Quality Evaluation of Reverberation in Audioband Speech Signals

Amaro A. de Lima†, Sergio L. Netto†, Luiz W. P. Biscainho†,
Fabio P. Freeland†, Bruno C. Bispo†*, Rafael A. de Jesus†**,
Ronald Schafer‡, Amir Said‡, Bowon Lee‡, and Ton Kalker‡

†PEE/COPPE, Federal University of Rio de Janeiro - Rio de Janeiro, Brazil,
‡HP Labs - Palo Alto, USA
{amaro,sergioln,wagner,freeland,brunobispo,rjesus}@lps.ufrj.br
{ron.schafer,amir_said,bowon.lee,ton.kalker}@hp.com

Abstract. Modern telepresence systems constitute a new challenge for quality assessment of multimedia signals. This paper focuses on the evaluation of the reverberation impairment for audioband speech signals. A review on the reverberation effect is presented, with emphasis given on the mathematical modeling of its components, including early reflections and late reverberation. A subjective test for evaluating the human perception of the reverberation phenomenon is completely described, from its conception to the final results. Analyses are provided comparing the average subjective grades to current quality-evaluation standards for speech and audio signals. It is verified how the reverberation perception can be mapped onto three main system characteristics: reverberation time (associated to the room acoustical properties), source-microphone distance, and room volume. Direct estimation of these parameters from the room impulse response is discussed. One established reverberation measure is then revisited in the audioband speech context, showing high correlation with the subjective grades previously obtained.

Key words: acoustic signal processing; speech communication; speech quality evaluation; reverberation modeling.

1 Introduction

In recent years, teleconference systems have evolved to telepresence systems, which provide high-quality video and speech signals, enabling a more realistic meeting experience. This superior level of service is commonly accomplished through a dedicated data network, which delivers the required high data rates for up to three HDTV signals and audioband (up to 24 kHz) speech signals. To ensure user satisfaction at high levels, the performance of the telepresence system is continuously supervised by quality monitors.

According to [1], the three classes of impairments that historically dominated the transmission quality of speech in telecommunications were: loudness loss (reduction in

---

* Presently, with Dept. of Electrical and Computer Eng., University of Porto, Portugal.

** Presently, with Petrobras (Brazilian petroleum holding), Brazil.

signal strength); noise (from the circuit itself and other interfering sources); and echo (where the talker experiences his/her voice after a given transmission delay).

After the introduction of digital technology, the common noise sources were almost eliminated and the importance of loudness loss was reduced, since noise removal allowed effortless signal amplification. More recently, teleconference/telepresence systems introduced other impairments that demand improvements in the models used to represent signal degradation within the system. In this scenario, the main impairments currently considered are: background noise (possibly generated by an air-conditioner, a computer, or any other source in room A or B); echo (signal from room A returning to this room through speaker-microphone coupling in room B); and reverberation (acoustical properties of rooms A or B being imposed to the signal that leaves or enters the rooms). Among these three impairments, reverberation is the most intricate one, thus deserving a deeper analysis description as proposed in this work. Other challenges to be considered come from the increase in the allowed speech bandwidth, which demands new objective methods for quality evaluation, since current state-of-the-art considers only 8- and 16-kHz speech signals [2, 3]

This work deals with the issue of evaluating reverberation levels in audioband speech signals in a telepresence context. The paper is organized as follows: in Section 2, the reverberation impairment is revisited in the light of the application at hand; in Section 3, classic reverberation models are discussed and a slight modification of Gardner's full reverberator is proposed; in Section 4, subjective tests for evaluating the human perception of the reverberation effect in audioband speech signals are described; Section 5 focuses objective reverberation assessment based on standard speech evaluation tools and the so-called articulation loss of consonant (ALcons) measure [4], extending the result of [5]; finally, Section 6 concludes the paper emphasizing its main contributions.

2 Reverberation Definition

Reverberation corresponds to the modification of a signal by the acoustic response of the enclosure in which the signal source is placed. Excessive reverberation reduces the intelligibility of speech and degrades the performance of acoustic echo cancelers in case of hands-free communication—as mentioned in ITU-T Software Tool Library [6], which includes some reverberation models.

