Signal Enhancement by Single Channel Source Separation

Bagus Tris Atmaja[1], Dhany Arifianto [2]

***Abstract*—Most gadgets and electronics devices are commonly equipped with single microphone only. This is difficult task in source separation world which traditionally required more sensors than sources to achieve better performance. In this paper we evaluated single channel source separation to enhance target signal from interred noise. The method we used is non-negative matrix factorization (NMF) that decompose signal into its components and find the matched signal to target speaker. As objective evaluation, coherence score is used to measure the perceptual similarity from enhanced to original one. It show the extracted has 0.5 of average coherence that shows medium correlation between both signals.**

Keywords— Signal enhancement, source separation, NMF, coherence.

I. Introduction

Signal enhancement has been applied widely in electronic devices, from land line phone to smartphone. Commonly, it used noise suppression method which attempt to remove background noise from target signal. Moreover, the active noise cancellation or active noise control (ANC) required more than one sensor for telephony, one to capture background noise and another to speak via telephone. However, most gadgets at this time, including smartphone, tablet or laptops is only equipped with single microphone. This makes not easy to implement such noise suppression method built in devices.

Source separation, in other side is challenging problem, especially in signal processing communities and applied widely including in audio/acoustic signal. This problem progressively improved from unblind to blind, from supervised to unsupervised, and from overdetermined to underdetermine source separation. In blind source separation, almost no prior information is used to decompose mixed signal likewise in unsupervised source separation. In the other hand, early source separation method used prior information like number of source and sensors, geometry information, time delay and others, this is previously stated unblind or supervised source separation.

Overdetermined source separation is condition where number of sensors more than number of sources. Mathematically, it will give more information compared to underdetermined source separation, where number of sensors less than number of sources. Underdetermined source separation is actively improved, inspired by human binaural ears, and continued developed to reduce the number of sensors. Currently, researchers [1] are investigate to expand underdetermined source separation to single channel source separation as breaking old source separation rule, number of sensor must be greater than sources.

The problem of underdetermined and single channel source separation is about how to decompose small matrix to reconstruct bigger matrix after decomposition. As the signal becomes matrix in computational method, the solution is mathematic matrix manipulation. The widely used matrix decomposition to solve this problem is known as non-negative matrix factorization (NMF).

Virtanen [1] proposed non-negative matrix factorization to solve underdetermined source separation. While their proposal is implementation of NMF for source separation, this paper evaluate NMF for audio signal speech enhancement from single channel recording. This work is the continuity of our previous work [2], while the previous one used two sensors, this research used one sensor only to extract target signal. For evaluation the performance of enhancement by single channel NMF source separation, we used coherence score. Coherence measurement is usually used to measure the quality of processed (enhanced) signal compared to original clean signal or coherence measures how well signal correlated to each other.

II. Method

A. Simulation

Data used in this research can be obtained by simulation and experiment to get mixture sound. For simulation, modeling sound mixture is done by convoluting sound signal with room impulse response which is called convolution reverb. Anechoic room is used to minimize noises and assume no noises in simulation. In this research we evaluated sine sweep sound to measure room impulse response. Measured impulse response then convoluted with source signal to reproduce sound signal in anechoic room.

Room impulse response can be measured by generating sine sweep (s(t)),

$$s(t) * h(t) = r(t) \quad (1)$$

where h(t) is measured impulse response. We use Aliki software to generate sine sweep, measure sine sweep response and room impulse response.

Measured impulse response can be used to produce simulated acoustic mixture by convoluting it with sound sources (s(t)) by the following model,

$$x(t) = \sum_{1}^{n} S_n(t) * h(t) \quad (2)$$

Bagus Tris Atmaja is with Departement of Engineering Physics, Institut Teknologi Sepuluh Nopember, Surabaya 60111, E-mail: bagus@ep.its.ac.id

Dhany Arifianto is with Departement of Engineering Physics, Institut Teknologi Sepuluh Nopember, Surabaya 60111, E-mail: dhany@ep.its.ac.id

where x (t) is simulated sound from convolution reverb. The waveform of this reproduced sound can be seen in figure 2 (left).

Data in the form of sound signal from multiple sources are obtained from simulation and experiment can be obtained through the following steps:

1. Generate sine sweep signals
2. Capture sine sweep response
3. Obtain impulse response
4. Convolute sound source with impulse response

While step 1 to step 3 is done with Aliki software, step 4 is done with Matlab.

The configuration for sweep signal generation is as follows: 48000 Hz of sampling rate, 0.1 seconds fade in frequency range from 20 Hz - 20 KHz, 0.003 fade out. Obtained impulse response with 48 kHz of sampling rate is downsampled to 16 kHz because our focus is audible sound and to match other signals.

