43**(4): 509–516.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Tian B, Reser D, Durham A, et al.: **Functional specialization in rhesus monkey auditory cortex.** *Science.* 2001; **292**(5515): 290–293.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Tobias PV: **The brain of *Homo habilis*: A new level of organization in cerebral evolution.** *J Hum Evol.* 1987; **16**(7–8): 741–761.
[Publisher Full Text](#)
- Tsunada J, Lee JH, Cohen YE: **Representation of speech categories in the primate auditory cortex.** *J Neurophysiol.* 2011; **105**(6): 2634–2646.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Turken AU, Dronkers NF: **The neural architecture of the language comprehension network: converging evidence from lesion and connectivity analyses.** *Front Syst Neurosci.* 2011; **5**: 1–20.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Ulrich G: **Interhemispheric functional relationships in auditory agnosia. An analysis of the preconditions and a conceptual model.** *Brain Lang.* 1978; **5**(3): 286–300.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Vaadia E, Benson DA, Hienz RD, et al.: **Unit study of monkey frontal cortex: active localization of auditory and of visual stimuli.** *J Neurophysiol.* 1986; **56**(4): 934–952.
[PubMed Abstract](#)
- Vanderhorst VG, Terasawa E, Ralston HJ 3rd, et al.: **Monosynaptic projections from the lateral periaqueductal gray to the nucleus retroambiguus in the rhesus monkey: implications for vocalization and reproductive behavior.** *J Comp Neurol.* 2000; **424**(2): 251–268.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Vanderhorst VG, Terasawa E, Ralston HJ 3rd: **Monosynaptic projections from the nucleus retroambiguus region to laryngeal motoneurons in the rhesus monkey.** *Neuroscience.* 2001; **107**(1): 117–125.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Viceic D, Fornari E, Thiran JP, et al.: **Human auditory belt areas specialized in sound recognition: a functional magnetic resonance imaging study.** *Neuroreport.* 2006; **17**(16): 1659–1662.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Vigneau M, Beaucousin V, Hervé PY, et al.: **Meta-analyzing left hemisphere language areas: phonology, semantics, and sentence processing.** *Neuroimage.* 2006; **30**(4): 1414–1432.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Vignolo LA, Boccardi E, Caverni L: **Unexpected CT-scan findings in global aphasia.** *Cortex.* 1986; **22**(1): 55–69.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Wallace MN, Johnston PW, Palmer AR: **Histochemical identification of cortical areas in the auditory region of the human brain.** *Exp Brain Res.* 2002; **143**(4): 499–508.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Warren JD, Griffiths TD: **Distinct mechanisms for processing spatial sequences and pitch sequences in the human auditory brain.** *J Neurosci.* 2003; **23**(13): 5799–5804.
[PubMed Abstract](#)
- Warren JD, Scott SK, Price CJ, et al.: **Human brain mechanisms for the early analysis of voices.** *Neuroimage.* 2006; **31**(3): 1389–1397.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Warren JD, Uppenkamp S, Patterson RD, et al.: **Separating pitch chroma and pitch height in the human brain.** *Proc Natl Acad Sci U S A.* 2003; **100**(17): 10038–10042.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Warren JD, Zielinski BA, Green GG, et al.: **Perception of sound-source motion by the human brain.** *Neuron.* 2002; **34**(1): 139–148.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Watkins KE, Dronkers NF, Vargha-Khadem F: **Behavioural analysis of an inherited speech and language disorder: comparison with acquired aphasia.** *Brain.* 2002; **125**(Pt 3): 452–464.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Wernicke C: **Der aphasische Symptomenkomplex.** Springer Berlin Heidelberg. 1974; 1–70.
[Publisher Full Text](#)
- Wich SA, Swartz KB, Hardus ME, et al.: **A case of spontaneous acquisition of a human sound by an orangutan.** *Primates.* 2008; **50**(1): 56–64.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Wood B, Richmond BG: **Human evolution: taxonomy and paleobiology.** *J Anat.* 2000; **197**(Pt 1): 19–60.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Woods DL, Herron TJ, Cate AD, et al.: **Functional properties of human auditory cortical fields.** *Front Syst Neurosci.* 2010; **4**: 155.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Woods TM, Lopez SE, Long JH, et al.: **Effects of stimulus azimuth and intensity on the single-neuron activity in the auditory cortex of the alert macaque monkey.** *J Neurophysiol.* 2006; **96**(6): 3323–3337.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Yin P, Mishkin M, Sutter M, et al.: **Early stages of melody processing: stimulus-sequence and task-dependent neuronal activity in monkey auditory cortical fields A1 and R.** *J Neurophysiol.* 2008; **100**(6): 3009–3029.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Zaidel E: **Auditory vocabulary of the right hemisphere following brain bisection or hemidecortication.** *Cortex.* 1976; **12**(3): 191–211.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Zatorre RJ, Bouffard M, Ahad P, et al.: **Where is ‘where’ in the human auditory cortex?** *Nat Neurosci.* 2002; **5**(9): 905–909.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Zatorre RJ, Bouffard M, Belin P: **Sensitivity to auditory object features in human temporal neocortex.** *J Neurosci.* 2004; **24**(14): 3637–3642.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Zhang SP, Davis PJ, Bandler R, et al.: **Brain stem integration of vocalization: role of the midbrain periaqueductal gray.** *J Neurophysiol.* 1994; **72**(3): 1337–1356.
[PubMed Abstract](#)

Open Peer Review

Current Referee Status:



Version 2

Referee Report 13 September 2017

doi:10.5256/f1000research.8315.r12035



Josef Rauschecker

Laboratory of Integrative Neuroscience and Cognition, Georgetown University, Washington DC, USA

The author has responded well to my criticisms, and the paper has become more readable. Even though one can still debate some of his suggestions (or speculations), I find the paper worthy of publication now, as it will help to enliven an already lively debate on the evolution of speech and language.

