1. Audibility
   1. Data Preprocessing

2.1.1 Data Structure

The whole data includes 6 columns, and the total observations are 672. The following is a code book for the dataset.

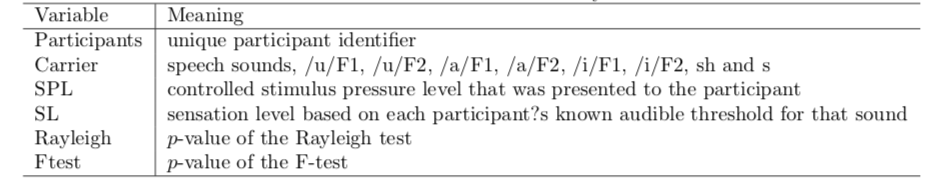


Table Code Book for the Audibility Dataset

2.1.2 Getting Categorical Data into Shape

Since “Carrier”, “SPL”, “SL” are all binary features, so I do a transformation to get them into 0-1 form so that the step of regression can be much more easily. (details in appendix)

* 1. Experiments on Accuracy of EFRs on Predicting Audibility of Speech Stimulus

2.2.1 Relationship between Accuracy of EFRs on Predicting Audibility of Speech Stimulus and Carriers or Frequency Groups

To explore the accuracy of EFRs on predicting audibility of each speech stimulus differ between carriers or frequency groups, we need to do some regression on the data.

We mainly consider the detectability (EFRs), which is a binary outcome, so I choose to use logistic regression since it fits the condition that response value is a binary one.

2.2.2.1 Carriers

2.2.2.1.1 Main Effect Analysis

(1) Model

We consider the following one factor regression models:



(2) Arguments and Parameters

|  |  |
| --- | --- |
| EFRs | representing the detectability of the stimulus |
| Carrier | speech sounds, which is a categorical data with 8 levels |
| Beta0 | accuracy of EFRs (detectability of the stimulus) for reference level |
| Beta(j) | difference in accuracy of EFRs (detectability of the stimulus) between j level of carrier and reference level of carrier |
| Beta0+ Beta(j) | accuracy of EFRs (detectability of the stimulus) at j level of carrier |

Table Arguments and Parameters for the Model

(3) Analysis

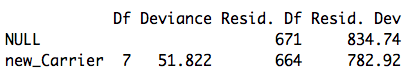


Table Anova Table for One Factor Model

Null model has a deviance of 834.74 on 671 degrees of freedom, the p-value is nearly asymptotic to 0, which doesn’t pass the goodness-of-test, so we reject the null hypothesis that all detectability is the same.

Introducing Carrier variable leads to substantial reduction of 127.29 deviance at only 7 degrees of freedom. So the variable of Carrier has significant effect on the detectability of the stimulus and accuracy of EFRs.

2.2.2.1.2 Effect of Eight Levels of Carriers on Accuracy of EFRs

In order to explore the effect of 8 levels in carriers on accuracy of EFRs, we need to fit the following model both the result of F-test and Rayleigh-test.



(1) F-test

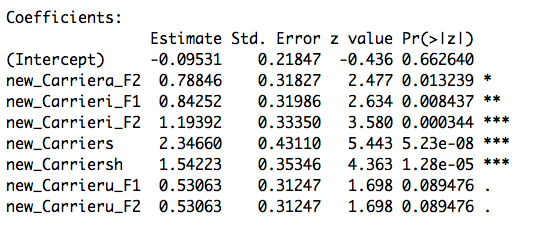


Table Model Fitting for EFRs~Carriers on F-test



Based on the result of model fitting for EFRs~Carriers on F-test above, we conclude the p-values for each 8 level in carriers so that we can judge the significance of each level in the model.

