ROOM IMPULSE RESPONSE VARIATION DUE TO THERMAL FLUCTUATION AND ITS IMPACT ON ACOUSTIC ECHO CANCELLATION

Gary W. Elko[1], *Eric Diethorn*[1] *and Tomas Gänsler*[2]

[1]Avaya Labs, 233 Mount Airy Rd, Basking Ridge, NJ 07920, USA
[2]Agere Systems, 555 Union Blvd, Allentown, PA 18109, USA
gwe@avaya.com, ejd@avaya.com, gaensler@agere.com

ABSTRACT

In the field of acoustic echo cancellation, it is well known that the amount of cancellation attainable in real-world systems is limited. Euphemistically this limitation has been referred to in the acoustic echo cancellation community as the “20 dB rule” [1]. This rule reflects the observation that, typically, one can obtain only about 20 to 30 dB of acoustic echo cancellation in actual physical settings.

The acoustic echo cancellation problem is generally approached as a linear systems identification problem. As such, one would expect that the acoustic coupling between the loudspeaker and microphone could be well identified and therefore result in very large cancellation values. Unfortunately, this is not the case as there are practical issues that significantly limit the amount of attainable cancellation. This paper investigates the effect of room thermal fluctuations on the loudspeaker-microphone impulse response. Both a theoretical model and actual experimental measurements are given that show thermal variations on the order of tenths of a degree Centigrade can lead to surprisingly large variations in the room impulse response. Thus, small thermal variations can be a limiting physical factor in acoustic echo cancellation performance.

1. INTRODUCTION

The acoustic echo cancellation problem is usually modelled as a linear systems identification problem. As such, the acoustic coupling between the loudspeaker and microphone can be well identified and therefore allow very large cancellation values. Unfortunately, this is not the case as there are the practical issues of room noise, time-varying coupling paths, distortion, amplifier noise, A/D quantization noise and improper double-talk detection schemes.

The impulse response between a loudspeaker and microphone varies depending on many factors, such as: transducer directivity, frequency response of the transducers, position, room geometry, acoustic treatment, temperature, humidity, and air flow, to name a few. The purpose of this paper is to quantify the effect of temperature variation on the room impulse response. It is clear that most rooms are inhomogeneous in temperature and that room temperature fluctuates because of heating and cooling systems, heating due to sunlight or electronic equipment and the diffusion of heat into and out of the room’s boundaries. It is with the observation that there are many sources of heat causing local and global variations in temperature that we investigate the effects of these variations on the room impulse response.

2. THERMAL EFFECTS ON ACOUSTIC PROPAGATION

As sound propagates in an inhomogeneous medium where the inhomogeneity is due to a variation in local temperature, the net effect will be a fluctuation in sound speed due to the local air density fluctuations along with secondary effects including absorption, refraction and diffraction. That sound speed changes with temperature is a result of the fluid dynamic equation that is typically referred to as “the equation of state” [2]. This equation relates the variables of fluid pressure and density. The equation of state can be generally written as

$$p_o + p' = p(\rho_o + \rho', s_o) \tag{1}$$

where p and ρ are the ambient fluid pressure and density respectively, and p' and ρ' are the fluctuating acoustic contributions. The specific entropy term, s_o, indicates that the entropy is constant and that the ambient medium is homogeneous. The linearized equation of state can be obtained by using a Taylor series expansion of (1),

$$p' = c^2\rho' \quad c^2 = \frac{\partial p}{\partial \rho} \tag{2}$$

where c is the speed of sound. For sound propagation in gases (the problem we are interested in for this paper), we can make the Laplace adiabatic assumption and write,

$$c^2 = \frac{\gamma p}{\rho} \tag{3}$$

where γ is the ratio of specific heats c_p/c_v, equal to 1.4 for air. If we finally apply the ideal gas equation $p = \rho RT$, we obtain the speed of sound as a function of absolute temperature, T,

$$c = \sqrt{\gamma RT} \tag{4}$$

where $R = c_p - c_v$ and is a constant equal to 287 for dry air. For dry air we can substitute in the appropriate constants and write

$$c \approx 20.03\sqrt{273.15 + T_C} \tag{5}$$

where T_C is the temperature in Celsius. Equation (5) is used in the numerical simulations performed and discussed in the next section.

