sensors

MDPI

Article

Parametric Estimations Based on Homomorphic Deconvolution for Time of Flight in Sound Source Localization System

Yeonseok Park [1], Anthony Choi [2] and Keonwook Kim [1,*]

[1] Division of Electronics & Electrical Engineering, Dongguk University-Seoul, Seoul 04620, Korea; dustjrdk@dongguk.edu

[2] Department of Electrical & Computer Engineering, Mercer University, 1501 Mercer University Drive, Macon, GA 31207, USA; choi_ta@mercer.edu

* Correspondence: kwkim@dongguk.edu; Tel.: +82-2-2260-3334

Received: 20 December 2019; Accepted: 6 February 2020; Published: 10 February 2020

Abstract: Vehicle-mounted sound source localization systems provide comprehensive information to improve driving conditions by monitoring the surroundings. The three-dimensional structure of vehicles hinders the omnidirectional sound localization system because of the long and uneven propagation. In the received signal, the flight times between microphones delivers the essential information to locate the sound source. This paper proposes a novel method to design a sound localization system based on the single analog microphone network. This article involves the flight time estimation for two microphones with non-parametric homomorphic deconvolution. The parametric methods are also suggested with Yule-walker, Prony, and Steiglitz-McBride algorithm to derive the coefficient values of the propagation model for flight time estimation. The non-parametric and Steiglitz-McBride method demonstrated significantly low bias and variance for 20 or higher ensemble average length. The Yule-walker and Prony algorithms showed gradually improved statistical performance for increased ensemble average length. Hence, the non-parametric and parametric homomorphic deconvolution well represent the flight time information. The derived non-parametric and parametric output with distinct length will serve as the featured information for a complete localization system based on machine learning or deep learning in future works.

Keywords: sound source localization; time of flight; homomorphic deconvolution; cepstrum; Yule-walker; Prony; Steiglitz-McBride; vehicle

1. Introduction

The sound source localization (SSL) system estimates the angle of arrival (AoA) for an acoustic source based on the received signal. SSL approaches are extensive, ranging from the physical rigid structures to the machine learning algorithms to design the spatial filter. The prevalent methods utilize the phase differences between the receivers for beamforming [1] which can be employed for such various applications as underwater warfare systems. Since the beamforming performance is proportional to the receiver quantity, numerous microphones are required for high precision AoA estimation. The beamforming constraints are challenged by the biomimetics methods. Humans can accurately localize sound sources in three-dimensional (3D) space by using the binaural correlation and structure profile. Various monaural and binaural sound localization systems have been proposed to imitate human-like hearing system [2–7]. Currently, researches are being conducted to understand the propagation on the dedicated structure with single or dual receivers in nature science and practical engineering.

Sound is the complementary to vision for navigating mobile objects. The acoustic information enhances the system safety since the sound can propagate over the obstacle by diffraction property to deliver the situation over the non-line-of-sight (NLOS) locations. Presently, the human and vehicle are cooperated to drive the transport sorely based on the vision information. Hence, indirect imminent endangerment cannot be realized until the human has the visual contact. For example, a car with emergency braking cannot be perceived by the indirect position observers. The squeal sound provides the situation. However, the system cannot recognize the direction to activate the pre-emptive safety devices which reduce or remove the impact of the secondary collisions. The sound information can be used for improving the safety of future autonomous transport system.

The driver barely obtains the acoustic information since the vehicle structure debilitates the propagation by airtight cabin. The acoustic perception on sound source and arrival direction are both required to understand the situation. The SSL system mounted on a vehicle could provide the comprehensive information to improve the driving conditions by monitoring the surroundings including the NLOS observation. The conventional SSL approaches recently employed for transport are as follows. The moving vehicle presences are identified by sound localization based on the arrival time difference between the microphones [8]. A sensing technique to localize an approaching vehicle is proposed by an acoustic cue from the spatial-temporal gradient method [9]. For the sequential movement events of vehicles, robust direction-of-arrival estimation is realized by the incoherent signal-subspace method based on a small microphone array [10].

The 3D structure of vehicle with size causes the problems to realize the omnidirectional SSL system because of the long and uneven propagation. In addition, the aerodynamic profile of the vehicle prevents to install the monaural and binaural localizer which use the pinna-like configuration. This paper proposes the novel method to design the SSL system for transport based on the analog microphone network. The single channel signal utilizes the deconvolution process and machine learning (or deep learning) for SSL. The comprehensive functional diagram is illustrated in Figure 1. The mixed signal on the microphone network bus contains the various time delay information to represent the time of flight (ToF) between microphones. The implicit ToF is estimated by the homomorphic deconvolution (HD) algorithm, which is established through the homomorphic system [11,12]. In order to use the ToF on machine learning stage, the HD algorithm can be modified for parametric methods as extracted features. The coefficients of the parametric model as well as distribution of the non-parametric method are essential clues to derive the AoA information as shown in Figure 2. The suggested vehicle SSL system is extensive to describe the complete proposition in single article. Therefore, this paper only describes the ToF estimation for two microphones with non-parametric and parametric HD algorithms. Future articles will include the SSL performance with machine learning based on the findings in this paper.

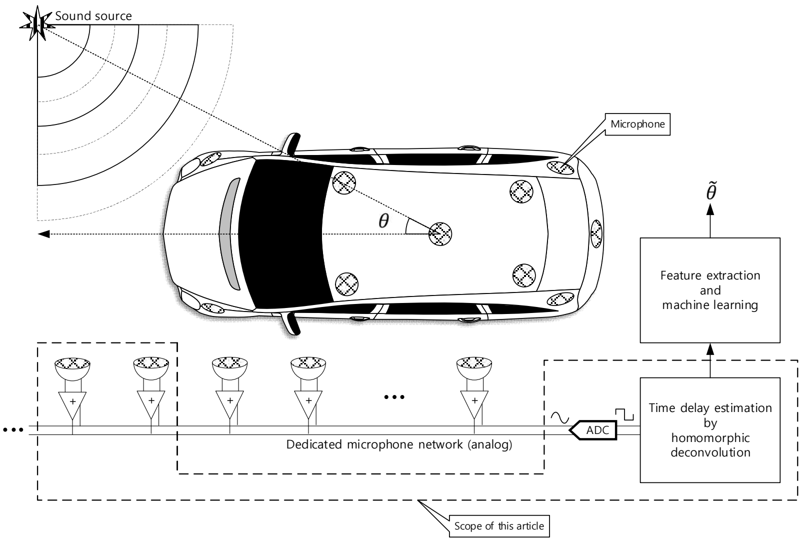

Sound source
Microphone
θ
$\tilde{\theta}$
Feature extraction and machine learning
Time delay estimation by homomorphic deconvolution
ADC
Dedicated microphone network (analog)
Scope of this article

Figure 1. Functional diagram of overall SSL system for transport. This paper presents the parametric and non-parametric homomorphic deconvolution for two-microphones situation. The car shape is illustrated by Nichkov Alexey [13].

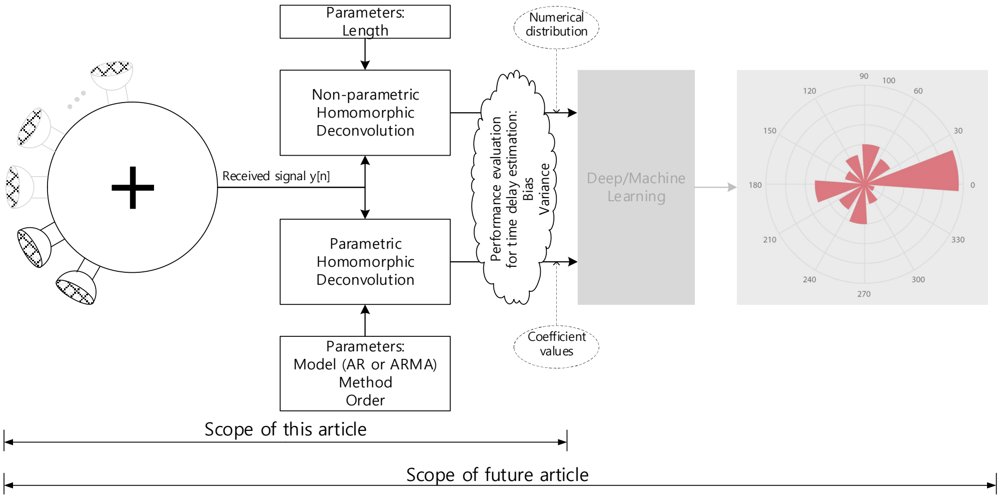


Figure 2. Overall methodology for vehicle sound source localization system. Note that the gray area will be investigated in the future article.

Numerous investigations have been conducted for SSL system and are summarized below. Dong et al. [14] combine with the arrivals of multi-sensor and inversion of the real-time average for iterative method based on the time differences to locate source coordinates with clear arrivals. SSL monitors the development of the crack in real time for the complex structure by searching the fastest wave path over the structure with irregular spaces in order to improve location accuracy [15]. An analytical SSL method was developed for a simplified algorithm and conditions without an iterative algorithm, without pre-measured velocity, without initial value, and without square root operations [16]. The SSL system based on the embedded microphone array for outdoor environments is presented for unmanned aerial vehicle application [17]. The 3D sound source locations are estimated by generating direct and reflection acoustic rays based on ray tracing along with Monte Carlo localization algorithm [18]. Speech localization with microphone arrays is devised from the sparse Bayesian learning method by using a hierarchical two-level inference [19]. An extensive review and classification of sound source localization techniques and popular tracking methodologies is organized by Rascon and Meza [20].

The recent studies initiated to employ the machine learning methods for SSL systems are as follows. Observe that this paper focuses on the ToF estimation. The SSL system based on the ToF and machine learning which will be realized in the future article as shown in Figure 2. The reverberation robust feature extraction method for machine leaning is proposed by Li and Chen [21] for sound source localization based on sound intensity estimation over a microphone array. The sound localization system is designed from deep neural networks using the frequency domain feature based on the integrating directional information and directional activator [22]. For near-field sound localization, a weighted minimum variance distortionless response algorithm is presented with a machine learning for the steered response power computation [23]. The multiple source localization on underwater horizontal arrays is presented by using deep neural networks with two-stage architecture for the directions and ranges in shallow water [24].

This paper achieves the aims proposed by the authors' previous SSL publications. The fundamental frequencies induced by the asymmetric horizontal pyramidal horns were arranged for the far-field monaural localization system by utilizing cepstral parameters [25]. The small-profile near-field monaural system was realized by the asymmetric vertical cylindrical pipes around a single microphone [26]. The reflective monaural localization system [27] placed the multiple plates for the direction-wise time delay to be estimated by homomorphic deconvolution. Other localization works on the subject by the authors are also related to and expanded during the research, such as azimuthal movement detection based on binaural architecture [28] and a target localization algorithm over a distributed acoustic sensor network [29]. Observe that the experiments are performed and evaluated within an identical anechoic chamber [30] to that used in the previous works.

2. Methodology

The ToF estimation for the two microphones can be described as the linear time-invariant (LTI) system with proper impulse response. The time distance between the receivers is contained over the two Kronecker delta functions (or delta functions) in the propagation impulse response. Without the noise condition, the sound source travels through the medium and last arrivals for microphones inscribe the time distance on the delta functions. The receiver structure presents the ToF based on the arrival angle; however, we assume that the ToF is given in this paper. As part of the sound localization, the system requires to obtain the ToF information with or without the sound source restoration. The received signal is the convolution sum (or integral) between the sound source and ToF impulse response. Therefore, the goal of this paper is the deconvolution problem between the two signals.

In order to realize the separation, the propagation function should be understood with parameters. The performance of the deconvolution depends on the signal and propagation condition since the deconvolution projects the problem domain into the maximally separable space. In this paper, the HD utilizes the homomorphic systems in cascade to remove the sound source and to derive the propagation function. The forward conversion of the homomorphic system is implemented by the real cepstrum to compress the geometric series from poles and zeros of the received signal model. In time domain, the propagation function shows the echo property as wide stride structure which provides relatively slow decreasing rate distribution after the real cepstrum. The distinct compression rates in cepstrum domain deliver the powerful tool to divide the signal and propagation. The simple window known as frequency-invariant linear filtering (FILF) performs the separation. The backward conversion of the homomorphic system by inverse cepstrum finally derives the propagation function for ToF. The real cepstrum extensively uses the discrete Fourier transform (DFT) or fast Fourier transform (FFT) for non-parametric estimation. The parametric ToF estimation is realized by the propagation function model with parametric estimation methods such as Yule-walker, Prony, and Steiglitz-McBride in last stage of HD. Note that the non-parametric technique produces the numerical distribution and parametric method estimates model coefficients.

Figure 3 demonstrates the overall signal propagation and estimation procedure. The wide or narrow band signal $x[n]$ is delivered to the both microphone with dT_s time difference. The T_s is the sampling period and the corresponding propagation function $h[n]$ is the $\delta[n] + \alpha\delta[n-d]$ with attenuation rate α. The received signal $y[n]$ is the convolution sum (or integral) as $x[n] * h[n]$. The first FFT and IFFT pair with absolute logarithm presents the real cepstrum to divide the signal and propagation distribution. The window function $w[n]$ extracts the propagation function. The other FFT and IFFT pair with exponential function reestablishes and estimates the non-parametric propagation impulse response $\widetilde{h}[n]$. The parametric estimation utilizes the regressive model to define the peaky distribution in propagation function as delta function. Originally, the estimation methods as Yule-walker, Prony, and Steiglitz-McBride is devised to present the spectral property of the signal. The parametric methods are applied to the estimation in the reverse direction. Therefore, the techniques should be extended by extension of signal and coefficient space into the complex number. The reverse usage will be explained in the following parametric section.

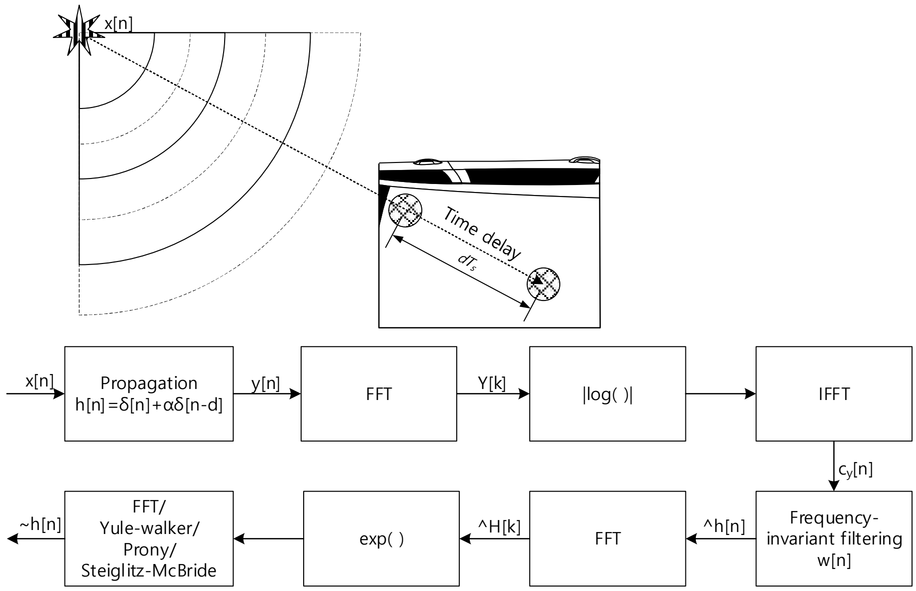


Figure 3. System architecture of time delay estimation based on the parametric and non-parametric homomorphic deconvolution.

