Corporate Development Collaboration Project II

Software Engineer

Signal processing/Machine learning/Acoustic engineering /Optimization problems/Blind Source Separation/Auditory Scene Analysis

Seoul National University of Science & Technology
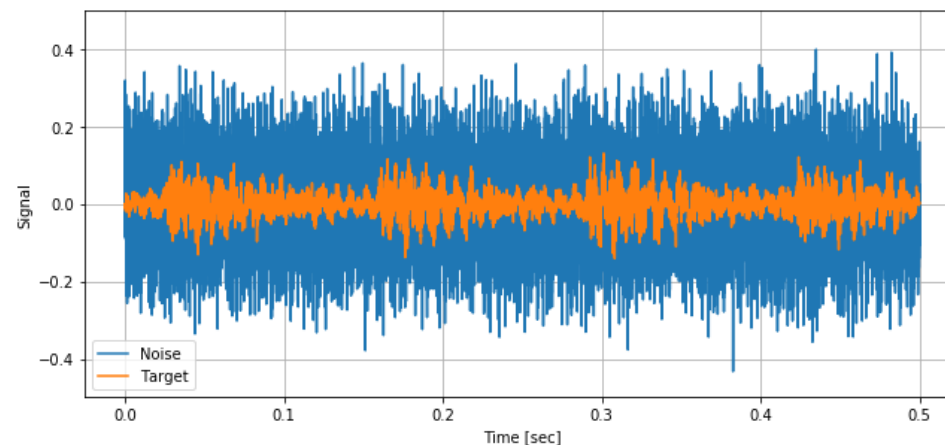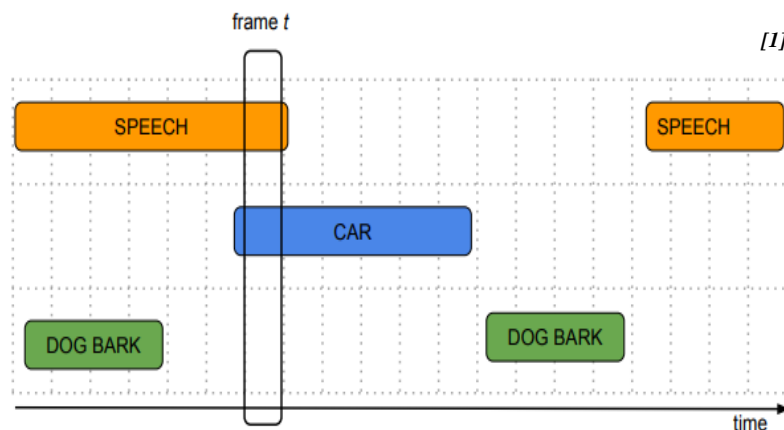Mechanical System Design Engineering
Sonmook Oh(Luke)

Realistic Environments & Complicated Circumstances

※**Which source is the target? & Where is the target source? & What is the target Source?**

"
Hard to recognize/**_localize_**/separate

- ❖ Noisy environment
- ❖ Reverberant room
- ❖ Multi sources
- ❖ Same/Similar direction(depth)

$$When, \boldsymbol{a}(f, \tau_n) \cong \boldsymbol{b}(f, \tau_n) \text{ or } \boldsymbol{a}(f, \tau_n) < \boldsymbol{b}(f, \tau_n) \quad (1)$$

$$\boldsymbol{x}(t, f) = \sum_{n=1}^{N} \boldsymbol{a}(f, \tau_n) S_n(t, f) + \sum_{n=1}^{M} \boldsymbol{b}(f, \tau_n) N_n(t, f) + \boldsymbol{c}(t, f) \quad (2)$$

$$\boldsymbol{a}(f, \tau_n) S_n(t, f) \quad (3)$$

$$\tau_n \quad (4)$$

When,
$$\boldsymbol{x}(t, f) = [x_1(t, f), x_2(t, f)]^T$$
$$\boldsymbol{a}(f, \tau_n) = [1, a_n e^{-2i\pi f \tau_n}]^T$$
$$\boldsymbol{b}(f, \tau_n) = [1, b_n e^{-2i\pi f \tau_n}]^T$$

