Measuring Dependencies between Biological Signals with Temporal Self-supervision, and its Limitations

Evangelos Sariyanidi[1,*], John D. Herrington[1,2], Lisa Yankowitz[1], Pratik Chaudhari[2], Theodore D. Satterthwaite[2], Casey J. Zampella[1], Robert T. Schultz[1,2], Russell T. Shinohara[2], Birkan Tunc[1,2]

[1] The Children's Hospital of Philadelphia [2]University of Pennsylvania

Abstract

Measuring the statistical dependence between observed signals is a primary tool for scientific discovery. However, biological systems often exhibit complex non-linear interactions that currently cannot be captured without a priori knowledge regarding the nature of dependence. We introduce a self-supervised approach, concurrence, which is inspired by the observation that if two signals are dependent, then one should be able to distinguish between temporally aligned vs. misaligned segments extracted from them. Experiments with fMRI, physiological and behavioral signals show that, to our knowledge, concurrence is the first approach that can expose relationships across such a wide spectrum of signals and extract scientifically relevant differences without ad-hoc parameter tuning or reliance on a priori information, providing a potent tool for scientific discoveries across fields. However, depencencies caused by extraneous factors remain an open problem, thus researchers should validate that exposed relationships truely pertain to the question(s) of interest.

1 Introduction

Measuring dependencies between biological signals is fundamental for understanding the complex interplay within and between molecular, neurobiological, and behavioral processes. The most common approach to quantifying statistical dependence is using linear model-based statistics, with the Pearson correlation coefficient being the dominant metric [1]. However, biological systems often exhibit interactions [2] that cannot be captured by linear models [3], such as cross-frequency coupling [4–6], threshold effects [7], phase shifts [8], feedback systems [7, 9], or multi-scale interactions [10].

While linear models cannot comprehensively capture statistical dependence, linear and non-linear models together can. That is, if two time signals x and y are dependent but uncorrelated, then there must be (non-linear) mathematical transformations f and g such that the transformed signals $f(x)$ and $g(y)$ are correlated [11]. However, the specific transformations that expose the dependence can be particular to each problem and difficult to identify when the compared signals are generated by complex or unknown mechanisms. The Hilbert-Schmidt Independence Criterion [12] –which can be considered as a generalization of distance correlation [13, 14]– or variants of canonical correlation analysis [15, 16] can, in principle, determine linear and non-linear dependence. However, these approaches are successful only if one can identify model parameters or kernels that expose the dependence [17, 18], which may not be possible or may require large samples [19, 20]. Alternatively, one may use analytical transformations such as Fourier or wavelet decomposition [21–23], but the generalizability of this approach is limited, as there is no single transformation that works for all signals [24, 25]. Moreover, analytical transformations pose family-wise error problems [26, 27] because they typically decompose each signal into multiple signals (e.g., frequency bands), and dependence can occur between any pair of decomposed signals (e.g., cross-frequency dependence). These issues are exacerbated when the compared signals are multi-dimensional and only a subset in

[*] sariyanide@chop.edu

Preprint.

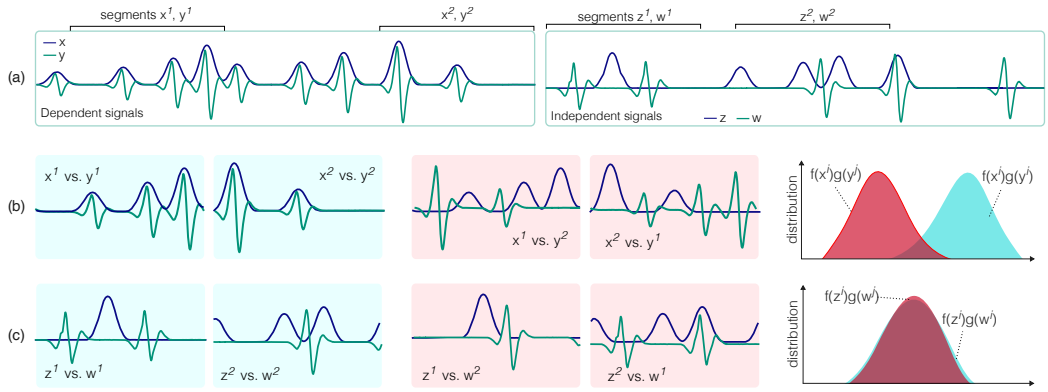

Figure 1: (a) A pair of dependent signals x and y, where x approximately resembles the integral of y; and a pair of independent signals w and z. (b) Concurrent segments from the dependent signals x and y have different characteristics from non-concurrent segments, therefore it is possible to find functions f and g such that when they are respectively applied to concurrent segments of x and y they produce a larger inner product compared to when they are applied to non-concurrent segments; for example, when f is the identity operator and g is intergral operator. (c) The concurrent and non-concurrent segments from independent signals exhibit similar characteristics, therefore it is not possible to find functions f and g that produce a different inner product on concurrent versus non-concurrent segments of z and w.

one set of signals depends on an unknown subset in the other. In sum, currently there is no tractable method that can detect or quantify the degree of dependence between a broad variety of biological signals when the dependence structure is not known a priori—presenting a major obstacle to scientific discovery.

We introduce a new approach, called concurrence, to quantify the statistical dependence between pairs of signals. The proposed approach is built on the following heuristic: if two signals are statistically dependent, then temporally aligned (i.e., concurrent) segments of the compared signals must be separable from segments that are temporally misaligned (i.e., not concurrent), as illustrated in Fig. 1.This separability criterion provides a straightforward recipe for automatically finding linear or non-linear transformations that expose dependence—namely, training a machine learning model that classifies between concurrent vs. non-concurrent segments extracted from signals (Fig. 1). We test this approach in experiments with three distinct biological signal types (i.e., fMRI, physiological, and behavioral data). Results indicate that the concurrence approach can expose a large range of dependencies without any ad-hoc modification, as the same parameters successfully detected dependence in all three signal types; and can handle noise and stochastic dependence (Fig. 2). Furthermore, our approach yields a score coined the concurrence coefficient, which is scaled between 0 and 1. Finally, concurrence is computationally efficient and we provide an open-source software package to implement it for research purposes.

