# REPORT- HOMEWORK 1

**-Nandan PC(nandanpc@usc.edu/ 3394 068 908)**
**-Sree Priyanka Uppu(uppu@usc.edu/ 6822 468 564)**

1. The following nonverbal and verbal cues are associated with the positive/negative and neutral opinions:
    - Negative Opinions: Negative opinions are associated mainly with violent bursts of anger. We observed that when people are expressing this kind of emotion they tend to speak fast, have a very high pitch, strong tone, deep release of breadths, lot of pauses as they have a loss of words, tend to gaze away, keep moving their face or body a lot and in some cases, eyes are narrowed. Also in facial expression the eyebrow cringes. Even though they smile their cheek doesn't raise indicating anger or sarcasm.
    - Positive Opinions: These are associated with happiness or laughter. We observed that when people are expressing this kind of emotion they tend to have a varying pitch from normal to high (when expressing something exciting the pitch rises from normal to high), smile or laugh more often, tend to tilt their head to about 45 degree angle when they laugh genuinely (in some cases) and then their gaze becomes direct. Facially their eyebrow raises along with smile cheek too raises expressing happy feeling.
    - Neutral Opinions: These are associated with calm feelings when giving a review or just a causal talk .We observed that when people are expressing this kind of emotion they tend to have a normal tone which is consistent, keep their face straight and not move their face a lot, and tend to have a direct gaze. There isn't any changes in their eyebrow nor do they smile a lot.

2. We calculated the Krippendorf's alpha on the Matlab where we passed the annotator scores of the data and used Nominal as the scale. (We used this scale over the others like, ordinal, interval, ratio, etc since the numbers assigned by the annotators have no real meaning other than differentiating between the classes) and we got a value of 0.5845. An alpha value of 0.6 and above implies moderate agreement among the annotators. In our case, we have a 0.5845,which means that there is a very low inter rater agreement and hence are not consistent. Without a proper establishment of intercoder reliability, the interpretation cannot be valid and can be subjected to doubt. If it was higher than 0.6, it was possible to have just 1 annotator, since they seem to agree and having one would have saved cost and man power.

3. Hypothesis:
    a. NAQ in acoustic file measures the tenseness in voice. According to our hypothesis, we get that NAQ is statistically significant (p= 0.00023) in differentiating between both positive and negative groups (from TTEST), as well as in differentiating among the positive, negative and zero groups (from ANOVA).

       **Results from the tests:**

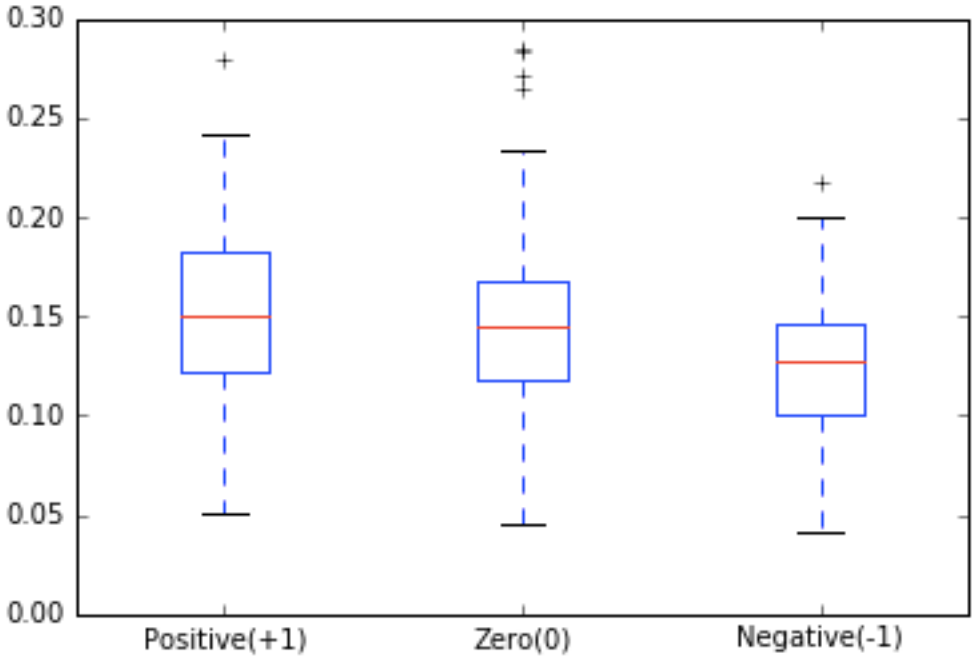
Anova: statistic=8.62719, pvalue=0.00023

