

Software Engineering Department  
Braude College

Capstone Project Phase A – 61998

**Interpersonal Synchrony and Leader–Follower Detection**

25-2-D-2

**Abstract**

This project will extend the InterSync system, a previous capstone platform designed to measure interpersonal movement synchrony. Our planned contribution is the integration of automated leader–follower detection capabilities, using Time-Lagged Cross-Correlation (TLCC) applied to pose-based motion data. The enhanced system will analyze specific body regions over time to infer which participant assumes the leading role during interaction. This extension is expected to enable more precise and temporally resolved analysis of movement coordination.

**Introduction**

Interpersonal synchrony refers to the real-time temporal coordination of behaviors between individuals during social interaction. Prior research has demonstrated its strong association with key social outcomes, including increased trust, rapport, and cooperative behavior. A particularly salient subtype is movement synchrony, where individuals unconsciously or deliberately align their bodily movements in space and time.

In 2023, the InterSync system was developed as part of a capstone project to quantify dyadic movement synchrony using computer vision techniques. The platform utilized state-of-the-art pose estimation tools, including MediaPipe and OpenPose, to extract body landmarks from video data and compute synchrony metrics based on joint-level motion trajectories. While this system successfully captured overall levels of coordination between participants, it lacked the ability to detect directional dynamics—namely, which individual was leading, and which was following during the interaction.

The planned project will extend the InterSync framework by incorporating a module for automated leader–follower detection.  
 Specifically, we will employ Time-Lagged Cross-Correlation (TLCC) to assess temporal dependencies between the movements of two individuals, allowing us to infer role asymmetries over time.  
The analysis will focus on specific Regions of Interest (ROIs), such as the hands or legs. It will apply a sliding-window approach to detect transitions in leadership across the interaction.

This enhancement will transform InterSync from a symmetric synchrony measurement tool into a dynamic framework capable of identifying influence patterns and role shifts.  
 By offering finer temporal resolution and direction-sensitive metrics, the extended system will hold potential for use in behavioral research and therapeutic monitoring.  
 It will provide not only a measure of how well participants are synchronized, but also insights into who initiates and who responds—an essential aspect of real-world human coordination.

**Literature Review**

## **Interpersonal Synchrony and Its Movement Subtype**

**Interpersonal Synchrony (IS)** refers to the precise timing and coordination of signals exchanged between individuals during social interaction [1]. This synchrony can manifest in various channels—such as motion, vocalization, or physiology—but the current review focuses specifically on movement synchrony, defined as the real-time matching of body motion between individuals.

A common example of movement synchrony is two people walking in step, or unconsciously mimicking each other’s gestures in conversation. This form of coordination can be quantitatively assessed using tools designed to compare time series data, such as Dynamic Time Warping (DTW) and cross-correlation. These methods enable researchers to evaluate not only the degree of alignment but also the temporal dynamics and directional influence between interacting partners.

Empirical studies have linked stronger movement synchrony to increased trust, empathy, and collaboration in experimental tasks [2], as well as enhanced therapeutic alliance in clinical contexts [3]. Disruptions in movement synchrony, on the other hand, have been observed in developmental and psychiatric conditions such as ADHD [6]. As such, this subtype of IS offers valuable insights for both theoretical models of social cognition and applied domains, including diagnostic evaluation and progress tracking.

## **The Mirror-Game Paradigm and Leader–Follower Dynamics**

A crucial distinction within movement synchrony is whether a leader–follower dynamic is present. The mirror game is a widely used research task used to study interpersonal synchrony [5]. In its original one-dimensional version, two participants move handles along parallel 55-cm rails while trying to stay in sync in real time, allowing millimeter-accurate measurement of temporal dynamics.

Building on this setup, recent studies have employed depth-sensing technologies to analyze hand movements in free space [9].

This framework enables two main interaction modes:  
 • Leader–Follower (LF): one leads, the other imitates.  
 • Joint Improvisation (JI): no designated leader.

Noy et al. (2011) found that during joint improvisation (JI), expert performers exhibit very small time delays (under 40 ms) and broader movement compared to leader–follower (LF) interactions. In LF, the follower often shows 2–3 Hz jitter. JI can also lead to *co-confident motion*—smooth, symmetric movement without a clear leader [5].

## **Comparative Appraisal of Contemporary Synchrony-Measurement Techniques**

Scholars have approached the study of movement synchrony with progressively more automated families of tools, each suited to distinct empirical contexts and burdened by its limitations.

Coding Interactive Behavior (CIB): A Manual Framework for Social Interaction Research  
 The manual coding method for social interactions is based on the CIB tool, developed by Ruth Feldman. This tool uses a global rating scale (1–5) applied to short video clips of natural interactions (for example, free play between a mother and child). It codes parameters such as maternal sensitivity, child involvement, dyadic reciprocity, or negative states. The tool includes over 40 codes, is adapted to diverse populations and cultures, and is widely used in developmental and clinical research [4].

Vision-based systems  
Automatic analysis of video recordings has become the leading technique for research conducted on a large scale or in natural environments. Two separate approaches have emerged in this field:

**(a) Motion-Energy Analysis (MEA).**

Context of Use: MEA is particularly well-suited for structured, dyadic settings-

where environmental conditions are stable: cameras remain fixed, lighting is consistent, and participants keep a physical distance.  
 These characteristics help ensure accurate detection of nonverbal synchrony using automatic video processing [7].

Limitations: In dynamic or less controlled environments—such as those with complex backgrounds or overlapping bodies—the resulting motion energy signals may mix between participants, making it hard to tell who is doing what.

**(b) Pose-estimation libraries (e.g., OpenPose, MediaPipe)**

Context of Use: Pose-estimation models are beneficial when researchers want to analyze the movement of specific body parts (like arms or legs).

Fujiwara and Yokomitsu (2021) found that pose-based synchrony measures are able to detect the same patterns found using MEA, especially those related to gender (for example, higher synchrony in female pairs) and personality traits (such as extraversion) [8].  
This means that pose-based methods can still capture meaningful differences in how people synchronize, depending on who they are and their typical social behavior.

Limitations: This level of detailed motion tracking comes with some downsides: the accuracy drops when people are blocked from view or wearing loose clothing, and getting near real-time performance usually requires a powerful graphics card (GPU).

