

Effect of quantization on beamforming in binaural hearing aids

Sriram Srinivasan

Ashish Pandharipande

Kees Janse

{sriram.srinivasan, ashish.p, kees.janse}@philips.com

Philips Research
High Tech Campus 36
5656AE Eindhoven, The Netherlands

ABSTRACT

We consider a beamforming scheme for binaural hearing aids and analyze the effects of quantization on its performance. A limited bandwidth wireless link is assumed between the binaural hearing aids for communication from one ear to the other ear, where signals are then processed using a generalized sidelobe canceller (GSC). We consider the case where observations are quantized before transmission, and study the performance of the GSC beamformer under quantization errors.

Keywords

Beamforming, binaural, hearing aids, quantization

1. INTRODUCTION

Improving speech intelligibility in the presence of interfering sources is an important aspect in the design of hearing aid signal processing algorithms [1]. Modern hearing aids attempt to achieve this goal through beamforming using two or more microphones and exploit the spatial diversity resulting from the different spatial positions of the desired and interfering sound sources [2].

The distance between the microphones on a single hearing aid is typically less than 1 cm, due to the small size of such devices for aesthetic reasons. This small spacing limits the gain that can be obtained from microphone array speech enhancement algorithms. A high speed wireless link between the hearing aids worn on the left and right ears has been recently introduced [3], and the two devices form a body area network. This enables beamforming using the microphones on the two hearing aids, which are separated by a larger distance corresponding to the distance between the two ears (around 20 cm).

To conserve bandwidth, it is useful to quantize the signal prior to transmission. In this paper, we study the effects of quantization on the performance of a beamforming algorithm. We assume that each

hearing aid has one microphone and that the right hearing aid quantizes and transmits its signal to the left hearing aid. To retain our focus only on the effect of source quantization errors, we further assume that the transmission channel is error-free.

If the power spectral density (PSD) of the desired source is known a priori, the two-microphone Wiener filter provides the optimal (in the mean squared error sense) estimate of the desired source. The effect of quantization errors in such a framework has been investigated in [4]. However, in practice the PSD is unknown. In this paper, we consider a particular data-independent beamformer, the generalized sidelobe canceller (GSC) [5], which does not require prior knowledge of the source PSD.

The GSC requires knowledge of the location of the desired source and of the microphone array configuration, both of which are available in hearing aid applications. The array configuration is fixed and known a priori. The location of the desired source in the hearing aid scenario is commonly assumed to be 0° (in front of the microphone array) [2]. In a free field (no reverberation), the two microphone GSC can cancel out an interfering sound source without distorting the desired signal, which is a desirable feature in hearing aids. Thus the GSC is well suited for hearing aid applications, and we study the impact of quantization errors on the GSC in this paper.

2. SIGNAL MODEL

Consider a two-element broadside microphone array spaced d meters apart as shown in Fig. 1. Consider a source located in the

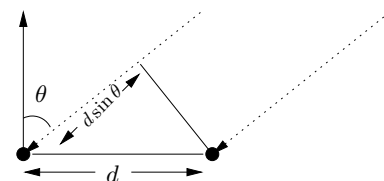


Figure 1: A two element broadside microphone array with a source in the far field.

far field at an angle θ , where θ corresponds to the angle between the vertical y -axis and a ray from the origin to the sound source. The signal received at the left microphone is delayed by $\frac{d \sin \theta}{c}$ seconds relative to the right microphone, where c is the speed of sound.

