Multiscale kinematic analysis reveals structural properties of change in evolving manual languages in the lab

Wim Pouw1, Mark Dingemanse1,3, Yasamin Motamedi4, & Asli Ozyurek1,2,3

1 Donders Institute for Brain, Cognition and Behaviour, Radboud University Nijmegen

2 Institute for Psycholinguistics, Max Planck Nijmegen

3 Center for Language Studies, Radboud University Nijmegen

4 Language and Cognition Lab, University College London

Author note

This work is supported by a Donders Fellowship awarded to Wim Pouw and Asli Ozyurek and is supported by the Language in Interaction consortium project ‘Communicative Alignment in Brain & Behavior’ (CABB).

Correspondence concerning this article should be addressed to Wim Pouw, Montessorilaan 3, 6525 HR Nijmegen. E-mail: [w.pouw@psych.ru.nl](mailto:w.pouw@psych.ru.nl)

Abstract

In the current approach we perform a multiscale quantification of kinematic changes of an evolving silent gesture system. Silent gestures consist of complex multi-articulatory movement that have so far proven elusive to quantify in a structural and reproducable way, and is primarily studied through human coders meticulously interpreting the referential content of gestures. Here we demonstrate that the signal’s objective form (kinematics of gesture) can however also be investigated for the emergence of linguistic constraints. Here we reanalyzed the video data from a gesture evolution experiment (Motamedi et al. 2019), which originally showed increases in systematicity in gesture’s content over time. We applied a signal-based approach, utilizing computer vision techniques to quantify kinematics. Then we performed a kinematic analysis, showing that over generations of language users, silent gestures became more efficient and less complex in their kinematics. We further detect systematicity of the gesture form on the level of their interrelations, which proved itself as a proxy of systematicity obtained via human observation data, and was related to the kinematic changes observed on the token level. Finally we demonstrate that gesture kinematic cultures emerge over generations, as chains of participants gradually diverged over iterations from other chains. We thereby show how gestures can come to embody the linguistic system at the level of interrelationships between gesture forms, paving a way towards a reproducible analysis of linguistic constraints in continuous multidimensional signals.

*Keywords:* language evolution, silent gesture, kinematics, systematicity

*Word count:* X

Multiscale kinematic analysis reveals structural properties of change in evolving manual languages in the lab

How do language scientists judge a communicative signal to be linguistic in nature? What properties must a communicative signal have? It can be contended that no *particular* property in a communicative signal is revealing about its systematic function. This is where the notion of arbitrariness comes from: when there is an arbitrary mapping between a signal’s form and its meaning, we cannot attune to the meaning of the signal in reference to its form alone. We can only attune to such a meaning when knowledgeable about how a signal’s form functions in relation to other signaling tendencies in relevant contexts.

Arbitrariness in our view is a relational concept and has many forms. Most communicative signals have forms that *seem* arbitrary to their meaning if we are ignorant about how they relate to physical aspects of the signaling modality (e.g., emotional stress and vocal fold tensioning), or when we are ignorant about how they relate to common action routines and event experiences as in the case of iconic reference. Similarly, some signals seem arbitrary when not properly related to linguistic constraints (e.g., syntactic rules of word class order). Importantly, arbitrariness is not something that is fixed, it can be *resolved* by relating a signals form to relevant aspects of the systems that constrain them. A signals form can be constrained by systems that operate on different timescales. Vocal fold tensioning due to emotional stress can be said to operate more directly on signaling, but we cannot say the same for linguistic constraints. In the linguistic case, the form of a communicative signal is constrained by a culturally shaped system operating on a timescale that transcends the level at which those constraints are exerted (Pattee & Rączaszek-Leonardi, 2012; Rączaszek-Leonardi & Scott Kelso, 2008).

Importantly, in this view, arbitrariness only means we have not properly related a signal’s form with the systems they are constrained by, and the timescales such systems operate on. Arbitrariness can lead us astray in this way by leading to the conclusion that a signals form *is* arbitrary and not revealing of linguistic competences. A paradigmatic example of this is the current mainstream to study manual gesture communication. Manual gestures are seldomly semiotically studied based on the signal’s measurable form, namely manual and whole-body postures in movement (i.e., the kinematic level). Instead gestures are studied as already categorizable expressions by researchers inferring meaning from their form. As such the kinematics are in one sense reduced (from continuous to discrete tokens) and another sense enriched (from movement to message) with meanings that are *projected by human coders*. This is indeed the only viable way if we agree that kinematics *are* arbitrary, as the very movement as such is not what is important, but what the movement signifies. This seems supported by the very fact that a gesture may mean something completely different while being similar in kinematics.

In the current approach we will assess whether there is a way to study communicative body movements revealing of the linguistic system they are possibly constrained by. Communicative body movements, or silent gestures, are an exemplary test case for the current view that linguistic constraints are relationally present in the signal’s form. This is because, as mentioned, a semiotic study of gesture kinematics is underdeveloped (but see Lepic, Börstell, Belsitzman, & Sandler, 2016; Östling, Börstell, & Courtaux, 2018), and relatedly, because gesture kinematics are highly complex signals, involving multiple body parts and continuous trajectories that are generally held to be difficult to parse into discrete meanings (see e.g., Sandler, 2018).

