Acoustic Scene Analysis Using Estimated Impulse Responses

Erik Larsen, Chris D. Schmitz,
Charissa R. Lansing, William D. O'Brien Jr., Bruce C. Wheeler and Albert S. Feng

Beckman Inst. for Adv. Sci. and Techn.
University of Illinois at Urbana-Champaign
405 N. Mathews Ave. Urbana, IL 61801
Email: elarsen@uiuc.edu

Abstract–Pre-processing for hearing aids, such as adaptive beamforming, is sensitive to characteristics of the acoustic environment, in particular, reverberation. We conjecture that knowledge of the acoustic environment can aid selection of an optimal signal processing strategy. In this paper we present a method for estimation of direct-to-reverberant energy ratio from partial room responses to an impulsive sound at one receiver location, and of source distance, average absorption, reverberation time, and room volume as intermediate results.

I. Introduction

Reverberation and noise are detrimental to speech intelligibility, especially in combination, and more so for the hearing impaired and elderly than for normal hearing listeners (Nàbělek [1], Helfer and Huntley [2]). For hearing aid users, intelligibility improvements can be obtained by using directional microphones (Hawkins and Yacullo [3]), which emphasize the target speech over reverberation and noise by virtue of higher sensitivity in the target direction. Higher directionality can be obtained by using fixed beamformers, which have shown substantial intelligibility gains over directional microphones (Desloge *et al.* [4], Merks [5]). Best performance in midly reverberant environments and with limited number of interferers can be obtained by adaptive beamformers; however, performance is compromised when the number of interferers increases and/or the environment has moderate to strong reverberation (Greenberg and Zurek [6]). In strong reverberation, the adaptive beamformer directional pattern defaults to the that of the underlying fixed beamformer. The existing data shows that there is little advantage of adaptive over fixed systems for direct-to-reverberant energy ratios of 0 dB or lower. This ratio depends on source distance, as well as other room acoustic parameters Kuttruff [7].

To maximize speech intelligibility in a listening environment, we expect that information of the room reverberation time (RT) and direct-to-reverberant energy ratio (D/R) (that can be blindly estimated for one or more passively received microphone signals) would be useful for selecting the optimal signal processing strategy. The RT estimation problem has been addressed in Ratnam *et al.* [8, 9]. In this paper results are reported from a preliminary study to estimate D/R from a partial room impulse response (RIR). As intermediate results, we can obtain estimates of source distance, room volume, RT, and average absorption coefficient.

Research is supported by grants from the NIDCD of the National Institutes of Health (R21DC-04840), Beckman Institute, Charles M. Goodenberger Fund, and Phonak AG.

II. Estimation of direct-to-reverberant energy ratio

A. *Definitions and preliminary remarks*

D/R is defined through the RIR $h(t)$ as

$$\text{D/R} = 10 \log \frac{\int_0^{T_d} h^2(t)\,\mathrm{dt}}{\int_{T_d}^{\infty} h^2(t)\,\mathrm{dt}} \ \ \text{dB}, \tag{1}$$

where T_d is the duration of the direct sound (usually taken as roughly 2 ms). If $h(t)$ is entirely known, computation of D/R is trivial. However, in the following we will assume that only the first part of $h(t)$ is known, more specifically, the first, say, 10 discrete early reflections. For the moment we assume that we know the delays $\{\tau_i'\}$ (relative to direct sound) and pressure amplitudes $\{a_i\}$ of N early discrete reflections (including the direct sound, for which $\tau_1' = 0$.) The actual delay of reflection i is $\tau_i = \tau_i' + \tau_d$, where τ_d is the actual delay of the direct sound, which is unknown. Note that in order to obtain this data, we first need to blindly estimate the RIR from a passively recorded signal, i.e. blind channel identification. Subsequently, the actual reflection delays and amplitudes need to be extracted from the RIR; we have made some progress into the latter using sparse spike inversion (O'Brien *et al.* [10]). We will not discuss these matters further here as they are not directly relevant to scene analysis algorithms, but are to be considered as pre-processors.

