Natural And Emotional Linguistic Text To Speech Synthesis

Uddhav Raj B150878CS
TGDK Sumanathilaka B150413CS
Venkatesh Raju B151040CS
Viggnah Selvaraj B150080CS

Abstract

A real time system to synthesise emotional and natural human-like sound from unstructured text was built. The whole system has three parts. The first part deals with extracting emotion out of text using text mining and LSTMs. The second part identifies the key features of speech for Speech Emotion Recognition. In the third part these attributes are modified in a neutral speech to give it an emotional and human-like base.

Part 1 : Emotion Detection From Text

What is it?

•••

To extract emotion out of text based on linguistic features.

Datasets

...

Dataset 1 : ISEAR

```
label,text,
joy,"On days when I feel close to my partner and other friends.
When I feel at peace with myself and also experience a close
contact with people whom I regard greatly.",
fear,"Every time I imagine that someone I love or I could contact a
serious illness, even death.",
anger,"When I had been obviously unjustly treated and had no possibility
of elucidating this.",
sadness,"When I think about the short time that we live and relate it to
the periods of my life when I think that I did not use this
short time.",
disgust,"At a gathering I found myself involuntarily sitting next to two
people who expressed opinions that I considered very low and
discriminating.",
shame,"When I realized that I was directing the feelings of discontent
with myself at my partner and this way was trying to put the blame
on him instead of sorting out my own feeliings.",
guilt,"I feel guilty when when I realize that I consider material things
more important than caring for my relatives.  I feel very
self-centered.",
joy,"After my girlfriend had taken her exam we went to her parent's
place.",
fear,"When, for the first time I realized the meaning of death.",
anger,"When a car is overtaking another and I am forced to drive off the
road.",
sadness,"When I recently thought about the hard work it takes to study, and
how one wants to try something else.  When I read a theoretical
book in English that I did not understand.",
```

Dataset 2 : Twitter tweets

```
3  140499585285111809: the moment when you get another follower and you cheer.     :: joy
4  145207578270507009: Be the greatest dancer of your life! practice daily positive habits.  #fun #freedom #habits
   :: joy
5  139502146390470656: eww.. my moms starting to make her annual rum cake for the whole ramdyal/ally family. fml fml
   fml the smelll.......     :: disgust
6  146042696899887106: If ur heart hurts all the time for tht person something isn't right where's the     :: joy
7  145492569609084928: I feel awful, and it's way too freaking early. Now off to leadership highschool...  :: joy
8  145903955229151232: So chuffed for safc fans! Bet me dar comes in mortz from the match  :: joy
9  142717613234069504: Making art and viewing art are different at their core!     :: fear
10 144183822873927680: Soooo dooowwwn!! Move on, get some sleep... Me deserve better. #forgetit #yawning   :: anger
11 145959995803058177: &quot;We're sorry, but the clip you selected isn't available from your location. Please
   select another clip.&quot; NO I REFUSE.     :: sadness
12 142918097010036736: Een ware Sint &quot;in-en uittocht&quot; achter de rug; Winkel IN, winkel UIT, winkel IN,
   winkel UIT!   :: surprise
13 142186403223179264: People know they can pull you  down and they dont give a care &     :: surprise
14 145703538276843520: My heart and soul @Jay_BabeBee is leaving me and I can't even see here  :: sadness
15 145947752709369857: chips and curry sauce   :: joy
16 138958921875464194: Soo if i hit youu , i garrentee i won't stopp . type to keep going till i make a bitch bleed
   foreal !   :: anger
```

Twitter tweets with intensity

A	B	C	D
D	TWEET	Emotion	Intensity
10942	@CorningFootball IT'S GAME DAY!!!! T MINUS 14:30 #relentless	anger	0.144
10943	This game has pissed me off more than any other game this year. My blood is boiling! Time to turn it off! #STLCards	anger	0.898
10944	@spamvicious I've just found out it's Candice and not Candace. She can pout all she likes for me (t/üòç	anger	0.271
10945	@moocowward @mrsajhargreaves @Melly77 @GaryBarlow if he can't come to my Mum'a 60th after 25k tweets then w	anger	0.646
10946	@moocowward @mrsajhargreaves @Melly77 @GaryBarlow if he can't come to my Mum'a 60th after 25k tweets then w	anger	0.583
10947	wanna go home and focus up on this game . Don't wanna rage at all	anger	0.375
10948	@virginmedia I've been disconnected whilst on holiday (t/üò§ but I don't move house until the 1st October (t/ü§î #furious	anger	0.625
10949	@virginmedia I've been disconnected whilst on holiday (t/üò§ but I don't move house until the 1st October (t/ü§î	anger	0.396
10950	I wanna see you smile I don't wanna see you make a frown	anger	0.25
10951	@shae_caitlin ur road rage gives me anxiety.	anger	0.438
10952	@eMilsOnWheels I'm furious (t/üò©(t/üò©(t/üò©	anger	0.708
10953	@EtherealMystic_ She was winning this war that had been raging on inside of his mind. The desire and love all rolling ir	anger	0.333
10954	They gonna give this KKk police bitch the minimum sentence..just wattch #angry	anger	0.877
10955	They gonna give this KKk police bitch the minimum sentence..just wattch	anger	0.708
10956	I just got murdered in madden. (t/ü§î	anger	0.417
10957	My nephew sees that i have a frown on my face and he tells me 'you're beautiful '!(t/üò¢(t/üíû(t/üò©	anger	0.229
10958	@CUTEFUNNYANIMAL @luvcaps19 My sister's dog does this. I think it's because she knows it'll provoke a reaction	anger	0.375
10959	new madden 16 video was gonna be up but xbox is being an ahole and not going through (t/üòÑ(t/üòÑ #struggles	anger	0.667

Methodology

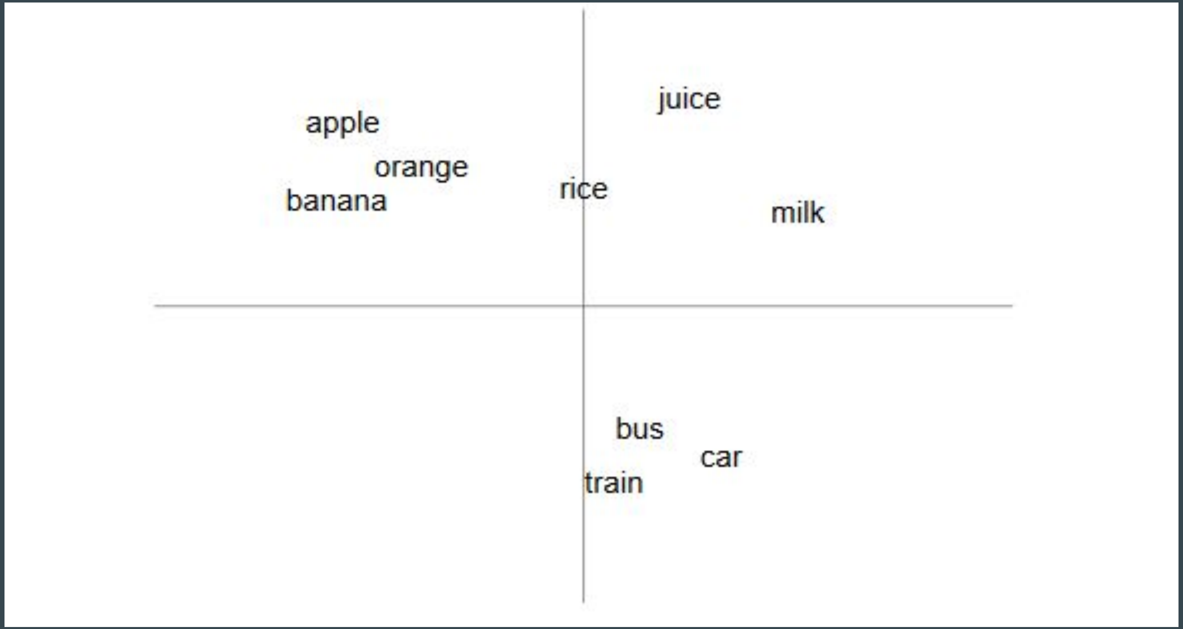
Our design was general enough to take into account two of the predominant emotional models: categorical and dimensional. We use a deep learning architecture to accurately identify emotions in datasets.

Preprocessing - Intermediate Representation

In order to convert the words into a suitable intermediate representation that can be fed into the neural network, we used word embeddings. We also tried incorporating latent sentiment features into these embeddings as in Bespalov et al.[9] in order to see performance improvements.

[9] D. Bespalov, B. Bai, Y. Qi, and A. Shokoufan- deh, "Sentiment classification based on super- vised latent n-gram analysis," in Proceedings of the 20th ACM International Conference on In- formation and Knowledge Management, CIKM '11, (New York, NY, USA), pp. 375–382, ACM, 2011.

“A word is known by the company it keeps”



Comparison

Traditional Method - Bag of Words Model	Word Embeddings
• Uses one hot encoding • Each word in the vocabulary is represented by one bit position in a HUGE vector. • For example, if we have a vocabulary of 10000 words, and “Hello” is the 4th word in the dictionary, it would be represented by: 0 0 0 1 0 0 . . . . 0 0 0 0 • Context information is not utilized	• Stores each word in as a point in space, where it is represented by a vector of fixed number of dimensions (generally 300) • Unsupervised, built just by reading huge corpus • For example, “Hello” might be represented as : [0.4, -0.11, 0.55, 0.3 . . . 0.1, 0.02] • Dimensions are basically projections along different axes, more of a mathematical concept.

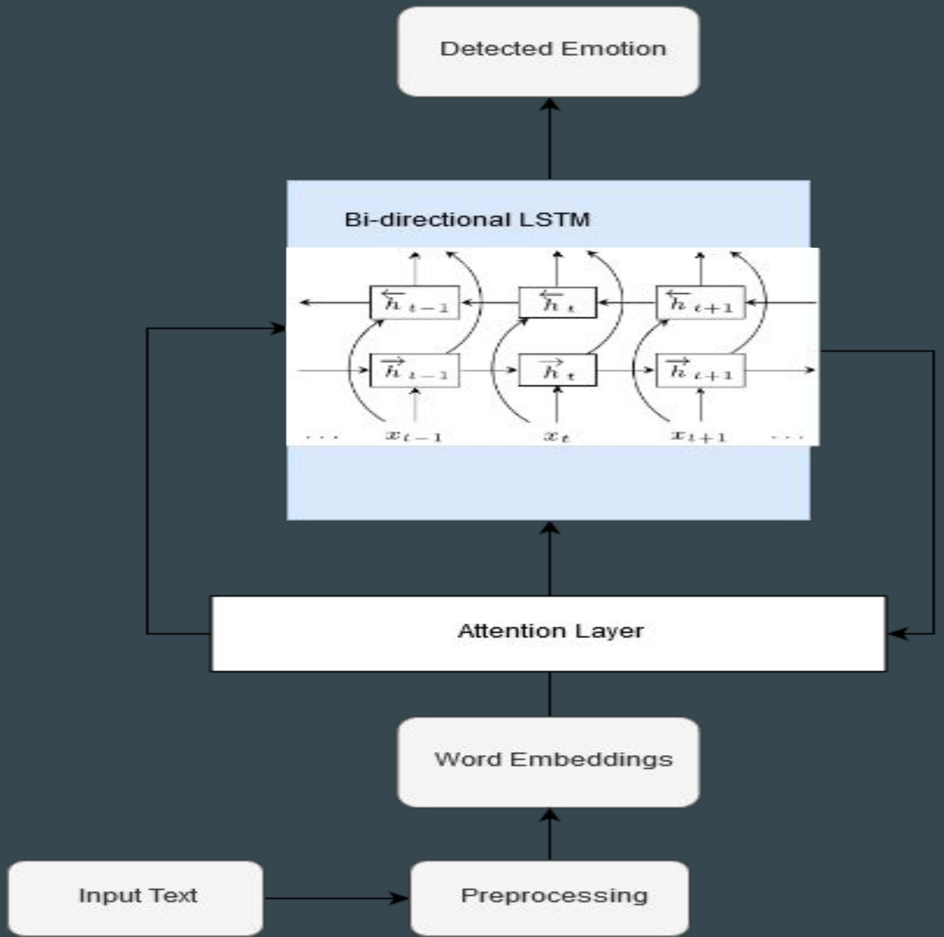
Model

We have implemented a recurrent neural network that takes into account temporal dependencies, but due to potential instabilities during optimization, we followed common choices that advocate the use of long short-term memory networks (LSTM). We extended the recurrent neural network architecture presented in Chen et al.[10] as in [1], to include dropout layers for regularization and used a weighted loss function to deal with class imbalance. We will also incorporate the attention mechanism[11], to increase the accuracy.

