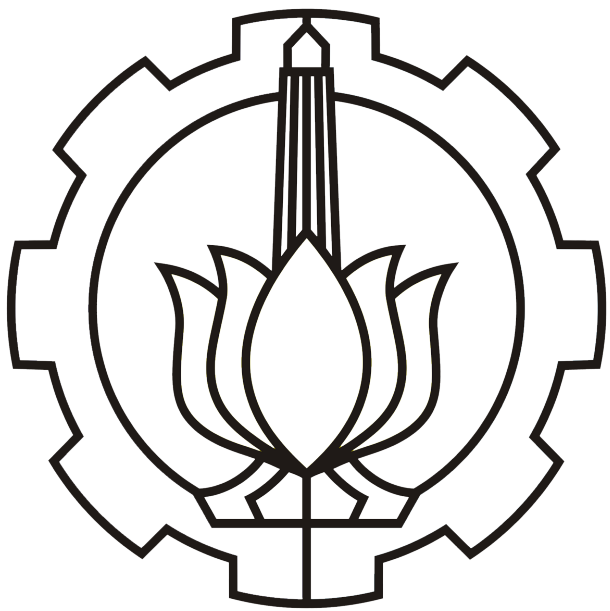
Speech Segregation Based-on Binaural Cue: Interaural Time Difference (ITD) and Interaural Level Difference (ILD)