0-7803-8104-1/03/$17.00 ©2003 IEEE

B. D/R estimation

The energy $E(0, \tau_N)$ of the $N-1$ reflections (excluding direct sound) and the energy $E(0,0)$ of the direct sound are

$$E(0, \tau_N) = \frac{1}{\rho c} \sum_{i=2}^{N} a_i^2, \quad (2)$$

$$E(0,0) = a_1^2 / \rho c, \quad (3)$$

using ρ for density of air and c for the speed of sound. We can estimate the total reverberant energy E_r because the decrease in energy of reflections, as well as the number of reflections, arriving after τ_N at the receiver are predictable from observations up to τ_N. To estimate E_r we first define $e(\tau)$ as expected energy of a reflection arriving with a delay τ and $D(\tau)$ as the expected density of reflections in units of s^{-1} arriving around delay τ. We have

$$e(\tau) = \frac{1}{\rho c} \frac{(1-\bar{\alpha})^{R(\tau)}}{(c\tau)^2}, \quad (4)$$

$$D(\tau) = \frac{\partial n}{\partial \tau} = \frac{4\pi}{V} c^3 \tau^2, \quad (5)$$

$$R(\tau) = \frac{c\tau}{\lambda}, \quad (6)$$

$$\lambda = \frac{4V}{S}. \quad (7)$$

The denominator in Eqn. 4 is due to geometrical spreading of the wave front, and the numerator accounts for energy absorbed during reflections. Eqn. 5 can be derived from the image model. We have used c for speed of sound, V for room volume, $\bar{\alpha}$ for energy absorption coefficient (averaging over frequency and angles of incidence), and λ for mean free path length of sound rays. We define $R(\tau)$ as the expected number of reflections a sound ray experiences τ s after the sound source emits the sound. Then we may compute the expected energy $U(\tau_a, \tau_b)$ of the RIR between delays τ_a and τ_b as

$$\begin{aligned} U(\tau_a, \tau_b) &= \int_{\tau_a}^{\tau_b} e(\tau) D(\tau) \, \mathrm{d}\tau, \quad (8) \\ &= -\frac{16\pi}{\rho c S \ln(1-\bar{\alpha})} \times \\ &\quad \left[(1-\bar{\alpha})^{R(\tau_a)} - (1-\bar{\alpha})^{R(\tau_b)}\right], \end{aligned}$$

which is valid for $\bar{\alpha} > 0$. From Eqn. 8 we find for the total reverberant energy E_r

$$\begin{aligned} E_r &= \frac{U(0, \tau_N)}{U(0, \tau_N)} E_r, \\ &= E(0, \tau_N) \frac{U(0, \infty)}{U(0, \tau_N)}, \\ &= \frac{1}{\rho c} \frac{\sum_{i=2}^{N} a^2(i)}{1 - (1-\bar{\alpha})^{R(\tau_N)}}, \quad (9) \end{aligned}$$

where we have substituted $E(0, \tau_N)$ for $U(0, \tau_N)$ and $U(0, \infty)$ for E_r. Then D/R follows as

$$\begin{aligned} \text{D/R} &= 10 \log \frac{a_1^2}{\sum_{i=2}^{N} a_i^2} \times \\ &\quad \left[1 - (1-\bar{\alpha})^{R(\tau_N)}\right] \text{ dB.} \quad (10) \end{aligned}$$

The structure of this D/R estimate can be seen as a pre-estimate taken directly from the $\{a_i\}$ and a correction factor that is computed based on expected values of room acoustic quantities.

C. Volume estimation

To actually compute D/R according to Eqn. 10, we need to know $\bar{\alpha}$ and $R(\tau_N)$. To compute the latter we need an estimate of λ, which in turn requires estimates of V and S. V can be estimated using only the delays $\{\tau_i\}$, as the expected density of reflections depends on V, refer to Eqn. 5. The estimated value for V is thus

$$V = \frac{4\pi}{3N} (c\tau_N)^3. \quad (11)$$

We can estimate S from V by considering that

$$S = \beta V^{2/3}, \quad (12)$$

where β is a parameter that depends on the geometry; its minimum value $\beta = 6$ is attained for a cube shape. To estimate the probability distribution of β, we computed its value for random rectangular room geometries (length of each side within 1:5 ratio of one another), for 10000 trials. The resulting distribution is monotically decreasing as a function of β. Its parameters are: mean 6.4, median 6.3, standard deviation 0.36, skewness 1.2, kurtosis 4.0. Although the distribution thus found is probably not an exact model for a distribution based on actual room geometries, we expect this to be a reasonable approximation.

