Audio Engineering Society

Convention Paper 8786

Presented at the 133rd Convention
2012 October 26–29 San Francisco, USA

Distance-Based Automatic Gain Control with Continuous Proximity-Effect Compensation

Walter Etter[1],

[1] *Bell Labs, Alcatel-Lucent, Murray Hill, NJ, 07974, USA*

Correspondence should be addressed to Walter Etter (walter.etter@alcatel-lucent.com)

ABSTRACT

This paper presents a method of Automatic Gain Control (AGC) that derives the gain from the sound source to microphone distance, utilizing a distance sensor. The concept makes use of the fact that microphone output levels vary inversely with the distance to a spherical sound source. It is applicable to frequently arising situations, in which a speaker does not maintain a constant microphone distance. In addition, we address undesired bass response variations caused by the proximity effect. Knowledge of the sound-source to microphone distance permits accurate compensation for both frequency response changes and distance-related signal level changes. In particular, a distance-based AGC can normalize these signal level changes without deteriorating signal quality, as opposed to conventional AGCs, which introduce distortion, pumping, and breathing. Provided an accurate distance sensor, gain changes can take effect instantaneously and do not need to be gated by attack and release time. Likewise, frequency response changes due to undesired proximity-effect variations can be corrected adaptively using precise inverse filtering derived from continuous distance measurements, sound arrival angles, and microphone directivity, no longer requiring inadequate static settings on the microphone for proximity-effect compensation.

1. INTRODUCTION

In voice transmission or recording, speakers often neglect to pay attention to the distance from mouth to microphone as they speak. As a result, microphone output levels can vary greatly. These variations are frequently normalized using closely related techniques known as automatic gain control [1], automatic level control [2], [3], or automatic volume control [4]. Typically, short attack times of a few tens of ms and long recovery times of the order of seconds are applied. Since these techniques reduce the dynamic range of the signal, the process of equalizing speech levels is also known as dynamic range compression [5]-[7]. To avoid amplification of noise, or to reduce the original noise level, dynamic compression can be complemented with dynamic expansion [5], [6], [8].

For many decades, researchers focused on reducing the artifacts produced by AGCs or dynamic range control. Research areas included dynamic distortion [5], [9], [10], harmonic distortion [11], aliasing [12], and quantization-related distortion [6]. The detrimental effect of compression was also confirmed in listening tests [13]; it was observed that sound quality decreases as the degree of compression increases. Furthermore, subjects reported that too much compression made music "restless" and "reverberant". In an attempt to achieve a realistic auditory experience in telepresence, the mentioned artifacts represent a significant obstacle.

The need for AGCs is commonly created by speech acquisition. Inadequate speech levels may be of acoustic origin, such as an undesired speaker-to-microphone distance, or of electric origin, such as an inappropriate microphone pre-amplifier gain setting. The goal of this paper is to find an AGC that avoids causes of acoustic origin.

While conventional AGCs derive the reference signal from the microphone signal itself, we present a distance-based AGC that derives its reference signal from the speaker-to-microphone distance. This principle completely avoids the drawback of conventional signal-based AGCs. In a static situation where the speaker-to-microphone distance remains unchanged, the distance-based AGC 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), regardless whether the speaker-to-microphone distance is fixed or variable.

Furthermore, the time-constants of a signal-based AGC always reflect a compromise. With shorter time constants, the signal gain reaches its target sooner, 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 AGC. The gain is adjusted immediately, without degrading sound quality.

A distance-based AGC requires continuous estimation of the speaker-to-microphone distance. Methods to estimate this distance can be separated into two categories: (1) methods that estimate the distance from the microphone signal itself and (2) methods that use external sensors. The former methods typically rely on the fact that the direct-to-reverberant sound energy ratio decreases as the distance increases and as such, make use of changes in signal features. For example, Georganti et al. [14] use linear prediction residual peaks and skewness of the spectrum to determine speaker-to-microphone distance.

If multiple microphones are available, additional properties can be exploited to estimate the distance to a microphone. Vesa [15] uses two microphones in a binaural setup and exploits the coherence between left and right signals. If at least three microphones are available, sound arrival angles can be determined from correlation [16] and then used for triangulation to determine the sound source location.

Typically, methods that estimate the speaker-microphone distance from the microphone signal itself lead to latency. For example, in [14] a block-size of two seconds is used, a choice based on performance versus block-size analysis. Other drawbacks that make these methods unsuited for distance-based AGC include the necessity to train the algorithm to each room [14], [15], or to each microphone position [15]. Moreover, they can only estimate the speaker-microphone distance in the presence of speech, and typically lack the accuracy required for a distance-based AGC.

For the above reasons, external sensor-based distance measurement is preferable. A distance sensor can be based on video input (face detection), acoustic input (ultrasound), mechanic input (accelerometer), magnetic or other input not perceivable by human senses [17]. Accurate close-range sensors are becoming available. However, current size and cost may still be an obstacle for integration in microphones.

This paper is organized as follows. In Section 2, we derive the gain for the AGC from the inverse-square-law and the critical distance. In Section 3, we specify the filter for the proximity-effect compensation. These theoretical results are then verified in Section 4 and compared with results from a conventional signal-based AGC. Finally, we summarize the results in Section 5.

