Rate-constrained multi-microphone noise reduction for hearing aid devices

Amini, J.

DOI
10.4233/uuid:54fa083f-5ddf-4b6c-b663-a1b61c6681f5

Publication date
2021

Document Version
Final published version

Citation (APA)
Amini, J. (2021). *Rate-constrained multi-microphone noise reduction for hearing aid devices*. [Dissertation (TU Delft), Delft University of Technology]. https://doi.org/10.4233/uuid:54fa083f-5ddf-4b6c-b663-a1b61c6681f5

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Rate-Constrained Multi-Microphone Noise Reduction for Hearing Aid Devices

RATE-CONSTRAINED MULTI-MICROPHONE NOISE REDUCTION FOR HEARING AID DEVICES

Proefschrift

ter verkrijging van de graad van doctor
aan de Technische Universiteit Delft,
op gezag van de Rector Magnificus prof. dr. ir. T.H.J.J. van der Hagen,
voorzitter van het College voor Promoties,
in het openbaar te verdedigen op dinsdag 13 april 2021 om 12:30 uur

door

Jamal AMINI

Elektrotechnisch ingenieur,
Technische Universiteit Delft, Delft, Nederland
geboren te Tehran, Iran.

Dit proefschrift is goedgekeurd door de promotoren

promotor: Prof. dr. ir. R. Heusdens
promotor: Dr. ir. R. C. Hendriks

Samenstelling promotiecommissie:

Rector Magnificus,	voorzitter
Prof. dr. ir. R. Heusdens,	Technische Universiteit Delft
Dr. ir. R. C. Hendriks,	Technische Universiteit Delft
Onafhankelijke leden:	
Prof. dr. J. Østergaard	Aalborg U., Denemarken
Prof. dr. ir. S. Doclo	U. of Oldenburg, Germany
Prof. dr. F.M.J. Willems	TU Eindhoven
Prof. dr. ir. A.J. van der Veen	Technische Universiteit Delft
Prof. dr. ir. J.H. Weber	Technische Universiteit Delft
Prof. dr. ir. A. Yarovoy	Technische Universiteit Delft, reservelid

This work was supported by the Netherlands Organisation for Scientific Research (NWO), and the hearing-aid company Oticon A/S, under the project entitled “Spatially Correct Multi-Microphone Noise Reduction Strategies Suitable for Hearing Aids”. We would like to thank Prof. dr. ir. Jesper Jensen and dr. Meng Guo for their significant contributions to this dissertation.

Nothing endures but change.

Heraclitus

CONTENTS

SUMMARY

Many people around the world suffer from hearing problems (In the Netherlands, around 11% of the population is considered hearing-impaired). To overcome their hearing problems, advanced technologies like hearing aid devices can be used. Hearing aids are meant to assist the hearing-impaired to improve the speech intelligibility and the quality of sounds that they intend to hear. Usually these include processors which are mainly designed to enhance the sound signals originating from the source of interest by reducing the environmental noise. Binaural hearing aids, on the other hand, can also help to preserve some spatial information from the acoustic scene, which can help the hearing aid user to hear the sounds from the correct locations. To construct the binaural hearing aid system, two hearing aids are needed to be placed in the left and the right ears, which can potentially communicate through a wireless link. In addition, one can think of additional assisting devices with microphones placed in the environment. One common way to reduce the noise is to use advanced binaural multi-microphone noise reduction algorithms, which aim at estimating some desired sources while reducing the power of the undesired sources. One typical method is to use spatial filtering, which aims at estimating the target signal by shaping the beam towards the location of the desired source while canceling/suppressing the other sources.

To perform binaural noise reduction, while assuming centralized processing, the signals recorded at remote microphones (for example from additional assisting devices or in the binaural hearing aid setup, the sound signals from the contralateral hearing aid) need to be transmitted to the central processor. Due to the power and bandwidth limitations, the data needs to be compressed before transmission. Therefore, the main question would be, at which rate the data should be compressed to have reasonably good noise reduction performance. This links the noise reduction problem to the data compression problem. Generally, the higher the data rate, the better the noise reduction performance. Therefore, there is a trade-off between the performance of the noise reduction algorithm and the data-rate at which the information is compressed. This problem is closely connected to the rate-distortion problem from an information-theoretic viewpoint. Studying the effect of data compression on the performance of noise reduction problems would be of great interest to reduce the power consumption of hearing assistive devices.

One way to incorporate data compression into the noise reduction problem is to perform quantization, which leads to a rate-constrained noise reduction problem. In the rate-constrained noise reduction, the goal is to estimate the desired sources based on the imperfect data. The observations from remote sensors are quantized and transmitted to the fusion center. The main challenge in the binaural rate-constrained noise reduction is to find the best quantization rates for the different sensors at different frequencies, given the physical constraints like bitrate and power constraints.

Another aspect of the rate-constrained noise reduction is to expand the network to

receive more information on the acoustic scene using additional assistive devices. Target source estimation using information from such assistive devices (rather than only binaural hearing aids) is shown to result in better noise reduction performance. Now the question is how to allocate the bitrates to the assistive devices as well. These assistive devices can be thought of as the remote embedded microphones on the cell-phones (mobile) or wearable microphones placed at the users' bodies. The binaural hearing aid system can thus be generalized to allow other assistive devices to contribute to noise reduction.

In this dissertation, we study and propose different rate-constrained multi- microphone noise reduction algorithms. We try to expand the notion of the binaural rate-constrained noise reduction to multi-microphone rate-constrained noise reduction for general wireless acoustic sensor networks (WASNs). The WASN in this case can include the binaural setup along with other assistive devices. We propose different algorithms to cover the main objectives of rate-constrained noise reduction problems. These objectives mainly include good target estimation (less environmental noise power) given the compressed data, good rate allocation strategies in WASNs, and preferably preserved spatial information of the sources in the acoustic scene to get the correct impression of the acoustic scene.

1

INTRODUCTION

All truly great thoughts are conceived by walking.

Friedrich Nietzsche

In the Netherlands, approximately 11% of the population is hearing-impaired [1, 2]. Due to their hearing loss, these people suffer from a worse speech intelligibility and have worse abilities to localize sound sources, especially in acoustically challenging situations. Hearing aid (HA) devices have shown to be effective solutions to help such people and can provide a better understanding of the speech signals in the acoustic scene. Typically, the HAs contain one or more embedded microphones to capture the acoustical information from the environment and then aim to improve the speech intelligibility with respect to one or more sources of interest while suppressing the environmental noise. This process is usually referred to as noise reduction [3]. A high-level view of the noise reduction scheme in HAs is shown in Figure 1.1. The acoustical information is recorded by the microphones. The processor filters the digitalized microphone observations (digitalized by the analog-to-digital converter (ADC)), and then, after taking into account the user-specific adjustments by the HA (e.g., hearing loss compensation), the noise reduced signal with increased intelligibility will be converted to the analog signal and played back by the embedded loudspeakers. Due to the size limitations of hearing aids, the microphones and loudspeakers are placed close to each other. Therefore, the loudspeaker's output signal can partially leak back into the microphone recordings. This artifact is known as the feedback problem. To overcome feedback, it is crucial to include a feedback cancellation system along with a central processor [4, 5].

The users' hearing loss can be (partially) compensated using the audiogram of the hearing-impaired person and by applying the proper amplification at the corresponding frequencies. In this thesis, we will not consider the feedback problem and the hearing loss compensation but assume that the HA will perfectly compensate for both the generated feedback, as well as the hearing loss.

Noise reduction algorithms [6–9] typically combine the different microphone observations to perform multi-microphone noise reduction (also often referred to as beam-

1

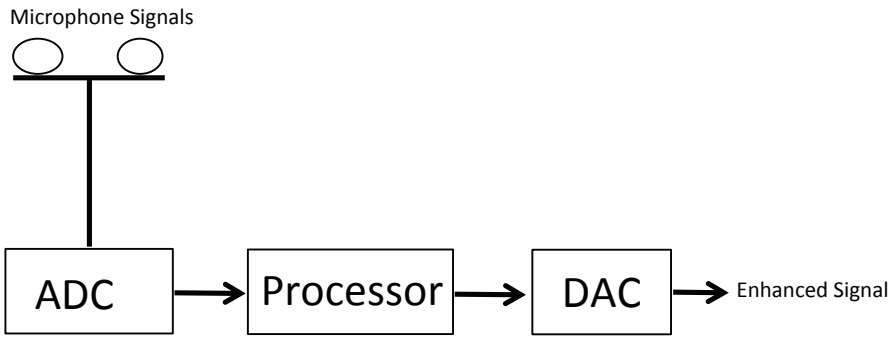


Figure 1.1: High-level schematic of the hearing aid noise reduction system.

forming). Roughly speaking, the more microphone observations are involved in the processing, the better the performance [10]. Compared to single-microphone algorithms, multi-microphone noise reduction allows for spatial as well as temporal filtering, generally leading to better performance.

For the placement and use of the microphones in the HAs, several setups are possible:

- Using multiple microphones per device with only local processing. For example, in the case of the HA, it is possible to have more than one microphone embedded in the HA, where none of the signals is shared with the contralateral device. This setup is often referred to as the monaural HA setup[11]. In this case, the observations are locally available by the processor and there is no need to transmit the observation to another processor.

- Using multiple microphones in two devices, where the signals are shared. For example, two HAs can collaborate through a wireless link to exchange information. Then the HAs can construct a binaural HA system. In this case, the observation from the contralateral device must be transmitted to the HA. This can lead to an increased amount of noise reduction, as spatial diversity can be better-exploited [12].

- Using multiple microphones in more than two devices. For example, the binaural HA setup can collaborate with additional assistive devices in the vicinity of the user. This can potentially provide better noise reduction performance, as the spatial diversity is increased and can be exploited even more, as the assistive devices might have more valuable spatial information about the sound sources in the acoustic scene [13]. In this case, it is required that also the microphone signals from the assistive devices are transmitted.

Although noise reduction performance is an important factor in multi-microphone noise reduction algorithms, localization performance by the user is equally well impor-

tant. Human sound localization is predominantly done by exploiting phase and amplitude differences between the two ears. As multi-microphone noise reduction algorithms are based on adjusting phase and amplitude differences between the microphone signals, multi-microphone algorithms will harm the localization performance of the user if no special counter-measures are taken. The spatial information of the sound sources should, therefore, be carefully taken into account to provide a natural impression of the acoustic scene [14]. Altogether, these phase (time) and amplitude differences that a source has with respect to the different microphones are usually referred to as the spatial cues [8]. Preservation of the spatial cues may lead to a more natural impression of the acoustic scene. Several beamforming algorithms, e.g. [14, 15], have been developed to explicitly try to preserve such spatial information. However, as this usually comes in the form of spatial constraints on the beamformer filter coefficients, this sacrifices the noise reduction performance [15]. Therefore, there is a trade-off between the amount of noise reduction and the preservation of the spatial cues.

In scenarios where there is more than one device involved in the noise reduction process, the observations need to be transmitted. The processing can be done in distributed form [16–18] or centrally using a fusion center (FC) [19]. In this thesis, we will mainly focus on central FC-based processing. This requires the necessary data to be available at the FC to be processed and finally to output an estimated desired signal. As these devices are typically battery-powered, the power consumption of such wireless devices should be considered when designing noise reduction algorithms. Among all processes, data transmission has a huge impact on the power consumption [20]. The rate that is used to transmit the remote microphone observations (those which are not locally available at the FC) should therefore be constrained in the noise reduction algorithms that are meant to be used in such (small) wireless acoustic sensor networks. [21].

Using higher rates for data compression, the observations will be more informative for the processor and, hence, the performance of the noise reduction will increase. Therefore, there is a trade-off between the rate of transmission of the data and the noise reduction performance. From an information-theoretic point of view, this is referred to as the rate-distortion trade-off [22–25]. Typically the noise reduction performance is defined by fidelity criteria or a distortion function. However, higher rates will consume more power. Therefore, there is a trade-off between the energy that is consumed for transmission of data, i.e., the rate of transmission, and the noise reduction performance. Looking at the problem from the noise reduction perspective, we can argue that there is a link between data compression and estimation of the desired signal, which turns to the notion of rate-constrained noise reduction [21]. There are several ways to compress the data before transmission. One common way to compress the data is quantization. Prior to transmission, the observation is quantized at a certain bit-rate. This raises the important question at which bit-rate the observations should be quantized to have a good trade-off between the estimation accuracy and consumption of the limited bit budget? In this dissertation, we will focus on answering this question from the multi-microphone noise reduction perspective.

1

1.1. MULTI-MICROPHONE NOISE REDUCTION BASED ON LINEAR ESTIMATION

In this section, we give a brief overview of the different existing multi-microphone noise reduction algorithms that are based on linear estimation. Generally, these algorithms try to estimate one or more sources of interest in the acoustic scene by combining the microphone observations. Linear estimation-based algorithms try to solve an optimization problem that aims at minimizing a distortion function of the estimation error between the source of interest and a linear constrained estimate of that signal.

The multi-channel Wiener filter (MWF) [26] is one of the most well-known linear estimators which tries to estimate sources of interest by minimizing the mean square error (MSE) between the source of interest and its estimate. The solution to this optimization problem is a vector of weights, say $\mathbf{w}$, which needs to be applied to the noisy microphone signals to project the observation onto a single estimated target signal. It is proven that the MWF has the best noise reduction performance in MSE sense among all other linear MSE-based methods [27]. However, if the prior distribution of the target signal is taken into account, better (non-linear) estimators can be derived if the prior is non-Gaussian, e.g., [29]. In the original MWF, the preservation of sources is not considered, meaning that the optimization problem only tries to minimize the MSE without imposing any constraint on preserving the target. In other words, the target signal may be distorted after applying the optimal weights to the noisy microphone signals, as there is no constraint in the optimization problem to keep the target signal un-distorted.

The minimum variance distortionless response (MVDR) [27, 30] is a well-known technique that aims at minimizing the output noise power (which can be reformulated as the MSE for a single source of interest), while keeping the target signal un-distorted by adding a distortionless constraint to the optimization problem. As a price, the noise reduction performance of the MVDR is worse than that of the MWF, as there is less degree of freedom for the MVDR to further minimize the noise power. A more generalized version of the MVDR is the linearly constrained minimum variance (LCMV) noise reduction technique, which allows us to include a set of linear constraints to the noise reduction problem. These constraints can be used to preserve specific sources, cancel specific sources, or, as will be discussed in Section 1.2, to preserve the spatial cues of specific sound sources in a binaural setting. Including additional constraints reduces the degrees of freedom for the algorithm even more. The noise reduction performance of the LCMV might, therefore, be even worse than that of the MVDR.

In the binaural setting, two HAs are considered. One for the left ear and one for the right ear. In such a setting, each HA outputs an estimate of the target signal. However, sound localization is to a large extent based on time (or phase) and magnitude differences between the two. Without carefully aligning the amplitude and phase differences between these two outputs, the spatial cues of the estimated sound source will be destroyed. One common binaural cue is the difference in arrival time of the sound source between the left and the right ears, which is called interaural time difference (ITD). Another important binaural cue is the level difference between the left and the right ear, of the sound source, which is called the interaural level difference (ILD). In the frequency domain, the ITD is transformed into the interaural phase difference (IPD). These bin-

aural cues provide spatial information of the sound sources in the acoustic scene. The human auditory system typically uses the IPDs of the low-frequency components (usually below 1.5 kHz) and the ILDs of components above 3 kHz [31]. In the next section, we will explain how binaural noise reduction algorithms can preserve the above-mentioned binaural cues.

1.2. BINAURAL MULTI-MICROPHONE NOISE REDUCTION BASED ON LINEAR ESTIMATION

Binaural HA systems consist of two HAs which can potentially collaborate through a wireless link, as shown in Figure 1.2. This can provide an extended microphone array, which can lead to better noise reduction. Typically the binaural multi-microphone noise reduction methods, which are based on linear estimation, consider two fusion centers (FCs), one in each ear, that aim at estimating two versions of the target signal (one for each ear), while reducing the environmental noise.

Over the last decade, several binaural multi-microphone noise reduction algorithms have been proposed [27, 32]. These algorithms can be categorized based on the objective function to be optimized (for example, MSE or output noise power) and based on the constraints which are designed to preserve the spatial cues of the sources (target signal or interferers). The types of constraints used in these algorithms can also be different. For example, spatial cues of the sources can be completely preserved, when equality constraints [8, 35–37] are applied. On the other hand, to approximately (not exactly) preserve the spatial cues, inequality constraints [14, 34]) are applied which can lead to better noise suppression compared to the case with equality constraints. In this dissertation, we will mainly focus on equality-constrained binaural multi-channel noise reduction filters.

Binaural multi-channel Wiener filter (MWF) is a well known mean square error-based noise reduction algorithm [38]. The algorithm tries to minimize the MSE of the target signal estimated at the left and the right reference microphones of the two FCs. No constraints are imposed in the optimization problem which may lead to a distorted target signal. To reduce the target distortion, the binaural speech distortion weighted MWF (BSD-MWF) method [39, 40], has been proposed which provides a parametric trade-off between the performance of noise reduction and the target distortion. However, this method will distort the binaural cues of the interferers.

To have an undistorted target signal at the two reference microphones, the binaural minimum variance distortionless response (BMVDR) beamformer [27] minimizes the output noise power under two linear distortionless constraints. However, imposing two constraints will reduce the degree of freedom, leading to less noise reduction performance than that of the binaural MWF. To preserve the spatial cues of multiple sources (desired source and multiple interfering signals), the binaural linearly constrained minimum variance (BLCMV) beamformer [33], is used, which includes additional constraints for preserving the interferers' interaural transfer function between the two ears. With certain considerations, the optimal BLCMV (OBLCMV) [8] can lead to better noise reduction performance, when comparing with the BLCMV. The OBLCMV beamformer, however, has less degrees of freedom compared to [35, 36]. In [35, 36] a method is proposed

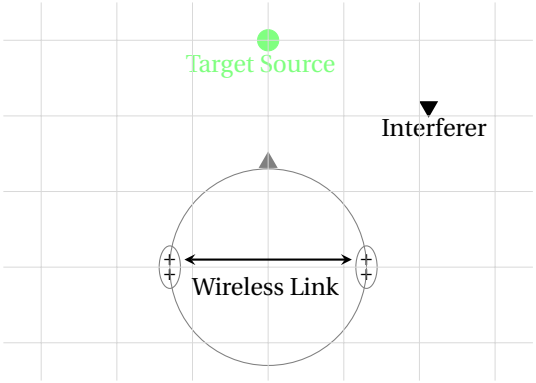


Figure 1.2: An Example Binaural Hearing Aid Setup.

which introduces a set of linear equality constraints (firstly introduced in [41]) to increase the degree of freedom of the optimization problem. Therefore, comparing with the BLCMV and the optimal BLCMV, the method enables preserving more interferers, for a given number of microphones.

An important aspect of binaural filters is the requirement that information from one HA is transmitted to the other HA (e.g. through a wireless link) in order to be combined with local observations. Typically, transmission capacities are limited due to limited battery life-time [21, 42], which necessitates data compression. Ideally, the algorithm trades off the transmission bit-rate of contralateral HA observations against the estimation error on the estimated target signal [21], which is remotely (i.e., indirectly after being filtered by the room channel) observable at the HAs. Therefore, it is crucial to study the effect of the data compression on the performance of the noise reduction algorithms. In the next section, we will mention existing algorithms, which take into account the rate of transmission in the noise reduction problem.

1.3. RATE-CONSTRAINED NOISE REDUCTION

In all binaural noise reduction methods, mentioned in the previous section, the two FCs of the binaural beamformers each estimate the target source with respect to their corresponding reference microphone. To calculate these estimates, both FCs are in need of the microphone recordings from all sensors. This means that observations form the contralateral devices, and potentially any other device included in the setup should be transmitted to the FCs. As the devices have a limited amount of resources (here transmission bandwidth) due to the limited battery lifetime, the total bit-rate used for transmission is constrained. Therefore in addition to the environmental noise in the signal model, the quantization/compression noise should also be included and noise reduction methods should be quantization aware. In [43] the effect of the uniform quantization on the BMVDR Method was studied and the assumptions on the second-order statistics of the environmental noise and the quantization noise, are investigated.

Looking at the problem from an information-theoretic perspective, the estimation (beamforming) problem can be seen as remote source coding [22–24]. The beamformer

with respect to the one of the HAs, combines the decoded transmitted signals from the contralateral HA with its local observations. These local observations can be thought of as side information. Therefore, more accurately, the binaural estimation problem can be viewed as remote source coding with side information at the decoder. This problem is referred to as Wyner-Ziv (WZ) coding [25], for directly observable sources, and for indirectly observable (i.e., remote) sources as remote WZ coding [46]. This will be explained in more detail in the next chapter. Based on the remote source coding scheme, a binaural rate-constrained beamforming problem is presented in [21], assuming jointly Gaussian random sources. The method provides an upper bound on the performance of the minimum MSE (MMSE)-based binaural rate-constrained beamforming algorithms, since it finds an optimal tradeoff between the transmission rate and the MSE between the target signal and its estimate. However, since the algorithm inevitably requires the knowledge of the (joint) statistics at both HAs, therefore, this limits the application of the method in practice. The joint statistics between the two HA observations need to be estimated, and this requires the realizations to be exchanged between the HAs. Moreover, the upper bound on the performance is derived assuming that there are only two processing nodes, that are, the two HAs. The optimal approach does not consider scenarios with more than two nodes, and this is still an open problem.

As practical alternatives to the optimal method in [21], sub-optimal rate-constrained noise reduction methods are proposed in [21], [47–49] in which functions of the observations from the contralateral HA are transmitted, projecting the multi-microphone signals onto a single signal. Unlike the optimal method in [21], these methods do not need the knowledge of the (joint) statistics. However, as these suboptimal methods blindly project the multi-microphone signals onto one signal, a significant mismatch will occur in the performance even at sufficiently high rates [47]. In [49], an MWF-based binaural noise reduction method is proposed, in which local estimates of the target signal are assumed to be iteratively exchanged error-free between the HAs without any rate constraint. Assuming that there is only one target signal, the performance of the iterative algorithm converges to that of the binaural MWF after sufficient transmissions between HAs, as shown in [49]. However, taking rate constraints into account in the iterative method [50], unlike the optimal method, the rate-constrained method is sub-optimal, since the quantization stage of the processing does not use side information aware coding scheme. Explaining such sub-optimal algorithms in a unified framework as done in [51], the sub-optimal approaches pre-filter the multi-microphone observation before quantization without knowing the joint statistics. This may help the process to be faster and simpler compared to the optimal method in [21]. However, in the pre-filtering stage, some important information may be lost as these sub-optimal approaches do not consider the joint statistics between the two HA observations, and thus, the performance will not approach that of the optimal algorithm, even at infinitely high rates. In fact, to keep the necessary information in the pre-filtering stage and to resolve this asymptotic sub-optimality issue, any knowledge (even incomplete) about the joint statistics may be helpful which motivates estimating the joint statistics.

To summarize the shortcomings of the existing optimal/sub-optimal rate-constrained noise reduction methods, the existing optimal/sub-optimal methods have the following limitations

- Asymptotic sub-optimality of the methods due to the blind projection of multiple observations onto a single observation before transmission (pre-filtering).
- Inevitable requirement of knowledge on the joint statistics at both HAs in the optimal method [21].
- Considering only two processing nodes in the optimal method. Scenarios with more than two processing nodes are not considered in [21].

1.4. GOAL OF THE DISSERTATION

The work which has been covered out in this dissertation was funded by the Netherlands Organisation for Scientific Research (NWO), and the hearing-aid company Oticon A/S, under the project entitled "Spatially Correct Multi-Microphone Noise Reduction Strategies Suitable for Hearing Aids". The project includes two sub-projects entitled “Spatially Optimal Multi- Microphone Noise Reduction Techniques" and "Rate-Constrained Multi-Microphone Noise Reduction Techniques”. In this dissertation, we mainly focus on the latter sub-project, to study and propose new methods to deal with the rate-constrained problem. Altogether, in this thesis we try to answer the following research questions:

1- What is the effect of the quantization on the noise reduction performance, and how do quantization related assumptions affect the performance of the quantization aware noise reduction?

2- As mentioned in the previous section, the optimal binaural rate-constrained method in [21] unavoidably requires the knowledge of the joint statistics at both processing nodes. Can we design a coding algorithm from an information-theoretic point of view, which can inherently estimate the joint statistics to be applied to provide an optimal solution at least for one processor?

3- Existing methods for rate-constrained binaural noise reduction consider only two processing nodes and some of them do not take the acoustic scene dependency into account. Can we generalize the binaural hearing set up with a smart rate allocation technique to enable more assistive devices to cooperate to improve the noise reduction performance?

4- Most of the existing rate-constrained problems do not take the preservation of spatial cues into account when designing the optimal rate allocation algorithms. Can we efficiently link the rate-constrained problem to have a spatially correct rate-constrained noise reduction system?

1.5. ORGANIZATION OF THE DISSERTATION AND CONTRIBUTIONS

In this section, we summarize the contributions of the dissertation in the following chapters.

1.5.1. CHAPTER 2

In this chapter, we present the necessary background literature that is required in order to read the remaining chapters of this thesis. We explain the fundamentals in linear estimation based noise reduction algorithms and describe the notion of the rate-distortion trade-off. We first present the signal model that we use in the remaining chapters. Then the mathematical formulation of the linear estimation based binaural multi-microphone noise reduction algorithms is presented. Finally, the theory of the rate-distortion trade-off will be explained in brief from the information-theoretic viewpoint and different rate-distortion trade-offs for different coding scenarios will be summarized.

1.5.2. CHAPTER 3

In this chapter, we study the effect of uniform quantization on the noise reduction performance based on the MVDR beamformer. The binaural setup will be considered as an example acoustic scene. Most of the content in this chapter is based on our proposed conference paper in [43]. We investigate the assumptions made on the second-order statistics of the environmental noise as well as those of the quantization noise. We also investigate the effect of dithering on the second-order statistics. This chapter tries to answer the first research question made in Section 1.4.

1.5.3. CHAPTER 4

As argued in the last part of Section 1.3, the inevitable requirement of knowledge of joint statistics at both HAs in the optimal method [21] motivates us to find a way to estimate the joint statistics in a rate-distortion sense. First, we present a unified framework to study the performance of the existing optimal and sub-optimal rate-constrained beamforming methods for binaural HAs, followed by an asymmetric sequential coding approach [51] for the transmission of the information from one HA to the other HA and vice versa. With this asymmetric source coding scheme, theoretically, we show how to estimate/retrieve the unquantized joint statistics between the microphones in the two HAs. An extension of the probability distribution preserving quantization method from [52, 53] to vector sources is proposed to retrieve the unquantized statistics and used to apply the optimal coding strategy from [21] in at least one HA, knowing the joint entropy between the pre-filtered signal and the side information at the decoder. We also resolve the asymptotic sub-optimality of the existing sub-optimal approaches with the proposed coding scheme, as the data is not blindly pre-filtered prior to transmission and important information will not be lost.

Altogether, the rate-constrained noise reduction framework proposed in this chapter tries to address the first and second limitations of the optimal/sub-optimal methods, mentioned in the last paragraph of Section 1.3 and tries to answer the second research question made in Section 1.4.

1.5.4. CHAPTER 5

The optimal algorithm from [21] considers only two processing nodes, which are the left and the right HAs. Scenarios in which there are some additional assistive devices to improve the noise reduction performance are thus not considered in [21]. To address

this issue, and to address the asymptotic suboptimality of the sub-optimal methods, the binaural HA problem can be approached from a more general perspective.

In this chapter, the general setup of a (small) WASN is considered based on our proposed method in [13], where joint statistics are only assumed to be known at the FC, instead of at every node as in [21]. The operational rate-constrained noise reduction framework, which we proposed in [13], estimates the optimal rate allocation across different frequencies and sensors using an operational rate-distortion trade-off [54]. Unlike [21], it allows considering scenarios with some assistive devices along with the binaural HA setup (thereby forming a small-size wireless acoustic sensor network (WASN) with more than two nodes). Furthermore, the performance of the operational rate-constrained noise reduction framework approaches that of the optimal algorithm in [21] at high rates without any mismatch, as the observations are not pre-filtered before quantization and the necessary information will not be removed. However, the exhaustive search, which is used in [13] to find the optimal allocation across sensors, becomes intractable when the size of the WASN grows. Therefore, this method is suitable for small-size networks only.

The method proposed in this chapter tries to address the first and the third limitations of the optimal/sub-optimal methods, mentioned in the last paragraph of Section 1.3, and tries to answer the third research question made in Section 1.4.

1.5.5. CHAPTER 6

The operational rate-distortion trade-off based noise reduction method that will be presented in Chapter 5 finds the optimal rate allocation across both the frequencies and the sensors. However, to find the best rate allocations across the sensors (nodes in WASN) an exhaustive search-based approach is proposed to be used which becomes intractable when the size of the microphone array (WASN) grows.

To address the scalability issue, we propose, in chapter 6, a rate-constrained noise reduction approach based on non-convex optimization, which is also published in [55]. This method jointly finds the best rate allocation and the best estimation weights across all frequencies and sensors for arbitrary sized WASNs. Based on the MSE criterion, the optimal estimation weights are found to be rate-dependent Wiener filters and the optimal rates are the solution to a filter-dependent "water filling" problem. An alternating optimization approach that is used in this method avoids an exhaustive search to find the best allocations and performs almost as good as the exhaustive search-based approach, in most practical scenarios, at the benefit of a much lower computational complexity. Therefore, these methods can perform in general (arbitrary-sized) WASNs as no exhaustive search is used.

The proposed method in this chapter tries to address the third limitation of the optimal /sub-optimal methods, mentioned in the last paragraph of Section 1.3 and tries to answer the third research question made in Section 1.4.

1.5.6. CHAPTER 7

The methods presented in Chapters 4, 5 and 6 find different rate-distortion trade-offs in the noise reduction problem based on the MSE criterion. However, when designing the rate-constrained noise reduction problems, these methods do not consider the preser-

vation of the spatial information. Although the performance of noise reduction might be optimal by minimizing the MSE, the spatial information may be destroyed and the estimated signals may sound unnatural and not spatially correct. Therefore, it is reasonable to incorporate the spatial information into rate-constrained noise reduction problems.

In this chapter, we propose a multi fusion center spatially correct rate-constrained noise reduction problem [56], to find the best rate allocation and the best estimation weights across all sensors and frequencies such that the spatial information of the sources is preserved. We focus mainly on the spatial cue preservation based on equality constraints and try to link the LCMV-based beamformers to data compression by including a set of linear constraints to the original rate-distortion problem. Unlike Chapter 6, here, there are two FCs, therefore, the objective function is to minimize the sum of the distortions of the target estimation at both hearing aids, while considering the total rate budget and simultaneously preserving the spatial information of the sources. Using an alternating optimization approach, the optimal estimation weights are found to be the rate-dependent LCMV filters, and the rates (for both fusion centers) are the solutions to two water-filling problems.

The proposed method in this chapter tries to answer the last (fourth) research question made in Section 1.4.

1.5.7. CHAPTER 8

In this chapter, we conclude the dissertation and discuss the future possibilities to continue the research on the rate-constrained noise reduction problem.

1.6. LIST OF PUBLICATIONS

PAPERS

JOURNALS

1. J. Amini, R. C. Hendriks, R. Heusdens, M. Guo and J. Jensen, “Asymmetric Coding for Rate-Constrained Noise Reduction in Binaural Hearing Aids,” in IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 27, no. 1, pp. 154-167, 2019.

2. J. Amini, R. C. Hendriks, R. Heusdens, M. Guo and J. Jensen, “Rate-Constrained Noise Reduction in Wireless Acoustic Sensor Networks,” in IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 28, pp. 1-12, 2020.

3. J. Amini, R. C. Hendriks, R. Heusdens, M. Guo and J. Jensen, “Spatially Correct Rate-Constrained Noise Reduction for Binaural Hearing Aids in Wireless Acoustic Sensor Networks,” in IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 28, pp. 2731-2742, 2020.

CONFERENCES

1. J. Amini, R. C. Hendriks, R. Heusdens, M. Guo and J. Jensen, “On the Impact of Quantization on Binaural MVDR Beamforming,” Speech Communication; 12. ITG Symposium, Paderborn, Germany, pp. 1-5, 2016.

2. J. Amini, R. Hendriks, R. Heusdens, M. Guo, and J. Jensen, "Operational rate- constrained beamforming in binaural hearing aids," in 26th European Signal Processing Conference (EUSIPCO), 2018.

SYMPOSIA

1. J. Amini, R.C. Hendriks, R. Heusdens, M. Guo and J. Jensen, "Rate-Constrained Beamforming in Binaural Hearing Aids," in Symposium on Information Theory and Signal Processing in the Benelux, Delft University of Technology, Delft, the Netherlands, May 11-12, 2017 (**best student presentation award**).

2. J. Amini, R.C. Hendriks, R. Heusdens, M. Guo and J. Jensen, "Operational Rate-Constrained Noise Reduction for Generalized Binaural Hearing Aid Setups," in Symposium on Information Theory and Signal Processing in the Benelux, University of Twente, Enschede, the Netherlands, May 31-June 1, 2018.

PATENT

- J. Jensen, M. Guo, R. Heusdens, R. Hendriks, and J. Amini, "Binaural beamformer filtering unit, a hearing system and a hearing device," U.S. Patent, No. 10,375,490, 2019.

REFERENCES

[1] E. Leegwater and W. L. van Bueren, "Gehoor in Nederland," TNS NIPO2005.

[2] L. van Thiel, "Gehoor Nederland 2010," TNS NIPO2010.

[3] M. Brandstein and D. Ward, *Microphone Arrays: Signal Processing Techniques and Applications,* Berlin, Germany: Springer Science and Business Media, 2001.

[4] J. M. Kates, "Feedback cancellation in hearing aids: results from a computer simulation,"in IEEE Transactions on Signal Processing, vol. 39, no. 3, pp. 553-562, 1991.

[5] A. Spriet, I. Proudler, M. Moonen and J. Wouters, "Adaptive feedback cancellation in hearing aids with linear prediction of the desired signal," in IEEE Transactions on Signal Processing, vol. 53, no. 10, pp. 3749-3763, 2005.

[6] P. Vary and R. Martin, *Digital Speech Transmission: Enhancement, Coding And Error Concealment*, John Wiley and Sons, 2006.

[7] P. C. *Loizou, Speech Enhancement: Theory and Practice, Second Edition*, 2013.

[8] D. Marquardt, E. Hadad, S. Gannot, and S. Doclo, "Optimal binaural LCMV beamformers for combined noise reduction and binaural cue preservation," in 2014 14th International Workshop on Acoustic Signal Enhancement (IWAENC), pp. 288–292, 2014.

[9] R. C. Hendriks; T. Gerkmann; J. Jensen, "DFT-Domain Based Single-Microphone Noise Reduction for Speech Enhancement: A Survey of the State of the Art," Morgan and Claypool, 2013.

[10] R. Sockalingam, M. Holmberg, K. Eneroth, and M. Shulte, "Binaural hearing aid communication shown to improve sound quality and localization," The Hearing Journal, vol. 62, no. 10, pp. 46–47, 2009.

[11] V. Hamacher, "Comparison of advanced monaural and binaural noise reduction algorithms for hearing aids," in 2002 IEEE International Conference on Acoustics, Speech, and Signal Processing, vol. 4, pp. 4008–4011, 2002.

[12] H. L. Van Trees, *Optimum Array Processing. Part IV of Detection, Estimation and Modulation Theory*, New York, NY: Wiley, 2008.

[13] J. Amini, R. Hendriks, R. Heusdens, M. Guo, and J. Jensen, "Operational rate-constrained beamforming in binaural hearing aids," in 26th European Signal Processing Conference (EUSIPCO), 2018.

[14] A. I. Koutrouvelis, R. C. Hendriks, R. Heusdens, and J. Jensen, "A convex approximation of the relaxed binaural beamforming optimization problem," IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 27, no. 2, pp. 321–331, 2019.

[15] T. J. Klasen, T. Van den Bogaert, M. Moonen, and J. Wouters, "Binaural noise reduction algorithms for hearing aids that preserve interaural time delay cues," IEEE Transactions on Signal Processing, vol. 55, no. 4, pp. 1579–1585, 2007.

[16] S. Markovich-Golan, S. Gannot, and I. Cohen, "Distributed multiple constraints generalized sidelobe canceler for fully connected wireless acoustic sensor networks," IEEE Transactions on Audio, Speech, and Language Processing, vol. 21, no. 2, pp. 343–356, 2013.

[17] A. Bertrand and M. Moonen, "Distributed adaptive node-specific signal estimation in fully connected sensor networkspart i: Sequential node updating," IEEE Transactions on Signal Processing, vol. 58, no. 10, pp. 5277–5291, 2010.

[18] R. Heusdens, G. Zhang, R. C. Hendriks, Y. Zeng, and W. B. Kleijn, "Distributed MVDR beamforming for (wireless) microphone networks using message passing," in IWAENC 2012; International Workshop on Acoustic Signal Enhancement, pp. 1–4, 2012.

[19] T. C. Lawin-Ore, S. Stenzel, J. Freudenberger, and S. Doclo, "Generalized multichannel Wiener filter for spatially distributed microphones," in Speech Communication; 11. ITG Symposium, pp. 1–4, 2014.

[20] T. M. Cover and J. A. Thomas, E. *lements of information theory*, Wiley- Interscience, 2006.

[21] O. Roy and M. Vetterli, "Rate-constrained collaborative noise reduction for wireless hearing aids," IEEE Transactions on Signal Processing, vol. 57, no. 2, pp. 645–657, 2009.

[22] T. Flynn and R. Gray, "Encoding of correlated observations," IEEE Transactions on Information Theory, vol. 33, no. 6, pp. 773–787, 1987.

[23] T. Berger, *Rate-distortion theory: A mathematical basis for data compression,* Englewood Cliffs, N.J.: Prentice-Hall, 1971.

[24] J. K. Wolf and J. Ziv, "Transmission of noisy information to a noisy receiver with minimum distortion," IEEE Transactions on Information Theory, vol. 16, no. 4, pp. 406–411, 1970.

[25] A. D. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," IEEE Transactions on Information Theory, pp. 1–10, 1976.

[26] L. W. Brooks and I. S. Reed, "Equivalence of the likelihood ratio processor, the maximum signal-to-noise ratio filter, and the Wiener filter," IEEE Transactions on Aerospace and Electronic Systems, vol. AES-8, no. 5, pp. 690–692, 1972.

[27] S. Doclo, W. Kellermann, S. Makino, and S. E. Nordholm, "Multichannel signal enhancement algorithms for assisted listening devices: Exploiting spatial diversity using multiple microphones," IEEE Signal Processing Magazine, vol. 32, no. 2, pp. 18–30, 2015.

[28] E. Hadad, S. Doclo and S. Gannot, "The Binaural LCMV Beamformer and its Performance Analysis," in IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 24, no. 3, pp. 543-558, 2016.

[29] R. C. Hendriks, R. Heusdens, U. Kjems and J. Jensen, "On Optimal Multichannel Mean-Squared Error Estimators for Speech Enhancement," in IEEE Signal Processing Letters, vol. 16, no. 10, pp. 885-888, 2009.

[30] O. L. Frost, "An algorithm for linearly constrained adaptive array processing," in Proceedings of the IEEE, vol. 60, no. 8, pp. 926-935, 1972.

[31] W. M. Hartmann, *How we localize sound,* 1999.

[32] B. Cornelis, S. Doclo, T. Van dan Bogaert, M. Moonen, and J. Wouters, "Theoretical analysis of binaural multimicrophone noise reduction techniques," IEEE Transactions on Audio, Speech, and Language Processing, vol. 18, no. 2, pp. 342–355, 2010.

[33] E. Hadad, S. Doclo and S. Gannot, "The Binaural LCMV Beamformer and its Performance Analysis," in IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 24, no. 3, pp. 543-558, 2016.

[34] A. I. Koutrouvelis, R. C. Hendriks, R. Heusdens, and J. Jensen, "Relaxed binaural LCMV beamforming," IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 25, no. 1, pp. 137–152, 2017.

[35] E. Hadad, D. Marquardt, D. Doclo, and S. Gannot, "Theoretical analysis of binaural transfer function mvdr beamformers with interference cue preservation constraints," IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 23, no. 12, pp. 2449–2464, 2015.

[36] A. I. Koutrouvelis, R. C. Hendriks, J. Jensen, and R. Heusdens, "Improved multimicrophone noise reduction preserving binaural cues," in IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 460–464, 2016.

[37] E. Hadad, S. Gannot, and S. Doclo, "Binaural linearly constrained minimum variance beamformer for hearing aid applications," in IWAENC International Workshop on Acoustic Signal Enhancement, pp. 1–4, 2012.

[38] S. Doclo, A. Spriet, J. Wouters, and M. Moonen, "Frequency-domain criterion for the speech distortion weighted multichannel wiener filter for robust noise reduction," Speech Communication, vol. 49, no. 7-8, pp. 636–656, 2007.

[39] S. Doclo and M. Moonen, "GSVD-based optimal filtering for single and multimicrophone speech enhancement," IEEE Transactions on Signal Processing, vol. 50, no. 9, pp. 2230–2244, 2002.

[40] T. J. Klasen, M. Moonen, T. Van den Bogaert and J. Wouters, "Preservation of interaural time delay for binaural hearing aids through multi-channel Wiener filtering based noise reduction," Proceedings. (ICASSP '05). IEEE International Conference on Acoustics, Speech, and Signal Processing, Philadelphia, PA, pp. iii/29-iii/32, 2005.

[41] D. Marquardt, E. Hadad, S. Gannot, and S. Doclo, "Theoretical analysis of linearly constrained multi-channel wiener filtering algorithms for combined noise reduction and binaural cue preservation in binaural hearing aids," IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 23, no. 12, pp. 2384–2397, 2015.

[42] S. Srinivasan, "Low-bandwidth binaural beamforming," Electronics Letters, vol. 44, no. 22, pp. 1292–1293, 2008.

[43] J. Amini, R. C. Hendriks, R. Heusdens, M. Guo and J. Jensen, "On the Impact of Quantization on Binaural MVDR Beamforming," Speech Communication; 12. ITG Symposium, Paderborn, Germany, pp. 1-5, 2016.

[44] D. Slepian and J. Wolf, "Noiseless coding of correlated information sources," IEEE Transactions on Information Theory, vol. 19, no. 4, pp. 471–480, 1973.

[45] S. C. Darper, "Successive structuring of source coding algorithms for data fusion, buffering and distribution in networks," Ph.D. dissertation, Massachusetts Institute of Technology, 2002.

[46] H. Yamamoto and K. Itoh, "Source coding theory for communication systems with a remote source," Trans. IECE Jpn, vol. E63, no. 6, pp. 700–706, 1980.

[47] S. Srinivasan and A. den Brinker, "Rate-constrained beamforming in binaural hearing aids," EURASIP Journal on Advances in Signal Processing, pp. 1–9, 2009.

[48] S. Srinivasan and A. C. den Brinker, "Analyzing rate-constrained beamforming schemes in wireless binaural hearing aids," in 2009 17th European Signal Processing Conference, pp. 1854–1858, 2009.

[49] S. Doclo, M. Moonen, T. Van den Bogaert, and J. Wouters, "Reducedbandwidth and distributed MWF-based noise reduction algorithms for binaural hearing aids," IEEE Transactions on Audio, Speech, and Language Processing, vol. 17, no. 1, pp. 38–51, 2009.

[50] S. Doclo, T. C. Lawin-Ore, and T. Rohdenburg, "Rate-constrained binaural MWF-based noise reduction algorithms," in Proc. ITG Conference on Speech Communication, Bochum, Germany, 2010.

[51] J. Amini, R. C. Hendriks, R. Heusdens, M. Guo and J. Jensen, "Asymmetric Coding for Rate-Constrained Noise Reduction in Binaural Hearing Aids," in IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 27, no. 1, pp. 154-167, 2019.

[52] M. Li, A. Ozerov, J. Klejsa, and W. B. Kleijn, "Asymptotically optimal distribution preserving quantization for stationary Gaussian processes," QC 20110829, 2011.

[53] M. Li, J. Klejsa, and W. B. Kleijn, "On distribution preserving quantization," arXiv preprint arXiv:1108.3728, 2011.

[54] Y. Shoham and A. Gersho, "Efficient bit allocation for an arbitrary set of quantizers," IEEE Transactions on Acoustics, Speech, and Signal Processing, vol. 36, no. 9, pp. 1445–1453, 1988.

[55] J. Amini, R. C. Hendriks, R. Heusdens, M. Guo and J. Jensen, "Rate-Constrained Noise Reduction in Wireless Acoustic Sensor Networks," in IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 28, pp. 1-12, 2020.

[56] J. Amini, R. C. Hendriks, R. Heusdens, M. Guo and J. Jensen, "Spatially Correct Rate-Constrained Noise Reduction for Binaural Hearing Aids in Wireless Acoustic Sensor Networks," in IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 28, pp. 2731-2742, 2020.

2

BACKGROUND

In this chapter, the goal is to describe some important tools in estimation and information theory to facilitate understanding of the algorithms in the remaining chapters. In this chapter, we also provide the signal model, including the underlying assumptions, which will be used in the proposed algorithms in the remaining chapters.

2.1. TIME DOMAIN SIGNAL MODEL

Let us consider an example acoustic scene shown in Figure 2.1, in which there are some microphones placed in random positions. We denote the number of microphones by M. We assume the source of interest is a point source, denoted by $s[n]$, where n indicates the discrete time-domain index. In addition, there are some interfering point noise sources, say $i_j[n]$, where j denotes the interferer's index. The signal captured at each mth microphone is denoted by $y_m[n]$. This signal consists of contributions. An important part of the microphone signal $y_m[n]$ consists of the spatially filtered version of the target signal, say $h_m[n] * s[n]$, where $h[n]$ denotes the room impulse response (RIR). The function $h_m[n]$ includes the delay in time at which the signal is received at the microphone, the reverberation which is due to the non-line of sight paths in which the signals may be received by the microphones (which is due to reflections in the acoustic scene), the attenuation factor to which the target signal may be affected due to the channel characteristics, and many other factors which may affect the signal before being captured by the microphones. Similar room impulse responses can be defined for the point noise sources. The microphone signal will also be corrupted by some uncorrelated additive internal noise, say $u_m[n]$. Putting all these parts together, the microphone signal can be modeled as

$$y_m[n] = h_m[n] * s[n] + \sum_{j=1}^{b} h_m^j[n] * i_j[n] + u_m[n]. \tag{2.1}$$

2.2. SIGNAL MODEL IN THE FREQUENCY DOMAIN

The speech signal can be thought of as samples of a random process which can be non-stationary in general. In fact, the speech production process starts with an excitation of the vocal cords which is subsequently filtered by the vocal tract. The shape of the vocal tract changes over time, which causes the non-stationarity in the speech signal. One way to reduce the non-stationarity effect is to segment the signal to different overlapping frames with up to 20 ms frame lengths. Therefore, given these quasi-stationary speech frames, it is typical to perform the processing in the frequency domain as the Fourier transform acts as a decorrelating transform on the time samples. The process in which the speech frames are transformed into the frequency domain is called the short-term Fourier transform (STFT) [1], which is what we will use in this work. Looking at (2.1), the convolution operator will be converted to the multiplication (after some approximations) in the frequency domain and the impulse responses are transformed into the frequency domain acoustic transfer functions (ATFs). For hearing aid applications, where the microphones are positioned on the HAs, the hearing aid user's head is also included in the setup. In this case, the acoustic transfer functions should also consider the effect of the head, which are also known as the head-related transfer functions (HRTFs) [2].