Mifta Nur Farid
Dhany Arifianto

VibrasticLab
Dept. of Engineering Physics ITS

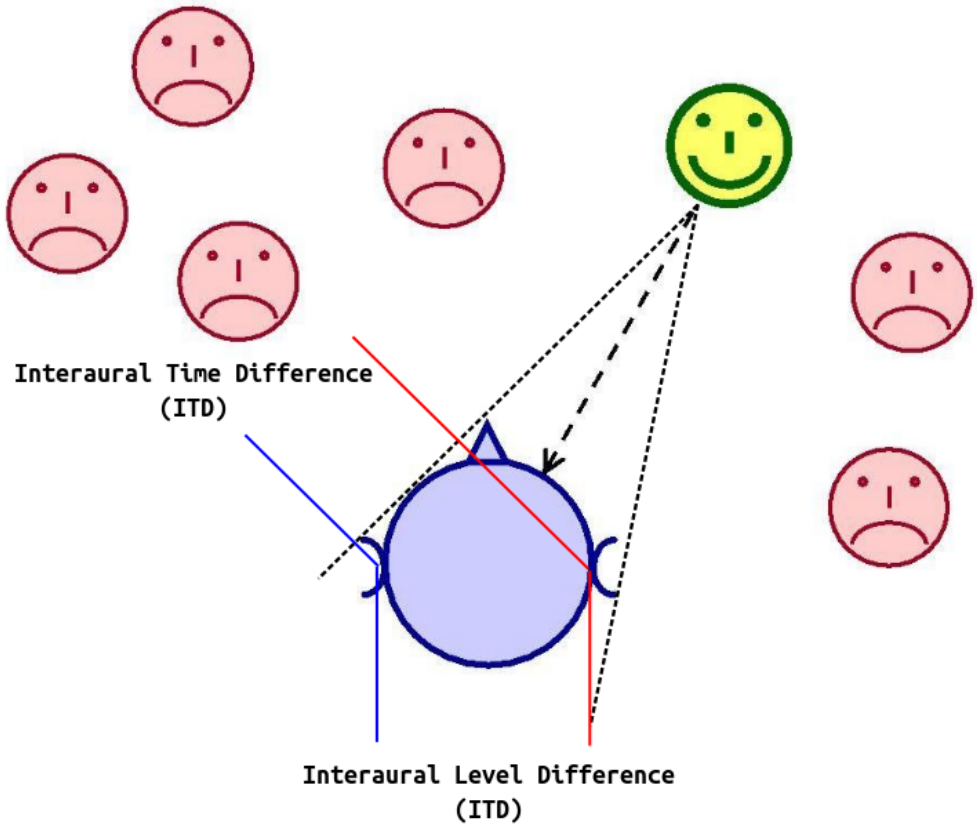
August 23, 2016

Motivation

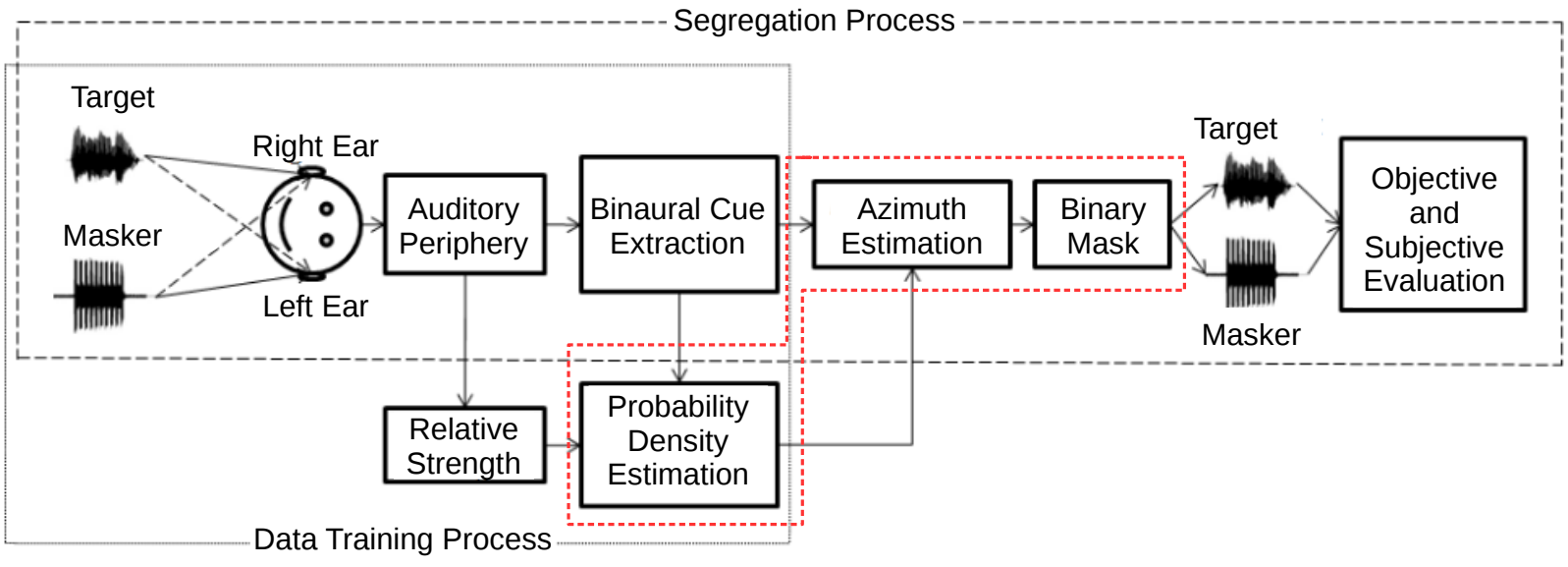
- Hearing loss
- Cause of hearing loss
  - Infection/ injury 17.1%
  - Born with hearing loss 4.4%
  - Noise damage 33.7%
  - Ageing 28%
  - Other 16.8%
- Hearing loss $\rightarrow$ hearing aids
- Hearing aids $\rightarrow$ amplify all sounds
- Most optimal hearing aids: Binaural hearing aids $\sim$ human auditory system

Human auditory system

- Human auditory ability $\rightarrow$ cocktail party effect
- Binaural hearing $\rightarrow$ localization $\rightarrow$ segregation

