applied sciences

MDPI

Communication

Late Reverberant Spectral Variance Estimation for Single-Channel Dereverberation Using Adaptive Parameter Estimator

Zhaoqi Zhang [1], Xuelei Feng [1] and Yong Shen [1,2,*]

[1] Institute of Acoustics, Nanjing University, Nanjing 210093, China; zhaoqizhang@smail.nju.edu.cn (Z.Z.); xlfeng@nju.edu.cn (X.F.)
[2] Shenzhen Research Institute of Nanjing University, Shenzhen 518000, China
* Correspondence: yshen@nju.edu.cn

Abstract: The estimation of the late reverberant spectral variance (LRSV) is of paramount importance in most reverberation suppression algorithms. This letter proposes an improved single-channel LRSV estimator based on Habets LRSV estimator by using an adaptive parameter estimator. Instead of estimating the direct-to-reverberation ratio (DRR), the proposed LRSV estimator directly estimates the parameter κ in a generalized statistical model since the experimental results show that even the κ calculated using measured ground truth DRR may not be the optimal parameter for the LRSV estimator. Experimental results using synthetic reverberant signals demonstrate the superiority of the proposed estimator to conventional approaches.

Keywords: dereverberation; single-channel; probability-based

Citation: Zhang, Z.; Feng, X.; Shen, Y. Late Reverberant Spectral Variance Estimation for Single-Channel Dereverberation Using Adaptive Parameter Estimator. *Appl. Sci.* **2021**, *11*, 8054. https://doi.org/10.3390/app11178054

Academic Editor: Edoardo Alessio Piana

Received: 6 July 2021
Accepted: 28 August 2021
Published: 30 August 2021

1. Introduction

Speech signals received within a room usually contain reverberation which impairs the intelligibility of speech in communication scenarios such as mobile phones and hearing aids. Reverberation will also degrade the recognition performance of automatic speech recognition systems. Hence, speech dereverberation is still an important issue nowadays.

Dereverberation techniques can be divided into reverberation cancellation [1] and reverberation suppression [2,3] depending on whether or not the acoustic impulse response (AIR) needs to be estimated during the dereverberation [4]. The major part of most reverberation suppression methods is the estimation of late reverberant spectral variance (LRSV), which remains a challenging task due to its high time variability [5]. Habets proposed a single-channel LRSV estimator [3] based on a generalized statistical model [6] to suppress late reverberation, and it still performs outstanding nowadays [5]. However, in Habets LRSV estimator, two parameters (i.e., the reverberation time T_{60} and the parameter κ which is related to direct-to-reverberation ratio (DRR)) should be given in advance or estimated online.

To the authors' knowledge, there are numerous reverberation time estimation methods, whereas there are few single-channel DRR online estimation methods [7]. Besides, according to practical experience, it may also be inappropriate to obtain κ indirectly by estimating DRR, because even the κ calculated via measured ground truth DRR may not be the optimal κ for the Habets estimator. A detailed discussion can be found in Section 5. Therefore, unlike other traditional methods using estimated DRR to calculate κ, the present work aims to propose a blind adaptive κ estimator which can improve the performance of the Habets LRSV estimator and makes it more practical. Inspired by the optimally-modified log-spectral amplitude (OM-LSA) algorithm [8], this letter differentiates between the direct sound presence/absence hypotheses and derives the conditional direct sound presence probability to give a time-varying recursive average on the estimated κ. The proposed κ

estimator is evaluated and compared with existed κ estimator [9] and κ calculated using measured ground truth DRR. The evaluation results show that the proposed κ estimator performs better than the conventional κ estimator or measured κ under all evaluation conditions. Besides, the quality of the dereverberated speech is also evaluated and compared to a method using recursive maximum-sparseness-power-prediction-model (MSPP) [10].

2. Problem Formulation

The reverberant signal results from the convolution of the anechoic speech signal and a causal AIR. The anechoic speech signal can be expressed in the Short-time Fourier Transform (STFT) domain by $S(k,l)$, where k and l are the frequency and frame indices, respectively. According to the convolutive transfer function (CTF) model [2], the reverberant speech signal $Z(k,l)$ can be expressed as Equation (1)

$$Z(k,l) = \sum_{l'=0}^{+\infty} H(k,l')S(k,l-l'), \tag{1}$$

where $H(k,l)$ represents the AIR and it can be split into three components as Equation (2)

$$H(k,l) = \begin{cases} H_d(k), & l = 0 \\ H_e(k,l), & 1 \le l \le N_e \\ H_l(k,l), & l > N_e \end{cases} \tag{2}$$

where $H_d(k)$ is the direct sound, $H_e(k,l)$ consists of early reflections, $H_l(k,l)$ represents later reflections, and N_e usually corresponds to approximately 20–50 ms. The late reverberant speech component $Z_l(k,l) = \sum_{l'=N_e+1}^{+\infty} H_l(k,l')S(k,l-l')$ mainly decreases the speech fidelity and intelligibility [4] and needs to be suppressed. Hence, the main challenge is to derive an estimator for the spectral variance of the late reverberant speech component (i.e., LRSV) $\lambda_l(k,l) = E\left[|Z_l(k,l)|^2\right]$, where $E[\cdot]$ denotes the expectation operator. Once $\lambda_l(k,l)$ is given, a spectral enhancement method [11] can be used to suppress the late reverberation.

