ON THE POOR ROBUSTNESS OF SOUND EQUALIZATION IN REVERBERANT ENVIRONMENTS

Biljana D. Radlovic *Robert C. Williamson* *Rodney A. Kennedy*

Department of Engineering
Faculty of Engineering and
Information technology
Australian National University
Canberra ACT 0200
Australia

Telecommunications Engineering Group
Research School of Information Sciences
and Engineering
The Institute of Advanced Studies
Australian National University
Canberra ACT 0200, Australia

ABSTRACT

This paper examines the sensitivity of sound equalization to source or microphone position changes in a reverberant room. It is demonstrated that even small displacements from the reference (equalization) point, of the order of a tenth of the acoustic wavelength, can cause large degradations in the equalized room response. The general theory developed in this paper, which implies that the sound equalization in practical environments may be an ill-posed problem, is verified by the simulation results averaged over different source and microphone positions.

1. INTRODUCTION

In many applications, when the speech signal is transmitted from source to microphone in a reverberant room, it is necessary to use an inverse filter in order to compensate for uneveness in the frequency response of the room, that is, to make the transfer function of the system be (approximately) equal to the desired one. The majority of literature concerned with the sound equalization in reverberant enclosures deals with the case where the source and microphone positions are assumed to be fixed, which significantly simplifies the problem of reverberation removal from the sound picked up by a microphone at some point in a room. However, an effective technique for sound equalization over a wider area in a room has not yet been found. Spatial limitations to the standard equalization techniques for speech dereverberation have been experimentally demonstated in [1].

In this paper we present an analysis of the robustness of equalization in reverberant rooms, for changing source or microphone positions. We are basically interested in the spatial extent of the zone where a significant reverberation reduction is obtained, which can be achieved when a sound picked up by a single microphone is equalized using a fixed room response inverse filter. For simplicity, only an ideal case of an exact inverse response filter is considered; it is understood that a similar analysis will apply to the approximate inverse.

In order to derive more general results, we make use of the statistical-average properties of sound transmission in rooms. The basic assumption of room statistics is that the sound pressure is fairly uniformly distributed throughout the room volume. For this to be true, the following conditions must be met [2]:

1. The linear dimensions of the room must be large relative to the wavelength. This condition is easily satisfied in almost all rooms, for frequencies of interest.

2. The average spacing of the resonance frequencies must be smaller than one-third of their bandwidth. In a room having a volume V and reverberation time T_{60}[1], this criterion can be met for all frequencies which exceed the Schroeder large room frequency given by $f_s = 2000(T_{60}/V)^{1/2}$.

3. Both source and microphone are in the interior of a room, at least a half-wavelength away from the walls.

Subject to these not very restrictive conditions, the frequency response between the source and receiver may be treated as a random function, the properties of which are determined by the room volume, reverberation time and magnitude of the sound pressure.

2. ROBUSTNESS RESULTS

We begin the theoretical consideration of our subject from the following simplifying assumption:

[1]The reverberation time T_{60} is the length of time for the sound intensity level in a room to drop by 60 dB after the source is shut off.

Let $\underline{G}_f$ be the complex steady-state frequency response between the sound source and the receiver, and $\underline{H}_f$ the frequency response of an inverse filter designed to equalize room response at the microphone point. We idealize our problem by assuming that the transmission path between the source and receiver is perfectly equalized, where *perfect* means equalizing both the amplitudes and the phases of the frequency response.

Moving the microphone will distort the frequency response of the equalized transmission path. A quantitative measure of this degradation can be based on a difference between the two system's transfer functions related to the reference and the displacement point.

Definition 1: Let $\underline{\widetilde{G}}_f$ be the frequency response between the source and the receiver placed some distance away from the equalization point, and $\underline{H}_f$ an exact inverse of $\underline{G}_f$. The *mean squared error* at frequency f due to the displacement of the receiving point is defined by

$$\mathbb{W}_f = E\{|\underline{\widetilde{G}}_f \underline{H}_f - 1|^2\} \tag{1}$$

where $E\{\cdot\}$ represents the expected value operator. The expectation is taken with respect both to the distribution of source locations (assumed uniform throughout the room volume but at least a half-wavelength away from the walls) and to the distribution of microphone positions (assumed uniformly distributed on the sphere of radius r centered at the reference location). We note that $\mathbb{W}_f$ goes to zero at the receiver reference point, where $\underline{\widetilde{G}}_f$ becomes equal to $\underline{G}_f$.

