On the Impact of Quantization on Binaural MVDR Beamforming

Jamal Amini[†], *Richard C. Hendriks*[†], *Richard Heusdens*[†], *Meng Guo*[⋆] *and Jesper Jensen*[⋆‡]

[†]Circuits and Systems (CAS) Group, Delft University of Technology, 2628 CD Delft, the Netherlands
[⋆]Oticon A/S, Kongebakken 9, 2765 Smørum, Denmark. [‡]Electronic Systems Department, Aalborg University, 9100 Aalborg, Denmark
Email: [†]{j.amini, r.c.hendriks, r.heusdens}@tudelft.nl, [‡⋆]{megu, jesj}@oticon.com

Abstract

Multi-microphone noise reduction algorithms in binaural hearing aids which cooperate through a wireless link have the potential to become of great importance in future hearing aid systems. However, limited transmission capacity of such devices necessitates the data compression of signals transmitted from one hearing aid to the contralateral one. In this paper we study the impact of quantization as a data compression scheme on the performance of the multi-microphone noise reduction algorithms. Using the binaural minimum variance distortionless response (BMVDR) beamformer as an illustration, we propose a quantization aware beamforming scheme which uses a modified cross power spectral density (CPSD) of the system noise including the quantization noise (QN). Moreover, several assumptions on the QN are investigated in the proposed method. Based on the output SNR, we compare different variations of the proposed method with the conventional BMVDR beamformer. The results confirm the improved performance of the proposed method.

Index Terms - Multi-microphone noise reduction, binaural hearing aids, MVDR beamforming, quantization, dithering.

1 Introduction

Hearing aid devices are designed to help hearing-impaired people to compensate their hearing loss. Among other things, they aim to improve the intelligibility of speech, captured by one or multiple microphones in the presence of environmental noise. A binaural hearing aid system consists of two hearing aids that potentially collaborate through a wireless link. Using collaborating hearing aids can help to preserve the spatial binaural cues, which may be distorted using traditional methods, and may increase the amount of noise suppression. This can be achieved by means of multi-microphone noise reduction algorithms, which generally lead to better speech intelligibility than the single-channel approaches [1]. An example of a binaural multi-microphone noise reduction algorithm is the binaural minimum variance distortionless response (BMVDR) beamformer [2, 3], which is a special case of binaural linearly constrained minimum variance (BLCMV)-based methods [4, 5]. The BMVDR consists of two separate MVDR beamformers which try to estimate distortionless versions of the desired speech signal at both left-sided and right-sided hearing aids while suppressing the environmental noise and maintaining the spatial cues of the target signal.

Using binaural algorithms requires that the signals recorded at one hearing aid are transmitted to the contralateral hearing aid through a wireless link. Due to the limited transmission capacity, it is necessary to apply data compression to the signals to be transmitted [6]. This implies that additional noise due to data compression (quantization) is added to the microphone signals before transmission. Typically, binaural beamformers do not take this additional compression noise into account. In [7], one binaural noise reduction scheme based on the generalized sidelobe canceller (GSC) beamformer under quantization errors was proposed. However, the quantization scheme used in [7] assumes that the acoustic scene consists of stationary point sources, which is not realistic in practice. The target signal typically is a non-stationary speech source. Moreover, the far field scenario assumed in [7] cannot support the real and practical analysis of the beamforming performance.

In this paper we study the impact of quantization as a data compression approach on the performance of binaural beamforming. We use the BMVDR beamformer as an illustration, but the findings can easily be applied to other binaural algorithms. Optimal beamformers rely on the statistics of all noise sources, including the quantization noise (QN). Fortunately, the QN statistics are readily available at the transmitting hearing aids. We propose a binaural scheme based on a modified noise cross-power spectral density (CPSD) matrix including the QN in order to take into account the QN. To do so, we introduce two assumptions: i) the QN is uncorrelated across microphones, and ii) the QN and the environmental noise are uncorrelated. The validity of these assumptions depends on the used bit-rate as well as the exact scenario. Under low bit-rate conditions, we show that using subtractive dithering the two assumptions always hold. Without dithering, the assumptions hold approximately for higher bit-rates. However as we show, for many practical scenarios the loss in performance due to not strict validity of these assumptions is negligible.

Based on the BMVDR as a binaural processor, and the binaural output signal-to-noise ratio (SNR) as the performance measure, we show that the modified BMVDR taking into account the QN outperforms significantly the case where the QN is not taken into account, especially at low bit-rates. In addition, the effect of the above-mentioned assumptions on the SNR performance are studied in detail.