I only have a few more issues with the References:

In all fairness, the paper by Rauschecker & Tian (2000) should be mentioned when introducing the concepts of auditory ventral and dorsal streams (AVS, ADS) on p. 4. The idea of parallel ventral and dorsal processing streams in the auditory system was first proposed and developed there and even earlier. I leave it up to the author, which one(s) of the papers below he wants to cite.

I also wonder if references suggested by the reviewers shouldn't generally be added to the overall reference list of the revised version. For instance, the papers by Bornkessel-Schlesewsky *et al.*, which seem highly relevant here, are still not cited, even though they are mentioned by two reviewers.

References

1. Rauschecker JP: Processing of complex sounds in the auditory cortex of cat, monkey, and man. *Acta Otolaryngol Suppl.* 1997; **532**: 34-8 [PubMed Abstract](#)
2. Rauschecker JP: Parallel processing in the auditory cortex of primates. *Audiol Neurootol.* **3** (2-3): 86-103 [PubMed Abstract](#)
3. Rauschecker JP: Cortical processing of complex sounds. *Curr Opin Neurobiol.* 1998; **8** (4): 516-21 [PubMed Abstract](#)
4. Rauschecker JP, Tian B: Mechanisms and streams for processing of "what" and "where" in auditory cortex. *Proc Natl Acad Sci U S A.* 2000; **97** (22): 11800-6 [PubMed Abstract](#) | [Publisher Full Text](#)

Competing Interests: No competing interests were disclosed.

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Version 1

Referee Report 18 January 2016

doi:10.5256/f1000research.6619.r11960



Michael A Arbib

Computer Science Department, University of Southern California, Los Angeles, CA, USA

I see this paper as a first draft of what could become an important contribution to neurally based approaches to the study of the evolution of the human brain's capacity for language. Its importance is three-fold:

1. It treats the ventral and dorsal streams for both the auditory and visual modalities.
2. It regards monkey calls not in terms of perception alone or production alone but rather in terms of their role in the interaction between two individuals in the context of their environment.
3. It places the ability to ask and answer questions at the heart of language use.

Below, I will offer several comments on how some shortcomings of the current version might be removed in future work by Poliva, but first a disclosure: I have emphasized the role of the two visual streams in relation to both the production and comprehension of language with an emphasis on the role of manual gesture and protosign in language evolution, and in terms of visual perception of what an utterance may be about (Arbib, 2013). By contrast, Bornkessel-Schlesewsky and Schlesewsky (2013) offer hypotheses on the roles of the auditory streams in the perception of sentences of a spoken language, linking them to neurolinguistic data from their lab and others. I have attempted a preliminary synthesis of these approaches (Arbib, 2015). More recently, they have co-authored a review of relevant data on the auditory streams in both monkey and human with the claim that no major evolutionary innovations were required in these streams to make language possible (Bornkessel-Schlesewsky, Schlesewsky, Small, & Rauschecker, 2015) – a claim with which I (and, I suspect, Poliva) would disagree. I hope to support the counter-claim in a forthcoming article in the *Journal of Neurolinguistics*. I believe Poliva's assessment of these articles would enrich his work, but now let me turn to other issues.

1. I endorse the key points of Amy Poremba's review: (i) The dorsal auditory stream was over-emphasized at the expense of assessing the role of the ventral stream and how these streams are integrated. (ii) Poremba notes the relevance of work from Mishkin's lab on auditory memory – see, e.g., Fritz, Mishkin, and Saunders (2005) which “raises the possibility that language is unique to humans not only because it depends on speech but also because it requires long-term auditory memory.” I would add that Aboitiz and his colleagues have emphasized the expansion of working memory capacity as a key element in evolving a language-ready brain (see Aboitiz, 2012, for a recent review of this approach). (iii) The leap from contact calls to “individuals ... capable of inventing new words and offspring ... capable of inquiring about objects in their environment and learning their names via mimicry” is essentially unbridged.
2. Since there are many monkey calls, it seems unclear why, if one is to use these calls as the core for evolving a brain with language, one should focus on contact calls alone. Including other calls might add more “evolutionary opportunities.” In this regard, note the argument of Seyfarth and Cheney that one may see the structure of language prefigured in the “rules” monkeys develop for social cognition (Cheney & Seyfarth, 2005; Seyfarth & Cheney, 2014).
3. I suspect that further work in language evolution will reveal a “mosaic” of innovations, some of which are apparent in different monkey or ape species. One may hope that studies of the brains of different species will reveal diverse cues that illuminate, perhaps, the convergent evolution of different tiles of the language-supporting neural mosaic of the human brain. Consider, for example, the capability for turn taking in geladas (Gustison, le Roux, & Bergman, 2012; Richman, 1987) and marmosets (Miller, Thomas, Nummela, & de la Mothe, 2015; Takahashi, Narayanan, & Ghazanfar,

2013) as just one of the diverse components of language-ready brain that are differentially evident in different species of nonhuman primate.