3. Brief Review of Habets Late Reverberant Spectral Variance Estimator

The underlying theory for the present work is based on the LRSV estimator derived by Habets [3]. The Habets method is based on a generalized statistical model which is an improvement on Polack's statistical model [4]. Using $H_r(k,l)$ represents early and late reflections. Then, the corresponding spectral variance can be written as Equation (3)

$$\lambda_h(k,l) = E\left[|H(k,l)|^2\right] = \begin{cases} \lambda_{h_d}(k), & l = 0 \\ \lambda_{h_r}(k,l), & l \ge 1 \end{cases} \tag{3}$$

$$\lambda_{h_d}(k) = E\left[|H_d(k)|^2\right], \lambda_{h_r}(k,l) = \kappa(k)\lambda_{h_d}(k)e^{\frac{-13.8lR}{T_{60}(k)f_s}},$$

where $T_{60}(k)$ is the frequency-dependent reverberation time, f_s denotes the sampling frequency, R is the discrete time shift, and $\kappa(k)$ is a prior parameter that is related to DRR.

Assuming that the direct component $Z_d(k,l) = H_d(k)S(k,l)$ and the reverberant component $Z_r(k,l) = \sum_{l'=1}^{+\infty} H_r(k,l')S(k,l-l')$ are uncorrelated, the corresponding spectral variance $\lambda_z(k,l) = E\left[|Z(k,l)|^2\right]$ can be expressed as the sum of the direct component spectral variance $\lambda_d(k,l) = E\left[|Z_d(k,l)|^2\right]$ and the reverberant component spectral variance $\lambda_r(k,l) = E\left[|Z_r(k,l)|^2\right]$, such that Equation (4)

$$\lambda_z(k,l) = \underbrace{\lambda_{h_d}(k)\lambda_s(k,l)}_{\lambda_d(k,l)} + \underbrace{\sum_{l'=1}^{+\infty} \lambda_{h_r}(k,l')\lambda_s(k,l-l')}_{\lambda_r(k,l)}, \tag{4}$$

where $\lambda_s(k,l)$ is the spectral variance of $S(k,l)$. The reverberant component $\lambda_r(k,l)$ can be further split into early reverberation $\lambda_e(k,l)$ and late reverberation $\lambda_l(k,l)$, as Equation (5)

$$\lambda_r(k,l) = \underbrace{\sum_{l'=1}^{N_e} \lambda_{h_r}(k,l')\lambda_s(k,l-l')}_{\lambda_e(k,l)} + \underbrace{\sum_{l'=N_e}^{+\infty} \lambda_{h_r}(k,l')\lambda_s(k,l-l')}_{\lambda_l(k,l)}, \tag{5}$$

and the main purpose is to derive an estimator for the LRSV $\lambda_l(k,l)$. Combining Equations (3) and (4), $\lambda_r(k,l)$ can be obtained by Equation (6)

$$\lambda_r(k,l) = \exp\left\{\frac{-13.8R}{T_{60}(k)f_s}\right\}[(1-\kappa(k))\lambda_r(k,l-1) + \kappa(k)\lambda_z(k,l-1)]. \tag{6}$$

Finally, according to Equations (3) and (5), $\lambda_l(k,l)$ can be obtained using $\lambda_r(k,l)$ as Equation (7)

$$\lambda_l(k,l) = \exp\left\{\frac{-13.8R(N_e-1)}{T_{60}(k)f_s}\right\}\lambda_r(k,l-N_e+1). \tag{7}$$

4. Parameter Estimation

In Habets LRSV estimator, two parameters (i.e., T_{60} and κ) should be given in advance. The reverberation time T_{60} can be determined by applying Schroeder's method to the AIR. The parameter κ is related to DRR and can be calculated [3] by solving Equation (8)

$$\kappa = \frac{1}{DRR}\frac{1-\exp\left\{\frac{-13.8R}{T_{60}f_s}\right\}}{\exp\left\{\frac{-13.8R}{T_{60}f_s}\right\}}, \tag{8}$$

where $DRR = \sum_{n=0}^{R-1} h^2(n) \Big/ \sum_{n=R}^{+\infty} h^2(n)$ and $h(n)$ represents AIR. The Habets LRSV estimator is often used without knowing those two parameters. The T_{60} estimation has been well investigated and numerous blind approaches can be found. However, the DRR estimation is less mature and there are few online single-channel estimation algorithms [7]. Therefore, the reverberation time T_{60} is assumed to be known in the following, and the present work focuses on the κ estimation. Most existed κ estimators [4,9] treat κ as a frequency-independent parameter. Hence, this letter also derives a fullband κ estimator which can make the LRSV estimator more practical and accurate.

