sensors

MDPI

Article

Study of Generalized Phase Spectrum Time Delay Estimation Method for Source Positioning in Small Room Acoustic Environment

Vladimir Faerman [1,*], Valeriy Avramchuk [2], Kirill Voevodin [2], Ivan Sidorov [3] and Evgeny Kostyuchenko [1]

[1] Laboratory for Acquisition, Processing and Manipulating Biological Signals, Institute of System Integration and Security, Tomsk State University of Control Systems and Radioelectronics, 40 Lenina Ave., 634050 Tomsk, Russia; key@fb.tusur.ru
[2] Department of Complex Information Security of Computer Systems, Faculty of Security, Tomsk State University of Control Systems and Radioelectronics, 40 Lenina Ave., 634050 Tomsk, Russia; avs@fb.tusur.ru (V.A.); Kirill.Voevodin2@infotecs.ru (K.V.)
[3] Irkutsk Supercomputer Center of SB RAS, 134, Lermontova, 664033 Irkutsk, Russia; ivan.sidorov@icc.ru
* Correspondence: fva@fb.tusur.ru; Tel.: +7-3822-41-43-26

Abstract: This paper considers the application of signal processing methods to passive indoor positioning with acoustics microphones. The key aspect of this problem is time-delay estimation (TDE) that is used to get the time difference of arrival of the source's signal between the pair of distributed microphones. This paper studies the approach based on generalized phase spectrum (GPS) TDE methods. These methods use frequency-domain information about the received signals that make them different from widely applied generalized cross-correlation (GCC) methods. Despite the more challenging implementation, GPS TDE methods can be less demanding on computational resources and memory than conventional GCC ones. We propose an algorithmic implementation of a GPS estimator and study the various frequency weighting options in applications to TDE in a small room acoustic environment. The study shows that the GPS method is a reliable option for small acoustically dead rooms and could be effectively applied in presence of moderate in-band noises. However, GPS estimators are far less efficient in less acoustically dead environments, where other TDE options should be considered. The distinguishing feature of the proposed solution is the ability to get the time delay using a limited number of the adjusted bins. The solution could be useful for passively locating moving emitters of narrow-band continual noises using computationally simple frequency detection algorithms.

Keywords: generalized phase spectrum; time delay estimation; indoor positioning; room acoustics; sensors array

check for updates

Citation: Faerman, V.; Avramchuk, V.; Voevodin, K.; Sidorov, I.; Kostyuchenko, E. Study of Generalized Phase Spectrum Time Delay Estimation Method for Source Positioning in Small Room Acoustic Environment. *Sensors* **2022**, *22*, 965. https://doi.org/10.3390/s22030965

Academic Editors: Jari Nurmi and Riccardo Carotenuto

Received: 18 November 2021
Accepted: 24 January 2022
Published: 26 January 2022

1. Introduction

The problem of time-delay estimation (TDE) is to measure the difference in the time of arrival of signals recorded by space-separated sensors. This task is relevant for many applications, including those which are related to signal source localization [1]. The position of the object can be determined on the straight line [2,3], on the plane [4,5], and in space [6–8] depending on the location and the number of sensors.

The use of TDE methods is typical for those areas of technology where there is a need for the passive location of objects emitting signals. The physical nature of the signal, however, is not essential. Among practical applications, we can highlight the pipeline leaks position determination [2,3], local mobile objects positioning [9], passive radio positioning [1], etc. In recent years, the problem of TDE has become more relevant in connection with the spread, on the Internet, of concepts and services providing contactless control of household appliances [10], automatic tracking of objects [7], as well as in the sensor systems of robotic devices [11]. A common problem in the implementation of each of the listed services is the need for signal sources spatial discrimination, which normally requires TDE. Also, it should be noted that the development of industrial Internet applications requires solving the TDE problem for the time synchronization of data coming from asynchronous and spatially distributed sensors [11].

TDE methods and algorithms form a broad subject area. At present different approaches for TDE are known. A number of reviews have been devoted to the classification and systematization of TDE algorithms for numerous and diverse applications, in particular [8,12–15]. This paper compares well-known but seldom used TDE algorithms based on estimating the phase shift (GPS TDE) between signals.

Even though the frequency-domain TDE technique was originally proposed by Piersol [16] and developed by Zhen and Zi-Quang [15] back in the 1980s, studies devoted to its applications are relatively rare. This could be because the practical implementation of the GPS TDE technique is not as straightforward as the implementation of GCC TDE. Efficient implementation requires unwrapped phase spectrum estimation and time lag extraction which can be performed in various ways. This applies some limitations on using well-described GPS TDE algorithms [14] for different practical tasks. With this paper, we will propose an implementation applicable for most typical TDE applications, such as pipeline leak locating [2] or acoustic intrusion detection [4].

Related studies considering TDE for sound source positioning in room acoustic environment have been carried out before, for instance, in [7,8]. However, GPS TDE or similar frequency-domain techniques were not considered there. Variations of CPS TDE are compared in [14] in the different applications of locating the acoustic source, but the single path propagation model was used to simulate a practical case. The single path propagation model is considered not accurate [7,8] for a small room reverberation environment, so the conclusions of [14] could not be extrapolated to this application without further research. In [17], a hardware implementation of an indoor positioning system based on the phase correlation TDE algorithm was proposed, however, only substitutional research was carried out within the framework of the signal processing.

2. Materials and Methods

The most studied and widespread TDE technique is based on cross-correlation functions computation (CCF) [2]. CCFs are calculated for different time series pairs of sampled microphone signals, based on the position of the maximum in a correlogram. An alternative to the TDE correlation methods are phase-frequency methods, suggested firstly in [17]. Unlike correlation methods which analyze signals in the time domain, phase methods operate with signals frequency-domain representations. This section is devoted to the phase methods of TDE.

This paper considers the simplest case with two sensors, shown in Figure 1. Obviously, two sensors are not enough for unambiguous signal source localization on a plane or in space [11]. Depending on the relative sensor's position and the position of the signal source, a pair of microphones may be sufficient to determine the direction towards the object. In general cases, at least three sensors are required to determine the position of the source in a room [16]. In this case, the signals of the sensors array can be processed both simultaneously and in pairs [8]. The latter means that the algorithm considered in the paper can be used to localize the signal source in a room using three or more microphones.

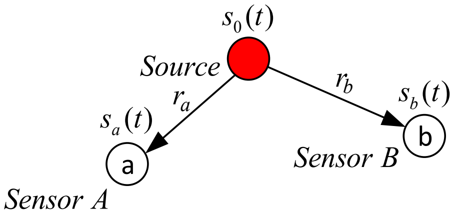


Figure 1. TDE with two sensors.

2.1. Ideal Propagation Model

The TDE task for sound source detecting in a room can be formalized in several ways [8]. Each method is a compromise between the signal propagation model accuracy and the complexity of the mathematical description of the problem. The main acoustic signal propagation models are [8]: ideal propagation model, multipath propagation model, and reverberation model. In this work, we consider that the simulated microphones are equally capable of efficiently registering signals coming from any direction.

The ideal propagation model assumes that there is only one path from the signal source to each of the microphones. Let $s_0(t)$ be the signal emitted by the source. Then the signals of the receivers will be

$$\begin{gathered} s_a(t) = \alpha_a \cdot s_0(t-\tau_a) + n_a(t), \\ s_b(t) = \alpha_b \cdot s_0(t-\tau_b) + n_b(t). \end{gathered} \tag{1}$$

where τ_a, τ_b are lag values; α_a, α_b are signal attenuation coefficients; $n_A(t)$, $n_B(t)$ are random uncorrelated additive microphone noises. The values of τ_a, τ_b are determined by the geometric distances r_a, r_b from the signal source to the corresponding receiver

$$\tau_a = \frac{r_a}{c},\ \tau_b = \frac{r_b}{c}, \tag{2}$$

where c is the sound speed. Attenuation of signals α_a, α_b can be caused by various factors, however, in the simplest ideal case, exclusively source beam pattern and the scattering of the sound wave are considered and, so

$$\alpha_a = \frac{k}{{r_a}^2},\ \alpha_b = \frac{k}{{r_b}^2}, \tag{3}$$

where k is a constant coefficient.

