



A statistical room impulse response model with frequency dependent reverberation time for single-microphone late reverberation suppression

Jan S. Erkelens, Richard Heusdens

Department of Mediamatics, Delft University of Technology, The Netherlands

j.s.erkelens@tudelft.nl, r.heusdens@tudelft.nl

Abstract

Single-channel late reverberation suppression algorithms need estimates of the late reverberance spectral variance (LRSV) in order to suppress the late reverberance. Often the LRSV estimators are derived from a statistical room impulse response (RIR) model. Usually the late reverberation is modeled as a white Gaussian noise sequence with exponentially decaying variance. The whiteness assumption means that the same decay constant is assumed for all frequencies. Since there is generally more absorption of sound energy with increasing frequency, there is a need for RIR models that take this into account. We propose a new statistical time-varying RIR model that consists of a sum of decaying cosine functions with random phases, with a frequency dependent decay constant. We show that the resulting LRSV estimators have the same form as existing ones, but with an inherent frequency dependency of the decay constant. Experiments with real measured RIRs, however, indicate that, for the purpose of reverberation suppression, using a frequency independent decay constant is often sufficiently good.

A common assumption in the derivation of LRSV estimators is that the direct signal and early reflections are uncorrelated with the late reverberation. We verify this assumption experimentally on measured RIRs and conclude that it is accurate.

Index Terms: speech enhancement, echo suppression, statistical room impulse response models, reverberation time.

1. Introduction

A large amount of work has been done on the development of single-channel additive-noise reduction algorithms [1–4], during the past three decades. More recently, there has been a growing interest in single-microphone reverberation cancellation and suppression, e.g., [5–12].

Lebart *et al.* [5] proposed to treat the late reverberation as a kind of nonstationary additive noise which has to be suppressed. From a statistical model of the RIR, an estimator of the *Late Reverberation Spectral Variance* (LRSV) was derived that was used in a spectral subtraction algorithm to suppress the late reverberation. This statistical RIR model depends only on the reverberation time T_{60} . Habets [8, 11] showed that the model of Lebart *et al.* will lead to overestimation of the LRSV when the *Direct-to-Reverberation-Ratio* (DRR) is larger than about 1. To avoid this overestimation, a more flexible RIR model was introduced. Just as in the case of Lebart's model, the resulting LRSV estimator follows a simple recursive equation. However, more parameters, such as the DRR, have to be estimated blindly. Several blind single-channel estimators of T_{60} and DRR have been proposed recently [8, 13–18].

In [12] it was shown how the LRSV can also be estimated from the long-term correlations induced by the RIR, without having to estimate neither reverberation time nor DRR.

However, this correlation-based method requires the RIR to be nearly time-invariant. In the derivations of the model-based LRSV estimators mentioned above time-invariance is actually assumed as well. Since in practice the RIRs are often time-varying, Erkelens and Heusdens [12] studied the time-varying case. Interestingly, it was shown that in that case the existing estimators may still be used, provided some mild conditions on the rate of change of T_{60} , the DRR, and the correlations between RIR tap values at different time indices are met.

However, all existing model-based LRSV estimators stem from statistical RIR models that use white Gaussian noise processes with a decaying variance. The corresponding decay constant is identical for each frequency bin and corresponds to the one that can be found with the Schroeder integral [19] of the time domain RIR sequence. Although this may be a good approximation for certain rooms, especially when working with narrowband speech signals, it is not always the case. For example, in [20] the reverberation time measured in 1/3 octave subbands shows much more variation for a corridor or a stairway than for an office or a kitchen. There is clearly a need for statistical RIR models and corresponding LRSV estimators that take the frequency dependency of reverberation time into account. In this paper, we propose a new statistical model that describes possibly time-varying RIRs as realizations of a sum of decaying cosines with random phases. This model adopts the form of the solution to the wave equation for rectangular rooms with damping [21]. The decay constant of each cosine depends on the frequency dependent reverberation time. We show that existing LRSV estimators follow approximately also from this model, if we insert a frequency dependent reverberation time.