Ttest: (between positive and negative groups)
statistic=3.86498, pvalue=0.00015

| | **Positive (+1)** | **Negative (-1)** | **Zero (0)** |
|---|---|---|---|
| **Mean** | 0.14824 | 0.12647 | 0.14697 |
| **Median** | 0.14960 | 0.12743 | 0.14483 |
| **Standard Deviation** | 0.04361 | 0.03395 | 0.04556 |

**Conclusion:**
*Higher value of NAQ indicates more tenseness. So, when people show Positive emotion, they tend to be tenser. And while, showing negative emotion, they tend to be less tense.*

**Box Plot:**



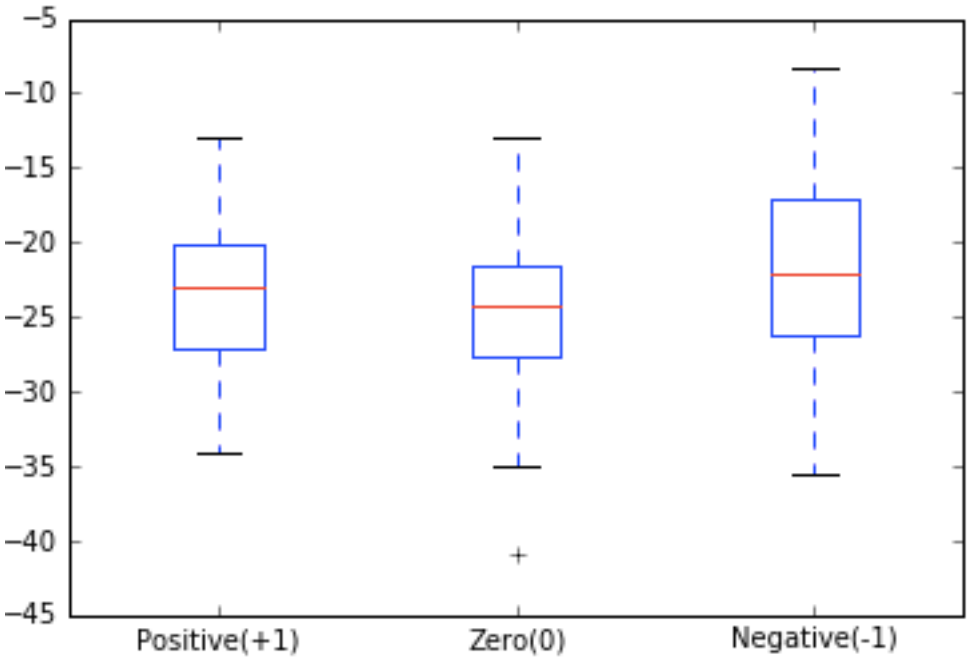

b. Energy: in acoustic file measures the tenseness in voice. According to our hypothesis, we get that Energy is statistically significant only for the top 25% of Energy values (decreasing energy values, measures how loud can each category can get to) in each category. The reason, I took in such a way is that when we exhibit emotion in a segment say, shout, we shout only for 3sec of the 7sec and that 3sec causes the annotators to mark the segment as -1. When we calculate mean of the entire segment, the higher values are not getting attributed to in the mean. Hence, by looking at top25% of values in each category we can get a better conclusion that p= 0.03125 and it is statistically significant in differentiating between both positive and negative groups (from TTEST), as well as in differentiating among the positive, negative and zero groups (from ANOVA).

