Automated Methods to Examine Nonverbal Synchrony in Dyads

Norah E. Dunbar NDUNBAR@UCSB.EDU
Department of Communication
University of California, Santa Barbara
Santa Barbara, CA, USA

Judee, K. Burgoon JUDEE@ARIZONA.EDU
Eller College of Management
Arizona University
Tucson, AZ, USA

Ken Fujiwara PSYKF@CCU.EDU.TW
Department of Psychology
National Chung Cheng University
Minhsiung, Chiayi, Taiwan (R.O.C.)

Abstract

Interpersonal synchrony is when two parties in an interaction engage similarly due to the rhythmic coordination of their behavioral patterns. The study of synchrony in communication and psychology dates back to the 1960s but has evolved over time. Historically, studying synchrony has involved the manual coding of nonverbal cues by trained human coders, such as counting the occurrence of a specific behavior or making subjective ratings about a speaker. However, its time-consuming nature has been a serious barrier to the development of the field and has made it difficult for new scholars to adopt the technique. Recent advances in automated coding techniques allow researchers to collect nonverbal behavioral data effectively and objectively, and in a much more efficient manner than laborious manual coding methods historically relied upon. This chapter will review some of the theoretical and methodological challenges in studying interpersonal synchrony and propose alternatives using automated computer vision techniques.

Keywords: Nonverbal, Synchrony, Automated coding

1. Introduction

Over 50 years of research on interactional synchrony has largely produced consensus on its existence, with a few dissenting voices (see, e.g., McDowall, 1978a,b). However, the conceptualizations, attributed origins, and functions in human interaction have not been as consensual, leading to some muddle in what is treated as interactional synchrony (IS). For some, engaging in matching behavior, which is behavioral similarity between interactants in the most general sense, or mirroring, which is identical matching of an interlocutor's visibly observable posture, gestures, facial expressions and the like, is sufficient to qualify as IS. However, most definitions of IS emphasize that synchrony involves a temporal component, a rhythmic coordination of behavioral patterns (Bernieri et al., 1988; Burgoon et al., 1995; Mathiot and Carlock, 1982). Delaherche et al. (2012) define synchrony as "the dynamic and reciprocal adaptation of the temporal structure of behaviors between interactive partners"

(p. 351). But the temporal coordination could be between identical body parts, such as two interlocutors both holding hands to chest in a protective gesture upon hearing about a terrible accident, or between speaker and listener both adopting left-hand gesturing, or merely between speaker and listener both making postural shifts and head tilts at the same time.

Burgoon et al. (1998) brought some clarity to the matter by distinguishing multiple forms of adaptation, including synchrony, that are often conflated with one another. These include matching, mirroring, complementarity, compensation, self-synchrony, interactional synchrony, convergence, divergence and accommodation. Often, these various forms of adaptation are considered as synchrony. Bernieri and Rosenthal (1991) brought additional clarity to the concept by arguing that IS involves three forms of coordination: congruent interaction rhythms between the parties; simultaneous co-occurrence of vocal, facial and/or body movements; and behavioral meshing, in which individuals' separate behavior patterns come together to form a single, unified whole. Although these criteria create a clearly distinguishable pattern of social interaction, methods for operationalizing IS rarely incorporate all three of these criteria. Thus, many different patterns fly under the IS flag in the literature.

Since the study of IS has a long history in the social sciences, it is fruitful to explore the historical ways of studying and measuring it, especially for those who might be new to the field or are coming from more technical disciplines such as machine learning who may be unaware of the history. Scholars studying IS have struggled with laborious paper-and-pencil or even computer-aided manual coding for decades (see Fujiwara et al., 2021a) and have benefitted greatly from advances in computer modeling and vision analysis (see Chetouani et al., 2017; Delaherche et al., 2012, for an introduction). In this paper, we review these historical methods and then summarize the work to date on more modern computer-based coding methods. While these techniques are not new, this paper can be a useful starting point for non-technical scholars looking to simplify their methods for examining synchrony.

2. Historical Perspectives on IS

Early research in synchrony was limited by the methodology that was available at the time. Schmidt et al. (2012) say that early methods that employed the temporal coding of specific actions film or video recordings of interactions to evaluate movement changes in the form of initiations and terminations of body part movements or vocal activity to judge whether temporal co-occurrence of actions was present. For example, Newtson and colleagues (Newtson, 1993; Newtson et al., 1977, 1987) placed a transparency over a still frame on a video screen and located 15 different body parts at 1.0 or 0.5 s intervals and tallied the number of changes per frame to establish a time series. This was an ingenious method but clearly very laborious and was not widely adopted by other researchers. Schmidt et al. argue these types of coding methods in addition to being difficult to employ, provide a rather coarse grain view of IS because they are limited in the number of behaviors that can be measured. Another alternative that has been used is a periodic rating approach in which third-party raters watching a video make gestalt judgments about synchrony at regular intervals (see Bernieri et al., 1994, for an example), but this also misses the variety of behaviors that can be examined using continuous methods. Julien (2005) used 30-second rating intervals to create his reciprocity framework, but it also was quite laborious in the frequent human judgments to be made.