2 Signal Model

Typically, a binaural hearing aid consists of two hearing aids which collaborate through a wireless link. Let us assume there are M_{L} and M_{R} microphone sensors embedded in the left-side and right-side hearing aids, respectively, with $M = M_{\mathrm{L}} + M_{\mathrm{R}}$. The beamforming in this paper is performed in the short-term Fourier transform (STFT) domain. Each microphone is assumed to capture the attenuated and delayed version of the target speech signal in the STFT domain, say $S[k,l]$, corrupted by r interfering point sources, $U_j[k,l],\ \ j = 1, ..., r$, and by the internal microphone noise, $V[k,l]$. Indices k and l denote the frequency and frame index, respectively. The signal model in the STFT domain is then given by

$$Y_i[k,l] = A_i[k,l]S[k,l] + \sum_{j=1}^{r} B_{ij}[k,l]U_j[k,l] + V_i[k,l], \quad (1)$$

where $i = 1, ..., M$ is the microphone index, A_i is the acoustic transfer function (ATF) from the target point source to the ith microphone, and B_{ij} is the ATF from the jth interferer to the ith microphone. Using a vector notation by stacking the $Y_i[k,l]$ across microphones, we get

$$\mathbf{y} = \mathbf{x} + \sum_{j=1}^{r} \mathbf{n}_j + \mathbf{v}, \quad (2)$$

where, $\mathbf{y} = [Y_1[k,l], ..., Y_M[k,l]]^{\mathrm{T}}$, $\mathbf{v} = [V_1[k,l], ..., V_M[k,l]]^{\mathrm{T}}$, $\mathbf{x} = \mathbf{a}S$, and $\mathbf{n}_j = \mathbf{b}_j U_j$. Note that $\mathbf{a} = [A_1[k,l], ..., A_M[k,l]]^{\mathrm{T}}$, and $\mathbf{b}_j = [B_{1j}[k,l], ..., B_{Mj}[k,l]]^{\mathrm{T}}$. The superscript "T" represents transpose operator. To simplify the notation, the frequency and frame indices k and l will be omitted. In this paper all point sources, including the target signal and interferes along with the internal microphone noise, are assumed to be mutually uncorrelated. Also, the ith internal microphone noise is assumed to be spatially uncorrelated zero-mean with variance σ_i^2. Without loss of generality we assume all internal microphone noises have the same constant variance, i.e, $\sigma_i^2 = \sigma^2$. Therefore, the CPSD matrix of the noisy signal vector $\mathbf{y}$, denoted by $\mathbf{\Phi_y}$, is written as

$$\mathbf{\Phi_y} = \mathbf{\Phi_x} + \overbrace{\sum_{j=1}^{r} \mathbf{\Phi}_{\mathbf{n}_j} + \mathbf{\Phi_v}}^{\mathbf{\Phi}}, \tag{3}$$

where,

$$\begin{aligned} \mathbf{\Phi_x} &= E[\mathbf{x}\mathbf{x}^{\mathrm{H}}] = \sigma_s^2 \mathbf{a}\mathbf{a}^{\mathrm{H}}, \\ \mathbf{\Phi}_{\mathbf{n}_j} &= E[\mathbf{n}_j\mathbf{n}_j^{\mathrm{H}}] = \sigma_{u_j}^2 \mathbf{b}_j\mathbf{b}_j^{\mathrm{H}}, \qquad j = 1,...,r, \end{aligned} \tag{4}$$

and $\mathbf{\Phi_v} = \sigma^2 I$. Note that $\sigma_s^2 = E[|S|^2]$ is the power spectral density (PSD) of the clean speech signal S. Similarly, $\sigma_{u_j}^2 = E[|U_j|^2]$ is PSD of the jth interfering signal U_j. $E[\cdot]$ and the superscript "H" denote the expectation and the conjugate transpose operators, respectively.

The estimated clean speech signal at the left and right reference microphones is obtained by weighted averaging of all received signals, i.e., $\hat{X}_{\mathrm{L}} = \mathbf{w}_{\mathrm{L}}^{\mathrm{H}}\mathbf{y}$ and $\hat{X}_{\mathrm{R}} = \mathbf{w}_{\mathrm{R}}^{\mathrm{H}}\mathbf{y}$, where $\hat{X}_{\mathrm{L}}$ and $\hat{X}_{\mathrm{R}}$ are the estimated clean signals at the left and right reference microphones, respectively, and $\mathbf{w}_{\mathrm{L}}$ and $\mathbf{w}_{\mathrm{R}}$ are the applied spatial filters. Notice that the use of $\mathbf{w}_{\mathrm{R}}$ and $\mathbf{w}_{\mathrm{L}}$ implies that $\mathbf{y}$ is assumed to be present at both hearing aids, i.e., the noisy microphone signals are exchanged. Wireless exchange of these signals will introduce additional noise due to quantization. In this paper we focus on a simple quantization scheme and investigate the impact of the additional QN on the beamformer performance as a function of the used transmission bit-rate.

