

Nonlinear Analysis of Nonverbal Behavior

Synchronization

Howon Lee

June 5, 2015

1 Abstract

There exists a literature dealing with time series analysis in the nonlinear sciences in the case of synchronization, or the adjustment of rhythms of oscillating objects to coincide with each other due to weak interaction. We attempt some of these methods in analyzing an instance of the phenomenon of interpersonal synchrony in human movement data. We also propose a mapping from time series to networks with a deterministic inverse, useful for exploratory analysis.

2 Introduction

Before defining what we mean by synchronization in this context, we note that there exists a long history of the study of synchronization outside of social science. Synchronization was first reported in physical systems by Huygens in 1665. He saw that two pendulum clocks in the boat that he was testing the clocks in were mysteriously ticking in unison, and found that the cause was a support beam which connected the clocks, which moved imperceptibly and therefore maintained the synchronization[2].

Centuries later, the physical and biological systems which were later found to synchronize include mechanical rotators, lasers, fireflies emitting synchronous patterns of light, synchronous contraction of heart cells, synchronization of human circadian rhythm to a solar light cycle, and synchronous interaction of human beings[1], [2].

Now, we should define the terms. Interpersonal synchrony is defined in the social science literature as individuals' temporal coordination during social interaction[3]. In the physical sciences, synchronization is defined as an adjustment of rhythms of oscillating objects due to weak interaction, with generalizations possible for chaotic systems, which depend upon a phase of the chaotic system existing[2].

Although social synchrony is one of many other synchronies, many of which do not necessarily serve a function of any kind, it is observed that social synchrony serves a function in human social groups. There exists evidence that synchronization acts as a cooperation-inducing mechanism[5], that it acts to induce rapport[6], and that it in actuality, independent of its effect on rapport, enhances the ability to pursue joint goals in tandem for those who are synchronized[7]. This suggests immediately that it must be measured in a systematic way, to study phenomena of that kind.

In the beginning of the social synchronization literature, most psychologists did not use the then-developing automated signals processing techniques for the detection and measurement of synchronization, instead using manual methods to detect and rate the presence or absence of synchronization, with trained raters and validated measuring systems[10]. Although these measurements have been validated[11], they depend upon human raters and therefore are less replicable and less convenient than automated systems.

Given that the signals created by an individual during social interaction have a phase, and social interaction can be construed as a weak interaction between the two individual systems, it should be clear that the definition of interpersonal synchrony given in the social psychological literature is a subset of the definition given in the physics literature. This has often been noted[3], and has therefore spawned a cross-disciplinary field wherein signals

processing techniques are used to measure interpersonal synchrony.

As a specific instance of a domain where signals processing techniques are used, a large problem in synchronization is the definition of the signal itself and its extraction from observations of social interaction. To this end, many methods have been used, including extraction of coordination of movement features and speech features[?], movement of single and dyadic body parts[?], image processing techniques[?] and video tracking techniques[?].

An important analogy exists between the collection of time series data about the physiology of the body in order to assess synchrony and the collection of time series data about the body in order to assess health. Indeed, there exists a literature on the synchronization properties and the time series analysis of heart rate variability which we have derived much inspiration from, not necessarily in the usage of correlational metrics, for which there is a literature dealing in social synchronization, but more sophisticated metrics for which there is not such a literature[15][16]. The essential premise is that we treat social synchronization as an instance of an abstract problem of physical synchronization, which justifies using measures used by other physiological measurements.

It is also important to note that there exists a literature which uses various nonlinear methods to track nonverbal synchronization. Recurrence analysis, which is also from nonlinear dynamical systems, provides a graphical representation of the coupling of systems represented by time series, and is used by [33], [34] and [35] for detecting postural synchronization. [36] modelled the interpersonal coordination of two people rocking together on rocking chairs side-by-side as a coupled dynamical system.

This project will attempt to use some of the already existing tools for the analysis of time series data on nonverbal tracking time series data, as well as apply some tools which have not previously been used to analyze synchronization in such time series data of social synchronization in a virtual world.

We can take the problem at hand as trying to find hidden structure in the unlabeled time series we are given, where the hidden structure is any present synchrony. Therefore, this can

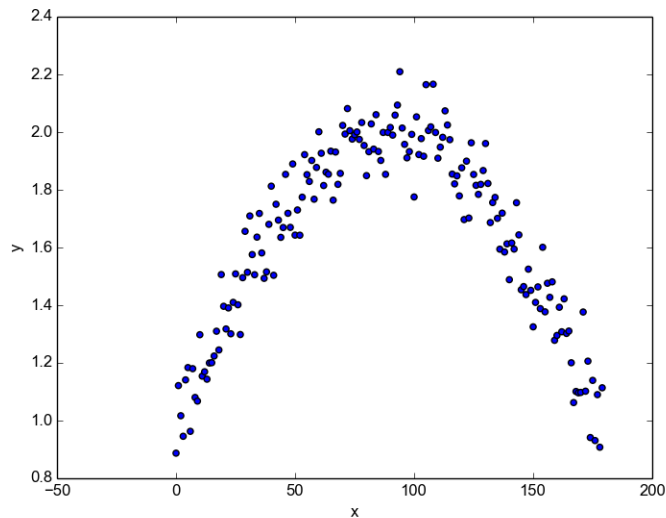


Figure 1: Example of simple nonlinear relation between two variables

also be construed as an unsupervised learning problem.[3].