Yet, beyond data collection itself, accurate evaluation of synchrony also depends on how the captured signals are processed and analyzed.  
The following section introduces key signal-analysis techniques that enable researchers to quantify synchrony over time with greater temporal precision.

Analytical Techniques for Signal-Based Synchrony Measurement  
Regardless of the data source (video, sensors, etc.), researchers rely on signal-processing techniques that analyze how movement signals change across both time and frequency. This is known as the time–frequency domain, which captures not only when actions occur, but also how quickly and in what rhythm.

In the previous InterSync project, four primary methods were employed to measure interpersonal synchrony, enabling flexible, accurate, and temporally robust analysis of body movements. Each technique offers unique advantages and is suited to different types of interactional contexts:

**Dynamic Time Warping (DTW)-** compares sequences even when they differ in length or speed, by stretching or compressing the time axis to find the optimal match. This method is particularly suitable for analyzing long or variable motion patterns but requires high computational resources.

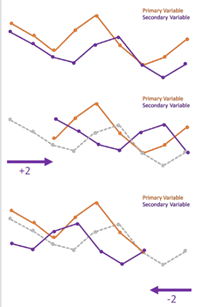
**Smith–Waterman** is an algorithm adapted from bioinformatics that performs local alignment of movement sequences, penalizing gaps or mismatches. It excels at identifying isolated segments of synchrony but is less appropriate for analyzing the overall course of an interaction.

**Simple Matching Synchrony Analysis-** provides a direct comparison between movement vectors based on a distance measure (e.g., Euclidean), and is well-suited for short or simple sequences. Its main advantage lies in processing speed and simplicity, but it is limited in handling temporal variations or complex synchrony patterns.

The TLCC method will be further elaborated, as it will serve as a central tool for analyzing synchrony and leader–follower roles in the planned project.

* **Time-Lagged Cross-Correlation (TLCC)**:

It is a fundamental technique in time-series analysis used to quantify the linear association between two temporal signals by calculating correlation coefficients across a range of temporal offsets, or lags (i.e., delays in time where one time series is systematically shifted forward or backward relative to the other). At each lag, a Pearson correlation coefficient is computed for the overlapping values. This procedure identifies the temporal alignment at which the two series exhibit the strongest co-movement, offering insights into the direction and strength of delayed dependencies [10].



**Figure 1. TLCC plots showing peak synchrony at different lags,**

**indicating temporal shifts between the primary and secondary variables.**

In movement synchrony research, this approach enables the detection of not only whether

two individuals are coordinated, but also how their coordination unfolds over time. Specifically,

the lag at which peak correlation occurs reveals leader–follower dynamics: a negative lag indicates

that one participant’s movements consistently follow those of the other, implying a follower role,

while a positive lag suggests temporal precedence and potential leadership.

This makes cross-correlation a valuable method for measuring not just the strength of coordination,

but also, for understanding who tends to lead and who tends to follow in time-based interactions.

Combining these methods enables multi-resolution measurement of synchrony—ranging from global coordination to the identification of micro-level joint movement processes.

## **The InterSync Pilot System: Foundational Platform for the Present Project**

The InterSync system was developed in 2023 by Dana Betesh and Semion Rudman as part of their capstone project.

**Project context and objective.**  
InterSync was a 2023 capstone initiative that developed an integrated and researcher-oriented workstation for the quantitative assessment of dyadic movement synchrony. Its design emphasized accessibility and relevance for developmental and clinical populations, such as individuals with ADHD or ASD. The present study builds upon that prototype and therefore begins with a concise conceptual overview of its architecture and empirical foundations.

**Core pipeline (conceptual overview).**

The InterSync system is built upon a streamlined pipeline designed to extract and evaluate dyadic movement synchrony from raw video data. Its architecture reflects a balance between computational efficiency and analytical depth, enabling precise motion tracking with minimal manual intervention. The key components of the pipeline are outlined below:

1. **Video acquisition:** The system processes video recordings of two participants engaged in a social interaction.
2. **Key-frame selection:** To reduce data redundancy and computational demands, a lightweight pixel-difference heuristic identifies frames that exhibit salient motion, allowing the system to focus on moments of interest.
3. **Landmark extraction:** MediaPipe Pose serves as the primary tool for extracting 33 two-dimensional body landmarks per frame. In instances of occlusion or complex limb configurations, OpenPose is used as a fallback to maintain detection robustness.
4. **Motion vector construction:** Frame-to-frame differences in landmark positions are used to compute motion vectors, capturing both joint-level dynamics and composite limb movements.
5. **Synchrony scoring:** The system evaluates the degree of coordination between participants using established similarity metrics. While the specific methods are abstracted here, the approach supports both global body-level and localized limb-level comparisons.
6. **Output and Visualization:** Final results are presented in an automatically generated PDF report, which includes annotated key frames, synchrony scores, and summary statistics for enhanced interpretability and further analysis.

**Reported limitations.**  
The prototype exhibited several limitations. Extended self-occlusions occasionally caused landmark dropouts, leading to artificially elevated synchrony scores. Loose or flowing garments interfered with detection accuracy, generating spurious joint positions. Real-time feedback remained dependent on GPU acceleration, and the system was limited to dyadic interactions, preventing application to more complex group dynamics or leader–follower structures.

**Relevance to the present study.**  
InterSync provides a functional foundation—such as key-frame pruning, MediaPipe/OpenPose-based landmark extraction, and multi-metric synchrony scoring—that the planned project intends to build upon.  
 Our intended contribution is to incorporate leader–follower identification within the synchrony metrics, extending the system's capabilities to include directional role analysis.