2. INVERSE-SQUARE-LAW AGC

2.1. Direct Sound Field

To approximate an acoustic wave field, two simple wave models are frequently used: the plane wave, originating from a plane source, and the spherical wave, originating from a point source. The spherical wave model can be applied to many practical sound sources. For example, the human voice closely produces a spherical wave in the frontal hemisphere [18]. Focusing on voice transmission, we assume a spherical sound field in the following derivation.

The sound pressure of a harmonic wave (i.e., a sine wave) can be denoted in complex form as

$$p(t) = \hat{p} \cdot e^{j(\omega t + \varphi_0)} \tag{1}$$

with sound pressure amplitude $\hat{p}$, frequency f, $\omega = 2\pi f$, and phase φ_0. Note, from the theory of Fourier series, we can represent any steady state wave as a linear superposition of sine waves. With the above notation, we can write the wave equation for a spherical wave [19] as

$$\frac{\delta^2 p}{\delta r^2} + \frac{2}{r}\frac{\delta p}{\delta r} = \frac{1}{c^2}\frac{\delta^2 p}{\delta t^2}$$

with radius r (i.e., the distance from the sound source), speed of sound c, and time t. Evaluating the 2nd derivative of p with respect to t, we obtain

$$\frac{\delta^2 p}{\delta r^2} + \frac{2}{r}\frac{\delta p}{\delta r} + k^2 p = 0 \tag{2}$$

where $k = \omega/c$ represents the wavenumber.

Eq. (2) is a homogenous differential equation 2nd order with the solution for the outward moving wave

$$p(r) = p_0 \frac{r_0}{r} e^{-jkr} \tag{3}$$

where p_0 denotes the sound pressure at radius r_0. Eq. (3) shows that if we double the radius r, the sound pressure drops to a half of its original value.

This property can also be derived in an alternate way from the inverse-square-law, which states that the sound intensity I (i.e., the sound energy per unit area) due to a spherical sound source is inversely proportional to the square of the distance from the sound source. Expressed in an equation for a point source emitting energy W, and an assumed sphere with radius r and resulting area $A = 4\pi r^2$ [20], we obtain

$$I = \frac{W}{A} = \frac{W}{4\pi r^2} \tag{4}$$

If the radius r is doubled, the sphere area is quadrupled. As a result, the intensity will drop to 1/4, i.e., the inverse of the squared distance ratio. Furthermore, since the average sound pressure $p = \hat{p}/\sqrt{2}$ is proportional to the square root of the sound intensity I according to

$$p = \sqrt{I \rho_0 c} \tag{5}$$

where ρ_0 is the density and $\rho_0 c$ the acoustic impedance of air, the sound pressure will drop to a half, i.e., 6dB, in accordance with Eq. (3).

To compensate for sound pressure variations caused by source-to-microphone distance variations, we can apply a gain inverse to the sound pressure ratio $p(r)/p_0$ given by Eq. (3). A distance-controlled AGC can be specified in this way as

$$G(r) = G_0 \frac{r}{r_0} e^{jkr} \tag{6}$$

where G_0 denotes the desired nominal gain at radius r_0. Note, $G(r)$ is complex, it adjusts not only the magnitude, but also the phase. Since the phase is linear, it represents simply a delay, the sound delay associated with the distance difference $r - r_0$. If we consider only the magnitude of the gain, we obtain

$$|G(r)| = G_0 \frac{r}{r_0} \tag{7}$$

Eqs. (6) and (7) realize an inverse-square-law or distance-based AGC, neglecting any diffuse sound field components.

2.2. Combined Direct and Diffuse Sound Field

Eq. (3) assumes free field conditions, in other words, a direct sound field only. 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 the reflected sound. To derive an equation for the AGC gain that also takes the diffuse sound field into account, we add the sound energy densities of these two sound fields. The sound energy density for a sound pressure p is given by

$$D = \frac{p^2}{\rho c^2} \tag{8}$$

For the direct sound, we obtain from Eqs. (4) and (5)

$$D = \frac{W}{4\pi r^2 c} \tag{9}$$

For the diffuse sound, the sound energy density is given in accordance with Beranek [20] by

$$D' = \frac{4W(1-\bar{\alpha})}{cS\bar{\alpha}} \tag{10}$$

where $\bar{\alpha}$ denotes the average sound absorption coefficient, and S the room surface area. The average absorption coefficient $\bar{\alpha}$ is the area-weighted average of the individual absorption coefficients α_i,

$$\bar{\alpha} = \frac{1}{S}\sum_i S_i \alpha_i$$

While the direct sound field depends on the radius, i.e., the distance to the sound source, the diffuse sound field does not. Adding the sound energy densities of direct and diffuse sound field, multiplying with $\rho_0 c^2$, and taking the square root results in the sound pressure

$$p(r) = \sqrt{W\rho_0 c\left(\frac{1}{4\pi r^2} + \frac{4(1-\bar{\alpha})}{S\bar{\alpha}}\right)} \tag{11}$$

With Eq. (11), we can determine the ratio of the sound pressures at radius r and r_0,

$$\frac{p(r)}{p(r_0)} = \sqrt{\frac{R + 1/r^2}{R + 1/r_0^2}} \tag{12}$$

where

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

Given the room dimensions, we can specify the room surface S and the room volume V. Fig. 1 illustrates the sound pressure ratio $p(r)/p(r_0)$ stated by Eq. (12). For $\bar{\alpha} = 1$, i.e., a direct sound field only, the pressure ratio is inversely proportional to the radius throughout the entire range, reflected by the straight $1/r$ line in the double-logarithmic plot. As $\bar{\alpha}$ decreases, the sound pressure level of the diffuse sound field increases. Furthermore, for a small radius, the direct sound pressure dominates, and vice versa, for a large radius, the diffuse sound pressure dominates. The radius or distance for which the sound pressure levels of direct and diffuse sound field are equal is called critical distance or reverberation distance. If the speaker-to-microphone distance stays well within the critical distance, we can neglect the diffuse sound field in the computation for the AGC gain and apply Eq.(7).