We assume there exists a desired sound source $S(t)$ located at an angle θ_s , and an interfering sound source $I(t)$ located at an angle θ_i , where t denotes the time index. Assuming there is no reverberation, the received microphone signals $X_l(t)$ and $X_r(t)$, at the left

The herein disclosed information is secret until 2008-03-13; eyes only for the appointed reviewers of this conference.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

BodyNets '08 Tempe, Arizona USA

Copyright 2008 ACM X-XXXXX-XX-X/XX/XX ...\$5.00.

and right sensors respectively, can be written as

$$X_l(t) = S(t - \tau_s) + I(t - \tau_i) + N_l(t) \quad (1)$$

$$X_r(t) = S(t) + I(t) + N_r(t), \quad (2)$$

where $\tau_s = d \sin(\theta_s)/c$, $\tau_i = d \sin(\theta_i)/c$. N_l and N_r include both sensor noise and background noise, which we assume to be uncorrelated. We assume S , I , N_l and N_r to be pairwise independent zero mean white Gaussian processes with variances σ_s^2 , σ_i^2 , $\sigma_{n_l}^2$, and $\sigma_{n_r}^2$, respectively. We assume $\sigma_{n_l}^2 = \sigma_{n_r}^2 = \sigma_n^2$. Thus, X_l and X_r are both zero mean white Gaussian with variance $\sigma_x^2 = \sigma_s^2 + \sigma_i^2 + \sigma_n^2$. In practice, the desired and interfering signals in hearing aid applications need not be white Gaussian but we retain this assumption for the purpose of studying quantization effects.

We consider the case where the quantities θ_i , σ_s^2 , σ_i^2 , and σ_n^2 are all unknown. Only the location of the desired source θ_s (e.g., 0° in the hearing aid scenario) and the distance d are assumed to be known. The GSC structure [5] depicted in Fig. 2 is well suited for this application as it only requires knowledge of θ_s and d . The fixed beamformer is a delay-and-sum beamformer that averages its two inputs after compensating for their relative delay τ_s with respect to the desired source. The blocking matrix subtracts the delay compensated signals resulting in a reference signal that is devoid of the desired signal, and forms the input to the adaptive interference canceller.

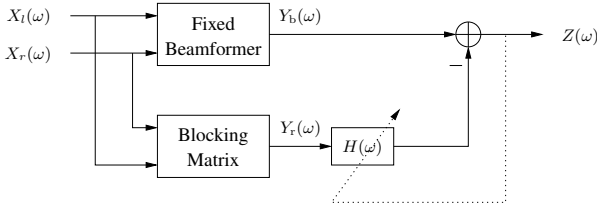


Figure 2: Frequency domain implementation of the GSC.

We assume that the hearing aid at the right ear quantizes and transmits its signal to the hearing aid at the left ear where the two are combined. Let \tilde{X}_r represent the reconstructed signal obtained after encoding and decoding X_r at a rate R bits per sample resulting in a distortion D . The *forward channel* w.r.t to the squared error criterion can be written as [6, pp. 100-101]

$$\tilde{X}_r = \alpha(X_r + W) \quad (3)$$

where $\alpha = (\sigma_x^2 - D)/\sigma_x^2$, and $W \sim \mathcal{N}(0, \sigma_x^2 D/(\sigma_x^2 - D))$. The rate-distortion relation for a white Gaussian source with variance σ_x^2 is given by [6, ch. 4]

$$R(D) = \frac{1}{2} \log_2 \left(\frac{\sigma_x^2}{D} \right), \quad (4)$$

so that the distortion D is obtained as $D = \sigma_x^2 2^{-2R}$. The signals X_l and \tilde{X}_r form the two inputs to the GSC.

If the signal variances σ_s^2 , σ_i^2 and σ_n^2 are known, more efficient quantization schemes may be designed, e.g., one could first estimate the desired signal (using a Wiener filter) from the noisy observation X_r at the right ear, and then quantize the estimate as in [4]. However, as the variances are unknown in our model, we quantize the noisy observation itself, and the resulting performance serves as a lower bound on the achievable performance.

3. THE GENERALIZED SIDELobe CANCELLER

We consider a frequency domain version of the GSC. We first look at the case when there is no quantization and the left hearing aid receives an error-free description of X_r . This corresponds to an upper bound in our performance analysis. We then consider the case when $X_r(t)$ is quantized at a rate R bits per sample.

