Population Variation and Equity
Volume 1 · Number 1 · 2026

Coordinate-Channel Contrasts Across Human Activities: An Edge-Signal Description of Public Smartphone Recordings

Aaron Crawford[1] · Timothy Murphy[2] · Jason Reed[3]

[1] Croatia · [2] Greece · [3] Romania

Original publication: November 3, 2026 | Article 661

Abstract

Overall motion energy does not describe how acceleration is distributed across coordinate channels. Using 10,299 public activity windows from thirty participants, we measured normalized squared differences on the three pairs of body-acceleration axes. After accounting for pair identity and log motion energy, upstairs and downstairs walking had contrast values 0.124 and 0.133 above ordinary walking. The corresponding subject-clustered intervals excluded zero, while the standing contrast remained uncertain. These results describe a chosen coordinate representation of preprocessed signals. They do not establish a causal activity effect, orientation invariance, activity-classifier accuracy, or the performance of an adaptive graph-edge estimator.

Descriptive analysis of an existing waist-phone dataset. The declared channel graph is a representation choice, not a discovered physical or anatomical network.

Coordinate-Channel Contrasts Across Human Activities: An Edge-Signal Description of Public Smartphone Recordings

Descriptive analysis of an existing waist-phone dataset. The declared channel graph is a representation choice, not a discovered physical or anatomical network.

Abstract

Overall motion energy does not describe how acceleration is distributed across coordinate channels. Using 10,299 public activity windows from thirty participants, we measured normalized squared differences on the three pairs of body-acceleration axes. After accounting for pair identity and log motion energy, upstairs and downstairs walking had contrast values 0.124 and 0.133 above ordinary walking. The corresponding subject-clustered intervals excluded zero, while the standing contrast remained uncertain. These results describe a chosen coordinate representation of preprocessed signals. They do not establish a causal activity effect, orientation invariance, activity-classifier accuracy, or the performance of an adaptive graph-edge estimator.

1. Move from channel levels to relationships between channels

An accelerometer records motion along coordinate axes. A total-energy summary reduces those channels to one magnitude, which can be useful while obscuring how channels co-vary. Pairwise differences retain a different aspect of the recording. Their behavior may vary across activities even when a model adjusts for overall motion energy. This article examines that descriptive possibility using a public smartphone resource with activity and participant labels.

The research question is whether a dimensionless coordinate-channel contrast differs across the six recorded activity categories after adjustment for channel pair and total motion energy. The term edge signal refers to a deliberately declared graph on the three coordinate channels. It does not imply that the channels are separate physical devices connected by observed links, or that the chosen pairwise difference is an invariant property of a person's movement.

Graph-edge estimation, including Yan and Kuruoglu, 2023, motivates studying quantities defined on relationships rather than only on nodes. Our experiment does not run that adaptive estimation method or reconstruct a line graph from measured infrastructure. It offers a bounded empirical example of an edge-defined outcome whose interpretation can be checked directly against its input channels. The result concerns representation and subgroup differences, not superiority over a graph-learning algorithm.

2. What the graph means in this study

The perspectives in Ortega et al., 2017 and Shuman et al., 2012 explain how graph representations can organize multivariate signals. Here the nodes are the x, y, and z body-acceleration coordinates, and all three unordered pairs are edges. We assign this complete graph before calculating outcomes. Its topology follows the available coordinate representation, rather than a learned similarity threshold or evidence about anatomical connections.

Imputation systems such as Cini et al., 2021, You et al., 2020, and Cao et al., 2018 show other uses of multivariate relationships when measurements are incomplete. The forecasting architectures in Li et al., 2017 and Yu et al., 2017 use spatio-temporal dependencies in different application settings. None is fitted here. Their relevance is to the broader distinction between defining a relationship and learning a model

over it. The present analysis makes the relationship explicit and examines its observed variation without adding a neural predictor.

Coordinate choice is a material limitation. A rotation can preserve total three-axis energy while changing the differences between particular axes. A normalized edge contrast therefore does not automatically remove orientation sensitivity. We do not estimate phone orientation or rotate every participant into a common anatomical frame. A future orientation-robust representation would need to specify that transformation and evaluate its consequences rather than assuming that normalization alone supplies invariance.