2.1. Homomorphic Deconvolution

The HD [31,32] in this paper consists of forward and backward real cepstrum to perform the deconvolution. The complex cepstrum [31,32] is shown in below. The $Y\left(e^{j\omega}\right)$ is the discrete-time Fourier transform (DTFT) of the received signal. The inverse DTFT of the complex logarithm on $Y\left(e^{j\omega}\right)$ is the complex cepstrum. The actual complex cepstrum is realized by the DFT which presents the principal values for phase. The additional procedure to unwarp the phase is necessary for proper cepstrum output. However, the complex cepstrum provides the capability to separate the sound source and propagation function approximately intact.

$$\hat{y}[n] = \tfrac{1}{2\pi}\int_{-\pi}^{\pi}\log\left(Y\left(e^{j\omega}\right)\right)e^{j\omega n}d\omega \tag{1}$$

The real cepstrum [31,32] employs the absolute logarithm before the inverse DTFT as below. Because of the omitted phase information, the real cepstrum demonstrates the limited performance to obtain the sound source from the deconvolution. The deconvolution system estimates the ToF based on the propagation function. Hence, this paper exploits the real cepstrum for HD.

$$c_y[n] = \frac{1}{2\pi}\int_{-\pi}^{\pi}\log\left|Y\left(e^{j\omega}\right)\right|e^{j\omega n}d\omega \tag{2}$$

The relationship between the complex and real cepstrum is below. Due to the absolute operation on the DTFT, the real cepstrum represents the even function property.

$$c_y[n] = \frac{\hat{y}[n] + \hat{y}^*[-n]}{2} \quad \because \left|Y\left(e^{j\omega}\right)\right| = \sqrt{Y\left(e^{j\omega}\right)Y^*\left(e^{j\omega}\right)} \tag{3}$$

The real cepstrum casts the sound source and propagation function on the distinct locations; therefore, the proper window can partition the propagation function from the other signals. Note that the sound source cannot be recovered from the real cepstrum in most of conditions. The n domain window $w[n]$ is applied as below for FILF.

$$\hat{h}[n] = c_y[n]w[n] \tag{4}$$

The inverse cepstrum is implemented by the inverse DTFT of the exponential $\hat{H}\left(e^{j\omega}\right)$ which is the $\widetilde{h}[n]$ DTFT. The derived $\widetilde{h}[n]$ corresponds to the propagation function; hence, the peak location except zero position determines the ToF. Observe that the cepstrum engages and disengages the domain by using the logarithm and exponential function with entrance and exit functions as Fourier and Z-transform. The analysis can be performed over the DTFT or z domain.

$$\widetilde{h}[n] = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{\hat{H}(e^{j\omega})} e^{j\omega n} d\omega \tag{5}$$

The numerical realization by computer utilizes the DFT and IDFT for real cepstrum. The periodicity is developed as below due to the DFT and IDFT property. Also, the symmetricity is produced because of the absolute operation for logarithm. Note that N is the DFT length.

$$\widetilde{c}_y[n] = \sum_{k=-\infty}^{\infty} c_y[n+kN]; \quad \widetilde{c}_y[n] = \widetilde{c}_y^{*}[N-n] \tag{6}$$

The real cepstrum based on the DTFT denotes the infinite duration according to the inherited property. Therefore, the periodic $\widetilde{c}_y[n]$ from the DFT is the time-aliased form of the $c_y[n]$. The increased DFT length N is recommended to avoid the data corruption due to the periodicity. The symmetricity plays the important role in the minimum phase system realization. The appropriate window $w[n]$ should consider the $\widetilde{c}_y[n]$ symmetricity for deconvolution based on the minimum- and maximum-phased decomposition. Appendix A delivers detailed analysis of the non-parametric HD with sound source and propagation model and example.

2.2. *Model Based Parametric Estimations*

The parametric estimation is involved in the last stage of the inverse cepstrum. The estimation output is the coefficient values of desired signal model. The evaluation of the signal model with estimated coefficients corresponds to the propagation impulse response. The HD provides the deconvolution work and the parametric estimation presents the domain transfer. The conventional parametric spectral estimation algorithms require the frequency domain signal model as rational function. Therefore, following section explains the estimation algorithms for forward transformation. The parametric HD exercises the estimation algorithm for backward transformation that will be described in the last part of below section. The conventional signal models are below.

- Autoregressive (AR) model

$$y[n] + a_1 y[n-1] + \cdots + a_N y[n-N] = x[n] \tag{7}$$

$$H(z) = \frac{1}{1 + a_1 z^{-1} + \cdots + a_N z^{-N}} \tag{8}$$

- Autoregressive moving average (ARMA) model

$$y[n] + a_1 y[n-1] + \cdots + a_N y[n-N] = b_0 x[n] + b_1 x[n-1] + \cdots + b_M x[n-M] \tag{9}$$

$$H(z) = \frac{b_0 + b_1 z^{-1} + \cdots + b_M z^{-M}}{1 + a_1 z^{-1} + \cdots + a_N z^{-N}} \tag{10}$$

The signal model is chosen by comprehending the desired response profile. The AR model uses the poles and the ARMA model utilizes the zeros and poles at z domain. The zeros close to the unit circle presents the valley shape distribution and the poles nearby the unit circle provides the peaky style response in the domain. Observe that group of poles or zeros could demonstrate the flat with slight fluctuation response. After the signal model is selected, the order and coefficients are estimated by the various algorithms. The estimation process requires the statistical signal processing which regards signals as stochastic processes. The expectation, (auto)covariance, etc. operations based on the stationary white noise are exercised with various assumptions.

The following is the fundamental requirements and operations for statistical signal processing. The signal $y[n]$ is necessary to be a random variable sequence with zero mean as below.

$$E\{y[n]\} = 0 \quad \text{for all } n \in \mathbb{Z} \tag{11}$$

where the $E\{\ \}$ represents the expectation operation over the ensemble realizations. The auto-covariance sequence (ACS) is presented as below.

$$r[k] = E\{y[n]y^*[n-k]\} \tag{12}$$

With above conditions, the signal $y[n]$ is a second-order stationary sequence. Due to the limited length of the signal, the biased estimation of ACS is below.

$$\widetilde{r}[k] = \frac{1}{N}\sum_{n=k+1}^{N} y[n]y^*[n-k] \quad \text{for } 0 \le k \le N-1 \text{ and } 1 \le n \le N \tag{13}$$

The biased ACS estimate is usually used for the processing because the biased $\widetilde{r}[k]$ sequence is guaranteed to be positive semidefinite for positive spectral estimation [33].

2.2.1. Yule-Walker (AR Model Based)

The Yule-Walker algorithm [34] figures out signal distribution based on the AR model. Below is the conventional AR model in time domain with input $e[n]$ which is the white noise with σ^2 variance.

$$y[n] + a_1 y[n-1] + \cdots + a_N y[n-N] = e[n] \tag{14}$$

Apply the $y^*[n-k]$ product and expectation in both side of the equation as below.

$$E\{y^*[n-k](y[n] + a_1 y[n-1] + \cdots + a_N y[n-N])\} = E\{y^*[n-k]e[n]\} \tag{15}$$

By definition of ACS, the left-hand side of the equation is modified as below.

$$r[k] + \sum_{j=1}^{N} a_j r[k-j] = E\{y^*[n-k]e[n]\} \tag{16}$$

The right-hand side of the equation is determined by the k value condition. For the positive k value, the $y^*[n-k]$ and $e[n]$ is uncorrelated; hence, the expectation is zero as below.

$$r[k] + \sum_{j=1}^{N} a_j r[k-j] = 0 \quad \text{for } k > 0 \tag{17}$$

For zero k value, the right-hand side of the equation is equivalent to the $E\{e^*[n]e[n]\}$ as σ^2 due to the zero mean white noise $e[n]$.

$$r[0] + \sum_{j=1}^{N} a_j r[-j] = \sigma^2 \quad \text{for } k = 0 \;\; r[k] = r^*[-k] \tag{18}$$

Equations (17) and (18) are organized in matrix form as below. The lines in the following matrix are segmentation boundaries for submatrix construction. Equation (19) is known as the Yule-Walker equation or Normal equation to develop the fundamental of many AR estimation methods.

$$\left[\begin{array}{c|ccc} r[0] & r[-1] & \cdots & r[-N] \\ \hline r[1] & r[0] & \cdots & r[-N+1] \\ \vdots & \vdots & \ddots & \vdots \\ r[N] & r[N-1] & \cdots & r[0] \end{array}\right] \left[\begin{array}{c} 1 \\ \hline a_1 \\ \vdots \\ a_N \end{array}\right] = \left[\begin{array}{c} \sigma^2 \\ \hline 0 \\ \vdots \\ 0 \end{array}\right] \tag{19}$$

The submatrix is demonstrated as below. The solution of the following equation provides the coefficient estimation. However, computational limitation cannot realize the equation directly.

$$\begin{bmatrix} r[0] & \cdots & r[-N+1] \\ \vdots & \ddots & \vdots \\ r[N-1] & \cdots & r[0] \end{bmatrix} \begin{bmatrix} a_1 \\ \vdots \\ a_N \end{bmatrix} = -\begin{bmatrix} r[1] \\ \vdots \\ r[N] \end{bmatrix} \tag{20}$$

The biased ACS estimation with limited length data presents the below equation for the feasible approach. Note that the square matrix for inversion operation is Toeplitz matrix.

$$\begin{bmatrix} \widetilde{a}_1 \\ \vdots \\ \widetilde{a}_N \end{bmatrix} = -\begin{bmatrix} \widetilde{r}[0] & \cdots & \widetilde{r}[-N+1] \\ \vdots & \ddots & \vdots \\ \widetilde{r}[N-1] & \cdots & \widetilde{r}[0] \end{bmatrix}^{-1} \begin{bmatrix} \widetilde{r}[1] \\ \vdots \\ \widetilde{r}[N] \end{bmatrix} \tag{21}$$

Based on the calculated coefficients, the spectral distribution can be represented by the subsequent rational function from DTFT.

$$\left|\widetilde{H}\left(e^{j\omega}\right)\right|^2 = \frac{\sigma^2}{\left|1 + \widetilde{a}_1 e^{-j\omega} + \cdots + \widetilde{a}_N e^{-j\omega N}\right|^2} \tag{22}$$

This paper assumes that the order of the estimation methods is given based on the signal condition.

2.2.2. Prony (ARMA Model Based)

The Prony algorithm [35] estimates the coefficients of ARMA model for various types of signal. The complexity is developed from the nonlinear property of the ARMA difference equation. The impulse response of the ARMA model linearizes the system to provide the matrix description. The solution of the matrix derives the parameter estimation for ARMA. Below is the ARMA model in difference equation form.

$$y[n] + a_1 y[n-1] + \cdots + a_N y[n-N] = b_0 x[n] + \cdots + b_M x[n-M] \tag{23}$$

The rational function in z domain presents the impulse response as below.

$$\frac{Y(z)}{X(z)} = \frac{b_0 + b_1 z^{-1} + \cdots + b_M z^{-M}}{1 + a^1 z^{-1} + \cdots + a_N z^{-N}} = \frac{B(z)}{A(z)} = H(z) = h_0 + h_1 z^{-1} + h_2 z^{-2} + \cdots \tag{24}$$

Note that the impulse response $h[n]$ is most likely to be infinite length sequence because of the rational function property. The product between the impulse response and denominator demonstrates the numerator of the rational function in z domain as below.

$$B(z) = H(z)A(z) \tag{25}$$

Observe that the length of a and b coefficients are finite. However, the h length is infinite in general as below.

$$b_0 + b_1 z^{-1} + \cdots + b_M z^{-M} = \left(h_0 + h_1 z^{-1} + h_2 z^{-2} + \cdots\right)\left(1 + a^1 z^{-1} + \cdots + a_N z^{-N}\right) \tag{26}$$

The above polynomial in z domain can be organized in matrix form as below. The lines in the following matrix are segmentation boundaries for submatrix construction.

$$\left[\begin{array}{c} b_0 \\ b_1 \\ \vdots \\ b_M \\ \hline 0 \\ \vdots \\ 0 \end{array}\right] = \left[\begin{array}{c|cccc} h_0 & 0 & 0 & \cdots & 0 \\ h_1 & h_0 & 0 & \cdots & 0 \\ h_2 & h_1 & h_0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ h_M & h_{M-1} & h_{M-2} & \cdots & \vdots \\ \hline \vdots & \vdots & \vdots & \cdots & \vdots \\ h_L & h_{L-1} & h_{L-2} & \cdots & h_{L-N} \end{array}\right] \left[\begin{array}{c} 1 \\ \hline a_1 \\ a_2 \\ \vdots \\ a_N \end{array}\right] \tag{27}$$

The submatrix is demonstrated in bold alphabet as below. The upper and lower case represent the matrix and vector, respectively.

$$\left[\begin{array}{c} \boldsymbol{b} \\ \hline \boldsymbol{0} \end{array}\right] = \left[\begin{array}{cc} \multicolumn{2}{c}{\boldsymbol{H_1}} \\ \hline \boldsymbol{h_1} & \boldsymbol{H_2} \end{array}\right] \left[\begin{array}{c} 1 \\ \hline \boldsymbol{a}^{\dagger} \end{array}\right] \tag{28}$$

The solution of the following equations provides the coefficient estimations as $\widetilde{\boldsymbol{a}}$ and $\widetilde{\boldsymbol{b}}$. Note that $\boldsymbol{a}^{\dagger}$ is the submatrix of $\boldsymbol{a}$ column vector from the second row to end.

$$\boldsymbol{0} = \boldsymbol{h_1} + \boldsymbol{H_2}\boldsymbol{a}^{\dagger} \tag{29}$$

$$\boldsymbol{b} = \boldsymbol{H_1}\boldsymbol{a}$$

Based on the calculated coefficients, the spectral distribution can be represented by the subsequent rational function from DTFT.

$$\left|\widetilde{H}\left(e^{j\omega}\right)\right|^2 = \frac{\left|\widetilde{b}_0 + \widetilde{b}_1 e^{-j\omega} + \cdots + \widetilde{b}_M e^{-j\omega M}\right|^2}{\left|1 + \widetilde{a}_1 e^{-j\omega} + \cdots + \widetilde{a}_N e^{-j\omega N}\right|^2} \tag{30}$$

Therefore, the Prony receives the $h[n]$ as the input to estimate the a and b coefficients for describing the unknown system model in best. Note that the longer $\boldsymbol{H}$ matrix could lead the solution to the least square sense due to the non-square matrix size.

2.2.3. Steiglitz-McBride (ARMA Model Based)

The Steiglitz-McBride method [36] is the iterative technique to identify an ARMA system by minimizing the mean square error between the system and model output. Figure 4 shows the simplified system model for the Steiglitz-McBride algorithm.

The Steiglitz-McBride method recognizes the target system $B(z)/A(z)$ by minimizing the energy of $e[n]$ to equalize the model system $\widetilde{B}(z)/\widetilde{A}(z)$ to the target. The impulse response of the target and model are denoted by $h_d[n]$ and $h[n]$, respectively. The system error $e[n]$ is represented as below.

$$e[n] = h_d[n] - h[n] \overset{Z}{\leftrightarrow} E(z) = \frac{B(z)}{A(z)} - \frac{\widetilde{B}(z)}{\widetilde{A}(z)} \tag{31}$$

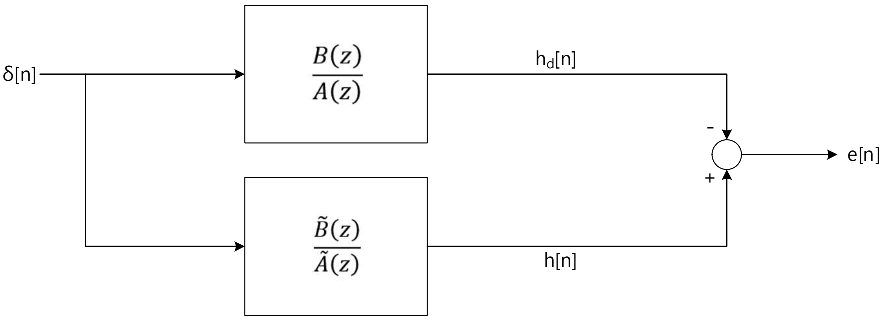


Figure 4. System model for Stiglitz-McBride algorithm.

The solution is given as below equation. The arguments of the minima are the points of the function at which the error $e[n]$ energy is minimized. Therefore, the difference between the target and model system is decreased to the least.

$$\widetilde{\gamma} = \underset{\gamma}{\operatorname{argmin}} \sum_n \left|e[n]\right|^2 \quad \text{where } \gamma = \{a_1, \ldots, a_N, b_0, \ldots, b_M\} \text{ from } \frac{\widetilde{B}(z)}{\widetilde{A}(z)} \tag{32}$$

The rational function of ARMA provides the recursive equation; hence, the solution of previous equation implies the highly nonlinear and intractable property. The modified system model as Figure 5 presents the linearized solution by iterative approach.

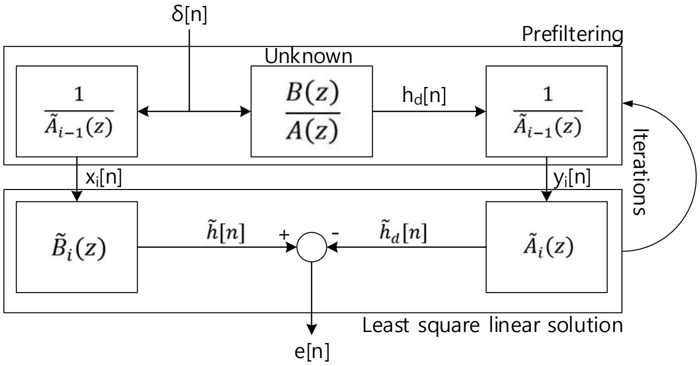


Figure 5. Modified system identification model for Stiglitz-McBride algorithm.