One might wonder why the existing methods of measurement of synchronization do not already serve for the purpose of measuring synchronization of body movements. They do not serve for the same reasons that they do not serve in physical time series. That is, they are fundamentally *linear* measures of the relation that they have to each other, meaning that measuring only correlation implicitly puts in the assumption that an affine transform relates the two signals to each other. It is possible to learn about higher order dependences if correlations of higher moments are used, but this becomes rapidly time-consuming very quickly[13].

One example from [30] which should help develop a theoretical intuition about nonlinearity and synchronization is the sine wave relation example. Consider the figure 2. According to a linear correlation measure, taking the whole domain of x into account at once, the two variables x and y cannot disprove the null hypothesis of independence ($p = 0.45$), but by inspection, one can say that actually $y = \sin(x) + \epsilon$, where ϵ is a small amount of Gaussian noise. But nonlinear measures *will* take this nonlinear relation into account. For example, mutual information will say that the mutual information between the two variables x and y

is 7.5 bits, which indicates that the variables determine each other to a great degree.

In the psychological literature dealing with referential behavior, or behavior which adapts to some process or set of events or plan [19], there exist two theoretical schools towards the modelling of the feedback systems which are presumed to allow synchronization with respect to a referent: information processing theories and dynamical systems theories. Information processing theories tend to deal with participant responses as discrete time series, and fit to data which is compatible with this paradigm (tapping studies, for example). Dynamical systems theories tend to deal with participant responses as trajectories in a phase space and deal with continuous movements as data [1].

There exist unifications [19] which use the formalism of control theory to include both sorts of models as special cases. In that same paper, some evidence of nonlinearity (specifically, difficulty finding a low correlation dimension and a small but positive Lyapunov exponent, with the caveat that the cardinality of the data was small) which could not be fitted with a linear autoregressive moving average approach was found. [21] notes that there are overfitting difficulties with manually incorporating a nonlinear perception detection threshold into a second-order autoregressive moving average model which is to be fit to data.

However, in the *measurement* of degrees of synchronization, or of the experimental discernment of synchronization state, it does not necessarily follow that avoiding the assumption of linearity is as intractable as fitting a nonlinear *model* would be.

Lagged correlation is the most common measure of synchrony in social science. We will examine those in this project as a first pass. Then, we will go into recurrence methods, spectral coherence, Hilbert transforms and transformation into networks.

2.1 Data Description

Much of the data is from a previous study in the lab,

— don't have a real description of the data handy! somewhat of an oversight

Notably, to reduce the data and to avoid the curse of dimensionality in some of our

analyses, a preprocessing step was done:

A problem with the data is that it was missing in places. Because of the differencing in preprocessing which we did, an interpolation would not have been prudent, because an interpolation would have claimed that the original time series had been steadily increasing, which is not necessarily warranted. Therefore the missing data was construed as 0's, because this would be a simple interpolation in the original time series, before preprocessing.

3 Time Domain Analysis

3.1 Correlation Analysis

When we say *correlation*, in most cases, we mean *Pearson's correlation*, or the *Pearson product-moment correlation coefficient*.

Three fundamentally correlational analyses were tried: time-lagged cross-correlation, Poincare plots and an ad hoc method based on taking the slopes of the change of shape of the data-containing ovals of Poincare plots.

One analysis commonly used with relaxation oscillators (like ECG data) is the Poincare map, not to be confused with the Poincare plot. However, because there are not clear demarcations as to where each oscillation is, the Poincare map is less suited to this dataset and will not be used[14].

3.1.1 Time-Lagged Cross-Correlation

This part is a re-analysis of the work from [4], but as stated before, only using the head angle. As in [4], Pearson's correlation was taken on a window of 400 frame data points(50 seconds) and then shifted for consecutive lag times according to [32].