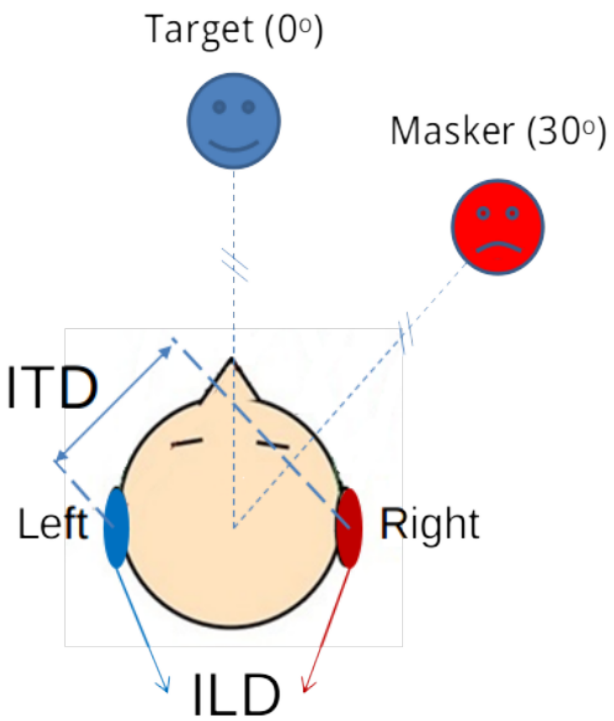


Architecture Model

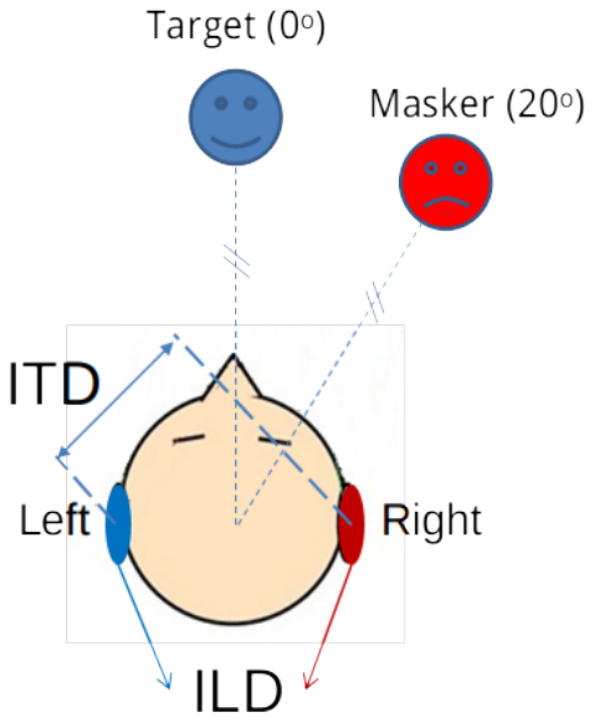


Experiment Setup

- Experiment Setup 1 and 2.



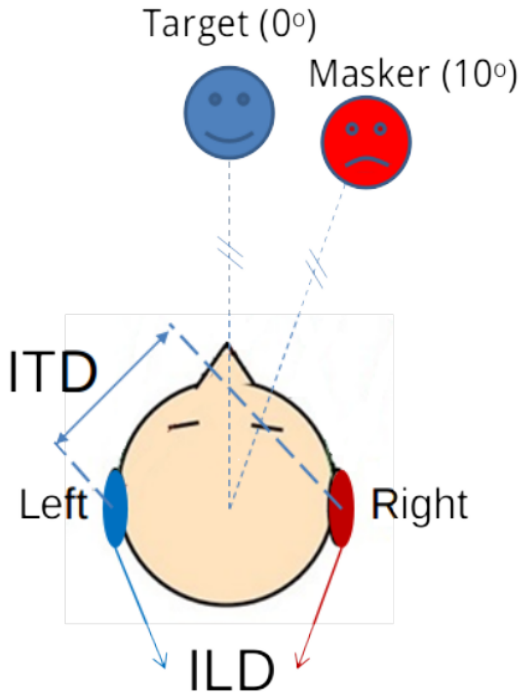
Experiment 1.



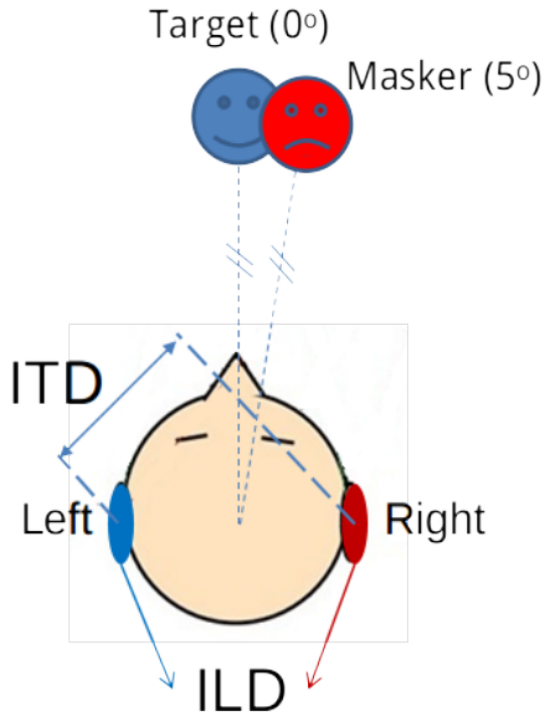
Experiment 2.

Experiment Setup

- Experiment Setup 3 and 4.