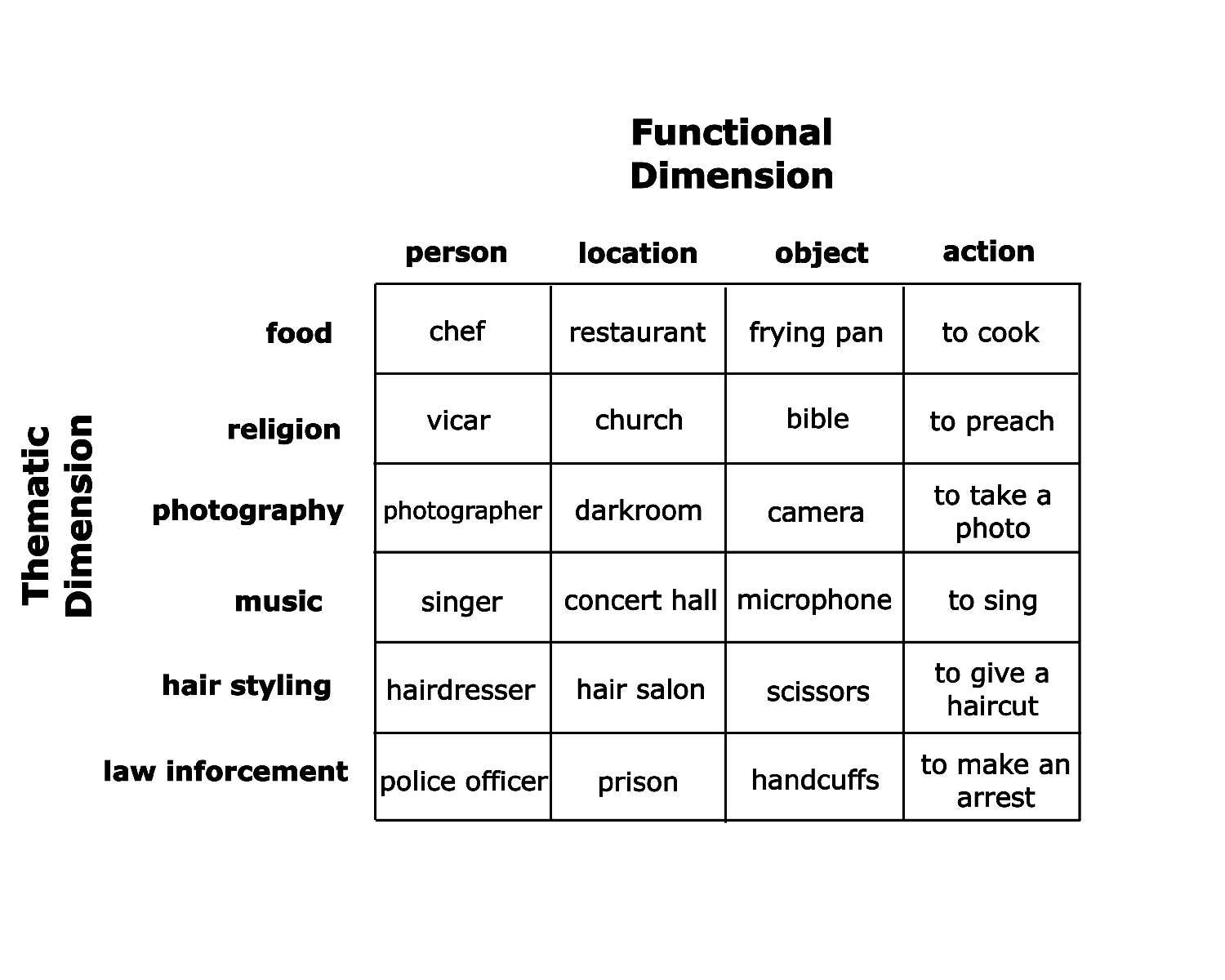
**Context of the present study**  
There is an interdisciplinary scientific effort to unveil the historical and/or necessary constraints that allow(ed) for the emergence of human language (e.g., Bickerton, 2009; Deutscher, 2005; Fitch, 2010, 2010; McNeilage, 2008; Tomasello, 2008). An important approach within this enterprise comes from experimental cognitive science (Scott-phillips & Kirby, 2010), where interactive communication processes likely to have contributed to language emergence are simulated in the lab (e.g., Kirby, Cornish, & Smith, 2008).