A room impulse response (RIR) is usually modeled as a finite-duration impulse response (FIR) that can be measured between the location of a specific source and that of the receiver. Thus, it is possible to imprint room-related acoustical modifications to a possibly anechoic source $s(k)$ by its convolution with a room impulse response $\mathrm{RIR}(k)$, i.e.,

$$s_{\mathrm{rev}}(k) = \sum_{l=0}^{N-1} \mathrm{RIR}(l)s(k-l), \tag{1}$$

where $s_{\mathrm{rev}}(k)$ is the resulting reverberated signal and N is the length of $\mathrm{RIR}(k)$. In order to compare the original and modified signals, their respective powers should be matched by an adequate level scaling.

The reverberation effect associated with a given RIR can be divided into two distinct sections: early reflections and late reverberation, as seen in Figure 1. The early reflections, which can be directly linked to the room geometry, correspond to the first 80∼100 ms and are commonly modeled by FIR filters. The late reverberation dominates over the early reflections after a certain transition point and extends itself to the end of the RIR duration. This final part of the RIR is nearly diffuse, which means that the magnitude and direction of the sound pressure can be considered as randomly distributed. Moreover, its magnitude typically decays exponentially over time. In practice, the late reverberation is commonly modeled by infinite-duration impulse response (IIR) filters.

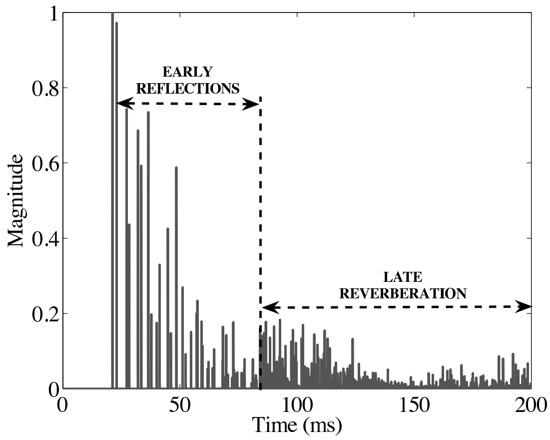

Fig. 1. Artificial RIR depicting the early and late reverberation sections.

An important parameter associated with the reverberation effect is the so-called reverberation time T_{60}, defined as the time that it takes for the sound-pressure-level of a steady-state excitation within a room to decrease 60 dB after being abruptly stopped. The reverberation time T_{60} for a given room can be estimated from Sabine's formula [7]:

$$T_{60} = 0.161(\mathrm{Vol}/S_{\mathrm{e}}), \tag{2}$$

where Vol is the room volume and S_{e} is the effective room area, determined as

$$S_{\mathrm{e}} = \sum_{i=1}^{L} (1 - r_i) S_i, \tag{3}$$

where S_i and r_i are the area and reflection coefficient, respectively, for each of the $i = 1, 2, \ldots, L$ room walls.

3 Reverberation Modeling

3.1 Early Reflections via Image Method

The early reflections are characterized by exponentially decaying echoes in the RIR, with associated gain and delay values. These parameters are commonly obtained using the image method [8], which consists of representing the room as a shoebox with the microphone and sound source at given predefined positions, as depicted in Figure 2. The walls then act like acoustic mirrors and produce several layers of virtual sources as reflections of both the original and the virtual sound sources. The delays are proportional to the distance of the sources (real and virtual) to the microphone. The gains depend on these same distances and on the reflection coefficients of all walls crossed by each signal path.

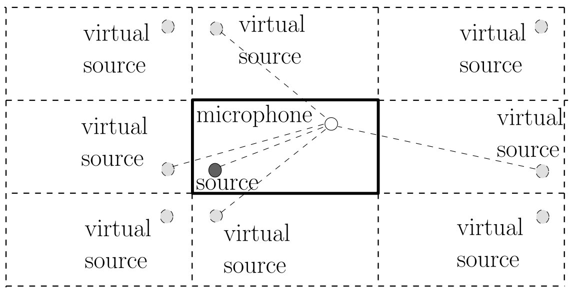

Fig. 2. Image method concept.