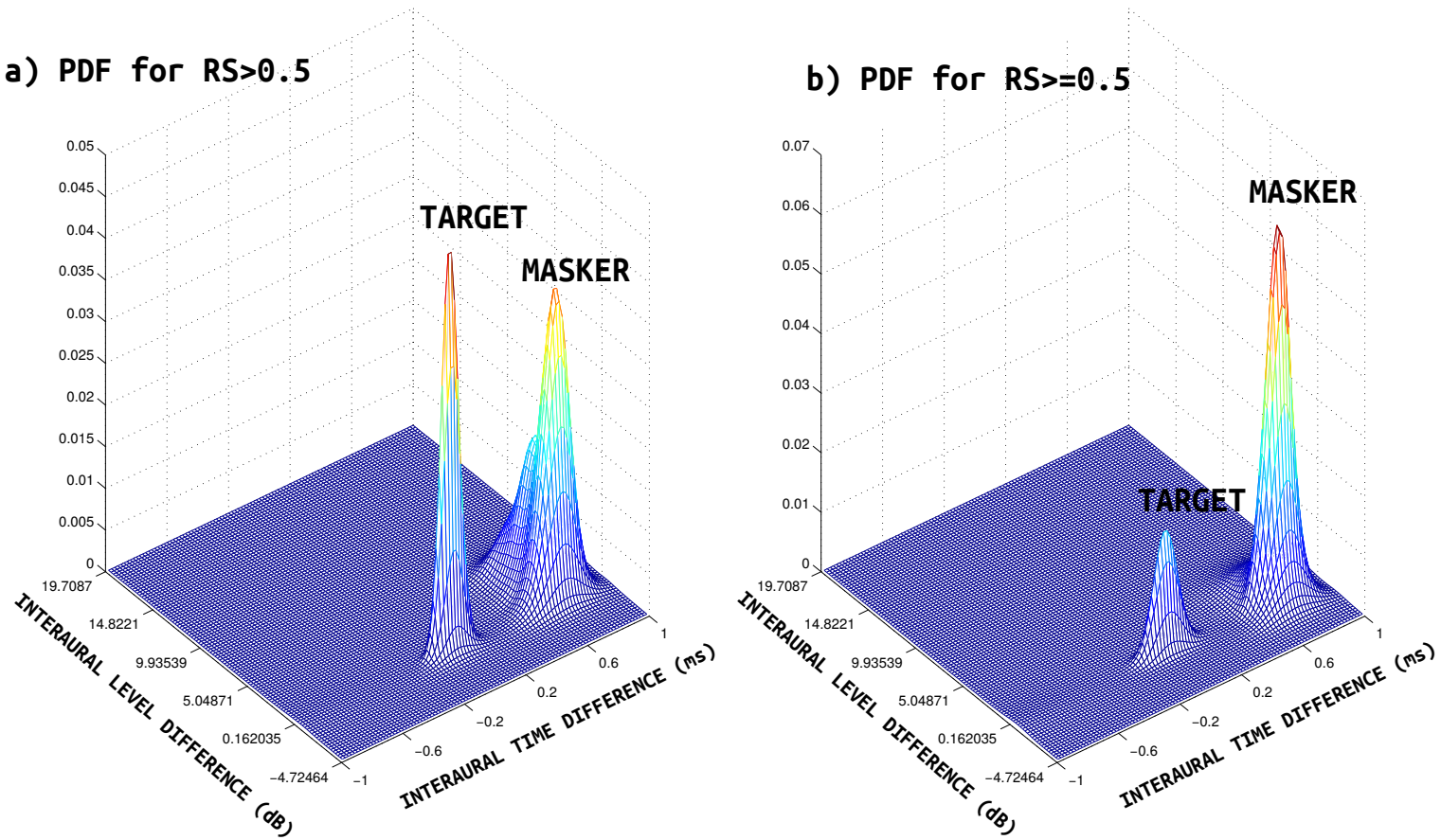
Experiment 3.



Experiment 4.

Probability Density Estimation

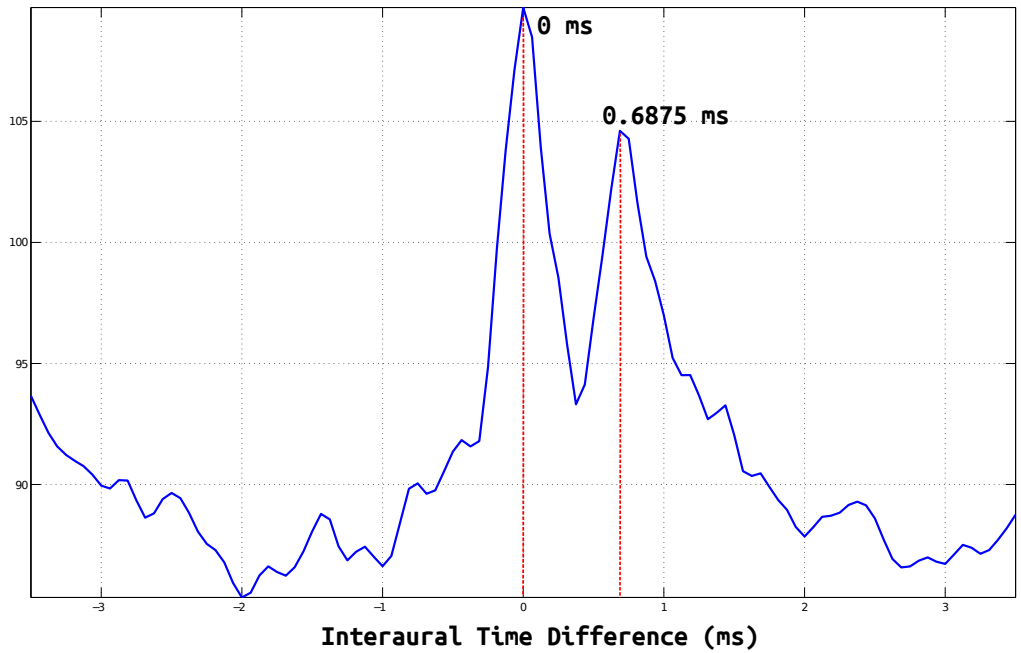
- Experiment setup 1 (SIR 10 dB, 500 Hz)



