

Project report Sta444–Fitting ARIMA model to Beijing Opera Data

Xiaoyang Liu

Due 5/3/2018

Dataset description. <http://www.isophonics.net/SingingVoiceDataset>

Research question.

Do human singing voices, in particular Beijing Opera, have a simple ARIMA model? If so, is there any difference between voices of different singers and between singers from two genders? How is the difference reflected in the model's fitted parameters?

Methods.

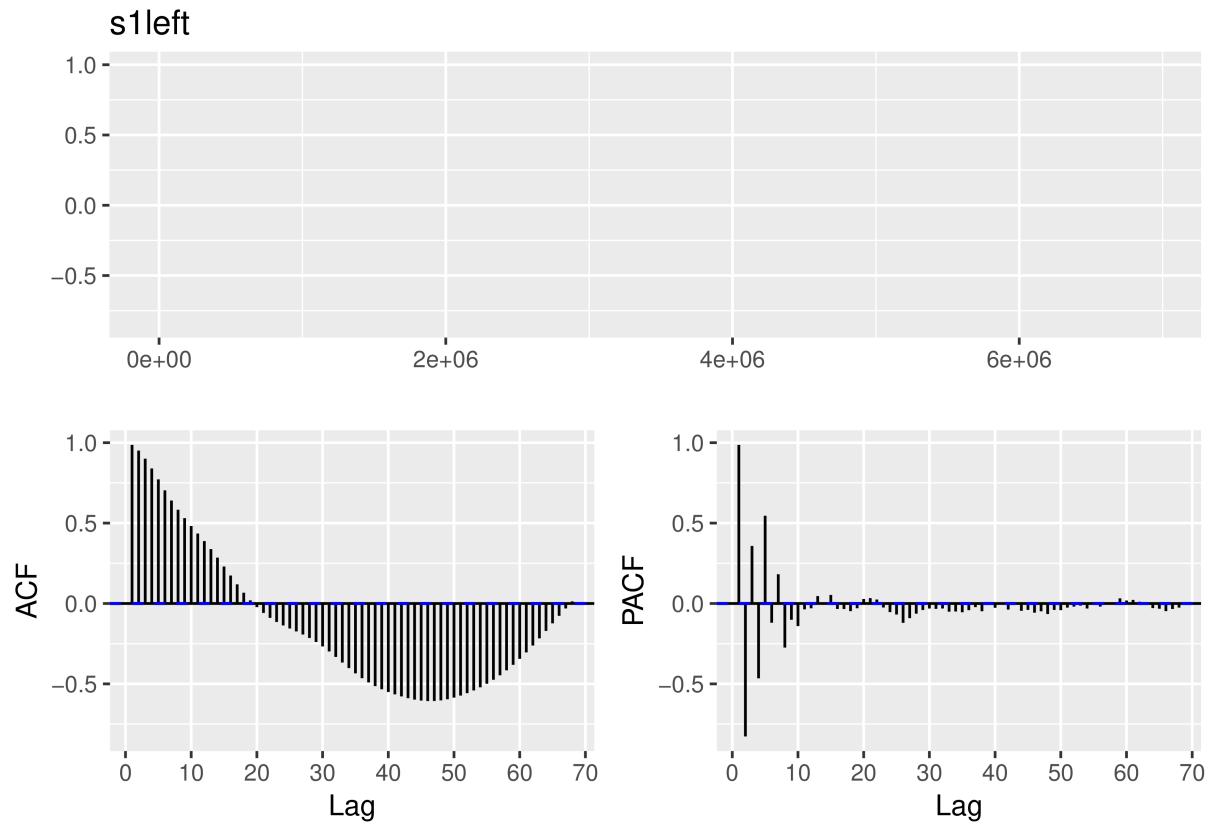
A total of 6 song segments are selected for analysis. They come from 4 different singers, 2 females and 2 males. Song segment #1 and #2 are from the same song by one female singer. A variety of ARIMA models are devised for each song segment and one best model is selected depending on a balance of its complexity and AIC, BIC values. Then the model and its parameters are compared between different segments in consideration of the singer's characteristic such as gender. It is controlled that all songs have positive emotion labelled and they are all Beijing opera. Segments are all 5 second long. They are selected to be as continuous as possible (no breaks in the audio due to singer breathing in etc). The songs are recorded at 44.1kHz sampling rate in .wav format. Only the best fitted models are included in this brief report, the supplementary models mentioned in Discussion and the full fitting procedure can be found in another file, 'Sta 444 Project full fitting procedure and supplementary models_xl149.pdf'.

Results.

Song segment #1 Singer. Female 1. Zhang Shuo Song type. Beijing Opera Song name. bei4 jiu1 chan2 Emotion. Positive.

```
song1 <- readWave('female1_pos_1.wav')
song1left <- song1@left
s1left <- song1left / 2^(song1@bit -1)

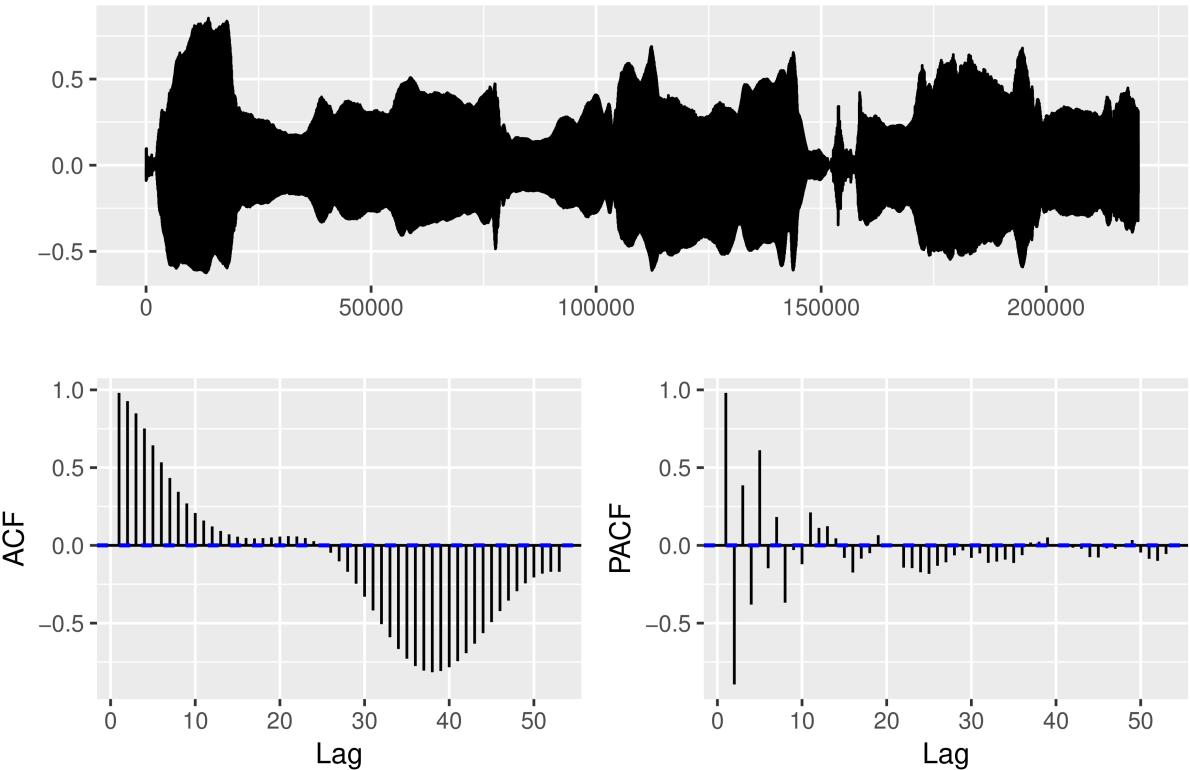
#full duration
forecast::ggttsdisplay(s1left, points=FALSE)
```



An echo pattern is observed in ACF. Don't see periodic pattern. Because the song duration varies from one song to another and in general 1-2 mins long of recording has too many data points. I will select the beginning 5 s duration of every song for analysis.

```
#5s duration at 44.1kHz sampling rate
forecast::ggtsdisplay(s1left[20000:240500], points=FALSE)
```