3.2 Schroeder's Late Reverberator

The first model for artificially producing late reverberation was presented in [9]. The basic units to implement the Schroeder's late reverberators are the comb and all-pass comb filters, whose diagrams are shown in Figure 3.

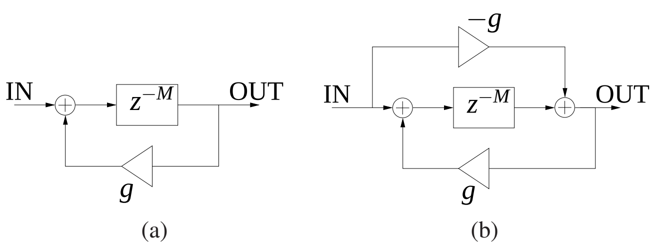

Fig. 3. Basic units for Schroeder's late reverberator: (a) Comb filter; (b) All-pass comb filter.

The transfer function of the comb filter is given by

$$H_c(z) = \frac{z^{-M}}{1 - gz^{-M}}, \tag{4}$$

where $g \leq 1$ is the feedback gain and M is the delay length in samples. The associated impulse response is an exponentially decaying sequence of impulses spaced M samples apart. The frequency response is shaped like a comb, with M periodic peaks that correspond to the pole frequencies. The relation among the comb filter parameters and the reverberation time is given by

$$\frac{20\log_{10}(g)}{MT_s} = \frac{-60}{T_{60}}, \tag{5}$$

where T_s is the sampling period. The parallel combination of comb filters with pure delays results in a frequency response that contains the peaks contributed by all individual comb filters, being the resulting echo density given by the sum of the individual densities.

The transfer function of the all-pass comb filter is given by

$$H_{ac}(z) = \frac{z^{-M} - g}{1 - gz^{-M}}. \tag{6}$$

Short delays M correspond to widely spaced frequency peaks, which yield an unpleasant characteristic timbre. By increasing the delay length, a higher peak density ensues, but at the cost of a decrease in echo density in the time domain. Such decrease will be perceived as a discrete set of echoes, rather than a smooth diffuse decay. The series combination of all-pass comb filters helps increasing echo density without affecting the magnitude response of the system.

The complete Schroeder's reverberator uses a parallel combination of comb filters cascaded with a serial combination of all-pass comb filters. Practical implementation guidelines for this device can be found in [10].

3.3 Feedback Delay Networks (FDN)

The feedback delay networks [11] constitute a generalization of unitary multichannel networks, which are N-dimensional counterparts of the all-pass comb filter, where N is the number of delay lines in the FDN diagram given in Figure 4. In this structure, the a_{ij} coefficients control the feedback level for the output of the jth delay to the input of the ith delay. This structure can generate much higher echo densities than the parallel comb filters, given a sufficient number of non-zero feedback coefficients and pure delay lengths. The choice of the delays is made according to Schroeder's suggestion [9].

In the FDN structure, the lowpass filters

$$H_i(z) = k_i \frac{1 - b_i}{1 - b_i z^{-1}}, \tag{7}$$

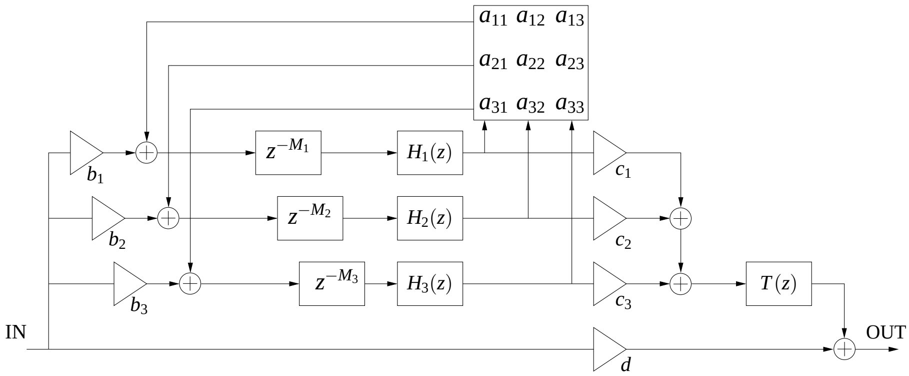

Fig. 4. Feedback delay networks ($N = 3$) for modeling the late reverberation.