3.1 No quantization

The output of the fixed delay-and-sum beamformer is given by

$$Y_b(\omega) = \mathbf{F}(\omega) \mathbf{X}(\omega), \quad (5)$$

where $\mathbf{F}(\omega) = \frac{1}{2} [1 \ e^{-j\omega\tau_s}]$, $\mathbf{X}(\omega) = [X_l(\omega) \ X_r(\omega)]^T$, ω is the angular frequency, and $X_l(\omega)$ and $X_r(\omega)$ correspond to the discrete Fourier transforms of $X_l(t)$ and $X_r(t)$ respectively. We can rewrite $Y_b(\omega)$ as

$$\begin{aligned} Y_b(\omega) &= \frac{1}{2} (X_l(\omega) + X_r(\omega)e^{-j\omega\tau_s}), \\ &= S(\omega)e^{-j\omega\tau_s} + \frac{1}{2} I(\omega)(e^{-j\omega\tau_i} + e^{-j\omega\tau_s}) \\ &\quad + \frac{1}{2} (N_l(\omega) + N_r(\omega)e^{-j\omega\tau_s}). \end{aligned} \quad (6)$$

$S(\omega)$, $I(\omega)$, $N_l(\omega)$, and $N_r(\omega)$ correspond to the discrete Fourier transforms of $S(t)$, $I(t)$, $N_l(t)$ and $N_r(t)$ respectively. The blocking matrix is given by $\mathbf{B}(\omega) = [1 \ -1]$, so that the input to the adaptive interference canceller $H(\omega)$ is obtained as

$$\begin{aligned} Y_r(\omega) &= \mathbf{B}(\omega) \mathbf{X}(\omega) \\ &= I(\omega)(e^{-j\omega\tau_i} - e^{-j\omega\tau_s}) + N_l(\omega) - N_r(\omega)e^{-j\omega\tau_s}. \end{aligned} \quad (7)$$

The adaptive filter $H(\omega)$ is updated such that the expected energy of the residual $Z(\omega)$ given by $\eta = \mathbb{E}\{|Y_b(\omega) - H(\omega)Y_r(\omega)|^2\}$ is minimized. Since $Y_r(\omega)$ does not contain the desired signal, minimizing η corresponds to minimizing the energy of the interferer in the residual. The optimal (Wiener) solution for $H(\omega)$ is given by

$$H_{\text{opt}}(\omega) = \frac{\mathbb{E}\{Y_b(\omega)Y_r^\dagger(\omega)\}}{\mathbb{E}\{Y_r(\omega)Y_r^\dagger(\omega)\}}, \quad (8)$$

where \dagger indicates complex conjugate transpose. The estimation error at the output of the GSC is given by

$$\begin{aligned} \xi(\omega) &= \mathbb{E}\{(S(\omega)e^{-j\omega\tau_s} - Z(\omega))(S(\omega)e^{-j\omega\tau_s} - Z(\omega))^\dagger\} \\ &= P_z(\omega) - \sigma_s^2, \end{aligned} \quad (9)$$

where

$$\begin{aligned} P_z(\omega) &= \mathbb{E}\{Z(\omega)Z^\dagger(\omega)\} \\ &= \mathbb{E}\{Y_b(\omega)Y_b^\dagger(\omega)\} - \mathbb{E}\{Y_b(\omega)Y_r^\dagger(\omega)\}H_{\text{opt}}^\dagger(\omega). \end{aligned} \quad (10)$$

3.2 Quantization at a rate R

The beamformer output in this case is given as

$$\begin{aligned} \tilde{Y}_b(\omega) &= \frac{1}{2} (X_l(\omega) + \tilde{X}_r(\omega)e^{-j\omega\tau_s}), \\ &= \frac{1}{2} (1 + \alpha)S(\omega)e^{-j\omega\tau_s} + \frac{1}{2} I(\omega)(e^{-j\omega\tau_i} + \alpha e^{-j\omega\tau_s}) \\ &\quad + \frac{1}{2} (N_l(\omega) + \alpha N_r(\omega)e^{-j\omega\tau_s}) + \frac{1}{2} \alpha W(\omega)e^{-j\omega\tau_s}, \end{aligned} \quad (11)$$