The paper is organized as follows. In Section 2 the new statistical RIR model and the corresponding LRSV estimators are introduced. The LRSV estimators are derived in Appendix A and evaluated experimentally in Section 3. The paper is concluded in Section 4.

2. RIR model and LRSV estimators

We assume the reverberant signal z to be the convolution of a source speech signal s and a, possibly time-varying, RIR h , as follows

$$z(n) = \sum_{l=0}^{\infty} h_n(l)s(n-l), \quad (1)$$

where n is the discrete-time sample index. Since absorption of sound energy is frequency dependent, the reverberation time generally decreases with increasing frequency [21]. An example can be found in [20]. This results in RIR spectra that are not flat, and there is a time-dependent correlation between the RIR samples. This time-dependent correlation is not described by current statistical RIR models in the literature. We propose a

new statistical RIR model with a frequency dependent decay. It will be used to derive LRSV estimators with a frequency dependent reverberation time. The proposed model looks as follows

$$h_n(l) = \begin{cases} 1 & : l = 0 \\ \sum_{j=1}^J \sigma_{j,n} e^{-\delta_{j,n} l} \cos(\omega_j l + \phi_{j,n}) & : l \geq 1. \end{cases} \quad (2)$$

Here the direct path is modeled by a discrete delta pulse and the reverberant part by a sum of decaying cosine functions. We allow the amplitude $\sigma_{j,n}$, decay constants $\delta_{j,n}$ and phases $\phi_{j,n}$ to be time-varying. We assume the phases to be uniformly distributed and independent for different values of j . Note that the reverberant part of this model has the form of the solution to the wave equation for rectangular rooms with damping [21]. Physically, the frequencies ω_j correspond to the modal frequencies of the room. From a practical point of view we assume the number of components J to be large enough to accurately model the late reverberation in a statistical sense. The decay rates are determined by the frequency dependent reverberation times T_{60} of the room, as follows

$$\delta_{j,n} = \frac{3 \ln(10)}{F_s T_{60}(j, n)}, \quad (3)$$

where F_s is the sampling frequency. We assume in this paper that the RIRs do not change much during half a DFT frame length¹. In Appendix A we then show that the LRSV λ_r in the DFT domain can be approximated by

$$\lambda_r(k, m) \approx e^{-2\bar{\delta}_{k,m} L} \{ \lambda_z(k, m-L) - \lambda_s(k, m-L) \}, \quad (4)$$

where k is the frequency bin index, m the time sample index, and L is the interval in samples after which the late reverberation part of the RIR is assumed to begin. The expression contains a decay parameter $\bar{\delta}(k, m)$, where the overline means an average over a narrow band of frequencies around the frequency corresponding to DFT bin k . The spectral variance of the reverberant signal is λ_z and that of the direct-path signal is λ_s . When the late reverberation is assumed uncorrelated from the sum of direct-path signal and early reflections, we have

$$\lambda_z(k, m) = \lambda_{de}(k, m) + \lambda_r(k, m), \quad (5)$$

where λ_{de} is the spectral variance of direct-path signal plus early reflections. We will experimentally verify (5) in Section 3. Using (5), (4) can be written in the alternative form [12]

$$\lambda_r(k, m) \approx e^{-2\bar{\delta}_{k,m} L} \{ \kappa_{k,m} \lambda_{de}(k, m-L) + \lambda_r(k, m-L) \}, \quad (6)$$

where the DRR-related parameter $\kappa_{k,m}$ is defined as $1 - \lambda_s(k, m) / \lambda_{de}(k, m)$. The estimator (6) has the advantage over (4) that the spectral variances λ_z and λ_s do not have to be estimated and we can use the decision-directed approach [22] for estimation of λ_{de} . For the model (2) we have (see Appendix A for details)

$$\kappa_{k,m} \approx \frac{\bar{\sigma}_{k,m}^2 \Phi(\omega_k)}{4 + \bar{\sigma}_{k,m}^2 \Phi(\omega_k)}, \quad (7)$$

where ω_k is here the k -th DFT frequency and $\Phi(\omega)$ is a function that depends on the decay rate and also on L , as follows

$$\Phi(\omega) = \left(\frac{1 - \alpha}{\beta - 1} \right)^2 + \left(\frac{1 - 2\alpha \cos(2\omega L) + \alpha^2}{\beta^2 - 2\beta \cos(2\omega) + 1} \right), \quad (8)$$

