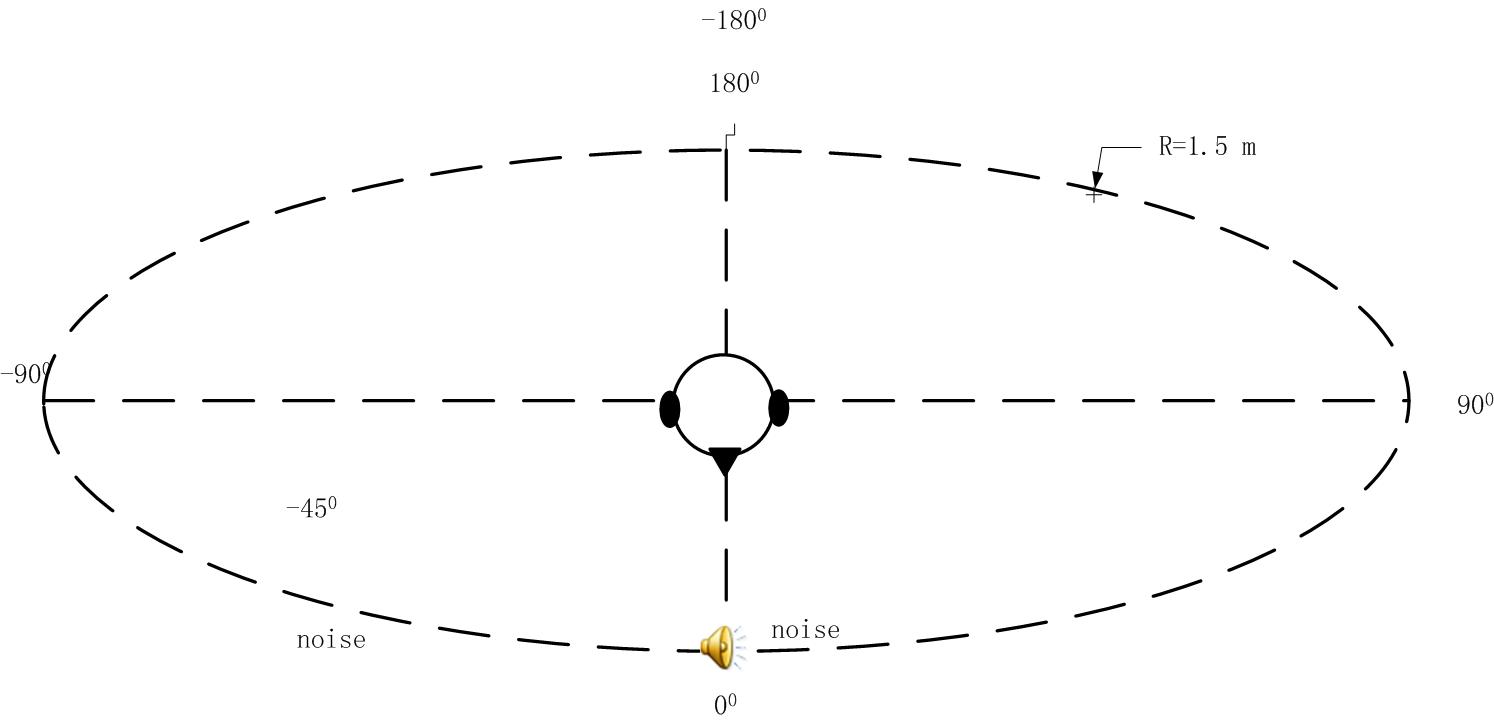
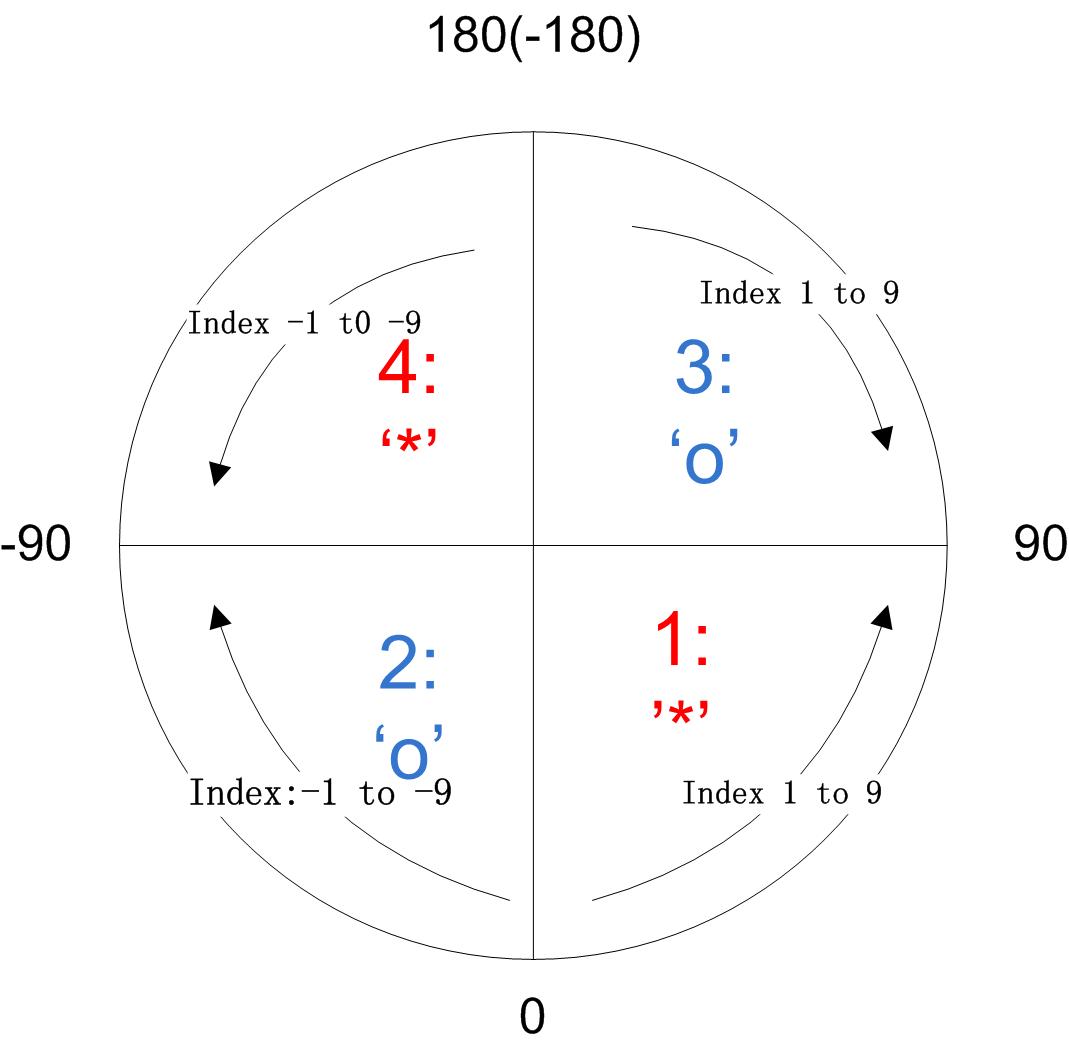
**1、Training and testing azimuths define**

