**Project Draft**

**Dana Betesh**

**Semion Rudman**

11/05/24

# Introduction

The problem we are looking to solve is the issue of exact measuring of Interpersonal synchronization – including solution for the overlap problem.

Interpersonal synchrony refers to the coordination and alignment of behaviors, emotions, and physiological responses between individuals during social interactions. It involves the temporal alignment of actions, such as movements, speech patterns, and emotional expressions, between two or more people. We can see it in various ways, such as walking in step with a friend or clapping hands in unison with an audience. This synchronization can occur consciously or unconsciously and plays a crucial role in establishing rapport, enhancing communication, and fostering social bonds between individuals. Despite its prevalence, the underlying mechanisms and significance of interpersonal synchronization in implicit social interactions have not been fully understood.

Interpersonal synchrony is measured for several reasons, as it serves as a subtle yet powerful bonding mechanism in social interactions. The coordination of body movement rhythms between individuals, known as interpersonal synchrony, is a common means of affiliation and bonding. It can be used to identify moments of connection, strengthen relationships, and investigate the qualities of relationships between individuals. Furthermore, it has been shown that to be deprived of social bonds can have significant psychological effects, such as increasing feelings of loneliness and experiences of depressive symptoms. Therefore, measuring interpersonal synchrony can provide insights into the dynamics of social interactions, the qualities of relationships, and the psychological impact of social bonds [1][2][3].

The measurement of interpersonal synchrony is also valuable for understanding the coordination of behavioral and physiological rhythms between individuals. Computational methods have been developed to describe the interpersonal dynamics of behavioral synchrony and complementarity over time, providing a framework for evaluating and quantifying impactful aspects of interpersonal interactions [4].

Several methods are used to measure interpersonal synchrony, including human observation and identification, computational algorithms, and data-driven approaches. Human observation and identification involve assessing synchrony through visual observation and perception of behavioral coordination [5].

To measure IS we need to observe people engaged in a shared activity – such as a ‘Mirror Game’. ‘Mirror game’, is a fundamental practice in improvisation theater and dance/movement therapy - played by two players imitating each other, producing coherent dance-like motion that seems choreographed. The game can be viewed as a simple paradigm in which two people create new improvised motion together [2]

Additionally, as it was already tested during previous tries to measure IS - mirror game can serve as a controlled experimental paradigm for studying interpersonal synchrony, allowing researchers to quantify the degree of coordination between individuals in terms of both temporal and spatial dimensions. By measuring the similarities between movements of two players, or generally speaking – similarity in the movements of people who try to imitate each other we can quantify the IS between them. The measurement is done through precise analysis of the movements executed by the participants. This analysis involves capturing data on the timing, spatial coordination, and form of the movements performed by the individuals engaged in the activity. [2]

From what we know – previous attempts at measuring IS through the means of mirror game had their downsides, such as measurement done based on motions in 2d, or on recordings – which in turn again reduce the measurement to two-dimensional source.

Our suggested solution is to measure interpersonal synchrony through the means of a ‘mirror game’ played by players or by any interacting parties and recorded on video. We are proposing a measure constructed based on similarity of vectors of motion by the players, reconstructed from the video file using ML and Pose Landmark detection - this way will be able to compare the movement based on a strict mathematical model and similarity of vectors.

Our solution will give us ability to measure the interpersonal synchrony more precisely and in a dynamic manner and eliminate one of the main obstacles of previous attempt at measuring interpersonal synchrony by means of analysis based on video recording - actions of one of the participants being obscured by a second one in a candid or dynamic recording. This is expected to happen when measuring an interpersonal synchrony between people not explicitly playing the mirror game but interacting in a natural way in a controlled environment.

# Literature Review

**Overview of Interpersonal Synchrony**

Interpersonal synchrony refers to the coordination of behaviors, emotions, and physiological responses between two or more individuals, where their movements or experiences overlap in time.

Interpersonal synchrony is not limited to behavioral synchrony but includes synchronization on neural, physiological, and affective levels[6]. It has been linked to empathy, prosocial behavior, social affiliation, and cooperation[7].

It can be observed in various forms such as tapping, walking, bouncing, and even in neural activities, where the coordination of neural activities during interpersonal synchrony has been proposed as a fundamental mechanism facilitating communication and affective co-regulation.

Research on interpersonal synchrony has shown its significant social influence, including its impact on self-esteem, affiliation, cooperation, and social-cognitive functioning[8]. The analysis of this phenomenon is complex, requiring the perception and integration of multimodal communicative signals. The evaluation of synchrony has received multidisciplinary attention because of its role in early development, language learning, and social connection[9]. Interpersonal synchrony has provided a framework for studying early human relationships and a growing body of research[10].

**Significance of Interpersonal Synchrony**

Interpersonal Synchrony plays an important role in fostering social bonds, rapport, and trust among individuals. A meta-analysis of 60 experiments found a medium effect of interpersonal synchrony on prosociality, revealing that it fosters prosocial attitudes and behaviors[6]. Additionally, research has shown that intentional synchrony has significant social influence, enhancing rapport, cooperation, and social-cognitive functioning[8]. It can be said that Interpersonal synchrony is a common feature of social interactions and plays a key role in promoting social bonding and pro-social behavioral outcomes across the lifespan[11].

Moreover, Interpersonal Synchrony facilitates non-verbal communication and effective collaboration. Interpersonal synchrony involves the matching of behaviors, movements, and gestures between individuals, creating a feeling of connection and understanding, which can be intentionally used in various settings such as therapy and education[12]. Whole-body synchrony has been linked to the therapeutic alliance between the therapist and patient, demonstrating the importance of measuring whole-body synchrony in clinical contexts[12].

There is also a non-insignificant impact of IS on mental health, including its role in reducing feelings of loneliness and depressive symptoms. The study on intentional synchrony versus asynchrony found that individuals felt better about themselves following a period of synchrony compared to asynchronous interaction, indicating a potential positive impact on self-esteem and self-evaluations[11]. Interpersonal synchrony has been associated with the encouragement of prosocial behaviors and the reduction of feelings of loneliness, making it an important factor in social interactions and mental well-being[12].

**Theoretical Frameworks**

These are some current theories that contribute to understanding how individuals align their behaviors, emotions, and physiological states with others, facilitating social interaction and connectivity, leading to Interpersonal synchrony.

