LaTeX submissions are mandatory. Submitting your assignment in another format will be graded no higher than R.

1 Group members

Cantao Su, Chenyu Li, Weihao Jiang, Yanhua Liao

2 In-class lab 2.

In this group lab we will be exploring emotional speech. For this we will take a subset from the audio portion of the Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS[1]). Overall, tthere are 1440 files (60 trials per actor x 24 actors = 1440). The RAVDESS contains 24 professional actors (12 female, 12 male) vocalizing two lexically-matched statements in a neutral North American accent. Speech emotions includes calm, happy, sad, angry, fearful, surprise, and disgust expressions. Each expression is produced at two levels of emotional intensity (normal, strong), with an additional neutral expression.

Speech emotions include calm, happy, sad, angry, fearful, surprise, and disgust expressions. Each expression is produced at two levels of emotional intensity (normal, strong), with an additional neutral expression.

The filename consists of a 7-part numerical identifier (e.g., 03-01-06-01-02-01-12.wav). These identifiers define the stimulus characteristics (in bold - those characteristics that are relevant for this lab):

- Modality (01 = full-AV, 02 = video-only, **03 = audio-only**).
- Vocal channel (**01 = speech**, 02 = song).
- Emotion (**01 = neutral**, 02 = calm, **03 = happy**, **04 = sad**, **05 = angry**, **06 = fearful**, 07 = disgust, **08 = surprised**).
- Emotional intensity (01 = normal, **02 = strong**). NOTE: There is no strong intensity for the 'neutral' emotion.
- Statement (01 = "Kids are talking by the door", **02 = "Dogs are sitting by the door"**).
- Repetition (01 = 1st repetition, 02 = 2nd repetition).
- Actor (01 to 24. Odd numbered actors are male, even numbered actors are female).

We will be exploring a small subset of recordings: 10 actors, one trial in neutral state (no emotion) and in five different emotions (anger, fear, happiness, sadness, surprise).

We will explore several simple measurements related to prosody and see if the results tell us anything and if our perceptual observations are comparable.

Preparation:

For your group lab you will need:

- Decide which of the four emotions above you would like to explore.
- Take the recordings of the neutral state speech and the emotional state speech (in the emotion you have chosen) from here.

2.1 Assignment 1

[10pts] *Simple prosody measurements in emotional speech*: We will look into several simple **measurements** that relate to fundamental frequency (f0) and rate:

1. Speech rate (number of syllables / total time) - use either this script by Nivja de Jong and Ton

[1] https://doi.org/10.1371/journal.pone.0196391

Wempe (2009), or the its newer version, you can find tutorials and descriptions on the website as well.

2. Articulation rate (number of syllables / phonation time) - use the same script.
3. In Praat: mean f0 over the whole phrase (select phrase – "Pitch -> get pitch"), you can also do that with your own R/Python script.
4. In Praat: f0 range in the whole phrase (select phrase – "Pitch -> get pitch maximum" and "get pitch minimum", and based on that calculate range), you can also do that with your own R/Python script.
5. f0 variance, average of the squared deviations from the mean of f0: $var = mean|f0 - mean(f0)|^2$ (select phrase – "get pitch listing", save as text file, clean it to keep only the numbers, calculate according to the formula. You can use Python, R, excel – whatever you wish [a])

Task:

1. Listen to the recordings you selected to work with.
2. Together with your group members, hypothesize which of the given measurement(s) could work best to differentiate between recordings of neutral and emotional speech.
3. Calculate all the measurements listed above for every recording (you should get 50 measured values for neutral speech and 50 measured values for emotional speech).
4. In your answer, report how you obtained the measurements and any difficulties you have faced in the process.
5. Calculate simple descriptive statistics (mean, median and standard deviation) and visualize the measurements (for example with a box plot) to compare neutral and emotional speech. There is no need to calculate descriptive statistics or plot anything per individual speaker.
6. Include the descriptive statistics and plots in your LaTeX report (see the previous lab book for an example of '`includegraphics`').
7. For each measurement describe the differences you see between neutral state and emotional state measurements you see (if there are any).
8. Has your hypothesis about the measurements from step 2 been confirmed? If not, what other measurement (if any) was the best for the differentiation between recordings of neutral and emotional speech?