The target speech was fixed on 0 degree.



To show the train azimuths directly, I divided the train 10 degree azimuths to four part index from 1 to 9 (-1 to -9) indicate the interference leave the target speech from 10 degree to 90 degree (-10 degree to -90degree)

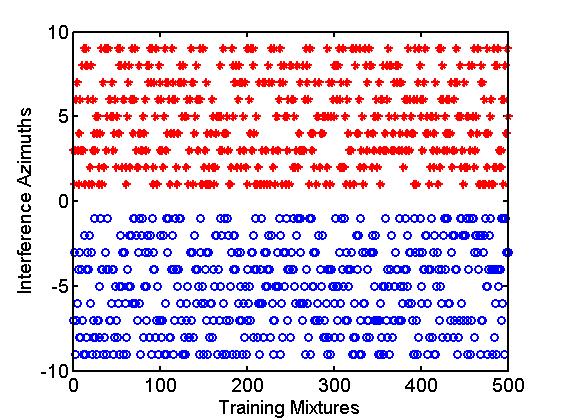


1) Three sources conditions (two noise sound sources)

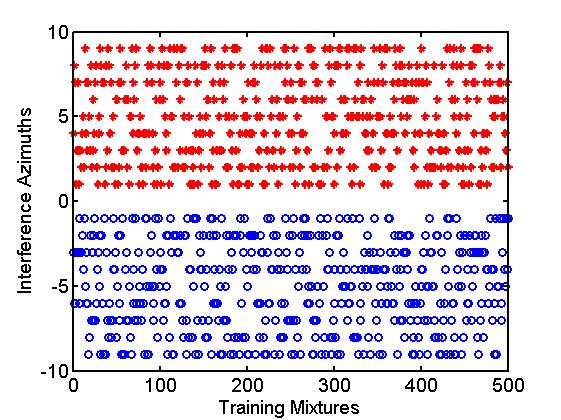
Train: I separate the azimuths to two parts with line from 0°to 180︒(-180°). I use two parts of azimuths: positive part 10°to 90°(part “1”) and negative part -10° to -90°(part “2”). We randomly choose one interference azimuth from positive part, and then choose another interference azimuth from the negative part.

To show the train azimuths directly, I divided the train 10 degree azimuths to four part index from 1 to 9 (-1 to -9) indicate the interference leave the target speech from 10 degree to 90 degree (-10 degree to -90degree)

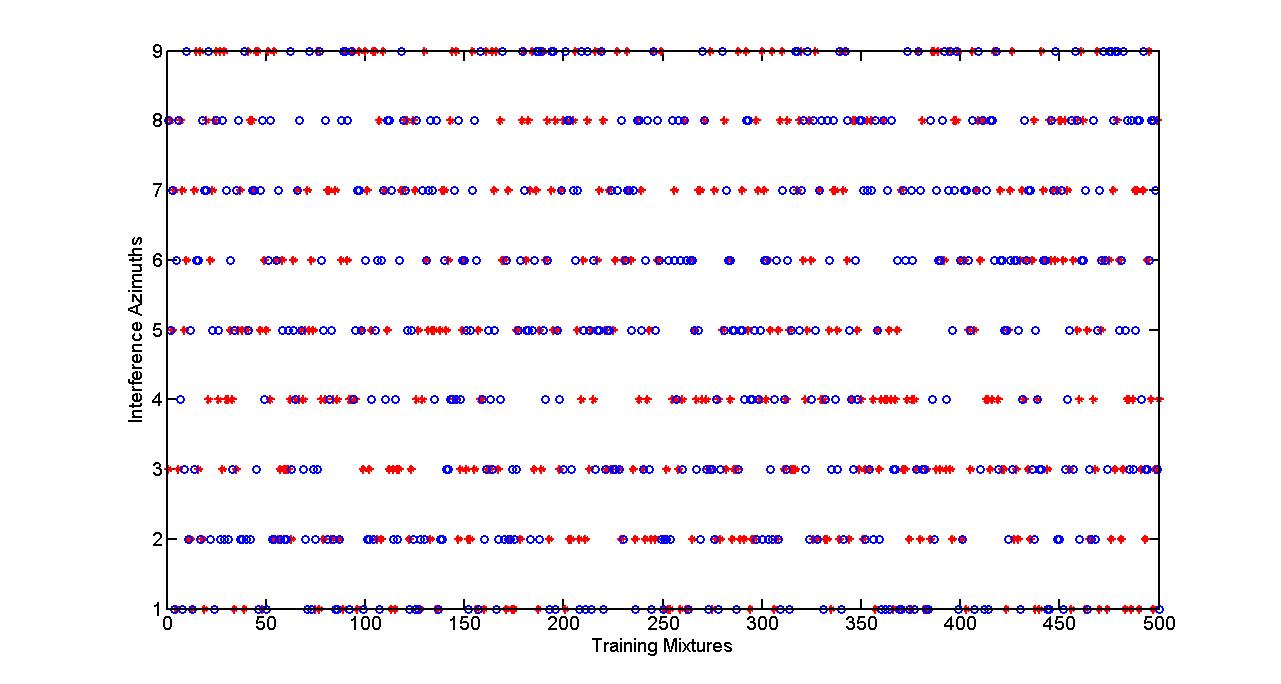
Then the T60=0 two noise sound sources azimuths locate as:



Then the T60=0.3 two noise sound sources azimuths locate as:

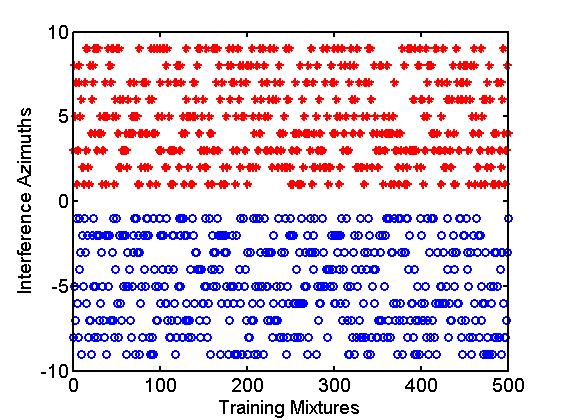


We also can show these positive and negative azimuths only with azimuth as example



We can see the distribution is almost equality

Then the T60=0.7 two noise sound sources azimuths locate as:



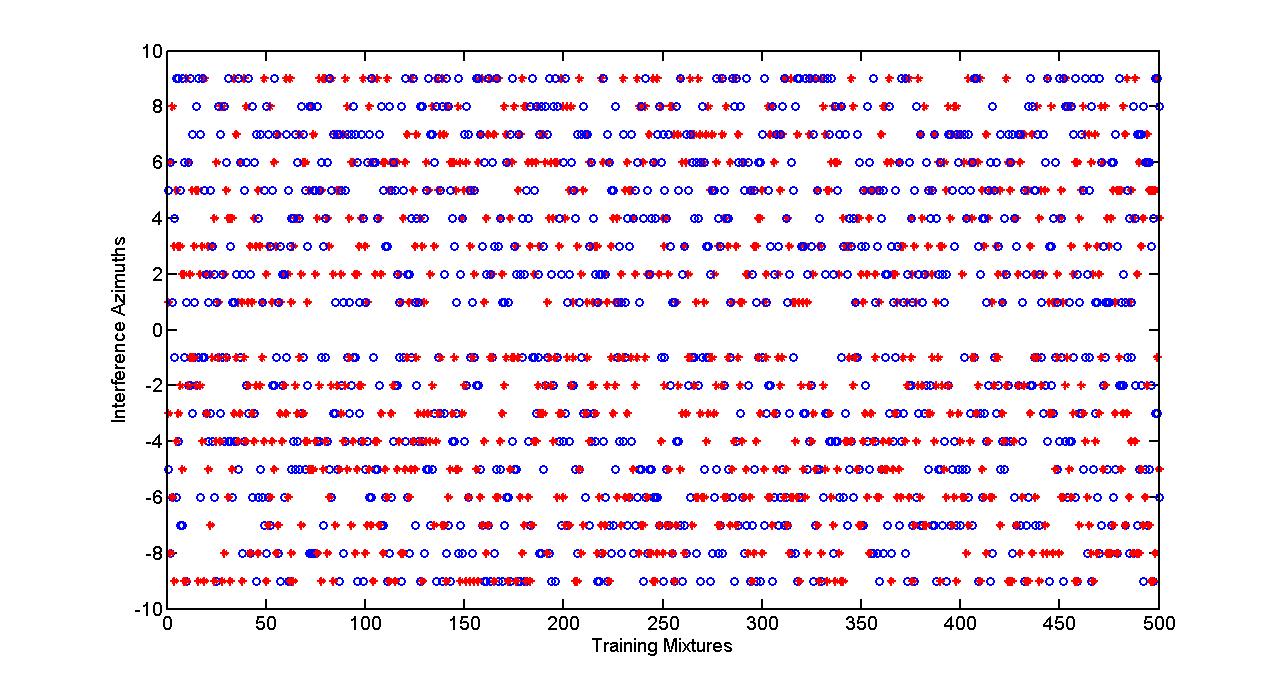
Test: The interferences are fixed on 45° and -45︒.