The definition given above can be generalized: multiplied by the squared amplitude of the source signal at frequency f, (1) estimates the power of an error signal which is the difference between the equalizer output signal and desired (source) signal.

Before going further, we define the quantity called the *wavenumber*, k, as $k = 2\pi/\lambda = 2\pi f/c$, where λ is the acoustic wavelength, and c is the velocity of sound, generally specified at $21°$C as 344 m/s.

Theorem 1 *Let R denote the distance from the source to the reference location, r the displacement from the equalization point, k the wavenumber, and γ the ratio of the direct to the reverberant sound energy density at the reference point. Then, the mean squared error at frequency f is*

$$\mathbb{W}_f \cong \frac{\gamma \frac{R}{2r} \ln|\frac{R+r}{R-r}| + 1}{\gamma + 1} - 2\frac{\sin(kr)}{kr} + 1 \tag{2}$$

with γ given by

$$\gamma = \frac{0.01V}{\pi R^2 (1-\bar{\alpha}) T_{60}} \tag{3}$$

where $\bar{\alpha}$ is the average absorption coefficient (the fraction of incident acoustic power absorbed by the room surfaces), and V is the volume of the room. □

In (3), absorption in the air is neglected for simplicity.

Proof: See the Appendix.

We make the following observations regarding this result:

1. The mean squared error depends merely on the $ratio$ of the room volume and reverberation time. This property generalizes the results derived herein to rooms of different shapes, volumes, and reverberation times.

2. If the displacement from the equalization point is small compared to the source-to-microphone distance, the first term in (2) approaches 1, meaning that direct field component has negligible effect on the error signal[2]. Because $\gamma \sim 1/R^2$, the same will be valid at larger distances from the source location, where (2) reduces to a simpler formula: $\mathbb{W}_f \cong 2 - 2\sin(kr)/(kr)$.

3. The greatest amount of distortion can be expected for high frequencies, where the term $\sin(kr)/kr$ falls off rapidly with increasing the distance from the equalization point.

In analogy, the same results would be derived with the microphone being fixed and source being moved from one point to another.

3. SIMULATION EXAMPLE

In order to validate the theoretical results derived in the preceding section, simulations have been performed for several rooms of different volumes and reverberation times, over a wide range of source-microphone positions. In calculating the frequency response of the room (source-to-microphone transfer function), we have used the image method [6].

In the example we present here, a rectangular room with dimensions 6.4 m, 5 m, and 4 m is considered. We assume that the reverberation time does not change with the frequency, and that all walls of the room have the same reflection coefficient β (the average absorption coefficient $\bar{\alpha} = 1 - \beta^2$).

First we calculated frequency response at some fixed distance from the source placed in the interior of the room. Then, moving the microphone in an arbitrary direction, we calculated the error at several points along a straight path of one wavelength (see (1)). The 200 simulations with different source-microphone positions were made, with both the source and microphone being displaced randomly between the runs. The average power of the error signal at frequency f was estimated by

$$\overline{\mathbb{W}}_f = \frac{1}{N}\sum_{n=1}^{N} \left|\frac{\underline{\widetilde{G}}_{fn}}{\underline{G}_{fn}} - 1\right|^2 \tag{4}$$

[2]This can be easily verified by using the l'Hôpital's rule, according to which: $\lim_{r\to 0}[(R/2r)\ln|(R+r)/(R-r)|] = \lim_{r\to 0}[(\partial(\ln|(R+r)/(R-r)|)/\partial r)/(\partial(2r/R)/\partial r)] = 1$.