4. Figure 1 shows dual stream connectivity between the auditory cortex and frontal lobe of monkeys and humans. What can be said about the intersection of the 2 streams in VLPFC? And what can be said about the interaction of DLPFC and VLPFC? Figure 2 depicts the “From Where to What model” via three stages of neuroanatomical modifications. It might be useful to first provide a diagram focusing on VVS and VDS (initial V for Visual) and discussing the relation in both anatomy and function of these paths with each other. It might also be helpful to present pieces of the model along with the exposition of the related data, postponing this integrative figure until the pieces are in place.
5. A valuable feature of Poliva’s model is its suggestion of how the response to an auditory call might initiate visual search as the basis for action (he emphasizes the mother emitting a call if the child is not seen; a related scenario would be movement toward the child if it were seen). This issue of integration of communication and action, which may (but need not) integrate audition with vision, is an important feature which too few studies take into account. My question is whether he unduly emphasizes cortical pathways involving the frontal eye fields and shortchanges subcortical interactions involving the superior colliculus (noting of course that these are open to cortical influences modulated by the basal ganglia).
6. In Figure 2, Poliva asserts: (i) “Approximately 2.5 million years ago, the *Homo* genus emerged as a result of [my italics] duplication of the IPS and subsequent duplication of its frontal projections” (a) Surely, many more changes led to the emergence of *Homo*. (b) At the end of Section 7, Poliva suggests the relevance of endocast data to this claim. Are there relevant data on apes that could help us assess this transition? (ii) “Since the auditory cortex targeted the more proximal of these duplicated parietal regions, a new pathway dedicated for auditory processing emerged (i.e., auditory dorsal stream; ADS.” But monkey data show an ADS, so what is the transition being suggested here? Picking up on the issue in (5), one needs to better understand the division of labor between ADS and subcortical mechanisms (as well as AVS, to reiterate Poremba’s point).
7. Poliva claims to review “evidence for a role of the ADS in the transition from mediating contact calls into mediating human speech” but simply cites data correlating ADS impairment with disorders like speech apraxia. Nothing in the data privileges contact calls over other vocal productions – and, anyway, clear articulation is a far cry [sic] from mechanisms supporting the role of syntax and semantics in language production and perception.
8. In relation to 6(i), Poliva notes the dual role of the parietal lobe in sensory-motor transformation of both audio-spatial and verbal information, and proposes that during *Hominin* evolution there was a cortical field duplication, of the IPS with further duplication of its projections to the VLPFC which resulted in a pathway dedicated for audio-vocal conversion. How would this serve people who employ a signed language? (Of course, those who advocate a gestural origin of language must face the complementary question of how visuo-manual pathways came to support audio-vocal signals – which they must do because other primates lack vocal learning, let alone the use of syntax and semantics in either domain.)
9. Poliva stresses that the ability to ask and answer questions is an essential feature of language use. I agree. Future work on language evolution should pay more attention to the challenge of explaining how this evolved. However the focus on modifying contact calls with prosodic

intonations seems to me too narrow (I may be wrong, but more argument would be needed) and (as Poremba observed) the account of the transition remains too sketchy. Poliva cites “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?”),” but one must be careful to distinguish these infant “communicative acts” from the ability to deploy grammar to formulate an open-ended repertoire of commands and questions using the structures of a language – let alone being able to marshal answers to questions of even modest complexity.

10. In any case, it seems mistaken to place exclusive emphasis on the role of ADS in the transition – one might thus assess the hypotheses of Bornkessel-Schlesewsky and Schlesewsky (2013) on the roles of both ADS and AVS (and frontal areas) in speech comprehension. However, a companion paper is promised: “Discussing the transition from exchanging low-level distress contact calls into complex vocal language, however, is beyond the scope of the present paper and a model for such transition is discussed [at] length in a sibling paper titled ‘Vocal Mimicry as the Sculptor of the Human Mind. A Neuroanatomically based Evolutionary Model of The Emergence of Vocal Language’ (Poliva, in preparation).” Perhaps it would be better if less were said about this topic in the present paper so that the implications of the evidence on ADS function and evolution could be better assessed for their merits irrespective of the contact call hypothesis.

References

1. Aboitiz F: Gestures, vocalizations, and memory in language origins. *Front Evol Neurosci*. 2012; **4**: 2 [PubMed Abstract](#) | [Publisher Full Text](#)
2. Arbib MA: Mirror Systems and the Neurocognitive Substrates of Bodily Communication and Language. In C. Müller, A. Cienki, E. Fricke, S. Ladewig, D. McNeill. *Body-Language Communication*. 2013. 445-460
3. Arbib MA: Towards a Computational Comparative Neuroprimatology: Framing the language-ready brain. *Phys Life Rev*. 2015. [PubMed Abstract](#) | [Publisher Full Text](#)
4. Bornkessel-Schlesewsky I, Schlesewsky M: Reconciling time, space and function: a new dorsal-ventral stream model of sentence comprehension. *Brain Lang*. 2013; **125** (1): 60-76 [PubMed Abstract](#) | [Publisher Full Text](#)
5. Bornkessel-Schlesewsky I, Schlesewsky M, Small SL, Rauschecker JP: Neurobiological roots of language in primate audition: common computational properties. *Trends Cogn Sci*. 2015; **19** (3): 142-50 [PubMed Abstract](#) | [Publisher Full Text](#)
6. Cheney D, Seyfarth R: Constraints and preadaptations in the earliest stages of language evolution. *The Linguistic Review*. 2005; **22** (2-4). [Publisher Full Text](#)
7. Fritz J, Mishkin M, Saunders RC: In search of an auditory engram. *Proc Natl Acad Sci U S A*. 2005; **102** (26): 9359-64 [PubMed Abstract](#) | [Publisher Full Text](#)
8. Gustison ML, le Roux A, Bergman TJ: Derived vocalizations of geladas (*Theropithecus gelada*) and the evolution of vocal complexity in primates. *Philos Trans R Soc Lond B Biol Sci*. 2012; **367** (1597): 1847-59 [PubMed Abstract](#) | [Publisher Full Text](#)
9. Miller CT, Thomas AW, Nummela SU, de la Mothe LA: Responses of primate frontal cortex neurons during natural vocal communication. *J Neurophysiol*. 2015; **114** (2): 1158-71 [PubMed Abstract](#) | [Publisher Full Text](#)
10. Richman B: Rhythm and melody in gelada vocal exchanges. *Primates*. 1987; **28** (2): 199-223 [Publisher Full Text](#)
11. Seyfarth RM, Cheney DL: The evolution of language from social cognition. *Curr Opin Neurobiol*. 2014; **28**: 5-9 [PubMed Abstract](#) | [Publisher Full Text](#)
12. Takahashi DY, Narayanan DZ, Ghazanfar AA: Coupled oscillator dynamics of vocal turn-taking in monkeys. *Curr Biol*. 2013; **23** (21): 2162-8 [PubMed Abstract](#) | [Publisher Full Text](#)

Competing Interests: No competing interests were disclosed.