From a study that discusses the role of the mirror neuron system in understanding others' emotions and intentions through the observation of their actions, comes a suggestion that the mirror neuron system is implicated in the automatic synchronization of actions between individuals, suggesting a neural basis for interpersonal coordination and empathy. [8]

Another work highlights the role of IS in enhancing social bonds, positing that coordinated movements and shared rhythms can enhance group cohesion and individual-to-group connectivity. The authors argue that, quote, "synchrony acts as a glue in social groups, enhancing feelings of similarity, closeness, and unity." [7]

Additionally, even in things like empathy IS can be found. It is suggested that empathy is facilitated by neural mechanisms that enable the synchronization of brain activity across individuals, where experiencing similar emotions or intentions can lead to IS [6]

**Connection between Interpersonal synchrony and ASD**

Interpersonal synchrony, the coordination of behavior during social interactions, plays a crucial role in social bonding and communication. Research indicates that individuals with autism spectrum disorder (ASD) often exhibit reduced levels of interpersonal synchrony compared to typical individuals[11][13][14]. Despite this, studies have shown that individuals with ASD can still synchronize with others, albeit to a lesser extent, especially in more simplified interactions[13]. The difficulties in interpersonal synchrony observed in ASD may stem from challenges in social orienting and perceptual access to stimuli necessary for synchronization[11][15]. Understanding the components contributing to successful non-verbal interactions are essential to identify potential interventions and improve social communication outcomes for individuals with ASD[11].

Interpersonal synchrony in individuals with autism is measured using various methods that assess the coordination of behavior during social interactions.

One approach to measuring interpersonal synchrony involves assessing the coordination of movements between individuals. Research has found that individuals with ASD show less coordination of movements, as evidenced by reduced changes in relative phase angles and less synchronization of pendulum swings with their parents[15]. Additionally, studies have highlighted the importance of intra-personal synchronization of communicative signals as a necessary prerequisite for successful interpersonal synchrony[15].

Furthermore, research has explored the presence and quality of interpersonal synchrony in individuals with ASD across various domains such as motor, conversational, physiological, and neural interactions. While synchrony is present in individuals with ASD, it is often reduced or atypical compared to interactions with typically developing individuals[16]. This reduction in synchrony may reflect differences in intra-personal mechanisms and timing functions that impact interpersonal coordination in individuals with ASD[15].

**Measurement Challenges and Methods**

The accurate measurement of IS involves navigating several obstacles, ranging from technical limitations to the inherent subtleties of synchronous behavior.

Capturing the full spectrum of IS requires attention to not only physical movements but also physiological responses and neural activity, highlighting the challenge of detecting multimodal synchrony. [9] In addition, the subtlety and variety of synchronous behaviors can complicate measurement, as synchrony often occurs below the level of conscious awareness, making it difficult to capture without sophisticated analytical techniques. [17]

The technological challenge of accurately capturing and analyzing synchronized behavior is emphasized across several sources. High-resolution data collection and analysis tools are required to measure the nuanced aspects of IS, from motion capture systems to advanced computational algorithms.

Despite these challenges, a variety of methods have been developed to measure IS, each with its own strengths and applications.

Some sources mention the use of human observers to identify instances of IS, stating that trained observers can reliably detect moments of synchrony [17], although this method is subject to human error and bias.

On the other hand, the application of computational algorithms for IS measurement is also discussed by some authors, where they note that the algorithms can analyze temporal and spatial aspects of synchronization, offering a more objective measure than human observation alone [9]

Some explore data-driven approaches, including machine learning models, which can uncover patterns of synchrony across large datasets, identifying subtle instances of IS that may not be immediately apparent.[12]

There authors reference novel measures for measuring IS like dynamic pose similarity (DPS) [5] which, for example, creates a time series for each participant that aggregates the position of fifteen joints with the directions of movement to receive a measure of IS in participants.

In the same article there is a promising discussion of the integration of multimodal data for a more comprehensive measure of IS. It is suggested that combining behavioral, physiological, and neural indicators of synchrony can provide a more holistic understanding of IS and its effects on social interaction.[12]

**Previous work and our planned contribution**

In our work we are going to focus on computation-based methods of measuring IS, based on analysis of videos recorded. Previously, similar work in measuring IS based on analysis of video recordings was done using the Motion Energy Analysis (MEA) system.

The Motion Energy Analysis (MEA) system is an objective automated method used to continuously monitor and quantify the amount of movement occurring in pre-defined regions of interest (ROI) during social interactions. MEA employs a frame-differencing algorithm that measures the differences between consecutive frames in the given ROI. The algorithm calculates the number of changes, providing an objective measure of motion dynamics during interactions[18]. MEA has been utilized in empirical studies of nonverbal behavior in clinical settings and healthy dyads to analyze movement patterns and nonverbal synchrony between individuals[18]. It is based on frame-differencing methods that evaluate differences in grayscale pixels frame by frame to quantify movement in predefined regions of interest, such as the head and torso[19]. Researchers have employed MEA to assess movement during speech tasks, psychotherapy sessions, and other social interactions to understand nonverbal synchrony and coordination between individuals[19][18].

Despite being used in studies before, there are certain limitations and considerations to using MEA for measuring IS. MEA requires specific conditions such as a static camera position and stable lighting which may limit its applicability in settings where these conditions are challenging to maintain[18]. Also, while MEA provides objective measures of movement dynamics, it does not quantify the direction of movement. ROI being measured is marked manually, but the algorithm used is unable to discern the direction and location of the movement within the ROI which means that interpreting these measures in the context of social interactions and nonverbal synchrony requires careful consideration and expertise to draw meaningful conclusions[20].

There was also work done in measuring IS by fellow students in Braude College, based on ML to try and circumvent the issues posed by use of MEA. But based on the results achieved there is still a place for improvement, as their solution was lacking in a critical aspect – failure of measuring of the synchrony in a case of overlap between the subjects of the recording analyzed, due to partial or full overlap between their extremities. Additionally the same issue was observed when the subjects of the recording were interacting in a dynamic manner, and partially obscuring each other.

To alleviate said issue we are proposing a modification to the solution found previously -Similarly to previous work we decided on using ML and video analysis to measure IS, but unlike previous work – instead of measuring frame-to-frame differences we chose to focus on calculating vectors that represent movements of the body, doing so separately for each part of the body and for each person taking part in the IS. These vectors, normalized, should allow a more direct comparison of movements, and can help circumvent both the problems posed by usage of MEA and issues found in previous solutions.

# Expected Achievements

Our project aims to advance the currently existing methods of measurement of Interpersonal Synchrony through computational video analysis. Building and expanding on existing methods and addressing known limitations, such as an ‘overlap’ issue, our project seeks to develop a new approach to quantifying IS in dynamic social interactions.

Our main expected achievement is the development of a new, machine learning-based method for measuring IS, including both developing the idea behind a new measurement algorithm and implementation of code, using ML libraries for human body pose detection, that will allow us to execute it. We want to achieve this by focusing on the calculation of movement vectors for each body part for every participant in the video-recorded interaction session, and by doing so to overcome the limitations of Motion Energy Analysis (MEA) system and previous attempts at achieving accurate measurements of IS using ML methods that were undertaken by our fellow students at Braude College. Additionally we aim to solve the existing overlap problem. This issue exists both for MEA and for ML analysis and is defined by bodies of the participants in the measurement process being overlapped in the recording - due to the recording being imperfectly set up or participants moving and additionally the video being a flat image without depth data.  
The factors described can lead to parts of the participants bodies being hidden and hard to correctly detect and analyze for Interpersonal synchrony score. Our solution, aims to alleviate this issue by creating vectors that represent the movement and by doing so - at least partially reconstruct the movements even when they are not clearly measurable

To achieve these goals we expect to focus on improving both precision and reliability of our solution compared to previously existing ones.

