CATEGORICAL PERCEPTION AND DISCREETNESS: AN EVOLUTIONARY AND NEUROCOGNITIVE PERSPECTIVE

ELIZABETH QING ZHANG 1[*1], EDWARD RUOYANG SHI 2[2]

[*]Corresponding Author: zqelizabeth@gmail.com

[1] Center for Language Evolution Studies, University Centre of Excellence IMSErt, Nicolaus Copernicus University, Torun, Poland

[2] Department of Catalan Philology and General Linguistics, University of Barcelona, Barcelona, Spain

As one of the design features of human language, discreteness is present at both phonological and syntactic levels. We hypothesize that categorical perception (CP) could have paved the way for discreteness across domains including language. Evidence from comparative and neurobiological perspectives are provided in this paper. Comparative studies on CP suggest that it has a phylogenetic root and is a combination of nature and nurture. Neurobiological studies suggest that the connection between auditory cortex and (pre)motor cortex, and between (pre)motor cortex and the basal ganglia could play an important role in CP and further discreteness. This paper discusses how domain general perceptual ability—categorical perception—could have contributed to discreteness, highlighting the role of sensory-motor system in the evolution of language.

1. Introduction

In this paper, we propose that Categorical Perception (CP) could have played a foundational role for discreteness of language, which is a prominent feature of human language (Hockett, 1960). We will approach discreteness from domain general perspective and highlight how it is salient in language. Then by reviewing CP of sounds in nonhuman animals, we argue that CP has its phylogenetic roots in terms of evolution. Following this, we explicate how CP could have been the basis for discreteness with neurological evidence focusing on auditory cortex, (pre)motor cortex and the basal ganglia. The current work discusses the role of perception in language evolution, which provides new avenue to explore the evolution of human language from sensory-motor system.

2. Discreteness

Discreteness is an essential concept in several areas, such as quantum physics and chemistry. As Hauser et al. (2002) pointed out, life is constructed on separate and

unblendable components (genes). Besides, an example in our daily lives is that time is calculated as discrete in the units of seconds. In terms of human cognition, it is generally established that there exist discrete units of mathematics, such as integers, and discrete units of music, such as melody and rhythm. In the domain of human language, as Hockett (1960) pointed out, discrete and compositional language differs from continuous and holistic nonhuman animal communication signals, which he considered as one of the design features of human language. But what Hockett has emphasized is phonological structure of language, namely syllables are made up of separate vowels and consonants at the phonological level. However, discreteness is also observed at syntactic levels. Through merging lexical items into grammatical phrases or sentences at the syntactic level, lexical elements are discretely dispersed and organized into grammatical phrases or sentences. These discrete atoms at various levels of language combine to form bigger structures based on the rules at that level. For example, you may find 6- and 15-word sentences, but not 7.5-word ones. Just as Searle (2008; p 176) defined, discreteness is "the property through which syntactical elements keep their identity when subjected to diverse syntactical processes".

3. Categorical perception and Discreteness: Comparative evidence

Categorical perception (CP) is a psychophysical process in which continuous inputs are perceived discretely across modalities (Harnad, 1987). The interactions between low-level perception and high-level cognition are revealed by this implicit segmentation of continuous physical inputs (Goldstone & Hendrickson, 2010). The mechanism of CP has been investigated utilizing psychophysical approaches such as labeling and discerning continuous stimulus paradigms (Harnad, 1987). The phenomenon of CP was initially identified in sound modality in human speech sounds (Liberman et al., 1957), where listeners were prone to perceive the b-d-g continuum into three distinct groups. Furthermore, research on infants has revealed that they appear to be endowed with the ability to discriminate different sounds in all languages (which is the nature part), but as they have more contact with one (or more for multilinguals) ambient language(s), they tend to group sounds that are not contrastive in their native language(s). For instance, after 8-10 months, Japanese newborns can no longer distinguish between /l/ and /r/, despite the fact that these two sounds are acoustically distant. Statistical learning (Maye et al., 2002) has been proposed to achieve such postnatal shaping CP of phonemes, which appears to be reliant on the requirement that speech sounds are discrete. Statistical learning is the process of learning the statistical distribution of environmental inputs via sheer exposure to the stimuli