**Expected Achievements: Leader–Follower Detection within the InterSync Framework**

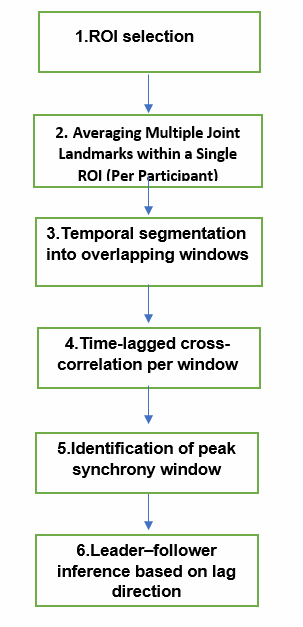
This research project continues the development of the InterSync framework. This open-source desktop platform already performs key-frame selection, extracts limb-specific motion features using MediaPipe landmarks, and computes dyadic synchrony scores based on temporal alignment.  
Our contribution is to expand this system by adding an automated module for leader identification, allowing the investigation of directionality in interpersonal interactions rather than relying only on overall similarity.  
 The module will use Time-Lagged Cross-Correlation (TLCC) applied to overlapping time windows to estimate which participant tends to lead and which one follows. The system will generate time-resolved leader–follower metrics and present the results through an interactive graphical interface, enabling researchers to track how roles shift during the interaction.

**Engineering Process**

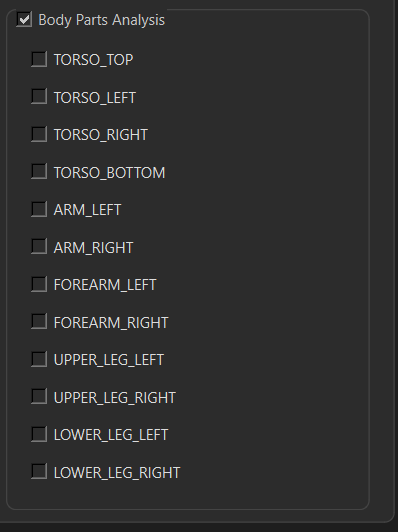
**Algorithm Concept Overview:**

The InterSync system is designed to support interpersonal movement synchrony by extracting body joint trajectories from video data using the MediaPipe pose estimation framework. It automatically detects 33 body landmarks per participant and computes motion vectors across time for each joint. While the previous system we are building upon provides full-body analysis and supports multiple synchrony metrics (e.g., Dynamic Time Warping (DTW)**,** Time-Lagged Cross-Correlation (TLCC), and Smith-Waterman alignment), we aim to extend its capabilities by isolating a specific joint and evaluating dynamic leader–follower relationships over segmented time windows.

**Leader–Follower Detection Pipeline**



The following steps describe the process by which we extended the capabilities of the InterSync system to enable dynamic, time-resolved analysis of leader–follower patterns based on motion data from a single Region of Interest (ROI), such as a hand.

**1. ROI Selection**  
The system computes motion trajectories for 33 anatomical landmarks per frame using the MediaPipe pose estimation model. From this full-body data, a single **Region of Interest (ROI)** is selected—i.e., a specific anatomical area (e.g., head, hand; not necessarily a single joint but potentially a combination of nearby parts like wrist and fingers) whose motion will serve as the basis for leader–follower analysis. This step enables focused, region-level investigation rather than global synchrony evaluation.

### **Figure 2. ROI Selection Interface in the InterSync System.**

The figure illustrates the Region of Interest (ROI) selection panel in the InterSync interface, enabling users to select specific body segments (e.g., torso, arm, leg) for focused synchrony analysis.

**2. Averaging Multiple Joint Landmarks within a Single ROI (Per Participant)**  
To enable localized synchrony analysis, we compute the motion trajectory of a predefined Region of Interest (ROI) by calculating the three-dimensional average of all anatomical landmarks associated with that region (e.g., the elbow and wrist for the forearm).

For each video frame and each participant, we calculate the mean value of the X-axis across all ROI landmarks and repeat this process independently for the Y and Z axes.  
 This results in a single 3D vector per frame per participant:

Where each component represents the average spatial position of the ROI along that axis.

These vectors form each participant’s ROI-based motion trajectory, which is then used as input for the next step—Time-Lagged Cross-Correlation (TLCC).  
 TLCC enables the detection of synchrony patterns over time and supports the inference of leader–follower dynamics between participants.