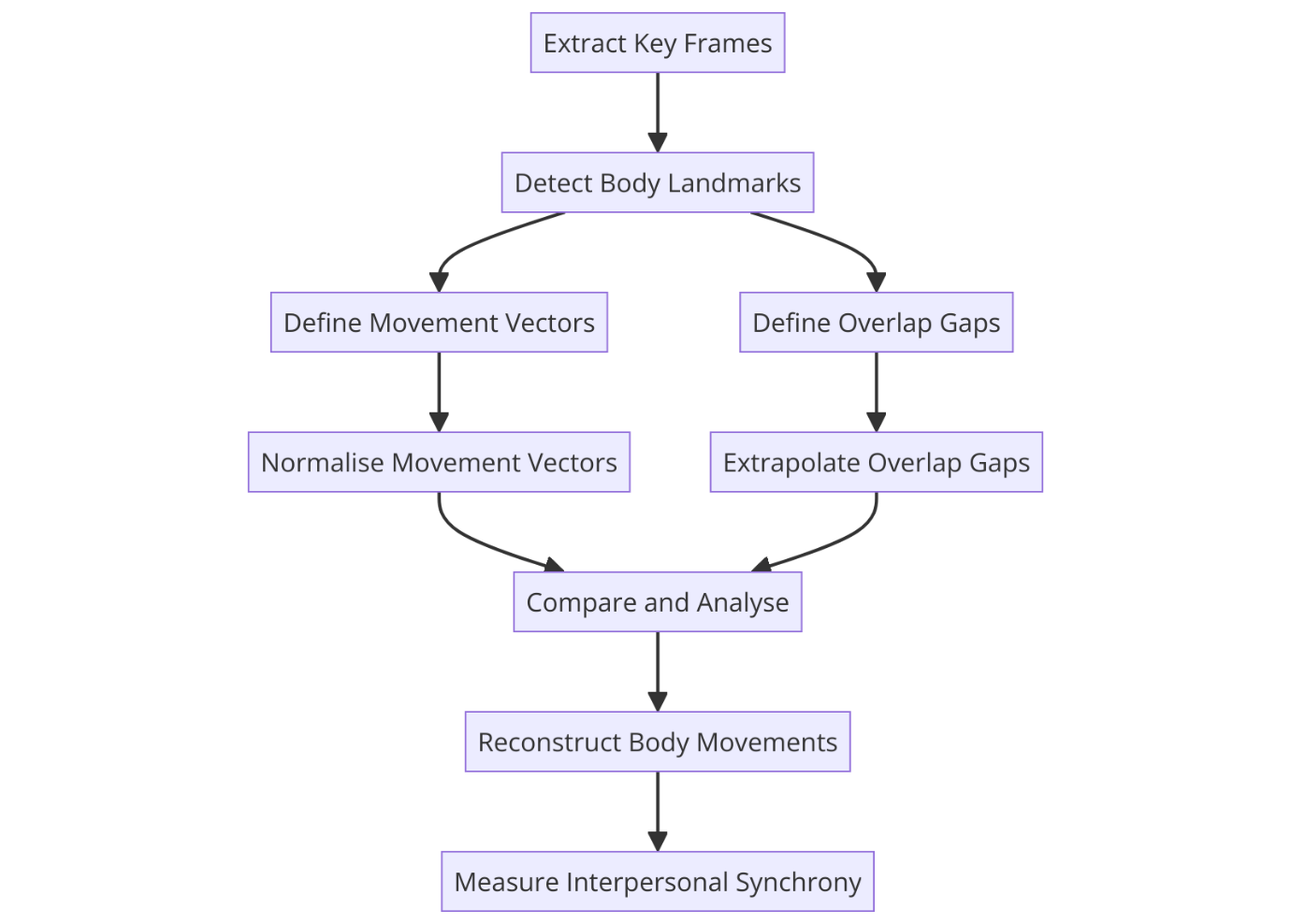
Additionally, by accurately measuring interpersonal synchrony in scenarios with partial or full overlap of participants and in dynamic, complex interactions, we are hoping to contribute to the understanding of interpersonal synchrony in real-world settings. Given the project's capability to assess interpersonal synchrony with improved accuracy, we hope to contribute a valuable tool for future research.

# Engineering Process

**Algorithm idea description:**

From a precision standpoint, we plan on creating movement vectors for body parts as follows:

* First we plan on extracting key frames from the video, with keyframes being starting and ending points of a movement segments and start / end points of overlaps
* Then we plan on performing detection of body landmark locations using a ML library for body pose detection on the keyframes found previously.
* Each pair of sequential ‘movement’ keyframes will supply us with ‘start’ and ‘end’ points for a movement that can be described by a normalized vector.
* Each pair of sequential ‘overlap’ keyframes will supply us with ‘start’ and ‘end’ points of a gap that needs to be extrapolated using previous movement and final body landmark locations
* Lastly, normalizing these vectors will allow us to discard local differences, and allow a direct comparison.
* Overall, this approach will hopefully allow us to reconstruct and compare a complete picture of body movements and measure interpersonal synchrony of recorded participants.



To do so we considered few of the more popular ML libraries to extract body locations from video. We provide following comparison regarding libraries examined:

**Key Frames Extraction:**

Key frames in videos are specific frames that define significant moments or changes within a video sequence and due to that provide a representative snapshot of the content. Key frames are used for various video processing tasks such as summarization, indexing, and compression, providing a concise but comprehensive overview of the video's overall dynamics.

We identified a number of different currently used methods to extract key frames from the video:

* **Frame Difference:**
  + Pixel Difference: calculating the absolute difference in pixel values between successive frames. A threshold-based selection process identifies frames that surpass predefined change criteria
  + Histogram Difference: comparing color or intensity histograms of consecutive frames. Significant histogram variations usually correlate with substantial scene changes, defining key frame selection
* **Optical Flow:** analysis of the motion vectors between frames. Frames that demonstrate notable vector magnitudes which indicate major movement or scene transitions—are selected
* **Machine Learning:**
  + Clustering: algorithms such as k-means used to group frames based on visual similarity. Frames that best represent each cluster's characteristics are selected as key frames
  + Deep Learning: models such as CNNs and RNNs are trained to extract temporally significant frames based on learned features from large datasets

For the purpose of our project, we are planning on using a ‘frame difference’ based method for key frame extraction, as the easiest one to implement.