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Reader Comment 21 Jan 2016

Oren Poliva, Bangor University, UK

I want to thank the reviewer for his positive review and for his insightful and constructive comments. Below are my responses:

I endorse the key points of Amy Poremba's review: (i) The dorsal auditory stream was over-emphasized at the expense of assessing the role of the ventral stream and how these streams are integrated.

Response: I agree with the reviewer that the article focuses on the ADS, and pay little attention to the AVS. I also agree that the AVS partakes an important role in the perception and production of human language, and that it interacts with the ADS. However, as I also previously responded to Poremba, in the present paper I propose a model for the emergence of speech and not language, and speech appears to be primarily or solely a function of the ADS. A possible course for the transition from speech to language and the role the AVS in such functions is discussed in detail in the second paper (mentioned in the article).

Poremba notes the relevance of work from Mishkin's lab on auditory memory – see, e.g., Fritz, Mishkin, and Saunders (2005) which “raises the possibility that language is unique to humans not only because it depends on speech but also because it requires long-term auditory memory.” I would add that Aboitiz and his colleagues have emphasized the expansion of working memory capacity as a key element in evolving a language-ready brain (see Aboitiz, 2012, for a recent review of this approach).

Response: I agree with the reviewer that expansion of auditory memory (or its ability to sustain interferences as shown by Scott, Mishkin & Yin, 2012) took an important part in the evolution of language. However, as I previously responded to Poremba, this change likely occurred after *Hominins* acquired volitional control over the vocal apparatus, and thus is beyond the scope of the present paper. This issue is also discussed in detail in the second paper.

Scott BH, Mishkin M, Yin P. Monkeys have a limited form of short-term memory in audition. *Proceedings of the National Academy of Sciences*. 2012 Jul 24;109(30):12237–41.

The leap from contact calls to “individuals ... capable of inventing new words and offspring ... capable of inquiring about objects in their environment and learning their names via mimicry” is essentially unbridged.....

In any case, it seems mistaken to place exclusive emphasis on the role of ADS in the transition – one might thus assess the hypotheses of Bornkessel-Schlesewsky and Schlewewsky (2013) on the

roles of both ADS and AVS (and frontal areas) in speech comprehension. However, a companion paper is promised: “Discussing the transition from exchanging low-level distress contact calls into complex vocal language, however, is beyond the scope of the present paper and a model for such transition is discussed [at] length in a sibling paper titled ‘Vocal Mimicry as the Sculptor of the Human Mind. A Neuroanatomically based Evolutionary Model of The Emergence of Vocal Language’ (Poliva, in preparation).” Perhaps it would be better if less were said about this topic in the present paper so that the implications of the evidence on ADS function and evolution could be better assessed for their merits irrespective of the contact call hypothesis.

Response: I agree with the reviewer that the article doesn’t delve enough into the transition from speech to vocal mimicry. As I responded to Poremba, and mentioned in the paper, this topic is discussed in detail in the second paper. As the primary concern of the present paper is the emergence of speech, I removed from the abstract and introduction any mentioning of the transition from speech to vocal mimicry based language, and limited its discussion to a short paragraph near the end of the paper.

Since there are many monkey calls, it seems unclear why, if one is to use these calls as the core for evolving a brain with language, one should focus on contact calls alone. Including other calls might add more “evolutionary opportunities.” In this regard, note the argument of Seyfarth and Cheney that one may see the structure of language prefigured in the “rules” monkeys develop for social cognition (Cheney & Seyfarth, 2005; Seyfarth & Cheney, 2014).

Response: The reviewer presents an interesting question when he suggests that contact calls might not be special. The paper he cites suggests that rule based alarm calls could serve as a potential precursor to human language. In my opinion, contact calls are a more likely candidate precursor to present day vocal conversation than alarm calls. Like present day vocal conversations, contact call are characterized with turn taking and require interaction between (at least) two participants. The content of contact calls is also similar to present day question answer dialogue (as if similar to the question ‘where are you?’ and the answer ‘I’m here, Where are you?’). Alarm calls in contrast, although context dependent and thus likely under cortical influence, do not require vocal response and thus don’t resemble conversation. Moreover, as I present in the paper, converging evidence suggests that both human speech and contact call exchange in non-human primates are processed in the ADS. As far as I’m aware of, no study provided evidence that alarm calls are processed in the ADS. (Given its dependence on observing emotive stimuli, I would assume that expressing alarm calls occurs through processing in the visual ventral stream and amygdala, and response to alarm calls occurs through the auditory ventral stream and amygdala.)

I suspect that further work in language evolution will reveal a “mosaic” of innovations, some of which are apparent in different monkey or ape species. One may hope that studies of the brains of different species will reveal diverse cues that illuminate, perhaps, the convergent evolution of different tiles of the language-supporting neural mosaic of the human brain. Consider, for example, the capability for turn taking in geladas (Gustison, le Roux, & Bergman, 2012; Richman, 1987) and marmosets (Miller, Thomas, Nummela, & de la Mothe, 2015; Takahashi, Narayanan, & Ghazanfar, 2013) as just one of the diverse components of language-ready brain that are differentially evident in different species of nonhuman primate.

Response: I admit I got confused from the reviewer’s comment. The reviewer argues that

turn taking occurs in gelada monkeys. The studies he cite however don't mention such behavior. The reviewer then proceed to cite turn taking vocal behavior in marmoset monkeys, as an alternative explanation to how humans developed turn taking in conversations. The reviewer, however, cite studies that explore turn taking in the exchange of contact calls, which further support the discussed model.

Figure 1 shows dual stream connectivity between the auditory cortex and frontal lobe of monkeys and humans. What can be said about the intersection of the 2 streams in VLPFC? And what can be said about the interaction of DLPFC and VLPFC?

