TEST 1

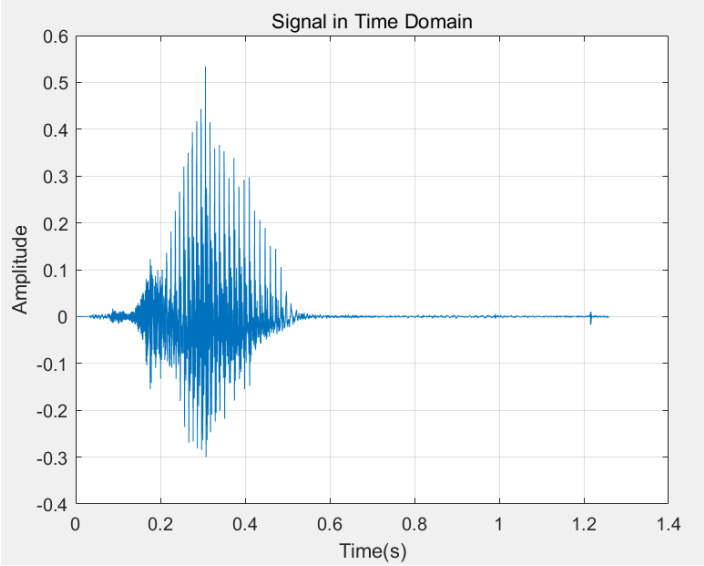
Under the condition of listening to the training and test sets only once, our accuracy is very low, at 25% and 12.5%, respectively. The main reason is the difficulty in remembering which speaker a heard voice belongs to. However, if we can listen to the training set multiple times and match each speaker in the test set, the accuracy improves significantly to 87.5% and 100%. However, this process is quite time-consuming.

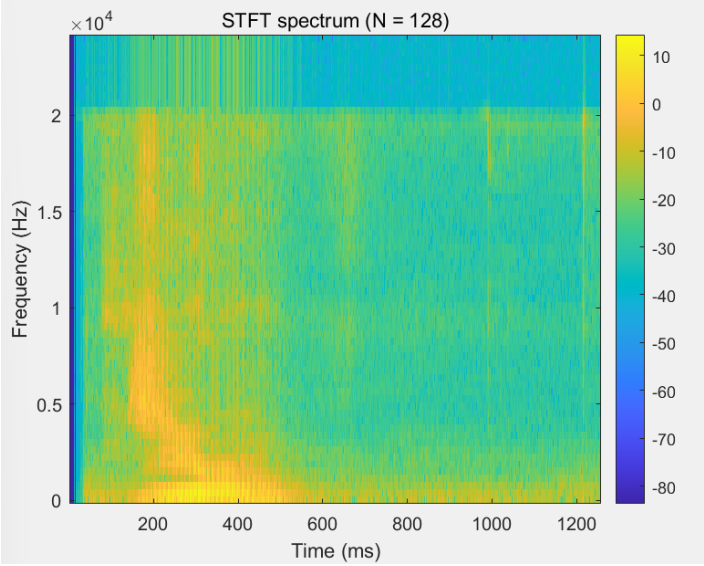
TEST 2

The test file is Zero_train1.wav.

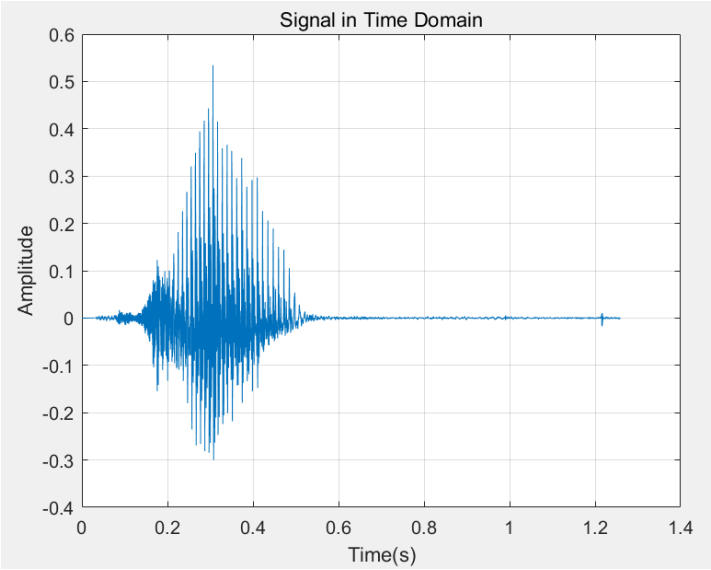
```
>> test2
Sampling rate of Zero_train1.wav : 48000 Hz
time of 512 samples: 5.33 ms
```