4.1. Proposed κ Estimator

Inspired by the OM-LSA algorithm [8], this letter proposed an adaptive κ estimator using a probability-based framework. Given two hypotheses, $H_0(l)$ and $H_1(l)$, which indicate, respectively, direct sound absence and presence in the lth frame, as in Equation (9)

$$\begin{aligned} H_0(l) &: Z(k,l) = Z_r(k,l), \\ H_1(l) &: Z(k,l) = Z_d(k,l) + Z_r(k,l). \end{aligned} \tag{9}$$

When the direct sound is absent, the desired κ can be directly estimated according to Equation (6). Accordingly, the proposed κ estimation strategy is to recursively average past

estimated κ during periods of direct sound absence, and hold the estimate during direct sound presence. Specifically, the proposed κ estimator is as follows in Equation (10)

$$\begin{aligned} H_0(l) &: \kappa(l+1) = \alpha_\kappa \kappa(l) + (1-\alpha_\kappa)\widehat{\kappa}(l), \\ H_1(l) &: \kappa(l+1) = \kappa(l), \end{aligned} \tag{10}$$

where α_κ denotes a smoothing parameter, and $\widehat{\kappa}(l)$ denotes the estimated κ in the lth frame. Under direct sound uncertainty, the frame conditional direct sound presence probability $p(l)$ can be employed by $p(l) \triangleq P(H_1(l)|Z(k,l), k = 0, 1, \ldots, K)$, and the recursive averaging can be carried out in Equation (11)

$$\begin{aligned} \kappa(l+1) &= p(l)\kappa(l) + (1-p(l))[\alpha_\kappa \kappa(l) + (1-\alpha_\kappa)\widehat{\kappa}(l)] \\ &= \widetilde{\alpha}_\kappa(l)\kappa(l) + (1-\widetilde{\alpha}_\kappa(l))\widehat{\kappa}(l), \end{aligned} \tag{11}$$

where $\widetilde{\alpha}_\kappa(l) = p(l) + \alpha_\kappa(1-p(l))$ is a time-varying smoothing parameter which is adjusted by the frame conditional direct sound presence probability $p(l)$.

Now, there are two remaining parts in the proposed κ estimator that need to be determined: (1) the frame conditional direct sound presence probability, $p(l)$; (2) the estimated κ in the lth frame, $\widehat{\kappa}(l)$.

4.1.1. Frame Conditional Direct Sound Presence Probability

Let us assume that the STFT coefficients, $Z_d(k,l)$ and $Z_r(k,l)$, are complex Gaussian variables. Then, applying Bayes rule [8], the conditional direct sound presence probability $p(k,l) \triangleq P(H_1(l)|Z(k,l))$ can be written as Equation (12)

$$p(k,l) = \left\{ 1 + \frac{q(l)}{1-q(l)} [1+\xi(k,l)] e^{-v(k,l)} \right\}^{-1}, \tag{12}$$

where $q(l) \triangleq P(H_0(l))$ is the *a priori* probability for direct sound absence, $\xi(k,l) \triangleq \frac{\lambda_d(k,l)}{\lambda_r(k,l)}$ is the *a priori* signal-to-reverberation ratio (SRR), $\gamma(k,l) \triangleq \frac{|Z(k,l)|^2}{\lambda_r(k,l)}$ is the *a posteriori* SRR, and $v(k,l) \triangleq \frac{\gamma(k,l)\xi(k,l)}{1+\xi(k,l)}$. Note that $\gamma(k,l)$ can be calculated directly whereas $q(l)$ and $\xi(k,l)$ need to be determined.

Considering that $\lambda_z(k,l)$ decays frame by frame during periods of direct sound absence $H_0(l)$, the a priori probability for direct sound absence $q(l)$ can be defined as Equation (13)

$$q(l) = \frac{1}{K} \sum_{k=0}^{K-1} u(\lambda_z(k,l-1) - \lambda_z(k,l)), \tag{13}$$

where $u(\cdot)$ is the unit step function. Then, the *a priori* SRR $\xi(k,l)$ can be obtained via recursive average as in Equation (14)

$$\xi(k,l) = \alpha_\xi \xi(k,l-1) + (1-\alpha_\xi) \max\left\{ \frac{\lambda_z(k,l)}{\lambda_r(k,l)} - 1, 0 \right\}, \tag{14}$$

where α_ξ is a smoothing parameter.

After $p(k,l)$ is determined, the frame conditional direct sound presence probability $p(l)$ can be regarded as an average of $p(k,l)$ over all frequency bins $p(l) = \frac{1}{K} \sum_{k=0}^{K-1} p(k,l)$.

4.1.2. Estimated κ in Each Frame

Under direct sound absence hypothesis $H_0(l)$, Equation (4) becomes $\lambda_z(k,l) = \lambda_r(k,l)$, and substituting it into Equation (6) yields Equation (15)

$$\lambda_z(k,l) = \exp\left\{ \frac{-13.8R}{T_{60}(k) f_s} \right\} [(1-\kappa)\lambda_r(k,l-1) + \kappa \lambda_z(k,l-1)]. \tag{15}$$

After some algebra, Equation (15) can be rewritten as Equation (16)

$$\kappa = \frac{\exp\left\{\frac{13.8R}{T_{60}(k)f_s}\right\}\lambda_z(k,l) - \lambda_r(k,l-1)}{\lambda_z(k,l-1) - \lambda_r(k,l-1)}. \tag{16}$$

Then, the estimated κ in the lth frame is determined in Equation (17) by averaging Equation (16) in the frequency domain

$$\widehat{\kappa}(l) = \frac{\sum_{k=0}^{K-1}\left[\exp\left\{\frac{13.8R}{T_{60}(k)f_s}\right\}\lambda_z(k,l) - \lambda_r(k,l-1)\right]}{\sum_{k=0}^{K-1}\left[\lambda_z(k,l-1) - \lambda_r(k,l-1)\right]}. \tag{17}$$

Note that the numerator and the denominator of Equation (16) are separately averaged in order to avoid division by zero.