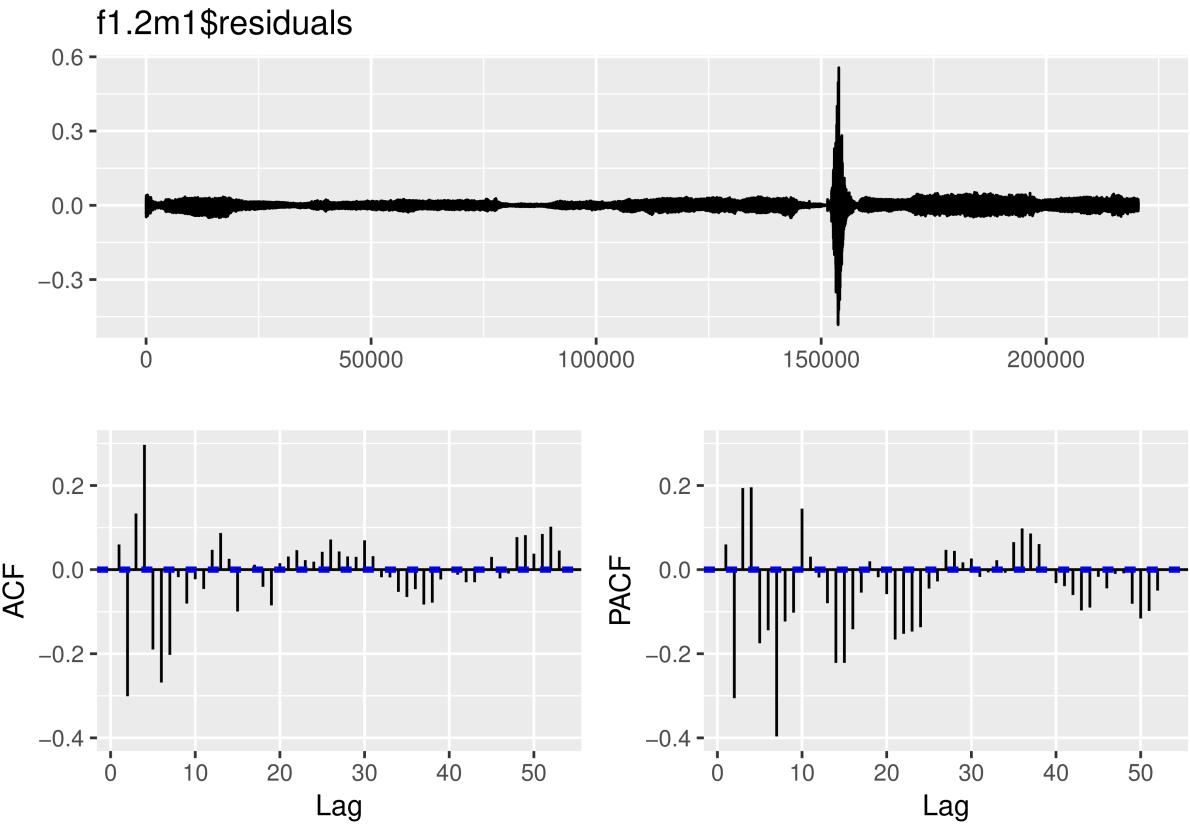
s1left[20000:240500]



After going through the fitting procedure (fully described in full analysis and omitted here), we arrive at a $(0,2,1)$ model describing song from 0.45 s to 5.45 s.

```
f1.2m1=forecast::Arima(s1left[20000:240500],order = c(0,2,1));
f1.2m1
```

```
## Series: s1left[20000:240500]
## ARIMA(0,2,1)
##
## Coefficients:
##         ma1
##     0.9502
## s.e. 0.0006
##
## sigma^2 estimated as 0.0002168: log likelihood=617264.8
## AIC=-1234526  AICc=-1234526  BIC=-1234505
forecast::ggtsdisplay(f1.2m1$residuals, points=FALSE)
```



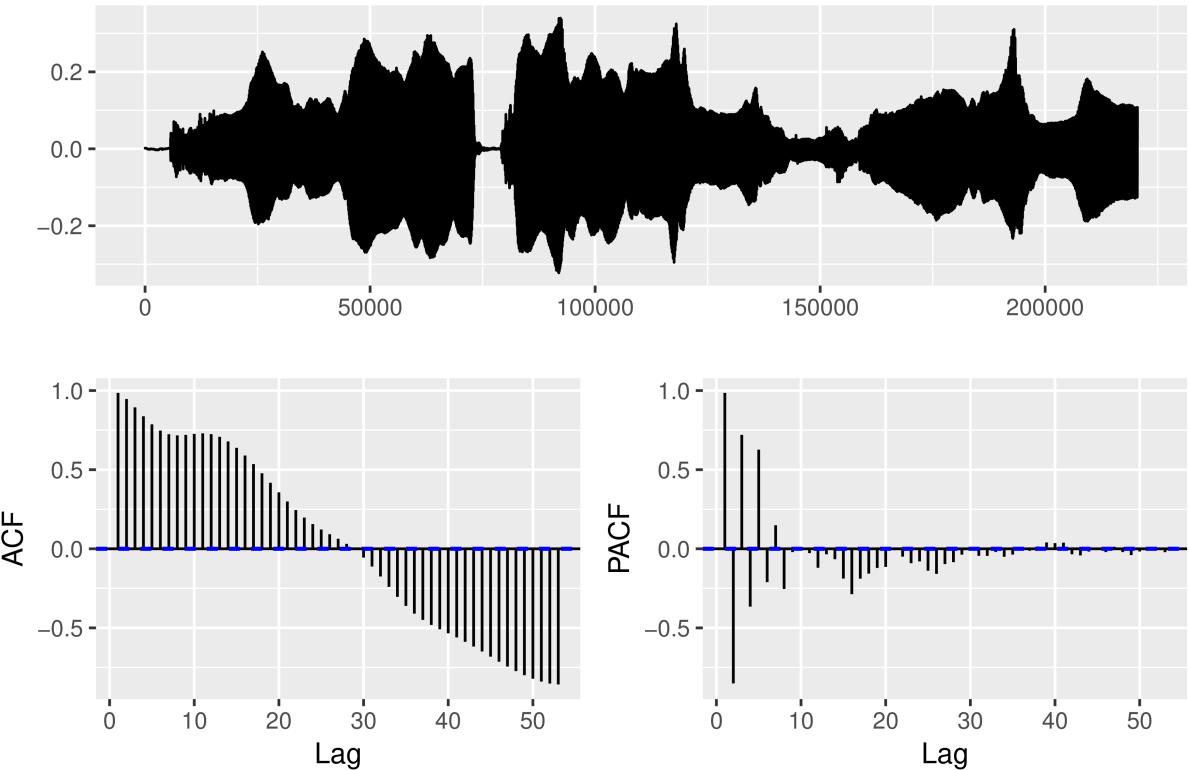
Brief summary. Its a (0,2,1) ARMA model.

Song segment #2 Singer. Female 1. Zhang Shuo Song type. Beijing Opera Song name. bei4 jiu1 chan2 Emotion. Postive.

This segment#2 (from 43.5 to 48.5 s)is from the same song as segment#1.

```
#5s duration at 44.1kHz sampling rate
forecast::ggttsdisplay(s1left[1920000:2140500], points=FALSE)
```