Azimuth Estimation

- Experiment setup 1 (SIR 10 dB, 500 Hz)

$$C(i,j,\tau) = \frac{\sum_{k=0}^{K-1}(l_i(j-k)-\bar{l}_i)(r_i(j-k-\tau)-\bar{r}_i)}{\sqrt{\sum_{k=0}^{K-1}(l_i(j-k)-\bar{l}_i)^2}\sqrt{\sum_{k=0}^{K-1}(r_i(j-k)-\bar{r}_i)^2}} \quad (1)$$



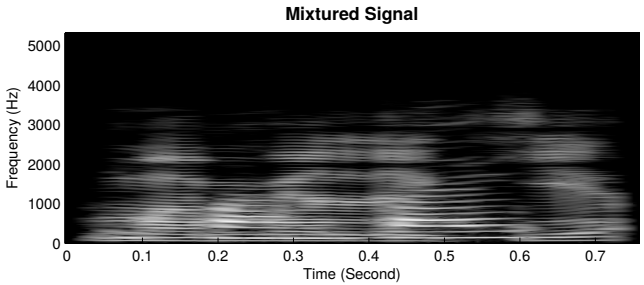
Binary mask estimation

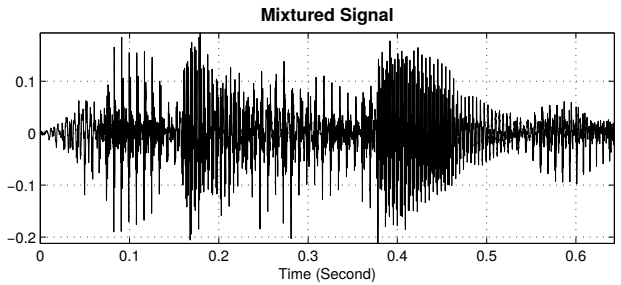
- From both binaural cue (ITD and ILD), binary mask can be estimated using

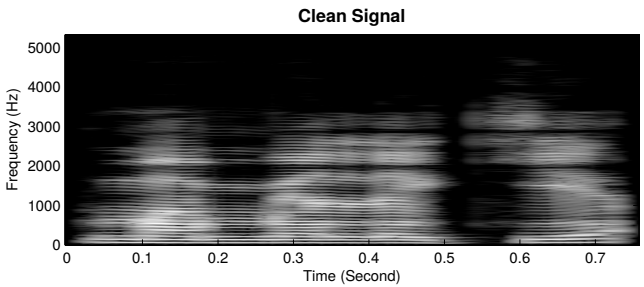
$$BM = \begin{cases} 1, & \text{if } \{(PDE\ RS > 0.5) > (PDE\ RS \leq 0.5)\} \\ 0, & \text{otherwise} \end{cases} \tag{2}$$

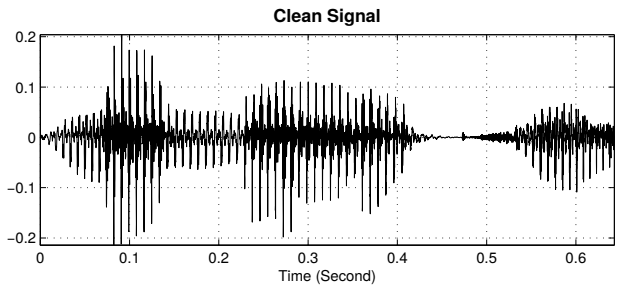
Results

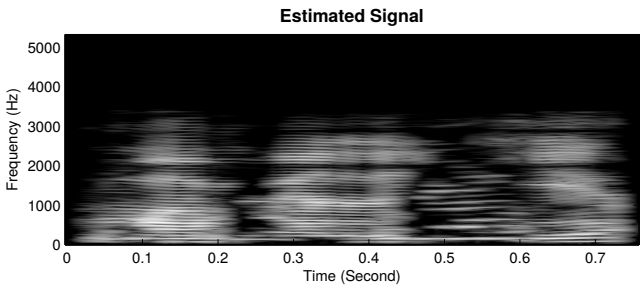
Comparison of spectrum and waveform between mixtured, clean and estimated signal

