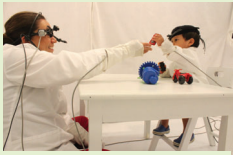


# Building an Interactive Visual Tool to Visualize Multimodal Data from Dyadic Social Interaction

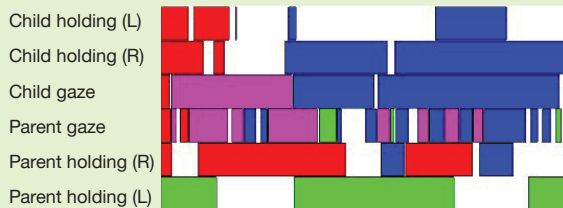
Ethan Vogelsang, Seth Foster, Dr. Chen Yu - Indiana University, Department of Psychological and Brain Sciences

## Introduction

- A trend in contemporary behavioral research is data-driven, relying on collecting, annotating and data-mining high-density multimodal data.
- A challenge in data-driven science is to find effective ways to analyze huge volumes of data, discover reliable patterns and ultimately acquire new knowledge.
- The goal of the current project is to develop an interactive MATLAB tool that aids to explore and query time series derived from raw multimedia data.
- It is characterized by an intuitive design that can quickly process data from multimodal behaviors to form a visual representation during key moments of the study.
- The current study is a naturalistic free-flowing interaction between a parent and child.
- Audio files were transcribed to find moments when parent talked about the toys (objects).
- Eye trackers were used to capture the first person views of the parent and child.
- Point-of-gaze was calculated at a rate of 30 Hz. This enables us to determine where the child was looking at any point during the interaction.



## An Example of Continuous Data Streams



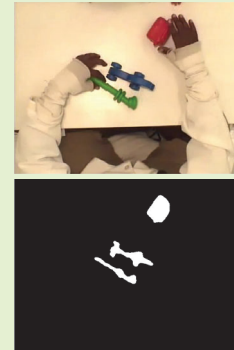
- We can ask questions such as: *How often does the child look at the object that they are holding?*
- The MATLAB tool aims to enable researchers to quickly examine the results of these types of questions.

Agent/Modality	Frequency	Resolution
Child Vision	30 Hz	640*480 px
Child Gaze	30 Hz	0.50 degrees
Child Head Motion	120 Hz	0.01 in
Child Hand Motion	120 Hz	0.01 in
Parent Vision	30 Hz	640*480 px
Parent Gaze	30 Hz	0.50 degrees
Parent Head Motion	120 Hz	0.01 in
Parent Hand Motion	120 Hz	0.01 in
Parent Speech	44.1k	N/A

## Design and Implementation

### Detecting Object Location with Image Processing

- As part of the free-flowing interaction, objects are moved around the table frequently.
- The spatial arrangement of objects may be an important consideration when analyzing behaviors that act on those objects.
- The top right image shows the spatial arrangement from a bird's eye point of view.
- An object detection algorithm is used to produce an object mask image (below), from which we can calculate the centroid of each individual object.
- The centroids allow us to represent the location as a single point on a graph.
- For behaviors that last many seconds, we can plot the mean centroid location during this time range.



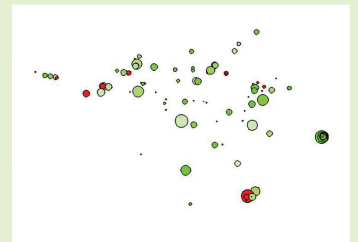
- Search for behavior by modality, e.g., *eye hand* will list all variables relating to eye gaze and holding objects with hands.
- The main plot, each data point represents the location of the target object for each event in the chosen behavior. The coordinates are from a bird's eye view.
- When user clicks one of the scatter plot dots, it will open the first person views during that moment of the interaction.

Attribute	Meaning
Size	Duration of the variable1 events. Longer events have larger circles.
XY-Position	Mean centroid of the target object during the variable1 events. This is from the bird's eye view. Points closer to the top are closer to the child.
Color	The proportion of time that variable2 occurs <i>during</i> variable1 moments. E.g. Proportion of time that child looks at the object while they are holding that object. Red indicates a proportion of zero. The full scale is shown below.

## Example Results

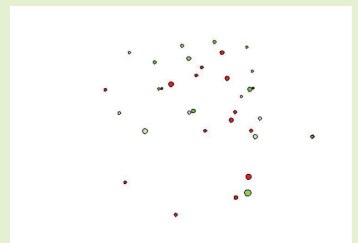
Variable1 (position/size):  
child holding object

Variable2 (color):  
child looking at object



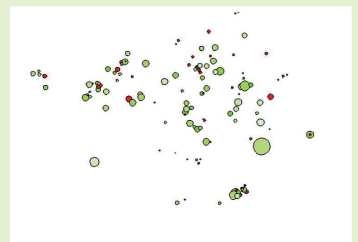
Variable1 (position/size):  
parent naming object

Variable2 (color):  
child looking at object



Variable1 (position/size):  
child looking at object

Variable2 (color):  
parent looking at object



## Summary

- We believe this visualization technique can be used to present data in an informative way that makes it easier for researchers to employ their own domain knowledge and expertise to detect new patterns that were previously overlooked or difficult to find, and to generate new hypotheses which will lead to new discoveries.
- We plan to extend the functionality to include up to 3 variables at a time. In this case, the user can specify how the coloring scheme will map to the different combinations of behaviors that can occur during the interaction.
- For example, we can analyze whether the child looks *and* holds a target object that is also being talked about by the parent. Analyzing cross-modal behaviors in this way can prove to be insightful.