2 The Concurrence Approach

Biological signals often contain dependencies that can be observed within relatively short time windows. For example, the dependence between respiration and cardiac activity occurs regularly and can be observed within a few seconds [28]. Multimodal behavioral signaling during conversations contains events indicative of social coordination, such as mimicry [29] and backchanneling [30], which also occur within a short time window. Dependencies between activity in different parts of the brain can be observed (e.g., using fMRI) instantaneously or within a short time delay, usually milliseconds. As such, one can focus on relatively small time chunks and still capture a broad range of dependencies.

Suppose that $x_{t,w}$ and $y_{t,w}$ are segments of time signals x and y, observed between the time points t and $t + w$. If both $x_{t,w}$ and $y_{t,w}$ contain (finite) responses to a common event, then they must be statistically dependent. Thus, there must exist transformations f and g such that transformed

versions of the segments, $f(x_{t,w})$ and $g(y_{t,w})$, are linearly correlated [11]. The crux of our approach is that, while $f(x_{t,w})$ and $g(y_{t,w})$ are expected to be correlated, $f(x_{t,w})$ and $g(y_{t',w})$ are, on average, uncorrelated if t' is a random time point, different from t. For example, if x and y are behavioral signals of two people in a conversation, t may indicate a time when one partner spontaneously mimics the behavior of the other, and it is unlikely that the same mimicry pattern is present at a random time t'.

Since extracting concurrent or non-concurrent segments from synchronized pairs of signals is trivial, this intuitive idea provides a straightforward manner of detecting linear or non-linear dependence fully automatically, without relying on an appropriate choice of kernel or other hyperparameters. Specifically, we quantify dependence through the concurrence coefficient, which is obtained by training a machine learning model to classify between randomly cropped concurrent versus non-concurrent segments in a dataset of signal pairs, and then calculating the normalized classification accuracy on another dataset $D = (x^1, y^1), (x^2, y^2), \ldots (x^N, y^N)$ as:

$$\text{concurrence coefficient} = 2 \times \max(\text{accuracy}, 0.5) - 1 \tag{1}$$

This dataset D must not overlap with the dataset used while training the machine learning model, lest the classifier may overfit and overestimate the dependence. Thus, one may use a cross-validation (CV) procedure and compute the average concurrence coefficient over the test sets of the CV folds, or compute the concurrence coefficient on an independent dataset.

The concurrence coefficient is bounded between 0 and 1, and its magnitude indicates the strength of dependence between the compared signal pairs (Fig. 2). The classifier that we use is a neural network that separately processes segments $x_{t,w}$ and $y_{t,w}$ through learned transformations f and g, which can be multi-dimensional even if the input is one-dimensional. This network design allows us to also quantify the dependence between an individual pair of segments $x_{t,w}$ and $y_{t,w}$ as the per-segment concurrence score (PSCS), which can be interpreted as the weighted average of the entries of the covariance matrix between the transformed segments, $f(x_{t,w})$ and $g(y_{t,w})$. While the concurrence coefficient can uncover whether and to what extent two biological processes (e.g., breathing rate and cardiac activity) are related in general (i.e., at the sample level), the PSCS can uncover whether this relationship is stronger for a specific individual, for individuals with a certain condition (e.g., anxiety), or for certain moments within the compared signals. Our experiments on real data include use cases for both the concurrence coefficient and the PSCS.

The concurrence approach is practically viable, because the transformations f and g are modeled with Convolutional Neural Networks (CNNs), which are universal approximators [31] and thus can model arbitrary non-linear functions. Also, thanks to recent advances in machine learning [32, 33], CNNs can now be trained successfully across a large variety of temporal analysis tasks without ad-hoc modifications [32], particularly when modeling short-term dependencies [33]. Our experiments with real and synthetic data verify that CNNs with the same parameters (Table 1) can detect a wide range of linear or non-linear dependence patterns between signals that have distinct frequency characteristics and are corrupted by large amounts of noise. Further, the training does not require an unrealistic sample size, as our experiments show that fewer than 100 signal pairs can be enough to expose the dependence.

2.1 Measuring Stochastic Dependence

Biological signals may depend on each other stochastically rather than deterministically, as they typically reflect an admixture of multiple processes, and only a subset of these processes may be dependent between compared signals. Another factor leading to non-deterministic dependence is measurement noise, which can affect the compared signals significantly, to the point that there is more noise than signal [34]. This section investigates the ability of the concurrence method to determine and quantify stochastic dependence.

Fig. 2A simulates stochastic dependence scenarios where the degree of dependence is controlled with a parameter ξ, which varies between 0 (independent signal pairs) and 1 (completely dependent signal pairs). Fig. 2F verifies that the concurrence coefficient can expose dependencies across a large range of stochasticity (i.e., ξ values). Of note, the concurrence coefficient is approximately linearly proportional to ξ when the segment size w is large enough to contain the entire event (e.g., w=100 for signals as in Fig. 2A) but not much larger. While the degree of dependence ξ is overestimated

with larger segments, this does not imply that the concurrence coefficient risks detecting spurious relationships (false positives), as the concurrence coefficient is approximately zero when $\xi = 0$, regardless of the segment size w. If available, information about the segment size within which a dependence event is expected to be observed should be used, especially if one aims to uncover whether a pair of biological processes are more strongly dependent than another pair, lest the comparisons may not be commensurate. Nevertheless, the exact w value is of little concern when the goal is to uncover whether two processes are dependent or not, as w values of different orders of magnitude can robustly expose dependence (Fig. 2F) and under no dependence, the estimated concurrence coefficient is zero.