3. The public recordings provide repeated activity windows

We use Reyes-Ortiz et al., 2013, documented in the resource paper Anguita et al., 2013. Thirty volunteers performed six activities while wearing a waist-mounted smartphone. The released dataset includes preprocessed inertial signals, activity labels, participant identifiers, and an official participant-based train/test partition. Its windows contain 128 samples over 2.56 seconds and overlap by fifty percent. These repeated and overlapping observations require a different interpretation from independently sampled individual movements.

The official training portion contains 7,352 windows and the test portion 2,947. We combine them to describe activity subgroups across all 10,299 available windows. This is not a held-out predictive evaluation. The original partition labels remain in the processed data, but no model is selected on one partition and advertised as validated on the other. Combining them supplies the descriptive sample relevant to the stated question while preserving its participant structure.

We extract the released body-acceleration signals along x, y, and z. These have undergone the creators' preprocessing and separation of body motion from gravity. They are not raw unfiltered accelerometer measurements. We do not redo that filtering, infer a new gravity component, or treat every small value during a static activity as an independently verified sensor error. The representation being described is the captured public body-acceleration release.

The activity labels are ordinary walking, walking upstairs, walking downstairs, sitting, standing, and laying. The last term follows the source label. Activity categories are observed descriptions, not randomized assignments in our secondary analysis. Participant behavior, placement, movement speed, and preprocessing can all affect measured channel relationships. The available data do not identify which of those contributes to each category difference.

4. Construct a contrast with an explicit denominator

For an axis pair within a window, we average the squared difference between its two body-acceleration signals. We divide that by the sum of the two channels' mean squared values in the same window. Every denominator is positive in the retained release. The resulting dimensionless contrast describes how much the channels differ relative to their combined energy. Three pairs produce three outcomes per window, yielding 30,897 edge-window rows.

The normalization is unchanged if both paired channels are multiplied by the same nonzero scale. This algebraic property is different from orientation invariance or robustness to arbitrary corruption. Changing only one channel, rotating the coordinate system, or adding an offset can alter the value. The empirical question is how the actual normalized values vary across activity groups, not whether a mathematically fixed scaling property can be discovered from the dataset.

A value near one occurs when the pair's cross-product averages near zero relative to its total energy. Lower or higher values can reflect positively or negatively aligned channel movement, respectively. This interpretation follows the definition and does not imply a causal interaction between coordinates. The two axes are measurements of the same movement system, and their apparent relation depends in part on the chosen frame and preprocessing.

For adjustment, we also calculate total motion energy as the mean sum of squares over all three body-acceleration channels and take its logarithm. The outcome and this covariate share underlying measurements. Adjusting for energy therefore creates a descriptive conditional comparison, not an independent instrument or a way to remove all common causes. That shared construction is recorded rather than hidden behind a generic control-variable label.

5. Model activity differences without claiming recognition accuracy

The regression outcome is normalized edge contrast. Predictors are activity indicators, edge-pair indicators, and log motion energy. Ordinary walking is the activity reference and the xy pair the edge reference. The specification is additive: it estimates common activity differences across the three pair types rather than a fully interacted activity-by-pair surface. Descriptive pair means are retained separately so that the additive approximation does not conceal all pair-specific variation.

Standard errors cluster by participant, allowing overlapping windows and repeated activities and edge rows from the same individual to share errors. We use thirty clusters, the recorded small-sample correction, and a t reference with twenty-nine degrees of freedom, following dependence concerns in Cameron and Miller, 2015. Heteroskedasticity-consistent variance in White, 1980 would not alone account for that participant repetition. The clustering scheme does not establish representativeness beyond the recruited volunteer group.