2) Five sources conditions (Four noise sound sources)

Train: I use all four parts of azimuths: positive part 10°to 90°(part ‘1’and ’3’) and negative part -10° to -90°(part ‘2’ and ‘4’). I randomly choice one azimuth in each part. As show in fellow figures.

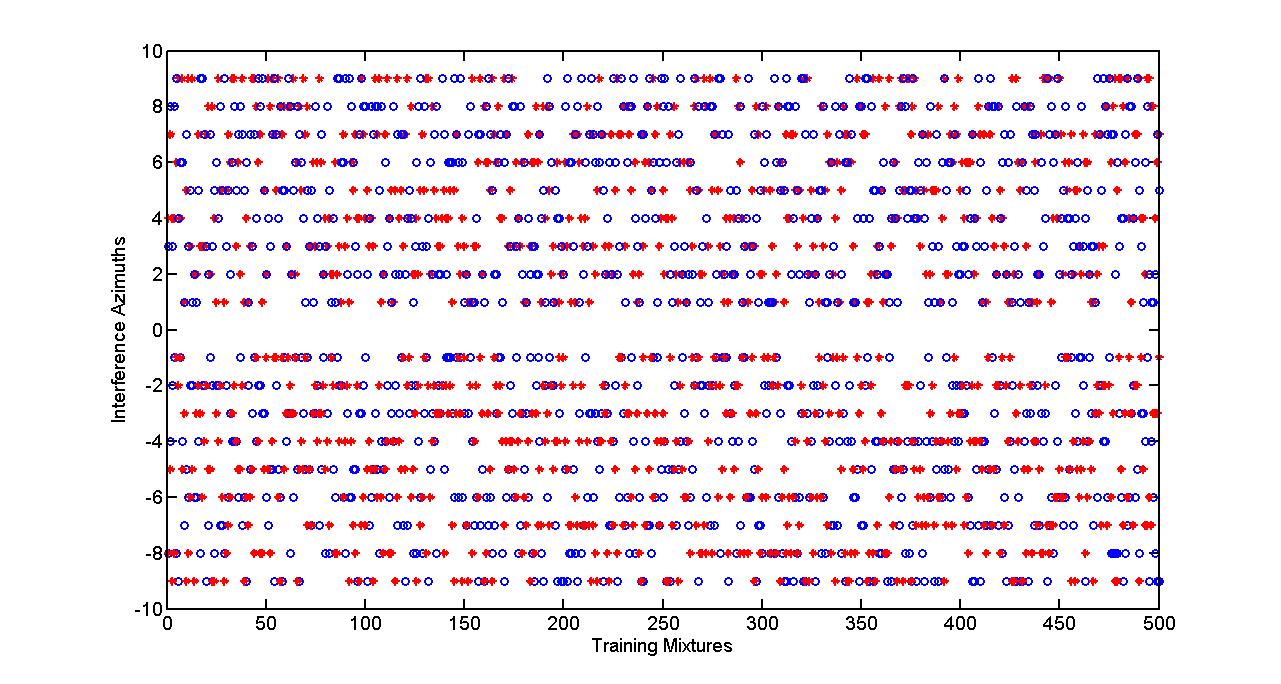
Test: The interferences were fixed on 45°,135°, -45°and -135°.