without instruction (Romberg & Saffran, 2010). The categories boundaries of the stimuli could be changed as a result of statistical learning. The finding of Japanese newborns no longer discriminating between /l/ and /r/ might be due to implicit statistical learning of the non-contrastive distribution of /l/ and /r/ in Japanese. Statistical learning may also be seen in the process of segmenting words from speech streams (Saffran et al., 1996), which underlies CP in the sense that discrete elements are extracted from continuous inputs. It is also worth mentioning that statistical learning is used in a variety of cognitive domains other than language (Bogaerts et al., 2020).

It is also worth emphasizing that CP is not exclusive to speech sounds, but rather a phenomenon observed in a variety of domains and modalities. If our hypothesis that discreteness arises from CP is correct, the discovery of CP in nonhuman animals shows that discreteness of language may have been derived from a preserved trait, namely CP. Comparative studies in nonhuman animals reveal that CP is phylogenetically anchored in early invertebrates and appears to be a combination of nature and nurture (Zhang et al., 2021). The psychophysical paradigm was also employed to investigate CP in nonhuman animals, and the participants were instructed to discriminate and categorize the stimuli, much as it did in human research. We present examples of CP in the sound modality in nonhuman animals.

Crickets have been observed to be able to distinguish between communication calls and predator ultrasounds at a sharp border when it comes to naturally produced noises (Wyttenbach et al., 1996). Female tungara frogs respond to mating sounds in a categorical manner (Baugh, Akre & Ryan, 2008). These findings imply that sound CP is linked to reproduction and survival. In addition, early research looked at how nonhuman animals categorize human speech sounds, which demands a high level of auditory learning capacity. Nonhuman primates were shown to be able to discriminate both consonants and vowels (e,g. Sinnott & Mosteller, 2001). Furthermore, the CP of sounds in birds was studied from the standpoints of both conspecific calls and human speech sounds. It was discovered as early in the 1980s that budgerigars not only had a low discrimination frequency corresponding to their contact calls, but also had a similar range of voice onset time to humans in the perceptual change of bilabial, alveolar, and velar continua (Dooling et al. 1987). The budgerigars used the same cues as humans to distinguish between vowel groups (Dooling & Brown, 1990). This shows that CP of speech sounds is related to auditory perception in general. Moreover, the vocal learning continuum hypothesis, with the most widely distributed species having auditory learning ability and a small number of species

having vocal production learning ability, is in some ways paralleled by the large-range distribution of categorical perception in sound modality in diverse animals (Jarvis, 2019).

How might CP of sounds be linked to the discreteness of language? To perceive the world in a discrete way seems to be a prerequisite for discrete production. Then how is the discreteness of sounds related to the discreteness of lexical items or words? In this context, one of the most intriguing and contentious questions is how lexical items or words emerged in evolution. Two scenarios have been considered: One is that words evolved directly from animal communication; and the other one is that words evolved from some other cognitive ability. Some researchers have defended that, if language developed from a propositional and holistic animal communication signal, there must have been a period when humans successfully extracted pieces from certain expressions that could be reused in other expressions (Tallerman, 2007), gradually forming compositionality. Another possible scenario is that words developed through vocal imitation of natural sounds. Word learning, according to Hauser et al. (2002), necessitates the ability of vocal imitation, which has been observed in nonhuman species such as songbirds (p. 1574). Recent research found that generations of vocal imitation of some natural sounds can lead to word-like outputs with category labels, implying that vocal imitation may be a major factor in the emergence of first words (Edmiston et al., 2018). Returning to our hypothesis that CP prepared the basis for discreteness, it has been shown that the discreteness of sounds and the discreteness of words are inextricably intertwined. When it comes to CP of speech sounds, or how to discriminate and categorize speech sounds in a given language, it appears that sequential statistics manifested in words (Transitional Probability reference), rather than acoustic features of the sounds, drive discrimination and categorization of the speech sounds (Saffran et al., 1996). Furthermore, data in infants suggest that perceptual statistical learning also plays a key role in word segmentation from speech streams (Romberg & Saffran, 2010). In this sense, the key of discreteness offered at the phoneme level appears to be influenced by how sound sequences in words are organized (Bidelman & Lee, 2015). This shows that discreteness may have been founded on categorical perception.