Fig. 2G shows that the proposed method can also handle measurement noise. The concurrence coefficient usually decreases with the signal-to-noise (SNR), yet it can uncover dependence even when the SNR is 0.10, which is lower than a worst-case estimate of noise in fMRI data (SNR=0.35) [34].

3 Experiments

We tested the concurrence method on three types of real data with single- and multi-dimensional biological signals and included cases of linear or non-linear dependence. Specifically, we experimented on brain imaging (fMRI), physiological (breathing and heart rate), and behavioral (facial expressions and head movements) signals.

3.1 Brain Imaging

Our experiments on fMRI signals aim to identify how strongly different brain regions are functionally connected. Pearson's correlation is the single-most commonly used metric for this purpose [38]. Fig. 3 compares the connectivity matrices obtained with Pearson's correlation (i.e., correlation matrix) and the concurrence coefficient (i.e., concurrence matrix) on a version of the Philadelphia Neurodevelopmental Cohort dataset [39] that was pre-processed as in prior work [40]. This dataset uses the parcellation scheme that divides each brain into 400 regions [41]. The concurrence coefficient is computed on segments of size $w = 30$ time points, which corresponds to approximately 90 seconds, whereas the entire signals included 120 time points. Thirty percent of the dataset (426 participants) was used to train the neural networks needed for the concurrence coefficients, and the results in both connectivity matrices were computed from the remaining 70%.

The overall similarity between the two connectivity matrices (Fig. 3C vs. Fig. 3D) is striking and suggests that the concurrence coefficient uncovers a dependence structure that has been validated in the field. Fig. 3G shows that there are no pairs of regions with a concurrence score less than 0.2, even though there are many pairs that are uncorrelated (i.e., Pearson r $\approx$ 0), suggesting that concurrence captured statistical dependencies that cannot be captured with correlation (e.g., Fig. 3B) as well as those that can (e.g., Fig. 3A). The fact that the concurrence coefficient exposed a dependence between all 400×199=79,600 pairs of brain regions verifies that the training of the neural network needed for the concurrence coefficient can be done robustly. We ran a permutation test to identify if the method detects spurious dependence (Type I error) by computing concurrence between signals of mismatched participants. The concurrence coefficient was closely distributed around zero (Fig. 3H), indicating no spurious relationships. The differences between the concurrence coefficient and Pearson's correlation exhibit a structured pattern across the seven brain networks (Fig. 3F), increasing progressively from lower-order affective (limbic), somatomotor and sensory (visual) networks to higher-order cognitive control (ventral attention, dorsal attention, default mode, frontoparietal) networks. This systematic increase suggests that linear correlation may not be capturing complex connectivity patterns that involve integrative processing or dynamic modulation.

3.2 Physiological data

We next investigate dependencies in a dataset of breathing and cardiac activity. While these two processes are known to be biologically linked [42], the correlation between respiration rate and electrocardiogram (ECG) signals is approximately zero (Fig. 4A). We applied the proposed method to a dataset of 60 pairs of temporally synchronized ECG and respiration rate signals, collected at the Children's Hospital of Philadelphia using Zephyr BioModule sensors. The duration of the tasks used

