DE GRUYTER MOUTON
DOI 10.1515/langcog-2013-0011 — Language and Cognition 2013; 5(2–3): 167–176

Francisco Aboitiz

How did vocal behavior “take over” the gestural communication system?

Abstract: In this commentary, I argue (i) that there are some peripheral homologies between the monkey and the human vocalization systems; (ii) that complex vocal learning can be achieved without need of a voluntary hand grasping circuit; and (iii) that in the monkey there are rudimentary circuits that can convey auditory information into Broca’s region, via the “ventral pathway” but also via the arcuate or the superior longitudinal fasciculi.

Francisco Aboitiz: Departamento de Psiquiatría, Facultad de Medicina, and Centro Interdisciplinario de Neurociencia, Pontificia Universidad Católica de Chile, Santiago, Chile
E-mail: faboitiz@puc.cl

1 Introduction

I will focus my commentary in two specific issues: one has to do with the arguments of (dis)continuity of the human and non-human vocalization systems, and the participation of hand grasping and vocalization mechanisms in early communication; and the second relates to the evolution of the neural circuitry involved in vocalizations in humans and non-human primates.

2 Hand control and vocalizations

Like other exponents of the mirror neuron hypothesis, Arbib emphasizes a discontinuity between human speech and primate vocalizations, and that the articulatory systems involved in speech rather derive from the mirror neuron circuits involved in gestural communication. He proposes an “expanding spiral” of coevolution between gestures and vocalizations that contributed to the increasing richness of communication, but places strong emphasis on the notion that an open-ended communication system originated in the gestural domain (protosigns), while being replaced by speech (protospeech and speech) only when the cortico-bulbar tract innervating hypoglossal motor neurons became functional.

However, many other processes had to take place concomitantly with the development of the corticobulbar tract. Modern speech requires a complex combination of peripheral modifications like the descent of the larynx (which is also present in other species; Fitch and Reby 2001), deformation of the mouth cavity and innervation of the tongue and vocal cords, among other things. Like in the origin of many complex adaptations, many of these modifications may have originally appeared in response to different selective processes, but converged and were co-opted for the development of speech.

I believe no one argues against homology in the peripheral organs involved in speech and primate vocalizations. Tongue, lips, larynx and vocal folds are clearly homologous, not only among primates but in other mammals as well. Moreover, some brainstem central pattern generators that coordinate communicative, rhythmic movements of these organs may also be homologous with those involved in human speech. Monkey lip smacking is independent of throat movements and has a frequency of about five cycles per second, the same frequency that is observed in human speech, which is also much faster than the chewing cycle (Ghazanfar et al. 2012). In addition, the developmental trajectory of this behavior is strikingly similar to that of human speech and different from that of chewing patterns, with slower, variable mouth movements in early stages, which progresses into faster and more stereotyped patterns during development (Morrill et al. 2012). Remarkably, the 5 Hz frequency of speech coincides with phase-locked theta band activity (5–7 Hz) observed in the human auditory cortex during speech perception (Howard and Poeppel 2012). The articulatory organization of speech patterns, partly based on the above mentioned 5 Hz vocal oscillations, is much faster than that of manual and body gestures, and allows the vocalizer to generate highly complex messages in a brief time, a capacity that is much less obvious in hand signing.

Therefore, at least some aspects of vocal articulation display interesting similarities between humans and monkeys. Note that this does not preclude the cooption of additional systems or motor programs like ingestive behaviors, or hand manipulation, to subserve the emerging new vocal system (Ferrari et al. 2003). As mentioned, this is what would be expected in the evolution of a complex innovation like speech. As the neural circuits controlling vocal behavior developed, they became much more flexible than signing mechanisms, transmitting more complex information in a shorter time, and allowing communication at a larger distance without need of sight. Thus, they rapidly became superior than gestures as communication channels.