3 BMVDR

The BMVDR beamformer is a special case of the BLCMV beamformer [4, 5], and consists of two separate MVDR beamformers

$$\begin{aligned} \mathbf{w}_{\mathrm{L}}^{\star} &= \underset{\mathbf{w}_{\mathrm{L}}}{\mathrm{argmin}} \quad \mathbf{w}_{\mathrm{L}}^{\mathrm{H}}\mathbf{\Phi}\mathbf{w}_{\mathrm{L}} \quad \text{s.t.} \quad \mathbf{w}_{\mathrm{L}}^{\mathrm{H}}\mathbf{a} = A_{\mathrm{L}}, \\ \mathbf{w}_{\mathrm{R}}^{\star} &= \underset{\mathbf{w}_{\mathrm{R}}}{\mathrm{argmin}} \quad \mathbf{w}_{\mathrm{R}}^{\mathrm{H}}\mathbf{\Phi}\mathbf{w}_{\mathrm{R}} \quad \text{s.t.} \quad \mathbf{w}_{\mathrm{R}}^{\mathrm{H}}\mathbf{a} = A_{\mathrm{R}}, \end{aligned} \tag{5}$$

where $\mathbf{\Phi}$ is the CPSD matrix of the noise, see (3). Solving (5), the optimal weight vectors are computed as

$$\mathbf{w}_{\mathrm{L}}^{\star} = \frac{\mathbf{\Phi}^{-1}\mathbf{a}}{\mathbf{a}^{\mathrm{H}}\mathbf{\Phi}^{-1}\mathbf{a}}\bar{A}_{\mathrm{L}}, \quad \mathbf{w}_{\mathrm{R}}^{\star} = \frac{\mathbf{\Phi}^{-1}\mathbf{a}}{\mathbf{a}^{\mathrm{H}}\mathbf{\Phi}^{-1}\mathbf{a}}\bar{A}_{\mathrm{R}}, \tag{6}$$

where $\bar{A}$ is the complex conjugate of a complex number A.

4 Quantization and Dithering

For simplicity, we assume that the data compression scheme is simply given by a uniform r-bit quantizer. Notice that the data is already finite and quantized at high rate (16 bits) at the corresponding hearing aid. The symmetric uniform quantizer maps the actual range of the signal, $x_{\min} \le x \le x_{\max}$, to the quantized range $x_{\min} \le \hat{x} \le x_{\max}$, where $x_{\max} = -x_{\min}$. The quantized value $\hat{x}$ can take one out of $K = 2^r$ different discrete levels. The amplitude range is subdivided into $K = 2^r$ uniform intervals of width $\Delta = (2x_{\max})/2^r$, where $x_{\max}$ is the maximum value of the signal to be quantized [8]. A well-known quantizer is the *midtread* quantizer with a staircase mapping function $f(x)$, defined as $f(x) = \hat{x} = \Delta\lfloor \frac{x}{\Delta} + \frac{1}{2} \rfloor$, where $\lfloor \quad \rfloor$ is the "floor" operation. The quantization error that we refer to in this paper as the QN is denoted by $e = \hat{x} - x$, and is determined by the value of the stepsize Δ. Under certain conditions [9, 10], e has a uniform distribution, that is,

$$p(e) = \begin{cases} \Delta^{-1}, -\frac{\Delta}{2} \le e \le \frac{\Delta}{2} \\ 0, \text{otherwise}, \end{cases} \tag{7}$$

with variance $\sigma_e^2 = \frac{\Delta^2}{12}$. One of the conditions when this happens, is when the characteristic function (CF), which is the Fourier transform of a probability density function, of the variable that is quantized is band-limited. In that case, the QN is uniform. However, the characteristic functions of many random variables are not band-limited (e.g., consider the Gaussian random variable). A less strict condition is that the characteristic function has zeros at frequencies $k\Delta^{-1}, \forall k$ except for $k = 0$. Alternatively, subtractive dithering can be applied, which can be used to guarantee that one of the above conditions is met.

