On Source Signal Segregation based on Binaural Inputs

By Bagus Tris Atmaja
NRP. 2410 201 006

Supervisor: Dr. D. Arifianto, Prof. T. Usagawa

Dept. of Engineering Physics
Faculty of Industrial Technology

Institut Teknologi Sepuluh Nopember

Introduction

- Hearing is one of the most important sense on human.
- The early study on cocktail party problem was elaborated by Helmholtz (1863)
- The early study on computational sound separation was proposed by P. Comon (Elsevier Sig. Proc., 1994)
- High quality sound separation using Information Maximization (Infomax) was proposed by Bell & Sejnowsky (Neural Com., 1995)
- The recent study by Kim. et. al., 2011 proposed a more realistic model on binaural hearing.

Motivation

- Computational sound separation is not easy task. It mimics how the brain works. (Wang, 2006)
- However, current method (Kim, 2011) does not consider the spacing between ears due to spatial aliasing.
- This research proposes FastICA with binary mask.
- The proposed method was evaluated with the other methods on sources segregation by evaluating coherence criterion and PESQ score.

Cocktail Party Phenomena

Problem Statement

- Compare some methods including proposed method on source separation problem for signal enhancement task based on binaural inputs.
- Measure the objective evaluation by means of coherence and PESQ.

Applications :

- Speech Recognition, Telecommunication,
- Hearing Aids, Machine Sound Separation etc.

How to separate sound sources?

- Independent Component Analysis (Bell & Sejnowsky, 1995)
- ICA with binary mask (Pedersen, 2008)
- Binaural Model using phase difference channel weighting (Kim, 2011)
- FastICA (Hyvarinen & Oja, 2000)
- FastICA with binary mask (Proposed Method)

Independent Component Analysis

- Known: ***x*** *m*-sensors and ***s*** *n*-sources

$$x(n) = As(n) + v(n)$$

- ICA can be defined to find ***s*** from ***x*** only
- In this research included noise ***v*** to close real problem, if **W=A**$^{-1}$ and s'(n)=y(n), then

$$y(n) = W\left(x(n) - v(n)\right)$$

- There are many methods to find **W,** in this research ICA was used using max likelihood

$$L(W) = \sum_{n=1}^{N} \sum_{i=1}^{M} \log p_i(\mathbf{W}_i \mathbf{x}(n)) + T \log |\det \mathbf{W}|$$

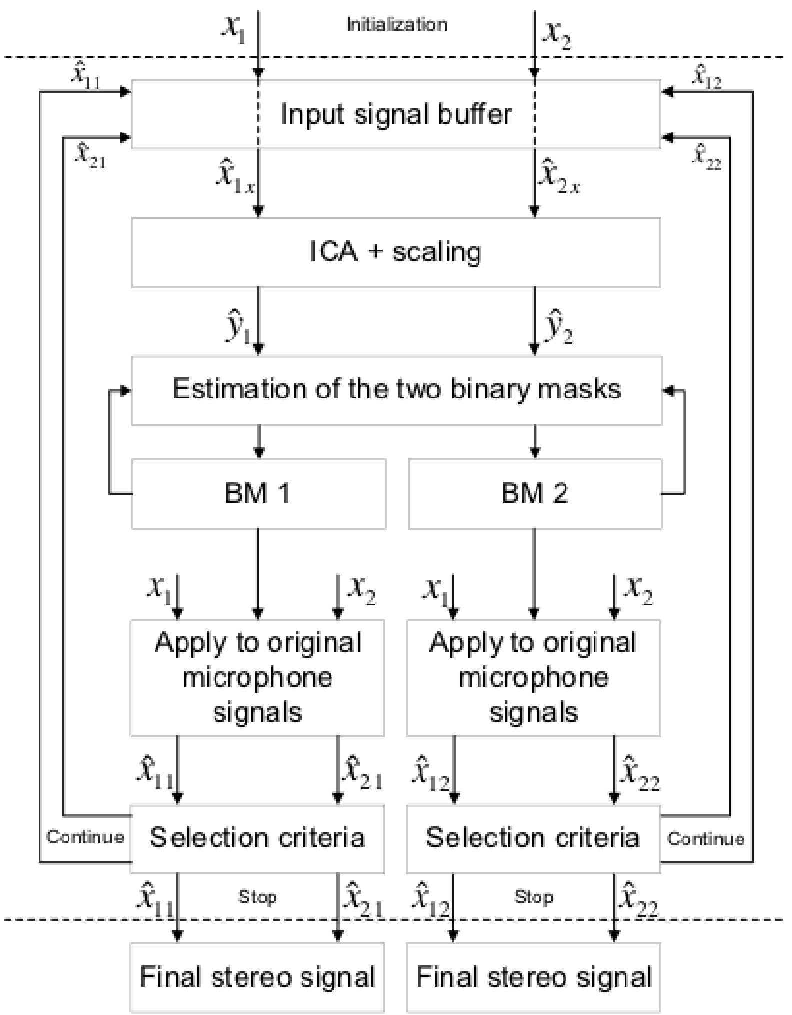
ICA with binary mask (ICABM)

Binary Mask → Applying the mask as a binary weight matrix to the mixture in the T-F domain

Mask → weighting (filtering) the mixture

$$m(n,k) = \begin{cases} 1 & if\, s_1(n,k) - s_2(n,k) > \theta \\ 0 & otherwise \end{cases}$$