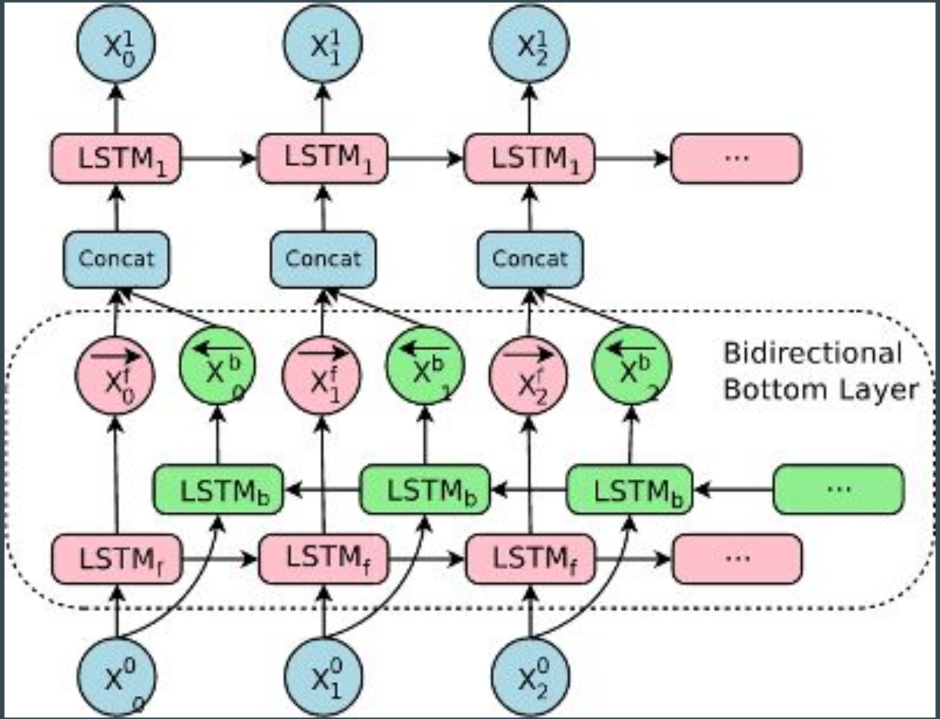
[1] B. Kratzwald, S. Ili, M. Kraus, S. Feuerriegel, and H. Prendinger, "Deep learning for affective computing: Text-based emotion recognition in decision support," Decision Support Systems, vol. 115, pp. 24 – 35, 2018.

[10] H. Chen, M. Sun, C. Tu, Y. Lin, and Z. Liu, "Neural sentiment classification with user and product attention," in Proceedings of the 2016.

Diagrammatic Representation



Inside the LSTM



Results

LSTM model on ISEAR Dataset

```
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
embedding_1 (Embedding)      (None, 100, 300)          2700000
_________________________________________________________________
lstm_1 (LSTM)                (None, 128)               219648
_________________________________________________________________
dense_1 (Dense)              (None, 7)                 903
_________________________________________________________________
activation_1 (Activation)    (None, 7)                 0
=================================================================
Total params: 2,920,551
Trainable params: 220,551
Non-trainable params: 2,700,000
_________________________________________________________________
None
Train on 6087 samples, validate on 677 samples
Epoch 1/40
6087/6087 [==============================] - 26s 4ms/step - loss: 1.8766 - acc: 0.2435 - val_loss: 1.7245 - val_acc: 0.3
575
Epoch 2/40
6087/6087 [==============================] - 21s 3ms/step - loss: 1.7210 - acc: 0.3498 - val_loss: 1.5600 - val_acc: 0.4
136
Epoch 3/40
6087/6087 [==============================] - 21s 3ms/step - loss: 1.6404 - acc: 0.3867 - val_loss: 1.4549 - val_acc: 0.4
727
Epoch 4/40
6087/6087 [==============================] - 21s 3ms/step - loss: 1.5768 - acc: 0.4124 - val_loss: 1.3977 - val_acc: 0.4
919
Epoch 5/40
6087/6087 [==============================] - 21s 3ms/step - loss: 1.5168 - acc: 0.4312 - val_loss: 1.3461 - val_acc: 0.5
```

```
6087/6087 [==============================] - 22s 4ms/step - loss: 0.9183 - acc: 0.6726 - val_loss: 0.9790 - val
_acc: 0.6484
Epoch 28/40
6087/6087 [==============================] - 29s 5ms/step - loss: 0.9068 - acc: 0.6783 - val_loss: 0.9944 - val
_acc: 0.6411
Epoch 29/40
6087/6087 [==============================] - 28s 5ms/step - loss: 0.8997 - acc: 0.6833 - val_loss: 0.9777 - val
_acc: 0.6514
Epoch 30/40
6087/6087 [==============================] - 29s 5ms/step - loss: 0.8847 - acc: 0.6887 - val_loss: 0.9855 - val
_acc: 0.6499
Epoch 31/40
6087/6087 [==============================] - 28s 5ms/step - loss: 0.8605 - acc: 0.6946 - val_loss: 0.9944 - val
_acc: 0.6440
Epoch 32/40
6087/6087 [==============================] - 28s 5ms/step - loss: 0.8636 - acc: 0.6949 - val_loss: 0.9799 - val
_acc: 0.6470
Epoch 33/40
6087/6087 [==============================] - 28s 5ms/step - loss: 0.8492 - acc: 0.6916 - val_loss: 0.9885 - val
_acc: 0.6588
Epoch 34/40
6087/6087 [==============================] - 26s 4ms/step - loss: 0.8255 - acc: 0.7120 - val_loss: 0.9956 - val
_acc: 0.6573
Epoch 35/40
6087/6087 [==============================] - 24s 4ms/step - loss: 0.8169 - acc: 0.7128 - val_loss: 0.9941 - val
_acc: 0.6484
Epoch 36/40
6087/6087 [==============================] - 34s 6ms/step - loss: 0.8227 - acc: 0.7092 - val_loss: 0.9861 - val
_acc: 0.6588
Epoch 37/40
6087/6087 [==============================] - 28s 5ms/step - loss: 0.8185 - acc: 0.7058 - val_loss: 0.9946 - val
_acc: 0.6558
Epoch 38/40
6087/6087 [==============================] - 28s 5ms/step - loss: 0.7901 - acc: 0.7201 - val_loss: 0.9939 - val
_acc: 0.6588
Epoch 39/40
6087/6087 [==============================] - 31s 5ms/step - loss: 0.7775 - acc: 0.7202 - val_loss: 0.9985 - val
_acc: 0.6588
Epoch 40/40
6087/6087 [==============================] - 29s 5ms/step - loss: 0.7644 - acc: 0.7296 - val_loss: 1.0098 - val
accuracy : 75.83
```

LSTM model on Twitter Tweets dataset

```
________________________________________________________________
Layer (type)                 Output Shape              Param #
================================================================
embedding_1 (Embedding)      (None, 100, 300)          6000300
________________________________________________________________
lstm_1 (LSTM)                (None, 128)               219648
________________________________________________________________
dense_1 (Dense)              (None, 6)                 774
________________________________________________________________
activation_1 (Activation)    (None, 6)                 0
================================================================
Total params: 6,220,722
Trainable params: 220,422
Non-trainable params: 6,000,300
________________________________________________________________
None
Train on 17048 samples, validate on 1895 samples
Epoch 1/40
17048/17048 [==============================] - 91s 5ms/step - loss: 1.4496 - acc: 0.4330 - val_loss: 1.2957 - val_acc: 0
.4908
Epoch 2/40
17048/17048 [==============================] - 92s 5ms/step - loss: 1.3241 - acc: 0.4805 - val_loss: 1.2374 - val_acc: 0
.5156
Epoch 3/40
17048/17048 [==============================] - 99s 6ms/step - loss: 1.2686 - acc: 0.5079 - val_loss: 1.2041 - val_acc: 0
.5240
Epoch 4/40
17048/17048 [==============================] - 84s 5ms/step - loss: 1.2258 - acc: 0.5245 - val_loss: 1.1603 - val_acc: 0
.5509
```

```
Epoch 34/40
17048/17048 [==============================] - 64s 4ms/step - loss: 0.8701 - acc: 0.6791 - val_loss: 1.0639 - val_acc: 0
.6290
Epoch 35/40
17048/17048 [==============================] - 64s 4ms/step - loss: 0.8628 - acc: 0.6827 - val_loss: 1.0513 - val_acc: 0
.6274
Epoch 36/40
17048/17048 [==============================] - 68s 4ms/step - loss: 0.8598 - acc: 0.6821 - val_loss: 1.0618 - val_acc: 0
.6359
Epoch 37/40
17048/17048 [==============================] - 68s 4ms/step - loss: 0.8575 - acc: 0.6848 - val_loss: 1.0614 - val_acc: 0
.6248
Epoch 38/40
17048/17048 [==============================] - 66s 4ms/step - loss: 0.8477 - acc: 0.6854 - val_loss: 1.0624 - val_acc: 0
.6290
Epoch 39/40
17048/17048 [==============================] - 68s 4ms/step - loss: 0.8437 - acc: 0.6892 - val_loss: 1.0632 - val_acc: 0
.6280
Epoch 40/40
17048/17048 [==============================] - 93s 5ms/step - loss: 0.8366 - acc: 0.6902 - val_loss: 1.0624 - val_acc: 0
.6285
 accuracy : 71.06
```

LSTM model on Twitter Tweets (old) dataset

```
Layer (type)                 Output Shape              Param #
=================================================================
embedding_1 (Embedding)      (None, 100, 300)          6000300
_________________________________________________________________
lstm_1 (LSTM)                (None, 128)               219648
_________________________________________________________________
dense_1 (Dense)              (None, 13)                1677
_________________________________________________________________
activation_1 (Activation)    (None, 13)                0
=================================================================
Total params: 6,221,625
Trainable params: 221,325
Non-trainable params: 6,000,300
_________________________________________________________________
None
Train on 32400 samples, validate on 3600 samples
Epoch 1/10
32400/32400 [==============================] - 146s 5ms/step - loss: 2.0807 - acc: 0.2764 - val_loss: 1.9840 - val_acc:
0.3011
Epoch 2/10
32400/32400 [==============================] - 128s 4ms/step - loss: 1.9564 - acc: 0.3281 - val_loss: 1.9171 - val_acc:
0.3328
Epoch 3/10
32400/32400 [==============================] - 120s 4ms/step - loss: 1.9042 - acc: 0.3483 - val_loss: 1.9007 - val_acc:
0.3411
Epoch 4/10
32400/32400 [==============================] - 120s 4ms/step - loss: 1.8732 - acc: 0.3575 - val_loss: 1.8805 - val_acc:
0.3411
Epoch 5/10
32400/32400 [==============================] - 121s 4ms/step - loss: 1.8558 - acc: 0.3636 - val_loss: 1.8713 - val_acc:
0.3533
```

```
Epoch 34/40
17048/17048 [==============================] - 64s 4ms/step - loss: 0.8701 - acc: 0.6791 - val_loss: 1.0639 - val_acc: 0
.6290
Epoch 35/40
17048/17048 [==============================] - 64s 4ms/step - loss: 0.8628 - acc: 0.6827 - val_loss: 1.0513 - val_acc: 0
.6274
Epoch 36/40
17048/17048 [==============================] - 68s 4ms/step - loss: 0.8598 - acc: 0.6821 - val_loss: 1.0618 - val_acc: 0
.6359
Epoch 37/40
17048/17048 [==============================] - 68s 4ms/step - loss: 0.8575 - acc: 0.6848 - val_loss: 1.0614 - val_acc: 0
.6248
Epoch 38/40
17048/17048 [==============================] - 66s 4ms/step - loss: 0.8477 - acc: 0.6854 - val_loss: 1.0624 - val_acc: 0
.6290
Epoch 39/40
17048/17048 [==============================] - 68s 4ms/step - loss: 0.8437 - acc: 0.6892 - val_loss: 1.0632 - val_acc: 0
.6280
Epoch 40/40
17048/17048 [==============================] - 93s 5ms/step - loss: 0.8366 - acc: 0.6902 - val_loss: 1.0624 - val_acc: 0
.6285

accuracy: 59.22
```

Pre-Processing

...