Secondly, Arbib correctly claims that a mirror neuron system may be required, but is not sufficient for the development of complex imitation, which is crucial for language learning and, in general, for our social behavior. In fact, monkeys and apes are not considered particularly good imitators (Whiten et al. 1996; Tennie et al. 2012; but see Horowitz 2003 and Taglialatela et al. 2012). On the contrary, songbirds and many other species of birds are able to imitate sounds and actions (Zentall 2006; Bolhuis et al. 2010). Unlike monkeys, songbirds display a segmented, hierarchically organized song structure and have been proposed to be able to learn syntactically recursive utterances (Gentner et al. 2006). However, they seem not to be able to generalize this learning to structurally similar sequences composed of different sounds (van Heijnningen et al. 2009; see also Beckers et al. 2012), implying that they are not able to grasp recursion as an abstract property. Nonetheless, children initially use phonetic clues to recognize structural patterns that later become assimilated as abstract rules (Mehler et al. 2006), suggesting that there is a stage in syntactic learning in which they also use sound patterns and memory to recognize structural regularities. Perhaps more striking for our purposes are the vocal and gestural imitative capacities of mammals like dogs (gestural imitation; Huber et al. 2009; Zentall 2006) and cetaceans (gestural and vocal; Herman 2002; Abramson et al. 2012). There are even reports of human voice imitation by belugas (Ridgway et al. 2012) and elephants (Stoeger et al. 2012). Certainly, there is the possibility that a mirror neuron system participates in these instances of imitation (Bonini and Ferrari 2011), although it is clear that these mammals are devoid of grasping abilities (with the possible exception of the elephant’s trunk).

The above impinges on Arbib’s proposal that a specific protosign stage took place before (and was possibly required for) the advent of protospeech and speech, which then “took over” the hand-based signing system. Hand-signing languages are usually claimed to support this view. However, these instances may as well reflect the fact that our brain is immensely plastic and that human communication is truly opportunistic, just like the case of Christy Brown, born with cerebral palsy, who learned to paint and write with his only controllable limb (this has been popularized in the movie “My Left Foot”). In my view, it is not clear what characteristics would the hypothetical protosign stage have had before the advent of speech, and to what extent it developed beyond simple pantomiming. Likewise, there are no convincing evolutionary relicts of the protosign stage during normal human development, which would be expected if this was a major transition in human evolution.

It is likely that vocalizations, and gestural communication in the form of pantomimes and other mechanisms like pointing to call attention, intimately co-evolved in early humans. But for the same reason, it seems to me difficult to separate a hand signing first, vocal signing second, sequence. Vocal plasticity was possibly critical in establishing mother-child bonds, and therefore was highly dependent on early social interactions. In later life, it became useful for making within-group alliances (Aboitiz et al. 2006, 2010). In this context, vocal imitations of sounds, like onomatopoeias, were possibly supported by mirror system circuits, and may have been an important factor in the acquisition of early meanings (Aboitiz 2012; Assaneo et al. 2011). Likewise, early signs, in the form of pantomimes and as ritualized cues to generate shared attention, were also probably very important in the development of a primitive semantics, making reference to objects or events in the near environment. The point that I am making here is that human communication has always been multimodal, with a continuous and strong interaction between gesture and voice.

3 Cortical control over the vocal apparatus

A central contention of the mirror neuron hypothesis concerns the notorious overlap of grasping mirror neurons and language circuits. Most grasping mirror neurons have been observed in the premotor area F5 (Brodmann's area 6v) and the rostral inferior parietal area; and "oral" mirror neurons (recognizing facial gestures like lip-smacking and feeding behaviors) are found in the lateral aspect of F5 (Belmalih et al. 2009; Gerbella et al. 2011; Rizzolatti and Craighero 2004), near the location of area 44 (see below). This has been confirmed by fMRI studies showing activation of areas F5, 45 and 46 in the frontal lobe, and in areas PFG and AIP of the parietal lobe, during grasping and imitation actions in both monkeys and humans (Nelissen et al. 2005, 2011). In these studies, projections to area F5 were described as originating in parietal areas PFG and AIP, which receive input from the superior temporal sulcus (STS), while input to area 45 was considered to originate directly in the ventral bank of the STS and from the lateral intraparietal area.

