Deception Detection - Design

Team 2 - Socratis, Daniel, Miguel, Tarekul, Maureen

Introduction

- Thermal analysis will not be efficient in the long run for detecting lies.
- Four facial analysis features, five vocal analysis features.
- Open source libraries, software, data, and tools.
- Attempt analysis not only from interview data, but existing public interviews.
- Possible high training complexity.

Background - Facial Analysis

Facial Micro-Expressions - Are involuntary facial changes (micro movements) that are not discernable to the human eye.

Eye Blink Patterns - Research states that people who lie tend to slow down their blink rate followed by a rapid increase of their blink rate.

Gaze Direction - People tend to avert their gaze when lying

Pupil Dilation - Relatively conclusive indicator of a person lying. Enlarged pupils indicate that the brain is working hard which is how a lie can best be executed.

Background - Vocal Analysis

Vowel Duration - The length of time that has gone by when a vowel is said.

Formant Frequencies - Is a concentration of acoustic energy around a particular frequency in the speech wave.

Amplitude at the first formant frequency - The volume at the first formant frequency.

Pitch Contour - Rise and fall of the voice pitch.

Pitch Amplitude - The volume of the pitch.

Voice Methodologies

- Step 1:

1. Multiply the frame (frame i) by a Hamming window:

$$y_i(j) = s(j).w(j) \quad \text{where } w(j) = (0.54 + 0.46\cos(2\pi j / 255))$$

2. Compute the energy of the frame:

$$E_i = \sum_{j=0}^{256} y_{ij}^2 \quad ,i = 1 \ldots, N$$

Where N is the total number of the frames in an utterance .

3. Compute the FFT for the frame:

$$Y_i(f) = \text{Real}\{\text{FT}(y_i[j])\} \quad .$$

4. Compute the logarithm of the magnitude spectrum of the frame :

$$Y_i(f) = 20\log|Y_i(f)| \quad .$$

5. Compute the Inverse FFT for the frame to get the output of the homomorphic operation (Cepstrum).

$$y_i(j) = \text{IFT}\{Y_i(f)\} \quad .$$

This output contains information of both the vocal tract (lower part of the time scale) and the excitation (high part of the time scale). This could be shown in Fig. 4.

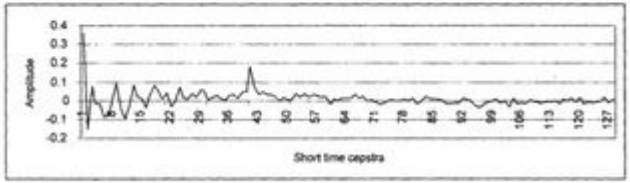


Fig. 4 typical output after IFFT via homomorphic analysis

6. Beginning from a high time sample location (40 for male, 20 for women and children) seek for the peak in the remainder of the frame.

$$X_i = \max_{\substack{j=40 \\ or\ 20}}^{128} y_i(j).$$

P_i = index (Xi).

$F P_i = i$.

$Ai = |X(i)|$.

Where P_i is the pitch period at frame i, $F P_i$ is the frame number whose pitch period is P_i and Ai is the amplitude of the signal at P_i.

- Step 2: For the whole voiced interval do the following :

1. Compute the mean energy for the utterance.

$$\overline{E} = (\sum_{i=1}^{N} E_i) / N \quad , \quad N = \text{total number of frames}.$$

2. Compute the maximum pitch Amplitude for the utterance.

$$A_{max} = \max_{i=1}^{N} X(A_i) \quad .$$

3. Compute the vowel duration in msec [11].

$$Vd = 0.23 \sum_{i=1}^{N} (V/2) \quad ,V = \begin{cases} 1 & \text{for v} \quad \geq s \\ 0 & \text{elsewhere} \end{cases}$$

Where s is a percentage threshold of Amax

4. Compute the mean of the pitch amplitudes:

$$\overline{AP} = (\sum_{i=0}^{m} A_i)/an \quad , A_i = \begin{cases} A_i & \text{for } P_i(A_{max}) + 5 > P_i(A_i) > P_i(A_{max}) - 5 \\ 0 & \text{elsewhere} \end{cases}$$

, an .... the number of P_i's that that lie in the range[11] :

$$(P_i(A_{max}) + 5) > P_i(A_i) > (P_i(A_{max}) - 5).$$

Voice Results

Vocal Measure	*Accuracy*
Mean energy	61%
Mean pitch amplitudes	78%
Max. pitch amplitudes	86%
Pitch contour	93%
Vowel duration	82%
H&W scores	74%