In examining synchrony over the past several decades, researchers have concluded that a host of behaviors, both verbal and nonverbal, is involved. Early on, Bernieri and Rosenthal (1991) defined behavior broadly because it may refer to specific muscle movements of the face or body, nonverbal gestures, vocalizations, body positions, and even mental states. This raises difficulty in studying synchrony because two partners may enmesh their behavior without enacting the very same behavior or even doing it at the same time. Again, Bernieri and Rosenthal (1991) explain it in this way: "Postures, for example, may be considered simultaneous (congruent) within the time frame of an entire interaction or interaction segment. When observing micromovement changes, the time frame for determining simultaneity may be less than 5/100 of a second" (p. 413). So, if one partner is excited and displays it with rapid gestures and increased voice volume and the other responds simultaneously with vigorous head nodding, the pair would appear synchronized to outsiders because of their behavioral meshing even if objectively they have not matched on any single behavior. Delaherche et al. (2012) note that while some studies examine synchrony on matched behaviors, others examine what they call "global motion" to capture more generalized synchrony.

If the interlocutors' behaviors match and are timed to the same vocal "metronome," this is known as *simultaneous synchrony*. It is what most people think of when referring to IS. However, speaker to listener synchrony, known as *concatenous synchrony*, more clearly captures the concept of identical behaviors between interlocutors, since while the speaker is gesturing, the auditor is not. The only way for the auditor and speaker to synchronize such behaviors is in a sequential speaker-speaker relationship, not speaker-listener relationship. This form of IS has received far less attention than simultaneous synchrony. Simultaneous synchrony almost certainly must entail a rhythmic pacing component inasmuch as it is unlikely to include identical communicative acts.

Rennung and Göritz (2016) argue that interpersonal synchrony can occur both intentionally, as well as incidentally, comparing it to what Knoblich et al. (2011) call *planned coordination* (strategic and intentional) and *emergent coordination* (a result of simple "perception-action coupling"). Whether or not synchrony occurs automatically or is done strategically will determine how we interpret it. If someone is synchronizing their movements with another person in order to build rapport and create connection with them, we usually think this is a positive social behavior and it has benefits for both their relationship and their future interactions (Tickle-Degnen and Rosenthal, 1990). But, if the rapport developed is meant to make it easier to deceive the partner, then the strategic use of synchrony is seen as having a nefarious intent (Dunbar et al., 2020). Whether or not we interpret synchronizing with a conversational partner as positive or negative has to do with the outcomes of the interaction and may not be determined in the moment.

Some of these aspects have been difficult to operationalize before the advent of automated measurement of timing and micro-level nonverbal behaviors. Researchers have had to make pragmatic decisions about how to segment interaction into fine-grained increments that can capture the rhythmicity of interactions. Because manual coding is labor-intensive (Murphy et al., 2019), researchers have had to resort to focusing on the most important behaviors or the ones that are easiest to code, which tend to be the macro-level behaviors or objective behaviors that can be counted, such as number of gestures in a time segment.

Although slow signals such as postural shifts occur as large time intervals, most macro-level measures typically are too crude to reflect the nuanced micro-level vocal, facial and gestural behaviors that reflect actual dynamic patterned behavior of IS. The demands to measure two people's behaviors over extended time periods was an inhibiting factor for researchers to investigate IS. The arrival of tools such as automated facial tracking (Yu et al., 2015), OpenFace (Baltrušaitis et al., 2016), OpenSmile (Eyben et al., 2010), OpenPose (Noori et al., 2019), and Theme (Magnusson et al., 2016) have made it possible to begin to measure the fine-grained patterns of nonverbal behavior that underpin IS.

3. Automated coding techniques of synchrony

We now turn our attention to summarizing the methods of measuring synchrony using automated methods which may be unfamiliar to non-technical audiences. This process will require two steps. First, behavior is captured using automated motion capture systems in order to produce a time series for analysis. Then, both the timing and the rhythm can be analyzed, and compared with chance levels to predict the degree of synchrony (Figure 1). These steps are explained in more detail below.