In a subtractively dithered topology, the quantizer input is comprised of a quantization system input x plus an additive random signal (e.g. uniformly distributed), called the dither signal, denoted by v which is assumed to be stationary and statistically independent of the signal to be quantized [10]. The dither signal is added prior to quantization and subtracted after quantization (at the receiver). For the exact requirements on the dither signal and the consequences on the dithering process, see [10]. In fact, subtractive dither assumes that the same noise process v can be generated at the transmitter and receiver and guarantees a uniform QN e that is independent of the quantizer input.

5 Quantization Aware Beamforming

In Sec.2 we assumed that the received signals at the microphones in one hearing aid are transmitted without error to the contralateral side and vice versa. This is not the case in practice. In order to take into account the QN in a beamfroming task, we introduce new noisy signal vectors available at both the left and right hearing aids, say $\mathbf{y}_{\mathrm{L}} = \mathbf{y} + \mathbf{e}_{\mathrm{L}}$ and $\mathbf{y}_{\mathrm{R}} = \mathbf{y} + \mathbf{e}_{\mathrm{R}}$, where $\mathbf{y}$ in defined in (2) and $\mathbf{e}_{\mathrm{L}} = [\mathbf{0}_{M_{\mathrm{L}}}^{\mathrm{T}}, \tilde{\mathbf{e}}_{\mathrm{L}}^{\mathrm{T}}]^{\mathrm{T}}$ with $\mathbf{0}_{M_{\mathrm{L}}}$ the M_{L}-dimensional vector of zeros and $\tilde{\mathbf{e}}_{\mathrm{L}}$ a vector with quantization errors of the signals transmitted from the right side to the left side . Similarly we define $\mathbf{e}_{\mathrm{R}} = [\tilde{\mathbf{e}}_{\mathrm{R}}^{\mathrm{T}}, \mathbf{0}_{M_{\mathrm{R}}}^{\mathrm{T}}]^{\mathrm{T}}$.

Taking into account the QN, the modified BMVDR beamformer is defined as

$$\begin{aligned} \mathbf{w}_{\mathrm{L}}^{\star} &= \underset{\mathbf{w}_{\mathrm{L}}}{\mathrm{argmin}} \quad \mathbf{w}_{\mathrm{L}}^{\mathrm{H}}\mathbf{\Phi}_{\mathrm{nL}}\mathbf{w}_{\mathrm{L}} \quad \text{s.t.} \quad \mathbf{w}_{\mathrm{L}}^{\mathrm{H}}\mathbf{a} = A_{\mathrm{L}}, \\ \mathbf{w}_{\mathrm{R}}^{\star} &= \underset{\mathbf{w}_{\mathrm{R}}}{\mathrm{argmin}} \quad \mathbf{w}_{\mathrm{R}}^{\mathrm{H}}\mathbf{\Phi}_{\mathrm{nR}}\mathbf{w}_{\mathrm{R}} \quad \text{s.t.} \quad \mathbf{w}_{\mathrm{R}}^{\mathrm{H}}\mathbf{a} = A_{\mathrm{R}}, \end{aligned} \tag{8}$$

where,

$$\mathbf{\Phi}_{\mathrm{nL}} = \mathbf{\Phi} + \mathbf{\Phi}_{e_{\mathrm{L}}}, \quad \mathbf{\Phi}_{\mathrm{nR}} = \mathbf{\Phi} + \mathbf{\Phi}_{e_{\mathrm{R}}}. \tag{9}$$

Here $\mathbf{\Phi}_{\mathrm{nL}}$ and $\mathbf{\Phi}_{\mathrm{nR}}$ are the modified CPSD matrices of the total noise including QN corresponding to the left and right beamformer, respectively. Note that $\mathbf{\Phi}_{e_{\mathrm{R}}} = E[\mathbf{e}_{\mathrm{R}}\mathbf{e}_{\mathrm{R}}^{\mathrm{H}}]$ and $\mathbf{\Phi}_{e_{\mathrm{L}}} = E[\mathbf{e}_{\mathrm{L}}\mathbf{e}_{\mathrm{L}}^{\mathrm{H}}]$ such that $\mathbf{\Phi}_{\mathrm{nL}}$ and $\mathbf{\Phi}_{\mathrm{nR}}$ can be reformulated as

$$\begin{aligned} \mathbf{\Phi}_{\mathrm{nL}} &= \mathbf{\Phi} + \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Phi}'_{e_{\mathrm{L}}} \end{bmatrix}, \qquad \mathbf{\Phi}'_{e_{\mathrm{L}}} \in \mathbf{R}^{\mathbf{M}_{\mathrm{R}} \times \mathbf{M}_{\mathrm{R}}}, \\ \mathbf{\Phi}_{\mathrm{nR}} &= \mathbf{\Phi} + \begin{bmatrix} \mathbf{\Phi}'_{e_{\mathrm{R}}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}, \qquad \mathbf{\Phi}'_{e_{\mathrm{R}}} \in \mathbf{R}^{\mathbf{M}_{\mathrm{L}} \times \mathbf{M}_{\mathrm{L}}}. \end{aligned} \tag{10}$$

