Using Automated Interactional Synchrony to Predict Complex Behaviors

Jessica Tang, Jasper Hu, Damilola Oshunyinka
Project Manager: Daniel Parr, Project Lead: Dr. Jana Schaich Borg

Introduction

Our project's primary goal is to curate a unique data set for studying how different aspects of social synchrony relate to social and psychiatric variables, such as trust, empathy, and scores on clinical social competence and autism scales. This work will help researchers better understand the human brain and treat psychiatric diseases.

- Social synchrony: the coordination or alignment of behavior or emotions between people

Objectives

The initial audio and video files come from pairs of participants who interacted with each other and filled out behavioral surveys. Our objectives:

1. Propose preprocessing steps for audio feature extraction, such as methods to help with noise reduction and signal smoothing.
2. Extract audio features.
3. Plot a correlogram to depict the correlation between all pairs of the 17 types of facial movements by type of social synchrony measure (Derivative Dynamic Time Warping, Pearson Correlation, Spearman Correlation, Motion Energy Analysis, Window Cross Correlation)

Methods

Audio Feature Extraction:

To understand synchrony between 2 interacting individuals, it is important to investigate the dynamics of the audio features between the individuals. We proposed to extract two auditory features: **pitch and loudness**.

Pitch:

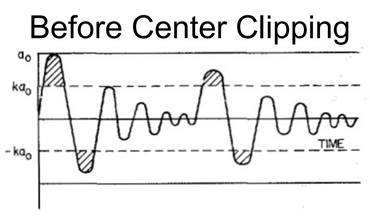
Preprocessing steps:

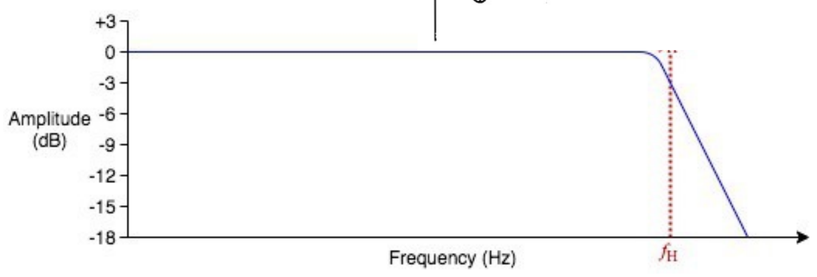
- Center clipping

Helps reduce noise

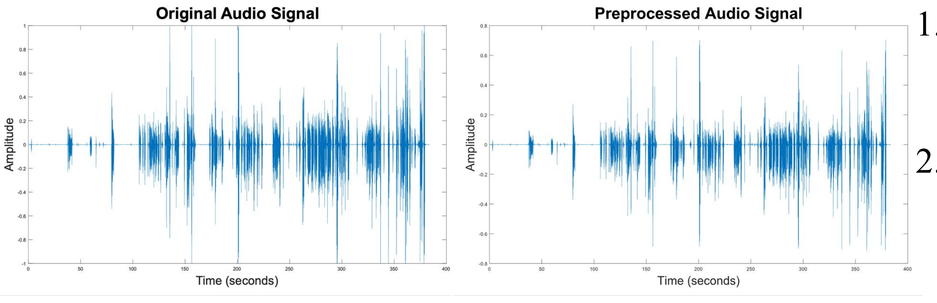
- Low pass filtering

Eliminates frequencies above the cutoff frequency (900 Hz)