Then the T60=0 four noise sound sources azimuths locate as:

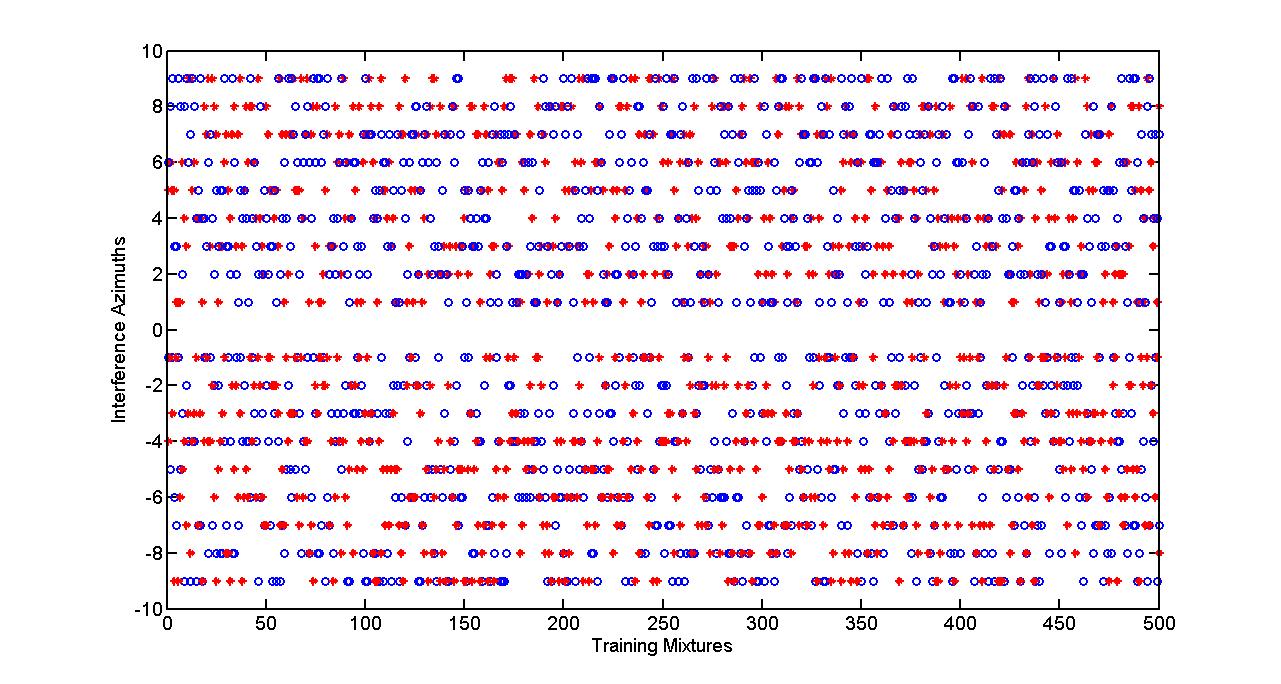


The positive and the negative will have two results, as for the front and rear of the head.