Preprocessing steps

1. Tokenization
2. Normalization
   i. Stemming eg: Running -> Run
   ii. Lemmatization eg: Better -> Good
3. Noise Removal
   i. Removing Twitter Username
   ii. Removing Web Links
4. Converting all letters to lower or upper case
5. Converting numbers into words

Isear dataset - Accuracy : **82.53**

```
6087/6087 [==============================] - 22s 4ms/step - loss: 0.9183 - acc: 0.6726 - val_loss: 0.9790 - val
_acc: 0.6484
Epoch 28/40
6087/6087 [==============================] - 29s 5ms/step - loss: 0.9068 - acc: 0.6783 - val_loss: 0.9944 - val
_acc: 0.6411
Epoch 29/40
6087/6087 [==============================] - 28s 5ms/step - loss: 0.8997 - acc: 0.6833 - val_loss: 0.9777 - val
_acc: 0.6514
Epoch 30/40
6087/6087 [==============================] - 29s 5ms/step - loss: 0.8847 - acc: 0.6887 - val_loss: 0.9855 - val
_acc: 0.6499
Epoch 31/40
6087/6087 [==============================] - 28s 5ms/step - loss: 0.8605 - acc: 0.6946 - val_loss: 0.9944 - val
_acc: 0.6440
Epoch 32/40
6087/6087 [==============================] - 28s 5ms/step - loss: 0.8636 - acc: 0.6949 - val_loss: 0.9799 - val
_acc: 0.6470
Epoch 33/40
6087/6087 [==============================] - 28s 5ms/step - loss: 0.8492 - acc: 0.6916 - val_loss: 0.9885 - val
_acc: 0.6588
Epoch 34/40
6087/6087 [==============================] - 26s 4ms/step - loss: 0.8255 - acc: 0.7120 - val_loss: 0.9956 - val
_acc: 0.6573
Epoch 35/40
6087/6087 [==============================] - 24s 4ms/step - loss: 0.8169 - acc: 0.7128 - val_loss: 0.9941 - val
_acc: 0.6484
Epoch 36/40
6087/6087 [==============================] - 34s 6ms/step - loss: 0.8227 - acc: 0.7092 - val_loss: 0.9861 - val
_acc: 0.6588
Epoch 37/40
6087/6087 [==============================] - 28s 5ms/step - loss: 0.8185 - acc: 0.7058 - val_loss: 0.9946 - val
_acc: 0.6558
Epoch 38/40
6087/6087 [==============================] - 28s 5ms/step - loss: 0.7901 - acc: 0.7201 - val_loss: 0.9939 - val
_acc: 0.6588
Epoch 39/40
6087/6087 [==============================] - 31s 5ms/step - loss: 0.7775 - acc: 0.7202 - val_loss: 0.9985 - val
_acc: 0.6588
Epoch 40/40
6087/6087 [==============================] - 29s 5ms/step - loss: 0.7644 - acc: 0.7296 - val_loss: 1.0098 - val
accuracy: 82.53
```

Twitter Tweets - Accuracy **79.81**

```
Epoch 34/40
17048/17048 [==============================] - 64s 4ms/step - loss: 0.8701 - acc: 0.6791 - val_loss: 1.0639 - val_acc: 0
.6290
Epoch 35/40
17048/17048 [==============================] - 64s 4ms/step - loss: 0.8628 - acc: 0.6827 - val_loss: 1.0513 - val_acc: 0
.6274
Epoch 36/40
17048/17048 [==============================] - 68s 4ms/step - loss: 0.8598 - acc: 0.6821 - val_loss: 1.0618 - val_acc: 0
.6359
Epoch 37/40
17048/17048 [==============================] - 68s 4ms/step - loss: 0.8575 - acc: 0.6848 - val_loss: 1.0614 - val_acc: 0
.6248
Epoch 38/40
17048/17048 [==============================] - 66s 4ms/step - loss: 0.8477 - acc: 0.6854 - val_loss: 1.0624 - val_acc: 0
.6290
Epoch 39/40
17048/17048 [==============================] - 68s 4ms/step - loss: 0.8437 - acc: 0.6892 - val_loss: 1.0632 - val_acc: 0
.6280
Epoch 40/40
17048/17048 [==============================] - 93s 5ms/step - loss: 0.8366 - acc: 0.6902 - val_loss: 1.0624 - val_acc: 0
.6285
accuracy: 79.81
```

Twitter Tweets - Accuracy **62.47**

```
Epoch 34/40
17048/17048 [==============================] - 64s 4ms/step - loss: 0.8701 - acc: 0.6791 - val_loss: 1.0639 - val_acc: 0
.6290
Epoch 35/40
17048/17048 [==============================] - 64s 4ms/step - loss: 0.8628 - acc: 0.6827 - val_loss: 1.0513 - val_acc: 0
.6274
Epoch 36/40
17048/17048 [==============================] - 68s 4ms/step - loss: 0.8598 - acc: 0.6821 - val_loss: 1.0618 - val_acc: 0
.6359
Epoch 37/40
17048/17048 [==============================] - 68s 4ms/step - loss: 0.8575 - acc: 0.6848 - val_loss: 1.0614 - val_acc: 0
.6248
Epoch 38/40
17048/17048 [==============================] - 66s 4ms/step - loss: 0.8477 - acc: 0.6854 - val_loss: 1.0624 - val_acc: 0
.6290
Epoch 39/40
17048/17048 [==============================] - 68s 4ms/step - loss: 0.8437 - acc: 0.6892 - val_loss: 1.0632 - val_acc: 0
.6280
Epoch 40/40
17048/17048 [==============================] - 93s 5ms/step - loss: 0.8366 - acc: 0.6902 - val_loss: 1.0624 - val_acc: 0
.6285
accuracy: 62.47
```

Dataset	Accuracy without preprocessing (%)	Accuracy with preprocessing(%)
Isear Dataset	75.83	82.53
Twitter Tweet	71.06	79.81
Twitter Tweet (OLD)	59.22	62.47

What do we have uptil now?

Text with the corresponding emotion

Part 2 : Speech Emotion Recognition

What is it?

To classify audio into their emotions on the basis of only speech features without taking care about the linguistic features

Dataset

...

Ryerson Audio-Visual Database for Emotional Speech and Song (RAVDESS)

The Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS) contains 7356 files (total size: 24.8 GB). The database contains 24 professional actors (12 female, 12 male), vocalizing two lexically-matched statements. Speech includes calm, happy, sad, angry, fearful, surprise, disgust and neutral expressions.

The Emotions Used

1. Anger
2. Calm
3. Disgust
4. Fearful
5. Happy
6. Neutral
7. Sad
8. Surprised

Tools Used For Feature Extraction

Feature Extraction from Vocal

- Praat
  - Pitch (Hz)
    - Max
    - Min
    - Median
  - Intensity : Audio Level Differences (W/m^2)
    - Max
    - Min
    - Median
  - Formant : Acoustic Energy around a particular frequency in speech (pa)
    - First : Lowest Frequency energy level
    - Second
    - Third : Highest Frequency energy level

Jaudio

Features:

1. Area method of moments
2. Spectral Centroid
3. Spectral Rolloff Point
4. RMSE
5. Spectral Density
6. Spectral Variability
7. Fraction of Low Energy Windows
8. Zero Crossings

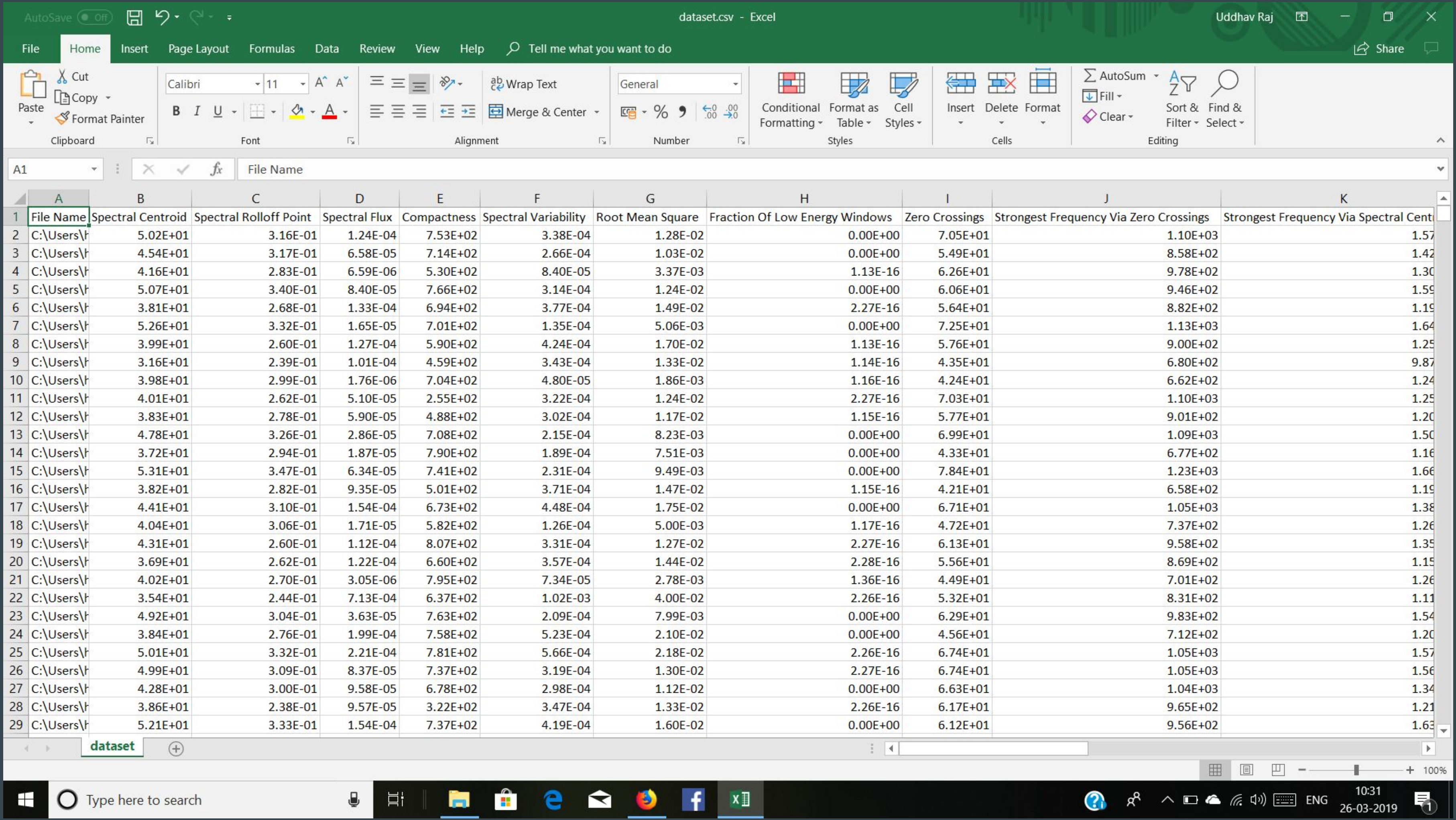
Librosa and PyAudio

Python Libraries. Features extracted:

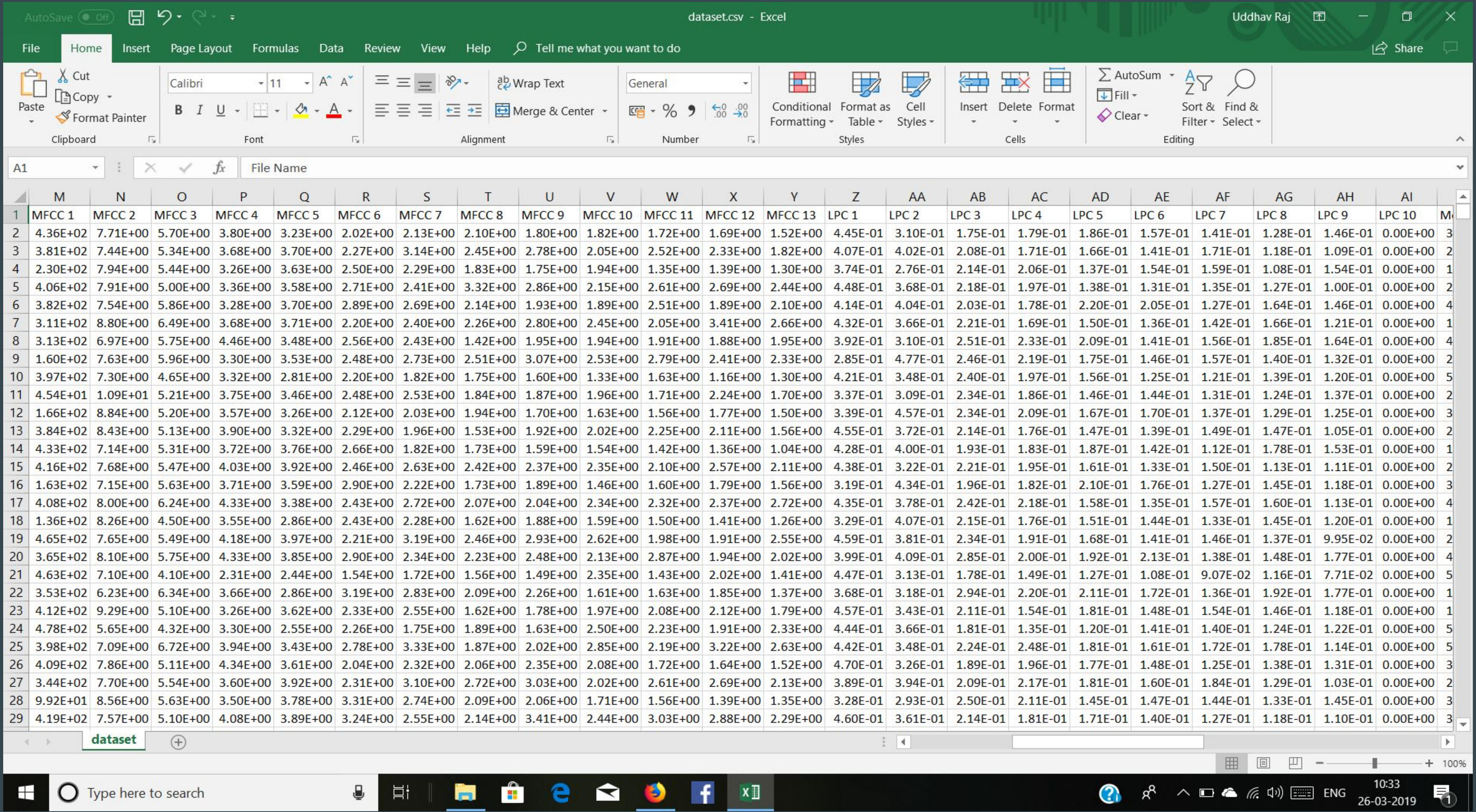
1. MFCC
2. LPCC
3. Strongest Frequency via Zero Crossings

Dataset after Feature Extraction

…



File Name	Spectral Centroid	Spectral Rolloff Point	Spectral Flux	Compactness	Spectral Variability	Root Mean Square	Fraction Of Low Energy Windows	Zero Crossings	Strongest Frequency Via Zero Crossings	Strongest Frequency Via Spectral Cent
C:\Users\h	5.02E+01	3.16E-01	1.24E-04	7.53E+02	3.38E-04	1.28E-02	0.00E+00	7.05E+01	1.10E+03	1.57
C:\Users\h	4.54E+01	3.17E-01	6.58E-05	7.14E+02	2.66E-04	1.03E-02	0.00E+00	5.49E+01	8.58E+02	1.42
C:\Users\h	4.16E+01	2.83E-01	6.59E-06	5.30E+02	8.40E-05	3.37E-03	1.13E-16	6.26E+01	9.78E+02	1.30
C:\Users\h	5.07E+01	3.40E-01	8.40E-05	7.66E+02	3.14E-04	1.24E-02	0.00E+00	6.06E+01	9.46E+02	1.59
C:\Users\h	3.81E+01	2.68E-01	1.33E-04	6.94E+02	3.77E-04	1.49E-02	2.27E-16	5.64E+01	8.82E+02	1.19
C:\Users\h	5.26E+01	3.32E-01	1.65E-05	7.01E+02	1.35E-04	5.06E-03	0.00E+00	7.25E+01	1.13E+03	1.64
C:\Users\h	3.99E+01	2.60E-01	1.27E-04	5.90E+02	4.24E-04	1.70E-02	1.13E-16	5.76E+01	9.00E+02	1.25
C:\Users\h	3.16E+01	2.39E-01	1.01E-04	4.59E+02	3.43E-04	1.33E-02	1.14E-16	4.35E+01	6.80E+02	9.87
C:\Users\h	3.98E+01	2.99E-01	1.76E-06	7.04E+02	4.80E-05	1.86E-03	1.16E-16	4.24E+01	6.62E+02	1.24
C:\Users\h	4.01E+01	2.62E-01	5.10E-05	2.55E+02	3.22E-04	1.24E-02	2.27E-16	7.03E+01	1.10E+03	1.25
C:\Users\h	3.83E+01	2.78E-01	5.90E-05	4.88E+02	3.02E-04	1.17E-02	1.15E-16	5.77E+01	9.01E+02	1.20
C:\Users\h	4.78E+01	3.26E-01	2.86E-05	7.08E+02	2.15E-04	8.23E-03	0.00E+00	6.99E+01	1.09E+03	1.50
C:\Users\h	3.72E+01	2.94E-01	1.87E-05	7.90E+02	1.89E-04	7.51E-03	0.00E+00	4.33E+01	6.77E+02	1.16
C:\Users\h	5.31E+01	3.47E-01	6.34E-05	7.41E+02	2.31E-04	9.49E-03	0.00E+00	7.84E+01	1.23E+03	1.66
C:\Users\h	3.82E+01	2.82E-01	9.35E-05	5.01E+02	3.71E-04	1.47E-02	1.15E-16	4.21E+01	6.58E+02	1.19
C:\Users\h	4.41E+01	3.10E-01	1.54E-04	6.73E+02	4.48E-04	1.75E-02	0.00E+00	6.71E+01	1.05E+03	1.38
C:\Users\h	4.04E+01	3.06E-01	1.71E-05	5.82E+02	1.26E-04	5.00E-03	1.17E-16	4.72E+01	7.37E+02	1.26
C:\Users\h	4.31E+01	2.60E-01	1.12E-04	8.07E+02	3.31E-04	1.27E-02	2.27E-16	6.13E+01	9.58E+02	1.35
C:\Users\h	3.69E+01	2.62E-01	1.22E-04	6.60E+02	3.57E-04	1.44E-02	2.28E-16	5.56E+01	8.69E+02	1.15
C:\Users\h	4.02E+01	2.70E-01	3.05E-06	7.95E+02	7.34E-05	2.78E-03	1.36E-16	4.49E+01	7.01E+02	1.26
C:\Users\h	3.54E+01	2.44E-01	7.13E-04	6.37E+02	1.02E-03	4.00E-02	2.26E-16	5.32E+01	8.31E+02	1.11
C:\Users\h	4.92E+01	3.04E-01	3.63E-05	7.63E+02	2.09E-04	7.99E-03	0.00E+00	6.29E+01	9.83E+02	1.54
C:\Users\h	3.84E+01	2.76E-01	1.99E-04	7.58E+02	5.23E-04	2.10E-02	0.00E+00	4.56E+01	7.12E+02	1.20
C:\Users\h	5.01E+01	3.32E-01	2.21E-04	7.81E+02	5.66E-04	2.18E-02	2.26E-16	6.74E+01	1.05E+03	1.57
C:\Users\h	4.99E+01	3.09E-01	8.37E-05	7.37E+02	3.19E-04	1.30E-02	2.27E-16	6.74E+01	1.05E+03	1.56
C:\Users\h	4.28E+01	3.00E-01	9.58E-05	6.78E+02	2.98E-04	1.12E-02	0.00E+00	6.63E+01	1.04E+03	1.34
C:\Users\h	3.86E+01	2.38E-01	9.57E-05	3.22E+02	3.47E-04	1.33E-02	2.26E-16	6.17E+01	9.65E+02	1.21
C:\Users\h	5.21E+01	3.33E-01	1.54E-04	7.37E+02	4.19E-04	1.60E-02	0.00E+00	6.12E+01	9.56E+02	1.63