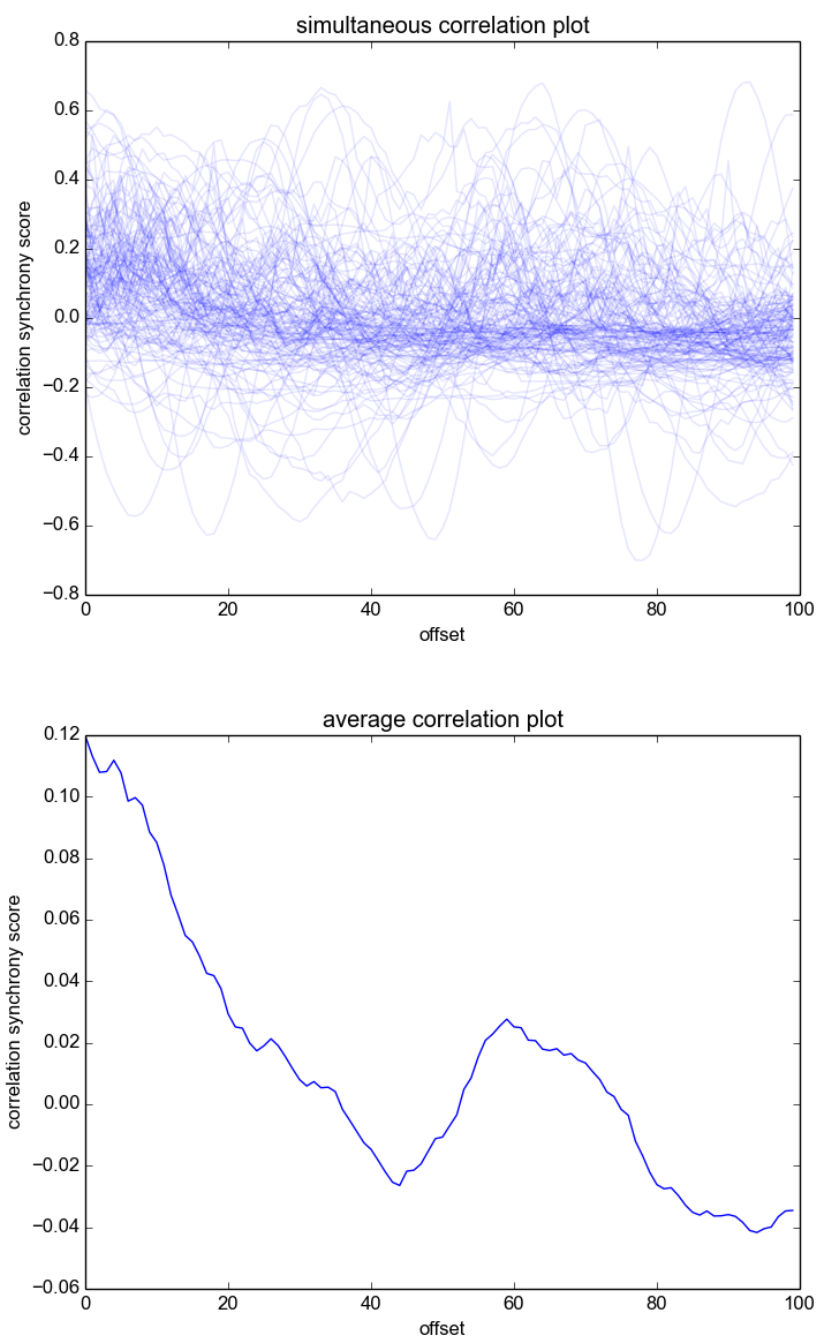


Figure 2: Lagged correlation of all data and pointwise averages of lagged correlation

Figure 3: Example Poincare plots from real data

3.1.2 Poincare Plots

3.1.3 Ad Hoc Method

3.2 Mutual Information Analysis

Mutual information is important as a measure because it measures the lack of independence of two random variables, as opposed to correlation, which merely measures whether two random variables covary linearly, and therefore is not sensitive to more complex relations as mutual information is.

In order to introduce mutual information, we must introduce Shannon information and Shannon entropy[28].

Shannon information is defined on a measurement drawn from some set (let the set be discrete, $X \in 1...n$, for ease of notation: there exist continuous analogues). Given that measurement from some set, the Shannon information in bits is defined as:

$$-\log_2 P_X(i)$$

where $P_X(i)$ is the probability that a measurement will find the system in state i . That is, a lower probability answer is more surprising, and therefore more information is found in it.

Shannon entropy, denoted $H(X)$ is defined on a data stream (a measurement stream) X . It is the average amount of information contained in every message received, and so is also measured in bits. An easy intuition about entropy can be gained by noting that a *fair* coin flipping process is the most entropic possible coin flip, because we are most surprised by it. Formally, entropy is:

$$H(X) = - \sum_i P_X(i) \log_2 P_X(i)$$

Information and entropy are also defined on joint probability distributions in an analogous

way. The information of a joint probability distribution on variables X and Y is:

$$-\log_2 P_{XY}(i, j)$$

And the entropy, therefore, is:

$$H(X, Y) = - \sum_i \sum_j P_{XY}(i, j) \log_2 P_{XY}(i, j)$$

Now, if we sample two random variables simultaneously, the relationship has a quantity defined upon it which is called the *mutual information*. Because it is defined on the relation, it has an auto-mutual information variant and a cross-mutual information variant, like correlation does. However, mutual information measures something different, which is the *information which one of the random variables gives about the other random variable*. It is defined as:

$$I(X, Y) = H(X) + H(Y) - H(X, Y)$$

And one sees that this quantity is symmetric with respect to X and Y . Note that two linearly related variables will have a high mutual information *in addition* to a high correlation coefficient.

Because we are dealing with real data, mutual information must be estimated from probability distributions represented as histograms. This is problematic in high-dimensional data, owing to the curse of dimensionality[22]. Also, it is the case that the measure itself depends upon an intelligent discretization, which does not make the individual buckets of the histogram too sparse with data[29]. We discretized with 64 buckets, which was recommended for a time series similar in size and range as ours by [29].

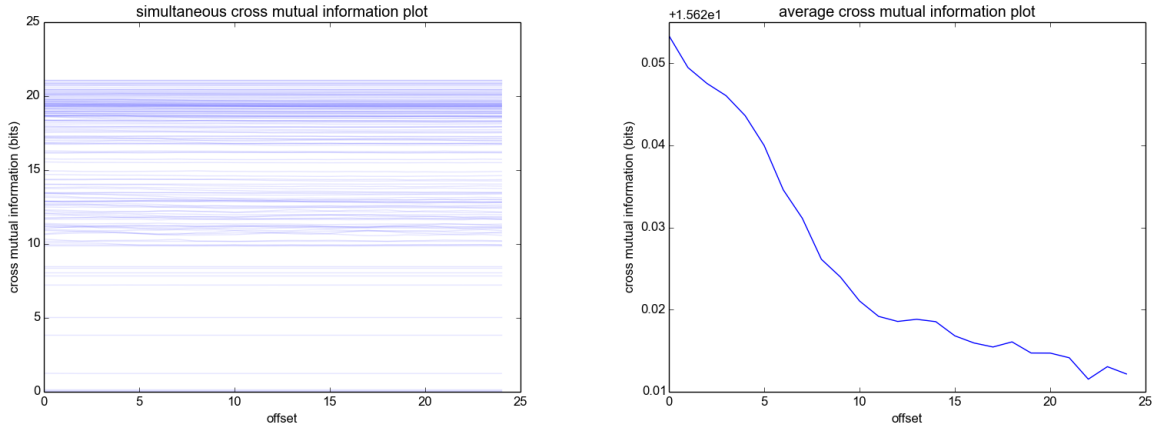


Figure 4: Lagged cmi

3.2.1 AMI and CMI

Auto mutual information (AMI) and cross mutual information (CMI) are the mutual information analogs to autocorrelation and cross correlation. They have the same advantages over the correlation variants as mere mutual information itself has to correlation. Like lagged correlation, mutual information is symmetric with respect to the compared time series.

