ADAPTIVE DISTANCE AND NEAR-FIELD COMPENSATION APPLIED TO MICROPHONES

Walter Etter

Bell Laboratories
Alcatel-Lucent
Murray Hill, NJ, 07974, USA
walter.etter@alcatel-lucent.com

ABSTRACT

In voice acquisition, variations of the microphone distance introduce not only level changes, but also frequency response changes due to the near-field effect. This paper presents a method for adaptive distance and near-field compensation based on the talker-to-microphone distance and the microphone polar pattern. If available, the microphone orientation and the critical distance associated with the room acoustic can be taken into account to further improve compensation accuracy. Aimed at teleconference use, the significance of the critical distance for compensation is discussed for office and conference rooms. An example for the performance of the algorithm is provided, in which a sensor is applied to continuously measure a varying microphone distance.

Index Terms— Near-field compensation, proximity effect, automatic gain control, microphone, critical distance

1. INTRODUCTION

In a conference room, high quality voice acquisition is often degraded by ambient noise and reverberation. These effects can be significantly reduced by picking up the voice signal close to the talker's mouth; for example, by using separate directional microphones for individual talkers. Such arrangements are widely used and realized, e.g., by mounting the microphone on an extension arm for close placement to the talker's mouth, or by installing boundary microphones on the surface of the table at individual seat positions. While close microphone placement does improve sound quality significantly, it introduces a new problem: for an assumed expected radial range of the talker position, a closer microphone will result in larger relative microphone distance variations, which, in turn, lead to larger microphone signal variations.

As a result of close microphone placement, the microphone signal requires considerable level equalization, typically performed by automatic gain control (AGC) or compression, both of which degrade sound quality.

For many decades, researchers focused on reducing artifacts produced by AGC and compression methods. Research addressed dynamic distortion [1], harmonic distortion [2], and aliasing [3], among others. The detrimental effect of compression has also been confirmed in formal listening tests [4]. These artifacts unnecessarily limit the sound quality and impede an attempt to achieve a realistic auditory experience in telepresence or virtual presence.

While conventional AGCs derive the reference signal from the microphone signal itself, we present a distance-based compensation that derives its reference signal from the talker-to-microphone distance. This principle completely avoids the drawback of conventional signal-based AGCs. For a static situation, where the talker-to-microphone distance remains unchanged, the distance-based compensation gain remains fixed, resulting in a stable sound source image. In contrast, a signal-based AGC still changes the gain constantly to adjust for a continuously changing reference signal (typically the signal's RMS or envelope), whether the talker-to-microphone distance is fixed or varying.

Furthermore, the time-constants of a signal-based AGC always reflect a compromise. With shorter time constants, the signal gain reaches its target faster, but speech quality suffers. With longer time constants, speech quality improves, but speech may no longer be adjusted quickly enough and signal levels may end up too high or too low for periods in order of seconds. In contrast to signal-based AGCs, no compromise is necessary for a distance-based level compensation. The gain is adjusted immediately, without degrading sound quality.

Like short-term sound pressure level variations, long-term variations contribute equally to a realistic auditory experience. When a talker intentionally raises his voice, a signal-based AGC reduces the signal level, whereas for a whisper, the level is increased. Both actions defy the intentions of the talker and stand in the way of a convincing auditory experience. Distance-based compensation does not suffer from these imperfections.

The concept of adaptive distance and near-field compensation was introduced recently [5], but evaluated only conceptually at a number of fixed pre-determined distances without using continuous distance measurements from a sensor. Furthermore, the distance compensation was based on the average absorption coefficient. However, for practical reasons, it would be preferable if such compensation were solely based on the critical distance (the radius or distance from a spherical sound source at which the sound pressure levels of direct and diffuse sound fields are equal). In the following sections, we present results that address these shortcomings.

The paper is organized as follows: Section 2 derives the gain for the distance compensation with the critical distance as the sole acoustic room parameter. Section 3 reviews distance-adaptive near-field compensation and introduces a normalized near-field compensation that accounts for pre-configured static proximity-effect compensation. Section 4 provides two room examples, the counterparts in a teleconference, and discusses the relevance for room-specific distance compensation. Finally, Section 5 shows an example for the compensation, where a variable talker-to-microphone distance is continuously measured by a distance sensor.