	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD	AE	AF	AG	AH	AI	
1	MFCC 1	MFCC 2	MFCC 3	MFCC 4	MFCC 5	MFCC 6	MFCC 7	MFCC 8	MFCC 9	MFCC 10	MFCC 11	MFCC 12	MFCC 13	LPC 1	LPC 2	LPC 3	LPC 4	LPC 5	LPC 6	LPC 7	LPC 8	LPC 9	LPC 10	M
2	4.36E+02	7.71E+00	5.70E+00	3.80E+00	3.23E+00	2.02E+00	2.13E+00	2.10E+00	1.80E+00	1.82E+00	1.72E+00	1.69E+00	1.52E+00	4.45E-01	3.10E-01	1.75E-01	1.79E-01	1.86E-01	1.57E-01	1.41E-01	1.28E-01	1.46E-01	0.00E+00	3
3	3.81E+02	7.44E+00	5.34E+00	3.68E+00	3.70E+00	2.27E+00	3.14E+00	2.45E+00	2.78E+00	2.05E+00	2.52E+00	2.33E+00	1.82E+00	4.07E-01	4.02E-01	2.08E-01	1.71E-01	1.66E-01	1.41E-01	1.71E-01	1.18E-01	1.09E-01	0.00E+00	2
4	2.30E+02	7.94E+00	5.44E+00	3.26E+00	3.63E+00	2.50E+00	2.29E+00	1.83E+00	1.75E+00	1.94E+00	1.35E+00	1.39E+00	1.30E+00	3.74E-01	2.76E-01	2.14E-01	2.06E-01	1.37E-01	1.54E-01	1.59E-01	1.08E-01	1.54E-01	0.00E+00	1
5	4.06E+02	7.91E+00	5.00E+00	3.36E+00	3.58E+00	2.71E+00	2.41E+00	3.32E+00	2.86E+00	2.15E+00	2.61E+00	2.69E+00	2.44E+00	4.48E-01	3.68E-01	2.18E-01	1.97E-01	1.38E-01	1.31E-01	1.35E-01	1.27E-01	1.00E-01	0.00E+00	2
6	3.82E+02	7.54E+00	5.86E+00	3.28E+00	3.70E+00	2.89E+00	2.69E+00	2.14E+00	1.93E+00	1.89E+00	2.51E+00	1.89E+00	2.10E+00	4.14E-01	4.04E-01	2.03E-01	1.78E-01	2.20E-01	2.05E-01	1.27E-01	1.64E-01	1.46E-01	0.00E+00	4
7	3.11E+02	8.80E+00	6.49E+00	3.68E+00	3.71E+00	2.20E+00	2.40E+00	2.26E+00	2.80E+00	2.45E+00	2.05E+00	3.41E+00	2.66E+00	4.32E-01	3.66E-01	2.21E-01	1.69E-01	1.50E-01	1.36E-01	1.42E-01	1.66E-01	1.21E-01	0.00E+00	1
8	3.13E+02	6.97E+00	5.75E+00	4.46E+00	3.48E+00	2.56E+00	2.43E+00	1.42E+00	1.95E+00	1.94E+00	1.91E+00	1.88E+00	1.95E+00	3.92E-01	3.10E-01	2.51E-01	2.33E-01	2.09E-01	1.41E-01	1.56E-01	1.85E-01	1.64E-01	0.00E+00	4
9	1.60E+02	7.63E+00	5.96E+00	3.30E+00	3.53E+00	2.48E+00	2.73E+00	2.51E+00	3.07E+00	2.53E+00	2.79E+00	2.41E+00	2.33E+00	2.85E-01	4.77E-01	2.46E-01	2.19E-01	1.75E-01	1.46E-01	1.57E-01	1.40E-01	1.32E-01	0.00E+00	2
10	3.97E+02	7.30E+00	4.65E+00	3.32E+00	2.81E+00	2.20E+00	1.82E+00	1.75E+00	1.60E+00	1.33E+00	1.63E+00	1.16E+00	1.30E+00	4.21E-01	3.48E-01	2.40E-01	1.97E-01	1.56E-01	1.25E-01	1.21E-01	1.39E-01	1.20E-01	0.00E+00	5
11	4.54E+01	1.09E+01	5.21E+00	3.75E+00	3.46E+00	2.48E+00	2.53E+00	1.84E+00	1.87E+00	1.96E+00	1.71E+00	2.24E+00	1.70E+00	3.37E-01	3.09E-01	2.34E-01	1.86E-01	1.46E-01	1.44E-01	1.31E-01	1.24E-01	1.37E-01	0.00E+00	2
12	1.66E+02	8.84E+00	5.20E+00	3.57E+00	3.26E+00	2.12E+00	2.03E+00	1.94E+00	1.70E+00	1.63E+00	1.56E+00	1.77E+00	1.50E+00	3.39E-01	4.57E-01	2.34E-01	2.09E-01	1.67E-01	1.70E-01	1.37E-01	1.29E-01	1.25E-01	0.00E+00	3
13	3.84E+02	8.43E+00	5.13E+00	3.90E+00	3.32E+00	2.29E+00	1.96E+00	1.53E+00	1.92E+00	2.02E+00	2.25E+00	2.11E+00	1.56E+00	4.55E-01	3.72E-01	2.14E-01	1.76E-01	1.47E-01	1.39E-01	1.49E-01	1.47E-01	1.05E-01	0.00E+00	2
14	4.33E+02	7.14E+00	5.31E+00	3.72E+00	3.76E+00	2.66E+00	1.82E+00	1.73E+00	1.59E+00	1.54E+00	1.42E+00	1.36E+00	1.04E+00	4.28E-01	4.00E-01	1.93E-01	1.83E-01	1.87E-01	1.42E-01	1.12E-01	1.78E-01	1.53E-01	0.00E+00	1
15	4.16E+02	7.68E+00	5.47E+00	4.03E+00	3.92E+00	2.46E+00	2.63E+00	2.42E+00	2.37E+00	2.35E+00	2.10E+00	2.57E+00	2.11E+00	4.38E-01	3.22E-01	2.21E-01	1.95E-01	1.61E-01	1.33E-01	1.50E-01	1.13E-01	1.11E-01	0.00E+00	2
16	1.63E+02	7.15E+00	5.63E+00	3.71E+00	3.59E+00	2.90E+00	2.22E+00	1.73E+00	1.89E+00	1.46E+00	1.60E+00	1.79E+00	1.56E+00	3.19E-01	4.34E-01	1.96E-01	1.82E-01	2.10E-01	1.76E-01	1.27E-01	1.45E-01	1.18E-01	0.00E+00	3
17	4.08E+02	8.00E+00	6.24E+00	4.33E+00	3.38E+00	2.43E+00	2.72E+00	2.07E+00	2.04E+00	2.34E+00	2.32E+00	2.37E+00	2.72E+00	4.35E-01	3.78E-01	2.42E-01	2.18E-01	1.58E-01	1.35E-01	1.57E-01	1.60E-01	1.13E-01	0.00E+00	4
18	1.36E+02	8.26E+00	4.50E+00	3.55E+00	2.86E+00	2.43E+00	2.28E+00	1.62E+00	1.88E+00	1.59E+00	1.50E+00	1.41E+00	1.26E+00	3.29E-01	4.07E-01	2.15E-01	1.76E-01	1.51E-01	1.44E-01	1.33E-01	1.45E-01	1.20E-01	0.00E+00	1
19	4.65E+02	7.65E+00	5.49E+00	4.18E+00	3.97E+00	2.21E+00	3.19E+00	2.46E+00	2.93E+00	2.62E+00	1.98E+00	1.91E+00	2.55E+00	4.59E-01	3.81E-01	2.34E-01	1.91E-01	1.68E-01	1.41E-01	1.46E-01	1.37E-01	9.95E-02	0.00E+00	2
20	3.65E+02	8.10E+00	5.75E+00	4.33E+00	3.85E+00	2.90E+00	2.34E+00	2.23E+00	2.48E+00	2.13E+00	2.87E+00	1.94E+00	2.02E+00	3.99E-01	4.09E-01	2.85E-01	2.00E-01	1.92E-01	2.13E-01	1.38E-01	1.48E-01	1.77E-01	0.00E+00	4
21	4.63E+02	7.10E+00	4.10E+00	2.31E+00	2.44E+00	1.54E+00	1.72E+00	1.56E+00	1.49E+00	2.35E+00	1.43E+00	2.02E+00	1.41E+00	4.47E-01	3.13E-01	1.78E-01	1.49E-01	1.27E-01	1.08E-01	9.07E-02	1.16E-01	7.71E-02	0.00E+00	5
22	3.53E+02	6.23E+00	6.34E+00	3.66E+00	2.86E+00	3.19E+00	2.83E+00	2.09E+00	2.26E+00	1.61E+00	1.63E+00	1.85E+00	1.37E+00	3.68E-01	3.18E-01	2.94E-01	2.20E-01	2.11E-01	1.72E-01	1.36E-01	1.92E-01	1.77E-01	0.00E+00	1
23	4.12E+02	9.29E+00	5.10E+00	3.26E+00	3.62E+00	2.33E+00	2.55E+00	1.62E+00	1.78E+00	1.97E+00	2.08E+00	2.12E+00	1.79E+00	4.57E-01	3.43E-01	2.11E-01	1.54E-01	1.81E-01	1.48E-01	1.54E-01	1.46E-01	1.18E-01	0.00E+00	1
24	4.78E+02	5.65E+00	4.32E+00	3.30E+00	2.55E+00	2.26E+00	1.75E+00	1.89E+00	1.63E+00	2.50E+00	2.23E+00	1.91E+00	2.33E+00	4.44E-01	3.66E-01	1.81E-01	1.35E-01	1.20E-01	1.41E-01	1.40E-01	1.24E-01	1.22E-01	0.00E+00	5
25	3.98E+02	7.09E+00	6.72E+00	3.94E+00	3.43E+00	2.78E+00	3.33E+00	1.87E+00	2.02E+00	2.85E+00	2.19E+00	3.22E+00	2.63E+00	4.42E-01	3.48E-01	2.24E-01	2.48E-01	1.81E-01	1.61E-01	1.72E-01	1.78E-01	1.14E-01	0.00E+00	5
26	4.09E+02	7.86E+00	5.11E+00	4.34E+00	3.61E+00	2.04E+00	2.32E+00	2.06E+00	2.35E+00	2.08E+00	1.72E+00	1.64E+00	1.52E+00	4.70E-01	3.26E-01	1.89E-01	1.96E-01	1.77E-01	1.48E-01	1.25E-01	1.38E-01	1.31E-01	0.00E+00	3
27	3.44E+02	7.70E+00	5.54E+00	3.60E+00	3.92E+00	2.31E+00	3.10E+00	2.72E+00	3.03E+00	2.02E+00	2.61E+00	2.69E+00	2.13E+00	3.89E-01	3.94E-01	2.09E-01	2.17E-01	1.81E-01	1.60E-01	1.84E-01	1.29E-01	1.03E-01	0.00E+00	2
28	9.92E+01	8.56E+00	5.63E+00	3.50E+00	3.78E+00	3.31E+00	2.74E+00	2.09E+00	2.06E+00	1.71E+00	1.56E+00	1.39E+00	1.35E+00	3.28E-01	2.93E-01	2.50E-01	2.11E-01	1.45E-01	1.47E-01	1.44E-01	1.33E-01	1.45E-01	0.00E+00	3
29	4.19E+02	7.57E+00	5.10E+00	4.08E+00	3.89E+00	3.24E+00	2.55E+00	2.14E+00	3.41E+00	2.44E+00	3.03E+00	2.88E+00	2.29E+00	4.60E-01	3.61E-01	2.14E-01	1.81E-01	1.71E-01	1.40E-01	1.27E-01	1.18E-01	1.10E-01	0.00E+00	3

