HOMEWORK 2 REPORT

Nandan PC
nandanpc@usc.edu
3394 068 908

Sree Priyanka Uppu
uppu@usc.edu
6822 468 564

1. Data preparation:

The way we extracted the 3 acoustic and 3 visual features is: from homework 1 we were able to identify the features that have a good p value (typically<0.05) to be a suitable set of features for running the model.

In case of acoustic features, we considered all the following 6 features: Peak Slope, NAQ, fundamental frequency, Energy, energy slope and spectral stationary. The goal is to summarize these features from their .feat files to match the segments in sentiment.annotation file. For each of the 5 features, we summarized them based on the mean, median, min, max and standard deviation. Next, we found theier respective p values from ANOVA test. From that analysis, we selected the following features for the model:

Feature	**Summarization measure**	**ANOVA p value**
NAQ	Mean	0.000231912
Energy	Max	0.001164012
Energy Slope	Median	0.056881484

For visual features, we considered the following 5 features : Gaze Up Down, Gaze Left Right, Face Roll, Mouth Open, Smile Level. Mouth Open feature is extracted from .shore files and all other features are extracted from OKAO output. All the features are summarized to match their frames in sentiment.annotation file. For each of the 5 feature, we summarized mean, median, min, max and standard deviation. Next, we found their respective p values from ANOVA test. From that analysis, we selected the following features for the model:

Feature	**Summarization measure**	**ANOVA p value**
Gaze Up Down	Min	0.035985733
Face Roll	Max	0.154137689
Mouth Open	Mean	0.10035937

2. Validation strategies (~1 page): Create the following tables to analyze the effect of validation strategies:

a. Accuracies of linear SVM for 4 Fold

We tested the accuracies using the **normalized data** as input and found the respective below accuracies for the same linear SVM with 4 fold:

Normalized Data

K value	Training Accuracy	Validation Accuracy	Test Accuracy
1	0.515151515152	0.424657534247	0.462686567164
2	0.515151515152	0.462686567164	0.424657534247
3	0.470149253731	0.522388059701	0.408450704225
4	0.513888888889	0.582089552239	0.393442622951
Average	0.503585293	0.497955428	0.422309357

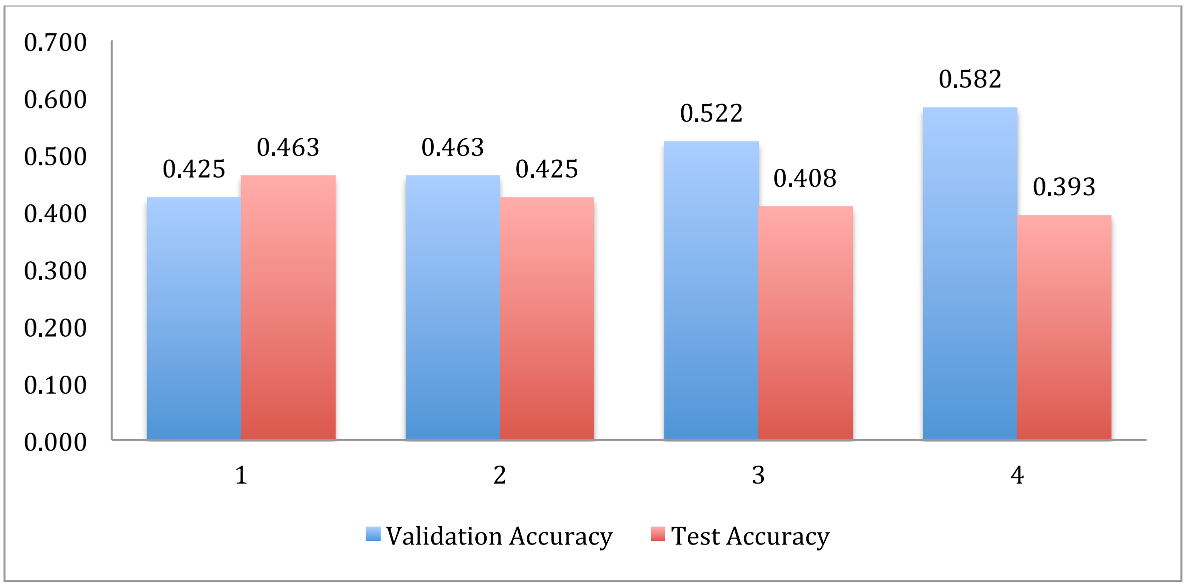


Fig 2a: Validation accuracy, test accuracy for each of the 4 folds in SVM Linear Kernel

b. Accuracies of linear SVM for 3 Fold

K value	Training Accuracy	Validation Accuracy	Test Accuracy
1	0.493975903614	0.377777777778	0.363636363636
2	0.493975903614	0.363636363636	0.377777777778
3	0.555555555556	0.474747474747	0.421686746988
Average	0.514502454	0.405387205	0.387700296

