NOISE AND LATE-REVERBERATION SUPPRESSION IN TIME-VARYING ACOUSTICAL ENVIRONMENTS

Jan S. Erkelens and Richard Heusdens

Signal & Information Processing Lab, Delft University of Technology, Delft, The Netherlands
Email: {j.s.erkelens, r.heusdens}@tudelft.nl

ABSTRACT

We consider single-channel blind late-reverberation suppression in noisy and time-varying acoustical environments. Existing estimators for the late reverberant spectral variance (LRSV) are derived assuming the room impulse responses (RIRs) to be time-invariant realizations of a stochastic process. In this paper, we go one step further and analyze time-varying RIRs. We show theoretically that existing LRSV estimators may still be used even when the individual RIR filter taps vary rapidly with time, provided that the reverberation time $T_{60}$ and direct-to-reverberation ratio (DRR) remain nearly constant during an interval of the order of a few frames. We show that these parameters can be taken frequency-independent in DFT-based enhancement algorithms. We estimate them blindly. Experiments with time-varying RIRs validate the analysis and show the importance of accurate estimation of the reverberant spectral variance. Experiments with additive non-stationary noise show the influence of $T_{60}$ and DRR estimation.

Index Terms— Speech enhancement, echo suppression.

1. INTRODUCTION

Noise and reverberation decrease the quality and intelligibility of speech [1]. Generally also the performance of automatic speech recognition systems decreases [2, 3]. It is therefore of much interest to develop processing algorithms that enhance speech degraded by additive and convolutive distortions. Recently, several blind single-microphone dereverberation techniques have been proposed in the literature, e.g., [3–9]. It has been shown [3] that blind reverberation cancelation is possible for time-invariant RIRs, but in many practical situations the RIRs may vary and there may be nonstationary additive noise as well. In those cases, a more feasible approach seems to treat the late reverberation as a kind of nonstationary additive noise and use a noise suppression algorithm to reduce it, e.g., [4–9]. The RIRs are modeled as realizations of nonstationary continuous-time random signals that decay according to the reverberation time $T_{60}$ of the room. These models allow for the derivation of recursive relations between the autocorrelation functions of the reverberant signal and the reverberation. With a short-term stationarity assumption about the speech signal, they transform to recursive relations for the spectral variances. The derivations are based on expected values of RIR autocorrelation functions, for fixed $T_{60}$ and Direct-to-Reverberation Ratio (DRR). This means that the derived LRSV estimators are correct when spatial averaging can be performed [9], but this requires multiple microphones. They are on average correct for single-microphone speech signals filtered with different, time-invariant, realizations of the RIRs.

In this paper, we consider time-varying RIRs. We use a discrete-time statistical RIR model where both the parameters that describe the global shape of the RIRs ($T_{60}$ and DRR) and the individual RIR filter tap values are allowed to change for each discrete time index. We show directly in the DFT domain that existing LRSV estimators can still be used under mild conditions on the correlations between RIR tap values at different time indices. This remains true even when the individual RIR filter taps vary rapidly with time, provided that the shape parameters $T_{60}$ and DRR remain nearly constant during an interval of the order of only a few frames. From our analysis it also follows that frequency-bin independent values of those parameters may be used for the LRSV estimators in the DFT domain.

This paper is organized as follows. In Section 2 a statistical time-varying RIR model is introduced and LRSV estimators are derived. Section 3 considers estimation of the shape parameters. Section 4 verifies the derivations experimentally and illustrates the performance of the estimators. The paper is concluded in Section 5.

2. SIGNAL MODELING AND LRSV ESTIMATORS

2.1. Time-domain modeling

We assume the observed noisy and reverberant speech signal $x$ to be the sum of a source speech signal $s$ convolved with a time-varying RIR $h$ and non-stationary additive noise $d$, independent of $s$:

$$x(n) = \sum_{l=0}^{\infty} h_n(l) s(n-l) + d(n) = z(n) + d(n),$$

where $n$ is the discrete-time sample index and $z$ is the noise-free reverberant signal. RIRs generally consists of a number of impulses for the early reflections and an exponentially decaying tail with a more noise-like appearance giving rise to the late reverberation. A simple RIR model is an i.i.d. Gaussian noise sequence with exponentially decaying variance [4]. This model is described by one parameter $T_{60}$. However, the LRSV estimator corresponding to this model may overestimate the LRSV in case of a DRR larger than 1. Habets [9] proposed a more sophisticated model described by $T_{60}$ and DRR. In theory, overestimation of the LRSV can be avoided when the DRR is estimated. For our analysis, we will use a slightly different RIR model, also described by two parameters, and, as we will see, closely related to Habets’s model. We model the direct path by a delta pulse and the reverberation with an i.i.d. noise sequence with exponentially decaying variance, as follows:

$$h_n(l) = \begin{cases} 1 & : l = 0 \\ r_n(l)e^{-\delta_n l} & : l \geq 1 \end{cases},$$

where $r_n(l)$ is a zero-mean i.i.d. Gaussian process with variance $\sigma_r^2(n) \leq 1$. In other words, the RIR is assumed to consist of a deterministic part that models the direct path, and a stochastic part that models the reverberation. The decay rate $\delta_n$ and $\sigma_r^2(n)$ are allowed to change slowly over time. Changes in source-microphone distance are modeled by changes in $\sigma_r^2$. The decay rate depends on $T_{60}$, and the DRR also on $\sigma_r^2$, as follows.

978-1-4244-4296-6/10/$25.00 ©2010 IEEE
\[ \delta_n = \frac{3 \ln(10)}{T_{60}(n) F_s}, \quad \text{DRR}(n) = e^{2\delta_n} - 1 \]

where \( F_s \) is the sampling frequency.

2.2. Spectral enhancement

Signal enhancement will be performed in the DFT domain. First, a standard noise suppression algorithm is applied to the noisy reverberant signal \( x \) in (1). The reverberant spectral variance \( \lambda_r(k,m) \) is now estimated by recursive smoothing, similarly as in [9]. The frequency bin index is \( k \), and \( m \) is the sample index at which \( R \) frame starts, that is, a multiple \( j R \) of the frame hop size \( R \) of the enhancement system. The estimated reverberant spectral variance is used to calculate the LRSV estimator (section 2.3) and the late reverberance is suppressed as if it were independent additive noise. For ease of notation we may drop in the following time and/or frequency indices when this does not cause confusion. Estimated quantities carry a hat.

2.3. LRSV estimation for time-varying RIRs

In Appendix A the following LRSV estimator is derived directly in the DFT domain for time-varying RIRs of the form (2):

\[ \lambda_r(k,m) \approx e^{-2\lambda_r L} \{ \lambda_s(k,m - L) - \lambda_s(k,m - L) \} \quad (4) \]

where \( \lambda_s \), \( \lambda_r \), and \( \lambda_s \) are the LRSV, the spectral variance of the reverberant speech, and the spectral variance of the direct-path signal, respectively. \( L \) is the interval after which the late reverberation is assumed to start. Since we assume the direct signal, the early reverberation, and the late reverberation to be uncorrelated, the variance of the reverberant signal is the sum of three contributions

\[ \lambda_r(k,m) = \lambda_s(k,m) + \lambda_s(k,m) + \lambda_r(k,m) \quad (5) \]

where \( \lambda_s \) is the spectral variance due to the early reverberation, i.e., due to the part of the RIR from \( l = 1 \) till \( l = L \). With (5), (4) becomes

\[ \lambda_r(k,m) \approx e^{-2\lambda_r L} \{ \lambda_s(k,m - L) + \lambda_r(k,m - L) \} \quad (6) \]

If we define \( \kappa_m = \lambda_s(k,m)/\{ \lambda_s(k,m) + \lambda_r(k,m) \} \), then an alternative estimator is

\[ \lambda_r(k,m) \approx e^{-2\lambda_r L} \{ \kappa_m \lambda_s k,m - L) + (1 - \kappa_m) \lambda_r(k,m - L) \} \quad (7) \]

If we assume that the speech signal is stationary during an interval of \( L + 1 \) samples and the RIR taps change little during a frame length [10], then \( \kappa \) will be a constant (depending only on \( T_{60} \) and the DRR):

\[ \kappa_m = \left( 1 + \frac{\text{DRR}(m)}{1 - e^{-2\lambda_r L}} \right)^{-1} \quad (8) \]

Note that we can then use the same \( T_{60} \) and \( \kappa \) values for all frequency bins. For more quickly changing RIRs, \( \kappa \) becomes frequency dependent [10], but \( T_{60} \) does not. Equation (7) is similar to the expression for the LRSV as derived in [9] from a time-invariant continuous-time statistical RIR model. This shows that the models in (2) and [9] are closely related. The estimator in [9] can also be derived directly from a discrete-time time-varying version of the latter model, using similar steps and assumptions as in Appendix A, but without having to assume stationarity of the speech signal. We will use (7) in the following, with a frequency-independent \( \kappa \).