In this section, we reformulate the signal model in the frequency domain. Assume that the target signal in the STFT domain is denoted by $S(f,l)$, where f denotes the frequency index, and l denotes the frame index. The noisy microphone signal in the frequency domain, with respect to the mth microphone, then is indicated by $Y_m(f,l)$. The ATF between the target signal and the mth microphone is denoted by $A_m(f)$. The ATF

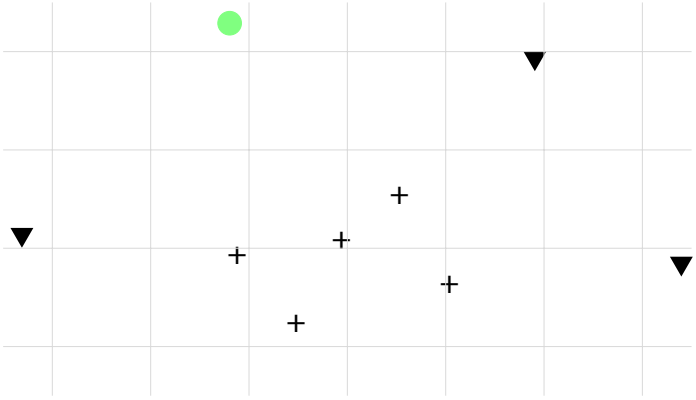


Figure 2.1: An example acoustic scene: a general microphone array is shown by the black "+" symbols, the target signal by the green circle, and the interferers by the black triangles.

between the tjth interfering signal and the mth microphone is denoted by $B_m^j(f)$. With this, the signal model in (2.1) can be rewritten in the STFT domain as

$$Y_m(f,l) = A_m(f)S(f,l) + \sum_{j=1}^{b} B_m^j(f,l)I_j(f,l) + U_m(f,l). \tag{2.2}$$

In the case of multi-microphone processing, it is more convenient to stack the microphone signals into a vector to come up with a signal model in vector notation. For this, and for simplicity, we will drop the frame index and the frequency index, as we perform all processing in the frequency domain per time frame and per frequency bin. For a specific frequency f and frame l, let the vector $\mathbf{y}$ be defined as $\mathbf{y} = [Y_1(f,l),\ldots,Y_M(f,l)]^{\mathrm{T}}$. Similarly, we define

$$\begin{aligned}\mathbf{a} &= [A_1(f),\ldots,A_M(f)]^{\mathrm{T}},\\ \mathbf{b}_j &= [B_1^j(f),\ldots,B_M^j(f)]^{\mathrm{T}},\\ \mathbf{u} &= [U_1(f,l),\ldots,U_M(f,l)]^{\mathrm{T}}.\end{aligned}$$

Stacking all the variables into vectors, the signal model with vector notation is given by

$$\mathbf{y} = \overbrace{\mathbf{a}S}^{\mathbf{x}} + \overbrace{\sum_{j=1}^{b} \mathbf{b}_j I_j + \mathbf{u}}^{\mathbf{n}} = \mathbf{x} + \mathbf{n}. \tag{2.3}$$

All noise components in (2.3) are together referred to as $\mathbf{n}$.

In this thesis, we assume that all components of (2.3) are mutually uncorrelated and zero mean. With this, we can write the second-order statistics as

$$\mathbf{\Phi}_{\mathbf{y}} = \mathbf{\Phi}_{\mathbf{x}} + \mathbf{\Phi}_{\mathbf{n}} \in \mathbb{C}^{M\times M}, \tag{2.4}$$

where,

$$\begin{aligned}\mathbf{\Phi_x} &= \Phi_S \mathbf{a}\mathbf{a}^{\mathrm{H}},\\ \mathbf{\Phi_n} &= \sum_{j=1}^{b} \Phi_{I_j} \mathbf{b}_j \mathbf{b}_j^{\mathrm{H}} + \Phi_U \mathbf{I}.\end{aligned} \tag{2.5}$$

The matrix $\mathbf{\Phi_y}$ is defined as $\mathbf{\Phi_y} = \mathrm{E}[\mathbf{y}\mathbf{y}^{\mathrm{H}}]$, the cross-power spectral density (CPSD) matrix of the microphone signal vector $\mathbf{y}$. The power spectral density of the scalar-valued signal S is defined as $\Phi_S = \mathrm{E}[SS^*]$, where the superscript $[\cdot]^*$ denotes the conjugate operator. Similarly for Φ_{I_j} and Φ_U. The superscripts $[\cdot]^{\mathrm{T}}$ and $[\cdot]^{\mathrm{H}}$ indicate transpose and Hermitian transpose operators on the vector and matrices. Here, for simplicity, we assumed equal microphone noise powers, say Φ_U, with respect to all microphones.

2.3. MULTI-MICROPHONE NOISE REDUCTION

In this part, we focus on multi-microphone noise reduction using linear estimation. Let us assume the goal is to estimate a source of interest, say S, given the noisy observation vector $\mathbf{y}$, at a fusion center (FC). In fact, the noisy microphone observations are transmitted to the FC, and are then combined to estimate S. Therefore the central estimator at the FC will output $\hat{S}$ as an estimate of the target signal, which is given by the linear combination of the noisy microphone observations as

$$\hat{S} = \mathbf{w}^{\mathrm{H}}\mathbf{y}, \tag{2.6}$$

where the vector $\mathbf{w}$ indicates the filter coefficients. In multi-microphone noise reduction based on linear estimation, one important question is how to estimate the filter coefficients $\mathbf{w}$. For this, we need a fidelity criterion to measure the similarity (or distortion) between the estimate $\hat{S}$ and the original target signal S. Although the goal in hearing aid applications is to improve the speech intelligibility [3, 4], we focus in this work for simplicity on the mean square error (MSE). However, this can easily be extended to information theoretical motivated intelligibility metrics as used in [5–7]. The MSE D between the target signal and its estimate can be defined as the averaged (over F frequency bins) power spectral densities of the error process $E = S - \hat{S}$, which is given by [8]

$$D = \frac{1}{F}\sum_{f=1}^{F} d(S,\hat{S}), \tag{2.7}$$

where,

$$\begin{aligned} d(S,\hat{S}) &= \mathrm{E}[|S-\hat{S}|^2] = \mathrm{E}[|S-\mathbf{w}^{\mathrm{H}}\mathbf{y}|^2]\\ &= \mathrm{E}[|S-\mathbf{w}^{\mathrm{H}}(\mathbf{a}S+\mathbf{n})|^2]\\ &= \mathrm{E}[|S-\mathbf{w}^{\mathrm{H}}\mathbf{a}S|^2] + \mathbf{w}^{\mathrm{H}}\mathrm{E}[\mathbf{n}\mathbf{n}^{\mathrm{H}}]\mathbf{w}\\ &= \Phi_S|1-\mathbf{w}^{\mathrm{H}}\mathbf{a}|^2 + \mathbf{w}^{\mathrm{H}}\mathbf{\Phi_n}\mathbf{w}. \end{aligned} \tag{2.8}$$

Under stationarity assumptions, the filtering process can be done independently for each frequency, meaning that the error function $d(S,\hat{S})$ can be minimized independently for each frequency f. Looking at (2.8) the error function includes two terms. 1) the residual error with respect to the target signal distortion which is given by $\Phi_S|1-\mathbf{w}^{\mathrm{H}}\mathbf{a}|^2$, and 2)

the residual environmental noise power which is given by $\mathbf{w}^{\mathrm{H}}\mathbf{\Phi}_{\mathbf{n}}\mathbf{w}$. Typically, the multi-microphone noise reduction algorithms differ in how to impose different constraints when minimizing $d(S(f),\hat{S}(f))$. In general, the estimation problem, in the frequency domain, can be formulated as

$$\begin{aligned} \min_{\mathbf{w}(f)\in\mathbb{C}^{M\times 1}} \quad & d(S(f),\hat{S}(\mathbf{w}(f))) \\ \text{subject to} \quad & \text{set of constraint functions.} \end{aligned} \tag{2.9}$$

In the following we will describe some important filtering algorithms based on the optimization framework in (2.9).

2.3.1. MULTI-CHANNEL WIENER FILTERING [9–11]

In this part, we explain the multi-channel Wiener filtering [9–11] on the optimization problem in (2.9). If the optimization problem in (2.9) is unconstrained, then the estimation process is called multi-channel Wiener filtering (MWF). The goal is here to minimize the error function without imposing any distortion-less response constraints, which in fact turns to the best linear minimum mean square estimation (LMMSE) [11].

Rewriting the optimization problem in (2.9) using (2.8) we have

$$\min_{\mathbf{w}\in\mathbb{C}^{M\times 1}} \quad \Phi_S|1-\mathbf{w}^{\mathrm{H}}\mathbf{a}|^2+\mathbf{w}^{\mathrm{H}}\mathbf{\Phi}_{\mathbf{n}}\mathbf{w}. \tag{2.10}$$

After solving the convex optimization problem (as the objective function is quadratic over $\mathbf{w}$ and the matrix $\mathbf{\Phi}_{\mathbf{n}}$ is positive semi-definite) in (2.10) over $\mathbf{w}$, the best Wiener filter coefficients are given by

$$\mathbf{w}^{\star}_{\mathrm{Wiener}}=\mathbf{\Phi}_{\mathbf{y}}^{-1}\mathbf{\Phi}_{\mathbf{y}S}, \tag{2.11}$$

where, $\mathbf{\Phi}_{\mathbf{y}}\in\mathbb{C}^{M\times M}$ is the CPSD of the noisy observations $\mathbf{y}$, and $\mathbf{\Phi}_{\mathbf{y}S}=\Phi_S\mathbf{a}\in\mathbb{C}^{M\times 1}$ is the CPSD vector between the noisy observation vector $\mathbf{y}$ and the scalar-valued target signal S.

Looking at (2.10), it is clear that the algorithm allows some distortion in the target estimation, while achieving the best noise reduction performance among all other MSE-based filters. In the next part, we introduce another important linearly constrained filter, which does not allow distortion in the target with as a result less degree of freedom for noise reduction.

2.3.2. LINEARLY CONSTRAINED MINIMUM VARIANCE FILTERING [12, 13]

In this part, we explain the linearly constrained minimum variance Filtering [12, 13] based on the optimization problem in (2.9).

We define the optimization problem as

$$\begin{aligned} \min_{\mathbf{w}\in\mathbb{C}^{M\times 1}} \quad & \Phi_S|1-\mathbf{w}^{\mathrm{H}}\mathbf{a}|^2+\mathbf{w}^{\mathrm{H}}\mathbf{\Phi}_{\mathbf{n}}\mathbf{w} \\ \text{subject to} \quad & \mathbf{w}^{\mathrm{H}}\mathbf{\Lambda}=\mathbf{f}^{\mathrm{H}}, \end{aligned} \tag{2.12}$$

where, $\mathbf{\Lambda}\in\mathbb{C}^{M\times d}$ is a constraint matrix which includes the ATFs w.r.t. the target signal(s) as well as the interfering signal, and $\mathbf{f}\in\mathbb{C}^{d\times 1}$ is a vector that controls which spatial information needs to be preserved. Together, $\mathbf{\Lambda}\in\mathbb{C}^{M\times d}$ and $\mathbf{f}\in\mathbb{C}^{d\times 1}$ can be used to formulate

linear constraints on the filter $\mathbf{w}$. The number of constraints is indicated by d, which usually is less than M (for the optimization problem to have enough degrees of freedom to have controlled noise reduction). As mentioned at the beginning of this section, the first term in the objective function in (2.12) is the residual distortion with respect to the target signal. The set of constraints in (2.12) typically includes the important distortion-less response constraint, which is given by

$$\mathbf{w}^{\mathrm{H}}\mathbf{a} = 1. \tag{2.13}$$

In fact, by imposing the distortion-less constraint, the estimation problem in (2.12) can be thought of as an unbiased estimator as $\mathrm{E}[\hat{S}] = \mathrm{E}[\mathbf{w}^{\mathrm{H}}\mathbf{y}] = \mathrm{E}[\mathbf{w}^{\mathrm{H}}\mathbf{a}S + \mathbf{w}^{\mathrm{H}}\mathbf{n}] = \mathrm{E}[S]$, under the assumption that the noise vector is zero-mean.

By imposing the distortion-less constraint in (2.13) to the problem in (2.12), the first term in the objective function which is $\Phi_S|1-\mathbf{w}^{\mathrm{H}}\mathbf{a}|^2$ will disappear, therefore the problem can be simplified as follows

$$\begin{aligned} \min_{\mathbf{w}\in\mathbb{C}^{M\times 1}} \quad & \mathbf{w}^{\mathrm{H}}\boldsymbol{\Phi}_{\mathbf{n}}\mathbf{w} \\ \text{subject to} \quad & \mathbf{w}^{\mathrm{H}}\boldsymbol{\Lambda} = \mathbf{f}^{\mathrm{H}}. \end{aligned} \tag{2.14}$$

Now, the problem in (2.14) is an LCMV problem and can be interpreted as follows: minimizing the residual noise power subject to a set of linear constraints. After solving the convex optimization problem in (2.14) (as the objective function is quadratic over $\mathbf{w}$ and $\boldsymbol{\Phi}_{\mathbf{n}}$ is positive semi-definite), the LCMV filter coefficients are derived as

$$\mathbf{w}^{\star}_{\mathrm{LCMV}} = \boldsymbol{\Phi}_{\mathbf{n}}^{-1}\boldsymbol{\Lambda}(\boldsymbol{\Lambda}^{\mathrm{H}}\boldsymbol{\Phi}_{\mathbf{n}}^{-1}\boldsymbol{\Lambda})^{-1}\mathbf{f}. \tag{2.15}$$

2.3.3. MINIMUM VARIANCE DISTORTION LESS RESPONSE FILTERING [14, 15]

In this part, we explain a special case of the LCMV filter, which is called minimum variance distortion-less response (MVDR). If the constraint matrix $\boldsymbol{\Lambda}$ includes one column, that is $\mathbf{a}$, with $\mathbf{f} = 1$, and therefore, the only constraint in the optimization problem in (2.14) will be the distortion-less constraint, that is $\mathbf{w}^{\mathrm{H}}\mathbf{a} = 1$. In this case, the optimization problem in (2.14) can be further simplified as

$$\begin{aligned} \min_{\mathbf{w}\in\mathbb{C}^{M\times 1}} \quad & \mathbf{w}^{\mathrm{H}}\boldsymbol{\Phi}_{\mathbf{n}}\mathbf{w} \\ \text{subject to} \quad & \mathbf{w}^{\mathrm{H}}\mathbf{a} = 1. \end{aligned} \tag{2.16}$$

The optimization problem in (2.16) results in the MVDR filter. After solving the optimization problem, the MVDR filter coefficients are derived as [14]

$$\mathbf{w}^{\star}_{\mathrm{MVDR}} = \frac{\boldsymbol{\Phi}_{\mathbf{n}}^{-1}\mathbf{a}}{(\mathbf{a}^{\mathrm{H}}\boldsymbol{\Phi}_{\mathbf{n}}^{-1}\mathbf{a})}. \tag{2.17}$$

The minimum output noise power then can be computed using (2.17) as

$$(\mathbf{w}^{\star}_{\mathrm{MVDR}})^{\mathrm{H}}\boldsymbol{\Phi}_{\mathbf{n}}(\mathbf{w}^{\star}_{\mathrm{MVDR}}) = (\mathbf{a}^{\mathrm{H}}\boldsymbol{\Phi}_{\mathbf{n}}^{-1}\mathbf{a})^{-1}. \tag{2.18}$$

The MVDR filter is only constrained to preserve the target. Therefore, the MVDR may have better noise reduction performance, compared to the LCMV. The MVDR method can be thought of as the best unbiased estimator when all the parameters are estimated/obtained error-free [11].

It can be shown [15] that there is a relation between the MVDR filter and the Wiener filter as

$$\mathbf{w}^{\star}_{\text{Wiener}} = \frac{\Phi_S}{\Phi_S + (\mathbf{a}^{\text{H}}\boldsymbol{\Phi}_{\mathbf{n}}^{-1}\mathbf{a})^{-1}}\mathbf{w}^{\star}_{\text{MVDR}}. \tag{2.19}$$

The relation in (2.19) can be interpreted as follows. The MWF filter can be achieved by the MVDR filter followed by a single-channel Wiener filter.

2.4. BINAURAL MULTI-MICROPHONE NOISE REDUCTION

In this section, we describe the acoustical scene for binaural hearing aid applications. Figure 2.2 shows the binaural setup.

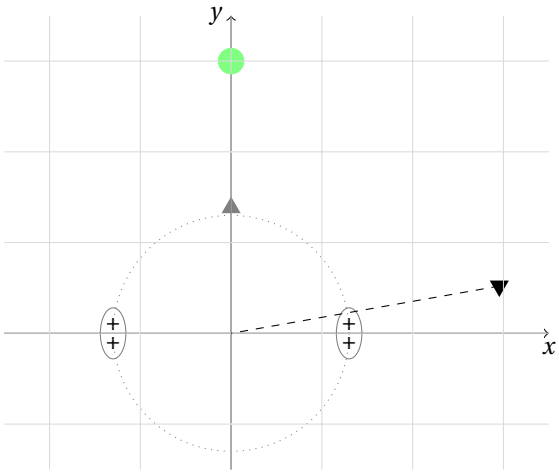


Figure 2.2: Typical acoustic scene. The two HA microphones, the target signal, and the interferer are indicated by the black "+", the green circle, and the black triangle, respectively.

The binaural hearing aid system consists of two hearing aids. There are M_{l} and M_{r} (in total $M = M_{\text{l}} + M_{\text{r}}$) microphones, which are assumed to be embedded in the left-side and the right-side HAs, respectively. The hearing aids can potentially communicate with each other through a wireless link to construct a binaural system. Therefore, there are in this case two fusion centers, one for each side. The microphones embedded in the hearing aids receive the information from the acoustic scene. In this binaural setup two processors filter the received data to estimate two versions of the target signal for each side. For example, for the left side processor, the information captured at the right side HA should be transmitted to the left side, then combined with the local left-side observation to estimate a version of the target signal with respect to a reference microphone at the left side HA. Therefore, two sets of filter coefficients need to be estimated.

Let $\mathbf{w}_{\text{l}} \in C^{M\times 1}$ and $\mathbf{w}_{\text{r}} \in C^{M\times 1}$ be the filter coefficients with respect to the left and the

right side FCs. The processors then output two versions of the target signal as

$$\begin{aligned}\hat{S}_{\mathrm{l}} &= \mathbf{w}_{\mathrm{l}}^{\mathrm{H}}\mathbf{y},\\ \hat{S}_{\mathrm{r}} &= \mathbf{w}_{\mathrm{r}}^{\mathrm{H}}\mathbf{y}.\end{aligned} \tag{2.20}$$

The estimation accuracy in binaural algorithms is an important factor. However, the preservation of binaural information also plays an important role when designing such binaural algorithms. Ideally, we want to get a natural impression of the acoustic scene after we process the data. Practically, different algorithms have a different impact on binaural information. In the following, we give some necessary measures to evaluate binaural information preservation.

2.4.1. BINAURAL CUES

In this section, we explain some important binaural spatial information measures. First, we explain the interaural time (phase) (IPD) difference. Second, we explain the interaural level differences (ILD). The IPD of a particular source is the time (phase) difference of the signal between arriving at the left and the right side HAs. ILDs are the level differences of a particular source between the left and the right HA. Mathematically speaking, IPDs and ILDs can be derived from the input and output interaural transfer functions. As an example we define here the input and output ITF for the filtered target source. Let $X_{\mathrm{l}} = A_{\mathrm{l}}S$ and $X_{\mathrm{r}} = A_{\mathrm{r}}S$. The input and output ITF are then defined as [16, 17]

$$\begin{aligned}\mathrm{ITF}_X^{\mathrm{in}}(f) &= \frac{X_{\mathrm{l}}}{X_{\mathrm{r}}} = \frac{A_{\mathrm{l}}}{A_{\mathrm{r}}},\\ \mathrm{ITF}_X^{\mathrm{out}}(f) &= \frac{\mathbf{w}_{\mathrm{l}}{}^{\mathrm{H}}\mathbf{x}}{\mathbf{w}_{\mathrm{r}}{}^{\mathrm{H}}\mathbf{x}} = \frac{\mathbf{w}_{\mathrm{l}}{}^{\mathrm{H}}\mathbf{a}}{\mathbf{w}_{\mathrm{r}}{}^{\mathrm{H}}\mathbf{a}}.\end{aligned} \tag{2.21}$$

For the ITFs for of the jth point noise source, the signal X and the transfer function A are replaced by I_j and B_j, respectively in (2.21). With this, the input and output ILDs are defined as the squared magnitudes of the input and output ITFs. That is

$$\mathrm{ILD}_X^{\mathrm{in}}(f) = |\mathrm{ITF}_X^{\mathrm{in}}(f)|^2, \qquad \mathrm{ILD}_X^{\mathrm{out}}(f) = |\mathrm{ITF}_X^{\mathrm{out}}(f)|^2, \tag{2.22}$$

and the input and output ITDs defined as the phase of the input and output ITFs. That is

$$\mathrm{IPD}_X^{\mathrm{in}}(f) = \angle\mathrm{ITF}_X^{\mathrm{in}}(f), \qquad \mathrm{IPD}_X^{\mathrm{out}}(f) = \angle\mathrm{ITF}_X^{\mathrm{out}}(f). \tag{2.23}$$

The ILD and ITD errors are then defined as

$$\begin{aligned}\mathrm{ER}_{\mathrm{ILD}_X^{\mathrm{out}}}(f) &= |\mathrm{ILD}_X^{\mathrm{out}}(f) - \mathrm{ILD}_X^{\mathrm{in}}(f)|,\\ \mathrm{ER}_{\mathrm{ITD}_X^{\mathrm{out}}}(f) &= \frac{|\mathrm{ITD}_X^{\mathrm{out}}(f) - \mathrm{ITD}_X^{\mathrm{in}}(f)|}{\pi}.\end{aligned} \tag{2.24}$$

Note that $0 \leq \mathrm{ER}_{\mathrm{ITD}_X^{\mathrm{out}}}(k) \leq 1$. In the following we explain one of the important binaural linear multi-microphone noise reduction techniques.

2.4.2. BINAURAL LCMV-BASED NOISE REDUCTION [13, 16, 18, 19]

One common approach for binaural multi-microphone noise reduction is to estimate the signal of interest at both left side and right side reference positions by combining all the available noisy observations into a single estimate for each HA. Similar to Section 2.3.2, here there are two processors which output two estimates of the target signal such that a fidelity criterion is satisfied and that the binaural information is preserved. The target signals at the left and right HA, i.e., S_{l} and S_{r}, respectively, are estimated as

$$\hat{S}_{\mathrm{l}} = \mathbf{w}_{\mathrm{l}}{}^{\mathrm{H}}\mathbf{y}, \quad , \hat{S}_{\mathrm{r}} = \mathbf{w}_{\mathrm{r}}{}^{\mathrm{H}}\mathbf{y}, \tag{2.25}$$

where $\mathbf{w}_{\mathrm{l}}{}^{\mathrm{H}} \in \mathbb{C}^{M}$ and $\mathbf{w}_{\mathrm{r}}{}^{\mathrm{H}} \in \mathbb{C}^{M}$ are the filter coefficients of the left and right filters, respectively. Minimizing the sum of the output noise powers, for both beamformers, the binaural linearly constrained beamforming problem can be formulated as [12]

$$\begin{aligned} \min_{\mathbf{w}} \quad & \mathbf{w}^{\mathrm{H}}\mathbf{\Phi}\mathbf{w} \\ \text{subject to} \quad & \mathbf{\Lambda}^{\mathrm{H}}\mathbf{w} = \mathbf{f}, \end{aligned} \tag{2.26}$$

where

$$\begin{aligned} \mathbf{w} &= [\mathbf{w}_{\mathrm{l}}^{\mathrm{T}}\mathbf{w}_{\mathrm{r}}^{\mathrm{T}}]^{\mathrm{T}} \in \mathbb{C}^{2M\times 1}, \\ \mathbf{\Phi} &= \begin{bmatrix} \mathbf{\Phi_n} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Phi_n} \end{bmatrix} \in \mathbb{C}^{2M\times 2M}, \end{aligned}$$

and $\mathbf{\Lambda} \in \mathbb{C}^{2M\times d}$ is the constraint matrix, with d the number of linear constraints. Different binaural LCMV-based beamformers can be constructed by changing the entries of $\mathbf{\Lambda}$. As mentioned in the previous chapter, to preserve the ITF of the interferers, one can add extra constraints on the interfering source, with respect to the left and right side ATFs, as in [13]. However, to increase the degrees of freedom of the method, in [18, 19] it is shown that we can combine the constraints on the left and the right side into a one constraint function per source which has more degrees of freedom, compared to independent constraints for the left and the right side beamformers. These additional degrees of freedom can then be used to cancel more interferers, given a fixed number of microphones. Following [18, 19], matrix $\mathbf{\Lambda}$ and vector $\mathbf{f}$ are given by

$$\begin{aligned} \mathbf{\Lambda} &= \begin{bmatrix} \mathbf{a} & \mathbf{0} & \mathbf{b}_1 B_1^{\mathrm{r}} & \dots & \mathbf{b}_b B_b^{\mathrm{r}} \\ \mathbf{0} & \mathbf{a} & -\mathbf{b}_1 B_1^{\mathrm{l}} & \dots & -\mathbf{b}_b B_b^{\mathrm{l}} \end{bmatrix} \in \mathbb{C}^{2M\times(b+2)}, \\ \mathbf{f}^{\mathrm{H}} &= [A_{\mathrm{l}}\ A_{\mathrm{r}}\ 0\ \dots\ 0] \in \mathbb{C}^{1\times(b+2)}. \end{aligned} \tag{2.27}$$

Solving the problem in (2.26), the optimal weights are computed as [18]

$$\mathbf{w}^{\star} = \mathbf{\Phi}^{-1}\mathbf{\Lambda}(\mathbf{\Lambda}^{\mathrm{H}}\mathbf{\Phi}^{-1}\mathbf{\Lambda})^{-1}\mathbf{f}, \tag{2.28}$$

and the optimal beamformer outputs are given by

$$\hat{S}_{\mathrm{l}}^{\star} = (\mathbf{w}_{\mathrm{l}}^{\star})^{\mathrm{H}}\mathbf{y}, \quad , \hat{S}_{\mathrm{r}}^{\star} = (\mathbf{w}_{\mathrm{r}}^{\star})^{\mathrm{H}}\mathbf{y}. \tag{2.29}$$

In order to compute the binaural outputs $\hat{S}_l^\star$ and $\hat{S}_r^\star$, the actual signal realizations should be available error-free at both HAs. However, due to limited battery power, and therefore, limited transmission power, in practice, the bit-rate used to represent the transmitted signals is constrained, which is denoted by r_m bits per sample (bps). Theoretically, there is a trade-off between the rate of transmission and accuracy (distortion) of the represented signals. Therefore, in the next section, we explain rate-distortion trade-off (lossy source coding) from an information-theoretic point of view and show scenarios that resemble the binaural hearing aid setup.

2.5. LOSSY SOURCE CODING: RATE-DISTORTION TRADE-OFF

We summarize important scenarios in lossy source coding and explain the algorithms and assumptions made in such scenarios. This is based on the material proposed in [20, 21].

2.5.1. DIRECT LOSSY SOURCE CODING THEORY

We start with a description of direct lossy source coding. Assume that we have a source that produces a sequence $X^n = \{X_1, X_2, , ..., X_n\}$ of independent identically distributed (i.i.d) variables with $X_i \sim p_{X_i}(x_i)$ where $X_i \in \mathscr{X}$. We will assume that the alphabet is finite, but most of the results can be extended to continuous random variables.

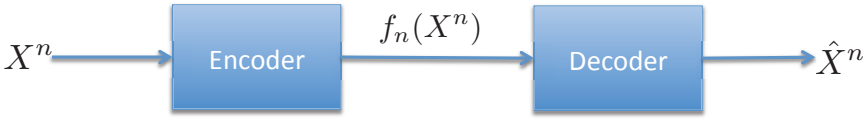


Figure 2.3: Lossy source coding scenario

The encoder, as illustrated in Figure 2.3, describes the source sequence X^n by an index $f_n(X^n)$, where $f_n : \mathscr{X}^n \rightarrow \{1, 2, ..., 2^{nR}\}$, and R denotes the total rate to represent X^n. The decoder estimates X^n by $\hat{X}^n = g_n(f_n(X^n))$, where $g_n : \{1, 2, ..., 2^{nR}\} \rightarrow \hat{\mathscr{X}}^n$. The distortion function is defined as

$$d = \sum_{x^n} \sum_{\hat{x}^n} p(x^n, \hat{x}^n) d(x^n, \hat{x}^n), \tag{2.30}$$

where, $p(x^n, \hat{x}^n) = p(x^n) q(\hat{x}^n | x^n)$ is the joint probability density function, $q(\hat{x}^n | x^n)$ is the conditional probability density function, and $d(x^n, \hat{x}^n)$ is an averaged distortion of a single-letter fidelity criterion $d(x_t, \hat{x}_t)$, for $t = 1, 2, ..., n$ and is defined as

$$d(x^n, \hat{x^n}) = \frac{1}{n} \sum_{t=1}^{n} d(x_t, \hat{x}_t). \tag{2.31}$$

The rate-distortion theory states that the minimum achievable rate at which the source output can be reproduced with maximum distortion D at the decoder corresponds to

the optimum of the following variational problem

$$\begin{aligned} R(D) = &\min_{q(\hat{x}^n|x^n)} \quad I(X^n;\hat{X}^n), \\ &\text{subject to} \quad \sum_{x^n}\sum_{\hat{x}^n} p(x^)q(\hat{x}^n|x^n)d(x^n,\hat{x}^n) < D. \end{aligned} \tag{2.32}$$

where, $I(\cdot;\cdot)$ is the mutual information function. Assuming independent identically distributed (i.i.d.) Gaussian sources and a squared-error criterion, the rate-distortion function for the problem in (2.32) can be computed as

$$R(D) = \begin{cases} \frac{1}{2}\log_2(\frac{\sigma_x^2}{D}), & 0 \le D \le \sigma_x^2 \\ 0, & D \ge \sigma_x^2, \end{cases} \tag{2.33}$$

where σ_x^2 is the variance of the source. The corresponding optimum "forward channel" [20] which can achieve the R-D relation in (2.33) is shown in Figure 2.4.

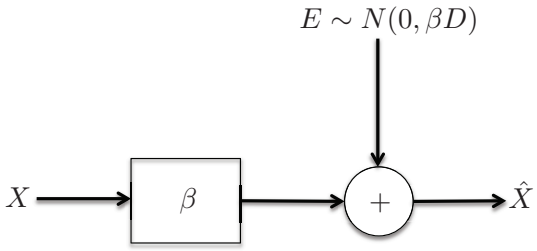


Figure 2.4: Optimum forward channel interpretation of the lossy source coding scheme in Figure 2.3, for i.i.d. Gaussian sources.

where, $\beta = 1 - \frac{D}{\sigma_X^2}$. If the sources are not Gaussians, the "forward channel" representation is not optimal according to the MSE criterion.

For stationary jointly Gaussian sources, the rate-distortion function with weighted MSE criterion is derived as [20]

$$\begin{aligned} R(\theta) &= \frac{1}{4\pi}\int_{-\pi}^{+\pi} \max\left(0, \log_2 \frac{|A|^2\Phi_X(\Omega)}{\theta}\right) d\Omega, \\ D(\theta) &= \frac{1}{2\pi}\int_{-\pi}^{+\pi} \min\left(\theta, |A|^2\Phi_X(\Omega)\right) d\Omega, \end{aligned} \tag{2.34}$$

where A is frequency dependent weight, Ω is the frequency, θ is the reverse-water-filling parameter, and Φ_X is the PSD of signal. The corresponding test channel is illustrated in Figure 2.5

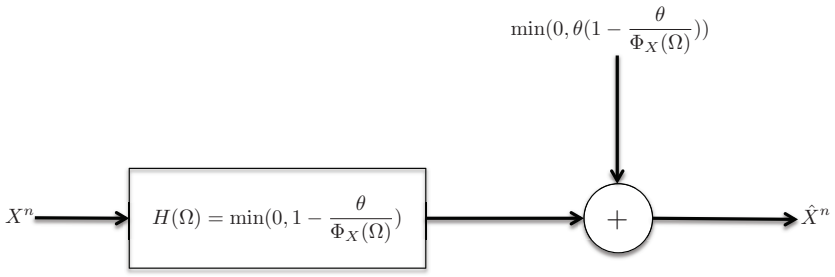


Figure 2.5: Optimum forward channel interpretation of the lossy source coding scheme in Figure 2.3, for jointly Gaussian sources.

2.5.2. REMOTE (NOISY) SOURCE CODING [20, 22]

The more generalized version of the problem depicted in Figure 2.3 is when the source is not directly observable, but remotely (indirectly) observed at the encoder, which is illustrated in Figure 2.6.

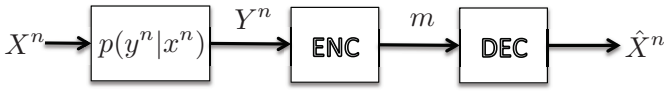


Figure 2.6: Indirect (remote) source coding scheme.

First, the desired source sequence X^n is distorted, and then, the noisy signal sequence Y^n is coded and transmitted to the receiver. We can interpret this problem in the hearing aid setup as the right side noisy observations are transmitted to the left side hearing aid. The encoder produces a code sequence $m = f_n(Y^n)$ which appears at bitrate R. Finally, the decoder reproduces an estimate of the desired signal X^n which is denoted by $\hat{X}^n = g_n(m)$. The averaged distortion $D(R)$ between X^n and $\hat{X}^n$ is defined as

$$D(R) = \mathrm{E}\left[d(X^n, \hat{X}^n)\right] = \mathrm{E}\left[\frac{1}{n}||X^n - \hat{X}^n||^2\right]. \tag{2.35}$$

The problem is to choose f and g in order to minimize the averaged distortion $D(R)$ between X^n and $\hat{X}^n$ which is defined as

$$D^*(R) = \inf_{f,g} \quad D(R) \tag{2.36}$$

Note that as f and g are functions of the transmission rate R, the distortion will be a function of rate.

This problem is considered in [20, 22] and a nice decomposition of the distortion function is derived in [22]. In [22] it concluded that the optimal distortion is achieved

if a function of the noisy observation, that is the conditional mean estimate $E[X^n|Y^n]$ of the desired signal, is transmitted to the receiver rather than the noisy observations themselves.

The conditional mean estimator of the source X^n, say U^n, is given by $U^n = E[X^n|Y^n]$, given the observation Y^n. In [22] it is proved that the distortion can be decomposed as

$$D(R) = \mathrm{E}\left[\frac{1}{n}||X^n - \hat{X}^n||^2\right] = \mathrm{E}\left[\frac{1}{n}||X^n - U^n||^2\right] + \mathrm{E}\left[\frac{1}{n}||U^n - \hat{X}^n||^2\right]. \tag{2.37}$$

Hence, the distortion is decomposed into two parts:

1) Distortion due to the estimation of the signal given the observations.

2) Distortion due to the quantization of the estimated signal to reproduce $\hat{X}^n$.

As U^n is computed, independent from the choices for f and g (independent of the rate), the optimal distortion can be found by taking the minimum only over the second part of the distortion function, that is

$$D^* = \mathrm{E}[\frac{1}{n}||X^n - U^n||^2] + \inf_{f,g} \quad \mathrm{E}[\frac{1}{n}||U^n - \hat{X}^n||^2]. \tag{2.38}$$

The corresponding optimum architecture for remote lossy source coding is illustrated in Figure 2.7.

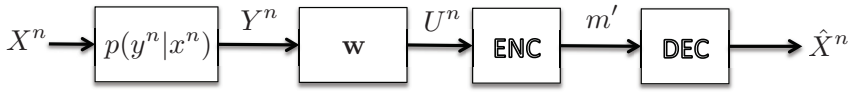


Figure 2.7: Optimum architecture for remote source coding.

Here, **w** is the conditional mean estimator filter. In fact, it is optimal to first estimate the desired signal then quantize it. The quantization step now simply corresponds to the direct lossy source coding problem, introduced in the previous subsection, with input signal U^n and the output signal $\hat{X}^n$.

For stationary jointly Gaussian sources, the rate-distortion function can be computed as [20]

$$\begin{aligned} R(\theta) &= \frac{1}{4\pi}\int_{-\pi}^{+\pi} \max\left(0, \log_2 \frac{|A|^2\Phi_U(\Omega)}{\theta}\right) d\Omega, \\ D(\theta) &= \frac{1}{2\pi}\int_{-\pi}^{+\pi} |A|^2\Phi_{X|Y}(\Omega)\,d\Omega + \frac{1}{2\pi}\int_{-\pi}^{+\pi} \min\left(\theta, |A|^2\Phi_U(\Omega)\right) d\Omega, \end{aligned} \tag{2.39}$$

where

$$\begin{aligned} \Phi_U &= \mathbf{\Phi}_{XY}\mathbf{\Phi}_Y^{-1}\mathbf{\Phi}_{YX}, \\ \Phi_{X|Y} &= \Phi_X - \mathbf{\Phi}_{XY}\mathbf{\Phi}_Y^{-1}\mathbf{\Phi}_{YX}. \end{aligned} \tag{2.40}$$

2.5.3. SOURCE CODING WITH SIDE INFORMATION (WYNER-ZIV CODING)

In this section, we explain the problem of source coding in a situation where the decoder/encoder has access to side information about the source, say Y^n, which is proposed in [23] and is illustrated in Figure 2.8.

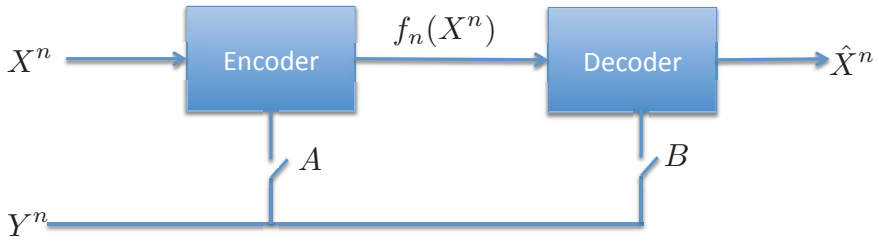


Figure 2.8: Direct source coding with side information.

Let us assume dependent source pairs (X^n, Y^n) producing a sequence $(X_1, Y_1), (X_2, Y_2), ..., (X_n, Y_n)$, where $(X_i, Y_i) \in \mathcal{X} \times \mathcal{Y}$, where Y^n is the side information. In this scenario, the encoder and decoder functions are

$$\begin{aligned} f_n &: \mathcal{X}^n \rightarrow I_n = \{1, 2, ..., 2^{nR}\}, \\ g_n &: \mathcal{Y}^n \times I_n \rightarrow \mathcal{X}^n. \end{aligned} \tag{2.41}$$

In fact, the decoded variable is denoted by $\hat{X}^n = g(Y^n, f(X^n))$.

With respect to the model shown in Figure 2.8, the problem is defined to find the rate R at which the source X^n is coded and transmitted such that the distortion between X^n and $\hat{X}^n$ does not exceed the certain value D, assuming the decoder has access to the side information Y^n, which is correlated to X^n. It seems that the optimal rate of transmission in this case should be lower compared to the case without the side information (Section 2.5.1).

Considering the system shown in Figure 2.8, three situations can happen

- Case 1: Switches A and B are closed, i.e., both the decoder and the encoder have access to the side information.
- Case 2: Switch A is open and switch B is closed, i.e, only the decoder has access to the side information.
- Case 3: Switch A is open and switch B is open, i.e, the direct lossy source coding in Section 2.5.1.

The case in which switch A is closed and switch B is open is not considered here. For

case 1, the rate-distortion problem is defined as

$$\begin{aligned} R_{X^n|Y^n}(D) = & \min_{q(\hat{x}^n|x^n,y^n)} \quad I(X^n;\hat{X}^n|Y^n), \\ \text{subject to} \quad & \sum_{x^n}\sum_{y^n}\sum_{\hat{x}^n} p(x^n,y^n)q(\hat{x}^n|x^n,y^n)d(x^n,\hat{x}^n) < D \\ & \sum_{\hat{x}^n} p(x^n,y^n,\hat{x}^n) = p(x^n,y^n). \end{aligned} \tag{2.42}$$

In case of lossless coding, it can be shown that the above problem asymptotically equals to that of the Slepian and Wolf coding scheme [24], i.e.,

$$R^*(0) = R_{X^n|Y^n}(0) = H(X^n|Y^n), \tag{2.43}$$

where $R^*(0)$ is the lower bound of transmission rate in the loss-less transmission scheme, which is the conditional entropy of the signal given side information.

Figure 2.9 shows the interpretation of the coding scheme in Figure 2.8, for i.i.d Guassian random variables, in which the encoder and the decoder are combined.

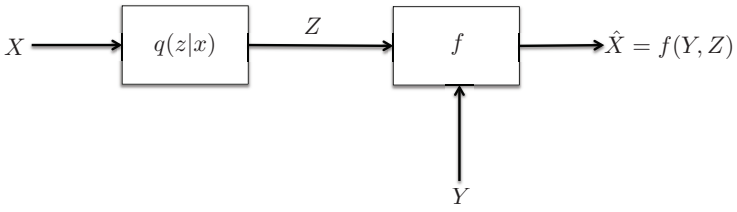


Figure 2.9: Corresponding test channel for Figure 2.8.

First, the signal is passed through a "test channel" with transition probability $q(z|x)$, where Z is output of the channel, then Z and the side information Y are combined to estimate the input signal, i.e., $\hat{X} = f(Y,Z)$.

For case 2, in which only the decoder has access to the side information, the problem is defined as

$$\begin{aligned} R^*(D) = & \min_{q(z^n|x^n)} \quad I(X^n;Z^n|Y^n) \\ \text{subject to} \quad & \mathrm{E}[d(X^n,\hat{X}^n)] < D, \end{aligned} \tag{2.44}$$

where Z is a auxiliary random sequence satisfying following conditions

1) $\sum_{z^n} p(x^n,y^n,z^n) = p(x^n,y^n)$.
2) Y^n, Z^n are conditionally independent given X, i.e., $I(X^n;Z^n|Y^n) = I(X^n;Z^n) - I(Y^n;Z^n)$ or $I(Y^n;Z^n|X^n) = 0$.
3) $\hat{X}^n = f(Y^n,Z^n)$.

Thanks to the data-processing theorem, $R^*(D) \geq R_{X^n|Y^n}(D)$ as

$$I(X^n;Z^n|Y^n) \geq I(X^n;\hat{X}^n|Y^n). \tag{2.45}$$

The equality holds if and only if $I(X;Z|\hat{X},Y^n) = 0$. For a Gaussian variable, the optimal choice of variables $q(z|x)$ and f shown in Figure 2.9 will be derived in a more general scenario in the next subsection.

2.5.4. REMOTE (NOISY) SOURCE CODING THEORY WITH SIDE INFORMATION (REMOTE W-Z CODING)

In this section, we explain the final coding scheme, based on the coding schemes which have been previously explained. It resembles the binaural hearing aid scenario with two processing nodes.

Figure 2.10 shows the scenario of remote source coding with side information, which is proposed in [25, 26].

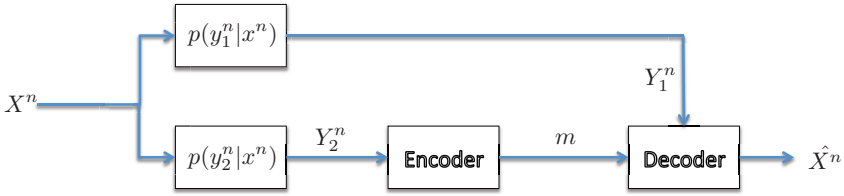


Figure 2.10: remote Wyner-Ziv Coding system.

The source X^n (n is large enough and realizations are i.i.d) is remotely observed via two noisy memory-less channels $p(y_2^n|x^n)$ and $p(y_1^n|x^n)$ at the encoder and decoder, respectively. Based on its observation Y_2^n, the encoder transmits a massage $m = f_n(Y_2^n)$ over a rate-constrained channel to the decoder. The decoder produces $\hat{X}^n$, an estimate of the source X^n, as a function of m and its side information Y_1^n. This scenario differs from Wyner-Ziv source coding because X^n is not directly observable at the encoder and it resembles the hearing aid scenario, where the target signal is first filtered by the acoustic scene (room) and then the degraded signal is received by the HAs.

We denote the rate-distortion function for this system by $R^*_{r-WZ}(D)$. Under certain conditions, the optimal rate can be computed by solving the following optimization problem

$$\begin{aligned} R^*_{r-WZ}(D) = \min_{q(z^n|y_2^n)} \quad & I(Y_2^n; Z^n|Y_1^n), \\ \text{subject to} \quad & \mathrm{E}[d(X^n, \hat{X}^n)] < D, \end{aligned} \tag{2.46}$$

where, $\hat{X}^n = f(Y_1^n, Z^n)$. The conditions to be satisfied are

1) $\sum_{z^n} p(x^n, y_1^n, y_2^n, z^n) = p(x^n, y_1^n, y_2^n)$.

2) Y_1^n, Z^n are conditionally independent given Y_2^n, i.e., $I(Y_2^n; Z^n|Y_1^n) = I(Y_2^n; Z^n) - I(Y_1^n; Z^n)$.

3) X^n, Z^n are conditionally independent given Y_1^n.

For i.i.d. Gaussian sources, where the observed signals are molded as $Y_1 = X + N_1$, and $Y_2 = X + N_2$, the system shown in Figure 2.10 can also be thought of as the model shown in Figure 2.11.

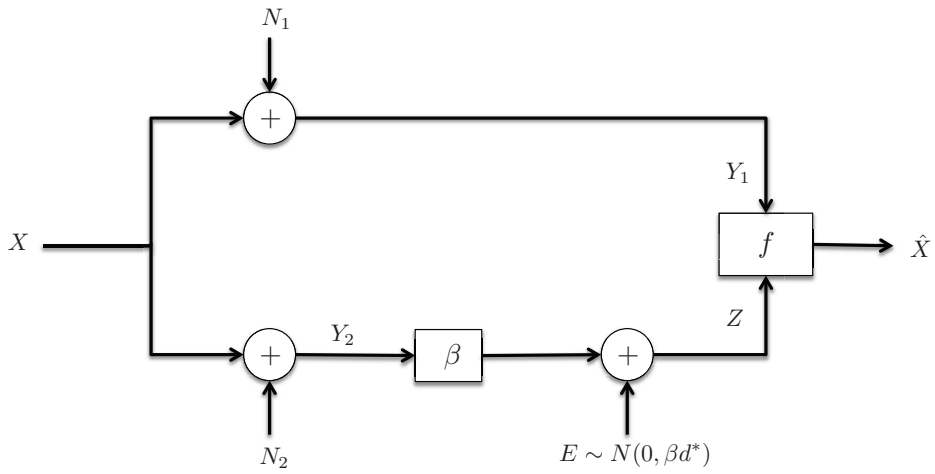


Figure 2.11: Interpretation of remote Wyner-Ziv coding system.

The auxiliary random variable is defined as $Z = \beta Y_2 + E$, where, $E \sim N(0, \beta d^*)$ is independent of Y_1 and Y_2. The optimal choices for the function f, scaling parameter β, and the direct quantization distortion value d^* will be derived later in this section.