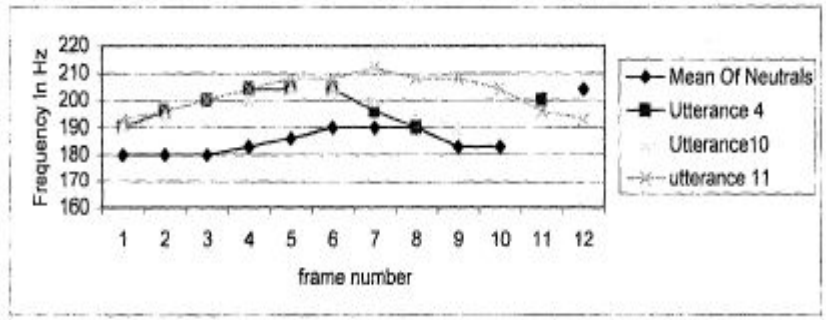


Fig. 5 Pitch contours plot for a suspect.

Spikes in frequencies are clear indications of deception

System Design

Preliminary data processing will be split amongst the group. Machine learning will be done by everyone for experience purposes.

Digital Signal Processing with ThinkDSP library

Facial Recognition with OpenCV library and Google Vision API

Machine Learning with TensorFlow

System Design

Algorithms:

Voice - (Mahmoud E. Gadallah, Matar A. Matar, Ayman F. Algezawi) calculations for speech analysis.

Hirsch and Wiegele scoring method to enhance poor reliability.

Face - (Vahid Kazemi, Josephine Sullivan) Facial landmark detection using a cascade of regression functions (uses 68 points to track jaw, mouth, nose, eyebrows, eyes)

Track Eye Blink using EAR method developed by Tereza Soukupova and Jan Cech

********** DEMO *********

Data Collection

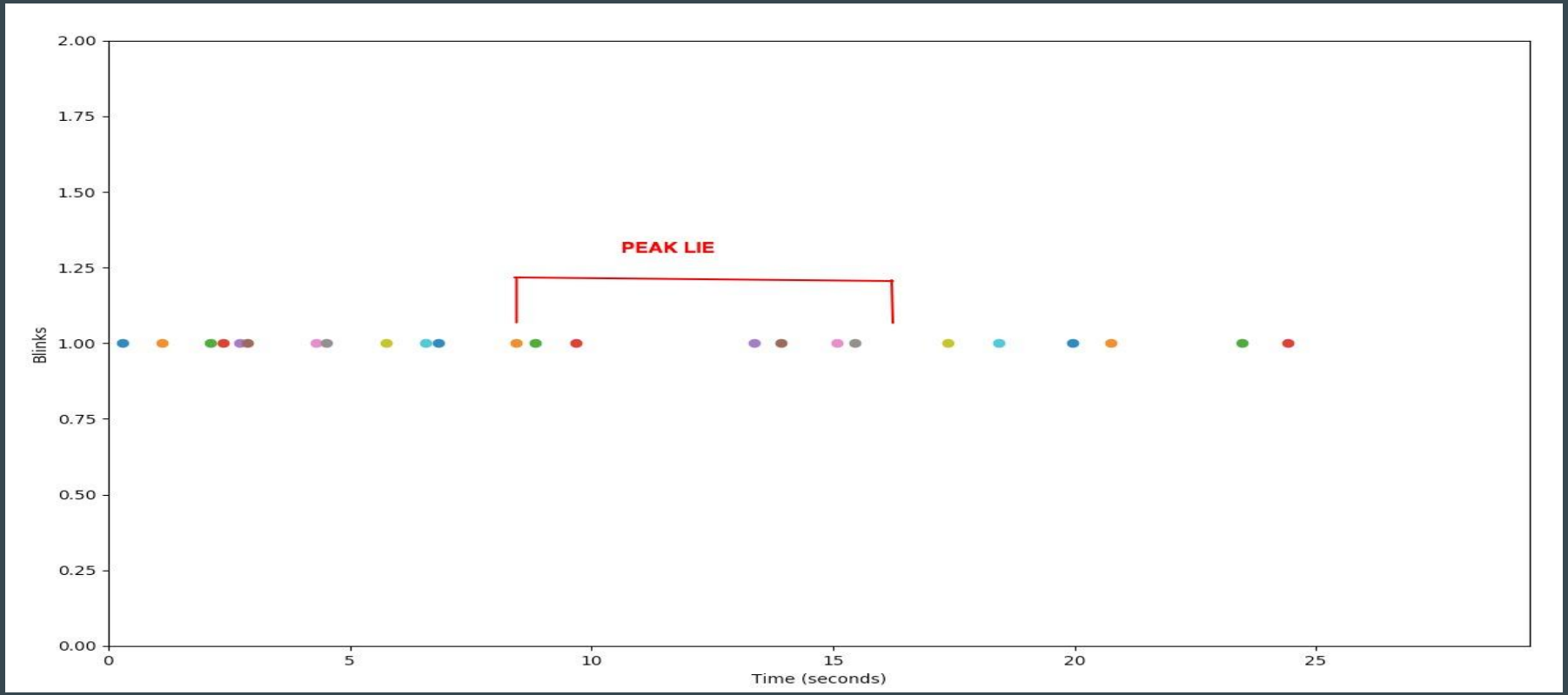
We will be using a set of approximately a set of 20+ questions in our data collection. The questions will be a mix of simple, critical thinking, and control questions.