This circuit shows an important overlap with the language-related circuits which involve among other regions the superior temporal lobe, the STS and the ventrolateral prefrontal cortex (VLPFC), including areas 44 and 45, which is considered to represent the core of Broca's region (Hickok and Poeppel 2007). Furthermore, a participation of inferior parietal regions in language was also suggested, based on early fMRI findings indicating activation of inferoparietal areas PG and PFG during verbal working memory tasks (see Aboitiz and García 1997). Some sudies supported this proposal by evidencing a direct inferoparietal projection from the inferior parietal lobe (IPL) to areas 44 and 45 in both the monkey and the human (Catani and ffytche 2005; Petrides and Pandya 2009; Frey et al. 2008; Dick and Tremblay 2012). Projections from the IPL to the VLPFC are proposed to run mainly via the superior longitudinal fasciculus (SLF), while the projections from the (posterior) superior temporal lobe are considered to correspond to the arcuate fasciculus (Frey et al. 2008). Interestingly, in the monkey many fibers from the arcuate fasciculus run into dorsal prefrontal regions and are involved in the spatial processing of sound sources, while the pathway directed to Broca's region becomes more prominent in the human (Frey et al. 2008; Petrides and Pandya 2009; Yeterian et al. 2012).

Supporting our original contention (Aboitiz and García 1997), the role of the inferior parietal lobe in verbal working memory and in conduction aphasia has been highlighted in several recent studies (Hickok and Poeppel 2007; Buchsbaum et al. 2011). However, perhaps more than a phonological storage compartment, recent interpretations consider the role of parietal regions in working memory as generating sensory-directed action plans that contribute to the stability of the phonological motor program, which would be consistent with mirror-neuron activity (Buchsbaum and D'Esposito 2008, 2011; Aboitiz 2012).

In addition to this "dorsal" language circuit, there is a "ventral" projection reaching the VLPFC via the internal capsule from the anterior superior temporal gyrus, the STS and the middle temporal gyrus, conveying multimodal auditory and visual information (Belin and Zatorre 2000). This projection also fits a "ventral" mirror neuron circuit involved in action recognition (Arbib 2010). The dorsal and ventral pathways for language are widely considered to have different functions, the dorsal being associated to articulatory processes and sequential processing, while the ventral is more involved in lexical-semantic aspects (Hickok and Poeppel 2007; Frey et al. 2008). We initially proposed that the ventral pathway was more conserved in evolution, being a distinct tract already in the macaque, while the dorsal pathway had undergone a fundamental development in human evolution (Aboitiz et al. 2006). This proposal soon became confirmed by comparative tractographic studies (Rilling et al. 2008, 2011).

There are, nonetheless, a few anatomical discrepancies between authors regarding the homologies between certain brain regions, particularly Broca's region, and the specific termination of some projections, namely the IPL connection with the ventral frontal cortex (see Aboitiz 2012 and Dick and Trembley 2012 for full discussion). It has been widely claimed that F5 in the monkey represents the homologue of human's Broca's region (Rizzolatti and Craighero 2004). However, area F5 most likely corresponds to the premotor area 6 ventralis, while areas 45 and 44 (Broca's region *sensu stricto*) are distinguishable as separate from area 6v/F5 both in the human and in the macaque (Petrides et al. 2005; Belmalih et al.

2009). In the monkey, area 44 has been considered an ourburst of the anterior area 6v/F5, and is involved in communicative and ingestive orofacial movements (Petrides et al. 2005). Hand and mouth mirror neurons have been described in area F5, near the border with area 44, but evidence of mirror neurons in area 44 is yet to come. The second point refers to the site of termination of inferior parietal areas, mainly PG and PFG, into the inferior frontal cortex. While some authors emphasize their projection to areas 44 and 45 (Petrides and Pandya 2009; Preuss and Goldman-Rakic 1991), others have focused on the termination into the premotor region 6v/F5 (Belmalih et al. 2009; Gerbella et al. 2010). Despite these disagreements, the overall evidence points to a strong inferior parietal projection to the ventral premotor and/or ventrolateral prefrontal cortices. These two regions are involved in aspects of orofacial control, and participate in both speech production and speech perception (Dick and Tremblay 2012), suggesting an inferior parietal role in speech control.

Thus, beside the ventral projection between anterior temporal regions and the VLPFC, a rudimentary dorsal projection conveying auditory information to the VLPFC may have already existed in the monkey. This may have been direct via the arcuate fasciculus, and/or indirect, relying in the IPL. This is supported by tract-tracing studies (Petrides and Pandya 2009; Yeterian et al. 2012), and also by functional analyses indicating activation of posterior temporal, ventrolateral prefrontal, and parietal regions with presentation of conspecific vocalizations in the macaque (Gil-da-Costa et al. 2006). Furthermore, a recent report indicates the presence of neurons controlling voluntary vocalizations in the macaque ventral premotor cortex (Coudé et al. 2011). This may work in parallel with the descending projection from the medial cortex, controlling emotional vocalizations (Jürgens and Alipour 2002). Further studies are strongly required on this line.