Pro-Con: perceptually good, poor coherence



Binaural model using PDCW

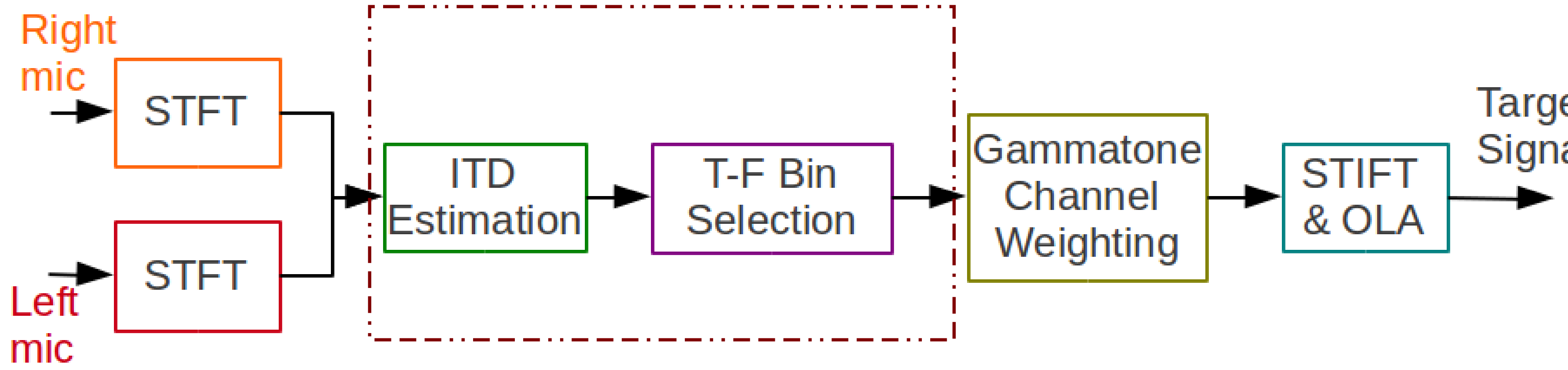
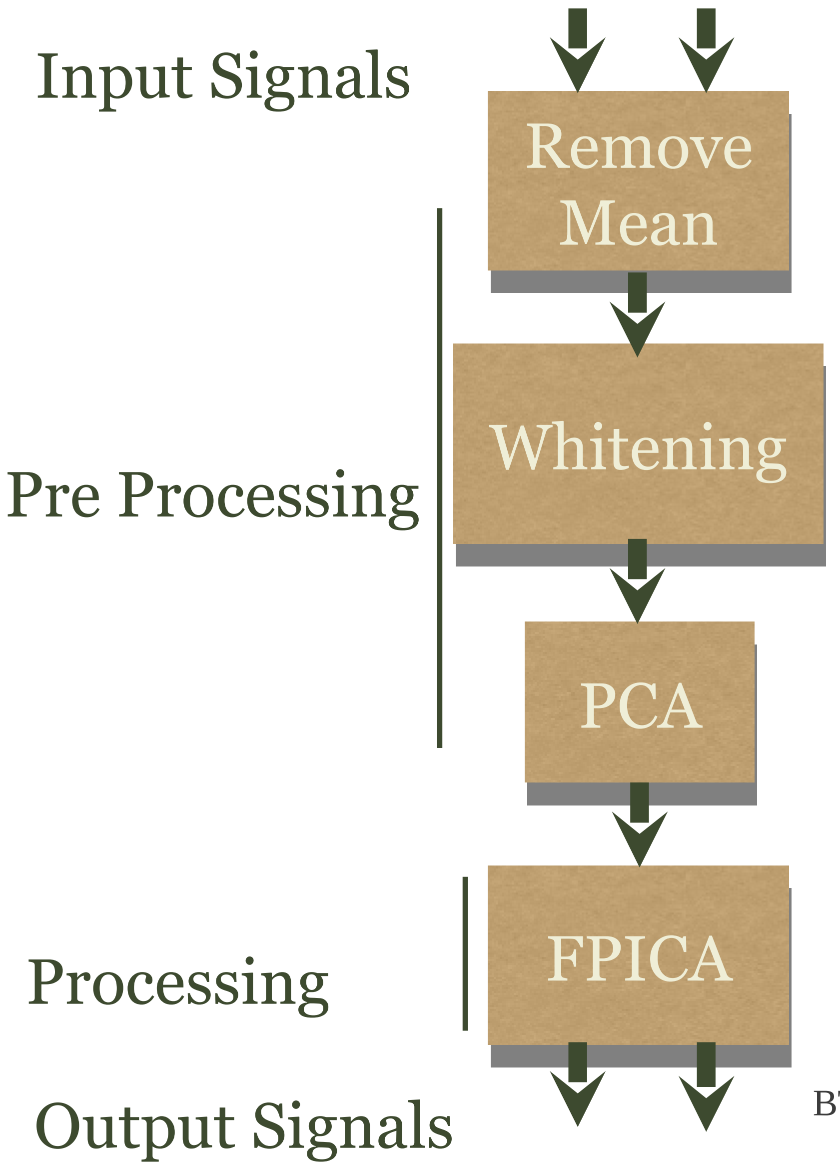


Fig. Block diagram binaural model using PDCW
(Kim et. al., 2011)

FastICA



In FastICA, separation matrix can be obtained by the following formula :