4. Categorical perception and Discreteness: Neurocognitive considerations

Neurobiological studies on CP and discreteness of language also support our proposal. Using birds as an example, HVCx (HVC is a letter-based term) cells in swamp sparrows have been demonstrated to respond robustly to auditory

categorical changes in note duration (Prather et al., 2009). In birds, HVC projects to many brain regions, with HVCx being the projection to AreaX (striatal area x) among others. HVC is a premotor nucleus that has been proposed to be analogous to Broca's area in humans and serve sensory-motor functions (Prather et al., 2017). As a result, HVCx could be analogous to the premotor-striatal connection, which has been linked to beat perception in humans (Grahn & Rowe, 2009). Furthermore, the auditory nuclei field L projecting to CLM (caudolateral mesopallium) and NCM (the caudal part of the medial nidopallium), both projecting to CMM (the caudal part of the medial mesopallium), present an analogous hierarchy to humans in which physical information is processed at the lower level while abstract concepts are encoded at the higher level in European Starlings (Jeanne et al., 2011). It is worth noting that NCM and CMM, which are similar to human auditory cortex, have also been documented for auditory memory (Bolhuis & Gahr, 2006). Moreover, Lampen et al. (2017) discovered that NCM and CMM in zebra finches are more responsive to rhythmic than arrhythmic songs, indicating that NCM and CMM are involved in auditory detection and discrimination.

The auditory cortex is also shown to be involved in CP of speech sounds in primates. Spiking activity from the superior temporal gyrus (STG) in rhesus monkeys were recorded (Tsunada et al., 2011). In humans, however, CP of sounds was mediated by not only the STG and superior temporal sulcus (STS) (Harinen & Rinne, 2013), but also the premotor and primary motor cortex (Chevillet et al., 2013) comparable to those found in the NCM and CMM of birds. Furthermore, Hickok & Peoppel (2004) proposed a dual-stream theory, with a ventral stream dealing with phoneme and lexical recognition and lexical combinations, and a dorsal stream dealing with sensorimotor transformation in speech output. The dorsal stream connects the STG with the premotor cortex, both regions were reported to be where CP takes place. If CP is the basis for discreteness, then such dorsal connection could have played a role in discreteness. The dorsal pathway has been shown to be necessary for vocal imitation, which has been shown to be one of the key factors for word emergence (Edmiston et al., 2018). In this sense, discreteness seems to rely on CP in both phonological level, namely phonemes, and syntactic level, namely words.

Beyond cortex, recent evidence has shown that the basal ganglia also relate to cognitive functions beyond motor domain which includes perceptual categorization (e.g., Ashby et al., 1998; Hochstenbach et al., 1998; Seger, 2008, for reviews). Though basal ganglia-mediating category learning and speech perception and learning have been largely studied independently, as Lim et al. (2014) reviewed, these separate lines of research share commonalities and there

is great potential in bridging efforts to understand speech perception and learning with general cognitive neuroscience approaches and neurobiological models of Learning. Due to the fact that speech perception required the integration of acoustic cues across various dimensions, explicit attempts to the properties of speech are difficult. It has been shown that explicit/directed attention to acoustic features are engaged in the process of directed categorization training (Logan et al., 1991). Further, in an fMRI study, Tricomi et al. (2006) showed that category training of non-native speech categories engages the basal ganglia (i.e., the striatum). These results suggest that the basal ganglia learning system are involved in promoting adult speech category learning, which is the nurture part of CP. This is also in parallel with the finding in swamp sparrow mentioned above that Area X is also involved in CP of note duration.

