

A note on the definition of signal-to-noise ratio of room impulse responses

Csaba Huszty^{1,*} and Shinichi Sakamoto²

¹Graduate School of Engineering, The University of Tokyo,
Ce 401 4-6-1 Komaba, Meguro-ku, Tokyo, 153-8505 Japan

²Institute of Industrial Science, The University of Tokyo,
Ce 401 4-6-1 Komaba, Meguro-ku, Tokyo, 153-8505 Japan

(Received 16 September 2011, Accepted for publication 10 October 2011)

Keywords: Room impulse response, SNR, PNR

PACS number: 43.20.Ye, 43.58.Gn, 43.55.Mc [doi:10.1250/ast.33.117]

1. Introduction

The signal-to-noise ratio (SNR) of room impulse responses (RIRs) and its calculation are subjects of continuous interest in room acoustic measurements. The current measurement standard [1] describes a ‘decay range’ for decay time estimations, but it is common for different representations of the same RIR, for example, the energy-time function (ETC) or the backward integrated energy decay function (EDC), to yield different decay ranges. In decay time estimations, the EDC is often used, on which the knee point shows a dependence on the integration length, while for the ETC, there is no such dependence. The lack of an accurate definition of SNR may lead to ambiguously interpretable results.

In this paper, a definition of two commonly used but different decay ranges and their corresponding signal-to-noise-like ratios is given and discussed first, and then a novel, easily usable method of calculating the SNR from a noise estimate is presented. The iterative truncation [2,3] and subtraction [4] methods for background noise correction and knee point determination, being the currently most widely adopted methods are compared with the proposed method, and it is found that the new method outperforms these two methods in accuracy when using the same input data.

Model fitting methods [5–7] are not considered in this paper since the determined SNR in these approaches is often not used for decay time estimation, they are computationally more expensive than the simpler methods, and finally, they rely on *a priori* information of the decay and their applicability depend on the appropriateness of the selected model.

2. SNR and PNR definitions

The determination of the SNR of a noisy signal depends on the purpose and definition, but it is commonly accepted in engineering applications that the SNR is a ratio of average powers, represented by a single number and expressed in decibels. The power for a stationary discrete signal $x[k]$ can be approximated as its empirical mean square value

$$P = 10 \log_{10} \frac{1}{N} \sum_{k=1}^N (x[k])^2, \quad (1)$$

where N denotes the number of samples. The empirical power of noise can be calculated similarly. SNR is then

$$\text{SNR} = 10 \log_{10} \frac{\frac{1}{N} \sum_{k=1}^N (s[k])^2}{\frac{1}{M} \sum_{k=1}^M (n[k])^2}, \quad (2)$$

where $s[k]$ is the noise-free signal sample and $n[k]$ is the noise sample. In practical RIR measurements however, determining the signal power is not evident, because the signal is not stationary but transient, therefore, different evaluation lengths yield different power values. Since the total energy of the RIR, $\sum_{k=1}^{\infty} (s[k])^2$, is finite even if the evaluation length is infinite (for example assuming an exponential decay envelope corresponding to a diffuse sound field), P_S in Eq. (1) has a zero limit with $N \rightarrow \infty$, or in practice, a decreasing P_S value is obtained with increasing length.

Correspondingly, in the literature, although often not expressed specifically, a peak-to-noise ratio (PNR) is used [8,9] for RIRs as the apparent SNR, which, following the rigorous definition of power ratios, cannot be considered to be the SNR. PNR is obtained as the noise power of a normalized RIR sample,

$$\text{PNR} = 10 \log_{10} \frac{1}{\frac{1}{N} \sum_{k=1}^N (n_n[k])^2}, \quad (3)$$

where $n_n[k]$ is the noise that is observed in a 0 dB normalized RIR. Here, the reference value of the dB scale is the full scale (of the measuring and storage device, i.e., a wave file), and the normalization is in accordance with the noisy RIR. Some authors use a measure called the impulse-to-noise ratio (INR) [10], which is essentially the PNR, so this name will not be used here.

The ISO standard for room acoustic measurements defines the decay curve as a “graphical representation of the decay of the sound pressure level” [1]. If this can be interpreted as being not only the ETC but also the EDC, then the dynamic range until noise dominates the measurement results, defined as the decay range, becomes ambiguous.