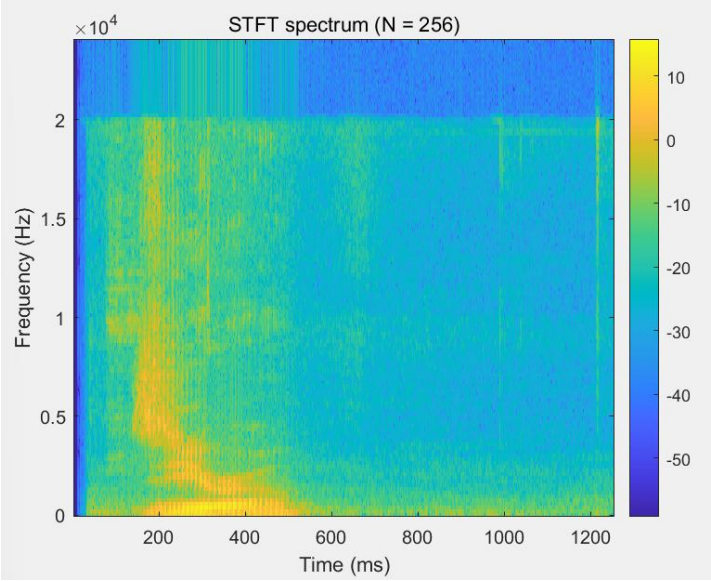
N=128



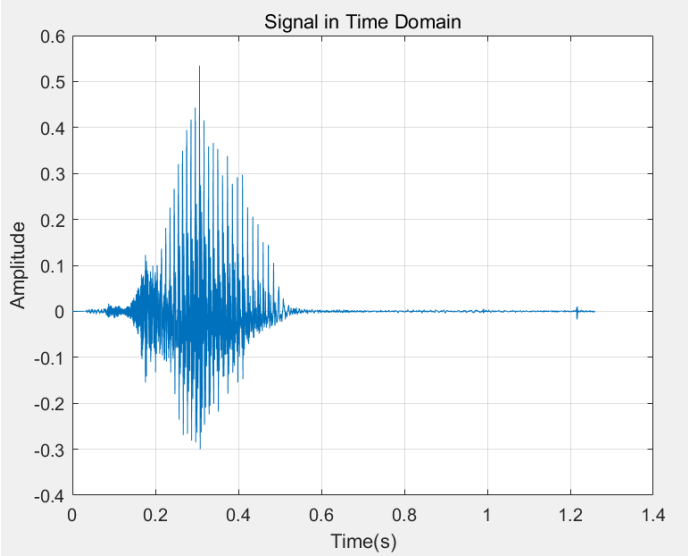


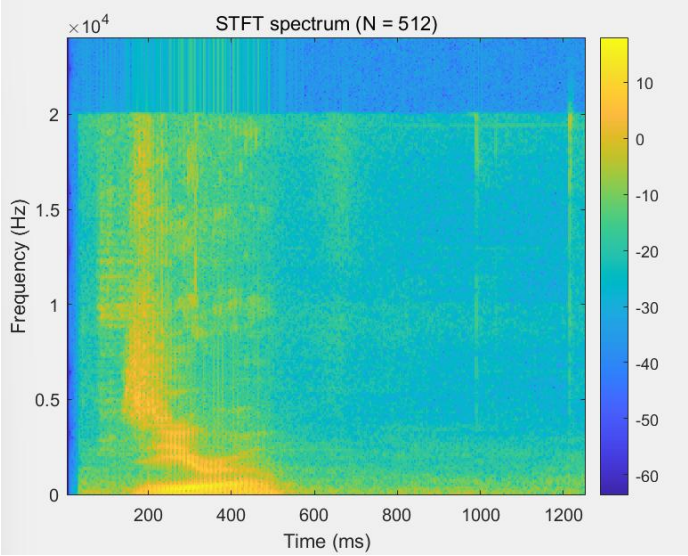
N=256


Signal in Time Domain
Amplitude
Time(s)
0.6
0.5
0.4
0.3
0.2
0.1
0
-0.1
-0.2
-0.3
-0.4
0
0.2
0.4
0.6
0.8
1
1.2
1.4