The rate-distortion function in this case is derived in [25] as

$$R(D) = \begin{cases} \frac{1}{2}\log_2\left(\frac{\sigma^2_{X|Y_1} - \sigma^2_{X|Y_1,Y_2}}{D - \sigma^2_{X|Y_1,Y_2}}\right), & \sigma^2_{X|Y_1,Y_2} \le D \le \sigma^2_{X|Y_1} \\ 0, & D \ge \sigma^2_{X|Y_1} \end{cases} \tag{2.47}$$

where, $\sigma^2_{X|Y}$ generally is the minimum mean-squared estimation error in X given Y.

Based on the model shown in Figure 2.11, the estimated source $\hat{X}$ is computed as

$$\hat{X} = E[X|Y_1, Z] = f(Y_1, Z) = \frac{d^* + \sigma^2_{X|Y_1,Y_2}}{\sigma^2_{N_1}} Y_1 + (1 + \frac{\sigma^2_{N_2}}{\sigma^2_{X|Y_1}}) Z \tag{2.48}$$

where the optimal choices of d^* and β are

$$\begin{aligned} d^* &= D - \sigma^2_{X|Y_1,Y_2} \\ \beta &= \frac{\sigma^2_{X|Y_1} - D}{\sigma^2_{X|Y_1} + \sigma^2_{N_2}}. \end{aligned} \tag{2.49}$$

REFERENCES

[1] T. Quatieri *Discrete-time speech signal processing: principles and practice (First. ed.)*. Prentice Hall Press, USA, 2001.

[2] H. Kayser, S. D. Ewert, J. Anem¨uller, T. Rohdenburg, V. Hohmann, and B. Kollmeier, "Database of multichannel in-ear and behind-the-ear head-related and binaural room impulse responses," EURASIP J. Adv. Signal Process, vol. 2009, pp. 6:1–6:10, 2009.

[3] C. H. Taal, R. C. Hendriks, Heusdens R., and J. Jensen, "An algorithm for intelligibility prediction of timefrequency weighted noisy speech," IEEE Transactions on Audio, Speech, and Language Processing, vol. 19, no. 7, pp. 2125–2136, 2011.

[4] J. Jensen, C. H. Taal,"An Algorithm for Predicting the Intelligibility of Speech Masked by Modulated Noise Maskers," IEEE/ACM Trans. Audio, Speech and Lang. Proc., 2016.

[5] W. B. Kleijn, J. B. Crespo, R. C. Hendriks, P. Petkov, B. Sauert and P. Vary, "Optimizing Speech Intelligibility in a Noisy Environment: A unified view," in IEEE Signal Processing Magazine, vol. 32, no. 2, pp. 43-54, 2015.

[6] W. B. Kleijn and R. C. Hendriks, "A Simple Model of Speech Communication and its Application to Intelligibility Enhancement," in IEEE Signal Processing Letters, vol. 22, no. 3, pp. 303-307, 2015.

[7] S. Khademi, R. C. Hendriks and W. B. Kleijn, "Intelligibility Enhancement Based on Mutual Information," in IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 25, no. 8, pp. 1694-1708, 2017.

[8] O. Roy and M. Vetterli, "Collaborating hearing aids," in Proceedings of MSRI Workshop on Mathematics of Relaying and Cooperation in Communication Networks, 2006.

[9] L. W. Brooks and I. S. Reed, "Equivalence of the likelihood ratio processor, the maximum signal-to-noise ratio filter, and the Wiener filter," IEEE Transactions on Aerospace and Electronic Systems, vol. AES-8, no. 5, pp. 690–692, 1972.

[10] S. Doclo and M. Moonen, "GSVD-based optimal filtering for single and multimicrophone speech enhancement," IEEE Transactions on Signal Processing, vol. 50, no. 9, pp. 2230–2244, 2002.

[11] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*, Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1993.

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

[13] 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, no. 3, pp. 543–558, 2016.

[14] P. Vary and R. Martin, *Digital Speech Transmission: Enhancement, Coding And Error Concealment*, John Wiley and Sons, 2006.

[15] M. Brandstein and D. Ward, *Microphone Arrays: Signal Processing Techniques and Applications,* Berlin, Germany: Springer Science and Business Media, 2001.

[16] D. Marquardt, E. Hadad, S. Gannot, and S. Doclo, "Optimal binaural LCMV beamformers for combined noise reduction and binaural cue preservation," in 14th International Workshop on Acoustic Signal Enhancement (IWAENC), pp. 288–292, 2014.

[17] B. Cornelis, S. Doclo, T. Van dan Bogaert, M. Moonen, and J. Wouters, "Theoretical analysis of binaural multimicrophone noise reduction techniques," IEEE Transactions on Audio, Speech, and Language Processing, vol. 18, no. 2, pp. 342–355, 2010.

[18] A. I. Koutrouvelis, R. C. Hendriks, J. Jensen, and R. Heusdens, "Improved multimicrophone noise reduction preserving binaural cues," in 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 460–464, 2016.

[19] E. Hadad, D. Marquardt, D. Doclo, and S. Gannot, "Theoretical analysis of binaural transfer function mvdr beamformers with interference cue preservation constraints," IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 23, no. 12, pp. 2449–2464, 2015.

[20] T. Berger, *Rate-distortion theory: A mathematical basis for data compression,* Englewood Cliffs, N.J.: Prentice-Hall, 1971.

[21] T. M. Cover and J. A. Thomas, E*lements of information theory,* Wiley- Interscience, 2006.

[22] J. K. Wolf and J. Ziv, "Transmission of noisy information to a noisy receiver with minimum distortion," IEEE Transactions on Information Theory, vol. 16, no. 4, pp. 406–411, 1970.

[23] A. D. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," IEEE Transactions on Information Theory, pp. 1–10, 1976.

[24] D. Slepian and J. Wolf, "Noiseless coding of correlated information sources," IEEE Transactions on Information Theory, vol. 19, no. 4, pp. 471–480, 1973.

[25] S. C. Darper, "Successive structuring of source coding algorithms for data fusion, buffering and distribution in networks," Ph.D. dissertation, Massachusetts Institute of Technology, 2002.

[26] H. Yamamoto and K. Itoh, "Source coding theory for communication systems with a remote source," Trans. IECE Jpn, vol. E63, no. 6, pp. 700–706, 1980.

3

On the Impact of Quantization on Binaural MVDR Beamforming

This chapter is published as "On the Impact of Quantization on Binaural MVDR Beamforming," by J. Amini, R. C. Hendriks, R. Heusdens, M. Guo and J. Jensen, in Speech Communication; 12. ITG Symposium, Paderborn, Germany, 2016, pp. 1-5.

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 chapter, we study the impact of quantization as a data compression approach on the performance of binaural beamforming, which is from [8]. 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 that of 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.

3.1. 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 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

$$Y_i[k,l] = A_i[k,l]S[k,l] + \sum_{j=1}^{r} B_{ij}[k,l]U_j[k,l] + V_i[k,l], \tag{3.1}$$

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

$$\mathbf{y} = \mathbf{x} + \sum_{j=1}^{r} \mathbf{n}_j + \mathbf{v}, \tag{3.2}$$

where, $\mathbf{y} = [Y_1[k,l],...,Y_M[k,l]]^{\mathrm{T}}$, $\mathbf{v} = [V_1[k,l],...,V_M[k,l]]^{\mathrm{T}}$ (with left side signals stacked on top), $\mathbf{x} = \mathbf{a}S$, 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. 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 $\mathbf{\Phi}_{\mathbf{y}}$, is written as

$$\mathbf{\Phi}_{\mathbf{y}} = \mathbf{\Phi}_{\mathbf{x}} + \overbrace{\sum_{j=1}^{r} \mathbf{\Phi}_{\mathbf{n}_j} + \mathbf{\Phi}_{\mathbf{v}}}^{\mathbf{\Phi}}, \tag{3.3}$$

where,

$$\begin{aligned} \mathbf{\Phi}_{\mathbf{x}} &= E[\mathbf{x}\mathbf{x}^{\mathrm{H}}] = \sigma_s^2 \mathbf{a}\mathbf{a}^{\mathrm{H}}, \\ \mathbf{\Phi}_{\mathbf{n}_j} &= E[\mathbf{n}_j \mathbf{n}_j^{\mathrm{H}}] = \sigma_{u_j}^2 \mathbf{b}_j \mathbf{b}_j^{\mathrm{H}}, \qquad j = 1,...,r, \end{aligned} \tag{3.4}$$

and $\mathbf{\Phi}_{\mathbf{v}} = \sigma^2 I$. 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

$\hat{X}_L$ and $\hat{X}_R$ are the estimated clean signals at the left and right reference microphones, respectively, and $\mathbf{w}_L$ and $\mathbf{w}_R$ are the applied spatial filters. Notice that the use of $\mathbf{w}_R$ and $\mathbf{w}_L$ 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. 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.

3.2. BMVDR

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

$$\begin{aligned} \mathbf{w}_L^\star &= \underset{\mathbf{w}_L}{\text{argmin}} \quad \mathbf{w}_L^H \mathbf{\Phi} \mathbf{w}_L \quad \text{s.t.} \quad \mathbf{w}_L^H \mathbf{a} = A_L, \\ \mathbf{w}_R^\star &= \underset{\mathbf{w}_R}{\text{argmin}} \quad \mathbf{w}_R^H \mathbf{\Phi} \mathbf{w}_R \quad \text{s.t.} \quad \mathbf{w}_R^H \mathbf{a} = A_R, \end{aligned} \tag{3.5}$$

where $\mathbf{\Phi}$ is the CPSD matrix of the noise, see (3.3). Solving (3.5), the optimal weight vectors are computed as

$$\mathbf{w}_L^\star = \frac{\mathbf{\Phi}^{-1}\mathbf{a}}{\mathbf{a}^H \mathbf{\Phi}^{-1} \mathbf{a}} \bar{A}_L, \quad \mathbf{w}_R^\star = \frac{\mathbf{\Phi}^{-1}\mathbf{a}}{\mathbf{a}^H \mathbf{\Phi}^{-1} \mathbf{a}} \bar{A}_R, \tag{3.6}$$

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

3.3. 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, $x_{\min} \leq x \leq x_{\max}$, to the quantized range $x_{\min} \leq \hat{x} \leq x_{\max}$, 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 [9]. 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 as the QN is denoted by $e = \hat{x} - x$, and is determined by the value of the stepsize Δ. Under certain conditions [10, 11], e has a uniform distribution, that is,

$$p(e) = \begin{cases} \Delta^{-1}, & -\frac{\Delta}{2} \leq e \leq \frac{\Delta}{2} \\ 0, & \text{otherwise,} \end{cases} \tag{3.7}$$

with variance $\sigma_e^2 = \frac{\Delta^2}{12}$, for small values of $Delta$. 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 quantization noise (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.

3.4. QUANTIZATION AWARE BEAMFORMING

In Section 3.1 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 beamforming 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 $\mathbf{y}$ is defined in (3.2) 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

$$\begin{aligned} \mathbf{w}_{\mathrm{L}}^{\star} &= \underset{\mathbf{w}_{\mathrm{L}}}{\operatorname{argmin}} \quad \mathbf{w}_{\mathrm{L}}^{\mathrm{H}} \mathbf{\Phi}_{\mathrm{nL}} \mathbf{w}_{\mathrm{L}} \quad \text{s.t.} \quad \mathbf{w}_{\mathrm{L}}^{\mathrm{H}} \mathbf{a} = A_{\mathrm{L}}, \\ \mathbf{w}_{\mathrm{R}}^{\star} &= \underset{\mathbf{w}_{\mathrm{R}}}{\operatorname{argmin}} \quad \mathbf{w}_{\mathrm{R}}^{\mathrm{H}} \mathbf{\Phi}_{\mathrm{nR}} \mathbf{w}_{\mathrm{R}} \quad \text{s.t.} \quad \mathbf{w}_{\mathrm{R}}^{\mathrm{H}} \mathbf{a} = A_{\mathrm{R}}, \end{aligned} \tag{3.8}$$

where,

$$\mathbf{\Phi}_{\mathrm{nL}} = \mathbf{\Phi} + \mathbf{\Phi}_{e_{\mathrm{L}}}, \quad \mathbf{\Phi}_{\mathrm{nR}} = \mathbf{\Phi} + \mathbf{\Phi}_{e_{\mathrm{R}}}. \tag{3.9}$$

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

$$\begin{aligned} \mathbf{\Phi}_{\mathrm{nL}} &= \mathbf{\Phi} + \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Phi}'_{e_{\mathrm{L}}} \end{bmatrix}, \qquad \mathbf{\Phi}'_{e_{\mathrm{L}}} \in \mathbf{R}^{\mathbf{M}_{\mathrm{R}} \times \mathbf{M}_{\mathrm{R}}}, \\ \mathbf{\Phi}_{\mathrm{nR}} &= \mathbf{\Phi} + \begin{bmatrix} \mathbf{\Phi}'_{e_{\mathrm{R}}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}, \qquad \mathbf{\Phi}'_{e_{\mathrm{R}}} \in \mathbf{R}^{\mathbf{M}_{\mathrm{L}} \times \mathbf{M}_{\mathrm{L}}}. \end{aligned} \tag{3.10}$$

Note that in (3.9) and (3.10) 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 (3.8) are functions of the bit-rate r. For simplicity we assume in this chapter 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.

3.5. VALIDITY OF ASSUMPTIONS

In (3.9) and (3.10) 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.

3.5.1. 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,

$$D=\frac{\sum_{i=1}^{M_\mathrm{R}}[|\mathbf{\Phi}'_{e_\mathrm{L}}|]_{ii}{}^2-\sum_{i=1}^{M_\mathrm{R}}\sum_{j=1}^{M_\mathrm{R}}[|\mathbf{\Phi}'_{e_\mathrm{L}}|]_{ij}{}^2}{\sum_{i=1}^{M_\mathrm{R}}\sum_{j=1}^{M_\mathrm{R}}[|\mathbf{\Phi}'_{e_\mathrm{L}}|]_{ij}{}^2}. \tag{3.11}$$

This measure can be interpreted as a normalized distance between the sum of all entries and the sum of diagonal entries of the matrix $\mathbf{\Phi}'_{e_\mathrm{L}}$. 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 $D_{\min}=\frac{M_\mathrm{R}a^2-M_\mathrm{R}{}^2a^2}{M_\mathrm{R}{}^2a^2}=\frac{1}{M_\mathrm{R}}-1$. In the best case where the signals are highly uncorrelated, the value D approaches zero. In general, $(\frac{1}{M_\mathrm{R}}-1)\leq D\leq 0$, the more negative, the larger off-diagonal entries. The closer to zero, the more diagonally dominant.

3.5.2. 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 $\mathbf{\Phi}_{e_\mathrm{L}}$ and $\mathbf{\Phi}$ should be equal to the CPSD matrix of the total noise, $\mathbf{\Phi}_\mathrm{nL}$ according to (3.9). 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 (3.9) as

$$E=\frac{\sum_{i=1+M_\mathrm{L}}^{M}\sum_{j=1+M_\mathrm{L}}^{M}[|\mathbf{\Phi}_\mathrm{nL}-\mathbf{\Phi}-\mathbf{\Phi}_{e_\mathrm{L}}|]_{ij}{}^2}{\sum_{i=1+M_\mathrm{L}}^{M}\sum_{j=1+M_\mathrm{L}}^{M}[|\mathbf{\Phi}_\mathrm{nL}|]_{ij}{}^2}. \tag{3.12}$$

3.6. 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.

3.6.1. SETUP AND SIMULATION PARAMETERS

A typical acoustic scene, which we use in this chapter, is illustrated in Figure 3.1. In the experiments the exact source positions are not necessarily the same as those in Figure 3.1. 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 $R = \sqrt{x^2 + y^2}$. We define θ in a way that zero degree corresponds to the front of the virtual head (like green circle in Figure 3.1). Four "+" symbols denote four virtual omnidirectional 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.

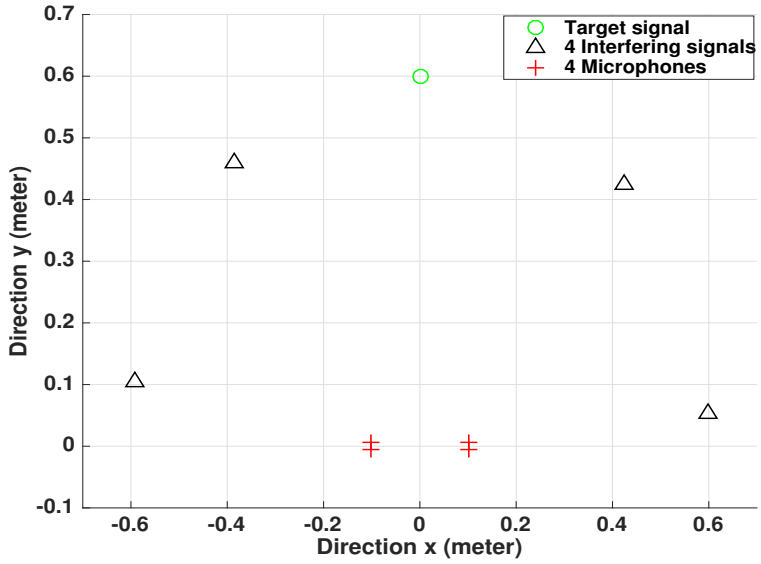


Figure 3.1: An example acoustic scene. Note that the exact source positions can be potentially different in the experiments.

3.6.2. VALIDATION OF ASSUMPTIONS: RESULTS

Based on the two measures, introduced in Section 3.5, we evaluate for which bit-rates the assumptions hold. Moreover, we apply dithering (Section 3.3) as a decorrelation process to assure that the assumptions on the QN in (3.9) and (3.10) 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 Figure 3.1. 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 (3.11). Therefore, we can isolate the effect of position dependency of the noise source on the total performance.

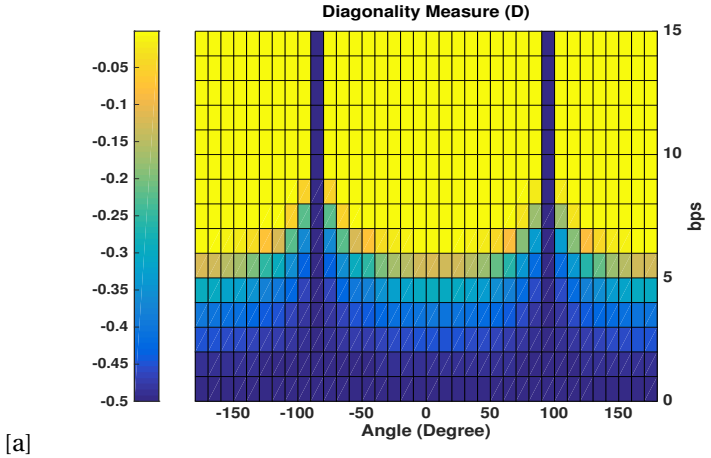
The results of the D measure in (3.11) in terms of the bit per sample (bps) and the angle, before and after dithering are shown in Figure 3.2. As shown in Figure 3.2a, at higher rates the assumption holds and the CPSD matrix of the QN ($\mathbf{\Phi}'_{e_{\mathrm{L}}}$) becomes more-and-more diagonal ($D \rightarrow 0$) with increasing rates. The results show that if the interfering source is positioned at either ±90 degrees (left or right side of the virtual head), the $\mathbf{\Phi}'_{e_{\mathrm{L}}}$ is fully correlated even at high rates, i.e., $D = -0.5$. After applying dithering, $\mathbf{\Phi}'_{e_{\mathrm{L}}}$ becomes diagonal at all rates and angles, as shown in Figure 3.2b.

Similarly, the results of the "correlation measure" (E in (3.12)) are shown in Figure 3.3 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 Figure 3.3a to the maximum value of 0.0013 in Figure 3.3b), even at low bit-rates. This means that after dithering the QN and environmental noise become almost uncorrelated at all rates and angles.

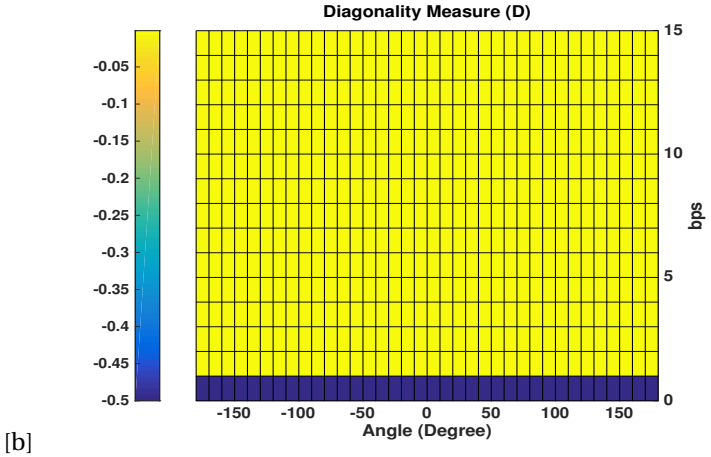
3.6.3. PERFORMANCE EVALUATION

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.
- Case 3) **Proposed method version 1 without dithering**: beamforming based on (3.8), 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 (3.8), i.e., taking into account the QN by estimating the modified total noise CPSD



[a]



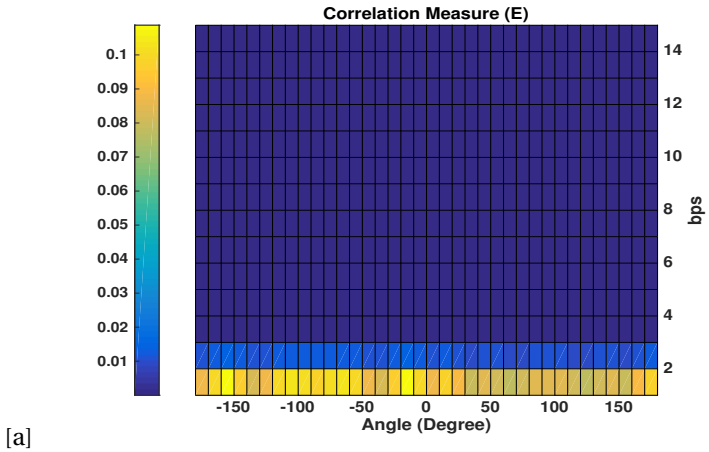
[b]

Figure 3.2: Diagonality measure:(a) without, and (b) with dithering

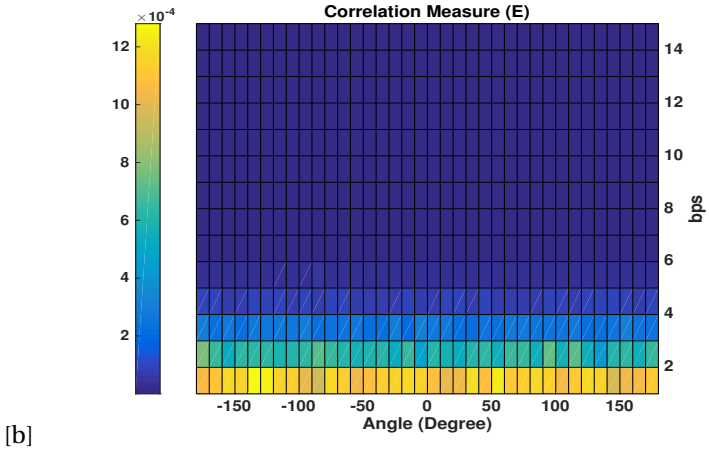
$\mathbf{\Phi}_{\mathrm{nL}}$ using (3.9). Therefore unlike the case 3, the QN and the environmental noise is assumed to be uncorrelated and no extra information need be to transmitted.

- Case 5) **traditional BMVDR**: beamforming based on (3.5), 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 Figure 3.4 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



[a]



[b]

Figure 3.3: Correlation measure:(a) without, and (b) with dithering

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 Figure 3.4, 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, 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.

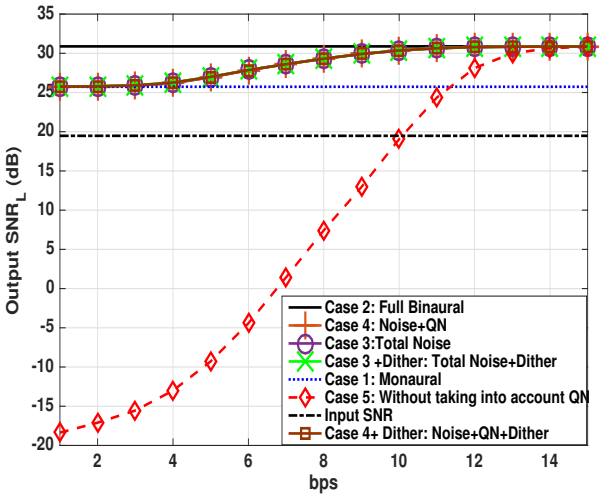


Figure 3.4: Output SNR performance for the left-sided reference microphone.

3.7. CONCLUSIONS

In this chapter 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, 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 binau-

ralmvdr beamformer," in International Workshop on Acoustic Signal Enhancement (IWAENC), 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, 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, 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, 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] J. Amini, R. C. Hendriks, R. Heusdens, M. Guo and J. Jensen, "On the Impact of Quantization on Binaural MVDR Beamforming," Speech Communication; 12. ITG Symposium, Paderborn, Germany, pp. 1-5, 2016.

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

[10] 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, 1977.

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

4

Asymmetric Coding for Rate-Constrained Noise Reduction in Binaural Hearing Aids

This chapter is published as "Asymmetric Coding for Rate-Constrained Noise Reduction in Binaural Hearing Aids," by J. Amini, R. C. Hendriks, R. Heusdens, M. Guo and J. Jensen, in IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 27, no. 1, pp. 154-167, Jan. 2019.

Nowdays hearing aids (HAs) include digital signal processing algorithms to improve the intelligibility of the signal of interest. The HAs record the acoustic field with one or multiple microphones and try to improve the intelligibility of the desired speech signal while reducing environmental noise [1]. Using a wireless link, the two HAs can collaborate with each other and construct a binaural HA system to increase noise suppression and potentially preserve some important binaural (spatial) cues [2]. This leads to the notion of binaural multi-microphone noise reduction which may have a better speech intelligibility compared to single-microphone solutions [3], [4]. In addition, unlike bilateral systems or independent dual-channel systems, where a set of two monaural systems operate independently from each other (no collaboration between the devices), the binaural system exploits the spatial diversity, using more observations, and interaural information to increase the speech intelligibility [5]. A well known binaural multi-microphone noise reduction algorithm is the binaural multichannel Wiener filter (MWF) [6–9]. The binaural MWF includes two separate MWF beamformers for the two HAs. Each MWF beamformer combines its local observations with those of the contralateral HA to estimate its own version of the target signal such that the mean square error (MSE) between the target signal and its estimate is minimized. The binaural MWF allows for more degree of freedom for noise reduction [5], among all the MMSE-based single-channel or dual-channel speech enhancement methods [10], as it exploits the full potential of binaural processing by exchanging the signals between the HA devices (more microphone observations, the better the noise reduction performance) and as it outputs the best MSE estimate of the target signal. Binaural filters require that noisy observations from one HA are transmitted to the other one (e.g. through a wireless link) in order to be combined with local observations. Typically, transmission capacities are limited due to limited battery life-time [11, 12], which necessitates data compression. Ideally, the algorithm trades off the transmission bit-rate of contralateral HA observations against the estimation error of the target signal [12, Sec. I], which is remotely (i.e., indirectly after being filtered by the room channel) observable at the HAs .

From an information theoretic viewpoint, such an estimator can be seen as remote source coding [13–15]. Since the beamformer needs to decode the transmitted signals and combine these with its local observations (which are available error-free as "side information"), the more accurate problem formulation would be that of remote source coding with side information at the decoder. For directly observable sources this problem is referred to as Wyner-Ziv (WZ) coding [16] and for remote (i.e., indirectly observable) sources as remote WZ coding [17]. In fact, in remote source coding problems the (remote) target signals are of interest, and not necessarily the (noisy) direct observations. In [12, Sec. III-A] this problem is considered, assuming jointly Gaussian random sources, and an optimal tradeoff between the transmission rate and the MSE between the target signal and its estimate is derived. The method provides an upper bound on the performance of the minimum MSE (MMSE)-based rate-constrained binaural beamforming algorithms. However, the requirement of knowing the (joint) statistics severely limits the application of the method in practice. In fact, the joint statistics between the two HA observations remain unknown in practice and can only be estimated if realizations are exchanged between the HAs.

Several sub-optimal approaches are proposed in [12, Sec. III-B], [18–20] in which lo-

cal functions of the contralateral observations are transmitted, projecting the multiple signals on to a single signal. These methods provide practical alternatives to the optimal method in [12, Sec. III-A], as they do not need the knowledge of the (joint) statistics. However, the blind projection of the multi-microphone observations to one signal tends to a significant asymptotic mismatch in the performance even at sufficiently high rates [18]. An iterative reduced bandwidth MWF-based beamformer is proposed in [20], where local estimates of the target signal are assumed to be exchanged error-free between the HAs without any rate constraint. It is shown in [20] that for a single target signal, the iterative approach converges to the binaural MWF after sufficient transmissions between HAs. However, when analyzing the rate-constrained scenario [21], the total rate is distributed over transmissions (iterations) which results in a poor final performance (after convergence) in terms of bit-rate.

Typically, the aforementioned sub-optimal approaches do not make use of the joint statistics between the two HA observations. As a result, before exchanging the signals, some information will be removed which may be necessary for the other side to cancel out the noise sources, leading to an asymptotic sub-optimality. By asymptotic sub-optimality, we mean that the performance does not approach the optimal performance for increasing rate. Therefore, any knowledge (even incomplete) of the (joint) statistics between the nodes may be crucial to keep the necessary information when filtering the information, resolving the asymptotic sub-optimality, and to provide a good tradeoff between the rate and the distortion. This motivates trying to estimate the joint statistics.

In this chapter, we study the performance of sub-optimal rate-constrained noise reduction techniques based on a unified encoding-decoding framework which can easily be translated to the existing sub-optimal schemes by changing certain parameters. Moreover, we propose an asymmetric sequential coding approach for the transmission of the information from one HA to the other HA (which we will refer to as *Link 1*) and vice versa (which we will refer to as *Link 2*). In addition, we propose an extension of the probability distribution preserving quantization method [22], to vector sources, to be used in *Link 1*. Using this distribution preserving quantization, the unquantized statistics can be retrieved and used to apply the optimal coding strategy [12, Sec. III-A] in *Link 2*. Based on the MSE criteria, the distortion gap between the monaural noise reduction approach, in which there is no communication between devices, and different sub-optimal/optimal noise reduction approaches are compared for both links. The results show that the proposed methods outperform the sub-optimal approaches in most practical scenarios and confirm the optimal asymptotic behavior of the proposed methods.

The chapter is organized as follows. In Sec. 4.1 the binaural HA problem is stated and the well-known information theoretic rate-distortion tradeoff is introduced. In Sec. 4.2 we state the rate-constrained noise reduction problem in a unified framework and the optimal and some sub-optimal approaches are explained. The proposed asymmetric 2-way coding scheme is presented in Sec. 4.3. The performance analysis of the proposed and existing methods is carried out in Sec. 4.4. Finally, Sec. **??** concludes the chapter.

4.1. PROBLEM STATEMENT

4.1.1. SIGNAL MODEL

A typical binaural HA system consists of two wireless collaborating HAs. Assume that the left-side and right-side HAs include M_1, and M_2 microphones, respectively, with $M = M_1 + M_2$ microphones in total. All microphones record a filtered version of the target speech signal which is denoted in the frequency domain by $S[k]$, corrupted by and additive noise $N[k]$, with k the discrete frequency bin index. The frequency-domain description of the noisy observation captured by the ith microphone is given by

$$Y_i[k] = A_i[k]S[k] + N_i[k], \tag{4.1}$$

where $i = 1,...,M$, A_i is the acoustic transfer function (ATF) between the target signal and the ith microphone. Stacking all noisy observations across the microphones in a vector, the signal model can be rewritten as

$$\mathbf{y} = \mathbf{x} + \mathbf{n}, \quad \mathbf{y} \in \mathbb{C}^M. \tag{4.2}$$

where $\mathbf{y} = [Y_1[k],...,Y_{M_1}[k], Y_{M_1+1}[k],...,Y_M[k]]^{\mathrm{T}}$ denotes the total M noisy microphone signals, $\mathbf{x} = \mathbf{a}S$, and

$$\mathbf{a} = [A_1[k],...,A_M[k]]^{\mathrm{T}}, \mathbf{n} = [N_1[k],...,N_M[k]]^{\mathrm{T}}.$$

Note that the frequency index k is omitted, when defining the signal vectors, for ease of notation. To distinguish between the left-side and the right-side noisy microphone observations, vectors $\mathbf{y}_1 \in \mathbb{C}^{M_1}$ and $\mathbf{y}_2 \in \mathbb{C}^{M_2}$ are defined, respectively, as $\mathbf{y}_1 = [Y_1[k],...,$ $Y_{M_1}[k]]^{\mathrm{T}}$ and $\mathbf{y}_2 = [Y_{M_1+1}[k],...,Y_M[k]]^{\mathrm{T}}$. The superscripts $(\cdot)^{\mathrm{T}}$ and $(\cdot)^{\mathrm{H}}$ denote transpose and conjugate transpose operators, respectively. All sources are assumed to be zero-mean and mutually uncorrelated. The cross-power spectral density (CPSD) matrix of the noisy signal vector $\mathbf{y}$, denoted by $\mathbf{\Phi}_{\mathbf{y}}$, can then be written as $\mathbf{\Phi}_{\mathbf{y}} = \mathbf{\Phi}_{\mathbf{x}} + \mathbf{\Phi}_{\mathbf{n}}, \mathbf{\Phi}_{(\cdot)} \in \mathbb{C}^{M\times M}$, with $\mathbf{\Phi}_{\mathbf{x}} = \Phi_S \mathbf{a}\mathbf{a}^{\mathrm{H}}$. Here, Φ_S is the power spectral density (PSD) of the clean speech signal S and $\mathbf{\Phi}_{\mathbf{n}} = E[\mathbf{n}\mathbf{n}^{\mathrm{H}}]$, where $E[\cdot]$ denotes the expectation.

The goal of the multi-microphone noise reduction algorithms is to estimate the clean speech signal while suppressing the environmental noise power. The binaural MWF [7, 8] consists of two filters (the left-side and the right-side filters). Let the left and right reference microphone indices be denoted by 1 and $M_1 + 1$, respectively. The filters estimate the target signal at the left-side and the right-side reference microphones, say $S_1 = A_1 S$ and $S_2 = A_{M_1+1}S$, respectively, by minimizing the MSE between the target signal and its estimates, say $\hat{S}_1$ and $\hat{S}_2$, respectively. Scalars A_1 and A_{M_1+1} denote the ATFs with respect to the corresponding reference microphones. Finally, the estimates are given by $\hat{S}_1^\star = E[S_1|\mathbf{y}]$ and $\hat{S}_2^\star = E[S_2|\mathbf{y}]$.

Computing each of these MWF outputs requires the availability of the error-free contralateral noisy signal realizations. In practice, only a compressed/quantized version of the contralateral noisy signals are available. These signals are compressed at a certain rate. Therefore, the problem can be viewed as a rate-constrained estimation task which will be described in the next section.

4.1.2. RATE-DISTORTION FUNCTION (RDF) [13, CH. 4]

Let $s^N = \{s[i]\}_{i=1}^N$ be a sequence of discrete stationary Gaussian source samples, where $s[i] \in \mathbb{C}$ and N is the number of samples. The sequence s^N can be thought of as a single microphone observation along the time-axis. Assume that the encoder maps the sequence with R bits per sample to a bit sequence. The decoder receives the bit sequence and produces the quantized sequence $\hat{s}^N = \{\hat{s}[i]\}_{i=1}^N$.

The direct rate-distortion problem is to find the minimum asymptotic achievable rate at which the sequence can be encoded such that the reconstruction error does not exceed a certain value D, as $N \to \infty$ [13]. The problem is solved in [13, Ch. 4] and a parametric rate-distortion tradeoff is found, analytically. To achieve the optimal tradeoff [13, Ch. 4], for a stationary Gaussian source, the optimal forward test channel interpretation is presented in [13, Ch. 4] and [18, Sec. 3], which is illustrated in Fig. 4.1, where the quantization procedure in the frequency domain can be thought of as a test channel with input $s(\Omega)$ and output $\hat{s}(\Omega)$. The channel noise $e(\Omega)$ is uncorrelated to the input source $s(\Omega)$. The quantization parameters are computed as [13]

$$\beta(\Omega) = \max(0, 1 - \frac{\theta}{\Phi_s(\Omega)}),\ \Phi_e(\Omega) = \max(0, \theta(1 - \frac{\theta}{\Phi_s(\Omega)})), \tag{4.3}$$

where Φ_s is the PSD of the sequence s^N, $N \to \infty$, and $\theta \in (0, \sup \Phi_s]$ denotes the "reverse water filling" threshold parameter [13, 23].

4.2. RATE-CONSTRAINED NOISE REDUCTION

Binaural rate-constrained noise reduction (RCNR) aims at estimating the target signal at the reference microphones, given some local observations and the quantized contralateral observations, such that the communication rate between HAs is minimized, satisfying a certain constraint. For the right-side beamformer, the local observation vector $\mathbf{y}_2$ acts as the "side information" and $\mathbf{y}_1$ as the contralateral observations. A similar argument holds for the left-side beamformer.

Most approaches like [12, 18] use the following structure of three stages. First, the contralateral observations are filtered prior to being quantized, as we are interested in information about the target source (S_1 or S_2) and not necessarily in those of the contralateral noisy observations themselves (filtering stage). Second, the filtered signals are quantized and transmitted to the other side (quantization stage). Finally, the target signal is estimated given the side information and the filtered-quantized contralateral observations (estimation stage). Existing approaches differ in how different operators for these three processing stages of filtering, quantization, and final estimation are chosen, which we will explain using the above-mentioned unified description.

4.2.1. OPTIMAL RATE-CONSTRAINED NOISE REDUCTION

Encoding the sources for a decoder which has access to the side information is known as the Wyner-Ziv (WZ) problem [16]. Based on the WZ coding, In [12, Sec. III-A] the problem is optimally solved for the multiple microphones per HA setup (binaural setup) and the optimal rate-distortion tradeoff is found analytically. For stationary Gaussian sources, the interpretation of the optimal RCNR system is illustrated in Fig. 4.2 [12, Sec.

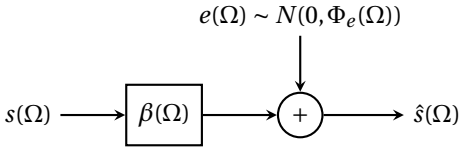


Figure 4.1: Forward test channel representation of lossy source coding.

III-A]. From the right-side beamformer's perspective, first, the left-side noisy signals ($\mathbf{y}_1$) are filtered as $Y_{12} = (\mathbf{w}_1^o)^{\mathrm{H}}\mathbf{y}_1$, using the joint statistics between the HA observations, with the optimal coefficients $\mathbf{w}_1^o$ computed as

$$\mathbf{w}_1^o = \mathbf{\Phi}_{\bar{\mathbf{y}}_1}^{-1}\mathbf{\Phi}_{\bar{\mathbf{y}}_1 S_2}, \quad \mathbf{w}_1^o \in \mathbb{C}^{M_1 \times 1} \tag{4.4}$$

where $\mathbf{\Phi}_{\bar{\mathbf{y}}_1}$ is the CPSD matrix of the direct innovation process $\bar{\mathbf{y}}_1$ defined as $\bar{\mathbf{y}}_1 = \mathbf{y}_1 - E[\mathbf{y}_1|\mathbf{y}_2]$ and $\mathbf{\Phi}_{\bar{\mathbf{y}}_1 S_2}$ is the CPSD vector between S_2 and $\bar{\mathbf{y}}_1$.

Second, using the WZ coding philosophy [16], the filtered signal Y_{12} will be optimally encoded, knowing the joint statistics and that the decoder has access to $\mathbf{y}_2$. The WZ-based decoder (Fig. 4.2) consists of an MMSE estimator, which estimates S_2, the target signal at the right-side reference microphone, given $\mathbf{y}_2$ and the quantized version of Y_{12}. See [12, Sec. III-A] for more details.

4.2.2. SUB-OPTIMAL RATE-CONSTRAINED NOISE REDUCTION

Achieving the optimal rate-distortion tradeoff as in [12, Sec. III-A] requires knowledge of the joint statistics between the noisy signals from both HAs, which are not available in practice. In [12, Sec. III-B] a sub-optimal method is presented in which a local estimate of the target signal, without using the correlation between the two HA observations, is transmitted to the contralateral device. However, in the presence of point noise sources, the performance does not approach the ideal binaural performance, not even asymptotically (at infinite bit-rate) and a significant loss will occur at high rates, as confirmed by experiments in Sec. 4.4. Two alternatives to the method presented in [12, Sec. III-B] were proposed in [18, 19]. We briefly explain these sub-optimal methods based on a unified communication scheme illustrated in Fig. 4.3.

Unlike [12, Sec. III-A], the sub-optimal filter $\mathbf{w}_1^s$, shown in Fig. 4.3, is only a function of the local observations. The above-mentioned sub-optimal methods differ from each other in how the filter $\mathbf{w}_1^s$ is chosen. For example, in [12, Sec. III-B], $\mathbf{w}_1^s$ denotes a filter which locally estimates the target signal, without any access to the side information.

The quantization stage in Fig. 4.3 can be represented by the forward test channel (Fig. 4.1) with input Y_{12} and output $\tilde{Y}_{12}$. The final MMSE estimate of the desired signal S_2 is given by $\hat{S}_2 = f(\mathbf{y}_2, \tilde{Y}_{12})$ and the corresponding MMSE by [18]

$$D_2(\theta) = \frac{1}{2\pi}\int_0^{2\pi} [\Phi_{S_2}(\Omega) - \mathbf{\Phi}_{S_2\tilde{\mathbf{y}}}(\Omega)\mathbf{\Phi}_{\tilde{\mathbf{y}}}^{-1}(\Omega)\mathbf{\Phi}_{\tilde{\mathbf{y}}S_2}(\Omega)]\,d\Omega, \tag{4.5}$$

where $\tilde{\mathbf{y}} = [\mathbf{y}_2^{\mathrm{T}}, \tilde{Y}_{12}]^{\mathrm{T}}$. Note that $\tilde{Y}_{12}$ depends on the rate of transmission.

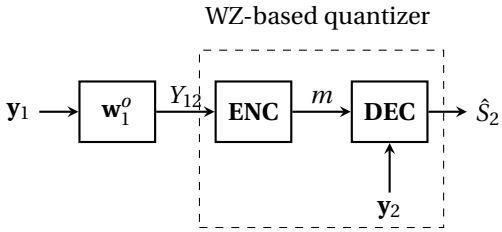


Figure 4.2: Optimal Rate-Constrained Beamforming.

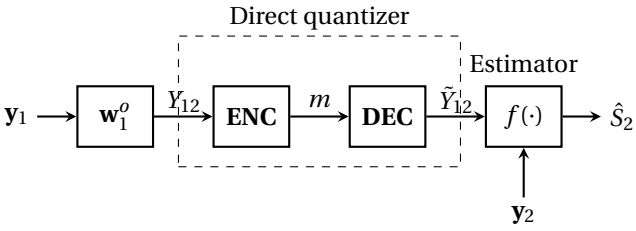


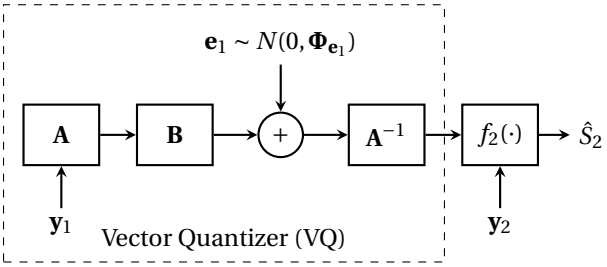
Figure 4.3: Sub-Optimal Rate-Constrained Beamforming.

4.3. ASYMMETRIC CODING FOR RCNR

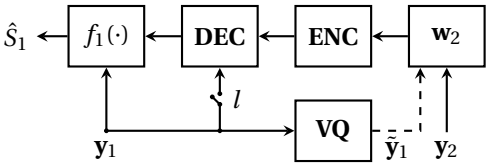
As described in Sec. 4.2.2, the main limitation of the sub-optimal methods is that they are not even asymptotically optimal, because of the blind filtering stage at the start of the communication chain. This results in a significant performance loss, as shown in Sec. 4.4.

Our proposed idea is to leave all contralateral observations active, at least in one direction (for example, in the transmission from left-to-right) in order to exploit some statistics which will be helpful for informed coding in the other direction. This brings the notion of vector source quantization into the estimation process in one link. Assuming stationarity in the time domain and sufficiently large sample sequences, the quantization can be performed in the frequency domain assuming independent frequency bins. However, it is noteworthy that microphones at different spatial positions generally capture the sources with different powers, resulting in spatially non-stationary signals. Therefore the optimal quantization for microphone signals is done in the frequency domain along the time axis and in the eigenvalue domain of the CPSDs along the space axes. This will become more clear in Sec. 4.3.1.

We propose a 2-way sequential coding scheme for communication between two HAs. This scheme is asymmetric in the sense that the quantization in one link is different from the other link. The proposed coding scheme is sequential meaning that some information is exploited in one link to be used in coding in another link. The scheme is illustrated in Fig. 4.4. The link where the communication starts is referred to as "*Link 1*" and vice versa as "*Link 2*". Let us assume the communication starts from the left HA to the right HA. In the following, we explain the proposed architecture.



(a) *Link 1*: from the left HA to the right HA.



(b) *Link 2*: from the right HA to the left HA.

Figure 4.4: Proposed Asymmetric 2-way Coding scheme.

4.3.1. *Link 1*: FROM LEFT-TO-RIGHT

Unlike the common RCNR techniques, in this method, the observation vector $\mathbf{y}_1$ is not projected onto a scalar signal (i.e., without filtering) for two reasons. First, we wish to resolve the asymptotic sub-optimality problem at high rates, and secondly, we wish to exploit the joint statistics at the right-side HA to reduce the redundancy in information transmission in *Link 2*. We introduce two methods based on the architecture in Fig. 4.4a.

Method 1: RDF FOR VECTOR SOURCES WITH MEMORY

In Sec. 4.1.2 we explained the RDF for a time-stationary Gaussian source which accounts for one sensor observation in time. In order to quantize more than one observation, an extension to vector sources is required, which is presented in [24]. Recall that in the scalar case (Sec. 4.1.2), the correlation matrix Φ_s is (for $N \to \infty$) diagonalizable by the Fourier transform and, hence, the RDF can be written in terms of the PSD of the stationary source, i.e., Φ_s [13]. Different from the scalar case in Sec. 4.1.2, the correlation matrices involving multiple microphone observations are not diagonalizable by the (spatial) Fourier transform. Therefore, the resulting RDF for vector sources of such (spatially) non-stationary sources is different from that of the scalar case in Sec. 4.1.2, and will be explained in the following.