Response: In the paper I describe two pathways connecting the auditory cortex with the prefrontal cortex. The prefrontal cortex is primarily ascribed with planning and problem solving. When detecting and responding to contact calls, the prefrontal cortex likely mediates high level processing, such as determining the best way to overcome an obstacle in order to reach the caller. In the present model, I attempt to demonstrate that the detection and production of contact calls occur in the same pathway as speech in humans, and on that account attribute a relationship between them. The role of the prefrontal cortex in such high level processing is not necessary for establishing this relationship and is thus beyond the scope of the present paper.

Figure 2 depicts the "From Where to What model" via three stages of neuroanatomical modifications. It might be useful to first provide a diagram focusing on VVS and VDS (initial V for Visual) and discussing the relation in both anatomy and function of these paths with each other. It might also be helpful to present pieces of the model along with the exposition of the related data, postponing this integrative figure until the pieces are in place....

In Figure 2, Poliva asserts: (i) "Approximately 2.5 million years ago, the *Homo* genus emerged as a result of [my italics] duplication of the IPS and subsequent duplication of its frontal projections" (a) Surely, many more changes led to the emergence of *Homo*. (b) At the end of Section 7, Poliva suggests the relevance of endocast data to this claim. Are there relevant data on apes that could help us assess this transition? (ii) "Since the auditory cortex targeted the more proximal of these duplicated parietal regions, a new pathway dedicated for auditory processing emerged (i.e., auditory dorsal stream; ADS." But monkey data show an ADS, so what is the transition being suggested here? Picking up on the issue in (5), one needs to better understand the division of labor between ADS and subcortical mechanisms (as well as AVS, to reiterate Poremba's point).... In relation to 6(i), Poliva notes the dual role of the parietal lobe in sensory-motor transformation of both audio-spatial and verbal information, and proposes that during *Hominin* evolution there was a cortical field duplication, of the IPS with further duplication of its projections to the VLPFC which resulted in a pathway dedicated for audio-vocal conversion. How would this serve people who employ a signed language? (Of course, those who advocate a gestural origin of language must face the complementary question of how visuo-manual pathways came to support audio-vocal signals – which they must do because other primates lack vocal learning, let alone the use of syntax and semantics in either domain.)

Response: I agree with the reviewer that in depth description of the visual streams and addition of evidence for the parietal duplication hypothesis could add more depth to the paper. However, reviewer 2 (Josef Rauschecker) argued that the section of the paper discussing the relationship between the auditory and visual streams is problematic, and overall disagreed with the parietal duplication hypothesis. Although I don't entirely agree with his perspective, given that the paper is already rich in hypotheses and evidence, I

chose to remove the sections discussing this topic from the paper. Possibly, the parietal duplication hypothesis will be presented in the future in its own paper.

A valuable feature of Poliva's model is its suggestion of how the response to an auditory call might initiate visual search as the basis for action (he emphasizes the mother emitting a call if the child is not seen; a related scenario would be movement toward the child if it were seen). This issue of integration of communication and action, which may (but need not) integrate audition with vision, is an important feature which too few studies take into account. My question is whether he unduly emphasizes cortical pathways involving the frontal eye fields and shortchanges subcortical interactions involving the superior colliculus (noting of course that these are open to cortical influences modulated by the basal ganglia)

Response: I agree with the reviewer that area LIP in the intraparietal sulcus likely guides eye movements via projections to the frontal eye field and the superior colliculi. Such connections from the area LIP to the superior colliculi were described in tracing studies (Lynch et al., 1985). However, to the best of my knowledge no study so far demonstrated that this parieto-collicular pathway carries auditory information. It would also be very difficult to demonstrate that auditory influence on the superior colliculus occurs via connections from area LIP and not via ascending connections from the inferior colliculi. Given the lack of evidence of an auditory parieto-collicular pathway I chose at this point not to include it in the revised paper.

Lynch, J. C., AMs Graybiel, and L. J. Lobeck. "The differential projection of two cytoarchitectonic subregions of the inferior parietal lobule of macaque upon the deep layers of the superior colliculus." *Journal of Comparative Neurology* 235.2 (1985): 241-254.

Poliva claims to review "evidence for a role of the ADS in the transition from mediating contact calls into mediating human speech" but simply cites data correlating ADS impairment with disorders like speech apraxia.

Response: In addition to the paragraph discussing the role of the ADS in speech production, I present throughout the paper many other studies that indirectly show a role of the human ADS in speech production, such as fMRI studies that compare speech production to the production of melodies (Hickok et al., 2003) and many studies that ascribe the ADS with a role in speech repetition (Hickok et al., 2007).

Nothing in the data privileges contact calls over other vocal productions

Response: Many studies have shown that the ADS (associated in human with speech production) has a special role in the detection and production of contact calls. For example:

"Further corroborating the involvement of the ADS in the perception of contact calls are intra-cortical recordings from the posterior insula (near area CM-A1) of the macaque, which revealed stronger selectivity for a contact call (coo call) than a social call (threat call; Remedios et al., 2009a). Contrasting this finding is a study that recorded neural activity from the anterior auditory cortex, and reported that the proportion of neurons dedicated to a contact call was similar to the proportions of neurons dedicated to other calls (Perrodin et al., 2011)."

Also:

“Consistently, a study that sacrificed marmoset monkeys immediately after responding to contact calls (phee calls) measured highest neural activity (genomic expression of cFos protein) in the posterior auditory fields (CM-CL), and VLPFC (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).”

– and, anyway, clear articulation is a far cry [sic] from mechanisms supporting the role of syntax and semantics in language production and perception.

Response: I agree with the reviewer that arguing that the ADS processes speech does not necessitate that the ADS process more complex linguistic functions such as semantics and syntax. This is why in the paper I only present a model for the emergence of speech. More complex linguistic functions and possible evolutionary course will be discussed in the second paper.

Poliva stresses that the ability to ask and answer questions is an essential feature of language use. I agree. Future work on language evolution should pay more attention to the challenge of explaining how this evolved. However the focus on modifying contact calls with prosodic intonations seems to me too narrow (I may be wrong, but more argument would be needed) and (as Poremba observed) the account of the transition remains too sketchy. Poliva cites “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?”),” but one must be careful to distinguish these infant “communicative acts” from the ability to deploy grammar to formulate an open-ended repertoire of commands and questions using the structures of a language – let alone being able to marshal answers to questions of even modest complexity.