3. NUMERICAL COMPUTER SIMULATIONS

This section describes a simple computer experiment to investigate the effect of temperature on room impulse responses. For rooms having a simple parallelepiped geometry, the impulse response between an omnidirectional source and receiver can be computed

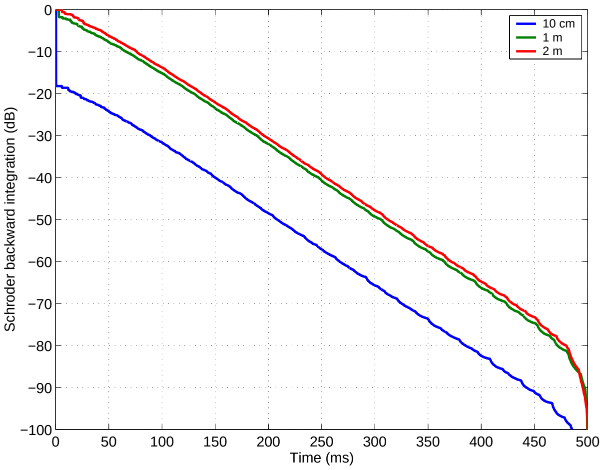

Fig. 1. Image model room impulse response with 10 cm, 1 m, and 2 m spacing between the microphone and loudspeaker.

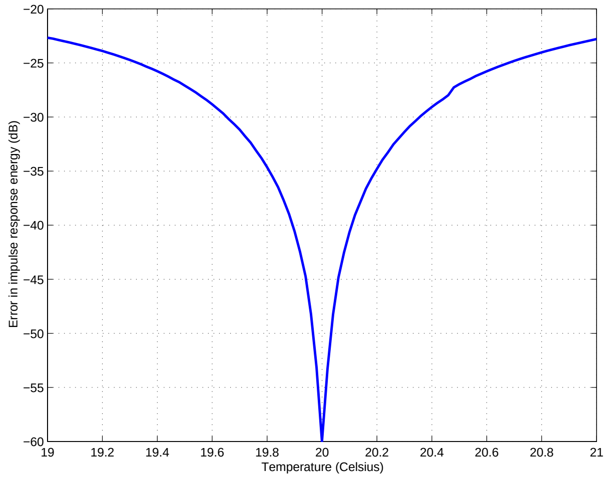

Fig. 2. Predicted misalignment error as a function of temperature for 10 cm microphone to loudspeaker spacing.

easily by using the image method initially described by Allen and Berkley [3]. One practical limitation in the Berkley-Allen method is the quantization rounding used in computing the impulse response. Temporal quantization of the sampling time interval is done to reduce the memory and computational requirements in the method. Each image contribution is computed exactly in time and then rounded to the closest sampling time. The effect of this quantization on the quality of a computed monophonic reverberation is imperceptible in most cases. A simple solution to this problem is to compute the image model response at a very high sampling rate and then downsample to a lower rate. For the numerical experiments a 1 MHz the sampling rate is chosen and the resulting impulse responses downsampled to 10 kHz. The only added computational cost in this approach is the one-time downsampling of the computed impulse response.

The parameters of the image method are set to closely match the room size and global acoustic absorption of a typical business office. The room acoustic energy decay is typically plotted using the backward integration method of Schroeder [4]. Schroeder showed that the backward integration technique results in a statistically stable estimate of reverberation decay. The backward integration method is defined mathematically as,

$$r(\tau) = \int_{\tau}^{\infty} p^2(t)dt \qquad (6)$$

where p(t) is the measured pressure decay of an acoustic signal that is shut-off after the room reaches “steady-state”. Figure 1 shows the Schroeder backward integrated reverberation decay for the cases of a 10 cm, 1 m, and 2 m loudspeaker microphone spacing. The rapid decay at 500 milliseconds is due to limiting the impulse response to this length and is characteristic of the backward integration method using truncated impulse responses.

The larger direct path for the 10 cm data, can be seen in Figure 1 as a rapid change at zero on the time axis. From this figure we can estimate that the direct to total reverberant ratio is close to 20 dB (by the 20 dB increase at the time origin).

The effect of temperature change on the impulse response is a result of the sound speed change. Since later reverberant field contributions propagate over a much longer distance by definition, they undergo the largest relative timing changes. Thus one would intuitively expect that the effect of temperature variability will be greater for larger microphone-to-loudspeaker distances and for more reverberant rooms.