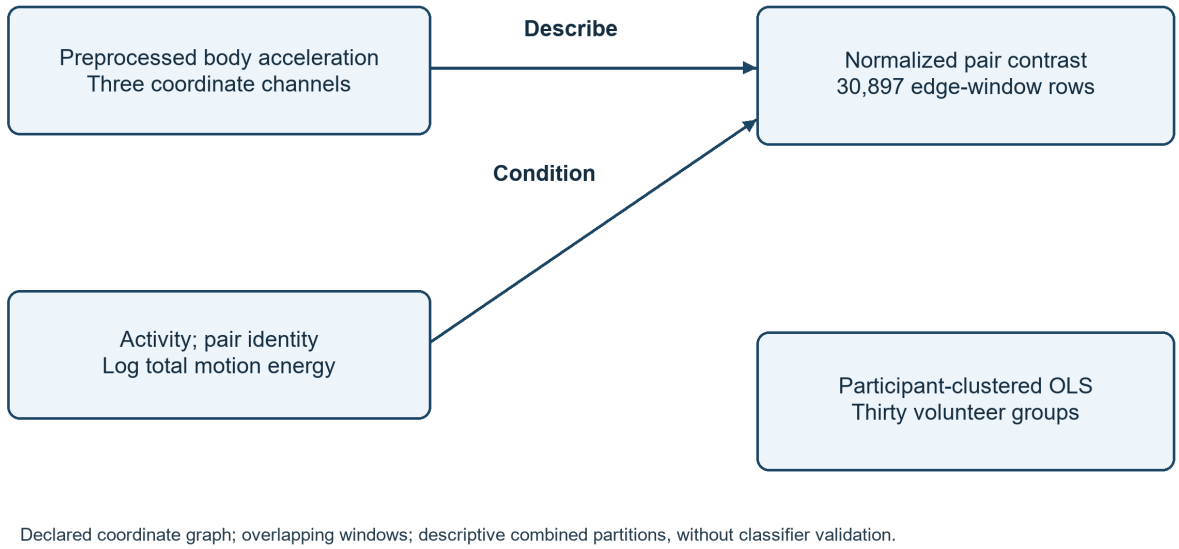


Figure 1. Preprocessed acceleration channels define three normalized pair contrasts, modeled by activity, pair identity, and overall motion energy.

Table 1. Selected adjusted activity, pair, and energy coefficients for the normalized contrast outcome.

Model	Term	Estimate (SE)	95% CI	N
M1	Upstairs vs walking	0.1238 (0.0179)	[0.0871, 0.1605]	30897
M1	Downstairs vs walking	0.1334 (0.0166)	[0.0994, 0.1673]	30897
M1	Sitting vs walking	0.0973 (0.0331)	[0.0297, 0.1650]	30897
M1	Standing vs walking	-0.0714 (0.0364)	[-0.1458, 0.0030]	30897
M1	Laying vs walking	0.1770 (0.0310)	[0.1136, 0.2404]	30897
M1	xz vs xy pair	0.0620 (0.0203)	[0.0205, 0.1035]	30897
M1	yz vs xy pair	-0.1966 (0.0217)	[-0.2411, -0.1522]	30897
M1	Log motion energy	0.0048 (0.0032)	[-0.0017, 0.0113]	30897
M1: fit	R-squared = 0.1877	-	-	30897

Model labels: M1 = Activity differences in normalized channel-edge contrast.

The model R-squared is 0.1877. This describes variation in a signal summary and should not be called activity-recognition accuracy. We predict no class labels and calculate no held-out confusion matrix. Mixed-effects ideas in Laird and Ware, 1982 could provide another model for repeated participant observations, but no such fit is reported here. The actual table comes from ordinary least squares with the specified participant-clustered covariance.

6. Stair activities retain differences after amplitude adjustment

Relative to walking, the upstairs activity coefficient is 0.12377, with standard error 0.01795 and interval 0.08706 to 0.16048. The downstairs coefficient is 0.13335, with standard error 0.01659 and interval 0.09942 to 0.16728. Under the fitted additive description, both have higher normalized channel contrast after adjustment for pair identity and total energy. We do not infer that stair travel causes these exact changes independently of speed, placement, and other movement differences.

Sitting has coefficient 0.09733, with interval 0.02968 to 0.16499. Standing has coefficient -0.07138, with interval -0.14577 to 0.00301, so its difference from walking is uncertain under this specification. Laying has coefficient 0.17696, with interval 0.11356 to 0.24037. These patterns show that the normalized coordinate relation is not simply a scalar ranking of how physically active a category sounds.

The log-motion-energy coefficient is 0.00480, with standard error 0.00317 and interval -0.00168 to 0.01128. Its conditional association is not clearly separated from zero. That does not establish that energy is irrelevant to movement or that normalization removes every amplitude dependence. It describes one coefficient after activity and pair adjustment in an outcome whose definition already includes a local energy denominator.