In this case, the TDE is performed to get the value $\tau_{ab} = \tau_b - \tau_a$ which is used further to determine the position of the sound source. Using the notations above and having redefined $t = t - \tau_b$, we can rewrite (1)

$$\begin{gathered} s_a(t) = \tfrac{k}{{r_a}^2} \cdot s_0(t + \tfrac{r_b - r_a}{c}) + n_a(t), \\ s_b(t) = \tfrac{k}{{r_b}^2} \cdot s_0(t) + n_b(t). \end{gathered} \tag{4}$$

Expression (4) does not consider the influence of several physical factors, such as reflection and absorption of sound in a room.

Later, in the course of computational experiments with the ideal scenario, we will take that $k = 1$, since the target signal-to-noise ratio (SNR) can be achieved exclusively by changing the noise intensity.

2.2. Reverberation Model

The problem of the ideal propagation model is that the assumptions made do not correspond to the acoustic conditions of the real-world enclosed room. Firstly, there are always several paths for sound propagation between the source and the receiver due to the presence of reflected waves. Secondly, the absorption of sound energy by room surfaces has a significant effect on the recorded signal.

In accordance with the reverberation model, the received signals are described as follows

$$\begin{gathered} s_a(t) = \int_0^T h_a(\tau) \cdot s_0(t-\tau) \cdot d\tau\ + n_a(t), \\ s_b(t) = \int_0^T h_b(\tau) \cdot s_0(t-\tau) \cdot d\tau\ + n_b(t). \end{gathered} \tag{5}$$

where $h_a\ (t)$, $h_b\ (t)$ are room impulse response (RIR) functions. The complexity of application of (5) is in the practical difficulty of RIR determination. Acoustic measurements [18] or mathematical methods can be used to solve this problem. The image model method, first proposed in [19], is the most widespread among the latter. Alternatively, statistical methods [20] or methods based on geometric acoustics and ray tracing [21] can be used. To create realistic sound signals in this work, the image model method was used in the implementation of Lehman, Johansson and Nordholm [22,23].

2.3. Basic Phase Shift TDE

The phase TDE algorithm is based on obtaining information about the delay value from the cross-phase spectrum Φ_{ab} of two signals. The algorithm for constructing the cross-phase spectrum is known from spectral analysis [14]. At the initial stage, the Fourier transforms $S_a(f_k)$ and $S_b(f_k)$ of the signals of each of the channels are determined

$$S_a(f_k) = F_D(s_a(t_i)),\ S_b(f_k) = F_D(s_b(t_i)), \tag{6}$$

where $s_a(t_i)$ and $s_b(t_i)$ are series of N real samples of $s_a(t)$ and $s_b(t)$ signals sampled with an interval Δ; F_D is the operator of short-time discrete Fourier transform (DFT); $S_a(f_k)$ and $S_b(f_k)$ are spectrums of the signals.

Further instantaneous cross-spectrum of signals $S_{ab}^{(q)}(f_k)$ are calculated

$$S_{ab}^{(q)}(f_k) = S_a^{(q)*}(f_k) \times S_b^{(q)}(f_k), \tag{7}$$

where superscript (q) indicates the time instant $t_q = \Delta{\cdot}N{\cdot}q$ of the beginning of the q-th time window; * is the element-wise complex conjugation; × is the element-wise product. The final measurement of the cross-spectrum $S_{ab}(f_k)$ is obtained by averaging the Q instantaneous spectrums

$$S_{ab}(f_k) = \frac{1}{Q}\sum_{q=0}^{Q-1} S_{ab}^{(q)}(f_k). \tag{8}$$

It should be noted that the application of (8) requires that the signal source remains stationary relatively to the receivers during the entire time of signal recording. If it is not, the spectral estimation $S_{ab}(f_k)$ would not be correct. However, this assumption is normally relevant for the cross-spectrum. If we consider that neither source nor sensors are moving, the phase shift for each particular harmonic component will remain the same for all Q instantaneous spectrums. Therefore, coherent accumulation is applied this way to reduce the impact of the additive random noise.

To retrieve the set of phases, the phase cross-spectrum $\Phi_{ab}(f_k)$ is finally calculated

$$\Phi_{ab}(f_k) = \mathrm{U}[\arg[S_{ab}(f_k)]], \tag{9}$$

where U is an operator of phase unwrapping [24]; arg is the operator for defining the argument of a complex number.

All harmonic components presented in $s_0(t)$ will also be present in $s_a(t)$ and $s_b(t)$. In this case, the phase difference between the k-th harmonic components of $s_a(t)$ and $s_b(t)$ is determined by $\tau_{ab}{\cdot}f_k$. Therefore, the estimation τ_{ab} can be obtained as the coefficient of proportionality in the line equation of the approximating $\Phi_{ab}(f_k)$.

The value $\hat{\tau}_{ab}$ can be determined, for example, based on the criterion for minimizing the squared error function [14]. Let the error e be determined as

$$e = \sum_k \Big(\Phi_{ab}(f_k) - \Big(\widehat{\tau}_{ab} \cdot 2\pi \cdot f_k + b_{ab}\Big)\Big)^2, \tag{10}$$

where b_{ab} is a constant term. Then

$$\begin{cases} \frac{de}{d\widehat{\tau}_{ab}} = -2 \cdot \sum\limits_k f_k \cdot \Big(\Phi_{ab}(f_k) - \widehat{\tau}_{ab} \cdot 2\pi \cdot f_k - b_{ab}\Big), \\ \frac{de}{db_{ab}} = -2\sum\limits_k \Big(\Phi_{ab}(f_k) - \widehat{\tau}_{ab} \cdot 2\pi \cdot f_k - b_{ab}\Big). \end{cases} \tag{11}$$

Equating the derivatives to zero in (11) results in

$$\widehat{\tau}_{ab} = \frac{\Delta \cdot N}{2\pi} \cdot \frac{D \cdot K - A \cdot C}{B \cdot K - A^2}, \tag{12}$$

where values A, C, B, D can be computed with the proposed scheme

$$A = \sum_k k;\ B = \sum_k k^2;\ C = \sum_k \Phi_{ab}(f_k);\ D = \sum_k k \cdot \Phi_{ab}(f_k). \tag{13}$$

An advantage of the algorithm based on the use of (12) and (13) is that non-adjacent spectral bins can be used for TDE. It is optimal to choose $k \in S$, where S is a set of the most essential harmonic components of the signal $s_0(t)$.

2.4. Generalized Phase Spectrum TDE

A modification of the method described in the previous subsection can be used to localize stationary signal sources. The modified method was initially proposed in [15] and was named GPS TDE.

A distinctive feature of the generalized method is the use of real-valued frequency weight function $W(f_k)$ which is used to determine $\hat{\tau}_{ab}$. Similarly to (10), the weighted error in this case are introduced

$$e = \sum_k \left[W(f_k) \cdot \left(\Phi_{ab}(f_k) - \left(\widehat{\tau}_{ab} \cdot 2\pi \cdot f_k + b_{ab} \right) \right) \right]^2. \tag{14}$$

Obtaining a calculation formula for $\hat{\tau}_{ab}$ could be carried out in the same way as in the previous subsection

$$\widehat{\tau}_{ab} = \frac{\Delta \cdot N}{2\pi} \cdot \frac{\Lambda \cdot \mathrm{K} - \mathrm{A} \cdot \Theta}{\mathrm{K} \cdot \mathrm{B} - \mathrm{A}^2}, \tag{15}$$

$$\mathrm{K} = \sum_k W(f_k), \mathrm{A} = \sum_k k \cdot W(f_k), \mathrm{B} = \sum_k k^2 \cdot W(f_k), \Theta = \sum_k \Phi_{ab}(f_k) \cdot W(f_k), \Lambda = \sum_k k \cdot \Phi_{ab}(f_k) \cdot W(f_k). \tag{16}$$

It is clear from (14) that the functions $W(f_k)$ should be chosen in the way that its value is high if the useful signal prevails over noises at the f_k frequency and differs little from zero in other cases. A set of five frequency weighting functions was investigated in [14]. Table 1 below shows the calculation formulas for these functions.

Table 1. Weight functions.