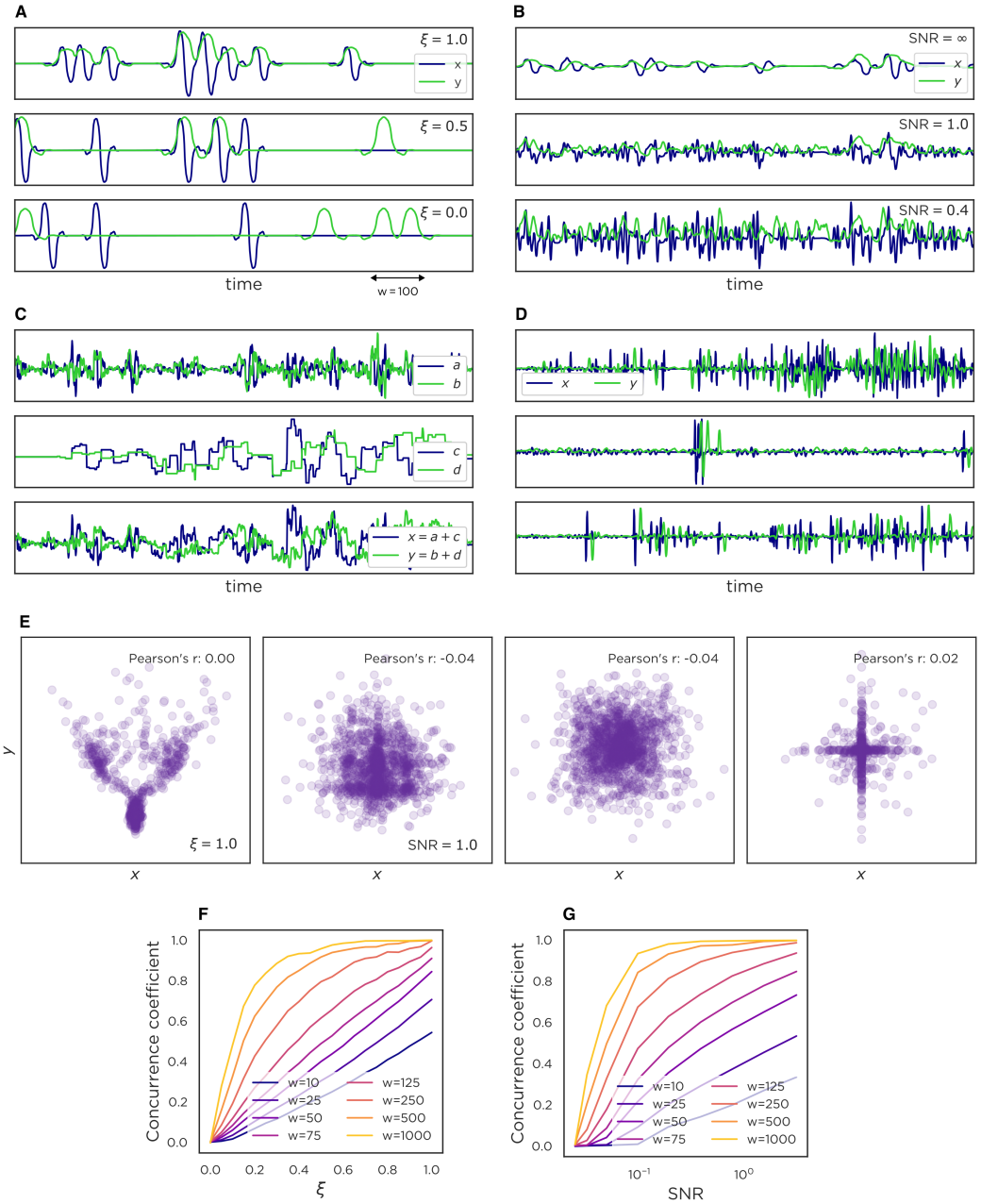

Figure 2: Synthesized signals with various structures of dependence that can be detected with the concurrence coefficient. (A) Examples of synthesized signals with deterministic dependence ($\xi = 1.0$), stochastic dependence ($\xi = 0.5$) and no dependence ($\xi = 0.0$). (B) A dependent pair of signals with varying degrees of added noise. (C) A pair of (multi-factored) dependent signals x and y that are the sum of two other pairs of dependent signals (a,b) and (c,d). (D) Examples of signals that are non-stationary (i.e., variance increases over time) and non-ergodic (i.e., the variance in each pair of signals is different). (E) Scatter plots of the signals from the four respective cases (A to D) indicate that the signals are uncorrelated albeit dependent. (F) Results from synthesized signals with various degrees of dependence (i.e., ξ) indicate that the concurrence coefficient increases with ξ regardless of the choice of segment size w. (G) Results from synthesized signals with varying degrees of signal-to-noise ratio (SNR) indicate that the concurrence coefficient can successfully expose dependance even if there is more noise than signal.

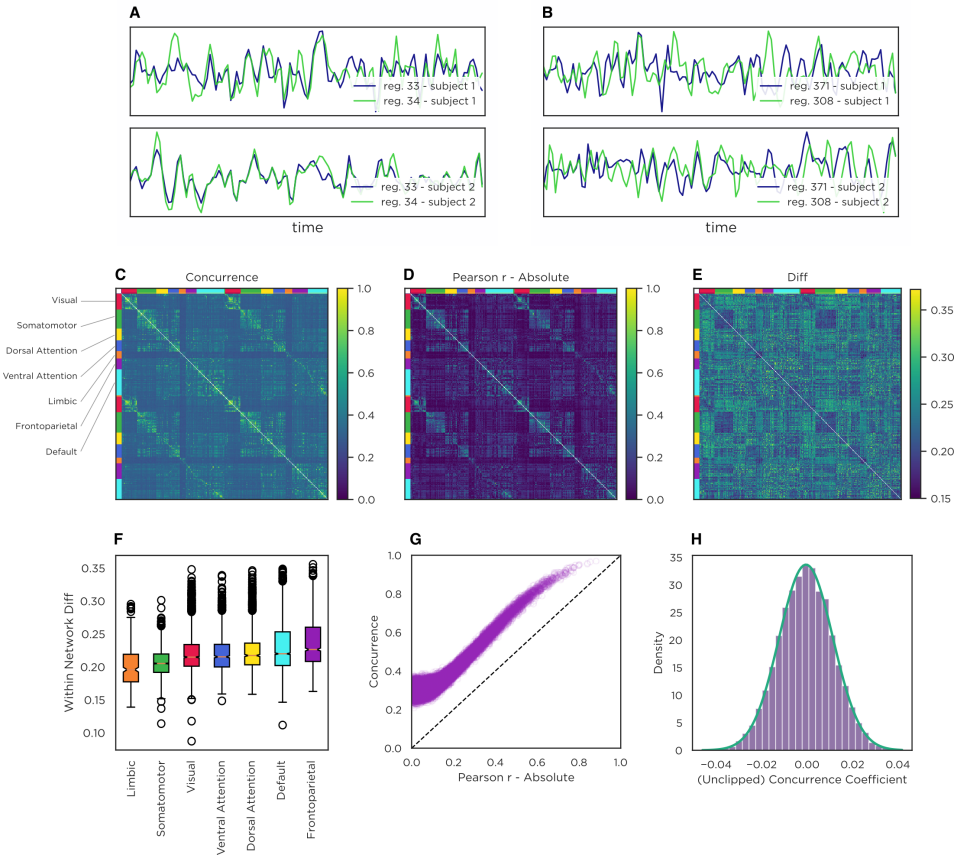

Figure 3: (A) Signals from two brain regions with correlated fMRI signals. (B) Signals from two brain regions with fMRI signals that are dependent (concurrence coefficient: 0.25) but approximately uncorrelated (Pearson's r: 0.02). (C) The connectivity matrix computed with the concurrence coefficient. (D) The connectivity matrix computed with (absolute) correlation values (Pearson's r). (E) The difference between the two concurrence- and correlation-based connectivity matrices. (F) The distribution of the difference between the concurrence- and correlation-based connectivity matrices, shown separately for the seven brain networks. (G) Alternative comparison of the Pearson's r vs. concurrence coefficients computed from all the brain region pairs. (H) The (unclipped) concurrence coefficient between 10,000 pairs of brain regions of mismatched participants.

for data collection ranged from 4 to 7 minutes. The data were split into four subject-independent cross-validation folds. The segment size w was equivalent to 5 seconds.