Control questions help us create a baseline data collection of a particular subject. The control question is meant to serve as basis when a subject is telling the truth. We use that as a measure of comparison to when we ask relevant simple/critical thinking questions, we can compare arousal.

When a subject is lying his/her arousal will be compared to the baseline arousal from the control questions. Our control questions give us truthful answers and arousal associated with truth.

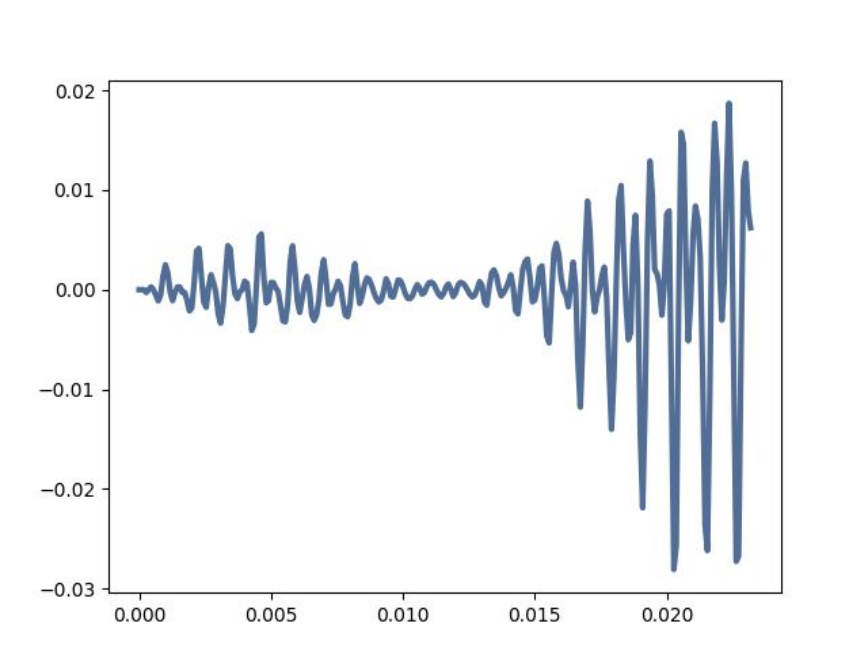
In addition to interviewing individuals, we will be using controversial president videos for preliminary and training data.

Initial Visualizations - Face (Test Run)

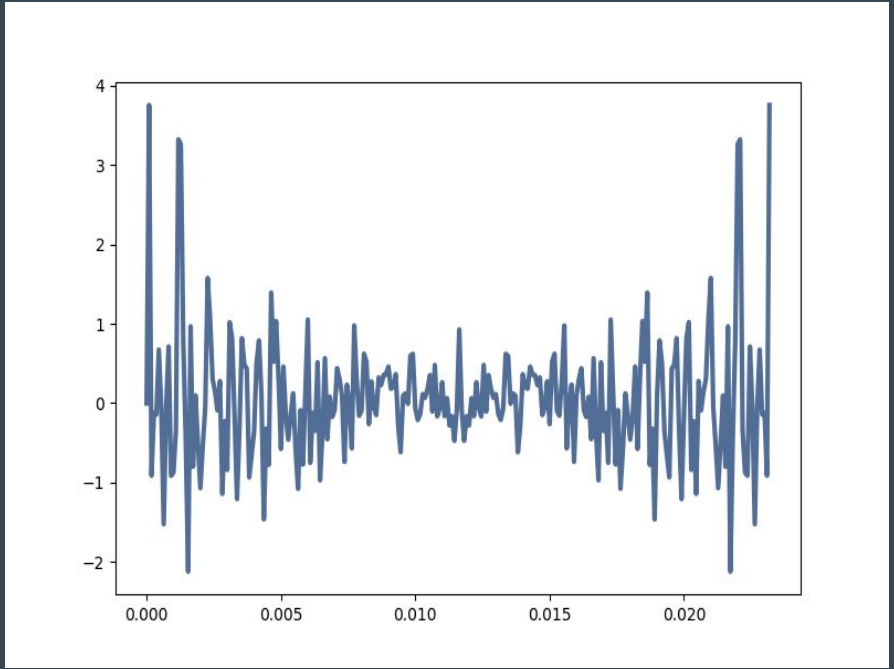


Initial Visualizations - Voice

Before processing:



After processing:



Discussion