¹ The case of more rapid RIR fluctuations is currently under study.

where β and α are defined as $\beta = e^{\bar{\delta}(k,m)}$ and $\alpha = \beta^{-L}$. The estimators (4) and (6) have the same form as those derived from an RIR model that describes the reverberance as white Gaussian noise with a decaying variance [12]. The difference is in the inherent frequency dependency of the attenuation factor $e^{-2\bar{\delta}L}$ in (4) and (6), and in the expression for κ (7).

3. Experimental evaluations

In the derivation of the LRSV estimator (6) and others [11, 12] it has been assumed that the signal consisting of the direct-path plus early reflections is uncorrelated with the late reverberance so that (5) holds. We will experimentally verify this equation as follows. For a number of measured RIRs from the AIR database [23], we will construct three signals. The first one, $z(n)$, is formed by filtering 1 minute of clean speech (silence frames removed) with an RIR. Both the clean speech and RIRs were sampled at 16 kHz sampling frequency. The second one, $z_{de}(n)$ is the result of filtering with only the first $L=512$ samples (32 ms) of that RIR and is therefore the sum of the direct-path signal plus the early reflections. For the third filtered signal $r(n)$ we set the first 512 samples of the RIR to zero and $r(n)$ is therefore the late-reverberance signal. The short time DFT spectra of these signals are computed in frames of 32 ms, and averaged. Now, we compute the Logarithmic Error (*LogErr*, see, e.g., [24]) between the average spectrum of $z(n)$, and the sum of the average spectrum of $z_{de}(n)$ and of $r(n)$. The results are shown in Table 1 and we see that equation (5) is very accurate. We also computed the time-domain correlation coefficient between $z_{de}(n)$ and $r(n)$. In all cases was this correlation coefficient very close to zero.

Table 1: *LogErr* between the average reverberant spectrum and the sum of the average spectrum of direct signal plus early reflections and the average spectrum of the late reverberance for different measured RIRs from [23].

RIR	T_{60} [s]	DRR [dB]	<i>LogErr</i> [dB]
Kitchen 1	0.42	11.2	0.04
Kitchen 2	0.52	4.6	0.14
Lecture hall 1	0.70	2.0	0.13
Lecture hall 2	0.83	-10.2	0.29
Corridor 1	1.25	11.0	0.03
Corridor 2	1.47	4.4	0.11

Table 2 shows the values of reverberation time T_{60} , the mean $\bar{T}_{60}(k)$ and standard deviation $\sigma_{T_{60}}$ of its frequency dependent equivalent, the mean $\bar{g}(k)$ and standard deviation σ_g of the attenuation factor $g = e^{-2\delta L}$, the mean $\bar{\kappa}(k)$ and standard deviation σ_κ of the parameter κ , and the DRR, for several measured RIRs. Reverberation time was measured directly from the RIRs, either from the time domain Schroeder integral or from the Schroeder integral in subbands. We used 8 subbands of width 1 kHz each, to find the frequency dependent reverberation times. In the LRSV estimation experiment described later in this section, we used for each DFT bin the value of the corresponding subband. We used the reverberant test speech signals of 3 minutes in length to measure $\kappa(k)$ in each DFT bin. We see that there is considerable variation across frequency in reverberation time, especially for "Corridor 3" and "Stairway". However, the variation in the attenuation factor $g(k)$ is much smaller. There is quite a lot of variation in $\kappa(k)$.

We compare now the performance of three LRSV estimators: the first one, $\lambda_{\kappa(k)}^{T_{60}}$ is (6) with the frequency independent

Table 2: Frequency independent RIR parameters and mean and standard deviations of frequency dependent RIR parameters for the RIRs used in the dereverberation experiments.