Note that in (9) and (10) we implicitly assume the QN to be uncorrelated to the environmental noise. If the quantization error is uniform, $\mathbf{\Phi}'_{e_{\mathrm{R}}}$ and $\mathbf{\Phi}'_{e_{\mathrm{L}}}$ are block-diagonal matrices with the elements corresponding to the theoretical variance $\sigma_e^2 = \Delta^2/12$. Note that the objective functions in the modified optimization problems in (8) are functions of the bit-rate r. For simplicity we assume in this paper that all signals are quantized at equal bit-rates. Finally, the beamformed estimates at left and right reference microphones are $\hat{X}'_{\mathrm{L}} = \mathbf{w}_{\mathrm{L}}^{\star\mathrm{H}}\mathbf{y}_{\mathrm{L}}$ and $\hat{X}'_{\mathrm{R}} = \mathbf{w}_{\mathrm{R}}^{\star\mathrm{H}}\mathbf{y}_{\mathrm{R}}$, respectively.

6 Validity of Assumptions

In (9) and (10) it is assumed that the QN ($\mathbf{e}_{\mathrm{L}}$ and $\mathbf{e}_{\mathrm{R}}$) is uncorrelated to the environmental noise ($\sum_{j=1}^{r} \mathbf{n}_j + \mathbf{v}$). In addition, by assuming $\mathbf{\Phi}'_{e_{\mathrm{L}}}$ and $\mathbf{\Phi}'_{e_{\mathrm{R}}}$ to be diagonal, it is also assumed that the QN is uncorrelated across microphones. In this section we introduce two measures to verify the validity of these assumptions.

For a given choice of quantizers, we expect the validity to depend on bit-rate and source position. Experiments will therefore be carried out as a function of source position and bit-rate. For simplicity we only focus on the left beamformer formulations. A similar analysis can be applied to the right beamformer.

6.1 Correlation of quantization noise across microphones

If the QN is truly uncorrelated across microphones, the noise correlation matrix is diagonal. To validate this assumption, we use the following "diagonality measure" of a matrix,

$$D = \frac{\sum_{i=1}^{M_R} [|\mathbf{\Phi}'_{e_L}|]_{ii}{}^2 - \sum_{i=1}^{M_R}\sum_{j=1}^{M_R} [|\mathbf{\Phi}'_{e_L}|]_{ij}{}^2}{\sum_{i=1}^{M_R}\sum_{j=1}^{M_R} [|\mathbf{\Phi}'_{e_L}|]_{ij}{}^2}. \tag{11}$$

This measure can be interpreted as a normalized distance between the sum of all entries and the sum of diagonal entries of the matrix $\mathbf{\Phi}'_{e_L}$. In the worst case, where the signals are highly correlated, all of the entries have the same value (for example value a for each entry) and the lower bound for this measure is $D_{\text{min}}=\frac{M_R a^2 - M_R{}^2 a^2}{M_R{}^2 a^2}=\frac{1}{M_R}-1$. In the best case where the signals are highly uncorrelated, the value D approaches zero. In general, $(\frac{1}{M_R}-1) \leq D \leq 0$, the more negative, the larger off-diagonal entries. The closer to zero, the more diagonally dominant.

6.2 Correlation between quantization noise and environmental noise

In case the environmental noise and the quantizer noise are uncorrelated, the sum of the two CPSD matrices $\mathbf{\Phi}_{e_L}$ and $\mathbf{\Phi}$ should be equal to the CPSD matrix of the total noise, $\mathbf{\Phi}_{\text{nL}}$ according to (9). To measure whether this assumption holds, we compare the normalized difference between the estimated values of the right side and the left side of the first equation in (9) as

$$E = \frac{\sum_{i=1+M_L}^{M}\sum_{j=1+M_L}^{M} [|\mathbf{\Phi}_{\text{nL}} - \mathbf{\Phi} - \mathbf{\Phi}_{e_L}|]_{ij}{}^2}{\sum_{i=1+M_L}^{M}\sum_{j=1+M_L}^{M} [|\mathbf{\Phi}_{\text{nL}}|]_{ij}{}^2}. \tag{12}$$

7 Experiments

In this section we present experimental results comparing the proposed method with other traditional beamformers that do not take QN into account. Moreover, we investigate the assumptions on the QN.