**Results from the tests:**
*TTEST*(between positive and negative groups): statistic=-2.16992, pvalue=0.03125
*ANOVA*: statistic=5.6826, pvalue=0.00381

| | **Positive (+1)** | **Negative (-1)** | **Zero (0)** |
|---|---|---|---|
| **Mean** | -23.63168 | -21.84045 | -24.43674 |
| **Median** | -22.98376 | -22.18391 | -24.18616 |
| **Standard Deviation** | 5.18640 | 6.04811 | 4.99024 |

**Conclusion:**
*Higher negative value of energy indicates it is less loud. So, when people show Positive emotion, they tend to be less loud. And while, showing negative emotion, they tend to be louder (higher value of Energy)*

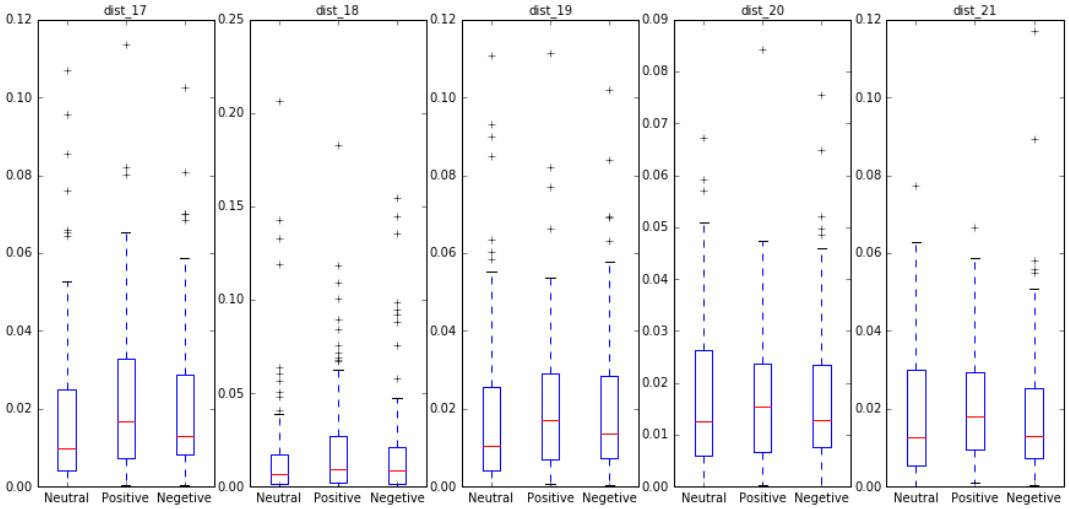
**BoxPlot:**




**4. Features related to Facial:**
1. Eye brow variation: CLM output has been used to analyse eye brow feature. Normal distance has been taken from point 30 i.e. wrt nose and all the individual eyebrow points from point 17 to 26. Then variation of distance within each segment of video is calculated. The box plot below shows how far the variance of distance taking complete eyebrow into account and individual data point into account.