Particularly, we are planning on using PeakUtils python package that provides peak detection functions for 1D data. We will be performing peak estimation based on frame difference which will be calculated based on pixel values between frames

**Currently used libraries for pose estimation:**

For the task of estimating human pose and extracting the vectors we looked at following popular python ML libraries suited for body landmark estimation:  
  
**OpenPose:** OpenPose is one of the first real-time multi-person systems to jointly detect human body, hand, facial, and foot keypoints (in total, 135 keypoints) on single images. It's open-source and supports multiple platforms.

**YOLO:**  You Only Look Once series of object detection models. The YOLO was chosen as one of the candidates for its speed and accuracy in detecting objects in images or video streams, and was found to be  a popular choice for a range of existing computer vision applications. Able to detect up to 16 keypoints

**PoseNet:** PoseNet is a machine learning model that allows for real-time human pose estimation in the browser. It can be used with TensorFlow, making it a viable tool for web-based work. Able to detect up to 16 keypoints

**MediaPipe:** Developed by Google, MediaPipe supports cross-platform, customizable ML solutions for live, streaming, and generally low latency media. The MediaPipe Pose solution is optimized for real-time applications and can work efficiently across different devices, with relatively low demand for hardware. Detects up to 32 keypoints

**Comparison of Body Pose Estimation Libraries**

When comparing OpenPose, YOLO Pose, PoseNet, and MediaPipe for body pose estimation, several factors needed to be considered, across multiple web-based sources:

1. Accuracy

* YOLOv7 Pose is highlighted for its real-time, multi-person keypoint detection model capable of providing highly accurate pose estimation results[21].
* MediaPipe offers accurate full-body pose estimation using pre-trained deep learning models for facial landmarks, hand tracking, and body poses[22].
* PoseNet is known for its ability to estimate either a single pose or multiple poses[24].

2. Multi-Person Pose Estimation:

* YOLO Pose is capable of multi-person keypoint detection[21].
* MediaPipe Holistic provides solutions for full-body pose detection, including multi-person scenarios[22].

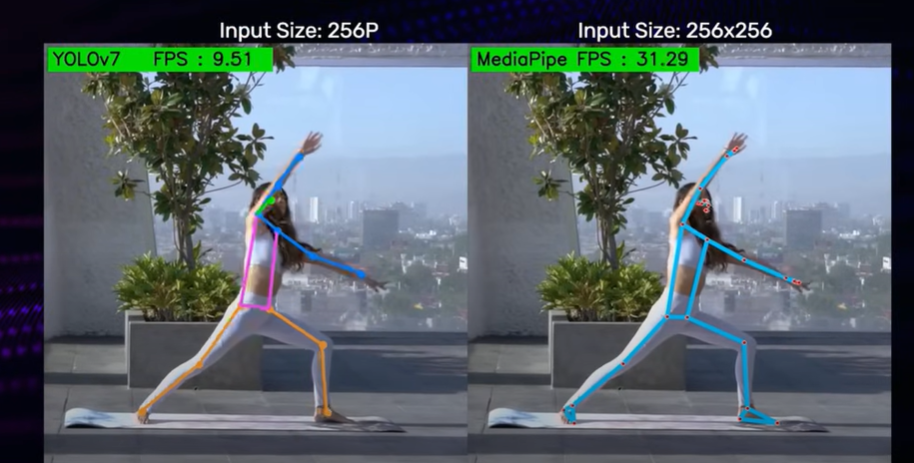
3. Speed:

* YOLO Pose is designed for real-time processing, emphasizing speed in keypoint detection[21].
* MediaPipe offers real-time capabilities for full-body pose estimation tasks[22].

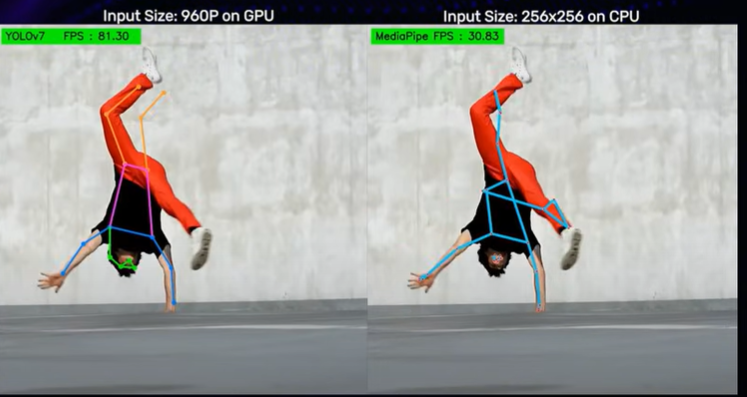
4. Hardware Demands:

* MediaPipe utilizes machine learning models that offer flexibility in hardware requirements[23].
* YOLO Pose is optimized for real-time performance but may have specific hardware demands due to its speed requirements[21].

**Render speed example difference in FPS:**

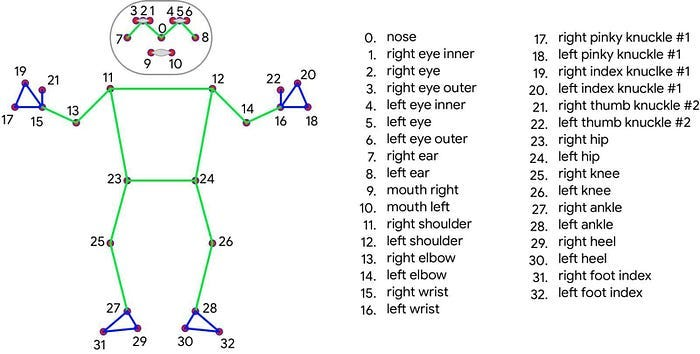
****

**Accuracy on ‘hard poses’**

****

For our project we will be working with Google’s MediaPipe ML library, as it is both well suited for our needs, and well documented and supported.

**Key body points as detected by MediaPipe:**

****

**Measuring Interpersonal Synchrony**

Following the extraction of movement vectors from video data, we will measure the IS based on vectors, using methodologies adapted from previous projects. The analysis will use either the Rolling-Window-Time-Lagged-Cross-Correlation (RWTLCC) or Dynamic Time Warping (DTW), depending on which is deemed more suitable after preliminary tests on vector-based synchronization data.

Rolling-Window-Time-Lagged-Cross-Correlation (RWTLCC) is utilized to compute correlations between two time-shifted streams of vector data. This method is particularly effective for identifying synchrony in datasets where exact temporal alignment does not occur naturally. By adjusting the time window and lag, RWTLCC can pinpoint the periods of highest synchrony, making it invaluable for analyzing the temporal dynamics of interpersonal interactions characterized by slight temporal discrepancies.

Dynamic Time Warping (DTW) offers a solution for comparing vector sequences that vary in length or are temporally offset. Through non-linear adjustments, DTW aligns these sequences to maximize similarity, using a cost function—typically Euclidean distance—to assess the disparity between points in the sequences. This method is critical for handling non-linear temporal variations and ensuring comprehensive synchronization analysis, especially when traditional linear methods fall short.

The decision to implement RWTLCC or DTW will be based on their ability to process specific attributes of the movement vector data gathered.