I have proposed that the amplification of a dorsal auditory pathway to the VLPFC, including both a direct projection via the arcuate fasciculus and an indirect component via the IPL, was a key event in human evolution. As the ability of vocal learning developed, the primitive dorsal auditory-vocal circuit started to expand, recruiting neighboring circuits involved in other aspects like gestures and also hand control. Specifically, I suggested that the IPL was recruited to support working memory for vocalizations, by specifying motor goals based on sensory (acoustic) information. Together with the amplification of a cortical projection to brainstem vocal motor neurons, this innovation marked the beginnings of plastic vocalizations in our lineage, with an expanded working memory capacity and the appearance of the so-called phonological loop, which allowed speakers to learn increasingly complex vocal utterances that became of the utmost social relevance in establishing social bonds (Aboitiz and García 1997; Aboitiz et al.

2006, 2012; Aboitiz 2012). Consistent with this proposal, a recent study indicates that monkeys display limited auditory short-term memory, and they are consequently unable to transfer this short-term information into long-term auditory memory (Scott et al. 2012).

4 Final comment

To me, the mirror neuron hypothesis does not provide a clear answer to the following question: How did speech "take over" gestural communication? That is, how did auditory-vocal circuits come to dominate a robust gesture-dedicated circuit, out of nearly nothing? The possibility of a continuity between some aspects of monkey vocalization and human speech, and the existence of a rudimentary dorsal auditory-vocal pathway in the monkey might help to make this point clearer.

References

Aboitiz, F. 2012. Gestures, vocalizations and memory in language origins. *Frontiers in Evolutionary Neuroscience* 4(2) doi: *10.3389/fnevo.2012.00002*.

Aboitiz, F., S. Aboitiz & R. García. 2010. The phonological loop: a key innovation in human evolution. *Current Anthropology* 51(Suppl.S1). S55–S65.

Aboitiz, F. & R. García. 1997. The evolutionary origin of the language areas in the human brain. A neuroanatomical perspective. *Brain Research Reviews* 25(3). 381–396.

Aboitiz, F., R. García, C. Bosman & E. Brunetti. 2006. Cortical memory mechanisms and language origins. *Brain and Language* 98(1). 40–56.

Abramson, J. Z., V. Hernández-Lloreda, J. Call & F. Colmenares. 2012. Experimental evidence for action imitation in killer whales (Orcinus orca). *Animal Cognition* [Epub ahead of print Aug 9].

Arbib, M. A. 2010. Mirror system activity for action and language is embedded in the integration of dorsal and ventral pathways. *Brain and Language* 112(1). 12–24.

Assaneo, M. F., J. I. Nichols & M. A. Trevisan. 2011. The anatomy of onomatopoeia. *PLoS One* 6(12). e28317.

Beckers, G. J., J. J. Bolhuis, K. Okanoya & R. C. Berwick. 2012. Birdsong neurolinguistics: songbird context-free grammar claim is premature. *Neuroreport* 23(3). 139–145.

Belin, P. & R. J. Zatorre. 2000. 'What,' 'where' and 'how' in auditory cortex. *Nature Neuroscience* 3(10). 965–966.

Belmalih, A., E. Borra, M. Contini, M. Gerbella, S. Rozzi & G. Luppino. 2009. Multimodal architectonic subdivision of the rostral part (area F5) of the macaque ventral premotor cortex. *Journal of Comparative Neurology* 512(2). 183–217.

Bolhuis, J. J., K. Okanoya & C. Scharff. 2010. Twitter evolution: converging mechanisms in birdsong and human speech. *Nature Reviews Neuroscience* 11(11). 747–759.

Bonini, L. & P. F. Ferrari. 2011. Evolution of mirror systems: a simple mechanism for complex cognitive functions. *Annals of the New York Academy of Sciences* 1225(4). 166–175.