where $b_i = 1 - 2/\left(1 + k_i^{(1-1/\varepsilon)}\right)$, with $\varepsilon = T_{60}(\pi)/T_{60}(0)$ and $k_i = 10^{-3M_iT_s/T_{60}(0)}$, keep the reverberation time of the low frequency components ($T_{60}(0)$) larger than that of the higher frequency components ($T_{60}(\pi)$). The tone-corrector filter

$$T(z) = g_t \frac{1 - b_t z^{-1}}{1 - b_t}, \tag{8}$$

where $g_t = \sqrt{(T_s/T_{60}(0)\sum M_i)}$ and $b_t = (1 - \sqrt{\varepsilon})/(1 + \sqrt{\varepsilon})$, compensates for the distortion in the frequency response envelope introduced by the $H_i(z)$ filters.

3.4 Gardner's Late Reverberator

Gardner's late reverberation model [7] considers three different schemes according to the reverberation time ranges, as indicated in Table 1.

Table 1. Reverberation time ranges × reverberator model.

Reverberation time (s)	**Reverberator model**
$0.38 \leq T_{60} \leq 0.57$	Small-room model
$0.58 \leq T_{60} \leq 1.29$	Medium-room model
$1.30 \leq T_{60} < \infty$	Large-room model

Gardner's model uses the idea of nested all-pass filters, which are based directly on the all-pass comb filter represented in Figure 3(b). The main idea consists of replacing the z^{-M} delay in the all-pass comb filter by one or more all-pass filters. For instance, in the single nested case, shown in Figure 5, the delay unit is replaced by an all-pass filter with gain g_{in} and delay $z^{-M_{\text{in}}}$, followed by a delay $z^{-(M_{\text{out}}-M_{\text{in}})}$, such that $M_{\text{out}} > M_{\text{in}}$. In the double nested structure, the usual delay is replaced by two all-pass filters in series

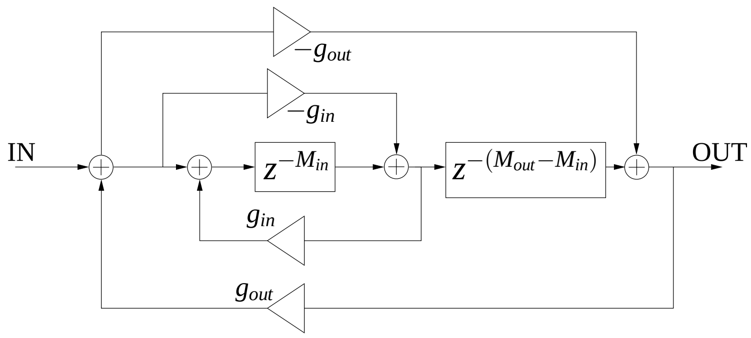

Fig. 5. Single nested all-pass filter for Gardner's late reverberator.

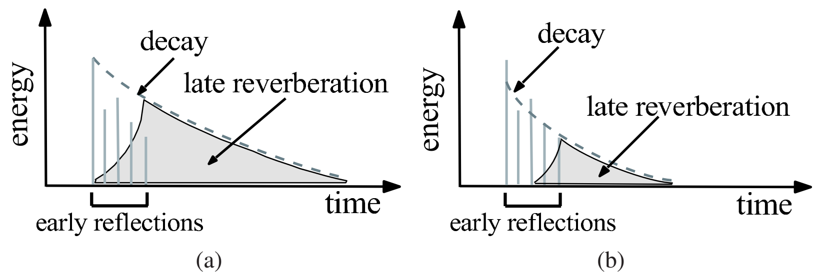

Fig. 6. Combining early reflections with late reverberation: (a) Gardner's approach; (b) Modified approach.

with gains $g_{\text{in}1}, g_{\text{in}2}$ and delays $z^{-M_{\text{in}1}}, z^{-M_{\text{in}2}}$, followed by a delay $z^{-(M_{\text{out}}-M_{\text{in}1}-M_{\text{in}2})}$, such that $M_{\text{out}} > (M_{\text{in}1} + M_{\text{in}2})$. The single and double nested all-pass figures are used to implement Gardner's late reverberation models included in [7] for the three distinct room types included in Table 1.