Equation (17) is similar to the conventional estimator Equation (18) [9]. However, Equation (17) is derived under direct sound absence hypothesis using Equation (6). Hence, the proposed estimator using a probability-based framework to update κ, rather than a simple heuristic used in conventional estimator. Further comparison can be found in Section 5.

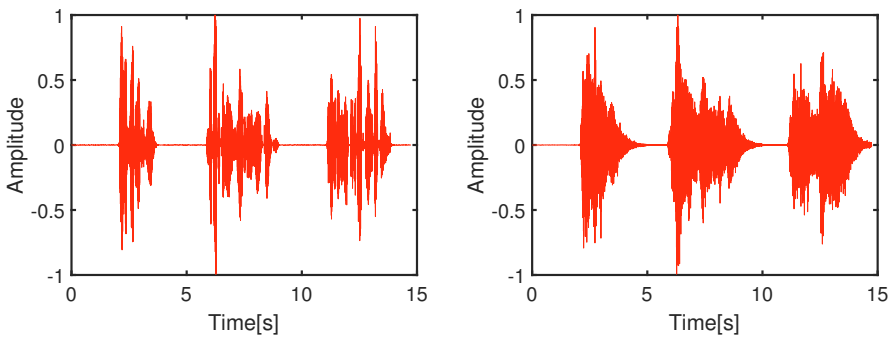
$$\widehat{\kappa}(l) = \frac{\exp\left\{\frac{13.8RN_e}{T_{60}f_s}\right\}\sum_{k=0}^{K-1}\lambda_z(k,l) - \sum_{k=0}^{K-1}\lambda_l(k,l-N_e)}{\sum_{k=0}^{K-1}\lambda_z(k,l-N_e) - \sum_{k=0}^{K-1}\lambda_l(k,l-N_e)}. \tag{18}$$

5. Performance Evaluation

In this section, the performance of the LRSV estimator using the proposed κ estimator is evaluated. The performance using κ obtained by other four different methods are also evaluated, including conventional κ estimator [9], the measured ground truth κ calculated with measured DRR and T_{60} (fullband and subband) according to Equation (8), and the scanning-optimal κ obtained by scanning method which scans κ successively from 0.05 to 1.5 at intervals of 0.01 . Besides, the quality of the dereverberated speech using proposed method is also evaluated and compared to a recent method using recursive MSPP [10].

5.1. Setup

The Signals to be processed in this letter are synthetic reverberant signals created by convolving original AIRs measured in a real hall with reverberation time of 2 s (from an open database [12]) with a male speaker signal of 15 s length. Six AIRs (referred to as $AIR_1 \sim AIR_6$) with different κ ranging from 0.12 to 1.54 are adopted. Figure 1 demonstrates the signal there was used in experiment with and without reverberation.



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

Figure 1. Plot of signal used in experiment with and without reverberation. (**a**) with reverberation; (**b**) without reverberation.

As mentioned in Section 4, the ground truth κ(fullband and 1/3-octave subband) is calculated using the measured DRR and T_{60} via Equation (8). Besides, the reverberation time T_{60} is assumed to be known. Hence, the T_{60} used in this work is directly determined in 1/3-octave subbands by applying Schroeder's method to AIRs. For evaluation purposes, the ground truth late reverberant speech component $z_l(n)$ is defined as the anechoic male speaker signal convolved with the tail of AIR starting 50 ms after the direct sound. Other parameters used in this paper are chosen empirically as $\kappa(0) = 1$, $\alpha_\kappa = 0.75$, and $\alpha_{\bar{\xi}} = 0.95$, similar to the reference [8]. All experiments are carried out in computer using MATLAB software.

The Log Spectral Distortion (LSD) [4] is adopted to evaluate the LRSV estimator by computing the root mean square(RMS) value of the difference between the estimated LRSV $\widehat{\lambda}_l(k,l)$ and the ground truth LRSV $\lambda_l(k,l)$, which is defined as Equation (19)

$$\begin{aligned} LSD_{late}(l) &= \sqrt{\frac{1}{K}\sum_{k=0}^{K-1}|e(k,l)|^2}, \\ e(k,l) &= L\left\{\widehat{\lambda}_l(k,l)\right\} - L\{\lambda_l(k,l)\}, \end{aligned} \tag{19}$$

where $L\{\cdot\} = \max\{10\lg|\cdot|, \delta\}$ is the log spectrum confined to 50 dB dynamic range and $\delta = \max_{k,l}\{10\lg|\cdot|\} - 50$. The mean LSD (refered to as $\overline{LSD}$) is obtained by averaging Equation (19) over all frames. In addition, the lower and upper semi-variance of error $e(k,l)$ were also calculated to evaluate the LRSV estimator [5] as Equation (20)

$$\begin{aligned} \sigma_l &= \sqrt{\frac{1}{|\tau_l|}\sum_{k,l\in\tau_l}(e(k,l)-\bar{e})^2}, \tau_l : e(k,l) \le \bar{e} \\ \sigma_u &= \sqrt{\frac{1}{|\tau_u|}\sum_{k,l\in\tau_u}(e(k,l)-\bar{e})^2}, \tau_u : e(k,l) > \bar{e} \end{aligned} \tag{20}$$

where $\bar{e} = \text{mean}_{k,l}\{e(k,l)\}$ is the mean value of $e(k,l)$.