dataset.csv - Excel

A1 File Name

	BE	BF	BG	BH	BI	BJ	BK
1	Area Method of Moments of MFCCs 5	Area Method of Moments of MFCCs 6	Area Method of Moments of MFCCs 7	Area Method of Moments of MFCCs 8	Area Method of Moments of MFCCs 9	Area Method of Moments of MFCCs 10	Emotion
2	1.42E+06	4.86E+04	8.39E+12	1.31E+07	4.53E+05	1.13E+09	Angry
3	2.52E+06	2.28E+05	8.20E+12	2.30E+07	2.09E+06	5.30E+09	Angry
4	1.64E+06	7.15E+04	9.05E+12	1.50E+07	6.56E+05	1.67E+09	Angry
5	2.86E+06	2.79E+05	8.39E+12	2.61E+07	2.55E+06	6.53E+09	Angry
6	2.31E+06	1.82E+05	8.83E+12	2.09E+07	1.66E+06	4.28E+09	Angry
7	3.15E+06	3.21E+05	8.64E+12	2.86E+07	2.91E+06	7.58E+09	Angry
8	1.83E+06	1.08E+05	9.46E+12	1.68E+07	9.99E+05	2.56E+09	Angry
9	3.25E+06	3.40E+05	8.61E+12	2.96E+07	3.10E+06	8.10E+09	Angry
10	1.21E+06	2.14E+04	8.57E+12	1.11E+07	2.04E+05	5.12E+08	Angry
11	1.29E+06	5.71E+04	9.30E+12	1.22E+07	5.35E+05	1.37E+09	Angry
12	1.09E+06	2.02E+04	8.68E+12	1.00E+07	1.90E+05	4.83E+08	Angry
13	2.00E+06	1.23E+05	1.07E+13	1.85E+07	1.14E+06	2.95E+09	Angry
14	1.36E+06	5.29E+04	8.46E+12	1.22E+07	4.70E+05	1.27E+09	Angry
15	2.34E+06	1.94E+05	8.94E+12	2.12E+07	1.76E+06	4.67E+09	Angry
16	1.28E+06	3.27E+04	9.02E+12	1.16E+07	3.00E+05	7.88E+08	Angry
17	2.85E+06	2.49E+05	1.03E+13	2.61E+07	2.28E+06	6.01E+09	Angry
18	1.04E+06	5.50E+03	8.76E+12	9.44E+06	5.06E+04	1.33E+08	Angry
19	2.75E+06	2.40E+05	1.01E+13	2.54E+07	2.20E+06	5.82E+09	Angry
20	2.46E+06	2.06E+05	9.66E+12	2.26E+07	1.89E+06	5.02E+09	Angry
21	1.50E+06	8.19E+04	8.43E+12	1.38E+07	7.55E+05	2.00E+09	Angry
22	1.65E+06	5.56E+04	1.11E+13	1.48E+07	4.92E+05	1.36E+09	Angry
23	1.90E+06	1.28E+05	9.71E+12	1.77E+07	1.19E+06	3.13E+09	Angry
24	2.17E+06	1.78E+05	8.88E+12	2.03E+07	1.66E+06	4.37E+09	Angry
25	4.54E+06	4.91E+05	1.14E+13	4.11E+07	4.45E+06	1.21E+10	Angry
26	1.88E+06	1.01E+05	1.08E+13	1.74E+07	9.38E+05	2.50E+09	Angry
27	2.76E+06	2.45E+05	1.02E+13	2.52E+07	2.24E+06	6.12E+09	Angry
28	1.84E+06	8.30E+04	1.14E+13	1.68E+07	7.61E+05	2.08E+09	Angry
29	3.09E+06	2.97E+05	1.02E+13	2.80E+07	2.69E+06	7.45E+09	Angry

dataset

Filename	first formant	second formant	third formant	Min inensity	Max intensity	Mean intenity	Min pitch	Max pitch	Mean pitch
03-01-05-01-01-01-01	709.5436947345565	1094.3941224218995	2782.4766292062995	68.74	68.74	68.74	211.36	508.02	260.70
03-01-05-01-01-01-02	502.7193917597737	2688.463996785892	2954.425527791287	57.35	57.35	57.35	200.43	391.74	280.38
03-01-05-01-01-01-03	455.73891426994885	2152.8344684453864	2756.1253736899243	42.08	42.08	42.08	199.72	550.63	301.31
03-01-05-01-01-01-04	429.5963851505157	1111.4440704503197	3005.5126239354636	60.83	60.83	60.83	199.75	402.99	282.78
03-01-05-01-01-01-05	1126.2350288280281	2508.189223544073	2683.37015874763	50.11	50.11	50.11	204.94	534.66	291.70
03-01-05-01-01-01-06	457.91747486191184	2612.871884920839	3097.0709668794643	50.28	50.28	50.28	206.08	350.05	253.85
03-01-05-01-01-01-07	934.9304798270165	2258.635749356093	3472.806082457927	52.62	52.62	52.62	199.74	574.57	263.22
03-01-05-01-01-01-08	301.598497611885	2021.204842194419	2926.36256071997	60.85	60.85	60.85	203.60	321.10	246.30
03-01-05-01-01-01-09	324.32745038818143	1709.4987943249716	2702.9985292177844	40.33	40.33	40.33	206.48	576.29	423.52
03-01-05-01-01-01-10	432.91641555393943	2078.4315414339594	2432.285046038958	54.87	54.87	54.87	199.87	269.60	230.74
03-01-05-01-01-01-11	752.440355530897	1951.0280929098833	2459.2991537109638	50.53	50.53	50.53	200.52	314.54	228.78
03-01-05-01-01-01-12	606.0985443285775	2515.3597363414774	2789.2764399259368	45.30	45.30	45.30	200.45	272.95	232.51
03-01-05-01-01-01-13	385.34173486893803	2185.5727048194303	2508.7133672926657	55.27	55.27	55.27	215.38	596.11	415.28
03-01-05-01-01-01-14	962.2401639187635	1233.4187608487946	2772.36203410932	41.88	41.88	41.88	199.76	483.25	314.52
03-01-05-01-01-01-15	386.5452303028606	2188.2056762271804	2816.4748941310418	62.05	62.05	62.05	201.65	558.83	288.36
03-01-05-01-01-01-16	1074.6363829505572	1377.2002115424161	2909.8350424012074	38.58	38.58	38.58	199.63	344.81	255.38
03-01-05-01-01-01-17	204.96929415736176	1425.7711076897835	2082.927537196665	46.88	46.88	46.88	219.40	323.79	277.75
03-01-05-01-01-01-18	895.0595152322939	1328.8662514354671	2981.6011957063233	35.80	35.80	35.80	199.68	345.33	252.33
03-01-05-01-01-01-19	373.52000263693924	2213.82538864467	2617.039739708075	50.75	50.75	50.75	200.58	434.53	267.85

Results

We were able to classify audio into emotions with 88% accuracy using Random Forest Classifier.

Firewall Authenticati | My Drive - Goog | Emotion Detecti | The Ryerson Audio-V | Desktop/start/ | Logistic Regressi | PCA | Preprocessing | Density_Plots

localhost:8888/notebooks/Desktop/start/PCA.ipynb

jupyter PCA Last Checkpoint: 12/13/2018 (autosaved)

Logout

File Edit View Insert Cell Kernel Widgets Help

Trusted | Python 3

Code

Out[228]:
```
array([0, 3, 2, 0, 3, 5, 7, 7, 6, 1, 3, 0, 6, 4, 0, 7, 3, 4, 1, 2, 6, 6,
       6, 2, 4, 3, 2, 2, 0, 2, 4, 6, 6, 3, 2, 4, 0, 1, 6, 0, 1, 6, 2, 1,
       4, 7, 0, 1, 7, 1, 6, 3, 7, 6, 3, 4, 2, 6, 5, 0, 2, 7, 7, 1, 7, 1,
       7, 0, 4, 6, 4, 1, 4, 0, 1, 1, 6, 1, 3, 3, 2, 0, 2, 6, 7, 6, 1, 0,
       0, 2, 1, 6, 4, 2, 2, 2, 2, 6, 5, 0, 3, 6, 3, 1, 4, 3, 7, 2, 6, 3,
       1, 4, 2, 2, 4, 4, 6, 3, 4, 5, 1, 1, 4, 0, 5, 2, 6, 2, 0, 3, 4, 5,
       3, 4, 1, 2, 6, 1, 7, 7, 3, 3, 3, 1, 0, 3, 1, 3, 4, 4, 0, 4, 6, 3,
       4, 2, 4, 1, 2, 1, 5, 1, 3, 2, 6, 0, 3, 0, 1, 3, 3, 4, 5, 3, 4, 7,
       0, 0, 7, 4, 4, 1, 4, 2, 4, 3, 6, 1, 2, 6, 2, 0, 3, 5, 2, 2, 7, 6,
       6, 7, 2, 6, 4, 7, 6, 7, 3, 7, 5, 5, 2, 4, 7, 7, 5, 2, 1, 6, 1, 0,
       4, 4, 1, 4, 0, 0, 2, 4, 7, 4, 6, 1, 5, 1, 4, 2, 7, 2, 7, 3, 3, 3,
       1, 6, 6, 1, 7, 0, 6, 0, 7, 1, 6, 1, 6, 0, 6, 3, 4, 1, 5, 2, 6, 4,
       6, 1, 7, 4, 7, 1, 2, 1, 5, 6, 3, 2, 4, 1, 6, 3, 7, 4, 0, 7, 1, 7,
       2, 7, 6, 3, 1, 0, 4, 7, 0, 2, 1, 7, 4, 4, 2, 7, 6, 0, 1, 3, 7, 3,
       4, 1, 0, 7, 1, 3, 1, 2, 0, 0, 6, 6, 1, 6, 6, 7, 2, 0, 6, 6, 2, 6,
       4, 6, 6, 1, 2, 7, 2, 3, 1, 6, 2, 3, 6, 2, 7, 5, 2, 4, 0, 4, 4, 1,
       4, 0, 5, 0, 2, 3, 4, 6, 1, 5, 6, 7, 1, 7, 7, 0, 6, 2, 3, 5, 1, 0,
       3, 3, 6, 1, 1, 0, 6, 3, 5, 7, 5, 3, 1, 3, 6, 7, 2, 2, 0, 0, 3, 1,
       1, 4, 1, 2, 5, 2, 0, 3, 1, 1, 3, 6, 5, 0, 3, 5, 2, 3, 3, 5, 4, 0,
       4, 6, 2, 2, 2, 6, 3, 2, 3, 3, 6, 6, 1, 3], dtype=int64)
```

In [229]:
```python
from sklearn import metrics
accuracy = metrics.accuracy_score(y_test, y_pred)*100
```

In [230]:
```python
accuracy
```

Out[230]:
```
88.42592592592592
```

In [ ]:

In [ ]:

Type here to search

ENG 10:41 26-03-2019

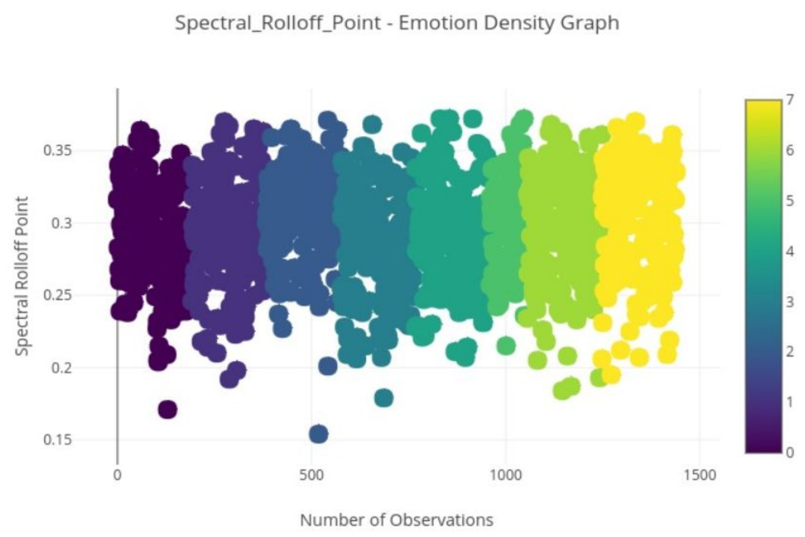
Why we did Speech Emotion Recognition or SER

To identify the features of speech which are important for classifying speech into various emotions so that in the later part of the project these features can be modified to give an emotional base to the audio.

After doing Feature Selection we identified a set of 27 features important for audio classification.