where $W(\omega)$ is the discrete Fourier transform of $W(t)$. Comparing (11) with (6), since $0 \leq \alpha \leq 1$, it can be seen that while the fixed beamformer preserves the desired source in the unquantized case,

there is attenuation of the desired source in the quantized case. The blocking matrix produces

$$\begin{aligned} \tilde{Y}_r(\omega) &= (1 - \alpha)S(\omega)e^{-j\omega\tau_s} + I(\omega)(e^{-j\omega\tau_i} - \alpha e^{-j\omega\tau_s}) \\ &\quad + N_l(\omega) - N_r(\omega)\alpha e^{-j\omega\tau_s} - \alpha W(\omega)e^{-j\omega\tau_s}. \end{aligned} \quad (12)$$

It is evident from (12) that due to the quantization, the reference signal $\tilde{Y}_r(\omega)$ is not completely free of the desired signal $S(\omega)$, which will result in some cancellation of the desired source in the interference cancellation stage. The adaptive interference canceller is given by

$$\tilde{H}_{\text{opt}}(\omega) = \frac{E\{\tilde{Y}_b(\omega)\tilde{Y}_r^\dagger(\omega)\}}{E\{\tilde{Y}_r(\omega)\tilde{Y}_r^\dagger(\omega)\}}, \quad (13)$$

where we have using (11) and (12)

$$\begin{aligned} E\{\tilde{Y}_b(\omega)\tilde{Y}_r^\dagger(\omega)\} &= \frac{\sigma_x^2 - \alpha^2(\sigma_x^2 + \sigma_w^2) + j2\alpha\sigma_i^2 \sin\omega(\tau_i - \tau_s)}{2}, \\ E\{\tilde{Y}_r(\omega)\tilde{Y}_r^\dagger(\omega)\} &= \sigma_x^2 + \alpha^2(\sigma_x^2 + \sigma_w^2) \\ &\quad - 2\alpha(\sigma_s^2 + \sigma_i^2 \cos\omega(\tau_i - \tau_s)), \end{aligned} \quad (14)$$

with $\sigma_x^2 = \sigma_s^2 + \sigma_i^2 + \sigma_n^2$. The estimation error can be written as

$$\begin{aligned} \tilde{\xi}(\omega) &= E\{(S(\omega)e^{-j\omega\tau_s} - \tilde{Z}(\omega))(S(\omega)e^{-j\omega\tau_s} - \tilde{Z}(\omega))^\dagger\} \\ &= \tilde{P}_z(\omega) - \alpha\sigma_s^2 + (1 - \alpha)\sigma_s^2(\tilde{H}_{\text{opt}}(\omega) + \tilde{H}_{\text{opt}}^\dagger(\omega)), \end{aligned} \quad (15)$$

where $\tilde{Z}(\omega) = \tilde{Y}_b(\omega) - \tilde{H}_{\text{opt}}(\omega)\tilde{Y}_r(\omega)$, and

$$\begin{aligned} \tilde{P}_z(\omega) &= E\{\tilde{Z}(\omega)\tilde{Z}^\dagger(\omega)\} \\ &= E\{\tilde{Y}_b(\omega)\tilde{Y}_b^\dagger(\omega)\} - E\{\tilde{Y}_b(\omega)\tilde{Y}_r^\dagger(\omega)\}\tilde{H}_{\text{opt}}^\dagger(\omega), \end{aligned} \quad (16)$$

with

$$\begin{aligned} E\{\tilde{Y}_b(\omega)\tilde{Y}_b^\dagger(\omega)\} &= \frac{1}{4}[\sigma_x^2 + \alpha^2(\sigma_x^2 + \sigma_w^2) \\ &\quad + 2\alpha(\sigma_s^2 + \sigma_i^2 \cos\omega(\tau_i - \tau_s))]. \end{aligned} \quad (17)$$

Using (13)-(17), we have

$$\tilde{\xi}(\omega) = \frac{\alpha^2(\sigma_x^2 + \sigma_w^2)(\sigma_x^2 - \sigma_s^2) + \sigma_s^2\sigma_x^2(1 - 2\alpha) + \alpha^2(\sigma_s^4 - \sigma_i^4)}{E\{\tilde{Y}_r(\omega)\tilde{Y}_r^\dagger(\omega)\}}. \quad (18)$$