As can be seen, there exist clear patterns which stay constant over the lagging process of which time series are more close to each other with cross-mutual information.

4 Frequency Domain Analysis

There is more than one way to represent time series data, even given only the data itself and not any underlying function that generates it. The most important alternate representation is the frequency domain representation, which is taken in finite data with the discrete Fourier transform. Later, we will show the use of the frequency representation in re-representing our data so that it shows a numerical current state of the phase, using the related Hilbert transform.

The frequency domain represents data in terms of the linear combination of simple sine

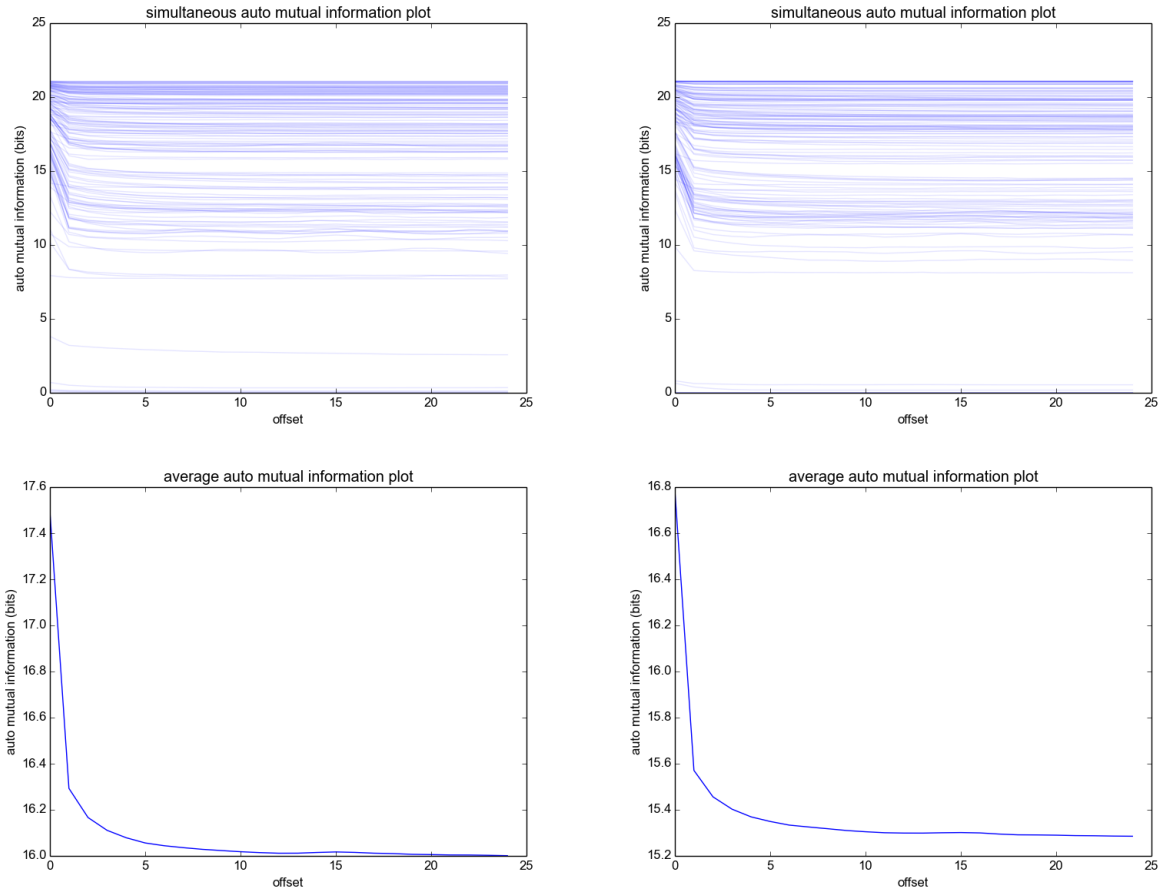


Figure 5: Lagged amis: from left to right, west and north, west and north simultaneous and average plot

and cosine waves: the analogy often used is that of a chord being hit on a piano, which can be represented with its time domain signal but can also be represented as the amplitude of the individual notes within the chord, which presents a better decomposition of the sound in and of itself. Likewise, we can attempt to decompose our time domain signals into the frequencies of the sinusoids which make it up.

The discrete Fourier transform is an essential tool for the further analyses below. It converts a finite discrete sample from the time to the frequency domain in a computable and fast way.

A measure of synchronization in the frequency domain is coherence, which is also a linear measure of synchronization, but not in the same basis as correlation.

4.1 Hilbert Transform and γ

A signal is called an *analytic signal* if it does not have any negative-frequency components. The Hilbert transform is a way to get the analytic *representation* for a signal in the time domain, which throws away the superfluous information which would be in a spectrum of that signal. This allows the expression of the analytic signal in terms of its time-variant magnitude and phase, allowing a study of the phase in an *unwrapped form*[18]. Why this is important is that it allows the measurement of phase synchronization, or synchronization of phase alone, apart from the more general phenomena of synchronization.