In this iterated learning approach to language evolution, agents are tasked to learn a novel set of signals, iteratively transmitted to later generations and/or used in communication by later generations (iterated learning + communication). Over many cycles of learning and use, the signals are affected by various transmission biases (e.g., Christiansen & Chater, 2016; Enfield, 2016). Processes of iterated learning and communication can simulate how structural properties such as systematicitity, learnability, and compositionality evolve from more simpler communication systems — a process that must have occurred in human language evolution too (Bickerton, 2009). In such simulations communicative tokens undergo cultural evolution constrained by population dynamic properties such as historicity (the system is constrained by past contingencies) and adaptivity (the system is able to tweak itself in service of its informative goals). Such population dynamics must have played out over long temporal and vast population scales, but through these iterated learning paradigms such processes are to some limited degree brought under experimental control. These evolving or emerging communicative systems can be constituted by a variety of different signal media, from simple discrete symbol sequences to more challenging continuous signals (Cornish, Dale, Kirby, & Christiansen, 2017; Ravignani, Delgado, & Kirby, 2016; Verhoef, Kirby, & de Boer, 2016).

As mentioned communicative body movements are challenging in this way, but insights from meticulous study by (psycho-)linguists have revealed important insights on how they evolve. For example, in dyadic social interaction, repeated gestural referrals to an object or a picture will lead to those gestures becoming more reduced in size (Gerwing & Bavelas, 2004; Namboodiripad, Lenzen, Lepic, & Verhoef, 2016). This is comparable to research in ‘pictionary’ paradigms where a concept is drawn out and to be interpreted by another player. After repeated trials of drawing, a reduction of the drawings’ complexity is observed, with smaller-sized and less iconic drawings as a result, while communicative accuracy increases over time (Fay, Garrod, Roberts, & Swoboda, 2010; Garrod, Fay, Lee, Oberlander, & Macleod, 2007).

Though, drifts from less or more iconicity are not fixed processes. When interaction between people is opened up, a whole new suit of social affordances arise. For example, while gestures may reduce in size and iconicity when some common ground is established, at any moment an interlocutor may request a clarification, soliciting large and iconic gestures per implicitly requested (Bavelas, Gerwing, Sutton, & Prevost, 2008; Holler & Wilkin, 2011). In such moments of interactional repair, common ground is calibrated and re-established. Indeed, establishing common ground is not something that is a linear phenomenon, but an interactive and dialectic process. These and many other interactive affordances turn out to be of central importance for smooth everyday language use (Dingemanse, Roberts, et al., 2015), and according to cultural evolutionary accounts of language, such local-scale processes of interaction and transmission between communicators are crucial for the emergence of any linguistic system (e.g., Enfield, 2016; Kirby & Christiansen, 2003; Kirby, Griffiths, & Smith, 2014; Raviv, Meyer, & Lev-Ari, 2019). A key question that drives cultural evolution research is which particular interactive constraints produce pressures for a certain communication system to adapt in one way or another, and how effective solutions are negotiated at the possible expense of other communicatively efficient solutions (Dingemanse et al., 2015).

In a recent iterated learning study with silent gesture (Motamedi, Schouwstra, Smith, Culbertson, & Kirby, 2019) two such possible constraints, transmission across generations and communication within generations, were studied simultaneously as well as separately. Learning occurred with a set of silent gesture-concept mappings (i.e., communicative tokens) communicated through five iterations of vertical transmissions, where gestures were transmitted from one participant to-be-reproduced by the next participant. Or, tokens would be communicated through five horizontal interactions in a director-matcher type task. These constraints - interaction and transmission - were first studied in combination in experiment 1, which is the focus for the current paper. An important aspect of the study was that every concept was characterisable along two dimensions: theme (e.g., food, religion) and function (e.g., person, location) (see Figure 1).

Figure 1. Concepts to be conveyed in gesture in Motamedi et al. 2019  


These dimensions provided possible axes for compressibility of the communicative tokens. After all, by combining 10 unique gestures one can pick out any referent (e.g., “to make an arrest”) from the 24 token meaning space, one gesture marking the functional category (e.g., “action”) and another gesture for the theme category (“justice”). To exemplify further, once confronted with communicating 24 meanings one can invent 24 unique gesture utterances, such as in the following videos for “to sing” (<https://osf.io/d8srx/>) and “singer” (<https://osf.io/974ke/>), which is challenging to do since they are both very much related. However, one can also start differentiating by functional category such that “microphone” is preceded by a general object marking gesture (“<https://osf.io/r3gcp/>”) and “singer” is preceded by a general person marking gesture (“<https://osf.io/ex4tv/>”), and then followed by the same thematic marking gesture conveying “music”. Not only do these general functional markers aid the disambiguation of related meanings, this expressive invention also allows for a systematic reemployment of functional markings for the whole meaning space through compositionality. Once you invent 4 functional marker gestures, and 6 thematic marker gestures, you can systematically recombine these to convey 24 meanings. The communicative system then has compressed its information density from 24 information units to 10 information units.

Motamedi and colleagues (2019) indeed qualitatively observed such signs of compression of the meaning space as the system developed. In early iterations of learning, large-sized iconic enactments were the most common way of gesturally depicting the referents. However, the occurrences of functional markers increased over generations, which represented meaning components reused across gestures. This kind of functional marking mainly targeted the thematic and function categories.