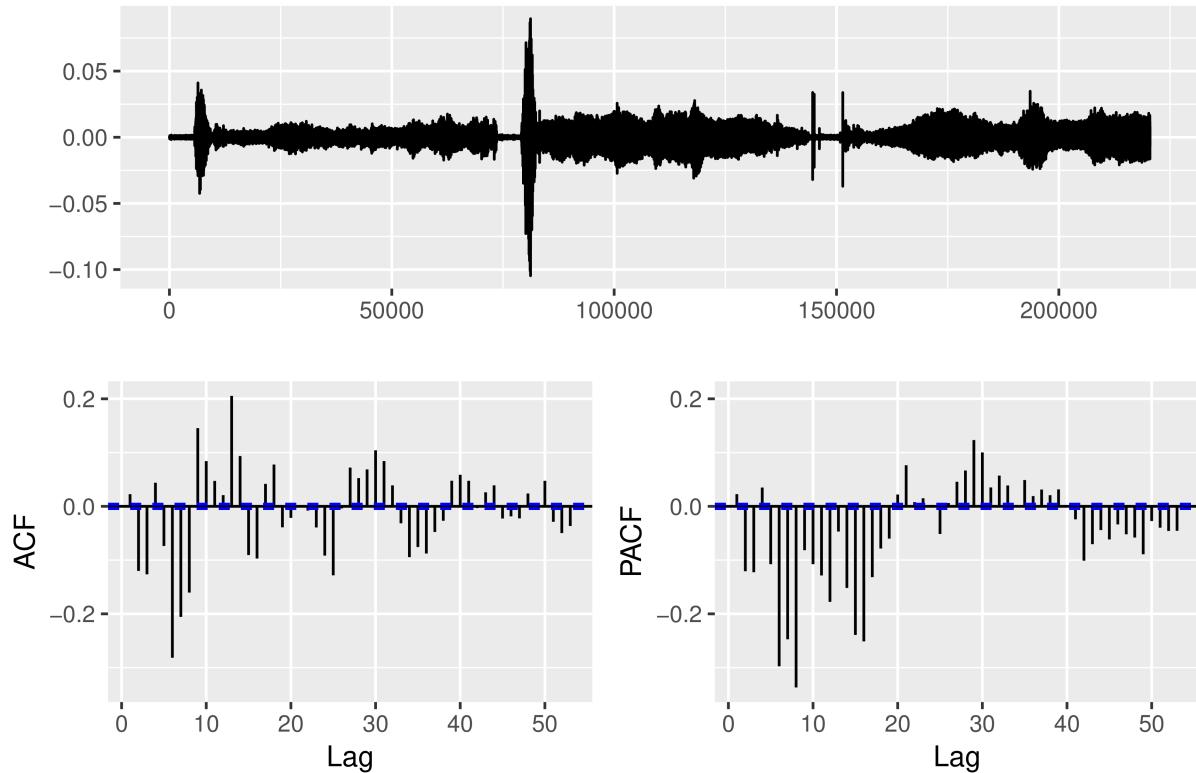
s1left[1920000:2140500]



```
f1s2.2m2=forecast::Arima(s1left[1920000:2140500],order = c(0,2,2));  
f1s2.2m2
```

```
## Series: s1left[1920000:2140500]  
## ARIMA(0,2,2)  
##  
## Coefficients:  
##          ma1     ma2  
##        1.4998  0.6350  
##  s.e.  0.0014  0.0016  
##  
## sigma^2 estimated as 1.978e-05:  log likelihood=881200.4  
## AIC=-1762395  AICc=-1762395  BIC=-1762364  
forecast::ggtsdisplay(f1s2.2m2$residuals, points=FALSE)
```

f1s2.2m2\$residuals



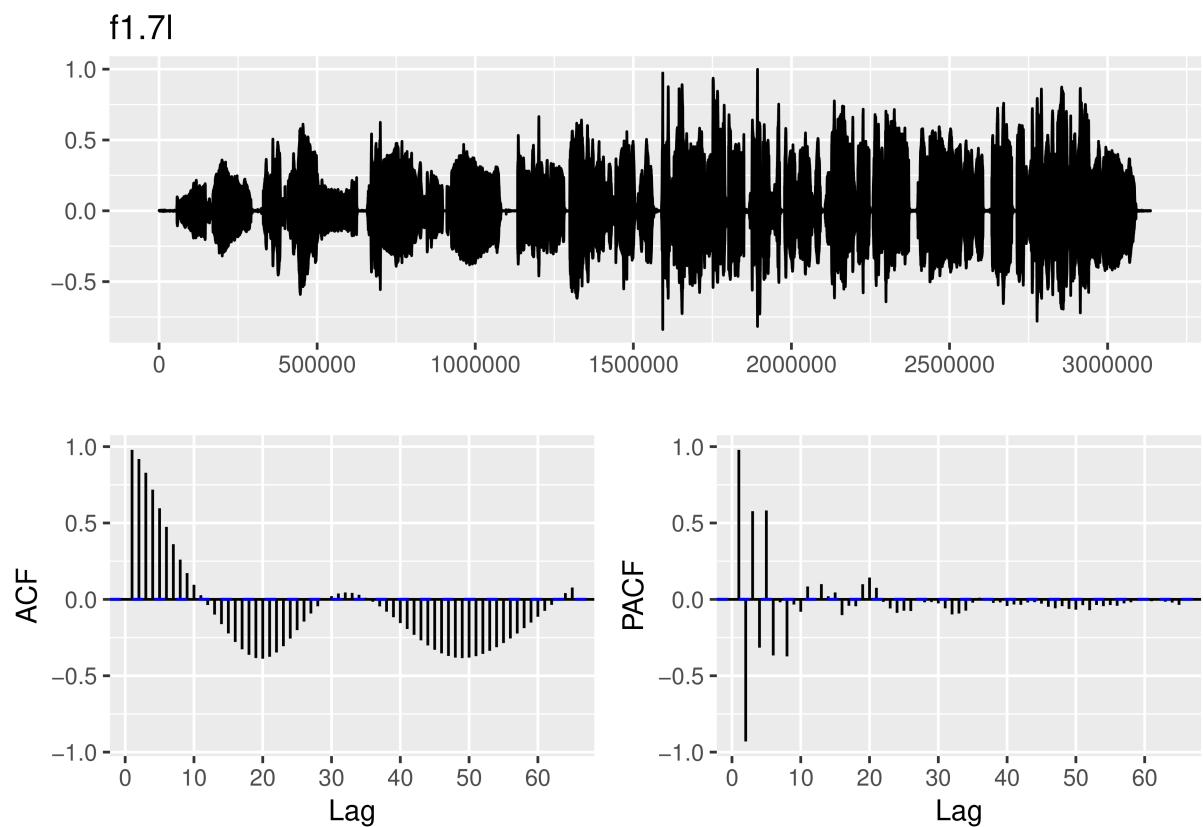
After going through the analysis fully described in project_full_analysis.pdf. Its concluded that (0,2,2) model has the best AIC and BIC.

Song segment #3 Singer. Female 1. Zhang Shuo Song type. Beijing Opera Song name. zhe4 cai2 shi4 Emotion. Postive.

Song segment #3 is sang by the same singer as segment#1 and #2. But its from a different beijing opera from #1 and #2.

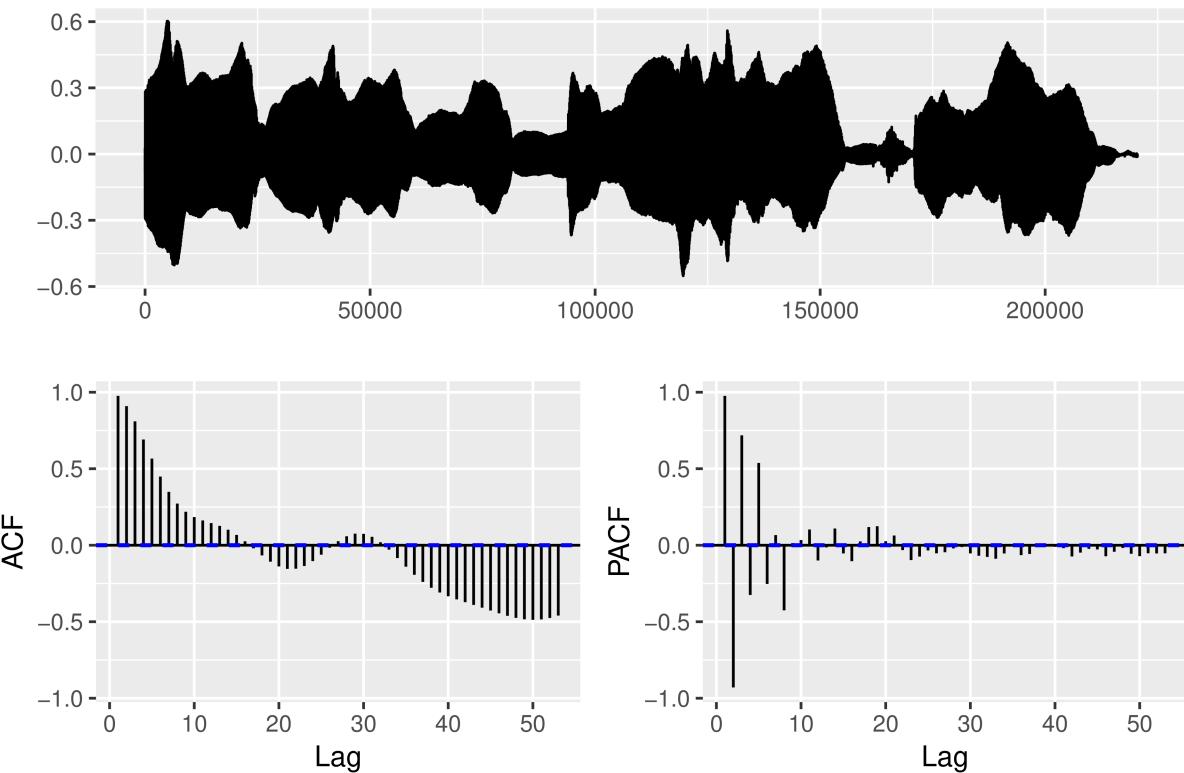
```
f1.7 <- readWave('female1_pos_7.wav')
f1.71 <- f1.7@left
f1.71 <- f1.71 / 2^(f1.7@bit -1) #converting to doubles

#full duration
forecast::ggttsdisplay(f1.71, points=FALSE)
```