Formally speaking, given that the original signal is $s(t)$, the analytic signal ζ is:

$$\zeta(t) = s(t) + js_H(t) = A(t)e^{j\phi(t)}$$

$$s_H(t) = \frac{1}{\pi}PV \int_{-\infty}^{\infty} \frac{s(\tau)}{t - \tau} d\tau$$

s_H is called the Hilbert transform; PV indicates that the integral is a Cauchy principal value integral. The Hilbert transform is thus formally equivalent to convolution of $s(t)$ with

$\frac{1}{\pi t}$, so in practice, it can easily be calculated from the fast Fourier transform.

5 Deterministic dual between discrete time series and ordered multigraph

5.1 Description

There exists a line of analysis in the complex network literature which aims to convert a digital time series into a network. This is advantageous for analysis of time series because there is a deep and well-developed theory of networks: so far, an important result is that different time series result in networks with distinct topological properties, and that these topological properties relate to the time series in some way in addition to being determined by them[8].

These methods are based upon a variety of features of the time series, such as correlational structure[25], visibility[26], phase space reconstruction[27] and many others, which all attempt to differentiate time series by mapping them into networks with divergent properties. Notably, the degree distribution of the visibility graph is claimed in [26] to be related directly to the Fourier power spectrum of the time series, because the visibility graph depends upon the relation of peaks of the time series to one another.

Upon hearing of the mapping from a time series into a network, it might be wondered at if there exists a mapping from a network back into a time series. There does exist some methods to apply the inverse mapping, but none of these are deterministic, meaning that the network topology constrains the set of time series that the mapping can produce from that network, but does not completely determine it[8]. This is an undesirable property because the time series cannot be deterministically reproduced from the graph: what is reproduced is only a time series with statistical properties akin to that of the previous time series.

A deterministic transformation from time series to networks and from networks to time

series can be made, however, if we allow duplication in the adjacencies of each vertex of the graph (making the graph a multigraph, by many definitions of multigraph[9]), and if we keep the order in which the vertices were travelled in the discrete time series. In addition to this, the start state of the time series must be stored.

In order to construct the labelled multigraph from a time series, the possible states of the time series, $\Sigma = 1...N$ are made into the vertices of the multigraph. Then, starting with the value of the series at $t = 1$, the vertices are traversed in the order that the corresponding states are traversed in the time series, and when one leaves a vertex V_1 to go to a vertex V_2 , the identity of V_2 is pushed onto the list (like a stack) for the adjacencies of V_1 (5.1). There is pseudocode for this algorithm in 5.1

In order to reconstruct the time series from a multigraph of this kind, therefore, we start from the start node again and consume the adjacency lists starting from the initial elements of the lists, to find the next node to jump to: this is detailed in 5.1.

The model is very reminiscent of models from symbolic dynamics, just as [8] mentioned that their analysis was reminiscent of models from symbolic dynamics, because a continuous system is being discretized into a sequence of symbols representing the state of the system, which turn into nodes in a graph. Therefore, another important consideration is the size of the buckets which the time series is to be discretized into. For the examples below we chose bucket size of 0.05 to discretize the time series, but this was not a principled decision and other methods, such as nearest neighbors, might be investigated, in an analogous way to the discretization of time series for determination of mutual information[28].

5.2 Implementation

We will investigate the network properties of three toy models of a one-dimensional time series: a simple sinusoid, a chaotic logistic map, and a fractional Brownian motion. We will also investigate them on an exemplary time series from the dataset.

The concept of a node degree has a natural analog in the stack graph, where the outdegree

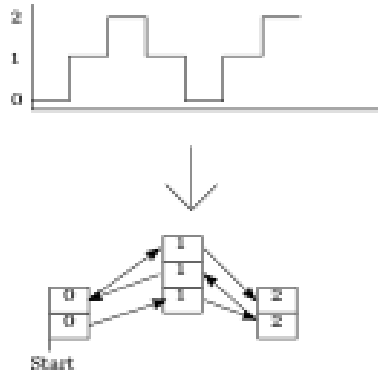


Figure 6: Example construction of stack graph

```
def time_series_to_stack_graph(data):
    stack_graph = graph in the form of list of lists
    for x, y in data bigrams:
        graph[x].append(y)
    return stack_graph, data[0] #must also have information about start
```

Figure 7: Construction of stack graph from time series

```
def stack_graph_to_time_series(stack_graph, start):
    curr_idx = start
    path = [start]
    while length(stack_graph[curr_idx]) > 0:
        curr_member = stack_graph[curr_idx].pop_head()
        curr_idx = curr_member
        path.append(curr_member)
    return path
```