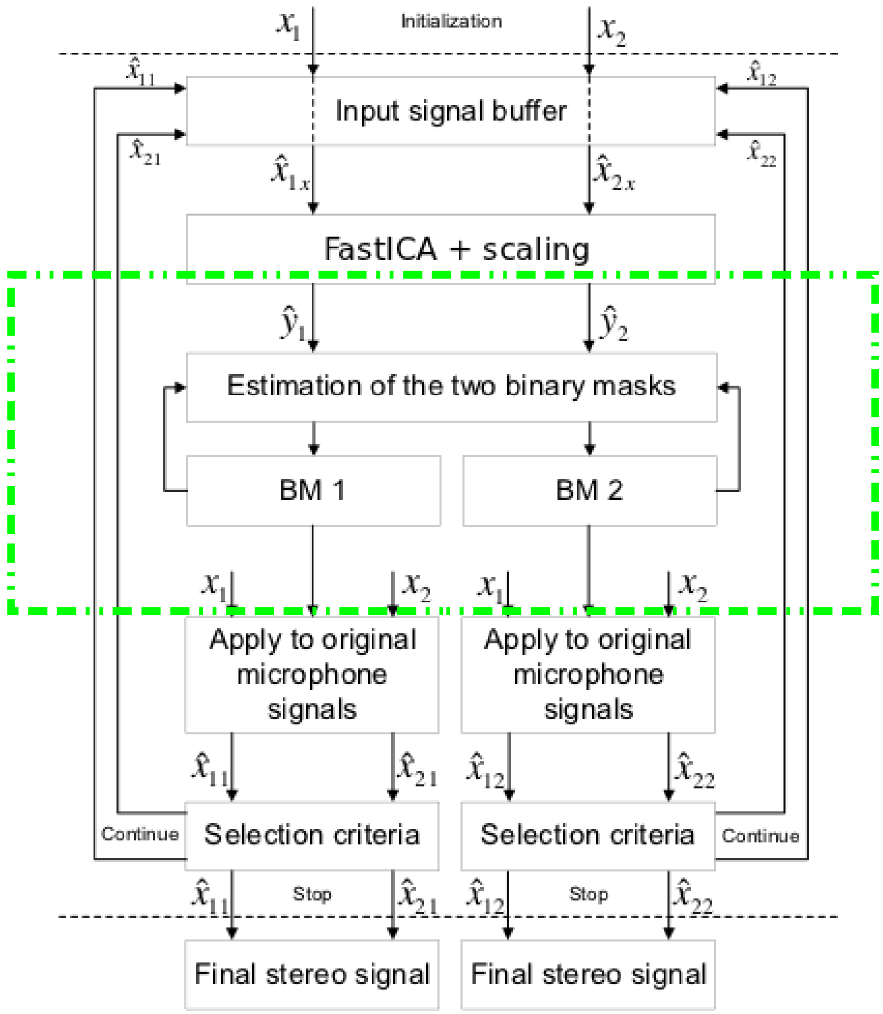
$$w^{+} = E\left\{xg\left(w^{T}x\right)\right\} - E\left\{g^{'}\left(w^{T}x\right)\right\}$$

$$w = \frac{w^{+}}{\left\|w^{+}\right\|}$$

Pro-Con: perceptually poor, good coherence, not yet implemented in binaural hearing

FastICA with binary mask (proposed method)

Binary mask → two tone suppression



Objective Evaluation

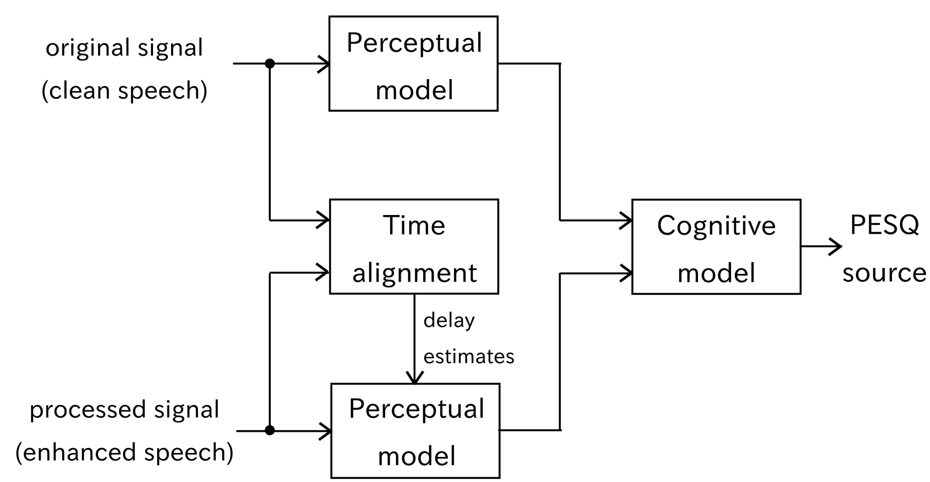
- Coherence Criterion → how well a signal correlated to other signal at each frequency