2. DISTANCE COMPENSATION

Focusing on voice transmission, we note that the spatial sound source characteristic of a human voice can be approximated by a spherical sound field [6]. If we consider a harmonic wave (i.e., a sine wave), we can express the sound pressure $p(r)$ resulting from a spherical sound source by its sound pressure $p(r_0)$ at radius r_0 as follows [7]:

$$\frac{p(r)}{p(r_0)} = \frac{r_0}{r} e^{-jkr}, \tag{1}$$

where k denotes the wavenumber ($k = \omega/c$) and c the speed of sound. Equation (1) shows that if we double the radius r, the sound pressure drops to half of its original value. Equation (1) assumes free-field conditions, i.e., only a direct sound field is generated. However, if the sound source is located in a room, two sound fields are produced: the direct sound field from the direct sound of the source, and the diffuse sound field from reflected sound.

The pressure ratio for a mixed direct and diffuse field can be expressed in terms of the average absorption coefficient $\bar{\alpha}$ and the room surface area S as

$$\frac{p(r)}{p(r_0)} = \sqrt{\frac{R + 1/r^2}{R + 1/r_0^2}} \tag{2}$$

according to [5], where

$$R = \frac{16(1-\bar{\alpha})\pi}{S\bar{\alpha}}. \tag{3}$$

Furthermore, the critical distance can be derived from the average absorption coefficient [5] with

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

Using (2), (3), and (4), we obtain

$$\frac{p(r)}{p(r_0)} = \sqrt{\frac{1/r_c^2 + 1/r^2}{1/r_c^2 + 1/r_0^2}}. \tag{5}$$

If $r_c \gg r_0$, we arrive at the following approximation:

$$\frac{p(r)}{p(r_0)} \approx \sqrt{r_0^2/r^2 + r_0^2/r_c^2}. \tag{6}$$

The first term in the sum is produced by the direct sound field, while the second term is produced by the diffuse sound field. The direct sound field contribution decreases with increasing distance, while the diffuse sound field contribution remains constant. To show the relationship to (1), we rewrite (6) as

$$\frac{p(r)}{p(r_0)} \approx \frac{r_0}{r}\sqrt{1 + r^2/r_c^2}. \tag{7}$$

To obtain the gain for the distance compensation, we take the inverse of the pressure ratio in (7), i.e.,

$$G(r) \approx \frac{r}{r_0}\sqrt{\frac{1}{1 + r^2/r_c^2}}. \tag{8}$$

Fig. 1 illustrates the resulting gain curves for parameter r_c.

The approximation in (6) can be applied in a wide range for $r_c \gg r_0$. Even for $r_c = 40$ cm and $r_0 = 20$ cm, an extreme case for the critical distance, the gain error caused by the approximation in (6) is still smaller than 1 dB.

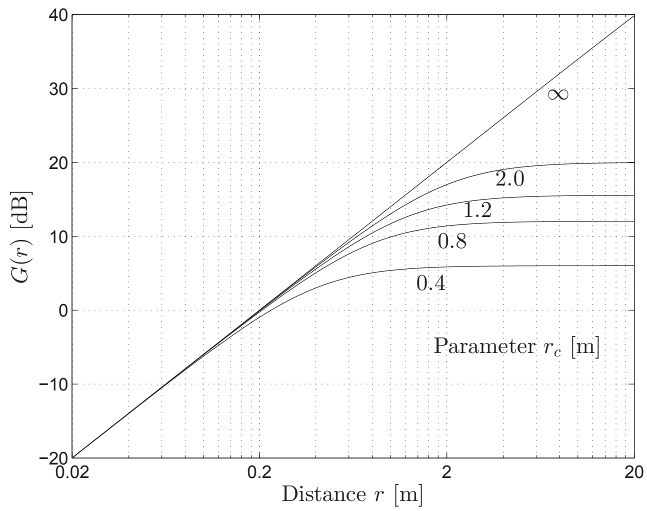

Figure 1: Desired distance compensation gain $G(r)$ for reference microphone distance $r_0 = 20$ cm and critical distance r_c.

If the talker-to-microphone distance stays well within the critical distance, i.e., $r \ll r_c$, we can neglect the diffuse sound field, in which case (8) reduces to $G(r) \approx r/r_0$.

3. NEAR-FIELD COMPENSATION

The near-field effect, also known as proximity effect, causes a gradual increase of low frequency output as a pressure-gradient microphone approaches a spherical sound source.