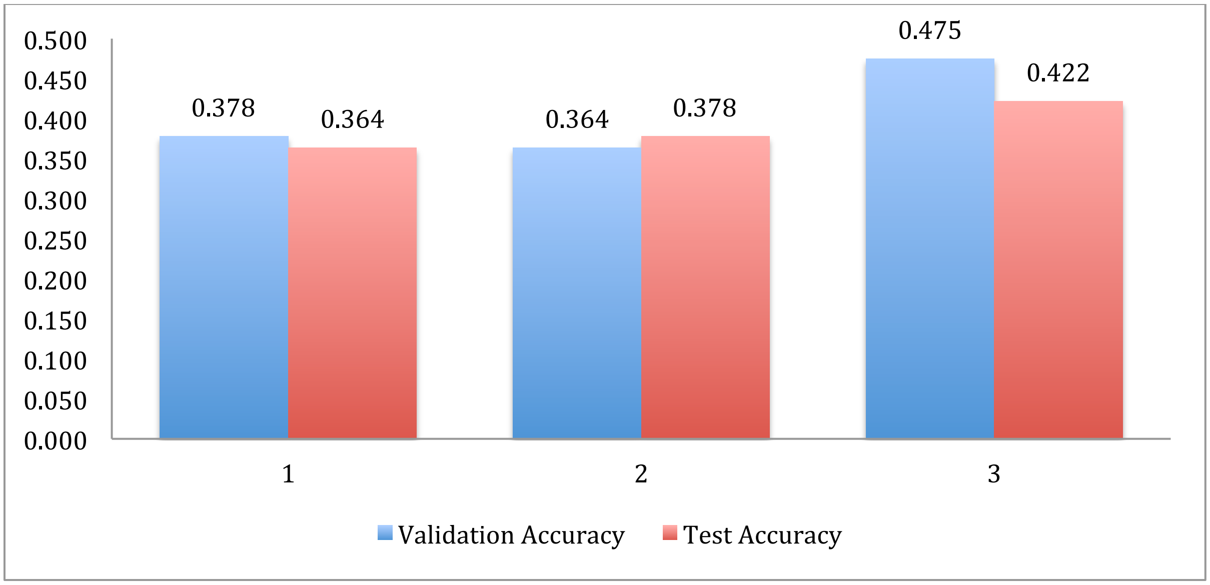


Fig 2b: Validation accuracy, test accuracy for each of the 3 folds in SVM Linear Kernel

Without respecting speaker independence:

K value	Training Accuracy	Validation Accuracy	Test Accuracy
1	0.47311827957	0.380434782609	0.393617021277
2	0.468085106383	0.391304347826	0.376344086022
3	0.531914893617	0.380434782609	0.41935483871