Method	Nomenclature	Formula
BCC	$W_{BCC}\ (f_k)$	$\lvert S_{ab}\ (f_k) \rvert / \max(\lvert S_{ab}\ (f_k) \rvert)$
PHAT	$W_{PHAT}\ (f_k)$	1
SCOT	$W_{SCOT}\ (f_k)$	$\gamma_{ab}\ (f_k)$
ML	$W_{ML}\ (f_k)$	$\gamma^2{}_{ab}\ (f_k)/[1 - \gamma^2{}_{ab}\ (f_k)]$
COH	$W_{COH}\ (f_k)$	$\gamma^2{}_{ab}\ (f_k)$

The coherence function $\gamma^2{}_{ab}\ (f_k)$ widely used for this purpose is calculated as

$$\gamma^2{}_{ab}\ (f_k) = \frac{\left| \sum\limits_{q=0}^{Q-1} \left(S_a{}^{(q)*}(f_k) \cdot S_b{}^{(q)}(f_k) \right) \right|^2}{\sum\limits_{q=0}^{Q-1} \left| S_a{}^{(q)} \right|^2 \cdot \sum\limits_{q=0}^{Q-1} \left| S_b{}^{(q)} \right|^2}. \tag{17}$$

It should be noted that the computational scheme proposed in this section differs from the one in [14]. Equation (15) allows the unwrapped phase spectrum to not pass through the origin, as far as we used coefficient b_{ab} in linear regression. This feature is practically important and will be addressed later. As far as $W(f_k)$ is based on spectral estimations, the generalized method should be applied carefully for signals that are non-stationary.

3. Results and Discussion

A series of computational experiments were carried out for a comparative evaluation of the algorithms. The human voice is commonly used for evaluation purposes in related studies [7,8]. Prior to the proposed study, we have tested algorithm performance for several speakers but did not find a significant difference in the results. Therefore, we have used the recording of one speaker and focused the study mainly on evaluating the impact of additive noise and multipath propagation in a reverberant environment.

A recording of a male speaker's voice with additive random noise was used to produce a set of test signals. The noise-free sound was synthesized based on the recorded voice by each of two means: in accordance with (4) and in accordance with (5).

Additive noises were generated by software, then scaled and summed with the preprocessed recording. The spectral noise density was equal in the range from 0 to 1000 Hz. Signals and noises outside of this frequency range were not considered in the experiments. A similar approach to preparing the set of test signals was used in [25].

Noises of the same intensity were applied to both channels. At the same time, the intensity of the noise was set in such a way as to provide the target SNR relative to the root-mean-square value of the signals recorded by the sensors for the entire time of each instance of the experiment. When applying (1), the delay was introduced by shifting one copy of the record relative to another by an integer number of sampling intervals (f_d = 44,100 Hz).

3.1. Experimental Setting

A set of stereo test records with a duration of about 20 s each were prepared for the study. The recording was analyzed in fragments of about 1 s during each instance of the experiment. At the same time, the analysis of each of the fragments was considered an independent experiment. The final estimations used to calculate the absolute error were obtained by averaging obtained values of the lag time.

The number of samples in each of the analyzed fragments was L = 40,960 (about 928.8 msec). The number of samples in the segment was taken to be N = 4096 (about 92.9 msec). Consequently, each piece of recording sound was fragmented into Q = 10 segments. When processing the results, the outputs corresponding to the segments of the recording, where pauses in speech predominated, were discarded.

Two different sets of frequency bins were used when applying (16). The first set contained frequency bins corresponding to the condition $f_k \in$ [100 Hz, 850 Hz]. The second set contained four non-overlapping frequency bands shown below. The choice of such frequency intervals was carried out in accordance with the form of power density spectrum of the raw signal shown in Figure 2. The presented characteristic was obtained by averaging all instantaneous power density spectrums with a window of N = 4096 samples. The position of the cut-off level was chosen empirically to optimize the TDE operation in the absence of reverberations. It should be noted that the power density spectrum for different speakers or even for different speech fragments by this speaker would not remain the same. However, the proposed procedure will remain applicable regardless.

3.2. Simulation of the Small Room Environment

As noted above, creating a realistic sound signal in accordance with (5) requires obtaining RIR functions h_a (t), h_b (t). The MATLAB program prepared by Eric Lehman [22] was used to obtain these characteristics. When calculating the RIR, the room parameters and the configuration of the sensors were specified as shown in Figure 3. The dimensions of the room were 5 × 3.5 × 2.25 m. The source has coordinates (1.5, 2.75, 1.8), and the microphones (4.5, 1.25, 1.8) and (4.5, 2.25, 1.8).

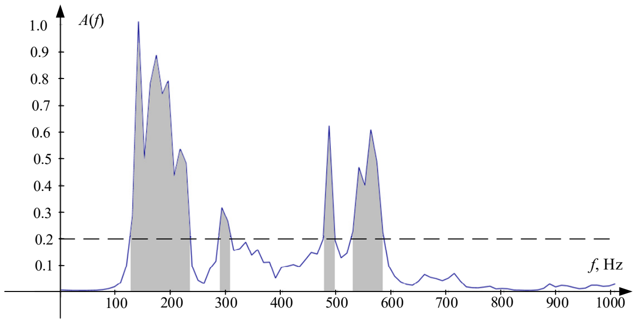


Figure 2. Raw signal power density spectrum. Frequency bins that are included in highlighted areas comprise the second set. Highlighted frequency bands are: 127–237 Hz, 285–305 Hz, 476–496 Hz, 531–580 Hz.

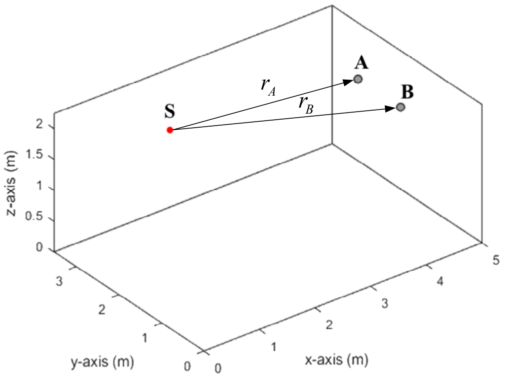


Figure 3. Source and microphones configuration in the model room. Source located in position S. Microphones are in positions A, B. Distances are r_A = 3.041 m, r_B = 3.354 m.

The reverberation time (T_{60}) was assumed to be 50 msec and 120 msec. The first value is compliant with the standards of a room intentionally designed for voice broadcasting. The second value is compliant with the requirements for verbal communication in an office space [26]. The synthesized RIRs are shown in Figure 4.