STFT spectrum (N = 256)
×10^4
Frequency (Hz)
Time (ms)
2
1.5
1
0.5
0
200
400
600
800
1000
1200
10
0
-10
-20
-30
-40
-50

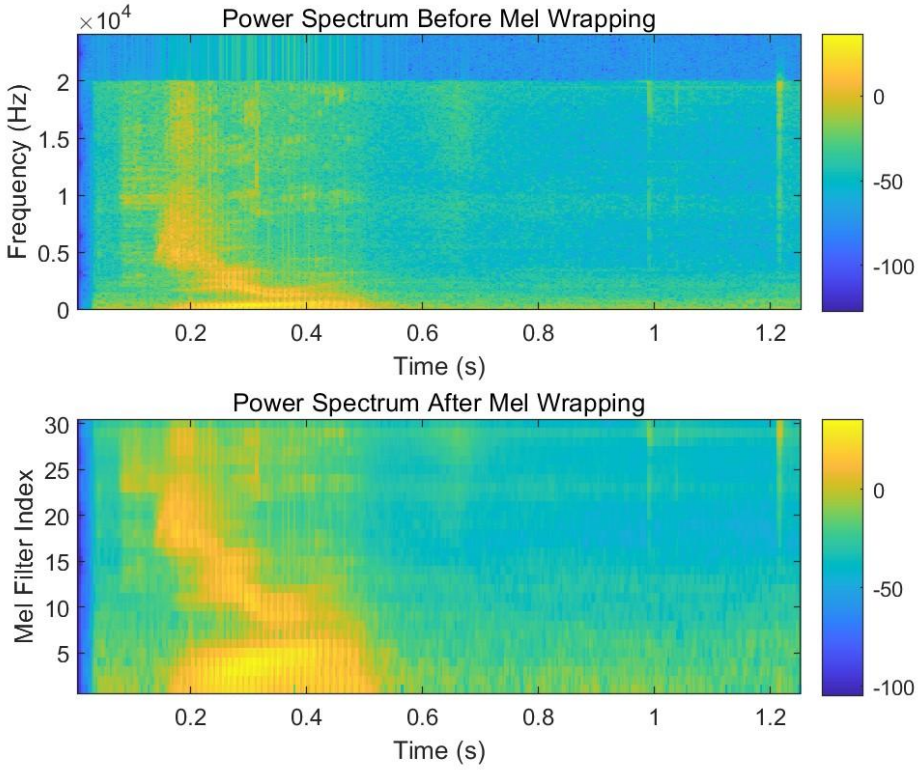
N=512

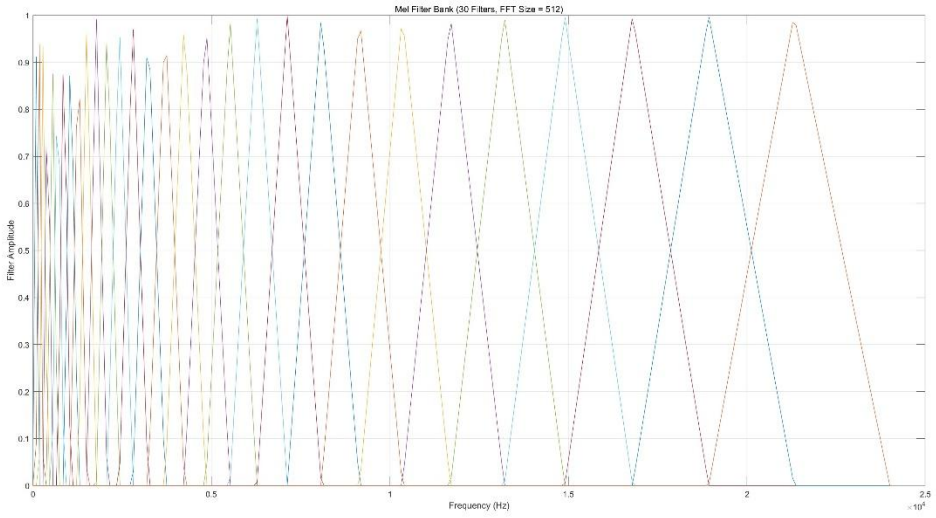

Signal in Time Domain
Amplitude
Time(s)
0.6
0.5
0.4
0.3
0.2
0.1
0
-0.1
-0.2
-0.3
-0.4
0
0.2
0.4
0.6
0.8
1
1.2
1.4