3.1. Generating time series movement data

An automated method for synchrony analysis requires continuous time series data with a constant time interval of sampling. In this regard, there are several options, which would be chosen depending on the research design and environment of data collection. When the researcher's interest is in the detailed movement of each body part, motion capture systems will be promising. Indeed, motion capture data offers great potential for synchrony research, including its application to character animation (Bente et al., 2020). However, laboratory-grade motion tracking systems, such as those provided by OptiTrack, Polhemus, and Vicon Motion Systems Ltd., have been very costly (Romero et al., 2017) so that they might not be accessible to many researchers. Instead, a depth camera (e.g., Microsoft's Kinect, Intel's RealSense) seems a low-cost alternative because it can detect the distance from the camera to objects and measure the 3D coordinates of each joint of the body. Such a low-cost system may not substitute the sophisticated motion capture system in some tasks (Romero et al., 2017), however, it successfully captures synchrony in a naturalistic conversation. (Won et al., 2014), for instance, collected time-series data of coordinate points with Kinect, and then calculated their synchrony score using angles of each joint part. Similarly, Kinect was used to capture the bodily movements of a small group consisting of three speakers and their synchrony (Fujiwara, 2016).

Alternately, the recent development of computer vision enables video-based tracking techniques, which has been a considerable driving force in research on synchrony. To date, two major options for the video-based tracking approach have been recognized: pixel or frame differencing (Paxton and Dale, 2013; Ramseyer, 2020) and OpenPose (Cao et al., 2017). As for the former technique, the Motion Energy Analysis (MEA) software, available on the Open Science Framework website (https://osf.io/gkzs3/), seems a promising option for non-technical scholars because it offers a user-friendly graphical interface and does not require a written code to perform an analysis (Ramseyer and Tschacher, 2021; Ramseyer, 2020). It automatically calculates the change in greyscale pixels between consecutive video

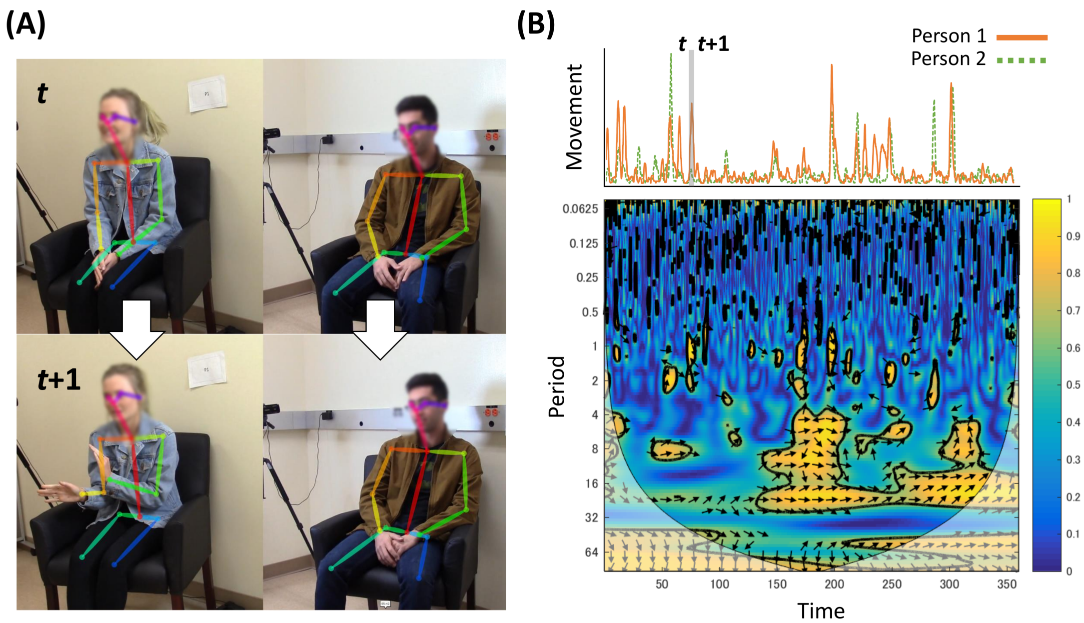

Figure 1: (A) Example of OpenPose. Each person's joint is estimated as coordinate points, which is overlaid on the image. Person on the left (Person 1) is moving her arms a lot from frame t to $t+1$, while person on the right (Person 2) is not moving much. Their movements are represented as differences in coordinate points. (B) Example of time series bodily movement data and its plane of cross-wavelet coherence. They share the same timeline (x-axis). The y-axis in the upper figure is the amount of movement. If the participant moves a lot (e.g., Person 1 from frame t to $t+1$), it will show a large value. The y-axis of the lower plane is the period (the inverse of the frequency, e.g., period 4 means 0.25 Hz). The upper region of the plane represents synchrony for fast rhythms, and the lower region for slow tempos. The intensity of their synchrony is represented by the color. In the example, the extent of synchrony increases in the middle and end of the conversation.

frames within a region of interest (ROI), which is encoded as movements when the only thing moving in the video is a participant. In other words, for the frame differencing technique, the video recorder must be fixed and neither background nor lighting conditions change. The entire body of each speaker is considered an ROI for the synchrony analysis of bodily movement (Dunbar et al., 2020; Fujiwara et al., 2019, 2020), however, specific body parts such as the head and body can also be separately investigated (Ramseyer and Tschacher, 2014). Although the pixel-based frame differencing methods offers simple "movement" information, its results have been extensively compared to those of 3D motion-tracking systems in several analytical contexts including social synchronization Romero et al. (2017) and multimodal synchronization Pouw et al. (2020), which supports its robustness as well as utility in a range of tasks. The current version of the MEA software (4.11) is available for both Mac and Windows operating systems, making it accessible to a wide range of users.