[a] Python hint:
$f0 = np.array(f0)$
$m = np.mean(f0)$
$v = np.mean((f0 - m) * *2))$

Answer

2.1.1 Your hypothesis, steps 1-2

In these datasets, we posit that **speech rate** serves as the most telling indicator of emotional shifts during moments of anger. Drawing from personal experiences and observations, when individuals find themselves in a state of anger or embroiled in a heated argument, their speech rate tends to exhibit a discernible uptick. For instance, sentences may flow out more rapidly, with phrases strung together in quick succession, and there is often a dearth of pauses or hesitations, as if the words rush forth unimpeded by self-regulation. This heightened speech rate can be likened to a verbal storm, reflecting the turbulence of emotions experienced during such moments of intense irritation or rage.

2.1.2 Measurements, descriptive statistics and plots, steps 3-6

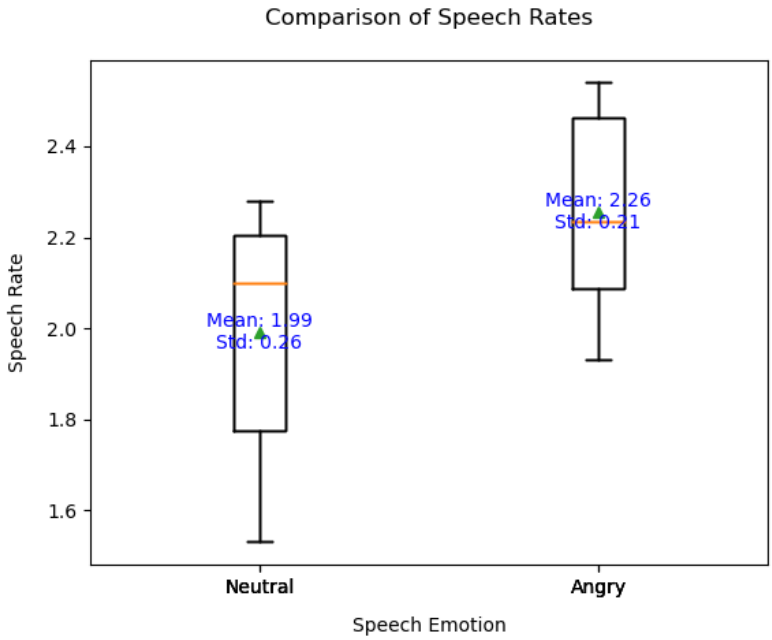

Figure 1: Comparison of Speech Rates for Neutral ($median = 2.1, mean = 1.99, sd = 0.26$) and Angry ($median = 2.2, mean = 2.26, sd = 0.21$)