**Software Application and User Interaction / GUI:**

For our project we found that the development of a user-accessible graphical user interface (GUI) is essential to ensure that the functionalities of the analysis tool are accessible to a broad spectrum of users, including if need be likes of researchers, therapists, and educators. A GUI should provide a visually intuitive way for users to easily upload the videos, initiate analyses, view and interpret results, and generate detailed reports. This interface also should simplify complex processes such as pose estimation and movement vector analysis by abstracting the underlying technical complexities into user-friendly operations.   
Furthermore, a GUI will facilitate immediate visual feedback and interactive elements, enhancing user engagement and understanding of the analysis process. By incorporating clear navigation, informational tooltips, and customizable settings, the GUI should empower users to effectively utilize the tool, regardless of their technical expertise.

Given that our project is in Python, we looked at common ways to create GUI for Python projects and arrived at two viable options – Tkinter and QT.

The Tkinter library is being bundled with Python for Windows and counts as the default Python GUI library, which represents a wrapper around the Tcl/Tk GUI toolkit. Additionally, as we found - Tkinter is a pure GUI library and not a GUI framework, which means it only provides a set of graphical components for building GUIs, while lacking in support for GUIs-driven data sources, databases, or for displaying multimedia.

QT on the other hand is a full framework, utilized in Python through bindings like PyQt and PySide, that provides a comprehensive toolkit for creating cross-platform applications. This framework is a wrapper around the native C++ toolkit Qt. Additionally, Qt goes beyond a simple GUI library to provide a full-fledged GUI framework, which means it not only offers an array of graphical components but also a support for multimedia, databases, and GUI-driven data sources. This makes it better suited for developing feature-rich applications.

Overall, the differences can be summarized as follows[25]:

|  |  |  |
| --- | --- | --- |
| Feature | Tkinter | Qt (via PyQt/PySide) |
| Licensing | Python Software Foundation License (free and open) | GPL for PyQt/open-source; LGPL for PySide/commercial licenses required for proprietary software |
| Cost | Free | Free in open-source context; commercial licenses available for proprietary development |
| API Updates | Stable, with updates tied to Python releases | Regular updates with new Qt releases |
| Tools and Utilities | Basic widget set | Comprehensive set including advanced graphics, networking, and database tools |
| Performance | Good for simple applications | High-performance, suitable for complex applications |
| Community Support | Large due to Python’s popularity but less for GUI | Very large with extensive resources and examples |
| Documentation | Sufficient for basic use, less detailed for advanced GUI topics | Extensive and detailed, with lots of tutorials and examples |
| API Style | Simple and easy to learn, but limited in capability | More complex but powerful and flexible |
| Ease of Use | Very easy to start with, best for simple GUIs | Steeper learning curve, best for feature-rich applications |
| Multimedia Support | Minimal, external libraries needed for beyond basic audio/video | Robust, built-in support for complex multimedia applications |
| Support for Data Sources | Limited, requires additional libraries for advanced features | Extensive, native support for a variety of data sources including SQL databases |

Based on the differences summarized in the table above – we decided to focus on Qt framework.  
The Qt framework, originally developed in C++ programming language and as such it is not natively usable in Python. To bridge this gap, binding libraries like PyQt and PySide have been developed. These bindings act as interfaces, translating Python code to work with the underlying C++ codebase of Qt. As these libraries differ in their API we comprised following comparison to decide on which one to use:

|  |  |  |
| --- | --- | --- |
| Feature | PySide2 (Qt for Python) | PyQt6 |
| Licensing | LGPL (suitable for both open-source and proprietary software) | GPL for open-source; Commercial license required for proprietary software |
| Cost | Free, no cost for commercial use unless modifying Qt itself | Commercial license requires purchase |
| API Updates | Slightly delayed updates after Qt releases | Quick updates following Qt releases |
| Tools and Utilities | Standard Qt tools | Includes tools like pyuic and pyrcc |
| Performance | Comparable to PyQt6 | Comparable to PySide2 |
| Community Support | Growing, backed by The Qt Company | Large, well-established, many resources |
| Documentation | Integrated with official Qt documentation | Extensive, but less streamlined |
| API Style | More Pythonic, follows Python naming conventions | Closer to Qt's C++ style |
| Ease of Use | Minor differences in API calls, more adjustments may be needed | Streamlined handling of UI and resource files |

**Expected Product:**

The goal of our project is to deliver a tool capable of accurately measuring Interpersonal Synchrony while overcoming existing limitations in the field. This tool will be based on the capabilities of machine learning and pose estimation technologies to provide a new approach to analyzing dynamic social interactions.

Our expectations from the final product are as follows:

1. Software Application for IS Analysis:

A working software application that integrates the developed machine learning-based method for measuring IS. This application should allow users to analyze video recordings of social interactions and receive detailed analyses of IS, including measurements of movement synchrony between individuals.

2. User-accessible GUI:

Usage of the Qt cross-platform application development framework paired with PyQT python bindings library to build user-accessible GUI for the software described.

3. Integration with MediaPipe for Pose Estimation:

Our choice to utilize Google's MediaPipe library for pose estimation is a critical component of our product. MediaPipe's real-time capabilities and accuracy in detecting up to 32 keypoints will provide the basis for our movement vector analysis, ensuring our tool can accurately capture and analyze body movements.

4. Comprehensive Documentation

We plan on providing comprehensive documentation, including user guide, technical specifications, and examples of use cases.

5. Reporting in PDF and CSV:

We plan on having an ability to generate customizable reports based on the analysis of interpersonal synchrony. The reports should be available both PDF and CSV formats

**Use Case diagram:**

Actors

Users: Possible users represent following groups – Researchers, Therapists, Educators

Use Cases

* Upload Video Recording

Actor: Researcher

Description: The user uploads a video recording of a social interaction to be analyzed for IS. The system accepts various video formats and prepares the video for analysis.

* Perform Pose Estimation:

Actor: Researcher - Include

Description: User prompts the system to perform pose estimation on the uploaded video to identify and track keypoints of participants' bodies throughout the recording, using the integrated MediaPipe library

* + Include: Set Video Analysis Settings

Description: User has an option to set settings used for the Pose Estimation process before the process start. If no settings applied – the system should use the default ones for the process.

* Analyze Interpersonal Synchrony:

Actor: Researcher

Description: User initiates the process of calculation of movement vectors for each body part of every participant, based on the pose estimation data. To analyze the degree and patterns of IS between individuals in the video.

* Generate Report:

Actor: Researcher

Description: After completing the analysis, the user can generate a customizable report summarizing the findings.