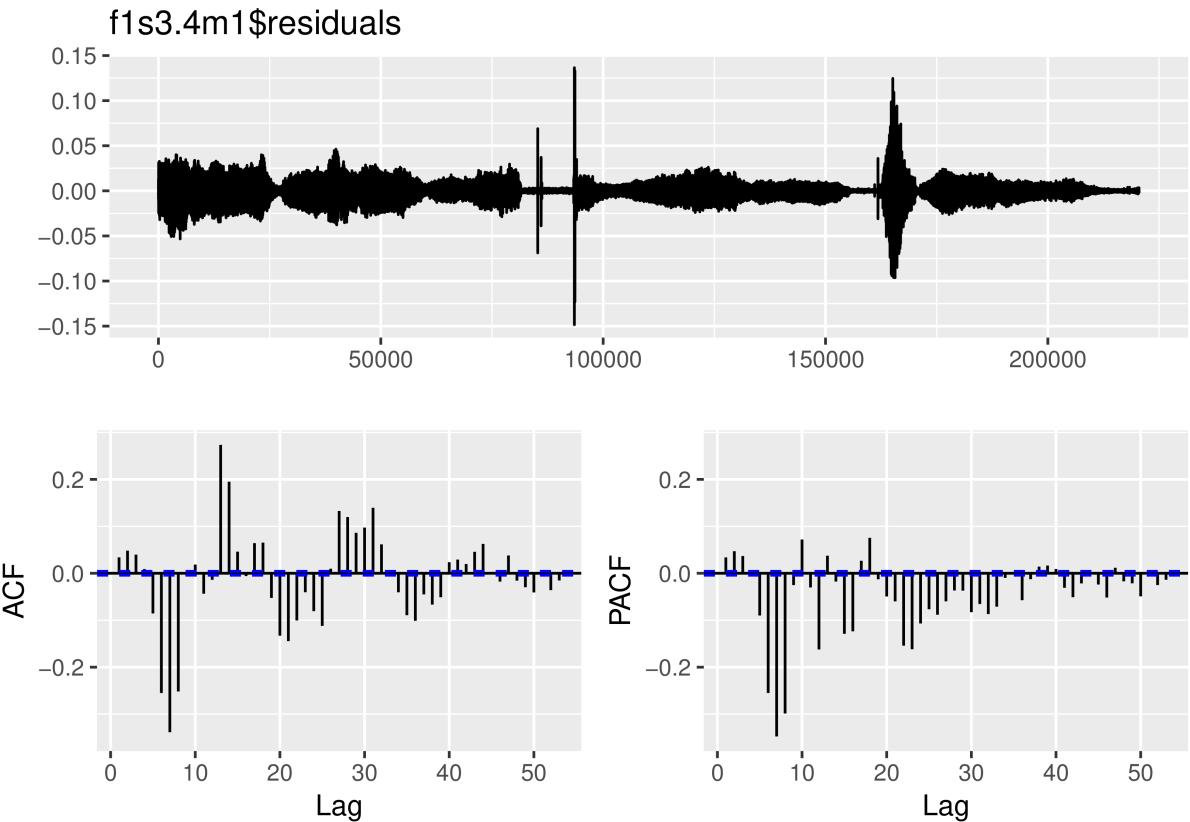
```
#5 s  
forecast::ggttsdisplay(f1.71[1350000:1570500], points=FALSE)
```

f1.7l[1350000:1570500]



```
f1s3.4m1=forecast::Arima(f1.7l[1350000:1570500],order = c(0,2,2));
f1s3.4m1
```

```
## Series: f1.7l[1350000:1570500]
## ARIMA(0,2,2)
##
## Coefficients:
##          ma1     ma2
##        1.4889  0.6158
## s.e.  0.0016  0.0017
##
## sigma^2 estimated as 4.233e-05:  log likelihood=797333.8
## AIC=-1594662  AICc=-1594662  BIC=-1594631
forecast::ggtsdisplay(f1s3.4m1$residuals, points=FALSE)
```



Select (0,3,1) which has the least number of parameters and comparable AIC BIC as (0,3,2) and (0,3,3) when d=3. But its found when d=2, (0,2,2) has better AIC and BIC than (0,3,1). So the final model is (0,2,2)

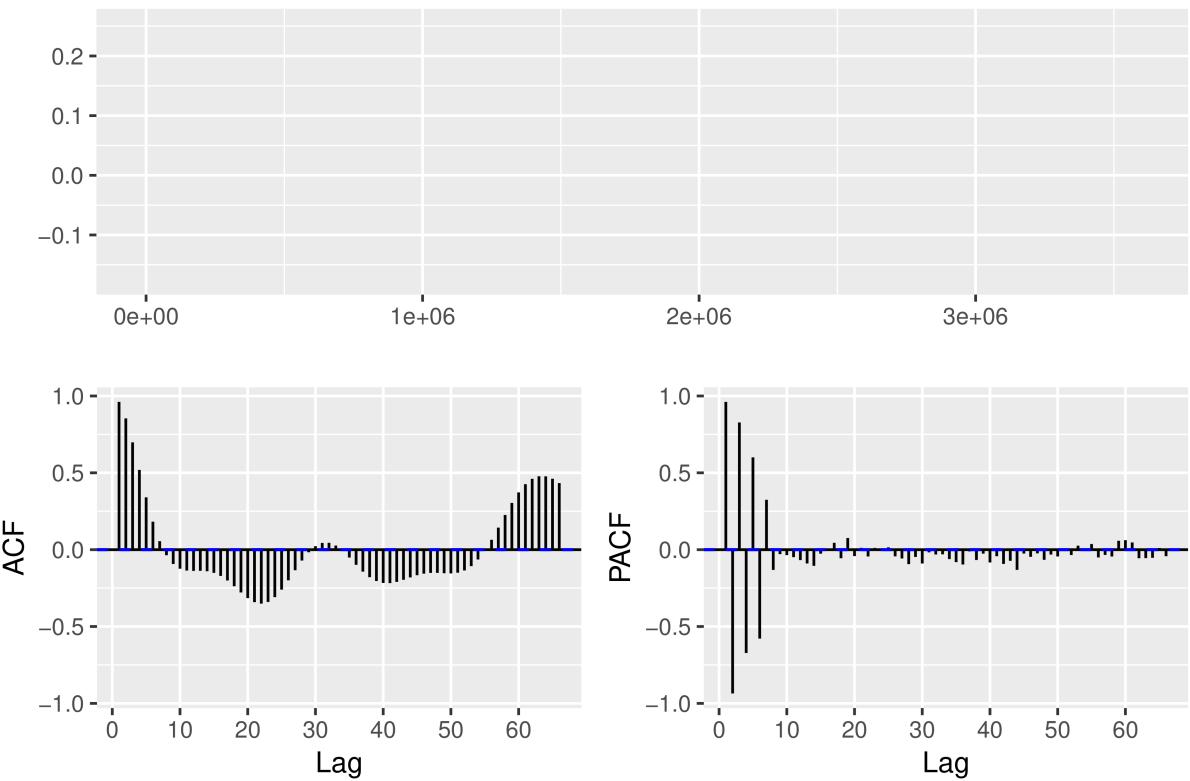
Song Segment #4 Singer. Female 10. Liang Sun Song type. Beijing Opera Song name. Mu Gui Ying Gua Shuai/Lady General Mu takes command Emotion. Postive.

