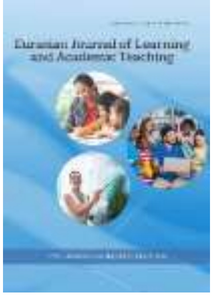


Linguistic aspects in the representation of kinematic

.

Kenjayeva Zemfira Alimjanovna
Senior teacher of UZSHWLU

ABSTRACT

Silent gestures consist of complex multi-articulatory movements but are now primarily studied through categorical coding of the referential gesture content. The relation of categorical linguistic content with continuous kinematics is therefore poorly understood. Here, we reanalyzed the video data from a gestural evolution experiment (Motamedi, Schouwstra, Smith, Culbertson, & Kirby, 2019), which showed increases in the systematicity of gesture content over time. We applied computer vision techniques to quantify the kinematics of the original data. Our kinematic analyses demonstrated that gestures become more efficient and less complex in their kinematics over generations of learners. We further detect the systematicity of gesture form on the level of the gesture kinematic interrelations, which directly scales with the systematicity obtained on semantic coding of the gestures. Thus, from continuous kinematics alone, we can tap into linguistic aspects that were previously only approachable through categorical coding of meaning. Finally, going beyond issues of systematicity, we show how unique gesture kinematic dialects emerged over generations as isolated chains of participants gradually diverged over iterations from other chains. We, thereby, conclude that gestures can come to embody the linguistic system at the level of interrelationships between communicative tokens, which should calibrate our theories about form and linguistic content

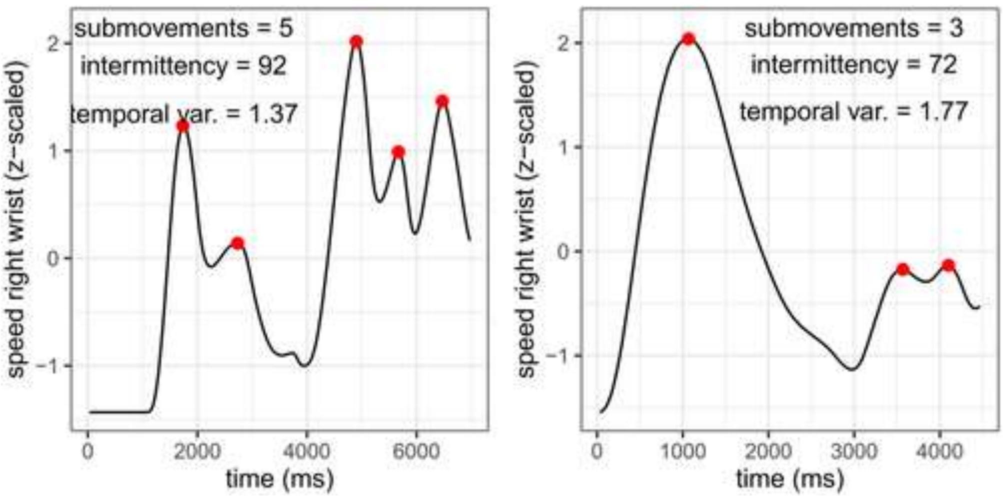
Keywords: Motion tracking, Kinematic properties, Gesture salience, Gesture segmentation, Kinematic entropy, Gesture kinematic culture

All known natural languages combine discrete categorical elements with continuous and dynamic properties (Bolinger, 1968). For a long time, the study of human communicative behavior has focused on aspects that best yield to analysis in terms of discrete categories such as lexical items, phonological building blocks, semantic categories, and their combinatorial properties. At the same time, language use is widely acknowledged to also feature more gradient and continuous streams of behavior that do not always easily yield to an analysis in terms of discrete symbol systems (Enfield, 2009; Kendon, 2004). Here, we investigate whether kinematic measures directly derived from continuous manual movements can capture the meaning space they are designed to communicate. Using communicative silent gestures as a test case, we show how continuous movements can be studied and are patterned as evolving dynamic systems.

Motion tracking was performed on each video recording with a sampling rate of 30 Hz. To extract movement traces, we used OpenPose (Cao et al., 2017), which is a pre-trained deep neural network approach for estimating human poses from video data (for a tutorial, see Pouw & Trujillo, 2019). We selected key points that were most likely to cover the gross variability in gestural utterances: positional x (horizontal) and y (vertical) movement traces belonging to left and right index fingers and wrists, as well as the nose. For all position traces and their derivatives, we applied a first-order 30 Hz low-pass Butterworth filter to smooth out high-frequency jitters having to do with sampling noise. We z-normalized and mean-centered position traces for each video to ensure that differences between subjects (e.g., body size) and within-subject differences in camera position at the start of the recording were inconsequential for our measurements.

Kinematic properties

We first selected five potential measures representative of the kinematic quality of the movements in terms of segmentation, salience, and temporality, namely, submovements, intermittency, gesture space, rhythm, and temporal variability (or rhythmicity). See Fig. 3 for two example time series from which most measures can be computed. All measures were computed for each key point's time series separately and then averaged so as to get an overall score for the multimodal utterance as a whole. Based on these exploratory measures, we eventually selected three measures tracking gesture salience (*gesture space*), gesture segmentation (*intermittency score*), and gesture temporality (*temporal variability*). We discuss the motivations for selecting each measure below. Correlations between these variables and distributions are shown in supplementary materials, Fig. S1.



Gesture salience

As a measure for gesture salience or reduction, we computed a gesture space measure. This was determined by extracting the maximum vertical amplitude of a key point multiplied by the maximum horizontal amplitude, that is, the area in pixels that has been maximally covered by the movement.