3.5 Gardner's Modified Full Reverberator

Figure 6(a) shows the approach proposed in [7] to combine the early reflections with the late reverberation portions of a particular RIR. In this method, an exponential decay corresponding to the desired reverberation time is adjusted to the peak energy of the first reflection and serves as a reference to the late reverberation, whose peak follows immediately after the end of the early reflections. This may cause an energy increase in the associated RIR right at the junction of its two portions, as indicated in Figure 6(a), which is uncommon in real RIRs.

Therefore, a modification is suggested in order to achieve a smoother RIR, as illustrated in Figure 6(b). The modified scheme equalizes the energy of the two RIR portions at their junction, while still forcing the late reverberation to fit under the exponential decay determined by the desired T_{60}. This alternative approach yielded a more natural perception of the reverberation behavior as judged by informal listening tests.

4 Subjective Evaluation of Reverberation

4.1 Test Specification

The reverberation effect in audioband high-quality speech signals was assessed subjectively.

Each original signal was recorded in a professional studio, digitally sampled at 48 kHz, using 24 bits per sample, and normalized to -26 dBov. Four different speakers (two female and two male) participated in the recording process, which was performed in a small room, acoustically treated, using different distances to the microphone: 10 cm for female-1, 30 cm for male-1 and male-2, and 100 cm for female-2. For these source-receiver locations the T_{60} of the room, estimated according to [12], was in the range 150$\sim$280 ms.

Each evaluated signal consisted of 2 Brazilian-Portuguese sentences, of duration between 2 and 3 seconds each, uttered by the same speaker. Sentences were separated in time by a mute interval of 500 ms and were preceded and succeeded by about 200$\sim$300 ms of silence. Before concatenation, the original recordings were convolved with a set of artificial RIRs that produce the reverberation effect. Considering the application at hand, such responses was chosen to yield an appropriate range of reverberation times: $T_{60} \in \{200, 300, 400, 500, 600, 700\}$ ms.

For all those RIRs, the early reflections were obtained via the image method, using a room of dimensions 4m-length, 3m-width, and 3m-height, assuming the distance source-microphone as $d = 1.8$ m. As regards the late reverberation, the FDN method was used to simulate $T_{60} = \{200, 300, 400\}$ ms, since informal listening tests indicated that this method lose naturalness above this range. The modified version of Gardner's method, which is conceived to simulate higher reverberation times, was used for $T_{60} = \{500, 600, 700\}$ ms. For each RIR, the corresponding T_{60} was evaluated using Schroeder's method [12], resulting in the predicted/obtained correspondence shown in Table 2.

Table 2. Desired and estimated reverberation times for the artificial RIRs used in the experiments.

Desired T_{60} (ms)	Estimated T_{60} (ms)
200	196
300	292
400	387
500	469
600	574
700	664

The subjective test comprised signals from the 4 speakers, with 3 repetitions of each desired T_{60} from the set $\{196, 292, 387, 469, 574, 664\}$. Additional 8 original signals were included in the test to serve as benchmark material, reaching a total of $[(4 \times 3 \times 6) + 8] = 80$ signals.

4.2 Test Results

A total of 26 listeners judged the overall "quality" of each signal using a grade range $1 \leq G \leq 5$, with 0.1 resolution. The mean opinion scores (MOS) for each T_{60} are shown in Figure 7, where one can notice that:

– Anchor refers to the original unprocessed signals, which in fact are non-anechoic;
– For the speech stimuli filtered by the artificial RIRs defined in Section 4.1 (see Table 2), longer reverberation times evoked lower MOS;
– The MOS attributed to the anchors is similar to that of the signals filtered through the RIR with $T_{60} = 196$ ms. This seems to indicate that the natural reverberation in the original speech stimuli sets an upper bound to MOS that is below the value that would be achievable if anechoic stimuli had otherwise been employed.