Pair identity matters as well. Relative to xy, xz has coefficient 0.06202 with interval 0.02052 to 0.10352, while yz has coefficient -0.19663 with interval -0.24111 to -0.15215. Thus the same activity labels do not erase coordinate-pair differences. A report pooling all three pairs without specifying their frame would lose a feature of the recorded representation that remains visible after the chosen controls.

7. Pair means reveal limits of a single activity coefficient

For walking, mean contrasts are 1.1020 for xy, 1.1131 for xz, and 0.7768 for yz. For upstairs walking they are 1.2672, 1.1952, and 0.9033, respectively. Downstairs walking gives 1.3397, 1.1864, and 0.8751. These are unconditional pair means. They are not expected to reproduce adjusted activity coefficients exactly because the regression also accounts for the observed energy distribution and imposes an additive pair structure.

Laying has mean values 0.9243 for xy, 1.2624 for xz, and 1.2499 for yz. This configuration differs visibly from the walking pattern rather than representing a uniform increase on every pair. The additive activity coefficient is therefore a compact conditional summary, not a claim that each coordinate pair responds identically. A later interaction model could explore that heterogeneity, but it was not selected after the present table to maximize a favored activity difference.

Agreement concepts in Bland and Altman, 1986, Lin, 1989, and Koo and Li, 2016 help distinguish these observed contrasts from reliability or interchangeability of sensors. The three channels are not independent instruments measuring a single scalar in identical units of interpretation. We estimate neither an agreement limit nor an intraclass reliability coefficient. Measurement-error reasoning in Carroll et al., 2006 remains relevant to placement, preprocessing, and recorded values even though the normalized outcome is dimensionless.

8. Robustness and missingness are additional questions

Robust-estimation frameworks in Huber, 1964, Hampel, 1974, and Rousseeuw, 1984 concern how unusual observations affect an estimator. A normalized pair statistic is not automatically robust in that sense. A small denominator or an exceptional channel can influence its value. Our retained signals have positive denominators, but we do not inject faults, trim values by their effect on the result, or claim protection against adversarial sensor perturbations.

Influence analysis following Cook, 1977 could identify participants or windows that strongly affect a coefficient. Such observations remain in the reported fit. Missing-data reasoning in Little and Rubin, 2002 also warns against treating complete released windows as proof that every attempted recording was available or that all real-world activities share the same data quality. The public matrix defines the available evidence; unrecorded acquisition failures are not reconstructed by assumption.

Validation guidance in Roberts et al., 2017 and selection concerns in Varma and Simon, 2006 would be essential for a later classifier using these edge features. Overlapping windows and shared participants can make random row splits overly optimistic. The present design avoids a predictive claim rather than claiming that descriptive adjustment solves that problem. A future recognition study should define participant-level transfer, lock features before evaluation, and compare performance on genuinely held-out individuals.

9. Use the edge description for the question it answers

Scoring concepts in Gneiting and Raftery, 2007 and accuracy-scale discussion in Hyndman and Koehler, 2006 reinforce the need to name the endpoint. This endpoint is a dimensionless relationship between coordinate signals, with uncertainty summarized across participants. It is not a forecast error, a probability of an activity, or a clinical mobility assessment. Its practical usefulness depends on whether a later task benefits from that relationship beyond simpler signal features.

The statistical interpretation principles in Wasserstein and Lazar, 2016 support retaining the uncertain standing and energy coefficients. Multiplicity considerations in Benjamini and Hochberg, 1995 would

matter if many pair definitions and activity contrasts were searched for favorable findings. We preserve the complete three-pair definition and all six categories. This makes the result an auditable description of a chosen representation, while leaving claims of optimal graph design or a universally useful edge feature untested.

Documentation under Gebru et al., 2021 records the combined official partitions, body-acceleration preprocessing, window overlap, participant clustering, edge definition, and shared energy covariate. Wilkinson et al., 2016 motivates preservation of the raw archive and derived rows. The empirical conclusion is that activity categories differ in this coordinate-channel contrast even after simple energy adjustment, with important pair-specific patterns. The contrast is informative about the released representation and requires additional validation before it can support recognition, physical-network, or orientation-invariant claims.