Subsequent equations show the proof of equivalence for system identification problem between the target and model. As the error norm approaches to zero, below equation demonstrates that the target and model system are became identical according to the Parseval's theorem.

$$\sum_n \left|e[n]\right|^2 = \frac{1}{2\pi j} \oint_c \left|H_d(z)\frac{\widetilde{A}_i(z)}{\widetilde{A}_{i-1}(z)} - \frac{\widetilde{B}_i(z)}{\widetilde{A}_{i-1}(z)}\right|^2 z^{-1}dz = \frac{1}{2\pi j} \oint_c \left|H_d(z)\widetilde{A}_i(z) - \widetilde{B}_i(z)\right|^2 \left|\frac{1}{\widetilde{A}_{i-1}(z)}\right|^2 z^{-1}dz \tag{33}$$

Hence, the optimal numerator and denominator coefficients can be derived by solving the below equation. The $\widetilde{A}_{i-1}(z)$ and $\widetilde{A}_i(z)$ are mutually related. Therefore, the iterative method should be employed.

$$\begin{aligned} \widetilde{\gamma} &= \underset{\gamma}{\operatorname{argmin}} \sum_n \left|e[n]\right|^2 = \underset{\gamma}{\operatorname{argmin}} \oint_c \left|H_d(z)\widetilde{A}_i(z) - \widetilde{B}_i(z)\right|^2 \left|\frac{1}{\widetilde{A}_{i-1}(z)}\right|^2 z^{-1}dz \\ &= \underset{\gamma}{\operatorname{argmin}} \oint_c \left|H_d(z)\widetilde{A}_i(z) - \widetilde{B}_i(z)\right|^2 z^{-1}dz \end{aligned} \tag{34}$$

As shown in Equation (33) and Figure 5, Equation (34) is converted to the linear problem as $\left|H_d(z)\widetilde{A}_i(z) - \widetilde{B}_i(z)\right|^2$ by the prefiltering with $1/\widetilde{A}_{i-1}(z)$. The prefiltering with desired response and delta function are given as below.

$$x_i[n] = -a_{(i-1)1}x_i[n-1] - a_{(i-1)2}x_i[n-2] - \cdots - a_{(i-1)N}x_i[n-M] + \delta[n] \tag{35}$$

$$y_i[n] = -a_{(i-1)1}y_i[n-1] - a_{(i-1)2}y_i[n-2] - \cdots - a_{(i-1)N}y_i[n-M] + h_d[n] \tag{36}$$

The prefiltering part of Figure 5 is demonstrated by Equations (35) and (36) for left-hand and right-hand side. The matrix representations below stand for the convolution sum. The least square linear solution part of Figure 5 is denoted by the below two matrix operations for left-hand and right-hand side.

$$\begin{bmatrix} x_i[0] & 0 & 0 & \cdots & 0 \\ x_i[1] & x_i[0] & 0 & \cdots & 0 \\ x_i[2] & x_i[1] & x_i[0] & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_i[M] & x_i[M-1] & x_i[M-2] & \cdots & x_i[0] \\ x_i[M+1] & x_i[M] & x_i[M-1] & \cdots & x_i[1] \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_i[L] & x_i[L-1] & x_i[L-2] & \cdots & x_i[L-M] \end{bmatrix} \begin{bmatrix} b_{i0} \\ b_{i1} \\ b_{i2} \\ \vdots \\ b_{iM} \end{bmatrix} = \begin{bmatrix} \widetilde{h}_i[0] \\ \widetilde{h}_i[1] \\ \widetilde{h}_i[2] \\ \vdots \\ \widetilde{h}_i[L] \end{bmatrix} \tag{37}$$

$$\begin{bmatrix} y_i[0] & 0 & 0 & \cdots & 0 \\ y_i[1] & y_i[0] & 0 & \cdots & 0 \\ y_i[2] & y_i[1] & y_i[0] & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ y_i[N] & y_i[N-1] & y_i[N-2] & \cdots & y_i[0] \\ y_i[N+1] & y_i[N] & y_i[N-1] & \cdots & y_i[1] \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ y_i[L] & y_i[L-1] & y_i[L-2] & \cdots & y_i[L-N] \end{bmatrix} \begin{bmatrix} a_{i0} \\ a_{i1} \\ a_{i2} \\ \vdots \\ a_{iN} \end{bmatrix} = \begin{bmatrix} \widetilde{h}_{di}[0] \\ \widetilde{h}_{di}[1] \\ \widetilde{h}_{di}[2] \\ \vdots \\ \widetilde{h}_{di}[L] \end{bmatrix} \tag{38}$$

For zero error norm condition, the above equations should be equal as below.

$$\widetilde{h}_i[n] = \widetilde{h}_{di}[n] \;\text{ for } e_i[n] = 0 \text{ and } 0 \le n \le L \tag{39}$$

With zero error norm condition, the above matrix can be combined as below.

$$\left[\begin{array}{c|ccccccc} -y_i[0] & 0 & \cdots & 0 & x_i[0] & 0 & \cdots & 0 \\ -y_i[1] & -y_i[0] & \cdots & 0 & x_i[1] & x_i[0] & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ -y_i[N] & -y_i[N-1] & \cdots & -y_i[0] & x_i[M] & x_i[M-1] & \cdots & x_i[0] \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ -y_i[L] & -y_i[L-1] & \cdots & -y_i[L-N] & x_i[L] & x_i[L-1] & \cdots & x_i[L-M] \end{array}\right] \left[\begin{array}{c} a_{i0} \\ \hline \vdots \\ a_{iN} \\ b_{i0} \\ \vdots \\ b_{iM} \end{array}\right] = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \tag{40}$$

The first column of the matrix and first row of the column vector are partitioned and reorganized as below.

$$\begin{bmatrix} 0 & \cdots & 0 & x_i[0] & 0 & \cdots & 0 \\ -y_i[0] & \cdots & 0 & x_i[1] & x_i[0] & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ -y_i[N-1] & \cdots & -y_i[0] & x_i[M] & x_i[M-1] & \cdots & x_i[0] \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ -y_i[L-1] & \cdots & -y_i[L-N] & x_i[L] & x_i[L-1] & \cdots & x_i[L-M] \end{bmatrix} \begin{bmatrix} a_{i1} \\ \vdots \\ a_{iN} \\ b_{i0} \\ \vdots \\ b_{iM} \end{bmatrix} = \begin{bmatrix} y_i[0] \\ y_i[1] \\ \vdots \\ y_i[N] \\ \vdots \\ y_i[L] \end{bmatrix} \tag{41}$$

The short representation of above matrix is below. Note that the i in the subscript indicates the i-th iterations.

$$\boldsymbol{H}_i \boldsymbol{c}_i = \boldsymbol{y}_i \tag{42}$$

The optimal parameters can be obtained by solving the linear solver such as QR solver to find minimum norm residual solution.

$$\boldsymbol{c}_i = \text{Linear Solver}\left(\boldsymbol{H}_i, \boldsymbol{y}_i\right) \tag{43}$$

Repeat the procedures for several times. Final $\boldsymbol{c}_i$ is the optimal $\widetilde{\boldsymbol{c}}$. In general, the iterative method converges to the optimal solution rapidly. The Prony method provides the good initial estimation of $\mathrm{a}_{(0)i}$ coefficients for Equation (35) and Equation (36). Based on the calculated coefficients $\widetilde{\boldsymbol{c}}$, the spectral distribution can be represented by the subsequent rational function from DTFT.

$$\left|\widetilde{H}\left(e^{j\omega}\right)\right|^2 = \frac{\left|\widetilde{b}_0 + \widetilde{b}_1 e^{-j\omega} + \ldots + \widetilde{b}_M e^{-j\omega M}\right|^2}{\left|1 + \widetilde{a}_1 e^{-j\omega} + \ldots + \widetilde{a}_N e^{-j\omega N}\right|^2} \tag{44}$$

Therefore, the Steiglitz-McBride algorithm receives the $h_d[n]$ as the input to estimate the prefiltering coefficients by using the Prony algorithm first. After the prefiltering, the linear solver generates the numerator and denominator coefficients. Repeat the prefiltering and linear solving until the error norm reaches to the minimum requirements.

3. Simulations

The non-parametric and parametric HD algorithms are theoretically explained at previous sections. In order to summarize the algorithm sequence, below equations are organized in the chronological order. Note that the numerical realizations by computer require the DFT (or FFT) for domain transfer. The $Y[k]$ is the FFT of the received signal with the power of two in order length L. The real cepstrum is realized as below.

$$c_y[n] = \frac{1}{N} \sum_{k=0}^{N-1} \log\left|Y[k]\right| e^{j\frac{2\pi}{N}kn} \tag{45}$$

Apply the FILF for minimum phase realization and low time delay removal. The FILF functions are described at Appendix A in detail.

$$\hat{h}[n] = c_y[n] w[n] \tag{46}$$

The $e^{\hat{H}[k]}$ denotes the exponential of $\hat{h}[n]$ FFT with length L. To maintain the consistency, forward FFT is applied to obtain the inverse transformation as below. The input conjugate to the FFT provides the conjugated inverse FFT outcome as below.

$$\widetilde{h}[n] = \frac{1}{L} \sum_{k=0}^{L-1} e^{\hat{H}[k]} e^{j\frac{2\pi}{L}kn} = \frac{1}{L} \left\{ \sum_{k=0}^{L-1} \left(e^{\hat{H}[k]}\right)^* e^{-j\frac{2\pi}{L}kn} \right\}^* = \frac{1}{L} \left\{ \mathrm{FFT}_L\left(\left(e^{\hat{H}[k]}\right)^*\right) \right\}^* \tag{47}$$

Since the $\widetilde{h}[n]$ magnitude corresponds to the time delay distribution, the absolute of scaled version as below represents the second microphone position by ToF. The maximum magnitude position of $\left|\widetilde{h}_{HD}[n]\right|$ demonstrates the time delay between the microphones.

$$\left|\widetilde{h}_{HD}[n]\right| = \frac{1}{L}\left|\mathrm{FFT}_L\left(e^{\hat{H}[k]}\right)^*\right| \tag{48}$$

The parametric HD follows the exactly identical procedures until $e^{\hat{H}[k]}$ computation. The Yule-walker, Prony, and Steiglitz-McBride algorithms are originally devised for forward domain transfer; therefore, the input conjugate to the parametric algorithms generate the conjugated parameters for estimation. Below equation is illustrated for Yule-walker method.

$$\left\{\widetilde{a}_1, \widetilde{a}_2 \ldots, \widetilde{a}_N, \widetilde{\sigma}^2\right\} = \mathrm{YW}\left(\left(e^{\hat{H}[k]}\right)^*\right) \tag{49}$$

Because the $\widetilde{h}_{\mathrm{YW}}[n]$ magnitude shows the time delay distribution, the absolute of rational function as below presents the ToF. The maximum magnitude location in n of $\left|\widetilde{h}_{\mathrm{YW}}[n]\right|$ indicates the time delay between the microphones.

$$\left|\widetilde{h}_{\mathrm{YW}}[n]\right|^2 = \frac{\widetilde{\sigma}^2}{\left|1 + \widetilde{a}_1 e^{-j\frac{2\pi}{L}n} + \widetilde{a}_2 e^{-j2\frac{2\pi}{L}n} + \ldots + \widetilde{a}_N e^{-jN\frac{2\pi}{L}n}\right|^2} \quad \{\widetilde{a}_1, \widetilde{a}_2 \ldots, \widetilde{a}_N\} \in \mathbb{C};\ \ 0 \le n \le L-1 \tag{50}$$

Below equation is illustrated for Prony method.

$$\left\{\widetilde{a}_1, \ldots, \widetilde{a}_N, \widetilde{b}_0, \ldots, \widetilde{b}_M\right\} = \mathrm{Prony}\left(\left(e^{\hat{H}[k]}\right)^*\right) \tag{51}$$

The maximum magnitude location in n of $\left|\widetilde{h}_{\mathrm{Prony}}[n]\right|$ indicates the ToF between the microphones.

$$\left|\widetilde{h}_{\mathrm{Prony}}[n]\right|^2 = \frac{\left|\widetilde{b}_0 + \widetilde{b}_1 e^{-j\frac{2\pi}{L}n} + \ldots + \widetilde{b}_M e^{-jM\frac{2\pi}{L}n}\right|^2}{\left|1 + \widetilde{a}_1 e^{-j\frac{2\pi}{L}n} + \ldots + \widetilde{a}_N e^{-jN\frac{2\pi}{L}n}\right|^2} \quad \left\{\widetilde{a}_1, \ldots, \widetilde{a}_N, \widetilde{b}_0, \ldots, \widetilde{b}_M\right\} \in \mathbb{C};\ \ 0 \le n \le L-1 \tag{52}$$

Below equation displays the Steiglitz-McBride method.

$$\left\{\widetilde{a}_1, \ldots, \widetilde{a}_N, \widetilde{b}_0, \ldots, \widetilde{b}_M\right\} = \mathrm{StMc}\left(\left(e^{\hat{H}[k]}\right)^*\right) \tag{53}$$

The maximum magnitude location in n of $\left|\widetilde{h}_{\mathrm{StMc}}[n]\right|$ delivers the ToF between the microphones.

$$\left|\widetilde{h}_{\mathrm{StMc}}[n]\right|^2 = \frac{\left|\widetilde{b}_0 + \widetilde{b}_1 e^{-j\frac{2\pi}{L}n} + \ldots + \widetilde{b}_M e^{-jM\frac{2\pi}{L}n}\right|^2}{\left|1 + \widetilde{a}_1 e^{-j\frac{2\pi}{L}n} + \ldots + \widetilde{a}_N e^{-jN\frac{2\pi}{L}n}\right|^2} \quad \left\{\widetilde{a}_1, \ldots, \widetilde{a}_N, \widetilde{b}_0, \ldots, \widetilde{b}_M\right\} \in \mathbb{C};\ \ 0 \le n \le L-1 \tag{54}$$

The data used for the simulation is generated from the standard Gaussian distribution with zero mean and unit variance. Two independent and identically distributed random variables as $e_s[n]$ and $e_m[n]$ correspond to the signal and noise source, respectively as below.

$$e_s[n] \sim \mathcal{N}(0,1) \ \text{ and } e_m[n] \sim \mathcal{N}(0,1) \tag{55}$$

The sound source $x[n]$ is provided by the filtering the $e_s[n]$ with infinite impulse response (IIR) filter based on the Butterworth algorithm as below. The $\mathrm{IIR}_{\mathrm{LPF}}$ function generates the corresponding coefficients and filter function performs the filter operation in time domain with coefficients and input data $e_s[n]$. The ω_c is the cutoff frequency which represents the half magnitude in filter response. Further, 10 is the filter order.

$$x[n] = \text{filter}(\text{IIR}_{\text{LPF}}(10, \omega_c), e_s[n]\ ) \tag{56}$$

The received signal $y[n]$ is produced with scaled and delayed signal and independent noise as below. The d is the delay in samples and α is the signal attenuation factor for second receiver. The β is the noise control factor.

$$y[n] = x[n] + \alpha x[n-d] + \alpha\beta e_m[n] \tag{57}$$

Due to the general physics basis, the α magnitude is less than one; however, the β magnitude can be greater than one to control the signal-to-noise ratio (SNR) which is given as below. Note that the variance of sound source $x[n]$ is equivalent to the cutoff frequency ω_c with given source $e_s[n]$.

$$\text{SNR (dB)} = 10\log_{10}\left(\frac{\omega_c + \alpha^2\omega_c}{(\alpha\beta)^2}\right) \tag{58}$$

The desired SNR can be achieved by adopting the following equation. The given SNR computes the required β value for noise magnitude. Also, the signal attenuation factor α is fixed for 0.8 in this simulation.