Response: I agree with the reviewer that adults often use complex syntax to ask questions. However, given that children (and occasionally adults) can express a question with a single word using intonations, suggests, in my opinion, that such question asking method could have preceded syntax, and thus indicate of an intermediate stage in the evolution of language. A transition from a single word question to syntax likely occurred at later evolutionary stages, and is thus beyond the scope of the present paper.

Competing Interests: No competing interests were disclosed.

Referee Report 24 December 2015

doi:10.5256/f1000research.6619.r7964

**Josef Rauschecker**

Laboratory of Integrative Neuroscience and Cognition, Georgetown University, Washington DC, USA

This is an interesting contribution to the literature on language evolution. The first two sections ('Introduction' and 'Models of Language Processing in the Brain...') are a joy to read. Later sections are more controversial and contain serious flaws that have to be brought up to speed with the current literature. These concerns are summarized here:

1) The terminology is quite fuzzy. For instance, when the author refers to 'perception' he seems to mean 'detection' or 'processing'. In most people's minds, and in most extant models of perception and action, perception is specifically tied to the ventral stream. Therefore, it can, almost by definition, not also be a property of the dorsal stream. This is best exemplified in the Abstract: The author states: '*I propose that the primary role of the auditory dorsal stream (ADS) in monkeys/apes is the perception and response to contact calls.*' This misstatement can be fixed by replacing 'perception' with 'detection'. Similarly, in a later sentence ('Perception of contact calls occurs by the ADS detecting a voice...'), 'Perception' can be substituted by 'Processing'. Thirdly, in the Abstract's second paragraph, the following sentence does not make any sense: 'Because the human ADS processes also speech production and repetition...'. Here, 'processes' needs to be replaced with 'performs'.

2) In the third section, the author first makes a strong case for a role of the ADS in auditory spatial processing, for encoding of sound location in memory and for use of this information in guiding eye movements. The published literature is well represented, though a key reference is missing here (Tian et al., Science, 2001). Then, in a surprising turnaround, the author suddenly concludes that 'audiospatial input is first converted into a visuospatial code and then processed via a visuospatial network'. The evidence cited stems from 15-year old studies of monkey area LIP, which is part of a visuospatial network; auditory signals, however, are relayed to a different part of IPS (area VIP; Lewis & VanEssen, 2000), for which corresponding studies have not been performed. Figure 2, which pertains to this section, reflects this misinterpretation: While the version on the left is neuroanatomically acceptable (with the only difference that parietal cortex is not just a visuospatial but a multisensory or amodal network, the versions in the center and on the right are incorrect on multiple grounds, most notably by postulating the 'duplication of the IPS [pivoting around an imaginary blue asterisk] and subsequent duplication of its frontal projections'. The assumptions about anatomical connections of the IPL with ventrolateral prefrontal cortex (VLPFC) are largely unsubstantiated and the characterization of VLPFC as a motor representation is plain wrong. I assume what the author may be referring to is ventral premotor cortex (PMv), which is indeed the terminal point of the auditory dorsal stream and is closely interfacing with Broca's area.

3) According to a third hypothesis put forward by the author, "*the Homo genus emerged as a result of duplicating the IPS and its frontal projections. This duplication resulted with area Spt and its projections to the VLPFC. In contrast to the visual dorsal stream that processes audiovisual spatial properties, the human ADS processes inner and outer speech.*" This hypothesis is seriously flawed, because both ADS and VDS process spatial properties and both process sensorimotor signals. In fact, they may be one and the same structure. Thus, there is no fundamental difference between visual and auditory processing that would require duplication of IPS or its projections or special evolution of speech (see Bornkessel et al., 2015).

4) Additional citations that the author should add:

DeWitt and Rauschecker, 2012 ¹

DeWitt and Rauschecker, 2013 ²

Bornkessel-Schlesewsky, et al., 2015 ³

Mesulam, et al., 2015 ⁴

Roux, et al., 2015 ⁵

References

1. DeWitt I, Rauschecker JP: Phoneme and word recognition in the auditory ventral stream. *Proc Natl Acad Sci U S A*. 2012; **109** (8): E505-14 [PubMed Abstract](#) | [Publisher Full Text](#)
2. DeWitt I, Rauschecker J: Wernicke's area revisited: Parallel streams and word processing. *Brain and Language*. 2013; **127** (2): 181-191 [Publisher Full Text](#)
3. Bornkessel-Schlesewsky I, Schlewsky M, Small SL, Rauschecker JP: Neurobiological roots of language in primate audition: common computational properties. *Trends Cogn Sci*. 2015; **19** (3): 142-50 [PubMed Abstract](#) | [Publisher Full Text](#)
4. Mesulam MM, Thompson CK, Weintraub S, Rogalski EJ: The Wernicke conundrum and the anatomy of language comprehension in primary progressive aphasia. *Brain*. 2015; **138** (Pt 8): 2423-37 [PubMed Abstract](#) | [Publisher Full Text](#)
5. Roux FE, Miskin K, Durand JB, Sacko O, Réhault E, Tanova R, Démonet JF: Electrostimulation mapping of comprehension of auditory and visual words. *Cortex*. 2015; **71**: 398-408 [PubMed Abstract](#) | [Publisher Full Text](#)

Competing Interests: No competing interests were disclosed.