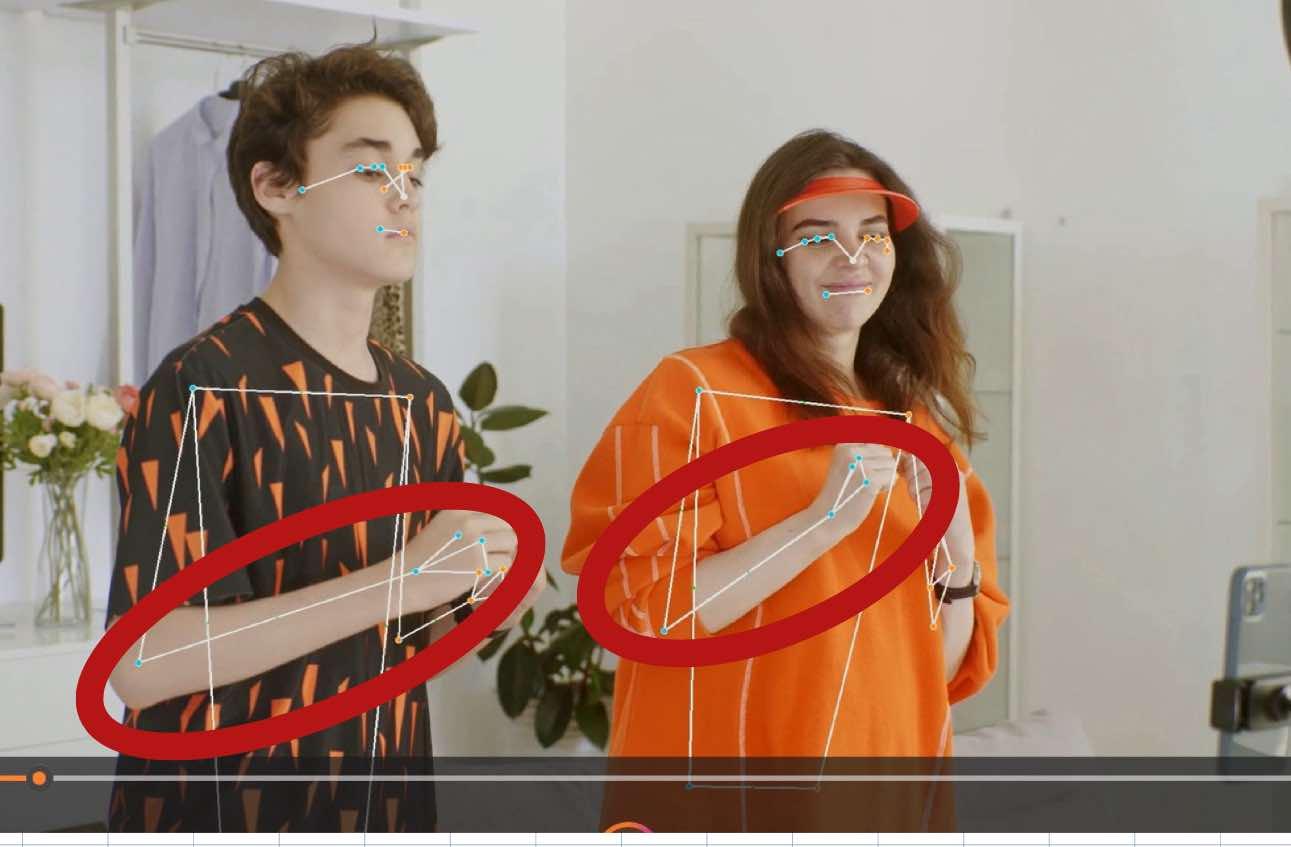


Figure 3. Full-Body Landmark Detection Using MediaPipe  
The image shows two participants with body landmarks detected via **MediaPipe**. The **forearms** are highlighted as **Regions of Interest (ROIs)**, where joint positions (e.g., elbow, wrist) are averaged per frame to generate a single motion trajectory per participant.

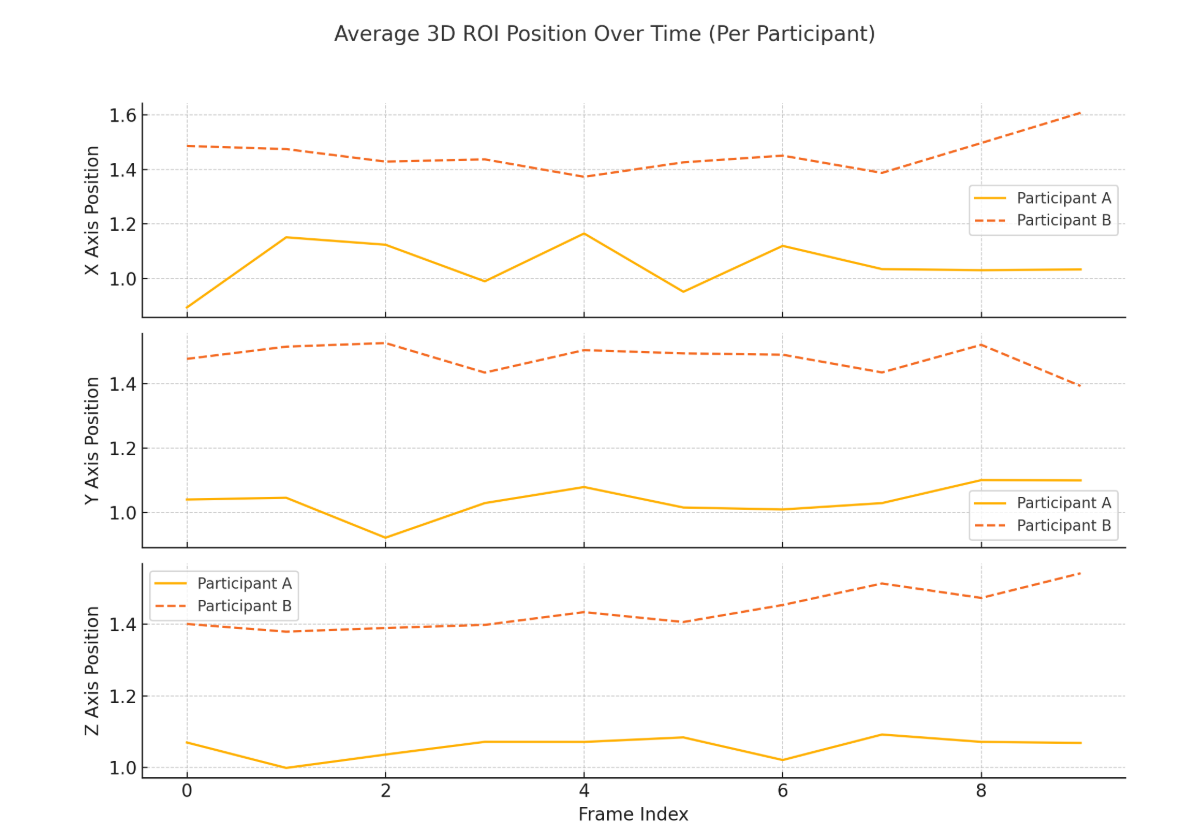


Figure 4. The graph shows the average X, Y, and Z positions of a selected body region (ROI) for each participant over time. Each line represents how that axis changes across frames. These axis-specific trends are used to analyze synchrony and detect leader–follower patterns using time-lagged cross-correlation (TLCC).

**3. Temporal Segmentation**

To enable synchrony analysis over time rather than across the entire video sequence, we propose to divide the 3D motion time series into fixed-length time windows. The length of each window will be defined in seconds, and the corresponding number of frames will be determined based on the video's frame rate (frames per second, FPS).

This segmentation will follow a sliding-window approach, where the window advances by a defined step size (*s* seconds) to produce overlapping segments. Overlapping windows are used to ensure smooth, continuous tracking of dynamic changes in synchrony and leader–follower roles, allowing the system to detect even brief transitions with higher temporal resolution.

Both the window length and step size (s) are user-defined parameters, allowing flexibility depending on the desired temporal resolution.

**4. Time-Lagged Cross-Correlation (TLCC) per Window**

Within each time window, we compute Time-Lagged Cross-Correlation (TLCC) between the movements of the two participants, based on a predefined Region of Interest (ROI). For every frame within the window, a 3D vector is calculated for each participant by averaging the X, Y, and Z coordinates of all anatomical landmarks in the ROI.

The correlation is planned to be computed using TLCC coefficients across a defined range of lags (e.g., from –*k* to +*k* frames), where *k* is a parameter determined by the temporal scale of interest and the frame rate of the video.