5. Conclusion

In this paper, we come up with the hypothesis that categorical perception (CP) could have laid the foundation for discreteness, one of the design features of language. By reviewing comparative studies on CP in nonhuman animals and humans, we found that CP has a phylogenetic root dating back to invertebrates which is closely related to reproduction and survival, and seems to be a combination of innateness and experience. In addition, by reviewing neurobiological studies, we show that tasks of CP activate cortical and subcortical areas including auditory and (pre)motor cortex as well as the basal ganglia, the connection between which could be insightful for locating domain general discreteness. The current work provides additional evidence for the important role of sensory-motor system in language evolution.

References

Ashby, F. G., Alfonso-Reese, L. A., & Waldron, E. M. (1998). A neuropsychological theory of multiple systems in category learning. *Psychological review,* 105(3), 442.

Baugh, A. T., Akre, K. L., & Ryan, M. J. (2008). Categorical perception of a natural, multivariate signal: Mating call recognition in túngara frogs. *Proceedings of the National Academy of Sciences of the United States of America,* 105(26), 8985‑8988.

Bogaerts, L., Frost, R., & Christiansen, M. H. (2020). Integrating statistical learning into cognitive science. *Journal of Memory and Language,* 115(August), 104167.

Bolhuis, J. J., & Gahr, M. (2006). Neural mechanisms of birdsong memory. *Nature Reviews Neuroscience*, 7(5), 347‑357.

Bidelman, G. M., & Lee, C. C. (2015). Effects of language experience and stimulus context on the neural organization and categorical perception of speech. *NeuroImage*, 120, 191‑200.

Chevillet, M. A., Jiang, X., Rauschecker, J. P., & Riesenhuber, M. (2013). Automatic phoneme category selectivity in the dorsal auditory stream. *Annals of Internal Medicine*, 158(6), 5208‑5215.

Dooling, R. J., & Brown, S. D. (1990). Speech perception by budgerigars (Melopsittacus undulatus): Spoken vowels. *Perception & Psychophysics*, 47(6), 568‑574.

Dooling, R. J., Brown, S. D., Park, T. J., Okanoya, K., & Soli, S. D. (1987). Perceptual organization of acoustic stimuli by budgerigars (Melopsittacus undulatus): I. Pure tones. *Journal of Comparative Psychology*, 101(2), 139‑149.

Edmiston, P., Perlman, M., & Lupyan, G. (2018). Repeated imitation makes human vocalizations more word-like. *Proceedings of the Royal Society B: Biological Sciences,* 285(1874).

Goldstone, R. L., & Hendrickson, A. T. (2010). Categorical perception. *Wiley Interdisciplinary Reviews: Cognitive Science,* 1(1), 69-78.

Grahn, J. A., & Rowe, J. B. (2009). Feeling the beat: premotor and striatal interactions in musicians and nonmusicians during beat perception. *The Journal of Neuroscience*, 29(23), 7540‑7548.

Harinen, K., & Rinne, T. (2013). Activations of human auditory cortex to phonemic and nonphonemic vowels during discrimination and memory tasks. *NeuroImage,* 77, 279‑287.

Harnad, S. (1987). Psychophysical and cognitive aspects of categorical perception: A critical overview. In Harnad, S. (ed.) *Categorical Perception: The Groundwork of Cognition.* New York: Cambridge University Press.

Hauser, M. D., Chomsky, N., & Fitch, W. T. (2002). The faculty of language: what is it, who has it, and how did it evolve?. *Science,* 298(5598), 1569-1579.

Hockett, C. F. (1960). The origin of speech. *Scientific American*, 203(3), 89‑96.

Hickok, G., & Poeppel, D. (2004). Dorsal and ventral streams: a framework for understanding aspects of the functional anatomy of language. *Cognition*, 92(1-2), 67-99.