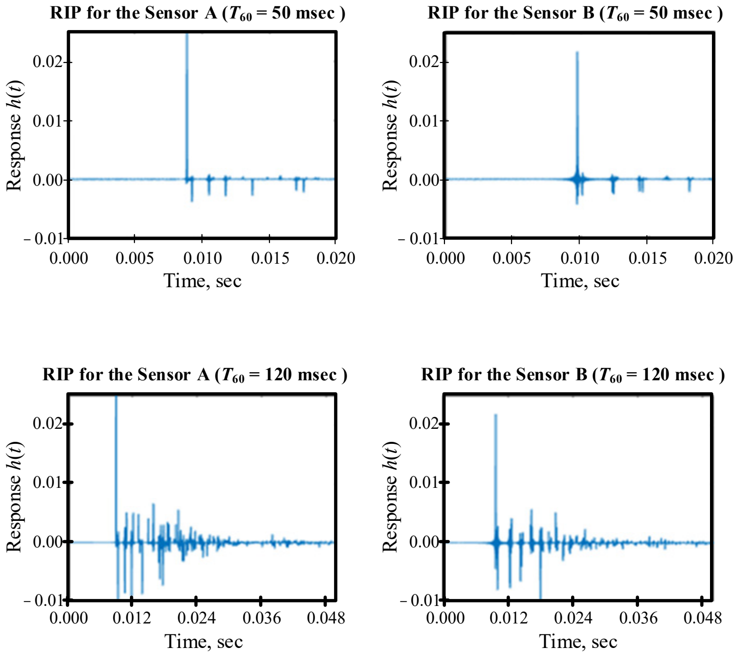


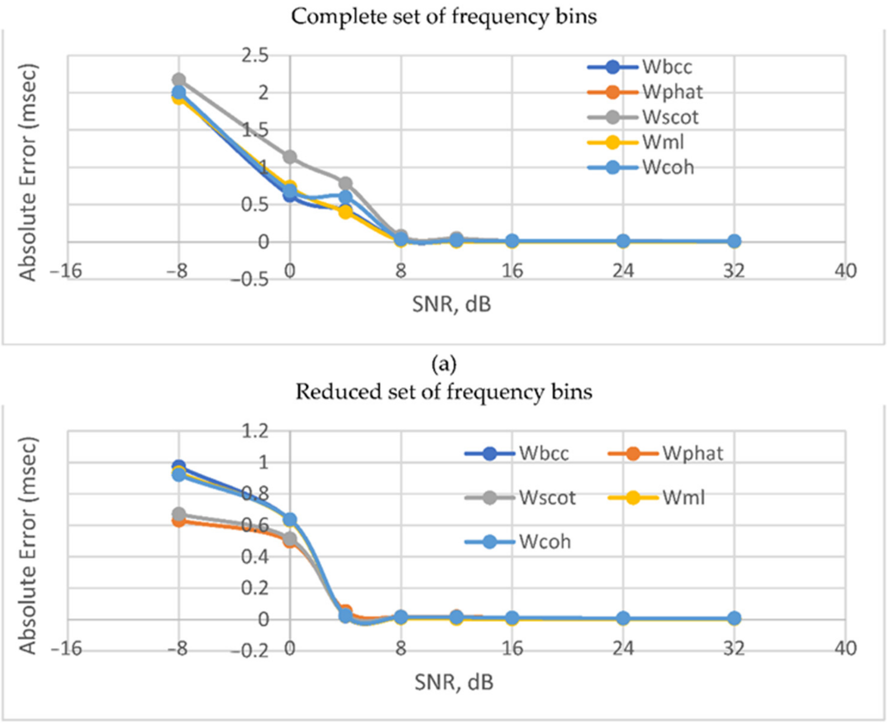
Figure 4. Room impulse responses for various reverberation times. True time delay is 0.923 msec.

3.3. Comparison of GPS TDE Methods in Anechoic Environment

Table 2 shows the absolute TDE errors for various weight functions and the ideal signal propagation model. Figure 5 shows the dependence of TDE error on SNR.

Table 2. Absolute error of GPS TDE with ideal propagation model.

Set	SNR	Mean Absolute Error (msec)				
	(dB)	W_{BCC} (f_k)	W_{PHAT} (f_k)	W_{SCOT} (f_k)	W_{ML} (f_k)	W_{COH} (f_k)
First	32	0.008	0.013	0.012	0.007	0.011
	24	0.007	0.014	0.016	0.007	0.013
	16	0.005	0.017	0.020	0.005	0.016
	12	0.008	0.049	0.031	0.008	0.023
	8	0.020	0.080	0.053	0.020	0.038
	4	0.425	0.782	0.626	0.399	0.600
	0	0.623	1.139	0.893	0.735	0.687
	−8	1.961	2.171	2.100	1.931	2.006
Second	32	0.005	0.008	0.009	0.005	0.009
	24	0.005	0.008	0.009	0.005	0.009
	16	0.004	0.012	0.011	0.003	0.012
	12	0.008	0.019	0.017	0.005	0.015
	8	0.011	0.016	0.018	0.010	0.015
	4	0.020	0.053	0.030	0.022	0.024
	0	0.634	0.497	0.514	0.631	0.637
	−8	0.973	0.631	0.672	0.934	0.921



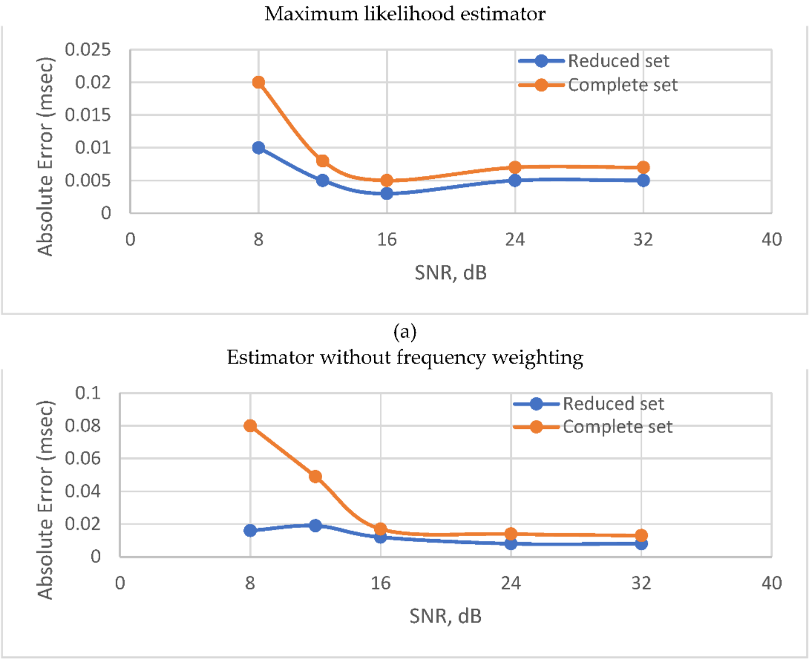
(a)

(b)

Figure 5. Absolute error vs SNR for anechoic room environment for: complete (**a**); and reduced (**b**) sets of frequency bins.

Figure 5 shows that the use of a reduced number of frequency bins in (15) and (16) provides greater accuracy while increasing the intensity of in-band noises. At the same time, the use of the second reduced frequency set allows you to reduce the threshold SNR to 4 dB over which sharp drop in the accuracy manifests.

Figure 6 shows the absolute TDE error for SNR ≥ 8 dB for W_{PHAT} and W_{ML}. When the noise intensity is not sufficient to go over the threshold, the estimators demonstrate the best possible performance in terms of accuracy regardless the noise level. When the SNR drops below the threshold level, the accuracy degrades gradually with the intensification of the noise. However, using a reduced set of frequency bins makes the contaminating effect of in-band noise less harsh. Notably, this is more obvious for W_{PHAT} than for W_{ML}. That can be explained by the fact that frequency weighting applied with ML estimator compensates for frequency bins where noise prevails over the signal. Despite the fact, that threshold SNR level appears in Figure 6 to be better for PHAT than for ML, the latter estimator surpasses the former in terms of accuracy in the single path scenario regardless of noise intensity. The frequency weighting function for the ML estimator is in Figure 7.


Maximum likelihood estimator

(a)

Estimator without frequency weighting

(b)

Figure 6. Absolute error vs SNR for anechoic room environment: (**a**) maximum likelihood weighting function (W_{ML}); (**b**) no weighting was applied (W_{PHAT}).