With meticulous hand coding of the different referential components of each silent gesture, it could further be quantitatively tested whether there was indeed systematicity emerging. The gesture coding included information about form of a particular gesture segment, such as the number of manual articulators used (1 or 2 hands), as well as the referential target of the gesture (e.g., hat; pan; turn page). Based on the full sequences of the referential components that were uniquely expressed in each gesture, Shannon entropy was computed, which expresses compressibility of the content of a signal, i.e., the amount of information that is needed to compress the signal. When a lot of referential components in the gesture utterance recur between other gestures that the participant produced, the gesture system has a more compressible structure and indicates systematic reuse of gestural components (e.g., Gibson et al., 2019). Dovetailing with the qualitative observations and other quantitative studies in this field (e.g., Verhoef et al., 2016), it was found that gesture-component entropy decreased over the generations. Furthermore, the gestures were coded for the amount of marking for the functional category, and this showed that such gestures occurred more often at later generations. Finally, average gesture duration - as a measure of communicative efficiency - did not reliably change over the generations, which ran counter to predictions that more mature communication systems tend towards maximal efficiency (Gibson et al., 2019), and that gestures tend to simplify in form when repeatedly used (e.g., Gerwing & Bavelas, 2004).

These results obtained in the lab resonate with findings from homesign (e.g., Haviland, 2013) and emerging sign languages (Senghas, Kita, & Özyürek, 2004). For example, it has been shown that in the expression of motion events first generation signers of Nicaraguan sign language performed more holistic presentations of path and manner, while in following generations manner and path were segmented. Such segmentations affords novel combinatoriality and therefore increases generativity of a language. It expresses the meaning space with fewer means similar to how participants studied by Motamedi et al. (2019) started to compress the meaning space by developing ways mark functional status across referents (e.g., “agent”, “action”).

# Current study

Here we build on data from this recent iterated learning paradigm with silent gestures (Motamedi et al., 2019). With computer vision (Cao, Simon, Wei, & Sheikh, 2017) we obtained motion traces of manual- and head gestures (see e.g., Lepic et al., 2016; Ripperda, Drijvers, & Holler, 2020). We then performed ‘gesture network analysis’ (Pouw & Dixon, 2019), which is a procedure that combines bivariate time series analysis (Dynamic Time Warping) with network analysis and visualization. Through this gesture network analysis we show that the study of gesture’s form can be revealing of linguistic constraints if we study the kinematic system as a whole, as proposed by the gesture network analysis approach.

We hypothesized the following. Over the iterations:

* gesture kinematics simplifies (token level).
* gesture kinematics relationships become more systematic, and this scales with (system level).
* changes in gesture kinematics are related to the systematicity of gesture (token-system level).
* idiosyncratic gesture cultures emerge as evidenced by drifts away from the community as a whole (inter-system level).

Importantly, 1) is based on previous research showing how gestures simplify over repeated use (Gerwing & Bavelas, 2004; Namboodiripad et al., 2016), though at present we know of no quantitative kinematic analysis of the simplification of gesture systems as will be performed here. We will use three kinematic properties that revealing of simplication, gesture size, gesture’s intermittency, and gesture rhythm. Whatever kinematic changes we observe, they are not in and of itself evidence for linguistic constraints. Only, 2) and 3) can be revealing of such constraints. Specifically, we will assess whether a Shannon-based Entropy measure computed on kinematic relationships show that the systems becomes more structured (i.e., compressible). By using a conceptually similar measure as has been applied on human coding (Motamedi et al., 2019), we can then also compare whether kinematics’ entropy is related to the content’s entropy. If so, we have a good evidence that linguistic constraints can be objectively studied from systematic changes in gesture form. If we further find that kinematic changes are related to order emerging on the system level, we show how the simplification of gesture kinematics is a co-constitutive (or an embodiment) of order on the system level. Finally, with 4), we show the true potential of a big data implementation to the gesture network approach, by assessing whether as a communicative gesture evolves within a chain of users, that community of users will drift away from other chains of users. In this way we can show how dialects emerge out of the unique historicity built within chains, leading chains to diverge on the whole. Order on the kinematic system level then, is order of an idiosyncratic kind.

# 5 References

Bavelas, J., Gerwing, J., Sutton, C., & Prevost, D. (2008). Gesturing on the telephone: Independent effects of dialogue and visibility. *Journal of Memory and Language*, *58*(2), 495–520. <https://doi.org/10.1016/j.jml.2007.02.004>

Bernstein, N. (1967). *The Co-ordination and Regulations of Movements* ([1st English ed.] edition). Pergamon Press.

Bickerton, D. (2009). *Adam’s Tongue*. New York: Hill & Wang.