Given a discrete-time sequence of zero-mean time-stationary vector Gaussian sources, say $\{\mathbf{s}[n]\}_{n=0}^{N-1}$, where $\mathbf{s}[\cdot] \in \mathbb{R}^{M\times 1}$ can be any vector source (like the noisy observations),

the cross correlation matrix is given by

$$\Sigma_{\mathbf{s}} = \begin{bmatrix} \Sigma_0 & \Sigma_{-1} & \dots & \Sigma_{-(N-1)} \\ \Sigma_1 & \ddots & \ddots & \Sigma_{-(N-2)} \\ \vdots & \ddots & \ddots & \vdots \\ \Sigma_{(N-1)} & \Sigma_{(N-2)} & \dots & \Sigma_0 \end{bmatrix} \tag{4.6}$$

where $\Sigma_{\mathbf{s}} \in \mathbb{R}^{NM\times NM}$ is a block-Toeplitz matrix. Matrices $\Sigma_i \in \mathbb{R}^{M\times M}$, $\quad i = -(N-1),\dots,(N-1)$ are of entries $[\Sigma_i]_{uv} = E[s_u[n+i]s_v[n]]$, for all $0 \le n \le N-1$. The scalars $s_u[\cdot]$ is the uth entry in $\mathbf{s}[\cdot]$ and $[\Sigma_i]_{uv}$ is the (statistical) cross correlation, $u,v = 1,\dots,M$. Stacking $\{\mathbf{s}[n]\}_{n=0}^{N-1}$ into a vector, say $\mathbf{s}_{\text{vec}} = [\mathbf{s}^T[0]...\mathbf{s}^T[N-1]]^T$, the rate-distortion tradeoff for stationary vector $\mathbf{s}_{\text{vec}}$ is given by [13]

$$\begin{aligned} R_{MN}(\theta) &= \sum_{i=1}^{MN} \max(0, \frac{1}{2}\log\frac{\lambda_i(\Sigma_{\mathbf{s}})}{\theta}) \\ D_{MN}(\theta) &= \sum_{i=1}^{MN} \min(\theta, \lambda_i(\Sigma_{\mathbf{s}})), \end{aligned} \tag{4.7}$$

where $\lambda_i(\Sigma)$ is the ith eigenvalue of a matrix Σ. From Szego's theorem [25], the asymptotic eigenvalue distribution of Toeplitz matrices corresponds to those of the PSD values in the frequency domain. The extensions to the Szego's theorem are proposed in [26] and [27] which state that for any Hermitian block-Toeplitz matrix (here $\Sigma_{\mathbf{s}}$) the asymptotic behavior of an arbitrary function of eigenvalues follows that of corresponding CPSD matrices in the frequency domain, i.e.,

$$\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{MN} F(\lambda_i(\Sigma_{\mathbf{s}})) = \frac{1}{2\pi}\int_{-\pi}^{\pi}\sum_{u=1}^{M} F(\lambda_u(\mathbf{\Phi}_{\mathbf{s}}(\Omega)))d\Omega, \tag{4.8}$$

where $\mathbf{\Phi}_{\mathbf{s}}$ is the CPSD matrix with respect to the vector sequence $\{\mathbf{s}[n]\}_{n=0}^{N-1}$ with elements $[\mathbf{\Phi}_{\mathbf{s}}]_{uv} = \sum_{k=-\infty}^{\infty}[\Sigma_k]_{u,v}e^{-j\Omega k}$. $F(\cdot)$ is an arbitrary function applied on the eigenvalues. Based on (4.8), asymptotically as $N\to\infty$, the RDF in (4.7) can be rewritten as

$$\begin{aligned} R^*(\theta) &= \lim_{N\to\infty}\frac{1}{N}R_{MN} \\ &= \frac{1}{2\pi}\int_{-\pi}^{\pi}\sum_{u=1}^{M}\max(0,\frac{1}{2}\log\frac{\lambda_u(\mathbf{\Phi}_{\mathbf{s}}(\Omega))}{\theta})d\Omega \\ D^*(\theta) &= \lim_{N\to\infty}\frac{1}{N}D_{MN} = \frac{1}{2\pi}\int_{-\pi}^{\pi}\sum_{u=1}^{M}\min(\theta,\lambda_u(\mathbf{\Phi}_{\mathbf{s}}(\Omega)))d\Omega, \end{aligned} \tag{4.9}$$

where the rate $R^*(\theta)$ is per source vector $\mathbf{s}[n] \in \mathbb{R}^{M\times 1}$. The entries of the vector $\mathbf{s}[n]$ can be thought of as spatial samples captured by M microphones. Microphone samples usually have different powers. In other words, matrices Σ_i, which can be thought of as the cross correlation matrix across microphone samples, are not necessarily Toeplitz. Therefore $\mathbf{\Phi}_{\mathbf{s}}$ will not be spatially diagonalizable by the spatial Fourier transform, not

even asymptotically, as the number of the microphones increases ($M \to \infty$) and the Karhunen-Loève transform (KLT) matrix is in need for optimal coding in (4.9).

Note that the distortion $D^*(\theta)$ is the MSE between the vector source $\mathbf{s}[n]$ and its quantized version, say $\tilde{\mathbf{s}}[n]$. We use this RDF for vector source quantization (the dashed box in Fig. 4.4a). In *Link 1*, we are interested in estimating S_2. Therefore, with appropriate translation of the RDF in (4.9), the following procedure is described based on Fig. 4.4a.

First, the observations $\mathbf{y}_1$ are spatially decorrelated as

$$\mathbf{z}_1 = \mathbf{A}\mathbf{y}_1,\ \mathbf{A} \in \mathbb{C}^{M_1 \times M_1} \tag{4.10}$$

where $\mathbf{A}$ is a matrix whose rows are the eigenvectors of the CPSD matrix $\mathbf{\Phi}_{\mathbf{y}_1}$. The CPSD matrix $\mathbf{\Phi}_{\mathbf{z}_1}$ of the decorrelated vector $\mathbf{z}_1$ is diagonal with diagonal elements $[\mathbf{\Phi}_{\mathbf{z}_1}]_{uu} = \lambda_u(\mathbf{\Phi}_{\mathbf{y}_1})$, where $\lambda(\cdot)$ is the eigenvalue operator. Second, $\mathbf{z}_1$ is quantized, achieving the RDF presented in (4.9) by replacing $\mathbf{\Phi}_{\mathbf{s}}$ in (4.9) with $\mathbf{\Phi}_{\mathbf{y}_1}$. It can be shown that the following quantization model is obtained

$$\tilde{\mathbf{z}}_1 = \mathbf{B}\mathbf{z}_1 + \mathbf{e}_1, \tag{4.11}$$

where $\tilde{\mathbf{z}}_1$ is interpreted as a transformed-quantized left-side noisy observation. As the elements of $\mathbf{z}_1$ are uncorrelated, the vector quantizer in (4.9) can be interpreted as M_1 test channels corresponding to Fig. 4.1. The Vector $\mathbf{e}_1$ can be thought of as M_1 test channel noises. Therefore matrices $\mathbf{B}, \mathbf{\Phi}_{\mathbf{e}_1} \in \mathbb{R}^{M_1 \times M_1}$ will be diagonal and the diagonal elements are computed based on (4.3), replacing $\Phi_s(\Omega)$ by $[\mathbf{\Phi}_{\mathbf{z}_1}]_{uu}(\Omega)$, respectively. Applying the inverse-decorrelation matrix $\mathbf{A}^{-1}$ to reproduce quantized left-side observations, we obtain

$$\tilde{\mathbf{y}}_1 = \mathbf{A}^{-1}(\mathbf{B}\mathbf{z}_1 + \mathbf{e}_1) = \mathbf{C}\mathbf{y}_1 + \mathbf{A}^{-1}\mathbf{e}_1, \tag{4.12}$$

where $\mathbf{C} = \mathbf{A}^{-1}\mathbf{B}\mathbf{A}$. Finally, the estimator f_2 estimates the target signal S_2 as $\hat{S}_2 = f_2(\mathbf{y}_2, \tilde{\mathbf{y}}_1) = E[S_2|\mathbf{y}_2, \tilde{\mathbf{y}}_1]$. The direct distortion per frequency between $\mathbf{y}_1$ and its quantized version $\tilde{\mathbf{y}}_1$ is given by $\mathrm{tr}\{\mathbf{\Phi}_{\mathbf{d}_1}(\Omega)\}$, where $\mathbf{\Phi}_{\mathbf{d}_1}$ is the CPSD matrix of the direct error process $\mathbf{d}_1 = \mathbf{y}_1 - \tilde{\mathbf{y}}_1$, and $\mathrm{tr}\{\cdot\}$ denotes the trace operator on a matrix. We have

$$\mathrm{tr}\{\mathbf{\Phi}_{\mathbf{d}_1}(\Omega)\} = \sum_{u=1}^{M} \min(\theta, [\mathbf{\Phi}_{\mathbf{z}_1}]_{uu}(\Omega)). \tag{4.13}$$

As $[\mathbf{\Phi}_{\mathbf{z}_1}(\Omega)]_{uu} = \lambda_u(\mathbf{\Phi}_{\mathbf{y}_1}(\Omega))$, the overall MSE over all frequencies corresponds to the distortion in (4.9) with $\lambda_u(\mathbf{\Phi}_{\mathbf{s}})$ replaced by $\lambda_u(\mathbf{\Phi}_{\mathbf{y}_1}) = [\mathbf{\Phi}_{\mathbf{z}_1}]_{uu}$. The final (remote) distortion corresponds to (4.5) with $\tilde{\mathbf{y}} = [\mathbf{y}_2^T \tilde{\mathbf{y}}_1^T]^T$.

(JOINT) STATISTICS ESTIMATION

Based on the quantized left-side signal model in (4.12), the second order statistics in the frequency domain can be written as

$$\begin{aligned} \mathbf{\Phi}_{\tilde{\mathbf{y}}_1} &= \mathbf{C}\mathbf{\Phi}_{\mathbf{y}_1}\mathbf{C}^H + \mathbf{A}^{-1}\mathbf{\Phi}_{\mathbf{e}_1}\mathbf{A}^{-H} \\ \mathbf{\Phi}_{\mathbf{y}_1} &= \mathbf{C}^{-1}(\mathbf{\Phi}_{\tilde{\mathbf{y}}_1} - \mathbf{A}^{-1}\mathbf{\Phi}_{\mathbf{e}_1}\mathbf{A}^{-H})\mathbf{C}^{-H}. \end{aligned} \tag{4.14}$$

To retrieve $\mathbf{\Phi}_{\mathbf{y}_1}$ we need to know $\mathbf{B}, \mathbf{A}$, and $\mathbf{\Phi}_{\mathbf{e}_1}$. The elements of the diagonal matrix $\mathbf{B}$

depend on the reverse water-filling parameter θ and the diagonal elements of $\mathbf{\Phi}_{\mathbf{z}_1}$. The scalar value θ is chosen by fixing the transmission bit-rate or the distortion. Moreover, using the backward test channel interpretation in [13, 23], for $\theta < [\mathbf{\Phi}_{\mathbf{z}_1}]_{uu}(\Omega)$ we have

$$[\mathbf{\Phi}_{\mathbf{z}_1}]_{uu}(\Omega) = [\mathbf{\Phi}_{\tilde{\mathbf{z}}_1}]_{uu}(\Omega) + \theta. \tag{4.15}$$

Equation (4.15) shows that at high rates (small θ) it is possible to retrieve the unquantized statistics of the transformed signal $\mathbf{z}_1$ from the quantized signal $\tilde{\mathbf{z}}_1$. Using the property in (4.15), $\mathbf{\Phi}_{\mathbf{z}_1}$ can be retrieved, estimating the quantized PSDs $[\mathbf{\Phi}_{\tilde{\mathbf{z}}_1}]_{uu}$ given the quantized realizations in the frequency domain, for a sufficiently small θ. Therefore, the matrix $\mathbf{B}$ can be computed at the decoder. Following a similar procedure, the diagonal matrix $\mathbf{\Phi}_{\mathbf{e}_1} = E[\mathbf{e}_1\mathbf{e}_1^H]$ can be computed. Computing $\mathbf{A}$ requires the data dependent KLT. As we often do not know this, we test in Sec. 4.4 the algorithm, next to the true $\mathbf{A}$ based on the KLT, also with an $\mathbf{A}$ based on the fixed discrete cosine transform (DCT). Finally, computing $\mathbf{B}$ and fixing $\mathbf{A}$, the matrix $\mathbf{C}$ is known and the local statistics $\mathbf{\Phi}_{\mathbf{y}_1} = E[\mathbf{y}_1\mathbf{y}_1^H]$ are retrieved at the decoder. Moreover, the joint statistics between the two side observations are retrieved as

$$E[\mathbf{y}_1\mathbf{y}_2^H] = \mathbf{C}^{-1}E[\tilde{\mathbf{y}}_1\mathbf{y}_2^H]. \tag{4.16}$$

Note that the statistics (joint or local) at a certain frequency Ω can be retrieved only if the matrix $\mathbf{C}(\Omega)$ is invertible and the PSD of the source at that frequency is positive. This implies that $\mathbf{B}(\Omega)$ should be invertible, as $\mathbf{A}(\Omega)$ is an orthogonal and invertible matrix. Since $\mathbf{B}(\Omega)$ is a diagonal matrix, all elements should be positive, implying that $\theta < \min_u [\mathbf{\Phi}_{\mathbf{z}_1}]_{uu}(\Omega)$, for a particular frequency. For $\mathbf{B}$ to be invertible in all frequencies, the condition is rewritten as $\theta < \min_\Omega \min_u [\mathbf{\Phi}_{\mathbf{z}_1}]_{uu}(\Omega)$. This condition is satisfied only at sufficiently high rates. In fact, at lower rates, the reverse-water filling algorithm tries to allocate more bit-rate to the frequency components with greater PSD values and zero bit rate to those with smaller (than the threshold θ) PSD values. In this case $\mathbf{B}$ becomes singular and, as a result, smaller PSDs cannot be retrieved at the decoder. In the next part of this section, another quantization method is proposed to address this limitation and guarantee the invertible $\mathbf{B}$ matrix for all frequencies.

Method 2: PDF PRESERVING SOURCE CODING FOR SOURCES WITH MEMORY

Reverse-water filling for vector sources (*Method 1*) cannot guarantee positive bit-rates for strictly positive PSDs. To keep all frequency components of the signal active after the quantization, a constrained source coding approach was proposed in [28]. This method imposes an extra constraint to the original lossy source coding problem such that the probability distribution of the signal is preserved after the quantization process. The distribution preserving RDF (DP-RDF) is given in [22, Proposition 1] for a time-stationary Gaussian process, which can be thought of as a single microphone observation. As there are multiple microphones per HA, we extend this result to multiple observations and find the DP-RDF for vector sources with memory. Moreover, we propose a conceptual test channel interpretation to achieve such a rate-distortion tradeoff.

Proposition 1: The DP-RDF for a discrete-time sequence of zero-mean time-stationary vector Gaussian sources, say $\{\mathbf{s}[n]\}_{n=0}^{N-1}$, where $\mathbf{s}[\cdot] \in \mathbb{R}^{M\times 1}$ with the corresponding block-

Toeplitz cross correlation matrix $\Sigma_{\mathbf{s}}$, and the Hermitian CPSD matrix $\mathbf{\Phi}_{\mathbf{s}}$ is given by

$$\begin{aligned} &R^{DP}(\mu) = \\ &\frac{1}{2\pi}\int_{-\pi}^{\pi}\sum_{u=1}^{M}\log_2\frac{\lambda_u(\mathbf{\Phi}_{\mathbf{s}}(\Omega))}{(\lambda_u(\mathbf{\Phi}_{\mathbf{s}}(\Omega))D_u(\mu,\Omega)-\frac{D_u^2(\mu,\Omega)}{4})^{\frac{1}{2}}}d\Omega \\ &D^{DP}(\mu) = \frac{1}{2\pi}\int_{-\pi}^{\pi}\sum_{u=1}^{M}D_u(\mu,\Omega)d\Omega, \end{aligned} \tag{4.17}$$

where

$$D_u(\mu,\Omega) = 2\,\lambda_u(\mathbf{\Phi}_{\mathbf{s}}(\Omega)) + \mu - (4\,\lambda_u^2(\mathbf{\Phi}_{\mathbf{s}}(\Omega)) + \mu^2)^{\frac{1}{2}}, \tag{4.18}$$

for $D_u(\mu,\Omega) < 2\lambda_u(\mathbf{\Phi}_{\mathbf{s}}(\Omega))$. The variable μ denotes a Lagrange parameter [22], which relates the rate to the distortion and satisfies the constraint on the distortion. Similar to "reverse water-filling" problems [13] with parameter θ, μ can be found by either fixing the total rate R^{DP} or the total distortion D^{DP}. The rate is per vector source $\mathbf{s}[\cdot] \in \mathbb{R}^{M\times 1}$. The distortion $D^{DP}(\mu)$ is the averaged MSE between the vector source $\mathbf{s}[\cdot]$ and the quantized source $\hat{\mathbf{s}}[\cdot]$. See App. 4-A for derivations.

Equation (4.17) represents the proposed extension to the DP-RDF for vector sources. Note that (4.17) is only valid for strictly positive CPSD eigenvalues ($\lambda_u(\mathbf{\Phi}(\Omega)) > 0$). For $\lambda_u(\mathbf{\Phi}(\Omega)) = 0$ the rate allocated to such frequency component will be zero. Unlike *Method 1* (reverse water filling), here all frequency components with strictly positive CPSD eigenvalues are allocated with positive rates. Therefore, the PSDs can be retrieved from the quantized signal vector at the decoder. Comparing the direct distortions in (4.17) and (4.9), the gap in distortions is 3 dB at zero bit-rate and it vanishes asymptotically [28]. However, as we are interested in estimating the source S_2, it is not clear if *Method 1* is better than *Method 2* with respect to the final distortion, as the joint statistics are not available at the encoder in both methods.

App. 4-B shows that the conceptual test channel interpretation, shown in the dashed box in Fig. 4.4a, achieves the DP-RDF in (4.17), but with different quantization parameters from those in *Method 1*. First $\mathbf{y}_1$, is spatially decorrelated ($\mathbf{z}_1 = \mathbf{A}\mathbf{y}_1$). Then the decorrelated signals are quantized using the proposed distribution preserving quantization, $\tilde{\mathbf{z}}_1 = \mathbf{B}\mathbf{z}_1 + \mathbf{e}_1$ with quantization parameters given as (see App. 4-B):

$$\begin{aligned} &\mathbf{B}(\Omega) = \mathrm{diag}\{\beta_1(\Omega),...,\beta_{M_1}(\Omega)\} = \\ &\mathrm{diag}\{1-\frac{D_1(\mu,\Omega)}{2\,\lambda_1(\mathbf{\Phi}_{\mathbf{y}_1}(\Omega))},...,1-\frac{D_M(\mu,\Omega)}{2\,\lambda_M(\mathbf{\Phi}_{\mathbf{y}_{M_1}}(\Omega))}\} \\ &\mathbf{\Phi}_{\mathbf{e}_1}(\Omega) = \mathrm{diag}\{\frac{\beta_1(\Omega)+1}{2}D_1(\mu,\Omega),...,\frac{\beta_M(\Omega)+1}{2}D_M(\mu,\Omega)\}. \end{aligned} \tag{4.19}$$

The quantized signal $\tilde{\mathbf{y}}_1$ is computed by applying the inverse transform matrix $\mathbf{A}^{-1}$ as $\tilde{\mathbf{y}}_1 = \mathbf{A}^{-1}\tilde{\mathbf{z}}_1$. Finally the target signal is estimated as $\hat{S}_2 = E[S_2|\tilde{\mathbf{y}}_1, \mathbf{y}_2]$.

The procedure to retrieve (joint) statistics is as follows. As the second order statistics of $\mathbf{y}_1$ are preserved, $\mathbf{\Phi}_{\mathbf{y}_1} = \mathbf{\Phi}_{\tilde{\mathbf{y}}_1}$ holds. $\mathbf{\Phi}_{\tilde{\mathbf{y}}_1}$ can be estimated using realizations in the frequency domain. By informing the decoder of the scalar parameter μ, the invertible matrices $\mathbf{B}$ and $\mathbf{\Phi}_{\mathbf{e}_1}$ in (4.19) are known at the decoder. Knowing $\mathbf{B}$ and fixing $\mathbf{A}$, $\mathbf{C}$ is known. Therefore, based on (4.16) the joint statistics are retrieved.

4.3.2. *Link 2*: FROM RIGHT-TO-LEFT

The goal in *Link 2* is to transmit the filtered-quantized right-side observations in order to estimate the target signal at the left-side reference microphone S_1. As the lossy (quantized) version of the left-side observation ($\tilde{\mathbf{y}}_1$) is available at the right-side, it acts as a (lossy) side information at the right-side encoder. We use this information to reduce the redundancy in the transmission of the information. This is done by the proposed coding architecture, illustrated in Fig. 4.4b. As shown, different coding algorithms can be obtained by changing the switch l and by using different methods in *Link 1*. Note that if for example, the switch l is open, i.e., $l = 0$, the realization of $\mathbf{y}_1$ will not be used at the decoder. The side information at the left-side decoder is $\mathbf{y}_1$. We describe some possible scenarios.

CASE A: CODING WITH QUANTIZED STATISTICS AND WITH $l = 0$

In this case we assume the *Method 1* is chosen in *Link 1*. The idea is to pre-filter the right-side observations $\mathbf{y}_2$ using quantized statistics retrieved from *Link 1* and directly quantize and transmit them to the other side. The sub-optimal filter coefficients $\mathbf{w}_2^a$ (compared to the optimal filter in (4.4)) are computed as $\mathbf{w}_2^a = \mathbf{\Phi}_{\tilde{\mathbf{y}}_2^a}^{-1}\mathbf{\Phi}_{\tilde{\mathbf{y}}_2^a S_1}$, where $\mathbf{\Phi}_{\tilde{\mathbf{y}}_2^a}$ is the CPSD matrix of the innovation process $\tilde{\mathbf{y}}_2^a = \mathbf{y}_2 - E[\mathbf{y}_2|\tilde{\mathbf{y}}_1]$. The filter coefficients are computed in a similar fashion to the one in optimal RCNR approach, described in Sec. 4.2.1, except that here only the lossy side information $\tilde{\mathbf{y}}_1$ is available and not the lossless $\mathbf{y}_1$. In this way, we try to reduce some information redundancy in estimating S_1 of $\mathbf{y}_2$ given $\tilde{\mathbf{y}}_1$.

The filtered scalar signal $Y_{21}^a = (\mathbf{w}_2^a)^{\mathrm{H}}\mathbf{y}_2$ is encoded for a decoder that has no access to the side information $\mathbf{y}_1$ (the switch l is open). This means that Y_{21}^a is directly (blindly) quantized, i.e., $\tilde{Y}_{21}^a = \beta_2^a Y_{21}^a + E_2^a$. The quantization parameters correspond to (4.3) with replacing Φ_s with $\Phi_{Y_{21}^a} = (\mathbf{w}_2^a)^{\mathrm{H}}\mathbf{\Phi}_{\mathbf{y}_2}\mathbf{w}_2^a$. In fact, $\tilde{\mathbf{y}}_1$ is used for estimating the filter coefficients $\mathbf{w}_2^a$, but not used in the coding process. Finally the MWF filter is applied to the total observations $\tilde{\mathbf{y}} = [\mathbf{y}_1^T\ \tilde{Y}_{21}^a]^T$ and the target signal S_1 is estimated as $\hat{S}_1 = f_1(\tilde{\mathbf{y}}) = E[S_1|\tilde{\mathbf{y}}]$.

CASE B: CODING WITH UNQUANTIZED STATISTICS AND WITH $l = 0$

In the previous case, we estimated the (joint) statistics based on the lossy side information $\tilde{\mathbf{y}}_1$ (*Method 1*). Therefore, the filtering coefficients are not estimated in an optimal manner as some frequency components are truncated, especially at lower rates. To estimate the optimal filter we do not need the actual realizations of the lossless side information $\mathbf{y}_1$. Instead, we only need to estimate the (joint) unquantized statistics from $\tilde{\mathbf{y}}_1$. To do so, *Method 2* is chosen in *Link 1*. The use of DP quantization enables retrieving the unquantized statistics $E[\mathbf{y}_1\mathbf{y}_2^H]$ and $E[\mathbf{y}_1\mathbf{y}_1^H]$ at all frequencies. The procedure to estimate the statistics is described in the Sec. 4.3.1. The optimal right-side filter coefficients $\mathbf{w}_2^o$ are computed, similar to (4.4), as $\mathbf{w}_2^o = \mathbf{\Phi}_{\tilde{\mathbf{y}}_2}^{-1}\mathbf{\Phi}_{\tilde{\mathbf{y}}_2 S_1}$, where $\mathbf{\Phi}_{\tilde{\mathbf{y}}_2} = \mathbf{\Phi}_{\mathbf{y}_2} - \mathbf{\Phi}_{\mathbf{y}_2\mathbf{y}_1}\mathbf{\Phi}_{\mathbf{y}_1}^{-1}\mathbf{\Phi}_{\mathbf{y}_1\mathbf{y}_2}$ is the CPSD matrix of the innovation process $\tilde{\mathbf{y}}_2 = \mathbf{y}_2 - E[\mathbf{y}_2|\mathbf{y}_1]$. The direct quantization and final estimation stages resembles those of case a with a different filtered signal $Y_{21} = (\mathbf{w}_2^o)^{\mathrm{H}}\mathbf{y}_2$.

CASE C: OPTIMAL CODING WITH UNQUANTIZED STATISTICS AND WITH $l = 1$

Like the Case b, in this case, again we use *Method 2* in *Link 1* since we want to preserve the statistics to compute optimal filter for *Link 2*. Following the optimal RCNR (when the

switch l is closed) the right-side processor encodes the filtered signal $Y_{21} = (\mathbf{w}_2^o)^{\mathrm{H}}\mathbf{y}_2$ for a decoder which has access to the side information $\mathbf{y}_1$. Therefore unlike case a and case b, here the quantization stage is a side information informed process (remote Wyner-Ziv quantizer). The filtered-quantized signal is given by $\tilde{Y}_{21} = \beta_2^o Y_{21} + E_2^o$. The optimal (remote Wyner-Ziv) quantization parameters β_2^o and $\Phi_{E_2^o}$ correspond to (4.3) with Φ_s replaced by $(\mathbf{w}_2^o)^{\mathrm{H}}\mathbf{\Phi}_{\tilde{\mathbf{y}}_2}\mathbf{w}_2^o$. It is important to note that here the quantization scaling factor β_2^o is now a function of the side information $\mathbf{y}_1$ (only a function of the statistics, not the realizations). This necessitates some extra information to be available at the decoder in order to decode indices which are computed knowing the fact that $\mathbf{y}_1$ would be available at the decoder. This extra information includes the joint entropy between $\mathbf{y}_1$ and $\tilde{Y}_{21}$. Knowing this information at the decoder, we can touch the performance bound of the optimal RCNR at least in one link (*Link 2*) in our proposed 2-way communication system. The final conditional mean estimator, which was included in the decoder box in the optimal RCNR architecture in Fig. 4.2, resembles that of the Case a, except with the different filtered-quantized signal $\tilde{Y}_{21}$.

4.4. PERFORMANCE EVALUATION

In this section, we compare the performance of the approaches, described in the previous sections, as a function of transmission bit-rate. We evaluate the methods based on two performance measures. The first performance measure presented in [12] and [18], is defined as the ratio of the MSE when there is no communication between the HAs to the one when the data is quantized before transmission. The output gains with respect to the two beamformers are given by

$$G_1(R_1) = \frac{D_1(0)}{D_1(R_1)}, \quad G_2(R_2) = \frac{D_2(0)}{D_2(R_2)}, \tag{4.20}$$

where $D_i(\cdot)$, $i = 1,2$ are defined in (4.5) but with different outputs $\hat{S}_i$ for different approaches. Note that the final outputs $\hat{S}_i$, $i = 1,2$ are functions of the corresponding bit-rates as the data is quantized. For example, $D_1(R_1)$ denotes the MSE between the target source at the left-side reference microphone S_1 and its estimate $\hat{S}_1$ when the data is quantized and transmitted at R_1 bit-rate from the right-side HA to the left one (*Link 2*). $D_1(0)$ denotes the left-side MSE when there is no communication between HAs. i.e., $R_1 = 0$.

Another performance measure which we refer to as "binaural gain" is proposed in [12] and is defined as the ratio of the sum of the MSEs with respect to the two HA reference microphones, when there is no communication between HAs to the one when the data is quantized and transmitted in both links at certain bit-rates, i.e.,

$$G_B(R_T) = \frac{D_1(0) + D_2(0)}{D_1(R_1) + D_2(R_2)}, \tag{4.21}$$

where $R_T = R_1 + R_2$ is the total rate budget for two links. The performance of the following approaches are compared throughout this section

- **B-MWF**: The full binaural MWF from [7], without quantization.

- **OPT**: Optimal approach from [12, Sec. III-A] (Sec. 4.2.1)
- **SIG**: Sub-optimal approach from [12]: An estimate of the target signal is transmitted to the contralateral processor (Sec. 4.2.2)
- **INT**: Sub-optimal approach from [18]: An estimate of the undesired (interfering) signal is transmitted to the contralateral processor (Sec. 4.2.2)
- **RAW**: Sub-optimal approach from [18]: A raw reference microphone signal is transmitted to the contralateral processor, without any pre-filtering (Sec. 4.2.2)
- **M1**: *Method 1* in *Link 1* (Sec. 4.3.1)
- **M2**: *Method 2* in *Link 1* (Sec. 4.3.1)
- **L2a**: Proposed sequential sub-optimal approach for *Link 2* using *Method 1* in *Link 1* (Case a, Sec. 4.3.2)
- **L2b**: Proposed sequential sub-optimal approach for *Link 2* using *Method 2* in *Link 1* (Case b, Sec. 4.3.2)
- **L2c**: Proposed sequential optimal approach for *Link 2* using *Method 2* in *Link 1* (Case c, Sec. 4.3.2)

The acoustic scene used for the experiments is illustrated in Fig. 4.5.

The four black "+" symbols indicate the microphones mounted on the virtual head. The planar distance between the two microphones per HA is 0.76 cm. The radius of the virtual head is set to 8.2 cm [29]. The Green circle denotes the desired (target) speech signal which is assumed to be fixed 0.8 m from the origin in front of the head for all experiments. The black triangles denote the interferers. The number and the position of the interferers vary in different experiments. Interferers are located randomly at different angles, say $\alpha = \tan^{-1}(\frac{y}{x}) - \frac{\pi}{2}$ and different distances from the origin ($(x,y) = (0,0)$), say $r = \sqrt{x^2+y^2}$. In this chapter zero degrees corresponds to the direction straight ahead of the HA user and the angles are computed counterclockwise. All point noise sources have flat PSDs $\Phi_{(.)}(\Omega)$ over the interval $\Omega \in [-\pi F_s, \pi F_s]$ where $F_s = 16$ kHz. ATFs in (4.1) are found via head-related transfer functions (HRTF)s from the database in [29]. The PSD of the target speech signal is estimated based on the Welch's method using 12.5 seconds of the recorded speech at 16 kHz sampling frequency from the "CMU-ARCTIC" [30] database, without considering voice activity detection (VAD) errors. We used 512-point frames with 50% overlap and discrete Fourier transform (DFT) size of 1024 for the PSD estimation process.

4.4.1. UNCORRELATED NOISE

In this scenario, per microphone, the target signal is degraded by additive white Gaussian noise (AWGN), uncorrelated to the signal as well as across microphones, having the same variance among all microphones. Note that there is no point noise source (interferer) here. The uncorrelated noise power is set such that the input signal to noise ratio (SNR) at the corresponding left-side and right-side reference microphones, say SNR_1

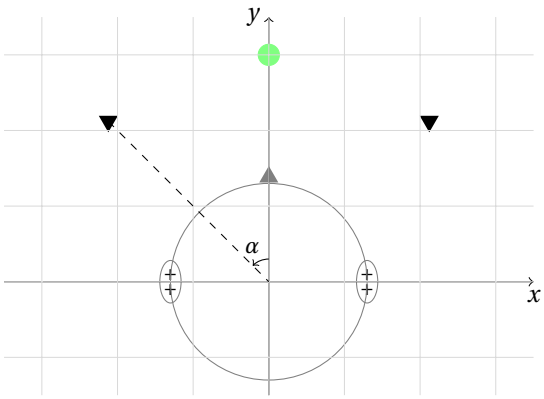
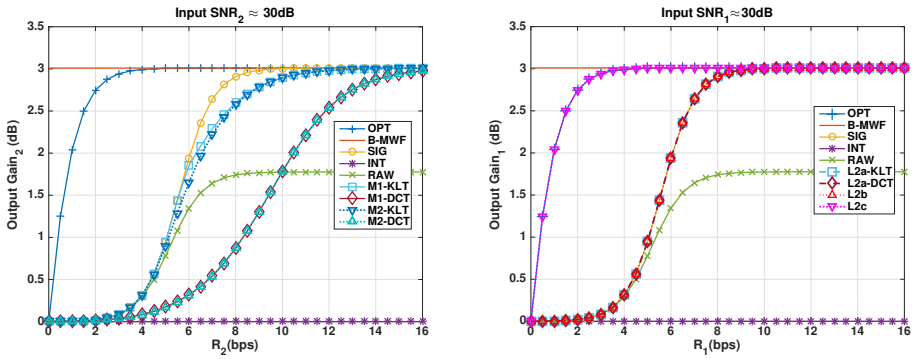


Figure 4.5: Typical acoustic scene. The Target signal, interferers, and microphones are denoted by the Green circle, the black triangles, and the black "+" symbols, respectively.

and SNR_2, respectively, be approximately 30 dB. Based on the performance measures in (4.20), the output gains (in dB) in terms of bit per sample (bps) are shown in Fig. 4.6 for *Link 1* and *Link 2*, for the above-mentioned approaches. By M1-DCT we mean that



(a) *Link 1*: from left to right. The gain is com-(b) *Link 2*: from right to left. The gain is computed with respect to the right-side beamformer.puted with respect to the left-side beamformer.

Figure 4.6: Output Gains in the presence of uncorrelated noise only.

the method M1 uses the fixed discrete cosine transform (DCT) matrix for the matrix $\mathbf{A}$ in (4.10), when spatially decorrelating the signals in the frequency domain. The similar explanation holds for M1-KLT, M2-KLT, and M2-DCT. Note that the method L2a-KLT in *Link 2* is sequentially related to the *Method 1* (M1-KLT) in *Link 1* as it uses the signal statistics quantized by *Method 1*. The similar relation holds between L2a-DCT and M1-DCT. L2b and L2c are related to the *Method 2* in *Link 1*. Their performances remain the same using KLT or DCT matrices in *Method 2*, as they use the retrieved (unquantized) statistics. These explanations hold also for other experiments in this section.

Based on Fig. 4.6 we can make the following observations

- Method SIG is asymptotically optimal as argued in [18] in the presence of uncorrelated noise only. As the noise components at the two sides are independent, no necessary information will be removed by estimating the desired part of the signal and sending it to the other side.

- Method INT has no gain compared the monaural setup. With a similar argument, estimating the noise on one side has no added information for the other side, resulting in no increase in the performance.

- M1 and M2 outperform methods RAW and INT since they are not even asymptotically optimal. In the presence of only uncorrelated noise, any extra observation (extra microphone signal) can help to increase the performance. Method RAW in [18] chooses only one microphone signal (out of two), which degrades the performance.

- L2a and L2b have almost the same performance as the SIG method since almost no redundancy in information is remained after locally estimating the target signal itself.

- L2c is an optimal coding scheme in *Link 2*. It takes the correlation between the filtered observation and the side information into account and encodes the filtered signals knowing the fact that the decoder can revive the correlated information which is reduced during the encoding process. This approach assumes the joint entropy between the filtered-encoded signals and the side information is available at the decoder.

4.4.2. CORRELATED AND UNCORRELATED NOISE

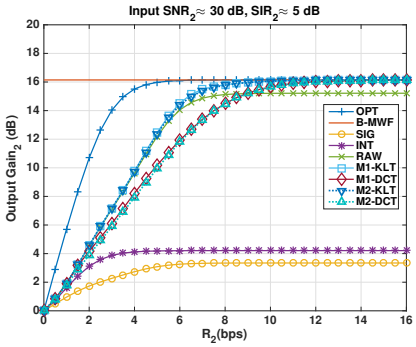
Two scenarios are considered in this section. First, one point noise source is added to the previous scenario at 30° and 0.8 m from the origin. The interfering signal power is set such that the input signal-to-interferer ratios (SIRs) with respect to the left and right reference microphones are approximately $\mathrm{SIR}_1 \approx 0$ dB, and $\mathrm{SIR}_2 \approx 5$ dB, respectively.

In the second scenario, four interferers are added at degrees $[-50°, -30°, 30°, 70°]$. The input SIRs at the corresponding reference microphones are $\mathrm{SIR}_1 \approx 0$ dB and $\mathrm{SIR}_2 \approx 0$ dB. The simulation results are shown in Fig. 4.7.

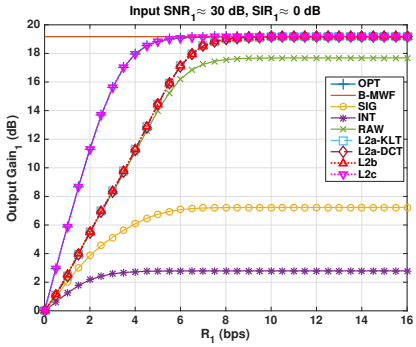
Based on Fig. 4.7 we can make the following observations

- As shown in Fig.s 4.7a and 4.7b, in the presence of one spatially correlated point noise source, the SIG method is not asymptotically optimal anymore. Some necessary (spatial) information about the interferer will be eliminated after the filtering stage before transmission. This information would be helpful for the left-side processor to cancel out the interferer [18]. In general, the loss in performance at high rates is significant.

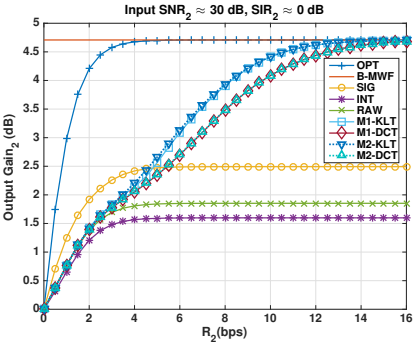
- In *Link 1* for highly correlated signals (one interferer), the methods M1-KLT and M2-KLT outperform all other approaches (when using the optimal KLT matrix). Using the DCT matrix also results in a good performance, especially at high rates.



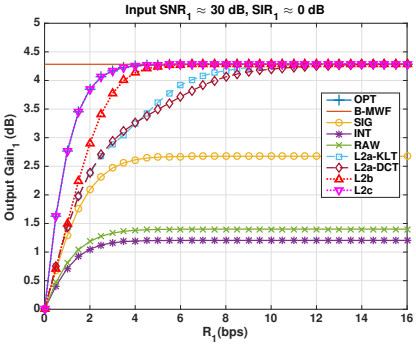
(a) One interferer: *Link 1*, from left to right.



(b) One interferer: *Link 2*, from right to left.



(c) Four interferers: *Link 1*, from left to right.



(d) Four interferers: *Link 2*, from right to left.

Figure 4.7: Output gains for observations with correlated noise sources.

- L2a uses the quantized left-side signals from *Link 1* (*Method 1*) to reduce the redundancy in information transmission, and hence, outperforms almost all existing sub-optimal approaches, especially at high rates. However, the use of quantized statistics results in a non-optimal filter in *Link 2* and degrades the performance, especially in the four interferers scenario, in comparison with the L2b and the L2c methods.

- L2b outperforms L2a, especially in the four interferers scenario. The use of *Method 2* in *Link 1* enables retrieval of the unquantized statistics (rather than quantized statistics), which helps to compute the optimal filter on the right side, and hence, results in a better estimation of the informative signal for the left-side beamformer. However, the filtered signal in both the L2a and the L2b methods is still correlated to the (left) side information $\mathbf{y}_1$ and the direct (blind) quantization of such a filtered signal does not take this correlation into account. As a result, the performance of both cases a and b are always worse than that of the case c. L2c touches the optimal performance as the side-information-aware quantization is used.

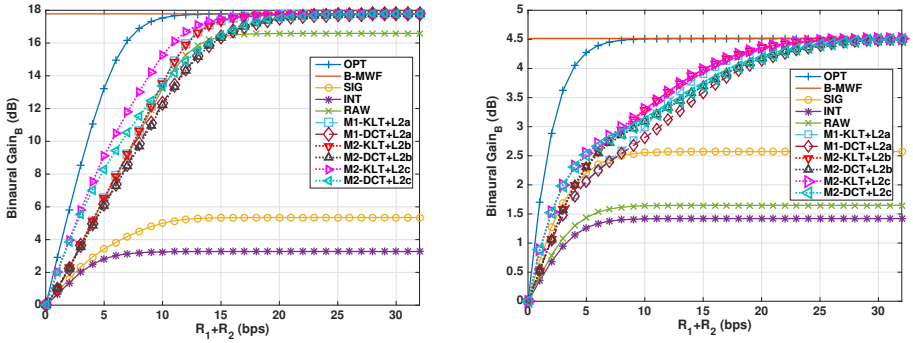
- The gap between M1 and the distribution preserving vector quantization method M2 is negligible in most scenarios and bit-rates. As the target estimation error is of interest (and not the direct estimation error between the noisy vector source $\mathbf{y}_i, i = 1,2$ and its corresponding quantized version $\tilde{\mathbf{y}}_i$), it is not clear if the M2 performance should always be worse than that of M1. In fact, as the blind quantization process in *Link 1* does not use the joint information, both quantizers are not aware that which frequency components are more important and not predictable on the other side. However, M2 preserves the statistics of the contralateral observations which include spatial cues of sources and perceptually may help to get a more natural impression of the sound field.

- For all MWF-based methods, the performance is a function of the correlation between the observations. Therefore, it is clear that the performance will change by changing the source position, as the correlation between the observations will change. However, this does not affect the generality of the proposed method. as well as that of the optimal method. The proposed method tries to estimate the joint statistics, without any assumption on the source positions, and use it in another link to reduce the redundancy in information transmission.

4.4.3. BINAURAL GAIN

In this section, we evaluate the methods based on the binaural gain measure in (4.21). We compute $G_B(R_T)$ for scenarios with correlated interferers, introduced in Sec. 4.4.2. In fact, the same distortions as those computed for the results in Fig. 4.7 are used for computing $G_B(R_T)$. Fig. 4.8 shows the binaural gain in terms of the total bit-rate budget $R_T = R_1 + R_2$, for the scenario in which there is one interferer (left: Fig. 4.8a), and the scenario in which there are four interferers (right: Fig. 4.8b), along with the target signal and the uncorrelated noise. For example, when computing the binaural gain of "M2-KLT+L2c", the MSE of M2-KLT in *Link 1* is added to the MSE for the method L2c. The sequential-asymmetric method M2-KLT+L2c has the best performance among all

other methods as it touches the optimal performance, at least in *Link 2*. However considering the loss in the performance of M2-KLT in *Link 1*, the binaural performance of M2-KLT+L2c is worse than that of the optimal performance [12, Sec. III-A] (which is optimal in both links). All proposed methods resolve the asymptotic sub-optimality issue of the existing sub-optimal methods and outperform them, especially at middle rates and high rates. The sequential nature of the proposed methods enables a smart use of the information at hand to reduce the bandwidth.



(a) One interferer + uncorrelated Noise.

(b) Four Interferers + Uncorrelated Noise.

Figure 4.8: Binaural output gains for observations with correlated noise sources.

4.5. CONCLUSION

In this chapter, we studied the performance of the optimal/sub-optimal binaural rate-constrained noise reduction (beamforming) approaches based on the unified framework which can be interpreted as filtering, quantization, and final estimation stages. Moreover, we proposed a two-way asymmetric coding scheme which retrieves the statistics between two HA observations from quantized signals in one link (*Link 1*) to be used in another link (*Link 2*) and addresses two main limitations of existing methods. The first limitation is the strict requirement of the complete knowledge of the joint statistics in the optimal approach. The second limitation is the asymptotic sub-optimality of the existing sub-optimal approaches. Based on two performance measures, the proposed results outperform those of sub-optimal approaches. Moreover, the results confirm the asymptotic optimality of the proposed method.

APPENDICES

4-A:DP-RDF FOR VECTOR SOURCES WITH MEMORY

We show the derivations that result in the proposed DP-RDF in (4.17) for vector sources with memory (multi-sensor observations) based on DP-RDF for a discrete-time independent scalar (one sensor) observation samples.

We are given a sequence of discrete-time zero-mean stationary vector Gaussian sources, denoted by $\{\mathbf{s}[n]\}_{n=0}^{N-1}$, where $\mathbf{s}[n] \in \mathbb{R}^{M\times 1}$, $n = 0,\dots,N-1$, with the corresponding block-Toeplitz cross correlation matrix $\Sigma_{\mathbf{s}} \in \mathbb{R}^{MN\times MN}$, defined in (4.6), and the Hermitian CPSD matrix $\mathbf{\Phi_s} \in \mathbb{C}^{M\times M}$. The sequence $\{\mathbf{s}[n]\}_{n=0}^{N-1}$ is stacked into the vector $\mathbf{s}_{\text{vec}} = [\mathbf{s}^T[0]...\mathbf{s}^T[N-1]]^T$. We define the following DP optimization problem based on the DP-RDF defined in [28]

$$\begin{aligned} \inf_{f(\tilde{\mathbf{s}}_{\text{vec}}|\mathbf{s}_{\text{vec}})} \quad & I(\mathbf{s}_{\text{vec}};\tilde{\mathbf{s}}_{\text{vec}}) \\ \text{subject to} \quad & E[|\mathbf{s}_{\text{vec}}-\tilde{\mathbf{s}}_{\text{vec}}|^2] \le D^{DP}, \\ & f(\mathbf{s}_{\text{vec}}) = f(\tilde{\mathbf{s}}_{\text{vec}}), \end{aligned} \tag{4.22}$$

where $I(\mathbf{x};\mathbf{y})$ is generally the mutual information between the random vector variables $\mathbf{x}$ and $\mathbf{y}$. The conditional distribution function of a random vector variable $\mathbf{x}$, given a random vector variable $\mathbf{y}$ is denoted by $f(\mathbf{x}|\mathbf{y})$. The problem in (4.22) tries to find the minimum rate R^{DP} at which the vector $\mathbf{s}_{\text{vec}}$ can be quantized such that the probabilty distribution of the source, say $f(\mathbf{s}_{\text{vec}})$, is preserved after the quantization, i.e., $f(\mathbf{s}_{\text{vec}}) = f(\tilde{\mathbf{s}}_{\text{vec}})$, and the MSE between $\mathbf{s}_{\text{vec}}$ and its quantized output $\tilde{\mathbf{s}}_{\text{vec}}$ does not exceed a certain value D^{DP}. Mutual information is invariant under unitary transformations [13] and the objective mutual information function $I(\mathbf{s}_{\text{vec}};\tilde{\mathbf{s}}_{\text{vec}})$ can be rewritten as a summation of separable functions [13, 23]

$$I(\mathbf{s}_{\text{vec}};\tilde{\mathbf{s}}_{\text{vec}}) = I(\mathbf{s}_{\text{dec}};\tilde{\mathbf{s}}_{\text{dec}}) = \sum_{i=1}^{MN} I(s_{\text{dec}}[i];\tilde{s}_{\text{dec}}[i]), \tag{4.23}$$

where $s_{\text{dec}}[i]$ is the ith element of the transformed vector $\mathbf{s}_{\text{dec}} = \mathbf{V}^{\text{H}}\mathbf{s}_{\text{vec}}$. Matrix $\mathbf{V}$ is derived by eigenvalue decomposition of the correlation matrix, i.e., $\Sigma_{\mathbf{s}} = \mathbf{V\Lambda V}^{\text{H}}$, where $\mathbf{\Lambda} = \text{diag}\{\lambda_1(\Sigma_{\mathbf{s}}),...,\lambda_{MN}(\Sigma_{\mathbf{s}})\}$. The second equality in (4.23) holds as the elements of $\mathbf{s}_{\text{dec}}$ are statistically independent. Note that as $\mathbf{V^{-1}} = \mathbf{V^{H}}$, we have $\tilde{\mathbf{s}}_{\text{vec}} = \mathbf{V}\tilde{\mathbf{s}}_{\text{dec}}$, where $\tilde{\mathbf{s}}_{\text{dec}}$ denotes the transformed-quantized vector signal. Using the unitary transformation, the reformulated problem is given by

$$\begin{aligned} \inf_{f(\tilde{\mathbf{s}}_{\text{dec}}|\mathbf{s}_{\text{dec}})} \quad & \sum_{i=1}^{MN} I(s_{\text{dec}}[i];\tilde{s}_{\text{dec}[i]}) \\ \text{subject to} \quad & \sum_{i=1}^{MN} D_i \le D^{DP}, \\ & f(s_{\text{dec}}[i]) = f(\tilde{s}_{\text{dec}}[i]), \end{aligned} \tag{4.24}$$

where $D_i = E[|s_{\text{dec}}[i]-\tilde{s}_{\text{dec}}[i]|^2]$. Note that the unitary transformation preserves the MSE, i.e., $E[|\mathbf{s}_{\text{vec}}-\tilde{\mathbf{s}}_{\text{vec}}|^2] = E[|\mathbf{s}_{\text{dec}}-\tilde{\mathbf{s}}_{\text{dec}}|^2] = \sum_{i=1}^{MN} E[|s_{\text{dec}}[i]-\tilde{s}_{\text{dec}}[i]|^2]$.