STFT spectrum (N = 512)
×10^4
Frequency (Hz)
Time (ms)
2
1.5
1
0.5
0
200
400
600
800
1000
1200
10
0
-10
-20
-30
-40
-50
-60

TEST 3

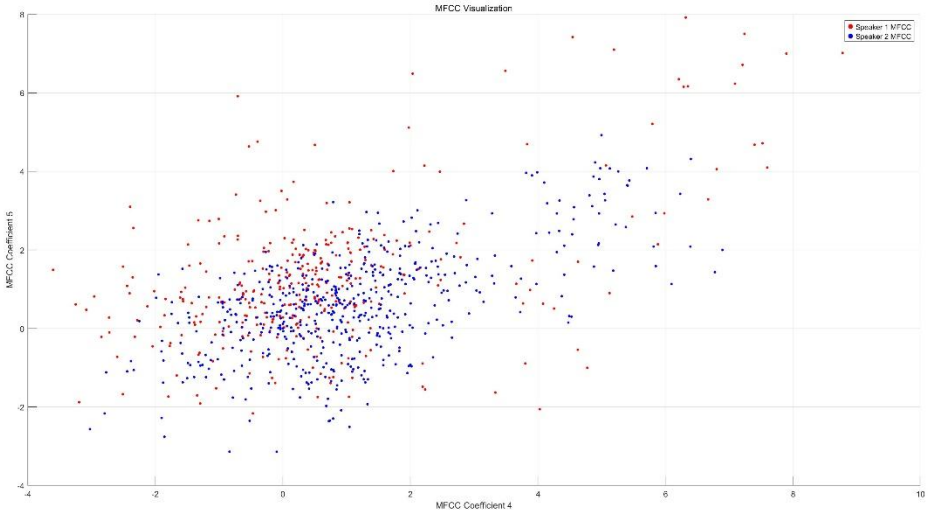
We can see that the filter bank exhibits slight distortions at the base of the triangle, and the tip appears somewhat asymmetric. However, the overall response remains very close to the triangular response. By comparing the power spectra before and after, we observe that after the mel-frequency wrapping step, the proportion of effective information has significantly increased. This indicates that this step effectively extracts the useful power components from the speech file while removing less meaningful regions, providing substantial convenience for the subsequent MFCC computation.





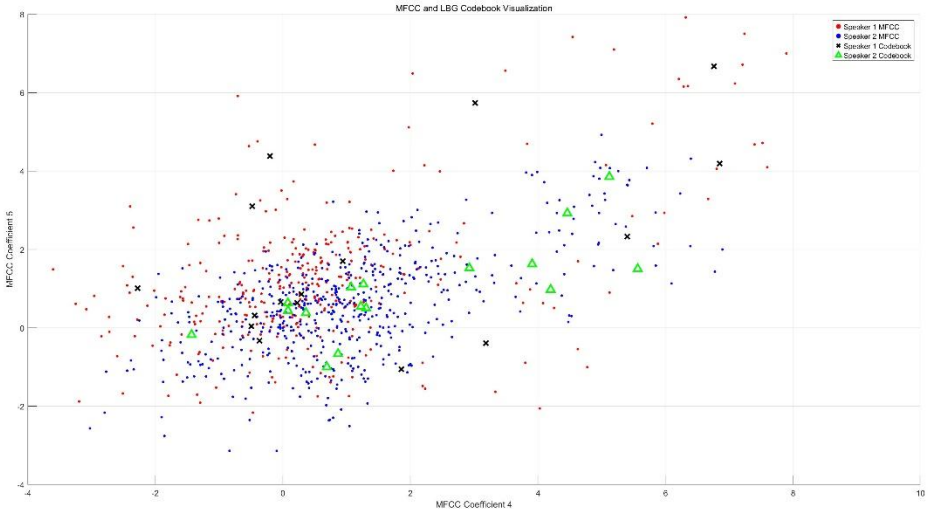
TEST 5

We can observe that when plotting a scatter plot using any two dimensions of the MFCC, each speaker indeed exhibits a clustering effect. However, the clustering trends between different speakers appear quite similar, raising concerns about whether the LBG algorithm can correctly distinguish between them. Therefore, we explored different methods to enhance the distinction between speakers, such as analyzing the fourth and fifth dimensions of the MFCC or increasing the FFT size.