The decay range on the EDC is dependent on the evaluated length, because the signal power is dependent on

*e-mail: csaba@iis.u-tokyo.ac.jp

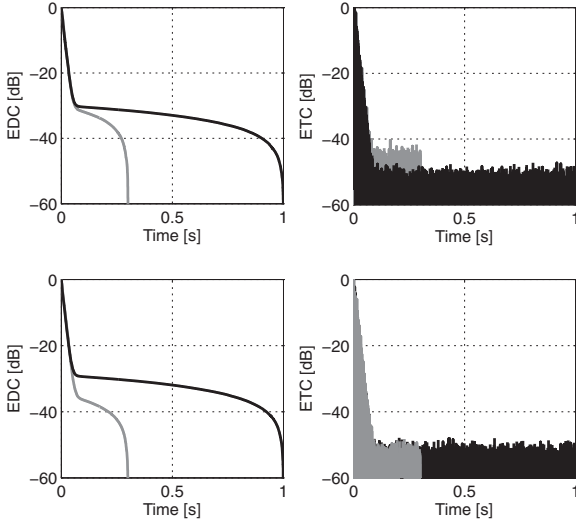


Fig. 1 Difference in the apparent signal-to-noise ratio (PNR) and real signal-to-noise ratio (SNR) in simulated RIRs having the same reverberation time but different evaluation lengths. Top: the same SNR based on the total power of RIR and noise yields a similar knee point on the EDC but different noise levels for the ETC. Bottom: the same level of noise of two different RIRs yield different EDC knee points, but a similar peak-to-peak level difference for the ETC.

the evaluated length, but the noise power is independent. Owing to the fact that both the signal and the noise are evaluated for the same length (i.e., $M = N$ in Eq. (2)), the power ratios are equal to the energy ratios; thus, the EDC indicates the (real) SNR at any given length. In contrast, the ETC shows no such dependence on evaluation length because the signal is considered to be normalized; thus, the ETC supports the direct observation of the PNR (Fig. 1).

The fact that the knee point on the EDC corresponds to the SNR of the RIR allows the effect of the noise to be 'rendered invisible' by forcing the knee point to be moved to lower values through the appropriate selection of evaluation length. This forms the basis of the truncation method [2].

Although the knee point of the EDC is not always visible, particularly at certain reverberation-time-to-length (RL) ratios—where the length is the time support of the available data—that point is usually more restrictive (lower) than the PNR range observed for the ETC (Fig. 2).

3. Determining the SNR from noise estimation

Determining the SNR of an RIR—corresponding to the usually more strict 'decay range' for the EDC—can be accomplished in different ways. One of the established methods is the subtraction method [4]: when the background-noise-corrected and original decay functions are compared, the time point where their difference reaches a given threshold can be used to approximate the SNR. Since the EDC is a monotonically decreasing function, the obtained value is unambiguous. Obviously, the method requires a good estimation of the stationary noise and a suitable selection of the difference threshold. Figure 3 shows an example of simulation results of the achievable accuracy.

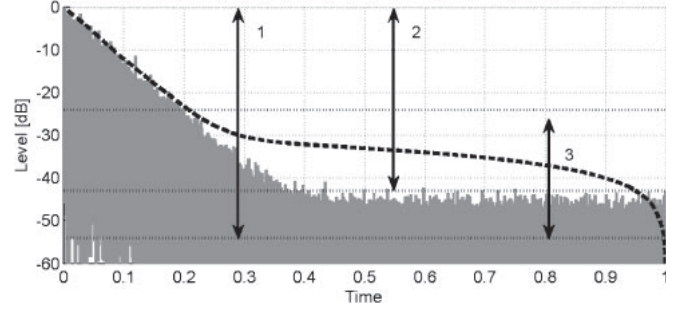


Fig. 2 Different ratios of levels in room impulse responses. 1: PNR = 54.5 dB, 2: peak-to-peak = 42.6 dB, 3: SNR = 30.0 dB. The ranges related to peaks show an apparently larger level difference than the SNR. Simulated decaying white noise and an independent Gaussian white noise sample was used to plot this example with SNR set to 30 dB.