In order to evaluate the robustness of the proposed estimator to noise, the white noise was added to synthetic reverberant signals with variable RSNR [5]

$$\text{RSNR} = \frac{\sum_{k,l}\lambda_d(k,l) + \lambda_r(k,l)}{\sum_{k,l}\lambda_v(k,l)} \tag{21}$$

where $\lambda_v(k,l)$ is the additive noise spectral variance.

5.2. Results and Analysis

Figure 2 depicts the mean LSD for Habets LRSV estimator using κ obtained by different methods, including the measured ground truth κ(fullband and subband), proposed κ estimator, conventional κ estimator and the scanning method.

As shown in Figure 2, an scanning-optimal κ can be obtained for each AIR as the corresponding $\overline{LSD_{late}}$ reaches a minimum during the scanning process, and it can be observed that such scanning-optimal κ is far from the measured ground truth κ, which alerts us that the measured κ may not be the optimal κ for Habets LRSV estimator. Although the measured fullband κ performs better for AIR_4 and the measured subband κ performs better for AIR_2, they perform poorly for other AIRs. As for the proposed κ estimator, the $\overline{LSD_{late}}$ value exhibits a minimum for three AIRs, and is close to the minimum for other AIRs. It suggests that the proposed κ estimator performs not only much better than the conventional κ estimator and measured ground truth κ (both fullband and subband), but even as well as the scanning-optimal κ obtained by scan method. It is worth mentioning that the scanning-optimal κ may not be the real optimal κ, but it still can be seen as an appropriate κ considering the experimental results.

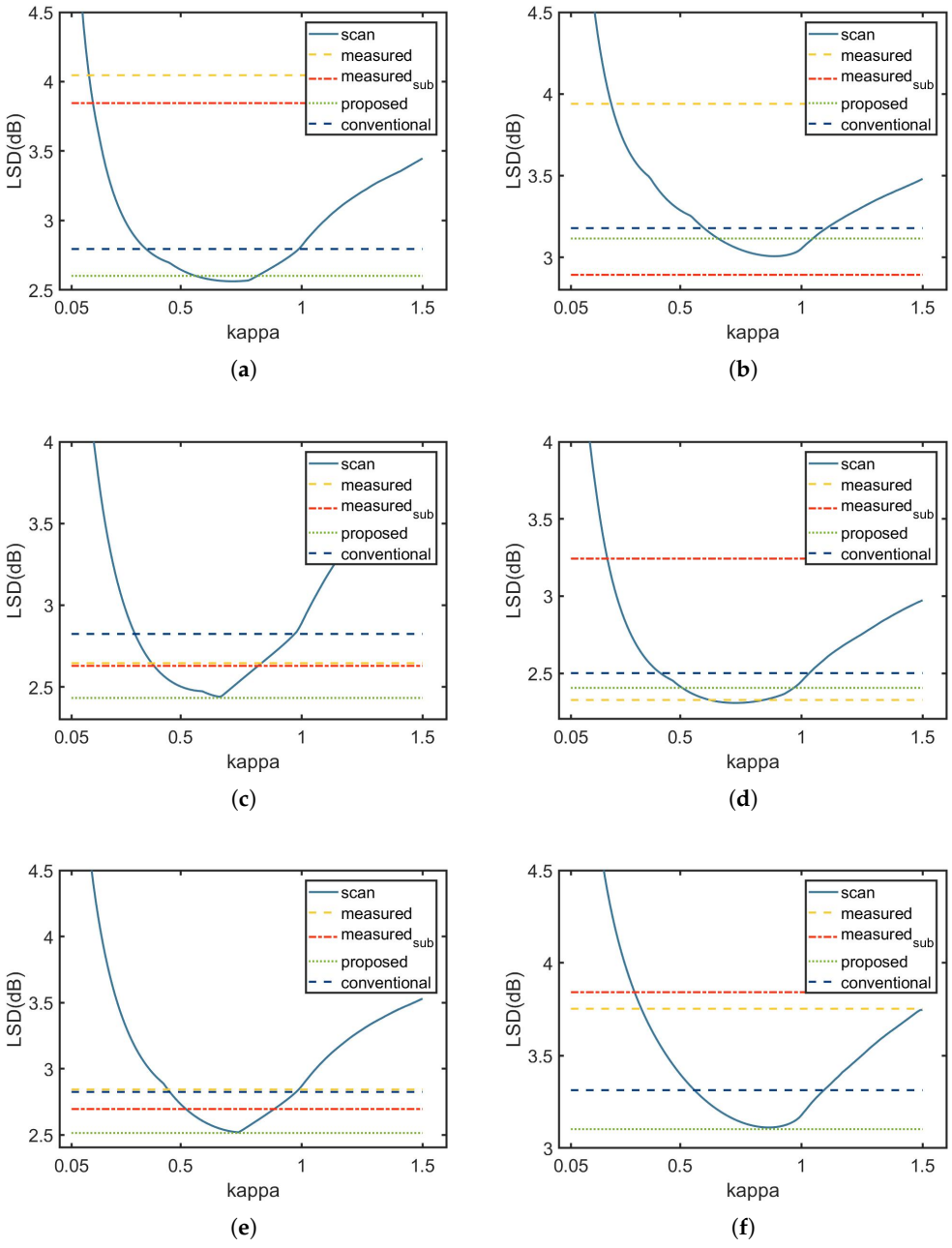


Figure 2. Plot of $\overline{LSD}$ as a function of κ for (**a**–**f**): $AIR_1 \sim AIR_6$. $\overline{LSD}$ for the measured ground truth κ(fullband and subband), proposed κ estimator and conventional κ estimator are presented as reference lines for comparison. (**a**) AIR_1; (**b**) AIR_2; (**c**) AIR_3; (**d**) AIR_4; (**e**) AIR_5; (**f**) AIR_6.