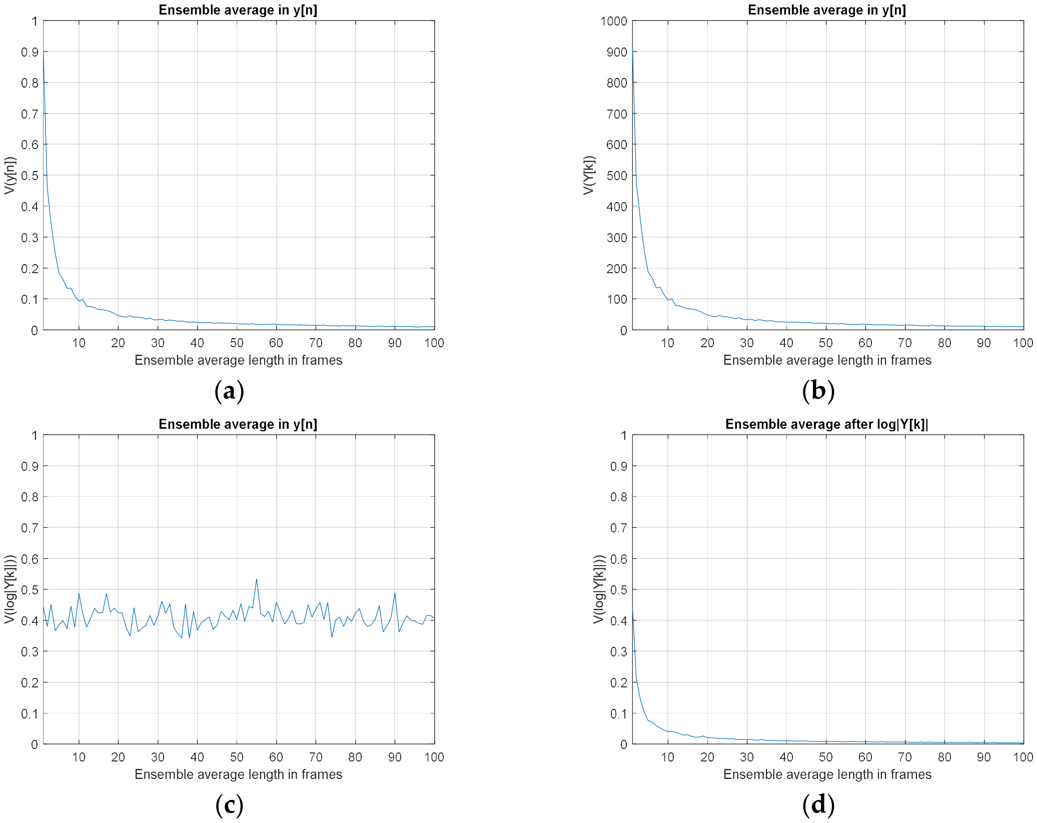
$$\beta = \sqrt{\frac{\omega_c + \alpha^2\omega_c}{\alpha^2 10^{\frac{SNR}{10}}}} \tag{59}$$

Once the received signal $y[n]$ is created from the given and derived parameters, the non-parametric and parametric HD algorithms can improve the SNR by using the ensemble average. The average over the framed data $y_k[n]$ enhances the noise variance in inversely proportional manner as shown below. Note that the average is performed at the identical time positions over the frames.

$$y[n] = \frac{1}{R}\sum_{k=0}^{R-1} y_k[n] \ \rightarrow\ \sigma_y^2 \approx \frac{\sigma_{y_k}^2}{R} \tag{60}$$

The ensemble average maintains the signal power and reduces the noise variance for increased SNR over conventional linear operations. However, the non-linear operations such as logarithm could show the strong narrow proportionality for the average. Figure 6 demonstrates the variances of HD algorithm up to $\log\left|Y[k]\right|$ stage for various ensemble average length R. The received signal $y[n]$ is pure noise based on the standard Gaussian distribution with zero mean and unit variance. The ensemble averaged $y[n]$ delivers the inversely proportional variance for $y[n]$ and $Y[k]$ against the ensemble length. However, the variance of $\log\left|Y[k]\right|$ sustains the constant value with fluctuation due to the non-linearity of logarithm operation. Therefore, instead of taking the average on the $y[n]$, performing the average after $\log\left|Y[k]\right|$ presents the inverse proportionality in variance which is illustrated in the last figure of below.

Based on the previously stated signal model and ensemble average, the received signal is generated for simulation. The FILF window $w[n]$ is determined to remove the 40 sample below and maximum phase realization. The detail $w[n]$ function and performance can be observed in Appendix A. In this paper, the computed non-parametric and parametric HD output $\widetilde{h}[n]$ is normalized to explore the maximum location n which corresponds to the ToF. Note that the non-parametric HD demonstrates the distribution by direct values. The parametric HD should evaluate the rational functions given in Equations (50), (52), and (54) with derived coefficient values from Yule-Walker, Prony, and Steiglitz-McBride.



(**a**) (**b**)

(**c**) (**d**)

Figure 6. Signal variance at each computational stage of HD algorithm with pure noise situation as $y[n] \sim \mathcal{N}(0,1)$. The frame size is 1024 samples: (**a**) variance of $y[n]$ with ensemble averaged $y[n]$; (**b**) variance of $Y[k]$ with ensemble averaged $y[n]$; (**c**) variance of $\log|Y[k]|$ with ensemble averaged $y[n]$; (**d**) variance of $\log|Y[k]|$ over ensemble averaged $\log|Y[k]|$.

Figure 7 denotes the estimated $\widetilde{h}[n]$ (normalized) with 20 dB SNR and 20 ensemble average length for various ToF distributions. The desired delays d in Equation (57) are scheduled from 50 to 200 with 10 sample intervals and illustrated with various colors in Figure 7. The individual peak in the figure $\widetilde{h}[n]$ distribution represents the simulation with given delay d. The non-parametric HD shows the near zero values below the 40-sample delay and the maximum peak delivers the exactly expected delay position with prominence. The wideband noise is observed over the whole range of samples above 40 samples in low profile. Note that the minimum phase realization can only deliver the clean and sharp abruption around the preferred window position. See Appendix A for further information.

The parametric HDs provide the poles and zeros from the signal models given in Equations (8) and (10). The second column in Figure 7 presents the pole-zero plots to deliver the pole locations with x mark and zero locations with o mark. The big circle in the pole-zero plot indicates the unit circle and the green dotted vector specifies the desired pole location from the given delay d. Note that the pole closer to the unit circle represents the peaky response on the evaluation as shown in Equations (50), (52), and (54). Therefore, the pole toward the unit circle is desirable for corresponding delay d position as $e^{2\pi d/L}$ where L is the FFT length. The prominent location in the parametric HD response is consistent with the green dotted vector. The closeness of the pole to the unit circle determines the sharpness of the response.

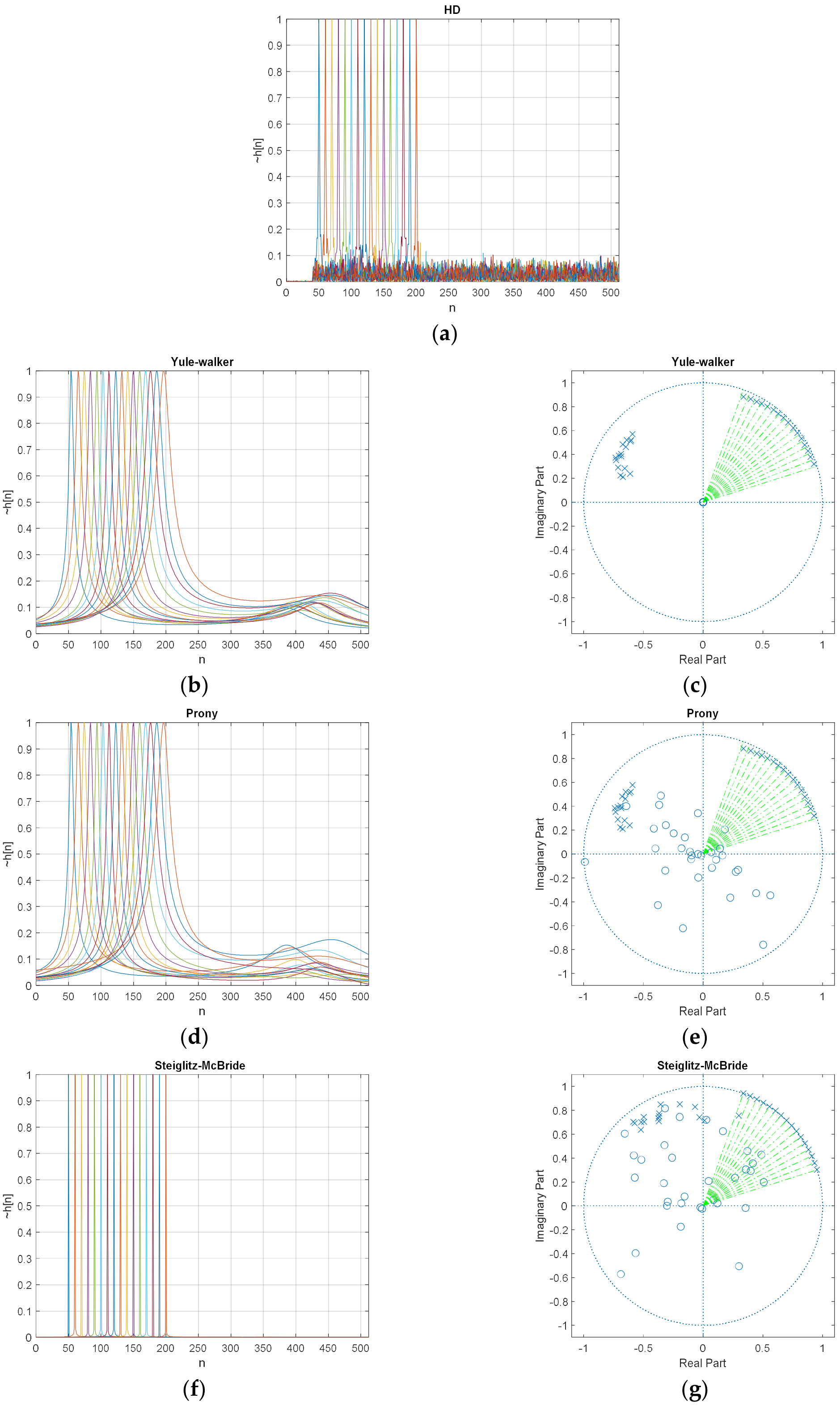


Figure 7. Estimated $\widetilde{h}[n]$ distribution (normalized) and pole-zero plot with 20 dB SNR & 20 ensemble average for ToF (50~200 in 10 sample interval) from: (**a**) non-parametric HD; (**b**) Yule-walker HD; (**c**) corresponding pole-zero plot; (**d**) Prony HD; (**e**) corresponding pole-zero plot; (**f**) Steiglitz-McBride HD; (**g**) corresponding pole-zero plot. The pole/zero locations are indicated by x/o mark, respectively. The green dotted vector specifies the desired pole location $e^{2\pi d/L}$ (FFT length L) from the given delay d.

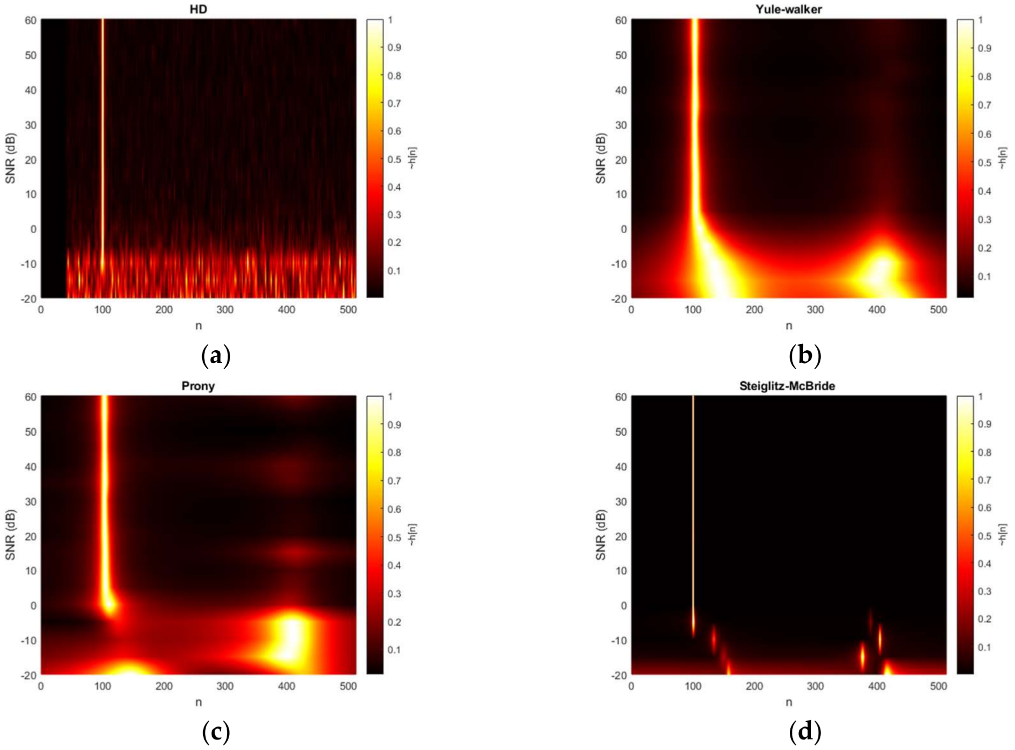
The parametric HD with Yule-walker method demonstrates the peak value at the desired position and the low hill points around 400 samples. The pole-zero plot shows the pole locations only since the Yule-walker method is derived from AR model as Equation (8) which contains the denominator coefficients for poles. The zeros are fixed in plot origin. The caution should be exercised for parametric methods in this paper because of the coefficient number sets. As we can observe the suspiciousness in the pole-zero plot, the coefficients are complex numbers. The real number coefficients present the symmetric poles and zeros distribution. However, the symmetric pole-zero location produces the biased estimation for low sample delays due to the mutual correlations. The green lines in pole-zero plot indicate the desired time delay locations and one pole of the Yule-walker method follows the green line direction. The other pole is placed and clustered in the perpendicular angle to reduce the correlation.

Similar to the Yule-walker method, the parametric HD with Prony method provides the peak location pointing the desired delay position in Figure 7. The low hill also can be detected with less consistency in position. The Prony method is derived from the ARMA model which includes the numerator and denominator for zeros and poles, respectively. The zero positions in the Prony method are distributed widely around the plot origin; therefore, the zeros barely contribute the magnitude modifications. Also, one pole of the Prony method follows the green line direction. The other pole is placed and clustered in the perpendicular angle to reduce the correlation as well.

The parametric HD with Steiglitz-McBride method shows sharp peak for desired delay position. Not only prominent peak but also low base values depict the best estimation configuration in HD algorithms. The Steiglitz-McBride method is also derived from the ARMA model. One pole is closely approached toward the unit circle in the pole-zero plot to deliver the sharp peak. The location of green lines and poles are well coordinated to indicate the proper time delay. Since the one pole contributes to the magnitude control significantly, the other pole and zeros are positioned widely in scattered manner. Observe that the poles in the Steiglitz-McBride method do not cross the unit circle for a stable response.

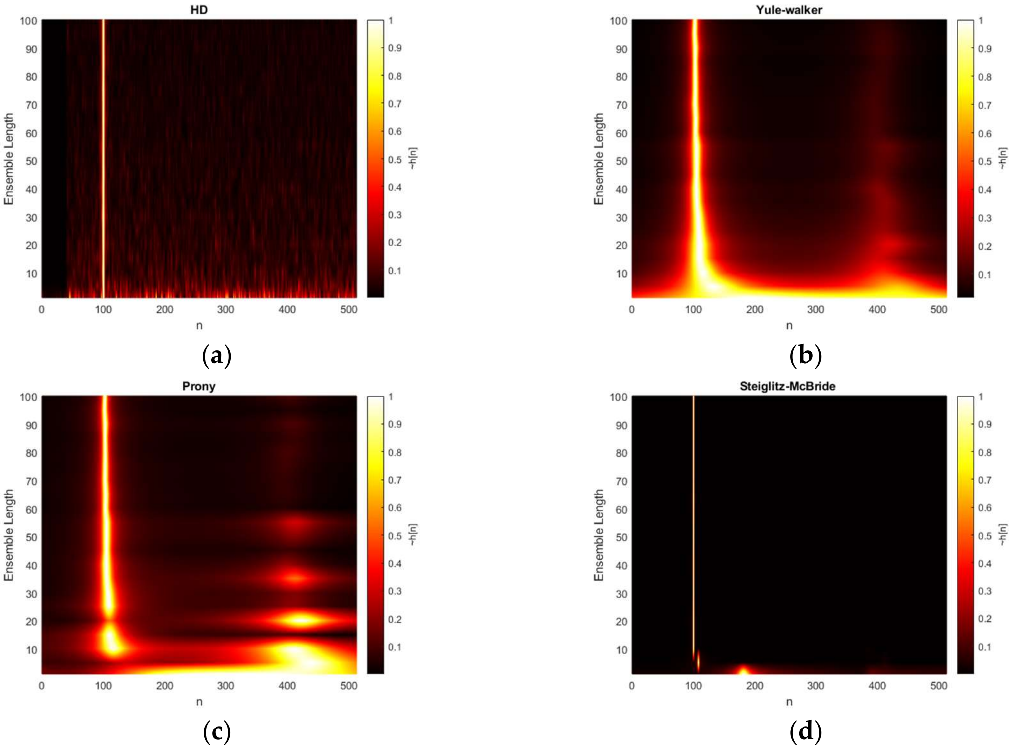
Figure 8 demonstrates the HD algorithm distributions for various SNR situations. The second microphone is positioned at 100-sample delay and the ensemble average length is arranged at 20 frames. The 2D plot in top viewpoint is generated by interpolation based on the simulated data for smooth texture. The non-parametric HD represents the peaky response below the 0 dB and the consistent response beyond the 0 dB. The single line in the 2D plot indicates the time delay in samples. The Yule-Walker and Prony algorithms also illustrate the weak performance below 0 dB condition with second strong peak around 400 samples. The consistent outputs are similarly examined beyond the 0 dB; however, the thickness of the pinnacle is wider than the non-parametric HD outcome. Therefore, certain bias and variance are expected in statistical performance for Yule-Walker and Prony method. The Steiglitz-McBride provides the thin line with low base values in Figure 8. The performance below 0 dB SNR is degraded as pointing to the improper locations.