RIR	T_{60} [s]	$T_{60}(k)$ [s]	$\sigma_{T_{60}}$ [s]	$g(k)$	σ_g	$\kappa(k)$	σ_κ	DRR [dB]
Corr. 3	1.48	1.26	0.41	0.68	0.09	0.14	0.16	6.8
Office	0.66	0.62	0.12	0.48	0.07	0.20	0.20	5.8
Stairway	1.68	1.47	0.40	0.73	0.07	0.19	0.19	4.5

T_{60} and the mean value of $\kappa(k)$ from Table 2. The second one, $\lambda_{\kappa(k)}^{T_{60}}$, also uses the frequency independent T_{60} , but the frequency dependent $\kappa(k)$, while the third one, $\lambda_{\kappa(k)}^{T_{60}(k)}$, uses the frequency dependent $T_{60}(k)$ and $\kappa(k)$. These estimators are compared both in clean conditions and with Office noise (from the NTT monaural noise database) with an SNR of 10 dB measured with respect to the reverberant speech. The sampling frequency was 16 kHz, the frame length was $N=512$ samples (32 ms), and we used $L=N$ in (6). Prior to the reverberation suppression, a noise suppression step was performed with a conventional noise suppression system. The suppression gain function for that algorithm was based on a generalized Gamma speech amplitude prior with parameters $\gamma=1$ and $\nu=1$ [25]. The noise spectra were tracked with our data-driven MMSE method [24]. LRSV estimation performance was measured in terms of LogErr where the sum of estimated noise variance and LRSV is evaluated (see also [12]). Table 3 shows the results. We see that using a frequency dependent value for T_{60}

Table 3: LogErr [dB] for the three LRSV estimators described in the main text that use frequency dependent and/or independent values for T_{60} and κ .

RIR	10 dB SNR Office noise			Noise free		
	$\lambda_{\kappa(k)}^{T_{60}}$	$\lambda_{\kappa(k)}^{T_{60}}$	$\lambda_{\kappa(k)}^{T_{60}(k)}$	$\lambda_{\kappa(k)}^{T_{60}}$	$\lambda_{\kappa(k)}^{T_{60}}$	$\lambda_{\kappa(k)}^{T_{60}(k)}$
Corr. 3	2.14	1.86	1.84	2.57	2.19	2.18
Office	2.35	2.14	2.15	2.82	2.58	2.60
Stairway	1.88	1.77	1.76	2.21	2.08	2.07

does not give better results than using a frequency independent value. This is no surprise, because the variation in the attenuation factors is rather small, as can be seen in Table 2. Using a frequency dependent value for κ is however slightly better than using the average value.

4. Concluding remarks

We have proposed a new statistical RIR model which describes the RIRs as a sum of decaying cosines with frequency dependent reverberation times. We show that this model leads to similar estimators as found in the literature, but with an inherent frequency dependency of the decay constants. Experiments show that in practice using a frequency independent reverberation time will often be sufficiently good, but we can expect improvements from using a frequency dependent parameter κ . Accurate blind estimation of this DRR-related parameter is, however, still an open problem.