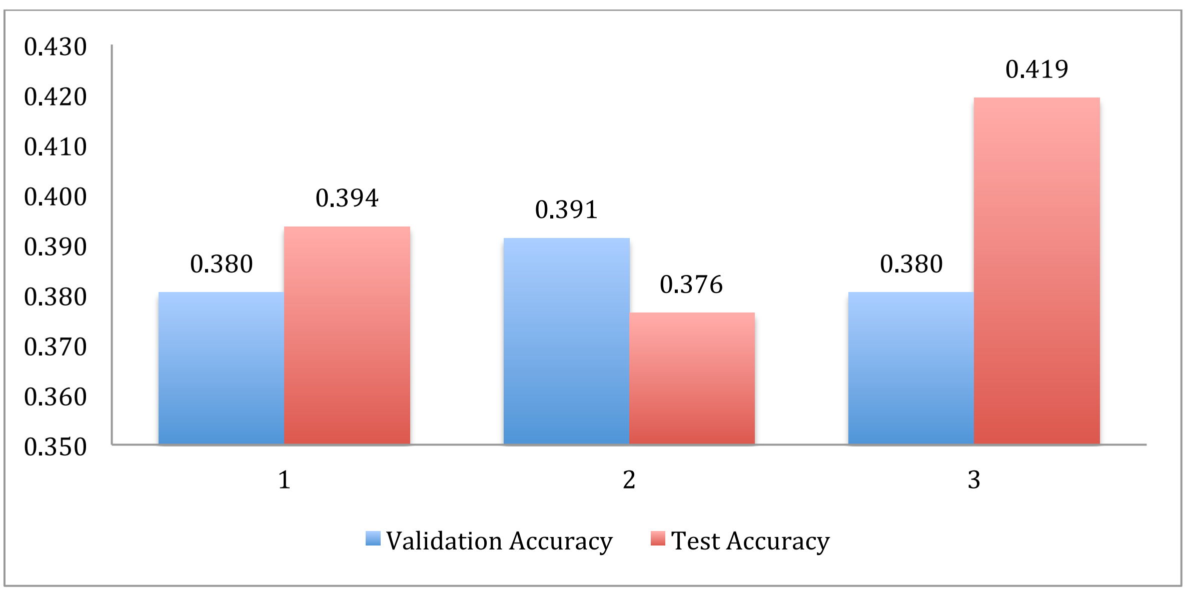


Fig 2c: Validation accuracy, test accuracy for each of the 3 folds in SVM Linear Kernel without respecting speaker independence

c. Do you expect the validation accuracies and test accuracies to be similar? Which validation strategy should give you better similarity between validation and test accuracies? Discuss the differences in these two tables.

No, we don’t expect the validation and test accuracy to be similar. Since, during the validation phase, we make the predictions by varying the hyper parameters for a classification algorithm and get the best hyper parameters to train the model and run it on test data set. Also, the model would have not encountered the test data before, so the accuracy will drop w.r.t the test accuracy when compared to validation accuracy.

In the current scenario, as we have small dataset we have better similarity between validation and test data for 3 fold validation. Lower ‘K’ is cheaper and more biased and larger ‘K’ is more expensive, less biased, but can suffer from large variability. Hence, choosing a good ‘K’ value is very important.

4 Fold gives a higher accuracy values in all the folds when compared to 3 fold strategy.

3. Different modalities (~0.5-1page):
a. Insert your graph comparing acoustic-only, visual-only and multimodal classifiers.

ACOUSTIC and VISUAL:

K value	Training Accuracy	Validation Accuracy	Test Accuracy
1	0.515151515152	0.424657534247	0.462686567164
2	0.515151515152	0.462686567164	0.424657534247
3	0.470149253731	0.522388059701	0.408450704225
4	0.513888888889	0.582089552239	0.393442622951
Average	0.503585293	0.497955428	0.422309357