As expected, the importance of the T_{60} parameter in characterizing reverberation is clearly demonstrated by the relationship between MOS and T_{60} shown in Figure 7. Nevertheless, one shall remember that T_{60} is not the sole determinant of the final subjective score [13].

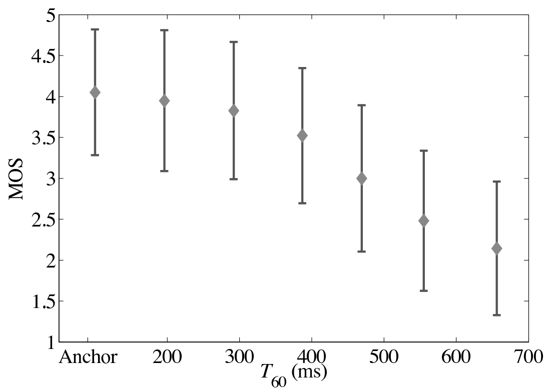

Fig. 7. MOS for each T_{60} (with confidence interval of one standard deviation).

5 Objective Evaluation of Reverberation

5.1 Standard Quality Assessment Methods

The subjective scores for all sentences, as mentioned in Section 4.2, were statistically correlated to the corresponding results from several objective quality-evaluation methods [14], such as PESQ [15], wideband-PESQ [16], P.563 [17], advanced and basic PEAQ [18], and Rnonlin [19]. The results are summarized in Table 3. It must be noted

Table 3. Statistical correlation ρ between subjective (MOS) and objective grades.

Objective method	ρ
PESQ	0.84
W-PESQ	0.79
P.563	0.45
PEAQ Basic	0.49
PEAQ Adv.	0.23
Rnonlin	0.83
PEMO-Q	0.55

that, except for PEAQ and Rnonlin, the aforementioned methods are intended for quality assessment of speech-band (4 or 8 kHz) signals. Therefore, comparisons with audioband subjective tests should be carefully made.

The relatively weak correlations in this table are somewhat expected, since none of the objective methods was designed to evaluate speech impairment by reverberation. Yet, the almost 85% correlation between MOS and PESQ measures is surprisingly good.

In addition, if one refers to each of the PEAQ internal model output variables (MOVs), the statistical correlation in some cases is also near the 85% mark, as indicated in Table 4.

Table 4. Statistical correlation ρ between MOS and PEAQ (basic and advanced) internal MOVs.

PEAQ Basic MOV	ρ	**PEAQ Adv. MOV**	ρ
1	0.02	1	−0.82
2	0.02	2	−0.73
3	−0.51	3	−0.50
4	−0.79	4	0.46
5	−0.61	5	−0.65
6	0.46	-	-
7	−0.75	-	-
8	−0.84	-	-
9	−0.71	-	-
10	−0.36	-	-
11	−0.55	-	-

The MOS attributed to each test sentence, organized by increasing order, is shown in Figure 8. Furthermore, it includes the corresponding objective grades yielded by the PESQ method, since they strongly correlate with MOS, as indicated in Table 3.

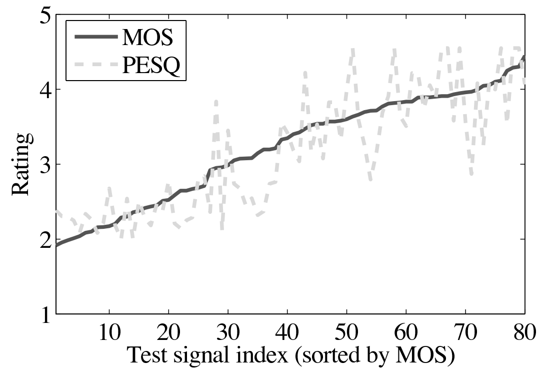

Fig. 8. MOS and PESQ grades for all test signals.

5.2 Articulation Loss of Consonants

Based on several experimental tests, Peutz [4] observed that increasing the source-microphone distance, d, decreased speech intelligibility, up to a critical distance, d_c, after which perception became constant. These experiments lead to the following measure closely related to the reverberation effect:

$$\text{ALcons}(\%) = \begin{cases} \left(\dfrac{200d^2T_{60}^2}{\text{Vol}} + zc\right), & \text{for } d \leq d_c \\ (9T_{60} + zc), & \text{for } d > d_c \end{cases}, \tag{9}$$

where zc denotes the correction value for the 0% score.