4. PERFORMANCE ANALYSIS

Using equations (9) and (18), we can study the performance of the GSC without quantization and with quantization at different bit rates. The output signal to interference plus noise ratio (SINR) is obtained as

$$\text{SINR}_{\text{out}}(\omega) = 10 \log_{10} \left(\frac{\sigma_s^2}{\tilde{\xi}(\omega)} \right). \quad (19)$$

We set $\theta_s = 0^\circ$, $\theta_i = 40^\circ$, $d = 0.2$ m and $c = 343$ m/s. The sampling frequency is 16 kHz so that a rate of 1 bit per sample corresponds to 16 kilobits per second (kbps). We consider different values for the signal-to-noise ratio (SNR), given by $10 \log_{10} \sigma_s^2/\sigma_n^2$, and the signal-to-interference ratio (SIR), given by $10 \log_{10} \sigma_s^2/\sigma_i^2$.

Figs. 3a and 3b plot the SINR improvement at different bit rates for an input SIR of 0 dB and input SNRs of 40 dB and 20 dB respectively (corresponding to input SINR of approx. 0 dB). Clearly performance improves as the bit rate increases. At frequencies where it is impossible to distinguish between the different spatial positions of the source and interferer (the sharp nulls in Fig. 3), there

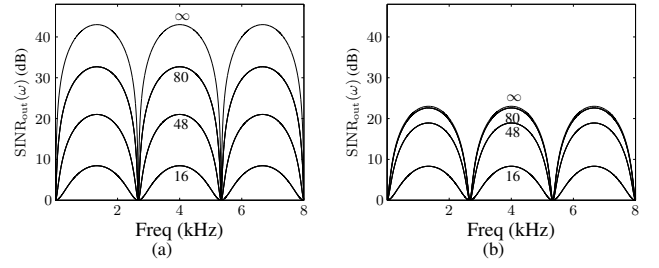


Figure 3: SINR after processing at bit rates of 16, 48, 80 kbps, and for the unquantized case. (a) input SIR 0 dB, input SNR 40 dB, corresp. to input SINR approx. 0 dB. (b) input SIR 0 dB, input SNR 20 dB, corresp. to input SINR approx. 0 dB.

is no improvement in SINR¹. The upper bound on the output SINR corresponds to the unquantized case and depends on the level of the uncorrelated noise, i.e., the SNR (the GSC can completely cancel out the interferer if there is no uncorrelated noise).

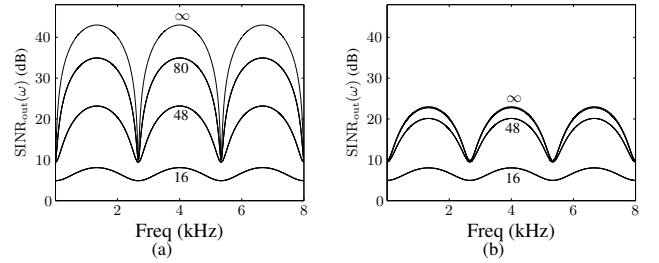


Figure 4: SINR after processing at bit rates of 16, 48, 80 kbps, and for the unquantized case. (a) input SIR 10 dB, input SNR 40 dB, corresp. to input SINR approx. 10 dB. (b) input SIR 10 dB, input SNR 20 dB, corresp. to input SINR approx. 10 dB.

Fig. 4 provides a similar plot but at a higher input SIR of 10 dB. Together with an SNR of 40 dB, this corresponds to an SINR of approximately 10 dB. At a bit rate of 16 kbps, we see that the GSC results in a decrease in performance. The decrease is higher at higher input SINRs. This effect is illustrated more clearly in Fig. 5 which corresponds to an input SINR of approx. 37 dB (SIR = 40 dB, SNR = 40 dB).