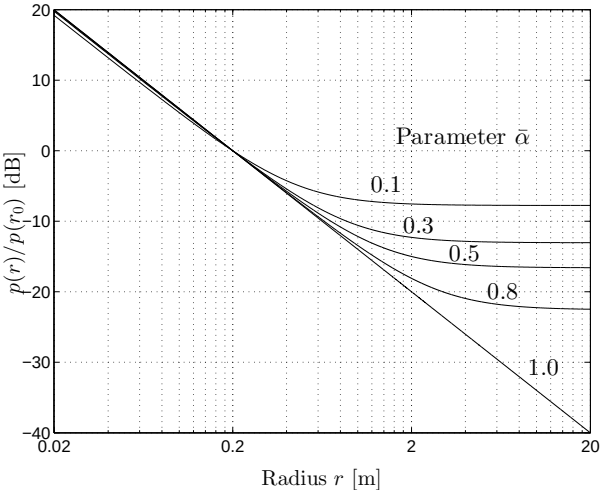

Fig. 1: Sound pressure ratio in room with average absorption coefficient $\bar{\alpha}$ for a reference microphone distance of $r_0 = 20$ cm and a room volume of $V = 90$ m^3 (combined direct and diffuse sound field).

Since the gain applied for the AGC is the inverse of the sound pressure ratio, Eq. (12), the AGC gain for the combined direct and diffuse sound fields becomes

$$G'(r) = G_0\sqrt{\frac{R + 1/r_0^2}{R + 1/r^2}} \tag{13}$$

In analyzing Eq.(13), the two extrema $\bar{\alpha} = 0$ and $\bar{\alpha} = 1$ for the average absorption coefficient are of particular interest. For $\bar{\alpha} = 1$ (complete absorption), we obtain $R = 0$, therefore

$$G'(r)\Big|_{\bar{\alpha}=1} = G_0\frac{r}{r_0} \tag{14}$$

For this case, we have only the direct sound field. Therefore, we obtain a result identical to Eq. (7). For the other extremum, $\bar{\alpha} = 0$ (no absorption), we

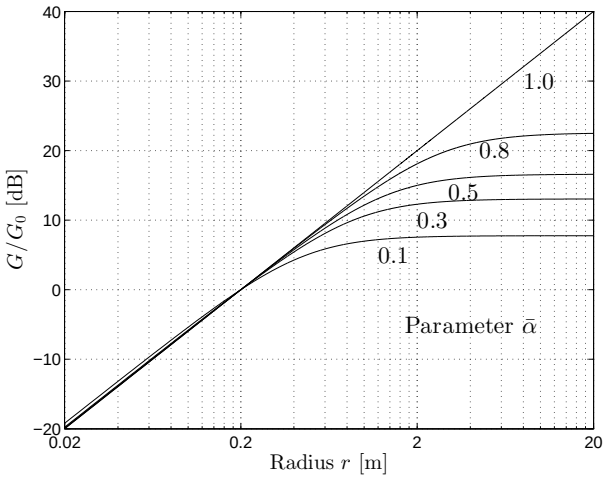

Fig. 2: Desired AGC gain in room with average absorption coefficient $\bar{\alpha}$ for a reference microphone distance of 20 cm and a room volume of 90 m^3 (combined direct and diffuse sound field).

obtain $R = \infty$, and

$$\lim_{\bar{\alpha}\to 0} G'(r) = G_0$$

With no absorption, i.e., completely reflecting walls, the sound pressure level will no longer depend on the distance to the sound source. Note, assuming no absorption by the medium, no sound energy would be lost in this case, and sound pressure would build up over time. In practice, however, a small amount of absorption will ensure an equilibrium.

The desired AGC gain in room conditions given in Eq.(13) is illustrated in Fig. 2. If we tolerate a maximum gain error of 3dB, which occurs at the critical distance r_c, we can approximate each gain curve by two linear portions. The first portion according to the inverse of the radius associated with the direct sound field, and the second portion a constant, associated with the diffuse field,

$$G'(r) \approx \begin{cases} G_0 \dfrac{r}{r_0} & ;\text{if } r \le r_c \\ G_0 \dfrac{r_c}{r_0} & ;\text{else} \end{cases} \tag{15}$$

For $r \le r_c$, we apply the gain required to adjust to the direct sound field; for $r > r_c$, we apply a constant gain identical to the gain required to adjust for the direct sound field at r_c.