One possible approach is to use the correlate function from SciPy. Signal module, which computes the cross-correlation between two time series over all possible lags. In this case, the function would be applied to a pair of vectors—one per participant—each representing the 3D motion trajectory (X, Y, Z) of the selected ROI throughout the time window. The correlation could then be computed separately for each axis and analyzed to identify synchrony patterns.

For each window, we intend to extract:

* **the maximum correlation value**, reflecting the degree of synchrony.
* **the lag at which this maximum occurs**, which may provide insight into momentary leader–follower relationships.

**5. Peak Synchrony Window Identification**

After computing TLCC across all time windows, we plan to identify the segment with the highest synchrony score, defined as the window exhibiting the maximum Pearson correlation coefficient across the tested lag range. This window will be considered the peak synchrony interval, representing the period of strongest temporal alignment between the participants’ movements at the selected Region of Interest (ROI).

### **6. Leader–Follower Inference**

To identify leader–follower roles, we examine the lag at which the maximum TLCC correlation occurs within each time window. This lag indicates the temporal shift between the participants' movements:

* A negative lag suggests that Participant 2 is following Participant 1.
* A positive lag suggests that Participant 1 is following Participant 2.
* A zero lag indicates simultaneous movement, with no clear leader or follower.

Rather than relying on a single window, we analyze lag patterns across a sequence of time windows. This means we track how the lag changes—or stays consistent—over time. A stable lag direction across several windows may point to consistent leadership, while frequent changes in lag direction may reflect dynamic shifts in leader–follower roles throughout the interaction.

**Processing Workflow**

### **1. System Workflow for Leader–Follower Detection in the InterSync Framework**

The following diagram will outline the core processing pipeline of the InterSync system—an existing project that we will extend and build upon to enable automated, time-resolved inference of leader–follower dynamics based on video-derived body movement data. The architecture will be designed to be modular, scalable, and optimized for ecological validity and frame-level precision.

1. **Video Input (Two Participants)**  
   The process begins with video recordings of two participants engaged in a naturalistic interaction. These recordings serve as the raw input for the system, providing the foundation for extracting body movement data and analyzing synchrony throughout the interaction.
2. **Frame Preprocessing – Key-frame Selection**  
   To reduce data redundancy and computational demands, a lightweight pixel-difference heuristic identifies frames that exhibit salient motion, allowing the system to focus on moments of interest.
3. **Landmark Extraction – MediaPipe (Primary) / OpenPose (Fallback)**  
   Body landmarks are extracted from each selected frame using the MediaPipe Pose framework, which identifies 33 anatomical points per participant. In cases of occlusion or tracking failure, OpenPose serves as a secondary extractor to maintain landmark continuity.
4. **Motion Vector Construction – Joint-level & ROI-level Analysis**  
   Motion trajectories are calculated by comparing landmark positions across consecutive frames. The system supports both joint-specific analysis and Region of Interest (ROI)-level aggregation, allowing for anatomically targeted examination of synchrony (e.g., hand or torso coordination).
5. **Time-Series Segmentation – Sliding Windows**  
   To detect transient synchrony events and leadership shifts, the motion signals are segmented into overlapping temporal windows. This makes it possible to closely examine how synchrony and leadership evolve at specific moments during the interaction.
6. **TLCC (Time-Lagged Cross-Correlation) per Window**  
   Within each time window, Time-Lagged Cross-Correlation (TLCC) is computed between the motion signals of the two participants, based on a single 3D vector per frame. This vector is obtained by averaging the X, Y, and Z coordinates of all landmarks within the selected Region of Interest (ROI). The resulting correlation values capture the temporal alignment between participants’ movements in this region. The lag at which the correlation peaks reflects both synchrony strength and directional influence.
7. **Leader–Follower Role Inference** Directionality of interaction is inferred from the sign of the lag at which the maximum TLCC correlation occurs within each time window:

* Lag > 0 → Participant 1 consistently leads.
* Lag < 0 → Participant 2 consistently leads.
* Lag = 0 → Participants move in mutual synchrony (co-motion).

By tracking these lag values across a sequence of time windows, the system can detect stable or shifting leadership dynamics throughout the interaction.

1. **Output & Visualization Module**  
   Analytical results are compiled into an automatically generated report featuring synchrony plots, annotated frames.

תמונה שמכילה טקסט, צילום מסך, גופן, מספר

התיאור נוצר באופן אוטומטי

**Project Requirements and Research Design Summary**

**1. System Requirements for Leader–Follower Detection Module**

**Functional Requirements**

The following functions are essential for the Leader–Follower system to support time-resolved detection of leader–follower dynamics based on video-derived motion data:

|  |  |
| --- | --- |
| **1** | The system will ingest video recordings of two participants engaged in social interaction (e.g., face-to-face, side-by-side, or mirroring scenarios). |
| **2** | The system extracts 33 anatomical body landmarks per participant using the MediaPipe pose estimation model. |
| **3** | The system allows the user to select a Region of Interest (ROI), such as the hand, for targeted motion analysis. |
| **4** | The system will compute the average 3D position (X, Y, Z) of all landmarks within the selected ROI for each participant and each frame. |
| **5** | The system will segment each participant’s motion trajectory into overlapping time windows of fixed duration. |
| **6** | The system shall compute the peak Pearson correlation coefficient within each time window and identify the window with the highest score as the peak synchrony interval. |
| **7** | The system generates a summary report including graphical outputs, role classification, and key metrics. |

**Non-Functional Requirements**

|  |  |
| --- | --- |
| **1** | The system should run efficiently on a regular personal computer and use only reliable, open-source software. |
| **2** | Easy-to-use graphical interface with clear steps for loading video, selecting the region of interest, and running analysis |
| **3** | The interaction process must follow a clear and minimal-click sequence (upload → analyze →report) |
| **4** | Analysis results must include accessible report formats, including text summaries and visual aids for non-expert users |

**2. Requirement Elicitation Process**

The system requirements were derived through a combination of literature review, system analysis, and expert input. Key influences included:

* Prior research on interpersonal synchrony and leader–follower dynamics.
* Functional review of the existing interSync system, which guided architectural reuse and clarified the need for directional analysis.
* Technical insights from pose estimation tools (e.g., MediaPipe, OpenPose) and observations from early pilot runs.