In the derivations it has been assumed that the correlation function of the stochastic processes \( r_n(l) \) in (2) along the \( n \) variable changes only slowly over time. It is important to realize that this assumption (15) does not mean that the RIR is nearly time-invariant.

In fact, \( h_n(l) \) could vary extremely fast as a function of \( n \). For example, suppose it varies so fast that we would have to model it by a completely different realization of the stochastic process \( r_n(l) \) for each \( n \). Then the correlation between \( h_n(l) \) and \( h_{n'}(l) \) would be 0 (for \( l > 0, n \neq n' \)), but constant, and hence (4)-(7) would be valid even though \( h_n(l) \) itself varies very fast in time. This is experimentally verified in Section 4.

3. ESTIMATION OF THE SHAPE PARAMETERS

The LRSV estimator (7) requires blind estimation of \( T_{60} \) and \( \kappa \). We estimate \( T_{60} \) from the negative-side variance of the slopes of the log-energy envelope of the (denoised) time domain signal frames, as proposed in [11]. An approximate ML method may also be used [12]. We use here frames of length \( N = 256 \) samples, a resolution of \( N/8 \), and 24 consecutive log-energy values to estimate the slopes. The sampling frequency was 8 kHz, so the slopes are calculated from an interval of about 100 ms. The negative-side variance is updated using recursive smoothing with a smoothing parameter equal to 0.998. To reduce the variance somewhat, values of the slopes corresponding to reverberation times smaller than 100 ms are discarded. Frames with an average SNR below \(-10\) dB (estimated in the prior noise suppression step) are also not taken into account. We used a data-driven method to map the resulting negative-side variance estimates to the decay rate. The training data was obtained by filtering one directory of TIMIT-TRAIN with RIRs generated by the model (2) for 20 values of \( T_{60} \) between 0.1 and 3 seconds after which noise suppression is applied. We found that the resulting mapping function is almost independent of \( \sigma^2 \) except for very low values where the decay rate of the anechoic speech becomes dominant. Therefore we used one fixed value \( \sigma^2 = 1 \) in the training.

Estimation of \( \kappa \) is more difficult than that of \( T_{60} \). For this reason, some authors simply use Lebart’s estimator [4, 7] (that is, (7) with \( \kappa \) fixed to 1). An adaptive estimator for \( \kappa \) has been proposed in [9] that tends to bring the LRSV estimates slowly towards the reverberant spectral power \( |Z(k,m)|^2 \). This lowers the value of \( \hat{\kappa} \) when \( \hat{\lambda}_r(k,m) \) exceeds \( |Z(k,m)|^2 \). However, during periods of speaker activity the true LRSV often is smaller than \( |Z|^2 \), so increasing \( \hat{\kappa} \) then would result in overestimation. To avoid this, we propose to make use of Lebart’s estimator to detect whether \( \hat{\kappa} \) is too small. Let \( \kappa_m \) be the value of \( \hat{\kappa} \) that would make \( \hat{\lambda}_r(L) \) equal to \( \hat{\lambda}_r(L) \), where the prime means that averaging over all frequency bins has been performed. We update \( \hat{\kappa} \) only when we detect that its value is either too large or too small, as follows

\[ \hat{\kappa}_{m+R} = \eta \hat{\kappa}_m + (1 - \eta) \min[\max[\kappa_m], 0], 1] \]

where \( R \) is the frame hop size and \( \eta = 0.95 \) is the smoothing parameter. The value of \( \hat{\kappa}_m \) is deemed too large when \( \hat{\lambda}_r(L) > \hat{\lambda}_r(L) \), while it is considered too small when \( \hat{\lambda}_r(L) < \hat{\lambda}_r(L) \) and \( \hat{\lambda}_r(L) > \hat{\lambda}_r(L) \) (where \( \hat{\lambda}_r(L) \) is Lebart’s estimator, averaged over all frequency bins). The value of \( \kappa_m \) is given by

\[ \kappa_m = \frac{e^{2\lambda_r L} \hat{\lambda}_r(L) - \hat{\lambda}_r(L)}{\hat{\lambda}_r(L) - \hat{\lambda}_r(L)} \]

After updating \( \hat{\kappa} \), the LRSV estimators \( \hat{\lambda}_r(L) \) and \( \hat{\lambda}_r(L) \) are limited to values smaller than \( \lambda_r(L) \). Note that the estimation of \( \kappa \) is influenced by errors in the estimated reverberation time and \( \hat{\lambda}_r \). For example, when \( T_{60} \) is underestimated, \( e^{-2\lambda_r L} \) is also too small, resulting in underestimation of the LRSV. The adaptation (9) may compensate somewhat by increasing \( \hat{\kappa} \).
4. EXPERIMENTAL RESULTS

In this section we first perform an experiment with simulated time-varying RIRs that corroborates the analysis of Appendix A. Experiments with measured time-invariant RIRs illustrate the performance of the estimators under noisy conditions.