I find that all of the levels are significant at 90% confidence level. And “i\_F2”, “s”, “sh” is the three most significant effects in the model.

|  |  |  |
| --- | --- | --- |
| level of carriers | Logit(pi) | pi |
| a\_F1 | -0.09531 | 0.4761905 |
| u\_F1 | 0.43532 | 0.6071433 |
| i\_F1 | 0.74721 | 0.6785705 |
| a\_F2 | 0.69315 | 0.6666673 |
| u\_F2 | 0.43532 | 0.6071433 |
| i\_F2 | 1.09861 | 0.7499996 |
| s | 2.25129 | 0.9047617 |
| sh | 1.44692 | 0.809524 |
| Average | / | 0.6875 |

Table Values of pi of Carriers on F-test

Based on the values of pi on F-test above, we can conclude that different levels of carriers will lead to a difference in accuracy in predicting the audibility of stimulus. The level “s” will have the highest accuracy and “a\_F1” will have lowest accuracy. And the average accuracy of all levels on F-test is 0.6875.

(2) Rayleigh-test

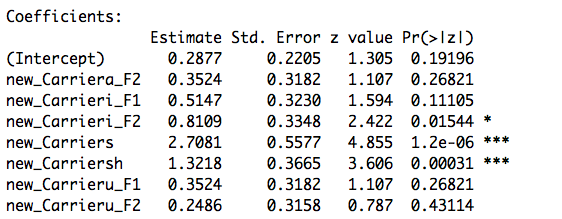


Table Model Fitting for EFRs~Carriers on Rayleigh-test

Based on the result of model fitting for EFRs~Carriers on Rayleigh-test above, we conclude p-values for 3 out of 8 levels are significant in the model, and they are: “i\_F2”, “s”, “sh”.

|  |  |  |
| --- | --- | --- |
| level of carriers | Logit(pi) | pi |
| a\_F1 | 0.2877 | 0.571433 |
| u\_F1 | 0.6400 | 0.6547535 |
| i\_F1 | 0.8023 | 0.6904663 |
| a\_F2 | 0.6400 | 0.6547535 |
| u\_F2 | 0.5363 | 0.6309513 |
| i\_F2 | 1.0986 | 0.7499977 |
| s | 2.9957 | 0.9523795 |
| sh | 1.6094 | 0.8333281 |
| Average | / | 0.71726 |

Table Values of pi of Carriers on Rayleigh-test

Based on the values of pi on Rayleigh-test above, we can conclude that different levels of frequency will also lead to a difference in accuracy in predicting the audibility of stimulus. The level “s” has the highest accuracy and “a\_F1” will have lowest accuracy. And the average accuracy of all levels on Rayleigh-test is 0.71726.

(3) Comparison

Compared with the result on the F-test, we can find that “i\_F2”, “s”, “sh” these 3 levels always have very small p-values based on the both two testing methods: F-test and Rayleigh-test, which means they are always the most significant effects in the model. Also, based on the Rayleigh-test, the average accuracy is higher than that on F-test.

From my perspective, the small difference between of two tests may caused by the refined classification on the carrier features.

2.2.2.2 Frequency Groups

(1) Model

In order to explore the difference of accuracy between frequency groups, we need to do a transformation for the dataset. Since the data has three levels of frequency: low, mid and high, we need to classify the Carrier feature into these three categories. Then, we can fit the model as follows again.



(2) Arguments and Parameters

|  |  |
| --- | --- |
| EFRs | representing the detectability of the stimulus |
| Frequency | frequency of sounds, which is a categorical data with 3 levels |
| Beta0 | accuracy of EFRs (detectability of the stimulus) for reference level |
| Beta(j) | difference in accuracy of EFRs (detectability of the stimulus) between j level of frequency and reference level of frequency |
| Beta0+ Beta(j) | accuracy of EFRs (detectability of the stimulus) at j level of frquency |