Recall that the polar pattern $R(\theta)$, a function of the sound incident angle θ, can be denoted according to [9] as

$$R(\theta) = a + b \cdot cos\theta, \tag{9}$$

where parameters $[a, b]$ (with $a \leq 1$, $b \leq 1$, and $a + b = 1$) specify the directivity of the microphone. The omnidirectional microphone is determined by [1, 0], the cardioid by [0.5, 0.5], the supercardioid by [0.37, 0.63], the hypercardioid by [0.25, 0.75], and the figure-8 by [0, 1]. Interpreting (9), we can represent any 1st order pressure-gradient microphone as a weighted sum of two microphone signals: the first from an omnidirectional microphone with weighting factor a, and the second from a figure-8 microphone with weighting factor b.

A general formulation of the proximity effect in terms of a, b, and θ is derived in [5] and stated as

$$H(kr) = a + b\cos\theta - j\frac{b\cos\theta}{kr}. \tag{10}$$

The corner frequency for this 1st order filter is

$$f_c = \frac{c}{2\pi r} \cdot \frac{b\cos\theta}{a + b\cos\theta}, \tag{11}$$

where c is the speed of sound.

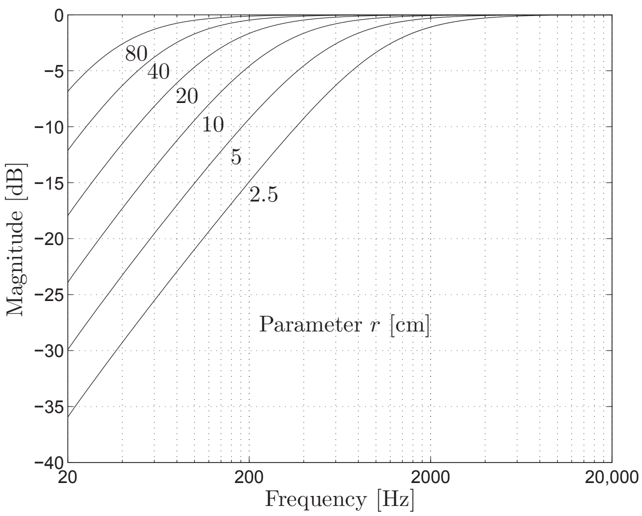

Figure 2: Magnitude response of a near-field compensation filter for a cardioid microphone shown at various talker-to-microphone distances r and fixed sound incident angle $\theta = 0^0$.

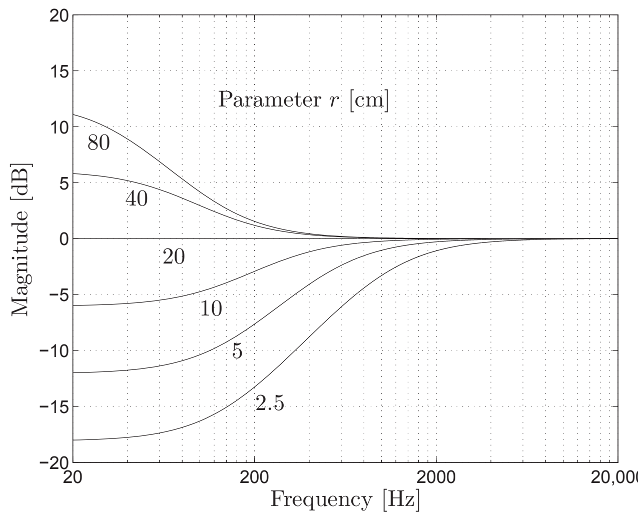

Figure 3: Magnitude response of a normalized near-field compensation filter for a cardioid microphone with pre-configured static proximity-effect filter for nominal distance r_n=20cm, shown at various talker-to-microphone distances r and fixed sound incident angle $\theta = 0^0$.

Taking the inverse of $H(kr)$, we can now specify the near-field compensation (NFC) filter as

$$C(kr) = \frac{kr}{kr(a + b\cos\theta) - jb\cos\theta}. \tag{12}$$

The magnitude response of this compensation filter for a cardioid microphone is shown in Fig. 2. As expected, a shorter microphone distance results in a higher corner frequency of the high-pass filter.