As the elements of $\mathbf{s}_{\text{dec}}$ are statistically independent, based on Lemma 3 in [22], the problem in (4.24) for a decorrelated vector $\mathbf{s}_{\text{dec}}$, can be solved as

$$R_{MN}^{DP}(\mu) = \begin{cases} \sum_{i=1}^{MN} \log_2 \frac{E[|s_{\text{dec}}[i]|^2]}{(E[|s_{\text{dec}}[i]|^2]D_i(\mu) - \frac{D_i^2(\mu)}{4})^{\frac{1}{2}}} & D_{MN}^{DP}(\mu) < 2\,\sigma^2 \\ 0 & D_{MN}^{DP}(\mu) \geq 2\,\sigma^2 \end{cases}$$
$$D_{MN}^{DP}(\mu) = \sum_{i=1}^{MN} D_i(\mu), \tag{4.25}$$

where

$$D_i(\mu) = 2\,E[|s_{\text{dec}}[i]|^2] + \mu - (4E[|s_{\text{dec}}[i]|^2] + \mu^2)^{\frac{1}{2}}, \tag{4.26}$$

with $E[|s_{\text{dec}}[i]|^2] = \lambda_i(\Sigma_{\mathbf{s}})$, $\sigma^2 = \sum_{i=1}^{MN} E[|s_{\text{dec}}[i]|^2]$ and μ a Lagrange variable relating the rate to the distortion [22]. The equation (4.26) is valid for $D_i(\mu) \leq 2\,E[|s_{\text{dec}}[i]|^2]$. In (4.25) $R_{MN}^{DP}(\mu)$ is the minimum achievable rate at which the source $\mathbf{s}_{\text{vec}}$ can be encoded and decoded with distortion not exceeding a certain value D_{MN}^{DP} such that its PDF is preserved after quantization. Note that the rate is per vector source $\mathbf{s}_{\text{vec}}$. $D_i(\cdot)$ is the corresponding MSE with respect to the ith element of the decorrelated vector source $\mathbf{s}_{\text{dec}}$.

For a given μ, R_{MN}^{DP} can be represented as a sum of non-linear functions of the eigenvalues of the block-Toeplitz matrix $\Sigma_{\mathbf{s}}$ (not a Toeplitz matrix as in [22]). Let the non-linear function be

$$F_R(\lambda_i(\Sigma_{\mathbf{s}}), \mu) = \log_2 \frac{\lambda_i(\Sigma_{\mathbf{s}})}{(\lambda_i(\Sigma_{\mathbf{s}})D_i(\mu) - \frac{D_i^2(\mu)}{4})^{\frac{1}{2}}}, \tag{4.27}$$

where $D_i(\cdot)$ is also a non-linear function of $\lambda_i(\Sigma_{\mathbf{s}})$, as shown in (4.26). We define $R^{DP}(\mu)$ as an asymptotic ($N \to \infty$) average of non-linear functions $F_R(\lambda_i(\Sigma_{\mathbf{s}}), \mu)$, which is given by

$$R^{DP}(\mu) = \lim_{N\to\infty} \frac{1}{N} R_{MN}^{DP}(\mu) = \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{MN} F_R(\lambda_i(\Sigma_{\mathbf{s}}), \mu) \tag{4.28}$$

With a similar argument, $D^{DP}(\mu)$ is defined as

$$D^{DP}(\mu) = \lim_{N\to\infty} \frac{1}{N} D_{MN}^{DP}(\mu) = \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{MN} D_i(\mu) \tag{4.29}$$

We use (4.8), the extension of the Szego's theorem, to find the corresponding rate-distortion tradeoff in the frequency domain. Substituting non-linear functions in (4.27) and (4.26) into (4.8), the corresponding equivalences of (4.28) and (4.29) are derived in the frequency domain, respectively, and consequently, the asymptotic DP-RDF for time-stationary Gaussian vector sources is given by

$$R^{DP}(\mu) = \\ = \frac{1}{2\pi} \int_{-\pi}^{\pi} \sum_{u=1}^{M} \log_2 \frac{\lambda_u(\mathbf{\Phi}_{\mathbf{s}}(\Omega))}{(\lambda_u(\mathbf{\Phi}_{\mathbf{s}}(\Omega))D_u(\mu,\Omega) - \frac{D_u^2(\mu,\Omega)}{4})^{\frac{1}{2}}} d\Omega$$
$$D^{DP}(\mu) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \sum_{u=1}^{M} D_u(\mu,\Omega) d\Omega, \tag{4.30}$$

where

$$D_u(\mu,\Omega) = 2\,\lambda_u(\mathbf{\Phi_s}(\Omega)) + \mu - (4\,\lambda_u^2(\mathbf{\Phi_s}(\Omega)) + \mu^2)^{\frac{1}{2}}, \tag{4.31}$$

for $D_u(\mu,\Omega) < 2\lambda_u(\mathbf{\Phi_s}(\Omega))$. The asymptotic rate $R^{DP}(\mu)$ in (4.30) is assumed to be set to zero for $D^{DP}(\mu) > \frac{1}{\pi}\int_{-\pi}^{\pi}\sum_{u=1}^{M}\lambda_u(\mathbf{\Phi_s}(\Omega))d\Omega$.

4-B: A TEST CHANNEL ACHIEVING DP-RDF

We show that the conceptual test channel achieving (4.17) is based on the vector quantizer model shown in Fig. 4.4a. We derive the respected quantization parameters $\mathbf{B}$ and $\mathbf{\Phi}_{\mathbf{e}_1}$ for *Method 2*. We introduce the following lemma.

Lemma 1- Let $\tilde{z}(\Omega) \in \mathbb{C}$ and $z(\Omega) \in \mathbb{C}$ be zero-mean Gaussian random variables representing frequency domain signals. Then there exist a real-valued linear operator (scaling factor) β and a zero-mean Gaussian random variable $e(\Omega)$, uncorrelated to $z(\Omega)$, such that

$$\tilde{z} = \beta z + e, \tag{4.32}$$

and $E[|\tilde{z}|^2] = E[|z|^2]$, i.e., variables $\tilde{z}$ and $z(\Omega)$ have the same PSDs.

Proof- Denote the cross PSDs between $\tilde{z}$ and z by $\Phi_{\tilde{z}z} = E[\tilde{z}z^*]$ and $\Phi_{z\tilde{z}} = E[z\tilde{z}^*]$ and PSD of e by Φ_e, where $(\cdot)^*$ denotes the conjugate operator. Based on (4.32), the (cross) PSD relations are given by

$$\Phi_{\tilde{z}} = \beta^2\Phi_z + \Phi_e, \tag{4.33a}$$

$$\Phi_{\tilde{z}z} = \Phi_{z\tilde{z}} = \beta\Phi_z, \tag{4.33b}$$

where we used the fact that z and e are uncorrelated, and that β is real. We define $D \triangleq E[|z-\tilde{z}|^2] \triangleq \Phi_{e'}$ where e' can be thought of as the error variable $e' = z - \tilde{z}$. As $\Phi_{\tilde{z}} = \Phi_z$, the distortion function D can be written as

$$D = \Phi_z + \Phi_z - 2\mathrm{Re}\{\Phi_{z\tilde{z}}\} = 2\Phi_z - 2\Phi_{z\tilde{z}}. \tag{4.34}$$

Solving (4.34) and (4.33b) for β we have

$$\beta = 1 - \frac{D}{2\Phi_z}, \tag{4.35}$$

and substituting (4.35) into (4.33a) for Φ_e, we have

$$\Phi_e = (\frac{1+\beta}{2})D. \tag{4.36}$$

The proof is complete.

Using Lemma 1, we derive the following distribution preserving quantization procedure for vector sources in the frequency domain, which achieves the DP-RDF in (4.30).

First, the left-side observations $\mathbf{y}_1$ are decorrelated as $\mathbf{z}_1 = \mathbf{A}\mathbf{y}_1$, using a unitary transformation matrix $\mathbf{A}$. Second, each element of the decorrelated vector $\mathbf{z}_1$, denoted by $Z_u(\Omega),;\ u = 1,..,M_1$, can be quantized in a probability distribution preserving manner based on the test channel model presented in Lemma 1 as $\tilde{Z}_u(\Omega) = \beta_u(\Omega)Z_u(\Omega) + E_u(\Omega)$. Let us denote the MSE $E[|Z_u(\Omega) - \tilde{Z}_u(\Omega)|^2]$ by $D_u(\Omega)$. Therefore, the distribution preserving quantization parameters $\beta_u(\Omega)$ and $\Phi_{E_u}(\Omega)$ correspond to (4.35) and (4.36), by replacing D and Φ_z with $D_u(\Omega)$ and $\Phi_{Z_u}(\Omega)$, respectively. Note that here the PSD of each element is preserved after the quantization, i.e., $E[|\tilde{Z}_u(\Omega)|^2] = E[|Z_u(\Omega)|^2] = [\Phi_{\mathbf{z}_1}]_{uu}$. We know from [22] and App. 4-A that the optimal choices for the distortions $D_u(\Omega)$ are derived by minimizing the sum-rate with respect to the constraint on total distortion, i.e.,

$\sum_{u=1}^{M_1} D_u(\Omega) \leq D^{DP}(\Omega)$. Therefore the optimal values for $D_u(\Omega)$, which are derived based on (4.31) by replacing $\lambda_u(\mathbf{\Phi_s}(\Omega))$ with $[\Phi_{\mathbf{z}_1}]_{uu}$, i.e.,

$$D_u(\mu,\Omega) = 2\,[\Phi_{\mathbf{z}_1}]_{uu} + \mu - (4[\Phi_{\mathbf{z}_1}]_{uu} + \mu^2)^{\frac{1}{2}}, \tag{4.37}$$

and hence, the optimal distribution preserving (DP) quantization of the decorrelated vector $\mathbf{z}_1$ is modeled by $\tilde{\mathbf{z}}_1 = \mathbf{B}\mathbf{z}_1 + \mathbf{e}_1$, where $\mathbf{B}$ is a diagonal matrix and the elements correspond to (4.35), replacing D and Φ_z with the optimal DP-MSE $D_u(\mu,\Omega)$ and $[\Phi_{\mathbf{z}_1}]_{uu}$, respectively. The vector $\mathbf{e}_1 = [E_1(\Omega),...,E_{M_1}(\Omega)]^{\mathrm{T}}$ will have the diagonal PSD matrix $\mathbf{\Phi}_{\mathbf{e}_1}$ which correspond to (4.36) with similar substitutions to those for $\mathbf{B}$. Finally, the decorrelated-quantized vector $\tilde{\mathbf{z}}_1$ will be transformed back to the original quantized vector $\tilde{\mathbf{y}}_1$ applying inverse-transform matrix $\mathbf{A}^{-1}$ ($\tilde{\mathbf{y}}_1 = \mathbf{A}^{-1}\tilde{\mathbf{z}}_1$).

Following the above-mentioned procedure, we summarize the achievability proof of the distortion given in (4.17) by defining two error variables $\mathbf{d}_1 = \mathbf{y}_1 - \tilde{\mathbf{y}}_1$ and $\mathbf{e}'_1 = \mathbf{z}_1 - \tilde{\mathbf{z}}_1$. The direct distribution preserving MSE between $\mathbf{y}_1$ and $\tilde{\mathbf{y}}_1$ is denoted by $D^{DP}(\mu)$. We have

$$\begin{aligned}
D^{DP}(\mu) &= \frac{1}{2\pi}\int_{-\pi}^{\pi} \mathrm{tr}\{\mathbf{\Phi}_{\mathbf{d}_1}\}d\Omega = \frac{1}{2\pi}\int_{-\pi}^{\pi} \mathrm{tr}\{\mathbf{\Phi}_{\mathbf{e}'_1}\}d\Omega && \text{(4.38a)}\\
&= \frac{1}{2\pi}\int_{-\pi}^{\pi} \mathrm{tr}\{(\mathbf{I}-\mathbf{B})\mathbf{\Phi}_{\mathbf{z}_1}(\mathbf{I}-\mathbf{B})^{\mathrm{H}} + \mathbf{\Phi}_{\mathbf{e}_1}\}d\Omega && \text{(4.38b)}\\
&= \frac{1}{2\pi}\int_{-\pi}^{\pi} \sum_{u=1}^{M_1} (1-\beta(\mu,\Omega))^2[\mathbf{\Phi}_{\mathbf{z}_1}]_{uu} + \\
&\qquad \frac{(1+\beta(\mu,\Omega))}{2} D_u(\mu,\Omega)d\Omega && \text{(4.38c)}\\
&= \frac{1}{2\pi}\int_{-\pi}^{\pi} \sum_{u=1}^{M_1} D_u(\mu,\Omega)d\Omega, && \text{(4.38d)}
\end{aligned}$$

which corresponds to the distortion function in (4.17), for the different vector source $\mathbf{y}_1$ (and not $\mathbf{s}$ in (4.17)). $D_u(\mu,\Omega)$ here corresponds to (4.37). Similar argument holds for the achievability proof of the parametric distribution preserving rate $R^{DP}(\mu)$.

REFERENCES

[1] V. Hamacher *et al.*, "Signal Processing in High-End Hearing Aids: State of the Art, Challenges, and Future Trends," EURASIP Journal on Advances in Signal Processing, no. 18, pp. 2915-2929, 2005.

[2] R. Sockalingam,M. HolmbergK. Eneroth, and M. Shulte, "Binaural hearing aid communication shown to improve sound quality and localization," The Hearing Journal, vol. 62, no. 10, pp. 46-47, 2009.

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

[4] V. Hamacher, “Comparison of advanced monaural and binaural noise reduction algorithms for hearing aids,” IEEE International Conference on Acoustics, Speech, and Signal Processing, vol. 4, pp. 4008-4011, 2002.

[5] S. Doclo, W. Kellermann, S. Makino, and S. E. Nordholm, “Multichannel signal enhancement algorithms for assisted listening devices: Exploiting spatial diversity using multiple microphones,” IEEE Signal Processing Magazine, vol. 32, no. 2, pp. 18–30, 2015.

[6] S. Doclo and M. Moonen, “GSVD-based optimal filtering for single and multimicrophone speech enhancement,” IEEE Transactions on Signal Processing, vol. 50, no. 9, pp. 2230–2244, 2002.

[7] S. Doclo, A. Spriet, J. Wouters, and M. Moonen, “Speech distortion weighted multichannel wiener filtering techniques for noise reduction,” Speech Enhancement, pp. 199–228, Berlin, Heidelberg: Springer, 2005.

[8] T. J. Klasen, T. Van den Bogaert, M. Moonen, and J. Wouters, “Binaural noise reduction algorithms for hearing aids that preserve interaural time delay cues,” IEEE Transactions on Signal Processing, vol. 55, no. 4, pp. 1579–1585, 2007.

[9] D. Marquardt, *Development and evaluation of psychoacoustically motivated binaural noise reduction and cue preservation techniques,* PhD Dessertation, Universiy of Oldenburg, 2015.

[10] T. Lotter, and P. Vary, “Dual-Channel Speech Enhancement by Superdirective Beamforming,” EURASIP Journal on Advances in Signal Processing, no. 1, pp. 1-14, 2006.

[11] S. Srinivasan, “Low-bandwidth binaural beamforming,” Electronics Letters, vol. 44, no. 22, pp. 1292–1293, 2008.

[12] O. Roy and M. Vetterli, “Rate-constrained collaborative noise reduction for wireless hearing aids,” IEEE Transactions on Signal Processing, vol. 57, no. 2, pp. 645–657, 2009.

[13] T. Berger, *Rate-distortion theory: A mathematical basis for data compression,* Englewood Cliffs, N.J.: Prentice-Hall, 1971.

[14] J. K. Wolf and J. Ziv, “Transmission of noisy information to a noisy receiver with minimum distortion,” IEEE Transactions on Information Theory, vol. 16, no. 4, pp. 406–411, 1970.

[15] T. Flynn, and R. Gray, “Encoding of correlated observations,” IEEE Transactions on Information Theory, vol. 33, no. 6, pp. 406–411, 1970.

[16] A. D. Wyner and J. Ziv, “The rate-distortion function for source coding with side information at the decoder,” IEEE Transactions on Information Theory, pp. 1–10, 1976.

[17] H. Yamamoto and K. Itoh, "Source coding theory for communication systems with a remote source," Trans. IECE Jpn, vol. E63, no. 6, pp. 700–706, 1980.

[18] S. Srinivasan and A. den Brinker, "Rate-constrained beamforming in binaural hearing aids," EURASIP Journal on Advances in Signal Processing, pp. 1–9, 2009.

[19] S. Srinivasan and A. C. den Brinker, "Analyzing rate-constrained beamforming schemes in wireless binaural hearing aids," in 2009 17th European Signal Processing Conference, pp. 1854–1858, 2009.

[20] S. Doclo, M. Moonen, T. Van den Bogaert, and J. Wouters, "Reducedbandwidth and distributed MWF-based noise reduction algorithms for binaural hearing aids," IEEE Transactions on Audio, Speech, and Language Processing, vol. 17, no. 1, pp. 38–51, 2009.

[21] S. Doclo, T. C. Lawin-Ore, and T. Rohdenburg, "Rate-constrained binaural MWF-based noise reduction algorithms," in Proc. ITG Conference on Speech Communication, Bochum, Germany, 2010.

[22] M. Li, A. Ozerov, J. Klejsa, and W. B. Kleijn, "Asymptotically optimal distribution preserving quantization for stationary Gaussian processes," QC 20110829, 2011.

[23] T. M. Cover and J. A. Thomas, E*lements of information theory,* Wiley- Interscience, 2006.

[24] V. Kafedziski, "Rate distortion of stationary and nonstationary vector Gaussian sources," in IEEE/SP 13th Workshop on Statistical Signal Processing, 2005, pp. 1054–1059, 2005.

[25] U. Grenander and G. Szego, *Toeplitz forms and their applications*, Berkeley: University of California Press, 1958.

[26] H. Gazzah, P. A. Regalia, and J. P. Delmas, "Asymptotic eigenvalue distribution of block Toeplitz matrices and application to blind SIMO channel identification," IEEE Transactions on Information Theory, vol. 47, no. 3, pp. 1243–1251, 2001.

[27] P. Tilli, "Singular values and eigenvalues of non-hermitian block Toeplitz matrices," Linear Algebra and its Applications, vol. 272, no. 1, pp. 59– 89, 1998.

[28] M. Li, J. Klejsa, and W. B. Kleijn, "On distribution preserving quantization," arXiv preprint arXiv:1108.3728, 2011.

[29] H. Kayser, S. D. Ewert, J. Anem¨uller, T. Rohdenburg, V. Hohmann, and B. Kollmeier, "Database of multichannel in-ear and behind-the-ear head-related and binaural room impulse responses," EURASIP J. Adv. Signal Process, vol. 2009, pp. 6:1–6:10, 2009.

[30] J. Kominek, A. W. Black, and V. Ver, "CMU arctic databases for speech synthesis," Tech. Rep., 2003.

5

Operational Rate-Constrained Beamforming in Binaural Hearing Aids

This chapter is published as "Operational Rate-Constrained Beamforming in Binaural Hearing Aids," by J. Amini, R. C. Hendriks, R. Heusdens, M. Guo and J. Jensen, in 26th European Signal Processing Conference (EUSIPCO), Rome, 2018, pp. 2504-2508.

Hearing aid (HA) devices are designed to increase the speech intelligibility. A typical way to improve the speech intelligibility is by means of multi-microphone noise reduction [1][2]. Modern HAs can collaborate through a wireless link to construct a binaural HA system. This considerably improves the potential of noise reduction, as effectively a larger microphone array can be used [3][4]. In addition, binaural HAs can collaborate with other assistive devices, and form a small wireless acoustic sensor network (WASN).

In such a small WASN, microphone recordings are received at the fusion center (FC), which estimates the target sources and suppresses the interferers. In this work, one of the two HAs is considered as an FC. A well-known binaural filter is the binaural multichannel Wiener filter (MWF) [5], which is based on constructing two monaural MWF beamformers. Each MWF tries to estimate the source of interest by linearly combining its locally recorded signals with those from the contralateral device such that the mean square error (MSE) between the target source and its estimate is minimized. Other binaural beamforming approaches, including [6] and [7], try to preserve some important spatial information of the target and interfering sources when minimizing the MSE.

To perform such binaural processing, the noisy observations need to be transmitted through wireless links to the FC. As the transmission capacities of such links are limited, the data must be quantized at a certain bit-rate [8]. This brings the notion of rate-constrained beamforming into the noise reduction problem. In [9] a binaural rate-constrained beamforming problem is introduced, assuming jointly Gaussian random sources, where an efficient trade-off between the transmission rate and the MSE between the target signal and its estimate is derived. However, this optimal framework is limited to only two processing nodes and is less practical due to the strong requirement that joint statistics are known at all processors and (infinitely) long-block vector quantizers are used. Transmission between binaural HAs and other assistive devices is thus not considered, nor how more practical implementations affect the performance. Different (sub-optimal) binaural rate-constrained approaches are proposed in [8] and [10], which provide more practical alternatives to the method in [9]. However, the performance of such methods depends heavily on the acoustic scene (e.g. target source location, spatial noise distributions, etc.) and it is typically far from optimal, even asymptotically, i.e., at sufficiently high rates.

In this work, the binaural HA problem is approached from a more general perspective. The general setup of a (small) WASN is considered here, where joint statistics are only assumed to be known at the FC, instead of at every node as in [9]. The binaural noise reduction problem is solved by minimizing a fidelity criterion, while satisfying a bit-rate constraint. To overcome the acoustic scene dependency, we consider a discrete set of processing candidates and a (discrete) set of operating resources (in this case bit-rates). This problem formulation of optimizing among a set of strategies under a rate constraint is related to operational rate-distortion optimization [11][12]. In [11] an elegant operational rate-distortion optimization method was proposed for rate allocation among an arbitrary set of quantizers. Most related approaches were inspired by the method in [11] for different applications such as optimal time segmentation of speech [13] or finding optimal time-varying wavelet packet bases for signal expansion [14].

We propose a new operational rate-constrained beamforming algorithm based on both strategy selection and rate allocation in the frequency domain. The Lagrange multi-

plier (LM) based technique [11] is used to allocate the rates and select the best strategies over frequency, while minimizing the sum of estimation error power spectral densities (PSDs). Unlike the theoretical approaches [8][9], the proposed method allows an arbitrary range of operating rates in each frequency bin. Moreover, it enables forming the set of processing candidates from existing algorithms and optimally choosing between different strategies in different frequency bins. The proposed method is evaluated based on the output MSE gap between the monaural (i.e., no communication) setup and the (rate-constrained) generalized binaural setup. The results show significant improvements in comparison with naive strategy selection and equal rate allocation across frequencies.

5.1. PROBLEM STATEMENT

The generalized binaural HA system that we consider consists of two wireless collaborating HAs with M_1 and M_2 microphones, respectively, and M_A assistive processors, which can collaborate with the HAs. The total number of microphones is thus $M = M_1 + M_2 + M_A$. In general, each assistive device can be equipped with multiple microphones. However, in this work, it is assumed for simplicity that each assistive processor is equipped with a single microphone. The clocks of the devices are assumed to be synchronized. All microphones receive a filtered version of the target speech signal, which is indicated in the short-time frequency transform (STFT) domain by $S[k]$, with $[k]$ denoting the frequency bin index. Notice that the time-frame index is neglected for notational convenience. The target speech is degraded by interfering noise, which might originate from, e.g., interfering point sources, diffuse noise, and microphone self-noise. The interfering noise observed at a particular microphone is indicated by $N_i[k]$, with $i = 1,\dots,M$ the microphone index. The signals $S[k]$ and $N_i[k]$, for $i = 1,\dots,M$ are assumed additive and mutually uncorrelated. Altogether we then have

$$Y_i[k] = A_i[k]S[k] + N_i[k], \tag{5.1}$$

where A_i is the acoustic transfer function (ATF) between the target signal and the ith microphone. The signal model can be rewritten in vector notation by stacking all noisy microphone coefficients in a vector, as

$$\mathbf{y} = \mathbf{x} + \mathbf{n}, \tag{5.2}$$

where $\mathbf{x} = \mathbf{a}S$, $\mathbf{y} = [Y_1[k],...,Y_M[k]]^{\mathrm{T}}$, and similarly for $\mathbf{n}$ and $\mathbf{a}$. Notice that we have left out the frequency bin index in (5.2) for notational convenience. The superscripts $(\cdot)^{\mathrm{T}}$ and $(\cdot)^{\mathrm{H}}$ denote transpose and conjugate transpose operators, respectively. The cross-power spectral density (CPSD) matrix $\mathbf{\Phi_y}$ of the vector $\mathbf{y}$ is given by $\mathbf{\Phi_y} = \mathbf{\Phi_x} + \mathbf{\Phi_n}$, where $\mathbf{\Phi_x} = \Phi_S\mathbf{a}\mathbf{a}^{\mathrm{H}}$, $\mathbf{\Phi_n} = E[\mathbf{n}\mathbf{n}^{\mathrm{H}}]$ with $\Phi_S = E[|S|^2]$ the PSD of the clean speech S, and with $E[\cdot]$ the expectation operator.

In this paper, our goal is to estimate the clean speech target signal at the FC. However, apart from the microphone signals acquired at the FC, the additional microphone signals are only available in quantized form. These signals are compressed at a certain operating rate, say R bits per sample (bps), which is considered as a (constrained) resource. Depending on this resource and the actual acoustic scene, different algorithm selections

are optimal. Therefore, we address the problem of operational rate-constrained beamforming in order to find the optimal beamforming strategy, given a set of candidate algorithms, satisfying the bit-rate as a resource constraint.

5.2. OPERATIONAL RATE-CONSTRAINED BEAMFOMRMING

Inspired by [11], in this section we propose operational rate-distortion optimization for beamforming based on both rate allocation and strategy selection across frequencies.

We are given a set $\mathscr{A} = \{A_1, A_2, \ldots, A_{N_A}\}$ of strategy candidates (could be different microphone configurations, different beamforming algorithms, and/or different coding schemes on the microphone signals) with cardinality $|\mathscr{A}| = N_A$. The goal is to optimally select the candidates and allocate the resources (bit-rates) in order to minimize a distortion, in this case, the MSE between the remote-source S and its estimate $\hat{S}$ in the frequency domain, while satisfying the constraints on the total rate budget, say $R_{\max}$. The proposed optimization problem is given by

$$\begin{aligned} &\min_{\boldsymbol{\alpha}\in\mathscr{A}'} \quad \min_{\mathbf{r}\in\mathscr{Q}} \quad D(\boldsymbol{\alpha},\mathbf{r}) \\ &\text{subject to} \quad R(\mathbf{r}) \le R_{\max}, \end{aligned} \tag{5.3}$$

where $\boldsymbol{\alpha} = [\alpha_1, \ldots, \alpha_{N_f}]^{\mathrm{T}}$ denotes a vector variable for possible choices of strategies for all N_f frequency bins. Similarly, $\boldsymbol{r} = [r_1, \ldots, r_{N_f}]^{\mathrm{T}}$ indicates a vector variable for possible operating rates to be allocated to the frequency components. The set of all possible strategy choices is given by $\mathscr{A}' = \{\boldsymbol{\alpha} \mid \alpha_k \in \mathscr{A}\}$, for $k = 1, \ldots, N_f$. The set $\mathscr{Q} = \{\mathbf{r} \mid r_k \in \mathscr{Q}_k\}$ consists of possible operating rates, where $\mathscr{Q}_k = \{p_k, \ldots, q_k\}, q_k > p_k \ge 0$, with representative cardinality $N_r = \max\{|\mathscr{Q}_1|, \ldots, |\mathscr{Q}_{N_f}|\}$, for all frequency bins. Note that p_k and q_k are the minimum and the maximum operating rates, respectively, for a particular frequency. $D(\boldsymbol{\alpha},\mathbf{r})$ is the averaged PSD of the estimation error, given the algorithm choices and rate allocation across frequencies and is given by

$$D(\boldsymbol{\alpha},\mathbf{r}) = \frac{1}{N_f}\sum_{k=1}^{N_f} d(\alpha_k, r_k), \tag{5.4}$$

where

$$d(\alpha_k, r_k) = E[|S[k] - \hat{S}[k]|^2 | \alpha_k, r_k], \tag{5.5}$$

which denotes the PSD of the estimation error in the kth discrete frequency bin, given the algorithm α_k and the quantization rate r_k. The cost function $R(\mathbf{r})$ is simply defined as the averaged rate over all bins and is given by

$$R(\mathbf{r}) = \frac{1}{N_f}\sum_{k=1}^{N_f} r_k. \tag{5.6}$$

The original problem in (5.3) is a (discrete) combinatorial optimization problem. Every possible solution is an operating point located in the 2-dimensional D-R coordinate system (D-R characteristics). Figure 5.1 illustrates an example D-R characteristic. The problem of finding the optimal operating point which satisfies the constraint in (5.3) is

untractable. One way to make the search problem tractable is to approximate the convex hull of the set of all possible solutions and select a point on the convex hull which satisfies the constraints [12]. Using the LM technique [11], the original problem in (5.3) is reformulated to the following Lagrangian form as

$$\min_{\boldsymbol{\alpha}\in\mathscr{A}'} \quad \min_{\mathbf{r}\in\mathscr{Q}} \quad D(\boldsymbol{\alpha},\mathbf{r}) + \lambda\, R(\mathbf{r}), \tag{5.7}$$

where λ is known as the Lagrange multiplier which satisfies $R(\mathbf{r}^*(\lambda)) \leq R_{\max}$. Substituting (5.4) and (5.6) into (5.7), we have

$$\min_{\boldsymbol{\alpha}\in\mathscr{A}'} \quad \min_{\mathbf{r}\in\mathscr{Q}} \quad \frac{1}{N_f}\sum_{k=1}^{N_f} d(\alpha_k, r_k) + \lambda\, \frac{1}{N_f}\sum_{k=1}^{N_f} r_k. \tag{5.8}$$

As the optimization objective function is separable across frequency, the problem can be further simplified to

$$\frac{1}{N_f}\sum_{k=1}^{N_f} \Big(\min_{r_k\in\mathscr{Q}_k} \min_{\alpha_k\in\mathscr{A}} (d(\alpha_k, r_k)) + \lambda\, r_k \Big). \tag{5.9}$$

After optimizing over α_k the problem can finally be reformulated as

$$\frac{1}{N_f}\sum_{k=1}^{N_f} \Big(\min_{r_k\in\mathscr{Q}_k} d^\star(r_k) + \lambda\, r_k \Big), \tag{5.10}$$

where $d^\star(r_k)$ is the minimum distortion per frequency with respect to the best strategy candidate choices, for a given rate r_k. Notice that for small N_A, $d^\star(r_k)$ can be found with exhaustive search. The final minimization problem can be solved by finding the operating point in the D-R curve which intersects first by the constant slope line $d_k + \lambda\, r_k = b$ with $b > 0$, for each frequency bin k [12]. This is illustrated in Figure 5.1. Alternatively, for small N_r, the best r_i values can be found by exhaustive search. The final step is to find a "good" λ satisfying the total rate budget constraint by iterating the same procedure in (5.10). For convex D-R relations, finding the optimal λ can be done using bisection algorithms [12][14]. However, as the D-R relations are not always convex, we use the method described in [11] (Variant 2) with a modified initialization formula, which is given by

$$\lambda^0 = \frac{1}{N_f}\sum_{k=1}^{N_f} [d^\star(\min(R_{\max}, q_k - 1)) - d^\star(\min(R_{\max}, q_k - 1) + 1)], \tag{5.11}$$

where λ^0 is the initial LM value, given a total rate budget $R_{\max}$ and q_k is the maximum operating rate at a particular frequency. More details about the method can be found in [11].

5.3. QUANTIZATION AWARE MWF BEAMFORMING

In this section, we describe an application of the presented theory to rate-constrained MWF beamforming using uniform quantizers in a small WASN.

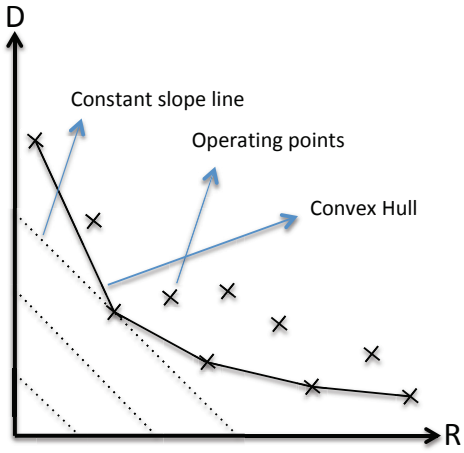


Figure 5.1: Geometric interpretation of the problem in (5.10)

Let us assume the left side HA acts as an FC. The goal is to estimate the target signal S at the left reference microphone, given local (left-side) information and remote quantized signals from other microphones. The remote signals are quantized through uniform quantization as follows. The signal x is quantized, and the quantized signal is denoted by $\tilde{x}$. Therefore, under certain assumptions [15][16], the quantization error $e = x - \tilde{x}$ is uniformly distributed with variance $\sigma_e^2 = \frac{\Delta^2}{12}$, where $\Delta = \frac{2x_{\max}}{2^R}$ is a step size, which depends on the range of the signal (maximum value $x_{\max}$) and the quantization rate R.

Let $\tilde{\mathbf{y}}_{\text{rem}}$ denote the concatenation of the STFT coefficients obtained from the quantized and transmitted remote microphone signals. The vector $\tilde{\mathbf{y}}_{\text{rem}}$ is then combined with the local information $\mathbf{y}_{\text{loc}}$ to construct the total observation vector $\mathbf{y}_{\text{tot}} = [\mathbf{y}_{\text{loc}}^{\text{T}} \tilde{\mathbf{y}}_{\text{rem}}^{\text{T}}]^{\text{T}}$. Finally, using the MWF beamformer, the estimated signal per frequency is given by $\hat{S} = \mathbf{w}^{\text{H}}\mathbf{y}_{\text{tot}}$, where $\mathbf{w}$ denotes the vector of optimal Wiener filter coefficients. The PSD of the MWF estimation error (for a particular frequency bin) is then given by

$$d(S,\hat{S}) = E[|S-\hat{S}|^2] = \Phi_S - \mathbf{\Phi}_{S\mathbf{y}_{\text{tot}}} \mathbf{\Phi}_{\mathbf{y}_{\text{tot}}}^{-1} \mathbf{\Phi}_{\mathbf{y}_{\text{tot}}S}, \tag{5.12}$$

where $\mathbf{\Phi}_{S\mathbf{y}_{\text{tot}}} = \mathbf{\Phi}_{\mathbf{y}_{\text{tot}}S}^{\text{H}}$ denotes the cross PSD vector between the target signal S and $\mathbf{y}_{\text{tot}}$, and $\mathbf{\Phi}_{\mathbf{y}_{\text{tot}}}$ denotes the cross PSD matrix of the vector $\mathbf{y}_{\text{tot}}$. The quantized signal vector $\tilde{\mathbf{y}}_{\text{rem}}$ is actually a function of the chosen strategy. Based on (5.12), distortions for different strategies and rates are computed. In this chapter, we consider a particular application of the presented theory, where the possible strategies consist of selection of local/remote signals and different bit-rate allocation schemes among these signals.

5.4. EXPERIMENTS

In this section we apply the method proposed in Section 5.2 to an example acoustic scene and perform simulations to evaluate the performance.

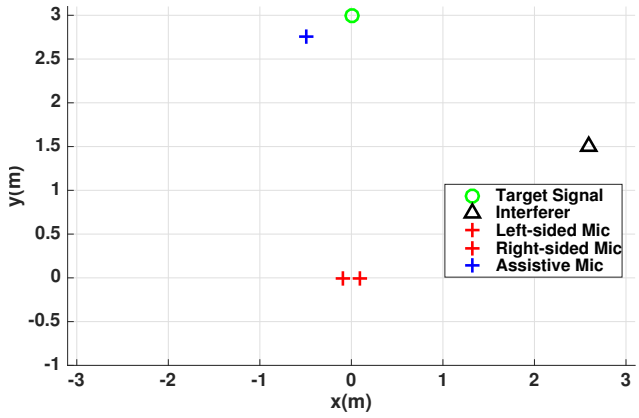


Figure 5.2: Generalized binaural HA setup

5.4.1. SETUP

The experimental setup is shown in Figure 5.2. Two red "+" symbols denote two microphones (one microphone per HA) located along the horizontal x-axis at a distance 10 cm from the origin ($(x, y) = (0, 0)$). The target speech signal, shown by the green circle, is located in front of the binaural HA system (at zero degrees) with a distance of 3 m from the origin. In this chapter, the angles are computed counter-clockwise and the straight looking direction corresponds to zero degrees. The blue "+" symbol shows the assistive wireless microphone located closer to the target speech signal, at $\theta = 10°$ and a distance of 3 m from the origin, where $\theta = \arctan(\frac{y}{x}) - \frac{\pi}{2}$. The interfering signal, which is denoted by the black triangle, is located at $-60°$ with a distance of 3 m from the origin. The point noise source (interfering signal) has a flat PSD $\Phi_{n_1}(\Omega)$ over the interval $\Omega \in [-\pi, \pi]$. Arround 10 s of the $F_s = 16$ kHz sampled speech of the "CMU-ARCTIC" [17] database are used for the PSD estimation (Φ_S) based on Welch's method. 512 discrete Fourier transform (DFT) coefficients, computed frame-by-frame from 50% overlapping speech frames, are used in the PSD estimation process. The cross PSD matrices are calculated using true ATFs [18] and corresponding estimated PSDs.

The reference microphone is chosen to be the microphone in the left-side HA (the FC). In addition to the target speech signal and the interferer, internal microphone noise is simulated and added, which is assumed to be uncorrelated between microphones. The signal-to-noise ratio (SNR) for the internal noise with respect to the target at the reference microphone is 40 dB. Similarly, the signal-to-interferer ratio (SIR) for the interferer is 0 dB.

5.4.2. STRATEGY CANDIDATE SET FOR SIMULATIONS

Based on the acoustic scene shown in Figure 5.2, we design the following strategy candidate set:

- 2CH: Rate-constrained MWF beamforming with two microphone signals, i.e., the left side (FC) and the right side microphone signals.
- 2CHa: Rate-constrained MWF beamforming with two microphone signals, i.e., the

left side and the assistive microphone signals.

- 3CH: Rate-constrained MWF beamforming with all three microphone signals. Note that opposed to the first two strategies, in this strategy multiple remote signals are selected. This implies that the total rate-budget now has to be allocated not only over frequency, but also over the two microphone signals.

When it happens that in one strategy (e.g., the candidate 3CH) there is more than one WASN node for which data needs to be quantized, then the candidate set is extended to cover all relevant rate allocations across microphones.

The number of all possible rate allocations across M microphones given N_r different operating rates ($0 \leq r \leq N_r - 1$) are computed as

$$|\mathscr{A}|^+ = \binom{M-1}{M-1} + \binom{M}{M-1} + \ldots + \binom{N_r+M-2}{M-1}. \tag{5.13}$$

The final set will be the union of the initial strategy set and the set which consists of all combinations across microphones. For example, in the candidate 3CH two quantized signals are transmitted to the FC, i.e., $M = 2$ in (5.13). In the experiments the same rate range $0 \leq r_k \leq 32$ is chosen for all frequencies, i.e., $N_r = 33$. Therefore the total number of combinations (strategy choices) will be 561.

5.4.3. EVALUATION

In this section, we compare variants of the proposed method with methods proposed in the literature. The following methods are compared:

- Full generalized binaural MWF: The MWF with all three microphone signals. This method serves as a performance bound assuming the signals are available at the infinite rate.
- Full binaural MWF: The MWF with both the left and right microphone signals. Similarly, this method serves as a performance bound for the binaural setup.
- Equal 2CH: The candidate 2CH. The rates are equally allocated over all frequencies.
- Equal 2CHa: The candidate 2CHa. The rates are equally allocated over all frequencies.
- Equal 3CH: The candidate 3CH. The rates are assumed to be equally allocated over all frequencies as well as across microphones.
- Proposed LM: The proposed method described in Section 5.2. The distortions are computed based on (5.12), for different algorithm choices and rates. Note that this method optimally allocates the rates over all frequencies, but equally across microphones, when a strategy is selected that involves multiple microphones.
- Proposed LM-modified (LM-M): This method is based on the Proposed LM, and optimally allocates the rates over all frequencies and across microphones, using the extended strategy set described in Section 5.4.2.

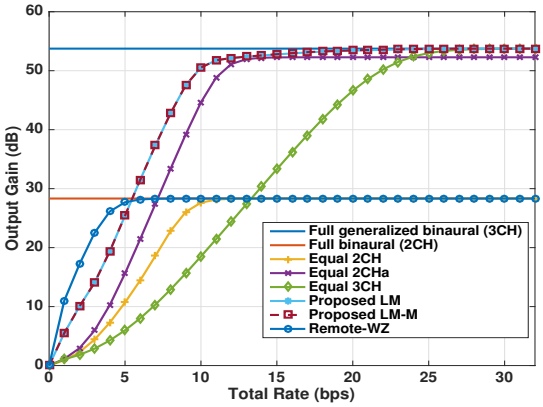


Figure 5.3: Output Gain (dB) versus total rate (bit per sample)

- Remote-Wyner-Ziv (WZ) [9]: The binaural rate-constrained beamforming presented in [9]. Note that only two HA microphones can be used in this method, joint statistics are needed at all processors (nodes) and long-block vector quantizers are impractical.

The performance measure is defined as the ratio of the MSE for the monaural configuration, i.e. when there is no communication with the FC, and the MSE achieved by the above-mentioned methods, and is given by

$$G = \frac{D(0)}{D(\boldsymbol{\alpha}, \mathbf{r})}. \tag{5.14}$$

The vectors $\boldsymbol{\alpha}$ and $\mathbf{r}$ are optimally chosen for the methods "proposed LM" and "proposed LM-M". For the other (reference) methods, $\boldsymbol{\alpha}$ is fixed as no selection is possible. Figure 5.3 shows the output gains G in dB as a function of the total bit-rate budget ($R_{\max}$). The performance of the 2CH-based methods saturates to that of full binaural MWF, as expected. The performance of the remote-WZ method is computed based on the theoretical upper bound, described in [9]. As shown, the performance curve of the remote-WZ method saturates as the assistive microphone is not considered in this method.

The proposed methods select the best microphone configurations and find optimal rate allocations over frequency. The performance curves of the proposed methods LM and LM-M almost coincide, for this specific scenario, as the proposed optimization problem mostly chooses the 2CH-based candidates at low and middle rates. However, at middle and high rates the proposed methods tend to select the 3CH or 2CHa candidates, and the proposed LM-M method performs slightly better than the LM method, as unequal (efficient) rate allocations are chosen across the right-side and the assistive microphone signals.

5.5. CONCLUSION

In this chapter, we proposed an operational rate-distortion based optimization problem for both strategy selection and rate allocation over frequency in (small) WASNs. Unlike

existing binaural beamforming algorithms, we considered a potential collaboration between the binaural HAs and some assistive wireless processors in a rate-distortion sense. The sensitivity of existing methods to the acoustic scene is addressed by introducing the strategy candidate set. The proposed framework was applied to the rate-constrained MWF beamforming problem. Assuming uniform quantizers efficient microphone configurations and rate allocations were found, meaning that the proposed algorithm can find an optimal rate allocation across frequency and microphones. The proposed methods were evaluated based on the MSE performance gap between the monaural configuration and the rate-constrained generalized binaural setup. The efficiency of the proposed method is demonstrated in simulation experiments with an example acoustic scene.

REFERENCES

[1] M. Brandstein and D. Ward (Eds.), *Microphone Arrays*, New York, NY, USA: Springer, 2001.

[2] P. Vary and R. Martin, *Digital Speech Transmission: Enhancement, Coding And Error Concealment*, John Wiley and Sons, 2006.

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

[4] R. Sockalingam, M. Holmberg, K. Eneroth, and M. Shulte, "Binaural hearing aid communication shown to improve sound quality and localization," The Hearing Journal, vol. 62, no. 10, pp. 46–47, 2009.

[5] T. J. Klasen, T. Van den Bogaert, M. Moonen, and J. Wouters, "Binaural noise reduction algorithms for hearing aids that preserve interaural time delay cues," IEEE Transactions on Signal Processing, vol. 55, no. 4, pp. 1579–1585, 2007.

[6] A. I. Koutrouvelis, R. C. Hendriks, R. Heusdens, and J. Jensen, "Relaxed binaural LCMV beamforming," IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 25, no. 1, pp. 137–152, 2017.

[7] 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, no. 3, pp. 543–558, 2016.

[8] S. Srinivasan and A. den Brinker, "Rate-constrained beamforming in binaural hearing aids," EURASIP Journal on Advances in Signal Processing, pp. 1–9, 2009.

[9] O. Roy and M. Vetterli, "Rate-constrained collaborative noise reduction for wireless hearing aids," IEEE Transactions on Signal Processing, vol. 57, no. 2, pp. 645–657, 2009.

[10] O. Roy and M. Vetterli, "Collaborating hearing aids," in Proceedings of MSRI Workshop on Mathematics of Relaying and Cooperation in Communication Networks, 2006.

[11] Y. Shoham and A. Gersho, "Efficient bit allocation for an arbitrary set of quantizers," IEEE Transactions on Acoustics, Speech, and Signal Processing, vol. 36, no. 9, pp. 1445–1453, 1988.

[12] K. Ramchandran and M. Vetterli, "Best wavelet packet bases in a ratedistortion sense," IEEE Transactions on Image Processing, vol. 2, no. 2, pp. 160–175, 1993.

[13] P. Prandoni and M. Vetterli, "R/D optimal linear prediction," IEEE Transactions on Speech and Audio Processing, vol. 8, no. 6, pp. 646– 655, 2000.

[14] Z. Xiong, K. Ramchandran, C. Herley, and M. T. Orchard, "Flexible tree-structured signal expansions using time-varying wavelet packets," IEEE Transactions on Signal Processing, vol. 45, no. 2, pp. 333–345, 1997.

[15] 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, no. 5, pp. 442–448, 1977.

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

[17] J. Kominek, A. W. Black, and V. Ver, "CMU ARCTIC databases for speech synthesis," Tech. Rep., 2003.

[18] E. A. P. Habets, "Room impulse response generator," https://www.audiolabs-erlangen.de/fau/professor/habets/software rirgenerator/, 2010.