Cao, Z., Simon, T., Wei, S.-E., & Sheikh, Y. (2017). Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields. In (pp. 7291–7299). Presented at the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Retrieved from <http://openaccess.thecvf.com/content_cvpr_2017/html/Cao_Realtime_Multi-Person_2D_CVPR_2017_paper.html>

Christiansen, M. H., & Chater, N. (2016). The Now-or-Never bottleneck: A fundamental constraint on language. *Behavioral and Brain Sciences*, *39*. <https://doi.org/10.1017/S0140525X1500031X>

Cornish, H., Dale, R., Kirby, S., & Christiansen, M. H. (2017). Sequence Memory Constraints Give Rise to Language-Like Structure through Iterated Learning. *PLoS ONE*, *12*(1), e0168532. <https://doi.org/10.1371/journal.pone.0168532>

Csárdi, G. (2019). Package ’igraph’ Network Analysis and Visualization (Version 1.2.4.1). Retrieved from <http://bioconductor.statistik.tu-dortmund.de/cran/web/packages/igraph/igraph.pdf>

Deutscher, G. (2005). *The Unfolding of Language: An Evolutionary of Mankind’s Greatest Invention*. New York: Metropolitan Books.

Dingemanse, M., Blasi, D. E., Lupyan, G., Christiansen, M. H., & Monaghan, P. (2015). Arbitrariness, Iconicity, and Systematicity in Language. *Trends in Cognitive Sciences*, *19*(10), 603–615. <https://doi.org/10.1016/j.tics.2015.07.013>

Dingemanse, M., Roberts, S. G., Baranova, J., Blythe, J., Drew, P., Floyd, S., … Enfield, N. J. (2015). Universal Principles in the Repair of Communication Problems. *PLoS ONE*, *10*(9). <https://doi.org/10.1371/journal.pone.0136100>

Donaldson, J. (2016). Tsne: T-Distributed Stochastic Neighbor Embedding for R (t-SNE) (Version 0.1-3). Retrieved from <https://CRAN.R-project.org/package=tsne>

Eagle, N., Macy, M., & Claxton, R. (2010). Network Diversity and Economic Development. *Science*, *328*(5981), 1029–1031. <https://doi.org/10.1126/science.1186605>

Enfield, N. J. (2016). *Natural causes of language: Frames, biases, and cultural transmission*. Berlin: Language Science Press. Retrieved from <https://langsci-press.org/catalog/book/48>

Fay, N., Garrod, S., Roberts, L., & Swoboda, N. (2010). The interactive evolution of human communication systems. *Cognitive Science*, *34*(3), 351–386. <https://doi.org/10.1111/j.1551-6709.2009.01090.x>

Fitch, W. T. (2010). *The Evolution of Language*. Cambridge University Press. Retrieved from <http://books.google.com?id=RProTk_Ag7gC>

Garrod, S., Fay, N., Lee, J., Oberlander, J., & Macleod, T. (2007). Foundations of representation: Where might graphical symbol systems come from? *Cognitive Science*, *31*(6), 961–987. <https://doi.org/10.1080/03640210701703659>

Gerwing, J., & Bavelas, J. (2004). Linguistic influences on gesture’s form. *Gesture*, *4*(2), 157–195. <https://doi.org/10.1075/gest.4.2.04ger>

Gibson, E., Futrell, R., Piantadosi, S. P., Dautriche, I., Mahowald, K., Bergen, L., & Levy, R. (2019). How Efficiency Shapes Human Language. *Trends in Cognitive Sciences*, *23*(5), 389–407. <https://doi.org/10.1016/j.tics.2019.02.003>

Giorgino, T. (2009). Computing and Visualizing Dynamic Time Warping Alignments in *R* : The **Dtw** Package. *Journal of Statistical Software*, *31*(7). <https://doi.org/10.18637/jss.v031.i07>

Haviland, J. B. (2013). The emerging grammar of nouns in a first generation sign language: Specification, iconicity, and syntax. *Gesture*, *13*(3), 309–353. <https://doi.org/10.1075/gest.13.3.04hav>

Hogan, N., & Sternad, D. (2009). Sensitivity of Smoothness Measures to Movement Duration, Amplitude and Arrests. *Journal of Motor Behavior*, *41*(6), 529–534. <https://doi.org/10.3200/35-09-004-RC>

Holler, J., & Wilkin, K. (2011). An experimental investigation of how addressee feedback affects co-speech gestures accompanying speakers’ responses. *Journal of Pragmatics*, *43*(14), 3522–3536. <https://doi.org/10.1016/j.pragma.2011.08.002>

Kelso, J. A. S., Tuller, B., & Harris, K. (1983). A “Dynamic Pattern” perspective on the control and coordination of movement. In *The production of speech*. Berlin: Springer-Verlag. Retrieved from <https://link.springer.com/chapter/10.1007/978-1-4613-8202-7_7>

Kelso, J. A., Tuller, B., Vatikiotis-Bateson, E., & Fowler, C. A. (1984). Functionally specific articulatory cooperation following jaw perturbations during speech: Evidence for coordinative structures. *Journal of Experimental Psychology. Human Perception and Performance*, *10*(6), 812–832.