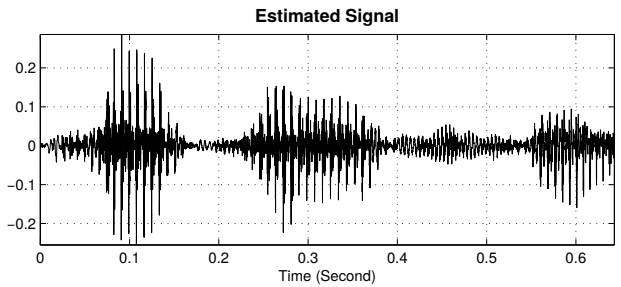










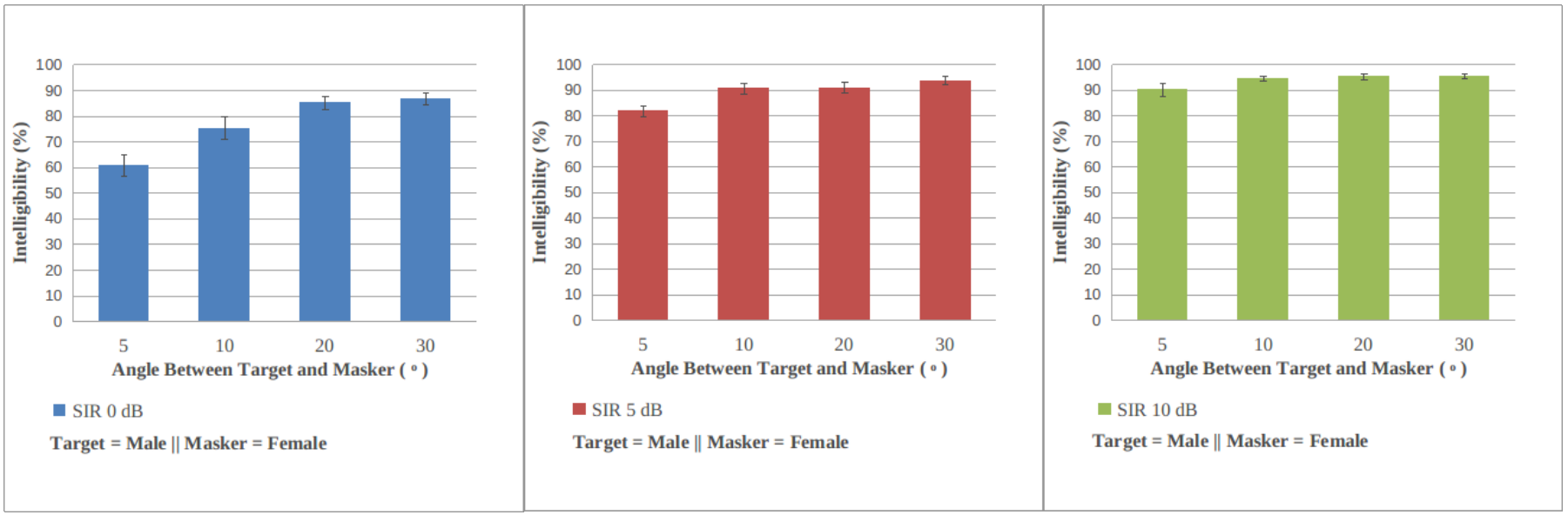


Subjective Evaluation

- Percent correct words
  - Ten respondents: 18 - 23 years old, have good hearing.
  - Testing is done at anechoic room.
  - 520 sentences stimuli, each sentece contain 4 - 8 words
  - The task of the respondent is listen stimuli without repetition and then repeated that stimuli by writing.
  - How many percent word is correct for entire words.