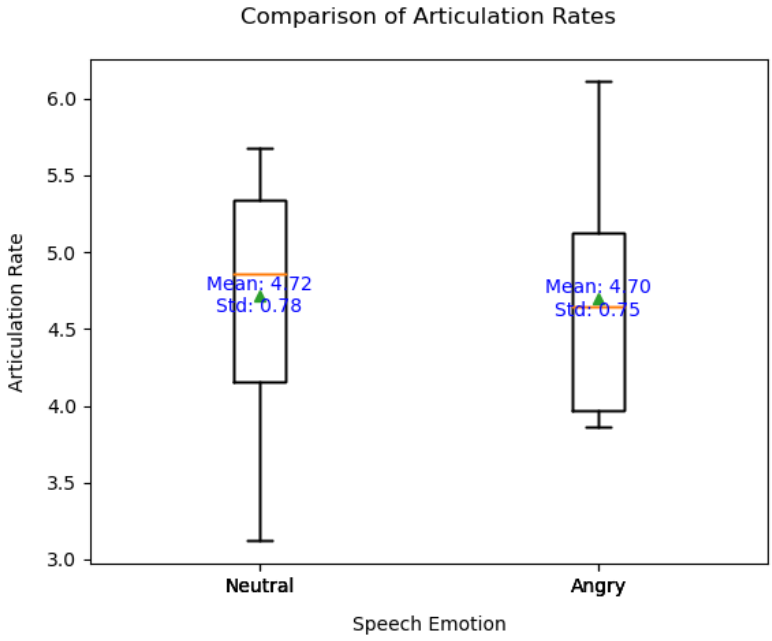

Figure 2: Comparison of Articulation Rates for Neutral ($median = 4.9, mean = 4.72, sd = 0.78$) and Angry ($median = 4.6, mean = 4.70, sd = 0.75$)

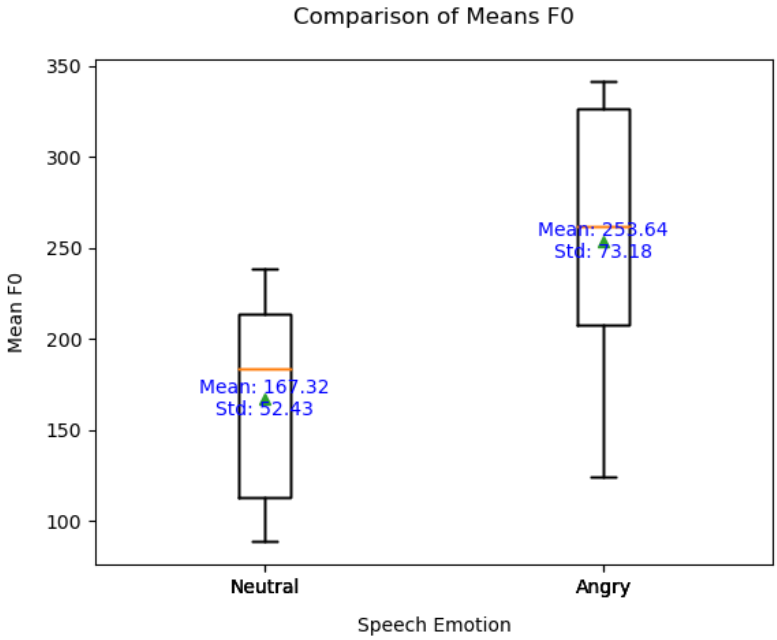

Figure 3: Comparison of Means F0 for Neutral ($median = 183.3, mean = 167.32, sd = 52.43$) and Angry ($median = 261.7, mean = 253.64, sd = 73.18$)

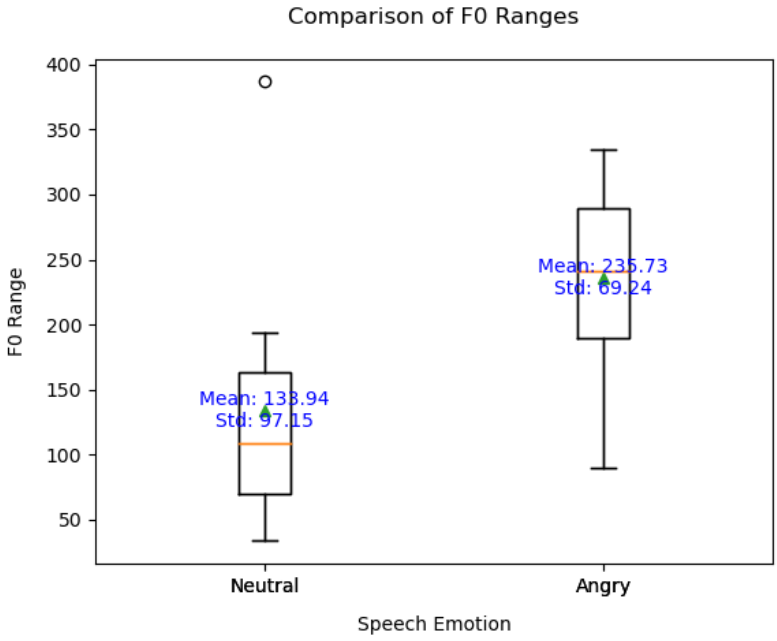

Figure 4: Comparison of f0 variance for Neutral ($median = 109.1, mean = 133.94, sd = 97.15$) and Angry ($median = 241.0, mean = 235.73, sd = 69.24$)