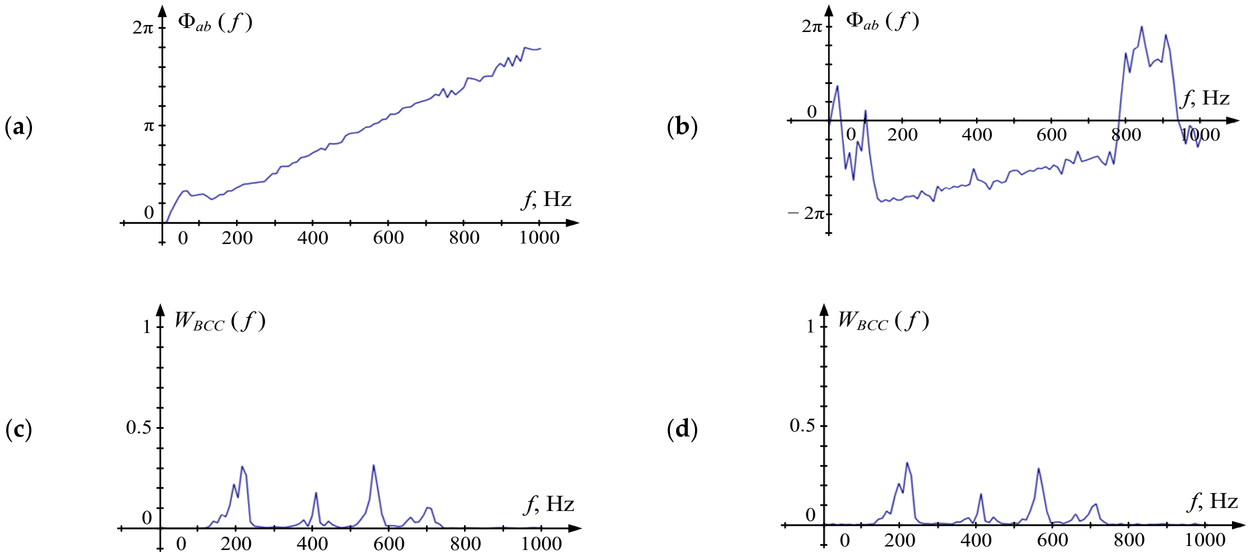


Figure 7. *Cont.*

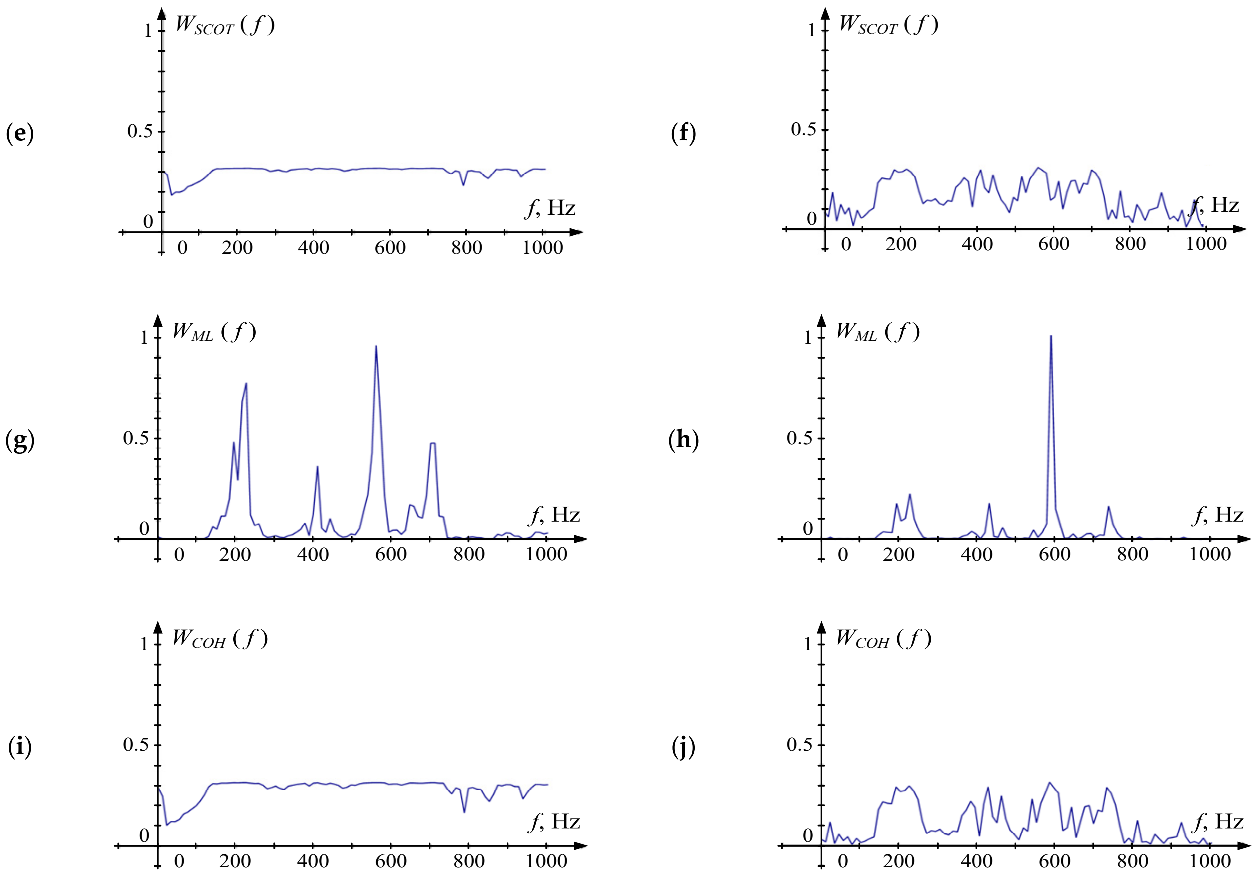


Figure 7. Sample phase cross spectrum Φ_{ab} (f_k) and weighting functions $W(f_k)$ for various SNR: (**a**,**b**) Φ_{ab} (f_k), (**c**,**d**) $W_{BCC}(f_k)$, (**e**,**f**) $W_{SCOT}(f_k)$, (**g**,**h**) $W_{ML}(f_k)$, (**i**,**j**) $W_{COH}(f_k)$. Figures (**a**,**c**,**e**,**g**,**i**) are obtained for SNR = 32 dB. Figures (**b**,**d**,**f**,**h**,**j**) are obtained for SNR = 4 dB. For W_{ML} (f_k) all values are normalized with the maximum value on the frequency band of interest.

Figure 7 shows the form of Φ_{ab} (f_k) and all $W(f_k)$ in the absence of noise (SNR = 32 dB) and their presence (SNR = 4 dB). A part of the curve that is close to linear shape is clearly distinguished at Φ_{ab}, in both cases, however, in the presence of noise, the corresponding frequency range is significantly narrower. It should be noted that Φ_{ab} in the absence of noise passes through the origin and behaves as described in [14]. However, when the signal is contaminated with the noise, Φ_{ab} is offset relative to the abscissa axis. This can be explained by the fact that there is no voice signal on frequencies up to 100 Hz, so the prevalence of the noise in this band results in an unpredictable offset of the unwrapped phase spectrum. That makes the estimation technique proposed in [14] not relevant for this task.

The shape of W_{SCOT} and W_{COH} is close to a line parallel to the time axis in the absence of noise. In the presence of noise, a high level of W_{SCOT} and W_{COH} is observed in the intervals where the cross-power spectrum $|S_{ab}|$ has high values. W_{BCC} form follows the shape of $|S_{ab}|$ and does not differ significantly in the presence of noise and their absence. Four areas of high values are visible at the W_{ML} corresponding to the Φ_{ab} regions that are best approximated by the line.

3.4. Comparison of GPS TDE Methods in Reverberant Environment

Tables 3 and 4 summarize the average TDE absolute errors for different weighting functions, reverberation model and different reverberation times.

Table 3. Absolute error of GPS TDE with reverberation model (T_{60} = 50 msec).

Set	**SNR**	**Mean Absolute Error (msec)**				
	(dB)	$W_{BCC}\,(f_k)$	$W_{PHAT}\,(f_k)$	$W_{SCOT}\,(f_k)$	$W_{ML}\,(f_k)$	$W_{COH}\,(f_k)$
First	32	0.081	0.012	0.009	0.054	0.009
	24	0.080	0.014	0.010	0.065	0.012
	16	0.063	0.019	0.017	0.059	0.019
	12	0.066	0.021	0.015	0.061	0.021
	8	0.072	0.027	0.017	0.073	0.022
	4	0.062	0.177	0.097	0.067	0.069
	0	0.272	0.505	0.407	0.247	0.324
	−8	1.735	1.775	1.746	1.736	1.747
Second	32	0.166	0.195	0.195	0.182	0.194
	24	0.165	0.190	0.192	0.178	0.192
	16	0.163	0.181	0.183	0.169	0.181
	12	0.161	0.171	0.171	0.163	0.168
	8	0.155	0.122	0.148	0.153	0.150
	4	0.165	0.151	0.156	0.157	0.150
	0	0.190	0.199	0.168	0.162	0.137
	−8	1.392	1.727	1.598	1.460	1.478

