A data-driven study of auditory iconicity

Andrea Gregor de Varda & Marco Marelli

A form of non-arbitrariness in language is iconicity, i.e. a consistent relationship between linguistic sounds and their referents that is defined not only by convention, but also by the sounds' and the objects' intrinsic qualities. While various research efforts have investigated vision-related linguistic iconicity (see for instance Kohler, 1929, 1947; Maurer et al., 2006; Ramachandran & Hubbard, 2001; Sapir, 1929), less attention has been paid to the other sensory modalities (although see Fontana, 2013; Gallace et al., 2011; Joo, 2020; Speed et al., 2021). This vision-centric approach is to some extent motivated by the perceptual predominance of the visual modality in humans (Lynott et al., 2020). However, some properties of auditory iconicity make it an interesting testbed for the study of non-abritrariness. In contrast with the other senses, auditory iconicity in oral languages takes place *within* a perceptual modality. Additionally, words with high auditory perceptual strength are considered the most iconic, as documented by both explicit iconicity ratings (Winter et al., 2017) and guessing studies (Dingemanse et al., 2016). Another piece of evidence for the peculiar status of auditory iconic words comes from linguistic typology, where it has been shown that auditory terms constitute the most prominent class of perceptual terms in the ideophonic lexicon across languages (McLean, 2021).

Given the increasing recognition of the role played by auditory iconicity in vocabulary structure, it is crucial to develop a valid measurement for the construct under invetigation. It is common practice to operationalize iconicity through subjective ratings; however, iconicity ratings have been criticized for having low construct validity. Thompson et al. (2020) have suggested that the participants' responses in rating studies might be based on semantics alone – and in particular on perceptual strength –, which would call for a different, and possibly objective measure of phonosemantic transparency. Additionally, ability of language users to assess whether a sound is iconic has been questioned, as it has been shown that native speakers have a positive bias when judging whether sounds fit their reference in their native language (Sutherland & Cimpian, 2015). Winter & Perlman (2021) provided a response to those criticisms, and rightfully noted that iconicity ratings have served an important purpose in explaining the distribution of iconic properties in the lexicon; however, they recognize that iconicity ratings should be complemented by other tools in order to grant a more comprehensive picture on the role of iconicity in language. In this presentation, we propose a data-driven alternative to iconicity ratings, where we operationalize auditory iconicity as the objective similarity between (i) the sound of a word and (ii) the natural sounds associated with its referent. Both spoken words and natural sounds are embedded in a shared vector format in one of three ways:

- Short-time Fourier transform (STFT), a mathematical procedure that transposes sound waves into individual frequencies and their amplitudes;
- Sound classification network, a neural network trained to label sounds;
- Speech recognition network, a neural network trained to recognize spoken words.

For all these three methods, the obtained sound representations embed natural sounds and word sounds into a shared vector space. This allows to estimate the objective similarity between word sounds and natural sounds, which we employ as a measure of auditory iconicity. We use these metrics to assess the pervasiveness of iconicity in the English auditory lexicon, and empirically validate them in explaining human intuitions, against a strong baseline of psycholinguistic variables associated with the construct under scrutiny. Our results show that word sounds are significantly more similar to

the natural sounds of their referents than to other unrelated natural sounds, demonstrating that imitation can be considered as a widespread mechanism underlying the structure of the English auditory vocabulary. Auditory mimicry thus plays a significant role in the phonological structure of the lexicon, and is not limited to onomatopoetic words. Furthermore, our computational measures are predictive of human judgements on form-meaning resemblance. This result challenges the criticism put forward by Thompson et al. (2020), who proposed that iconicity ratings only reflect subjective intuitions on perceptual strength: our data-driven measurements are significant predictors of human ratings in four out of six conditions even when perceptual strength is included as a covariate in the model. At the same time, our study demonstrates that data-driven alternatives to human judgments do exist, and can be employed to study subtle and elusive phenomena such as iconicity.

References

Dingemanse, M., Schuerman, W., Reinisch, E., Tufvesson, S., & Mitterer, H. (2016). What sound symbolism can and cannot do: Testing the iconicity of ideophones from five languages. *Language, 92*(2)

Fontana, F. (2013). Association of haptic trajectories to takete and maluma. In *International workshop on haptic and audio interaction design* (pp. 60–68).

Gallace, A., Boschin, E., & Spence, C. (2011). On the taste of "bouba" and "kiki": An exploration of word–food associations in neurologically normal participants. *Cognitive neuroscience, 2*, 34-46.

Joo, I. (2020). Phonosemantic biases found in leipzig-jakarta lists of 66 languages. *Linguistic Typology, 24*, 1-12

Kohler, W. (1929). Gestalt psychology. liveright. Oxford, England.

Kohler, W. (1947). Gestalt psychology: An introduction to new concepts in modern psychology.

Lynott, D., Connell, L., Brysbaert, M., Brand, J., & Carney, J. (2020). The lancaster sensorimotor norms: multidimensional measures of perceptual and action strength for 40,000 english words. *Behavior Research Methods, 52*(3), 1271–1291.

McLean, B. (2021). Revising an implicational hierarchy for the meanings of ideophones, with special reference to japonic. *Linguistic Typology, 25*(3), 507–549.

Ramachandran, V. S., & Hubbard, E. M. (2001). Synaesthesia–a window into perception, thought and language. *Journal of Consciousness Studies, 8*(12), 3–34.

Sapir, E. (1929). A study in phonetic symbolism. *Journal of experimental psychology, 12*(3), 225.

Speed, L. J., Atkinson, H., Wnuk, E., & Majid, A. (2021). The sound of smell: Associating odor valence with disgust sounds. *Cognitive Science, 45*(5).

Sutherland, S. L., & Cimpian, A. (2015). An explanatory heuristic gives rise to the belief that words are well suited for their referents. *Cognition, 143*, 228–240.

Thompson, A. L., Akita, K., & Do, Y. (2020). Iconicity ratings across the japanese lexicon: A comparative study with english. *Linguistics Vanguard, 6*(1)

Winter, B., & Perlman, M. (2021). Iconicity ratings really do measure iconicity, and they open a new window onto the nature of language. *Linguistics Vanguard, 7*(1).

Winter, B., Perlman, M., Perry, L. K., & Lupyan, G. (2017). Which words are most iconic?: Iconicity in english sensory words. *Interaction Studies, 18*(3), 443–464.