Using (8) and (12), we arrive at the total compensation filter by multiplying the contribution of distance and near-field compensations:

$$G_{tot}(kr) = G(r) \cdot C(kr). \tag{13}$$

A directional microphone may have been designed for use at a nominal distance r_n with compensation for the proximity effect at that and only that specific distance. To still provide a wide distance-range proximity-effect or near-field compensation, a normalized NFC filter of the form

$$\tilde{C}(kr) = C(kr)/C(kr_n) \tag{14}$$

can be utilized. Fig. 3 shows the magnitude response of a normalized NFC filter for a nominal distance $r_n = 20$ cm.

4. TELECONFERENCE ROOM ACOUSTIC

Table 1 presents two hypothetical room examples: a conference room and an office, typical counterparts in a teleconference. With these examples, we illustrate the relevance of the critical distance for distance compensation in teleconference applications.

Given the reverberation time T, we can use the Eyring-Norris formula [7],

$$T = 0.161\frac{V}{-S\ln(1-\bar{\alpha})}, \tag{15}$$

where V denotes the room volume and S the total surface area, to determine the average absorption coefficient $\bar{\alpha}$,

$$\bar{\alpha} = 1 - e^{-0.161V/(S \cdot T)}. \tag{16}$$

Parameter	Room 1 (Conference Room)	Room 2 (Office)
Size (L x W x H)	8.3m x 7.4m x 2.8m	4.8m x 4.3m x 2.6m
Volume V	172 m^3	54 m^3
Area S	211 m^2	89 m^2
Reverb time T	300 ms / 430 ms	203 ms / 330 ms
Absorption $\bar{\alpha}$	0.36 / 0.26	0.38 / 0.26
Critical dist. r_c	1.52 m / 1.22 m	1.04 m / 0.78 m

Table 1: Examples for conference room and office with parameters shown for acoustically treated (left) and untreated (right) rooms.

The critical distance is derived by setting the energy density of the direct and the diffuse sound field equal [5], which results in

$$r_c = \frac{1}{4}\sqrt{\frac{S\bar{\alpha}}{\pi(1-\bar{\alpha})}}. \tag{17}$$

A measured reverberation time T available, we can calculate $\bar{\alpha}$ from (16), and then use (17) to calculate r_c.

To assess the range for expected critical distance values, we consider both acoustically treated and untreated versions of the rooms in table 1. For the treated, i.e., acoustically well designed versions, we require that they meet ITU-R BS.1116-1 [8]. This standard provides the following recommendations for the reverberation time:

$$T = \frac{1}{4}\left(\frac{V}{V_0}\right)^{1/3}, \tag{18}$$

where $V_0 = 100\text{m}^3$. With appropriate acoustical treatment of the rooms, we can satisfy (18), and obtain the critical distances shown in table 1 (left entry).

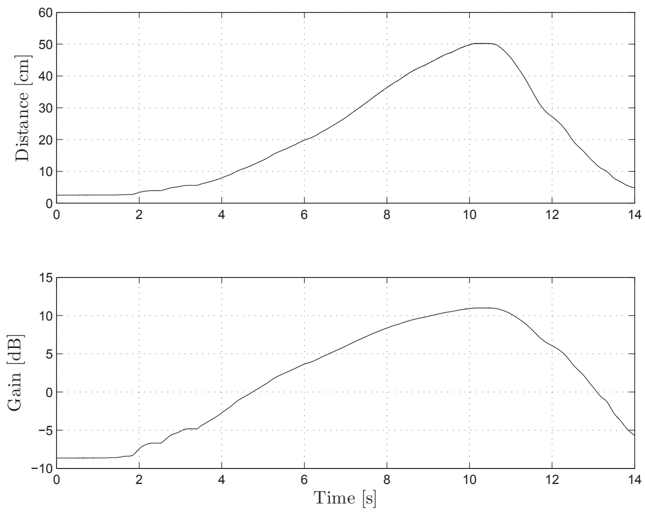

Figure 4: Estimated microphone distance (top) and computed adaptive compensation gain (bottom).