5. References

- [1] J. Benesty, S. Makino, and J. Chen (Eds.), *Speech Enhancement*. Springer, 2005.
- [2] Y. Ephraim and I. Cohen, "Recent advancements in speech enhancement," in *The Electrical Engineering Handbook*. CRC Press, 2006.
- [3] P. Vary and R. Martin, *Digital Speech Transmission*. John Wiley & Sons, 2006.
- [4] P. C. Loizou, *Speech Enhancement. Theory and Practice*. CRC Press, 2007.
- [5] K. Lebart, J. M. Boucher, and P. N. Denbigh, "A new method based on spectral subtraction for speech dereverberation," *Acta Acustica*, vol. 87, no. 3, pp. 359 – 366, 2001.
- [6] P. A. Naylor and N. D. Gaubitch, "Speech dereverberation," *Proc. IWAENC*, 2005.
- [7] T. Nakatani, K. Kinoshita, and M. Miyoshi, "Harmonicity-based blind dereverberation for single-channel speech signals," *IEEE Trans. Audio, Speech and Lang. Proc.*, vol. 15, no. 1, pp. 80 – 95, January 2007.
- [8] E. A. P. Habets, "Single- and multi-microphone speech dereverberation using spectral enhancement," *Ph.D. Thesis, Eindhoven Univ. of Techn., The Netherlands*, 2007.
- [9] B. Yegnanarayana and P. S. Murthy, "Enhancement of reverberant speech using LP residual signal," *IEEE Trans. Speech, Audio Proc.*, vol. 8, no. 3, pp. 267 – 281, may 2000.
- [10] T. Nakatani, T. Yoshioka, K. Kinoshita, M. Miyoshi, and B.-H. Juang, "Real-time speech enhancement in noisy reverberant multi-talker environments based on a location-independent room acoustics model," in *Proc. ICASSP*, 2009, pp. 137–140.
- [11] E. A. P. Habets, S. Gannot, and I. Cohen, "Late reverberant spectral variance estimation based on a statistical model," *IEEE Signal Proc. Lett.*, vol. 16, no. 9, pp. 770 – 773, September 2009.
- [12] J. S. Erkelens and R. Heusdens, "Correlation-based and model-based blind single-channel late-reverberation suppression in noisy time-varying acoustical environments," *IEEE Trans. Audio, Speech and Lang. Proc.*, vol. 18, no. 7, pp. 1746 – 1765, 2010.
- [13] R. Ratnam, D. L. Jones, B. C. Wheeler, W. D. O'Brien Jr., C. R. Lansing, and A. S. Feng, "Blind estimation of reverberation time," *J. Acoust. Soc. Am.*, vol. 114, no. 5, pp. 2877–2892, 2003.
- [14] M. Unoki and S. Hiramatsu, "MTF-method of blind estimation of reverberation time in room acoustics," in *Proc. EUSIPCO*, 2008.
- [15] J. Y. C. Wen, E. A. P. Habets, and P. A. Naylor, "Blind estimation of reverberation time based on the distribution of signal decay rates," in *Proc. ICASSP*, 2008, pp. 329–332.
- [16] T. H. Falk and W.-Y. Chan, "Temporal dynamics for blind measurement of room acoustical parameters," *IEEE Trans. Instrumentation and Measurement*, vol. 59, no. 4, pp. 978 – 989, 2010.
- [17] H. W. Löllman, E. Yilmaz, M. Jeub, and P. Vary, "An improved algorithm for blind reverberation time estimation," in *Proc. IWAENC*, August 2010.
- [18] H. W. Löllman and P. Vary, "Estimation of the frequency dependent reverberation time by means of warped filter-banks," in *Proc. ICASSP*, 2011, pp. 309–312.
- [19] M. R. Schroeder, "New method of measuring reverberation time," *J. Acoust. Soc. Am.*, vol. 37, no. 6, pp. 1187–1188, June 1965.
- [20] M. Jeub, M. Schäfer, H. Krüger, C. Nelke, C. Beaugeant, and P. Vary, "Do we need dereverberation for hand-held telephony?" in *Proc. 20th Int. Congress on Acoust.*, Sydney, Australia, 2010.
- [21] H. Kuttruff, *Room Acoustics*, 4th ed. Spon press, 2000.
- [22] Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator," *IEEE Trans. Acoust., Speech, Signal Proc.*, vol. ASSP-32, no. 6, pp. 1109–1121, December 1984.
- [23] M. Jeub, M. Schäfer, and P. Vary, "A binaural room impulse response database for the evaluation of dereverberation algorithms," in *Proc. 16th International Conference on Digital Signal Processing (DSP)*, Santorini, Greece, 2009.
- [24] J. S. Erkelens and R. Heusdens, "Tracking of nonstationary noise based on data-driven recursive noise power estimation," *IEEE Trans. Audio, Speech and Lang. Proc.*, vol. 16, no. 6, pp. 1112–1123, August 2008. [Online]. Available: www.mathworks.com/matlabcentral/fileexchange/27311.
- [25] J. S. Erkelens, R. C. Hendriks, R. Heusdens, and J. Jensen, "Minimum mean-square error estimation of discrete Fourier coefficients with generalized Gamma priors," *IEEE Trans. Audio, Speech and Lang. Proc.*, vol. 15, no. 6, pp. 1741–1752, August 2007.