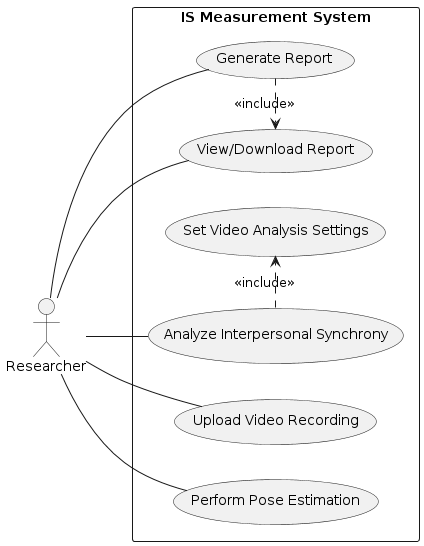
* View/Download Report:

Actor: Researcher

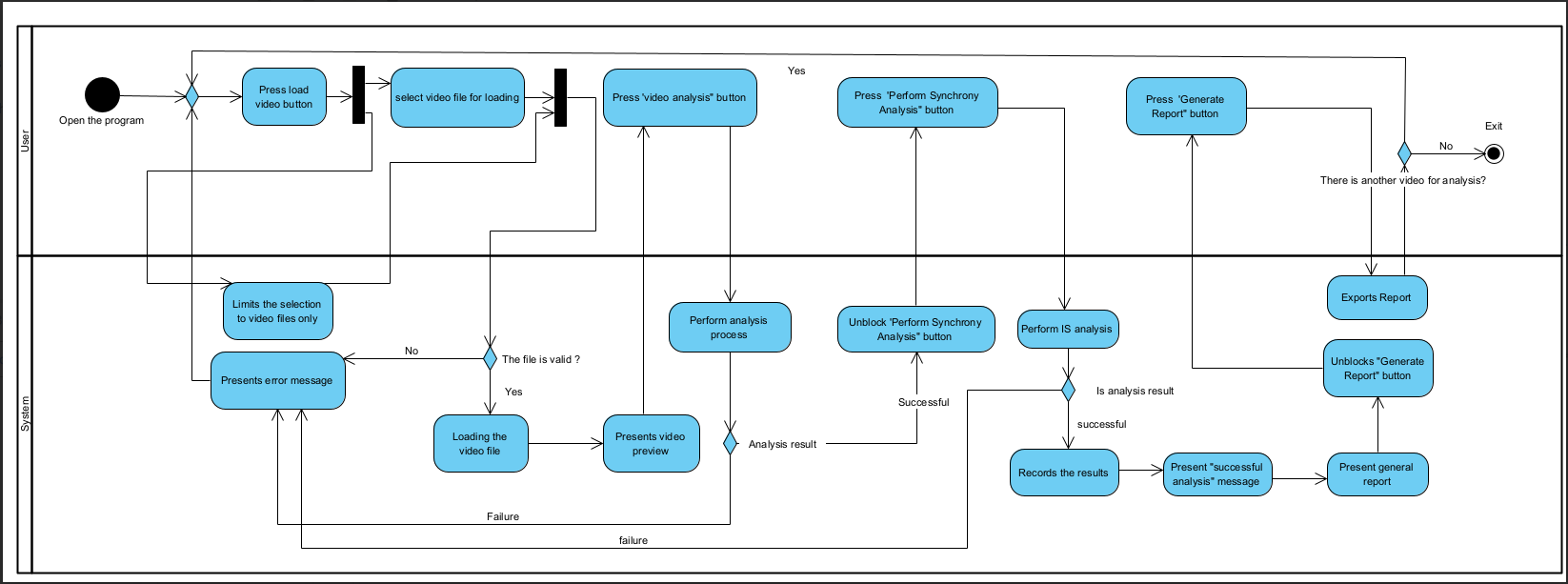
Description: Users can view the generated reports directly within the application interface or download them for offline access, sharing, or further analysis.

System Boundary

The system boundary encapsulates the software application for IS analysis, including its integration with MediaPipe for pose estimation, the analysis algorithm, and the reporting feature. It provides a platform for users to upload video recordings, perform IS analysis, and generate detailed reports, supported by documentation for users.



User Expected Interaction Activity Diagram:

Following activity diagram describes the user interaction flow for our intended program:

User Interaction Description:

1. User Opens the program.
2. User Opens Load Dialog
3. User Selects the Video File
   1. Add filter to limit the selection to video files supported.
4. System presents a video preview to User (?)
5. Optional – user can select video segment by giving time codes.
   1. Enter time code for ‘Analysis start’ (Optional – if no, defaults to 00:00:00 / ‘video\_start’, limited to ‘video\_end’)
   2. Enter time code for ‘Analysis end’ (Optional – if no, defaults to ‘video\_end’, limited to ‘video\_start’)
6. User presses a button to begin the video analysis process.
7. System performs the video keyframe extraction process.
8. As a result of a successful video extraction the System records following data:
   1. Keyframes, defined for our purposes as:
      1. Frames where a motion starts for ‘Actor A’ or ‘Actor B’
      2. Frames where a motion stops for ‘Actor A’ or ‘Actor B’
      3. Frames where a body location point disappears for ‘Actor A’ or ‘Actor B’
      4. Frames where a body location point appears for ‘Actor A’ or “Actor B’
   2. For each Keyframe – the System records the actor for which the data is relevant
   3. For each Keyframe – the System records the time code of the frame
9. System reconstructs body motion vectors for a video given at (3) and based on the data saved at (8), using the algorithm provided. Possible results for step (9):
   1. Analysis failed – video rejected. System shows to User a message with reason / Error, as example:
      1. Video failed to be opened by program
      2. No Persons detected
      3. Video too short
      4. Etc.?
   2. Analysis successful – video accepted and successfully analyzed. System shows to User a ‘Success’ message.
10. System unblocks for user the button – ‘Perform Synchrony Analysis’
11. User presses the button ‘Perform Synchrony Analysis’
12. System performs the IS analysis based on the provided Algorithm
13. System records the results of the IS analysis
14. System shows to User a message:
    1. In case of success – the system shows a ‘Success’ message
    2. In case of fail – the system shows ‘Fail’ message, possibly an ‘Error’ message
15. System presents a general report as a result of the analysis
16. System unblocks for user the ‘Generate Report’ buttons
17. (Optional) User can generate reports

**GUI Prototype:**

Based on the identified User Interaction process, and according to our decision to use Qt Framework with Python bindings we drafted a GUI prototype as follows:

This GUI prototype was created in the ‘Qt Designer’ development tool [26], provided by Qt Group as part of the Qt Framework ecosystem.