Buchsbaum, B. R., J. Baldo, K. Okada, K. F. Berman, N. Dronkers, M. D'Esposito & G. Hickok. 2011. Conduction aphasia, sensory-motor integration, and phonological short-term memory – an aggregate analysis of lesion and fMRI data. *Brain and Language* 119(3). 119–128.

Buchsbaum, B. R. & M. D'Esposito. 2008. The search for the phonological store: from loop to convolution. *Journal of Cognitive Neuroscience* 20(5). 762–778.

Buchsbaum, B. R., D. Ye & M. D'Esposito. 2011. Recency Effects in the Inferior Parietal Lobe during Verbal Recognition Memory. Front Hum Neurosci. 5(59). doi:10.3389/fnhum.2011.00059

Catani, M. & D. H. ffytche. 2005. The rises and falls of disconnection syndromes. *Brain* 128(10). 2224–2239.

Coudé, G., P. F. Ferrari, F. Rodà, M. Maranesi, E. Borelli, V. Veroni, F. Monti, S. Rozzi & L. Fogassi. 2011. Neurons controlling voluntary vocalization in the macaque ventral premotor cortex. *PloS ONE* 6(11). e26822. doi:10.1371/journal.pone.0026822

Dick, A. S. & P. Tremblay. 2012. Beyond the arcuate fasciculus: consensus and controversy in the connectional anatomy of language. *Brain*. Advance online access.

Ferrari, P. F., V. Gallese, G. Rizzolatti, and L. Fogassi. 2003. Mirror neurons responding to the observation of ingestive and communicative mouth actions in the monkey ventral premotor cortex. *European Journal of Neuroscience* 17(8). 1703–1714.

Fitch, W. T. & D. Reby. 2001. The descended larynx is not uniquely human. Proceedings of the Royal Society. Biological Sciences 268(1477). 1669–1675.

Frey, S., J. S. W. Campbell, G. B. Pike & M. Petrides. 2008. Dissociating the human language pathways with high angular resolution diffusion fiber tractography. *Journal of Neuroscience* 28(45). 11435–11444.

Gentner, T. Q., K. M. Fenn, D. Margoliash & H. C. Nusbaum. 2006. Recursive syntactic pattern learning by songbirds. *Nature* 440(7088). 1204–1207.

Gerbella, M., A. Belmalih, E. Borra, S. Rozzi & G. Luppino. 2010. Cortical connections of the macaque caudal ventrolateral prefrontal areas 45A and 45B. *Cerebral Cortex* 20(1). 141–168.

Gerbella, M., A. Belmalih, E. Borra, S. Rozzi & G. Luppino. 2011. Cortical connections of the anterior (F5a) subdivision of the macaque ventral premotor area F5. *Brain Structure and Function* 216(3–4). 43–65.

Ghazanfar, A. A., D. Y. Takahashi, N. Mathur & W. T. Fitch. 2012. Cineradiography of monkey lip-smacking reveals putative precursors of speech dynamics. *Current Biology* 22(13). 1176–1182.

Gil-da-Costa, R., A. Martin, M. A. Lopes, M. Muñoz, J. B. Fritz & A. R. Braun. 2006. Species-specific calls activate homologs of Broca's and Wernicke's areas in the macaque. *Nature Neuroscience* 9(8). 1064–1070.

Herman, L. M. 2002. Vocal, social, and self-imitation by bottlenosed dolphins. In: Kerstin Dautenhahn K and Chrystopher L. Nehaniv (eds.), *Imitation in animals and artifacts,* 63–108. Cambridge: MIT Press.

Hickok, G. & D. Poeppel. 2007. The cortical organization of speech processing. *Nature Reviews Neuroscience* 8(5). 393–402.

Horowitz, A. C. 2003. Do humans ape? Or do apes human? Imitation and intention in humans (*Homo sapiens*) and other animals. *Journal of Comparative Psychology* 117(3). 325–336.

Howard, M. F. & D. Poeppel. 2012. The neuromagnetic response to spoken sentences: co-modulation of theta band amplitude and phase. *Neuroimage* 60(4). 2118–2127.

Huber, L., F. Range, B. Voelkl, A. Szucsich, Z. Viranyi & A. Miklosi. 2009. The evolution of imitation: what do the capacities of nonhuman animals tell us about the mechanisms of imitation? *Philosohical Transactions of the Royal Society of London B Biological Sciences* 364(1528). 2299–2309.