The average concurrence coefficient on the test folds was 0.50 ($p<0.001$), indicating that the concurrence approach successfully detects the relationship between respiration rate and ECG signals. The PSCS can generally distinguish between compared signals that are temporally aligned or not (Fig. 4B–D), validating that concurrence can identify relationships (or lack thereof) that are difficult to determine visually. Fig. 4E plots the PSCSs from temporally aligned segments vs. the root mean square of successive differences (RMSSD) derived from the ECG signal of each interaction. That the PSCS is generally larger when the RMSSD is low may suggest that the trained algorithm predicts a stronger relationship between ECG and respiration rate when the latter is increased.

3.3 Behavioral Data

Finally, we apply the concurrence method to the analysis of facial and head movement behavior occurring in the context of a dyadic conversation task. The social behaviors of two conversation partners are expected to be dependent, due to well-established phenomena like nonconscious mimicry [29] or (nonverbal) backchanneling [30]. However, determining whether and to what degree behavioral

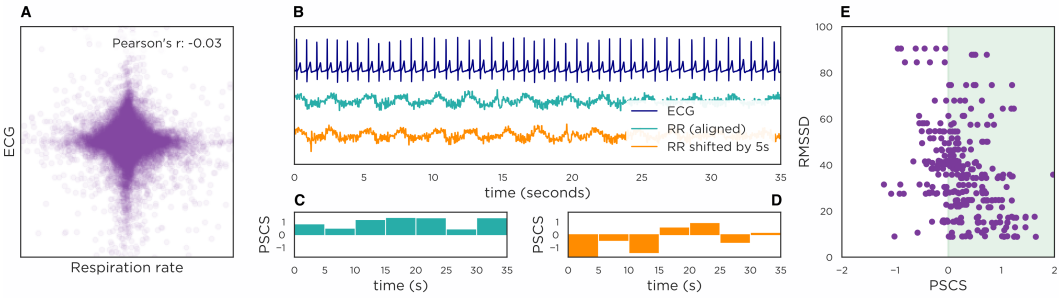

Figure 4: (A) Scatter plot and correlation between the respiration rate (RR) and ECG signal. (B) A sample ECG signal plotted against the synchronized (i.e., time-aligned) RR signal and the temporally misaligned RR. (C) The per-segment concurrence scores (PSCSs) between the temporally aligned ECG and RR are positive, which indicate that the PSCS correctly predicts that the segments are temporally aligned. (D) The PSCSs between the temporally misaligned ECG and RR are generally negative. (E) The PSCS for (temporally aligned) ECG and RR signals against the RMSSD, computed on a dataset of 30 participants for multiple segments per participant.

signals are dependent has proven challenging, as behavior is captured with multi-dimensional signals (Fig. 5A,B), and any subset of signals from one conversation partner may depend on the signals of the other partner through an unknown and possibly non-linear relationship.

We conduct experiments on a dataset of 199 participants (aged 5 to 40 years) engaged in a 3-5-minute semi-structured face-to-face conversation [43]. We quantify social behaviors (i.e., facial expressions and head movements) in each conversation partner with 82-dimensional signals (79 for facial expressions and 3 for head movements) [44]. The concurrence coefficient for $w = 4$ seconds is 0.49 ($p<0.001$), indicating that the behavior signals of the conversation partners are dependent. Moreover, the PSCS allows us to investigate differences within different subsamples. For example, Fig. 5C shows that the PSCS increases with age (Spearman's $r = 0.61$, $p < 0.001$), indicating that younger school-age participants tend to have less behavioral coordination with a conversation partner than older participants. Additional analyses on a subsample of 12-18 year-olds (N=42) with and without an autism diagnosis (matched on age and sex) indicate that autistic adolescents have reduced coordination with conversation partners relative to neurotypical adolescents (Cohen's D: 0.8; $p = 0.003$; Fig. 5D). Together, these results demonstrate that the concurrence method exposes clinically relevant differences in spontaneous social behavioral coordination, without any a priori information about the structure of the coordination.

4 Limitations of Measuring Dependence with Self-supervision

While the proposed method showed no propensity to discovering spurious dependencies (Fig. 2), it is not uncommon that signals generated for scientific analyses contain responses to common, extraneous events. These common events render the compared signals statistically dependent, yet since they are extraneous, they are often unrelated to the research question that is being investigated.

This phenomenon can be observed, for example, in a comparison of EEG signals versus facial behavior. While one may be interested in analyzing these two modalities to uncover potential relationships between mental states and facial behavior, the EEG sensors may be sensitive to activities that are not related to the brain. For example, eyeblinks are known to generate powerful electrical charges that leave a strong signature on the EEG signals. While pre-processing algorithms may be able to reduce the effect of blinking, it is possible that they leave a residue that can still be detected by a potent algorithm for detecting statistical dependencies.

An alternative approach for accounting extraneous factors is using correction procedures. These procedures are typically used with traditional approaches to measuring statistical dependencies, and their success is yet to be demonstrated for the proposed concurrence algorithm, or alternative algorithms based on (deep) learning, which can expose more nuanced dependencies that can remain on the compared variables even after correction.