Whereas frame differencing methods capture human motion in a plane or a region, OpenPose (Cao et al., 2017; Noori et al., 2019) captures motion using points. It incorporates computer vision and deep learning so that automatically detects the 2D coordinates of the face as well as the joint parts of the human body, such as the nose, neck, shoulders, hands, and legs. Its output is thus similar to that of motion capture systems or depth cameras, but OpenPose only estimates the 2D coordinates if it is run on single video footage. It may be noted that, unlike the frame differencing technique, the coordinate data obtained in JSON format needs to be converted to distance using the formula $\sqrt{(X_{t+1} - X_t)^2 + (Y_{t+1} - Y_t)^2}$. R (the *jsonlite* package) or Python (a built-in package called *json*) will likely work well for JSON format data. Despite the additional effort of preprocessing the data, the advantage of this pose-estimation cannot be underestimated. OpenPose offers highly precise bodily coordinate data with reduced sensitivity to background noise and lighting conditions compared to the frame-differencing technique, still, its performance should be hampered with low-quality videos and/or other extraneous factors such as the clothing worn by the target (Fujiwara and Yokomitsu, 2021). Although OpenPose may generate a missing value for the targeted joint most likely due to occlusion, it still gives an estimate of the joint if information for other joints is available. It also calculates the confidence score so that the user can eliminate unstable estimations if necessary. The currently available version of the software is pre-trained, so it is not necessary to train the algorithm to estimate the coordinates of body parts.

The primary advantage of those video-tracking approaches lies in their simplicity. Unlike a motion capture system, no markers nor special sensors (e.g., an infrared camera) are required to obtain the bodily movement data. Instead, video footage taken with a single camera is sufficient. The sampling rate is 25 to 30 Hz, the same as the frame rate of the video, which should be fine enough to study human nonverbal communication. As for synchrony analysis, it is beneficial that both techniques allow for the movement of multiple persons to be tracked simultaneously. Furthermore, there is another significant utility in that the timing of onset and offset is perfectly matched to each individual if all speakers are filmed. Since these techniques employ unique methods to capture bodily movement, a recent study directly compared both video-tracking techniques (i.e., MEA and OpenPose) for synchrony with the same video clips of dyadic conversations (Fujiwara and Yokomitsu, 2021). The results were highly comparable in terms of the difference from chance, the gender differences, and the association with personality traits. Given both techniques provide essentially the same information about bodily movement under the same condition, the choice of which technique is better must depend on the video recording constraints. For dyadic conversations with a fixed camera and set lighting conditions, MEA may be preferable for its user-friendly graphic interface and straightforward output. However, if the targets are near each other or overlap, MEA may not be the best option. In such cases, OpenPose may be a promising alternative. Moreover, if researchers wish to focus on synchrony in specific body parts (e.g., only the hands or head; Dunbar et al., 2014, 2020) or specific combinations thereof (e.g., one's hand and the other's head), they may find OpenPose more useful.

3.2. Time series analysis for synchrony

3.2.1. The convergence of timing

After getting time-series data, time series analysis will be performed to evaluate the extent of synchrony. In general, time-series analyses are used for a broader range of purposes in various fields of research including forecasting. Under the context of research on IS, time series of two speakers are supposed to be analyzed in terms of their convergence of timing and rhythm (Bernieri and Rosenthal, 1991). There is no one way to compute synchrony, and many authors have used different methods in the past. As for a way to investigate the convergence of timing, cross-correlation, a simple extension of Pearson's correlation to time-series data, has been widely used (e.g., Schoenherr et al., 2019; Tschacher et al., 2014). While cross-correlation is intuitively easy to understand, it also has the drawbacks of assuming stationarity for the entire time series (Delaherche et al., 2012). Stationarity refers to a property of time-series data, and a stationary time series is one whose statistical properties including mean, variance, and autocorrelation are all constant over time. The importance of stationarity is similar to the assumption of a normal distribution in classical statistical testing because many analytical tools and statistical models rely on it. However, stationarity is not likely ensured in an unstructured dyadic conversation since speakers usually change the speed of their speech and the tempo of their movements during their conversation. To account for the possible non-stationarity of the time series, previous studies have employed segment-wise computation (Schoenherr et al., 2019). By using a segmented short-time window, the assumption of stationarity would only be adapted to the local segment, not the whole time series, which can be more reasonable. However, due to the nature of the segment, researchers have to decide on four major parameters: window size, window increment, maximum lag, and lag increment (Boker et al., 2002), each of which has a great impact on the results (Schoenherr et al., 2019).