7.1 Setup and Simulation Parameters

A typical acoustic scene, which we use in this paper, is illustrated in Fig. 1. In the experiments the exact source positions are not necessarily the same as those in Fig. 1. For all experiments there is one target speech, shown by green circle in Fig.1, recorded at 16 kHz sampling frequency with duration of around 12.5 seconds. Four stationary interfering signals, shown by black triangles in Fig.1, are present at different angles, say $\theta = \tan^{-1}(\frac{y}{x}) - \frac{\pi}{2}$, and different distances form the origin $((x,y) = (0,0))$, say $R = \sqrt{x^2+y^2}$. In this paper, we define θ in a way that zero degree corresponds to the front of the virtual head (like green circle in Fig. 1). Four "$+$" symbols denote four virtual omni-directional microphones, two of them at the left virtual hearing aid and two of them at the right one. Two microphones at each hearing aid form a linear array in direction of y-axis having a distance of 1.2 cm. The distance between two hearing aids (two linear arrays) is 20 cm. The beamforming is performed independently on 512 DFT points frame signals shifted by 256 points (50% overlapping). The output SNR performance is measured at the left reference microphone position, averaged over all frequency bins and time frames. The CPSD matrix of the noise is calculated from the known true ATFs of the interferes and estimated PSDs using Welch's method.

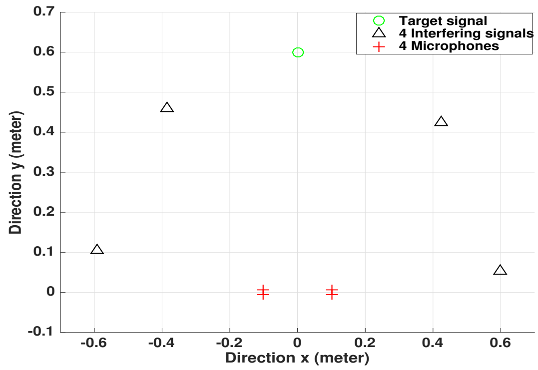


Figure 1: An example acoustic scene. Note that the exact source positions can be potentially different in the experiments.

7.2 Validation of Assumptions: Results

Based on the two measures, introduced in Sec.6, we evaluate for which bit-rates the assumptions hold. Moreover, we apply dithering (Sec.6) as a decorrelation process to assure that the assumptions on the QN in (9) and (10) are valid for all positions and bit-rates. All experiments in this sub-section are carried out as a function of the position of one of the noise sources in terms of angles with respect to the microphone array with a distance 2m from the origin. All three other fixed interfering sources are located at $\{(R,\theta)|(0m,0°),(2m,-90°),(2m,90°)\}$ and the target signal is positioned at $(2m,90°)$. Note that the source positions are different form those in Fig. 1. We use this setup for two reasons

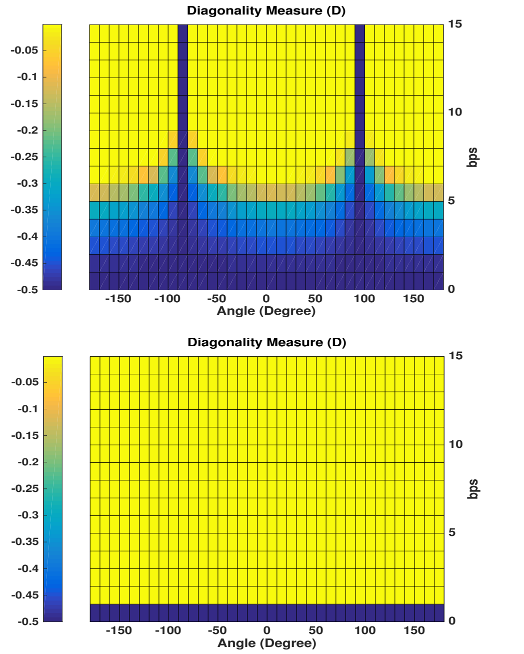
- If four microphones and four interfering signals are present in the acoustic scene, then the cross-PSD of the noise is full rank and invertible.
- the positions of the three interfering signals are symmetric with respect to that of each hearing aid, i.e., identical versions of these signals received at each hearing aid microphones such that they have no effect on the diagonality measure in (11). Therefore, we can isolate the effect of position dependency of the noise source on the total performance.