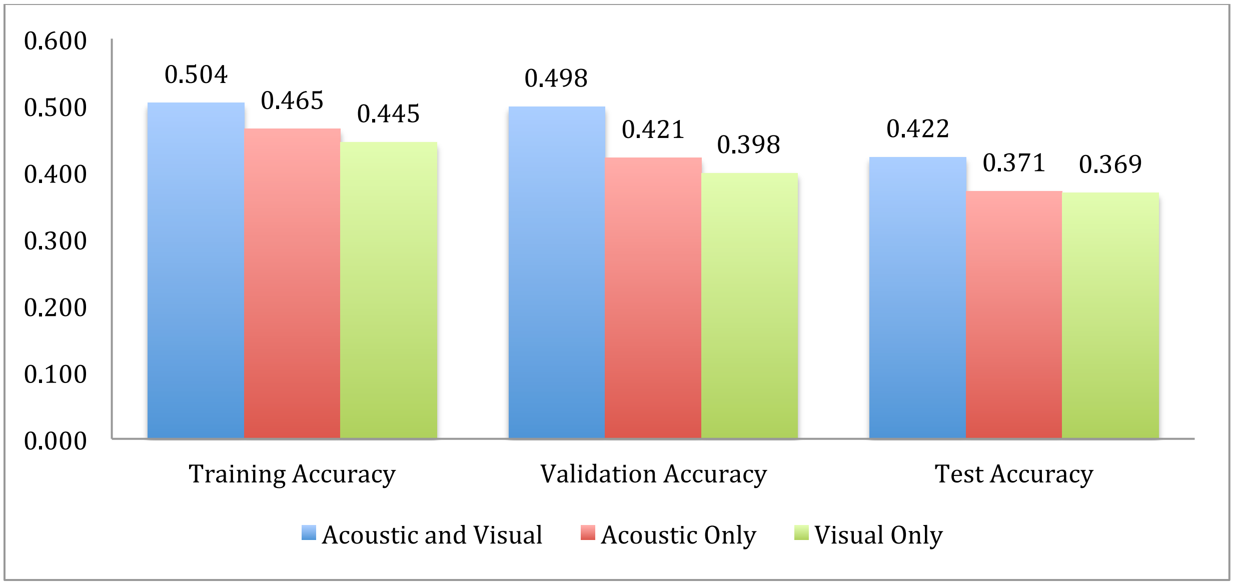
ACOUSTIC ONLY

K value	Training Accuracy	Validation Accuracy	Test Accuracy
1	0.469696969697	0.356164383562	0.388059701493
2	0.469696969697	0.388059701493	0.356164383562
3	0.432835820896	0.432835820896	0.380281690141
4	0.486111111111	0.507462686567	0.360655737705
Average	0.464585218	0.421130648	0.371290378

VISUAL ONLY:

K value	Training Accuracy	Validation Accuracy	Test Accuracy
1	0.477272727273	0.369863013699	0.388059701493
2	0.462121212121	0.417910447761	0.369863013699
3	0.388059701493	0.432835820896	0.323943661972
4	0.451388888889	0.373134328358	0.393442622951
Average	0.444710632	0.398435903	0.36882725

GRAPHS:



****X axis: Training Data, Validation Data and Test Data*
****Y axis: Average Accuracy for all 4 folds*

Fig 3a. Summarized accuracies vs Models(Acoustic and Visual, Acoustic only, Visual Only) across the Training, Validation and Test Data.

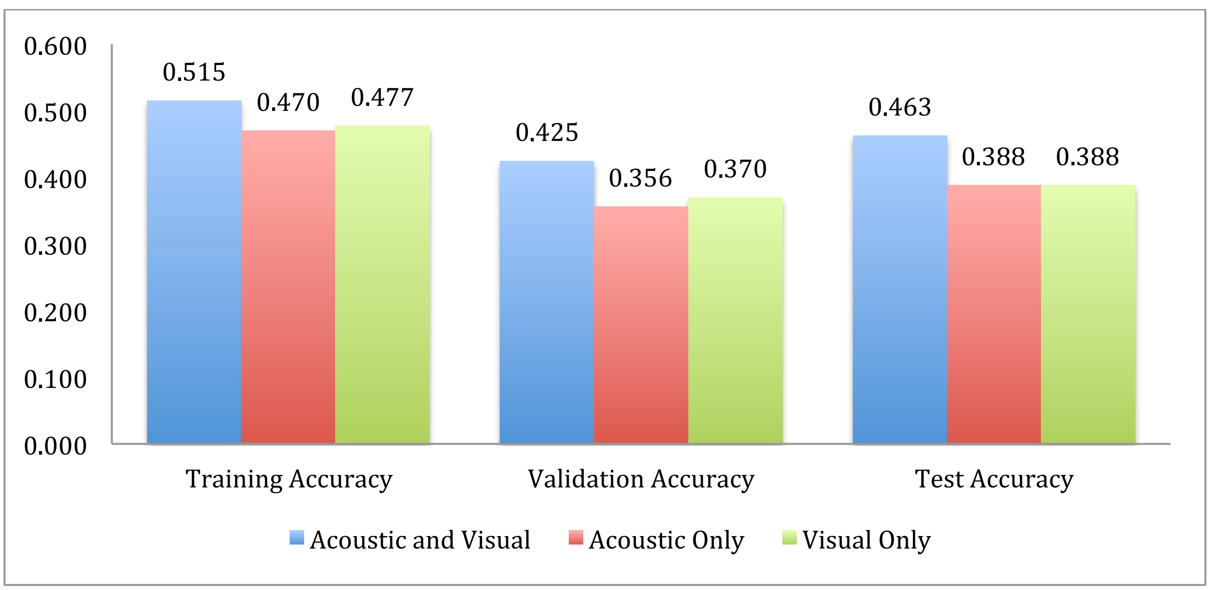


Fig 3b. Accuracies calculated when K=1 in 4 fold vs Models(Acoustic and Visual, Acoustic only, Visual Only) across the Training, Validation and Test Data.