a. Left eye brow individual points

**Results from the tests:**

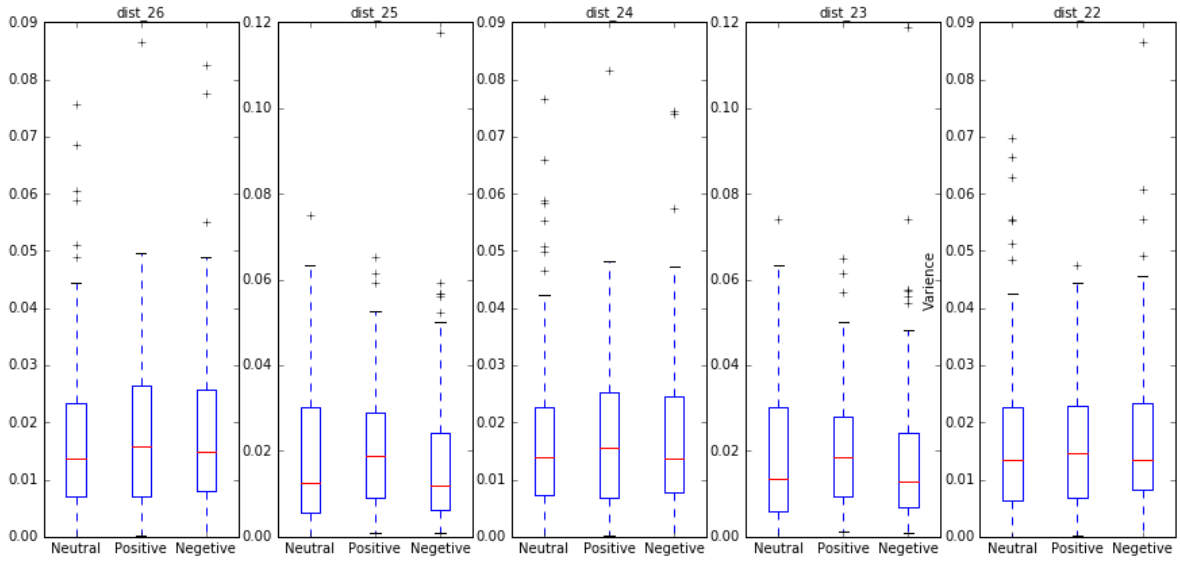



| neutral | dist_17 | dist_18 | dist_19 | dist_20 | dist_21 |
|---|---|---|---|---|---|
| Mean | 0.019225 | 0.01785 | 0.019103 | 0.01715 | 0.019628 |
| Std | 0.022813 | 0.033326 | 0.022748 | 0.015552 | 0.01845 |
| | | | | | |
| Positive | dist_17 | dist_18 | dist_19 | dist_20 | dist_21 |
| Mean | 0.022354 | 0.022361 | 0.022098 | 0.017088 | 0.021659 |
| Std | 0.020154 | 0.032274 | 0.019836 | 0.013779 | 0.015543 |
| | | | | | |
| Negative | dist_17 | dist_18 | dist_19 | dist_20 | dist_21 |
| Mean | 0.020729 | 0.02019 | 0.020601 | 0.016854 | 0.019625 |
| Std | 0.019414 | 0.030415 | 0.01945 | 0.014827 | 0.018664 |

**dist_17 is the distance between facial point 17 and 30.*

| | Dist 17 | Dist 18 | Dist 19 | Dist 20 | Dist 21 |
|---|---|---|---|---|---|
| ANOVA | 0.4999 | 0.43914 | 0.46246 | 0.01094 | 0.39911 |
| P-Value | 0.60711 | 0.64503 | 0.63021 | 0.98911 | 0.67129 |

## b. Right eye brow individual points :




| neutral | dist_22 | dist_23 | dist_24 | dist_25 | dist_26 |
|---|---|---|---|---|---|
| Mean | 0.017507 | 0.019863 | 0.017836 | 0.01941 | 0.01789 |
| Std | 0.015995 | 0.017977 | 0.01616 | 0.018018 | 0.015865 |
| | | | | | |
| Positive | dist_22 | dist_23 | dist_24 | dist_25 | dist_26 |
| Mean | 0.016697 | 0.021438 | 0.017977 | 0.020897 | 0.018627 |
| Std | 0.011952 | 0.015204 | 0.014576 | 0.014599 | 0.015563 |
| | | | | | |
| Negative | dist_22 | dist_23 | dist_24 | dist_25 | dist_26 |
| Mean | 0.017829 | 0.019201 | 0.017596 | 0.018385 | 0.018433 |
| Std | 0.015188 | 0.017836 | 0.014966 | 0.017516 | 0.015661 |

| | Dist 22 | Dist 23 | Dist 24 | Dist 25 | Dist 26 |
|---|---|---|---|---|---|
| ANOVA | 0.15047 | 0.42155 | 0.01548 | 0.53311 | 0.05257 |
| P-Value | 0.86037 | 0.65644 | 0.98463 | 0.58737 | 0.94879 |