Table Arguments and Parameters for the Model

(3) Analysis

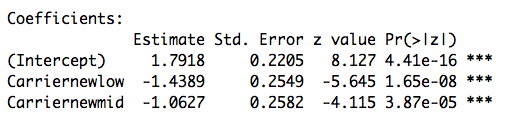


Table Model Fitting for EFRs~frequency on F-test

Based on the results of model fitting for EFRs~ frequency on F-test above, we conclude that these three levels all have significant effect on the accuracy of EFRs. The high frequency will lead to the highest accuracy of EFRs. The low frequency will lead to the lowest accuracy of EFRs.

|  |  |  |
| --- | --- | --- |
| level of frequency | Logit(pi) | pi |
| low | 0.3528 | 0.5872964 |
| mid | 0.7291 | 0.6746077 |
| high | 1.7918 | 0.8571478 |
| average | / | 0.70635 |

Table Values of pi of Frequency on F-test

Based on the table above, we can make sure that higher the frequency, the higher the accuracy. And the average accuracy is 0.70635.

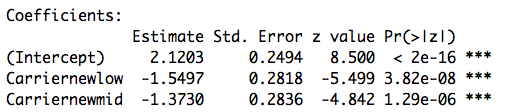


Table Model Fitting for EFRs~frequency on Rayleigh-test

Based on the results of model fitting for EFRs~ frequency on Rayleigh-test above, we also conclude that these three levels all have significant effect on the accuracy of EFRs based on the small p-values. The higher the frequency is, the higher the accuracy of EFRs will be.

|  |  |  |
| --- | --- | --- |
| level of frequency | Logit(pi) | pi |
| low | 0.5705 | 0.6388785 |
| mid | 0.7472 | 0.6785683 |
| high | 2.1203 | 0.8928606 |
| average | / | 0.7368 |

Table Values of pi of Frequency on Rayleigh-test

Based on the table above, we can make sure that higher the frequency, the higher the accuracy. And the average accuracy is 0.7368.

(3) Comparison

Comparing two testing methods, the results are almost the same, and it may be caused by the robustness of classification of frequency. But the average of accuracy on Rayleigh-test is still a little bit higher than that on F-test.

2.2.2 Performance between the F-test and the Rayleigh in Predicting Audibility

To make a general analysis of the accuracy performance of EFRs of predicting the audibility of stimulus between the F-test and the Rayleigh, we can use the table about types of error and statistical power respectively on F-test and Rayleigh.

2.2.2.1 F-test in Predicting Audibility

|  |  |  |
| --- | --- | --- |
| predicted  reality | True(Audible) | FALSE(inaudible) |
| TRUE(audible) | 0.6339286 | 0.3660714 |
| Flase(inaudible) | 0.04464286 | 0.9553571 |

Table F-test Accuracy

2.2.2.2 Rayleigh in Predicting Audibility

|  |  |  |
| --- | --- | --- |
| predicted  reality | True(Audible) | FALSE(inaudible) |
| TRUE(audible) | 0.6785714 | 0.3214286 |
| Flase(inaudible) | 0.08928571 | 0.9107143 |

Table Rayleigh-test Accuracy

2.2.2.3 Combination of F-test and Rayleigh in Predicting Audibility

|  |  |  |
| --- | --- | --- |
| predicted  reality | True(Audible) | FALSE(inaudible) |
| TRUE(audible) | 0.6125 | 0.3875 |
| Flase(inaudible) | 0.02678571 | 0.9732143 |

Table Combination of F-test and Rayleigh-test Accuracy

2.2.2.4 Conclusions

Making a comparison of the two tables, we draw the following conclusions:

1. Rayleigh-test has higher accuracy of predicting audibility on the audible stimulus.
2. F-test has higher accuracy of predicting audibility on the inaudible stimulus.
3. If we combine the results of F-test and Rayleigh-test, then the accuracy of prediction on both audible and inaudible stimulus will be higher than based on one test.
4. The combination of two test will make the accuracy of predicting audibility on the inaudible stimulus higher, but accuracy of predicting audibility on the audible stimulus lower.
5. Once we have a higher accuracy of predicting audibility on the inaudible stimulus, accuracy of predicting audibility on the audible stimulus will become lower. They are negative related.
   1. Minimum SL for Detectability