This hybrid approach ensured that the requirements are both theoretically grounded and practically feasible.

**3. Research and Development Process**

The development of the Leader–Follower system will follow an iterative, research-driven workflow designed to align empirical findings with computational implementation.  
The process involved the following key stages:

* **Feature operationalization**: These behavioral insights will be translated into algorithmic logic by selecting TLCC (Time-Lagged Cross-Correlation) as the primary synchrony metric, applied across 3D motion trajectories extracted from video.
* **Prototype construction**: Building on the existing use of MediaPipe for extracting 33 body landmarks per participant per frame, we will extend the pipeline to compute average motion within selected Regions of Interest (ROIs) and apply windowed Time-Lagged Cross-Correlation (TLCC) analyses.
* **Result integration**: The output will be formatted as time-stamped synchrony graphs to support interpretability for researchers and end users.

This combined research-development pipeline will ensure scientific validity while producing a working software prototype.

**4. Anticipated Challenges**

The development of the Leader–Follower system involves several expected challenges:

* Landmark instability – Occlusions or unusual poses may cause MediaPipe to miss or misplace body points.
* Role ambiguity – Rapid lag fluctuations may complicate consistent leader–follower detection.

These challenges underscore the need for a modular, adaptable design that can support robustness across scenarios**.**

**5. Technological Stack for Future Implementation**

The planned implementation of the Leader–Follower system will employ several software libraries and tools selected for their robustness, compatibility, and relevance to motion-based synchrony analysis:

* **MediaPipe Pose (Google)**: Used to extract 33 body landmarks per frame, serving as the foundation for motion tracking.
* **OpenCV (Python)**: will be employed for video handling, frame extraction, and visual annotation of landmarks and trajectories.
* **NumPy and SciPy**: will be utilized for numerical operations, trajectory calculations, and time-lagged cross-correlation (TLCC) computation.
* **Matplotlib**: will be used to visualize synchrony results, including motion paths and TLCC graphs across axes.
* **Python 3.9**: The core programming language, selected for its versatility, extensive library support, and scientific computing ecosystem.
* **Jupyter Notebook**: will serve as the development and testing environment, enabling step-by-step debugging, visual inspection, and parameter tuning.

These tools will be integrated into a modular pipeline that will support efficient motion data processing and interpretable synchrony analysis.

**6. Success Criteria**

To assess whether the leader–follower system effectively meets its design objectives, the following success metrics and evaluation strategies were established:

* **Interpretability**: Output reports must include clearly visualized synchrony graphs, lag distributions, and per-window leader classifications.
* **Generalizability**: The system should function across different body regions (e.g., hand, head, leg).

Meeting these criteria would indicate that the system not only performs its technical function but does so reliably, efficiently, and with scientific validity.

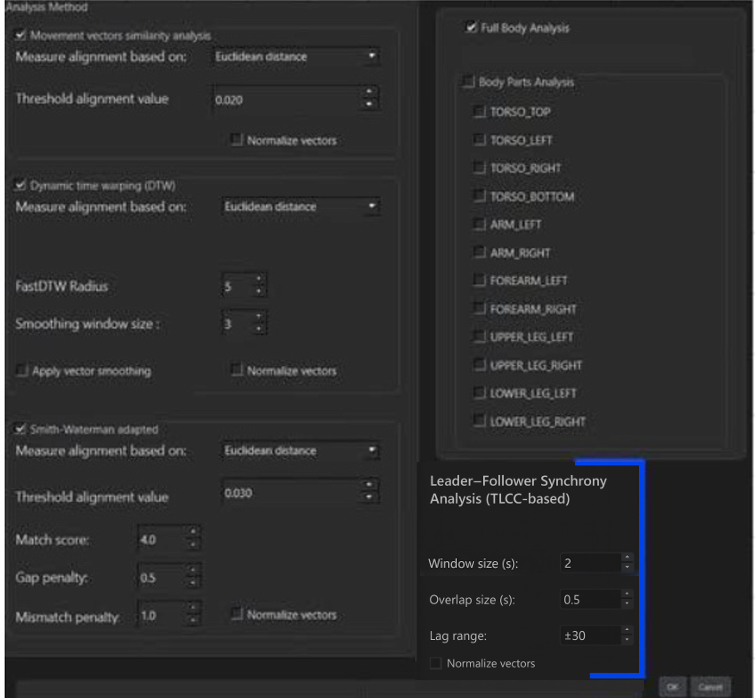
**7. Evaluation Plan**

* Dataset Exploration: We plan to search for publicly available video datasets that contain ground truth annotations of leader–follower roles defined at the window level. If such datasets are identified, we intend to use them to evaluate the system’s accuracy in later stages of development.
* If no appropriate dataset is found

1. *Manual Annotation of Selected Cases*:  
   We intend to manually annotate a set of selected test videos by visually determining who appears to lead or follow in each time window. These annotations will later be used to evaluate the system's role inference performance.
2. *Design of Controlled Video Scenarios*:  
    We also plan to create controlled video recordings in which two participants are explicitly instructed on who should lead and who should follow during specific segments. These recordings will be used to verify whether the system correctly infers the expected roles.

**GUI Prototype:**

Based on the previous InterSync project, this is the modification we are going to make:



Interface Enhancements – *Leader–Follower Synchrony Analysis (TLCC-based)*

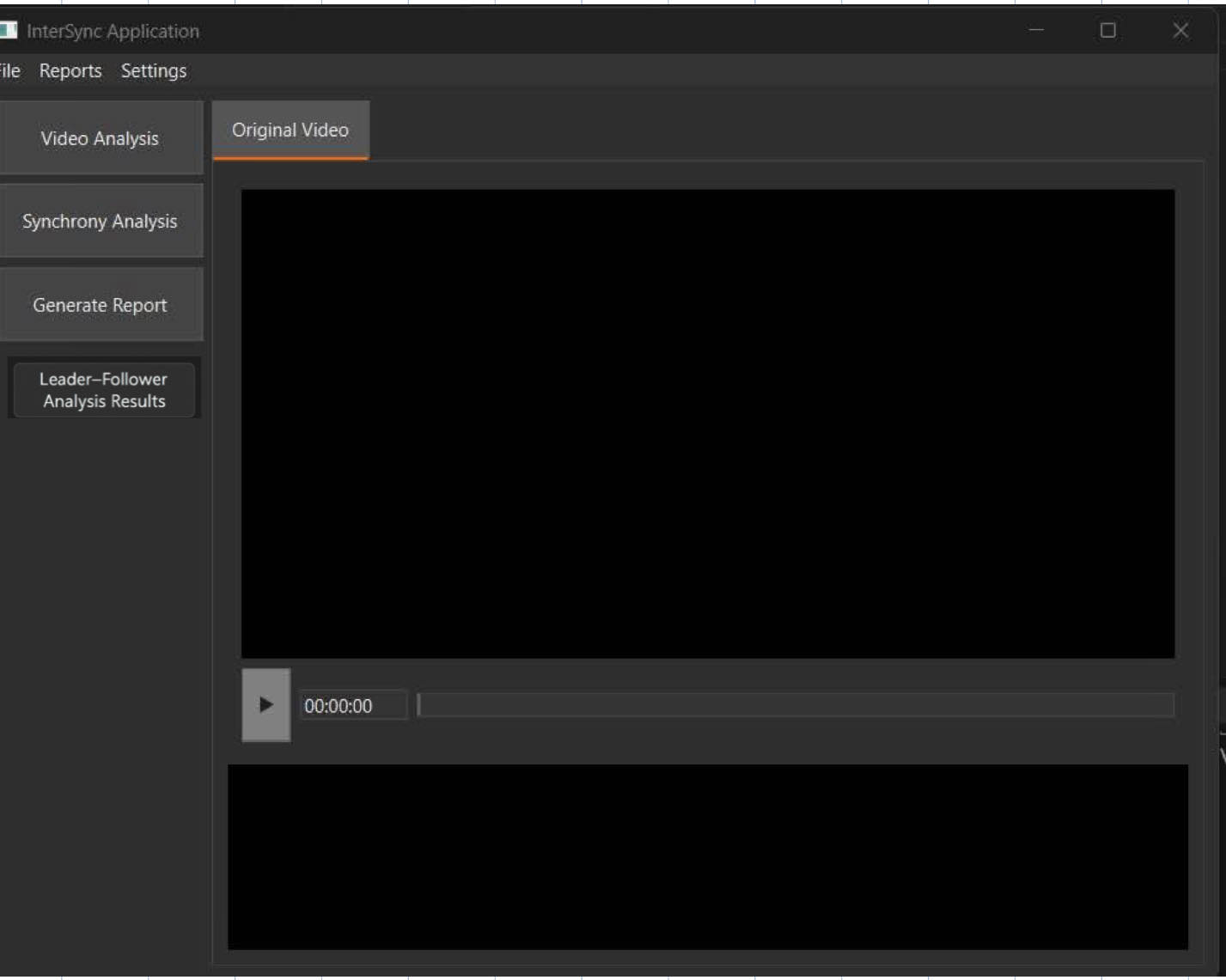
As part of the system extension to support dynamic leader–follower role detection, a new configuration panel was added under the *Leader–Follower Synchrony Analysis (TLCC-based)* section. This panel enables users to customize key parameters that control the application of Time-Lagged Cross-Correlation (TLCC) within sliding time windows.

Description of the newly added fields:

Window size (s): Specifies the duration (in seconds) of each time window used for TLCC computation. Short windows (e.g., 2-3seconds) allow for detection of rapid changes in leader–follower dynamics but may be more sensitive to noise. In contrast, longer windows (e.g., 20-30seconds) provide a more stable and focused observation but may overlook brief behavioral shifts.

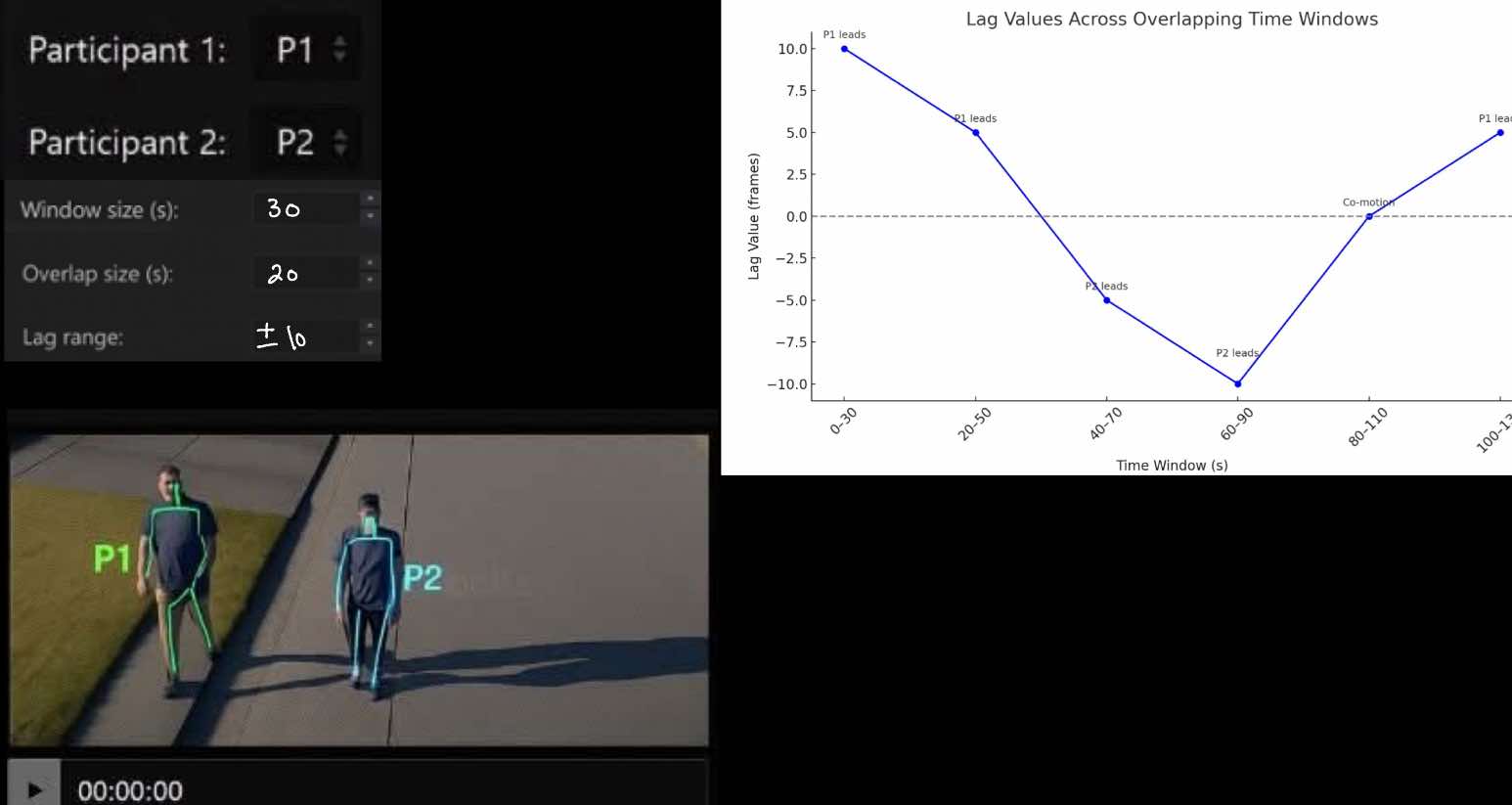
Overlap size (s): Defines the step size (in seconds) by which the analysis window advances between consecutive TLCC computations. Smaller values (e.g., 0.5–1 second) result in higher overlap, enabling the continuous tracking of leadership transitions. Larger overlap values reduce computational load but may miss subtle behavioral changes.