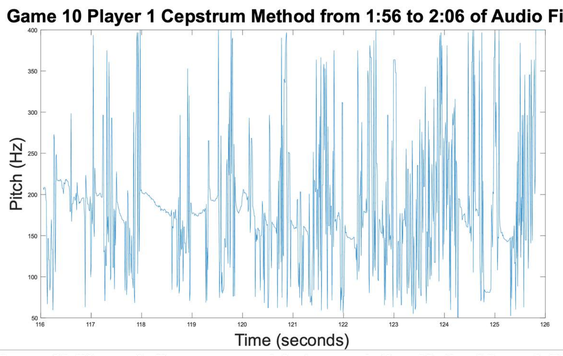


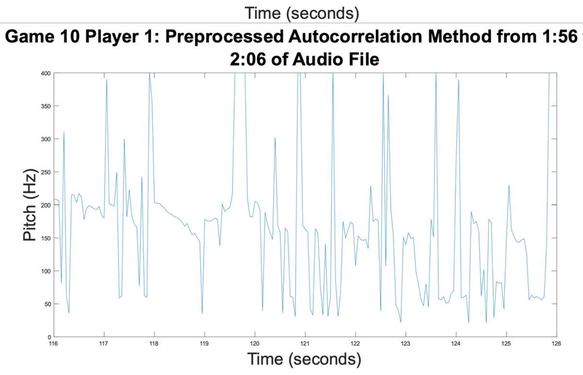
Low Pass Filtering



Before and After Preprocessing Steps

Methods (continued)





Cepstrum Method

- Uses full- band speech signal
- Does not use preprocessing steps

VS.

Autocorrelation Method

- Uses low-passed speech signal
- Uses preprocessing steps

(using one participant's audio file, we looked at a ten second period where she was speaking and extracted pitch using the two methods)

Loudness:

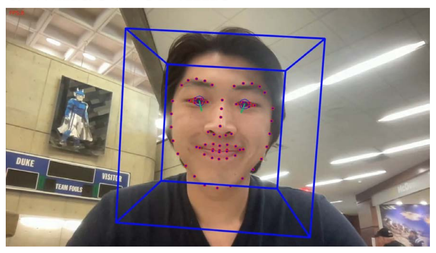
No Preprocessing steps

2 extraction methods used:

1. Extracting decibels from audio file
2. Extracting power through finding the average squared value of amplitude within a window of time

Video Feature Extraction:

- Action Units: anatomically distinct sets of facial muscle movements that describe/ categorize facial expressions (ex: raising lips)
- OpenFace: an open-source deep-neural network framework used for facial action units recognition
- Smooth signal to reduce noise: extract significant points in time for sparse signals
- Matching Pursuit: represent sparse signals with set of basis functions to remove noise with high frequency and low amplitude



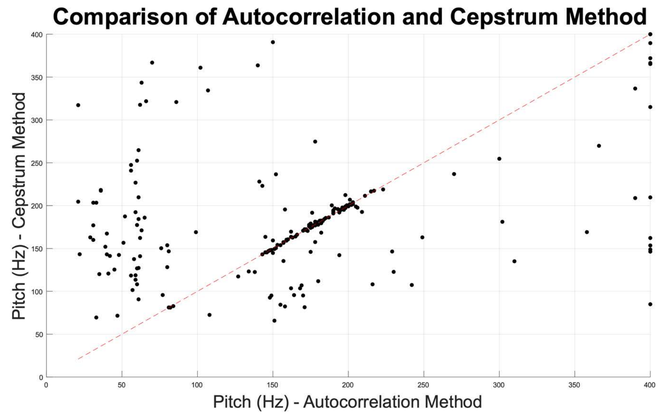
Upper Face Action Units					
AU 1	AU 2	AU 4	AU 5	AU 6	AU 7
Inner Brow Raiser	Outer Brow Raiser	Brow Lowerer	Upper Lid Raiser	Cheek Raiser	Lid Tightener
*AU 41	*AU 42	*AU 43	AU 44	AU 45	AU 46
Lid Droop	Slit	Eyes Closed	Squint	Blink	Wink
Lower Face Action Units					
AU 9	AU 10	AU 11	AU 12	AU 13	AU 14
Nose Wrinkler	Upper Lip Raiser	Nasolabial Deepener	Lip Corner Puller	Cheek Puffer	Dimpler
AU 15	AU 16	AU 17	AU 18	AU 20	AU 22
Lip Corner Depressor	Lower Lip Depressor	Chin Raiser	Lip Puckerer	Lip Stretcher	Lip Funneler
AU 23	AU 24	*AU 25	*AU 26	*AU 27	AU 28
Lip Tightener	Lip Pressor	Lips Part	Jaw Drop	Mouth Stretch	Lip Suck