Kirby, S., & Christiansen, M. H. (2003). From Language Learning to Language Evolution. In M. H. Christiansen & S. Kirby (Eds.), *Language Evolution* (pp. 272–294). Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199244843.003.0015>

Kirby, S., Cornish, H., & Smith, K. (2008). Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language. *Proceedings of the National Academy of Sciences*, *105*(31), 10681–10686. <https://doi.org/10.1073/pnas.0707835105>

Kirby, S., Griffiths, T., & Smith, K. (2014). Iterated learning and the evolution of language. *Current Opinion in Neurobiology*, *28*, 108–114. <https://doi.org/10.1016/j.conb.2014.07.014>

Lepic, R., Börstell, C., Belsitzman, G., & Sandler, W. (2016). Taking meaning in hand: Iconic motivations in two-handed signs. *Sign Language & Linguistics*, *19*(1), 37–81. <https://doi.org/10.1075/sll.19.1.02lep>

Lupton, L. K., & Zelaznik, H. N. (1990). Motor Learning in Sign Language Students. *Sign Language Studies*, *1067*(1), 153–174. <https://doi.org/10.1353/sls.1990.0020>

Maaten, L. van der, & Hinton, G. (2008). Visualizing Data using t-SNE. *Journal of Machine Learning Research*, *9*(86), 2579–2605. Retrieved from <http://jmlr.org/papers/v9/vandermaaten08a.html>

McNeilage, P. (2008). *The origin of speech*. New York: Oxford University Press.

Motamedi, Y., Schouwstra, M., Smith, K., Culbertson, J., & Kirby, S. (2019). Evolving artificial sign languages in the lab: From improvised gesture to systematic sign. *Cognition*, *192*, 103964. <https://doi.org/10.1016/j.cognition.2019.05.001>

Mueen, A. K., & Keogh, E. (2016). Extracting optimal performance from dynamic time warping. In *Proceedings of the 22Nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 2129–2130). [https://doi.org/10.1145/ 2939672.2945383](https://doi.org/10.1145/%202939672.2945383)

Muller, M. (2007). *Information retrieval for music and motion*. Heidelberg, Germany: Springer.

Namboodiripad, S., Lenzen, D., Lepic, R., & Verhoef, T. (2016). Measuring conventionalization in the manual modality. *Journal of Language Evolution*, *1*(2), 109–118. <https://doi.org/10.1093/jole/lzw005>

Ortega, G., & Özyürek, A. (2020). Types of iconicity and combinatorial strategies distinguish semantic categories in silent gesture across cultures. *Language and Cognition*, *12*(1), 84–113. <https://doi.org/10.1017/langcog.2019.28>

Östling, R., Börstell, C., & Courtaux, S. (2018). Visual Iconicity Across Sign Languages: Large-Scale Automated Video Analysis of Iconic Articulators and Locations. *Frontiers in Psychology*, *9*. <https://doi.org/10.3389/fpsyg.2018.00725>

Padden, C. A., Meir, I., Hwang, S.-O., Lepic, R., Seegers, S., & Sampson, T. (2013). Patterned iconicity in sign language lexicons. *Gesture*, *13*(3), 287–308. <https://doi.org/10.1075/gest.13.3.03pad>

Pattee, H. H., & Rączaszek-Leonardi, J. (2012). *LAWS, LANGUAGE and LIFE: Howard Pattee’s classic papers on the physics of symbols with contemporary commentary*. Springer Science & Business Media. Retrieved from <http://books.google.com?id=raEQodcVdYQC>

Pouw, W., & Dixon, J. A. (2019). Gesture networks: Introducing dynamic time warping and network analysis for the kinematic study of gesture ensembles. *Discourse Processes*. <https://doi.org/10.1080/0163853X.2019.1678967>

Pouw, W., & Trujillo, J. P. (2019). *Materials Tutorial Gespin2019 - Using video-based motion tracking to quantify speech-gesture synchrony*. Retrieved from <10.17605/OSF.IO/RXB8J>

Ravignani, A., Delgado, T., & Kirby, S. (2016). Musical evolution in the lab exhibits rhythmic universals. *Nature Human Behaviour*, *1*(1, 1), 1–7. <https://doi.org/10.1038/s41562-016-0007>

Raviv, L., Meyer, A., & Lev-Ari, S. (2019). Larger communities create more systematic languages. *Proceedings of the Royal Society B: Biological Sciences*, *286*(1907), 20191262. <https://doi.org/10.1098/rspb.2019.1262>

Rączaszek-Leonardi, J., & Scott Kelso, J. A. (2008). Reconciling symbolic and dynamic aspects of language: Toward a dynamic psycholinguistics. *New Ideas in Psychology*, *26*(2), 193–207. <https://doi.org/10.1016/j.newideapsych.2007.07.003>

Ripperda, J., Drijvers, L., & Holler, J. (2020). Speeding up the detection of non-iconic and iconic gestures (SPUDNIG): A toolkit for the automatic detection of hand movements and gestures in video data. *Behavior Research Methods*, *52*(4), 1783–1794. <https://doi.org/10.3758/s13428-020-01350-2>

Sandler, W. (2018). The Body as Evidence for the Nature of Language. *Frontiers in Psychology*, *9*. <https://doi.org/10.3389/fpsyg.2018.01782>

Scott-phillips, T. C., & Kirby, S. (2010). Language evolution in the laboratory. *Trends in Cognitive Sciences*, *14*, 411–417. <https://doi.org/10.1016/j.tics.2010.06.006>

Senghas, A., Kita, S., & Özyürek, A. (2004). Children creating core properties of language: Evidence from an emerging sign language in nicaragua. *Science*, *305*(5691), 1779–1782. <https://doi.org/10.1126/science.1100199>

Tomasello, M. (2008). *The origins of human communication*. Cambdride, MA: MIT press.