This segment is sung by a new female singer.

```
f10.3 <- readWave('female10_pos_3.wav')
f10.31 <- f10.3@left
f10.31 <- f10.31 / 2^(f10.3@bit -1) #converting to doubles

#full duration
forecast::ggttsdisplay(f10.31, points=FALSE)
```

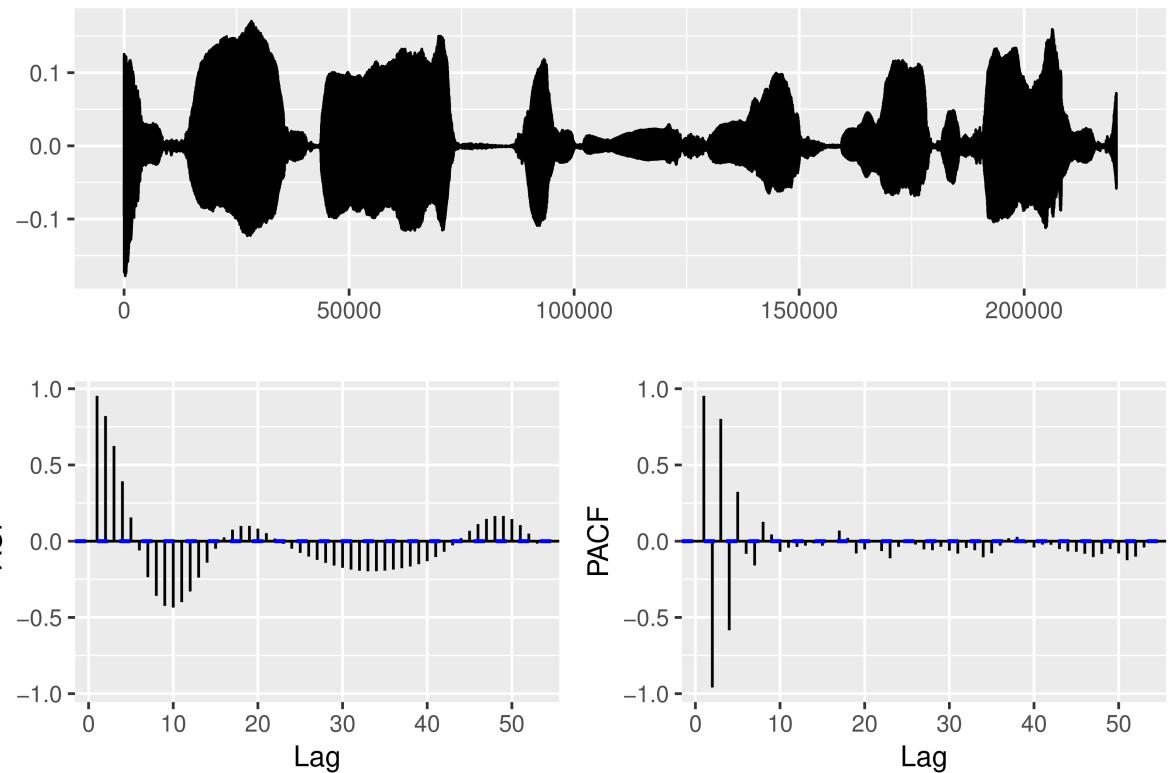
f10.31



The first 1 million datapoints are just white noise. The singing starts at around 22 s into the recording.

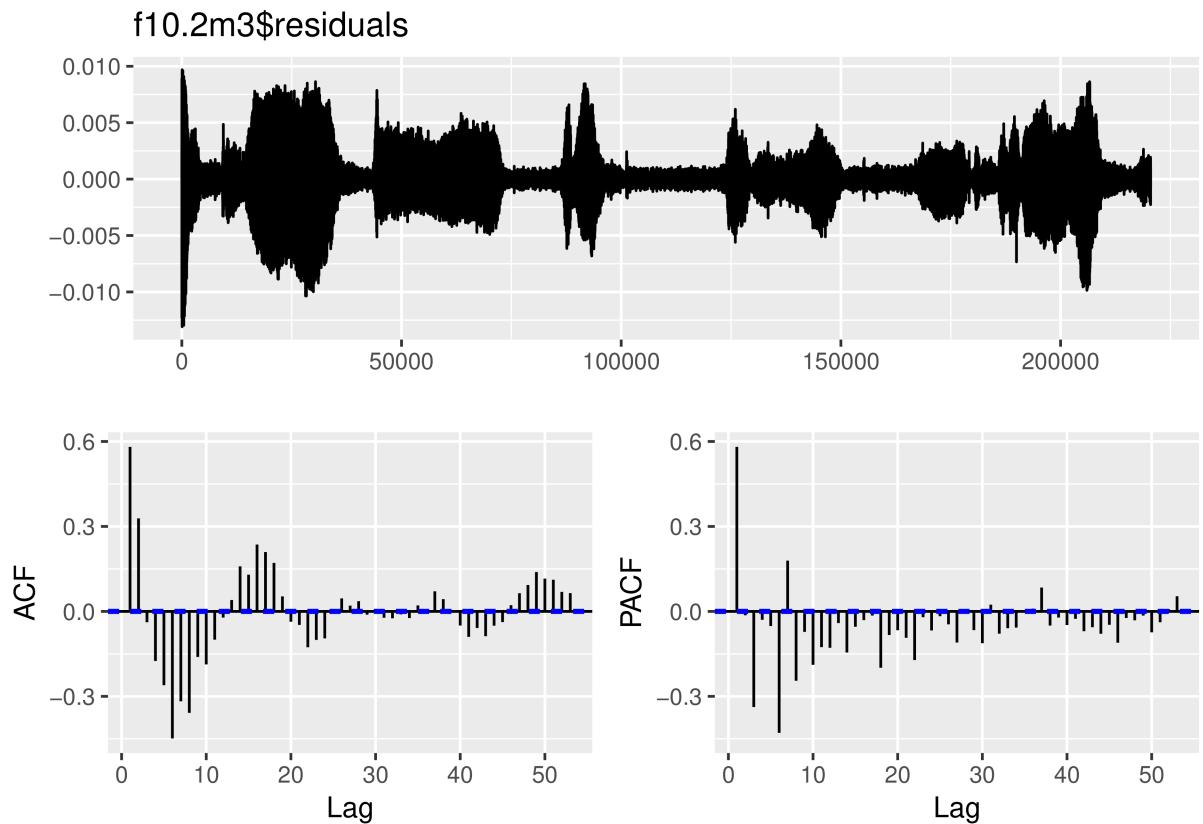
```
#5s duration  
forecast::gtsdisplay(f10.31[1020000:1240500], points=FALSE)
```

f10.3l[1020000:1240500]



```
f10.2m3=forecast::Arima(f10.3l[1020000:1240500],order = c(0,2,2));  
f10.2m3
```

```
## Series: f10.3l[1020000:1240500]  
## ARIMA(0,2,2)  
##  
## Coefficients:  
##          ma1     ma2  
##        1.7334  0.9155  
## s.e.  0.0008  0.0007  
##  
## sigma^2 estimated as 1.98e-06:  log likelihood=1134953  
## AIC=-2269899  AICc=-2269899  BIC=-2269868  
forecast::ggtstdisplay(f10.2m3$residuals, points=FALSE)
```



Eventually the best fit model is (0,2,2).

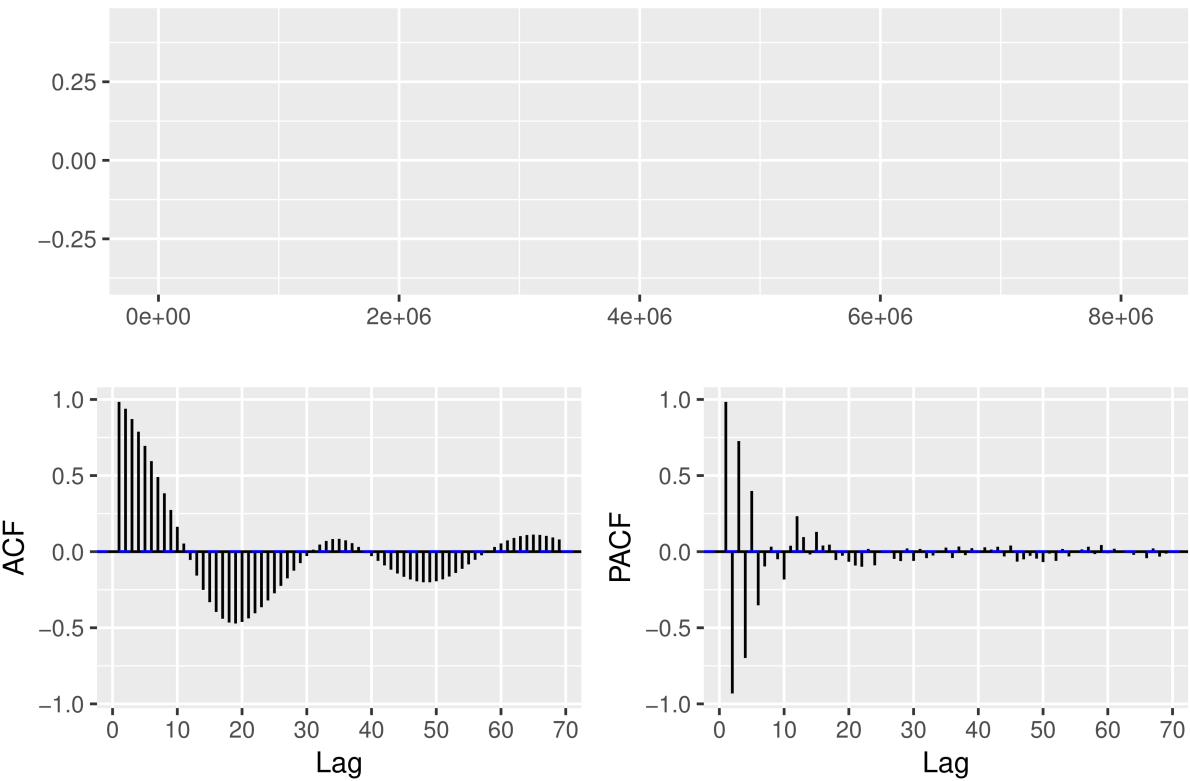
Song Segment#5 Singer 3 profile. Male 12. Bidong Guo Song type. Beijing Opera Song name. Kong Cheng Ji/The stratagem of the empty city Emotion. Postive.