In order to quantify the effect of the impulse response change, the following misalignment error measure is used:

$$J = \frac{\left[\mathbf{h} - \hat{\mathbf{h}}\right]^T \left[\mathbf{h} - \hat{\mathbf{h}}\right]}{\mathbf{h}^T \mathbf{h}} \qquad (7)$$

where $\mathbf{h} = [h_0, h_1, ...h_{L-1}]^T$ is the impulse response of length L at some reference temperature, and $\hat{\mathbf{h}}$ is the computed impulse response at some other temperature.

Numerical results computed with the image method at temperatures between 19° C and 21° C are investigated. The 20° C reference temperature was arbitrarily selected, and the misalignment for this case (by definition $-\infty$) set to -60 dB for graphical purposes. Numerical temperature experiments for two different microphone loudspeaker spacings are shown in Figs. 2 and 3. Figure 2 shows the results for the case of 10 cm loudspeaker-to-microphone spacing and 3 shows the results for a 2 m microphone-to-loudspeaker spacings. As expected, one can see that as the distance decreases, the misalignment becomes smaller. It should also be noted that the results show that relatively large changes occur in the room impulse response as a function of temperature. For instance, in Fig. 3, the misadjustment error is only 6 dB smaller than the actual filter tap weight power for the room at 1°C warmer or cooler. Larger scale fluctuation of temperature in rooms typically occurs at a relatively slow rate. However, it can be reasonably argued that small localized fluctuations in temperature can be relatively rapidly changing. By the term relative, we mean time scales that are on the order of the length of the room impulse response. Figure 3 shows that even small changes in temperature can result in large misadjustment errors. For instance, the data for the 2 m spacing shows that a 0.1° C change results in a misadjustment that is only 25 dB below the tap weight power.

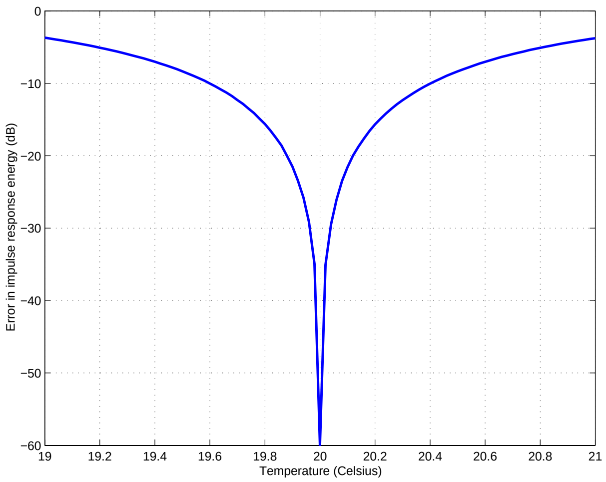

Fig. 3. Predicted misalignment error as a function of temperature for 2 m microphone to loudspeaker spacing.

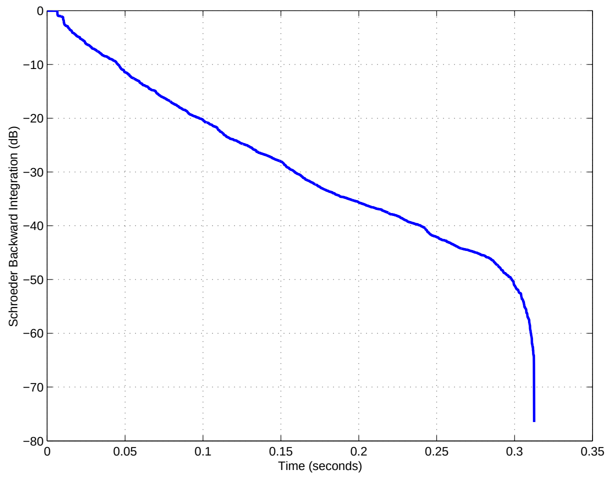

Fig. 4. Schroeder backward integrated room impulse response with 2 m spacing between the microphone and loudspeaker.

4. EXPERIMENTAL MEASUREMENTS

In order to verify the computer model results, sequential measurements of real room impulses were made. Although this is not the corresponding experiment as presented in the numerical results, thermal fluctuations should be noticeable by a time varying misadjustment, where the misadjustment increases as the measurement time grows. Sequential measurements should also give us some idea of the time constants involved in room temperature changes.