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Author Response 05 Jan 2016

Oren Poliva, Bangor University, UK

This is an interesting contribution to the literature on language evolution. The first two sections ('Introduction' and 'Models of Language Processing in the Brain...') are a joy to read. Later sections are more controversial and contain serious flaws that have to be brought up to speed with the current literature. These concerns are summarized here:

1) The terminology is quite fuzzy. For instance, when the author refers to 'perception' he seems to mean 'detection' or 'processing'. In most people's minds, and in most extant models of perception and action, perception is specifically tied to the ventral stream. Therefore, it can, almost by definition, not also be a property of the dorsal stream. This is best exemplified in the Abstract: The author states: '*I propose that the primary role of the auditory dorsal stream (ADS) in monkeys/apes is the perception and response to contact calls.*' This misstatement can be fixed by replacing 'perception' with 'detection'. Similarly, in a later sentence ('Perception of contact calls occurs by the ADS detecting a voice...'), 'Perception' can be substituted by 'Processing'. Thirdly, in the Abstract's second paragraph, the following sentence does not make any sense: 'Because the human ADS processes also speech production and repetition...'. Here, 'processes' needs to be replaced with 'performs'.

Response: As far as I understand it, perception refers to all elements of the external world that reach our awareness. In accordance with this definition, through the AVS we perceive the identity of sounds and through the ADS we perceive the location of sounds. As human speech production is also processed in the ADS, I would expect that we also perceive elements of speech preparation through the ADS. A good example is a study that reported of patients who were electrically stimulated in the left inferior parietal lobule and consequently believed they produced sounds, when in fact they didn't (Desmurget et al.,

2009). This study can be argued to demonstrate perception of speech preparation in the ADS. Nonetheless, considering that different researchers might have different definitions for perception, I replaced instances that describe perception with detection wherever it was applicable.

Desmurget M, Reilly KT, Richard N, Szathmari A, Mottolese C, Sirigu A. Movement intention after parietal cortex stimulation in humans. *Science*. 2009 May 8;324(5928):811–3.

2) In the third section, the author first makes a strong case for a role of the ADS in auditory spatial processing, for encoding of sound location in memory and for use of this information in guiding eye movements. The published literature is well represented, though a key reference is missing here ([Tian et al., *Science*, 2001](#)). Then, in a surprising turnaround, the author suddenly concludes that 'audiospatial input is first converted into a visuospatial code and then processed via a visuospatial network'. The evidence cited stems from 15-year old studies of monkey area LIP, which is part of a visuospatial network; auditory signals, however, are relayed to a different part of IPS (area VIP; [Lewis & VanEssen, 2000](#)), for which corresponding studies have not been performed. Figure 2, which pertains to this section, reflects this misinterpretation: While the version on the left is neuroanatomically acceptable (with the only difference that parietal cortex is not just a visuospatial but a multisensory or amodal network, the versions in the center and on the right are incorrect on multiple grounds, most notably by postulating the 'duplication of the IPS [pivoting around an imaginary blue asterisk] and subsequent duplication of its frontal projections'. According to a third hypothesis put forward by the author, "*the Homo genus emerged as a result of duplicating the IPS and its frontal projections. This duplication resulted with area Spt and its projections to the VLPFC. In contrast to the visual dorsal stream that processes audiovisual spatial properties, the human ADS processes inner and outer speech.*" This hypothesis is seriously flawed, because both ADS and VDS process spatial properties and both process sensorimotor signals. In fact, they may be one and the same structure. Thus, there is no fundamental difference between visual and auditory processing that would require duplication of IPS or its projections or special evolution of speech (see Bornkessel et al., 2015).

Response: Although I don't entirely agree with the reviewer's perspective in this regard, given that the paper is already rich in evidence and hypotheses, I removed the sections (last paragraph of section 3 and section 7) discussing these hypotheses in the revised version. Also, I removed figure 2 from the revised version, and accordingly modified the manuscript to accommodate this change.

3) ...I assume what the author may be referring to is ventral premotor cortex (PMv), which is indeed the terminal point of the auditory dorsal stream and is closely interfacing with Broca's area.

Response: Thinking back, I agree with the reviewer that referring to this region as the 'ventral premotor cortex' is more accurate. The reason I referred to this region as the ventrolateral prefrontal cortex is to be consistent with previous papers (e.g., [Romansky et al., 1999](#)). As it possible that the area most often referred to as Broca's area encompasses both parts of the ventrolateral prefrontal cortex and ventral premotor cortex, in the revised manuscript I replaced the term 'ventrolateral prefrontal cortex' with its anatomical equivalent, the 'inferior frontal gyrus'.

Romanski LM, Bates JF, Goldman-Rakic PS. Auditory belt and parabelt projections to the

prefrontal cortex in the rhesus monkey. J Comp Neurol. 1999 Jan 11;403(2):141–57.

As a final note, I want to thank the reviewer for his time and effort, and hope he finds the revised version even more enjoyable to read.

Competing Interests: No competing interests were disclosed.

Referee Report 17 July 2015

doi:[10.5256/f1000research.6619.r8933](https://doi.org/10.5256/f1000research.6619.r8933)



Amy Poremba

Department of Psychology, University of Iowa, Iowa City, IA, USA

This contribution is a wide-ranging theory of how speech evolved in humans, which incorporates the dorsal and ventral auditory processing streams, but primarily focused on the auditory dorsal stream.

There are several large leaps in the proposed trajectory for language evolution such as, “eventually, individuals were capable of inventing new words and offspring were capable of inquiring about objects in their environment and learning their names via mimicry.” While the first part of the overall proposed theory is well supported, these latter stages are under-supported by current knowledge, particularly when moving to discussing individuals that became capable of enunciating novel calls (e.g., last paragraph of introduction); (some publications that may be helpful, comment by [Meguerditchian et al., 2014](#); original article, [Ackermann et al., 2014](#)). The steps proposed for inventing new words and inquiring about objects are likely to require a large number of processes and the theory does not specify what those steps might be. Overall, Poliva’s theory as set forth does generate some interesting, testable, hypotheses as demonstrated in section 9, and the leaps in the logical flow do not negate these as the hypotheses are more closely related to the current knowledge base.