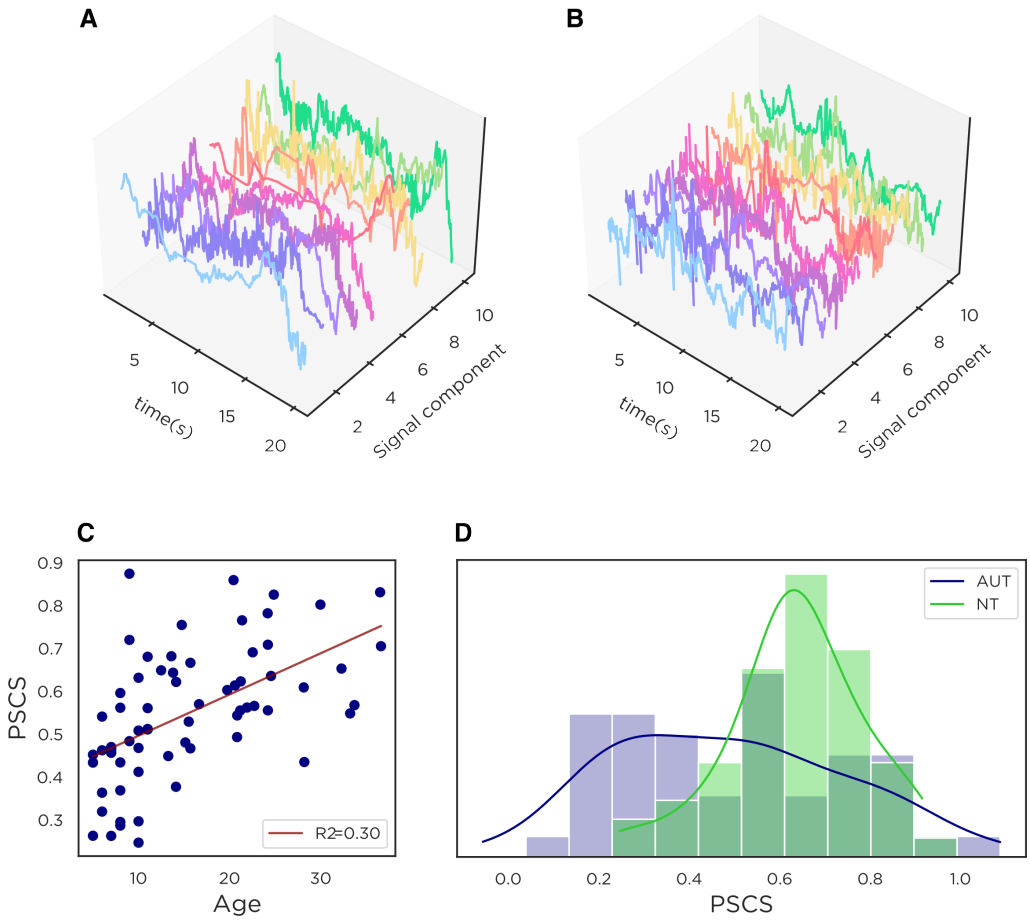

Figure 5: Results from experiments with (multi-dimensional) behavioral signals. (A) Facial behavior signals from a study participant in a dyadic conversation. For clarity, we show only 10 out of the 82 signal components. (B) Facial behavior signals of the conversation partner of the participant in (A). (C) The average per-segment concurrence score (PSCS) per neurotypical (NT) individual versus age, indicating that social coordination increases with age (Spearman's r = 0.61, $p < 10^{-8}$). (D) The distribution of average PSCS per individual is lower with autism (AUT) compared to the NT group, indicating reduced social coordination in AUT.

5 Conclusion

This paper introduced a new approach for measuring statistical dependence, namely, concurrence. We showed that this self-supervised can become a standard way of quantifying statistical dependence between time series, as it readily detects a wide range of linear or non-linear dependencies with an off-the-shelf implementation, even from modestly sized samples and noisy data, without requiring empirical (hyper)parameter tuning; and showed no propensity to false discoveries (Type I errors). Future research can further enhance this framework by theoretically establishing the link between statistical dependence and concurrence, while integrating advances in machine learning can ensure that its theoretical potential can be fully actualized. Ensuring whether the statistical dependencies exposed by the proposed method remains an open problem, therefore the users of the method should ensure that the signals do not contain common responses to extraneous factors, or they should perform post-hoc analyses to ensure that the dependencies that are exposed truly pertain to the research question of interest.

References

[1] Dag Tjøstheim, Håkon Otneim, and Bjørn Støve. Statistical dependence: Beyond pearson's ρ. *Statistical Science*, 37, 2022.

[2] Natalia B. Janson. Non-linear dynamics of biological systems. *Contemporary Physics*, 53: 137–168, 2012.

[3] Fei He and Yuguo Yang. Nonlinear system identification of neural systems from neurophysiological signals. *Neuroscience*, 458:213–228, 2021.

[4] Richard C. Schmidt and Barbara O'Brien. Evaluating the dynamics of unintended interpersonal coordination. *Ecological Psychology*, 9:189–206, 1997.

[5] Alexandre Hyafil, Anne-Lise Giraud, Lorenzo Fontolan, and Boris Gutkin. Neural cross-frequency coupling: Connecting architectures, mechanisms, and functions. *Trends in Neurosciences*, 38:725–740, 2015.

[6] Chung-Han Yeh, Cheng Zhang, Weidong Shi, Meng-Ting Lo, Gerd Tinkhauser, and Ashwani Oswal. Cross-frequency coupling and intelligent neuromodulation. *Cyborg and Bionic Systems*, 4:0034, 2023.

[7] Edward Beltrami and John Jesty. Mathematical analysis of activation thresholds in enzyme-catalyzed positive feedbacks: application to the feedbacks of blood coagulation. *Proceedings of the National Academy of Sciences of the USA*, 92:8744–8748, 1995.