These measurements must be made with care since there are many pitfalls that can obscure the effect that we are trying to measure. Measurement of the impulse response of the room is a standard linear systems identification problem. In order to simplify the experiment, the measured impulse response includes the loudspeaker and microphone responses. The inclusion of the loudspeaker and microphone convolutional effects are not expected to be significant since their impulse responses are relatively short compared to the room. Also, although the electroacoustic transducer responses do change with temperature, these changes are much smaller than the room acoustic changes.

A DSP Technology Siglab Model 20-22 is used to measure the system impulse responses. The Siglab unit uses a repetitive chirp signal as an excitation signal. The chirp signal is ideal since it has a very low peak-to-RMS power ratio, which is desirable for sampled data systems that have analog-to-digital (AD) and digital-to-analog (DA) conversion quantization and dynamic range limitations. The Siglab system has 16-bit AD and DA converters giving a measurement dynamic range in excess of 96 dB. A repetitive chirp signal can be used to average-out the effects of both electrical and acoustic measurement noise. However, since relatively rapid time scale measurements are desired, the amount of averaging has to be limited. Measurements for this paper were made using two realizations giving a 3 dB increase in signal-to-noise for the measurements. Another measurement requirement in using a repetitive chirp stimulus to estimate the impulse response is that the chirp length must be longer that the impulse response of the system under test. The Siglab system has a maximum chirp length of 8192 points. The Schroeder backward integration of the room impulse response is shown in Fig. 4. Figure 4 shows that the reverberation decay is about -50 dB at 300 milliseconds. Thus, the room reverberation time (length of time for the room impulse response to decay 60 dB) is about 360 milliseconds. With a block size of 8192 and a sampling rate of 12.8 kHz, we can accurately measure impulse responses that are less than 650 milliseconds. Since the room used for the experiments has a reverberation time of 360 milliseconds, we fall well within the measurement length limitation. Care must be taken in interpreting the measured data since it includes the filtering effect of the loudspeaker and microphone. A FOSTEX 6301B personal monitor loudspeaker was used in the measurements which has a low-frequency limit of about 100 Hz. The measured results are only reliable in the range where the loudspeaker has adequate signal output relative to the room background noise.

Figure 4 shows another issue that must be dealt with in real measurements: the effect of background noise. Figure 4 is computed by the Schroeder method of backward integration. This method is highly susceptible to room background noise and, as such, many extensions have been suggested to make it more robust. A standard technique is to truncate the impulse response at a point where the logarithmic decay remains approximately linear (exponential decay). The effect of room noise results in a levelling off of the decay as the decay level falls into the room noise. Figure 4 shows that the noise is about 45 dB below the measurement signal. This is a rough guess since we can see that the decay slope is starting to level off at about -35 dB decay point. A rule-of-thumb for the Schroeder method is that decay slopes start to show deviation from linear decay at values about 10 dB higher than the noise floor. Thus, we can conclude that the noise floor is about 45 dB below the measurement level. It should be pointed out that this approximation is based on experimental observation, readers more interested in this issue should refer to papers on this subject [5]. Measurements of the room noise indicate an SNR of 52 dB, where the noise includes the effects of nonlinear harmonic distortion from the loudspeaker and power-line harmonics. It is inter-

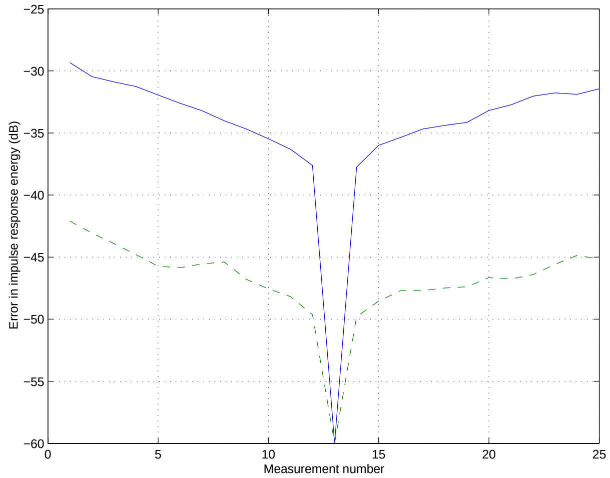

Fig. 5. Measured sequential variation in the room impulse response for 10 cm (dash) and 2 m (solid) microphone-loudspeaker spacings. The reference measurement is the center measurement.