B. Experiment

An experiment is conducted to validate simulation data. To evaluate modeled acoustic mixture, sound recording is recorded. Two speeches with same utterances from female and male speech are recorded along with background noises.

Experiment set up consists of two loudspeakers, one microphone and a personal computer. The distance between sound sources (loudspeakers) to microphone is 100 cm, and space between loudspeakers is 75 cm. For recording, we use Audacity running on Linux-based operating system. In experiment, 16 kHz of sampling frequency and 16 bit PCM is used to record sound from two loudspeakers. The two different sound files are transmitted into two loudspeaker as simulated by convolution reverb. The recorded sound waveform is compared to simulation data. The room dimension is 3.5 x 3 x 3 (length x wide x height, in meter).

C. Separation Principle

Source separation is the core of this work after modeling sound mixture. Separation principle consist of the following steps:

1. STFT
2. NMF (Non-negative Matrix Factorization)
3. FILTER (Masking)
4. ISTFT

The flow of those steps can be organized by the following block diagram. That Figure shows the decomposition of signal **x** into $\mathbf{x}_1$, $\mathbf{x}_2$ and $\mathbf{x}_s$. This explains how single channel source separation works.

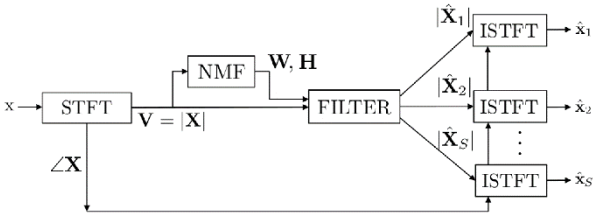

Figure 1 Diagram block of single channel source separation using KL-NMF [3]

III. Results and Discussion

By modeling sound mixture and conducting experiment from single channel, signal enhancement can be done by source separation method. This single channel source separation utilize Kullback-Leibler divergence NMF (KL-NMF). The first result is to analyze the simulation and experiment data (convolution reverb) by comparing both signals as shown in Figure 2. The reproduced signal from model is similar to recorded signal from experiment. However, the similarities between both decreased along with time axis or there is time lag between both.

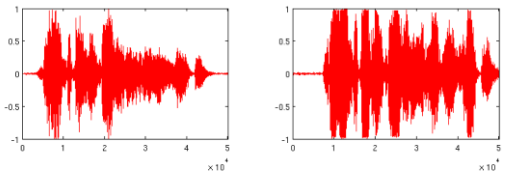

Figure 2 Waveform of simulation (left) and experiment data

Measuring signal similarities also can be done more precisely by cross-correlating signal from convolution reverb to recorded sound. The present of high peak shows correlation between them. The difference between peak and zero x-axis is time lag of both signal i.e. -3786 samples or -0.24 second which signal from experiment appears first from simulation data.

For the extraction, we use matrix decomposition with Non-negative Matrix Factorization (NMF) algorithm. The signals from multiple sources is captured by single sensor and using NMF algorithm, it can be decomposed to extract the first source. Figure 3 shows the spectrogram (plot of time-frequency) original and extracted signal (mixed signal is not shown). It is found that average coherence between both is 0.5 or there is medium correlation between original and extracted signal. The coherence is obtained after aligning signals via cross correlation.

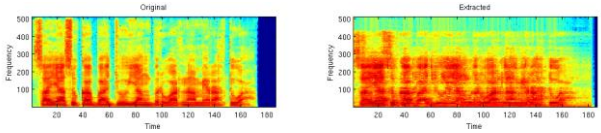

Figure 3. Spectrogram of original (left) and enhanced signal

IV. CONCLUSION

This research evaluate modeling sound mixture in anechoic room compared to experiment data and its separation of target signal from multiple sources. The model shows similarity of experiment data from waveform and measured by cross-correlation and coherence between extracted and original signal. On the experiment data, the amplitude shows bigger than simulated signal while the extracted signal has more sample points (x-axis) compared to original one.

References

[1] Ozerov, Alexey, and Cédric Févotte. "Multichannel nonnegative matrix factorization in convolutive mixtures for audio source separation." Audio, Speech, and Language Processing, IEEE Transactions on 18.3 (2010): 550-563

[2] B.T. Atmaja, D. Arifianto, Y. Chisaki, T. Usagawa. "Signal Enhancement by Using Sound Separation Methods Based On Binaural Inputs." Basic Science Int. Conf., 1-5. 1 (3), 2012.

[3] Nicholas Bryan, Dennis Sun, and Eunjoon Cho, "Single-Channel Source Separation Tutorial Mini-Series" [online] 2015.