NEURONAL RECYCLING AS SPEECH ORIGINS

AXEL G. EKSTRÖM[*1]

*Corresponding Author: axeleks@kth.se
[1]KTH Speech, Music & Hearing, Stockholm, Sweden

The neuronal recycling hypothesis posits that cultural inventions invade evolutionarily adapted neural circuitry and facilitate predictable neurological development in humans. I present an account of phonological development as derived from that same set of proposed principles. So applied, the hypothesis predicts (1) observable neural biases and abilities of communicative learning continuous across primates, (2) patterns of cross–culturally expectable phonological development in human infants, and (3) principled manifestation of phonologies across human cultures. I argue that all three are supported by empirical work in primatology, neurolinguistics, developmental psychology, and phonetics. Speech is invented anew in every speaking child, but representations of speech-centric behavior are consistently allocated to the same neural architecture across individuals – structures that house related functionality in non–human primates. Development of speech production and perception also follow predictable trajectories across cultures. Relevant scientific findings thus support the view that human infants are equipped with neural architecture evolutionarily prepared for the development of vocal–communicative behavior, which skew their ontogenetic development and constrain their ultimate expression. Implications for speech ontogenesis are discussed. The present work adds to the understanding of systems of speech as products of cultural evolution.

1. Introduction

All humans who have ever held a conversation have solved the same problem in ontogeny. They have learned to speak. Yet, despite their importance, the mechanisms by which speech is learned by the growing child remain poorly understood, and in urgent need of hypothesizing. In this work, I address the ontogenetic development of speech, and the evolutionary roots from which learned experience drives emergent speech behavior. I am not here concerned with aspects of social cognition (e.g., intentionality), but the biological architecture that underlies spoken language in particular. A view of speech from evolution must base its assumptions on principles of the natural world, without resorting to abstract theorizing. I address the fundamental question of how

abilities of speech may develop in the individual, without taking out significant loans that theories of evolution cannot hope to cash. The present stated goal, rather, is an account of speech ontogenesis that derives speech, and its neural, and perceptual features, from more basic elements of non–speech behavior.

The process by which phylogenetically adaptive neural structures are coopted for new uses has been described by various researchers (see Andersson, 2010). For one, Dehaene (2009) has argued that preexisting evolutionarily adaptive architectures in the human brain enact constraints upon the development of writing systems – cultural inventions too recent to have driven any significant evolutionary selection pressure – such that they exhibit principally emergent features across cultures. Accordingly, reading in the brain predictably occupies the same neural regions across individuals (i.e., the visual word form area). This set of theoretical assumptions is referred to as the neuronal recycling hypothesis (hereafter NR). By analogy with reading, I posit that speech development is similarly non-arbitrary but follows a predictable protocol. It is not claimed that speech as such is solely a learned behavior; rather, speech capacities likely represent a relatively recent evolutionary heritage in *Homo* (see Ghazanfar et al., 2012; MacNeilage, 1998; Lieberman, 2017; Lieberman et al., 2019). However, speech is also invented anew by every developing child, and the processing and production of speech is localized to the same neural architecture across individuals, languages, and cultures, but selectively sensitive to language-specific phonologies and linguistic cues. As cultural mutually agreed–upon conventions, human spoken languages possess multiple distinct common features, facilitating formalization of linguistic codes and usage. The NR framework helps explain several puzzling findings in speech-centric science, including observed constraints on individual linguistic development in ontogeny, and the apparently limited size of phonological repertoires inherent to particular languages. The text thus also places human phonology in a cultural evolution framework, such that systems of speech, themselves products of culture, exhibit influence over human linguistic development. The present account argues that the fallout of such organization in turn tunes and organizes the neural circuitry of the infant and child exposed to it.

2. Neuronal recycling for speech

Applied to speech, NR can be postulated as a set of three hypotheses (Dehaene, 2009; Christiansen & Müller, 2015). In the follow sections, support for each hypothesis is discussed at length. First, NR predicts (1) phylogenetic continuity across related species. For speech, such a supposition touches on a rich and

growing research tradition on the linguistic capacities of non–human primates (hereafter primates). Further, NR anticipates (2) predictable linguistic development in infants across cultures. By now, the trajectory of children's acquisition of speech capacities has been described in significant detail, allowing for a composite illustration. Finally, NR posits (3) limited variability in expression, resulting from the predispositions of adapted neurological structures toward learning prelusive behavior. Systems of speech are indeed characterized by a range of cross-cultural commonalities. Regardless of language, culture, and geography, speech-centric behavior is allocated to the same neural architecture but develops selective sensitivities to language-specific phonological elements.