2.1.3 (Visual) analysis of the measurements and conclusions, steps 7-8

Analysis of Measurements:

Analyzing the box plots of **speech rate**, we observe that the descriptive statistics for the two emotions

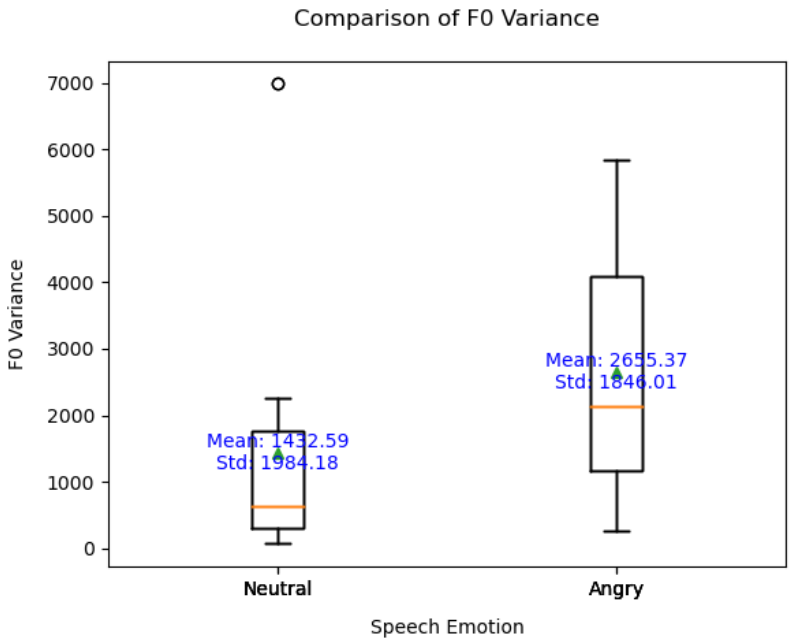

Figure 5: Comparison of f0 variance for Neutral ($median = 628.5, mean = 1432.59, sd = 1984.18$) and Angry ($median = 2136.2, mean = 2655.37, sd = 1846.01$)

are quite close. However, there are slight differences in the mean and median values, with anger exhibiting slightly higher rates. Furthermore, the standard deviation for the neutral condition is marginally greater, though the disparities are relatively minor. Notably, the anger condition also features a higher maximum value.

Turning our attention to the **articulation rate** plot, we note that the median, mean, and standard deviation values are in close proximity. Nevertheless, there are distinct disparities in the maximum and minimum values, especially with regards to the minimum value. The articulation rate for the angry condition is notably higher than that for the neutral condition.

Upon examining the **mean F0** plot, a substantial variance among the various indicators becomes apparent. The mean and median values for the angry condition are significantly elevated by approximately 100Hz compared to the neutral condition. In contrast, the standard deviation for the neutral condition undergoes a marked reduction in comparison to the angry state. While the minimum values for both conditions are nearly identical, the maximum value for the angry condition is 100Hz higher.

Analyzing the **F0 range** chart reveals a pattern consistent with the preceding mean F0 findings. Notably, the interquartile range, which signifies the frequency range between the upper and lower quartiles, appears to be narrower, indicating that the F0 range for each instance of anger is relatively concentrated. This phenomenon could be attributed to the inclusion of both male and female voices in the audio, as women typically exhibit higher F0 values and a broader vocal range. This factor contributes to the expansion of the maximum mean F0 value.