Results

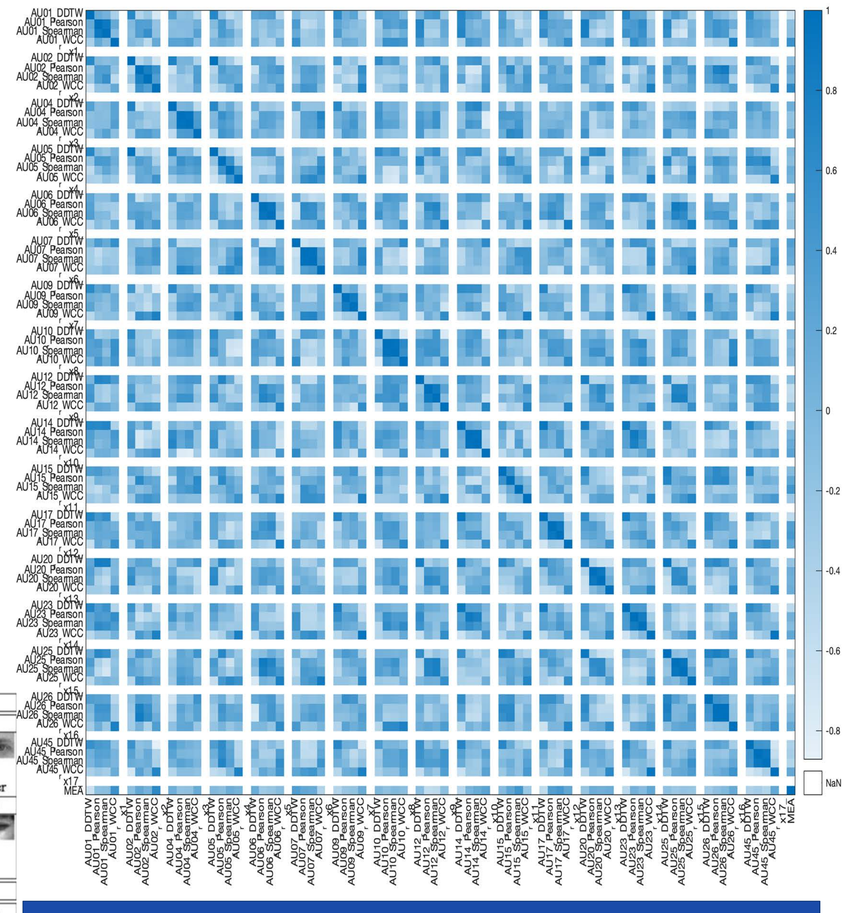
Pitch Extraction:

1. The pitch preprocessing steps for the Autocorrelation Method led to an overall smaller scaled amplitude range for the waveform.
2. Using the Autocorrelation Method with center clipping and low pass filtering yields less noise, as shown through the audio signal and pitch estimation graph.
3. The correlation between the two pitch extraction methods are weak. The r-value is 0.3264.

Results (continued)



The correlogram depicts the correlation between various synchrony measures on different action units across all game sessions.



Conclusion

- We defined preprocessing steps and methods for feature extraction. After comparing both methods of pitch extraction, we conclude that the Autocorrelation Method suits our project's purpose best.
- Our data was successfully cleaned for further analysis. Previously undeployable OpenFace code was successfully re-written.
- DDTW seems to produce similar results as Pearson and Spearman correlations, while WCC seems to produce values that differ from the other methods the most.

We hope to aid future researchers better understand interactional synchrony in regards to audio and video features.

References

(1) Arthur. "Audio Eq: What Is a Low-Pass Filter & How Do Lpfs Work?" *My New Microphone*, 30 Jan. 2021, mynewmicrophone.com/audio-eq-what-is-a-low-pass-filter-how-do-lpfs-work/. (2) M. Sondhi, "New methods of pitch extraction," in IEEE Transactions on Audio and Electroacoustics, vol. 16, no. 2, pp. 262-266, June 1968, doi: 10.1109/TAU.1968.1161986. (3) Zhi, R., Liu, M. & Zhang, D. A comprehensive survey on automatic facial action unit analysis. Vis Comput 36, 1067–1093 (2020). https://doi.org/10.1007/s00371-019-01707-5