Figure 9 presents the HD algorithm distributions for various ensemble average length conditions. The second microphone is located at 100-sample delay and the SNR is prescribed at 0 dB. The non-parametric HD represents the peaky response for short ensemble length and the consistent response for overall situations. Note that the steep cliff below 40-sample delay is caused by the FILF window. The Yule-walker and Prony also demonstrate the weak performance below 20 frames in ensemble length with second strong peak around 400-sample delay. The Yule-Walker algorithm delivers the coherent performance improvement in terms of mainlobe and sidelobe profile for increased ensemble length. However, the Prony algorithm provides the intermittent performance fluctuation up to the 40 ensemble length. The thickness of the dominant peak in Yule-walker and Prony method indicates the narrower outline for the increased ensemble length; therefore, the statistical performance by bias and variance are expected to be enhanced gradually. The Steiglitz-McBride denotes the very fine line with low base values in Figure 9. The performance degradation is only observed for one and five frame lengths for ensemble average.



(**a**) (**b**) (**c**) (**d**)

Figure 8. Estimated $\widetilde{h}[n]$ distribution (normalized) with given SNR & 20 ensemble average from: (**a**) non-parametric HD top-view (interpolated); (**b**) Yule-walker HD top-view (interpolated); (**c**) Prony HD top-view (interpolated); (**d**) Steiglitz-McBride HD top-view (interpolated).



(**a**) (**b**) (**c**) (**d**)

Figure 9. Estimated $\widetilde{h}[n]$ distribution (normalized) with given ensemble average length & 0 dB SNR from: (**a**) non-parametric HD top-view (interpolated); (**b**) Yule-walker HD top-view (interpolated); (**c**) Prony HD top-view (interpolated); (**d**) Steiglitz-McBride HD top-view (interpolated).

Figure 10 shows the extension of previous ensemble length simulation for high SNR scenario. The all conditions are identical except the 20 dB SNR. The non-parametric HD depicts the consistent response for overall situations including the non-average situation. The Yule-Walker and Prony present the narrow mainlobe without significant degradation except single-frame processing situation. The mainlobe thickness in Yule-Walker and Prony method produces the narrower profile for the increased ensemble length as well. No intermittent performance fluctuation is observed in the Prony method except non-averaging processing. The Steiglitz-McBride method represents the laser line with low base values in Figure 10. The performance degradation is only observed for one frame length for ensemble average.

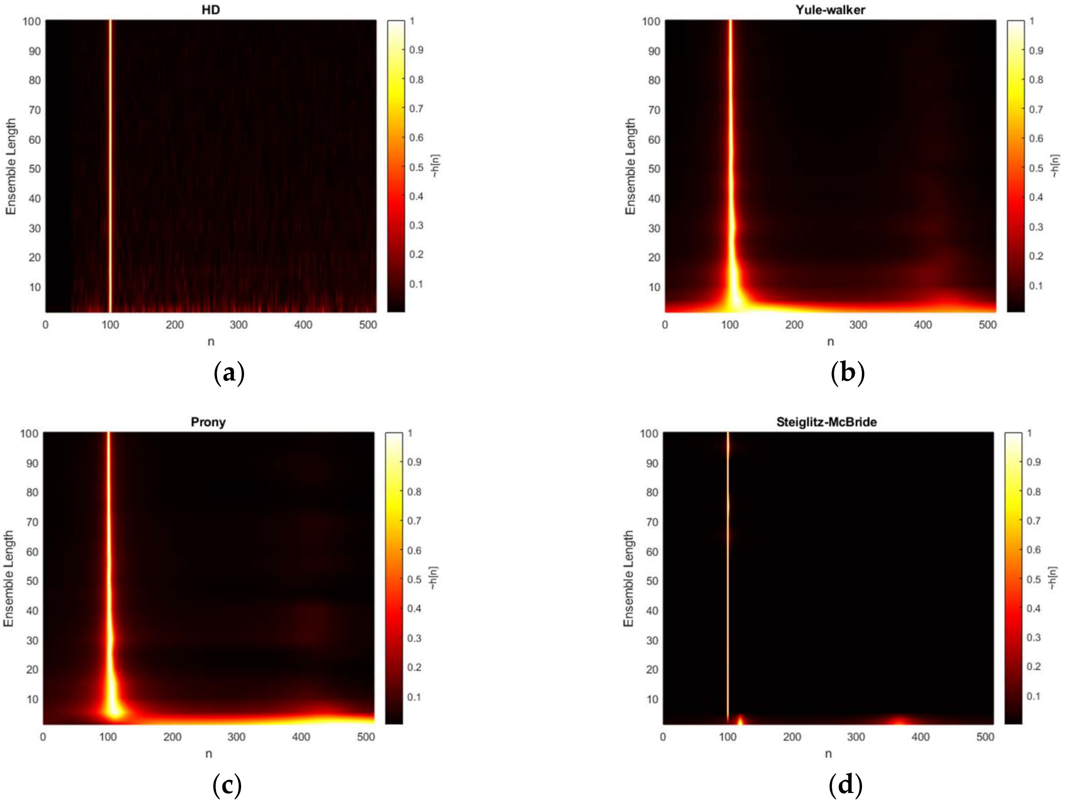
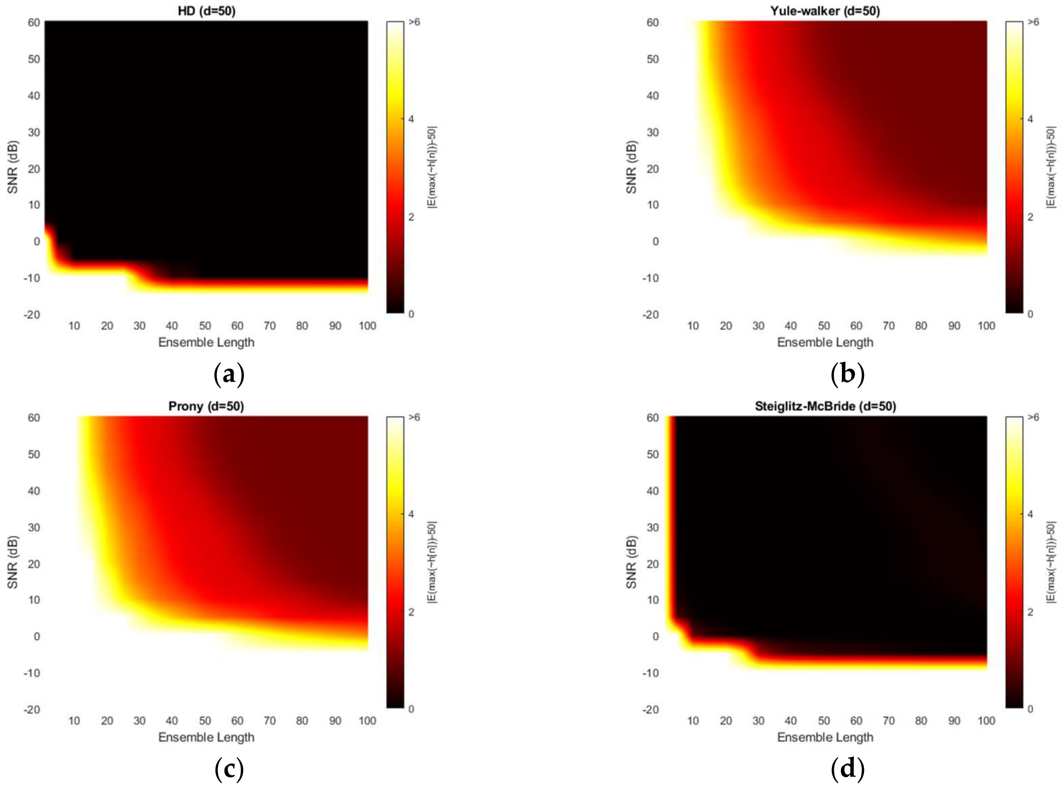


Figure 10. Estimated $\widetilde{h}[n]$ distribution (normalized) with given ensemble average length & 20 dB SNR from: (**a**) non-parametric HD top-view (interpolated); (**b**) Yule-walker HD top-view (interpolated); (**c**) Prony HD top-view (interpolated); (**d**) Steiglitz-McBride HD top-view (interpolated).

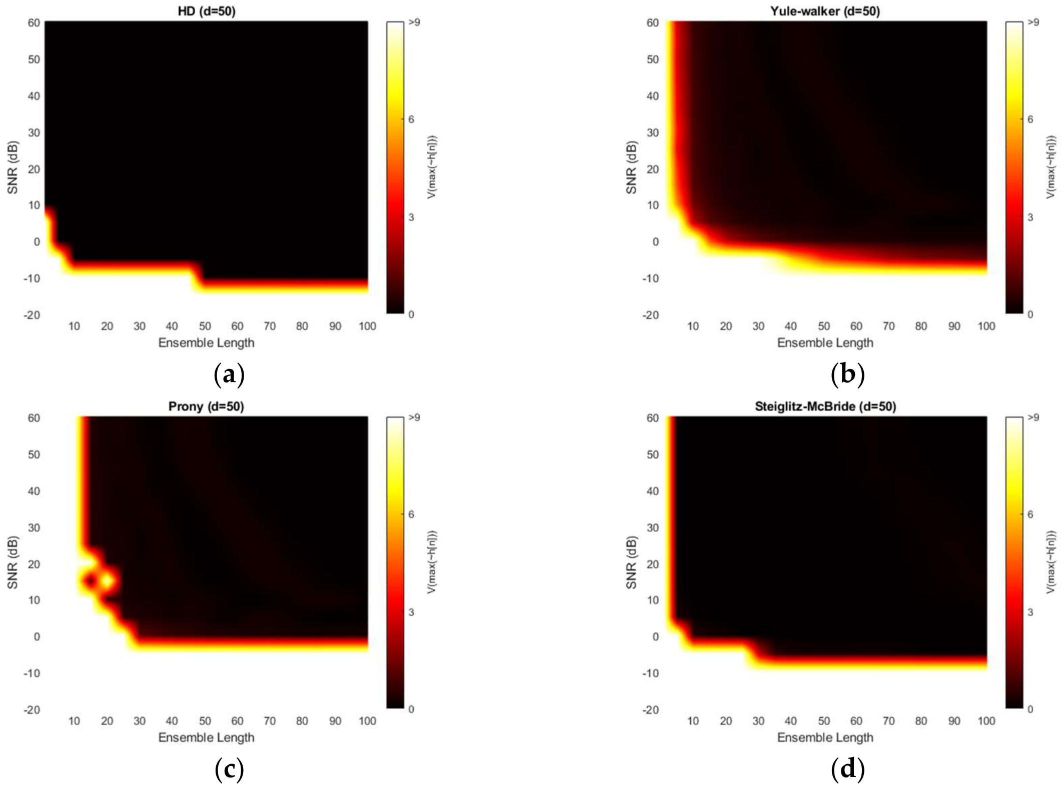
Figure 11 illustrates the absolute estimation error for 50-sample delay with various SNR and ensemble length. The non-parametric and Steiglitz-McBride method show the widespread dark black area which indicates the near zero bias. For near microphone placement (d = 50), the non-parametric and Steiglitz-McBride method provide the reduced biased estimation above the 0 dB SNR and 10 frame ensemble average. However, the Yule-Walker and Prony method represent the relatively high estimation error even for the high SNR and long average length. The observed least bias is around 0.5 sample in absolute value.

Figure 12 demonstrates the variance for 50-sample delay with various SNR and ensemble length. Note that the variance above the 9 is normalized for investigating the small-scale distribution. All four methods provide the near zero variance for high SNR and long average length. This observation is consistent with the previous simulation outcomes which correspond to the mainlobe thickness. From a variance perspective, the non-parametric and Steiglitz-McBride method present a similarly reliable performance. The Prony method describes the least performance due to the inconsistent performance.



(**a**) (**b**) (**c**) (**d**)

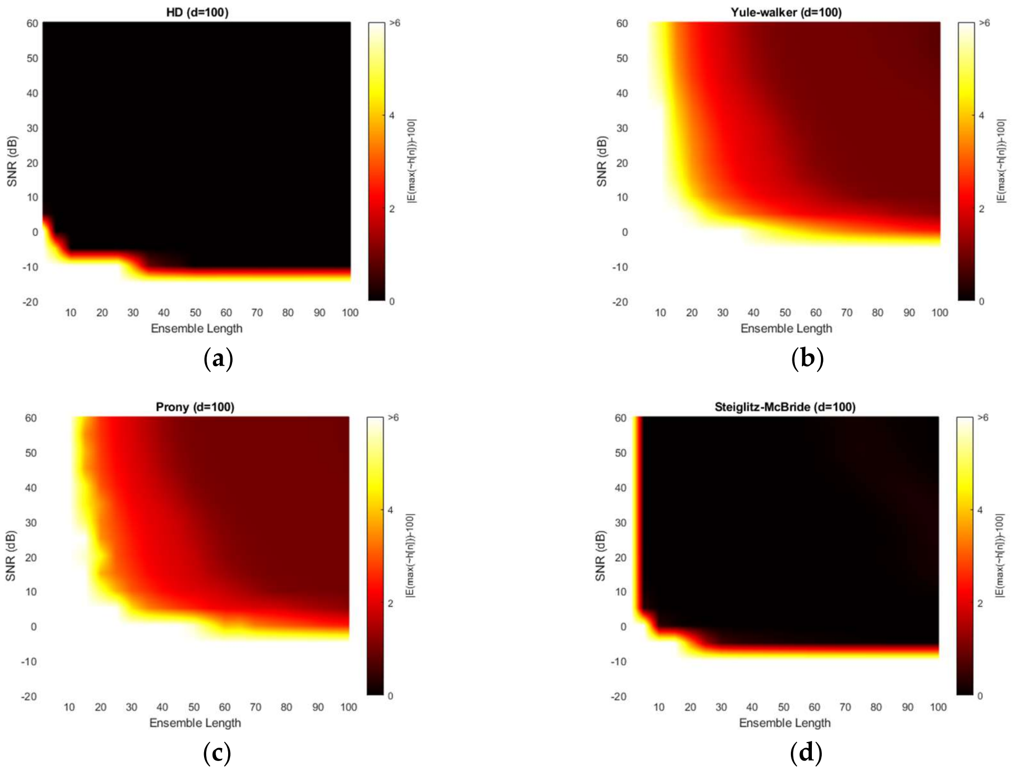
Figure 11. Average estimation error with given ensemble average length & SNR for 50-sample delay: (**a**) non-parametric HD; (**b**) Yule-walker HD; (**c**) Prony HD; (**d**) Steiglitz-McBride HD.



(**a**) (**b**) (**c**) (**d**)

Figure 12. Variance of estimation with given ensemble average length & SNR for 50-sample delay: (**a**) non-parametric HD; (**b**) Yule-walker HD; (**c**) Prony HD; (**d**) Steiglitz-McBride HD.

Figure 13 illustrates the absolute estimation error for 100-sample delay with various SNR and ensemble length. No significant differences are observed in the estimation bias from 50-sample delay performance. The non-parametric and Steiglitz-McBride method present the widespread near zero bias. The Yule-Walker and Prony method denote the relatively high estimation error for overall situations. The observed least bias is still around 0.5 sample in absolute value for middle distance microphone location (*d* = 100).