6

Rate-Constrained Noise Reduction in Wireless Acoustic Sensor Networks

This chapter is published as "Rate-Constrained Noise Reduction in Wireless Acoustic Sensor Networks," by J. Amini, R. C. Hendriks, R. Heusdens, M. Guo and J. Jensen, in IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 28, pp. 1-12, 2020.

Wireless acoustic sensor networks (WASNs) can provide increased spatial diversity [1, 2], leading to better noise reduction performance compared to single-microphone noise reduction systems. As a realistic example, consider binaural hearing aids (HAs), potentially extended with additional assistive devices, collaborating with each other thro ugh a wireless link [3]. Thanks to the increased number of microphones as well as the increased spatial diversity, they can enhance the speech intelligibility and quality for hearing-impaired listeners [4, 5]. This can be achieved by performing the noise reduction (estimation) process in a distributed way, e.g., [6–8] or by aggregating the microphone observations of the network nodes at a fusion center (FC) followed by estimation of the source of interest and suppression of the environmental noise. In the case of an FC, in practice, one of the nodes in the network (e.g., one of the HAs) could be selected as the FC.

One common approach for noise reduction is the multi-channel Wiener filter (MWF) [9], which is the linear minimum mean square error (MMSE) estimator [10, 11]. Although the original typical MWF considers situations where all microphones are integrated into the same device, many examples exist, where the microphones are distributed over multiple wirelessly connected devices. A well-known example is the binaural MWF [11–14], where the microphone recordings of both HAs are combined to calculate two target signal estimates, one for each ear of the user. Another more general example can be found in [15] where an MWF-based filter is proposed for spatially distributed microphones. Note that in all these methods, the microphone signals are assumed to be available error free at the fusion center.

To limit the scope of this work, we consider the situation where the processing of the microphone signals in the WASN is performed in an FC. To combine the observations at the FC, the actual (realization of the) microphone signals must be transmitted to the FC. As the transmission powers of the devices may be limited due to limited battery life-time, the data needs to be compressed/quantized at a certain data rate. The process of quantization, however, introduces errors in the representation of the microphone signals, and therefore errors in the final target signal estimation. This introduces a trade-off between the data rate and the estimation accuracy (or error) [16], which links the noise reduction problem to the data compression problem.

Several rate-constrained beamforming (noise reduction) algorithms have been introduced in the literature to consider the rate of transmission as a resource constraint in the beamforming process, e.g., [16–19]. Assuming all sources to be jointly Gaussian random processes and using Wyner-Ziv coding [20, 21], a binaural rate-constrained beamformer has been proposed in [17, Sec. III-A]. This beamformer is limited to two devices (i.e., two HAs), which efficiently trades off the data rate against the beamforming performance. The method inevitably assumes that the joint statistics (for example cross-correlations) between the two HAs are known in both devices, which is limiting in practice. Moreover, an infinitely long sequence with a sophisticated decoder is needed to implement the proposed framework, which essentially provides a bound on the possible performance. Finally, this method is limited to the case of only two processing nodes (potentially with multiple microphones per node). The more generalized setup, which may include assistive devices is not considered in this method. Unlike [17, Sec. III-A], sub-optimal rate-constrained beamformers have been proposed in [17, Sec. III-B], [16, 18, 19], which

do not suffer from the requirement that the joint statistics should be known. Typically, these approaches also only consider two collaborating devices. Although these methods are simpler and computationally less expensive than [17, Sec. III-A], they combine all the observations from one device (HA), say, device A, into a single-channel observation, without considering the correlation of the HA observations with the observations from the other HA, say device B, and transmit it to the other device (which serves as an FC). With such a sub-optimal combination, important information may get lost and the performance does not approach the optimal performance, not even asymptotically, at infinitely high data rates [16]. In fact, due to the local combination of the multiple realizations into a single realization, the acoustic scene dependency is not taken into account in the existing sub-optimal approaches.

Assuming the WASN consists of more than two devices (e.g., two hearing aids and multiple additional assistive devices), in this work, we obtain a generalized rate- constrained noise reduction formulation, which can be interpreted as a chief executive officer (CEO) problem (as in information theory), first introduced in [22]. The FC can be thought of as a CEO and the microphones as agents. Each agent records a version of the signal of interest to be transmitted to the FC. As the devices in the WASN have limited battery life-time, and that the power usage is proportional to the data rate (measured in bits) [23], there will be a limited bit rate available for transmitting/receiving the information to/from the agents. Agents should be prioritized (for the estimation task) based on the importance of the information they may have about the target signal. In addition, in our setup, as microphone signals may have generally non-flat power spectral densities, the rate-constrained estimation problem should be frequency dependent. Therefore, depending on the acoustic scene, it is reasonable to share the total data rate across different agents and different frequency components. In [24] a similar problem is studied for rate allocation and strategy selection in an operational rate-constrained beamforming task, given discrete sets of strategy candidates and operating rates. The method uses a discrete optimization algorithm, based on the Lagrange multiplier technique [25], to select the best candidates and operating rates in different frequencies. However, because of the discrete nature of the optimization problem, an exhaustive search is necessary for the rate allocation across agents, which is practically affordable only for a small-size microphone array.

In this chapter, we propose a joint quantization-estimation algorithm for the rate-constrained noise reduction task. We consider a linear estimation task at the FC and propose an optimization problem to both, allocate the total bit rate budget to different microphones in different frequencies (i.e., the quantization part), as well as to find the best filter weights (i.e., the estimation part), minimizing a rate-constrained estimation error. Unlike [24] which treated the problem sequentially with separate quantization and estimation tasks, in this work we consider the joint quantization-estimation problem. Moreover, unlike the exhaustive search for rate allocation across microphones proposed in [24], which is only good for small microphone arrays, we propose to optimize the rate allocations across frequency and space (i.e., devices). The proposed solution is scalable to arbitrarily big microphone arrays. For an MSE criterion, under certain assumptions, the optimal weights are found to be rate-constrained Wiener filter coefficients and the optimal rate allocation is the solution to a reverse "water-filling" problem.

An MSE-based performance measure and an instrumental speech intelligibility measure are used to evaluate the proposed framework and the proposed method outperforms equal/random rate allocation strategies. Moreover, the proposed method performs almost as good as the optimal non-polynomial discrete optimization that involves the infeasible exhaustive search [24], in most practical scenarios.

The chapter is organized as follows. In Sec. 6.1.1 the acoustical signal model is stated and the linear estimation task is introduced in Sec. 6.1.2. The quantization aware beamforming problem is introduced in Sec. 6.1.3. In Sec. 6.1.4 the proposed rate-constrained noise reduction problem formulation is presented in a unified framework and the proposed solution is described in Sec. 6.2. The performance analysis of the proposed and existing methods is carried out in Sec. 6.3. Finally, Sec. 6.4 concludes the chapter.

6.1. PROBLEM STATEMENT

6.1.1. SIGNAL MODEL

We consider a microphone array consisting of M microphones, assumed to be embedded in different devices (i.e., HAs and/or assistive devices) placed at potentially different locations in space. Devices (agents) only communicate with an FC (and not with each other). Only the FC has access to the joint statistics. Each device can be equipped with more than one microphone. In this chapter, it is assumed that for each device, the unprocessed microphone signals will be transmitted to the FC without pre-filtering stages, i.e., the microphone signals per device are not combined (pre-filtered) to a single signal. All microphones capture, in addition to the interferers, their version of the target speech signal, filtered by the acoustic channel, which is characterized by the room impulse response. In the short-time frequency transform (STFT) domain, we denote the target signal by $S_i \in \mathbb{C}$, with i the discrete frequency bin index. For notational convenience, the time-frame index is left out. The target speech is degraded by interfering noise, which might originate from, e.g., interfering point sources, diffuse noise, and/or microphone self-noise. The interfering noise observed at a particular microphone and at a particular frequency is indicated by $N_{ij} \in \mathbb{C}$, with $j = 1, \ldots, M$ being the microphone index. The signals S_i and N_{ij}, are assumed to be additive and mutually uncorrelated. Therefore, the microphone signal model can be written as

$$Y_{ij} = A_{ij} S_i + N_{ij} \in \mathbb{C}, \tag{6.1}$$

where $A_{ij} \in \mathbb{C}$ is the acoustic transfer function (ATF) between the target signal and the jth microphone. The signal model can be rewritten in vector notation by stacking all microphone signals in a vector, as

$$\mathbf{y}_i = \mathbf{a}_i S_i + \mathbf{n}_i = \mathbf{x}_i + \mathbf{n}_i \in \mathbb{C}^M, \tag{6.2}$$

where

$$\mathbf{y}_i = [Y_{i1}, ..., Y_{iM}]^{\mathrm{T}},$$

and similarly for $\mathbf{a}_i$ and $\mathbf{n}_i$, where the superscript $(\cdot)^{\mathrm{T}}$ denotes the transpose operator on vectors/matrices. Since the signals S_i and N_{ij} are assumed to be uncorrelated, the power spectral density (PSD) matrix $\mathbf{\Phi}_{\mathbf{y}_i} = \mathrm{E}[\mathbf{y}_i \mathbf{y}_i^{\mathrm{H}}]$ of the vector $\mathbf{y}_i$ is given by

$$\mathbf{\Phi}_{\mathbf{y}_i} = \mathbf{\Phi}_{\mathbf{x}_i} + \mathbf{\Phi}_{\mathbf{n}_i} \in \mathbb{C}^{M \times M}, \tag{6.3}$$

where

$$\mathbf{\Phi}_{\mathbf{x}_i} = \mathrm{E}[\mathbf{x}_i \mathbf{x}_i^{\mathrm{H}}] = \Phi_{S_i} \mathbf{a}_i \mathbf{a}_i^{\mathrm{H}}, \quad \mathbf{\Phi}_{\mathbf{n}_i} = \mathrm{E}[\mathbf{n}_i \mathbf{n}_i^{\mathrm{H}}], \tag{6.4}$$

with $\Phi_{S_i} = \mathrm{E}[|S_i|^2] \in \mathbb{R}$ the PSD of the clean speech, and $\mathrm{E}[\cdot]$ the expectation operator. The conjugate transpose operator on complex vectors/matrices is indicated by the superscript $(\cdot)^{\mathrm{H}}$.

6.1.2. LINEAR ESTIMATION TASK

One way to increase speech intelligibility and quality of noisy signals is spatial filtering. The goal is to estimate the signal of interest at the FC by combining all the noisy observations into one single signal, such that a fidelity criterion is satisfied. In this chapter, we consider linear estimation, i.e., S_i is estimated as $\hat{S}_i = \mathbf{w}_i^{\mathrm{H}} \mathbf{y}_i \in \mathbb{C}$, with $\mathbf{w}_i \in \mathbb{C}^M$ the weight vector. Minimizing the MSE, the best linear MSE estimator weights, say $\mathbf{w}_i^\star$, are given by the MWF [10]

$$\mathbf{w}_i^\star = \mathbf{\Phi}_{\mathbf{y}_i}^{-1} \mathbf{\Phi}_{\mathbf{y}_i S_i}, \quad i = 1, \ldots, F, \tag{6.5}$$

where F is the number of frequency bins and $\mathbf{\Phi}_{\mathbf{y}_i S_i} \in \mathbb{C}^M$ is the CPSD vector between the observation vector $\mathbf{y}_i$ and the source S_i, which is given by $\mathrm{E}[\mathbf{y}_i S_i^*] = \mathbf{a}_i \mathrm{E}[|S_i|^2]$. The superscript $(\cdot)^*$ denotes the conjugate operator. Therefore, the optimal estimate, denoted by $\hat{S}_i^\star$, is given by $\hat{S}_i^\star = \mathbf{w}_i^{\star \mathrm{H}} \mathbf{y}_i$. Finally, the minimum MSE is computed as

$$D = \frac{1}{F} \sum_{i=1}^{F} \mathrm{E}[|S_i - \hat{S}_i^\star|^2] = \frac{1}{F} \sum_{i=1}^{F} \Phi_{d_i}, \tag{6.6}$$

with

$$\begin{aligned} \Phi_{d_i} &= \mathrm{E}[|S_i - \hat{S}_i^\star|^2] \\ &= \mathrm{E}[|S_i - \mathbf{w}_i^{\star \mathrm{H}} \mathbf{y}_i|^2] \\ &= \Phi_{S_i} - \mathbf{\Phi}_{\mathbf{y}_i S_i}^{\mathrm{H}} \mathbf{\Phi}_{\mathbf{y}_i}^{-1} \mathbf{\Phi}_{\mathbf{y}_i S_i}, \quad i = 1, \ldots, F. \end{aligned}$$

To compute the MWF output $\hat{S}_i^\star$, the noisy signal realizations should be available error-free at the FC. In practice, only a compressed/quantized version of the contralateral noisy signals are available. These signals are compressed at a certain rate, say r_{ij} bits per sample (bps). This leads to a modified signal model including quantization noise, as explained in the next subsection.

6.1.3. QUANTIZATION AWARE BEAMFORMING

As mentioned in the previous part of this section, the microphone signals are compressed prior to transmission to the FC. In this chapter, we assume that the signals are being quantized using a uniform quantizer, which will be briefly explained in the following.

Let us consider an arbitrary signal x that is quantized, and the quantized version is denoted by $\tilde{x}$, with quantization noise $e = x - \tilde{x}$. Under high bit rate assumptions or by applying subtractive dithering to the signal to be quantized (at lower rates) [26, 27], the quantization error (noise) e will be uncorrelated to the signal x and will be uniformly distributed with variance $\sigma_e^2 = \frac{\Delta^2}{12}$. Here $\Delta = \frac{2x_{\max}}{2^r}$ is a step size, which depends on the range of the signal (maximum absolute value $x_{\max}$) and the quantization rate r. Applying

this to the beamforming task, the quantization noise is taken into account and the signal model in (6.1) can be modified as

$$\tilde{Y}_{ij} = Y_{ij} + E_{ij} = A_{ij}S_i + N_{ij} + E_{ij} \in \mathbb{C}, \tag{6.7}$$

where $\tilde{Y}_{ij}$ is the quantized noisy signal and E_{ij} is the quantization noise. Similar to (6.2), using vector notation, we then have

$$\tilde{\mathbf{y}}_i = \mathbf{y}_i + \mathbf{e}_i = \mathbf{a}_i S_i + \mathbf{n}_i + \mathbf{e}_i \in \mathbb{C}^M, \tag{6.8}$$

where the quantization noise vector $\mathbf{e}_i = [E_{i1}, E_{i2}, \cdots, E_{iM}]^{\mathrm{T}}$ is assumed to be uncorrelated to the microphone signal vector $\mathbf{y}_i$, which is valid under the above-mentioned assumptions [26, 27]. Therefore, the CPSD matrix of the quantization noise vector $\mathbf{e}_i$ will be diagonal with elements

$$\Phi_{E_{ij}} = \frac{\Delta^2}{12} = \frac{(Y_{ij}^{\max})^2}{3 2^{2r_{ij}}} = \frac{k_{ij}}{2^{2r_{ij}}}, \tag{6.9}$$

where $k_{ij} = \frac{(Y_{ij}^{\max})^2}{3}$. At the FC, the signal of interest S_i is estimated, given the compressed noisy microphone signals $\tilde{\mathbf{y}}_i$, as

$$\hat{S}_i = \mathbf{w}_i^{\mathrm{H}} \tilde{\mathbf{y}}_i. \tag{6.10}$$

The estimator $\hat{S}_i$ is a function of the estimation parameters $\mathbf{w}_i$ and the rates r_{ij}. In the next part of this section, we will propose a problem formulation to address the problem of finding the above-mentioned parameters, by minimizing the estimation error.

6.1.4. RATE-DISTORTION TRADE-OFF IN NOISE REDUCTION PROBLEMS

As argued in the previous part of this section, at the FC, signals are available at a certain operating rate, say r_{ij} (bps). In fact, the receiver at the FC has a limited total capacity, say R_{tot}, due to limitations on transmission capabilities, to communicate with its agents [22] (here, microphones). Depending on this resource R_{tot} and the actual acoustic scene, different rate allocations across frequency and space are optimal [24]. In this work, we address the problem of rate-constrained noise reduction in order to find the optimal rate allocation to each microphone signal at each specific frequency bin. We propose the following joint quantization-estimation problem.

PROPOSED PROBLEM FORMULATION

We are given a set of operating rates $\mathcal{Q} = \{\mathbf{R} \,|\, 0 \leq r_{ij} \leq \infty\}$, where the matrix

$$\mathbf{R} = \begin{bmatrix} r_{11} & r_{12} & \dots & r_{1M} \\ r_{21} & r_{22} & \dots & r_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ r_{F1} & r_{F2} & \dots & r_{FM} \end{bmatrix} \in \mathbb{R}^{F \times M}.$$

includes rates r_{ij} to be allocated to each frequency bin i and microphone j. Let the distortion function $D(\mathbf{R})$ be defined as the averaged (over frequency) power spectral density

of the estimation error, given the rates, that is

$$D(\mathbf{R}) = \frac{1}{F}\sum_{i=1}^{F} d(\mathbf{r}_i), \tag{6.11}$$

where

$$d(\mathbf{r}_i) = \mathrm{E}[|S_i - \hat{S}_i|^2 | \mathbf{r}_i], \quad \mathbf{r}_i \in \mathbb{R}^M,$$

denotes the PSD of the estimation error at the ith discrete frequency bin, given the rate vector $\mathbf{r}_i = [r_{i1}, \ldots, r_{iM}]^{\mathrm{T}}$, which is the ith row of the matrix $\mathbf{R}$ and includes the rates allocated to the different microphones for the specific frequency i. Furthermore, let $R(\mathbf{R})$ simply be defined as the sum-rate over all bins and microphones, given by

$$R(\mathbf{R}) = \sum_{i=1}^{F}\sum_{j=1}^{M} r_{ij}. \tag{6.12}$$

Then, the problem is defined as minimizing the estimation error, while satisfying the total budget R_{tot} on the rates. That is

$$\begin{aligned} \min_{\mathbf{R}\in\mathcal{Q}} \quad & D(\mathbf{R}) \\ \text{subject to} \quad & R(\mathbf{R}) \le R_{\mathrm{tot}}. \end{aligned} \tag{6.13}$$

Assuming that the joint statistics are known only at the FC, and using (6.8) and (6.10), the distortion function $d(\mathbf{r}_i)$ can be further parameterized as a function of the estimator weights $\mathbf{w}_i$ as

$$\begin{aligned} d(\mathbf{r}_i, \mathbf{w}_i) &= \mathrm{E}[|S_i - \hat{S}_i|^2 | \mathbf{r}_i] \\ &= \mathrm{E}[|S_i - \mathbf{w}_i^{\mathrm{H}} \tilde{\mathbf{y}}_i|^2 | \mathbf{r}_i] \\ &= \mathrm{E}[|S_i - \mathbf{w}_i^{\mathrm{H}} \mathbf{a}_i S_i - \mathbf{w}_i^{\mathrm{H}} \mathbf{n}_i - \mathbf{w}_i^{\mathrm{H}} \mathbf{e}_i|^2 | \mathbf{r}_i] \\ &= |1 - \mathbf{w}_i^{\mathrm{H}} \mathbf{a}_i|^2 \Phi_{S_i} + \mathbf{w}_i^{\mathrm{H}} \mathbf{\Phi}_{\mathbf{n}_i} \mathbf{w}_i + \mathbf{w}_i^{\mathrm{H}} \mathbf{\Phi}_{\mathbf{e}_i}(\mathbf{r}_i) \mathbf{w}_i. \end{aligned} \tag{6.14}$$

The diagonal matrix $\mathbf{\Phi}_{\mathbf{e}_i}(\mathbf{r}_i)$ is the CPSD matrix of the quantization noise with elements given by (6.9). Based on (6.9) and the fact that $\mathbf{\Phi}_{\mathbf{e}_i}(\mathbf{r}_i)$ is diagonal, the distortion function $d(\mathbf{r}_i, \mathbf{w}_i)$ can be rewritten as

$$d(\mathbf{r}_i, \mathbf{w}_i) = |1 - \mathbf{w}_i^{\mathrm{H}} \mathbf{a}_i|^2 \Phi_{S_i} + \mathbf{w}_i^{\mathrm{H}} \mathbf{\Phi}_{\mathbf{n}_i} \mathbf{w}_i + \sum_{j=1}^{M} \frac{|w_{ij}|^2 \, k_{ij}}{2^{2\, r_{ij}}}. \tag{6.15}$$

We define the weight matrix $\mathbf{W} \in \mathbb{C}^{F\times M}$ as

$$\mathbf{W} = \begin{bmatrix} \mathbf{w}_1^{\mathrm{T}} \\ \mathbf{w}_2^{\mathrm{T}} \\ \vdots \\ \mathbf{w}_F^{\mathrm{T}} \end{bmatrix} = \begin{bmatrix} w_{11} & w_{12} & \ldots & w_{1M} \\ w_{21} & w_{22} & \ldots & w_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ w_{F1} & w_{F2} & \ldots & w_{FM} \end{bmatrix} \in \mathbb{C}^{F\times M},$$

i.e., the ith row of $\mathbf{W}$ contains the beamformer coefficients for frequency bin i. Substituting (6.15) into (6.11), and then into the original problem formulation (6.13), the

reformulated problem can be rewritten as

$$\begin{aligned} \min_{\mathbf{R},\mathbf{W}} \quad & \frac{1}{F}\sum_{i=1}^{F}\left(|1-\mathbf{w}_i^{\mathrm{H}}\mathbf{a}_i|^2\Phi_{S_i}+\mathbf{w}_i^{\mathrm{H}}\mathbf{\Phi}_{\mathbf{n}_i}\mathbf{w}_i+\sum_{j=1}^{M}\frac{|w_{ij}|^2\,k_{ij}}{2^{2\,r_{ij}}}\right) \\ \text{s.t.} \quad & \sum_{i=1}^{F}\sum_{j=1}^{M} r_{ij} \le R_{\text{tot}}, \\ & r_{ij} \ge 0. \end{aligned} \tag{6.16}$$

Note that the estimation error function in (6.15) includes three terms: 1) the target signal distortion, i.e., $|1-\mathbf{w}_i^{\mathrm{H}}\mathbf{a}_i|^2\Phi_{S_i}$ 2) the residual noise power, i.e., $\mathbf{w}_i^{\mathrm{H}}\mathbf{\Phi}_{\mathbf{n}_i}\mathbf{w}_i$ and 3) the residual quantization noise, i.e., $\sum_{j=1}^{M}\frac{|w_{ij}|^2\,k_{ij}}{2^{2\,r_{ij}}}$. The first two terms are only functions of the weights and the last term is jointly a function of both the weights and the quantization rates. In fact, as the last term in (6.15) is a summation of "quadratic-over-nonlinear" functions, which are non-convex functions, the problem in (6.16) is a non-convex optimization problem. However, fixing $\mathbf{W}$ or $\mathbf{R}$, the problem will be convex in the remaining variable (component-wise convex).

6.2. PROPOSED SOLUTION

In the following, we propose a solution to the non-convex problem in (6.16), presented in the previous section. The third term in (6.15), which is a summation of "quadratic-over-nonlinear" functions, causes the non-convexity in the objective function. Nevertheless, we can write the necessary Karush-Kuhn-Tucker (KKT) conditions [28] for the problem in (6.16) to find the necessary optimality conditions. It can be shown (see Appendix 6-A) that the solution to (6.16) lies on the boundary of the feasibility set defined by the global budget constraint (first constraint in (6.16)). As a consequence, we can replace the inequality constraint on the total bit budget by an equality constraint. With this, the Lagrangian function is given by

$$\begin{aligned} L(\mathbf{R},\mathbf{W},\lambda,\mathbf{V}) = & \frac{1}{F}\sum_{i=1}^{F}[\,|1-\mathbf{w}_i^{\mathrm{H}}\mathbf{a}_i|^2\Phi_{S_i}+\mathbf{w}_i^{\mathrm{H}}\mathbf{\Phi}_{\mathbf{n}_i}\mathbf{w}_i \\ & +\sum_{j=1}^{M}\frac{|w_{ij}|^2\,k_{ij}}{2^{2\,r_{ij}}}\,]+\lambda\Big(\sum_{i=1}^{F}\sum_{j=1}^{M} r_{ij}-R_{\text{tot}}\Big)-\sum_{i=1}^{F}\sum_{j=1}^{M} v_{ij}r_{ij}, \end{aligned} \tag{6.17}$$

where the matrix $\mathbf{V}\in\mathbb{R}^{F\times M}$ consists of non-negative entries v_{ij} which denote the Lagrangian multipliers, responsible for the element-wise non-negativity constraints, i.e., $r_{ij}\ge 0$. The Lagrangian multiplier λ is to assure the total rate constraint is met with equality.

In the following proposition, the solution to the KKT conditions w.r.t. the problem in (6.16) and the Lagrangian equation (6.17) is given as a system of equations.

Proposition. *Minimizing the constrained problem in (6.16) based on the Lagrangian*

function in (6.17), the parametric optimal weights and the optimal rates are given as

$$\begin{cases} 1) & \mathbf{w}_i^\star(\mathbf{r}_i^\star) = \mathbf{\Phi}_{\tilde{\mathbf{y}}_i}^{-1}\mathbf{\Phi}_{\tilde{\mathbf{y}}_i S_i}(\mathbf{r}_i^\star), \\ 2) & r_{ij}^\star(\lambda'^\star, w_{ij}^\star) = \max(\frac{1}{2}\log_2(\frac{|w_{ij}^\star|^2\, k_{ij}}{\lambda'^\star}), 0), \end{cases} \tag{6.18}$$

where $i = 1,\ldots,F$, $j = 1,\ldots,M$, *and* $\lambda'^\star = \frac{\lambda^\star}{2\,ln2}$ *is a parameter, which satisfies the equality constraint*

$$\sum_{i=1}^{F}\sum_{j=1}^{M} r_{ij}(\lambda'^\star) = R_{\text{tot}}.$$

Proof. See Appendix 6-A. □

Note that the rates are zero-valued for $\lambda'^\star \geq |w_{ij}^\star|^2\, k_{ij}$. The operator $\max(\cdot,0)$ assures that the rates are non-negative, satisfying the second set of inequality constraints in (6.16).

Looking at the system of equations in (6.18), the optimal weights $\mathbf{w}_i^\star$ are the rate-dependent multi-channel Wiener filter coefficients (first set of equations) and the optimal rates $r_{ij}^\star$ are the solution to the weighted reverse water-filling problem. In fact, the set of Wiener equations are responsible for the target estimation part and the rate equation for the quantization part (rate allocation). It is clear from (6.18) that the rate allocation is done across both frequencies and microphones, depending on both the microphone signal power (which is related to k_{ij}) and the contribution of components to the estimation process (which is related to $|w_{ij}^\star|^2$). The frequencies and devices that contribute most to the target estimation will be allocated more bits. Similar to the classical water-filling problems [23, 29], the components for which $|w_{ij}^\star|^2\, k_{ij} \leq \lambda'^\star$ will be allocated zero bits.

One way to solve (6.18) is to apply alternating optimization [30]. First, the rates are initialized as $\mathbf{R}^0$, for example by an equal rate allocation where all components start to be allocated equal rates. Second, the optimal weight functions are computed, given $\mathbf{R}^0$, to find the updated weight matrix $\mathbf{W}^1$, where $\mathbf{W}^n$ denotes the updated matrix variable at nth iteration. Then the updated weights $\mathbf{W}^1$ are used to compute the updated rates $\mathbf{R}^1$. In this way, the equations are computed iteratively until a certain stopping criterion is met. As explained in Sec. 6.1.4, since the objective function in (6.16) is component-wise convex in the variables $\mathbf{W}$ and $\mathbf{R}$, as argued in [30, 31], any limit point (solution after sufficient iterations) is a critical point. Note that since the objective function is not jointly convex in $\mathbf{W}$ and $\mathbf{R}$, this critical point is not necessarily globally optimal. However, as confirmed by the simulation experiments in Sec. 6.3, the performance of the proposed method is almost as good as the (non-tractable) exhaustive search (for rate allocation across microphones) [24], for some representative example acoustic scenarios.

6.3. PERFORMANCE EVALUATION

In this section, we perform simulations in several example acoustical scenarios to evaluate the performance of the proposed and existing approaches, as a function of the total communication rate R_{tot}.

In addition to predicted intelligibility by means of the short-time objective intelligibility (STOI) measure [32], we use the performance measure introduced in [17] and [16], which is defined as the ratio of the target signal estimation MSE, when there is no communication between the agents and the FC, say $D(0)$, to the MSE when the data is quantized before transmission, say $D(R)$. The output gain with respect to the beamformer (FC) is given by

$$G_{\mathrm{FC}}(R) = \frac{D(0)}{D(R)}, \tag{6.19}$$

where $D(\cdot)$ is the MSE introduced in (6.11). $D(0)$ denotes the distortion when the devices do not communicate with the FC ($R_{\mathrm{tot}} = 0$). In this case, the distortion is computed based on the local observations at the FC only.

6.3.1. EXAMPLE GENERALIZED BINAURAL HA SETUP

The first example acoustic scene is illustrated in Fig. 6.1. The binaural HA system includes two HA microphones (one per HA), denoted by the black "+" symbols, and are located with a distance of 10 cm w.r.t. the origin ($(x_o, y_o) = (0,0)$), along the horizontal x-axis. The green circle indicates the target speech source, located in front of the HA system ($\theta = 0°$), at a distance of 3 m from the origin. In this work, the location angles are computed counter-clockwise starting from the look direction. There is an assistive wireless microphone in this setup which is denoted by the blue "×" symbol, placed closer to the target speech at an angle $\theta = 15°$ and a distance of 2.8 m from the origin. The black triangle indicates the interfering signal, located at a distance of 3 m from the origin at an angle $\theta = -80°$, with a signal-to-interferer ratio (SIR) of 0 dB. In addition, simulated internal microphone noise is added to the microphone signals. The internal noise is assumed to be uncorrelated across microphones and is added with a signal-to-noise ratio (SNR) of 40 dB w.r.t. the target signal at the reference point.

In this experiment, without loss of generality, the FC is chosen to be the left side HA. Therefore, the left side microphone signal is considered as the reference local observation and the two other microphone signals as the agents' observations. The PSD of the target speech Φ_S is estimated based on Welch's method, using a 512-points discrete Fourier transform (DFT), computed frame-by-frame from 50% overlapping speech frames, using around 10 s of the $F_s = 16$ kHz sampled speech signals taken from the "CMU-ARCTIC" database [33]. A flat PSD $\Phi_{n_1}(\omega)$ over the interval $\omega \in [-\pi, \pi]$ is assumed for the point noise source (interfering signal). Under the free-field assumption, the ATFs are generated using Habets' model [34], in a non-reverberant environment. The non-reverberant environment is chosen to get a more clear understanding of the effect of the number and location of the point noise sources on rate allocation behavior. Finally, the generated ATFs and the estimated PSDs are used to calculate the corresponding cross PSD matrices.

Based on the setup, the performance of the following approaches are compared in this section:

- **Equal Rate Allocation (2 Mics)**: Only the left-side and the right-side microphones (two microphones in total) are selected in this case (and thus not the assistive microphone). Therefore, there is only one microphone signal (from the right HA)

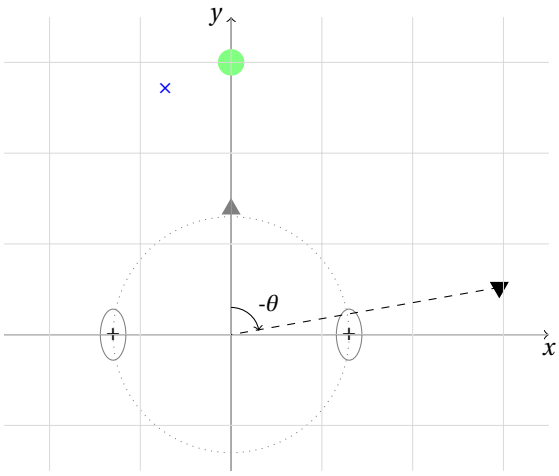


Figure 6.1: Typical acoustic scene. The two HA microphones, the assistive microphone, the target signal, and the interferer are indicated by the black "+", the blue "×", the green circle, and the black triangle, respectively.

which needs to be quantized before transmission. In this case, the rates are equally allocated over all frequencies.

- **Equal Rate Allocation (3 Mics)**: All three microphones are selected in this case. The rates are assumed to be equally allocated over all frequencies as well as across all microphones.
- **Discrete Optimization OPT [24]**: This method is based on discrete optimization, and optimally allocates the rates over all frequencies and across microphones. Note that, in this method, an exhaustive search is done to find the best allocations across the microphones, which is computationally very expensive and not tractable for big microphone arrays.
- **Proposed (2 Mics)**: The proposed method described in Sec. 6.2. In this case, only the binaural setup (2-Microphone setup) is considered, meaning that the assistive microphone signal is not used. Therefore, the rate allocation is optimized only across frequency.
- **Proposed (3 Mics)**: The proposed method described in Sec. 6.2. In this case, all microphones are used. Therefore, the rate allocation is optimized across both frequency and across microphones.
- **Remote Wyner Ziv (WZ)[17]**: The binaural rate-constrained beamforming presented in [17, Sec. III-A]. Note that only two processing nodes, i.e., in this setup two HAs, can be used in this method, joint statistics are needed at all processors (nodes) and impractical long-block vector quantizers are assumed.

OUTPUT GAINS

In this part, we compare the above-mentioned approaches based on the performance measure in (6.19). Fig. 6.2 shows the output gain G_{FC} in dB as a function of the nor-

malized (over frequency) total bit rate budget. The horizontal dash-dotted line denotes the performance of the 2-microphone MWF [11, 13], based on both the left and right microphone signals. It is assumed here that the right side observation is available (at an infinite rate) at the FC, i.e., without quantization noise. This method serves as a performance bound for the binaural setup. Similarly, the horizontal dashed line denotes the performance of the 3-microphone MWF [11, 13], where all microphone signals are used at an infinite rate. As shown, the performance of all methods approaches to the corresponding horizontal lines, at sufficiently high rates. The proposed method outperforms significantly the equal allocation strategies, as the rate allocation is optimized over frequency. The performance of the remote WZ method is computed based on the theoretical upper bound, described in [17]. As shown, the performance curve of the remote WZ method is upper-bounded by the 2-microphone MWF, as the assistive microphone is not considered in this method.

In this example setup, the proposed (3 Mics) method performs almost as good as the optimal discrete optimization method, which uses an exhaustive search to find the best allocations across microphones. Please note that, based on the complexity analysis which will be explained in Sec. 6.3.3, the computational complexity of the optimal discrete optimization method grows dramatically by increasing the number of the microphones. However, for the setup in Fig. 6.1 (with only three microphones) we could perform the exhaustive search for comparison. On average, the proposed alternating optimization approach needs less than 10 iterations to converge to a solution.

RATE ALLOCATIONS ACROSS FREQUENCY

Based on the results, shown in Fig. 6.2, the rate distribution for each agent as a function of frequency and total bit rate is shown in Fig. 6.3. As shown in Fig. 6.3b, with a very small total rate, only lower frequency components are allocated non-zero rates. The effect of very high-frequency components on the final target estimation is negligible compared to the low-frequency components, as they have small PSD values, and therefore less rate is allocated. As the total rate increases, more high-frequency components can contribute to the estimation process.

Comparing Fig. 6.3a and Fig. 6.3b, for a small total rate, the right side microphone is barely used as the assistive microphone signal contains more information about the target signal (since it is located closer to the target source, based on Fig. 6.1). Therefore, more rate is allocated to the assistive microphone. As the total rate (total budget) increases, the right side microphone starts to contribute to the estimation process on its most important frequency components. The sinusoidal behavior of the rate distribution in Fig. 6.3a (at middle total rate values) is related to the shape of the squared value of the filtering weights ($|w_{ij}|^2$) over frequency.

6.3.2. EXAMPLE GENERAL WASN CONFIGURATION

In this simulation experiment, we consider the second example acoustic scene, illustrated in Fig. 6.4. Five microphones are randomly located in space. The black triangles denote the interferers of which the number and location vary in different scenarios, which will be described later in this section. There is one target speech signal (Green circle) at $(2\,\mathrm{m}, 30°)$. In this section, we consider the following three scenarios.

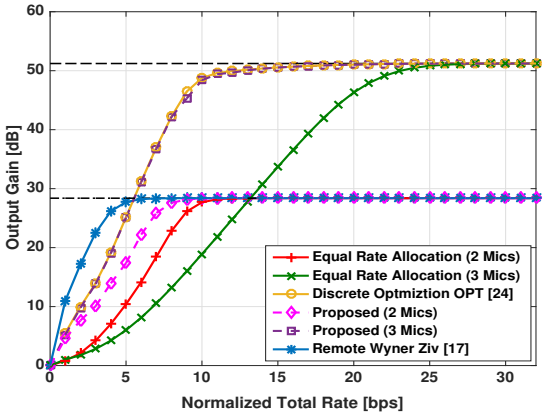
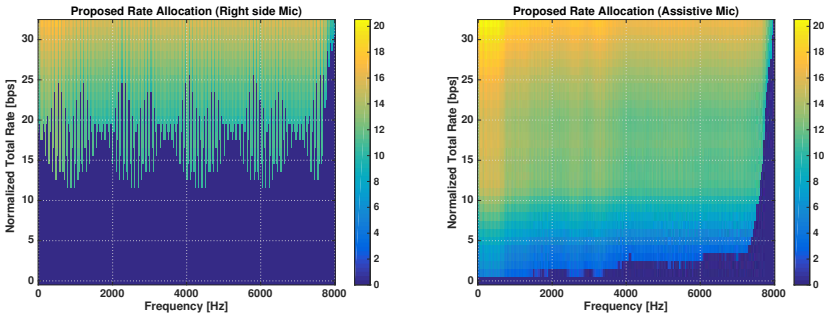


Figure 6.2: Output Gain [dB] versus total rate [bit per sample] based on a generalized binaural setup in Fig. 6.1.



(a) Right-side microphone contribution. (b) Assistive microphone contribution.

Figure 6.3: Rate distributions as a function of frequency and normalized total budget.

- **Scenario 1:** Only one interferer (point noise source).
- **Scenario 2:** Four interferers (point noise sources).
- **Scenario 3:** Four interferers along with diffuse noise.

The FC is assumed to be located at the origin as a reference point (no local observations). For all scenarios, the interfering signals' power is chosen such that the SIR w.r.t. the target signal at the FC is 0 dB. In all experiments, uncorrelated internal noise is added to the microphone signals at 40 dB SNR w.r.t. the FC. For all sources, the ATFs and the power spectral densities are estimated/computed in a similar way as in the previous setup, in a non-reverberant environment.

Based on the setup, shown in Fig. 6.4, the following methods are compared:

- **Discrete Optimization SUB [24]**: This method is based on discrete optimization, and optimally allocates the rates over all frequencies. However, it assumes an

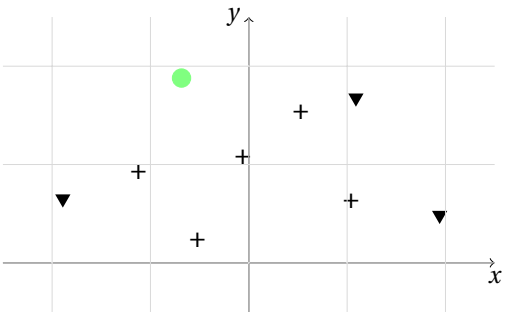


Figure 6.4: An example acoustic scene: a general microphone array is shown by the black "+" symbols.

equal rate allocation across microphones, as the optimal exhaustive search is very expensive and not tractable for big microphone arrays.

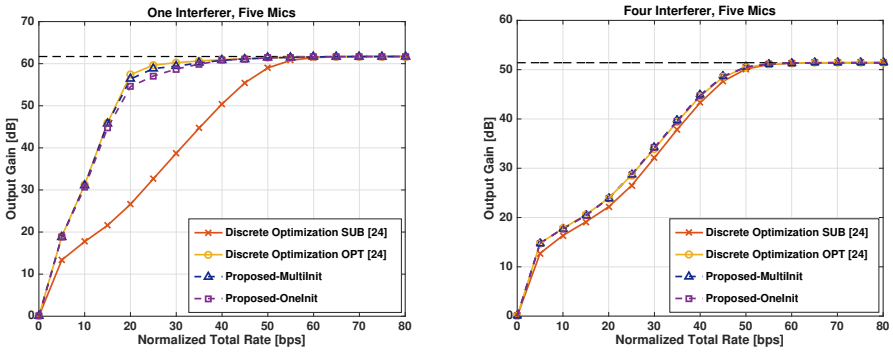
- **Discrete Optimization OPT [24]**: This method is based on discrete optimization, and optimally allocates the rates over all frequencies and across microphones. Based on our experiments and the complexity analysis, described in the Sec. 6.3.3, the exhaustive search used in this approach becomes intractable for more than five microphones.

- **Proposed**: The proposed method described in Sec. 6.2.

CORRELATED POINT NOISE SOURCES

In this case, the scenarios 1 and 2 are considered. Scenario 1 contains only one interferer located at (2 m, −60°). Scenario 2 contains four interferers located at (2 m, {−80°, −60°, 40°, 85°}). Similar to Fig. 6.2, the output gains G_{FC} in dB as a function of total bit rate budget, are shown Fig. 6.5. Please note that at each normalized total bit budget, the budget will be distributed (maximally) across five microphones. For example, if the normalized total budget is 30 bps, it means that on average 30 bps may be allocated across five agents, and not necessarily six bps per agent. The dashed line denotes the performance of the 5-microphone MWF (which is an upper bound on the performance of the MSE-based methods), assuming all microphone signals are available at the FC, without quantization noise.

The proposed algorithm is based on alternating optimization which needs to be initialized. In the proposed-OneInit method, the algorithm is initialized based on reverse water filling on the power of the signals, assuming equal weights for all components. As we are not (theoretically) necessarily guaranteed to converge to the globally optimal solution, in the proposed-MultiInit method, we also test the algorithm with multiple initializations. Initially, the total rate is randomly distributed to the components and the alternating optimization is carried out for each random initializations. The procedure is repeated and the allocation which results in a minimum distortion among all random initializations is selected. The proposed method with multiple initializations is very close, in performance, to the optimal discrete optimization approach. However, even with single initialization (proposed-OneInit) the performance of the proposed-OneInit



(a) Scenario 1: One Interferer

(b) Scenario 2: Four Interferers

Figure 6.5: Output Gain [dB] versus total rate [bit per sample] based on the second setup in Fig. 6.4.

method is not far from the optimal method. As shown in Fig. 6.5a, the proposed method performs significantly better than the sub-optimal discrete optimization method, as the optimal rates are also optimized across the agents. The remote Wyner Ziv approach is not included in the comparison, as it cannot consider more than two nodes, and therefore, it is not suitable for a general WASN setup.

In scenario 2, in Fig. 6.5b, instead of one point source, the scenario contains four interfering point sources. Increasing the number point sources has an interesting effect compared to the case of a single point source as in Fig. 5a. The performance gap between the sub-optimal approach, where the equal rate allocation is done across microphones, and the optimal methods is reduced. This can be explained as follows. Under mild differences in target signal powers captured by microphones, increasing the number of point sources, will reduce the spatial correlation (coherence) factor and makes the microphone signals more equally important in the target estimation process. Furthermore, in this case, all proposed and optimal curves are almost on top of each other, meaning that the proposed method managed to nearly achieve the optimal performance.

DIFFUSE NOISE

In this scenario, there is a simulated diffuse noise along with four interferers. The diffuse noise is simulated as a cylindrical source array around the microphone array, for which the estimated spatial coherence function reasonably resembles the theoretical spatial coherence function between the microphone signals. Four interferers are located at $(2\,\mathrm{m}, \{-80^\circ, -60^\circ, 40^\circ, 85^\circ\})$. The powers of the sources are chosen such that the input signal to point noise and diffuse noise ratio (SIDR) is approximately 0 dB at the FC.

Fig. 6.6 shows the output gains G_{FC} in dB as a function the total bit rate. The results show little difference between all competing methods, as almost the same (power-wise) impression of the environmental noise is received by each agent, and the observations become spatially less correlated. The sub-optimal discrete optimization, which is simple and fast, is therefore a suitable approach in this scenario. All proposed methods and the optimal method are almost on top of each other, and are asymptotically optimal

meaning that the performance approaches that of the 5-microphone MWF method at a sufficiently high rate.

As mentioned in Sec. 6.1.4, the joint statistics need to be known only at the FC. Assuming that the statistics do not change rapidly over the number of consecutive frames, a piece of over-head information, which is needed to inform the agents about their allocated rates, can be averaged out over the frames, and hence, does not affect the proposed solution.

6.3.3. COMPUTATIONAL COMPLEXITY

In this part, we compare the methods from a complexity point of view. The computational complexity of the competing methods in the previous part is listed in Table 6.1, for a given total rate R_{tot}. Variable q denotes the number of all possible choices for the integer bit rate assigned to each frequency. Note that q generally may depend on the number of microphones M so that it may increase by increasing the number of microphones. The set $\mathscr{A}$ includes all possible allocations of the rate across microphones, for each frequency. When computing the cardinality $|\mathscr{A}|$, it is assumed that the rate (per frequency) can vary from zero bit to $(q-1)$ bits. In the optimal discrete optimization method (Discrete Optimization OPT), the exhaustive search is done over the set $\mathscr{A}$ to find the best bit allocation across microphones. In the sub-optimal discrete optimization method (Discrete Optimization SUB), the total bit rate (for each frequency) is distributed equally across microphones, therefore, the exhaustive search is not necessary. The computational complexity of the proposed method is based on (6.18) for K iterations. As shown, the proposed and sub-optimal methods have polynomial complexity order w.r.t. M and F. For the proposed method, for $\log(MF) \gg M^2$ the second term in the complexity order is dominant, therefore, the complexity will be of order $O(MF\log(MF))$ for one iteration ($K=1$). For a small M, the complexity is comparable to that of an FFT (complexity of order $F\log F$). In this case, the proposed method does not have a significant extra complexity, compared to FFT computations, which are unavoidable in frequency-domain noise reduction algorithms.

The complexity (in logarithmic scale) as a function of the number of microphones (M) is shown in Fig. 6.7, for $F = 512$, $q = 32M$, and $K = 15$ iterations over (6.18). As shown, the optimal method is computationally much more expensive than the other two methods. As shown in the simulations in the previous subsections, the proposed method is very close to the optimal method in terms of performance, although with much lower complexity.

In scenarios with highly correlated microphone signals (for example, scenario 1), there is a big performance gain in optimizing rate allocation across microphones (compared to the sub-optimal method). However, in scenarios with multiple sources and diffuse noise, the microphone signals become less correlated implying that the sub-optimal discrete optimization method becomes closer to the optimal discrete optimization method in terms of performance, with lower complexity.

6.3.4. SPEECH INTELLIGIBILITY

In this section, we compare the competing methods in terms of speech intelligibility. Although all competing methods are based on optimizing the MSE criteria (and not based

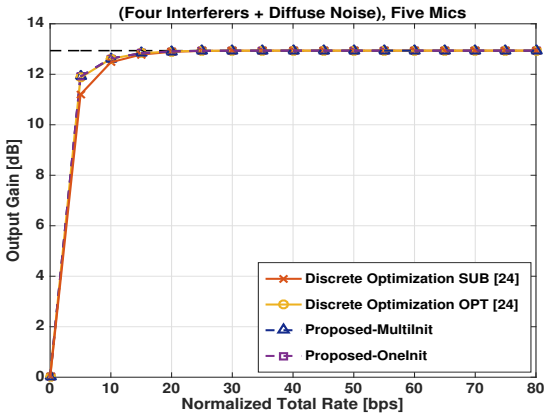


Figure 6.6: Scenario 3: Diffuse noise + four interferers.