The segment is sung by a male singer.

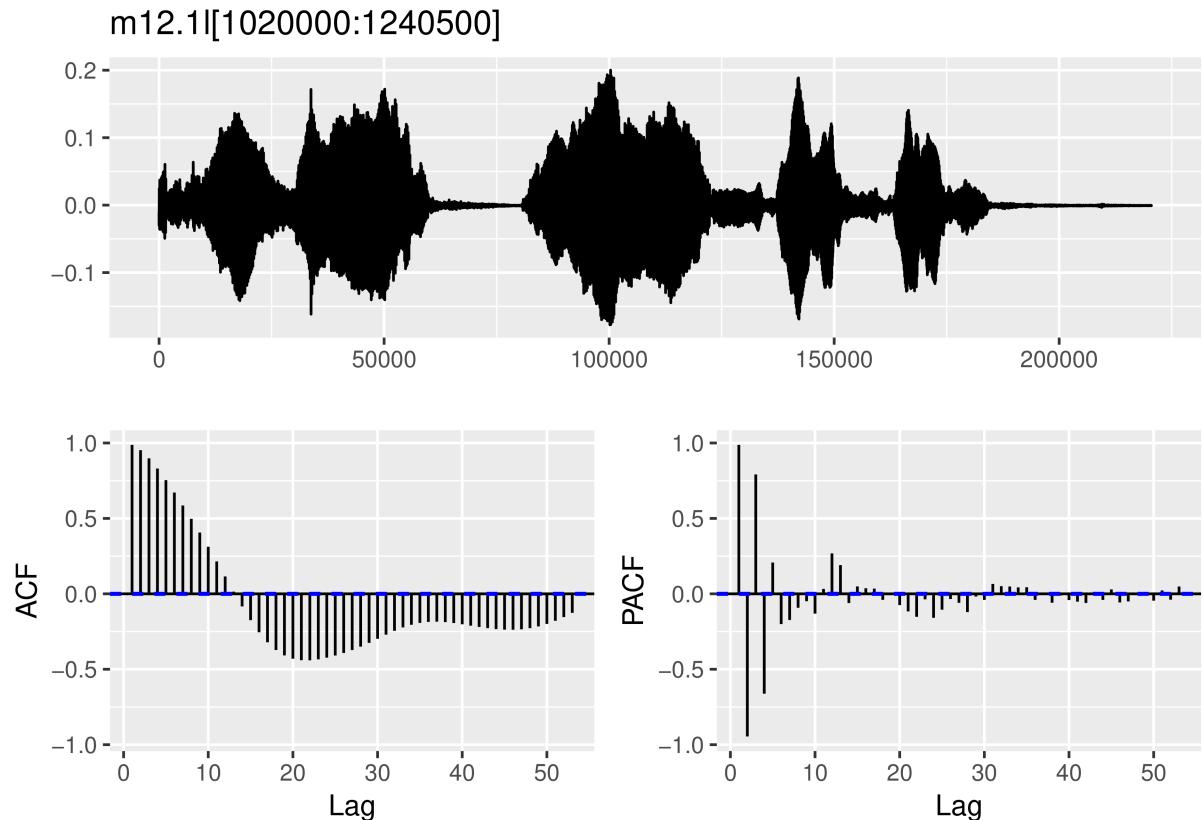
```
m12.1 <- readWave('male12_pos_1.wav')
m12.11 <- m12.1@left
m12.11 <- m12.11 / 2^(m12.1@bit -1) #converting to doubles

#full duration
forecast::ggtsddisplay(m12.11, points=FALSE)
```

m12.11



```
#5 s  
forecast::ggttsdisplay(m12.11[1020000:1240500], points=FALSE)
```



```
m12.3m2=forecast::Arima(m12.11[661500:882000],order = c(0,3,1));
m12.3m2
```

```
## Series: m12.11[661500:882000]
## ARIMA(0,3,1)
##
## Coefficients:
##          ma1
##          0.9010
## s.e.  0.0009
##
## sigma^2 estimated as 4.958e-06:  log likelihood=1033768
## AIC=-2067533  AICc=-2067533  BIC=-2067512
```

```
m12.3m3=forecast::Arima(m12.11[661500:882000],order = c(0,3,2));
m12.3m3
```

```
## Series: m12.11[661500:882000]
## ARIMA(0,3,2)
##
## Coefficients:
##          ma1      ma2
##          1.3316  0.5881
## s.e.  0.0016  0.0016
##
## sigma^2 estimated as 3.696e-06:  log likelihood=1066143
## AIC=-2132280  AICc=-2132280  BIC=-2132249
```

```

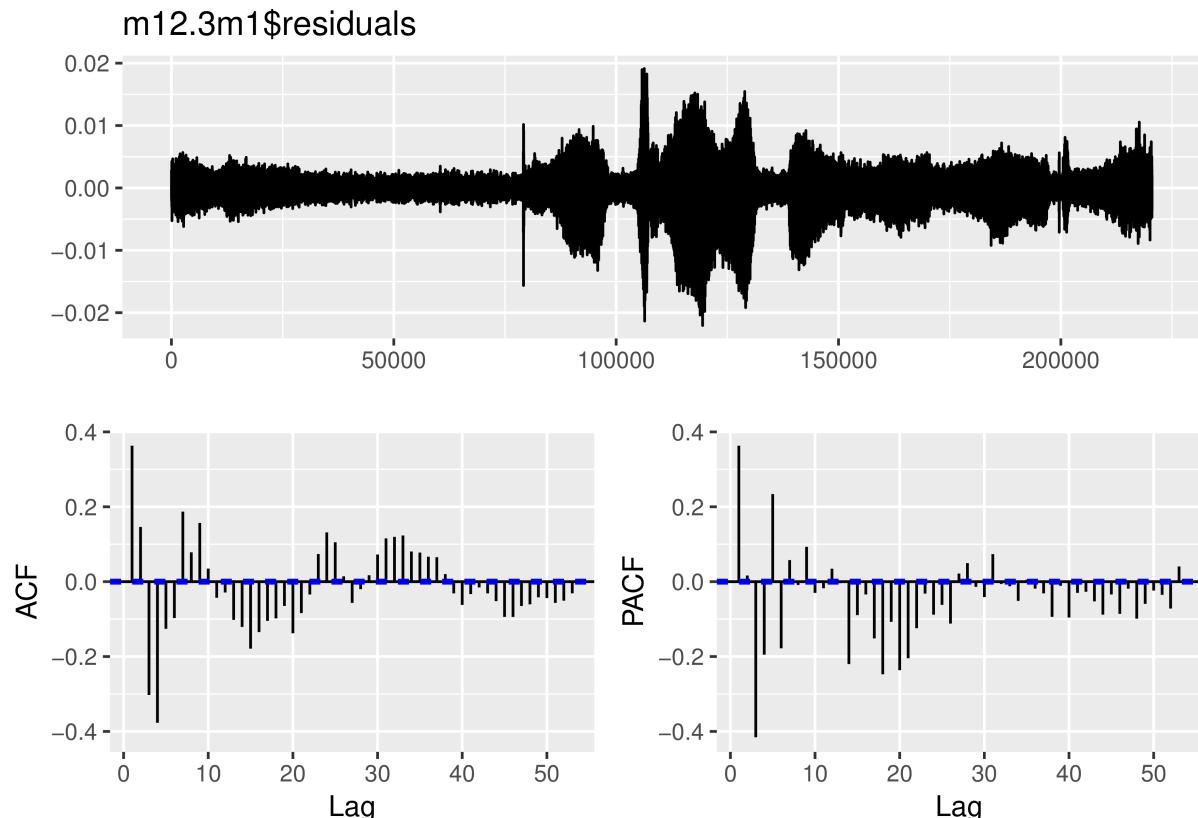
m12.3m1=forecast::Arima(m12.11[661500:882000],order = c(0,2,2));
m12.3m1

## Series: m12.11[661500:882000]
## ARIMA(0,2,2)
##
## Coefficients:
##          ma1     ma2
##        1.5978  0.807
## s.e.  0.0011  0.001
##
## sigma^2 estimated as 3.589e-06:  log likelihood=1069389
## AIC=-2138772  AICc=-2138772  BIC=-2138741

```

(0,3,2) gives minor improvement of both AIC and BIC over (0,3,1). But (0,2,2) is slightly better than (0,3,2) even with one less parameter.

```
forecast::gtsdisplay(m12.3m1$residuals, points=FALSE)
```



The best fit model for segment #5 is (0,2,2).