Jeanne, J. M., Thompson, J. V., Sharpee, T. O., & Gentner, T. Q. (2011). Emergence of learned categorical representations within an auditory forebrain circuit. *Journal of Neuroscience*, 31(7), 2595‑2606.

Kuhl, P. K., Stevens, E., Hayashi, A., Deguchi, T., Kiritani, S., & Iverson, P. (2006). Infants show a facilitation effect for native language phonetic perception between 6 and 12 months. *Developmental Science,* 9(2), F13-F21.

Lampen, J., McAuley, J. D., Chang, S.-E., & Wade, J. (2017). Neural activity associated with rhythmicity of song in juvenile male and female zebra finches. *Behavioural Processes,* 163, 45-52.

Liberman, A. M. (1957). Some results of research on speech perception. *The Journal of the Acoustical Society of America,* 29(1), 117-123.

Lim, S. J., Fiez, J. A., & Holt, L. L. (2014). How may the basal ganglia contribute to auditory categorization and speech perception?. *Frontiers in neuroscience*, 8, 230.

Logan, J. S., Lively, S. E., & Pisoni, D. B. (1991). Training Japanese listeners to identify English/r/and/l: A first report. *The Journal of the Acoustical Society of America,* 89(2), 874-886.

Maye, J., Werker, J. F., & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition*, 82(3), B101-B111.

Prather, J. F., Nowicki, S., Anderson, R. C., Peters, S., & Mooney, R. (2009). Neural correlates of categorical perception in learned vocal communication. *Nature Neuroscience*, 12(2), 221‑228.

Prather, J. F., Okanoya, K., & Bolhuis, J. J. (2017). Brains for birds and babies: Neural parallels between birdsong and speech acquisition. *Neuroscience and Biobehavioral Reviews*, 81, 225‑237.

Hochstenbach, J., van Spaendonck, K. P., Cools, A. R., Horstink, M. W., & Mulder, T. (1998). Cognitive deficits following stroke in the basal ganglia. *Clinical Rehabilitation*, 12(6), 514-520.

Jarvis, E. D. (2019). Evolution of vocal learning and spoken language. *Science*, 366(6461), 50‑54.

Romberg, A. R., & Saffran, J. R. (2010). Statistical learning and language acquisition. *Wiley Interdisciplinary Reviews: Cognitive Science*, 1(6), 906‑914.

Saffran, J. R., Newport, E. L., & Aslin, R. N. (1996). Word segmentation: The role of distributional cues. *Journal of Memory and Language*, 35(4), 606‑621.

Searle, J. R. (2008). What is language: Some preliminary remarks. *Explorations in Pragmatics: Linguistic, Cognitive and Intercultural Aspects*, 7‑37.

Seger, C. A. (2008). How do the basal ganglia contribute to categorization? Their roles in generalization, response selection, and learning via feedback. *Neuroscience & Biobehavioral Reviews*, 32(2), 265-278.

Sinnott, J. M., & Mosteller, K. W. (2001). A comparative assessment of speech sound discrimination in the Mongolian gerbil. *The Journal of the Acoustical Society of America*, 110(4), 1729‑1732.

Tallerman, M. (2007). Did our ancestors speak a holistic protolanguage? *Lingua*, 117(3), 579‑604.

Tricomi, E., Delgado, M. R., McCandliss, B. D., McClelland, J. L., & Fiez, J. A. (2006). Performance feedback drives caudate activation in a phonological learning task. *Journal of cognitive neuroscience,* 18(6), 1029-1043.

Tsunada, J., Lee, J. H., & Cohen, Y. E. (2011). Representation of speech categories in the primate auditory cortex. *Journal of Neurophysiology,* 105(6), 2634‑2646.

Wyttenbach, R. A., May, M. L., & Hoy, R. R. (1996). Categorical perception of sound frequency by crickets. *Science*, 273(5281), 1542‑1544.

Zhang, Q., Lei, L., & Gong, T. (2021). Categorical Perception as a Combination of Nature and Nurture. Proceedings of the Annual Meeting of the Cognitive Science Society, 43.