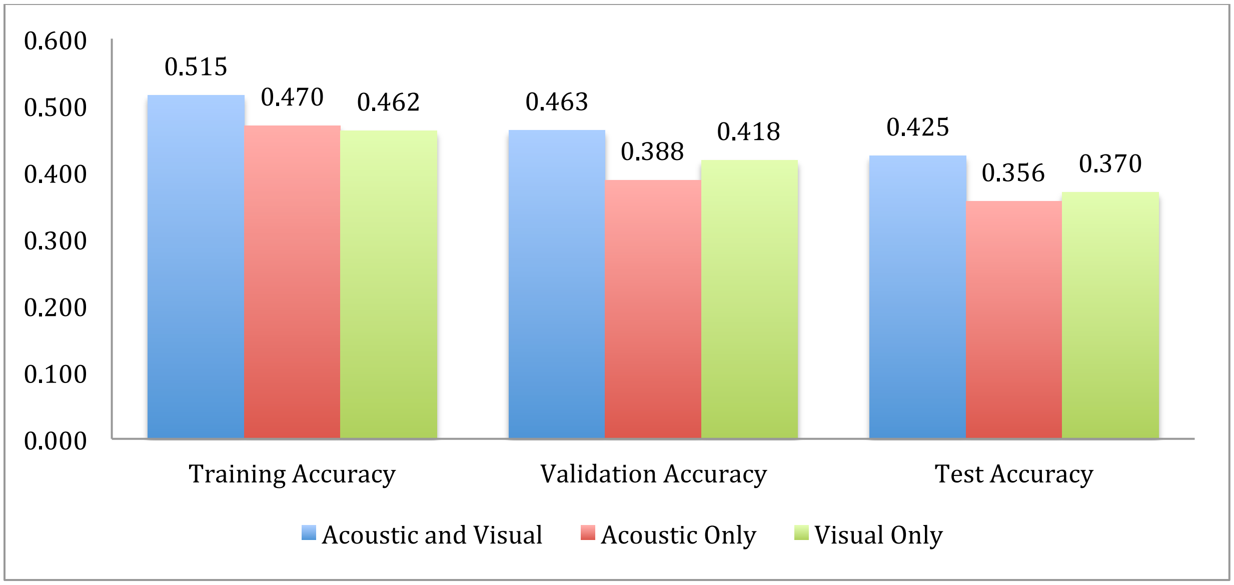


Fig 3c. Accuracies calculated when K=2 in 4 fold vs Models(Acoustic and Visual, Acoustic only, Visual Only) across the Training, Validation and Test Data.

