Audio and Visual Synchronization for Data Visualization

Michael Castellanos[1], Jack Weissburg[2], Blake Williams[3], Laura Vitale[4], Michael Scordilis[5], Daniel Messinger[6]

[1]University of Miami, [2]University of Wisconsin-Madison, [3]University of Vermont, [4]University of Miami, [5]University of Miami, [6]University of Miami

Introduction

- Collection of high-frequency behavioral data (movement, orientation, first-person audio). Presents challenges for analysis and comprehension.
- Data visualization is essential for the comprehension of complex behavioral interactions.
- Combined audio and visual representations, although rarely used, may be an important research tool.

Data Collection

- Audio of each interaction between all participants was recorded using the LENA recorders.
- Children wore first-person LENA recorders. These recorders were subject to drift, indicating the need for audio alignment.
- Real-time tracking of child location was collected using the Ubisense Dimension 4 location system. Children wore vests with left and right tags recording location at 6Hz using radio frequency ID.

LENA Recorder

Vest (rear view) with RFID tags shown

Audio Alignment

- To align the audio, a modified version of the cross-correlation method to align the first-person audio recordings.
- Cross-correlation on the 4.5 hour audio files, on average, was performed in the frequency domain to ease computational overhead.
- The earliest recording of that day is chosen as the "master LENA" and the rest of the recordings are synchronized to that recording.

Audio files before alignment

Audio files after alignment

Data Processing Pipeline for Visualization

- The audio data is aligned using a cross correlation method.
- Ubisense tag data was extrapolated to every 1/10th second using the UL processor.
- The audio files are then reduced to just social contact events where the individual is speaking, per the recording.
- The movie and audio files are then combined into a single file.(Due to sensitive information, the actual speech has been replaced with sine tones)

UbiSense/ADEX Data Branch

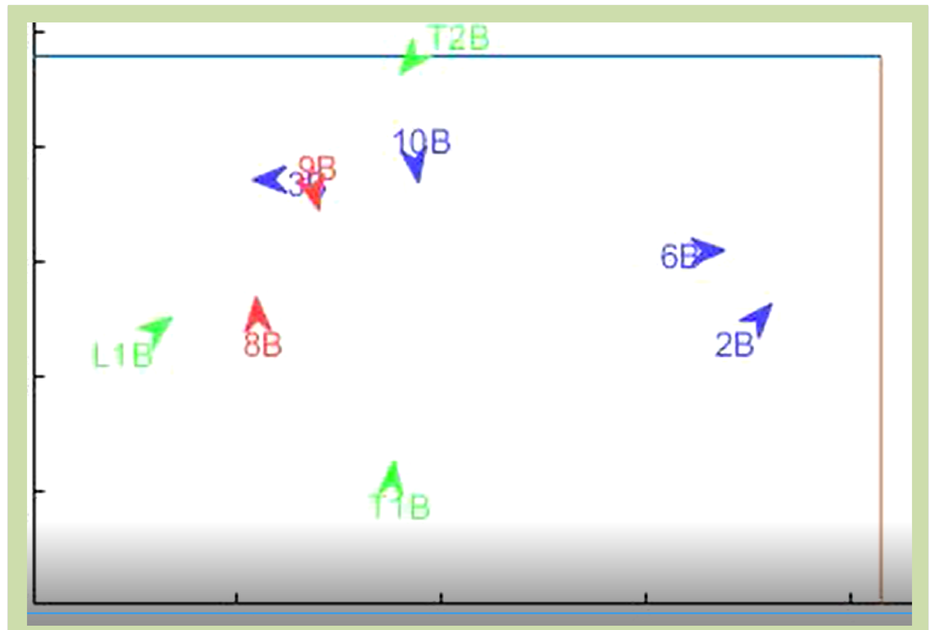
Ubisense

LENA ADEX

Ubisense Location Data

Visual Studio Platform for UL Processor

Matlab was used for movie generation



LENA Data Branch

LENA Pro

LENA foundation that produces the recorders

Spyder IDE for python code to align audio

Audacity software to output the final recordings.

Visualization

- Shows orientation of every person in the room.
- Utilizing two tags per child allowed heading to be determined. For each child, heading and location were found from the left and right tags.
- Ubisense tag data was extrapolated to every 1/10th second. Gaps in individual tag data longer than 60s were recorded as missing.
- Please refer to the laptop on the table for a real-time demo of the synchronized visualization.
- When individuals are engaged in social contact, the individual who is speaking color will get darker.

Person is not speaking

Person is speaking

Conclusions

- Using the generated audio/visual movies is both a good method of representing the data in a human-perceivable way, as well as validating the results that were found during data analysis.
- The movies act as an integrity checkpoint for any validation that occurs to verify the data. I.e. some speech event occurs at a given timestamp, one can now go and check if the speech event did in fact occur at that time.

Acknowledgements

- This material is based upon work supported by the National Science Foundation under Grant No. CNS-1659144.
- We would like to thank Chaoming Song, PhD, and all others involved in the *IBSS-L: Continuous Measurement of Children's Behavior and the Development of Social Dynamics* (1620294) project.
- We thank Burt Rosenberg, principal investigator to the REU site: *Scientific Computing for Structure in Big or Complex Datasets.*

m.castellanos29@umiami.edu