Some of these features are:

1. Spectral Rolloff Point
2. Pitch and Frequency
3. MFCCs, Method of Moments 1




```
In [28]: X = dataset.iloc[:, 34:35].values
```

```
In [29]: graph_labels = ["Number of Observations", "Method of Moments 1", "Method_Of_Moments_1 - Emotion Density Graph"]

         scatter_with_color_dimension_graph(list(X),y, graph_labels)
```

Firewall Authenticati | My Drive - Goog | Emotion Detecti | The Ryerson Audio-V | Desktop/start/ | Logistic Regressi | PCA | Preprocessing | Density_Plots

localhost:8888/notebooks/Desktop/start/Density_Plots.ipynb

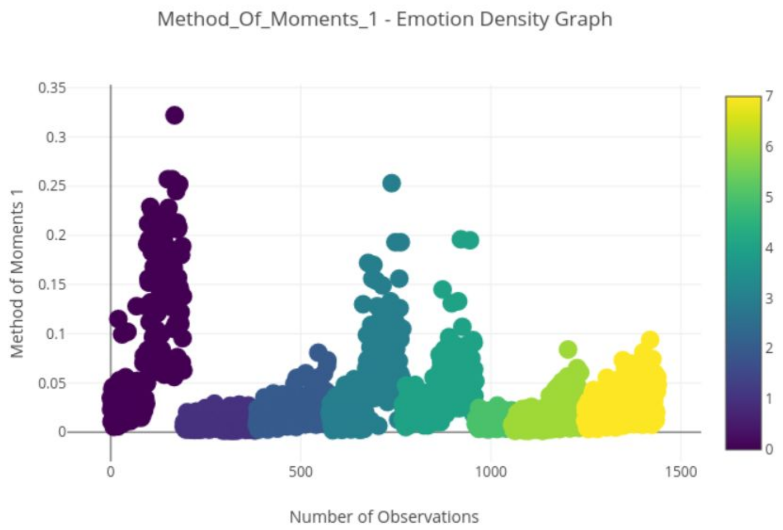
jupyter **Density_Plots** Last Checkpoint: 12/10/2018 (autosaved)

Logout

File Edit View Insert Cell Kernel Widgets Help

Trusted | Python 3

Code



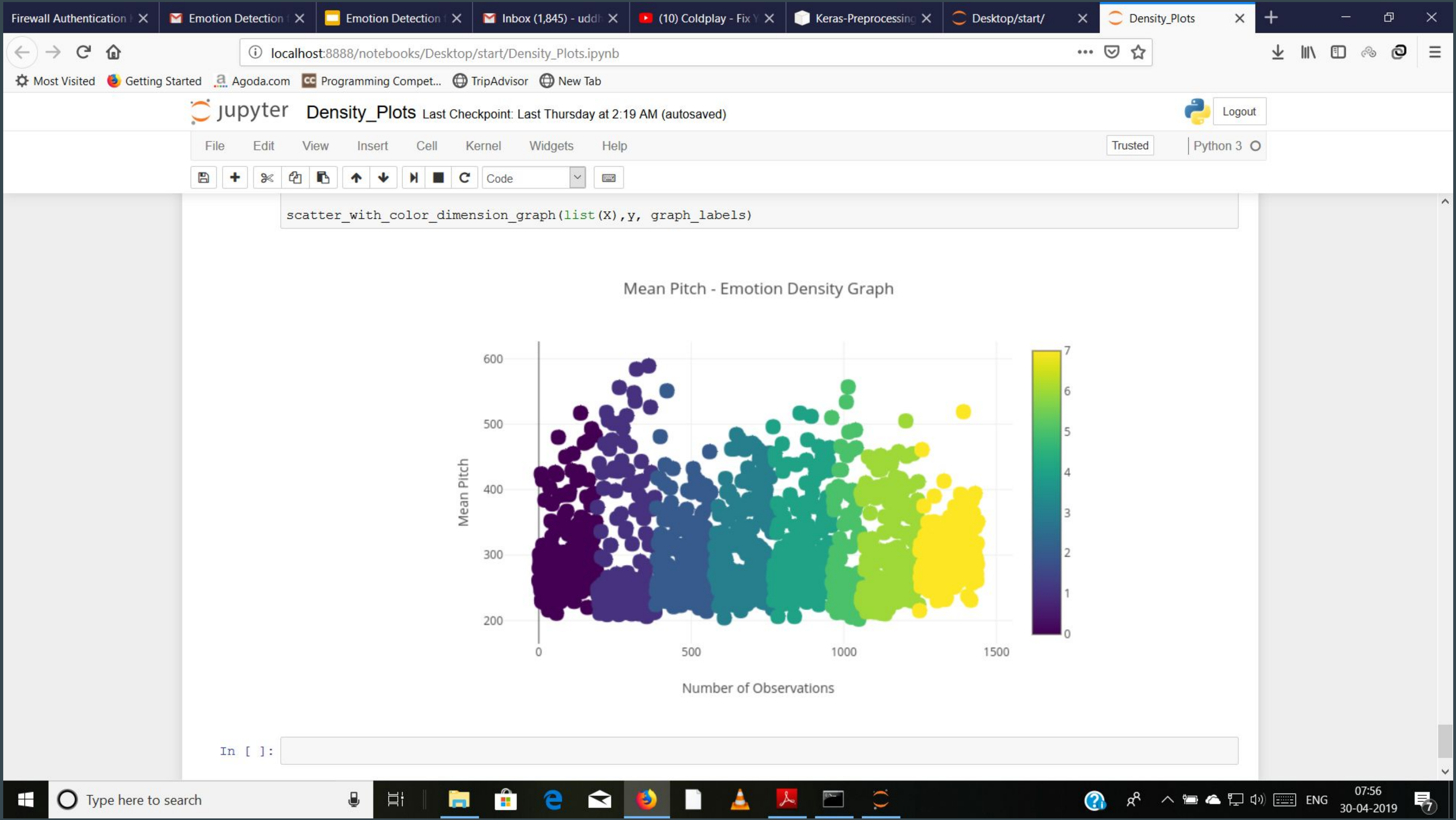
In [30]:
```
X = dataset.iloc[:, 60].values
```

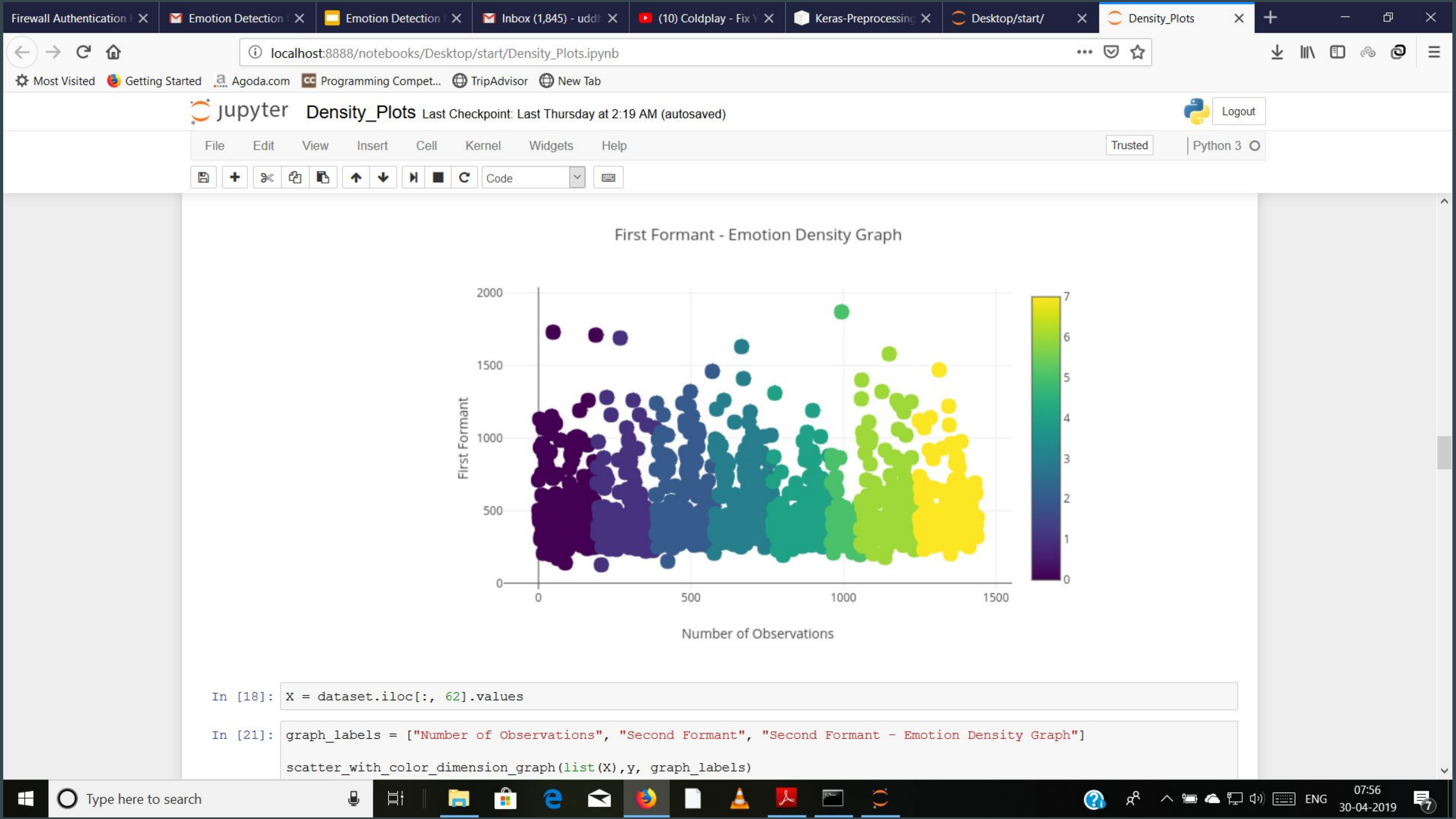
In [31]:
```
graph_labels = ["Number of Observations", "Method of Moments 10", "Method_Of_Moments_10 - Emotion Density Graph"]

scatter_with_color_dimension_graph(list(X),y, graph_labels)
```

Type here to search

ENG 10:44 26-03-2019


localhost:8888/notebooks/Desktop/start/Density_Plots.ipynb
jupyter Density_Plots Last Checkpoint: Last Thursday at 2:19 AM (autosaved)
File Edit View Insert Cell Kernel Widgets Help
Trusted
Python 3
Logout
scatter_with_color_dimension_graph(list(X),y, graph_labels)
Mean Pitch - Emotion Density Graph
Mean Pitch
600
500
400
300
200
0
500
1000
1500
Number of Observations
0
1
2
3
4
5
6
7
In [ ]:
Type here to search
ENG
07:56
30-04-2019


localhost:8888/notebooks/Desktop/start/Density_Plots.ipynb
jupyter Density_Plots Last Checkpoint: Last Thursday at 2:19 AM (autosaved)
File Edit View Insert Cell Kernel Widgets Help
Trusted
Python 3
Code
First Formant - Emotion Density Graph
First Formant
Number of Observations
0
500
1000
1500
2000
0
1
2
3
4
5
6
7
In [18]: X = dataset.iloc[:, 62].values
In [21]: graph_labels = ["Number of Observations", "Second Formant", "Second Formant - Emotion Density Graph"]
scatter_with_color_dimension_graph(list(X),y, graph_labels)
Type here to search
ENG
07:56
30-04-2019

What do we have uptil now?