Experimental set-up

We used 3 minutes of telephone-bandwidth speech (Fs = 8 kHz) from TIMIT, without intervening pauses. RIRs were either simulated with the model (2) or downsampled versions of the measured RIRs from the AIR database [13]. The gain function for spectral amplitude estimation assumes a generalized Gamma speech prior with parameters γ = 1 and ν = 1 [14]. The noise signals were taken from the NTN monaural noise database. The noise variance is tracked with the method in [15]. The enhancement system uses a frame length N of 256 samples, and 50% overlap between frames (R = N/2). We set L = N in the LRSV estimator (7).

In the first experiment we validate the theoretical derivations in Appendix A that predict that (7) is a good estimator also in case of time-varying RIRs. We will compare the performance on time-invariant and time-varying RIRs with the same T60 and DRR. The RIRs were generated with the model (2). The time-invariant RIR was 1 realization, while for the time-varying RIR we used a different realization of the stochastic process \( r_n(l) \) for every \( n \). T60 and the DRR were kept constant. We used the theoretical values of \( T60 \) and \( \kappa \) (8) in the LRSV estimator (7). No initial noise suppression step was performed. We computed the segmented logarithmic estimation error (\( \text{LogErr} \), see, e.g., [15]) between the LRSV estimators and an estimate of the true LRSV that serves as a reference. All estimators are averaged over all frequency bins before computing \( \text{LogErr} \).

Table 1 shows the results. The reference LRSV was obtained by temporally smoothing the late reverberant spectra. The late reverberant signal results from filtering the anechoic speech with just the tail of the RIR. The reference was computed from time-invariant RIRs. To reduce its variance, it was averaged over 20 realizations of a RIR.

Table 1. \( \text{LogErr} \) [dB] for LRSV estimation for time-invariant (TI) and time-varying (TV) RIRs with various \( T60 \) and DRR values.

<table>
<thead>
<tr>
<th>( T60 ) [s]</th>
<th>DRR= -10 dB</th>
<th>DRR= 0 dB</th>
<th>DRR= 5 dB</th>
</tr>
</thead>
<tbody>
<tr>
<td>TI</td>
<td>TV</td>
<td>TI</td>
<td>TV</td>
</tr>
<tr>
<td>0.2</td>
<td>0.09</td>
<td>0.27</td>
<td>0.09</td>
</tr>
<tr>
<td>0.4</td>
<td>1.02</td>
<td>0.25</td>
<td>1.11</td>
</tr>
<tr>
<td>0.8</td>
<td>0.87</td>
<td>0.19</td>
<td>0.58</td>
</tr>
<tr>
<td>1.6</td>
<td>0.64</td>
<td>0.20</td>
<td>0.47</td>
</tr>
</tbody>
</table>

We see that (7) performs well on both time-invariant and time-varying RIRs. It actually performs best on the time-varying RIRs. The difference probably lies in the estimation of \( \lambda_z \). For the time-varying RIRs, there is much less correlation in the spectra of the late reverberant signal than in the spectra of the early reverberant signal. \( \text{LogErr} \) decreases with increasing \( T60 \) probably because (7) provides more smoothing of the \( \lambda_z \) for larger \( T60 \).

The next experiment illustrates the performance of LRSV estimators on speech filtered with measured RIRs from the AIRD database [13]. The results are shown in Tables 2 and 3. Performance figures for modeled RIRs are similar [10]. We compare (7) with \( T60 \) and \( \kappa \) estimated blindly (denoted by \( \lambda_{\hat{z}} \)) with Lebart’s estimator with \( T60 \) estimated blindly (denoted by \( \lambda_{\hat{1}} \)). As references we also include (7) using the \( T60 \) and \( \kappa \) values estimated directly from the RIR (denoted by \( \lambda_{\hat{w}} \)), and the method in [8]. That method is a completely blind one that only makes use of long-term correlations in the DFT domain induced by the reverberation. It does not need any explicit estimates of \( T60 \) or \( \kappa \).