In addition to Eqs. (12) and (13), it is desirable to have an analytical expression for the sound pressure ratio and the AGC gain as a function of the critical distance r_c. These equations can be derived in the following way. To determine the critical distance r_c, we set the energy density of the direct sound Eq. (9) equal to the energy density of the diffuse sound Eq. (10), and solve for r_c,

$$r_c = \frac{1}{4}\sqrt{\frac{S\bar{\alpha}}{\pi(1-\bar{\alpha})}} \tag{16}$$

Alternatively, we can determine r_c from the standard deviation of the energy spectral response [21]. Likewise, we can measure the direct sound field pressure p_0 at a radius r_0 close to the source, and the diffuse field sound pressure p_D far from the source, both through sound level measurements, then apply $r_c = r_0 \cdot p_0/p_D$, an equation which can be derived from Eq. (3). With Eqs. (12) and (16), we can now write the pressure ratio in the following form,

$$\frac{p(r)}{p(r_0)} = K\sqrt{1+\left(\frac{r_c}{r}\right)^2} \tag{17}$$

whereby the factor K is given by

$$K = \left(1+\frac{1}{Rr_0^2}\right)^{-\frac{1}{2}} \tag{18}$$

Eq. (17) has the magnitude form of an inverse first order low-pass with respect to parameter r, apparent also from Fig. 1.

Taking the inverse of the right term in Eq. (17), we can now express the AGC gain in terms of the critical distance,

$$G'(r) = G_0\left(K\sqrt{1+\left(\frac{r_c}{r}\right)^2}\right)^{-1} \tag{19}$$

which is a 1^{st} order low-pass with a cut-off or transition radius of r_c, and a gain further depending on the room absorption and the desired reference distance. For simplicity, we assume a purely spherical sound source throughout this paper. However, a directivity factor γ can readily be included by using the effective critical distance $r_c^* = r_c\sqrt{\gamma}$ [22].

3. PROXIMITY-EFFECT COMPENSATION

The proximity effect is a gradual increase of the low frequency output as a pressure-gradient microphone approaches a sound source. It occurs for pressure-gradient, i.e., directional microphones only, and only if exposed to a wave field with curvature components, i.e., spherical components. It does not occur for pressure microphones, nor does it occur for any microphone type in a plane wave field.

Beranek [20] states that the force acting to move the diaghragm of a pressure-gradient microphone is

$$F_D = -A\frac{\delta p}{\delta r}\Delta l \cos\theta \tag{20}$$

where A denotes the area of the diaphragm, p the sound pressure, θ the angle between the direction of the sound wave and the normal of the diaphragm, and Δl the effective distance between the two sides of the diaphragm.

Using Eq. (3), we can determine the pressure gradient in Eq. (20) for a spherical sound wave,

$$\frac{\delta p}{\delta r} = -p_0\frac{r_0}{r}e^{jkr}\left(\frac{1}{r}+jk\right) \tag{21}$$

The first term in the sum is sometimes called the *near-field gradient* and the second term the *far-field gradient* [23]. The near-field gradient is frequency independent and caused by the amplitude decrease with distance, whereas the far-field gradient is frequency-dependant and caused by phase difference. Most relevant here, the first term causes the proximity effect.

The near-field term is only relevant for low frequencies. If we disregard the low frequency range for a moment, we observe that the pressure gradient as given in Eq. (21) increases proportional to the frequency, as a result of the contribution of $k = \omega/c$. For this reason, a basic pressure-gradient device requires equalization with a term proportional to $1/\omega$, a task that can be achieved either electronically by an analog/digital filter or mechanically by a mass-controlled ribbon or diaghragm [24].

To relate the gradient Eq. (21) to the velocity, we equalize the gradient in a suitable way, i.e., we multiply both sides of Eq. (21) with the term $1/j\omega\rho_0$,

$$\frac{\delta p}{\delta r}\frac{1}{j\omega\rho_0} = p_0\frac{r_0}{r}e^{jkr}\frac{1}{\rho_0 c}\left(1-\frac{j}{kr}\right) \tag{22}$$

Recall that the velocity can be invoked via the linear Euler equation,

$$-\nabla p = \rho_0\frac{\delta v}{\delta t} \tag{23}$$

where ∇ denotes the gradient, v the particle velocity, and ρ_0 the static air density. Using spherical coordinates and taking the derivative of the velocity of an assumed harmonic wave, Eq. (23) becomes

$$-\frac{\delta p(r,t)}{\delta r} = j\omega\rho_0 v(r,t) \tag{24}$$

Inserting Eq. (21) into Eq. (24) and resolving for the velocity, we obtain

$$v(r,t) = p_0\frac{r_0}{r}e^{jkr}\frac{1}{\rho_0 c}\left(1-\frac{j}{kr}\right) \tag{25}$$

While Eq. (22) is derived from the requirement for equalization of the pressure gradient and as such, represents the desired output of a gradient microphone, Eq. (25) is derived from the linear Euler equation. Both equations being equal, we can also state that the output of a gradient microphone is proportional to the velocity. Hence, the gradient microphone is also called velocity microphone.

From Eq. (25) we observe that the velocity of a spherical wave is complex in the near-field, i.e., pressure and velocity are not in phase. Furthermore, for $k \gg 1$, i.e., distances large compared to the wavelength, the spherical wave field approaches a plane wave field governed by the specific acoustic impedance $Z = \rho_0 c$ and the relationship $p = Z \cdot v$.

