The DNN use the momentum. It is 0.5. It is correct. I checked it with Yuxuan.

**1、Training and testing azimuths define**

The target speech was fixed on 0 degree.



1) Three sources conditions (two noise sound sources)

Train: I separate the azimuths to two parts with line from 0°to 180︒(-180°), I get two parts of azimuths:0°to 170° positive part and -10° to -180° negative part. We randomly choose one interference azimuth from positive part, and then choose another interference azimuth from the negative part.

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

2) Five sources conditions (Four noise sound sources)

Train: I separate the all azimuths to four parts: 0°to 80°,90 to 170,-10°to -90°and -100 to -180°. I randomly choice one location from 0°to 80 °for the first interference. Then I choice the one location from 90°to 170 °for the second interference. The third azimuth choice from -10°to -90°. The fourth azimuth choice from -100° to -180°.

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

**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)

Notes：

1) MESSL only has ratio masks. He used a sigmoid function to generate the ratio mask as





 In order to make a fair comparison, I generate the binary mask use



Actually, I tested MESSL with the binary masks and the ratio masks in several conditions, the SNR results were similar.