$$C_{x,y}(k) = \left| \frac{Wxy(k)}{W_{xx}(k)W_{yy}(k)} \right|$$

Value/Score:
0 ~ 1

$$Averaged\,Coherence = \sqrt{\frac{C_{xy}C'_{xy}}{N_C}}$$

- PESQ → Perceptual evaluation of speech quality

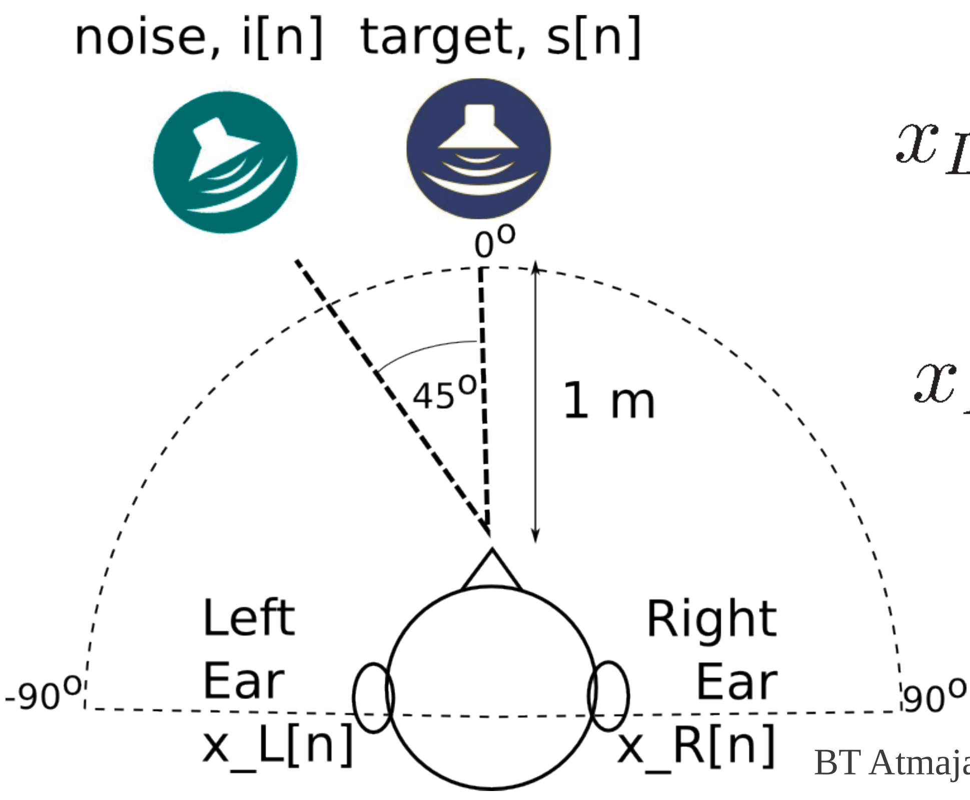


Value/Score:
0.5 ~ 4.5

Simulation

- How to make simulation data?