(**a**) (**b**)

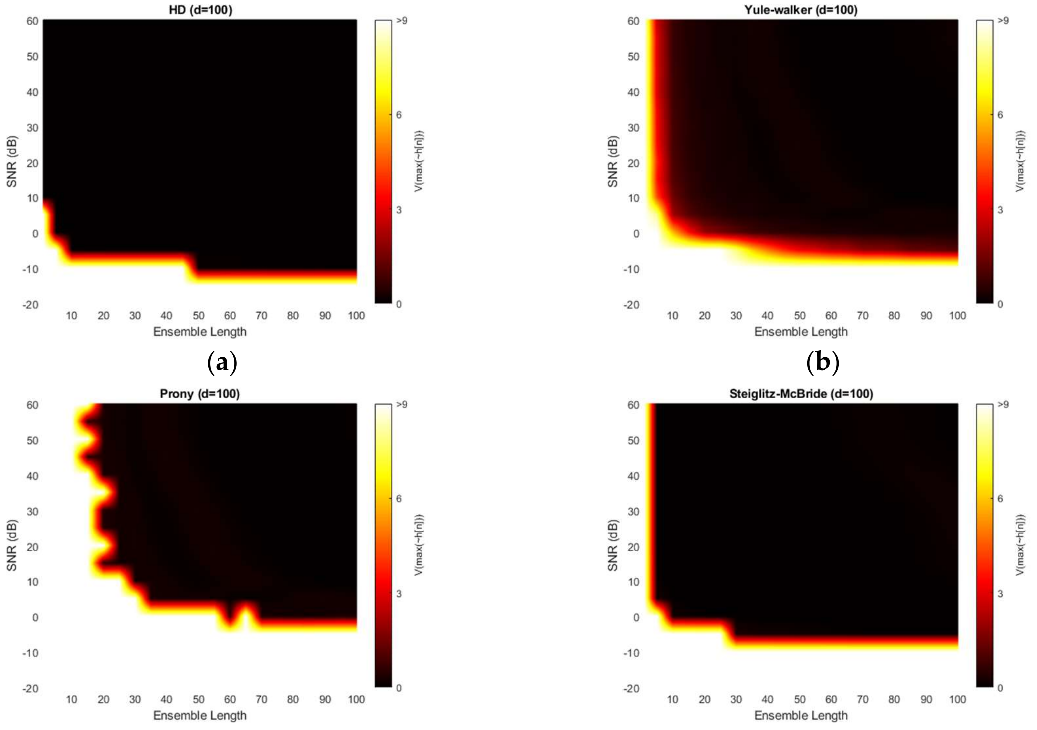
(**c**) (**d**)

Figure 13. Average estimation error with given ensemble average length & SNR for 100-sample delay: (**a**) non-parametric HD; (**b**) Yule-walker HD; (**c**) Prony HD; (**d**) Steiglitz-McBride HD.

Figure 14 depicts the variance for 100-sample delay with various SNR and ensemble length. Similar to the 50-sample delay simulation, all four methods provide the near zero variance for high SNR and long average length. Except the Prony method, the variance distribution is almost identical to the near microphone simulation (*d* = 50) counterpart. The Prony method demonstrates the wider white region and further uneven edges for high variance. The Prony method delivers the worse performance in middle distance microphone location (*d* = 100) for variance perspective.

Figure 15 presents the absolute estimation error for 200-sample delay with various SNR and ensemble length. The non-parametric and Steiglitz-McBride method illustrate the consistent performance for far microphone location (*d* = 200) as well. No considerable differences are visible from previous simulation outcomes. However, the Yule-Walker and Prony methods provide lower bias for the high SNR and long average length compare to the pervious simulations. Also, the black line is perceived in Yule-Walker and Prony methods. The local low bias does not signify good estimation performance because of the high variance in the given area as shown in Figure 16.

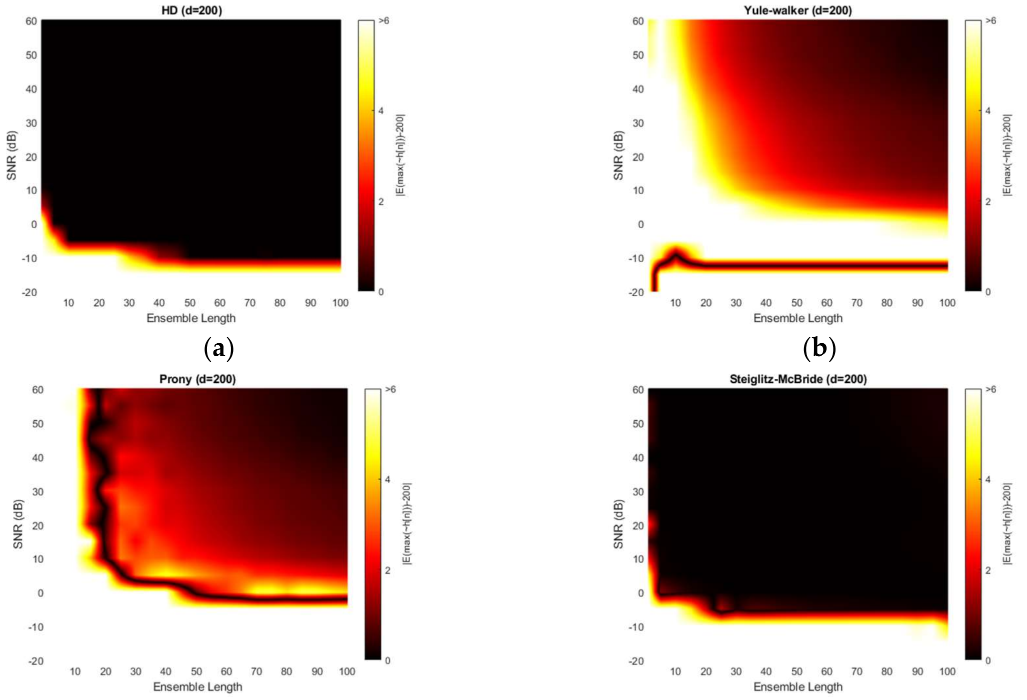
Figure 16 shows the variance for 200-sample delay with various SNR and ensemble length. Similar to the previous simulation outcomes, all four methods provide the near zero variance for high SNR and long average length. Except the Yule-Walker and Prony method, the variance distribution is almost identical to the near and middle microphone simulation counterpart. The Yule-walker method produces the slightly increased white area. The Prony method illustrates the wider and uneven area for high variance. The bias distribution from Figure 15 exhibits the decreased bias river in the low SNR and/or short average length area which corresponds to the high variance region. The reliable estimator requires the low bias as well as low variance.



(**a**) (**b**)

(**c**) (**d**)

Figure 14. Variance of estimation with given ensemble average length & SNR for 100-sample delay: (**a**) non-parametric HD; (**b**) Yule-walker HD; (**c**) Prony HD; (**d**) Steiglitz-McBride HD.



(**a**) (**b**)

(**c**) (**d**)

Figure 15. Average estimation error with given ensemble average length & SNR for 200-sample delay: (**a**) non-parametric HD; (**b**) Yule-walker HD; (**c**) Prony HD; (**d**) Steiglitz-McBride HD.

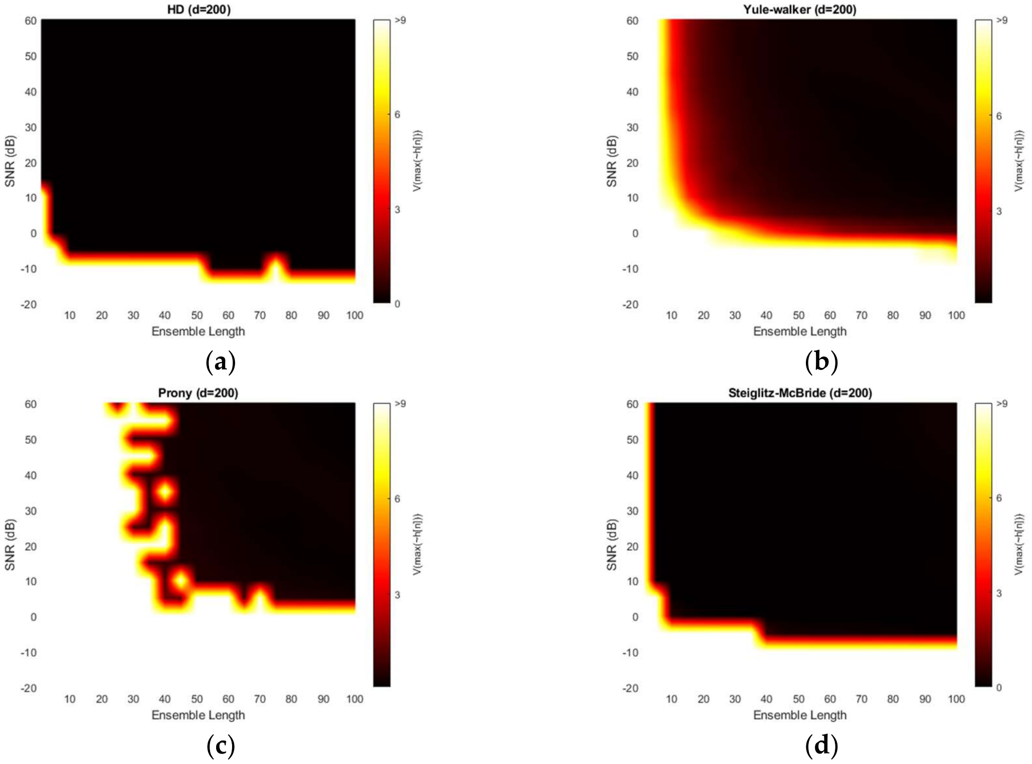


Figure 16. Variance of estimation with given ensemble average length & SNR for 200-sample delay: (**a**) non-parametric HD; (**b**) Yule-walker HD; (**c**) Prony HD; (**d**) Steiglitz-McBride HD.

4. Results

The acoustic experiments are performed and evaluate in an anechoic chamber that has been validated to demonstrate partial conformance with ISO 3745 [37] for the 250 Hz–16 kHz one-third octave band in a free-field chamber and for the 1–16 kHz one-third octave band in a hemi-free-field chamber [30]. The non-parametric and parametric HD algorithms are analyzed with the free-field chamber mode, which contains fully covered surfaces for all directions with acoustic wedges. Observe that the experiment configuration indicates the two microphones and one source with proper distance, and the HD algorithms denote the forward and inverse real cepstrum algorithm with or without parametric estimations. As shown in Figure 17, the directional alignment for microphones and transmitter is guided by the line laser (GLL 3-80 P, Bosch, Gerlingen, Germany) located above the sound source speaker. The first microphone is located at the direct-front direction 1.00 m away from the speaker. The distance between the microphones is controlled by the automatic relocation system, shown in Figure 17, based on the ball screw and stepping motor. The linear motion from the ball screw repositions the second microphone via employing the microprocessor (MSP430F5529LP, Texas Instruments, Dallas, TX, USA) based open loop control. The plastic parts are realized by the 3D printer (Replicator 2, MakerBot, Brooklyn, NY, USA) from polylactic acid (PLA) filament as illustrated in Figure 17.

The MATLAB programming operates the microphones (C-2, Behringer, Tortola, British Virgin Islands), computer-connected audio device (Quad-Capture, Roland, Hamamatsu, Japan), analog mixer (MX-1020, MPA Tech, Seoul, Republic of Korea), speaker (HS80M, Yamaha, Hamamatsu, Japan) and automatic relocation system simultaneously. The MATLAB system object with the audio stream input/output (ASIO) driver processes the real-time audio in terms of generation, reception, and execution. In the designated distance, the audio is recorded for 20 s with 48 kHz sampling rate. The first and last one second data is discarded to reduce the interruption by computational and environmental conditions. Therefore, the overall 18 s of recorded data is utilized for individual distance experiments. The individual data frame is organized by 1024 samples and the new ensemble average process is initiated after the 10 frames later. The rest of the frames are overlapped for data consistency.

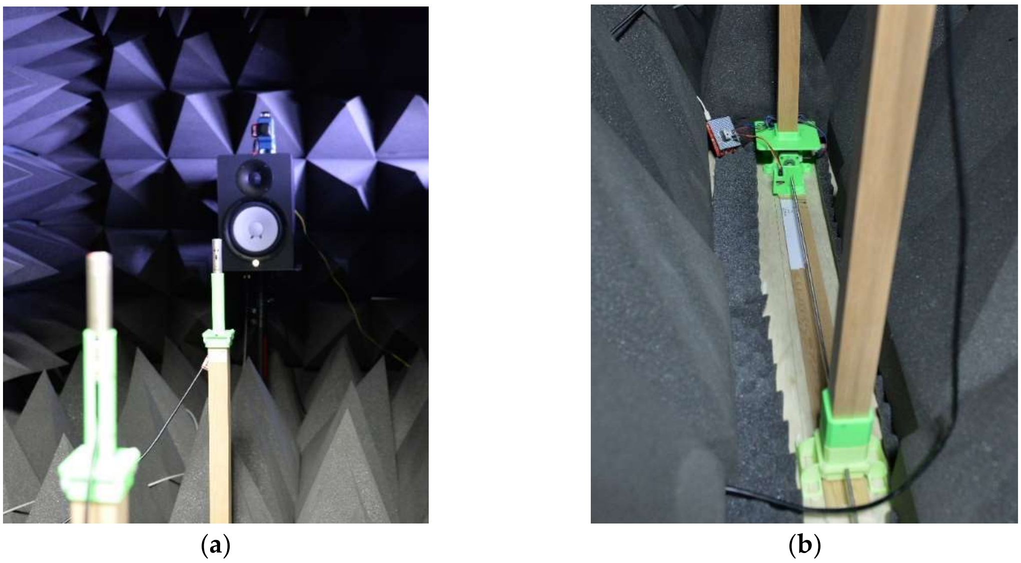

(**a**) (**b**)

Figure 17. Acoustic experiment in the anechoic chamber: (**a**) two microphones & speaker configuration with laser guidance; (**b**) automatic relocation system to control microphone distance.

In order to coordinate the experiments and simulations, the physical distance is converted for the discrete samples to measure the ToF. The resolution (samples/centimeter) Δ can be calculated from the sampling frequency and sound speed (at 25 °C) as shown below.

$$\Delta = \frac{f_s}{c} = \frac{48000}{34613} = 1.3868 \left(\frac{\text{samples}}{\text{centimeter}} \right) @\, 25°\text{C} \tag{61}$$

The recorded distances are arranged from 20 to 80 in 10 cm intervals as shown in Table 1. The corresponding distance in samples are computed by multiplying the resolution with the physical distances. Note that the calculated values are rounded to the nearest integer value. If there is one sample error in ToF estimation, 0.7211 cm $(1/\Delta)$ distance is away from the true position. The measurement accuracy is subject to the environment parameters such as temperature and sound speed.

Table 1. Conversion table for target distance in samples at 25 °C.

Target Distance(cm)	20	30	40	50	60	70	80
Target Distance (Samples)	28	42	55	69	83	97	111

In the experiments, the FILF window $w[n]$ is decided to eliminate the 25 sample below and maximum phase realization. Observe that the computed non-parametric and parametric HD output $\widetilde{h}[n]$ is normalized. Figure 18 denotes the estimated $\widetilde{h}[n]$ with 100 ensemble average length for given ToF situations. The non-parametric HD demonstrates the near zero values below the 25-sample delay and the maximum peak represents the expected delay position with distinction. The Yule-Walker method presents the peak at the desired position and the minor hill around 400 samples. The Prony method provides the peak location pointing to the desired delay. The slightly increased hill can be identified with a certain inconsistency in position. The Steiglitz-McBride method illustrates sharp peak for desired delay position with low base values.