Figure 8: Construction of time series from stack graph

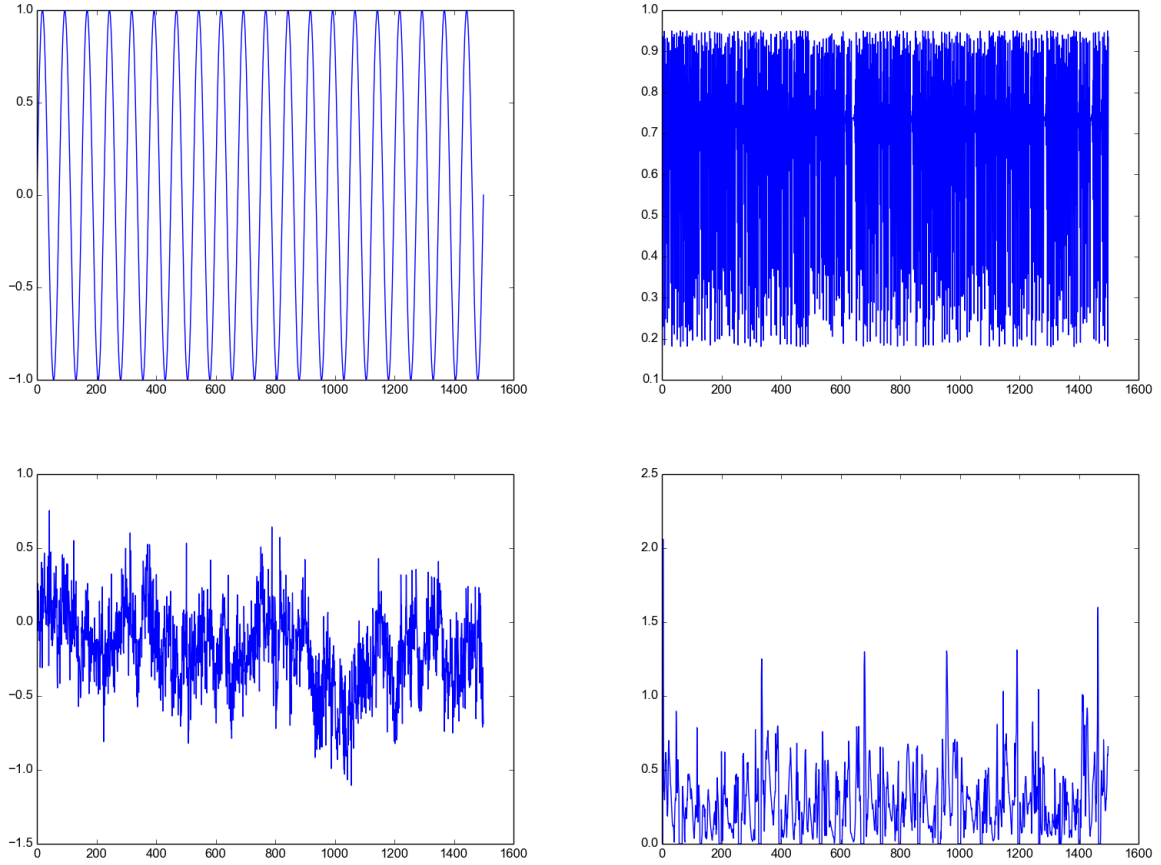


Figure 9: Plotted series in time domain. From left to right: sinusoid, logistic map with full chaos ($r=3.8$), fractional Brownian motion, real data

number for each node is the size of the corresponding stack (as opposed to the cardinality of the set of outgoing edges, which is a set and doesn't have duplicates). However, the clustering coefficient and average shortest path length of the stack graph do not have as immediately obvious meanings and therefore we took those statistics on a graph induced on the stack graph, where each stack had duplicates removed and therefore turned into a set of tails on the outdegree. This ends up being something like the recurrence matrix construction[31] of a network from the time series, and is an information-losing operation.