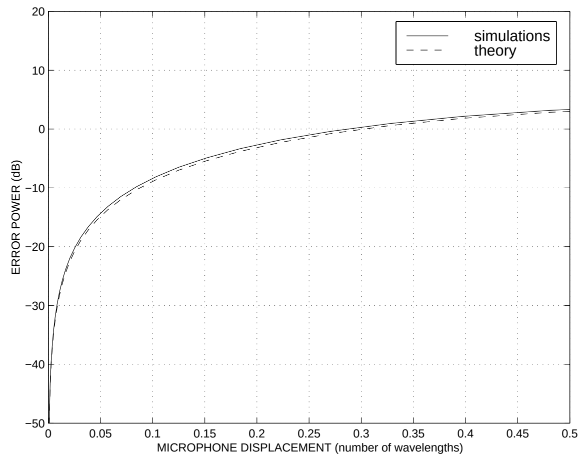

Figure 1: Reverberant power as a function of displacement from the equalization point, at frequency 2000 Hz. The room dimensions are $(6.4 \times 5 \times 4)$ m^3, wall reflection coefficients 0.84, and direct-to-reverberant energy ratio -8.4 dB (source-to-microphone distance $R = 3$ m).

where $\underline{G}_{fn}$ denotes the complex room response at the reference location (associated with the nth simulation), $\underline{\widetilde{G}}_{fn}$ is the room response at some distance r from the reference point, and $N = 200$.

In Fig. 1, the solid line represents the averaged trend of the error signal for many source and microphone locations; such an average is in very good agreement with what one calculates from (2) (dashed line). These results, as well as the other simulation results not presented in this paper, demonstrate that the zone of equalization, where more than 10 dB of reverberation reduction is obtained, is a sphere of a diameter of about $\lambda/7$, centered at the reference point.

4. CONCLUSION

The purpose of this paper was to investigate the robustness of equalization using a room response inverse filter, with respect to source or microphone location changes. The theory and simulation results presented above demonstrate that standard equalization techniques are merely a point-wise solution which not only fails in removing unwanted reverberation but may even further degrade speech quality away from the equalization point.

5. APPENDIX

We use the following relation between the frequency domain Green's function (frequency response function), $\underline{G}_f$, and the complex sound pressure at some point in a room, $\underline{P}_f$ [5, pages 555,591]:

$$\underline{P}_f = -ik\rho c \underline{S}_f \underline{G}_f \tag{5}$$

where factors ρ and $\underline{S}_f$, which will be defined later in this Appendix, are independent of the receiver location. Thus, the equation (1) can be written in the form

$$\mathbb{W}_f = E\left\{ \frac{\underline{\widetilde{P}}_f \underline{\widetilde{P}}_f^*}{\underline{P}_f \underline{P}_f^*} - \frac{\underline{\widetilde{P}}_f}{\underline{P}_f} - \frac{\underline{\widetilde{P}}_f^*}{\underline{P}_f^*} + 1 \right\} \tag{6}$$

where the symbol * represents the complex conjugate. Let us first multiply and divide the second and the third term in (6) by $\underline{P}_f^*$ and $\underline{P}_f$, respectively.

To calculate the expectation of each term in the sum in (6), we use the Taylor expansion [3, pages 246-7], according to which if g is the function of $random$ variables with the mean values $E\{x_j\} = \bar{x}_j$, $j = 1, \ldots, n$, then $g(x_1, x_2, \ldots, x_n)$, which we write as $g(x)$ for brevity, can be expressed in the form: $g(x) = g(\bar{x}) + \sum_{j=1}^{n} g_j'(\bar{x})(x_j - \bar{x}_j) + \hat{g}(x)$, where $\hat{g}$ is a function of order 2, i.e., all its partial derivatives up to the first order vanish at $(\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_n)$. Thus, to the first order of approximation, $E\{g(x)\} = g(\bar{x})$. With this approximation, we find

$$\mathbb{W}_f \cong \frac{E\{\underline{\widetilde{P}}_f \underline{\widetilde{P}}_f^*\}}{E\{\underline{P}_f \underline{P}_f^*\}} - \frac{E\{\underline{\widetilde{P}}_f \underline{P}_f^*\}}{E\{\underline{P}_f \underline{P}_f^*\}} - \frac{E\{\underline{\widetilde{P}}_f^* \underline{P}_f\}}{E\{\underline{P}_f^* \underline{P}_f\}} + 1. \tag{7}$$