2.1 Primate speechlike behavior

Articulate speech is a novel invention in the hominid lineage, its biomechanical components having evolved gradually. Because humans are the only (extant) species capable of generative speech, a proper cross–species continuity may initially appear impossible, or even undesirable. However, recent research on the communicative abilities and vocal capacities of primates allows for careful inference with relevance for the evolution of speech. Because language does not fossilize, bioanthropological research has instead focused on identifying biological precursors to speech. While the anatomy of the vocal tract itself is highly conserved in mammalian species (Negus, 1949) and thus unlikely to have contributed to the evolution of speech, much discussion has centered on uniquely human adult anatomy likely to enable speech (e.g., Lieberman et al., 2001; Lieberman, 2017). Of potentially greater relevance to understanding ontogenetic emergence of phonology, however, are studies of vocalizations themselves.

While primate communicative repertoires often include either vocalization or jaw–driven speech–like motor behavior, such as lip–smacking (e.g., Ghazanfar et al., 2012), the ability to couple the two is typically absent. Further, the acoustic structure of primate calls appears largely fixed, such that call behaviors develop reliably regardless of social experience (e.g., Cheney et al., 1992). A proper discussion of anatomical and neural limitations precluding primates from human or humanlike speech is beyond the scope of the present text and remains the subject of some controversy (see e.g., Lieberman, 2017). As a behavioral composite of voluntary mandibular and laryngeal control, speech is contingent on cortical coupling (i.e., overlap) between jaw and larynx muscle primary motor somatotopic representations (MacNeilage, 1998; Brown et al., 2021). In humans, the laryngeal motor cortex – the primary cortical center for control of vocalization – is located in the ventral primary motor cortex; in primates, the same region is

located in the ventral premotor cortex. Thus, the primary neural region of vocalization control has undergone a dorsal shift with the evolution of the *Homo* genus, likely contributing to the evolution of speech (Simonyan & Horwitz, 2011). Further, comparative neuroscience suggests that neural architecture commonly associated with human speech have correlates in the brains of other hominids, with species such as Chimpanzees (*Pan troglodytes*) possessing neuroanatomical homologues to Broca's area (Brodmann areas 44, 45), which appears to activate during vocal signaling (see e.g., Taglialatela et al., 2011).

Finally, there is substantial continuity in vocal perception across the primate lineage. Research on the call repertoires of primates suggests that modification of produced calls remains relatively limited across the lifespan, with more flexibility in perception than production (e.g., Cheney et al., 1992). Such findings are echoed by observations of human infants, who develop the ability to perceive and differentiate ingroup language-specific phonemes around the age of six months – long before they begin producing it reliably and willingly. Neurologically, speech perception and language–specific perception in the Broca's area also develop during the first year of life (Imada et al., 2006), whereas the articulatory vocal production organs take several months yet to bring under voluntary control.

2.2. A view from ontogeny

In human infants, mandibular oscillatory motor movements later coopted for syllable production are exclusively utilized for voiceless sucking (i.e., feeding) behavior, and are not coupled with motor movement of the vocal tract. Vocalizations observed during this stage include crying and are – unlike speech proper – innate, involuntary, and likely contingent on the same neural circuitry that underlies innate call repertoires observed in other hominids (Jürgens, 2009). Acoustically, infants' crying aligns more closely with primate vocalizations, than with adult human speech (Lieberman et al., 2019). Indeed, stemming from anatomical limitations, the infant vocal apparatus is likely incapable of producing anything resembling the range of speech sounds available to the adult speaker. Only around the age of six months does this begin to change, with the pre–verbal infant's invention of babbling, a form of emergent protospeech characterized by articulate non–word speech sounds (i.e., 'vocal play'; see Vihman, 2013). While there has been significant debate over the relationship of babbling to speech proper, researchers have long since mapped its developmental trajectory.

Human infants typically begin to babble by no later than six months of age. Around the same time, infants also begin to display an ability to discriminate perceptually between different vowel and consonant sounds (see Oller, 2000). At

this point in development, infants exhibit the emergent ability to voluntarily open and close the vocal tracts. NR further predicts the observed development of speech–centric behavior in the growing infant (for an overview, see Vihman, 2013). The innate mandibular–oscillatory motor behavior utilized for suckling provide a scaffold for the earliest occurring vowel-like "cooing" sounds, which are predictably produced by human infants in the first few months (Oller, 2000). While languages exhibit significant variability in vowel systems (Maddieson, 1984), such early pseudo vowels roughly correspond to the adult /a/ and /u/, which appear almost universally across languages. Similarly, phonemes composed from bilabial plosives such as /maː/ or /baː/ are more readily produced and acquired by the individual infant early in development, around the age of 10 months. It is thus not surprising that even unrelated languages have similar–sounding words for 'mother' – tellingly, some variation of /maː/ or /məˈmɑː/. This stands in contrast to, for example, phonemes composed from fricative alveolars such as /taː/ or /kaː/, which require the use of the tongue tip and tongue body, respectively; substantial articulatory maneuverability and muscle dexterity need develop before the production of such sounds becomes possible. That is, a developmental threshold must be passed before relatively complex speech sounds can be achieved.