Lag range: Indicates the full range of frame shifts (e.g., ±30 frames) evaluated when searching for peak correlation. This sets the boundaries for allowable lag during alignment.



### Button: Leader–Follower Analysis Results This button will be designed to display the results of directional synchrony analysis (*Leader–Follower*) that will have been precomputed using the Time-Lagged Cross-Correlation (TLCC) algorithm, based on parameters entered by the user in the interface. Pressing this button will not initiate computation but will instead open a dedicated display interface that presents the outputs in a clear, dynamic, and visual format.

What will be displayed upon clicking:



### **Lag Graph by Time Window :**

This graphical display represents the expected leader–follower pattern we aim to identify in TLCC-based analysis.

* **The X-axis** represents overlapping time windows (e.g., 0–30, 20–50, 40–70…).
* **The Y-axis** represents lag values in frames:
  + **Positive lag** → Participant 1 (P1) was leading.
  + **Negative lag** → Participant 2 (P2) was leading.
  + **Lag = 0** → Simultaneous movement (no clear leader).
* **The line in the graph** shows how leadership shifts dynamically across time.

### Conclusion:

Initially, P1 leads, followed by a shift to P2 leadership, and later returns to P1. The overlap between windows ensures smooth tracking of role transitions while capturing nuanced temporal changes.

#### P1 / P2 Video Tagging

* A screenshot from the video will be shown, displaying Participant 1 and Participant 2 with pose tracking overlays labeled P1 / P2, providing a visual reference to support the graphical interpretation.

#### Display of Analysis Parameters

The analysis parameters entered in the interface will be shown at the bottom of the display:

* + Which participant was defined as Participant 1 and Participant 2
  + The selected values for window length (s), overlap size (s), and lag range

**References**

**[1]** Delaherche, E., Chetouani, M., Mahdhaoui, A., Saint-Georges, C., Viaux, S., & Cohen, D. (2012).  
Interpersonal synchrony: A survey of evaluation methods across disciplines. IEEE Transactions on Affective Computing, 3(3), 349–365.  
<https://doi.org/10.1109/T-AFFC.2012.12>

**[2]** Rennung, M., & Göritz, A. S. (2016).  
Prosocial consequences of interpersonal synchrony: A meta-analysis. In: Zeitschrift für Psychologie, 224(3), 168–189.  
<https://doi.org/10.1027/2151-2604/a000252>

**[3]** Ramseyer, F., & Tschacher, W. (2011).  
Nonverbal synchrony in psychotherapy: Coordinated body movement reflects relationship quality and outcome. Journal of Consulting and Clinical Psychology, 79(3), 284–295.  
<https://doi.org/10.1037/a0023419>

**[4]** Feldman, R. (n.d.). Coding Schemes & Interventions. Ruth Feldman Lab.

[Coding Schemes/Interventions - Ruth Feldman Lab](https://ruthfeldmanlab.com/coding-schemes-interventions/)

**[5]** Noy, L., Dekel, E., & Alon, U. (2011).  
The mirror game as a paradigm for studying the dynamics of two people improvising motion together. Proceedings of the National Academy of Sciences, 108(52), 20947–20952.  
<https://doi.org/10.1073/pnas.1108155108>

**[6]** Gvirts Problovski, H. Z., Lavi, D., Yozevitch, R., Sherman, M., Hagay, Y., & Dahan, A. (2021).  
Impairments of interpersonal synchrony evident in attention deficit hyperactivity disorder (ADHD). Acta Psychologica, 212, 103210.  
<https://doi.org/10.1016/j.actpsy.2020.103210>

**[7]** Ramseyer, F. T. (2020).  
Motion Energy Analysis (MEA): A primer on the assessment of motion from video [Preprint]. American Psychological Association.  
<https://doi.org/10.1037/cou0000407>

**[8]** Fujiwara, K., & Yokomitsu, K. (2021). Video-based tracking approach for nonverbal synchrony: A comparison of Motion Energy Analysis and OpenPose. Behavior Research Methods, 53, 2700-2711. <https://doi.org/10.3758/s13428-021-01612-7>

[9] Yozevitch, R., Dahan, A., Seada, T., Appel, D., & Gvirts, H. (2023). Classifying interpersonal synchronization states using a data-driven approach: Implications for social interaction understanding. Scientific Reports, 13, Article 11150. <https://doi.org/10.1038/s41598-023-37316-5>

[10] Esri. (2023). *How Time Series Cross Correlation works*. ArcGIS Pro documentation. Retrieved from <https://pro.arcgis.com/en/pro-app/latest/tool-reference/space-time-pattern-mining/how-time-series-cross-correlation-works.htm>