  Convolution between sound data and HRTF from KEMAR

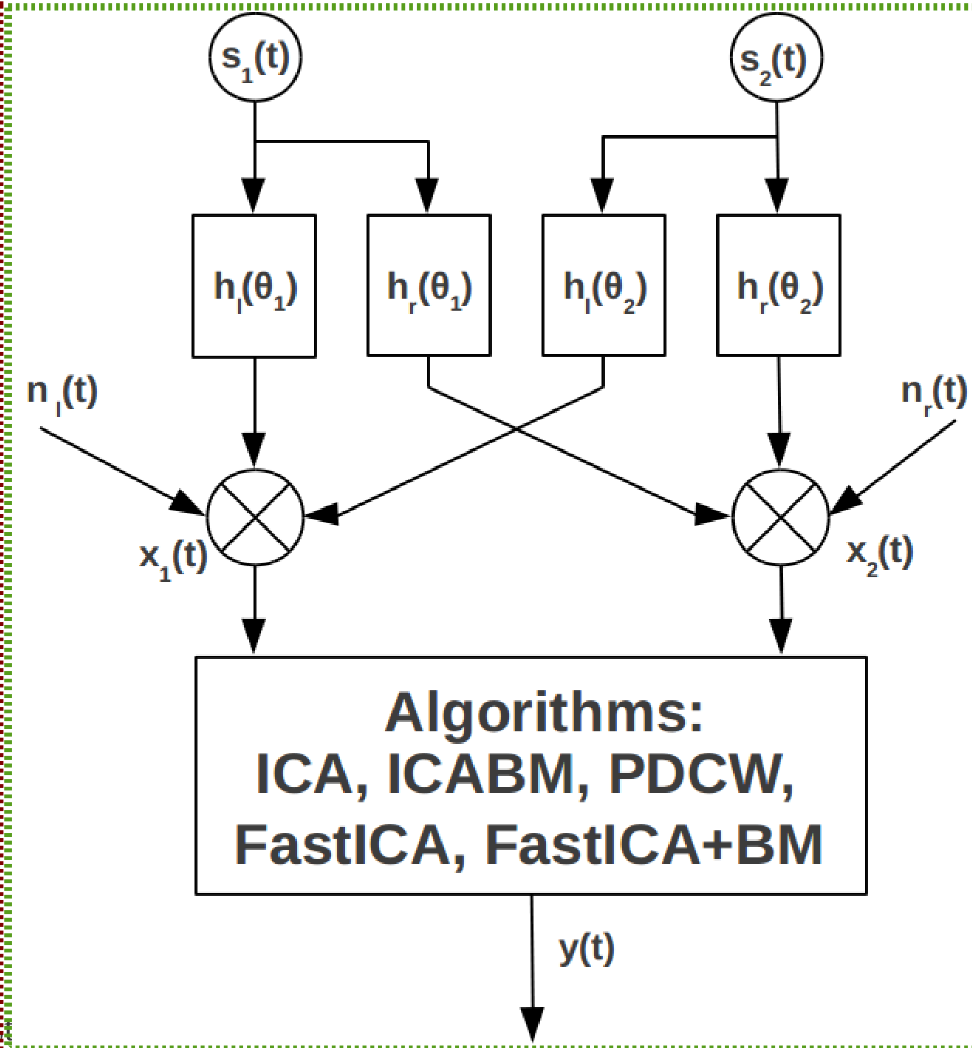


$$x_L(n) = \sum_i h_l(\theta_i) * s_i + n_l(n)$$

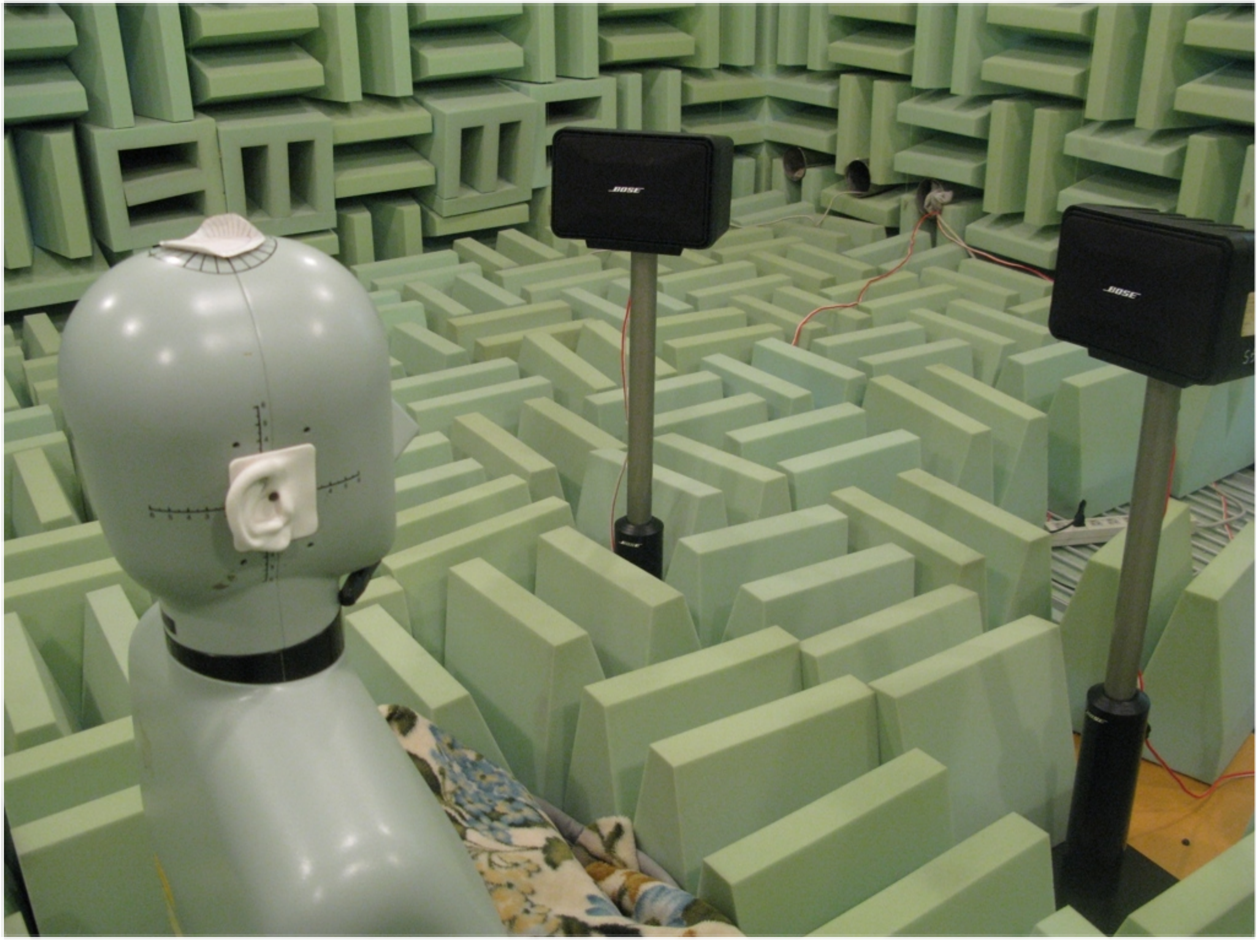
$$x_R(n) = \sum_i h_r(\theta_i) * s_i + n_r(n)$$

Simulation

Variable	Variation
Azimuth	90, 75, 60, 45, 30, 15, 0, -15, -30, -45, -60, -75, -90 (degree)
Elevation	10, 0, -10 (degree)
Fs	48, 44, 22, 16, 8 kHz
HRTF	MIT, Nagoya University
SIR	-20, -10, 0, 10, 20 dB
SNR	0, 5, 10, 15, 20, 25 dB



Experiment – Set Up



Result : Simulation Vs Experiment

(0^0, -45^0)