The term $1 - j/kr$ in Eq. (25) causes the proximity effect. For convenience, we introduce

$$H_0(kr) = 1 - \frac{j}{kr} \tag{26}$$

to refer to the transfer function of the proximity effect. Note however, Eq. (26) is only valid for a pure pressure-gradient microphone (figure-8). To derive a general transfer function for the proximity effect, we use angle θ as specified in association with Eq. (20). Recall that the polar pattern $R(\theta)$ for a first-order microphone can be denoted according to [24] as

$$R(\theta) = a + b \cdot cos\theta \tag{27}$$

whereby the parameters $a \leq 1$ and $b \leq 1$ (with $a+b=1$) specify the directivitiy of the microphone. The omnidirectional microphone is determined by $a = 1$, $b = 0$; the cardioid by $a = 0.5$, $b = 0.5$; the supercardioid by $a = 0.37$, $b = 0.63$; the hypercardioid by $a = 0.25$, $b = 0.75$; and the figure-8 by $a = 0$, $b = 1$. In other words, we can represent a 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. Accordingly, we can now specify the proximity effect in terms of the factors a and b,

$$H(kr) = a + b\cos\theta H_0(kr) \tag{28}$$

Since the omnidirectional microphone has no proximity effect, its contribution is frequency-independent.

Inserting $H_0(kr)$ from Eq. (26) into and Eq. (28), we obtain

$$H(kr) = a + b\cos\theta - j\frac{b\cos\theta}{kr} \tag{29}$$

The corner frequency for this 1st order filter is

$$f_c = \frac{c}{2\pi r}\cdot\frac{b\cos\theta}{a+b\cos\theta} \tag{30}$$

where c is the speed of sound. The proximity effect is often stated in its magnitude form [24], but only for the special case of $a = 0$, $b = 1$ (pure pressure-gradient microphone) and $\theta = 0$, i.e., $|H(kr)||_{\theta=0,b=1} = \sqrt{1+1/(kr)^2}$. With Eq. (29), we can now generalize this magnitude response,

$$|H(kr)| = \sqrt{(a+b\cos\theta)^2 + \left(\frac{b\cos\theta}{kr}\right)^2}$$

To illustrate the effect of the parameters r, b, and θ, we first plot the magnitude frequency response of the proximity effect in Fig. 3 for various speaker-to-microphone distances r, at fixed directivity $b = 1$, i.e., for a pure pressure-gradient microphone, and fixed angle $\theta = 0$. Fig. 4 shows the proximity effect for various pressure-gradient microphones at fixed speaker-to-microphone distances. Finally, Fig. 5 shows the proximity effect for various sound arrival angles θ at fixed speaker-to-microphone distance and fixed directivity (cardioid).

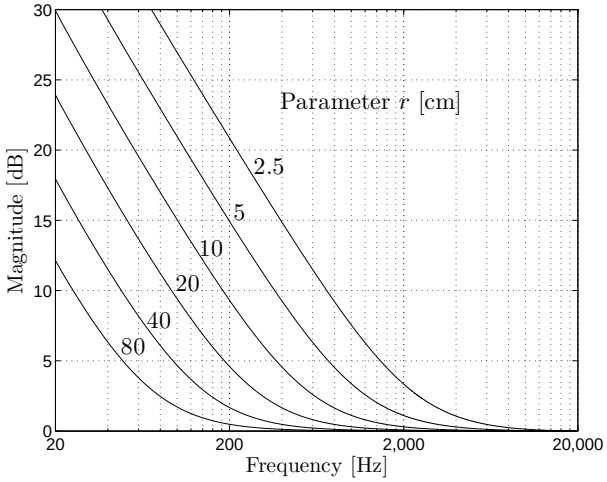

Fig. 3: Proximity effect magnitude response $|H(kr)|$ for a pure pressure-gradient microphone (figure-8) at various speaker-to-microphone distances r and fixed angle $\theta = 0^0$.

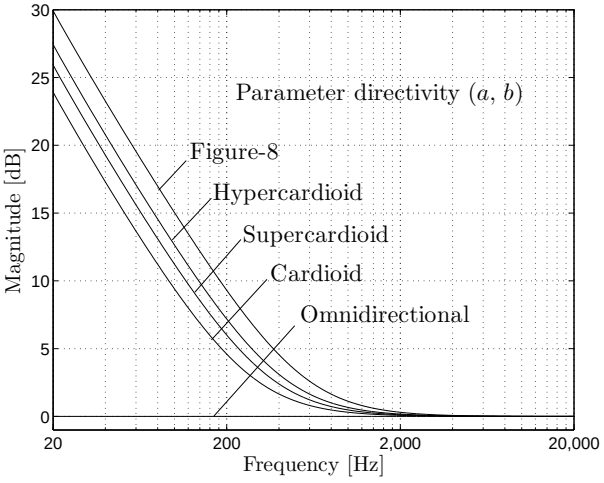

Fig. 4: Proximity effect magnitude response $|H(kr)|$ for various pressure-gradient microphones at fixed speaker-to-microphone distance $r = 0.1$m and fixed angle $\theta = 0^0$.