References

1. Ortega, Antonio; Frossard, Pascal; Kovačević, Jelena; Moura, José M. F.; Vandergheynst, Pierre. (2017). Graph Signal Processing: Overview, Challenges and Applications. arXiv:1712.00468. https://arxiv.org/abs/1712.00468.
2. Li, Yaguang; Yu, Rose; Shahabi, Cyrus; Liu, Yan. (2017). Diffusion Convolutional Recurrent Neural Network: Data-Driven Traffic Forecasting. arXiv:1707.01926. https://arxiv.org/abs/1707.01926.
3. Gneiting, Tilmann; Raftery, Adrian E. (2007). Strictly Proper Scoring Rules, Prediction, and Estimation. Journal of the American Statistical Association. https://doi.org/10.1198/016214506000001437.
4. Koo, Terry K., and Mae Y. Li. "A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research." *Journal of Chiropractic Medicine*, vol. 15, no. 2, 2016, pp. 155-163. https://doi.org/10.1016/j.jcm.2016.02.012.
5. Varma, Sudhir and Richard Simon. “Bias in error estimation when using cross-validation for model selection.” *BMC Bioinformatics*, vol. 7, no. 1, 2006. https://doi.org/10.1186/1471-2105-7-91.
6. White, Halbert. “A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity.” *Econometrica*, vol. 48, no. 4, 1980, pp. 817-838. https://doi.org/10.2307/1912934.
7. Anguita, Davide; Ghio, Alessandro; Oneto, Luca; Parra, Xavier; Reyes-Ortiz, Jorge Luis. (2013). A Public Domain Dataset for Human Activity Recognition Using Smartphones. ESANN 2013, 437-442. https://www.esann.org/sites/default/files/proceedings/legacy/es2013-84.pdf.
8. Carroll, Raymond J., David Ruppert, Leonard A. Stefanski, and Ciprian M. Crainiceanu. “Measurement Error in Nonlinear Models.” 2006. https://doi.org/10.1201/9781420010138.
9. Yu, Bing; Yin, Haoteng; Zhu, Zhanxing. (2017). Spatio-Temporal Graph Convolutional Networks: A Deep Learning Framework for Traffic Forecasting. arXiv:1709.04875. https://arxiv.org/abs/1709.04875.
10. Cook, R. Dennis. (1977). Detection of Influential Observation in Linear Regression. Technometrics. https://doi.org/10.1080/00401706.1977.10489493.
11. Laird, Nan M. and James H. Ware. “Random-Effects Models for Longitudinal Data.” *Biometrics*, vol. 38, no. 4, 1982, pp. 963. https://doi.org/10.2307/2529876.
12. Yan, Yi; Kuruoglu, Ercan Engin. (2023). Adaptive Spatio-temporal Estimation on the Graph Edges via Line Graph Transformation. arXiv:2311.00656. https://arxiv.org/abs/2311.00656.
13. Huber, Peter J.. (1964). Robust Estimation of a Location Parameter. The Annals of Mathematical Statistics. https://doi.org/10.1214/aoms/1177703732.
14. Cao, Wei; Wang, Dong; Li, Jian; Zhou, Hao; Li, Lei; Li, Yitan. (2018). BRITS: Bidirectional Recurrent Imputation for Time Series. arXiv:1805.10572. https://arxiv.org/abs/1805.10572.

15. Wilkinson, Mark D.; Dumontier, Michel; Aalbersberg, IJsbrand Jan; Appleton, Gabrielle; Axton, Myles; Baak, Arie; Blomberg, Niklas; Boiten, Jan-Willem; da Silva Santos, Luiz Bonino; Bourne, Philip E.; Bouwman, Jildau; Brookes, Anthony J.; Clark, Tim; Crosas, Mercè; Dillo, Ingrid; Dumon, Olivier; Edmunds, Scott; Evelo, Chris T.; Finkers, Richard; Gonzalez-Beltran, Alejandra; Gray, Alasdair J.G.; Groth, Paul; Goble, Carole; Grethe, Jeffrey S.; Heringa, Jaap; 't Hoen, Peter A.C; Hooft, Rob; Kuhn, Tobias; Kok, Ruben; Kok, Joost; Lusher, Scott J.; Martone, Maryann E.; Mons, Albert; Packer, Abel L.; Persson, Bengt; Rocca-Serra, Philippe; Roos, Marco; van Schaik, Rene; Sansone, Susanna-Assunta; Schultes, Erik; Sengstag, Thierry; Slater, Ted; Strawn, George; Swertz, Morris A.; Thompson, Mark; van der Lei, Johan; van Mulligen, Erik; Velterop, Jan; Waagmeester, Andra; Wittenburg, Peter; Wolstencroft, Katherine; Zhao, Jun; Mons, Barend. (2016). The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data. https://doi.org/10.1038/sdata.2016.18.