Another solution for the (non-)stationarity issue is using a non-linear method that extracts synchronous patterns occurring in varying lags throughout the time series. For instance, dynamic time warping (Berndt and Clifford, 1994) calculates a distance between two time series using the "warping" sequences, where the time scale of one sequence is locally shrunk or extended to ensure maximum alignment of the relevant part between the two time series. Because of the unique feature of using a flexible time sequence, it can measure the convergence of timing by handling the time lags that occur at various lengths (Van Der Zee et al., 2021). As software to perform dynamic time warping, the packages *dtw* for R and *dtw-python* for Python are available. A recent study employed OpenPose and dynamic time warping to examine synchrony between humans and non-human partners (i.e., avatars) in negotiation settings. The results demonstrated that the human participants synchronized their movements with the movements of non-human negotiation partners beyond the level of chance, and the greater extent of synchrony (i.e., the smaller distance calculated by dynamic time warping) was associated with the reported greater affiliation with the non-human partner and more altruistic or avatar-friendly decisions in the negotiation (Fujiwara et al., 2021b).

3.2.2. The convergence of rhythm

As a way to examine the convergence of rhythm, a spectrum analysis that deconstructs a complex time series into the properties of its rhythmic components has been used (Dunbar et al., 2020; Fujiwara et al., 2020). In a spectrum analysis, a spectral power that indicates the magnitude at each component frequency can be calculated for each time series. Moreover, if there are two time series, a cross-spectrum analysis can provide a coherence measure. Coherence, which ranges on a scale from 0 to 1, is a measurement of similarity between the two time series at each frequency component. A coherence of 1 reflects a perfect rhythmic match between the two movements, whereas 0 reflects no match. Coherence is a non-negative metric such that the case of anti-phase synchrony where the speakers move in the opposite timing (with the same rhythm) is also calculated as higher intensity of synchrony. This is similar to how previous studies use the absolute value of cross-correlation analysis (Schoenherr et al., 2019), where the negative correlation is considered as one pattern of synchrony. Overall, this measurement may be interpreted as Pearson's correlation, however, it indicates the convergence of rhythm in the frequency domain (Fujiwara and Daibo, 2016; Fujiwara et al., 2020; Schmidt et al., 2012, 2014).

In recent synchrony research, the wavelet transform is recognized as a promising option of the spectrum analysis for synchrony in unstructured conversation because it does not require constant properties (i.e., stationarity) in each time series (Issartel et al., 2006, 2015). The wavelet approach uses a predetermined form of wavelet (i.e., mother wavelet) that is supposed to be localized into the analyzed time series. The size of the mother wavelet changes according to the analyzed frequency band; it is dilated if the analyzed frequency band is low (i.e., slower rhythm), and contracted if the analyzed frequency band is high (i.e., faster rhythm). The multi-scale property works like a microscope with adjustable resolution, thus the wavelet approach enables the precise detection of signal properties in a very complex signal and is applicable to a wide range of motor signals (Issartel et al., 2006, 2015), which shows superiority over short-time windowed techniques such as the segment-wise Fourier transform. Furthermore, relative phase information, the pattern of synchrony, can also be obtained via the wavelet transform method. The front-to-back relationship on the time axis, illustrated as a difference in phase, tells us whose move is preceding the opponent's. If a research hypothesis focuses on who is in sync with whom (e.g., leader–follower interaction), relative phase should be a useful metric. To perform the cross-wavelet analysis, the wavelet toolbox is provided for MATLAB (https://github.com/grinsted/wavelet-coherence; Grinsted et al., 2004), and the *biwavelet* package for R features the same function as MATLAB. For Python, the *PIWavelet* module is available.

The recent evidence has shown speakers exhibit the greater extent of synchrony (i.e., the rhythmic convergence) when they are highly engaged in their interaction(Dunbar et al., 2020), especially in the rhythm of 0.5–1.5 Hz, the rhythm of once every 0.67–2 sec. (Fujiwara et al., 2021a). Synchrony in the particular frequency band is more salient in conversation between two females where the speakers are more socially oriented (Fujiwara and Yokomitsu, 2021) and also contributes to building rapport compared to synchrony in other frequency bands (Fujiwara et al., 2020). It should also be noted that synchrony in the rhythm of 0.5–1.5 Hz offers highly comparable results with the manually coded measure of synchrony (Fujiwara et al., 2021a). Synchrony in this frequency was mainly related to gestural synchrony, whereas the slower tempo of synchrony pertained to postural synchrony and laughing along the partner (Fujiwara et al., 2021a). Indeed, the function of synchrony in human communication may differ depending on the frequency band. For example, in a context of communication involving deception, the deceiver appeared to synchronize his or her movements with the opponent at a faster rhythm (Dunbar et al., 2020). Synchrony in a different rhythm is still a new point of focus and needs further investigation.