For the application at hand, which is high-quality teleconference systems, estimating the source-microphone distance from a RIR is not an involved job in an artificial scenario: the parameter d can be estimated by dividing the delay time t_d of the RIR peak level (which corresponds to the direct signal) by the sound speed. For the test signals described in Subsection 4.1 (emulating $d = 1.8$ m), this simple methodology yields an average estimate of $\mu_d = 1.84$ m with a relative standard deviation of $\frac{\sigma_d}{\mu_d} = 0.002$, which indicates a very accurate estimate of d in the low-noise scenario of interest.

As mentioned in Section 2, the room geometry is directly related to the number and energy of all sound reflections that compose the reverberation effect. In the ALcons measure, all aspects of the room geometry are synthesized by its volume, Vol, which, if not known *a priori*, can also be estimated from the RIR, as described below:

Algorithm V1: Assuming a spherical room geometry, one may write that [20]

$$\text{Vol} = \frac{4\pi}{3N}(cT)^3, \tag{10}$$

where c is the sound speed and N is the number of sound reflections within the RIR over a time interval T.

Algorithm V2: A more elaborate estimate for the room volume is given by [21]

$$\mathrm{Vol} = \frac{E_d}{E_r}\left(\frac{4\pi d^2 c T_{60}}{6\ln(10)}\right), \tag{11}$$

where E_d/E_r is the direct-to-reverberant energy ratio. Following [22], the direct (E_d) and reverberant (E_r) energy levels can be estimated by the corresponding energy values within the intervals $[(t_d-1),(t_d+1.5)]$ ms and $[(t_d+1.5),\infty)$ ms, respectively.

For room volumes $\mathrm{Vol} = \{36,140,240,324,450\}$ m^3 emulated by the artificial late reverberation effect, and keeping $d = 1.8$ m fixed, Algorithms V1 and V2 give, for the six different values of T_{60}, the average results provided in Table 5. From this table, one notices that both algorithms, although not precisely matching, do present a strong linear relation with the theoretical volume values. This relationship validates the usage of either algorithm in estimating the ALcons score, as verified below.

Table 5. Comparison of estimation algorithms V1 and V2 for room volume Vol (m^3): mean (μ) and relative standard deviation (γ).

Vol	V1 (μ, γ)	V2 (μ, γ)
36	58, 0.15	123, 0.26
140	78, 0.20	235, 0.30
240	111, 0.24	310, 0.45
324	168, 0.32	427, 0.25
450	231, 0.35	620, 0.33

The three ALcons parameters (T_{60}, d, and Vol) were estimated for all test signals described in Section 4.1, allowing one to determine the expression $\{[d^2(T_{60})^2]/\mathrm{Vol}\}$. The resulting ALcons grades were then statistically correlated with the subjective scores available for those signals. The correlation indices for both volume estimating algorithms reached $\rho = 92\%$. This high value, well above the best results in Table 3, indicates that the ALcons metric may constitute an effective objective method for reverberation assessment in audioband speech signals.

6 Conclusion

This paper addressed the problem of quality evaluation of audioband (24 kHz) speech signals with respect to the reverberation effect. Mathematical models were reviewed and the most important reverberation aspects for the application at hand were indicated. Subjective listening tests were designed and performed to quantify via MOS the human perception of speech impairment by reverberation. Correlation between objective and subjective quality measures have been computed in order to verify the potential ability of standard quality-evaluation methods and the ALcons measure in predicting the subjective quality of speech signals spoiled by reverberation. The $\rho = 92\%$ correlation level between MOS and the ALcons indicates that this measure may be efficiently employed in the application of interest.