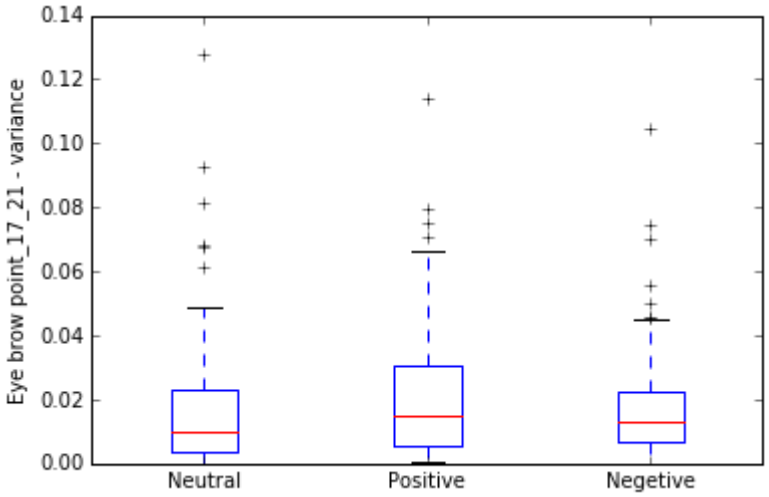
**Conclusion :**

From the above result observe the variance of right and left eye brow individual points p-value for ANOVA test is significantly higher than critical value. This indicating the each individual points taken separately doesn't clearly distinguish neutral , positive and negative videos.

c. So we considered taking all points in a eyebrow and check how does all contribute in distinguishing the neutral, positive and negative videos.

**Results from the tests:**

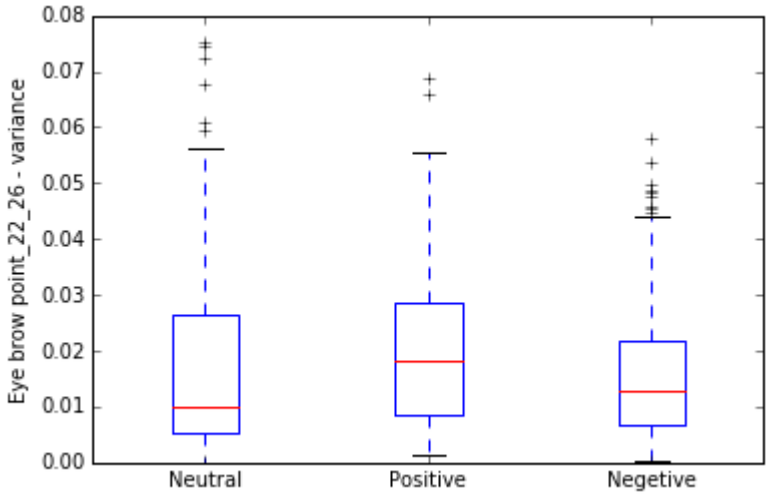
Left Eyebrow Variance




| Point 17 – 21 | neutral | Positive | Negative |
|---|---|---|---|
| Mean | 0.018087 | 0.022113 | 0.018015 |
| Std | 0.022301 | 0.021137 | 0.017538 |

**ANOVA**
**Statistic =1.2140807734837786,**
**Pvalue =0.29855819814326062**

Point 22-26 (Right Eye brow – variance )




| Point 22-26 | neutral | Positive | Negative |
|---|---|---|---|
| Mean | 0.018376 | 0.020621 | 0.016997 |
| Std | 0.019487 | 0.01484 | 0.013843 |