Trujillo, J. P., Vaitonyte, J., Simanova, I., & Özyürek, A. (2019). Toward the markerless and automatic analysis of kinematic features: A toolkit for gesture and movement research. *Behavior Research Methods*, *51*(2), 769–777. <https://doi.org/10.3758/s13428-018-1086-8>

Tyrone, M. E., & Mauk, C. E. (2016). The Phonetics of Head and Body Movement in the Realization of American Sign Language Signs. *Phonetica*, *73*(2), 120–140. <https://doi.org/10.1159/000443836>

Verhoef, T., Kirby, S., & de Boer, B. (2016). Iconicity and the Emergence of Combinatorial Structure in Language. *Cognitive Science*, *40*(8), 1969–1994. <https://doi.org/10.1111/cogs.12326>

Wang, M., Kong, L., Zhang, C., Wu, X., & Li, L. (2018). Speaking rhythmically improves speech recognition under "cocktail-party" conditions. *The Journal of the Acoustical Society of America*, *143*(4), EL255. <https://doi.org/10.1121/1.5030518>

Wilbur, R. B. (1990). An experimental investigation of stressed sign production. *International Journal of Sign Linguistics*, *1*(1).

# 6 Supplemental

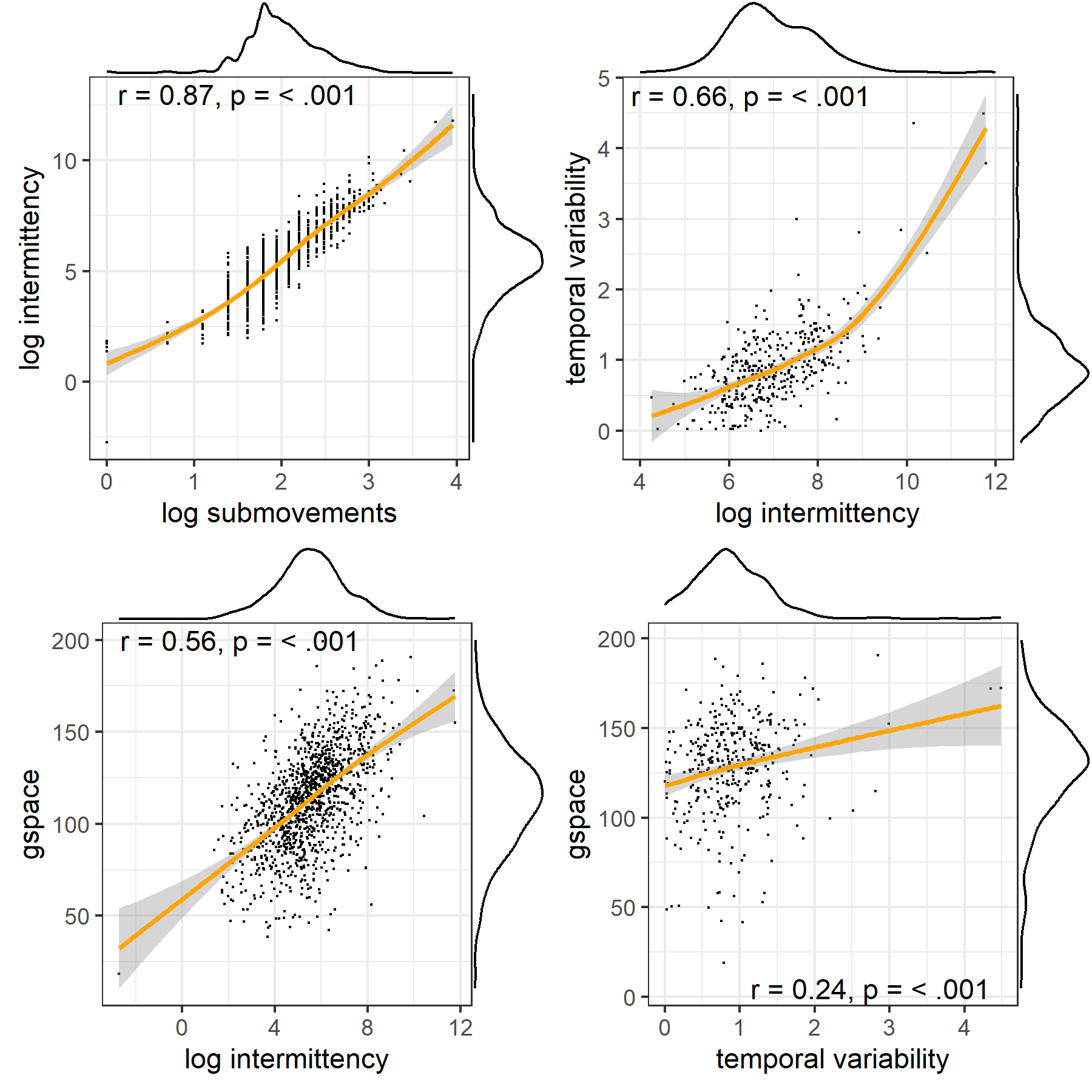