3.3. Comparison with chance synchrony

The earlier synchrony research (Cappella, 1981) suggested that certain coordinated or simultaneous movements could occur by chance, and a statistical control for this "baseline" of chance synchrony would be needed. Relatedly, the automated coding introduced above is so powerful that there may be an infeasible risk of falsely detecting noise as genuine synchrony. Therefore, once researchers calculated a synchrony score using automated coding methods, it should be compared to a baseline to show that the obtained synchrony is not a product of chance. As a baseline of chance synchrony, artificial interactions created using randomly shuffled pairs or data shuffling within a time series are commonly used (Bernieri et al., 1988; Fujiwara and Yokomitsu, 2021; Moulder et al., 2018). The former is known as the pseudo-synchrony experimental paradigm (Bernieri and Rosenthal, 1991), where two time series from the genuine pair who actually engaged in their interaction are isolated and recombined in random order. In pseudo pairs, each participant retains time series movement information, however, the extent of synchrony is supposed to be lowered because the pairs were not engaged in actual interaction and no coordination occurred between them. Because of this nature, the pseudo-synchrony paradigm should only be applied if the length of the time series to be shuffled is identical. It is noted that it may not be the best option if all the interactions shared a similar behavioral pattern at a similar point in time, including dancing or structured conversation with a specific topic. In such cases, the difference between the genuine and pseudo pair is supposed to be unclear due to the "similar" behaviors embedded into the structured interaction (Moulder et al., 2018).

The technique with data shuffling within a time series is known as surrogate data generation, which is considered a time series equivalent of a randomization/permutation test (Moulder et al., 2018) because it eliminates all time-dependent properties of the interacting time series. Still, the generated surrogate data keeps the averaged information available from the entire series (e.g., *Mean*, *SD*). Since this method shuffles the data within a series, it can be used even if the time length of each conversation is different.

4. Conclusion

In this paper, we have reviewed the historical perspective of IS and the recent development of its automatic coding methods. IS has fascinated many researchers for a long time, and therefore there is still diversity in what is considered IS. In addition, the traditional manual coding effort has been a bottleneck in the development of research. The emerging automated methods offer reliable data in a cost-efficient way, which may be appealing to researchers. However, it is fair to note that no standard operationalization of these measures exists under the current development phase of the field. We believe the automated approach does not negate the traditional manual coding and rating, but it will serve as its complement. With automated coding techniques, the research on IS can be further developed.

References

T. Baltrušaitis, P. Robinson, and L.-P. Morency. Openface: An open source facial behavior analysis toolkit. In *2016 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 1–10, 2016.

G. Bente, E. Novotny, D. Roth, and A. Al-Issa. Beyond stereotypes: Analyzing gender and cultural differences in nonverbal rapport. *Frontiers in Psychology*, 11(2):599703, 2020.

D. J. Berndt and J. Clifford. Using dynamic time warping to find patterns in timeseries. In *Proceedings of the Workshop on Knowledge Discovery in Database*, pages 359–370, 1994.

F. J. Bernieri and R. Rosenthal. Interpersonal coordination: Behavior matching and interactional synchrony. In R. S. Feldman and B. Rime, editors, *Fundamentals of nonverbal behavior: Studies in emotion & social interaction*, pages 401–432. Cambridge University Press, 1991.

F. J. Bernieri, J. S. Reznick, and R. Rosenthal. Synchrony, pseudosynchrony, and dissynchrony: Measuring the entrainment process in mother–infant interactions. *Journal of Personality and Social Psychology*, 54(2):243–253, 1988.

F. J. Bernieri, J. M. Davis, R. Rosenthal, and C. R. Knee. Interactional synchrony and rapport: Measuring synchrony in displays devoid of sound and facial affect. *Personality and Social Psychology Bulletin*, 20(3):303–311, 1994.

S. M. Boker, J. L. Rotondo, M. Xu, and K. King. Windowed cross-correlation and peak picking for the analysis of variability in the association between behavioral time series. *Psychological Methods*, 7(3):338–355, 2002.

J. K. Burgoon, L. A. Stern, and L. Dillman. *Interpersonal adaptation: Dyadic interaction patterns*. Cambridge University Press, 1995.

J. K. Burgoon, A. Ebesu, C. White, P. Koch, E. Alvaro, and T. Kikuchi. The many faces of interaction adaptation. In M. T. Palmer and G. A. Barnett, editors, *Progress in communication sciences*, volume 14, pages 191–220. Ablex, 1998.

Z. Cao, T. Simon, S. E. Wei, and Y. Sheikh. Realtime multi-person 2d pose estimation using part affinity fields. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 7291–7299, 2017.

J. N. Cappella. Mutual influence in expressive behavior: Adult-adult and infant adult dyadic interaction. *Psychological Bulletin*, 89(1):101–132, 1981.

