The phoneme as cognitive technology

Caleb Everett[1], Steven Moran[1,2], and Axel G. Ekström[3*]

[*]Corresponding Author: axeleks@kth.se
[1]Department of Anthropology, University of Miami, Coral Gables, FL, US
[2]Institute of Biology, University of Neuchâtel, Neuchâtel, Switzerland
[3]Division of Speech, Music & Hearing, KTH Royal Institute of Technology, Stockholm, Sweden

1. Introduction

While the number and characteristics of vowels and consonants are highly variable across the world's spoken languages (Moran & McCloy, 2019), all speakers make consistent and deliberate use of a relatively narrow set of contrastive speech sounds, i.e., phonemes. Such remarkable ubiquity is suggestive of extensive benefits to their speakers. Importantly, however, the phoneme is not a naturally occurring phenomenon. Rather, we argue that phonological systems constitute cognitive tools, i.e., that they support, guide, and extend speaker cognitive capacities (Everett, 2017). We make several claims toward this point.

2. Information rate

Compared with the communication systems of nonhuman animals, the consistent and socially deliberated use and reuse of phonemes enables rapid information transmission rates through syllabic speech (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967). Based on their work constructing reading machines for the blind, Liberman and colleagues noted that if spoken language were produced letter by letter (or phoneme by phoneme), then speech rates would be significantly reduced. In real-life speech, however, speech is always coarticulated – the production of a phoneme is continually affected by its context. Phonemes in a spoken language may serve as points of reference, even as speech sounds are distorted by linguistic and extra-linguistic (e.g., emotional speech) contexts (Lindblom, 1990). The cultural "invention" of the phoneme in human evolution and society, thus, enabled the rapid information transmission rates universally observed across human languages (Coupé, Oh, Dediu, & Pellegrino, 2019).

3. The phoneme as developmental scaffold

Native-language input provides developing human infants with acoustic-perceptual goals (Kuhl, Williams, Lacerda, Stevens, & Lindblom, 1996), the

replication or quasi-replication of which, serve as a marker of emergent social consciousness and identity (Tomasello, 2003). This development takes place in all developing human infants within the first few years of life (Vihman, 2014), likely exploiting existant subcortical neuromotor systems (see review in Ekström, 2022). Experimental evidence from perceptual abilities by human infants suggests that already by six months of age, humans begin selectively discriminating between language-specific phonemes; this selective perception eventually develops into a so-called "perceptual magnet" (Kuhl, 1991), effectively serving as a prototype for its category.

4. Perceptual overlap

Non-human animal oral tracts afford the capacity to non-uniformly affect formants. Recent work indicates that orangutan "long calls" are readily perceived as phonemic by listeners (Ekström, Moran, Sundberg, & Lameira, 2023). We propose that exposure to systems of speech sounds bias human perceptual systems toward selective perception of environmental sounds as speech-like, including the calls of other animals. Accordingly, vocalizations produced even by distantly related animals such as domestic cats (*Felis catus*) are uniformly transcribed – across even unrelated languages – as a consonant-vowel-consonant or consonant-vowel-vowel sequence corresponding, e.g., to /miauw/. We argue that the reason animal vocalizations may be perceived as essentially "word–like" (Nicastro & Owren, 2003) is contingent on phonemic learning.

5. Implications for evolutionary phonology

Treating contrasting sounds of speech as products of culture may also open up novel discussions of its evolution. Recent work in archaeology and anthropology point to distinctions between findings, such that tools of relatively low complexity cannot be used to infer cultural transmission (Snyder, Reeves, & Tennie, 2022). We suggest that similar distinctions may be made with regards to systems of speech in human evolution, such that a possible "early" system could be independently invented by individuals and groups, while more complex systems required cultural transmission (Benítez-Burraco & Kempe, 2018). This suggestion may be explored through computational modeling (Kirby & Hurford, 2002). Our view complements cognitive linguistic perspectives on human perception and consciousness by emphasizing the sounds of speech themselves.

Acknowledgements

SM and AE were funded by the Swiss National Science Foundation (PCEFP1_186841). The results of this work and the tools used will be made more widely accessible through the national infrastructure Språkbanken Tal under funding from the The Swedish Research Council (2017-00626, AE).

References

Benítez-Burraco, A., & Kempe, V. (2018). The emergence of modern languages: has human self-domestication optimized language transmission? *Frontiers in Psychology*, *9*, 551.

Coupé, C., Oh, Y. M., Dediu, D., & Pellegrino, F. (2019). Different languages, similar encoding efficiency: Comparable information rates across the human communicative niche. *Science Advances*, *5*(9), eaaw2594.

Ekström, A. G. (2022). Motor constellation theory: A model of infants' phonological development. *Frontiers in Psychology*, *13*, 996894.

Ekström, A. G., Moran, S., Sundberg, J., & Lameira, A. R. (2023). Prequel: Supervised phonetic approaches to analyses of great ape quasi-vowels. In R. Skarnitzl & J. Volín (Eds.), *Proceedings of the 20th International Congress of Phonetic Sciences* (pp. 3076–3080). Prague, Czech Republic.

Everett, C. (2017). *Numbers and the making of us: Counting and the course of human cultures.* Harvard University Press.

Kirby, S., & Hurford, J. R. (2002). The emergence of linguistic structure: An overview of the iterated learning model. In A. Cangelosi & D. Parisi (Eds.), *Simulating the evolution of language* (pp. 121–147). Springer London.

Kuhl, P. K. (1991). Human adults and human infants show a "perceptual magnet effect" for the prototypes of speech categories, monkeys do not. *Perception & Psychophysics*, *50(2)*, 93–107.

Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., & Lindblom, B. (1996). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science*, *255(5044)*, 606–608.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review*, *74*(6), 431–461.

Lindblom, B. (1990). Explaining phonetic variation: A sketch of the h & h theory. In W. J. Hardcastle & A. Marchal (Eds.), (pp. 403–439). Springer.

Moran, S., & McCloy, D. (2019). *Phoible 2.0.* Jena: Max Planck Institute for the Science of Human History.

Nicastro, N., & Owren, M. J. (2003). Classification of domestic cat (*Felis catus*) vocalizations by naive and experienced human listeners. *Journal of Comparative Psychology*, *117(1)*, 44.

Snyder, W. D., Reeves, J. S., & Tennie, C. (2022). Early knapping techniques do not necessitate cultural transmission. *Science Advances*, *8(27)*, eabo2894.

Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition.* Cambridge, MA: Harvard University Press.

Vihman, M. M. (2014). *Phonological development: The first two years.* Wiley.