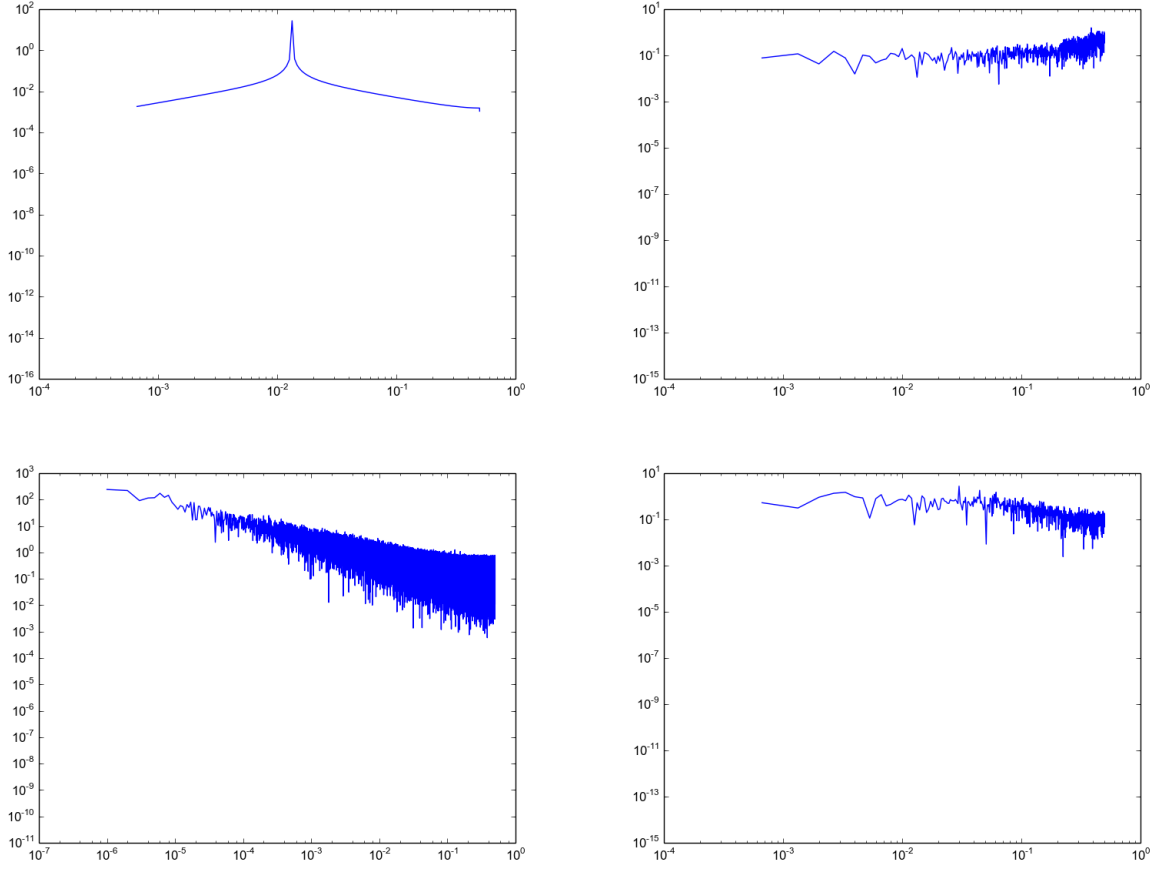


Figure 10: Plotted series in frequency domain. From upper left to lower right: sinusoid, logistic map with full chaos ($r=3.8$), fractional Brownian motion, real data

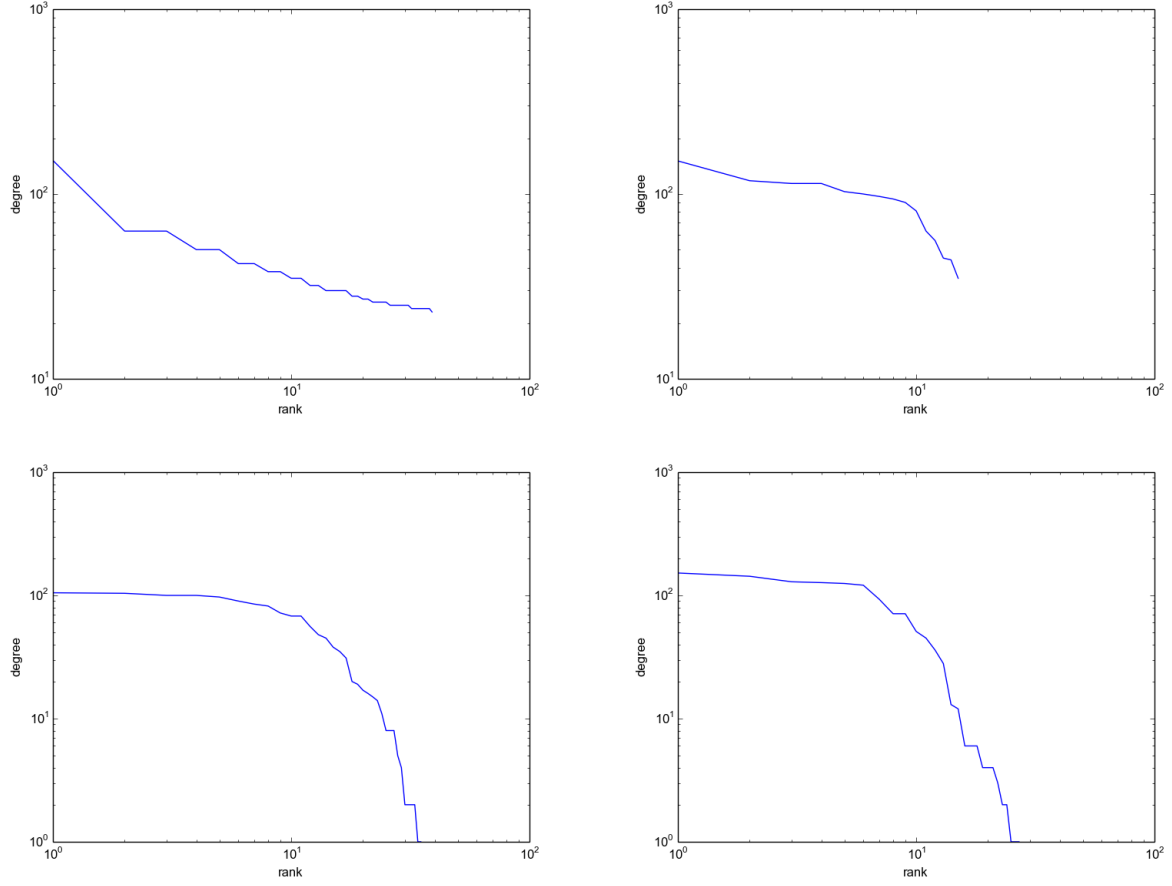


Figure 11: Equivalent degree distribution in stack graph. From upper left to lower right: sinusoid, logistic map with full chaos ($r=3.8$), fractional Brownian motion, real data

Measure	Sinusoid	Logistic Map	FBM	Real Data
Clustering Coefficient	0.425	0.092	0.768	0.572
Average Shortest Path Length	7.38	2.67	1.85	2.27

Figure 12: Other properties of stack graph, after flattening

References

- [1] Repp, B. H. (2005). Sensorimotor synchronization: a review of the tapping literature. *Psychonomic bulletin & review*, 12(6), 969-992.
- [2] Pikovsky, A., Rosenblum, M., & Kurths, J. (2001). A universal concept in nonlinear sciences. *Self*, 2, 3.
- [3] Delaherche, E., Chetouani, M., Mahdhaoui, A., Saint-Georges, C., Viaux, S., & Cohen, D. (2012). Interpersonal synchrony: A survey of evaluation methods across disciplines. *Affective Computing, IEEE Transactions on*, 3(3), 349-365.
- [4] Won, A. S., Bailenson, J. N., Stathatos, S. C., & Dai, W. (2014). Automatically detected nonverbal behavior predicts creativity in collaborating dyads. *Journal of Nonverbal Behavior*, 38(3), 389-408.
- [5]
- [6]
- [7]
- [8] Campanharo, A. S., Sirer, M. I., Malmgren, R. D., Ramos, F. M., & Amaral, L. A. N. (2011). Duality between time series and networks. *PloS one*, 6(8), e23378.
- [9]
- [10]
- [11]
- [12] Loomis, J. M., Blascovich, J. J., & Beall, A. C. (1999). Immersive virtual environment technology as a basic research tool in psychology. *Behavior Research Methods, Instruments, & Computers*, 31(4), 557-564.