1. Text with the corresponding emotion
2. Features important for classifying audio into their corresponding emotions

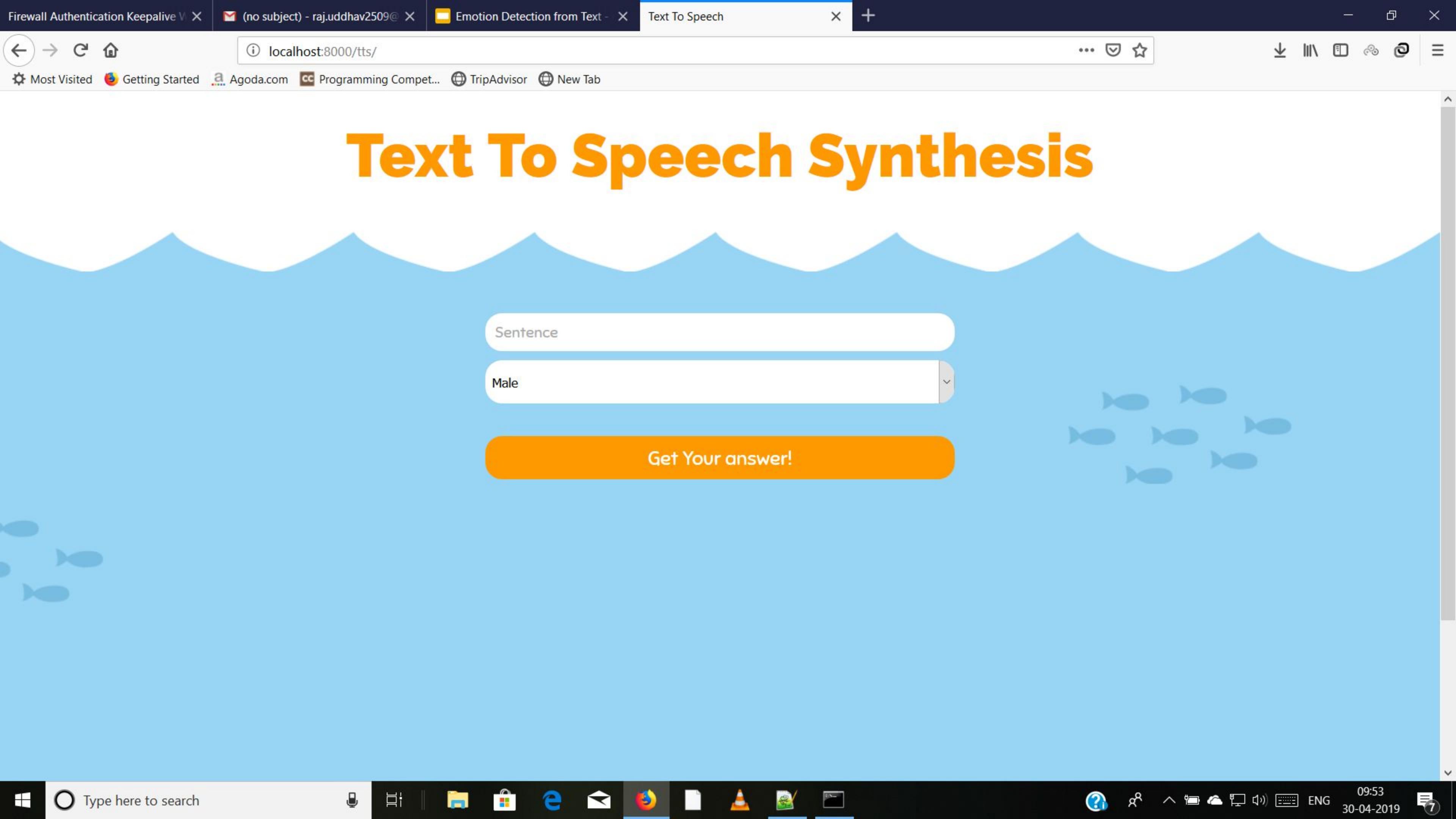
Part 3 : Giving The Neutral Audio an Emotional Base

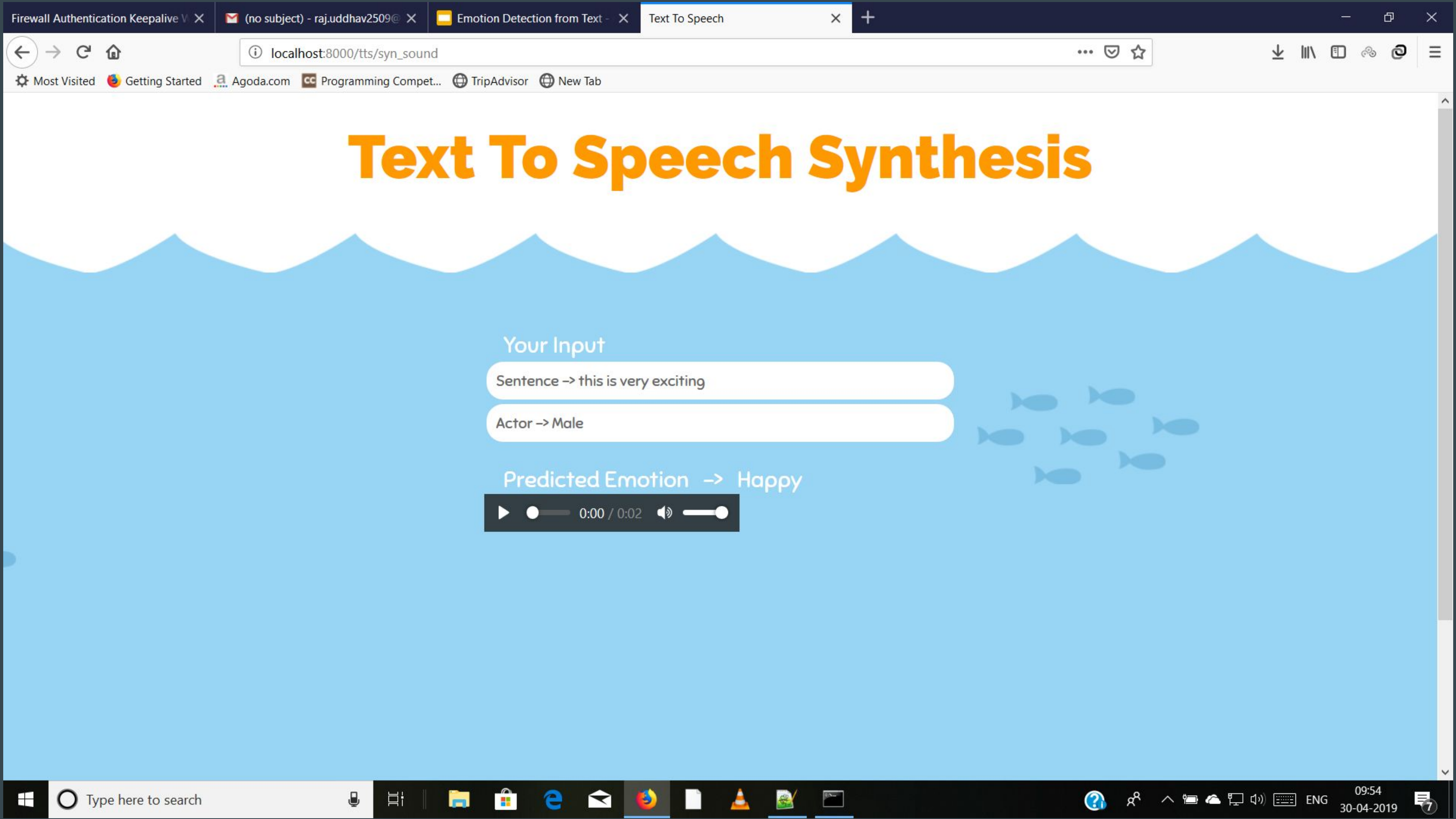
Tools used to modify features

1. Pyttsx - A python library to generate a neutral audio
2. Ffmpeg - A command line tool
3. pydub - A python library Modifying pitch and frequency
4. Librosa and pyaudio - Modifying spectral rolloff point and some other features

Results

The Interface


Firewall Authentication Keepalive W
(no subject) - raj.uddhav2509@
Emotion Detection from Text -
Text To Speech
localhost:8000/tts/
Most Visited
Getting Started
Agoda.com
Programming Compet...
TripAdvisor
New Tab
Text To Speech Synthesis
Sentence
Male
Get Your answer!
Type here to search
ENG
09:53
30-04-2019


Firewall Authentication Keepalive
(no subject) - raj.uddhav2509@
Emotion Detection from Text -
Text To Speech
localhost:8000/tts/syn_sound
Most Visited
Getting Started
Agoda.com
Programming Compet...
TripAdvisor
New Tab
Text To Speech Synthesis
Your Input
Sentence -> this is very exciting
Actor -> Male
Predicted Emotion -> Happy
0:00 / 0:02
Type here to search
ENG
09:54
30-04-2019

Some Samples

How to measure accuracy?

A big problem since there is no known method of knowing “How natural a voice is sounding”

What we did

10 samples for each of the 8 emotions were generated and feature extraction was performed on each of these samples. (All the features that we used in SER)

Then the same classifier that we trained in SER was used to classify these samples. And the classifier accurately predicted the emotion of these samples.

Conclusion

...

We were successful in designing and implementing a system which was capable of doing Text To Speech Synthesis.

For this we implemented three systems:

1. Emotion Detection From Text
2. Speech Emotion Recognition
3. Emotional Touch To Neutral Audio

We integrated all three of these and implemented the final product.

Applications

…

1. A boon for the Blind - People who are blind can listen to their favourite book in a very humanly voice.

2. An E-Book Reader - To make learning more interactive specially for small children.

3. Speech Emotion Recognition can be used in Call Centres to detect how satisfied was a customer with the executive handling the call.

4. Emotion Detection From Text can be used to remove hate speeches from social media sites.

Future Work

Two important parameters we found from speech emotion recognition was area method of moments and MFCCs, but we were not able to directly manipulate these features. Manipulation of these features may lead to more emotional sounding speech. Longer training times on GPUs and larger datasets will definitely improve the quality of the speech generated by our system.

Tacotron and Wavenet

Mel Spectrogram - Audio Representation

A Mel Spectrogram represents an acoustic time-frequency representation of a sound: the power spectral density $P(f, t)$. It is sampled into a number of points around equally spaced times t_i and frequencies f_j (on a Mel frequency scale).

The mel frequency scale is defined as:

$$\text{mel} = 2595 * \log10\,(1 + \text{hertz} / 700),$$

and its inverse is:

$$\text{hertz} = 700 * (10.0^{\text{mel} / 2595.0} - 1).$$

Tacotron

Tacotron is a deep neural network architecture that is capable of speech synthesis. It is an end-to-end TTS system with a sequence-to-sequence recurrent network that predicts mel spectrograms. It can be directly trained from data and can achieve reasonable sound quality.

Wavenet

WaveNet is a deep generative model of audio data that operates directly at the waveform level. WaveNets combine causal filters with dilated convolutions to allow their receptive fields to grow exponentially with depth, which is important to model the long-range temporal dependencies in audio signals.

Combining the two

Tacotron takes as input <Text, Audio> pairs and outputs mel spectrograms. Wavenet converts the mel spectrograms output by Tacotron to audio waveforms. This completes the synthesis of audio from text in an end-to-end fashion.

Future Work

Very recent, cutting edge research in text-to- speech suggests that an end to end system for generating speech from text, is able to generate higher quality audio, but at the expense of a much longer training and generation time[11]. But these architectures do not explicitly take an emotion label as input, they only take text and the corresponding audio. We feel that by modifying this architecture to incorporate emotion label as an input also, we would be able to generate emotional audio in an end to end fashion.

[11]Y. Wang, R. J. Skerry-Ryan, D. Stanton, Y. Wu, R. J. Weiss, N. Jaitly, Z. Yang, Y. Xiao, Z. Chen, S. Bengio, Q. V. Le, Y. Agiomyrgiannakis, R. Clark, and R. A. Saurous, "Tacotron: A fully end-to-end text-to- speech synthesis model," CoRR, vol. abs/1703.10135, 2017.

Thank You