Gesture segmentation

We first computed a submovement measurement similarly implemented by Trujillo, Vaitonyte, Simanova, and Özyürek (2019). Submovements are computed with a basic peak finding function which identifies and counts maxima peaks in the movement speed time series. We set the minimum inter-peak distance at eight frames, and minimum height = −1 (z-scaled; 1 std.), minimum rise = 0.1 (z-scaled).

Kinematic entropy

Entropy is a measure that quantifies the compressibility of data structures and has been used to gauge the combinatorial structure of communicative tokens in the field of language evolution (e.g., Verhoef et al., 2016; for theoretical grounding, see Gibson et al., 2019). In the original experiment, Motamedi et al. (2019) computed entropy from the gesture content codings, which captured recurrent information units between gestures. In our case, entropy quantifies the degree to which there are similar or more diverse edge lengths (i.e., similar/diverse levels of dissimilarity "D" between combinations of two gesture trajectories). If they are more similar, lower entropy reflects that communicative tokens relate to each other in more structural ways. So our measure of network entropy gauges how compressible kinematic interrelationships are, which is conceptually related to the systematic recurrence of information units between the human judged gesture content.

The network entropy measure we used (see Eagle, Macy, & Claxton, 2010) is almost identical to a classic Shannon entropy calculation used in the original study to quantify the systematicity of the gesture's content (Motamedi et al., 2019), where $Entropy\ H(X) = -\sum p(X)\log p(X)$.

The only difference is that our measure is computed on the distances for each node relative to the shortest path to the other nodes and then normalized by the number of gesture distances. So our measure quantifies the topological diversity of the gesture relationships, where a lower score indicates more similar relationships and a higher score indicates a more randomly distributed set of relationships. Specifically, for each gesture node, we compute the diversity of kinematic distances to other gestures, using a scaled Shannon entropy measure

Gesture kinematic culture. To assess whether there is a kinematic culture emerging such that gestures in a specific chain are over the generations becoming more similar in kinematics as compared to gestures from another chain, we leverage cluster performance measures. For each generation, we assess whether gestures from a particular chain are also likely to cluster in a super-ordinate kinematic space that includes all gestures performed across generations (i.e., gestures produced in chains 1 through 5). Clustering can be quantified in several ways. In our analysis, we report on two well-known cluster performance measures: Dunn index and Silhouette width (Yadav, Tomar, & Agarwal, 2013). In general, cluster performance measures relate within-cluster distances between nodes (minimal when clusters are stable) to between-cluster distances (maximal when clusters are stable), though they vary in how they compute the within and between distances. The Dunn index quantifies the compactness of the clusters assigned (chains in our case) and relates the minimum distance between centroids of each cluster to maximal distance between points, where higher values indicate better clustering. However, this Dunn index measure only yields five data points in our case, one for each chain, which makes it hard to perform a statistical test. Therefore, we will also compute a token level measure of Silhouette width, which, for each token, relates the mean distance to other tokens within its cluster to the minimum distance between a member of a neighboring cluster.

Conclusion

Human communicative behavior tends to combine categorical elements and continuous properties, but for technological as well as theoretical reasons, the categorical elements of evolving linguistic systems have long received more attention than their continuous aspects. Here, we have contributed to the study of multimodal language and cognition by considering the gesture kinematics of evolving gestural systems. We have used computer vision techniques to analyze the kinematic properties of evolving gestural systems, showing that over generations of learners, the dynamics of head and upper limb movements become simpler, increase in systematicity, and give rise to kinematic dialects. Our kinematic measures help characterize fine-grained levels of linguistic organization that remain out of reach of content-based discretized coding approaches, providing novel insights that corroborate and complement prior approaches. Our findings provide an unprecedented view of how gestures become structured and increasingly language-like as they evolve, in ways that are directly related to the coordination and simplification of bodily movements. While considerations of communicative efficiency and systematicity have so far been mostly based on analyses of discrete symbol systems like written words and text corpora, our work shows how hallmark features of linguistic systems may be grounded directly in the biomechanical properties of dynamically evolving systems of continuous signals.

References:

1. Beecks, C., Hassani, M., Brenger, B., Hinnell, J., Schüller, D., Mittelberg, I., & Seidl, T. (2016). Efficient query processing in 3D motion capture gesture databases. *International Journal of Semantic Computing*, 10(01), 5– 25. https://doi.org/10.1142/S1793351X16400018
2. Beecks, C., Hassani, M., Hinnell, J., Schüller, D., Brenger, B., Mittelberg, I., & Seidl, T. (2015). Spatiotemporal similarity search in 3D motion capture gesture streams. In C. Claramunt, M. Schneider, R. C. -W. Wong, L. Xiong, W. -K. Loh, C. Shahabi & K. -J. Li (Eds.), *Advances in spatial and temporal databases* (pp. 355– 372). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-22363-6_19
3. Bernstein, N. (1967). *The co-ordination and regulations of movements* (1st English ed.). Oxford, England: Pergamon Press.
4. Bolinger, D. (1968). *Aspects of language*. New York: Harcourt, Brace, and World.Google Scholar
5. Börstell, C., & Lepic, R. (2020). Spatial metaphors in antonym pairs across sign languages. Sign Language & Linguistics, 23(1-2), 112– 141. https://doi.org/10.1075/sll.00046.borGoogle Scholar
6. Bosker, H. R., & Peeters, D. (2020). Beat gestures influence which speech sounds you hear. *Proceedings of the Royal Society B: Biological Sciences*, 288(1946), 20202419. http://doi.org/10.1098/rspb.2020.2419Google Scholar