Table 2. \( \text{LogErr} \) [dB] values for LRSV estimators for RIRs from [13] under different noisy conditions. \( \lambda_{\hat{w}} \) is using the \( T60 \) and DRR values estimated directly from the RIRs, [8] is a blind correlation-based method, \( \lambda_{\hat{1}} \) is Lebart’s estimator with \( T60 \) estimated blindly, and \( \lambda_{\hat{2}} \) is (7) with \( T60 \) and \( \kappa \) estimated blindly.

<table>
<thead>
<tr>
<th>( T60 ) [s]</th>
<th>DRR</th>
<th>( \kappa )</th>
<th>( 20 \text{ dB SNR Car noise})</th>
<th>( 10 \text{ dB SNR Mall noise})</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \lambda_{\hat{w}} )</td>
<td>( \lambda_{\hat{1}} )</td>
<td>( \lambda_{\hat{2}} )</td>
<td>( \lambda_{\hat{1}} )</td>
<td>( \lambda_{\hat{2}} )</td>
</tr>
<tr>
<td>0.43</td>
<td>2.1</td>
<td>1.24</td>
<td>1.39</td>
<td>1.32</td>
</tr>
<tr>
<td>0.78</td>
<td>2.0</td>
<td>1.22</td>
<td>1.36</td>
<td>2.44</td>
</tr>
</tbody>
</table>

Table 3. \( \text{SegSIR} \) [dB] values for the LRSV estimators under the experimental conditions of Table 2.

<table>
<thead>
<tr>
<th>( T60 ) [s]</th>
<th>DRR</th>
<th>( \kappa )</th>
<th>( 20 \text{ dB SNR Car noise})</th>
<th>( 10 \text{ dB SNR Mall noise})</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \lambda_{\hat{w}} )</td>
<td>( \lambda_{\hat{1}} )</td>
<td>( \lambda_{\hat{2}} )</td>
<td>( \lambda_{\hat{1}} )</td>
<td>( \lambda_{\hat{2}} )</td>
</tr>
<tr>
<td>0.43</td>
<td>3.1</td>
<td>2.34</td>
<td>2.82</td>
<td>2.80</td>
</tr>
<tr>
<td>0.78</td>
<td>2.0</td>
<td>2.22</td>
<td>4.04</td>
<td>3.39</td>
</tr>
</tbody>
</table>

For the computation of \( \text{LogErr} \), we have to correct the reference for the effects of the initial noise suppression step. We multiplied the spectral amplitudes of the late reverberant signal by the gain values used for reverberant spectral amplitude estimation in the noise suppression step, before computing the reference LRSV. Since we can’t average here over more realizations of the RIRs, only temporal smoothing was applied. As a second performance measure, improvements in segmental signal-to-interference ratio (\( \text{SegSIR} \)) after reverberation suppression are listed in Table 3.

Table 2 shows that the correlation-based method [8] achieves lower \( \text{LogErr} \) values for the highest DRRs than the model-based estimators \( \lambda_{\hat{1}} \) and \( \lambda_{\hat{2}} \). It even outperforms \( \lambda_{\hat{w}} \) for \( T60=0.43 \) s / DRR=3.1 dB. This is possible because \( \lambda_{\hat{w}} \) uses estimated values of \( \lambda_z \) after noise suppression. The performance of \( \lambda_{\hat{w}} \) (and the other model-based estimators) is therefore limited by the statistics (bias and variance) of the \( \lambda_z \) estimator as well. As expected, Lebart’s estimator does not perform very well for positive DRR values. Estimating \( \kappa \) leads to some improvements, but we found that it is difficult to adapt to low \( \kappa \) values. The model-based estimators achieve somewhat higher \( \text{SegSIR} \) values than the correlation-based method (Table 3), but this is at least partly because they tend to overestimate the LRSV and therefore cause more suppression. Nevertheless, we recommend model-based LRSV estimators for the following reason: although the correlation-based method performs well for time-invariant RIRs, it is sensitive to changes in the individual RIR filter taps, and may therefore fail to suppress the reverberation in case of time-varying RIRs. The model-based estimators don’t have this problem.

5. CONCLUSION

We have analyzed time-varying RIRs and shown that existing LRSV estimators may be applied also in this case. Our experiments show the importance of accurate estimation of the reverberant spectral variance. Estimation of the DRR deserves more attention, also because it may vary quickly during speaker movements.