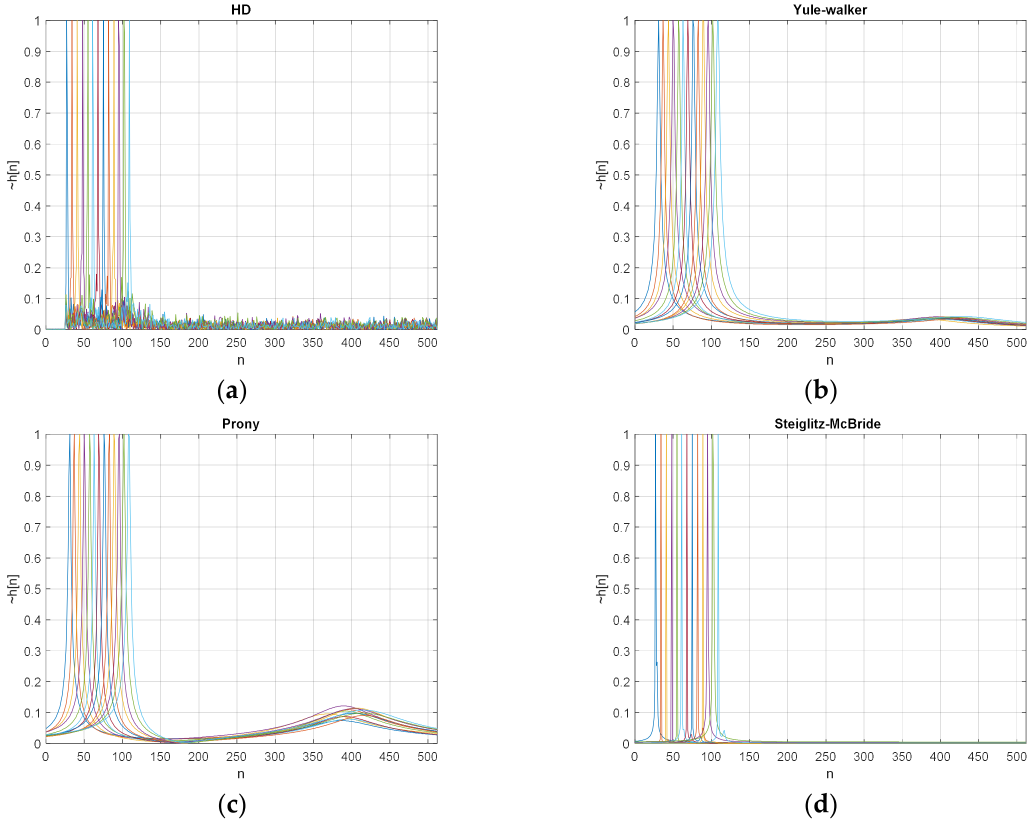


Figure 18. Estimated $\widetilde{h}[n]$ distribution (normalized) with 100 ensemble average for ToF (20~80 cm in 10 cm interval) from: (**a**) non-parametric HD; (**b**) Yule-walker HD; (**c**) Prony HD; (**d**) Steiglitz-McBride HD.

Table 2 provides the statistical performance for short ensemble average length as 20 frames. The non-parametric and Steiglitz-McBride method shows the low bias and variance in overall. The consistent discrepancy between the measurements and estimations are detected at 55, 97 and 111 sample distance in target. We assume that the estimated values are correct in the situations because of the susceptible measurement conditions. The Yule-Walker and Prony methods represent the high bias with variance. The majority situation shows the consistent high estimation error due to the low variance except the Prony at 111 sample distance. A single iteration of the 83 Prony results generates the significant second mainlobe for estimation around 400 sample distance which contributes to the high variance.

Table 2. Experiment results as mean (upper row) and variance (lower row) for 20 ensemble length.

Target	28	42	55	69	83	97	111
HD	28 0	42 0	56 0	69 0	83 0	96.01 0.0100	110 0
Yule-walker	34.96 0.4499	47.57 0.5413	61.28 0.4467	73.34 0.2751	86.52 0.4478	99.18 0.4182	111.57 0.7852
Prony	34.98 0.4628	47.57 0.5413	61.27 0.4655	73.25 0.2889	86.45 0.3476	98.99 0.4023	115.01 1164.0364
Steiglitz-McBride	28.94 0.0573	42 0	56 0	69 0	83 0	96.67 0.2220	110 0

Table 3 demonstrates the statistical performance for middle ensemble average length as 50 frames. The non-parametric HD delivers the zero bias and zero variance based on the assumption in previous paragraph. The Yule-Walker and Prony methods produce the substantially reduced bias and variance than the short ensemble length counterparts. For both methods, the estimation biases exhibit high values in low time delay conditions and the variances depict the coherent values for entire range. The variance abruption in Prony method is removed and stabilized in 50 ensemble length. The Steiglitz-McBride method establishes the similar performance as the short ensemble length with minor degradation at 28- and 97-sample delay cases.

Table 3. Experiment results as mean (upper row) and variance (lower row) for 50 ensemble length.

Target	28	42	55	69	83	97	111
HD	28 0	42 0	56 0	69 0	83 0	96 0	110 0
Yule-walker	32.59 0.2454	45.25 0.1899	59.01 0.1138	71.05 0.0481	84.36 0.2593	97.22 0.2019	109.96 0.2391
Prony	32.63 0.2373	45.26 0.1960	59 0.1013	71.03 0.0247	84.38 0.2373	97.13 0.1867	109.91 0.2581
Steiglitz-McBride	29.39 1.3543	42 0	56 0	69 0	83 0	96.42 0.7285	110 0

Table 4 delivers the statistical performance for long ensemble average length as 100 frames. The non-parametric and Steiglitz-McBride illustrate the zero bias and zero variance. The Yule-Walker and Prony method generate the improved bias and stabilized variance. The mean value of both methods further approaches the measured number than the previous experiments. Therefore, long ensemble average length provides the better statistical performance for non-parametric and parametric HD algorithms. Observe that the 100-frame ensemble length is equivalent to the 2.1333 s data.

Table 4. Experiment results as mean (upper row) and variance (lower row) for 100 ensemble length.

Target	28	42	55	69	83	97	111
HD	28 0	42 0	56 0	69 0	83 0	96 0	110 0
Yule-walker	31.84 0.1362	44.49 0.2533	58.04 0.0389	70.08 0.0746	83.69 0.2155	96.55 0.2512	109.33 0.2252
Prony	31.85 0.1268	44.53 0.2523	58.04 0.0389	70.07 0.0631	83.75 0.1917	96.36 0.2335	109.25 0.1917
Steiglitz-McBride	28 0	42 0	56 0	69 0	83 0	96 0	110 0

The introduced non-parametric and parametric HD algorithms well derive the ToF information in most of situations. Note that the output of non-parametric HD in this experiment is represented by 1024 samples. However, the output of Prony HD and Steiglitz-McBride HD is denoted by 5 coefficients as $a_{1,}$ a_2, b_0, b_1, and b_2 according to the ARMA model shown in Equation (10). Also, the output of Yule-Walker HD is indicated by two coefficients as a_1 and a_2 along with the AR model demonstrated in Equation (8). The given length of output will be served as the featured information for complete SSL system based on the machine learning (or deep learning) in future works.

5. Conclusions

This paper presents a novel time of flight estimation method based on the non-parametric and parametric homomorphic deconvolutions. The non-parametric homomorphic deconvolution is realized by the forward and inverse real cepstrum with frequency-invariant linear filtering. The proper window configuration removes the low time delay components and maximum phase realization. The numerical distribution of non-parametric homomorphic deconvolution produces the likelihood of time of flight information between the two microphones. Therefore, the time location corresponding to the maximum value indicates the estimation. The parametric methods replace the last inverse Fourier transform of non-parametric homomorphic deconvolution by parametric estimation algorithms as Yule-walker, Prony, and Steiglitz-McBride. The derived complex number coefficients for the parametric method properly specify the time of flight position by maximum value. Observe that the parametric methods require the evaluation process for numerical distribution. The simulations for various signal-to-noise ratio and ensemble average length present non-statistical and statistical performance outcomes. The non-parametric homomorphic deconvolution and Steiglitz-McBride methods illustrate the consistent distribution with low bias and variance overall. The Yule-Walker and Prony methods are vulnerable to the simulation conditions. The high signal-to-noise ratio and long ensemble length generally denote the better statistical performance in both methods. The experiments in anechoic chamber also delivers the comparable results as the simulation. The increased ensemble length shows the zero bias and variance for non-parametric homomorphic deconvolution and Steiglitz-McBride method. The Yule-Walker and Prony methods display the performance improvement in terms of bias and variance for longer length conditions.

Sound localization for the transport is a challenging topic because of the complexity induced by the structure and environment. This paper builds upon the fundamentals for sound localization by time of flight estimation based on non-parametric and parametric homomorphic deconvolution. Future works will offer optimal microphone array configuration with homomorphic deconvolution algorithm to maximize the localization performance by using the machine learning and deep learning approaches. Non-parametric homomorphic deconvolution is appropriate for the applying the deep learning method. The parametric methods are designed for the machine learning algorithms because the features are extracted by signal modelling and corresponding coefficients. This paper verifies the algorithm functionality by statistical simulations and experiments. However, the algorithm performance could be appended by the different perspective from machine/deep learning algorithms. In other words, the least performance method in this paper could be proved for improved realization by chance. The extensive simulation and experiments are required to obtain the receiver structure, homomorphic deconvolution, estimation order, learning method, and etc. Note that the problem size by receiver number is not correlated to the localization system complexity at all. Future papers will discuss the feasible localization system in detail based on this article outcome.

Author Contributions: Conceptualization, Y.P. and K.K.; methodology, K.K.; software, K.K.; validation, Y.P., K.K. and A.C.; formal analysis, Y.P.; investigation, K.K.; resources, Y.P.; data curation, Y.P.; writing—original draft preparation, K.K.; writing—review and editing, Y.P. and A.C.; visualization, Y.P.; supervision, K.K.; project administration, A.C.; funding acquisition, K.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded in part by NASA, grant number NNX15AP85H.

Acknowledgments: This work was supported by the Dongguk University Research Fund of 2020.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

The equations below present the non-parametric HD based on the real cepstrum in z domain. The received signal $Y(z)$ is arrived at two receivers with time delay d samples as shown in $H(z)$. The sound source $X(z)$ is described by poles α_k and zeros β_k inside the unit circle as well as the zeros γ_k outside the unit circle.

$$\begin{gathered} Y(z) = X(z)H(z) = \frac{A\prod_{k=1}^{M_i}\left(1-\beta_k z^{-1}\right)\prod_{k=1}^{M_o}\left(1-\gamma_k^{-1} z^{-1}\right)}{\prod_{k=1}^{N}\left(1-\alpha_k z^{-1}\right)}\left(1-\mu z^{-d}\right) \\ |\alpha_k| < 1,\ \left|\beta_k\right| < 1,\ \left|\gamma_k\right| < 1,\ \left|\mu\right| < 1 \end{gathered} \tag{A1}$$

Organize the equation into that below to represent the sequences composed of a stable sum exponential.

$$\begin{aligned} Y(z) &= \frac{A\prod_{k=1}^{M_i}\left(1-\beta_k z^{-1}\right)\prod_{k=1}^{M_o}\gamma_k^{-1}z^{-M_o}\left(\gamma_k z-1\right)}{\prod_{k=1}^{N}\left(1-\alpha_k z^{-1}\right)}\left(1-\mu z^{-d}\right) \\ &= \frac{Az^{-M_o}\prod_{k=1}^{M_o}\left(-\gamma_k^{-1}\right)\prod_{k=1}^{M_i}\left(1-\beta_k z^{-1}\right)\prod_{k=1}^{M_o}\left(1-\gamma_k z\right)}{\prod_{k=1}^{N}\left(1-\alpha_k z^{-1}\right)}\left(1-\mu z^{-d}\right) \end{aligned} \tag{A2}$$

Factorize into a product of minimum phase and maximum phase signals as below.

$$Y(z) = z^{-M_o}X_{min}(z)X_{max}(z)H(z) \tag{A3}$$

$$X_{min}(z) = \frac{A\prod_{k=1}^{M_i}\left(1-\beta_k z^{-1}\right)}{\prod_{k=1}^{N}\left(1-\alpha_k z^{-1}\right)} \tag{A4}$$

$$X_{max}(z) = \prod_{k=1}^{M_o}\left(-\gamma_k^{-1}\right)\prod_{k=1}^{M_o}\left(1-\gamma_k z\right) \tag{A5}$$

Apply logarithm of $Y(z)$ is below. The product of terms is transformed to the sum of logarithmic terms.

$$\begin{aligned} \log(Y(z)) = \hat{Y}(z) = \log|A| + \log\left(z^{-M_o}\right) + \textstyle\sum_{k=1}^{M_o}\log\left|\gamma_k^{-1}\right| + \textstyle\sum_{k=1}^{M_i}\log\left(1-\beta_k z^{-1}\right) \\ + \sum_{k=1}^{M_o}\log\left(1-\gamma_k z\right) + \log\left(1-\mu z^{-d}\right) - \sum_{k=1}^{N}\log\left(1-\alpha_k z^{-1}\right) \end{aligned} \tag{A6}$$

The Taylor series for the natural logarithm is given below known as Mercator series [38].

$$\log(1-z) = -\sum_{n=1}^{\infty}\frac{z^n}{n} \quad \text{for } |z| < 1 \tag{A7}$$

Employ the Mercator series for each logarithmic term as below.

$$-\sum_{m=1}^{\infty}\frac{\beta_k^m}{m}\delta[n-m] \overset{z}{\leftrightarrow} \log\left(1-\beta_k z^{-1}\right) = -\sum_{m=1}^{\infty}\frac{\beta_k^m z^{-m}}{m} \quad \text{for } |z| > \left|\beta_k\right| \tag{A8}$$

$$-\sum_{m=1}^{\infty}\frac{\gamma_k^m}{m}\delta[n+m] \overset{z}{\leftrightarrow} \log\left(1-\gamma_k z\right) = -\sum_{m=1}^{\infty}\frac{\gamma_k^m z^m}{m} \quad \text{for } |z| < \left|\gamma_k^{-1}\right| \tag{A9}$$

The inverse Z-transform from the Mercator series provides the complex cepstrum $\hat{y}[n]$ as below. It is important to note that the magnitude of each term is decreased by power of poles/zeros at least as fast as $1/|n|$. However, the term induced by time delay d is decaying at the rate of $1/|n|$ in every d sample stride. Therefore, the $\hat{h}[n]$ represents significantly slow decline rate.

$$\begin{aligned} \hat{y}[n] = \left\{\log|A| + \sum_{k=1}^{M_o}\log\left|\gamma_k^{-1}\right|\right\}\delta[n] + \sum_{m=1}^{\infty}\left(-\sum_{k=1}^{M_i}\frac{\beta_k^m}{m} + \sum_{k=1}^{N}\frac{\alpha_k^m}{m}\right)\delta[n-m] \\ - \sum_{m=1}^{\infty}\frac{\mu^m}{m}\delta[n-dm] - \sum_{m=1}^{\infty}\left(\sum_{k=1}^{M_o}\frac{\gamma_k^m}{m}\right)\delta[n+m] \end{aligned} \tag{A10}$$

The relation between the real cepstrum $c_y[n]$ and complex cepstrum $\hat{y}[n]$ is below.

$$c_y[n] = \frac{\hat{y}[n] + \hat{y}[-n]}{2} \tag{A11}$$

The derived real cepstrum is below.

$$\begin{aligned} c_y[n] = \left\{\log|A| + \sum_{k=1}^{M_o}\log\left|\gamma_k^{-1}\right|\right\}\delta[n] + \frac{1}{2}\sum_{m=1}^{\infty}\left(-\sum_{k=1}^{M_i}\frac{\beta_k^m}{m} + \sum_{k=1}^{N}\frac{\alpha_k^m}{m}\right)\delta[n\mp m] \\ -\frac{1}{2}\sum_{m=1}^{\infty}\frac{\mu^m}{m}\delta[n\mp dm] - \frac{1}{2}\sum_{m=1}^{\infty}\left(\sum_{k=1}^{M_o}\frac{\gamma_k^m}{m}\right)\delta[n\pm m] \end{aligned} \tag{A12}$$

The FILF is defined as below. The window function $w[n]$ simply removes the undesired low sample delays (below l in this example) as well as maximum phase realization which is represented by the anti-causal components. Note that the HD based on the DFT denotes the high half of the sequence as the anti-causal components due to the circular property of the DFT.

$$\hat{h}[n] = c_y[n]w[n] \tag{A13}$$

$$w[n] = 2u[n-l] \;\; \text{where } l \leq d \tag{A14}$$

After the FILF, the slow decaying exponential sequence is survived from the filtering. The anti-causal components are eliminated by the window as well.

$$\hat{h}[n] \approx -\sum_{m=1}^{\infty} \frac{\mu^m}{m}\delta[n-dm] \tag{A15}$$

Observe as below that the other sequences are decreasing the magnitude at the rate of $1/|n|$ for rapid converging.

$$\frac{1}{2}\sum_{m=1}^{\infty}\left(-\sum_{k=1}^{M_i}\frac{\beta_k^m}{m} + \sum_{k=1}^{N}\frac{\alpha_k^m}{m}\right)\delta[n-m] \approx 0 \;\; for\; m \geq l \tag{A16}$$

$$-\frac{1}{2}\sum_{m=1}^{\infty}\left(\sum_{k=1}^{M_o}\frac{\gamma_k^m}{m}\right)\delta[n-m] \approx 0 \;\; for\; m \geq l \tag{A17}$$