**ANOVA**

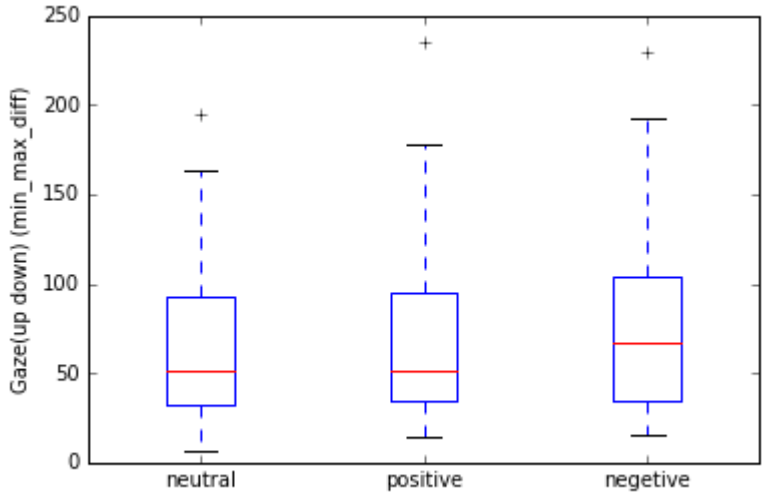
**Statistic =1.2141167871146616**
**P-value =0.29854753960381114**

**Conclusion:**

Again when taking all the eyebrow points combined p-values are more than critical values. Hence we can conclude from tests that eye brow feature confirms null hypothesis and we cannot consider eye brow raise to be used to distinguish between neutral , positive and negative videos.

2. Feature : Gaze(Up Down) : From the OKAO output we have taken column 125 for gaze updown. We observed there were significant difference in change in gaze while we viewed the data. So we took maximum and minimum head gaze difference in each segment of the video and computed the results.

Results from the tests :




| Gaze Level(up and down) | Neutral | Positive | Negative |
|---|---|---|---|
| Mean | 66.01 | 68.55 | 75.16 |
| Standard Deviation | 43.17 | 45.49 | 45.94 |

**ANOVA :**

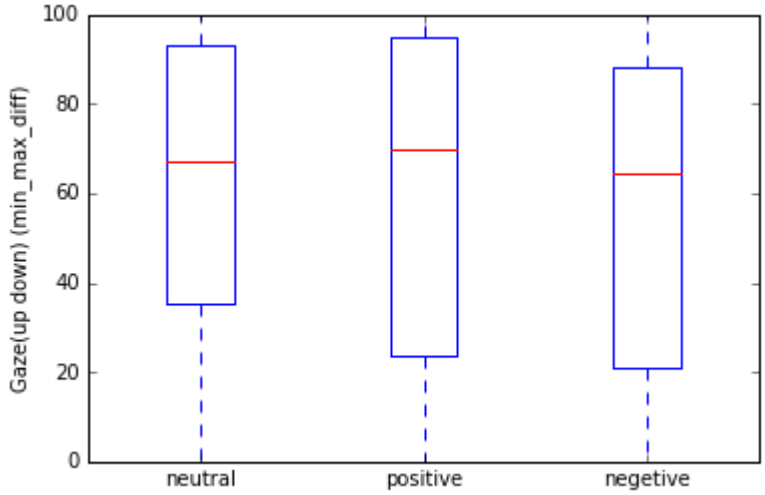
**statistic=1.07495,**
**p-value=0.342732)**

**Conclusion :**

Although the box plot indicates there is slight difference in variance between negative and positive , neutral videos , p-value for the ANOVA is greater than critical value. Hence we cannot consider eye gaze to be feature which statistically distinguish between different groups of videos.

**5. Smile Level:** To analyze smile level we have taken column 134 of OKAO output. We have considered this output since smile was observed more in positive and negative videos. For smile we have taken min max difference of the smile level values.