Combining Eqns. 6, 7, 11, and 12, we find that

$$R(\tau_N) = \frac{\beta}{4} \left(\frac{3N}{4\pi}\right)^{1/3}. \quad (13)$$

To compute $R(\tau_N)$ we actually do not need a room volume estimate; knowing the number of reflections N suffices.

D. Distance and absorption estimation

The last parameter required for estimating D/R is $\bar{\alpha}$, the average energy absorption coefficient at the wall boundaries. We can estimate this quantity from the amplitudes $\{a_i\}$ of the reflections, although we will need to compensate for the decay due to geometrical spreading. To properly account for this, an estimate of source distance is required. We can get a lower bound estimate of source distance by considering that

$$a_1 = \frac{g}{c\tau_d} \quad (14)$$

$$a_i \leq \frac{g}{c(\tau_d + \tau_i')}, \; i \geq 2, \quad (15)$$

where g is the (unknown) source strength. The actual amplitude a_i in Eqn. 15 will be smaller due to energy absorption upon reflection. From this we find that

$$\tilde{\tau}_{d,i} \geq \frac{a_i}{a_1 - a_i} \tau_i'. \quad (16)$$

An estimate $\tilde{\tau}_{d,i}$ can be obtained for each observed reflection i, and as final estimate the maximum $\tilde{\tau}_{d,i}$ thus obtained will be closest to the actual τ_d:

$$\hat{\tau}_d = \max(\tilde{\tau}_{d,i}). \quad (17)$$

Multiplying $\hat{\tau}_d$ by c provides an estimate of source distance. From this we can compute source strength as

$$g = a_1 c \hat{\tau}_d. \tag{18}$$

Now that τ_d has been estimated, we can estimate $\bar{\alpha}$, because if the $\{a_i\}$ are smaller than would be expected based on geometrical decay for a source at an estimated distance of $c\hat{\tau}_d$, the additional decay must be due to absorption upon reflection. The additional decay is equal to the pressure reflection coefficient β_r, which is related to α through $\alpha = 1 - \beta_r^2$; $\tilde{\beta}_{r,i}$ can be estimated for each observed reflection i as

$$\tilde{\beta}_{r,i} = \frac{a_i}{g} c(\hat{\tau}_d + \tau_i'). \tag{19}$$

Each $\tilde{\beta}_{r,i}$ thus obtained can be converted to α_i and finally to its mean value of $\bar{\alpha}$.

E. *Simulation results*

To illustrate performance of these methods in typical circumstances, we simulated impulse responses in a room of $5 \times 8 \times 3$ m, with $\bar{\alpha} = 0.4$ (not the same for all walls). The direct sound and the first 9 reflections were used in the estimation procedures. Random source and receiver positions within the room were selected on 1000 trials, the results of which appear in Figs. 1 (D/R estimation), 2 (distance estimation), 3 ($\bar{\alpha}$ estimation), 4 (room volume estimation), and 5 (reverberation time estimation). The D/R estimate (Fig. 1) is fairly accurate, the variance being a few dB, which should be acceptable in all but the most demanding applications. The correction factors for the reverberant tail were 4.2 ± 0.6 dB; this means that neglecting to use this factor would result in a significant overestimation of D/R ratio. The variance in the estimates is mainly due to the variation in energy of the early reflections. The distance estimates (Fig. 2) are biased to low values, especially for moderate and large distances; also, the spread is quite large. This underestimation of source distance also leads to underestimation of room volume (Fig. 4) and reverberation time (Fig. 5). In contrast, $\bar{\alpha}$ estimates are not seriously biased (Fig. 3), although the variance of estimates is quite large. Note that distance estimates will be more accurate in case of lower $\bar{\alpha}$ values, and this will have a positive effect on estimates of V and RT as well.