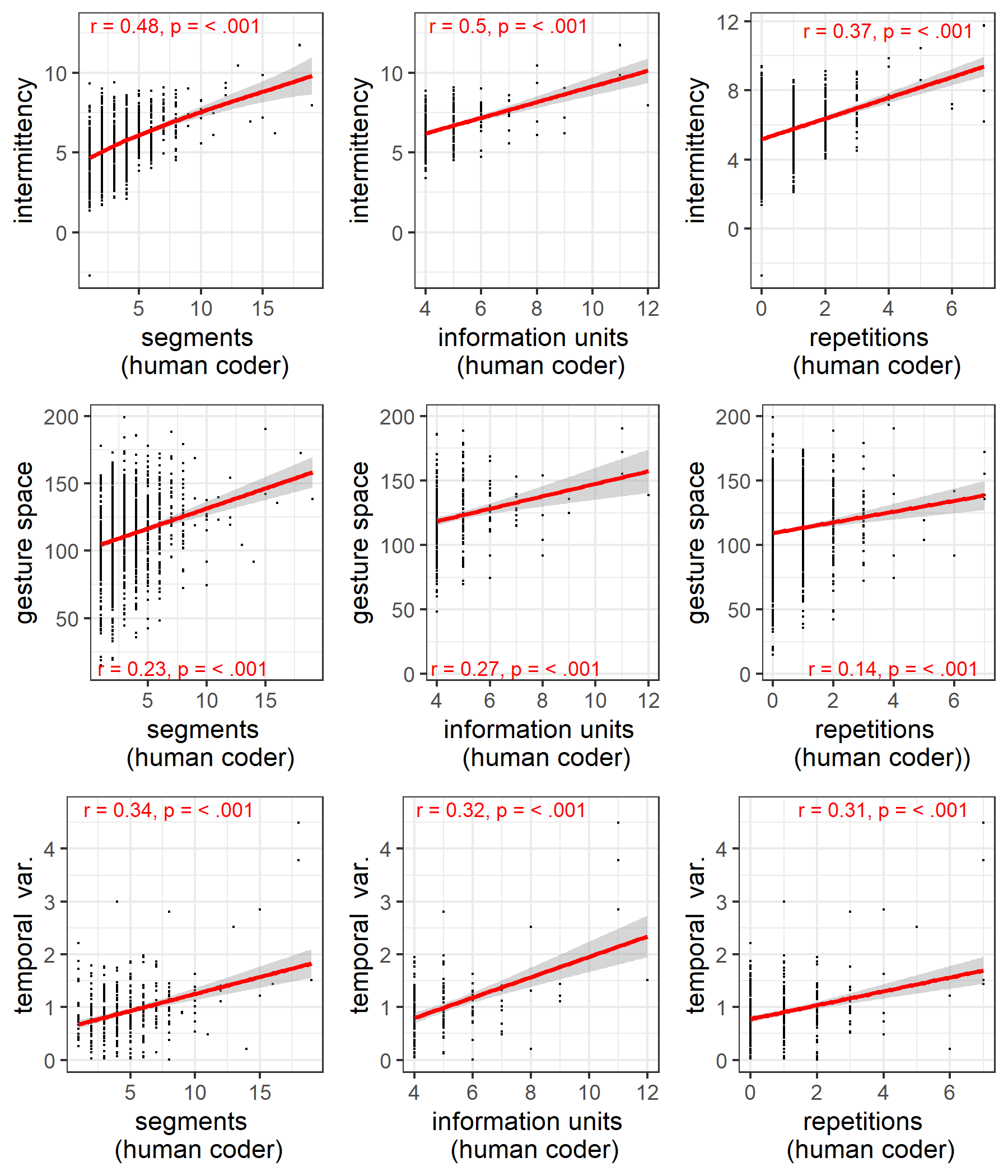
Figure S1. Correlations and distributions for kinematic measures per trial  *Note Figure S1*. Left upper panel, correlations and distributions are shown for intermittency and submovement. Given their high correlation we will use intermittency score for our final analysis. Other correlations are shown for the selected measures, rhythmiticy, gesture space and intermittency.

Figure S2. Correlations of kinematic measures with human-coded gesture information  
 *Note figure S2.* On the horizontal axes the human-coded number of gesture segments, unique information units, and the number repetitions (of information units) are shown. On the vertical axes our automatic kinematic measures are shown: intermittency, gesture space, and temporal variability. The findings show that our measures are a proxy for human judgments, such that more intermittent kinematics reflect more information units (repeater and/or unique). Larger gesture space is related to more information units. Lower temporal variability in kinematics is further associated with more information units.