Target

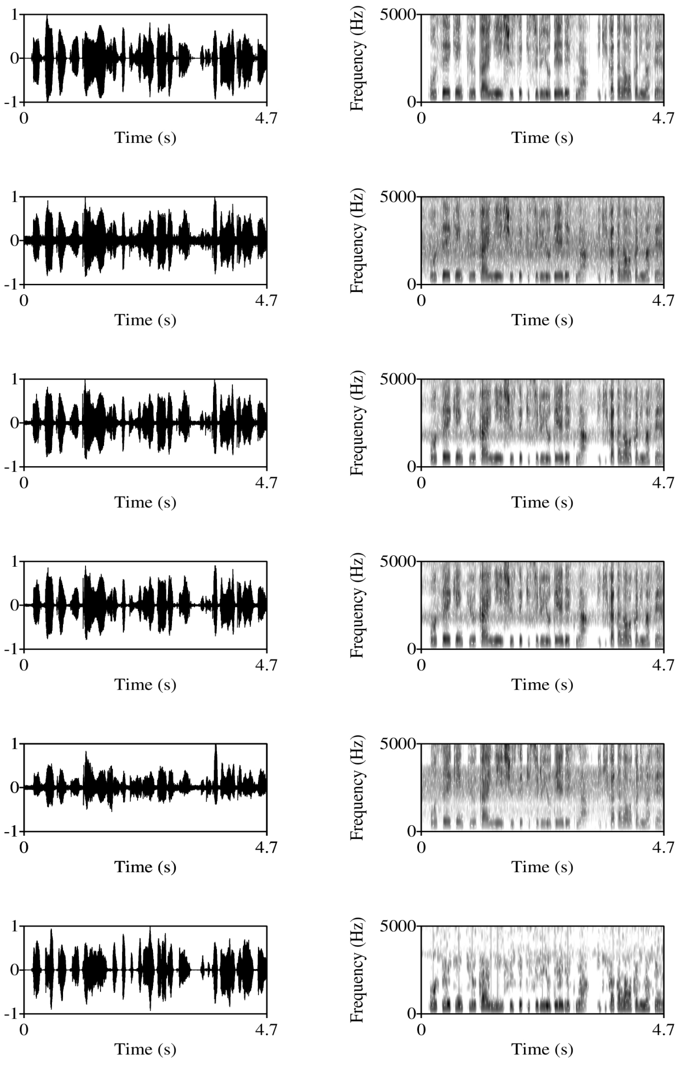
Left

Right

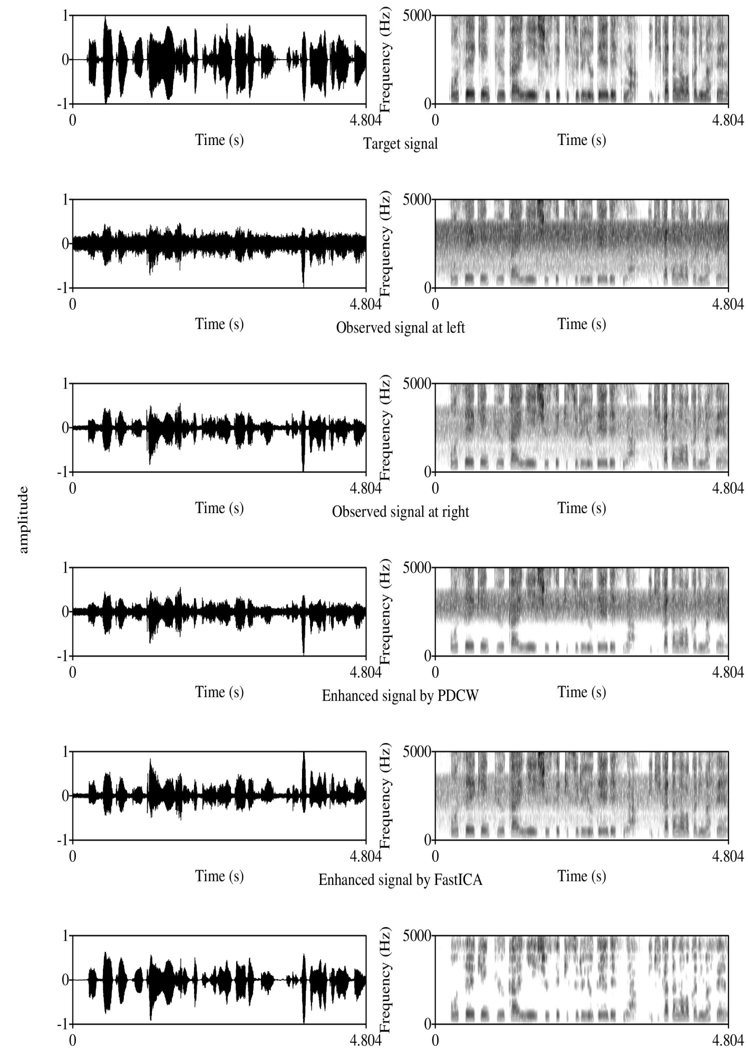
PDCW

FastICA

ICABM



Result of Simulation



Result of Experiment

Result : Simulation Vs Experiment

Methods	Simulation	Experiment
PDCW	0.542	0.28
FastICA	0.669	0.351
ICABM	0.539	0.277

Result : Types of Interference

Female Speech Vs White Noise Interference

Method	Coherence	PESQ
ICA	0.724	1.939
ICABM	0.683	1.945
PDCW	0.578	1.906
FastICA	0.724	1.938
FastCA+BM	0.72	1.905

Result : Types of Interference

Female Speech Vs Male Speech Interference

Method	Coherence	PESQ
ICA	0.735	2.078
ICABM	0.715	2.495
PDCW	0.554	1.562
FastICA	0.734	2.075
FastCABM	0.715	2.457

Result : Types of Interference

Female Speech Vs Male Speech & White Noise Interference

Method	Coherence	PESQ
ICA	0.677	1.748
ICABM	0.656	2.023
PDCW	0.483	1.332
FastICA	0.676	1.748
FastCA+BM	0.676	2.009

Result : Effect of Various SIR

Result based on Coherence Criterion

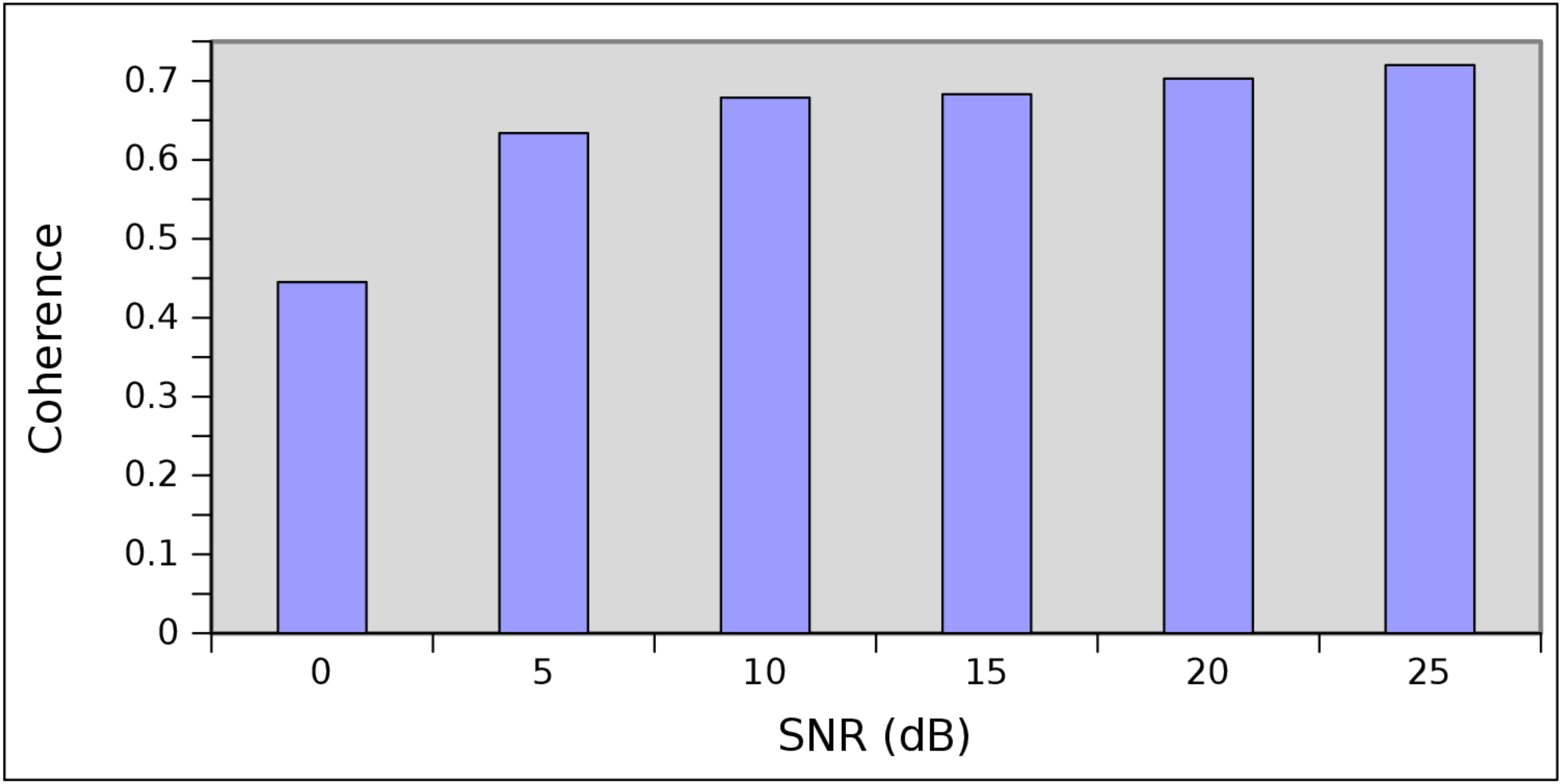
Methods	Signal to Interference Ratio (SIR)				
	-20 dB	-10 dB	0 dB	10 dB	20 dB
ICA	0.598	0.598	0.597	0.633	0.633
ICABM	0.603	0.608	0.394	0.325	0.325
PDCW	0.513	0.500	0.409	0.213	0.213
FastICA	0.598	0.598	0.597	0.632	0.632
FastICABM	0.631	0.609	0.315	0.418	0.471