Song segment #6 Singer Profile. Male 13 Tingfeng Zhang Song type. Beijing Opera Song name. Si Lang Tan Mu/Yang Silang Visits His Mother Emotion. Positive.

Its sung by a second male singer.

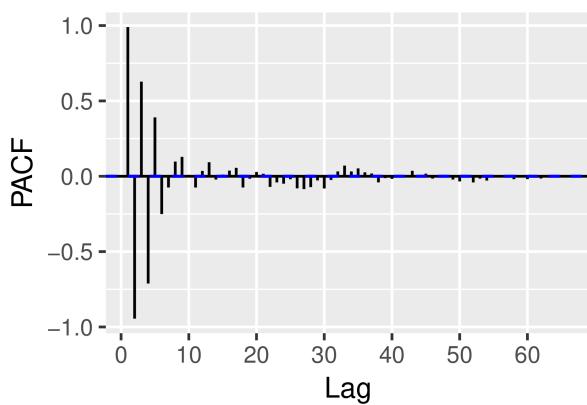
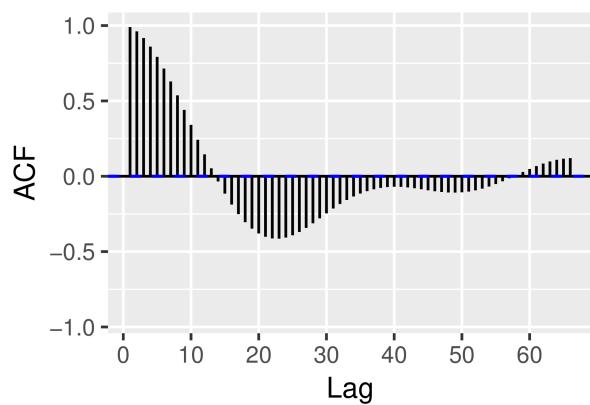
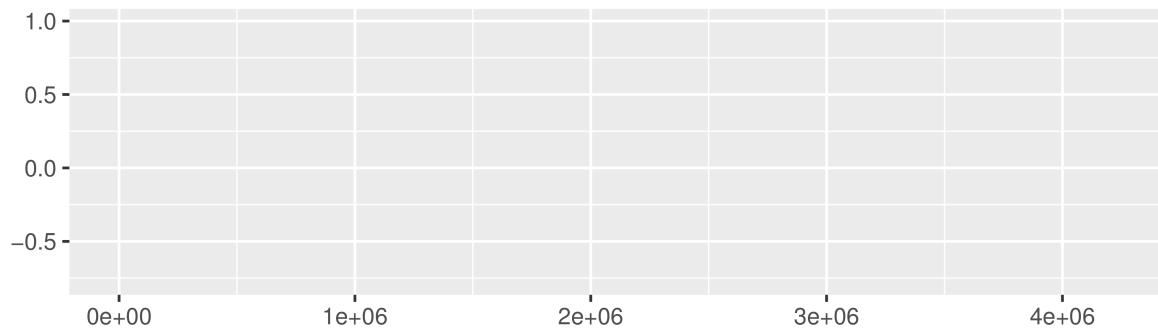
```

m13.3 <- readWave('male13_pos_3.wav')
m13.3l <- m13.3@left
m13.3l <- m13.3l / 2^(m13.3@bit -1) #converting to doubles

```

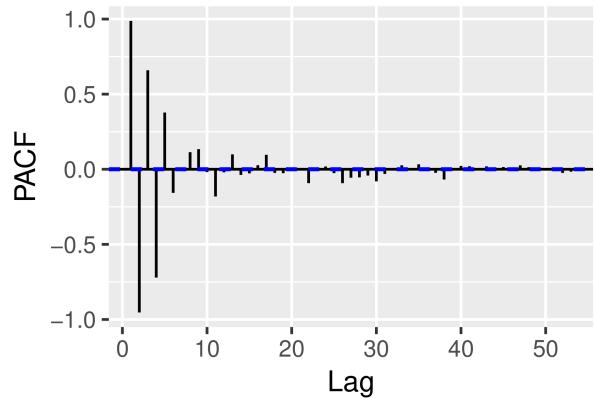
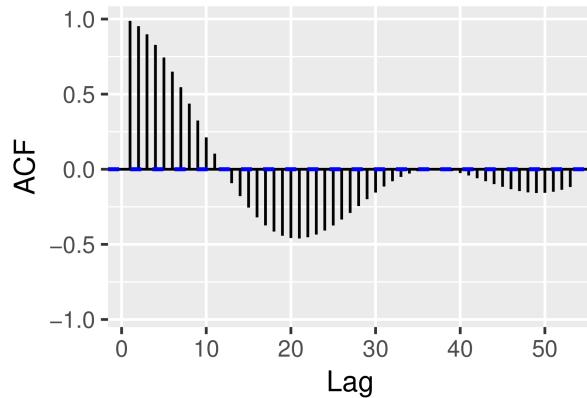
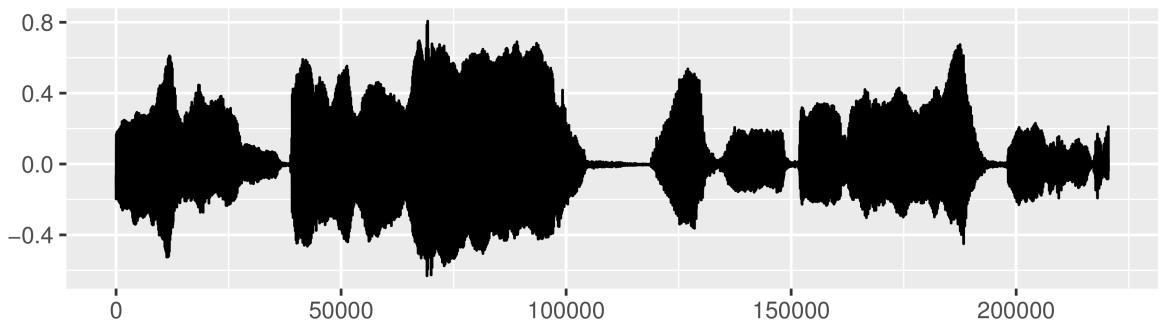
```
#full duration  
forecast::ggtsdisplay(m13.31, points=FALSE)
```