M. Chetouani, E. Delaherche, G. Dumas, and D. Cohen. Interpersonal synchrony: From social perception to social interaction. In J. K. Burgoon, N. Magnenat-Thalmann, M. Pantic, and A. Vinciarelli, editors, *Social signal processing*, pages 202–212. Cambridge University Press, 2017.

E. Delaherche, M. Chetouani, A. Mahdhaoui, C. Saint-Georges, S. Viaux, and D. Cohen. Interpersonal synchrony: A survey of evaluation methods across disciplines. *IEEE Transactions on Affective Computing*, 3(3):349–365, 2012.

N. E. Dunbar, M. L. Jensen, D. C. Tower, and J. K. Burgoon. Synchronization of nonverbal behaviors in detecting mediated and non-mediated deception. *Journal of Nonverbal Behavior*, 38(3):355–376, 2014.

N. E. Dunbar, H. Giles, Q. S. Bernhold, A. Adams, M. Giles, N. Zamanzadeh, K. Gangi, S. Coveleski, and K. Fujiwara. Strategic synchrony and rhythmic similarity in lies about ingroup affiliation. *Journal of Nonverbal Behavior*, 44(1):153–172, 2020.

F. Eyben, M. Wöllmer, and B. Schuller. Opensmile: The munich versatile and fast open-source audio feature extractor. In *Proceedings of the 18th ACM International Conference on Multimedia*, pages 1459–1462, 2010.

K. Fujiwara. Triadic synchrony: Application of multiple wavelet coherence to a small group conversation. *Applied Mathematics*, 7(14):1477–1483, 2016.

K. Fujiwara and I. Daibo. Evaluating interpersonal synchrony: Wavelet transform toward an unstructured conversation. *Frontiers in Psychology*, 7:516, 2016.

K. Fujiwara and K. Yokomitsu. Video-based tracking approach for nonverbal synchrony: A comparison of motion energy analysis and openpose. *Behavior Research Methods*, 53(6): 2700–2711, 2021.

K. Fujiwara, M. Kimura, and I. Daibo. Gender differences in synchrony: Females in sync during unstructured dyadic conversation. *European Journal of Social Psychology*, 49(5): 1042–1054, 2019.

K. Fujiwara, M. Kimura, and I. Daibo. Rhythmic features of movement synchrony for bonding individuals in dyadic interaction. *Journal of Nonverbal Behavior*, 44(1):173–193, 2020.

K. Fujiwara, Q. S. Bernhold, N. E. Dunbar, C. D. Otmar, and M. Hansia. Comparing manual and automated coding methods of nonverbal synchrony. *Communication Methods and Measures*, 15(2):103–120, 2021a.

K. Fujiwara, R. Hoegen, J. Gratch, and N. E. Dunbar. Synchrony facilitates altruistic decision making for non-human avatars. *Computers in Human Behavior*, 2021b. doi: 10.1016/j.chb.2021.107079.

A. Grinsted, J. C. Moore, and S. Jevrejeva. Application of the cross wavelet transform and wavelet coherence to geophysical time series. *Nonlinear Processes in Geophysics*, 11(5/6): 561–566, 2004.

J. Issartel, L. Marin, P. Gaillot, T. Bardainne, and M. Cadopi. A practical guide to time-frequency analysis in the study of human motor behavior: the contribution of wavelet transform. *Journal of Motor Behavior*, 38(2):139–159, 2006.

J. Issartel, T. Bardainne, P. Gaillot, and L. Marin. The relevance of the cross-wavelet transform in the analysis of human interaction - a tutorial. *Frontiers in Psychology*, 6: 566, 2015.

D. Julien. A procedure to measure interactional synchrony in the context of satisfied and dissatisfied couples' communication. In V. Manusov, editor, *The sourcebook of nonverbal measures: Going beyond words*, pages 199–208. Lawrence Erlbaum Associates Publishers, 2005.

G. Knoblich, S. Butterfill, and N Sebanz. Psychological research on joint action: Theory and data. In B. H. Ross, editor, *The psychology of learning and motivation: Advances in research and theory*, pages 59–101. Elsevier Academic Press, 2011.

M. Magnusson, J. K. Burgoon, and M. Casarrubea, editors. *Discovering hidden temporal patterns in behavior and interaction: T-pattern detection and analysis with THEME*. Springer, 2016.

M. Mathiot and E. Carlock. On operationalizing the notion of rhythm in social behavior. In M. Davis, editor, *Interaction rhythms: Periodicity in communicative behavior*, pages 175–194. Human Sciences Press, 1982.

J.J. McDowall. Microanalysis of filmed movement: The reliability of boundary detection by observers. *Environmental Psychology and Nonverbal Behavior*, 3(2):77–88, 1978a.

J.J. McDowall. Interactional synchrony: A reappraisal. *Journal of Personality and Social Psychology*, 36(9):963–975, 1978b.

