

# Using Videos for Political Science: In Case of Campaign Speech and Emotion

Hyun Woo Lim

Washington University in St. Louis

## Motivation

People perceive information not only from text but also from audio and image in communication, but despite their importance, existing research leaves this question unanswered: Do non-textual features of speech affect voters via emotional expression?

- Emotions are important media of political message and powerful tools in communication
- Texts are not enough to capture the non-textual effect of communication
- Difficulties in systematically measuring speaker's emotion from video data

# Introduction

Consider global warming in speech. The way to convey the exact same word can vary in the context of the speech. However, we are unsure about how various textual and non-textual features of communication affect reception of the information and the perception of speaker. Therefore, with Chris Lucas, I collect 2016 Congressional campaign speech video data from Youtube to use in an experiment in which we can test the effect of emotional expression on those two outcomes. In order to analyze comprehensive aspects of speech in video, we need to analyze text, audio and images from video.

## Overview of literature

- The role of emotions in politics: Valentino et al. 2011, Brader 2005
- Computer vision: Knox, Lucas and de Benedictis-Kessner 2016, Torres 2018
- Audio: El Ayadi, Kamel and Karray 2011
- Text: Tumasjan et. al. 2010

## Method

## Download videos from Youtube

- Crop Youtube videos of 2016 U.S. Congressional elections winners by keywords ("candidate name + campaign speech").
- Download top 3 results of each winners (Total about 60GB).

## Video text

- Cut extracted audio into pieces
- Run Google Speech API for each audio blocks
- Transcribe the audio blocks and merge them

## Video Audio

- Extract audio from video
- Analyze the audio: Amplitude, pitch, length etc.

## Video Images

- Recognize the face of a politician from video
- Use pre-trained face classifier for detecting emotions (Goodfellow 2013)
- Capture and analyze the facial emotions of the politician for each video frames

## Video text

The split audio file can be transcribed in Google Speech API. Text sources are useful data and there are various models to analyze text, but without a video and audio component of speech, we can only understand limited information from the communication. For example, the same word can be presented in different ways with different emotions that text data itself cannot capture that difference.



Figure 3: Frequent words in the speech

## Video audio

Speaker's emotions and non-textual implication of message can vary. These variations in audio can be measured by amplitude, pitch, and speed of speech. Although these two audio files are from the same speaker, the amplitude of speech tone differs by the context and message of the speech. For an experiment, I can manipulate the speed, tone or pitch of speech.