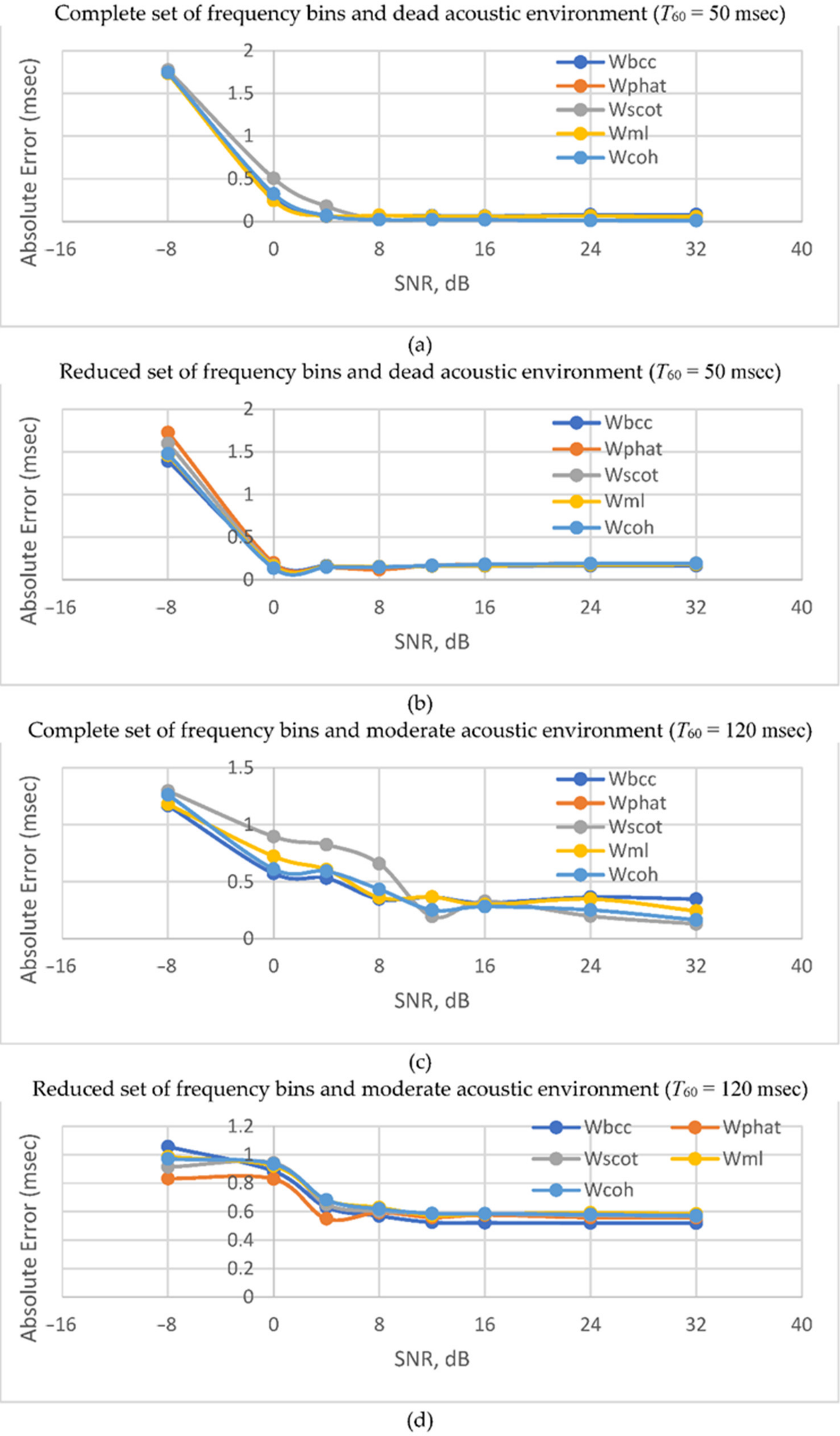
Table 4. Absolute error of GPS TDE with reverberation model (T_{60} = 120 msec).

Set	**SNR**	**Mean Absolute Error (msec)**				
	(dB)	$W_{BCC}\,(f_k)$	$W_{PHAT}\,(f_k)$	$W_{SCOT}\,(f_k)$	$W_{ML}\,(f_k)$	$W_{COH}\,(f_k)$
First	32	0.346	0.129	0.146	0.243	0.164
	24	0.364	0.198	0.208	0.347	0.251
	16	0.315	0.327	0.320	0.299	0.282
	12	0.365	0.194	0.212	0.366	0.251
	8	0.348	0.658	0.578	0.361	0.433
	4	0.531	0.825	0.768	0.606	0.592
	0	0.572	0.897	0.842	0.723	0.611
	−8	1.169	1.297	1.291	1.181	1.263
Second	32	0.519	0.558	0.567	0.584	0.571
	24	0.520	0.559	0.574	0.592	0.580
	16	0.522	0.577	0.586	0.583	0.586
	12	0.525	0.560	0.586	0.574	0.587
	8	0.570	0.594	0.604	0.629	0.619
	4	0.631	0.550	0.651	0.682	0.683
	0	0.884	0.829	0.942	0.921	0.935
	−8	1.057	0.832	0.914	0.985	0.971

Figure 8 shows that in the presence of reflected signals, the ML estimator is inferior in accuracy to the SCOT and COH estimators, especially in the absence of additive noises. At the same time, the accuracy turns out to be significantly lower than in the previous case. This can be explained by the correlation of the signals with their reflected copies. In the presence of reverberations and intense noises, none of the functions show any accuracy advantage. The latter makes it useful to apply the BPS TDE method (PHAT) as the simplest one.

The use of the second set of frequency bins provides an advantage in accuracy only in conditions of noise dominance (SNR $\leq$ 0 dB). The use of the complete set of frequency bins provides significantly better accuracy in other cases.

Figure 8 shows the dependence of TDE error on SNR graphically.


Complete set of frequency bins and dead acoustic environment (T_{60} = 50 msec)

(a)

Reduced set of frequency bins and dead acoustic environment (T_{60} = 50 msec)

(b)

Complete set of frequency bins and moderate acoustic environment (T_{60} = 120 msec)

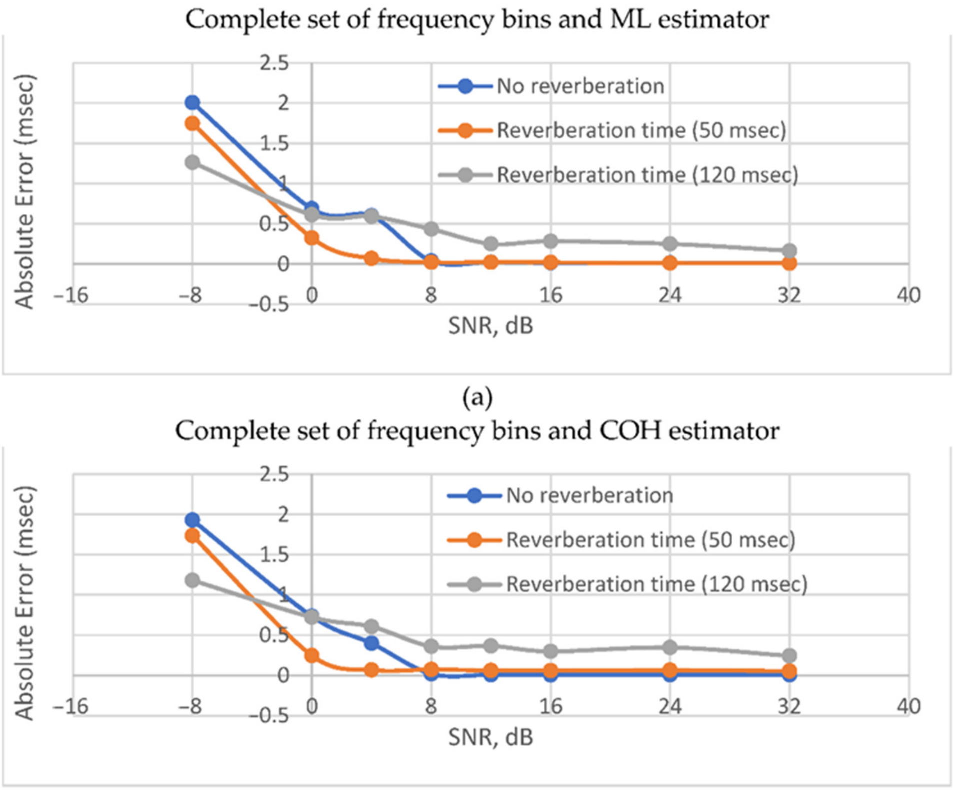
(c)

Reduced set of frequency bins and moderate acoustic environment (T_{60} = 120 msec)

(d)

Figure 8. Absolute error vs SNR for reverberant room environment. For subfigures (**a**,**b**) T_{60} = 50 msec. For figures (**c**,**d**) T_{60} = 120 msec. Reduced set was used for (**b**,**d**). Complete set was used for (**a**,**c**).