III. Discussion and Future Research

We have presented a method for estimating direct-to-reverberant energy ratio (D/R) from a partial impulse response, without knowledge of direct sound delay. The estimate consists of a pre-estimate based on observed amplitudes of direct sound and a few early reflections, and a correction factor which is based on expected values of room acoustic parameters, derived from the observed early reflections. We illustrated the method by means of an example using 9 reflections, which showed that nearly unbiased estimates are obtained over a wide D/R range, with a spread of a few dB. As an intermediate result, source distance can be estimated, but the estimates are markedly underestimated for moderate and large distances. Also, estimates of average energy absorption

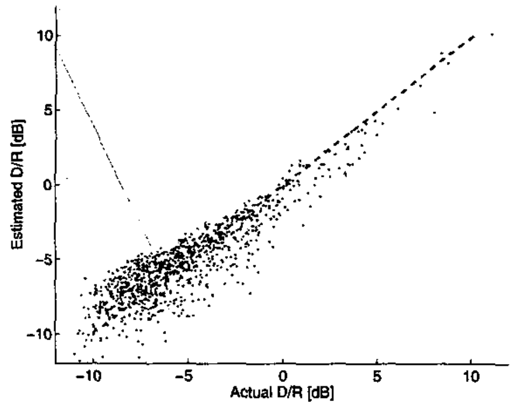

Figure 1: Actual D/R ratio (abscissa) vs. estimated D/R ratio (ordinate).

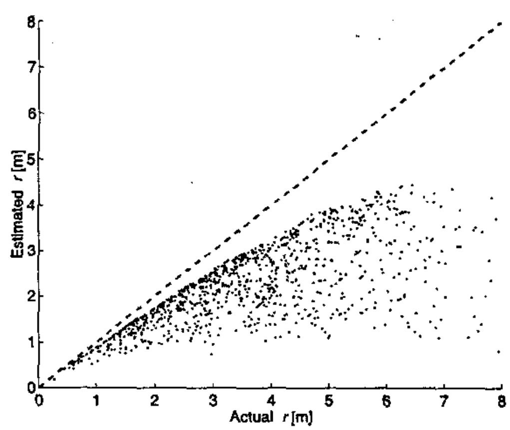

Figure 2: Actual distance (abscissa) vs. estimated distance (ordinate).

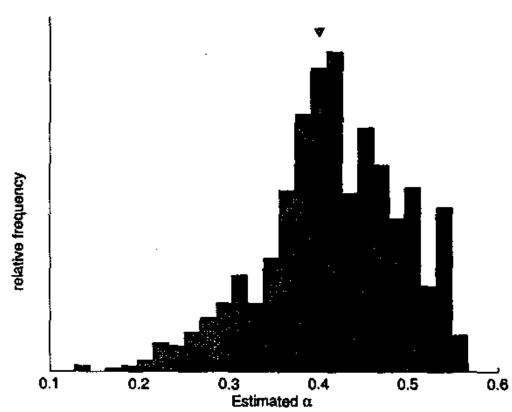

Figure 3: Histogram of $\bar{\alpha}$ estimates. The true value is indicated by the triangle.

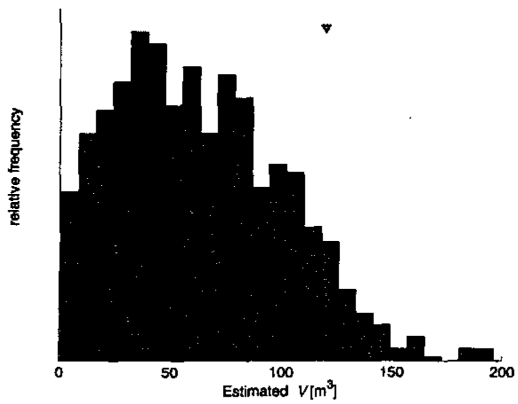

Figure 4: Histogram of room volume estimates. The true value is indicated by the triangle.