References

1. Perkins, M.E., Dvorak, C.A., Lerich, B.H., Zebarth, J.A.: Speech Transmission Performance Planning in Hybrid IP/SCN Networks. IEEE Commun. Mag., 37(7), 126–131 (1999)
2. Falk, T.H., Yuan, H., Chan, W.-Y.: Spectro-Temporal Processing for Blind Estimation of Reverberation Time and Single-Ended Quality Measurement of Reverberant Speech, In: 8^{th} Interspeech - Eurospeech. Antwerp, Belgium (2007)
3. Falk, T.H., Chan, W.-Y.: A Non-Intrusive Quality Measure of Dereverberated Speech. In: 11^{th} Intl. Workshop for Acoustic Echo and Noise Control. Seattle, USA (2008)
4. Peutz, V.M.A.: Articulation Loss of Consonants as a Criterion for Speech Transmission in a Room. J. Audio Eng. Soc., 19, 915–919 (1971)
5. de Lima, A.A., Freeland, F.P., Esquef, P.A.A., Biscainho, L.W.P., Bispo, B.C., de Jesus, R.A., Netto, S.L., Schafer, R., Said, A., Lee, B., Kalker, A.: Reverberation Assessment in Audioband Speech Signals for Telepresence Systems. In: Int. Conf. Signal Process. Multimedia Applic., pp. 257–262. Porto, Portugal (2008)
6. ITU-T Rec. G.191: Software Tools for Speech and Audio Coding Standardization, International Telecommunication Union (2005)
7. Gardner, W. G.: The virtual acoustic room. MSc Thesis, MIT, Cambridge, USA (1992)
8. Allen, J.B., Berkley, D.A.: Image Method for Efficiently Simulating Small-Room Acoustics. J. Acoustic. Soc. Am., 65(4), 943–950 (1979)
9. Schroeder, M.R.: Natural Sounding Artificial Reverberation. J. Audio Eng. Soc., 10(3), 219–233 (1962)
10. Gardner, W. G.: Reverberation Algorithms. In: Kahrs, M., Brandenburg, K. (eds.). Applications of Digital Signal Processing to Audio and Acoustics, pp. 85–131. Kluwer, New York, USA (1998)
11. Jot, J.-M., Chaigne, A.: Digital Delay Networks for Designing Artificial Reverberators. In: 90th Audio Eng. Soc. Conv., Preprint 3030. Paris, France (1991)
12. Schroeder, M.R.: New Method of Measuring Reverberation Time. J. Acoustic. Soc. Am., 37, 409–412 (1965)
13. Allen, J.B.: Effect of Small Room Reverberation on Subjective Preference. J. Acoustic. Soc. Am., 71:S5, (1982)
14. de Lima, A.A., Freeland, F.P., de Jesus, R.A., Bispo, B.C., Biscainho, L.W.P., Netto, S.L., Said, A., Kalker, A., Schafer, R., Lee, B., Jam, M.: On the Quality Assessment of Sound Signals. In: IEEE Int. Symp. on Circuits and Systems, pp. 426–429. Seattle, USA (2008)
15. ITU-T Rec. P.862: Perceptual Evaluation of Speech Quality (PESQ): Objective Method for End-to-End Speech Quality Assessment of Narrow Band Telephone Networks and Speech Codecs. ITU (2001)
16. ITU-T Rec. P.862.2: Wideband Extension to Recommendation P.862 for the Assessment of Wideband Telephone Networks and Speech Codecs. ITU (2005)
17. ITU-T Rec. P.563: Single-ended Method for Objective Speech Quality Assessment in Narrow-band Telephony Applications. International Telecommunication Union (2004)
18. ITU-R Rec. BS.1387: Method for Objective Measurements of Perceived Audio Quality. International Telecommunication Union (1998)
19. Tan, C., Moore, B.C.J., Zacharov, N., Mattila, V.: Predicting the Perceived Quality of Nonlinearly Distorted Music and Speech Signals. J. Audio Eng. Soc., 52(7/8), 699–711 (2004)
20. Larsen, E., Feng, A., Jones D.L.: Room volume and room dimension estimation. US Patent 2006/0126858 A1 (2006)
21. Kuster, M.: Reliability of Estimating the Room Volume from a Single Room Impulse Response. J. Acoustic. Soc. Am., 124, 982–993 (2008)
22. Zahorik, P.: Assessing Auditory Distance Perception Using Virtual Acoustics. J. Acoustic. Soc. Am., 111, 1832–1846 (2002)