Data Maps for Video Interaction Analysis

Nikolas Martelaro - Center for Design Research

Problem & Motivation

Analyzing interaction in video can be time-consuming and tedious. Within my own physical interaction design work, I have often spent hours watching videos in order to extract interesting interactions. Most of the time spent watching videos is skimming over uninteresting, low interaction sequences. However, as data-logging from physical objects becomes more widespread, there is an opportunity to use these data streams to aid exploration of video. Using the data as a map for exploring the video can help speed analysis time and allow for more time examining and understanding the interactions.

Approach

I used a video and associated dataset from a driving interaction study. One sample drive with a 45 min video speed data, and interaction data from a remote controlled text-to-speech system was used. I originally intended to try various distorted time-axis displays, shown in Figure 1. I found that accurate time scale and easy point selection were the more critical for an analyst. The final prototype uses linear time scales and focuses using tool-tips. The software automatically highlights areas of high interaction using threshold filters. Data highlights are linked across the graphs to help understand how data and interactions are related. The video is linked to the data so that you clicking data scrubs to that moment in the video.

Results

The current prototype allows for fast scrubbing of the video by using the data as a guide. This allows someone to move quickly and explore interesting moments in time. The data highlighting is responsive and helps to see how different interactions in the data overlap. One can quickly work across data streams to find interesting areas of the video to explore and avoid areas of non-interest.

Future Work

The tool can be extended to support analysis by adding features such as note taking and brushing to let the user highlight data. This type of interface could also be useful for live data and video streams by automatically highlighting areas for future review.

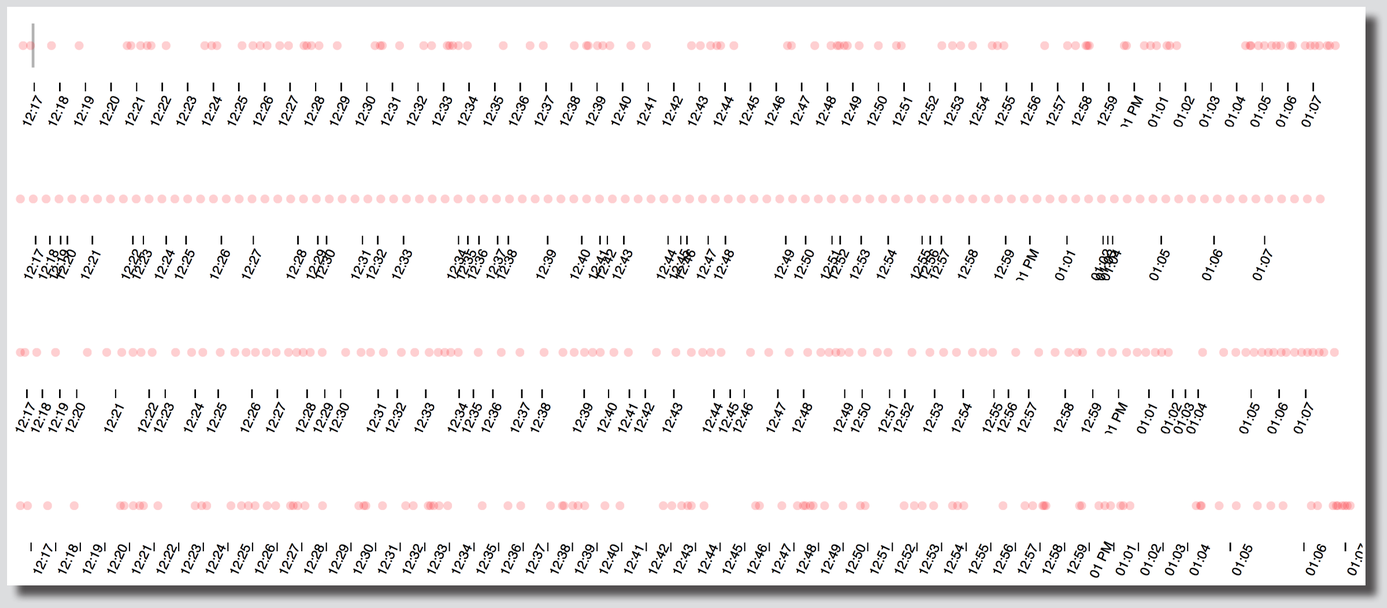

Figure 1 - Remote controlled text-to-speech timelines with different scalings.
1) *Linear time scale* - this time scale gave an accurate view of the data and showed were moments of high interaction were based on the density of data points.

2) *Equal spacing of data points scale* - This scale equally spaced the data points an allowed for easy selection of the data. The time axis warping showed the inverse of the data density (spread out times show lots of interaction). While this makes points easy to select, it is harder to understand where interesting sections of interaction are located.

3) *Backspaced smoothing time scale* - This time scale moved points with large distances between others backward, smoothing out the amount of whitespace. This made point selection somewhat easier and still showed some data density. Unfortunately, the mix of the warped time axis and overlap in the data leaves one wishing for either better selection or accurate data display.

4) *Dynamic x-axis fish-eye focusing* - This scale used a simple, dynamic fish-eye focus to spread the data out when hovering over a point. While this helped spread the data locally, the dynamic motion made clicking points more difficult and the warped time scale could make some areas seem more interesting because they became compressed.

Although the warped time scales were an interesting exploration of making the data more easily explored, they ultimately were more difficult to use and lead to inaccurate understanding of the data. After testing these myself and with other researchers in my lab, I found that easy selection of points was the most important aspect of the interaction for the tool. The final prototype reflects this by using a linear time scale with point focusing to select individual points.

Figure 2 - Current prototype
The current prototype allows exploration of the video using the linked datasets. Users click on data points in either the speed or speech interaction graph to scrub to that point in the video. Areas of high interaction are highlighted across graphs to allow exploration across data types. Individual points are brought into focus using tool-tips to allow accurate data selection.