Table 6.1: Computational complexity order

Method	Complexity
Discrete Optimization OPT [24]	$O(M^3 F \lvert\mathscr{A}\rvert)$
Discrete Optimization SUB [24]	$O(M^3 F q)$
Proposed	$O(M^3 F K + MF\log(MF) K)$

$$\lvert\mathscr{A}\rvert = \binom{M-1}{M-1} + \binom{M}{M-1} + \ldots + \binom{q+M-2}{M-1}$$

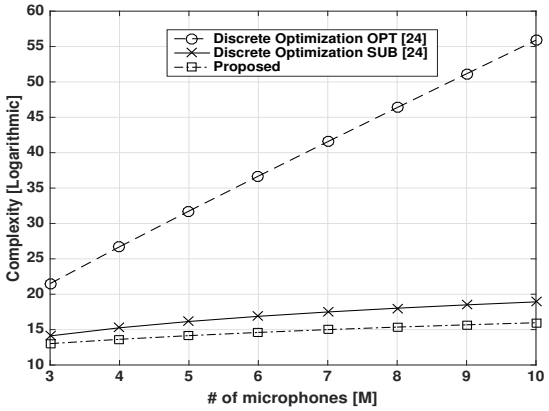


Figure 6.7: Computational Complexity as a function of number of microphones [M].

on speech intelligibility criteria) it is reasonable to see how they affect the speech intelligibility as a function of the bit rate.

In this chapter, we choose the STOI measure [32] to evaluate the proposed method. Scenario 3 (as in the Sec. 6.3.2) is chosen here based on the example acoustic scene

shown in Fig. 6.4, which includes a simulated diffuse noise along with four interferers located at $(2\,\text{m}, \{-80°, -60°, 40°, 85°\})$. The SIDR w.r.t. the FC is set to 0 dB and the SNR is set to 40 dB. Uniformly distributed random realizations are added to the microphone signals as quantization noises. The variances of the quantization noises are computed using the corresponding optimized rate allocations for different methods.

The STOI measure as a function of the total rate is shown in Fig. 6.8. As shown, all curves approach (at high total rates) to the black dashed line which is the asymptotic STOI value when there is no quantization noise. Comparing Fig. 6.8 with Fig. 6.6, in this specific scenario, the STOI gaps between the sub-optimal discrete optimization method and the optimal methods are very low. In fact, under uniform quantization assumptions, small output gain differences between the competing methods at different total rates may not cause significant speech intelligibility gaps. As shown in Fig. 6.8, the proposed method performs as good as the optimal discrete optimization method in terms of the STOI objective measure, at much lower complexity.

6.4. CONCLUSION

In this chapter, we proposed an MMSE-based rate-constrained noise reduction framework in wireless acoustic sensor networks (WASN) to jointly weight the contribution of the remote-microphone signals to the linear estimation task and allocate the bit rates across both frequency and spatial components (microphones). We introduced a joint estimation-compression optimization problem based on a rate-distortion trade-off to constrain the total rate at the fusion center. We proposed a solution to the component-wise convex estimation-compression problem based on alternating optimization. We found that the optimal estimation weights are actually the rate-constrained Wiener coefficients and the optimal rates are solutions to a filter-dependent reverse watering-filling problem. Based on the MSE criterion and the STOI intelligibility criterion, the performance of the proposed method is in most scenarios almost as good as the exhaustive search-based method, with lower complexity.

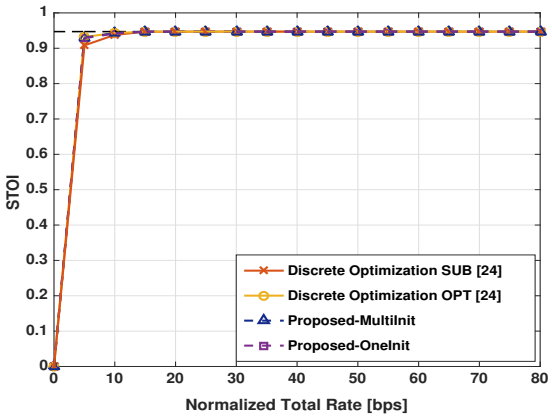


Figure 6.8: STOI as a function of the total rate [bps] for Scenario 3: diffuse noise + four interferers.

APPENDICES

6-A: DERIVATIONS OF THE SOLUTION PROPOSED IN SEC. 6.2 (6.18)

In this section, we derive the necessary equations to solve the optimization problem, introduced in (6.16). Given the Lagrangian objective function in (6.17), the necessary KKT conditions for optimality are then given by

$$L_{\mathbf{w}_i^*} = \mathbf{\Phi}_{\mathbf{x}_i}\mathbf{w}_i - \Phi_{S_i}\mathbf{a}_i + \mathbf{\Phi}_{\mathbf{n}_i}\mathbf{w}_i + \mathbf{\Phi}_{\mathbf{e}_i}\mathbf{w}_i = 0, \tag{6.20a}$$

$$L_{p_{ij}} = \frac{-|w_{ij}|^2\, k_{ij}\, 2\ln 2}{2^{2r_{ij}}} + \lambda - \nu_{ij} = 0, \tag{6.20b}$$

$$\sum_{i=1}^{F}\sum_{j=1}^{M} r_{ij} \le R_{\text{tot}}, \tag{6.20c}$$

$$(\sum_{i=1}^{F}\sum_{j=1}^{M} r_{ij} - R_{\text{tot}})\,\lambda = 0, \tag{6.20d}$$

$$\lambda \ge 0, \tag{6.20e}$$

$$r_{ij} \ge 0, \tag{6.20f}$$

$$r_{ij}\nu_{ij} = 0, \tag{6.20g}$$

$$\nu_{ij} \ge 0. \tag{6.20h}$$

We state that the optimal solution to this problem lies on the boundary of the budget constraint (6.20c). The proof of this statement is straightforward. Let us assume that an optimal solution, say $(\mathbf{W}^\star, \mathbf{R}^\star)$, is found such that $\mathbf{R}^\star$ lies strictly inside the feasibility set (and not on the boundary), with the corresponding objective distortion D^1. As the rates are constrained to be non-negative, one can increase the rates by a constant matrix, say $\mathbf{C}$, with non-negative entries to reach $\mathbf{R}^2 = \mathbf{R}^\star + \mathbf{C}$ such that the new solution, say $(\mathbf{W}^\star, \mathbf{R}^2)$ with a corresponding distortion D^2, still lies inside the set. As the distortion is a monotonically decreasing function over the rates, this implies $D^2 < D^1$. This shows that it is possible to increase rates until the full budget is used. Therefore, the third equation in the KKT conditions (6.20c) will be an equality constraint, and the fourth equation (complementary slackness over λ (6.20d)) and the fifth equation (6.20e) will be redundant.

We solve the KKT equations and find the optimal Lagrangian multiplier (λ) as a function of optimal weights. The first equation (6.20a) is actually the partial derivative with respect to the complex conjugate vector $\mathbf{w}_i^*$ [35], i.e.,

$$\begin{aligned} L_{\mathbf{w}_i^*} &= \mathbf{\Phi}_{\mathbf{x}_i}\mathbf{w}_i - \Phi_{S_i}\mathbf{a}_i + \mathbf{\Phi}_{\mathbf{n}_i}\mathbf{w}_i + \mathbf{\Phi}_{\mathbf{e}_i}\mathbf{w}_i \\ &= (\mathbf{\Phi}_{\mathbf{x}_i} + \mathbf{\Phi}_{\mathbf{n}_i} + \mathbf{\Phi}_{\mathbf{e}_i})\,\mathbf{w}_i - \Phi_{S_i}\mathbf{a}_i \\ &= \mathbf{\Phi}_{\tilde{\mathbf{y}}_i}\,\mathbf{w}_i - \Phi_{S_i}\mathbf{a}_i \\ &= \mathbf{\Phi}_{\tilde{\mathbf{y}}_i}\,\mathbf{w}_i - \mathbf{\Phi}_{\tilde{\mathbf{y}}_i S_i} = 0, \end{aligned} \tag{6.21}$$

where the superscript $\{\cdot\}^*$ denotes the complex conjugate operator on matrices/vectors. The solution to (6.21) are, in fact, the multi-channel Wiener filter coefficients, given the optimal rate vector $\mathbf{r}_i^\star = [r_{i1}^\star, \ldots, r_{iM}^\star]^{\mathrm{T}}$, given by

$$\mathbf{w}_i^\star(\mathbf{r}_i^\star) = \mathbf{\Phi}_{\tilde{\mathbf{y}}_i}^{-1}\mathbf{\Phi}_{\tilde{\mathbf{y}}_i S_i}(\mathbf{r}_i^\star) \in \mathbb{C}^{M\times 1}, \qquad i = 1, \ldots, F. \tag{6.22}$$

To find the optimal rates, we solve (6.20b) for v_{ij} and substitute it into (6.20g) (complementary slackness), i.e.

$$r_{ij}\left(\frac{-|w_{ij}|^2\, k_{ij}\, 2\ln 2}{2^{2r_{ij}}}+\lambda\right)=0, \qquad i=1,\ldots,F. \tag{6.23}$$

Equality in (6.23) holds either by setting r_{ij} or $v_{ij}=\lambda-\frac{|w_{ij}|^2\, k_{ij}\, 2\ln 2}{2^{2r_{ij}}}$ to be zero. Considering the last three equations in (6.20) together with (6.23), the optimal rate value is zero, i.e., $r_{ij}=0$ when $v_{ij}>0$, which implies $\frac{\lambda}{2\ln 2}>|w_{ij}|^2\, k_{ij}$. Otherwise, the optimal r_{ij} will be strictly positive when $v_{ij}=0$, which implies $\frac{\lambda}{2\ln 2}\leq|w_{ij}|^2\, k_{ij}$, and we have

$$r^{\star}_{ij}(\lambda'^{\star},w^{\star}_{ij})=\begin{cases}\frac{1}{2}\log_2\left(\frac{|w^{\star}_{ij}|^2\, k_{ij}}{\lambda'^{\star}}\right) & \lambda'^{\star}\leq|w^{\star}_{ij}|^2\, k_{ij},\\ 0 & \lambda'^{\star}>|w^{\star}_{ij}|^2\, k_{ij},\end{cases} \tag{6.24}$$

which simply can be rewritten as

$$r^{\star}_{ij}(\lambda'^{\star},w^{\star}_{ij})=\max\left(\frac{1}{2}\log_2\left(\frac{|w^{\star}_{ij}|^2\, k_{ij}}{\lambda'^{\star}}\right),0\right), \tag{6.25}$$

where $i=1,\ldots,F,\ \ j=1,\ldots,M$ with $\lambda'^{\star}=\frac{\lambda^{\star}}{2\ln 2}$ a rate reverse water filling parameter [23, 29]. In other words, the solution in (6.24) can be interpreted as if the equation (6.20b) is solved for r_{ij}, setting $v_{ij}=0$, and the result is projected onto the non-negative orthant, i.e., $r_{ij}\geq 0$. Finally, to find an optimal $\lambda'^{\star}$ which satisfies the equality budget constraint (the equation (6.20c) with equality), i.e.,

$$\sum_{i=1}^{F}\sum_{j=1}^{M} r^{\star}_{ij}(\lambda'^{\star},w^{\star}_{ij})=R_{\text{tot}}, \qquad i=1,\ldots,F. \tag{6.26}$$

we start by introducing a set $\mathscr{S}$ that contains the indices of components which are assumed to be allocated with positive rates

$$\mathscr{S}=\{(i,j)|\frac{|w^{\star}_{ij}|^2\, k_{ij}}{\lambda'^{\star}}>0\}, \qquad i=1,\ldots,F, \tag{6.27}$$

where $i=1,\ldots,F, j=1,\ldots,M$. Given the set S, the budget constraint can be rewritten as

$$\sum_{(i,j)\in\mathscr{S}}\left(\frac{1}{2}\log_2\left(\frac{|w^{\star}_{ij}|^2\, k_{ij}}{\lambda'^{\star}}\right)\right)=R_{\text{tot}}, \qquad i=1,\ldots,F. \tag{6.28}$$

Taking the logarithm of both sides of (6.28) and solving for λ' we have

$$\lambda'^{\star}=\frac{\left(\prod_{(i,j)\in\mathscr{S}}|w^{\star}_{ij}|^2\, k_{ij}\right)^{\frac{1}{|\mathscr{S}|}}}{2^{\left(\frac{2R_{\text{tot}}}{|\mathscr{S}|}\right)}}, \qquad i=1,\ldots,F. \tag{6.29}$$

To find the set $\mathscr{S}$, we use the water-filling procedure [23] as follows.

Algorithm 1: Linear Water-filling for optimal λ'

1 Sort the coefficients $|w_{ij}^{\star}|^2 \, k_{ij}$ in descending order into set $\mathscr{P}$.
2 **Initialize** an empty set $\mathscr{S} = \emptyset$, $\lambda'_{\text{opt}} = -\infty$:
3 **Pick** the first element in $\mathscr{P}$.
4 **If** λ'_{opt} is less than the picked value
5 Add the corresponding index into $\mathscr{S}$;
6 Compute (6.29) and update λ'_{opt};
7 **Else**
8 Stop and return $\mathscr{S}$ and λ'_{opt} (Optimal value is found).
9 **Repeat** 3-8 until all members of $\mathscr{P}$ are picked.

REFERENCES

[1] H. L. Van Trees, *Optimum Array Processing. Part IV of Detection, Estimation and Modulation Theory,* New York, NY: Wiley, 2008.

[2] M. Brandstein and D. Ward, *Microphone Arrays: Signal Processing Techniques and Applications,* Berlin, Germany: Springer Science and Business Media, 2001.

[3] H. Kayser, S. D. Ewert, J. Anem¨uller, T. Rohdenburg, V. Hohmann, and B. Kollmeier, "Database of multichannel in-ear and behind-the-ear head-related and binaural room impulse responses," EURASIP J. Adv. Signal Process, vol. 2009, pp. 6:1–6:10, 2009.

[4] R. Sockalingam, M. Holmberg, K. Eneroth, and M. Shulte, "Binaural hearing aid communication shown to improve sound quality and localization," The Hearing Journal, vol. 62, no. 10, pp. 46–47, 2009.

[5] D. Marquardt, "Development and Evaluation of Psychoacoustically Motivated Binaural Noise Reduction and Cue Preservation Techniques," PhD Dessertation, Universiy of Oldenburg, 2015.

[6] S. Markovich-Golan, S. Gannot, and I. Cohen, "Distributed multiple constraints generalized sidelobe canceler for fully connected wireless acoustic sensor networks," IEEE Transactions on Audio, Speech, and Language Processing, vol. 21, no. 2, pp. 343–356, 2013.

[7] A. Bertrand and M. Moonen, "Distributed adaptive node-specific signal estimation in fully connected sensor networkspart i: Sequential node updating," IEEE Transactions on Signal Processing, vol. 58, no. 10, pp. 5277–5291, 2010.

[8] R. Heusdens, G. Zhang, R. C. Hendriks, Y. Zeng, and W. B. Kleijn, "Distributed MVDR beamforming for (wireless) microphone networks using message passing," in IWAENC 2012; International Workshop on Acoustic Signal Enhancement, pp. 1–4, 2012.

[9] L. W. Brooks and I. S. Reed, "Equivalence of the likelihood ratio processor, the maximum signal-to-noise ratio filter, and the Wiener filter," IEEE Transactions on Aerospace and Electronic Systems, vol. AES-8, no. 5, pp. 690–692, 1972.

[10] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*, Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1993.

[11] S. Doclo and M. Moonen, "GSVD-based optimal filtering for single and multimicrophone speech enhancement," IEEE Transactions on Signal Processing, vol. 50, no. 9, pp. 2230–2244, 2002.

[12] S. Doclo, A. Spriet, J. Wouters, and M. Moonen, "Speech Distortion Weighted Multichannel Wiener Filtering Techniques for Noise Reduction," pp. 199–228, Berlin, Heidelberg: Springer, 2005.

[13] T. J. Klasen, T. Van den Bogaert, M. Moonen, and J. Wouters, "Binaural noise reduction algorithms for hearing aids that preserve interaural time delay cues," IEEE Transactions on Signal Processing, vol. 55, no. 4, pp. 1579–1585, 2007.

[14] S. Doclo, W. Kellermann, S. Makino, and S. E. Nordholm, "Multichannel signal enhancement algorithms for assisted listening devices: Exploiting spatial diversity using multiple microphones," IEEE Signal Processing Magazine, vol. 32, no. 2, pp. 18–30, 2015.

[15] T. C. Lawin-Ore, S. Stenzel, J. Freudenberger, and S. Doclo, "Generalized multichannel Wiener filter for spatially distributed microphones," in Speech Communication; 11. ITG Symposium, pp. 1–4, 2014.

[16] S. Srinivasan and A. den Brinker, "Rate-constrained beamforming in binaural hearing aids," EURASIP Journal on Advances in Signal Processing, pp. 1–9, 2009.

[17] O. Roy and M. Vetterli, "Rate-constrained collaborative noise reduction for wireless hearing aids," IEEE Transactions on Signal Processing, vol. 57, no. 2, pp. 645–657, 2009.

[18] S. Srinivasan and A. C. den Brinker, "Analyzing rate-constrained beamforming schemes in wireless binaural hearing aids," in 2009 17th European Signal Processing Conference, pp. 1854–1858, 2009.

[19] S. Doclo, M. Moonen, T. Van den Bogaert, and J. Wouters, "Reducedbandwidth and distributed MWF-Based noise reduction algorithms for binaural hearing aids," IEEE Transactions on Audio, Speech, and Language Processing, vol. 17, no. 1, pp. 38–51, 2009.

[20] A. D. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," IEEE Transactions on Information Theory, pp. 1–10, 1976.

[21] H. Yamamoto and K. Itoh, "Source coding theory for communication systems with a remote source," Trans. IECE Jpn, vol. E63, no. 6, pp. 700–706, 1980.

[22] T. Berger, Z. Zhang, and H. Viswanathan, "The CEO problem [multiterminal source coding]," IEEE Transactions on Information Theory, vol. 42, pp. 887902, 1996.

[23] T. M. Cover and J. A. Thomas, *Elements of Information Theory,* Wiley-Interscience, 2006.

[24] J. Amini, R. Hendriks, R. Heusdens, M. Guo, and J. Jensen, "Operational rate-constrained beamforming in binaural hearing aids," in 26th European Signal Processing Conference (EUSIPCO), 2018.

[25] Y. Shoham and A. Gersho, "Efficient bit allocation for an arbitrary set of quantizers," IEEE Transactions on Acoustics, Speech, and Signal Processing, vol. 36, no. 9, pp. 1445–1453, 1988.

[26] 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, no. 5, pp. 442–448, 1977.

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

[28] S. Boyd and L. Vandenberghe, *Convex Optimization,* Cambridge University Press, New York, NY, USA, 2004.

[29] T. Berger, *Rate-Distortion Theory: A Mathematical Basis for Data Compression,* Englewood Cliffs, N.J.: Prentice-Hall, 1971.

[30] A. Beck, "On the convergence of alternating minimization for convex programming with applications to iteratively reweighted least squares and decomposition schemes," SIAM Journal on Optimization, vol. 25, no. 1, pp. 185–209, 2015.

[31] L. Grippo and M. Sciandrone, "On the convergence of the block nonlinear Gauss-Seidel method under convex constraints," Operations Research Letters, vol. 26, no. 3, pp. 127–136, 2000.

[32] C. H. Taal, R. C. Hendriks, Heusdens R., and J. Jensen, "An algorithm for intelligibility prediction of timefrequency weighted noisy speech," IEEE Transactions on Audio, Speech, and Language Processing, vol. 19, no. 7, pp. 2125–2136, 2011.

[33] J. Kominek, A. W. Black, and V. Ver, "CMU arctic databases for speech synthesis," Tech. Rep., 2003.

[34] E. A. P. Habets, "Room impulse response generator," https://www.audiolabs-erlangen.de/fau/professor/habets/software/rirgenerator/, 2010.

[35] D. H. Brandwood, "A complex gradient operator and its application in adaptive array theory," IEE Proceedings F - Communications, Radar and Signal Processing, vol. 130, no. 1, pp. 11–16, 1983.

7

Spatially Correct Rate-Constrained Noise Reduction For Binaural Hearing Aids in Wireless Acoustic Sensor Networks

This chapter is published as "Spatially Correct Rate-Constrained Noise Reduction for Binaural Hearing Aids in Wireless Acoustic Sensor Networks," by J. Amini, R. C. Hendriks, R. Heusdens, M. Guo and J. Jensen, in IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 28, pp. 2731-2742, Oct. 2020, doi: 10.1109/TASLP.2020.3028264.

Multi-microphone noise reduction techniques, e.g., [1, 2], can be used to increase the speech quality and intelligibility of hearing aids (HAs). One way to use multi-microphone noise reduction techniques in modern HAs is to enable the left-ear and right-ear mounted HAs to collaborate through a wireless link, leading to a binaural HA setup. The binaural HA system provides increased spatial diversity and may result in better noise suppression, compared to the case where the monaural HAs perform noise reduction independently [3, 4]. In addition to better noise suppression, multi-microphone processing in the binaural HA setup can preserve binaural spatial information if taken care of, see e.g., [5–7]. These spatial information preserving noise reduction algorithms typically aim to preserve the interaural level differences (ILDs) and the interaural time differences (ITDs) of the relevant signal components. ILDs and ITDs are known to help humans determine the perceived location of the sound sources [6].

A common approach to achieve multi-microphone noise reduction is to combine the spatial observations captured by the microphones at a fusion center (FC) to estimate the sources of interest, while reducing the amount of environmental noise [2]. In the binaural HA setup, it is often considered that there are two FCs, one at each HA. Over the last decade, several binaural multi-microphone noise reduction algorithms have been proposed (see e.g., [6, 8] for overview). Typically they differ in the objective function they optimize and whether they can preserve the spatial cues of the target source, interferers, and the diffuse noise component. They can also differ in the types of constraints used to preserve the spatial cues. Equality constraints (see e.g., [5, 9–11]) are used to preserve exactly the spatial cues of the sources, while inequality constraints (see e.g., [12, 13]) are used to approximately preserve the spatial cues of the sources. The latter category can typically achieve a larger amount of noise suppression. In this chapter, we will focus on equality-constrained binaural multi-channel noise reduction filters.

A well known binaural minimum mean square error (MMSE)-based noise reduction algorithm is the binaural multi-channel Wiener filter (MWF) [14], which aims at minimizing the MSE of the target signal estimated at the reference microphones of the two FCs without imposing any source preserving constraints. This may result in significant noise reduction, but a distorted target signal. In contrast to the binaural MWF, the binaural minimum variance distortionless response (BMVDR) beamformer [8] minimizes the output noise power under two linear distortionless constraints that preserve the target signal at the two reference microphones leading to preservation of the binaural cues of the target source. These two constraints, however, reduce the noise reduction performance of the BMVDR, compared to the binaural MWF. Another example is the binaural linearly constrained minimum variance (BLCMV) beamformer [5, 15], which can preserve the ILDs and ITDs of the source of interest and multiple interferers. As another example, the optimal BLCMV (OBLCMV)[9] can achieve better noise reduction, compared to the BLCMV, however, can only preserve the ILD and ITD of one interferer. An LCMV-based approach is proposed in [10, 11] which tries to increase the degree of freedom of the optimization problem by introducing a set of linear equality constraints (firstly introduced in [16]) to enable preserving more interferers, for a given number of microphones, compared to the BLCMV and the optimal BLCMV. Most of the binaural LCMV-based methods differ in how the set of linear constraints is designed.

In all the above-mentioned methods, the two FCs of the binaural beamformers each

estimate the target source with respect to their corresponding reference microphone. To calculate these estimates, both FCs are in need of the microphone recordings from all sensors. This means that observations from the contralateral devices, and potentially any other device included in the setup, should be transmitted to the FCs. As the devices have a limited amount of resources (here transmission bandwidth) due to the limited battery lifetime, the total bit-rate used for transmission should be constrained. Several methods have been proposed in the literature to cope with this problem [17–20]. In [19] a binaural rate-constrained noise reduction approach is proposed which finds the optimal trade-off between the rate of transmission and the amount of noise reduction. The method finds the bound on the performance in case there are only two processing nodes. In the present context, these two processing nodes are the HAs. Scenarios with more than two nodes are not considered in [19]. Besides this, the inevitable requirement of the knowledge of the, generally time varying, joint statistics of all microphone signals at both HAs and using impractical infinitely long vector quantization limit the application of the method in practice. As alternatives to the optimal solution, several sub-optimal methods have been presented [21–23]. In [24], such algorithms were described in a unified framework. These sub-optimal methods try to pre-filter the observation before quantization without knowing the joint statistics, which enables the process to be faster and simpler. For example, this pre-filtering could be done to obtain a local estimate of the target or the interferer by combining the local microphone signals at the corresponding device. However, the pre-filtering stage combines the multi-microphone observations into a single observation, which may lead to a loss of some important information that needs to be known to retrieve the signals at high rates. As a result, even at an infinitely high rate of transmission, some important information may be lost and the performance will not approach that of the optimal algorithm presented in [19], not even asymptotically.

To address the aforementioned limitations, an operational rate-constrained noise reduction framework was proposed in [25], which estimates the optimal rate allocation across different frequencies and sensors using an operational rate-distortion trade-off [26]. Unlike [19], it allows considering scenarios with some additional assistive devices along with the binaural HA setup , thereby forming a small-size wireless acoustic sensor network (WASN) with more than two nodes. Furthermore, for the two-node case, the performance of the algorithm in [25] approaches that of the optimal algorithm in [19] at high rates without any mismatch, as the observations are not pre-filtered before quantization and necessary information will not be removed. However, the exhaustive search, which is used in [25] to find the optimal allocation across sensors, becomes intractable when the size of the WASN grows. Therefore, this method is suitable for small-size networks only. To address this scalability issue, another approach based on non-convex optimization was proposed in [27]. This method jointly finds the best rate allocation and the best estimation (beamforming) weights across all frequencies and sensors for arbitrary sized WASNs. Based on the MSE criterion, the optimal estimation weights are found to be rate-dependent Wiener filters and the optimal rates are the solution to a filter-dependent “water filling" problem. An alternating optimization approach which is used in this method avoids an exhaustive search to find the best allocations and performs almost as good as the exhaustive search-based approach, in most practical sce-

narios, at the benefit of a much lower computational complexity [27].

The above-mentioned methods deal with the rate-distortion trade-off in the noise reduction problem based on the MSE criterion. However, these methods do not take into account the preservation of spatial information (cues) when dealing with rate- constrained noise reduction problems. The noise reduction performance is optimal when minimizing the MSE, but the spatial information may be destroyed and the estimated signals may sound unnatural and spatially incorrect. Therefore, this raises the question of how to incorporate spatial information preservation into the rate-constrained noise reduction problem proposed in [27].

In this chapter, inspired by [27], we propose and solve a multi fusion-center spatially correct rate-constrained noise reduction problem, to find the best rate allocation and the best estimation (beamforming) weights across all sensors and frequencies such that the spatial information of the sources is preserved. The method links the LCMV-based beamformers to data compression by including a set of linear constraints to the original rate-distortion problem. Unlike [27], here, there are two FCs, therefore, the objective function is to minimize the sum of the distortions of the target estimation at both hearing aids, while considering the total rate budget and simultaneously preserving the spatial information of the sources. Using an alternating optimization approach, the optimal estimation weights are found to be the rate-dependent LCMV filters, and the rates for both fusion centers are the solutions to two water-filling problems. The performance of the proposed method is evaluated using output signal-to-noise ratio (SNR) gain measures, and ILD and ITD error measures. Simulation results show that the proposed method outperforms the methods with equal/random rate allocation strategies.

7

7.1. PROBLEM STATEMENT

7.1.1. SIGNAL MODEL

In this chapter, a generalized binaural hearing aid system is considered, which consists of two collaborating hearing aids along with a number of additional assistive devices. We assume that these assistive devices can only communicate with the two HAs and not with each other. In total $M = M^{\mathrm{L}} + M^{\mathrm{R}} + M^{\mathrm{A}}$ microphones are assumed to be embedded in the HAs and the assistive devices, including M^{L} microphones for the left HA, M^{R} microphones for the right HA, and M^{A} microphones for additional assistive devices. It is assumed here that no pre-filtering is applied to the unprocessed microphone signals to be transmitted to the FC, i.e., the microphone signals per device are not combined (pre-filtered) to a single signal.

Each microphone records a version of the target speech signal filtered by the position dependent room impulse response. The recorded target signal is degraded by a number of interfering point sources present in the room, diffuse noise and/or microphone self noise. The target signal, in the short-time Fourier transform (STFT) domain, is denoted by $S_k \in \mathbb{C}$, where k denotes the discrete frequency index. The interfering point sources are indicated by $I_{ki} \in \mathbb{C}$, where i denotes the point noise source index. All other sources of noise captured at a particular microphone are indicated by $U_{km} \in \mathbb{C}$, with m the microphone index. All sources are assumed to be additive and mutually uncorrelated.

Let the subscript $(\cdot)_m$ denote the microphone index. The signal model can then be

written as

$$Y_{km} = A_{km}S_k + \overbrace{\sum_{i=1}^{b} B_{kmi}I_{ki} + U_{km}}^{N_{km}}, \tag{7.1}$$

where $A_{km} \in \mathbb{C}$ is the acoustic transfer function (ATF) between the target signal and the mth microphone, and $B_{kmi} \in \mathbb{C}$ is the acoustic transfer function (ATF) between the ith point noise source and the mth microphone. The number of interferers is denoted by b.

Stacking all microphone signals in a vector, the signal model can be rewritten in vector notation as

$$\mathbf{y}_k = \overbrace{\mathbf{a}_k S_k}^{\mathbf{x}_k} + \overbrace{\sum_{i=1}^{b} \mathbf{b}_{ki}I_{ki} + \mathbf{u}_k}^{\mathbf{n}_k} = \mathbf{x}_k + \mathbf{n}_k, \tag{7.2}$$

where

$$\begin{aligned}
\mathbf{y}_k &= [(\mathbf{y}_k^{\mathrm{L}})^{\mathrm{T}}, (\mathbf{y}_k^{\mathrm{A}})^{\mathrm{T}}, (\mathbf{y}_k^{\mathrm{R}})^{\mathrm{T}}]^{\mathrm{T}}, \\
\mathbf{y}_k^{\mathrm{L}} &= [Y_{k1}, \ldots, Y_{kM^{\mathrm{L}}}]^{\mathrm{T}}, \\
\mathbf{y}_k^{\mathrm{A}} &= [Y_{k(M^{\mathrm{L}}+1)}, \ldots, Y_{k(M^{\mathrm{L}}+M^{\mathrm{A}})}]^{\mathrm{T}}, \\
\mathbf{y}_k^{\mathrm{R}} &= [Y_{k(M^{\mathrm{L}}+M^{\mathrm{A}}+1)}, \ldots, Y_{kM}]^{\mathrm{T}},
\end{aligned}$$

and similarly for $\mathbf{a}_k$, $\mathbf{b}_{ki}$ and $\mathbf{n}_k$. Let $\mathbf{y}_k^{\mathrm{L}}$, $\mathbf{y}_k^{\mathrm{A}}$, and $\mathbf{y}_k^{\mathrm{R}}$ denote the microphone signal vectors captured by the left side HA microphones, assistive microphones, and the right side microphones, respectively. The superscript $(\cdot)^{\mathrm{T}}$ denotes the transpose operator on vectors/matrices, and the power spectral density (PSD) matrix $\mathbf{\Phi}_{\mathbf{y}_k} = \mathrm{E}[\mathbf{y}_k\mathbf{y}_k^{\mathrm{H}}]$ of vector $\mathbf{y}_k$ is given by

$$\mathbf{\Phi}_{\mathbf{y}_k} = \mathbf{\Phi}_{\mathbf{x}_k} + \mathbf{\Phi}_{\mathbf{n}_k}, \tag{7.3}$$

where

$$\begin{aligned}
\mathbf{\Phi}_{\mathbf{x}_k} &= \mathrm{E}[\mathbf{x}_k\mathbf{x}_k^{\mathrm{H}}] = \Phi_{S_k}\mathbf{a}_k\mathbf{a}_k^{\mathrm{H}}, \\
\mathbf{\Phi}_{\mathbf{n}_k} &= \sum_{i=1}^{b} \Phi_{I_{ki}}\mathbf{b}_{ki}\mathbf{b}_{ki}^{\mathrm{H}} + \mathrm{E}[\mathbf{u}_k\mathbf{u}_k^{\mathrm{H}}],
\end{aligned} \tag{7.4}$$

and where $\Phi_{I_{ki}} = \mathrm{E}[|I_{ki}|^2] \in \mathbb{R}$ is the PSD of the ith interferer, $\Phi_{S_k} = \mathrm{E}[|S_k|^2] \in \mathbb{R}$ is the PSD of the clean target speech, and $\mathrm{E}[\cdot]$ denotes the expectation operator. The conjugate transpose operator on complex vectors/matrices is denoted by the superscript $(\cdot)^{\mathrm{H}}$.

7.1.2. LINEARLY CONSTRAINED ESTIMATION

A binaural beamformer estimates the signal of interest at both left side and right side reference positions by combining all the available noisy observations into a single estimate for each HA. Notice that in this chapter we do not only consider the presence of the two HAs, but also the presence of additional assistive microphones. The two resulting beamformer outputs are constructed such that a fidelity criterion is satisfied and the binaural information is preserved. The target signals at the left and right HA, i.e., S_k^{L} and S_k^{R}, respectively, are estimated as

$$\hat{S}_k^{\mathrm{L}} = (\mathbf{w}_k^{\mathrm{L}})^{\mathrm{H}}\mathbf{y}_k, \quad , \hat{S}_k^{\mathrm{R}} = (\mathbf{w}_k^{\mathrm{R}})^{\mathrm{H}}\mathbf{y}_k, \tag{7.5}$$

where $\mathbf{w}_k^{\mathrm{L}} \in \mathbb{C}^M$ and $\mathbf{w}_k^{\mathrm{R}} \in \mathbb{C}^M$ are the filter coefficients of the left and right beamformers, respectively. Minimizing the sum of the output noise powers, for both beamformers, the binaural linearly constrained beamforming problem can be formulated as [5]

$$\begin{aligned} \min_{\mathbf{w}_i} \quad & \mathbf{w}_k^{\mathrm{H}} \mathbf{\Phi}_k \mathbf{w}_k \\ \text{subject to} \quad & \mathbf{\Lambda}_k^{\mathrm{H}} \mathbf{w}_k = \mathbf{f}_k, \end{aligned} \tag{7.6}$$

where

$$\begin{aligned} \mathbf{w}_k &= [\mathbf{w}_k^{\mathrm{L}^{\mathrm{T}}} \mathbf{w}_k^{\mathrm{R}^{\mathrm{T}}}]^{\mathrm{T}} \in \mathbb{C}^{2M \times 1}, \\ \mathbf{\Phi}_k &= \begin{bmatrix} \mathbf{\Phi}_{\mathbf{n}_k} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Phi}_{\mathbf{n}_k} \end{bmatrix} \in \mathbb{C}^{2M \times 2M}, \end{aligned}$$

and $\mathbf{\Lambda}_k \in \mathbb{C}^{2M \times d}$ is the constraint matrix, with d the number of linear constraints. Different binaural LCMV-based beamformers can be constructed by changing the entries of $\mathbf{\Lambda}_k$. In this chapter, we use the methodology from [10, 11], having an increased amount of degrees of freedom compared to [9]. These additional degrees of freedom can then be used to cancel more interferers, given a fixed number of microphones. Following [10, 11] matrix $\mathbf{\Lambda}_k$ and vector $\mathbf{f}_k$ are given by

$$\begin{aligned} \mathbf{\Lambda}_k &= \begin{bmatrix} \mathbf{a}_k & \mathbf{0} & \mathbf{b}_1 B_{k1}^R & \dots & \mathbf{b}_b B_{kb}^R \\ \mathbf{0} & \mathbf{a}_k & -\mathbf{b}_1 B_{k1}^L & \dots & -\mathbf{b}_b B_{kb}^L \end{bmatrix} \in \mathbb{C}^{2M \times (b+2)}, \\ \mathbf{f}_k^H &= [A_k^L \; A_k^R \; 0 \; \dots \; 0] \in \mathbb{C}^{1 \times (b+2)}, \end{aligned} \tag{7.7}$$

respectively. Solving the problem in (7.6), the optimal weights are computed as [10]

$$\mathbf{w}_k^\star = \mathbf{\Phi}_k^{-1} \mathbf{\Lambda}_k (\mathbf{\Lambda}_k^H \mathbf{\Phi}_k^{-1} \mathbf{\Lambda}_k)^{-1} \mathbf{f}_k, \tag{7.8}$$

and the optimal beamformer outputs are given by

$$\hat{S}_k^{\mathrm{L}\star} = (\mathbf{w}_k^{\mathrm{L}\star})^{\mathrm{H}} \mathbf{y}_k, \quad , \hat{S}_i^{\mathrm{R}\star} = (\mathbf{w}_k^{\mathrm{R}\star})^{\mathrm{H}} \mathbf{y}_k. \tag{7.9}$$

In order to compute the binaural outputs $\hat{S}_k^{\mathrm{L}\star}$ and $\hat{S}_k^{\mathrm{R}\star}$, the actual signal realizations $\mathbf{y}_k$ should be available error-free at both HAs. However, due to limited battery power, and therefore, limited transmission power, in practice, the bit-rate, denoted by r_{km} bits per sample (bps), which is used to represent the transmitted signals must be constrained. Using a fixed bit-rate over frequencies and microphones can be shown to be sub-optimal, see e.g., [27]. Instead, the bit-rate dependent quantization noise should be included in the signal model, and optimized for.

7.1.3. QUANTIZATION AWARE ESTIMATION

In this sub-section, we introduce bit-rate dependent quantization noise in the signal model in (7.1). In this chapter, we assume that the microphone signals from all nodes in the WASN are being quantized using a uniform quantizer before transmission to the

corresponding FC (HA). Note that for each FC, the local observations at the FC are assumed to be quantized at the highest possible resolution, such that additional quantization noise on microphone signals at the FC can be neglected. In other words, only quantization noise with respect to the observations from other nodes in the WASN will be considered.

Consider an arbitrary signal denoted by x and its quantized version denoted by $\tilde{x}$, with quantization noise $q = x - \tilde{x}$. If subtractive dithering is applied to the signal to be quantized at lower rates or under high bit rate assumptions [28, 29], the quantization error q will be uniformly distributed and uncorrelated to signal x. In this case, the variance of the quantization noise is given by [28] $\sigma_q^2 = \frac{\Delta^2}{12}$, where $\Delta = \frac{2x_{\max}}{2^r}$ is the quantization step size, which depends on the range of the signal (maximum absolute value $x_{\max}$) and the quantization rate r.

Taking into account the quantization noise, the signal model for each side can be modified as

$$\begin{aligned}\tilde{Y}_{km}^{\mathrm{L}} &= Y_{km} + Q_{km}^{\mathrm{L}} = A_{km}S_k + \overbrace{\sum_{i=1}^{b} B_{kmi}I_{ki} + U_{km}}^{N_{km}} + Q_{km}^{\mathrm{L}},\\ \tilde{Y}_{km}^{\mathrm{R}} &= Y_{km} + Q_{km}^{\mathrm{R}} = A_{km}S_k + \overbrace{\sum_{i=1}^{b} B_{kmi}I_{ki} + U_{km}}^{N_{km}} + Q_{km}^{\mathrm{R}},\end{aligned} \tag{7.10}$$

where Q_{km}^{L} and Q_{km}^{R} denote the quantization noise w.r.t. the left and right side FCs, with $\tilde{Y}_{km}^{\mathrm{L}}$ and $\tilde{Y}_{km}^{\mathrm{R}}$ being the quantized microphone signals for the left and right side FCs, respectively. Using vector notation, we have

$$\begin{aligned}\tilde{\mathbf{y}}_k^{\mathrm{L}} &= \mathbf{y}_k + \mathbf{q}_k^{\mathrm{L}} = \mathbf{x}_k + \mathbf{n}_k + \mathbf{q}_k^{\mathrm{L}},\\ \tilde{\mathbf{y}}_k^{\mathrm{R}} &= \mathbf{y}_k + \mathbf{q}_k^{\mathrm{R}} = \mathbf{x}_k + \mathbf{n}_k + \mathbf{q}_k^{\mathrm{R}},\end{aligned} \tag{7.11}$$

where the quantization noise vector $\mathbf{q}_k^{\mathrm{L}} = [Q_{k1}^{\mathrm{L}}, Q_{k2}^{\mathrm{L}}, \cdots, Q_{kM}^{\mathrm{L}}]^{\mathrm{T}}$ is uncorrelated to the microphone signal vector $\mathbf{y}_k$, under the above-mentioned assumptions [28, 29], and similarly for $\mathbf{q}_k^{\mathrm{R}}$. Note that the bit-rates at which the left side signals are quantized are not necessarily the same as those at which the right side signals are quantized and transmitted to the left side FC. Under the above assumptions, and using $\Delta = \frac{2Y_{km}^{\mathrm{L,max}}}{2^{r_{km}^{\mathrm{L}}}}$, the CPSD matrix of the quantization noise vector $\mathbf{q}_k^{\mathrm{L}}$ will be diagonal with elements

$$\Phi_{Q_{km}^{\mathrm{L}}} = \frac{\Delta^2}{12} = \frac{(Y_{km}^{\mathrm{L,max}})^2}{3 2^{2r_{km}^{\mathrm{L}}}} = \frac{k_{km}^{\mathrm{L}}}{2^{2r_{km}^{\mathrm{L}}}}, \tag{7.12}$$

where $k_{km} = \frac{(Y_{km}^{\mathrm{L,max}})^2}{3}$. Similar expressions can be derived for the right side beamformer.

Applying the above mentioned quantization approach to the beamforming task, versions of the signal of interest S_k^{L} and S_k^{R} are estimated, given the quantized noisy microphone signals $\tilde{\mathbf{y}}_k^{\mathrm{L}}$ and $\tilde{\mathbf{y}}_k^{\mathrm{R}}$, as

$$\hat{S}_k^{\mathrm{L}} = (\mathbf{w}_k^{\mathrm{L}})^{\mathrm{H}}\tilde{\mathbf{y}}_k^{\mathrm{L}}, \quad , \hat{S}_k^{\mathrm{R}} = (\mathbf{w}_k^{\mathrm{R}})^{\mathrm{H}}\tilde{\mathbf{y}}_k^{\mathrm{R}}. \tag{7.13}$$

The beamformer outputs $\hat{S}_k^{\mathrm{L}}$ and $\hat{S}_k^{\mathrm{R}}$ depend on $\mathbf{w}_k^{\mathrm{L}}$, $\mathbf{w}_k^{\mathrm{R}}$, and on the rates r_{km}^{L} and r_{km}^{R}, respectively.

7.2. PROPOSED SPATIALLY CORRECT RATE-CONSTRAINED NOISE REDUCTION

In this sub-section, we propose and solve an optimization problem to jointly optimize the rates and the estimation weights across the sensors and frequencies. The FCs at the left and right HA have a limited total channel capacity of $R_{\mathrm{tot}}^{\mathrm{L}}$ and $R_{\mathrm{tot}}^{\mathrm{R}}$ bps, respectively, to receive information from the other nodes in the network, as argued in [30]. In addition to the transmission rate, in this chapter, we also take into account the preservation of spatial information, beneficial for binaural hearing aids. Altogether, in this chapter, we address the problem of joint rate-constrained noise reduction and spatial cue preservation to find the optimal filter coefficients and rate allocation for all sensors and frequencies.

7.2.1. PROBLEM FORMULATION

Let K indicate the number of frequency bins. Let the rate matrix $\mathbf{R}^{\mathrm{L}}$ be defined as

$$\mathbf{R}^{\mathrm{L}} = \begin{bmatrix} \mathbf{r}_1^{\mathrm{L}\,\mathrm{T}} \\ \mathbf{r}_2^{\mathrm{L}\,\mathrm{T}} \\ \vdots \\ \mathbf{r}_K^{\mathrm{L}\,\mathrm{T}} \end{bmatrix} = \begin{bmatrix} r_{11}^{\mathrm{L}} & r_{12}^{\mathrm{L}} & \dots & r_{1M}^{\mathrm{L}} \\ r_{21}^{\mathrm{L}} & r_{22}^{\mathrm{L}} & \dots & r_{2M}^{\mathrm{L}} \\ \vdots & \vdots & \ddots & \vdots \\ r_{K1}^{\mathrm{L}} & r_{K2}^{\mathrm{L}} & \dots & r_{KM}^{\mathrm{L}} \end{bmatrix},$$

which includes rates r_{km}^{L} to be allocated to frequency bin k and microphone signal m, for the left side FC. Please note that, here, the kth row of the matrix $\mathbf{R}^{\mathrm{L}}$ is defined as $\mathbf{r}_k^{\mathrm{L}\,\mathrm{T}} = [(\mathbf{r}_k^{\mathrm{LL}})^{\mathrm{T}}, (\mathbf{r}_k^{\mathrm{LA}})^{\mathrm{T}}, (\mathbf{r}_k^{\mathrm{LR}})^{\mathrm{T}}]^{\mathrm{T}}$, where $(\mathbf{r}_k^{\mathrm{LA}})^{\mathrm{T}}$ includes the rates at which the assistive microphones must be quantized and transmitted to the left side FC, and $(\mathbf{r}_k^{\mathrm{LR}})^{\mathrm{T}}$ includes the rates at which the right-side HA microphone signals must be quantized and transmitted to the left side FC, at kth frequency. A similar definition holds for the right side rate matrix $\mathbf{R}^{\mathrm{R}}$.

The weight matrix $\mathbf{W}^{\mathrm{L}}$ is similarly defined as

$$\mathbf{W}^{\mathrm{L}} = \begin{bmatrix} \mathbf{w}_1^{\mathrm{L}\,\mathrm{T}} \\ \mathbf{w}_2^{\mathrm{L}\,\mathrm{T}} \\ \vdots \\ \mathbf{w}_K^{\mathrm{L}\,\mathrm{T}} \end{bmatrix} = \begin{bmatrix} w_{11}^{\mathrm{L}} & w_{12}^{\mathrm{L}} & \dots & w_{1M}^{\mathrm{L}} \\ w_{21}^{\mathrm{L}} & w_{22}^{\mathrm{L}} & \dots & w_{2M}^{\mathrm{L}} \\ \vdots & \vdots & \ddots & \vdots \\ w_{K1}^{\mathrm{L}} & w_{K2}^{\mathrm{L}} & \dots & w_{KM}^{\mathrm{L}} \end{bmatrix},$$

which includes the left side beamformer coefficients w_{km}^{L}. A similar definition holds for the the right side beamformer coefficient matrix $\mathbf{W}^{\mathrm{R}}$.