TEST 6

Fortunately, after training with the LBG algorithm, the resulting codebook, or centroid points, exhibit a reasonable degree of distinction between different speakers. This suggests that we have a good chance of using the codebook to differentiate speakers in the test set.



TEST 7

The training and test files are from GivenSpeech_Data. The recognition rate is 100%.

```
>> speaker_identification
Test radio s1.wav is predicted to be: s1
Test radio s2.wav is predicted to be: s2
Test radio s3.wav is predicted to be: s3
Test radio s4.wav is predicted to be: s4
Test radio s5.wav is predicted to be: s5
Test radio s6.wav is predicted to be: s6
Test radio s7.wav is predicted to be: s7
Test radio s8.wav is predicted to be: s8
>>
```

TEST 8

When using the new test set of voice signals processed by notch filters, the following results show that test radio s3.wav is sometimes recognized incorrectly and sometimes recognized correctly. However, all other test radios can be recognized correctly, indicating that the system is robust.

```
>> speaker_identification
Test radio s1.wav is predicted to be: s1
Test radio s2.wav is predicted to be: s2
Test radio s3.wav is predicted to be: s8
Test radio s4.wav is predicted to be: s4
Test radio s5.wav is predicted to be: s5
Test radio s6.wav is predicted to be: s6
Test radio s7.wav is predicted to be: s7
Test radio s8.wav is predicted to be: s8
>> speaker_identification
Test radio s1.wav is predicted to be: s1
Test radio s2.wav is predicted to be: s2
Test radio s3.wav is predicted to be: s3
Test radio s4.wav is predicted to be: s4
Test radio s5.wav is predicted to be: s5
Test radio s6.wav is predicted to be: s6
Test radio s7.wav is predicted to be: s7
Test radio s8.wav is predicted to be: s8
```

TEST 9

The following results show that test radio Zero_test7 is incorrectly identified as Zero_train11, and the rest of the results are correct. Therefore, the accuracy is 94.4%, which is lower than the accuracy obtained in test 7.

```
>> speaker_identification
Test radio Zero_test1.wav is predicted to be: Zero_train1
Test radio Zero_test10.wav is predicted to be: Zero_train10
Test radio Zero_test11.wav is predicted to be: Zero_train11
Test radio Zero_test2.wav is predicted to be: Zero_train2
Test radio Zero_test3.wav is predicted to be: Zero_train3
Test radio Zero_test4.wav is predicted to be: Zero_train4
Test radio Zero_test6.wav is predicted to be: Zero_train6
Test radio Zero_test7.wav is predicted to be: Zero_train11
Test radio Zero_test8.wav is predicted to be: Zero_train8
Test radio Zero_test9.wav is predicted to be: Zero_train9
Test radio s1.wav is predicted to be: s1
Test radio s2.wav is predicted to be: s2
Test radio s3.wav is predicted to be: s3
Test radio s4.wav is predicted to be: s4
Test radio s5.wav is predicted to be: s5
Test radio s6.wav is predicted to be: s6
Test radio s7.wav is predicted to be: s7
Test radio s8.wav is predicted to be: s8
```

TEST 10a

Question 1

The following results show that Twelve_test16, Zero_test12, and Zero_test7 are misidentified, indicating that the system uses "twelve" to identify the speaker more accurately than "zero".