Figure 3 shows the averaged log error obtained using all RIRs for varying RSNR. As the RSNR decreases, all estimators show a more and more positive bias, which means the LRSV estimator performs worse with background noise and should be used after a denoising algorithm. However, the 'length' of the whisker bars of the proposed κ estimator is always shorter than other methods. In other words, the proposed κ estimator yields lower variance, which suggests that the proposed κ estimator is more robust with background noise.

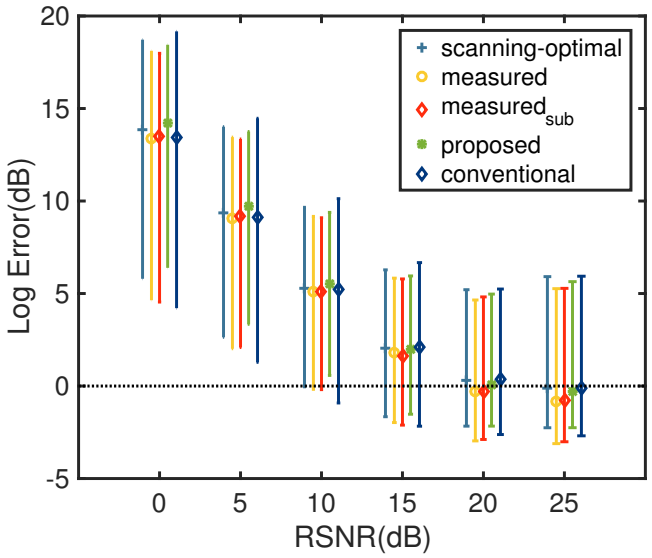


Figure 3. Mean and standard deviation of log error $e(k,l)$ for different RSNR. The means are indicated by symbols (circle, cross, etc.), and the semi-variances are indicated by whisker bars.

Figure 4 compares the measured ground truth κ with the scanning-optimal κ, and as depicted in it, the scanning-optimal κ is not obviously related to the measured ground truth κ. It precludes us from simply applying a bias correction to the measured κ, which is sometimes used in practical.

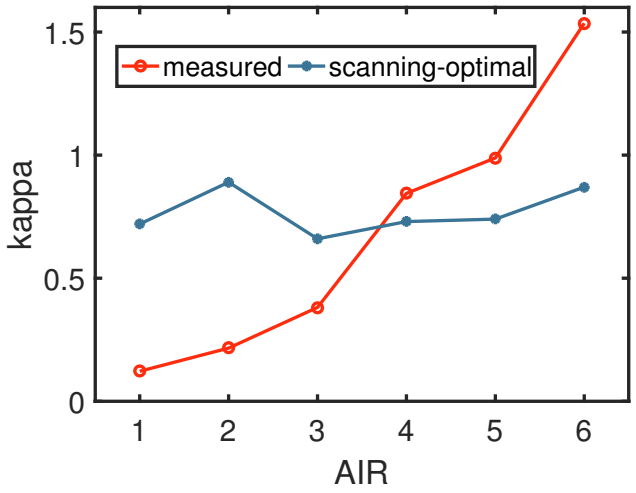


Figure 4. Plots of the measured ground truth κ and the scanning-optimal κ for each AIR.

The reason for the mismatch between the measured ground truth κ and the scanning-optimal κ may be that the generalized statistical model is a simplified approximation of AIR, which causes the error of estimation in Equation (6) and the error will vary with the anechoic speech signal $\lambda_s(k,l)$. Hence, in order to compensate that error, the value of κ needs to be modified, which makes the measured ground truth κ not the scanning-optimal κ for Habets LRSV estimator. To prove the above viewpoint, 13 different anechoic speech signals of 15 s length are used to obtain corresponding scanning-optimal κ for each AIR. The results are shown in Figure 5.