Neurologically, then, at this stage of development, the biomechanical equipment necessary for organizing speech into consonant 'frames', and consonant 'content' (MacNeilage, 1998) has likely begun to take place in the infant; however, the process of adapting to the ambient soundscape and local phonology (i.e., available phonemes inherent to the language spoken by present ingroup members; Goldstein et al., 2003) is not yet completed. Rather, the infant has reached a stage of orosensory exploration, where possible speech sounds – made available by combinations of laryngeal and mandibular motor activity – can be mapped onto perceptual phonemes (Kuhl, 2000). The resulting composite emphasizes their relationship in perceptual memory, enabling successful encoding across time. Thus, while a set of phonemes may be 'given', infants are born with a great range of possible developments, which are subsequently pruned through cultural learning and ingroup interaction. Such plasticity is greater in children than adult humans, facilitating acquisition of speech-centric behavior. Thus, in adult language learners, foreign speech sounds. For example, the Japanese language lacks the phoneme /l/ as observed in e.g., English, and as a result, adult Japanese speakers readily experience difficulty perceiving and producing it (see Yamada & Tohkura, 1992). Crucially, while being able to reliably discriminate /l/ from /r/ at six months of age, at eight months, Japanese infants have apparently lost this ability (Tsuhima et al., 1994). That is, cultural

learning via attention to, and interaction with, adult caregivers have formed the basis for later-in-life language-specific phonological speech production (Goldstein et al., 2003). That is, much of the necessary recycling takes place in the early years of development, after which much of the plasticity required for achieving articulatory fluency with regard to a given phonology is exhausted.

2.3. Principled manifestation

NR predicts constraints on cultural–environmental factors, such that culture influences the development of evolutionarily adaptive brain circuits – but not unlimitedly so. All languages are indeed organized through similar constraints on composition (e.g., grammar; but for a nuanced discussion, see Christiansen & Chater, 2008) regulating their combinatorial capacities; and content (i.e., words as units of transmission). Additionally, in another homologue to writing systems, representations of speech–centric behavior and processing is also predictably allocated to the same neural architectures (e.g., Brodmann areas 22, 44, and 45) in all human brains. Yet, human infants are born without language, and need to develop substantially before language acquisition becomes cognitively feasible.

As culturally agreed-upon systems of communication, languages exhibit significant cross-cultural similarities. Whereas all (spoken) languages constitute phonologically combinatorial systems, attempts have been made to formalize their structure according to phonological universals. The most successful of these posits that spoken language is typically organized in consonant/vowel cycles (MacNeilage, 1998). Further, spoken languages contain a limited pool of speech sounds. While phonemic libraries vary significantly between languages (Maddieson, 1984), the organization and character of those libraries are naturally constrained by the mechanical apparatus of speech. Namely, the human articulators – the tongue, upper and lower lips, alveolar ridge, hard palate, velum, uvula, pharyngeal wall, and glottis – allow for a large though ultimately limited set of articulate speech sounds. The resulting phonemic repertoire, averaging ~40 phonemes in a given language (Maddieson, 1984), however, is constrained in not only the dimension of production – resulting from biological and anatomical limitations of the vocal cords, vocal tract, and phonemic articulators – but also that of perception, operationalized as limitations on cognitive processing and memory, with classic work on optimal maximization of vowel space by Liljencrants and Lindblom (1972) indicating a crucial role of perceptual contrast in the emergent organization of phonetic structure of language vowel systems. Relatedly, Christianson and Chater (2008) have argued for a set of cognitive constraints on language evolution, shaped by limitations of the human brain. The

particular repertoire of any given spoken language should thus be seen as resulting from a combination of economic-organizational and articulatory limiting factors. Whereas the most prominent limitations on phonemic expression are biological and anatomical, principles of economy of memory impose further limitations on the emergent phonemic repertoire recruited into systems of speech.