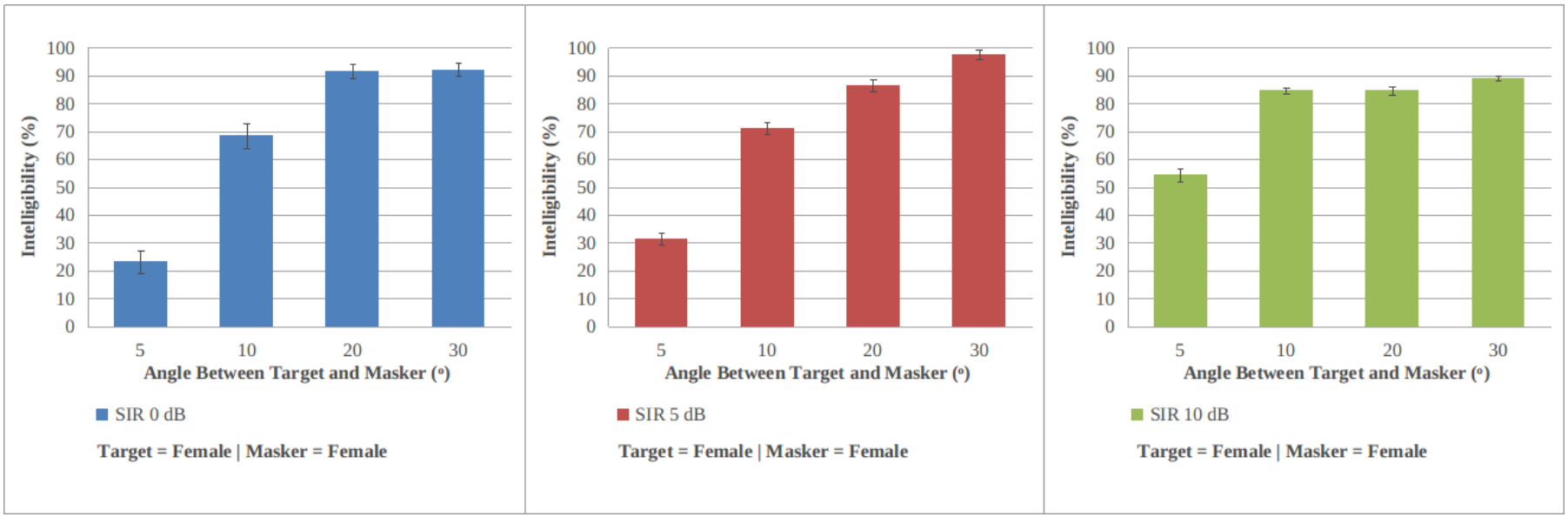
Subjective Evaluation

- Target is male speaker and masker is female speaker



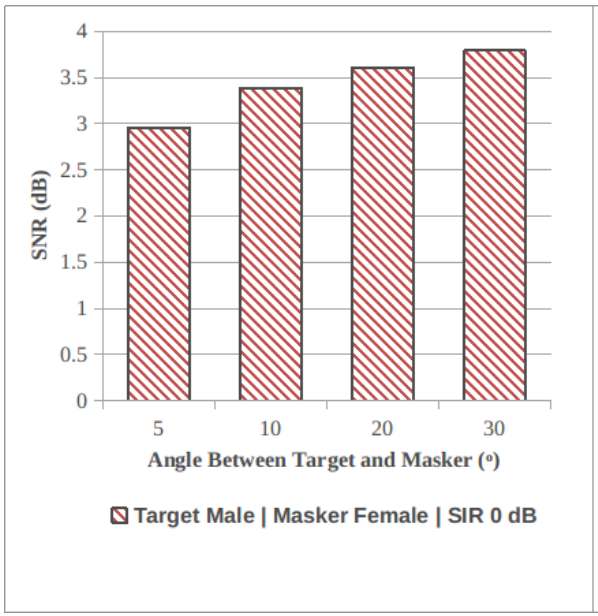
Subjective Evaluation

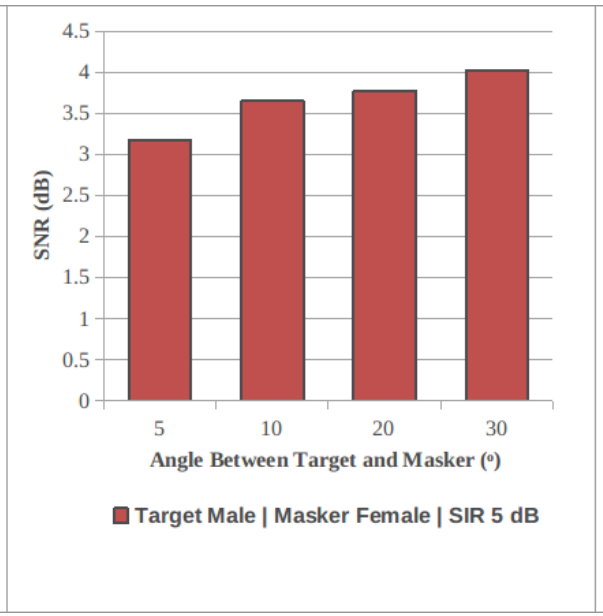
- Target and masker are both female speaker