- [13]
- [14]
- [15]
- [16]
- [17] Autebert, J. M., Berstel, J., & Boasson, L. (1997). Context-free languages and pushdown automata. In Handbook of formal languages (pp. 111-174). Springer Berlin Heidelberg.
- [18] Gabor, D. (1946). Theory of communication. Part 1: The analysis of information. Journal of the Institution of Electrical Engineers-Part III: Radio and Communication Engineering, 93(26), 429-441.
- [19] Pressing, J. (1999). The referential dynamics of cognition and action. Psychological Review, 106(4), 714.
- [20] Vorberg, D., & Schulze, H. H. (2002). Linear phase-correction in synchronization: Predictions, parameter estimation, and simulations. Journal of Mathematical Psychology, 46(1), 56-87.
- [21] Bellman, R. (1956). Dynamic programming and Lagrange multipliers. Proceedings of the National Academy of Sciences of the United States of America, 42(10), 767.
- [22] Mates, J. (1994). A model of synchronization of motor acts to a stimulus sequence. Biological cybernetics, 70(5), 463-473.
- [23] Hary, D., & Moore, G. P. (1987). Synchronizing human movement with an external clock source. Biological cybernetics, 56(5-6), 305-311.
- [24] Zhang, J., & Small, M. (2006). Complex network from pseudoperiodic time series: Topology versus dynamics. Physical Review Letters, 96(23), 238701.

- [25] Lacasa, L., Luque, B., Ballesteros, F., Luque, J., & Nuo, J. C. (2008). From time series to complex networks: The visibility graph. *Proceedings of the National Academy of Sciences*, 105(13), 4972-4975.
- [26] Gao, Z., & Jin, N. (2009). Complex network from time series based on phase space reconstruction. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 19(3), 033137.
- [27] Kraskov, A., Stgbauer, H., & Grassberger, P. (2004). Estimating mutual information. *Physical review E*, 69(6), 066138.
- [28] Jeong, J., Gore, J. C., & Peterson, B. S. (2001). Mutual information analysis of the EEG in patients with Alzheimer's disease. *Clinical Neurophysiology*, 112(5), 827-835.
- [29] Anscombe, F. J. (1973). Graphs in statistical analysis. *The American Statistician*, 27(1), 17-21.
- [30] Donner, R. V., Zou, Y., Donges, J. F., Marwan, N., & Kurths, J. (2010). Recurrence networks: a novel paradigm for nonlinear time series analysis. *New Journal of Physics*, 12(3), 033025.
- [31] Paxton, A., & Dale, R. (2013). Frame-differencing methods for measuring bodily synchrony in conversation. *Behavior research methods*, 45(2), 329-343.
- [32] Webber, C. L., & Zbilut, J. P. (1994). Dynamical assessment of physiological systems and states using recurrence plot strategies. *Journal of applied physiology*, 76(2), 965-973.
- [33] Camurri, A., Varni, G., & Volpe, G. (2009, September). Measuring entrainment in small groups of musicians. In *Affective Computing and Intelligent Interaction and Workshops, 2009. ACII 2009. 3rd International Conference on* (pp. 1-4). IEEE.
- [34] Shockley, K., Santana, M. V., & Fowler, C. A. (2003). Mutual interpersonal postural constraints are involved in cooperative conversation. *Journal of Experimental Psychology: Human Perception and Performance*, 29(2), 326.

- [35] Richardson, M. J., Marsh, K. L., Isenhower, R. W., Goodman, J. R., & Schmidt, R. C. (2007). Rocking together: Dynamics of intentional and unintentional interpersonal coordination. *Human movement science*, 26(6), 867-891.
- [36] Delaherche, E., & Chetouani, M. (2010, October). Multimodal coordination: exploring relevant features and measures. In *Proceedings of the 2nd international workshop on Social signal processing* (pp. 47-52). ACM.
- [37] Oullier, O., De Guzman, G. C., Jantzen, K. J., Lagarde, J., & Scott Kelso, J. A. (2008). Social coordination dynamics: Measuring human bonding. *Social neuroscience*, 3(2), 178-192.
- [38] Ashenfelter, K. T., Boker, S. M., Waddell, J. R., & Vitanov, N. (2009). Spatiotemporal symmetry and multifractal structure of head movements during dyadic conversation. *Journal of Experimental Psychology: Human Perception and Performance*, 35(4), 1072.
- [39] Campbell, N. (2008, December). Multimodal Processing of Discourse Information; The Effect of Synchrony. In *ISUC* (pp. 12-15).

As a specific instance of a domain where signals processing techniques are used, a large problem in synchronization is the definition of the signal itself and its extraction from observations of social interaction. To this end, many methods have been used, including extraction of coordination of movement features and speech features[?], movement of single and dyadic body parts[?], image processing techniques[?] and video tracking techniques[?].