[8] Paul H. Tiesinga and Terrence J. Sejnowski. Mechanisms for phase shifting in cortical networks and their role in communication through coherence. *Frontiers in Human Neuroscience*, 4:196, 2010.

[9] Claudio Cosentino and Declan Bates. *Feedback Control in Systems Biology*. CRC Press, 2011.

[10] Zhilin Qu, Alan Garfinkel, James N. Weiss, and Mikko Nivala. Multi-scale modeling in biology: How to bridge the gaps between scales? *Progress in Biophysics and Molecular Biology*, 107: 21–31, 2011.

[11] Alfréd Rényi. On measures of dependence. *Acta Mathematica Hungarica*, 10:441–451, 2005.

[12] Arthur Gretton, Kenji Fukumizu, Choon H. Teo, Le Song, Bernhard Schölkopf, and Alexander J. Smola. A kernel statistical test of independence. In *Advances in Neural Information Processing Systems 20 (NIPS 2007)*, 2007.

[13] Gábor J. Székely, Maria L. Rizzo, and Nail K. Bakirov. Measuring and testing dependence by correlation of distances. *The Annals of Statistics*, 35:2769–2794, 2007.

[14] Dino Sejdinovic, Bharath Sriperumbudur, Arthur Gretton, and Kenji Fukumizu. Equivalence of distance-based and rkhs-based statistics in hypothesis testing. *The Annals of Statistics*, 41: 2263–2291, 2013.

[15] Jakob Verbeek, Sam Roweis, and Nikos Vlassis. Non-linear cca and pca by alignment of local models. In *Advances in Neural Information Processing Systems 16*. MIT Press, 2003.

[16] Galen Andrew, Raman Arora, Jeff Bilmes, and Karen Livescu. Deep canonical correlation analysis. In *Proceedings of the 30th International Conference on Machine Learning (ICML)*, 2013.

[17] Wei-Yi Hua and Debashis Ghosh. Equivalence of kernel machine regression and kernel distance covariance for multidimensional phenotype association studies. *Biometrics*, 71:812–820, 2015.

[18] Arthur Gretton, Dino Sejdinovic, Heiko Strathmann, Sivaraman Balakrishnan, Massimiliano Pontil, Kenji Fukumizu, and Bharath K. Sriperumbudur. Optimal kernel choice for large-scale two-sample tests. In *Advances in Neural Information Processing Systems 25*. Curran Associates, Inc., 2012.

[19] Xiaoqin Zhuang, Zening Yang, and Dietmar Cordes. A technical review of canonical correlation analysis for neuroscience applications. *Human Brain Mapping*, 41:3807–3833, 2020.

[20] Scott Marek, Brenden Tervo-Clemmens, Finnegan J. Calabro, Daniel F. Montez, Benjamin P. Kay, Alexander S. Hatoum, Mary R. Donohue, William Foran, Ryan L. Miller, Timothy J. Hendrickson, Shane M. Malone, Suyash Kandala, Eric Feczko, Oscar Miranda-Dominguez, Amy M. Graham, Eric A. Earl, Alex J. Perrone, Miguel Cordova, Owen Doyle, Lily A. Moore, Grace M. Conan, Juan Uriarte, Kyle Snider, Benjamin J. Lynch, Jordan C. Wilgenbusch, Taylor Pengo, Angela Tam, Jiayu Chen, D. J. Newbold, Alexander Zheng, Nicholas A. Seider, Ahria N. Van, Asuka Metoki, R. J. Chauvin, Timothy O. Laumann, Deanna J. Greene, Steven E. Petersen, Hugh Garavan, Wesley K. Thompson, Thomas E. Nichols, B. T. T. Yeo, Deanna M. Barch, Beatriz Luna, Damien A. Fair, and N. U. F. Dosenbach. Reproducible brain-wide association studies require thousands of individuals. *Nature*, 603:654–660, 2022.

[21] Robert E. Greenblatt, Mary E. Pflieger, and Alexei E. Ossadtchi. Connectivity measures applied to human brain electrophysiological data. *Journal of Neuroscience Methods*, 207:1–16, 2012.

[22] Katsunori Fujiwara and Ikuo Daibo. Evaluating interpersonal synchrony: Wavelet transform toward an unstructured conversation. *Frontiers in Psychology*, 7, 2016.

[23] Richard C. Schmidt, Stephanie Morr, Paula Fitzpatrick, and Michael J. Richardson. Measuring the dynamics of interactional synchrony. *Journal of Nonverbal Behavior*, 36:263–279, 2012.

[24] Stéphane Mallat. *A Wavelet Tour of Signal Processing: The Sparse Way*. Academic Press, 3 edition, 2009.

[25] Martin Vetterli, Jelena Kovačević, and Vivek K. Goyal. *Foundations of Signal Processing*. Cambridge University Press, 2014.

[26] David Maraun and Jürgen Kurths. Cross wavelet analysis: significance testing and pitfalls. *Nonlinear Processes in Geophysics*, 11:505–514, 2004.

[27] Mark A. Kramer, Adriano B. L. Tort, and Nancy J. Kopell. Sharp edge artifacts and spurious coupling in eeg frequency comodulation measures. *Journal of Neuroscience Methods*, 170: 352–357, 2008.

[28] Maria A. Russo, David M. Santarelli, and David O'Rourke. The physiological effects of slow breathing in the healthy human. *Breathe*, 13:298–309, 2017.

[29] Jessica L. Lakin, Victoria E. Jefferis, C. M. Cheng, and Tanya L. Chartrand. The chameleon effect as social glue: Evidence for the evolutionary significance of nonconscious mimicry. *Journal of Nonverbal Behavior*, 27:145–162, 2003.

[30] Lisa Shelley and Frank Gonzalez. Back channeling: Function of back channeling and l1 effects on back channeling in l2. *Linguistics Portfolio*, 2, 2013.