Result : Effect of Various SIR

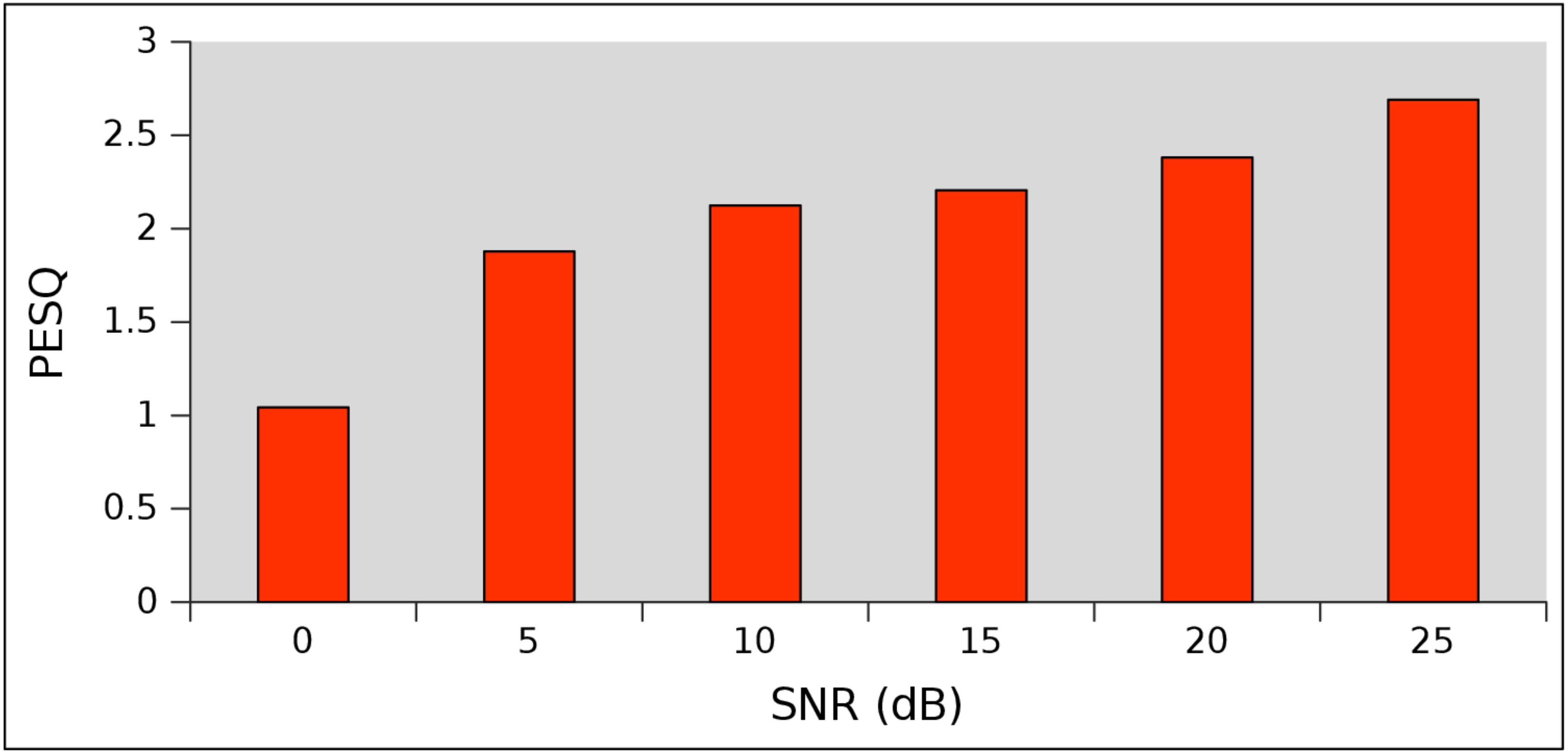
Result based on PESQ Score

Methods	Signal to Interference Ratio (SIR)				
	-20 dB	-10 dB	0 dB	10 dB	20 dB
ICA	1.180	1.180	1.184	1.378	1.378
ICABM	1.185	2.077	1.548	0.692	0.692
PDCW	1.169	1.167	1.190	0.991	0.991
FastICA	1.180	1.180	1.184	1.379	1.379
FastICABM	1.268	2.112	1.282	0.935	1.268