[1] Sharath Adavanne, Giambattista Parascandolo, Pasi Pertila, Toni Heittola, Tuomas Virtanen(2016). SOUND EVENT DETECTION IN MULTICHANNEL AUDIO USING SPATIAL AND HARMONIC FEATURES. Detection and Classification of Acoustic Scenes and Events 2016
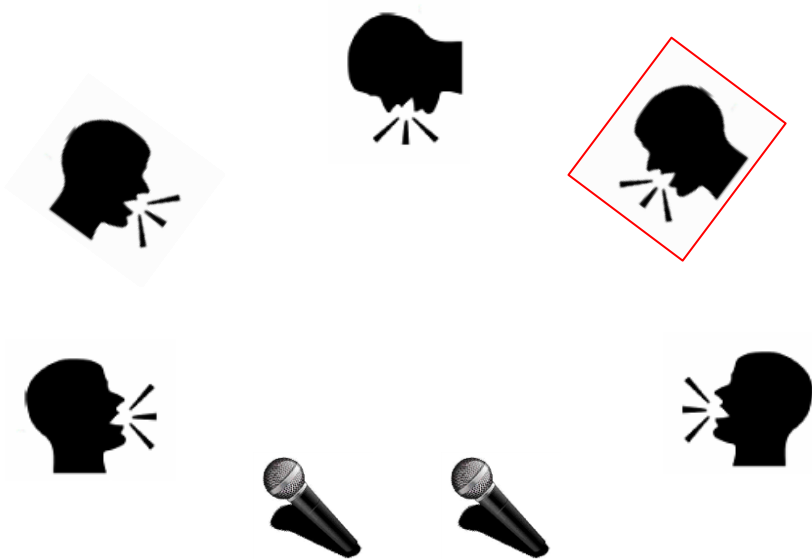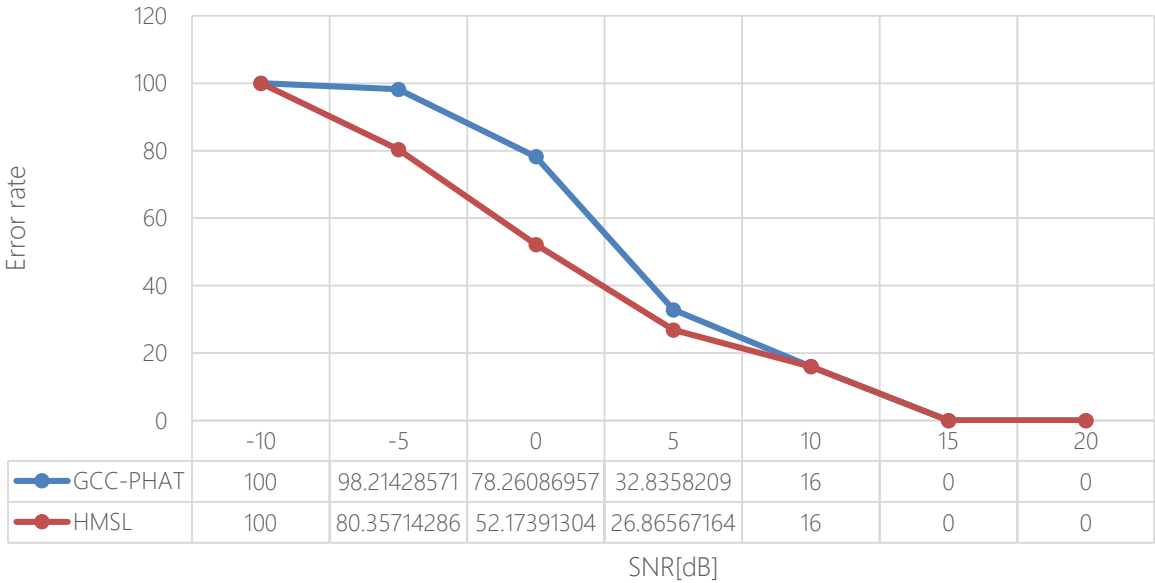
※ Multiple Sound Sources Localization by Automatically Finding Target Source Using NMF with Clustering and HMM

Source Counting and Rough Angle Estimation

Directional Clustering and Binary Masking

Observed matrix (Spectrogram) ≈ **Basis matrix (Spectral patterns)** × **Activation matrix (Time-varying gain)**

Training process — Ex. Musical scale
Sample sounds of target signal $Y_{\text{target}} \approx F$ Q Supervised basis matrix (spectral dictionary)

Separation process
Mixed signal $Y \approx F$ (Fixed) G + H U — Optimize
Target signal — Other signal

Separation by Supervised NMF

Final Direction Estimation

The Degree of Source is 31.00272
The Ratio of Similarity is 0.37153
=====================END=====================

Phoneme model
HMM
$b_1(x_t)$, $b_4(x_t)$, $b_3(x_t)$
Feature
$p(x_t \mid s_i) = \sum_m w_{im} N(x_t; \mu_{im}, \sigma_{im}^2)$
$x_1\,x_2\,x_3\,x_4\,x_5\,x_6\,x_7\,x_8\,x_9 \quad x_t$ — frame
Frame shift
State $|S_1| |S_2| |S_3|$
Speech signals — time

Target Selection by Supervised HMM

[Algorithm]
❖ GCC-Phat (Generalized Cross Correlation-Phase transform)
❖ K-means clustering
❖ NMF (Nonnegative Matrix Factorization)
❖ HMM (Hidden Markov Model)

1) D. Kitamura, H. Saruwatari, H. Kameoka, Y. Takahashi, K. Kondo and S. Nakamura, "Multichannel Signal Separation Combining Directional Clustering and Nonnegative Matrix Factorization with Spectrogram Restoration," in IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 23, no. 4, pp. 654-669, April 2015.

Experiment

※**Experiment Result**

Error rate in white noise

	-10	-5	0	5	10	15	20
GCC-PHAT	100	98.21428571	78.26086957	32.8358209	16	0	0
HMSL	100	80.35714286	52.17391304	26.86567164	16	0	0

SNR[dB]

Error rate in speech noise

	-35	-30	-25	-20	-15	-10	-5	0	5	10	15
GCC-PHAT	100	100	65	52.7174	36.4964	12.5	0	0	0	0	0
HMSL	100	80.7018	77.0833	36.9565	24.0876	10.9375	0	0	0	0	0

SNR[dB]