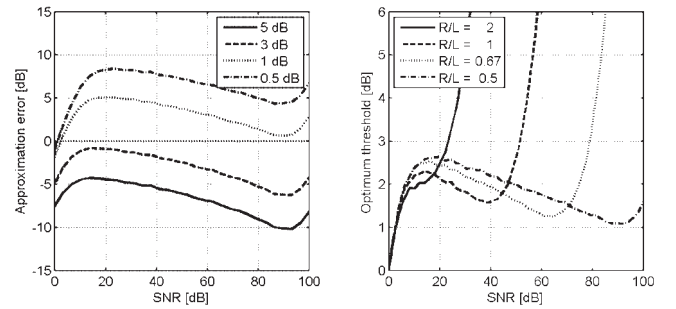


Fig. 3 Determination of the SNR from the EDC using Chu's subtraction method. The value of the subtracted EDC at the time point where the difference between the noisy and subtracted EDC reaches a given threshold determines the SNR. The last 90% of the RIR was considered to be noise. Left: error in the determined SNR in dB when $R/L = 0.5$ using different difference threshold levels. Right: optimum thresholds for determining the SNR accurately for given RL ratios.

Another method, the iterative truncation method [3] is based on the ETC. The iteration yields a time point on the ETC where the noise begins to dominate the RIR, at which point the RIR is truncated to correct for noise. For SNR estimations, the time point is substituted into the EDC, yielding an SNR estimate. Results are shown in Fig. 4, where increasing error with increasing SNR can be observed.

In the following, an alternative method is proposed, which is based on analytical considerations using the EDC.

The EDC using backward integration [11] and energy normalization is obtained as

$$D(t) = 10 \log_{10} d(t) = 10 \log_{10} \left(\frac{\int_t^\infty h^2(\tau) d\tau}{\int_0^\infty h^2(\tau) d\tau} \right) \quad (4)$$

$$= 10 \log_{10} \left(1 - \frac{\int_0^t h^2(\tau) d\tau}{E} \right)$$

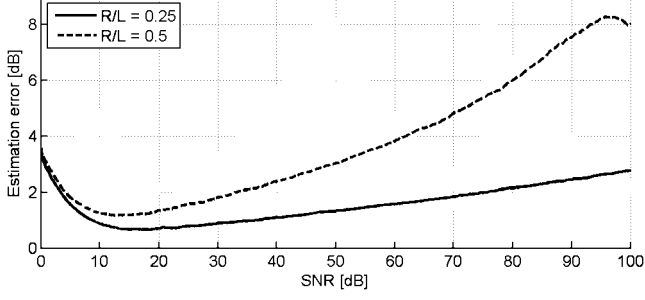


Fig. 4 Determination of the SNR using Lundeby's iterative method. The truncation point value yielded by the iterative method was substituted into the EDC to read its value and compared with the real SNR. Errors in the determined SNR in dB at given RL ratios are plotted. The RIR length was 1 s and was that of a white Gaussian random noise sample multiplied by a single decay constant.

where E denotes the total energy of the RIR $h(t)$, and $d(t)$ is the decay function on a linear scale.

Let the noisy RIR be formulated in terms of a noise-free decay $s(t)$ and an additive noise term $n(t)$. Then the integral in Eq. (4) can be written as

$$\begin{aligned} \int_0^t h^2(\tau) d\tau &= \int_0^t (s + n)^2(\tau) d\tau \\ &= \int_0^t s^2(\tau) d\tau + \int_0^t n^2(\tau) d\tau + 2 \int_0^t s(\tau) n(\tau) d\tau. \end{aligned} \quad (5)$$

In this noisy case, the upper integration limit cannot be infinite in general, as the backward integration requires square integrable functions, which, in stationary noise, is achieved only at a finite length. Hereafter, let $d(t)$ denote the finite noisy energy decay function on the linear scale. Then,

$$\begin{aligned} 1 - d(t) &= \frac{\int_0^t h^2(\tau) d\tau}{E} \\ &= \frac{\int_0^t s^2(\tau) d\tau + \int_0^t n^2(\tau) d\tau + 2 \int_0^t s(\tau) n(\tau) d\tau}{E_S + E_N + 2E_{SN}}, \end{aligned} \quad (6)$$

where $E_S = \int_0^L h^2(\tau) d\tau$ is the signal energy, $E_N = \int_0^L n^2(\tau) d\tau$ is the noise energy, $E_{SN} = \int_0^L s(\tau) n(\tau) d\tau$ is the signal-noise product integral, and L denotes a finite length. From this, it follows that the SNR is

$$\begin{aligned} \frac{E_S}{E_N} &= \frac{1}{1 - d(t)} \cdot \frac{1}{E_N} \cdot \left(\int_0^t s^2(\tau) d\tau + \int_0^t n^2(\tau) d\tau \right. \\ &\quad \left. + \int_0^t s(\tau) n(\tau) d\tau \right) - 1 - 2 \frac{E_{SN}}{E_N}. \end{aligned} \quad (7)$$