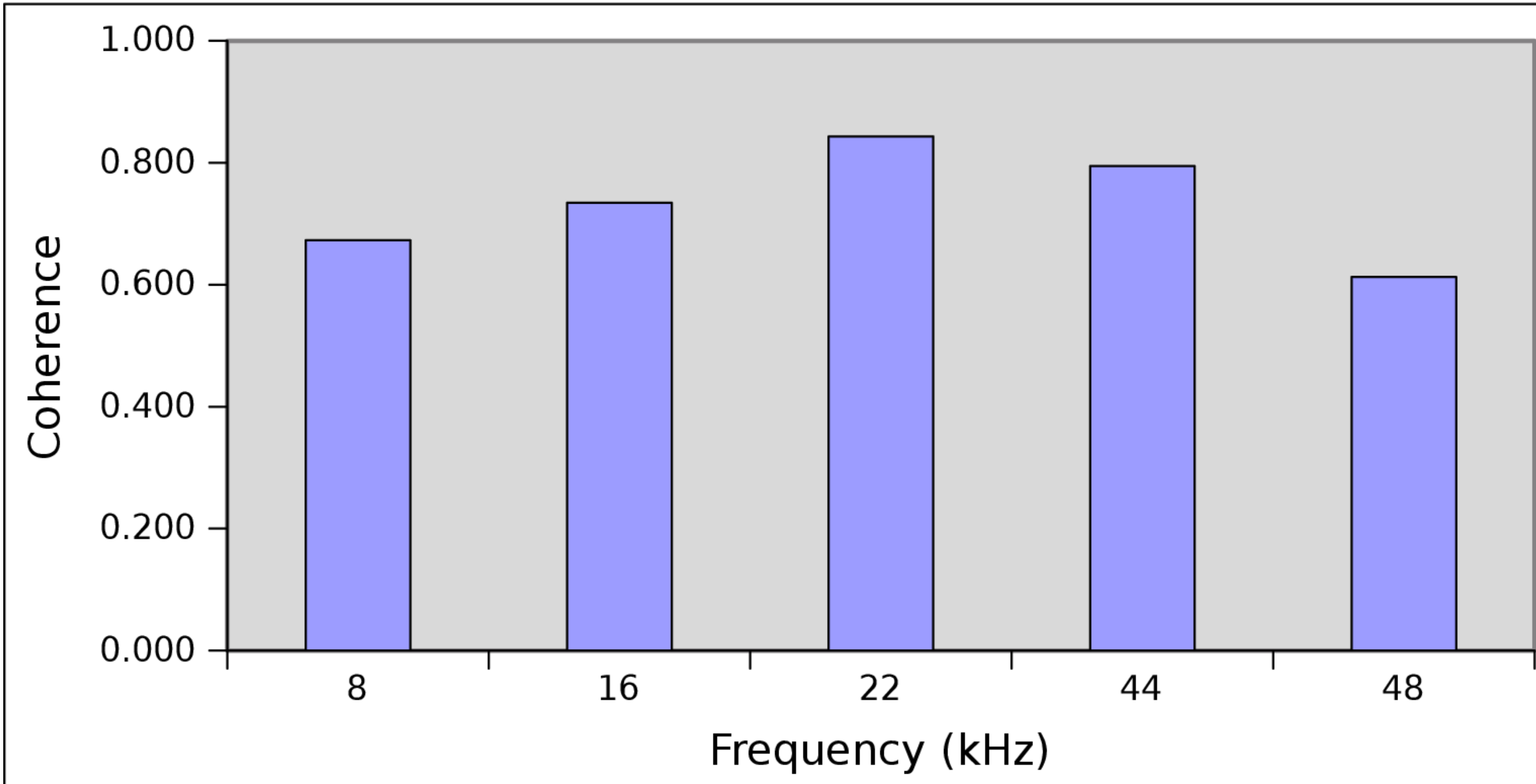
Result : Effect of Various SNR (white noise)



Result : Effect of Various SNR (white noise)



Result : Effect of Various Fs



Conclusions

- Mixed sounds can be separated by using some, in this research we use ICA, ICABM, PDCW, and FastICA. We propose FastICA with binary mask to solve the lack of ICABM and FastICA. This method perform best in different SIR of -20 dB and -10 dB. Those data included noise.
- Coherence criterion and PESQ score were used to evaluate separation result. Coherence was good to extract characteristic of estimated signal while PESQ suitable for perceptual application purpose.

References

- I.-T. R. P.862, “Perceptual evaluation of speech quality (pesq): An objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs,” 2001.
- D. Wang and G. J. Brown, eds., Computatinal Auditory Scene Analysis: Principles, Algorithms and Application. John Wiley and Sons.
- C. Kim, K. Kumar, B. Raj, , and R. M. Stern, “Signal separation for robust speech recognition based on phase difference information obtained in the frequency domain,” INTERSPEECH, pp. 2495–2498, 2009.
- A. Hyvarinen and E. Oja, “Independent component analysis: Algorithms and applications,” Neural Networks, vol. 13(4-5), pp. 411–430, 2000.
- A. Hyvarinen, “Independent component analysis,” vol. 2, pp. 94–128, 2001.
- M. S. Pedersen, D. Wang, J. Larsen, and U. Kjems, “Two-microphone separation of speech mixtures,” IEEE TRANSACTIONS ON NEURAL NETWORKS, vol. 19(3), pp. 475–492, 2008.
- B. T. Atmaja, T. Usagawa, Y. Chisaki, and D. Arifianto, “On performance of sound separation methods including binaural processors,” in Student meeting of Acoustic Society of Japan, Kyushu-Chapter, 2011.
- A. Hyvarinen, “Fast and robust fixed-point algorithms for independent component analysis,” IEEE Trans. on Neural Networks, vol. 10(03), pp. 626–634,1999.
- Etc.

Thank You

ありがとうございました
ご意見あるいはご討論
宜しくお願いします

Institut Teknologi Sepuluh Nopember

Kumamoto University

Machine Sound Separation & Identification