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

Mehler, J., M. Nespor, M. Shukla & M. Peña. 2006. Why is language unique to humans? *Novartis Foundation Symposium* 270. 251–292.

Morrill, R. J., A. Paukner, P. F. Ferrari & A. A. Ghazanfar. 2012. Monkey lipsmacking develops like the human speech rhythm. *Developmental Science* 15(4). 557–568.

Nelissen, K., E. Borra, M. Gerbella, S. Rozzi, G. Luppino, W. Vanduffel, G. Rizzolatti & G. A. Orban. 2011. Action observation circuits in the macaque monkey cortex. *Journal of Neuroscience* 31(10) 3743–3756.

Nelissen, K., G. Luppino, W. Vanduffel, G. Rizzolatti & G. A. Orban. 2005. Observing others: multiple action representation in the frontal lobe. *Science* 310(5746). 332–336.

Petrides, M., G. Cadoret & S. Mackey. 2005. Orofacial somatomotor responses in the macaque monkey homologue of Broca's area. *Nature* 435(7046). 1235–1238.

Petrides, M. & D. N. Pandya. 2009. Distinct parietal and temporal pathways to the homologues of Broca's area in the monkey. *PLoS Biology* 7(8). e1000170. doi:10.1371/journal.pbio.1000170

Preuss, T. & P. S. Goldman-Rakic. 1991. Ipsilateral cortical connections of granular frontal cortex in the strepsirhine primate Galago, with comparative comments on anthropoid primates. *Journal of Comparative Neurology* 310(4). 507–549.

Ridgway, S., D. Carder, M. Jeffries & M. Todd. 2012. Spontaneous human speech mimicry by a cetacean. *Current Biology* 22(20). R860–R861.

Rilling, J. K., M. F. Glasser, T. M. Preuss, X. Ma, T. Zhao, X. Hu & T. E. J. Behrens. 2008. The evolution of the arcuate fasciculus revealed with comparative DTI. *Nature Neuroscience* 11(4). 426–428.

Rilling, J. K., M. F. Glasser, S. Jbabdi, J. Andersson & T. M. Preuss. 2011. Continuity, divergence, and the evolution of brain language pathways. *Frontiers in Evolutionary Neuroscience* 3(11). doi:10.3389/fnevo.2011.00011

Rizzolatti, G. W. & L. Craighero. 2004. The mirror-neuron system. *Annual Reviews in Neuroscience* 27. 169–192.

Scott, B. H., M. Mishkin & P. Yin. 2012. Monkeys have a limited form of short-term memory in audition. *Proceedings of the National Academy of Science U.S.A.* 109(30). 12237–12241.

Stoeger, A. S., D. Mietchen, S. Oh, S. de Silva, C. T. Herbst, S. Kwon & W. T. Fitch. 2012. An asian elephant imitates human speech. *Current Biology* 22(22). 2144–2148.

Taglialatela, J. P., L. Reamer, S. J. Schapiro & W. D. Hopkins. 2012. Social learning of a communicative signal in captive chimpanzees. *Biological Letters* 8(4). 498–501.

Tennie, C., J. Call & M. Tomasello. 2012. Untrained chimpanzees (*Pan troglodytes schweinfurthii*) fail to imitate novel actions. *PLoS One* 7(8). e41548. doi:10.1371/journal.pone.0041548

van Heijningen, C. A. A., J. de Visser, W. Zuidema & C. ten Cate. 2009. Simple rules can explain discrimination of putative recursive syntactic structures by a songbird species. *Proceedings of the National Academy of Science U.S.A.* 106(48). 20538–20543.

Whiten, A., D. M. Custance, J. C. Gómez, P. Teixidor & K. A. Bard. 1996. Imitative learning of artificial fruit processing in children (Homo sapiens) and chimpanzees (*Pan troglodytes*). *Journal of Comparative Psychology* 110(1). 3–14.

Yeterian, E. H., D. N. Pandya, F. Tomaiuolo & M. Petrides. 2012. The cortical connectivity of the prefrontal cortex in the monkey brain. *Cortex* 48(1). 58–81.

Zentall, T. R. 2006. Imitation: definitions, evidence, and mechanisms. *Animal Cognition* 9(4). 335–353.