Then the T60=0.3 four noise sound sources azimuths locate as:



Then the T60=0.7 four noise sound sources azimuths locate as:

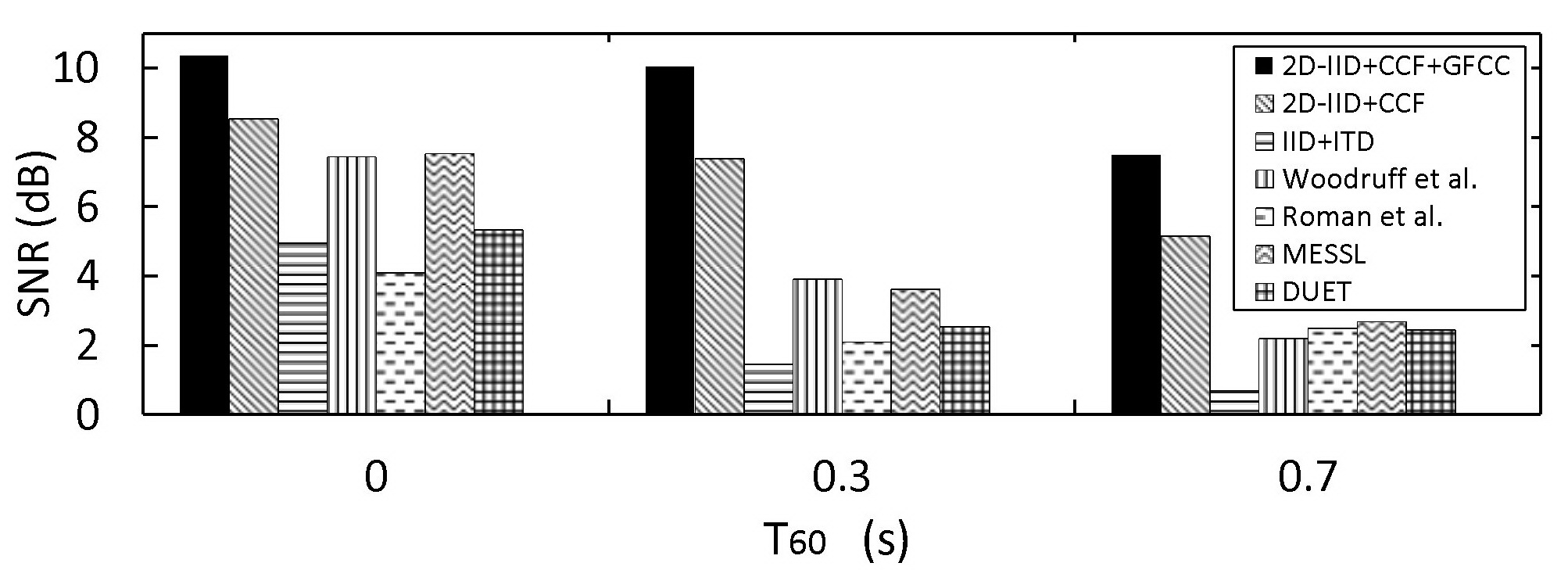


**2、Noise choice**

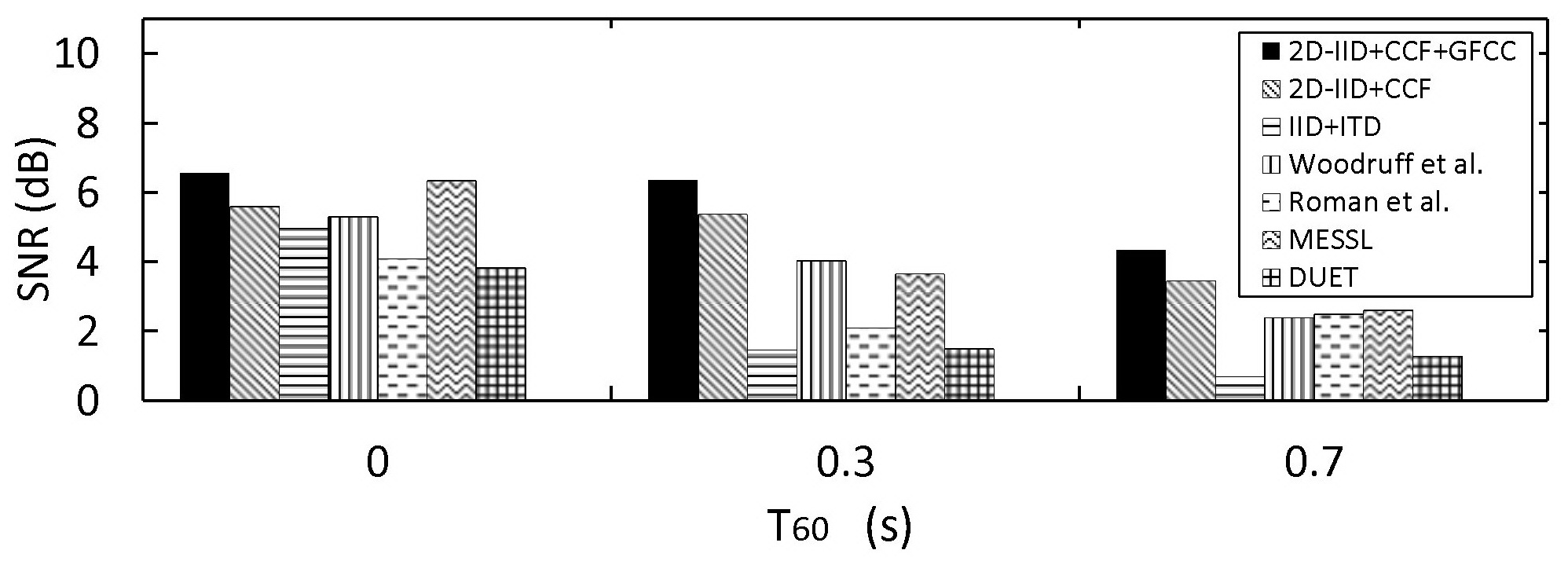
1) Training and test noise: I separated the original “babble” noise (3min55sec) to two parts. One is “babbleTR” (1 min 46 sec) to train the system. Another part called “babbleTE” (2 min 08 sec) to test the system.

2) Noise choice: When I get one target speech, I calculate the length of the target speech. Then I divide the “babbleTR” or “babbleTE” into several parts, which have the same length as the target speech. Than I randomly choice one part from the noise sound parts, and put it to one interference azimuth. I repeat to do it until get the need noise sound sources. We can notice, with different target speech length, the noise separation will be different.

**3、About traditional clean speech based SNR results**



IBM\_SNR results (0dB)



Clean speech base SNR (0dB)