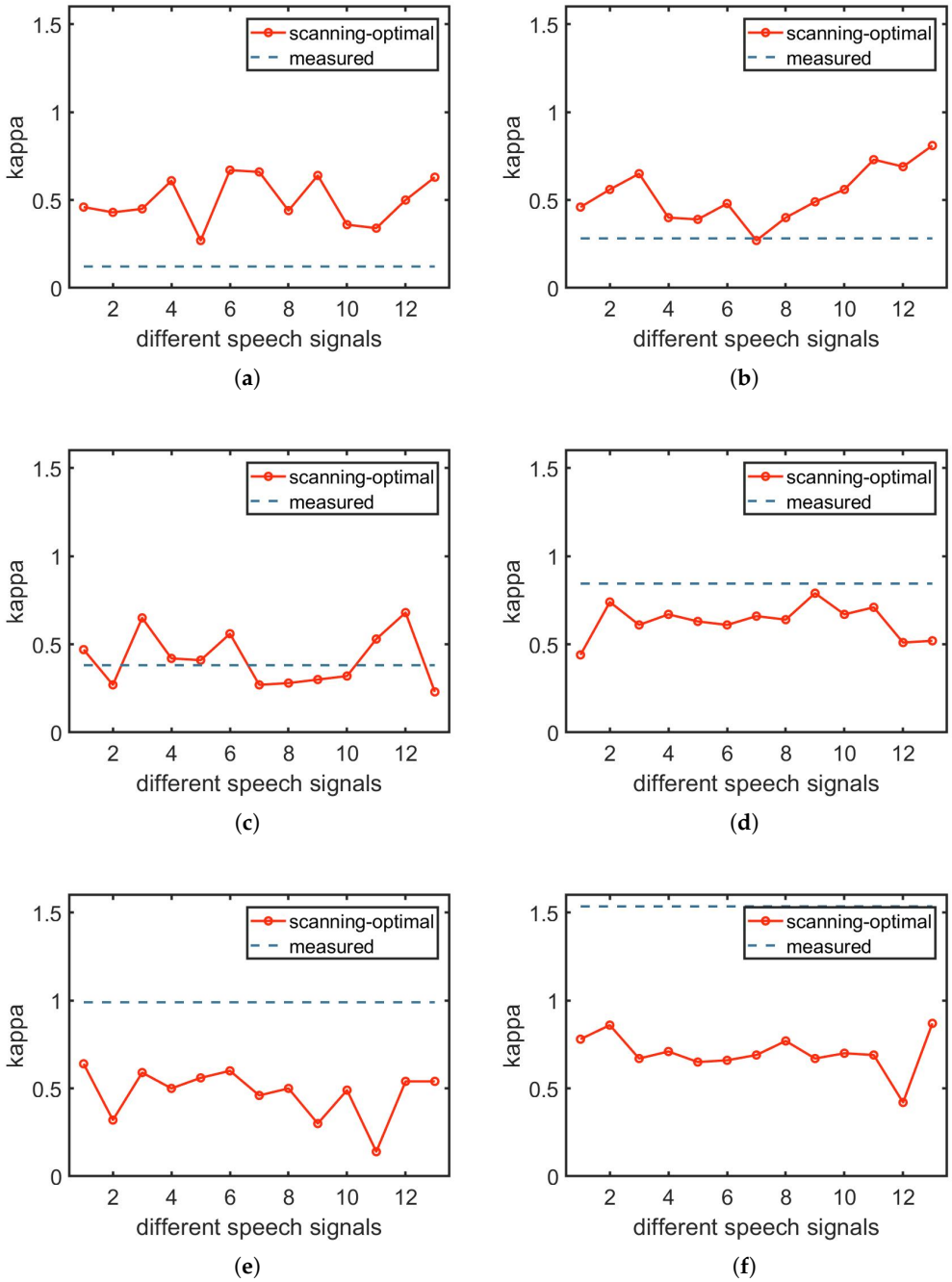


Figure 5. Plot of scanning-optimal κ using different speech signals for (**a**–**f**): $AIR_1 \sim AIR_6$. The measured κ is also presented as reference line for comparison. (**a**) AIR_1; (**b**) AIR_2; (**c**) AIR_3; (**d**) AIR_4; (**e**) AIR_5; (**f**) AIR_6.

It can be seen that for different speech signals, the scanning-optimal κ changes randomly and can differ by up to 0.54, which reveals that the optimal κ for Habets LRSV estimator may be not only related to DRR but also related to the speech signal. In other words, it may be less effective to obtain κ indirectly via the blind DRR estimation algorithm. On the contrary, estimating κ directly as the proposed method did may achieve better performance. However, this letter only uses 13 different anechoic speech signals of 15 s length each, along with six AIRs, which is not enough to prove this hypothes, further research is needed using more speech signals and more AIRs.

5.3. Speech Dereverberation

Furthermore, the quality of the dereverberated speech using the estimated LRSV is evaluated, and the log-spectral amplitude gain function [11] is adopted to suppress the late reverberant speech component. Besides, a method using recursive MSPP [10] is also evaluated as a reference. The measures are the segmental SRR and LSD (averaged over all frames) between the estimated and true early speech component [3,4], the short-time objective intelligibility(STOI) [13] and perceptual evaluation of speech quality(PESQ) [14]. The results are averaged over all AIRs and presented in Table 1.

Table 1. Improvement of objective speech quality measures.

	$\textbf{Measured}_{\textbf{full}}$	$\textbf{Measured}_{\textbf{sub}}$	$\textbf{Optimal}_{\textbf{scan}}$	**Proposed**	**Conventional**	**MSPP**
ΔSRR	7.81	7.73	7.80	8.61	6.80	7.60
ΔLSD	−3.47	−3.46	−3.57	−3.77	−3.29	−3.82
ΔSTOI	0.0678	0.0693	0.0765	0.0785	0.0676	0.0362
ΔPESQ	0.18	0.17	0.20	0.24	0.16	0.04

It can be observed that the proposed method achieves best performance in three measures and only performs slightly worse in LSD, which validates the superiority of the proposed estimator to conventional approaches. It also indicates that the LRSV estimator using proposed method performs even better than that using the measured ground truth κ(fullband and subband). It is worth mentioning that a single measure is not convincing, so this letter used four measures to jointly judge the performance of the proposed method. Hence, although MSPP has lower score than proposed method in LSD, considering all four measures, we still believe that the proposed algorithm is superior.