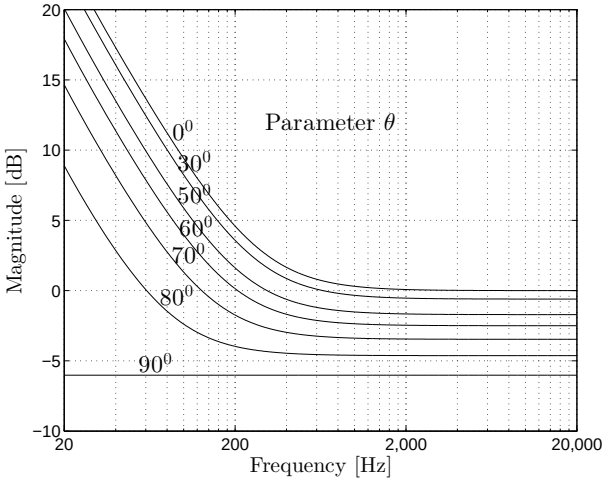

Fig. 5: Proximity effect magnitude response $|H(kr)|$ for various sound arrival angles θ at fixed speaker-to-microphone distance $r = 0.1$m and a fixed directivity pattern, a cardioid.

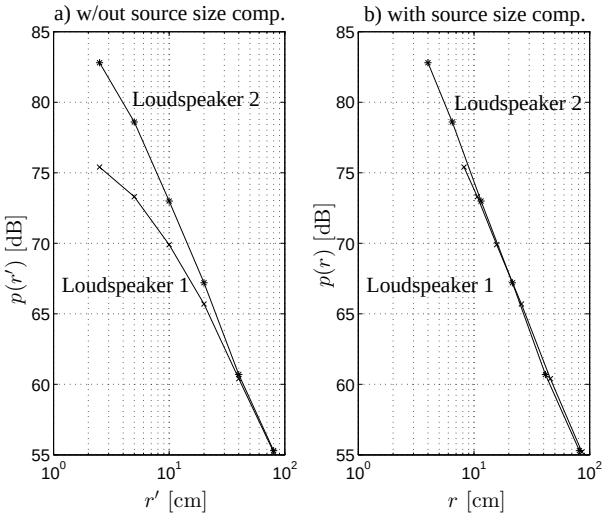

Fig. 6: Sound pressure levels displayed without (a) and with (b) source-size compensation for loudspeaker 1 (10cm Ø) and loudspeaker 2 (4cm Ø).

Taking the inverse of $H(kr)$, Eq. (29), we can now specify the compensation filter for the proximity effect, i.e.,

$$C(kr) = \frac{kr}{kr(a + b\cos\theta) - jb\cos\theta} \tag{31}$$

The magnitude responses of these filters are apparent from Fig. 3 - 5, the curves in these plots simply need to be mirrored at the 0-dB magnitude line. If the speaker-to-microphone distance r and the type of pressure-gradient microphone, (i.e., cardioid, supercardioid, hypercardidod) is known, we can apply the corresponding compensation filter $C(kr)$.

Using both Eqs. (19) and (31), we arrive at

$$G_{tot}(kr) = G'(r) \cdot C(kr) \tag{32}$$

the distance-based AGC with proximity-effect compensation.

In practice, sound sources are not infinitely small. That is, if we record voice in centimeter distance from the mouth, we cannot assume a sound field curvature corresponding to a point source with center right at the mouth. A simple way to model a finite sound source is to account for the sound source dimensions. Assuming a sphere for the sound source (e.g., a human head), we can associate a sound source radius r_s with it and write

$$r = r' + r_s \tag{33}$$

where r is the radius used in Eqs. (7)-(32) and r' the measured distance from the sound source (e.g., the mouth).

4. EXPERIMENTAL VERIFICATION

To verify the concept of distance-based AGC, we first investigate the relevance of the source size in Eq. (33) using two loudspeakers of different sizes, a 10-cm full-range speaker (Yamaha MS101II) here referred to as loudspeaker 1, and a 4-cm full-range speaker (HP USB mini) referred to as loudspeaker 2. Playing white noise on each loudspeaker individually, we recorded the sound with an omnidirectional microphone (Earthworks M23) sequentially placed at six logarithmically-spaced distances, r'=2.5cm, 5cm, 10cm, 20cm, 40cm, and 80cm. Fig. 6 shows the sound pressure levels calculated from the recorded signals and displayed with and without source-size compensation. Recall from Eq. (3) that a point source is characterized by a 6dB sound pressure roleoff per doubled distance (or 20dB per decade). Considering now Fig. 6a), we note that the smaller

loudspeaker behaves, as expected, in a wide range (5cm...80cm) similar to a point source. In order to apply source-size compensation (Fig.6b), necessary in particular for the larger loudspeaker 1, we first need to determine the source radius r_s. We may assume that we can set r_s to half the loudspeaker enclosure width (i.e., 14cm/2 and 5cm/2), which predominantly determines the horizontal wave field. To verify this assumption, we find the source radius that is optimal in a least square sense. Using r_s as a curve fitting parameter, we determine r_s such that it minimizes the sum of the squared errors with respect to the best positioned line with a -60dB/dec slope. This way, we obtain r_s =5.7cm for loudspeaker 1 and r_s =1.5cm for loudspeaker 2, results that are indeed similar to the corresponding loudspeaker-enclosure widths. Fig. 6b) shows the sound pressure levels when plotted against the source-size compensated radius (i.e., an offset to the abscissa), both curves closely approximating a line and hence modeling a point source. Note, the sound pressure level curves derived from half the enclosure width (not shown in Fig. 6) deviate only marginally (<1dB) from the optimal case.

Since Eq. (12) predicts a sound-pressure role-off as shown in Fig. 1, we can conclude from comparing Fig. 6b) with Fig. 1 that the critical distance r_c is greater than 80 cm in the measurements associated with 6b). Therefore, we can simply use Eq. (7), instead of Eq. (13), (15), or (19) to calculate the distance-based AGC gain for these recordings. In fact, for telepresence use-cases, we often have the speaker within $r <$ 80cm, in which case the assumption $r < r_c$ is usually valid, making the distance-based AGC calculation straight-forward (Eq. (7)).