Apply the Z-transform and use the Mercator series to combine into the logarithmic term as below.

$$\hat{H}(z) \approx -\sum_{m=1}^{\infty}\frac{\mu^m z^{-dm}}{m} = -\sum_{m=1}^{\infty}\frac{\left(\mu z^{-d}\right)^m}{m} = \log\left(1-\mu z^{-d}\right) \tag{A18}$$

Employ the exponential function over the given $\hat{H}(z)$ to obtain the simple polynomial.

$$e^{\hat{H}(z)} \approx e^{\log\left(1-\mu z^{-d}\right)} = \left(1-\mu z^{-d}\right) \tag{A19}$$

The inverse Z-transform provides the estimated propagation function $\widetilde{h}[n]$ as below.

$$\widetilde{h}[n] \approx \delta[n] - \mu\delta[n-d] \tag{A20}$$

In this paper, the last computational stage of HD algorithm is realized by forward transformation instead of inverse transformation as shown below. Therefore, the final stage of HD can be selected from FFT, Yule-walker, Prony, and Steiglitz-McBride algorithms for non-parametric and parametric methods.

$$\widetilde{h}[n] \overset{\text{FFT}}{\leftrightarrow} e^{\hat{H}[k]} \text{ for } N = 2^k \;\; k \in \mathbb{N} \tag{A21}$$

The implementation of inverse transformation based on the forward can be derived as below. The conjugated input to the forward transformation generates the conjugated inverse transformation with scale.

$$\widetilde{h}[n] = \frac{1}{N}\sum_{k=0}^{N-1} e^{\hat{H}[k]}e^{j\frac{2\pi}{N}kn} = \frac{1}{N}\left\{\sum_{k=0}^{N-1}\left(e^{\hat{H}[k]}\right)^{*}e^{-j\frac{2\pi}{N}kn}\right\}^{*} = \frac{1}{N}\left\{\text{FFT}_N\left(e^{\hat{H}[k]}\right)^{*}\right\}^{*} \tag{A22}$$

The magnitude of the HD algorithm is required to determine the ToF. Therefore, the absolute value of the output is given as below.

$$\left|\widetilde{h}[n]\right| = \frac{1}{N}\left|\text{FFT}_N\left(e^{\hat{H}[k]}\right)^{*}\right| \tag{A23}$$

Below is the example by numbers. The sound source $X(z)$ consists of poles and zeros at 0.9, 0.7, $1.25e^{-j\frac{\pi}{4}}$, $1.25e^{j\frac{\pi}{4}}$. The propagation function $H(z)$ shows the time delay at 30 samples with 0.8 magnitude decreasing.

$$Y(z) = X(z)H(z) = \frac{\left(1-0.7z^{-1}\right)\left(1-1.25e^{-j\frac{\pi}{4}}z^{-1}\right)\left(1-1.25e^{j\frac{\pi}{4}}z^{-1}\right)}{\left(1-0.9z^{-1}\right)}\left(1-0.8z^{-30}\right)$$

Organize the equation to represent the sequences composed of a stable sum exponential.

$$Y(z) = \frac{\left(1-0.7z^{-1}\right)1.25^2z^{-2}\left(1-0.8e^{j\frac{\pi}{4}}z\right)\left(1-0.8e^{-j\frac{\pi}{4}}z\right)}{\left(1-0.9z^{-1}\right)}\left(1-0.8z^{-30}\right)$$

The corresponding real cepstrum is below. Note that the maximum phase realization is represented by anti-causal components.

$$\begin{gathered}\hat{y}[n] = 2\log(1.25)\delta[n] + \sum_{m=1}^{\infty}\sum_{m=1}^{\infty}\left(-\tfrac{0.7^m}{m} + \tfrac{0.9^m}{m}\right)\delta[n-m] \\ -\sum_{m=1}^{\infty}\tfrac{0.8^m}{m}\delta[n-30m] - \sum_{m=1}^{\infty}\left(\tfrac{0.8^m e^{j\frac{\pi}{4}m}}{m} + \tfrac{0.8^m e^{-j\frac{\pi}{4}m}}{m}\right)\delta[n+m]\end{gathered}$$

The real cepstrum is below based on the above equation.

$$\begin{gathered}c_y[n] = 2\log(1.25)\delta[n] + \tfrac{1}{2}\sum_{m=1}^{\infty}\left(-\tfrac{0.7^m}{m} + \tfrac{0.9^m}{m}\right)\delta[n\mp m] \\ -\tfrac{1}{2}\sum_{m=1}^{\infty}\tfrac{0.8^m}{m}\delta[n\mp 30m] - \sum_{m=1}^{\infty}\left(\tfrac{0.8^m}{m}\cos\left(\tfrac{\pi}{4}m\right)\right)\delta[n\pm m]\end{gathered}$$

Apply the FILF as below.

$$\hat{h}[n] = c_y[n]w[n]$$

The window function $w[n]$ is designed to remove the time delay below 25 samples which maintains the given propagation function $h[n]$.

$$w[n] = 2u[n-l] \;\text{ where } l \le d,\; l = 25,\; d = 30$$

Employ the FFT for inverse cepstrum as below.

$$\hat{H}[k] = \mathrm{FFT}_{512}\left(-\frac{1}{2}\sum_{m=1}^{3}\frac{0.8^m}{m}\delta[n-30m]\right)$$

Use the exponential function and FFT to obtain the final HD output.

$$\widetilde{h}[n] = \frac{1}{N}\left\{\mathrm{FFT}_{512}\left(e^{\hat{H}[k]}\right)^*\right\}^*$$

Figure A1 demonstrates the computational procedures for given numerical example above. The complex cepstrum illustrates the causal component from minimum phase realization and anti-causal component from maximum phase realization. The real cepstrum combines the complex cepstrum with time reflection for even symmetric distribution. Also, the window function $w[n]$ and FILF output are also depicted to remove the sound source $x[n]$. The HD output $\widetilde{h}[n]$ estimates the time delay by investigating the distribution.

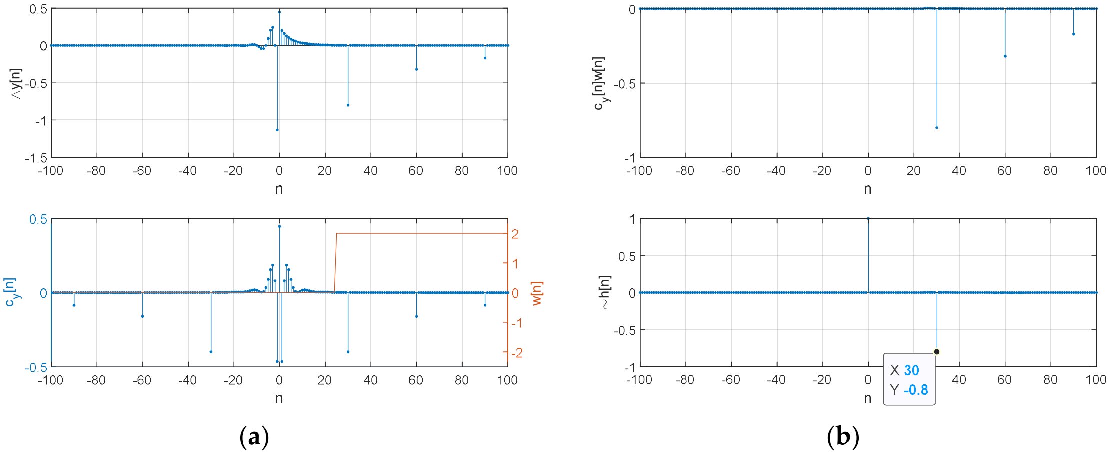


Figure A1. HD computational procedures for numerical example: (**a**) Complex cepstrum (above) and real cepstrum (below); (**b**) FILF output (above) and estimated propagation function (below).

References

1. Van Veen, B.; Buckley, K. Beamforming: a versatile approach to spatial filtering. *IEEE ASSP Mag.* **1988**, *5*, 4–24. [CrossRef]
2. Nakashima, H.; Mukai, T. 3D Sound Source Localization System Based on Learning of Binaural Hearing. In Proceedings of the 2005 IEEE International Conference on Systems, Man and Cybernetics; Institute of Electrical and Electronics Engineers (IEEE), Waikoloa, HI, USA, 12 October 2005; Volume 4, pp. 3534–3539.
3. Kumon, M.; Shimoda, T.; Kohzawa, R.; Mizumoto, I.; Iwai, Z. Audio servo for robotic systems with pinnae. In Proceedings of the 2005 IEEE/RSJ International Conference on Intelligent Robots and Systems, Edmonton, AB, Canada, 2–6 August 2005; Institute of Electrical and Electronics Engineers (IEEE): Piscataway, NJ, USA, 2005; pp. 1881–1886.
4. Hwang, S.; Park, Y.; Park, Y.-S. Sound direction estimation using an artificial ear for robots. *Robot. Auton. Syst.* **2011**, *59*, 208–217. [CrossRef]
5. Tomoko, S.; Toru, N.; Makoto, K.; Ryuichi, K.; Ikuro, M.; Zenta, I. Spectral Cues for Robust Sound Localization with Pinnae. In Proceedings of the 2006 IEEE/RSJ International Conference on Intelligent Robots and Systems, Beijing, China, 9–15 October 2006; pp. 386–391.
6. Saxena, A.; Ng, A.Y. Learning sound location from a single microphone. In Proceedings of the 2009 IEEE International Conference on Robotics and Automation, Kobe, Japan, 12–17 May 2009; Institute of Electrical and Electronics Engineers (IEEE): Piscataway, NJ, USA, 2009; pp. 1737–1742.
7. Kumon, M.; Noda, Y. Active soft pinnae for robots. In Proceedings of the 2011 IEEE/RSJ International Conference on Intelligent Robots and Systems, San Francisco, CA, USA, 25–30 September 2011; pp. 112–117.
8. Jang, Y.; Kim, J.; Kim, J. The development of the vehicle sound source localization system. In Proceedings of the 2015 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA), Hong Kong, 16–19 December 2015; Institute of Electrical and Electronics Engineers (IEEE): Piscataway, NJ, USA, 2015; pp. 1241–1244.
9. Mizumachi, M.; Kaminuma, A.; Ono, N.; Ando, S. Robust Sensing of Approaching Vehicles Relying on Acoustic Cues. *Sensors* **2014**, *14*, 9546–9561. [CrossRef] [PubMed]
10. Liu, H.; Li, B.; Yuan, X.; Zhou, Q.; Huang, J. A Robust Real Time Direction-of-Arrival Estimation Method for Sequential Movement Events of Vehicles. *Sensors* **2018**, *18*, 992. [CrossRef]
11. Oppenheim, A.V. Generalized superposition. *Inf. Control.* **1967**, *11*, 528–536. [CrossRef]
12. Gold, B.; Rader, C.M. *Digital Processing of Signals*; McGraw-Hill: New York, NY, USA, 1969.
13. Alexey, N. All World VISIO. Available online: https://surrogate-tm.github.io/digitall/index_en.html (accessed on 28 November 2019).
14. Dong, L.; Zou, W.; Li, X.; Shu, W.; Wang, Z. Collaborative localization method using analytical and iterative solutions for microseismic/acoustic emission sources in the rockmass structure for underground mining. *Eng. Fract. Mech.* **2019**, *210*, 95–112. [CrossRef]

15. Hu, Q.; Dong, L. Acoustic emission source location and experimental verification for two-dimensional irregular complex structure. *IEEE Sens. J.* **2019**. [CrossRef]
16. Dong, L.; Shu, W.; Li, X.; Han, G.; Zou, W. Three Dimensional Comprehensive Analytical Solutions for Locating Sources of Sensor Networks in Unknown Velocity Mining System. *IEEE Access* **2017**, *5*, 11337–11351. [CrossRef]
17. Hoshiba, K.; Washizaki, K.; Wakabayashi, M.; Ishiki, T.; Kumon, M.; Bando, Y.; Gabriel, D.; Nakadai, K.; Okuno, H.G. Design of UAV-Embedded Microphone Array System for Sound Source Localization in Outdoor Environments. *Sensors* **2017**, *17*, 2535. [CrossRef]
18. An, I.; Lee, H.; Jo, B.; Choi, J.-W.; Yoon, S.-E. Robust Sound Source Localization considering Similarity of Back-Propagation Signals. 2019. Available online: http://a.xueshu.baidu.com/usercenter/paper/show?paperid=1k4u0cm0ps230cu06r100t200s276435 (accessed on 5 February 2020).
19. Xenaki, A.; Boldt, J.B.; Christensen, M.G. Sound source localization and speech enhancement with sparse Bayesian learning beamforming. *J. Acoust. Soc. Am.* **2018**, *143*, 3912–3921. [CrossRef]
20. Rascon, C.; Meza, I.; Ruiz, I.V.M. Localization of sound sources in robotics: A review. *Robot. Auton. Syst.* **2017**, *96*, 184–210. [CrossRef]
21. Li, Y.; Chen, H. Reverberation Robust Feature Extraction for Sound Source Localization Using a Small-Sized Microphone Array. *IEEE Sens. J.* **2017**, *17*, 6331–6339. [CrossRef]
22. Takeda, R.; Komatani, K. Sound source localization based on deep neural networks with directional activate function exploiting phase information. In Proceedings of the 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Shanghai, China, 20–25 March 2016; Institute of Electrical and Electronics Engineers (IEEE): Piscataway, NJ, USA, 2016; pp. 405–409.
23. Salvati, D.; Drioli, C.; Foresti, G.L. A weighted MVDR beamformer based on SVM learning for sound source localization. *Pattern Recognit. Lett.* **2016**, *84*, 15–21. [CrossRef]
24. Huang, Z.; Xu, J.; Gong, Z.; Wang, H.; Yan, Y. Multiple Source Localization in a Shallow Water Waveguide Exploiting Subarray Beamforming and Deep Neural Networks. *Sensors* **2019**, *19*, 4768. [CrossRef] [PubMed]
25. Kim, K.; Kim, Y. Monaural Sound Localization Based on Structure-Induced Acoustic Resonance. *Sensors* **2015**, *15*, 3872–3895. [CrossRef]
26. Kim, Y.; Kim, K. Near-Field Sound Localization Based on the Small Profile Monaural Structure. *Sensors* **2015**, *15*, 28742–28763. [CrossRef]
27. Park, Y.; Choi, A.; Kim, K. Monaural Sound Localization Based on Reflective Structure and Homomorphic Deconvolution. *Sensors* **2017**, *17*, 2189. [CrossRef]
28. Kim, K.; Choi, A. Binaural Sound Localizer for Azimuthal Movement Detection Based on Diffraction. *Sensors* **2012**, *12*, 10584–10603. [CrossRef]
29. Kim, K. Lightweight Filter Architecture for Energy Efficient Mobile Vehicle Localization Based on a Distributed Acoustic Sensor Network. *Sensors* **2013**, *13*, 11314–11335. [CrossRef]
30. Kim, K.-W. Design and Analysis of Experimental Anechoic Chamber for Localization. *J. Acoust. Soc. Korea* **2012**, *31*, 225–234. [CrossRef]
31. Rabiner, L.R.; Schafer, R.W. *Theory and Applications of Digital Speech Processing*; Pearson: London, UK, 2011.
32. Oppenheim, A.V.; Schafer, R.W. *Discrete-Time Signal Processing*; Prentice Hall: Upper Saddle River, NJ, USA, 1989.
33. Stoica, P.; Moses, R.L. *Introduction to Spectral Analysis*; Prentice Hall: Upper Saddle River, NJ, USA, 1997.
34. Yule, G.U., VII. On a method of investigating periodicities disturbed series, with special reference to Wolfer's sunspot numbers. *Philos. Trans. R. Soc. Lond. Ser. A Contain. Pap. Math. Phys. Character (1896–1934)* **1927**, *226*, 267–298.
35. Parks, T.W.; Burrus, C.S. *Digital Filter Design*; Wiley: Hoboken, NJ, USA, 1987.
36. Steiglitz, K.; McBride, L.E. A technique for the identification of linear systems. *IEEE Trans. Autom. Control.* **1965**, *10*, 461–464. [CrossRef]
37. International Organization for Standardization. *Acoustics—Determination of Sound Power Levels of Noise Sources Using Sound Pressure—Precision Methods for Anechoic and Hemi-Anechoic Rooms*; ISO 3745:2003; ISO: Geneva, Switzerland, 2003.
38. Zwillinger, D. *CRC Standard Mathematical Tables and Formulae*, 31st ed.; CRC Press: Boca Raton, FL, USA, 2002.