Results from the test:

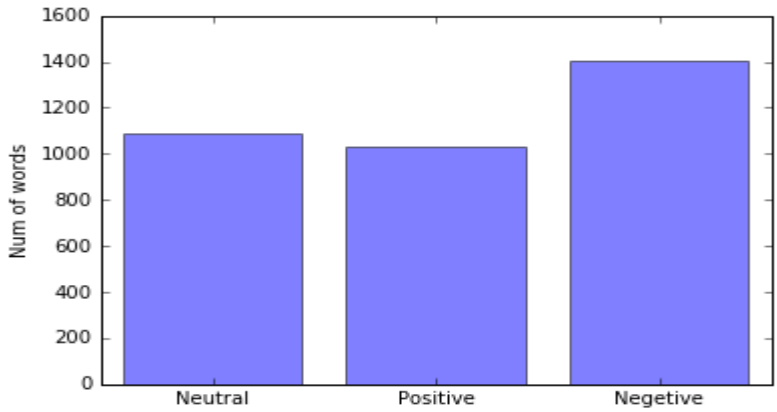



| | Neutral | Positive | Negative |
|---|---|---|---|
| Mean | 62.82954545454545 | 58.79545454545455 | 56.44230769230769 |
| Standard Deviation | 32.61591938750405 | 35.44922192972926 | 34.84719122641631 |

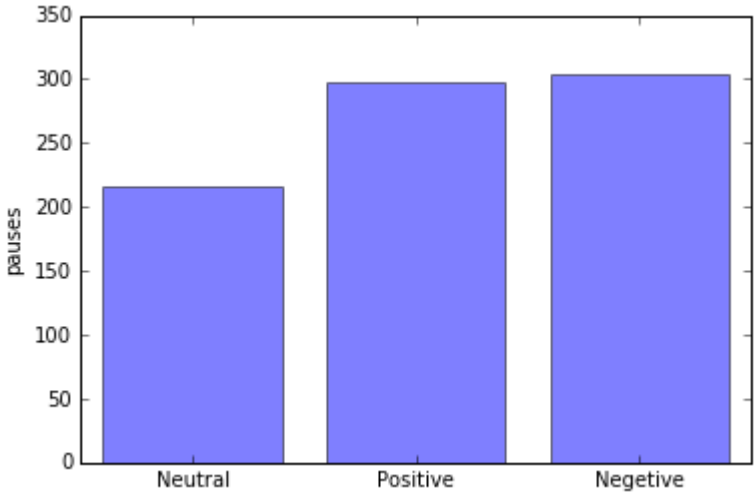
**ANOVA**
**statistic=0.832197,**
**pvalue=0.436177**

**Conclusion :**
Again we see that the p-value for smile level too well above the critical value.
Hence we cannot take smile level too as a feature to distinguish different types of videos.

**6. Feature Number of words spoken:** We observed there was significantly more words spoken in negative videos over positive or neutral videos. We wanted to verify that observation. So we extracted dialogs from each segment of videos.
Here are the results. Number of words spoken :



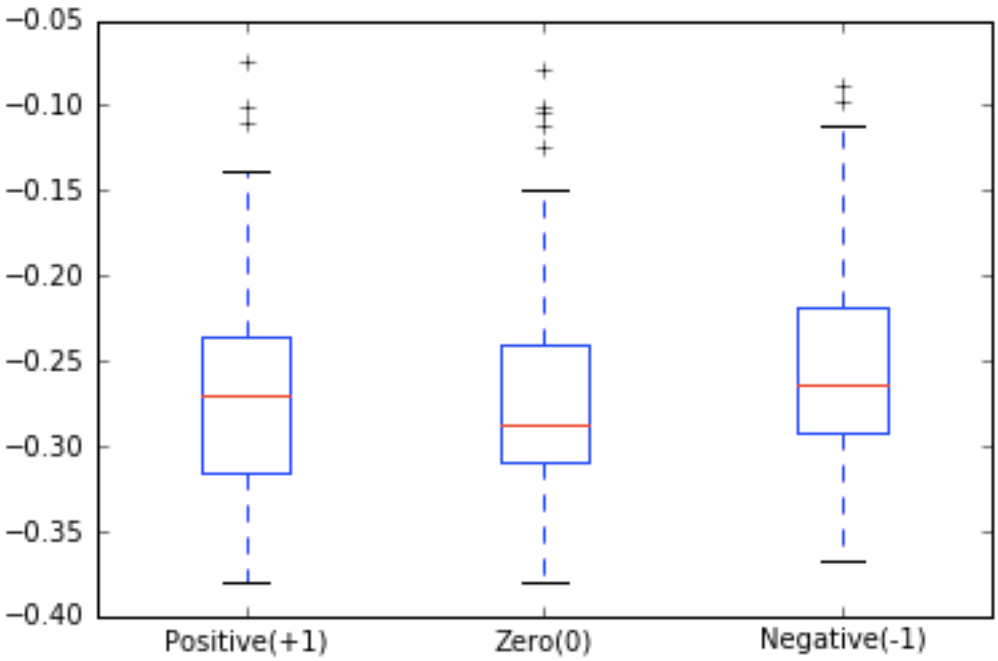
Number of Pauses :