R. G. Moulder, S. M. Boker, F. Ramseyer, and W. Tschacher. Determining synchrony between behavioral time series: An application of surrogate data generation for establishing falsifiable null-hypotheses. *Psychological Methods*, 23(4):757–773, 2018.

N. A. Murphy, J. A. Hall, M. A. Ruben, D. Frauendorfer, M. Schmid Mast, K. E. Johnson, and L. Nguyen. Determining synchrony between behavioral time series: An application of surrogate data generation for establishing falsifiable null-hypotheses. *Personality and Social Psychology Bulletin*, 45(7):983–993, 2019.

D. Newtson. The dynamics of action and interaction. In L. B. Smith and E. Thelen, editors, *A dynamic systems approach to development: Applications*, pages 241–264. MIT Press, 1993.

D. Newtson, G. Engquist, and J. Bois. The objective basis of behavior units. *Journal of Personality and Social Psychology*, 35(12):847–862, 1977.

D. Newtson, J. Hairfield, J. Bloomingdale, and S. Cutino. The structure of action and interaction. *Social Cognition*, 5(3):191–237, 1987.

F. M. Noori, B. Wallace, M. Z. Uddin, and J. Torresen. A robust human activity recognition approach using openpose, motion features, and deep recurrent neural network. In M. Felsberg, P.-E. Forssén, I.-M. Sintorn, and J. Unger, editors, *Image Analysis*, pages 299–310. Springer International Publishing, 2019.

A. Paxton and R. Dale. Frame-differencing methods for measuring bodily synchrony in conversation. *Behavior Research Methods*, 45(2):329–343, 2013.

W. Pouw, J. P. Trujillo, and J. A. Dixon. The quantification of gesture–speech synchrony: A tutorial and validation of multimodal data acquisition using device-based and video-based motion tracking. *Behavior Research Methods*, 52(2):723–740, 2020.

F. Ramseyer and W. Tschacher. Nonverbal synchrony of head-and body-movement in psychotherapy: Different signals have different associations with outcome. *Frontiers in Psychology*, 5:979, 2014.

F. Ramseyer and W. Tschacher. Nonverbal synchrony in psychotherapy: Coordinated body-movement reflects relationship quality and outcome. *Journal of Consulting and Clinical Psychology*, 79(3):284–295, 2021.

F. T. Ramseyer. Motion energy analysis (mea): A primer on the assessment of motion from video. *Journal of Counseling Psychology*, 67(4):536–549, 2020.

M. Rennung and A. S. Göritz. Prosocial consequences of interpersonal synchrony: A meta-analysis. *Zeitschrift für Psychologie*, 224(3):168–189, 2016.

V. Romero, J. Amaral, P. Fitzpatrick, R. C. Schmidt, A. W. Duncan, and M. J. Richardson. Can low-cost motion-tracking systems substitute a polhemus system when researching social motor coordination in children? *Behavior Research Methods*, 49(2):588–601, 2017.

R. C. Schmidt, S. Morr, P. Fitzpatrick, and M. J. Richardson. Measuring the dynamics of interactional synchrony. *Journal of Nonverbal Behavior*, 36(4):263–279, 2012.

R. C. Schmidt, L. Nie, A. Franco, and M. J. Richardson. Bodily synchronization underlying joke telling. *Frontiers in Human Neuroscience*, 8:633, 2014.

D. Schoenherr, J. Paulick, S. Worrack, B. M. Strauss, J. A. Rubel, B. Schwartz, A. K. Deisenhofer, W. Lutz, U. Stangier, and U. Altmann. Quantification of nonverbal synchrony using linear time series analysis methods: Lack of convergent validity and evidence for facets of synchrony. *Behavior Research Methods*, 51(1):361–383, 2019.

L. Tickle-Degnen and R. Rosenthal. The nature of rapport and its nonverbal correlates. *Psychological Inquiry*, 1(4):285–293, 1990.

W. Tschacher, G. M. Rees, and F. Ramseyer. Nonverbal synchrony and affect in dyadic interactions. *Frontiers in Psychology*, 5:1323, 2014.

S. Van Der Zee, P. Taylor, R. Wong, J. Dixon, and T. Menacere. A liar and a copycat: nonverbal coordination increases with lie difficulty. *Royal Society Open Science*, 8:200839, 2021.

A. S. Won, J. N. Bailenson, S. C. Stathatos, and W. Dai. Automatically detected nonverbal behavior predicts creativity in collaborating dyads. *Journal of Nonverbal Behavior*, 38 (3):389–408, 2014.

X. Yu, S. Zhang, Z. Yan, F. Yang, J. Huang, N. E. Dunbar, M. L. Jensen, J. K. Burgoon, and D. N. Metaxas. Is interactional dissynchrony a clue to deception? insights from automated analysis of nonverbal visual cues. *IEEE Transactions on Cybernetics*, 45(3): 492–506, 2015.