[31] Ding-Xuan Zhou. Universality of deep convolutional neural networks. *Applied and Computational Harmonic Analysis*, 48:787–794, 2020.

[32] Zhiguang Wang, Weizhong Yan, and Tim Oates. Time series classification from scratch with deep neural networks: A strong baseline. arXiv:1611.06455, 2016. URL `https://doi.org/10.48550/arXiv.1611.06455`.

[33] Shaojie Bai, J. Zico Kolter, and Vladlen Koltun. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. arXiv:1803.01271, 2018. URL `https://doi.org/10.48550/arXiv.1803.01271`.

[34] Marijke Welvaert and Yves Rosseel. On the definition of signal-to-noise ratio and contrast-to-noise ratio for fmri data. *PLoS ONE*, 8:e77089, 2013.

[35] Thomas E. Nichols and Andrew P. Holmes. Nonparametric permutation tests for functional neuroimaging: a primer with examples. *Human Brain Mapping*, 15:1–25, 2002.

[36] Anderson M. Winkler, Gerard R. Ridgway, Gwenaël Douaud, Thomas E. Nichols, and Stephen M. Smith. Faster permutation inference in brain imaging. *NeuroImage*, 141:502–516, 2016.

[37] F. Kazi-Aoual, S. Hitier, R. Sabatier, and J.-D. Lebreton. Refined approximations to permutation tests for multivariate inference. *Computational Statistics & Data Analysis*, 20:643–656, 1995.

[38] Zheng-Quan Liu, Andrea I. Luppi, Jesper Y. Hansen, Yue E. Tian, Andrew Zalesky, B. T. Thomas Yeo, Ben D. Fulcher, and Bratislav Mišić. Benchmarking methods for mapping functional connectivity in the brain. bioRxiv 2024.05.07.593018, 2024. URL `https://doi.org/10.1101/2024.05.07.593018`.

[39] Graham L. Baum, Zhen Cui, David R. Roalf, Rastko Ciric, Richard F. Betzel, Barron Larsen, Matthew Cieslak, Philip A. Cook, Chun H. Xia, Tyler M. Moore, Kosha Ruparel, David J. Oathes, Aaron F. Alexander-Bloch, Russell T. Shinohara, Armin Raznahan, Raquel E. Gur, Ruben C. Gur, Danielle S. Bassett, and Theodore D. Satterthwaite. Development of structure–function coupling in human brain networks during youth. *Proceedings of the National Academy of Sciences*, 117:771–778, 2020.

[40] Theodore D. Satterthwaite, John J. Connolly, Kosha Ruparel, Monica E. Calkins, Chad Jackson, Mark A. Elliott, David R. Roalf, Rachel Hopson, Karthik Prabhakaran, Melissa Behr, Hong Qiu, Frederick D. Mentch, Rosetta Chiavacci, Patrick M. A. Sleiman, Ruben C. Gur, Hakon Hakonarson, and Raquel E. Gur. The philadelphia neurodevelopmental cohort: A publicly available resource for the study of normal and abnormal brain development in youth. *NeuroImage*, 124:1115–1119, 2016.

[41] Alexander Schaefer, Ru Kong, Evan M. Gordon, Timothy O. Laumann, Xi-Nian Zuo, Avram J. Holmes, Simon B. Eickhoff, and B. T. Thomas Yeo. Local-global parcellation of the human cerebral cortex from intrinsic functional connectivity mri. *Cerebral Cortex*, 28:3095–3114, 2018.

[42] Edgar D. Adrian, Detlev W. Bronk, and G. Phillips. Discharges in mammalian sympathetic nerves. *The Journal of Physiology*, 74:115–133, 1932.

[43] Evangelos Sariyanidi, Casey J. Zampella, Evan DeJardin, John D. Herrington, Robert T. Schultz, and Bünyamin Tunç. Comparison of human experts and ai in predicting autism from facial behavior. In *CEUR Workshop Proceedings*, volume 3359, pages 48–57, 2023.

[44] Evangelos Sariyanidi, Casey J. Zampella, Robert T. Schultz, and Bünyamin Tunç. Inequality-constrained 3d morphable face model fitting. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46:1305–1318, 2024.

[45] Xinyu Zhang, Zhen Zhao, Theodoros Tsiligkaridis, and Marinka Zitnik. Self-supervised contrastive pre-training for time series via time-frequency consistency. In *Advances in Neural Information Processing Systems 35*, 2022.

[46] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, and Ilya Sutskever. Learning transferable visual models from natural language supervision. arXiv:2103.00020, 2021. URL `http://arxiv.org/abs/2103.00020`.

[47] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Advances in Neural Information Processing Systems 30*, 2017.

[48] Yuxuan Liang, Haixu Wen, Yuxuan Nie, Yuxuan Jiang, Mingkun Jin, Dawei Song, Shirui Pan, and Qingsong Wen. Foundation models for time series analysis: A tutorial and survey. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 6555–6565. Association for Computing Machinery, 2024. doi: 10.1145/3637528.3671451.

Kernel size of first conv. layer	5
Kernel size of other conv. layers	3
Step size (stride) at conv. first layer	3^*
Step size (stride) at other conv. layers	2^*
Number of blocks (B)	3
Number of output channels at 1st conv. layer (C)	512
Number of output channels at bth conv. layer	$512/2^{(b-1)}$
Dropout rate	0.25
Optimizer	Adam[51]
Number of iterations	100
Learning rate	10^{-4}

Table 1: Parameters of the CNNs that we use. *The step sizes larger than 1 effectively downsample the input, and if the segment size w is too small, the output of a convolutional kernel may be empty. To avoid the latter, one may need to reduce the stride sizes accordingly.