16. Hampel, Frank R.. (1974). The Influence Curve and its Role in Robust Estimation. Journal of the American Statistical Association. https://doi.org/10.1080/01621459.1974.10482962.

17. Roberts, David R., et al. "Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure." *Ecography*, vol. 40, no. 8, 2017, pp. 913-929. https://doi.org/10.1111/ecog.02881.

18. Little, Roderick J. A., and Donald B. Rubin. *Statistical Analysis with Missing Data*. Wiley, 2002. https://doi.org/10.1002/9781119013563.

19. Wasserstein, Ronald L.; Lazar, Nicole A.. (2016). The ASA Statement on p -Values: Context, Process, and Purpose. The American Statistician. https://doi.org/10.1080/00031305.2016.1154108.

20. Reyes-Ortiz, Jorge; Anguita, Davide; Ghio, Alessandro; Oneto, Luca; Parra, Xavier. (2013). Human Activity Recognition Using Smartphones [Dataset]. UCI Machine Learning Repository. https://doi.org/10.24432/C54S4K.

21. Bland, J. Martin, and Douglas G. Altman. "Statistical Methods for Assessing Agreement between Two Methods of Clinical Measurement." *The Lancet*, vol. 327, no. 8476, 1986, pp. 307-310. https://doi.org/10.1016/S0140-6736(86)90837-8.

22. Shuman, David I; Narang, Sunil K.; Frossard, Pascal; Ortega, Antonio; Vandergheynst, Pierre. (2012). The Emerging Field of Signal Processing on Graphs: Extending High-Dimensional Data Analysis to Networks and Other Irregular Domains. arXiv:1211.0053. https://arxiv.org/abs/1211.0053.

23. Gebru, Timnit, et al. "Datasheets for datasets" *Communications of the ACM*, 2021. https://doi.org/10.1145/3458723.

24. Cameron, A. Colin, and Douglas L. Miller. "A Practitioner's Guide to Cluster-Robust Inference." *Journal of Human Resources*, vol. 50, no. 2, 2015, pp. 317-372. https://doi.org/10.3368/jhr.50.2.317.

25. Cini, Andrea; Marisca, Ivan; Alippi, Cesare. (2021). Filling the G_ap_s: Multivariate Time Series Imputation by Graph Neural Networks. arXiv:2108.00298. https://arxiv.org/abs/2108.00298.

26. Hyndman, Rob J. and Anne B. Koehler. "Another look at measures of forecast accuracy." *International Journal of Forecasting*, vol. 22, no. 4, 2006, pp. 679-688. https://doi.org/10.1016/j.ijforecast.2006.03.001.

27. Benjamini, Yoav and Yosef Hochberg. "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing." *Journal of the Royal Statistical Society Series B: Statistical Methodology*, vol. 57, no. 1, 1995, pp. 289-300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x.

28. You, Jiaxuan; Ma, Xiaobai; Ding, Daisy Yi; Kochenderfer, Mykel; Leskovec, Jure. (2020). Handling Missing Data with Graph Representation Learning. arXiv:2010.16418. https://arxiv.org/abs/2010.16418.

29. Rousseeuw, Peter J.. (1984). Least Median of Squares Regression. Journal of the American Statistical Association. https://doi.org/10.1080/01621459.1984.10477105.

30. Lin, Lawrence I-Kuei. (1989). A Concordance Correlation Coefficient to Evaluate Reproducibility. Biometrics. https://doi.org/10.2307/2532051.