m13.31



```
forecast::ggtsdisplay(m13.31[50000:270500], points=FALSE)
```

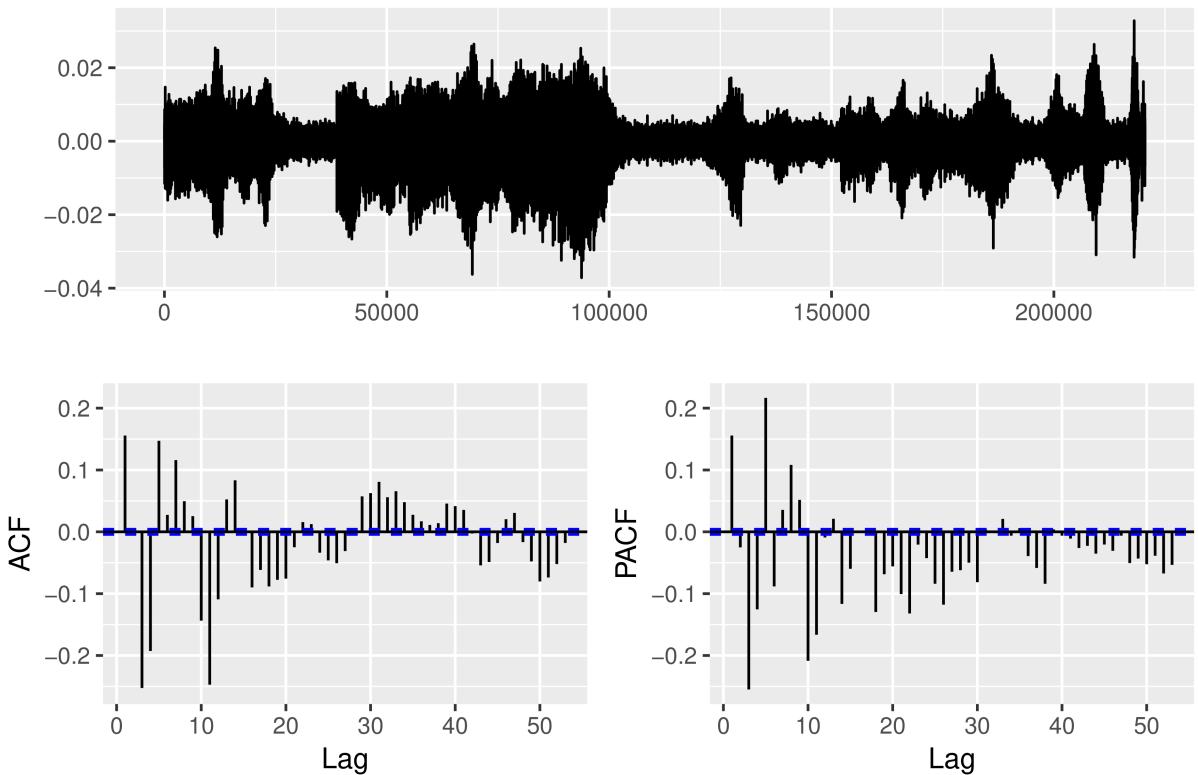
m13.3l[50000:270500]



```
m13.2m3=forecast::Arima(m13.3l[50000:270500],order = c(0,2,2));  
m13.2m3
```

```
## Series: m13.3l[50000:270500]  
## ARIMA(0,2,2)  
##  
## Coefficients:  
##          ma1     ma2  
##        1.4105  0.6865  
## s.e.  0.0014  0.0013  
##  
## sigma^2 estimated as 1.194e-05:  log likelihood=936872.5  
## AIC=-1873739  AICc=-1873739  BIC=-1873708  
forecast::ggtsdisplay(m13.2m3$residuals, points=FALSE)
```

m13.2m3\$residuals



Decide on $(0,2,2)$.

Discussion.

A total of 6 song segments from 4 different singers and 5 different songs are fitted with a ARIMA model. In general, it is observed that the best fit usually requires between second order of differencing. And the MA term is usually first or second order while AR term is always not necessarily. Looking at the final residual plot for each song segment, there are inevitably some acoustic structure left unexplained by the model but overall the residuals are smaller than about 5% of its original amplitude. So to some extent, a ARIMA model can explain the sound structure found in these Beijing Opera. The results from model fitting can be summarized below.

```

para <- matrix(c('female1','bei4 jiu1 chan2','(0,2,1)',-1234526,-1234505,0.9502,'none','female1','bei4
colnames(para) <- c("Singer","Song","Selected Model","AIC","BIC","Theta1","Theta2")
rownames(para) <- c("SongSeg#1","SongSeg#2","SongSeg#3","SongSeg#4","SongSeg#5","SongSeg#6")
para <- as.table(para)

para

##           Singer   Song          Selected Model AIC
## SongSeg#1 female1 bei4 jiu1 chan2      (0,2,1) -1234526
## SongSeg#2 female1 bei4 jiu1 chan2      (0,2,2) -1762395
## SongSeg#3 female1 zhe4 cai2 shi4      (0,2,2) -1594662
## SongSeg#4 female10 Lady General Mu takes command (0,2,2) -2269899
## SongSeg#5 male12 The stratagem of the empty city (0,2,2) -2138772
## SongSeg#6 male13 Yang Silang Visits His Mother   (0,2,2) -1873739

```

```

##          BIC      Theta1 Theta2
## SongSeg#1 -1234505 0.9502 none
## SongSeg#2 -1762395 1.4998 0.635
## SongSeg#3 -1594631 1.4889 0.6158
## SongSeg#4 -2269868 1.7334 0.9155
## SongSeg#5 -2138741 1.5978 0.807
## SongSeg#6 -1873708 1.4105 0.6865

```

All models mentioned below and not shown in the above table can be found in ‘Sta 444 Project full fitting procedure and supplementary models_xl149.pdf’

Comparing song segments from the same song and same singer.

Segment#1 and #2 are from the same song. The selected models for #1 and #2 both have a second order differencing term. And even though selected model segment#1 has only first order MA term, (0,2,2) for segment#1 actually gives a slightly lower AIC and BIC than (0,2,2), with $\theta_1 = 1.2168$, $\theta_2 = 0.2758$. Its just because the improvements are minor, a simpler model (0,2,1) is eventually preferred for segment#1. On the other hand, if we fit segment#2 with model (0,2,1), the fitted parameters would be $\theta_1 = 0.9439$, which is very close to the θ_1 value for segment#1 and AIC of -1690149, BIC =-1690129 (these additional models can be found in project_full_analysis.pdf). So performance of (0,2,1) is also comparable to (0,2,2) on segment#2. Considering that segment#1 is from about 0.5 to 5.5 seconds in the song and segment#2 is from 43.5 to 48.5s from the song and they are clearly audibly different, it suggests that segments from the same song can be modeled by a single ARIMA model, despite the different lyrics that might be sung at different times.

Comparing song segments by the same singer but different songs.

Segment#2 and #3 are by the same singer but from different songs. And they both have the best model to be (0,2,2) and the fitted parameters are surprisingly close to each other as well. $\theta_1 = 1.4998$, $\theta_2 = 0.635$ for segment#2 and $\theta_1 = 1.4889$, $\theta_2 = 0.6158$ for segment#3.

When we compare segment#1 to #3 by looking up parameters of (0,2,1) fitted to segment#3. Its found that $\theta_1 = 0.9385$ which is very close to θ for segment#1. Segment#2 (0,2,1) model has $AIC = -1520259$, $BIC = -1520239$, which are also comparable to that of (0,2,2).

These observations can be summarized as that despite that segments#1, #2 and #3 are from different parts of the same song or even from different songs. So long the singer is the same person, they can all be modeled by a single model relatively well in this case (0,2,2) (not necessarily implying that a single model would be the best for all segment but its at least comparable to the best model of that particular segment in terms of AIC and BIC).

Comparing song segments among the same gender.

Segment#1 is by female singer #1 and segment#4 is by female singer #10. The selected model for each segment differs from each other by 1 order in MA term. And if we fit (0,2,1) to segment#4 (found in project_full_analysis). It can be found that $\theta_1 = 0.9985$ with $AIC=-2058517$ and $BIC=-2058496$, which is above 10% greater than the AIC/BIC given by (0,3,1). And the θ_1 value of segment#4's (0,2,1) model differs from $\theta_1 = 0.9502$ of segment#1 by a fair amount. And the best fit model (0,2,2) of segment#4 have parameters $\theta_1 = 1.7334$, $\theta_2 = 0.9155$, which differ by quite a bit from those in (0,2,2) of segment#2 and segment#3.

Segment#5 is by male singer #12 and segment#6 is by male singer #13. Both of them again have the best fit model to be ARIMA (0,2,2). And in accordance with observations of famle singers, even they have the same model, the parameters values of θ_1, θ_2 differ from each other by quite a bit.

These observations suggest that even different female (or male) singers's voice can be fitted well by ARIMA (0,2,2) model. The values of θ_1, θ_2 depend on the individual voice characteristic of the singer.

Comparing song segments across genders.

As its briefly mentioned above, it seems regardless of genders, all beijing opera song data can be fitted well with ARIMA (0,2,2). However, a different singer will have a unique set of parameter values associated with θ_1 and θ_2 . So gender difference can not be reflected by ARIMA model.

Conclusion

After fitting 6 different song segments that are each 5 seconds long, its found that ARIMA (0,2,2) model fits generally well for all song segments regardless of the singer's gender and individuality. While gender difference can't be captured by ARIMA (0,2,2) model, individual differences are reflected in values of the fitted parameters namely θ_1, θ_2 . So long the segment came from the same singer, the lyrics of the segment or the song that segment is from don't contribute significantly to the values of the fitted parameter. As shown by models of segment#1, #2, #3, even segment #3 is in a different song from segment#1 and #2, its model parameter values still closely match those of segment#1 and #2 (provided that its fitted with the same model (0,2,2)). And segment #4, sang by a different female singer, show different fitted values from those of segment#1, #2, #3. The remaining two segments #5 and #6 can all be fitted with ARIMA (0,2,2) but their fitted values differ from all other segments because they each have a unique male singer. In brief, even gender difference can't be reflected in the model, but individuality can be distinguished from fitted parameter values θ_1, θ_2 . And surprisingly, ARIMA (0,2,2) works well for all segments regardless of singer, specific songs, lyrics and gender.