In the context of the **F0 variance** box plot, substantial disparities exist between the mean and median values for both the neutral and angry emotional states, with deviations exceeding 1000 Hz. It's worth noting that the standard deviation of the neutral state is greater than that of the angry state. Furthermore, the interquartile range for the angry state is notably broader compared to the neutral state. Even when considering their minimum values, the range for F0 variance in the angry state is nearly twice that of the neutral state.

Conclusions of Hypothesis:

å Our data clearly demonstrates a significant difference in the fundamental frequency range (F0 range) between the emotions of anger and neutrality. The fundamental frequency (F0) is a crucial acoustic parameter in sound wavefor

Measurements	Speech Rate		Articulation Rate		Mean F0		F0 Range		F0 Variance	
	Neutral	Angry	Neutral	Angry	Neutral	Angry	Neutral	Angry	Neutral	Angry
Aud. 1	1.50	2.10	4.00	4.10	112.30	258.20	60.00	180.30	245.30	4315.70
Aud. 2	2.30	2.00	5.60	3.90	115.00	157.50	70.70	218.20	2253.40	5200.60
Aud. 3	2.20	2.50	5.10	6.10	238.70	218.50	193.70	233.30	6982.80	2427.00
Aud. 4	1.70	2.50	4.10	5.60	176.60	123.90	386.80	89.60	2090.70	3443.40
Aud. 5	2.20	1.90	5.70	3.90	215.10	204.60	173.40	177.10	286.60	5837.30
Aud. 6	2.00	2.30	4.60	4.30	218.10	341.30	120.90	334.50	789.00	1110.70
Aud. 7	1.60	2.50	3.10	5.20	110.20	339.90	68.80	293.20	352.90	1845.30
Aud. 8	2.30	2.20	5.40	3.90	189.90	265.30	134.10	277.60	606.00	1340.00
Aud. 9	2.20	2.40	5.20	5.00	88.60	339.60	33.60	248.70	68.40	266.20
Aud. 10	1.90	2.20	4.40	5.00	208.60	287.70	97.40	304.80	650.90	767.50
Mean	2.00	2.30	4.70	4.70	167.30	253.60	133.90	235.70	1432.60	2655.40
Median	2.10	2.20	4.90	4.60	183.30	261.70	109.10	241.00	628.50	2136.20
SD	0.26	0.21	0.78	0.75	52.43	73.18	97.15	69.24	1984.18	1846.01

Table 1: Table with Subheaders

In the context of anger, speech typically becomes impulsive and less controlled. This impulsivity can lead to a reduction in the smoothness of speech, resulting in rapid and emotionally charged articulation. These factors collectively impact the fundamental frequency (F0), leading to frequent fluctuations in F0.

Anger is often accompanied by intense emotional expression. When individuals are angry, they tend to express themselves with heightened emotion, resulting in noticeable changes in pitch. Elevated pitch, characterized by high or rapid tonal variations, serves to emphasize the intensity of anger. Consequently, F0 tends to significantly increase to convey the emotional fervor.

Furthermore, increased muscular tension in the face and throat is a common occurrence during anger. This heightened tension affects the vibration of the vocal cords, causing a rise in pitch. The combination of high pitch and tense vocalization may lead to rapid fluctuations in the frequency of vocal cord vibrations, further increasing F0.

In summary, the observable variations in F0 during states of anger can be attributed to the demand for emotional expression, physiological changes in vocal cord tension, and alterations in speech fluency. The uncontrolled and rapid nature of speech during anger can result in frequent F0 fluctuations, collectively serving as a distinctive marker of this emotional state.

References:

De Jong, N.H. & Wempe, T. (2009). Praat script to detect syllable nuclei and measure speech rate automatically. Behavior research methods, 41 (2), 385 - 390.

De Jong, N.H., Pacilly, J., & Heeren, W. (2021). PRAAT scripts to measure speed fluency and breakdown fluency in speech automatically, Assessment in Education: Principles, Policy & Practice, 28:4, 456-476, DOI: 10.1080/0969594X.2021.1951162