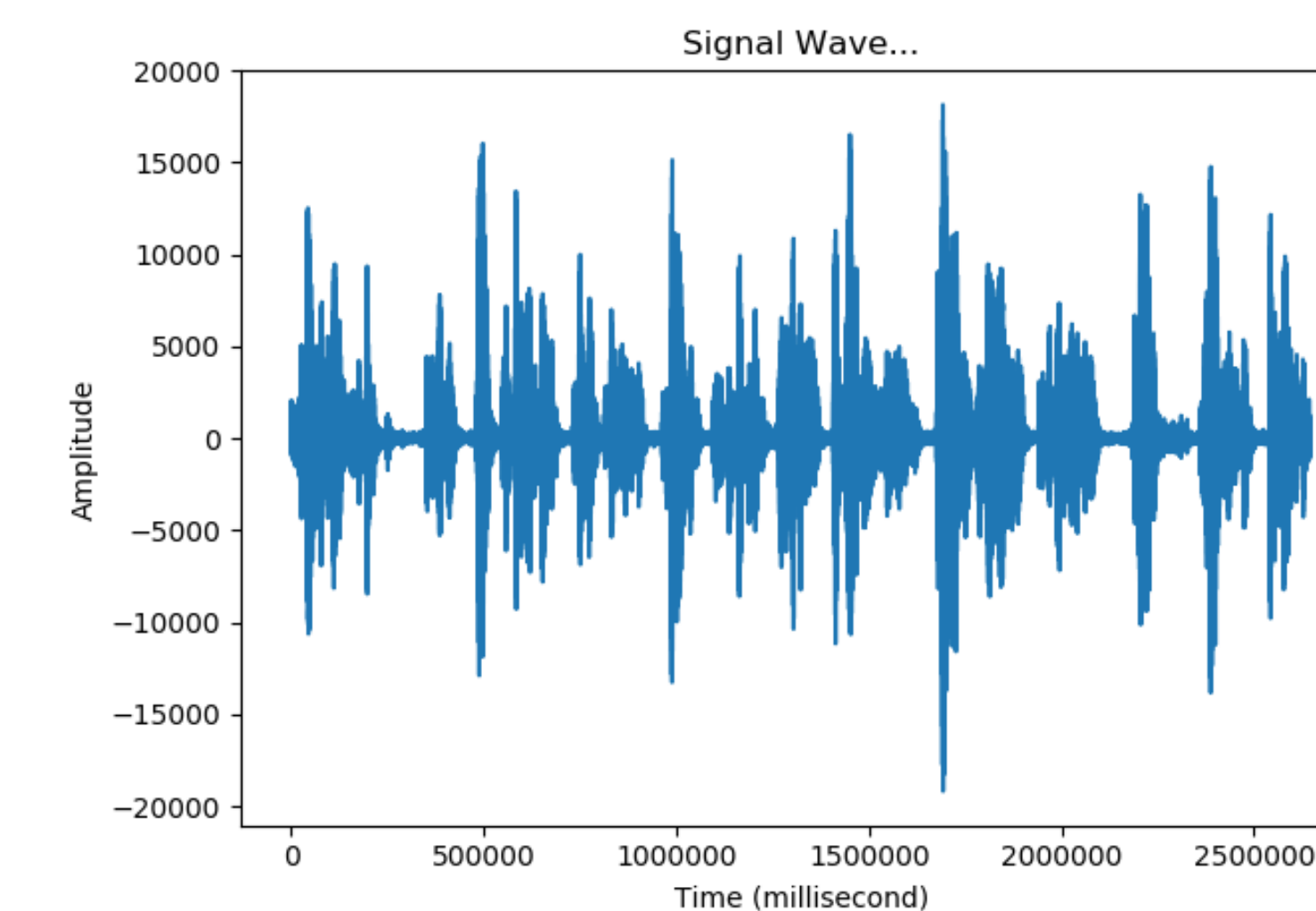


Figure 1: Amplitude from criticizing message

Text of figure 2: Even some very significant challenges we should have dealt with but we didn't do it why because hyper partisan politics prevents us from getting anything done.