Figure 9 shows the results of using GPS TDE for various acoustic conditions of the environment. It is clear from the figure that the reverberation time increase leads to a drastic increase in the error both in the presence and absence of noise. However, with the dominance of noise over the signal, the presence of reflected copies has a positive effect on accuracy. However, even if this is the case, the TDE error remains unacceptably high for a significant part of practical applications.



(a)

(b)

Figure 9. Absolute error vs SNR for various reverberation times and the complete set of frequency bins: (**a**) W_{ML}; and (**b**) W_{COH} frequency weighting functions were applied.

Figure 10 shows the form of Φ_{ab} (f_k) and all $W(f_k)$ for different values of reverberation time (T_{60}). All graphics in Figures 7 and 10 are obtained for one and the same fragment of the original signal. It can be seen from the form of Φ_{ab} that an increase in the reverberation time leads to a distortion of the frequency response form and a decrease in the estimate accuracy. At the same time, the distortions observed for W_{SCOT} and W_{COH} are not as significant as they were in the absence of reverberations and the presence of noises. This can be explained by the fact that the reflected signals are mutually correlated, and their presence does not contribute to a significant decrease in the level of signal coherence. The correlation of the reflected signals also affects at the shape $|S_{ab}|$ and, therefore, at the W_{BCC} form. The W_{ML} form also changes significantly with an increase in the reverberation time, while the regions of high values also correspond to the linear sections Φ_{ab}. At T_{60} = 120 msec, the number of such sections becomes smaller which negatively affects the accuracy.

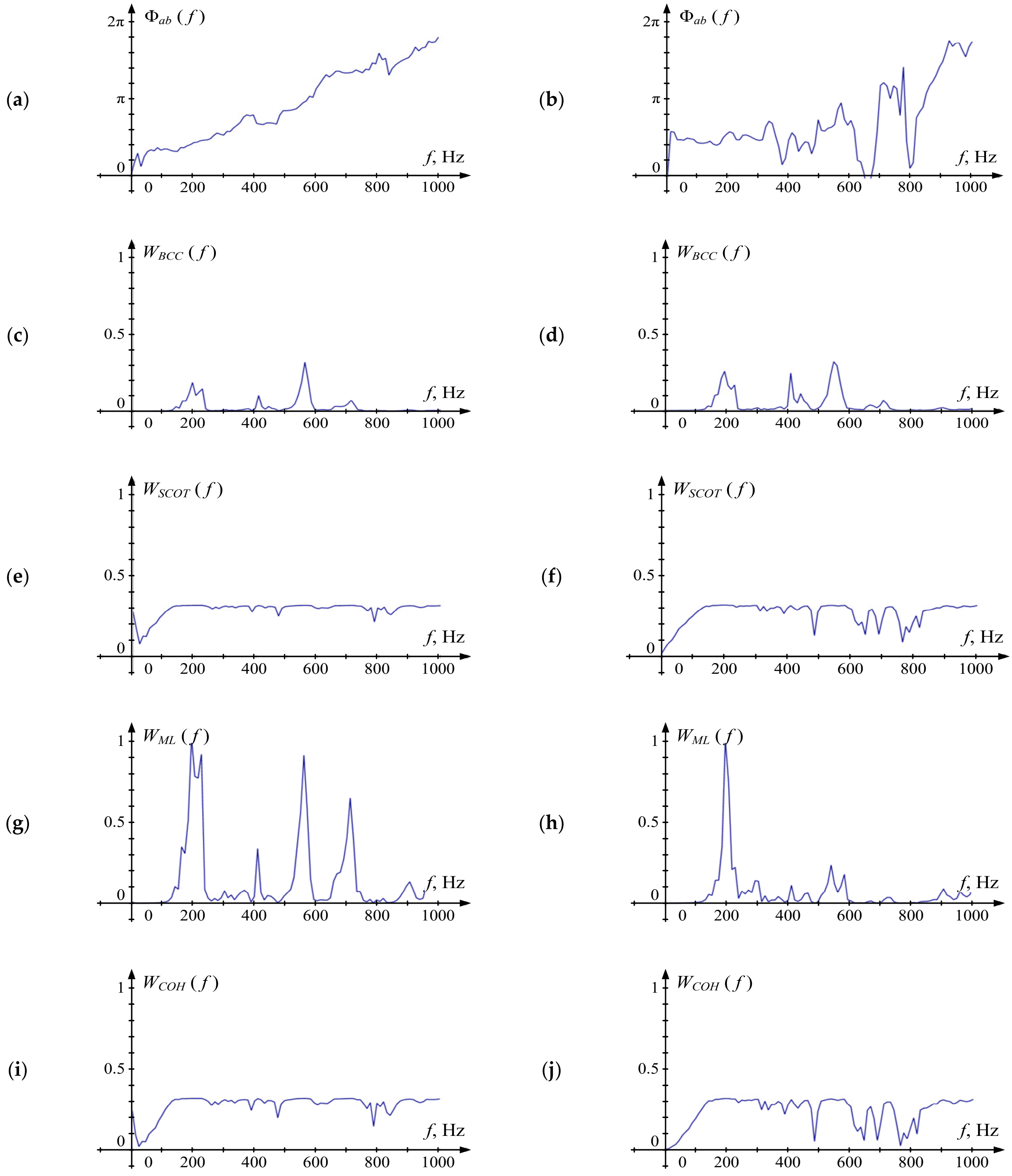


Figure 10. Sample phase cross spectrum Φ_{ab} (f_k) and weighting functions $W(f_k)$ for various reverberation times: (**a**,**b**) Φ_{ab} (f_k), (**c**,**d**) $W_{BCC}(f_k)$, (**e**,**f**) $W_{SCOT}(f_k)$, (**g**,**h**) $W_{ML}(f_k)$, (**i**,**j**) $W_{COH}(f_k)$. Figures (**a**,**c**,**e**,**g**,**i**) are obtained for T_{60} = 50 msec. Figures (**b**,**d**,**f**,**h**,**j**) are obtained for T_{60} = 120 msec. For W_{ML} (f_k) all values are normalized with the maximum value on the frequency band of interest.

4. Conclusions

This study investigated GPS TDE in relation to the problem of localizing a sound source in a small room. The suggested TDE algorithm is based on the analysis of the phase response form which makes it possible to estimate the time by analyzing an arbitrary set of spectral bins.

To assess the algorithm's applicability and efficiency, a series of computational experiments were performed to simulate the speaker positioning within a small room. To simulate room acoustics, the image model implemented by Lehman and Johanson [23] was used. During the course of the experiment, the SNR at the signal receivers was varied, as well as the room reverberation time.

The fundamental applicability of the suggested algorithm was shown due to the performed experiment. In the absence of noises and echo, GPS TDE demonstrates an accuracy comparable to the sampling error at f_d = 44,100 Hz (about 0.01 s). A decrease in accuracy is expected in the absence of echo but at an increase in the intensity of additive noise. However, narrowing of the frequency range over which TDE is performed helps to maintain accuracy under moderate noises (SNR > 4 dB). The best accuracy characteristics are provided by the ML GPS estimator.

When an echo occurs, TDE accuracy downgrades significantly. The reflected signals are correlated, and, therefore, introduce extra noise to the correlogram. In this case, the use of a reduced set of spectral bins affects the accuracy negatively. Even with insignificant reverberations, corresponding to an acoustical very dead room and the absence of noises, the ML GPS estimator demonstrates a relatively low accuracy. The SCOT and COH GPS estimators show the best results. In conditions of higher reverberations, the TDE error increases significantly in comparison to the ideal case and makes the use of the GPS method ineffective. In practice, however, the influence of echo can be lower, as real-world microphones are not omnidirectional.