Question 2

The accuracy of the whole system is 91.7%.

```
>> speaker_identification
Test radio Twelve_test1.wav is predicted to be: Twelve_train1
Test radio Twelve_test10.wav is predicted to be: Twelve_train10
Test radio Twelve_test11.wav is predicted to be: Twelve_train11
Test radio Twelve_test12.wav is predicted to be: Twelve_train12
Test radio Twelve_test13.wav is predicted to be: Twelve_train13
Test radio Twelve_test14.wav is predicted to be: Twelve_train14
Test radio Twelve_test15.wav is predicted to be: Twelve_train15
Test radio Twelve_test16.wav is predicted to be: Zero_train17
Test radio Twelve_test17.wav is predicted to be: Twelve_train17
Test radio Twelve_test18.wav is predicted to be: Twelve_train18
Test radio Twelve_test19.wav is predicted to be: Twelve_train19
Test radio Twelve_test2.wav is predicted to be: Twelve_train2
Test radio Twelve_test3.wav is predicted to be: Twelve_train3
Test radio Twelve_test4.wav is predicted to be: Twelve_train4
Test radio Twelve_test6.wav is predicted to be: Twelve_train6
Test radio Twelve_test7.wav is predicted to be: Twelve_train7
Test radio Twelve_test8.wav is predicted to be: Twelve_train8
Test radio Twelve_test9.wav is predicted to be: Twelve_train9
Test radio Zero_test1.wav is predicted to be: Zero_train1
Test radio Zero_test10.wav is predicted to be: Zero_train10
Test radio Zero_test11.wav is predicted to be: Zero_train11
Test radio Zero_test12.wav is predicted to be: Zero_train13
Test radio Zero_test13.wav is predicted to be: Zero_train13
Test radio Zero_test14.wav is predicted to be: Zero_train14
Test radio Zero_test15.wav is predicted to be: Zero_train15
Test radio Zero_test16.wav is predicted to be: Zero_train16
Test radio Zero_test17.wav is predicted to be: Zero_train17
Test radio Zero_test18.wav is predicted to be: Zero_train18
Test radio Zero_test19.wav is predicted to be: Zero_train19
Test radio Zero_test2.wav is predicted to be: Zero_train2
Test radio Zero_test3.wav is predicted to be: Zero_train3
Test radio Zero_test4.wav is predicted to be: Zero_train4
Test radio Zero_test6.wav is predicted to be: Zero_train6
Test radio Zero_test7.wav is predicted to be: Zero_train17
Test radio Zero_test8.wav is predicted to be: Zero_train8
Test radio Zero_test9.wav is predicted to be: Zero_train9
```

TEST 10b

Question 3

Whether "eleven" or "five" is used to identify the speaker, it can be correctly identified. The accuracy of the system is 100%.

Question 4

The accuracy is higher than that of zero/twelve.

```
>> speaker_identification
(Eleven)Test radio s1.wav is predicted to be: s1
(Eleven)Test radio s10.wav is predicted to be: s10
(Eleven)Test radio s11.wav is predicted to be: s11
(Eleven)Test radio s12.wav is predicted to be: s12
(Eleven)Test radio s13.wav is predicted to be: s13
(Eleven)Test radio s14.wav is predicted to be: s14
(Eleven)Test radio s15.wav is predicted to be: s15
(Eleven)Test radio s16.wav is predicted to be: s16
(Eleven)Test radio s17.wav is predicted to be: s17
(Eleven)Test radio s18.wav is predicted to be: s18
(Eleven)Test radio s19.wav is predicted to be: s19
(Eleven)Test radio s2.wav is predicted to be: s2
(Eleven)Test radio s20.wav is predicted to be: s20
(Eleven)Test radio s21.wav is predicted to be: s21
(Eleven)Test radio s22.wav is predicted to be: s22
(Eleven)Test radio s23.wav is predicted to be: s23
(Eleven)Test radio s3.wav is predicted to be: s3
(Eleven)Test radio s4.wav is predicted to be: s4
(Eleven)Test radio s5.wav is predicted to be: s5
(Eleven)Test radio s6.wav is predicted to be: s6
(Eleven)Test radio s7.wav is predicted to be: s7
(Eleven)Test radio s8.wav is predicted to be: s8
(Eleven)Test radio s9.wav is predicted to be: s9
```

```
>> speaker_identification
(Five)Test radio s1.wav is predicted to be: s1
(Five)Test radio s10.wav is predicted to be: s10
(Five)Test radio s11.wav is predicted to be: s11
(Five)Test radio s12.wav is predicted to be: s12
(Five)Test radio s13.wav is predicted to be: s13
(Five)Test radio s14.wav is predicted to be: s14
(Five)Test radio s15.wav is predicted to be: s15
(Five)Test radio s16.wav is predicted to be: s16
(Five)Test radio s17.wav is predicted to be: s17
(Five)Test radio s18.wav is predicted to be: s13
(Five)Test radio s19.wav is predicted to be: s19
(Five)Test radio s2.wav is predicted to be: s2
(Five)Test radio s20.wav is predicted to be: s20
(Five)Test radio s21.wav is predicted to be: s21
(Five)Test radio s22.wav is predicted to be: s22
(Five)Test radio s23.wav is predicted to be: s23
(Five)Test radio s3.wav is predicted to be: s3
(Five)Test radio s4.wav is predicted to be: s4
(Five)Test radio s5.wav is predicted to be: s5
(Five)Test radio s6.wav is predicted to be: s6
(Five)Test radio s7.wav is predicted to be: s7
(Five)Test radio s8.wav is predicted to be: s8
(Five)Test radio s9.wav is predicted to be: s9
```