**A screenshot of a computer

Description automatically generated**

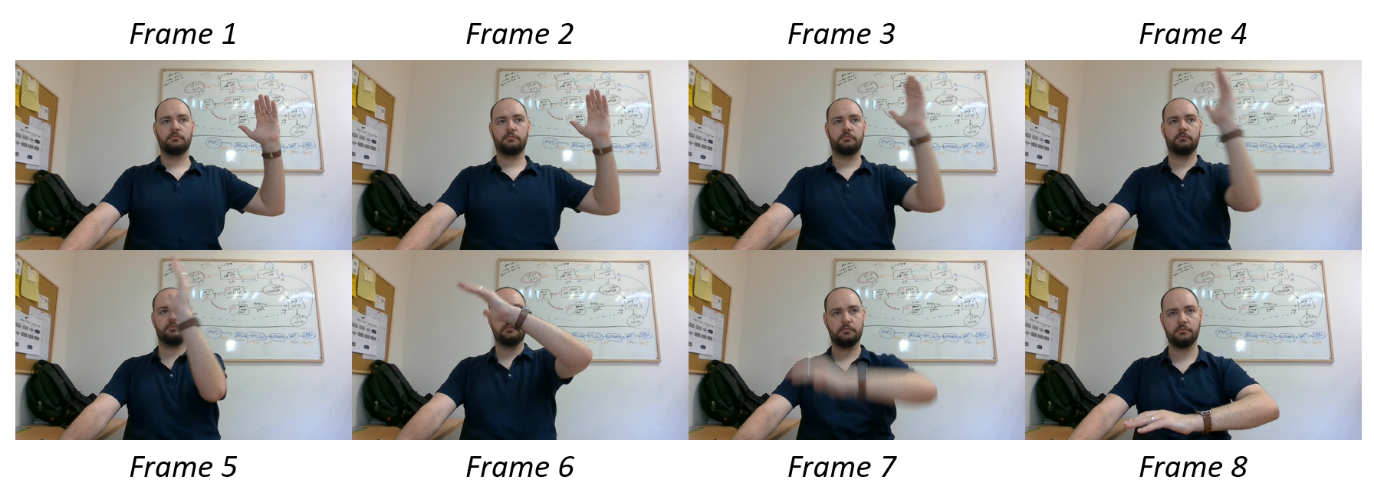
User Interaction

Area

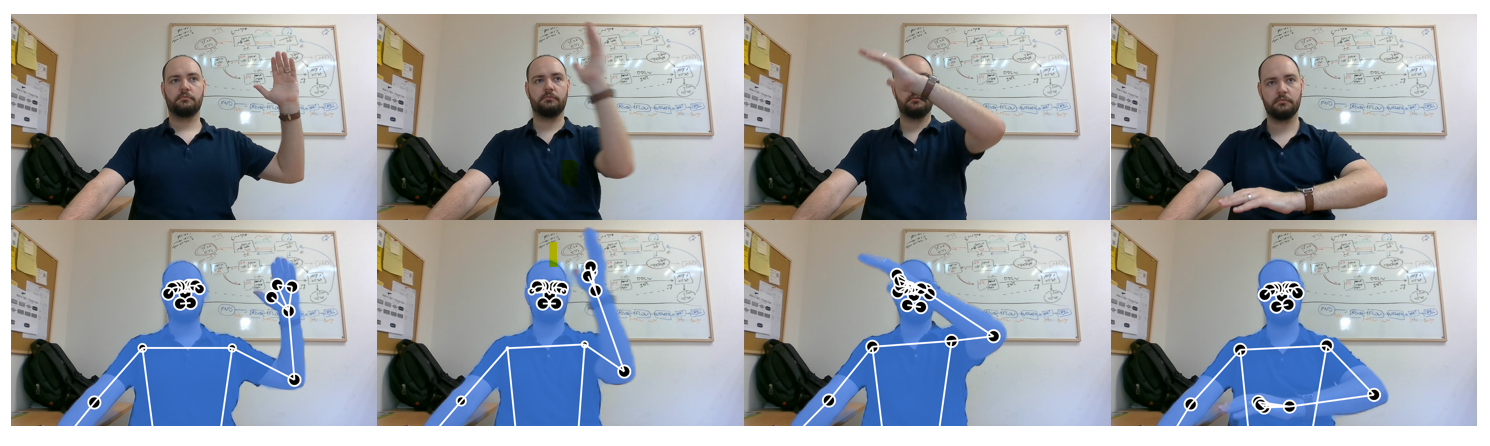
System Log and Messages

Multimedia Preview

**Pose Estimation Algorithm Prototype:**

An example of key frames extracted from a video, based on a frame difference peak estimation, with a­­ hand motion from left to right, then from top to bottom. The image is flipped on the vertical axis due to camera mirroring.  


Pose landmark analysis test performed on selected keyframes, using MediaPipe library:



# Testing Plan

The testing plan for our program is as follows:

|  |  |  |
| --- | --- | --- |
| **Use Case** | **Test** | **Expected Result** |
| **User Opens the Program** | Test GUI Initialization | The program GUI initializes and loads correctly without errors. |
| **Upload Video Recording** | Test Open File Dialog | The "Open File" dialog appears when initiated by the code. |
|  | Test File Type Filter | Only supported video file types are selectable in the dialog |
|  | Test File Opening and Loading | The selected video file loads correctly into the program |
|  | Test Video Handling | The program can handle various video sizes and lengths without errors. |
|  | Test Video Preview | The video preview displays correctly. |
|  | Test Video Playback Controls | Video playback controls (play, pause) function correctly. |
| **Set Video Analysis Settings** | Test Default Settings | Default settings are applied if no custom settings are specified. |
|  | Test Default Time Codes | Defaults to 00:00:00 for start and video end for end if not specified. |
|  | Test Time Code Setting for Estimation Start | Start time code is limited by end time code |
|  | Test Time Code Setting for Estimation End | End time code is limited by start time code |
| **Perform Pose Estimation** | Test Keyframe Extraction in Single Motion Single Person Video | Keyframes are extracted correctly according to frame difference for a single motion, single person video |
|  | Test Keyframe Extraction in Multiple Motions Single Person Video | Keyframes are extracted correctly according to frame difference for multiple motions, single person video |
|  | Test Keyframe Extraction in Single Motion Multiple Persons Video | Keyframes are extracted correctly according to frame difference for a single motion, multiple persons video |
|  | Test Keyframe Extraction in Multiple Motions Multiple Persons Video | Keyframes are extracted correctly according to frame difference for multiple motions, multiple persons video |
|  | Test Pose Landmark Detection for a Single Person Video | The system identifies and accurately tracks pose landmarks for a given frame using the MediaPipe library. |
|  | Test Pose Landmark Detection for a Multiple Persons Video | The system identifies and accurately tracks pose landmarks for a given frame using the MediaPipe library. |
|  | Test Keyframe Data Recording | The system correctly records the actor and time code for each keyframe. |
|  | Test Keyframe Vector Extraction | Vectors are extracted and calculated correctly where motions start/stop and points appear/disappear. |
|  | Test Pose Estimation Performance and Speed | Pose estimation completes efficiently without significant delays. |
| **Analyze Interpersonal Synchrony** | Test Movement Vector Calculation | Movement vectors for each body part of every participant are correctly calculated. |
|  | Test Synchrony Analysis Using RWTLCC or DTW | IS analysis using RWTLCC or DTW method works as expected. |
|  | Test Analysis Failure Handling | System displays appropriate error messages for failed analyses (e.g., no persons detected). |
|  | Test Analysis Success Handling | System displays a success message for successful analyses. |
| **Generate Report** | Test Report Generation | The system generates a report summarizing analysis results. |
|  | Test Report Customization | Generated reports are customizable based on user preferences. |
|  | Test Report Formats | Reports are available in both PDF and CSV formats. |
| **View/Download Report** | Test Report Viewing | Reports can be viewed within the application interface. |
|  | Test Report Downloading | Reports can be downloaded correctly as a file in a correct format |
|  | Test Report Integrity | Downloaded reports maintain format and content integrity. |
| **User Interaction Flow** | Test Button Enabling/Disabling | System unblocks buttons at appropriate steps (e.g., 'Perform Synchrony Analysis'). |
|  | Test General Report Presentation | System presents a general report after analysis. |
|  | Test Generate Report Button | System enables 'Generate Report' button after analysis. |