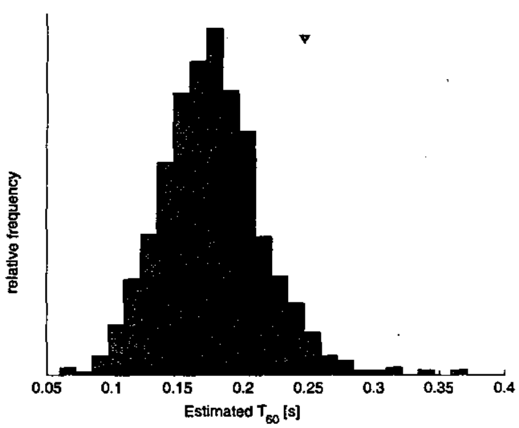

Figure 5: Histogram of reverberation time estimates. The true value is indicated by the triangle.

coefficient, room volume, and reverberation time can be obtained. The latter two estimates are biased downward due to the underestimation of the source distance.

It is thus imperative to find improved methods for estimating source distance r_s. An appealing option is to use the critical distance r_c, for which D/R=0 dB, which is given by

$$r_c = \frac{1}{4}\sqrt{\frac{Q\bar{\alpha}S}{\pi(1-\bar{\alpha})}}, \tag{20}$$

where Q is the source directivity factor (which is unknown). S is obtained from V through Eqn. 12, but to accurately estimate V, the delay of the direct sound must be known (Eqn. 11). This is equivalent to knowing r_s, but this is the quantity we are trying to estimate. A better way to estimate r_c might be through the power spectrum of the impulse response. Jetzt [11] showed that D/R is uniquely and monotonically related to the standard deviation of this power spectrum. Thus, if r_c and D/R are known r_s can be estimated, because the decay of D/R with distance is about 6 dB per distance doubling:

$$r_s = r_c \times 2^{-(\mathrm{D/R})/6}. \tag{21}$$

In a hearing aid application, often two microphones will be available (one at each ear). We are currently investigating the benefit of combining impulse responses from both microphones, such as enabling direction-of-arrival estimates for each reflection. When applied to (noisy) measurement data, the two microphone signals offer additional advantage in that two independent looks of the data are available.

Acknowledgment

The authors wish to thank Advay Shirsalkar, Tanvi Patel, and Deepti Narla from the University of Illinois for their assistance in evaluation of some of the presented methods.

References

[1] A.K. Nábělek. Identification of vowels in quiet, noise, and reverberation: relationships with age and hearing loss. *J. Acoust. Soc. Am.*, 84(2):476–484, 1988.

[2] K.S. Helfer and R.A. Huntley. Aging and consonant errors in reverberation and noise. *J. Acoust. Soc. Am.*, 90(4):1786–1796, 1991.

[3] D.B. Hawkins and W.S. Yacullo. Signal-to-noise ratio advantage of binaural hearing aids and directional microphones under different levels of reverberation. *J. Speech Hear Disorders*, 40:278–286, 1984.

[4] J.G. Desloge, W.M. Rabinowitz, and P.M. Zurek. Microphone-array hearing aids with binaural output — Part I: Fixed-processing systems. *IEEE Trans. Speech Audio Proc.*, 5(6):529–542, 1997.

[5] I.L.D.M. Merks. *Binaural application of microphone arrays for improved speech intelligibility in a noisy environenment.* PhD thesis, Delft University of Technology, 2000.

[6] J.E. Greenberg and P.M. Zurek. Microphone arrays, signal processing techniques and applications. In M. Brandstein and D. Ward, editors, *Mircophone-array hearing aids*, chapter 11. Springer (New York, NY), 2001.

[7] H. Kuttruff. *Room Acoustics*. Elsevier (Essex, UK), 3rd edition, 1991.

[8] 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.*, 114(5):2877–2892, 2003.

[9] R. Ratnam, D.L. Jones, and W.D. O'Brien Jr. Fast algorithms for blind estimation of reverberation time. *IEEE Sign. Proc. Lett. (in press).*

[10] M.S. O'Brien, A.N. Sinclair, and S.M. Kramer. Recovery of a sparse spike time series by l_1 norm deconvolution. *IEEE Trans. on Sign. Proc.*, 42(12):3353–3365, 1994.

[11] J.J. Jetzt. Critical distance measurements of rooms from the sound energy spectral response. *J. Acoust. Soc. Am.*, 65(5):1204–1211, 1979.