3. Conclusions

I have presented an account of speech ontogenesis, based on principles of neurobiology, phonetics, and observations of phonological and linguistic development. It is the view presented here that systems of speech, as learned cultural practices enact significant pressure on phonological and neurological development. While the text is primarily concerned with evolutionary, developmental, and neurological aspects of phonological activity, it aligns with a cultural evolution framework. Exposure to language-specific phonology prunes the infant listener for continued interaction and future speech sound production and perception. Speech systems – motor as well as orosensory – are subject to extensive such pruning throughout infancy and early childhood in preparation and accommodation for native-language speech sounds. Language-specific speech behavior can thus be seen as a product of recycling by cultural systems of speech.

References

Anderson, M. L. (2010). Neural reuse: A fundamental organizational principle of the brain. *Behavioral and brain sciences*, *33*(4), 245-266.

Brown, S., Yuan, Y., & Belyk, M. (2021). Evolution of the speech-ready brain: The voice/jaw connection in the human motor cortex. *Journal of Comparative Neurology*, *529*(5), 1018‑1028.

Cheney, D. L., Owren, M. J., Dieter, J. A., & Seyfarth, R. M. (1992). 'Food'calls produced by adult female rhesus (Macaca mulatta) and Japanese (M. fuscata) macaques, their normally‑raised offspring, and offspring cross‑fostered between species. *Behaviour*, *120*(3‑4), 218‑231.

Christiansen, M. H., & Chater, N. (2008). Language as shaped by the brain. *Behavioral and brain sciences*, *31*(5), 489-509.

Christiansen, M. H., & Müller, R. A. (2015). Cultural recycling of neural substrates during language evolution and development. *The cognitive neurosciences V*, 675‑682.

Dehaene, S. (2009). Reading in the brain. *New York*.

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

Goldstein, M. H., King, A. P., & West, M. J. (2003). Social interaction shapes babbling: Testing parallels between birdsong and speech. *Proceedings of the National Academy of Sciences*, *100*(13), 8030–8035.

Imada, T., Zhang, Y., Cheour, M., Taulu, S., Ahonen, A., & Kuhl, P. K. (2006). Infant speech perception activates Broca's area: a developmental magnetoencephalography study. *Neuroreport*, *17*(10), 957–962.

Jürgens, U. (2009). The neural control of vocalization in mammals: a review. *Journal of Voice*, *23*(1), 1-10.

Kuhl, P. K. (2000). A new view of language acquisition. *Proceedings of the National Academy of Sciences*, *97*(22), 11850-11857.

Lieberman, D. E., McCarthy, R. C., Hiiemae, K. M., & Palmer, J. B. (2001). Ontogeny of postnatal hyoid and larynx descent in humans. *Archives of oral biology*, *46*(2), 117–128.

Lieberman, P. (2017). Comment on "Monkey vocal tracts are speech-ready". *Science Advances*, *3*(7), e1700442.

Lieberman, P., Harris, K. S., Wolff, P., & Russell, L. H. (2019). *Newborn infant cry and nonhuman–primate vocalizations* (pp. 41–53). De Gruyter Mouton.

Liljencrants, J., & Lindblom, B. (1972). Numerical simulation of vowel quality systems: The role of perceptual contrast. *Language*, 839-862.

MacNeilage, P. F. (1998). The frame/content theory of evolution of speech production. *Behavioral and brain sciences*, *21*(4), 499–511.

Maddieson, I. (1984). *Patterns of sounds*. Cambridge: Cambridge University Press.

Negus, V. (1949). *The comparative anatomy and physiology of the larynx*. Heinemann.

Oller, D. K. (2000). *The emergence of the speech capacity*. Psychology Press.

Simonyan, K., & Horwitz, B. (2011). Laryngeal motor cortex and control of speech in humans. *The Neuroscientist*, *17*(2), 197–208.

Taglialatela, J. P., Russell, J. L., Schaeffer, J. A., & Hopkins, W. D. (2011). Chimpanzee vocal signaling points to a multimodal origin of human language. *PloS one*, *6*(4), e18852.

Tsushima, T., Takizawa, O., Sasaki, M., Shiraki, S., Nishi, K., Kohno., et al. (1994). Discrimination of English /r-l/ and /w-y/ by Japanese infants at 6-12 months: language-specific developmental changes in speech perception abilities. *1994 International Conference on Spoken Language Processing*, pp. 1695-1698. Yokohama: Japan.

Vihman, M. M. (2013). *Phonological development: The first two years*. John Wiley & Sons.

Yamada, R. A., & Tohkura, Y. I. (1992). The effects of experimental variables on the perception of American English/r/and/l/by Japanese listeners. *Perception & psychophysics*, *52*(4), 376-392.