2.3.1 Global Minimum of SL

To find the minimum of SL needed in order for the EFR to detect a response, we should first fit the following two models respectively on F-test and Rayleigh-test:





1. F-test

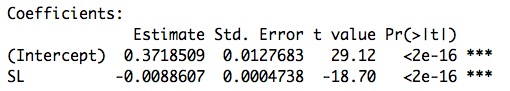


Table Global Minimum of SL on F-test

1. Rayleigh-test

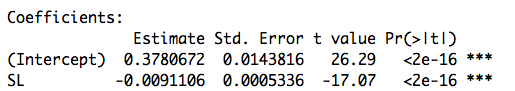


Table Global Minimum of SL on Rayleigh-test

1. Conclusions

|  |  |
| --- | --- |
| method | Minimum of SL |
| f-test | 37.02498 |
| Rayleigh-test | 36.0094 |

Table Comparison between F-test and Rayleigh-test

We can see that minimum of SL on Rayleigh-test is a bit smaller than that on F-test.

2.3.2 Relationship between Minimum of SL and Carrier or Frequency Groups

To find the relationship between minimum of SL needed in order for the EFR to detect a response and Carrier, we should fit the following two models respectively on F-test and Rayleigh-test:





* + - 1. Carrier

1. F-test

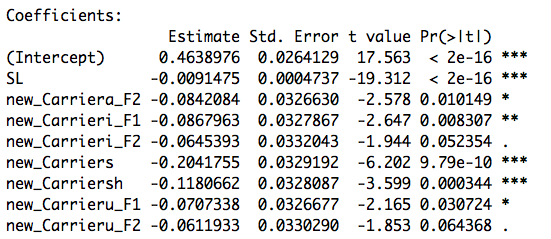


Table Relationship between Minimum of SL and Carrier on F-test

1. Rayleigh-test

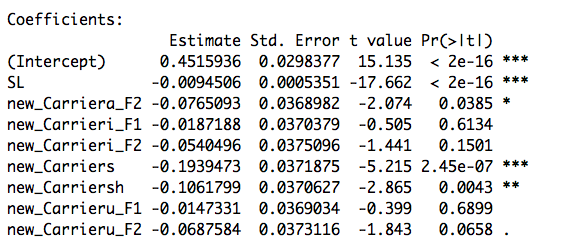


Table Relationship between Minimum of SL and Carrier on Rayleigh-test

1. Conclusions

Based on the above results of two tests, we can draw a table to find the minimum of SL based on different levels of Carrier:

|  |  |  |
| --- | --- | --- |
| Method | Carrier | Minimum of SL |
| F-test | a\_F1 | 45.24707 |
|  | a\_F2 | 36.04145 |
|  | i\_F1 | 35.75854 |
|  | i\_F2 | 38.19167 |
|  | u\_F1 | 37.51449 |
|  | u\_F2 | 38.55745 |
|  | s | 22.92671 |
|  | sh | 32.34014 |
| average of f-test |  | **35.8222** |
| Rayleigh-test | a\_F1 | 42.49398 |
|  | a\_F2 | 34.39828 |
|  | i\_F1 | 40.51329 |
|  | i\_F2 | 36.77481 |
|  | u\_F1 | 40.93503 |
|  | u\_F2 | 35.21843 |
|  | s | 21.97176 |
|  | sh | 31.25872 |
| average of Rayleigh-test |  | **35.4455** |

Table Comparison of Carrier between F-test and Rayleigh-test

Based on the above table, I use three different colors to differ three levels of carrier: low, mid and high frequencies. The F1 carriers are low frequency dominant, the F2 carriers are mid frequency dominant and the fricatives (sh and s) are high frequency dominant.

In general, we can find that with the higher carrier, and will have the smallest minimum of SL. The lower the carrier is, the larger the minimum of SL.