A. Derivations

In this appendix we outline the main steps in the derivation of the LRSV estimators given in Section 2. If we assume that the RIR in (2) varies slowly during the length of one DFT frame, then the DFT $Z(k, m)$ of a frame of the reverberant signal z that starts at sample index m is given by [12]

$$Z(k, m) \approx \sum_{l=0}^{\infty} \tilde{h}_m(l) S(k, m-l), \quad (9)$$

where the $S(k, l)$ are the DFT coefficients of the source signal s and $\tilde{h}_m(l) = h_{m+N/2}(l)$ with N the frame length in samples. The DFT coefficients of all reverberation (early plus late), Z_{er} , are found by subtracting the direct path contribution

$$Z_{er}(k, m) = Z(k, m) - S(k, m) \approx \sum_{l=1}^{\infty} \tilde{h}_m(l) S(k, m-l). \quad (10)$$

The reverberance spectral variance λ_{er} is now given by

$$\lambda_{er}(k, m) = E_s E_\phi |Z_{er}(k, m)|^2 = \lambda_z(k, m) - \lambda_s(k, m), \quad (11)$$

where the last equality follows from the random phase assumption made in (2). E_s is the expectation operator over the speech process, and E_ϕ evaluates the expectation over the random phases in (2). If we substitute (2) and (10) into (11), we can see that we have to take expectations over products of cosine functions. Since the phases are assumed to be independent, these expectations are easily evaluated to

$$E_\phi \cos(\omega_j l + \tilde{\phi}_{j,m}) \cos(\omega_{j'} l' + \tilde{\phi}_{j',m}) = \begin{cases} 0 & : j \neq j' \\ 1/2 \cos(\omega_j(l-l')) & : j = j', \end{cases} \quad (12)$$

where $\tilde{\phi}_{j,m} = \phi_{j,m+N/2}$. Equation (12) means that the double summation over frequency indices j and j' reduces to a single summation over j . We now arrive at

$$\lambda_{er}(k, m) \approx E_s \sum_{l=1}^{\infty} \sum_{l'=1}^{\infty} \sum_{j=1}^J \left[\frac{\tilde{\sigma}_{j,m}^2}{2} e^{-\tilde{\delta}_{j,m}(l+l')} \times \cos(\omega_j(l-l')) S(k, m-l) S^*(k, m-l') \right], \quad (13)$$

with $\tilde{\sigma}_{j,m} = \sigma_{j,m+N/2}$ and $\tilde{\delta}_{j,m}$ defined similarly.

We will argue that the biggest contribution to $\lambda_{er}(k, m)$ comes from frequencies ω_j near the DFT frequency of bin k . Figure 1(a) shows the phase angle $\psi(k, l)$ in a short time interval (i.e., for increasing values of l) of $S(k, l)$ for the frequency bin k corresponding to 1500 Hz. We see that the phase changes almost linearly with time. Figure 1(b) shows the positive half of the spectrum of $e^{i\psi(k,l)}$ obtained from about 2 seconds of the signal shown in (a). The speech signal was sampled at 16 kHz and the DFT frame size was 32 ms. Similar plots can be made for the other frequency bins. The effect of the nearly linear behavior of the phases of $S(k, l)$ on the summations over l and l' in (13) is to *emphasize* the contributions of the ω_j near the DFT frequency $2\pi k/N$ and to attenuate the others. Therefore (13) can be approximated by

$$\lambda_{er}(k, m) \approx E_s \sum_{l=1}^{\infty} \sum_{l'=1}^{\infty} \left[\frac{\tilde{\sigma}_{k,m}^2}{2} e^{-\tilde{\delta}_{k,m}(l+l')} \times \cos(\omega_k(l-l')) S(k, m-l) S^*(k, m-l') \right], \quad (14)$$