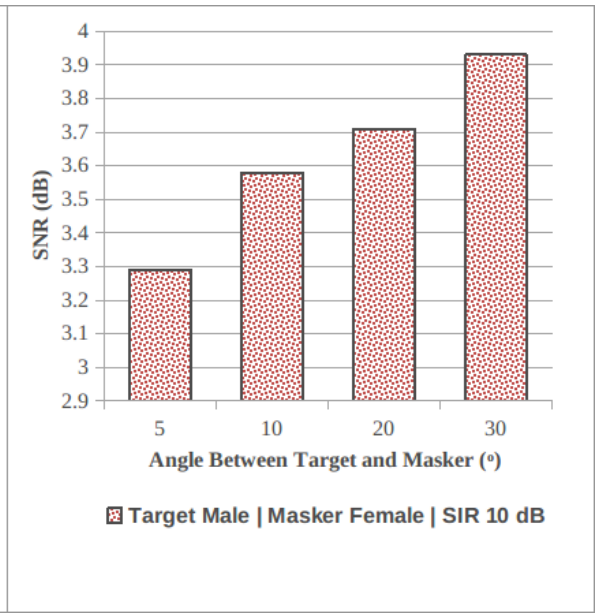


Objective Evaluation

- SNR target is male speaker and masker is female speaker

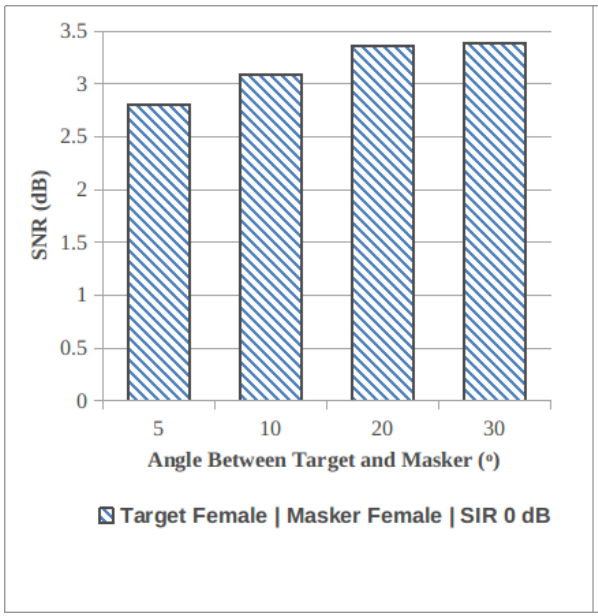


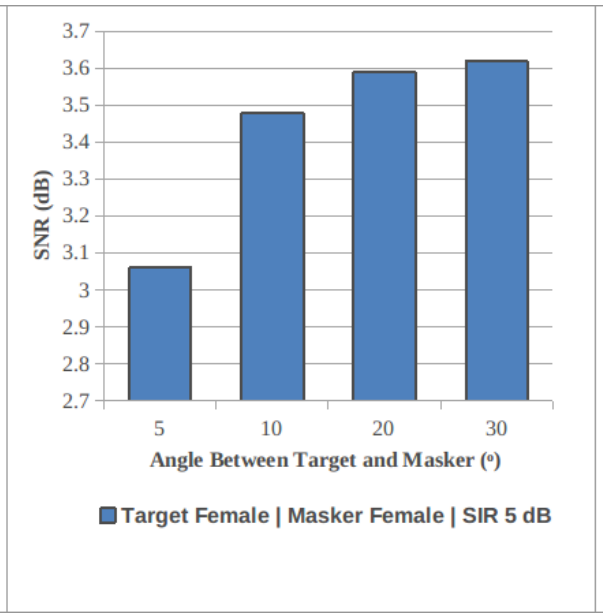


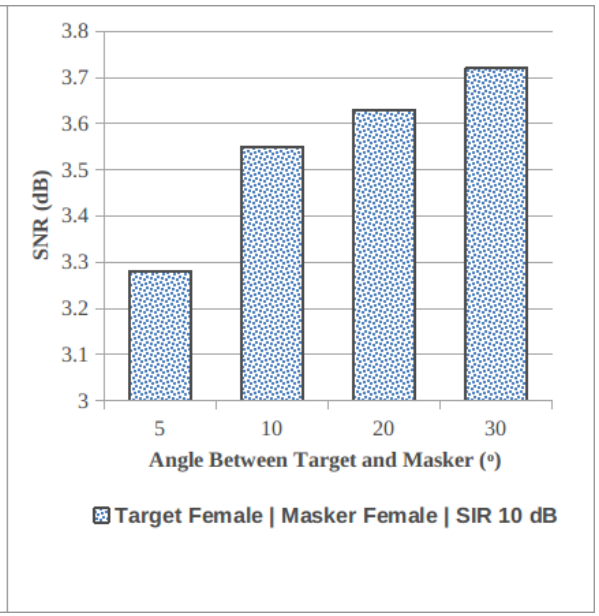


Objective Evaluation

- SNR target and masker are both female speaker







Conclusion

- Sound segregation perform well with speech intelligibility percent correct word 86% and 3dB SNR.
- The larger angle between target and masker then speech intelligibilty of separation result is increase.
- The large SIR between target and masker then speech intelligibilty of separation result is increase.