The total sound pressure at frequency f, at some point away from the source, may be expressed as: $\underline{P}_f = \underline{P}_{fd} + \underline{P}_{fr}$, where $\underline{P}_{fd}$ and $\underline{P}_{fr}$ are the direct and reverberant sound pressure components, respectively. Under the same conditions as in the Introduction, the direct and reverberant sound pressure are uncorrelated at the point of observation and, therefore, all cross terms in the sum in (7) will vanish, leaving only the following factors

$$\begin{aligned} \mathbb{W}_f \cong & \frac{E\{\underline{\widetilde{P}}_{fd} \underline{\widetilde{P}}_{fd}^* + \underline{\widetilde{P}}_{fr} \underline{\widetilde{P}}_{fr}^*\}}{E\{\underline{P}_{fd} \underline{P}_{fd}^* + \underline{P}_{fr} \underline{P}_{fr}^*\}} - \frac{E\{\underline{\widetilde{P}}_{fd} \underline{P}_{fd}^* + \underline{\widetilde{P}}_{fr} \underline{P}_{fr}^*\}}{E\{\underline{P}_{fd} \underline{P}_{fd}^* + \underline{P}_{fr} \underline{P}_{fr}^*\}} \\ & - \frac{E\{\underline{P}_{fd} \underline{\widetilde{P}}_{fd}^* + \underline{P}_{fr} \underline{\widetilde{P}}_{fr}^*\}}{E\{\underline{P}_{fd} \underline{P}_{fd}^* + \underline{P}_{fr} \underline{P}_{fr}^*\}} + 1. \end{aligned} \tag{8}$$

The reverberant-field mean-square pressure can be defined as

$$E\{\underline{P}_{fr} \underline{P}_{fr}^*\} = E\{\underline{\widetilde{P}}_{fr} \underline{\widetilde{P}}_{fr}^*\} = \frac{4\rho c \Pi (1 - \bar{\alpha})}{S\bar{\alpha}} \tag{9}$$

where Π is the power of the acoustic source, and ρ is the density of air in the room.

With (3) (see [4, page 582]) and (9), returning to (8), we find

$$\mathbb{W}_f \cong \frac{E\{\underline{\widetilde{P}}_{fd}\underline{\widetilde{P}}^*_{fd}\}\frac{S\bar{\alpha}}{4\rho c\Pi(1-\bar{\alpha})}+1}{\gamma+1} - \frac{E\{\underline{\widetilde{P}}_{fd}\underline{P}^*_{fd}\}\frac{S\bar{\alpha}}{4\rho c\Pi(1-\bar{\alpha})}+\frac{E\{\underline{\widetilde{P}}_{fr}\underline{P}^*_{fr}\}}{E\{\underline{P}_{fr}\underline{P}^*_{fr}\}}}{\gamma+1} - \frac{E\{\underline{P}_{fd}\underline{\widetilde{P}}^*_{fd}\}\frac{S\bar{\alpha}}{4\rho c\Pi(1-\bar{\alpha})}+\frac{E\{\underline{P}_{fr}\underline{\widetilde{P}}^*_{fr}\}}{E\{\underline{P}_{fr}\underline{P}^*_{fr}\}}}{\gamma+1}+1. \quad (10)$$

Next we calculate $E\{\underline{\widetilde{P}}_{fd}\underline{\widetilde{P}}^*_{fd}\}$ starting from the following equations: The free-space Green's function is defined as

$$\underline{g}_f = \frac{1}{4\pi R}e^{ikR}. \quad (11)$$

At a distance R from the source point, for a given source strength $\underline{S}_f$, the direct sound pressure is of the form [5, page 311]

$$\underline{P}_{fd} = -ik\rho c\underline{S}_f\underline{g}_f. \quad (12)$$

Some distance r away from the equalization point, at an angle θ to the reference direction, the sound pressure is given by

$$\underline{\widetilde{P}}_{fd} = -ik\rho c\underline{S}_f\underline{\widetilde{g}}_f. \quad (13)$$

We may write thus [5, page 582]

$$E\{\underline{\widetilde{P}}_{fd}\underline{\widetilde{P}}^*_{fd}\} = (k\rho c)^2|\underline{S}_f|^2E\{\underline{\widetilde{g}}_f\underline{\widetilde{g}}^*_f\} = 4\Pi\rho c\pi E\{\underline{\widetilde{g}}_f\underline{\widetilde{g}}^*_f\}. \quad (14)$$