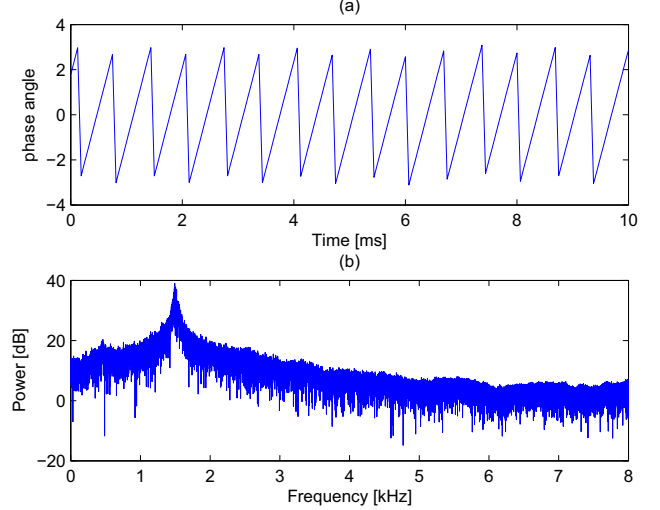


Figure 1: (a) Phase angle $\psi(k, l)$ of clean speech DFT coefficient as a function of time for the DFT bin corresponding to 1500 Hz and (b) Positive half of the spectrum of $e^{i\psi(k,l)}$ for 2 seconds of speech sampled at 16 kHz.

where ω_k is now the DFT frequency $2\pi k/N$, and $\tilde{\sigma}_{k,m}^2$ and $\tilde{\delta}_{k,m}$ must be considered as weighted sums/averages of the $\tilde{\sigma}_{j,m}^2$ and $\tilde{\delta}_{j,m}$ in a narrow frequency band around ω_k . This forms a justification for using a frequency dependent decay constant in the LRSV estimators, as was done in [8, 11], for example.

Note that the early reverberance spectral variance (ERSV) $\lambda_e(k, m)$ and the LRSV $\lambda_r(k, m)$ follow similar expressions as (14), but with sums that run from 1 to L and $L+1$ to ∞ , respectively. Now, similar to what was done in [5, 8, 12], we can evaluate $\lambda_{er}(k, m-L)$ and relate it to $\lambda_r(k, m)$. If we assume that $\tilde{\sigma}_{k,m}^2$ and $\tilde{\delta}_{k,m}$ change little over a time period of L samples, then the result is simply

$$\lambda_{er}(k, m-L) = e^{2\tilde{\delta}_{k,m}L} \lambda_r(k, m), \quad (15)$$

which, using (11), immediately leads to (4).

If we make the additional assumption that the speech is stationary during an interval of L samples, then we can find a closed-form expression for the ERSV. With the assumption of a linearly changing phase, we then have, for $l, l' \leq L$

$$E_s S(k, m-l) S^*(k, m-l') \approx e^{2\pi i k/N(l'-l)} \lambda_s(k, m). \quad (16)$$

The $\cos(\omega_k(l-l'))$ terms can be written as the sum of two complex exponentials. Therefore we can compute the sums in the ERSV over l, l' from 1 to L explicitly, using the expression for the geometric series. The result is

$$\lambda_e(k, m) \approx \frac{\tilde{\sigma}_{k,m}^2}{4} \Phi(\omega_k) \lambda_s(k, m), \quad (17)$$

where $\Phi(\omega)$ was given in (8). From the random phase assumption made in (2), we have $\lambda_{de}(k, m) = \lambda_s(k, m) + \lambda_e(k, m)$ and the expression for κ in (7) follows.

In the derivations we have assumed slow time variations in the RIR such that (9) holds. The case of faster variations is more difficult, but we can easily show that if the RIR changes so quickly that the $\phi_{j,m}$ become uncorrelated for different m , even for the same j , then the reverberation will behave as non-stationary additive *white* noise. Time variations in the RIR tend to decorrelate the reverberation (see also [12]).