It is notable that the SNR in the above formulation is a function of time, but it can be seen that the same value is obtained for every t if the estimates for $s[t]$ and $n[t]$ are perfect, or, in other words, if the signals are *a priori* known. This, unfortunately, is not the case in practice. Moreover, knowing these values would directly yield the SNR; however, the above equation still provides an effective way of

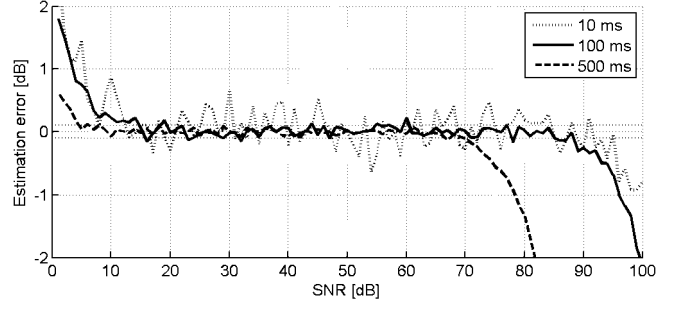


Fig. 5 Estimation of the SNR from the EDC by a direct method with formulas Eqs. (7) and (8) for different lengths of noise sample. Each point in the graph corresponds to an independent white-Gaussian-random-noise-based RIR with $R = 2$ s and an independent white Gaussian noise sample scaled to produce a given power-based SNR. The error shown in the graph is the difference between the generated SNR and the mean SNR[i]. $R/L = 0.5$ was set. The last 10 ms of the RIR contained -128.3 dB of energy relative to the full length, the last 100 ms, -115.3 dB, and the last 500 ms, -90 dB, explaining the dip toward high SNRs.

determining the knee point on the EDC by adopting the following approximative assumptions,

$$s[k] \doteq h[k], \quad k \in [1 \dots N] \quad (8)$$

$$n[k] \doteq [h[q], h[q], \dots, h[q]], \quad q \in [cN \dots N], \quad (9)$$

such that the noise-free sampled $s[k]$ is approximated with the noisy sampled RIR values $h[k]$, while $n[k]$ is a repeated concatenated sequence of noise samples taken from a representative region of the RIR, for example, its end part where only noise can be observed (here, $c = 0.9$ means that the last 10% of the RIR is treated as consisting solely of noise), from the available $N = L \cdot f_s$ samples, where f_s is the sampling frequency. This small noise sequence $h[q]$ is repeated until the length of N samples is reached, to provide the same length for the signal and the noise in the SNR calculation. The value of c can be chosen according to the best available information on the length and reverberation time during a practical measurement, but other parts of the RIR, such as the propagation delay part, can also be used if they are representative of a noise estimate. If no assumption can be made for c , using the above value of $c = 0.9$ will yield results with the accuracy shown as follows. This simplistic procedure yields useful results if the noise assumption is correct, so in this respect, the proposed method has similar requirements to the subtraction method, but uses a different formulation to obtain the SNR estimate more accurately.