Also, we can conclude that among F1 and F2 level, there are three different forms: “a”, “i”, and “u”. We find that for F1 level, these three kinds have a distinct difference in minimum of SL than that in F2 level.

Finally, the average minimum of SL on Rayleigh-test is still a little bit smaller than that on F-test.

* + - 1. Frequency Groups

To find the relationship between minimum of SL needed in order for EFR to detect a response and frequency groups, we should fit the following two models respectively on F-test and Rayleigh-test:





1. F-test

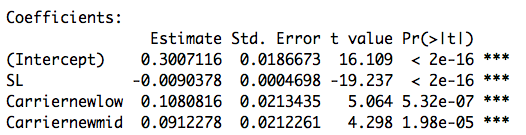


Table Relationship between Minimum of SL and Frequency Groups on F-test

1. Rayleigh-test

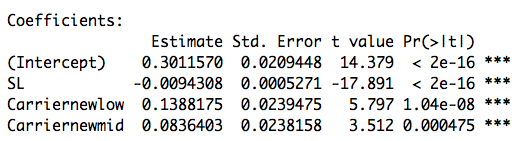


Table Relationship between Minimum of SL and Frequency Groups on Rayleigh-test

1. Conclusions

Based on the above results of two tests, we can draw a table to find the minimum of SL based on different levels of frequency groups:

|  |  |  |
| --- | --- | --- |
| Method | Frequency Groups | Minimum of SL |
| F-test | low | 39.69917 |
|  | mid | 37.83436 |
|  | high | 27.74034 |
| Rayleigh-test | low | 41.35116 |
|  | mid | 35.50042 |
|  | high | 26.63157 |

Table Comparison of Frequency Groups between F-test and Rayleigh-test

Based on the result in the table above, we find that based on the both of two testing, lowest level of frequency will always lead to the largest minimum of SL since in that way the criterion of finding the minimum of SL is stricter than other two.

* 1. Limitation and Improvement

2.4.1 Limitation

For the model in 2.2 as the following form, we should check the regression assumption that error terms are independent with each other and if the model is adequate.



2.4.2 Remedies

We fit additive models in the following forms

1. Additive Model



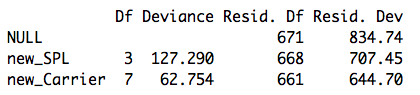


Table Deviance of Two Factor Model

The additive model has a deviance of 190.04 at only 10 degrees of freedom. So the model provides a good description of the data.

1. Two Factor Model with Interaction



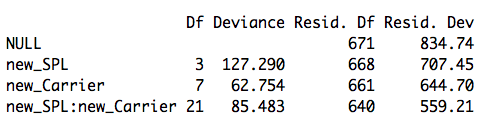


Table Deviance of Two Factor Model with Interaction

The two factor model with interaction has a deviance of 275.53 at only 31 degrees of freedom. So the interaction between SPL and carrier has a significant effect on the model and thus the model provides a good description of the data.

|  |  |
| --- | --- |
| Beta0 | intercept for detectability of the stimulus |
| Beta(i) | difference in detectability of the stimulus between i level of SPL and reference level of SPL |
| Alpha(j) | difference in detectability of the stimulus between j level of carrier and reference level carrier |
| Gamma(I,j) | difference in detectability of the stimulus between the interaction of j level of carrier and i level of SPL and the reference level of interaction |

Table Parameters for the Additive Model

2.4.3 Conclusions

|  |  |  |
| --- | --- | --- |
| Model | Deviance | d.f. |
| Null | 834.74 | 671 |
| *One Factor Model* |  |  |
| SPL | 707.45 | 668 |
| Carrier | 782.92 | 664 |
| *Two Factor Model* |  |  |
| SPL+Carrier | 644.7 | 661 |
| SPL+Carrier+ SPL\*Carrier | 559.21 | 640 |

Table Deviance for the Whole Model

Our conclusion is that the model with two factors and interaction is the most accurate and adequate model.