esting to note that the harmonic distortion terms are more than 50 dB below the fundamental at 1 kHz. It should not be expected that loudspeakers used in a speakerphone would have distortion values this low. We have used a relatively low-distortion loudspeaker for these measurements since we want to observe the effects of room impulse response change and high-distortion systems could obscure these measurements. For 2 m spacing between the microphone and the loudspeaker the computed SNR is 34 dB. With the use of two averages in the chirp impulse response measurements we can expect the SNR to be about 37 dB. Since the measured impulse response is essentially covered in one-half the block size of the measurement, truncating the data to the first-half of the time domain results in another SNR gain of 3 dB giving about 40 dB of expected SNR for measurement from 2 m.

Measurements of the misadjustment in sequential measurements for 10 cm and 2 m spacings are shown in Fig. 5. The measurements were made in rapid succession using a chirp time record of 640 milliseconds (8192 samples at 12.8 kHz). Each measurement consists of 2 time averages and the Siglab box discards the first chirp of an average. Also, there is some additional delay due to transferring the data from the Siglab unit to the host PC and delay to setup the next measurement. These two delays add approximately 500 milliseconds to the measurement period. Therefore each measurement shown in Fig. 5 take approximately 2 seconds to complete. For the 10 cm microphone distance in Figure 5, we do indeed find experimental data that the impulse response is varying in time and that the variation increases with time. If we examine the misadjustment error level and compare with the image model results for this case as shown in Fig. 2, it appears the the local temperature is fluctuating at about 0.05° C (around 45 dB down). If we select the same level of fluctuation for the 2 m distance, and examine Fig. 3, we could conclude that the level of misadjustment for this case would be approximately -30 dB. If we look at the measured response, the actual measured misadjustment level is approximately -30 dB. Agreement between the image model and the experimentally measured misadjustment is remarkable.

One caveat to the above discussion is that we really do not know the actual temperature fluctuation in the room and we have assumed it from the measured data. Another issue that remains unclear, is how the actual room temperature is dynamically changing. Clearly one can not expect the temperature fluctuations in a room to be homogeneous. In order to understand this, we would need to do a more careful experiment where we can accurately measure and control the temperature in a room that can be brought into thermal equilibrium (thermally homogeneous). Another issue is how much misadjustment is due to transducer changes. The microphone response is relatively insensitive to temperature. However, the loudspeaker will have some temperature variation due to voice coil heating. The experiments made above attempted to remove this variable by driving the loudspeaker for a few minutes before beginning the measurements.

5. CONCLUSIONS

This paper has shown that small temperature variations in rooms can lead to relatively large changes in the impulse response between a loudspeaker and microphone. The effect of temperature variation diminishes as the distance between the microphone and loudspeaker becomes smaller. This latter result should not be too surprising, since sound speed changes caused by temperature variations more significantly alter the propagation times of longer reverberation paths.

The thermal effect on the impulse response is demonstrated both computationally, using the image method, and experimentally. The overall conclusion is that one can expect that thermal fluctuations are yet another issue that affect real-world acoustic echo cancellation performance. For quiet rooms, this effect could be the dominant factor limiting maximum attainable acoustic echo return loss enhancement (AERLE). If a threshold for the maximum AERLE can be set due to knowledge of thermal variability and/or room background noise, one can establish a minimum filter length whereby adding extra length will not increase the canceler performance. Knowledge of the minimum filter length set by these practical limitations allows for the best echo canceler possible with a minimum of computational cost.

6. REFERENCES

[1] M. E. Knappe, and R. A. Goubran, "Steady State Performance Limitations of Full-Band Acoustic Echo Cancelers", ICASSP 1994, Adelaide, Australia, Vol. 2, pp. 73-76.

[2] P. M. Morse, K. U. Ingard, **Theoretical Acoustics** Princeton University Press, 1968.

[3] J. B. Allen, and D. A. Berkley, "Image Method for Efficiently Simulating Small Room Acoustics," J. Acoust. Soc. Am. 65-4, 943-949, 1967.

[4] M. R. Schroeder, "New method of measuring reverberation time, J. Acoust. Soc. Am. 37, 409412 1965.

[5] N. Xiang, "Evaluation of Reverberation Times using a Nonlinear Regression Approach", J. Acoust. Soc. Am., 98, 2112-2121, 1995.