Conclusion:
From the above result we confirm that number of words spoken in a negative videos is more than in neutral and positive videos.

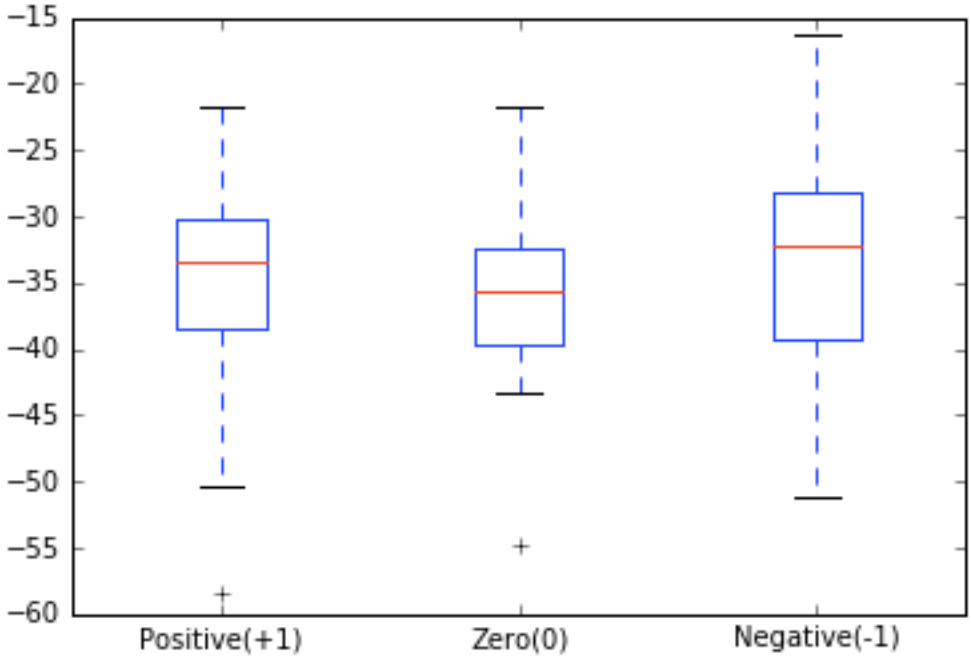
**Other Acoustic features considered:**
We tested the following hypothesis and got statistically insignificant values:

1. Peak Slope in Acoustic:
    a. TTEST(between positive and negative groups): statistic=-1.271348, pvalue=0.20516
    b. ANOVA: (statistic=1.39302, pvalue=0.250058)




2. Energy in Acoustic:
    a. TTEST(between positive and negative groups): statistic=-1.19867,pvalue=0.23215

b. ANOVA(statistic=2.72055, pvalue=0.067598)




CONCLUSION: More variation of Energy is observed when there is a negative emotion associated. This can be because, when people give out violent bursts of rage their anger varies and hence, we observe such a high variation.

3. Fundamental Frequence in Acoustic:
   a. TTEST(between positive and negative groups): (statistic=1.09753, pvalue=0.273797)
   b. ANOVA: (statistic=0.83783, pvalue=0.433742)