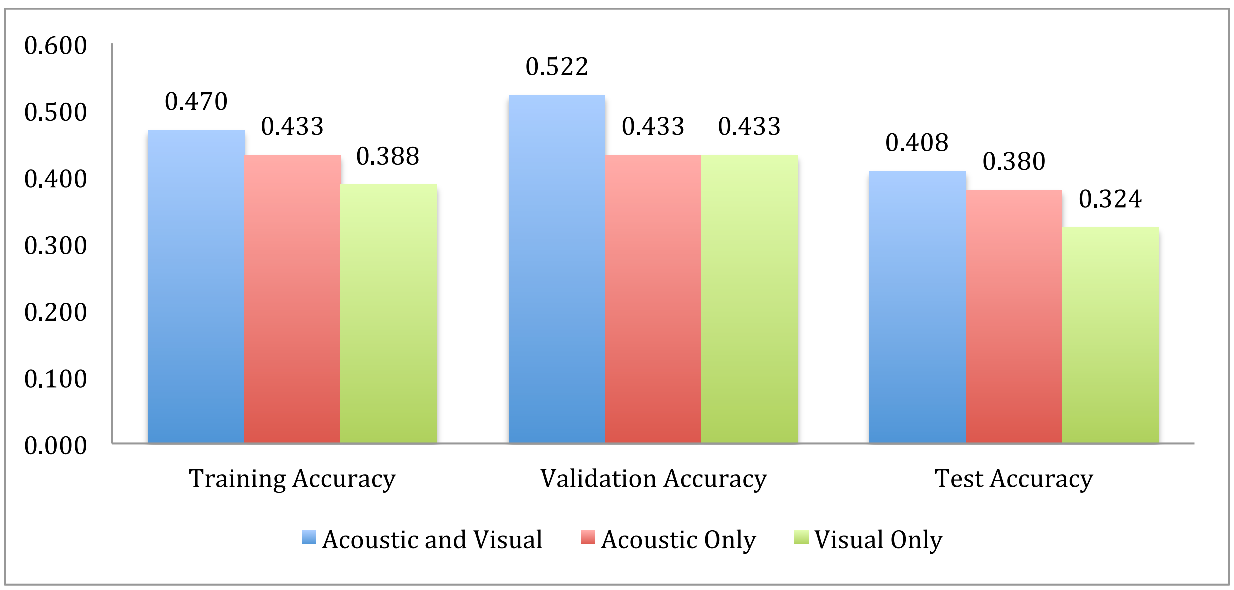
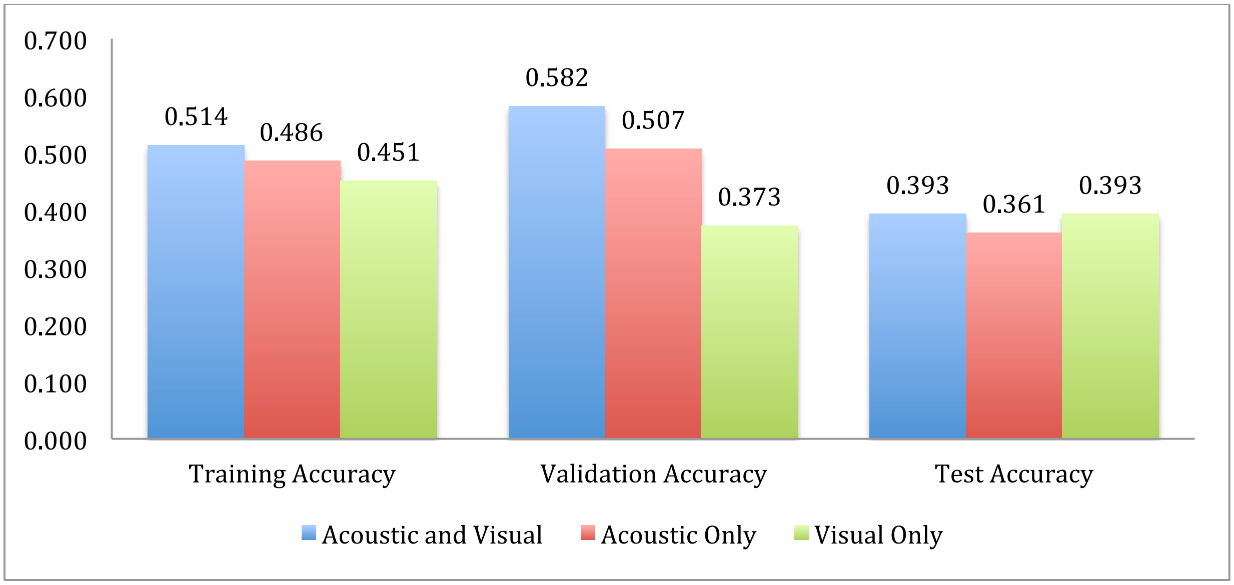


Fig 3d. Accuracies calculated when K=3 in 4 fold vs Models(Acoustic and Visual, Acoustic only, Visual Only) across the Training, Validation and Test Data.



<u>Fig 3e. Accuracies calculated when K=4 in 4 fold vs Models(Acoustic and Visual, Acoustic only, Visual Only) across the Training, Validation and Test Data.</u>

b. If some of these pairs are statistically significant (based on the two-sample t-test), include these differences in the graph (using stars *).

K	T Test values p value from Acoustic and Visual Data, Acoustic only data	T Test values p value from Acoustic and Visual Data, Visual only data
1	0.01850135	0.00000755
2	0.00000188	0.00161039
3	0.22321405	0.00000000
4	0.50565419	0.00010714
Average p Value	0.186842866	0.000431269

We observe that there is a significant co-relation between the Multimodal data results and Visual only data results.