**Table the HIT-FA results of the compare systems.**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **T60(s)** | **2D-IID+CCF+GFCC** | **2D-IID+CCF** | **IID+ITD** | **Woodruff et al.** | **Roman et al.** | **MESSL** | **DUET** |
| **0** | **72.60** | **68.17** | **39.36** | **60.16** | **27.90** | **41.57** | **39.58** |
| **0.3** | **70.95** | **65.23** | **26.41** | **36.59** | **25.18** | **21.68** | **18.97** |
| **0.7** | **64.69** | **58.42** | **23.05** | **20.97** | **20.12** | **14.11** | **15.52** |

To compare more clearly, I list the full HIT FA results of the binaural feature and MESSL at 0 dB S3 condition, and T60=0s.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Systems | HIT(%) | FA(%) | HIT-FA(%) | ACC(%) | RSR |
| **MESSL** | 57.40 | 15.83 | 41.57 | 76.83 | 85.60 |
| **2D-IID+CCF** | 77.91 | 9.75 | 68.17 | 87.91 | 89.91 |

RSR: retained-speech ratio[Wang, DeLiang. "Time-frequency masking for speech separation and its potential..." (2008).]

I found, in the first condition (comparison figure). The HIT-FA and the IBM-SNR is big gap between the MESSL and 2D-IID+CCF. But the RSR is similar. I thought this is reason that clear-SNR is similar between the MESSL and 2D-IID+CCF.