6. Conclusions

This work improves Habets LRSV estimator by proposing an adaptive κ estimator. We differentiate between the direct sound presence/absence hypotheses, and derive the frame conditional direct sound presence probability $p(l)$ using Bayes rule. Under the direct sound absence hypothesis, the estimated κ in the lth frame $\widehat{\kappa}(l)$ is given under the assumption of $\{\lambda_z(k,l) = \lambda_r(k,l)\}|H_0(l)$. Finally, $\kappa(l)$ is recursive averaged with a time-varying smoothing parameter $\widetilde{\alpha}_\kappa(l)$ which is adjusted by the frame conditional direct sound presence probability $p(l)$.

The proposed κ estimator has been evaluated and compared to conventional κ estimator and a recursive MSPP method proposed in recent years. Experimental results show that the LRSV estimator using the proposed κ estimator outperforms other methods. It is also found that the ground truth κ calculated using measured DRR is not the optimal κ for the LRSV estimator since the optimal κ may be affected by speech signals. It suggests us estimate κ directly and adaptively rather than using the blind DRR estimation algorithm to obtain κ, which may be a less effective approach. However, further research is needed to prove this hypothesis.

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

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

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

References

1. Yang, W.; Huang, G.; Chen, J.; Benesty, J.; Cohen, I.; Kellermann, W. Robust Dereverberation With Kronecker Product Based Multichannel Linear Prediction. *IEEE Signal Process. Lett.* **2021**, *28*, 101–105. [CrossRef]
2. Braun, S.; Schwartz, B.; Gannot, S.; Habets, E.A.P. Late reverberation PSD estimation for single-channel dereverberation using relative convolutive transfer functions. In Proceedings of the 2016 IEEE International Workshop on Acoustic Signal Enhancement (IWAENC), Xi'an, China, 13–16 September 2016; pp. 1–5.
3. Habets, E.A.P.; Gannot, S.; Cohen, I. Late Reverberant Spectral Variance Estimation Based on a Statistical Model. *IEEE Signal Process. Lett.* **2009**, *16*, 770–773. [CrossRef]
4. Naylor, P.A.; Gaubitch, N.D. *Speech Dereverberation*, 1st ed.; Springer Publishing Company Incorporated: Manhattan, NY, USA, 2010.
5. Braun, S.; Kuklasiński, A.; Schwartz, O.; Thiergart, O.; Habets, E.A.P.; Gannot, S.; Doclo, S.; Jensen, J. Evaluation and Comparison of Late Reverberation Power Spectral Density Estimators. *IEEE/ACM Trans. Audio Speech Lang. Process.* **2018**, *26*, 1056–1071. [CrossRef]
6. Habets, E.A.P.; Gannot, S.; Cohen, I. Speech dereverberation using backward estimation of the late reverberant spectral variance. In Proceedings of the 2008 IEEE 25th Convention of Electrical and Electronics Engineers in Israel, Eilat, Israel, 3–5 December 2008; pp. 384–388.
7. Eaton, J.; Gaubitch, N.D.; Moore, A.H.; Naylor, P.A. Estimation of Room Acoustic Parameters: The ACE Challenge. *IEEE/ACM Trans. Audio Speech Lang. Process.* **2016**, *24*, 1681–1693. [CrossRef]
8. Cohen, I. Optimal speech enhancement under signal presence uncertainty using log-spectral amplitude estimator. *IEEE Signal Process. Lett.* **2002**, *9*, 113–116. [CrossRef]
9. Erkelens, J.S.; Heusdens, R. Noise and late-reverberation suppression in time-varying acoustical environments. In Proceedings of the 2010 IEEE International Conference on Acoustics, Speech and Signal Processing, Dallas, TX, USA, 14–19 March 2010; pp. 4706–4709.
10. Herzog, A.; Habets, E.A.P. Blind Single-Channel Dereverberation Using a Recursive Maximum-Sparseness-Power-Prediction-Model. In Proceedings of the 2018 16th International Workshop on Acoustic Signal Enhancement (IWAENC), Tokyo, Japan, 17–20 September 2018; pp. 356–360.
11. Wolfe, P.; Godsill, S. Efficient Alternatives to the Ephraim and Malah Suppression Rule for Audio Signal Enhancement. *EURASIP J. Adv. Signal Process.* **2003**, *2003*, 910167. [CrossRef]
12. Merimaa, J.; Peltonen, T.; Lokki, T. Concert Hall Impulse Responses Pori, Finland. 2005. Available online: http://www.acoustics.hut.fi/projects/poririrs/ (accessed on 27 August 2021).
13. Taal, C.H.; Hendriks, R.C.; Heusdens, R.; Jensen, J. An Algorithm for Intelligibility Prediction of Time–Frequency Weighted Noisy Speech. *IEEE Trans. Audio Speech Lang. Process.* **2011**, *19*, 2125–2136. [CrossRef]
14. Rix, A.W.; Beerends, J.G.; Hollier, M.P.; Hekstra, A.P. Perceptual evaluation of speech quality (PESQ)-a new method for speech quality assessment of telephone networks and codecs. In Proceedings of the 2001 IEEE International Conference on Acoustics, Speech, and Signal Processing, Proceedings (Cat. No.01CH37221), Salt Lake City, UT, USA, 7–11 May 2001; Volume 2, pp. 749–752.