To obtain recorded speech signals, we removed the loudspeaker and a male speaker took its position. The speaker announced the microphone distance with the phrase "The microphone is now at __ cm". Three microphones were used to record this case: a cardioid (AKG Perception 170), an omnidirectional microphone (Earthworks M23), both mounted on a slider with a distance scale to measure the distance to the speaker; and an additional control microphone (AKG Perception 170), set up at fixed distance to verify that the speaker spoke at the same volume for each microphone distance.

Fig. 7 shows the recorded signals for the cardioid (a) and omni (b) at the six different distances, each distance represented by a 3.4 s interval in the plots. The processed signals are shown in (c)-(f). For both AGC gain computation and proximity-effect compensation, r_s was set to 2.5cm (Eq. (33)). Furthermore, with the condition $r < r_c$ being confirmed (see Fig. 6b), it was adequate to apply the simpler computation for the distance-based AGC gain, i.e., Eq. (7). Signal (c) is the output of the distance-based AGC for the cardioid (using Eq.(7)); signal (d) is the output for the omni (using Eq.(7)); signal (e) is the output of the distance-based AGC with proximity-effect compensation for the cardioid (using Eqs. (7), (29), (32)); and signal (f) is obtained from processing the omni signal with a conventional signal-based AGC (Ableton Live 8) at a preferred setting of 10ms attack and 1s release time. Due to the proximity effect, the cardioid (a) shows a larger level increase at close distances than the omni (b). Since sound arrival was from the front, the maximum proximity effect occurred. After processing with the distance-

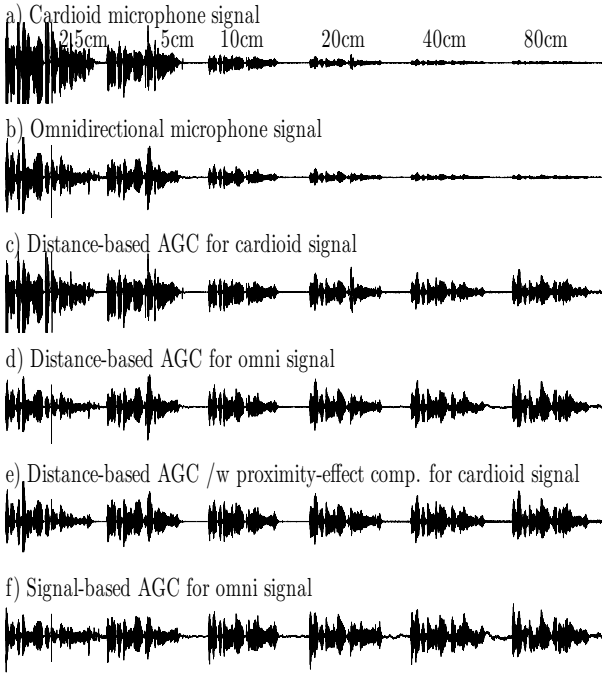

Fig. 7: Recorded speech signals (a), (b); output of distance-based AGC (c)-(e); and output of conventional signal-based AGC (f).

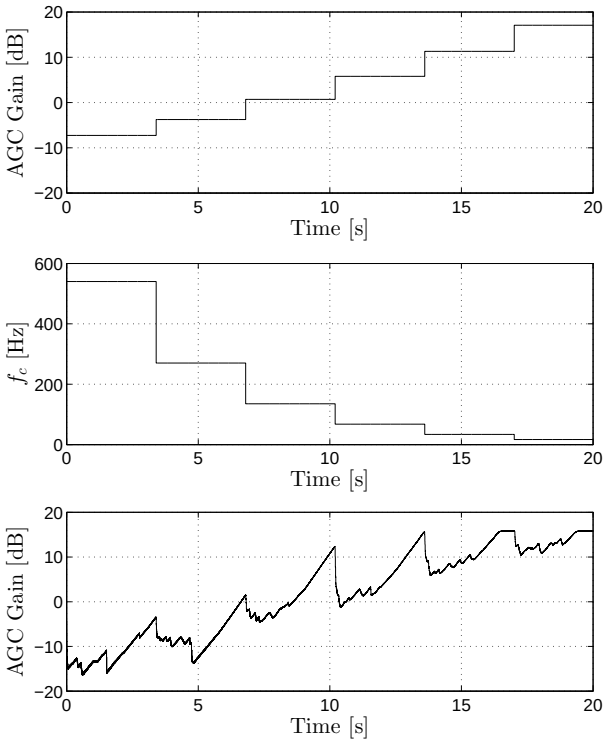

Fig. 8: From top to bottom: Gain of distance-based AGC, corner frequency f_c of proximity-effect compensation filter, gain of conventional signal-based AGC.

based AGC, the cardioid signal (c) still shows a level bias at close distances due to the bass boost of the proximity effect, whereas the omni signal (d) provides equal levels for all distances. If the cardioid signal is processed with both the distance-based AGC and the proximity-effect compensation (e), the signal levels become uniform. As expected in plot (f), the omni signal processed by a conventional signal-based AGC also provides equal levels, however on cost of distorted short-term dynamics.