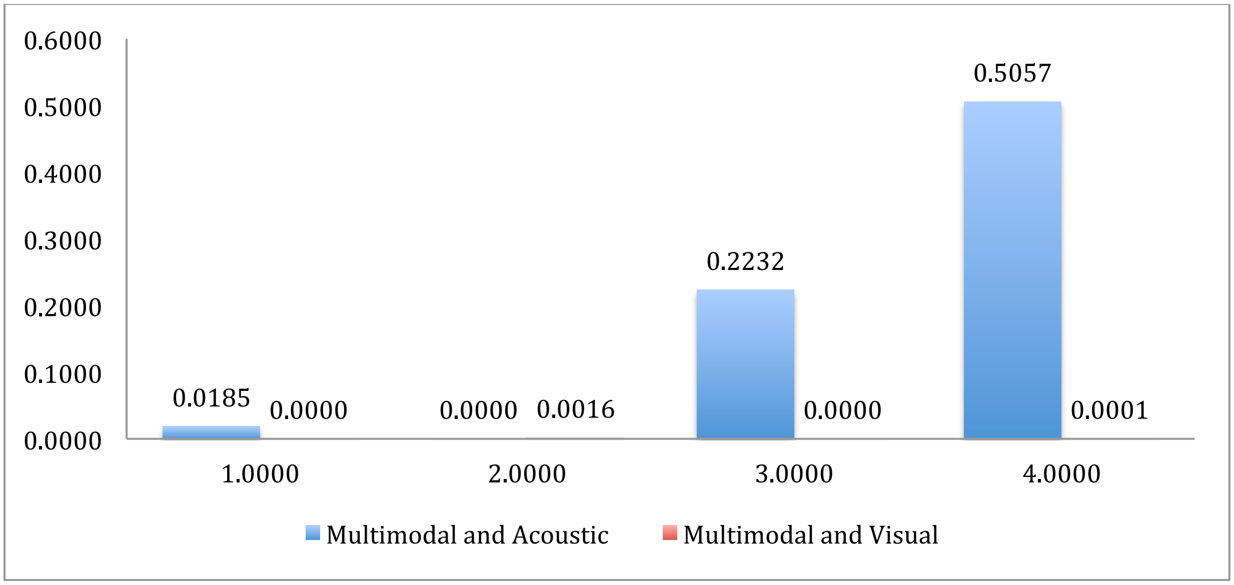


Fig3f: Comparison of p Values for each fold between Multimodal & Acoustic ONLY data and Multimodal & Visual ONLY data

c. **Discuss any differences between modalities. If possible, make some references to the qualitative observations you made during the homework**

Through the series of experiments, we observed that Multimodal Data(using both visual and acoustic) gave us the best results followed by the Visual only data gave us better results when compared to the Acoustic only data. In the course we learnt that multimodal gives us the best results when compared to single modalities and this has been proved through this experiment.

4. **Multiple Classifiers (~0.5-1page):**
   a. **Insert a graph comparing the three classifiers you trained and evaluated. Include any statistically significant differences in your graph.**

SVM Linear Kernel:

K value	Training Accuracy	Validation Accuracy	Test Accuracy
1	0.515151515152	0.424657534247	0.462686567164
2	0.515151515152	0.462686567164	0.424657534247
3	0.470149253731	0.522388059701	0.408450704225
4	0.513888888889	0.582089552239	0.393442622951
Average	0.503585293	0.497955428	0.422309357

SVM RBF Kernel:

K value	Training Accuracy	Validation Accuracy	Test Accuracy
1	1.0	0.328767123287671	0.358208955223
2	0.401515151515151	0.358208955223880	0.328767123287
3	1.0	0.298507462686567	0.295774647887
4	1.0	0.358208955223880	0.295081967213
Average	0.850378787879	0.335923124105	0.319458173403

Normalized Data SVM RBF Kernel:

K value	Training Accuracy	Validation Accuracy	Test Accuracy
1	1.0	0.3561643835616	0.35820895522388
2	0.9621212121212	0.3582089552238	0.39726027397260
3	1.0	0.3283582089552	0.25352112676056
4	0.9444444444444	0.3731343283582	0.44262295081967
Average	0.976641414141	0.353966469025	0.362903326694

Neural Network:

K value	Training Accuracy	Test Accuracy
1	0.536585365854	0.44776119403
2	0.577889447236	0.438356164384
3	0.557213930348	0.366197183099
4	0.587677725118	0.508196721311
Average	**0.544042857**	**0.467857143**

Naïve Bayes Classifier:

K value	Training Accuracy	Test Accuracy
1	0.49756097561	0.492537313433
2	0.542713567839	0.369863013699
3	0.517412935323	0.535211267606
4	0.530805687204	0.409836065574
Average	**0.519047619**	**0.432142857**

Graph:

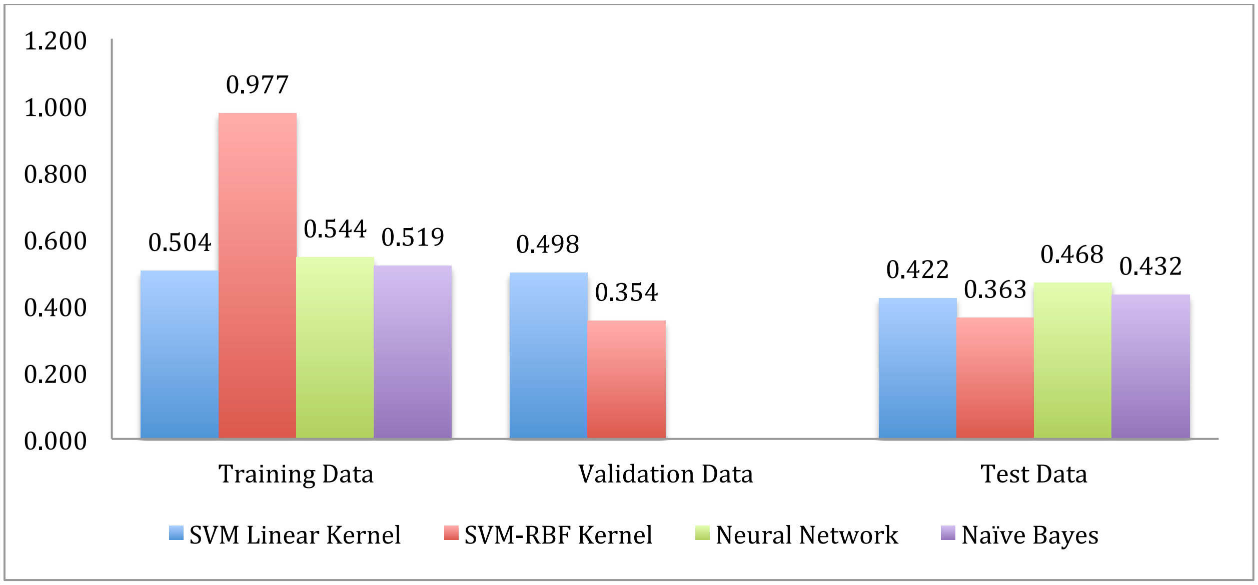


Fig 4a: Accuracies VS Comparison among different algorithms: SVM- Linear Kernel, SVM-RBF Kernel(Normalized Data) , Neural Network and Naïve Bayes across the Training , Validation and Test Data

b. Find which classifier performs the best and write one paragraph describing the Methodology you used to train, validate and test this classifier. You should include enough details so that a reader can recreate your experiment.

For the experiments we have used following classifiers SVM with linear and rbf kernel, Neural Networks and Naïve Bayes. The methodology we followed for splitting the data for SVM classifiers is as follows. For four fold validation data is split into 4 parts. 50% Train(2 parts), 25% Validation(1 Part) and 25% Test(1 Part) and we considered speaker independence. Also, the data has been normalized for building the models.

We train the model using train data and for getting the hyper parameters(C for Linear SVM kernel and Gamma for RBF kernel) we tested it on the validation data. We ran experiment in validation data for hyper parameter C – [10E-2, 10E-1, 10E0, 10E1, 10E2] for linear kernel and gamma- [1e-1, 1, 1e1], C – [10E-2, 10E-1, 10E0, 10E1, 10E2] RBF kernel. The best parameter, which gave highest accuracy, was selected and was used for the test data in each fold. For 3 fold validation data split into 3 parts. One part is held out as a test data, one part of training and one part validation. We repeated the same process for of running SVM classifiers with linear and RBF kernels. The accuracy for SVM with Linear kernel = **0.422309357** and for RBF kernel = **0.319458173403** and RBF kernel with normalized data = **0.362903326694**.

For Neural Networks, multilevel perceptron model(MLP) was used. Four fold validation technique was used to train the model. Data is split into 4 folds where 1 fold of data is held out for testing data. Other 3 folds where used as training data. Neural network was trained with the following parameters : solver= 'lbfgs', alpha=1e-5, hidden_layer_sizes=(3,) and number of random states=5. For each hidden layer sizes(3,5,7,9,11,13,15,17,19,21) parameter the model was tested with test data and we found that hidden_layer_size=3 gave us the best result. Best accuracy from each fold was used to get the average accuracy of "**0.467857143".**

For Naïve Bayes we have followed fourfold validation technique. 1 fold held out as test data and other 3 as training data. Accuracy for each fold was obtained and averaged to get final accuracy of "**0.432142857"**

c. What would make one of the classifier better in this setup? What would you expect if you had more data? Which classifier would you use in this case?

On an average, the Neural Nets are giving us better accuracies as proven above. But, if we had a larger dataset we feel Neural Networks would have given us even better performance and a higher accuracy. As learnt, more data implies more training and which in turn gives better predictions. If we have large amount of data we can use techniques like Deep Learning, Naïve Bayes, etc.