A simulation using artificially generated RIRs was used to determine the accuracy of the estimation. Figure 5 shows the case of $R/L = 0.5$. The estimation accuracy is limited, on one hand, by the energy of the RIR contained in the noise sample, and, on the other hand, by the length of the noise sample, i.e., the noise sample should be long enough to be representative of the actual noise. It can be concluded that the longer the noise sample, the worse the SNR conditions that can be accurately handled, assuming stationary noise. However, the

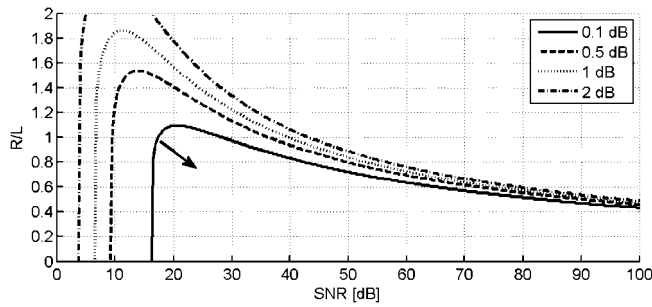


Fig. 6 Estimation accuracy of the proposed method at various RL ratios and SNR values. Contours of different accuracy thresholds are plotted. The arrow points toward the region where accuracy below a given threshold is ensured. The noise sample in this simulation was chosen so that $c = 0.9$.

shorter the noise sample, the higher the error deviation and the higher the maximum estimable SNR. It is also notable that with a suitably chosen noise sample length, the estimation error for artificial RIRs was found to be within approximately 0.1 dB by using this simple method.

Figure 6 shows the expected accuracy for a finite space of RL ratios and SNR values. It can be seen that for the strict case of low SNR values, the allowed range of RL ratios is still high, in some cases exceeding 1, which means that RIRs shorter than their expected reverberation times can also be used for estimating the SNR with the proposed method. When the SNR is very high, the RL ratio is required to be lower than 0.5.

4. Conclusions

Numerous works in the literature describe the use of the SNR in the context of RIRs without a rigorous definition, and the current measurement standard also uses an ambiguous term for the decay range: different decay ranges are obtained from the backward-integrated energy decay function (EDC) and the energy time function (ETC). The EDC indicates the SNR, defined as a power ratio of the impulse response and noise at a given length. As the RIR is a nonstationary signal,

increasing length results in decreasing power values and thus decreasing SNR values. ETC, on the other hand, indicates the peak-to-noise ratio (PNR). Since backward integration is often used in decay time estimations, an easily applicable analytic method to calculate the SNR from a noise estimate was proposed. The accuracy of the proposed method, similarly to the other methods, is dependent on the accuracy of noise estimation, but for the same inputs, the proposed method is more accurate.

References

- [1] ISO 3382-1:2009 "Acoustics—Measurement of room acoustic parameters—Part 1: Performance spaces," International Organization for Standardization (2009).
- [2] R. Kürer and U. Kurze, "Integrationsverfahren zur Nachhallauswertung (Integration procedure for evaluating reverberation)," *Acustica*, **19**, 313–322 (1967).
- [3] A. Lundebj, T. E. Vigran, H. Bietz and M. Vorländer, "Uncertainties of measurements in room acoustics," *Acustica*, **81**, 344–355 (1995).
- [4] W. T. Chu, "Comparison of reverberation measurements using Schroeder's impulse method and decay-curve averaging method," *J. Acoust. Soc. Am.*, **63**, 1444–1450 (1978).
- [5] M. Karjalainen, P. Antsalo, A. Makivirta, T. Peltonen and V. Valimäki, "Estimation of modal decay parameters from noisy response measurements," *J. Audio Eng. Soc.*, **50**, 867–878 (2002).
- [6] N. Xiang, "Evaluation of reverberation times using a nonlinear regression approach," *J. Acoust. Soc. Am.*, **98**, 2112–2121 (1995).
- [7] N. Xiang and P. M. Goggans, "Evaluation of decay times in coupled spaces: Bayesian parameter estimation," *J. Acoust. Soc. Am.*, **110**, 1415–1424 (2001).
- [8] N. Xiang and S. Li, "A pseudo-inverse algorithm for simultaneous measurements using multiple acoustical sources," *J. Acoust. Soc. Am.*, **121**, 1299–1302 (2007).
- [9] T. Jäsa and N. Xiang, "Efficient estimation of decay parameters in acoustically coupled-spaces using slice sampling," *J. Acoust. Soc. Am.*, **126**, 1269–1279 (2009).
- [10] T. Zakinthinos and D. Skarlatos, "The effect of ceramic vases on the acoustics of old Greek orthodox churches," *Appl. Acoust.*, **68**, 1307–1322 (2007).
- [11] M. R. Schroeder, "New method of measuring reverberation time," *J. Acoust. Soc. Am.*, **37**, 409–412 (1965).