The results of the D measure in (11) in terms of the bit per sample (bps) and the angle, before and after dithering are shown in Fig. 2a and Fig. 2b. As shown, at higher rates the assumption holds and the CPSD matrix of the QN ($\mathbf{\Phi}'_{e_L}$) becomes more-and-more diagonal ($D \to 0$) with increasing rates. The results show that if the interfering source is positioned at either $\pm$90 degrees (left or right side of the virtual head), the $\mathbf{\Phi}'_{e_L}$ is fully correlated even at high rates, i.e., $D = -0.5$. After applying dithering, $\mathbf{\Phi}'_{e_L}$ becomes diagonal at all rates and angles, as shown in Fig. 2b. Similarly, the results of the "correlation measure" (E in (12)) are shown in Fig. 3a and Fig. 3b in terms of the bps and the angle, before and after dithering, respectively. As shown, the error E decreases as bit-rate increases. After applying dithering the error decreases significantly (from the maximum value of 0.109 in Fig. 3a to the maximum value of 0.0013 in Fig. 3b), even at low bit-rates. This means that after dithering the QN and environmental noise become almost uncorrelated at all rates and angles.

7.3 Performance Evaluation

We compare the results of the following cases in terms of the output SNR for the left-sided reference microphone.

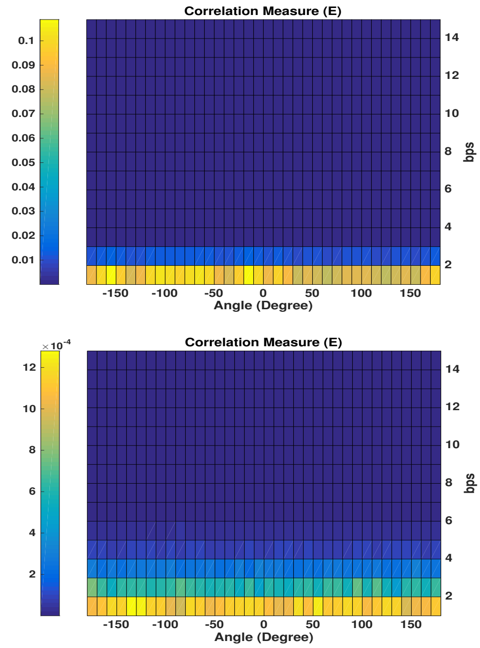
- Case 1) **monaural beamformer**: there is no transmission from one side to the contralateral side, i.e., no wireless link.
- Case 2) **full binaural beamformer**: All microphone signals are assumed to be available without error at the contralateral hearing aid.



[a]

[b]

Figure 2: Diagonality measure:(a) without, and (b) with dithering



[a]

[b]

Figure 3: Correlation measure:(a) without, and (b) with dithering

- Case 3) **Proposed method version 1 without dithering**: beamforming based on (8), i.e., taking into account the QN by estimating the modified total noise CPSD $\mathbf{\Phi}_{\mathrm{nL}}$. Note that in this case as the QN is not assumed to be uncorrelated to the environmental noise, and extra information (directly estimated $\mathbf{\Phi}_{\mathrm{nL}}$) should be transmitted.
- Case 4) **Proposed method version 2 without dithering**: beamforming based on (8), i.e., taking into account the QN by estimating the modified total noise CPSD $\mathbf{\Phi}_{\mathrm{nL}}$ using (9). Therefore unlike the case 3, the QN and the environmental noise is assumed to be uncorrelated and no extra information need be to transmitted.
- Case 5) **traditional BMVDR**: beamforming based on (5), i.e., without taking into account the QN.

We also evaluate the cases 3 and 4 with dithering. The output SNR performance with respect to the left reference microphone is shown in Fig. 4 in terms of bit-rate. Note that the x-axis is number of bits per samples varying from 1 to 15 integers (15 integer points). In this experiment Four interferes are located at $\{(2m,-85^\circ),(2m,-45^\circ),(2m,40^\circ),(2m,80^\circ)\}$, and the target speech is located at $(R,\theta)=(2m,0^\circ)$ (different positions from those in Fig. 1). The input SNR at left reference microphone is approximately 20dB (black dash-dot line). As shown, the cases 3 and 4 in which the QN has been taken into account, outperforms significantly the case 5 (red dashed line) without taking into account the QN, especially for low bit-rates. Note that the SNR performance of the cases 3 and 4 with and without dithering are always in between those of the cases 1 (blue dotted line) and case 2 (black solid top line). At very low rates the SNR values of those cases are close to that of the monaural beamforming (case 1). In fact, the modified BMVDR ignores the noisy low-bit signals so that it is actually acting as a monaural MVDR. As rate increases the SNR approaches to that of the full binaural beamforming (case 2).