As this is a theory of “From where to what,” missing for me was a better description of how the dorsal and ventral streams might interact in this theory. Calls still need to be “recognized” as auditory objects and imaging and recording studies have indicated the ventral stream does process this type of information. The ventral stream was given much less prominence and described in the appendix. It would be nice to include a paragraph or two on how the two systems may work together or how the ventral stream object identification comes to participate or interact with word formation and questions about objects.

In the first paragraph of the introduction, curiosity toward the unknown may be related to non-human primates’ tendency to pick novel objects from known objects. This is also true in many lower animals. The development of curiosity of objects that are absent from our environment as Poliva suggests must also be related to memory development. One must be able to remember that objects exist and have detailed memories in order to determine if an object is indeed missing. There are aspects of work by Mishkin and colleagues suggesting that the lack of robust, or expansive, long-term auditory memory may relate to the absence of complex communication systems in non-human primates, such as rhesus macaques. Clearly, visual memory is much more extensive and robust than auditory memory and the sign language that other non-human primates have demonstrated may be related to the robust nature of visual memory. The issue of memory mechanisms necessary for identifying that auditory objects are indeed missing from the environment, and how these may differ and interact between auditory and visual systems, should at least be mentioned in passing.

Competing Interests: No competing interests were disclosed.

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Author Response 05 Jan 2016

Oren Poliva, Bangor University, UK

This contribution is a wide-ranging theory of how speech evolved in humans, which incorporates the dorsal and ventral auditory processing streams, but primarily focused on the auditory dorsal stream.

There are several large leaps in the proposed trajectory for language evolution such as, “eventually, individuals were capable of inventing new words and offspring were capable of inquiring about objects in their environment and learning their names via mimicry.” While the first part of the overall proposed theory is well supported, these latter stages are under-supported by current knowledge, particularly when moving to discussing individuals that became capable of enunciating novel calls (e.g., last paragraph of introduction); (some publications that may be helpful, comment by [Meguerditchian et al., 2014](#); original article, [Ackermann et al., 2014](#)). The steps proposed for inventing new words and inquiring about objects are likely to require a large number of processes and the theory does not specify what those steps might be. Overall, Poliva’s theory as set forth does generate some interesting, testable, hypotheses as demonstrated in section 9, and the leaps in the logical flow do not negate these as the hypotheses are more closely related to the current knowledge base.

Response: I agree with the reviewer that the final evolutionary stages show a leap and are not strongly substantiated with evidence. As mentioned in the paper, in depth discussion of these stages is presented in a sibling paper, which is currently in writing. Nonetheless, in the revised manuscript, I made more effort to describe possible transition to mimicry. Moreover, I removed discussing this issue from the abstract and introduction, as it is not the primary concern of the present paper.

As this is a theory of “From where to what,” missing for me was a better description of how the dorsal and ventral streams might interact in this theory. Calls still need to be “recognized” as auditory objects and imaging and recording studies have indicated the ventral stream does process this type of information. The ventral stream was given much less prominence and described in the appendix. It would be nice to include a paragraph or two on how the two systems may work together or how the ventral stream object identification comes to participate or interact with word formation and questions about objects.

Response: As mentioned above, in the revised manuscript I downplayed the role of the ADS in object naming and mimicry, and limited the discussion to speech. In depth discussion into the role of the AVS in these functions will be presented in the sibling paper. There was simply too many hypotheses and topics to cover, which made it impossible to include them all in a single paper.

In the first paragraph of the introduction, curiosity toward the unknown may be related to

non-human primates' tendency to pick novel objects from known objects. This is also true in many lower animals.

Response: The hypothesis that curiosity to novel objects prompted our curiosity to the unknown is an interesting alternative hypothesis. Humans, however, since the beginning of written history, were also documented with another curiosity: desire to explore unknown places. In the present paper, I present evidence that the primary drive for the emergence of speech was by lost infants and mothers seeking to reunite. This model seems to explain both the emergence of speech and our unique curiosity for the unknown and is thus parsimonious. Presenting an alternative explanation would entail evidence for a different evolutionary course, and is thus beyond the scope of the present paper. Saying that, I'll be very interested to read about evidence for an evolutionary course that explains the curiosity to the unknown from this perspective. In the present model, I argue that the first question ever asked was "where are you". It leaves me wondering that if the curiosity to the unknown was prompted by curiosity to novel objects, then what could have been the first question?

The development of curiosity of objects that are absent from our environment as Poliva suggests must also be related to memory development. One must be able to remember that objects exist and have detailed memories in order to determine if an object is indeed missing. There are aspects of work by Mishkin and colleagues suggesting that the lack of robust, or expansive, long-term auditory memory may relate to the absence of complex communication systems in non-human primates, such as rhesus macaques. Clearly, visual memory is much more extensive and robust than auditory memory and the sign language that other non-human primates have demonstrated may be related to the robust nature of visual memory. The issue of memory mechanisms necessary for identifying that auditory objects are indeed missing from the environment, and how these may differ and interact between auditory and visual systems, should at least be mentioned in passing.

Response: The hypothesis that expansion of auditory memory contributed to the development of language is very interesting and I do appreciate that the reviewer brought this research to my attention. However, in the present paper I only describe an evolutionary course up to the advent of the first conversation. Enhancement of auditory memory likely occurred in later stages of language development, and is thus beyond the scope of the present paper.

As a final note, I want to thank the reviewer for her insightful comments and opinions, and hope that she enjoys the revised version of the paper.

Competing Interests: No competing interests were disclosed.

Discuss this Article

Version 1

Author Response 26 Aug 2015

Oren Poliva, Bangor University, UK

Thank you for your comment. Why do you think the model presented in this book is relevant to the present speech evolution model?

Competing Interests: No competing interests were disclosed.

Reader Comment 15 Aug 2015

Andrew Freinkel, Stanford University School of Medicine (Emeritus), USA

It's striking to me that the author made no mention of the spectacularly important work of Julian Jaynes in his book "The Origin of Consciousness In the Breakdown of the Bicameral Mind." If Dr. Poliva is unaware of Jaynes's work, he may find it of interest.

Competing Interests: No competing interests were disclosed.