Fig. 8 displays the inner workings of the algorithms that produced the results in Fig. 7. The distance-based AGC results in a step-wise gain, since for each distance the gain is at a fixed level, unlike the signal-based AGC, where the gain constantly changes and reacts to individual speech syllables, even though the distance changes only every 3.4 seconds. For the distance-based AGC, Fig. 8 also shows the corner frequency of the applied proximity-effect compensation filter. As expected, the corner frequency decreases as the distance increases. Informal listening tests confirmed that neither level-changes nor spectral changes (i.e., bass boost) are heard in the signals processed by the distance-based AGC with proximity-effect compensation, unlike for the original microphones signals that are disturbing to listen to, due to the large level change of ∼20dB (from 2.5cm to 80cm), and in the case of the directional microphone, due to the additional bass boost of >20dB (at 100 Hz) for close talking distances (2.5cm and 5cm).

The processed signals shown in Fig. 7 (c)-(e) confirmed that the theoretical results in Sections 2 and 3 apply well to recorded voice. In addition, we have applied the method described here to teleconference voice recordings, in which the speaker constantly changes his distance to a fixed table microphone, with his head position being tracked by a Polhemus Patriot. These result will be published separately.

5. CONCLUSION

Based on the inverse-square-law, we derived the equation for distance-based AGC accounting for both the direct and the diffuse sound field. We represented the AGC gain as a 1st order low-pass with respect to the distance from the sound source, with a cut-off or transition distance equivalent to the critical distance. We provided a simple approximation for the AGC gain based on a comparison with the estimated critical distance.

For the application to directional microphones, we derived a general proximity-effect transfer function, which accounts for the speaker-microphone distance, the sound arrival angle, and the microphone directivity. The inverse function is then applied for proximity-effect compensation.

In the experimental verification, we found that the distance-based AGC and the proximity-effect compensation both work well in practice. As accurate sensors for distance (and orientation) measurement become available at an acceptable cost, distance-based AGC with proximity-effect compensation will become applicable to speech acquisition in numerous fields.

6. REFERENCES

[1] T.G. Stockham, Jr., "The application of generalized linearity to automatic gain control," *IEEE Trans. Audio and Electroacoustics*, vol. 16, no. 2, pp. 267-270, June 1968.

[2] A. Kaiser and B.B. Bauer, "A new automatic level control for monophonic and stereophonic broadcasting," *IRE Trans. Audio*, pp. 171-173, Nov./Dec. 1962.

[3] D.C. Connor and R.S. Putnam, "Automatic Audio Level Control," *J. Audio Eng. Soc.*, vol. 16, no.3, July 1968.

[4] F. Felber, "An automatic volume control for preserving intelligibility," *IEEE Sarnoff Symposium*, 2011.

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

[6] G.W. McNally, "Dynamic range control of digital audio signals," *J. Audio Eng. Soc.*, vol. 32, no. 5, pp. 316-327, May 1984.

[7] E.F. Stikvoort, "Digital dynamic range compressor for audio," *J. Audio Eng. Soc.*, vol. 34, no. 1/2, pp. 3-9, Jan./Feb. 1986.

[8] W. Etter and G.S. Moschytz, "Noise reduction by noise-adaptive spectral magnitude expansion," *J. Audio Eng. Soc.*, vol. 42, no. 5, pp.341-349, May 1994.

[9] D. Mapes-Riordanm and W.M. Leach, Jr., "The design of a digital signal peak limiter for audio signal processing," *J. Audio Eng. Soc.*, vol. 36, no. 7/8, pp. 562-574, July/Aug. 1988.

[10] W.R. Aiken and C.F. Swisher, "An improved automatic level control device," *J. Audio Eng. Soc.*, vol. 16, no. 4, pp.447-449, Oct. 1968.

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

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

[13] 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, pp. 10-18, Jan./Feb. 1986.

[14] E. Georganti, T. S. van de Par, A. Harma, and J. Mourjopoulos, "Speaker distance detection using a single microphone," *IEEE Trans. Audio, Speech, and Language Process.*, vol. 19 , no. 7, pp. 1949-1961, Sept. 2011.

[15] S. Vesa, "Binaural speech source distance learning in rooms," *IEEE Trans. Acoust., Speech, Lang. Process.*, vol. 17, no. 8, pp. 1498-1507, Nov. 2009.

[16] C.H. Knapp and G.C. Carter, "The generalized correlation method for estimation of time delay," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 24, no. 4, pp. 320-327, Aug. 1976.

[17] J.P. Rolland, L.D. Davis, and Y. Bailoot, "A survey of tracking technology for virtual environments," in *Fundamentals of wearbale computers and augmented reality*, chapt. 3, Eds., Barfield, Caudell, Mahwah, NJ, 2000.

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

[19] L.E. Kinsler, et. al., "Fundamentals of acoustics," Wiley, 4th ed., New York, 2000, p. 127.

[20] L.L. Beranek, "Acoustics," McGraw-Hill, New York, 1954, pp. 49, 312.

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

[22] E. Zwicker and M. Zollner, "Elektroakustik," Springer, 2nd ed., Berlin, 1987, p. 199.

[23] J. Blauer and N. Xiang, "Acoustics for engineers," Springer, Berlin, 2008, p. 49.

[24] R.A. Rayburn, "Eargle's microphone book," Focal, 3rd ed., 2012, pp. 78, 64.