6. REFERENCES


**A. DERIVATION OF AN LRSV ESTIMATOR FOR TIME-VARYING ROOM IMPULSE RESPONSES**

Here we outline the main steps in the derivation of (4). We start from the following expression of the reverberant signal in the time domain using the RIR model (2)

\[
Z(n) = \sum_{l=-\infty}^{n} h_n(l) s(l) = s(n) + \sum_{l=-\infty}^{n-1} h_n(l) s(l). \tag{11}
\]

The DFT of one frame of \( z \) is given by

\[
Z(k, m) = \sum_{n=0}^{N-1} w(n) z(m+n) e^{-2 \pi i kn / N}, \tag{12}
\]

with \( w(m) \) the analysis window.

The reverberant spectral variance \( \lambda_s(k, m) \) is obtained by taking the expectation of \( |Z(k, m)|^2 \) over the source signal \( s \) and the random processes \( r \) in (2), i.e.,

\[
\lambda_s(k, m) = E_r E_r |Z(k, m)|^2;
\]

\[
\lambda_s(k, m) = \sum_{n=0}^{N-1} \sum_{l=-\infty}^{\infty} h_n(q-l) h'_q(q'-l') s(l)s(l') \tag{13}
\]

where \( q = m+n, q' = m+n' \), and \( \lambda_s(k, m) = e^{-2 \pi i kn / N} \). As in [4] and [9], we assume the processes \( s \) and \( r \) to be independent. From (11)–(13) we can see that the direct path contribution to \( \lambda_s(k, m) \) equals \( \lambda_s(k, m) \). From the assumptions made about the RIR in (2), we have, for \( l > 0 \):

\[
E_r \{ r_q(l) r_q(l') \} = \begin{cases} \sigma_q^2(q) & q = q' \& l = l' \\ 0 & l \neq l' \end{cases} \tag{14}
\]

We make one additional assumption about the correlations across time:

\[
E_r \{ r_q(l) r_q(q-l) \} = E_r \{ r_q(l) r_q(q-l) \}, \quad |q-q'| \leq N, \tag{15}
\]

This assumption will later allow us to relate the reverberant spectral variance, at time \( m - L \) to the LRSV at time \( m \). Note that (14) and (15) include the assumption \( \sigma^2_q(m) \approx \sigma^2_q(m + L) \). With (14), the double summation over \( l \) and \( l' \) in (13) reduces to the following single summation when the expected values are taken

\[
\sum_{l=-\infty}^{\infty} E_r \{ h_n(q-l) h'_q(q-l) \} R_{ss}(l, n' - n + l), \tag{16}
\]

where \( R_{ss} \) is the time-domain autocorrelation function of \( s \). As mentioned before, the term with \( l = q = m + n \) leads to the direct path contribution \( \lambda_s(k, m) \). The rest of the terms can be split up into two summations: one ranging from \( l = m + n - L \) to \( l = m + n + 1 \) and the other from \( l = -\infty \) to \( l = m + n - L - 1 \), leading to the early reverberance spectral variance \( \lambda_e(k, m) \) and the LRSV \( \lambda_r(k, m) \) contributions, respectively. We will make the assumption that the shape parameter \( \delta \) remains approximately constant during a time interval \( [m - L, m + N] \). Then

\[
E_r \{ h_n(q-l) h'_q(q-l) \} \approx e^{-2 \pi i q^2 / \delta} E_r \{ r_q(l) r_q(q-l) \} \tag{17}
\]

Now, as was done in [4, 9], we will evaluate \( \lambda_e(k, m - L) \). The summation in (16) then runs from \( l = -\infty \) to \( l = m + n - L \) and the term with \( l = m + n - L \) leads to \( \lambda_e(k, m - L) \). The other terms have the same values of \( l \) as in the summation leading to \( \lambda_s(k, m) \). This allows us to relate \( \lambda_s(k, m - L) \) to \( \lambda_e(k, m) \) as follows. The right hand side of (17) now becomes

\[
e^{-2 \pi i q^2 / \delta} e^{-2 \pi i L / \delta} E_r \{ r_q(l) r_q(q-l) \} \tag{18}
\]

Invoking (15) and assuming \( \delta_m \approx \delta_m - L \), we see that (18) is approximately equal to \( e^{2 \pi i L / \delta_m} \) times (17). Putting everything together, we can relate \( \lambda_s(k, m - L) \) to \( \lambda_e(k, m) \):

\[
\lambda_s(k, m - L) \approx \lambda_e(k, m - L) + e^{2 \pi i L / \delta_m} \lambda_e(k, m), \tag{19}
\]

from which (4) follows directly.