To ensure an improvement in SINR after processing, the following condition must be met

$$\tilde{\xi}(\omega) \leq \sigma_i^2 + \sigma_n^2. \quad (20)$$

If we assume $\sigma_n^2 = 0$, the above condition can be simplified as

$$2^{-2R}(1 + C(\omega)2^{-2R}) \leq 1, \quad (21)$$

where

$$C(\omega) = \frac{1}{2} \frac{(\sigma_s^2/\sigma_i^2)^2 - 1}{1 - \cos\omega(\tau_i - \tau_s)}. \quad (22)$$

It is evident from (21) and (22) that a high bit rate is required at a high input SIR σ_s^2/σ_i^2 , which is consistent with intuition and our observations, e.g., in Fig. 5a.

¹An optimal rate allocation strategy would assign fewer bits to these frequencies resulting in a non-uniform bit allocation across frequency.

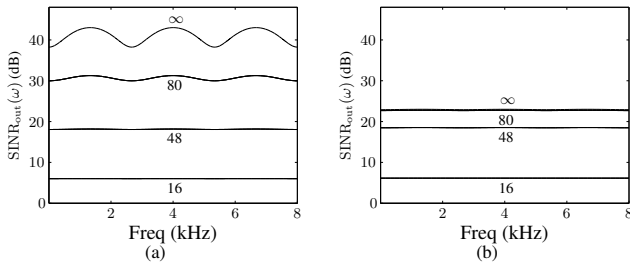


Figure 5: SINR after processing at bit rates of 16, 48, 80 kbps, and for the unquantized case. (a) input SIR 40 dB, input SNR 40 dB, corresp. to input SINR approx. 37 dB. (b) input SIR 40 dB, input SNR 20 dB, corresp. to input SINR approx. 20 dB.

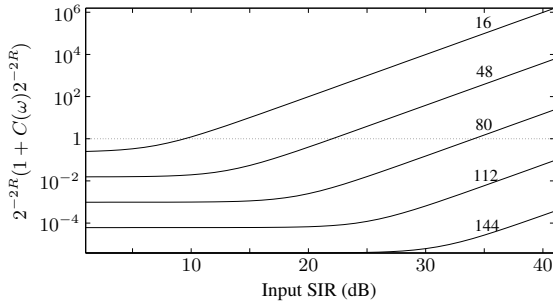


Figure 6: $2^{-2R}(1+C(\omega)2^{-2R})$ (c.f. eq. 21) at different bit rates (kbps) at 4 kHz. SINR improves after processing for values smaller than one (y-axis in log scale).

The quantization noise corresponding to low bit rates results in a high level of the desired signal (relative to the interferer) being present in $Y_r(\omega)$, which is then cancelled by the adaptive filter $H(\omega)$. In fact, as observed in [7], in the absence of uncorrelated noise, the SIR at the output of the adaptive interference canceller is the negative (on a log scale) of the SIR in $Y_r(\omega)$. At high input SIRs, even a small amount of desired signal leakage results in a high SIR in $Y_r(\omega)$, which in turn results in a low SIR at the output.

One approach to avoid cancellation of the desired signal is to adapt the filter only when the desired signal is not active [8]. For signals such as speech that contain a number of pauses, such an update scheme is possible by using a speech activity detector [9]. In this case, the adaptive filter becomes

$$\begin{aligned} \tilde{H}_{\text{opt}}^s(\omega) &= \frac{E\{\tilde{Y}_b(\omega)\tilde{Y}_r^\dagger(\omega)\}}{E\{\tilde{Y}_r(\omega)\tilde{Y}_r^\dagger(\omega)\}} \Big|_{\sigma_s^2=0} \\ &= \frac{1}{2} \frac{(1-\alpha^2)(\sigma_i^2 + \sigma_n^2) - \alpha^2\sigma_w^2 + j2\alpha\sigma_i^2 \sin \omega(\tau_i - \tau_s)}{(1+\alpha^2)(\sigma_i^2 + \sigma_n^2) + \alpha^2\sigma_w^2 - 2\alpha\sigma_i^2 \cos \omega(\tau_i - \tau_s)} \end{aligned} \quad (23)$$

Fig. 7 compares the output SINR at a bit rate of 16 kbps using $\tilde{H}_{\text{opt}}^s(\omega)$ (dotted) and $\tilde{H}_{\text{opt}}(\omega)$ (solid) in (15). Clearly, there is an improvement in performance when the filter update is controlled by a signal activity detector.