**Citations:**

[1]. https://www.frontiersin.org/articles/10.3389/fpsyg.2018.01560/full

Feniger-Schaal, R., Hart, Y., Lotan, N., Koren-Karie, N., & Noy, L. (2018). The body speaks: Using the mirror game to link attachment and non-verbal behavior. *Frontiers in psychology*, *9*, 1560.

[2]. <https://www.pnas.org/doi/full/10.1073/pnas.1108155108>

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.

[3]. <https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2021.749710/full>

Fu, D., Incio-Serra, N., Motta-Ochoa, R., & Blain-Moraes, S. (2021). Interpersonal physiological synchrony for detecting moments of connection in persons with dementia: A pilot study. *Frontiers in psychology*, *12*, 749710.

[4]. <https://www.researchgate.net/publication/233807151_Interpersonal_Synchrony_A_Survey_of_Evaluation_Methods_across_Disciplines>

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.

[5]. <https://link.springer.com/article/10.1007/s10919-022-00410-9>

Novotny, E., & Bente, G. (2022). Identifying signatures of perceived interpersonal synchrony. *Journal of Nonverbal Behavior*, *46*(4), 485-517.

[6] https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5137339/

Rennung, M., & Göritz, A. S. (2016). Prosocial consequences of interpersonal synchrony. Zeitschrift für Psychologie.

[7] https://www.nature.com/articles/s41598-019-50960-0

Galbusera, L., Finn, M. T., Tschacher, W., & Kyselo, M. (2019). Interpersonal synchrony feels good but impedes self-regulation of affect. Scientific reports, 9(1), 14691.

[8] https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4168669/

Lumsden, J., Miles, L. K., & Macrae, C. N. (2014). Sync or sink? Interpersonal synchrony impacts self-esteem. Frontiers in psychology, 5, 1064.

[9] https://www.cambridge.org/core/books/abs/social-signal-processing/interpersonal-synchrony-from-social-perception-to-social-interaction/50D491B6C3AB7767858C80CF612C28A5

Chetouani, M., Delaherche, E., Dumas, G., & Cohen, D. (2017). 15 Interpersonal Synchrony: From Social Perception to Social Interaction. Social signal processing, 202.

[10] https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2019.02078/full

Markova, G., Nguyen, T., & Hoehl, S. (2019). Neurobehavioral interpersonal synchrony in early development: The role of interactional rhythms. Frontiers in Psychology, 10, 2078.

[11] https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2022.897015/full

Bowsher-Murray, C., Gerson, S., Von dem Hagen, E., & Jones, C. R. (2022). The components of interpersonal synchrony in the typical population and in autism: A Conceptual analysis. Frontiers in Psychology, 13, 897015.

[12] https://www.nature.com/articles/s41598-023-37316-5

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(1), 11150.

[13] https://molecularautism.biomedcentral.com/articles/10.1186/s13229-019-0305-1

Georgescu, A. L., Koeroglu, S., Hamilton, A. F. D. C., Vogeley, K., Falter-Wagner, C. M., & Tschacher, W. (2020). Reduced nonverbal interpersonal synchrony in autism spectrum disorder independent of partner diagnosis: a motion energy study. Molecular autism, 11, 1-14.

[14] https://www.nature.com/articles/s41598-023-42006-3

Plank, I. S., Traiger, L. S., Nelson, A. M., Koehler, J. C., Lang, S. F., Tepest, R., ... & Falter-Wagner, C. M. (2023). The role of interpersonal synchrony in forming impressions of autistic and non-autistic adults. Scientific Reports, 13(1), 15306.

[15] https://www.frontiersin.org/articles/10.3389/frobt.2019.00073/full

Bloch, C., Vogeley, K., Georgescu, A. L., & Falter-Wagner, C. M. (2019). INTRApersonal synchrony as constituent of INTERpersonal synchrony and its relevance for autism spectrum disorder. Frontiers in Robotics and AI, 6, 73.

[16] http://www.dscn.umd.edu/DSCN/papers/McNaughton\_Redcay\_2020\_Postprint\_InterpersonalSynchronyinASD.pdf

McNaughton, K. A., & Redcay, E. (2020). Interpersonal synchrony in autism. Current psychiatry reports, 22, 1-11.

[17] http://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2016.00516/full

Fujiwara, K., & Daibo, I. Evaluating interpersonal synchrony: Wavelet transform toward an unstructured conversation. Front Psychol. 2016; 7 (516)

[18] https://www.researchgate.net/publication/342664554\_Motion\_energy\_analysis\_MEA\_A\_primer\_on\_the\_assessment\_of\_motion\_from\_video

Ramseyer, F. T. (2020). Motion energy analysis (MEA): A primer on the assessment of motion from video. Journal of counseling psychology, 67(4), 536

[19] https://www.nature.com/articles/s41537-022-00283-3

Lopes-Rocha, A. C., Corcoran, C. M., Andrade, J. C., Peroni, L., Haddad, N. M., Hortêncio, L., ... & Loch, A. A. (2022). Motion energy analysis during speech tasks in medication-naïve individuals with at-risk mental states for psychosis. Schizophrenia, 8(1), 73.

[20] https://pubmed.ncbi.nlm.nih.gov/28587814/

Dean, D. J., Samson, A. T., Newberry, R., & Mittal, V. A. (2018). Motion energy analysis reveals altered body movement in youth at risk for psychosis. Schizophrenia research, 200, 35-41.

[21]<https://learnopencv.com/yolov7-pose-vs-mediapipe-in-human-pose-estimation/>

[22]<https://www.toolify.ai/ai-news/amazing-ai-pose-detection-with-python-and-mediapipe-84599>

[23]<https://developers.google.com/mediapipe/solutions/vision/pose_landmarker>

[24]<https://viso.ai/deep-learning/pose-estimation-ultimate-overview/>

[25] <https://www.pythonguis.com/faq/pyqt-vs-tkinter/#an-introduction-to-python-gui-libraries-in-python>

[26] <https://doc.qt.io/qt-6/qtdesigner-manual.html>