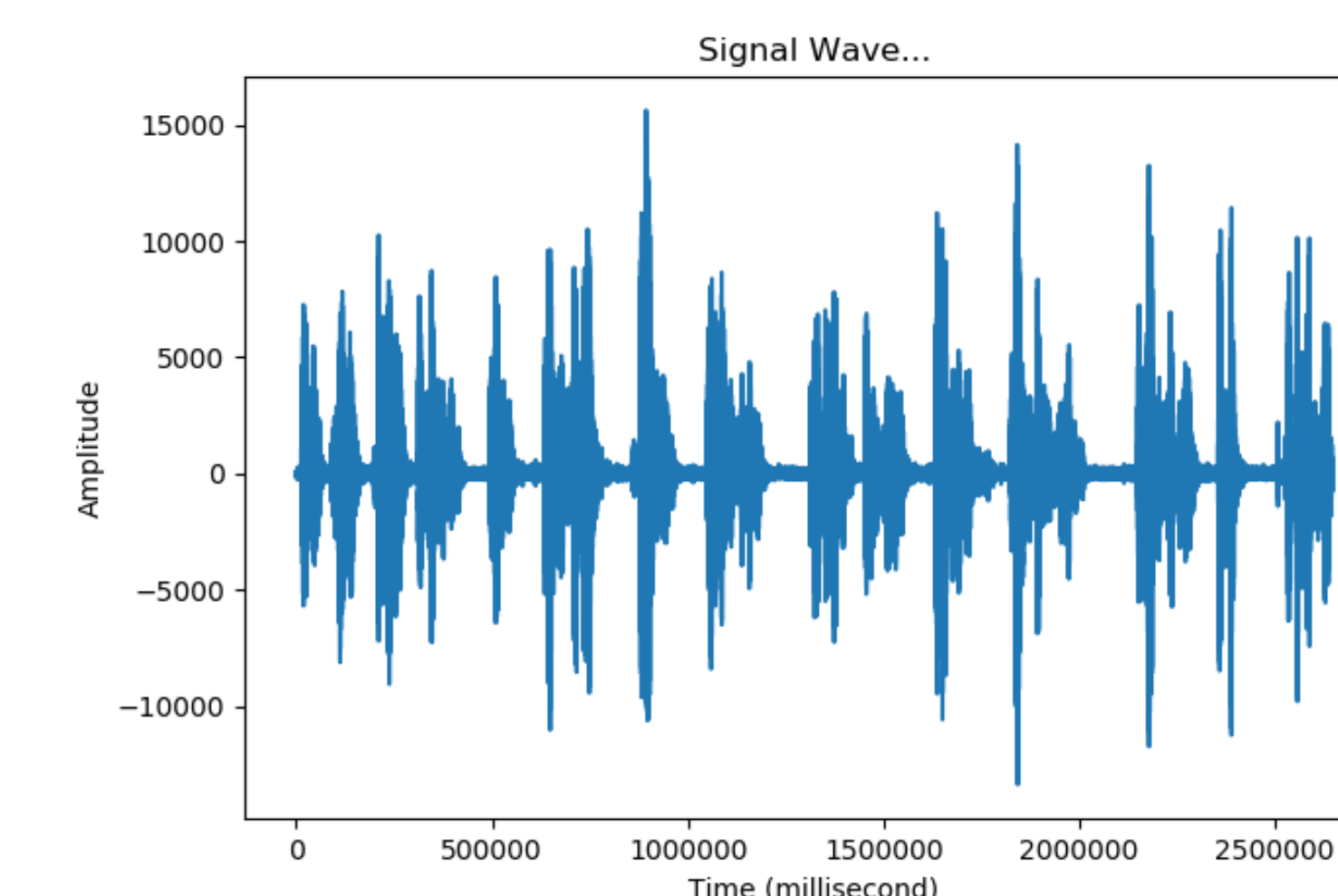


Figure 2: Amplitude from positive message

Text of figure 3: One of the most sacred fundamental ideals of this country the American dream the notion that if you work hard and you play by the rules you and your family can live a better life.

## Video image: detecting emotions of a speaker

I used **Emotion** Python package written by Octavio, Valdenegro-Toro, and Plöger (2017). The strength of their package is that it can analyze speaker’s facial emotions in real time for every frames based on general convolutional neural network (CNN). The trained classifier for emotion used data set by Goodfellow (2013), and the authors reported that their classifier has an accuracy of 66% in the FER-2013 emotion dataset comparing to human classifier. Among many videos, I chose one video to analyze its features in detail.