As shown in Fig. 4, the four lines according to the four cases 3 and 4 with and without dithering fall almost exactly on top of each other. It means that the SNR gaps between those cases are negligible (maximum gap is less than 0.1 dB). In fact, at very low rates (1-3bps) the QN is large which means a smaller contribution of the transmitted signals to the output beamformed signal. Therefore, although the assumptions might not hold exactly, but the impact of the invalidity of the assumptions on the output signal is very small. As assumptions tend to be valid at higher rates the gaps between those four cases approach zero.

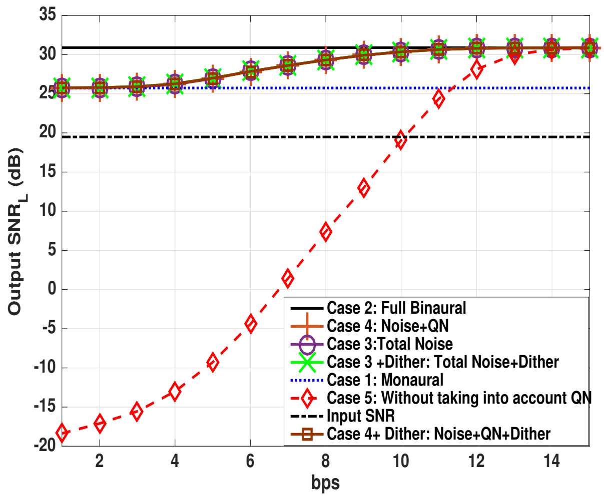


Figure 4: Output SNR performance for the left-sided reference microphone.

8 Conclusions

In this paper we studied the impact of quantization on binaural multi-microphone noise reduction algorithms. As an illustration we proposed a new scheme of quantization aware BMVDR beamforming. The new approach is based on the modified CPSD matrix of the noise including the QN. Assumptions on the QN, which are introduced in sec.6, were investigated experimentally. We conclude that applying dithering as a decorrelation process can guarantee the validity of the assumptions for all bit-rates and source positions. Based on the output SNR performance, the proposed speech enhancement method outperformed significantly the traditional BMVDR, especially for low bit-rates. In addition, different versions of the proposed method with and without applying dithering were evaluated. Generally speaking, in many practical scenarios the output SNR gaps between the proposed method with dithering and the one without dithering are negligible.

References

[1] K. Eneman et al, “Evaluation of signal enhancement algorithms for hearing instruments,” in *16th European Signal Processing Conference*, pp. 1–5, Aug 2008.

[2] S. Haykin and K. J. R. Liu, “Handbook on array processing and sensor networks,” pp. 269–302, 2010.

[3] S. Markovich-Golan, S. Gannot, and I. Cohen, “A reduced bandwidth binauralmvdr beamformer,” in *International Workshop on Acoustic Signal Enhancement (IWAENC)*, Sep 2010.

[4] E. Hadad, S. Gannot, and S. Doclo, “Binaural linearly constrained minimum variance beamformer for hearing aid applications,” in *Acoustic Signal Enhancement; Proceedings of IWAENC 2012; International Workshop on*, pp. 1–4, Sept 2012.

[5] E. Hadad, S. Doclo, and S. Gannot, “The binaural lcmv beamformer and its performance analysis,” *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 24, pp. 543–558, March 2016.

[6] O. Roy and M. Vetterli, “Rate-constrained collaborative noise reduction for wireless hearing aids,” *IEEE Transactions on Signal Processing*, vol. 57, pp. 645–657, Feb 2009.

[7] S. Srinivasan, A. Pandharipande, and K. Janse, “Beamforming under quantization errors in wireless binaural hearing aids,” *EURASIP Journal on Audio, Speech, and Music Processing*, vol. 2008, no. 1, pp. 1–8, 2008.

[8] P. Vary and R. Martin, “Digital speech transmission: Enhancement, coding and error concealment,” *John Wiley-Sons*, 2006.

[9] A. Sripad and D. Snyder, “A necessary and sufficient condition for quantization errors to be uniform and white,” *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 25, pp. 442–448, Oct 1977.

[10] S. P. Lipshitz, R. A. Wannamaker, and J. Vanderkooy, “Quantization and dither: A theoretical survey,” *Audio Eng. Soc.*, vol. 40, pp. 355–375, 1992.