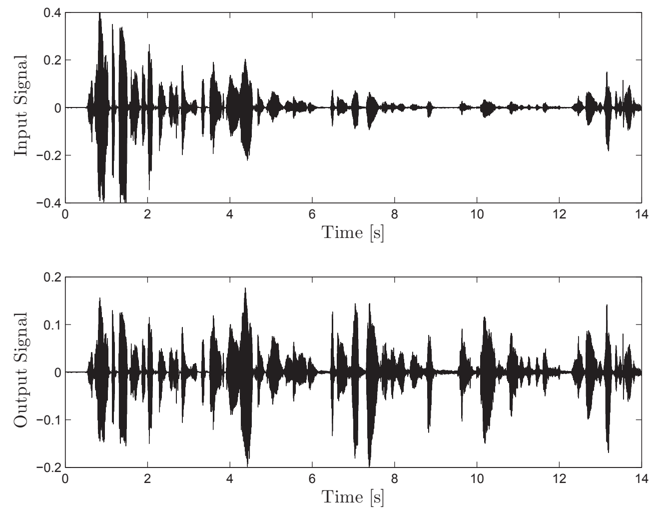

Figure 5: Input signal (top) and output signal (bottom) of compensation.

Table 1 also shows the case of acoustically untreated rooms with insufficient absorption (right entry). These examples do not meet the reverberation time constraint of (18).

Based on the critical distance and the expected talker-to-microphone distance variations, we are able to make a decision as to whether we need to take the diffuse sound field in (8) into account. For example, if the expected maximum for the talker-to-microphone distance is 1 m (a reasonable assumption for a setup with individual table microphones, but unlikely for a setup with a single table microphone), we can neglect the diffuse part in room 1, for both the acoustically treated and untreated room. For room 2, however, we can neglect it only for the acoustically treated room.

If the room parameters are unknown, the critical distance can still be estimated from the microphone signal [10]. If neither estimate is available, it is best to take a conservative assumption on the lower end, for example r_c=0.8 m. This way, the effective distance range will possibly be reduced, while the close vicinity of the microphone is still actively controlled.

5. EXPERIMENTAL RESULTS

For brevity, the simple case of an omnidirectional microphone is shown here, for which the NFC contribution in (13) vanishes. The talker-to-microphone distance was continuously varied, while being estimated with a Polhemus Patriot motion tracker. From the estimated distance r and applied source size compensation [5], the distance compensation gain was computed (see Fig. 4) and applied to the input signal (that is, the microphone signal) to produce the output signal (see Fig. 5). Informal listening tests confirmed the superior sound quality of the compensation.

6. CONCLUSION

With professional audio quality in mind, we have proposed adaptive distance and near-field compensation and shown its application to sound acquisition. As accurate close-range sensors for distance measurement become widely available at acceptable cost and size, adaptive distance and near-field compensation will be applicable to speech acquisition in numerous fields.

7. REFERENCES

[1] B. A. Blesser, “Audio dynamic range compression for minimum perceived distortion,” *IEEE Trans. Audio and Electroacoustic*, vol. 17, no.1, March 1969, pp. 22-32.

[2] F. Floru, “Attack and release time constants in feedback compressors” *J. Audio Eng. Soc.*, vol. 47, no. 10, Oct. 1999, pp. 788-804.

[3] P. Kraght, “Aliasing in digital clipper and compressors,” *J. Audio Eng. Soc.*, vol. 48, no. 11, Nov. 2000, pp. 1060-1065.

[4] W. M. Wagenaars, A. J. M. Houtsma, and R. A. J. M. Van Lieshout, “Subjective evaluation of dynamic compression in speech,” *J. Audio Eng. Soc.*, vol. 34, no. 1/2, Jan./Feb. 1986, pp. 10-18.

[5] W. Etter, “Distance-based automatic gain control with continuous proximity-effect compensation,” presented at *the 133rd Convention of the Audio Eng. Soc.*, Oct. 26-29, 2012.

[6] J. L. Flanagan, “Analog measurement of sound radiation from the mouth,” *J. Acoust. Soc. Am.*, vol. 32, no. 12, Dec. 1960, pp. 1613-1620.

[7] L. E. Kinsler, et. al., “Fundamentals of acoustics,” Wiley, 4th ed., New York, 2000.

[8] –, ”Methods for the subjective assessment of small impairments in audio systems including multichannel sound systems,” Recommendation ITU-R BS.1116-1, 1997.

[9] R. A. Rayburn, “Eargle’s microphone book,” Focal, 3rd ed., 2012, pp. 78, 64.

[10] J. J. Jetzt, “Critical distance measurement of rooms from the sound energy spectral response,” *J. Acoust. Soc. Amer.*, vol. 65, no. 5, May 1979, pp. 1204-1211.