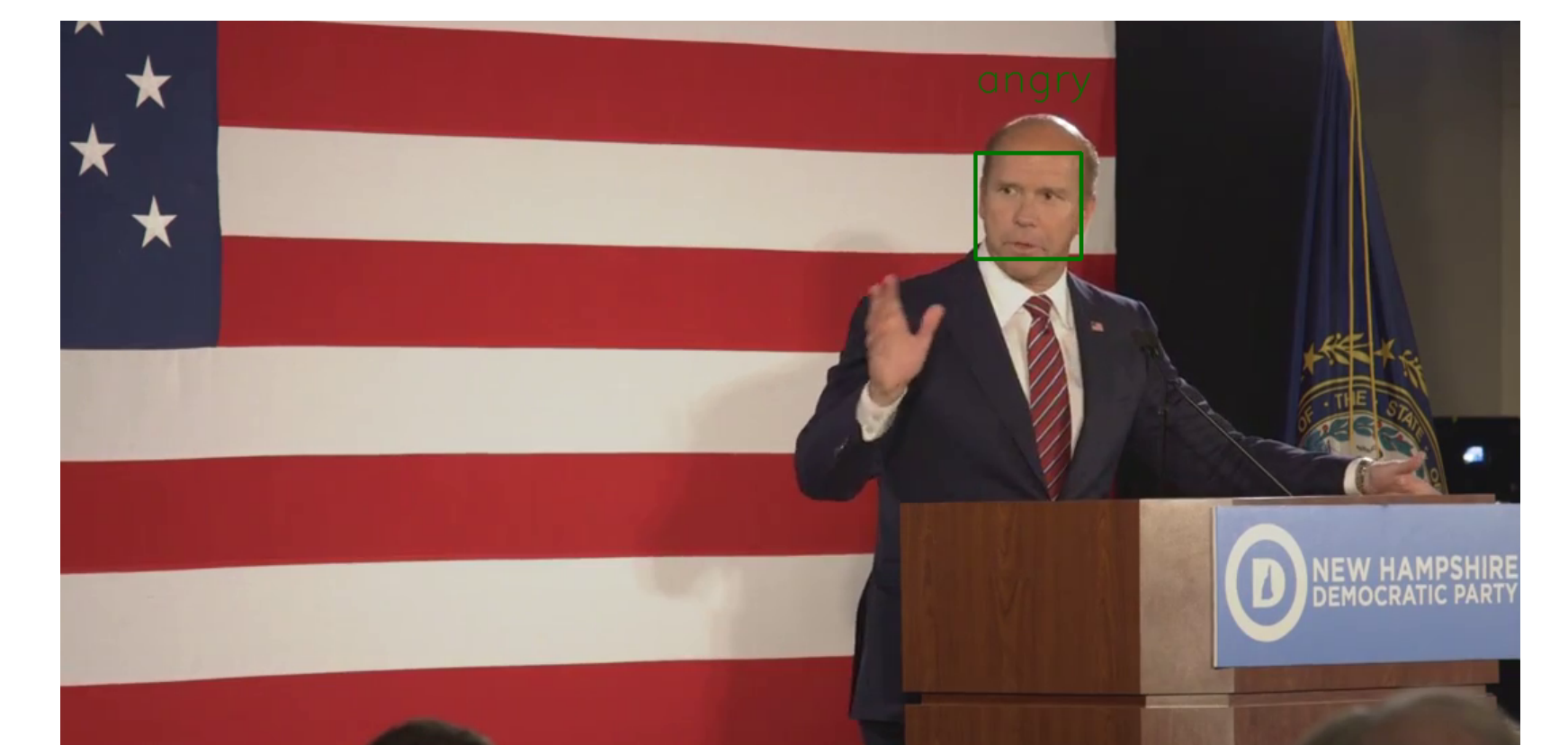


Figure 4: Real time video analysis on facial expression

Emotion	Frequency
Angry	162
Disgust	21
Fear	20
Happy	779
Neutral	165
sad	363
Surprise	14
Total length of video	19:17

Table 1: Captured emotions of a speaker from video

## Conclusion

We will conduct the experiments to identify the effect of emotion in speech via textual and non-textual ways. For instance, 1) we can use similar text with different audio and visual setting to identify high and low emotional expression. 2) We can test how emotional and contextual differences affect receivers' perception and knowledge. 3) With better and more Congressional campaign speech videos, we could build better speaker's emotion classifier and identify different candidates' campaign styles and their effects.