Even though the suggested method is inferior to analogs in a few aspects, its advantage remains high computational efficiency. The suggested computational scheme, when using a relatively small number of adjacent frequency samples for TDE, allows the use of Goertzel's algorithm instead of FFT [27]. This is essential for embedded computers with memory constraints. Additionally, the use of well-known implementations of the Goertzel algorithm designed for phase detection [28] will make it possible to re-evaluate the spectral characteristics of the signal with new data arrival. The latter is useful for solving the problem of positioning a mobile acoustic source. Further studies will be devoted to the testing of this hypothesis.

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

Funding: This research was funded by Ministry of Education and Science of the Russian Federation within the framework of scientific projects carried out by teams of research laboratories of educational institutions of higher education subordinate to the Ministry of Science and Higher Education of the Russian Federation, project number FEWM-2020-0042 (AAAA-A20-120111190016-9) [29].

Data Availability Statement: For the experiments, a model of a room's acoustic environment was used to synthetize test data. The model is implemented by Eric Lehman as MATLAB program and can be downloaded here http://www.eric-lehmann.com/ (last accessed on 17 November 2021).

Acknowledgments: We want to thank the organizers of the XV International Scientific and Technical Conference «Actual Problems of Electronic Instrument Engineering» for the provided opportunity to present this research. Also, we would like to express our deep gratitude to Irkutsk Supercomputer Center of SB RAS for providing their outstanding expertise and the access to HPC-cluster «Akademik V.M. Matrosov».

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Juang, B.H.; Chen, T. Highlights of Statistical Signal and Array Processing. *IEEE Signal Proc. Mag.* **1998**, *15*, 21–64. [CrossRef]
2. Fuchs, H.V.; Riehle, R. Ten Years of Experience with Leak Detection by Acoustic Signal Analysis. *Appl. Acoust.* **1991**, *33*, 1–19. [CrossRef]
3. Kousiopoulos, G.-P.; Papastavrou, G.-N.; Kampelopoulos, D.; Karagiorgos, N.; Nikolaidis, S. Comparison of Time Delay Estimation Methods Used for Fast Pipeline Leak Localization in High-Noise Environment. *Technologies* **2020**, *8*, 27. [CrossRef]
4. Zu, X.; Guo, F.; Huang, J.; Zhao, Q.; Liu, H.; Li, B.; Yuan, X. Design of an Acoustic Target Intrusion Detection System Based on Small-Aperture Microphone Array. *Sensors* **2017**, *17*, 514. [CrossRef] [PubMed]
5. Ren, E.; Ornelas, G.C.; Loeliger, H.-A. Real-Time Interaural Time Delay Estimation Via Onset Detection. In Proceedings of the 2021 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 2021, Toronto, ON, Canada, 6–11 June 2021; pp. 1988–2005. [CrossRef]
6. Carter, C. Time Delay Estimation for Passive Sonar Signal Processing. *IEEE Trans. Acoust. Speech* **1981**, *29*, 463–470. [CrossRef]
7. Dvorkind, T.G.; Gannot, S. Time Difference of Arrival Estimation of Speech Source in a Noisy and Reverberant Environment. *Signal Process.* **2005**, *85*, 177–204. [CrossRef]
8. Chen, J.; Benesty, J.; Huang, A. Time Delay Estimation in Room Acoustic Environments: An Overview. *EURASIP J. Adv. Signal Process.* **2006**, *26503*, 1–19. [CrossRef]
9. Potortì, F.; Palumbo, F.; Crivello, A. Sensors and Sensing Technologies for Indoor Positioning and Indoor Navigation. *Sensors* **2020**, *20*, 5924. [CrossRef]
10. Narayana Murthy, B.H.; Yegnanarayana, B.; Radiri, S.R. Time Delay Estimation from Mixed Multispeaker Speech Signals Using Single Frequency Filtering. *Int. J. Circuits Syst. Signal Process.* **2020**, *39*, 1988–2005. [CrossRef]
11. Trifa, V.M.; Koene, A.; Moren, J.; Cheng, G. Real-Time Acoustic Source Localization in Noisy Environments for Human-Robot Multimodal Interaction. In Proceedings of the 16th IEEE International Symposium on Robot and Human Interactive Communication, RO-MAN 2007, Jeju, Korea, 26–29 August 2007; pp. 1988–2005. [CrossRef]
12. Althoubi, A.; Alshahrani, R.; Peyravi, H. Delay Analysis in IoT Sensor Networks. *Sensors* **2021**, *21*, 3876. [CrossRef]
13. Faerman, V.A.; Avramchuk, V.S. Comparative Study of Basic Time Domain Time-Delay Estimators for Locating Leaks in Pipelines. *Int. J. Netw. Distrib. Comput.* **2020**, *8*, 49–57. [CrossRef]
14. Brennan, M.J.; Gao, Y.; Josephn, P.F. On the Relationship between Time and Frequency Domain Methods in Time Delay Estimation for Leak Detection in Water Distribution Pipes. *J. Sound Vib.* **2007**, *304*, 213–223. [CrossRef]
15. Zhen, Z.; Zi-qiang, H. The Generalized Phase Spectrum Method for Time Delay Estimation. In Proceedings of the IEEE International Conference on Conference: Acoustics, Speech, and Signal Processing ICASSP ′84, San Diego, CA, USA, 19–21 March 1984; pp. 459–462. [CrossRef]
16. Piersol, A.G. Time Delay Estimation Using Phase Data. *IEEE Trans. Acoust. Speech* **1981**, *29*, 471–477. [CrossRef]
17. Mannay, K.; Ureña, J.; Hernández, Á.; Villadangos, J.M.; Machhout, M.; Aguili, T. Evaluation of Multi-Sensor Fusion Methods for Ultrasonic Indoor Positioning. *Appl. Sci.* **2021**, *11*, 6805. [CrossRef]
18. Carini, A.; Cecchi, S.; Orcioni, S. Robust Room Impulse Response Measurement Using Perfect Periodic Sequences for Wiener Nonlinear Filters. *Electronics* **2020**, *9*, 1793. [CrossRef]
19. Allen, J.B.; Berkley, D.A. Image Method for Efficiently Simulating Small-Room Acoustics. *J. Acoust. Soc. Am.* **1979**, *65*, 943–950. [CrossRef]
20. Liu, J.; Yang, G.-Z. Robust Speech Recognition in Reverberant Environments by Using an Optimal Synthetic Room Impulse Response Model. *Speech Commun.* **2015**, *67*, 65–77. [CrossRef]
21. Alpkocak, A.; Sis, M.K. Computing Impulse Response of Room Acoustic Using the Ray-Tracing Method in Time Domain. *Arch. Acoust.* **2010**, *35*, 505–519. [CrossRef]
22. Lehmann, E.; Johansson, A.; Nordholm, S. Reverberation-Time Prediction Method for Room Impulse Responses Simulated with the Image-Source Model. In Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA'07), New Paltz, NY, USA, 21–24 October 2007; pp. 159–162. [CrossRef]
23. Lehmann, E.; Johansson, A. Prediction of Energy Decay in Room Impulse Responses Simulated with an Image-Source Model. *J. Acoust. Soc. Am.* **2008**, *124*, 269–277. [CrossRef]
24. Detmold, W.; Kanwar, G.; Wagman, L. Phase Unwrapping and One-Dimensional Sign Problems. *Phys. Rev. D* **2018**, *98*, 074511. [CrossRef]
25. Bedard, S.; Champagne, B.; Stephenne, A. Effects of Room Reverberation on Time-Delay Estimation Performance. In Proceedings of the ICASSP ′94 IEEE International Conference on Acoustics, Speech and Signal Processing, Adelaide, Australia, 19–22 April 1994; pp. 261–264. [CrossRef]
26. Levy, S.M. *Construction Calculations Manual*, 1st ed.; Butterworth-Heinemann: Oxford, UK, 2012; pp. 503–544.
27. Sysel, P.; Rajmic, P. Goertzel Algorithm Generalized to Non-Integer Multiples of Fundamental Frequency. *EURASIP J. Adv. Signal Process.* **2012**, *56*, 1–8. [CrossRef]

28. Yeh, C.-Y.; Hwang, S.-H. Efficient Detection Approach for DTMF Signal Detection. *Appl. Sci.* **2019**, *9*, 422. [CrossRef]
29. HPC-Cluster «Akademik V.M. Matrosov» Official Webpage. Available online: https://hpc.icc.ru/en/hardware/ (accessed on 18 November 2021).