The function $\underline{\widetilde{g}}_f$ in (14) can be defined by using the cosine law

$$\underline{\widetilde{g}}_f = \frac{1}{4\pi(R^2+r^2-2Rr\cos\theta)^{1/2}}e^{ik(R^2+r^2-2Rr\cos\theta)^{1/2}}. \quad (15)$$

Because all directions of the microphone displacement are assumed to be equally probable over the solid angle, $\cos\theta$ is distributed uniformly over the interval -1 to $+1$, and the expectation of $\underline{\widetilde{g}}_f\underline{\widetilde{g}}^*_f$ can be found as:

$$E\{\underline{\widetilde{g}}_f\underline{\widetilde{g}}^*_f\} = \frac{1}{2}\int_{-1}^{1}\frac{d(\cos\theta)}{(4\pi)^2(R^2+r^2-2Rr\cos\theta)} = \frac{1}{(4\pi)^2 2Rr}\ln\left|\frac{R+r}{R-r}\right|. \quad (16)$$

With (14) and (16), one obtains:

$$E\{\underline{\widetilde{P}}_{fd}\underline{\widetilde{P}}^*_{fd}\} = \frac{\Pi\rho c}{4\pi R^2}\frac{R}{2r}\ln\left|\frac{R+r}{R-r}\right|. \quad (17)$$

Having determined the first term in (10), we proceed with the similar mathematical analysis to determine $E\{\underline{\widetilde{P}}_{fd}\underline{P}^*_{fd}\}$. In this case, we find

$$E\{\underline{\widetilde{g}}_f\underline{g}^*_f\} = \frac{1}{2}\int_{-1}^{1}\frac{e^{ik(\sqrt{R^2+r^2-2Rr\cos\theta}-R)}d(\cos\theta)}{(4\pi)^2R\sqrt{R^2+r^2-2Rr\cos\theta}} = \frac{1}{(4\pi R)^2}\frac{\sin(kr)}{kr} \quad (18)$$

leading to

$$E\{\underline{\widetilde{P}}_{fd}\underline{P}^*_{fd}\} = \frac{\Pi\rho c}{4\pi R^2}\frac{\sin(kr)}{kr}. \quad (19)$$

Finally, returning again to the expression (10), we utilize the well-known formula for the normalized correlation function of the complex sound pressure amplitude at two points separated by a distance r in space [4]

$$\frac{E\{\underline{\widetilde{P}}_{fr}\underline{P}^*_{fr}\}}{E\{\underline{P}_{fr}\underline{P}^*_{fr}\}} = \frac{\sin(kr)}{kr}. \quad (20)$$

Because the second term in (10) turns out to be a real function of r, the same result must be valid for the third term in (10). These results can now be inserted in (10) to obtain an approximation solution to the power of the error signal at single frequency.

Acknowledgment: This work was partially supported by the Australian Research Council.

6. REFERENCES

[1] J. Mourjopoulos, “On the variation and invertibility of room impulse response functions”, *Journal of Sound and Vibration*, vol. 102, no. 2, pp. 217-228, 1985.

[2] M.R. Schroeder, “Frequency correlation functions of frequency responses in rooms”, *J.Acoust.Soc.Amer.*, vol. 34, no. 12, pp. 1819-1823, 1962.

[3] M. Kendall and A. Stuart, *The advanced theory of statistics (vol.1)*. Charles Griffin Co.Ltd, London, 1976, 4th ed.

[4] R.K. Cook, R.V. Waterhouse, R.D. Berendt, S. Edelman and M.C. Thompson, “Measurement of correlation coefficients in reverberant sound fields”, *J.Acoust.Soc.Amer.*, vol. 27, no. 6, pp. 1072-1077, 1955.

[5] P.M. Morse and K.U. Ingard, *Theoretical Acoustics*. McGraw-Hill, London, 1968.

[6] J.B.Allen and D.A.Berkley, “Image method for efficiently simulating small-room acoustics”, *J.Acoust.Soc.Amer.*, vol. 65, no. 4, pp. 943-950, 1979.