5. CONCLUSIONS

The GSC is a data-independent beamformer that only requires knowledge of the location of the desired source and the distance between the microphones, both of which are available in hearing aid applications. At low input SINRs (e.g., 0 dB), the GSC results

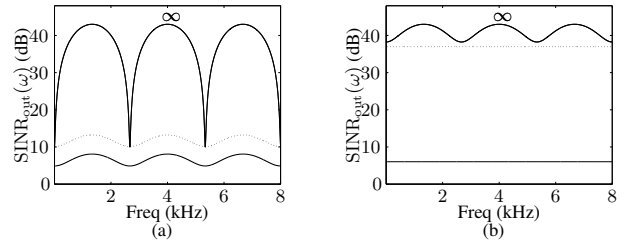


Figure 7: SINR after processing at a bit rate of 16 kbps when the adaptive filter is always updated (solid) and only when desired signal is inactive (dotted). (a) input SIR 10 dB, input SNR 40 dB, corresp. to input SINR approx. 10 dB. (b) input SIR 40 dB, input SNR 40 dB, corresp. to input SINR approx. 37 dB.

in a significant SINR improvement even at moderate bit rates (e.g., 16 kbps). A higher bit rate is required at higher input SINRs. In the noiseless case, a condition on the minimum bit rate required to ensure a gain in SINR has been derived. Some of the performance loss due to quantization at low bit rates can be offset by updating the adaptive interference canceller only when the desired signal is inactive, e.g., during speech pauses. Topics for future work include incorporating the effect of reverberation, and a study of the impact of quantization errors on binaural cues such as interaural level and time differences, which are important to preserve the auditory scene and have been shown to improve intelligibility.

6. REFERENCES

- [1] S. Kochkin, “MarkeTrak V: ‘Why my hearing aids are in the drawer’: the consumers’ perspective,” *The Hearing Journal*, vol. 53, no. 2, pp. 34–42, 2000.
- [2] V. Hamacher, *et al.*, “Signal processing in high-end hearing aids: State of the art, challenges, and future trends,” *EURASIP Journal on Applied Signal Processing*, no. 18, pp. 2915–2929, 2005.
- [3] Oticon Epoq: True binaural sound processing in new Oticon Epoq signals paradigm shift in hearing care.” *Press release*, Apr. 2007. [Online]. Available: http://www.oticon.com/com/Information/PressReleases/downloads/epoq_april2007.pdf.
- [4] O. Roy and M. Vetterli, “Collaborating Hearing Aids,” in *MSRI Workshop on Mathematics of Relaying and Cooperation in Communication Networks*, Apr. 2006.
- [5] L. Griffiths and C. Jim, “An alternative approach to linearly constrained adaptive beamforming,” *IEEE Trans. Antennas and Propag.*, vol. 30, no. 1, pp. 27–34, Jan. 1982.
- [6] T. Berger, *Rate distortion theory: a mathematical basis for data compression*, ser. Information and System Sciences Series, T. Kailath, Ed. Prentice Hall, 1971.
- [7] B. Widrow, *et al.*, “Adaptive noise cancelling: Principles and applications,” *Proc. IEEE*, vol. 63, no. 12, pp. 1692–1716, Dec. 1975.
- [8] D. van Compernelle, “Switching adaptive filters for enhancing noisy and reverberant speech from microphone array recordings,” in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*, vol. 2, Apr. 1990, pp. 833–836.
- [9] A. Koul and J. E. Greenberg, “Using intermicrophone correlation to detect speech in spatially separated noise,” *EURASIP Journal on Applied Signal Processing*, vol. 2006, pp. Article ID 93 920, 14 pages, 2006.