Inspired by [27], we propose to formulate a spatially correct noise reduction problem, which tries to minimize a sum-distortion function given by

$$D(\mathbf{R}^{\mathrm{L}}, \mathbf{R}^{\mathrm{R}}, \mathbf{W}^{\mathrm{L}}, \mathbf{W}^{\mathrm{R}}) = D(\mathbf{R}^{\mathrm{L}}, \mathbf{W}^{\mathrm{L}}) + D(\mathbf{R}^{\mathrm{R}}, \mathbf{W}^{\mathrm{R}}), \tag{7.14}$$

where

$$D(\mathbf{R}^{\mathrm{L}},\mathbf{W}^{\mathrm{L}})=\frac{1}{K}\sum_{k=1}^{K}d(\mathbf{r}_k^{\mathrm{L}},\mathbf{w}_k^{\mathrm{L}})=\frac{1}{K}\sum_{k=1}^{K}\mathrm{E}[|S_k^{\mathrm{L}}-\hat{S}_k^{\mathrm{L}}|^2|\mathbf{r}_k^{\mathrm{L}},\mathbf{w}_k^{\mathrm{L}}],$$

$$D(\mathbf{R}^{\mathrm{R}},\mathbf{W}^{\mathrm{R}})=\frac{1}{K}\sum_{k=1}^{K}d(\mathbf{r}_k^{\mathrm{R}},\mathbf{w}_k^{\mathrm{R}})=\frac{1}{K}\sum_{k=1}^{K}\mathrm{E}[|S_k^{\mathrm{R}}-\hat{S}_k^{\mathrm{R}}|^2|\mathbf{r}_k^{\mathrm{R}},\mathbf{w}_k^{\mathrm{R}}].$$

Here, $d(\mathbf{r}_k^{\mathrm{L}},\mathbf{w}_k^{\mathrm{L}})$ denotes the PSD of the estimation error at the kth discrete frequency bin for the left side fusion center, and similarly for $d(\mathbf{r}_k^{\mathrm{R}},\mathbf{w}_k^{\mathrm{R}})$.

To address the rate-constrained noise reduction problem, we need constraint functions over the rates. Let $R(\mathbf{R}^{\mathrm{L}})$ simply be defined as the sum-rate over all frequency bins and microphones with respect to the left HA, given by

$$R(\mathbf{R}^{\mathrm{L}})=\sum_{k=1}^{K}\sum_{m=M^{\mathrm{L}}+1}^{M}r_{km}^{\mathrm{L}}. \tag{7.15}$$

and similarly for $R(\mathbf{R}^{\mathrm{R}})$.

To address the spatially correct noise reduction problem, we use the set of linear equality constraints defined in the previous section as

$$\mathbf{\Lambda}_k^{\mathrm{H}}\mathbf{w}_k=\mathbf{f}_k,\quad k=1,\cdots,K, \tag{7.16}$$

where,

$$\mathbf{w}_k=[(\mathbf{w}_k^{\mathrm{L}})^{\mathrm{T}},(\mathbf{w}_k^{\mathrm{L}})^{\mathrm{T}}]^{\mathrm{T}}.$$

Then, the proposed problem is defined as minimizing the estimation error, while satisfying the above-mentioned constraints. That is

$$\begin{aligned}\min_{\mathbf{R}^{\mathrm{L}},\mathbf{R}^{\mathrm{R}},\mathbf{W}^{\mathrm{L}},\mathbf{W}^{\mathrm{R}}}\quad & D(\mathbf{R}^{\mathrm{L}},\mathbf{W}^{\mathrm{L}})+D(\mathbf{R}^{\mathrm{R}},\mathbf{W}^{\mathrm{R}})\\ \text{subject to}\quad & R(\mathbf{R}^{\mathrm{L}})\le R_{\mathrm{tot}}^{\mathrm{L}},\\ & R(\mathbf{R}^{\mathrm{R}})\le R_{\mathrm{tot}}^{\mathrm{R}},\\ & \mathbf{\Lambda}_k^{\mathrm{H}}\mathbf{w}_k=\mathbf{f}_k,\quad k=1,\cdots,K.\end{aligned} \tag{7.17}$$

The distortion function $D(\mathbf{R}^{\mathrm{L}},\mathbf{W}^{\mathrm{L}})=\frac{1}{K}\sum_{k=1}^{K}d(\mathbf{r}_k^{\mathrm{L}},\mathbf{w}_k^{\mathrm{L}})$ is parameterized as a function of the estimator weights and allocated rates with $d(\mathbf{r}_k^{\mathrm{L}},\mathbf{w}_k^{\mathrm{L}})$ defined as

$$\begin{aligned}d(\mathbf{r}_k^{\mathrm{L}},\mathbf{w}_k^{\mathrm{L}})&=\mathrm{E}[|S_k^{\mathrm{L}}-\hat{S}_k^{\mathrm{L}}|^2|\mathbf{r}_k^{\mathrm{L}},\mathbf{w}_k^{\mathrm{L}}]\\ &=\mathrm{E}[|S_k^{\mathrm{L}}-(\mathbf{w}_k^{\mathrm{L}})^{\mathrm{H}}\tilde{\mathbf{y}}_k^{\mathrm{L}}|^2]\\ &=\mathrm{E}[|S_k^{\mathrm{L}}-(\mathbf{w}_k^{\mathrm{L}})^{\mathrm{H}}\mathbf{a}_kS_k-(\mathbf{w}_k^{\mathrm{L}})^{\mathrm{H}}\mathbf{n}_k-(\mathbf{w}_k^{\mathrm{L}})^{\mathrm{H}}\mathbf{q}_k^{\mathrm{L}}|^2]\\ &=|A_k^{\mathrm{L}}-(\mathbf{w}_k^{\mathrm{L}})^{\mathrm{H}}\mathbf{a}_k|^2\Phi_{S_k}+(\mathbf{w}_k^{\mathrm{L}})^{\mathrm{H}}\overbrace{[\mathbf{\Phi}_{\mathbf{n}_k}+\mathbf{\Phi}_{\mathbf{q}_k^{\mathrm{L}}}(\mathbf{r}_k^{\mathrm{L}})]}^{\mathbf{\Phi}_k^{\mathrm{L}}(\mathbf{r}_k^{\mathrm{L}})}\mathbf{w}_k^{\mathrm{L}},\end{aligned} \tag{7.18}$$

and similarly for the right side distortion function $D(\mathbf{R}^{\mathrm{R}},\mathbf{W}^{\mathrm{R}})$. Assuming a distortion-less response in the target signal direction, i.e., using the constraint $(\mathbf{w}_k^{\mathrm{L}})^{\mathrm{H}}\mathbf{a}_k=A_k^{\mathrm{L}}$, which is

included in the linear equality constraints in (7.16), (7.17), and the fact that $\mathbf{\Phi}_{\mathbf{q}_k^{\mathrm{L}}}(\mathbf{r}_k^{\mathrm{L}})$ is diagonal (see (7.12)), the distortion function $d(\mathbf{r}_k^{\mathrm{L}}, \mathbf{w}_k^{\mathrm{L}})$ can be rewritten as

$$d(\mathbf{r}_k^{\mathrm{L}}, \mathbf{w}_k^{\mathrm{L}}) = (\mathbf{w}_k^{\mathrm{L}})^{\mathrm{H}} \mathbf{\Phi}_{\mathbf{n}_k} \mathbf{w}_k^{\mathrm{L}} + \sum_{m=M^{\mathrm{L}}+1}^{M} \frac{|w_{km}^{\mathrm{L}}|^2 \, k_{km}^{\mathrm{L}}}{2^{2\, r_{km}^{\mathrm{L}}}}. \tag{7.19}$$

A similar expression can be written for the right side beamformer. Stacking both the variables for the left and the right FCs into matrices, we have

$$\mathbf{w}_k = [(\mathbf{w}_k^{\mathrm{L}})^{\mathrm{T}}, (\mathbf{w}_k^{\mathrm{R}})^{\mathrm{T}}]^{\mathrm{T}} \in \mathbb{C}^{2M \times 1},$$
$$\mathbf{\Phi}_k = \begin{bmatrix} \mathbf{\Phi}_k^{\mathrm{L}} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Phi}_k^{\mathrm{R}} \end{bmatrix} \in \mathbb{C}^{2M \times 2M}.$$

It is natural to assume positive rates, $r_{km} \geq 0$ (e.g. $r_{\min} = 0$ and $r_{\max} = \infty$). Therefore, the reformulated problem can further be written as

$$\begin{aligned} \min_{\mathbf{R}^{\mathrm{L}}, \mathbf{R}^{\mathrm{R}}, \mathbf{W}} \quad & \frac{1}{K} \sum_{k=1}^{K} [\mathbf{w}_k^{\mathrm{H}} \mathbf{\Phi}_k(\mathbf{r}_k^{\mathrm{L}}, \mathbf{r}_k^{\mathrm{R}}) \mathbf{w}_k] \\ \text{s.t.} \quad & \sum_{k=1}^{K} \sum_{m=M^{\mathrm{L}}+1}^{M} r_{km}^{\mathrm{L}} \leq R_{\mathrm{tot}}^{\mathrm{L}}, \\ & \sum_{k=1}^{K} \sum_{m=1}^{M^{\mathrm{L}}+M^{\mathrm{A}}} r_{km}^{\mathrm{R}} \leq R_{\mathrm{tot}}^{\mathrm{R}}, \\ & r_{km}^{\mathrm{L}} \geq 0, \quad r_{km}^{\mathrm{R}} \geq 0, \\ & \mathbf{\Lambda}_k^{\mathrm{H}} \mathbf{w}_k = \mathbf{f}_k, \end{aligned} \tag{7.20}$$

where the objective function includes the distortion function in (7.19), and also, includes a similar distortion function for the right-side FC. The function in (7.19) includes two terms: 1) the residual noise power $(\mathbf{w}_k^{\mathrm{L}})^{\mathrm{H}} \mathbf{\Phi}_{\mathbf{n}_k} \mathbf{w}_k^{\mathrm{L}}$, which is a quadratic (convex) function of the weights and 2) the residual quantization noise $\sum_{m=M^{\mathrm{L}}+1}^{M} \frac{|w_{km}^{\mathrm{L}}|^2 \, k_{km}^{\mathrm{L}}}{2^{2\, r_{km}^{\mathrm{L}}}}$, which is a summation of "quadratic-over-nonlinear" functions, which are non-convex. Therefore the problem in (7.20) is a non-convex optimization problem. However, fixing either $\mathbf{W}$ or $\mathbf{R}$, the problem will be convex in the remaining variable.

7.2.2. PROPOSED SOLUTION

Although the problem formulated in (7.20) is non-convex, we can still find the necessary optimality conditions by writing the Karush-Kuhn-Tucker (KKT) conditions [31]. Considering the first and second inequality rate constraint functions in (7.20), it can be shown that the rate solutions actually lie on the boundary of the feasibility sets defined by the global rate budget constraints which are the first and the second constraints in (7.20) [27].

We solve the KKT conditions and the solution will be given in the following proposition.

Proposition. *The solution to the problem in (7.20) is given by*

$$\begin{cases} 1)\ \mathbf{w}_k^{\star}(\mathbf{r}_k^{\mathrm{L}\star}, \mathbf{r}_k^{\mathrm{R}\star}) = \mathbf{\Phi}_k^{-1}\mathbf{\Lambda}_k(\mathbf{\Lambda}_k^H\mathbf{\Phi}_k^{-1}\mathbf{\Lambda}_k)^{-1}\mathbf{f}_k, \\ 2)\quad r_{km}^{\mathrm{L}\star}(\lambda_{\mathrm{L}}'^{\star}, w_{km}^{\mathrm{L}\star}) = [\frac{1}{2}\log_2(\frac{|w_{km}^{\mathrm{L}\star}|^2\, k_{km}^{\mathrm{L}}}{\lambda_{\mathrm{L}}'^{\star}})]^+, \\ 3)\quad r_{km}^{\mathrm{R}\star}(\lambda_{\mathrm{R}}'^{\star}, w_{km}^{\mathrm{R}\star}) = [\frac{1}{2}\log_2(\frac{|w_{km}^{\mathrm{R}\star}|^2\, k_{km}^{\mathrm{R}}}{\lambda_{\mathrm{R}}'^{\star}})]^+, \end{cases} \tag{7.21}$$

where $\lambda_L'^{\star} = \frac{K\lambda_L^{\star}}{2\,ln2}$ *and* $\lambda_R'^{\star} = \frac{K\lambda_R^{\star}}{2\,ln2}$ *are parameters, which satisfy the following equality constraints, respectively*

$$\sum_{k=1}^{K}\sum_{m=M^{\mathrm{L}}+1}^{M} r_{km}^{\mathrm{L}}(\lambda_{\mathrm{L}}'^{\star}) = R_{\mathrm{tot}}^{\mathrm{L}},$$

$$\sum_{k=1}^{K}\sum_{m=1}^{M^{\mathrm{L}}+M^{\mathrm{A}}} r_{km}^{\mathrm{R}}(\lambda_{\mathrm{R}}'^{\star}) = R_{\mathrm{tot}}^{\mathrm{R}}.$$

Proof. See Appendix 7-A. □

The rates are non-zero valued for $\lambda_{\mathrm{L}}'^{\star} \leq |w_{km}^{\mathrm{L}\star}|^2\, k_{km}^{\mathrm{L}}$ and $\lambda_{\mathrm{R}}'^{\star} \leq |w_{km}^{\mathrm{R}\star}|^2\, k_{km}^{\mathrm{R}}$ and are zero-valued otherwise. The non-linear operator $[\cdot]^+$ projects all negative valued rates to zero and the positive valued rates will remain unchanged, satisfying the set of inequality constraints in (7.20) ($r_{km}^{\mathrm{L}} \geq 0, \quad r_{km}^{\mathrm{R}} \geq 0,$).

As shown in the proposition, the optimal weights $\mathbf{w}_k^{\star}$ are the rate-constrained BLCMV coefficients, which, as a special case of the BLCMV coefficients, can be expressed as the BMVDR solutions. Note that, in general, $\mathbf{\Phi}_k^{-1}$ is a function of the bit-rates $\mathbf{r}_k^{\mathrm{L}\star}$ and $\mathbf{r}_k^{\mathrm{R}\star}$. The optimal rates r_{km}^{L} and r_{km}^{R} are the solution to the weighted reverse water filling problem. In other words, looking at the system of equations in (7.21), it turns out that to allocate the rates, we need to follow the reverse water filling approach while using the BLCMV filter coefficients. As explained, the BLCMV filters, when there is no quantization, can guarantee the preservation of the spatial cues of the target signal. Also here in (7.21), it is possible to preserve the spatial cues of the target signal, even when imperfect data, which is quantized at finite rate, is received by the corresponding beamformer and used to compute $\mathbf{\Phi}_k^{-1}$. Unlike the original water filling problem, where the rate allocation depends only on the microphone signal power, here, the rate allocation not only depends on the microphone signal power but also on the importance of the corresponding frequency component of the microphone signal to the estimation process. That is, the frequency bins which are more important in the target estimation stage, i.e., more informative, will be allocated more bits.

To solve the system of equations in (7.21), a similar approach as in [27] is used. The approach is based on alternating optimization, where the system is initialized with, for example, equal rate allocation across all components for both the left and right FCs, say $\mathbf{R}_0^{\mathrm{L}}$ and $\mathbf{R}_0^{\mathrm{R}}$, respectively. Then the weight equation is computed based on the equal rates and the weight matrix $\mathbf{W}_1$ is updated. Then, the rates will be updated based on the computed weights to $\mathbf{R}_1^{\mathrm{L}}$ and $\mathbf{R}_1^{\mathrm{R}}$. This process will be repeated until a certain stopping criterion is met. As the problem in (7.20) is component-wise convex, it is shown in [32] that any limit point, which is the solution after sufficient iterations. is a critical point. This

means that the obtained critical point is not necessarily globally optimal. However, as shown in [27], based on MSE and STOI measures, for certain types of noise reduction methods, the performance is almost as good as the method which uses an exhaustive search, but at the benefit of much lower computational complexity.

SPECIAL CASES OF THE PROPOSED SOLUTION

In Table 7.1, we highlight several special cases of the proposed solution in (7.21). As shown, (A) if the rate budgets go to infinity, then the solution will be equal to the joint BLCMV (JBLCM) filters [10, 11], using (7.7). (B) If the rate budgets go to infinity, and the matrix $\mathbf{\Lambda}_k$ is given by

$$\begin{aligned}\mathbf{\Lambda}_k &= \begin{bmatrix}\mathbf{a}_k & \mathbf{0}\\ \mathbf{0} & \mathbf{a}_k\end{bmatrix} \in \mathbb{C}^{2M\times 2},\\ \mathbf{f}_k^H &= [A_k^{\mathrm{L}}\, A_k^{\mathrm{R}}] \in \mathbb{C}^{1\times 2}.\end{aligned} \tag{7.22}$$

then the solution will become equal to the BMVDR filters [8]. (C) If the rate budgets are finite numbers, and the above-mentioned $\mathbf{\Lambda}_k$ in (7.22) is used, then the weight solution will be the rate-constrained BMVDR filters, which we refer to as "Proposed alternating optimization (AO)-BMVDR" in the next section. (D) Finally, when the rate budgets are finite, solving the equations in (7.21) and using (7.7) will lead to the proposed method, which we refer to as "ProposedAO-JBLCMV".

7.3. PERFORMANCE EVALUATION

In this section, we evaluate the proposed method as a function of the total bit rate budget by carrying out simulations in different acoustic scenarios. The proposed method will be compared to some existing methods using the binaural output SNR, and the ILD and ITD error measures, which will be defined in the next part of this section. In the evaluation, we will consider two different acoustic scenarios discussed in Sections 7.3.2 and 7.3.3, respectively.

7.3.1. PERFORMANCE MEASURES

We use the definitions presented in [6, 9, 10] for binaural input and output SNRs and ITD and ILD errors.

BINAURAL SNRS

The binaural input SNR and the binaural output SNR are defined as [9]

$$\begin{aligned}\mathrm{SNR}_{\mathrm{in}}(k) &= 10\log_{10}\Big(\frac{\mathbf{e}_{\mathrm{L}}^{\mathrm{T}}\Phi_{\mathbf{x_k}}\mathbf{e}_{\mathrm{L}} + \mathbf{e}_{\mathrm{R}}^{\mathrm{T}}\Phi_{\mathbf{x_k}}\mathbf{e}_{\mathrm{R}}}{\mathbf{e}_{\mathrm{L}}^{\mathrm{T}}\mathbf{\Phi}_k^{\mathrm{L}}\mathbf{e}_{\mathrm{L}} + \mathbf{e}_{\mathrm{R}}^{\mathrm{T}}\mathbf{\Phi}_k^{\mathrm{R}}\mathbf{e}_{\mathrm{R}}}\Big),\\ \mathrm{SNR}_{\mathrm{out}}(k) &= 10\log_{10}\Big(\frac{(\mathbf{w}_k^{\mathrm{L}})^{\mathrm{H}}\Phi_{\mathbf{x_k}}\mathbf{w}_k^{\mathrm{L}} + (\mathbf{w}_k^{\mathrm{R}})^{\mathrm{H}}\Phi_{\mathbf{x_k}}\mathbf{w}_k^{\mathrm{R}}}{(\mathbf{w}_k^{\mathrm{L}})^{\mathrm{H}}\mathbf{\Phi}_k^{\mathrm{L}}\mathbf{w}_k^{\mathrm{L}} + (\mathbf{w}_k^{\mathrm{R}})^{\mathrm{H}}\mathbf{\Phi}_k^{\mathrm{R}}\mathbf{w}_k^{\mathrm{R}}}\Big),\end{aligned} \tag{7.23}$$

where k denotes the frequency index, and

Table 7.1: Special cases of the proposed solution in (7.21).

Method	Total Rate	Constraint Matrix $\mathbf{\Lambda}$
(A): JBLCMV [10, 11]	$\mathbf{R}_{\text{tot}}^{\text{L}} \to \infty$ $\mathbf{R}_{\text{tot}}^{\text{R}} \to \infty$	$\mathbf{\Lambda}_k$ as in (7.7)
(B): BMVDR [8]	$\mathbf{R}_{\text{tot}}^{\text{L}} \to \infty$ $\mathbf{R}_{\text{tot}}^{\text{R}} \to \infty$	$\mathbf{\Lambda}_k$ as in (7.22)
(C): ProposedAO-BMVDR	$\mathbf{R}_{\text{tot}}^{\text{L}}$ is finite $\mathbf{R}_{\text{tot}}^{\text{R}}$ is finite	$\mathbf{\Lambda}_k$ as in (7.22)
(D): ProposedAO-JBLCMV	$\mathbf{R}_{\text{tot}}^{\text{L}}$ is finite $\mathbf{R}_{\text{tot}}^{\text{R}}$ is finite	$\mathbf{\Lambda}_k$ as in (7.7)

$$\begin{aligned}\mathbf{e}_{\text{L}}^{\text{T}} &= [1,0,\ldots,0] \in \mathbb{R}^{M},\\ \mathbf{e}_{\text{R}}^{\text{T}} &= [\underbrace{0,\ldots,0}_{M^{\text{L}}+M^{\text{A}}},1,0\ldots,0] \in \mathbb{R}^{M}.\end{aligned}$$

The performance measure we use is defined as the binaural SNR gain, $\text{SNR}_{\text{gain}}(k)$, and is given by

$$\text{SNR}_{\text{gain}}(k) = \text{SNR}_{\text{out}}(k) - \text{SNR}_{\text{in}}(k). \tag{7.24}$$

ILD AND ITD ERRORS

To define the ILD and ITD errors, we first define the input and output interaural transfer functions (ITFs) w.r.t. the source of interest as [6, 10]

$$\begin{aligned}\text{ITF}_{X}^{\text{in}}(k) &= \frac{X_k^{\text{L}}}{X_k^{\text{R}}} = \frac{A_k^{\text{L}}}{A_k^{\text{R}}},\\ \text{ITF}_{X}^{\text{out}}(k) &= \frac{\mathbf{w}_k^{\text{L}\,\text{H}}\mathbf{x}_k}{\mathbf{w}_k^{\text{R}\,\text{H}}\mathbf{x}_k} = \frac{\mathbf{w}_k^{\text{L}\,\text{H}}\mathbf{a}_k}{\mathbf{w}_k^{\text{R}\,\text{H}}\mathbf{a}_k}.\end{aligned} \tag{7.25}$$

Note that to find the ITFs for the interferers, the signal X_k and the transfer function A_k should be replaced by I_{ki} and B_{ki}, respectively, in (7.25). With this, the input and output ILDs are defined as the squared magnitudes of the input and output ITFs. That is

$$\text{ILD}_{X}^{\text{in}}(k) = |\text{ITF}_{X}^{\text{in}}(k)|^2, \qquad \text{ILD}_{X}^{\text{out}}(k) = |\text{ITF}_{X}^{\text{out}}(k)|^2, \tag{7.26}$$

and the input and output ITDs defined as the phase of the input and output ITFs. That is

$$\text{ITD}_{X}^{\text{in}}(k) = \angle\text{ITF}_{X}^{\text{in}}(k), \qquad \text{ITD}_{X}^{\text{out}}(k) = \angle\text{ITF}_{X}^{\text{out}}(k). \tag{7.27}$$

The ILD and ITD errors are then defined as

$$\begin{aligned} \mathrm{ER}_{\mathrm{ILD}_X^{\mathrm{out}}}(k) &= |\mathrm{ILD}_X^{\mathrm{out}}(k) - \mathrm{ILD}_X^{\mathrm{in}}(k)|, \\ \mathrm{ER}_{\mathrm{ITD}_X^{\mathrm{out}}}(k) &= \frac{|\mathrm{ITD}_X^{\mathrm{out}}(k) - \mathrm{ITD}_X^{\mathrm{in}}(k)|}{\pi}. \end{aligned} \tag{7.28}$$

Note that $0 \leq \mathrm{ER}_{\mathrm{ITD}_X^{\mathrm{out}}}(k) \leq 1$. Please note that, in this chapter, all defined measures will be rate-constrained, meaning that the measures are computed for a given total bit budgets $R_{\mathrm{tot}}^{\mathrm{L}}$ and $R_{\mathrm{tot}}^{\mathrm{R}}$, which will become more clear in the simulation results.

7.3.2. Example Binaural HA Setup using Head-Related Transfer Functions

Acoustic Scene 1

The first acoustic scene is based on the setup described in [33] and depicted in Fig. 7.1. The green circle in Fig. 7.1 denotes the target speech source, which is positioned at 3 m distance from the origin ((0,0)), in front of the binaural HA system. The binaural HA system consists of two HAs with two microphones per HA, with thus $M = 4$ microphones in total, mounted on a virtual head and denoted by the red "+" symbol. The zero degree corresponds to the looking direction of the virtual head and the angles are computed counterclockwise. The planar distance between the two microphones per HA is 0.76 cm and the radius of the typical head is 8.2 cm [33]. Interferers are indicated by the black triangles, assumed to be located at different positions in space, with a spatial resolution of 5°. The number and location of the interferers may vary in different experiments. Uncorrelated flat PSD noise is also added to the microphone signals at an SNR of 40 dB with respect to the corresponding reference microphones to simulate internal microphone noise.

The left and right side HAs are considered as two FCs. For example, for the left side FC, the observations recorded at its microphones are thought as the local observations and the contralateral right side microphone signals are quantized and transmitted to the left side FC. A similar explanation holds for the right side FC. Welch's method is used to estimate the PSD of the target speech, using 512-discrete Fourier transform (DFT) points, which is computed frame-by-frame using 50% overlapping speech frames. Around 12s of recorded sampled speech (at $F_s = 16$ KHz) from the "CMU-ARCTIC" data base [34] is used for the PSD estimation process. The head-related transfer functions (HRTFs) from the database in [33], with a spatial resolution of 5°, are used in this experiment. For the point noise sources, flat PSDs $\Phi_{I_k}(\omega)$ over the interval $\omega \in [-\pi, \pi]$ are considered. The cross-PSD matrices with respect to the target signal and the noises are computed using the estimated/computed PSDs and the HRTFs.

Competing Methods

The following methods are chosen as reference methods: a) **EQ-BMVDR:** the rate- constrained BMVDR. In this approach, we assume equal rate allocation across all sensors and frequencies, i.e., no optimization is done here. Note that when there is no quantization noise, this approach is equal to the BMVDR beamformer [8]. b) **EQ-JBLCMV:** The rate-constrained variation of the method proposed in [10, 11]. The equal rate allocation across all sensors and frequencies is considered in this approach. Note that when

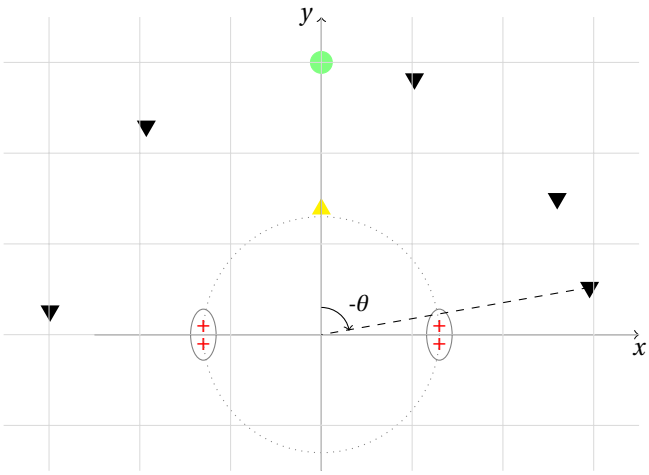


Figure 7.1: Example acoustic scene. The target signal, the interferers, and the four HA microphones (two microphones per HA) are denoted by the green circle, the black triangles , and the red "+", respectively.

there is no quantization noise, which happens at infinitely high rates, this method will be the same as the one proposed in [10, 11]. c) **ProposedAO-BMVDR:** In this approach, the special case of the proposed alternating optimization (AO) method described in Sec. 7.2.2 will be used to allocate the rates in the BMVDR beamforming setup. The constraint matrix Λ will simply have two columns, taking into account the distortion-less response constraints with respect to the target signal. d) **ProposedAO-JBLCMV:** In this approach, the proposed method described in Sec. 7.2.2 will be used to allocate the rates with the constraint matrix Λ mentioned in (7.7). Please note that to run the proposed algorithm, as well as the competing methods, the ATFs and the joint statistic are assumed to be known. Under stationary assumptions, and assuming that the spectral shape of the signal does not rapidly change over time, the over-head cost which is needed to inform the transmitters, on which bit-rate they should transmit the data, can be averaged out over consecutive frames.

SIMULATION RESULTS

In this section, we will compare the methods described in the previous sub-section based on the measures introduced in Sec. 7.3.1. We consider the acoustical setup, shown in Fig. 7.1 with five interferers located at $(3\text{m}, \{-80°, -60°, -20°, 40°, 85°\})$. The signal to interferer ratio (SIR) with respect to both FCs are set to approximately 0 dB. Fig. 7.2 shows the SNR gains as a function of total bit budget for the above-mentioned scenario. Please note that in Fig. 7.2 and all the remaining results in this chapter, the total bit-rate is normalized by the number of frequency samples, which is 512. The black horizontal dashed-line shows the upper bound on the performance of the BMVDR beamforming when there is no quantization noise, i.e., at infinitely high rates. Similarly, the black dashed-dotted horizontal line shows the upper bound on the performance of the JBLCMV beamforming at infinitely high rates. In fact, the BMVDR performs better than the JBLCMC in terms of SNR as it has more degrees of freedom for noise reduction, at the

cost of losing some binaural information, which will be shown later in this section. The performance of the both the “EQ-BMVDR" and the “ProposedAO-BMVDR" approach that of the BMVDR at high rates without any mismatch. As shown, the proposed method significantly outperforms the methods with equal rate allocation as the alternating optimization approach is used to jointly optimize the rates and weights. A similar argument holds for the “ProposedAO-JBLCMV". The performance of the “ProposedAO-JBLCMV" is always worse than that of the “ProposedAO-BMVDR" as less degrees of freedom remain for the noise reduction, compared to BMVDR beamforming.

To see how the methods affect the preservation of the binaural spatial information, we compute the ILD and ITD errors, introduced in (7.28). The ILD and ITD errors are shown in Fig. 7.3. In this chapter, the ILD and ITD errors are averaged among the target signal and the interferes.

The black dashed-line in both figures shows the asymptotic ILD and ITD errors for BMVDR beamforming, at infinitely high rates. Please note that the BMVDR method cannot preserve the spatial information with respect to the interferers, therefore there will be always ILD and ITD errors remaining in the processed signal. However, the JBLCMV beamformer can preserve the spatial information for up to $2M-3$ interferers, therefore, there is no ILD or ITD error with respect to the JBLCMV-based methods here. As shown in (7.21), in the proposedAO-JBLCMV method, as the weights are actually computed by the LCMV equations, it can also preserve the spatial information of $2M-3$ (which is five for $M=4$) interferers. As shown in Fig. 7.3a, in this specific scenario, the proposedAO-BMVDR method can perform better than the EQ-BMVDR method in terms of ILD errors at most total rates. However, as the problem proposed in (7.20) does not aim at optimizing the ILD or ITD errors, in general, it is not guaranteed to perform better than the equal rate allocation. The ILD and ITD errors w.r.t. both methods will approach that of the BMVDR beamforming at sufficiently high rates.

7.3.3. Example Generalized Binaural HA Setup Using Body-Related Transfer Functions

Acoustic Scene 2

In this section, we will compare the methods based on the generalized binaural HA setup from [35]. In addition to the binaural HA setup with four microphones as in Sec. 7.3.2, here, there is an assistive microphone, assumed to be mounted on the HA user's body (close to the left wrist). Therefore, this example includes five microphones. We use the body-related transfer functions (BRTFs) generated from the database presented in [35]. These impulse responses are measured with an adult human in an acoustically treated laboratory ($T_{60}\approx 200$ ms). All sources are assumed to be located at a planar distance of 2 m from the HA user. The target speech source is assumed to be located in front of the HA user and the six interferers are assumed to be located at $(2\text{m}, \{-15°, -30°, -60°, 30°, 60°, 90°\})$ with SIR set approximately to 0 dB w.r.t. both the left side and the right side reference microphones. Uncorrelated flat PSD noise is also added to the microphone signals with the SNR set to 40 dB to simulate internal microphone self noise. The PSD of the target speech and the other sources are estimated/assumed in the same fashion as described in the previous example setup in Sec. 7.3.2.

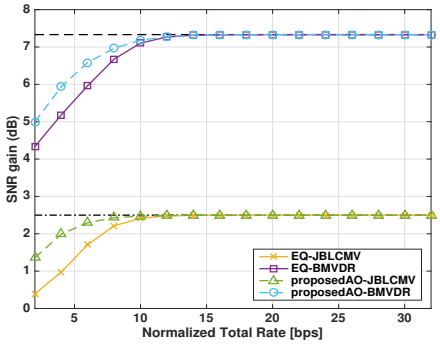
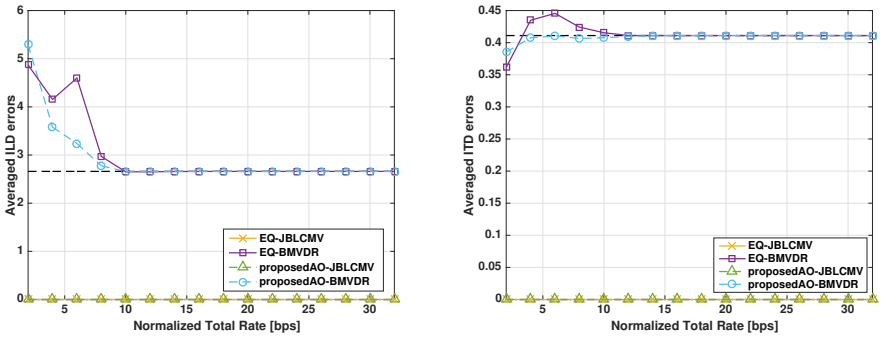


Figure 7.2: SNR gain [dB] versus total rate [bit per sample] based on a binaural setup in Fig. 7.1 (Acoustic Scene 1).



(a) ILD errors (Acoustic Scene 1).

(b) ITD errors (Acoustic Scene 1).

Figure 7.3: ILD and ITD errors versus total rate [bit per sample] based on the setup in Fig. 7.1 (Acoustic Scene 1).

SIMULATION RESULTS

The SNR gain is shown in Fig. 7.4

Similar to Sec. 7.3.2, The black horizontal dashed and the black dash-dotted lines denote the asymptotic BMVDR beamforming and JBLCMV beamforming SNR gains, respectively, at infinitely high rates. The performance of both “EQ-BMVDR" and “Proposed AO-BMVDR" follow a similar trend as in Fig. 7.2. Note that in this section, in addition to the generalized setup where there are five microphones (four microphones for the binaural HA setup and one additional assistive microphone), we also show the simulation results for the same acoustic scene, but with four microphones (without the assistive microphone), to show the benefit of having extra assistive microphone to increase the SNR gains. The methods which are based on the generalized setup are denoted by “x-5Mics", and the methods that are based on the binaural setup are denoted by “x-4Mics".

As shown in Fig. 7.4, with four microphones, the performance is always less than

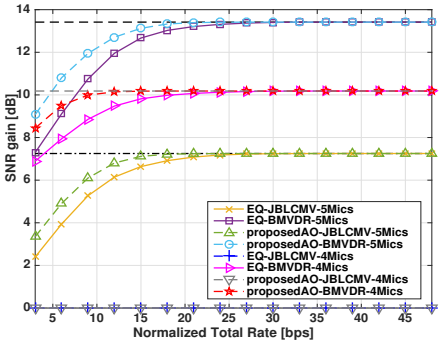


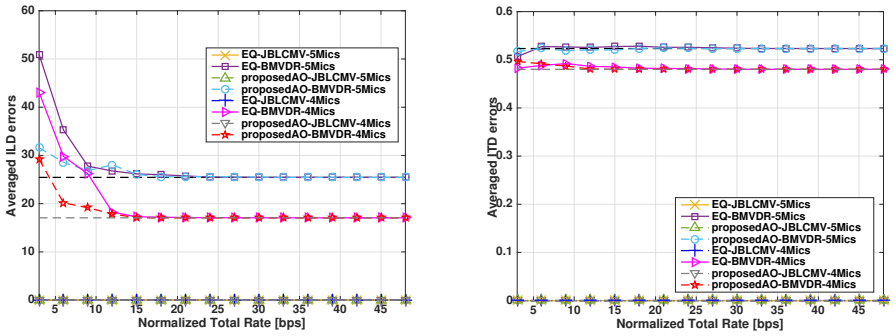
Figure 7.4: SNR gain [dB] versus total rate [bit per sample] based on the generalized binaural setup using BRTFs (Acoustic Scene 2).

the case with five microphones. In fact, with six interferers, in this simulation with four microphones, all JBLCMV-based methods spend all their degrees of freedom to preserve the spatial cues of the sources and hence, there is no control over the noise reduction (i.e., no SNR gain in this case). However, the BMVDR-based methods with four microphones still have control over the amount of noise reduction. Using the proposed alternating optimization method allows for optimal rate allocation for generalized-extended binaural setups where the additional assistive microphone can help to increase the averaged SNR gain, compared to the binaural configuration with four microphones.

The ILD and ITD errors based on the generalized setup with five microphones, as well as for the binaural setup with four HA microphones, are shown in Fig. 7.5. As shown, All JBLCMV-based methods can guarantee the preservation of the spatial cues (the yellow, green, blue, and gray-colored curves lie on top of each other with zero ILD and ITD errors), where the BMVDR-based methods suffer from spatial cue errors. Especially, the BMVDR-based methods with five microphones, focus more on the noise reduction task, and therefore, they have slightly more ILD and ITD errors compared to the case with four microphones.

With a similar explanation as in Sec. 7.3.2, the proposedAO-BMVDR, and the EQ-BMVDR methods are not able to preserve the spatial cues for all interferers as they do not impose any constraints to preserve the spatial cues of the interferers. As shown in Fig. 7.5b the proposedAO-BMVDR and the EQ-BMVDR methods have similar ITD errors at almost all rates, meaning that, if a certain amount of ITD error is of interest, then there is no need to send the high rate realizations to the FC, and hence, the observation can be quantized at lower rates and then transmitted. However, this argument is scenario-dependent.

Please note that similar to [27], here the proposed framework does not suffer from the scalability issue and can be applied to the more generalized scenarios including any number of microphones which can be located in random positions.



(a) ILD errors (Acoustic Scene 2). (b) ITD errors (Acoustic Scene 2).

Figure 7.5: ILD and ITD errors versus total rate [bit per sample] based on the generalized binaural HA setup (Acoustic Scene 2).

$$L(\mathbf{R}^{\mathrm{L}},\mathbf{R}^{\mathrm{R}},\mathbf{W}^{\mathrm{L}},\mathbf{W}^{\mathrm{R}},\lambda_{\mathrm{L}},\lambda_{\mathrm{R}},\mathbf{V}^{\mathrm{L}},\mathbf{V}^{\mathrm{R}},\mathbf{M}) = \frac{1}{K}\sum_{k=1}^{K}\mathbf{w}_k^{\mathrm{H}}\mathbf{\Phi}_k\mathbf{w} + \lambda_{\mathrm{L}}\left(\sum_{k=1}^{K}\sum_{m=M^{\mathrm{L}}+1}^{M}[r_{km}^{\mathrm{L}}] - R_{\mathrm{tot}}^{\mathrm{L}}\right) + \lambda_{\mathrm{R}}\left(\sum_{k=1}^{K}\sum_{m=1}^{M^{\mathrm{L}}+M^{\mathrm{A}}}[r_{km}^{\mathrm{R}}] - R_{\mathrm{tot}}^{\mathrm{R}}\right)$$
$$- \sum_{k=1}^{K}\sum_{m=M^{\mathrm{L}}+1}^{M}[v_{km}^{\mathrm{L}}r_{km}^{\mathrm{L}}] - \sum_{k=1}^{K}\sum_{m=1}^{M^{\mathrm{L}}+M^{\mathrm{A}}}[v_{km}^{\mathrm{R}}r_{km}^{\mathrm{R}}] + \sum_{k=1}^{K}\left(\mathrm{Re}\{\boldsymbol{\mu}_k\}^{\mathrm{T}}\mathrm{Re}\{\mathbf{\Lambda}_k^{\mathrm{H}}\mathbf{w}_k\} - \mathrm{Re}\{\boldsymbol{\mu}_k\}^{\mathrm{T}}\mathrm{Re}\{\mathbf{f}_k\}\right)$$
$$+ \sum_{k=1}^{K}\left(\mathrm{Im}\{\boldsymbol{\mu}_k\}^{\mathrm{T}}\mathrm{Im}\{\mathbf{\Lambda}_k^{\mathrm{H}}\mathbf{w}_k\} - \mathrm{Im}\{\boldsymbol{\mu}_k\}^{\mathrm{T}}\mathrm{Im}\{\mathbf{f}_k\}\right). \tag{7.29}$$

7

7.4. CONCLUSION

In this chapter, we proposed a spatially correct rate-constrained noise reduction problem which jointly finds the best rate allocation and estimation weights across all frequencies and sensors. The problem is based on the modified rate-distortion trade-off where the optimization problem is modified to incorporate the preservation of binaural cues, which is an important factor for increasing the speech intelligibility for hearing aid users. Solving the proposed optimization problem, based on the set of linear cue preservation constraints, the estimation (beamformer) weights are found to be the rate-dependent LCMV filters, and the rates are the solutions to the set of water filling problems. We chose two different acoustic scenes to evaluate the performance of the proposed methods: 1) The binaural HA setup with four microphones using HRTFs. 2) The generalized binaural HA setup with five microphones using BRTFs, where an additional assistive microphone is collaborating with HAs. We compared the BMVDR-based methods with the JBLCMV-based methods. The performance of the proposed method is evaluated using SNR gains and ILD and ITD errors. The results showed that the proposed method outperforms the methods with naive/equal choices of rates. In addition, as shown in Fig. 7.2 and Fig. 7.4, the BMVDR-based methods perform better than JBLCMV-based methods in terms of SNR in both scenarios as there is more degree of freedom for noise reduction, at the

cost of losing some spatial information of the sources. This behavior is consistent across different scenarios.

APPENDICES

7-A: DERIVATIONS OF THE PROPOSED SOLUTION IN (7.21)

The solution to the optimization problem in (7.20) is given by (7.21). In this section, we show the derivations leading to (7.21). We solve the KKT conditions, derived based on the problem in (7.20).

The Lagrangian function is given by (7.29). The matrix $\mathbf{M}$ includes the multipliers $\boldsymbol{\mu}_k$, i.e., $\mathbf{M} = [\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_K]$, and matrices $\mathbf{V}^{\mathrm{L}}$ and $\mathbf{V}^{\mathrm{R}}$ includes entries v^{L}_{km} and v^{R}_{km}, respectively. Given that

$$\begin{aligned} \mathrm{Re}\{\boldsymbol{\Lambda}_k^{\mathrm{H}}\mathbf{w}_k\} &= \frac{\boldsymbol{\Lambda}_k^{\mathrm{H}}\mathbf{w}_k + \boldsymbol{\Lambda}_k^{\mathrm{T}}\mathbf{w}_k^*}{2}, \\ \mathrm{Im}\{\boldsymbol{\Lambda}_k^{\mathrm{H}}\mathbf{w}_k\} &= \frac{\boldsymbol{\Lambda}_k^{\mathrm{H}}\mathbf{w}_k - \boldsymbol{\Lambda}_k^{\mathrm{T}}\mathbf{w}_k^*}{2i}, \end{aligned} \tag{7.30}$$

the KKT condition w.r.t. the Lagrangian function in (7.29) is given by

$$L_{\mathbf{w}_k^*} = \frac{1}{K}\boldsymbol{\Phi}_k\mathbf{w}_k + \frac{\boldsymbol{\Lambda}_k\mathrm{Re}\{\boldsymbol{\mu}_k\}}{2} - \frac{\boldsymbol{\Lambda}_k\mathrm{Im}\{\boldsymbol{\mu}_k\}}{2i} = 0, \tag{7.31a}$$

$$L_{r^{\mathrm{L}}_{km}} = \frac{-2\ln 2|w^{\mathrm{L}}_{km}|^2\, k^{\mathrm{L}}_{km}}{K 2^{2r^{\mathrm{L}}_{km}}} + \lambda_{\mathrm{L}} - v^{\mathrm{L}}_{km} = 0, \tag{7.31b}$$

$$L_{r^{\mathrm{R}}_{km}} = \frac{-2\ln 2|w^{\mathrm{R}}_{km}|^2\, k^{\mathrm{R}}_{km}}{K 2^{2r^{\mathrm{R}}_{km}}} + \lambda_{\mathrm{R}} - v^{\mathrm{R}}_{km} = 0, \tag{7.31c}$$

$$\sum_{k=1}^{K}\sum_{m=M^{\mathrm{L}}+1}^{M} r^{\mathrm{L}}_{km} \le R^{\mathrm{L}}_{\mathrm{tot}}, \tag{7.31d}$$

$$\sum_{k=1}^{K}\sum_{m=1}^{M^{\mathrm{L}}+M^{\mathrm{A}}} r^{\mathrm{R}}_{km} \le R^{\mathrm{R}}_{\mathrm{tot}}, \tag{7.31e}$$

$$\left(\sum_{k=1}^{K}\sum_{m=M^{\mathrm{L}}+1}^{M} r^{\mathrm{L}}_{km} - R^{\mathrm{L}}_{\mathrm{tot}}\right)\lambda_{\mathrm{L}} = 0, \tag{7.31f}$$

$$\left(\sum_{k=1}^{K}\sum_{m=1}^{M^{\mathrm{L}}+M^{\mathrm{A}}} r^{\mathrm{R}}_{km} - R^{\mathrm{R}}_{\mathrm{tot}}\right)\lambda_{\mathrm{R}} = 0, \tag{7.31g}$$

$$\lambda_{\mathrm{L}} \ge 0, \quad \lambda_{\mathrm{R}} \ge 0, \tag{7.31h}$$

$$r^{\mathrm{L}}_{km} \ge 0, \quad r^{\mathrm{R}}_{km} \ge 0, \tag{7.31i}$$

$$r^{\mathrm{L}}_{km} v^{\mathrm{L}}_{km} = 0, \quad r^{\mathrm{R}}_{km} v^{\mathrm{R}}_{km} = 0, \tag{7.31j}$$

$$v^{\mathrm{L}}_{km} \ge 0, \quad v^{\mathrm{R}}_{km} \ge 0. \tag{7.31k}$$

$$\boldsymbol{\Lambda}_k^{\mathrm{H}}\mathbf{w}_k = \mathbf{f}_k. \tag{7.31l}$$

First, we solve the KKT conditions w.r.t. the estimation weights $\mathbf{w}_k$. Solving (7.31a) for $\mathbf{w}_k$, we have

$$\mathbf{w}_k^\star = K\boldsymbol{\Phi}_k^{-1}\boldsymbol{\Lambda}_k\left(\frac{\mathrm{Re}\{\boldsymbol{\mu}^\star\} + i\mathrm{Im}\{\boldsymbol{\mu}^\star\}}{2}\right) = \frac{K}{2}\boldsymbol{\Phi}_k^{-1}\boldsymbol{\Lambda}_k\boldsymbol{\mu}^\star. \tag{7.32}$$

Substituting (7.32) into the linear constraint (7.31l) and solving (7.31l), the optimal $\boldsymbol{\mu}^\star$ is

given by

$$\boldsymbol{\mu}^{\star} = \frac{2}{K}(\boldsymbol{\Lambda}_k^{\mathrm{H}}\boldsymbol{\Phi}_k^{-1}\boldsymbol{\Lambda}_k)^{-1}\mathbf{f}_k. \tag{7.33}$$

Finally, substituting (7.33) back into (7.32), the optimal weights are given by

$$\mathbf{w}_k^{\star}(\mathbf{r}_k^{\mathrm{L}\star}, \mathbf{r}_k^{\mathrm{R}\star}) = \boldsymbol{\Phi}_k^{-1}\boldsymbol{\Lambda}_k(\boldsymbol{\Lambda}_k^{\mathrm{H}}\boldsymbol{\Phi}_k^{-1}\boldsymbol{\Lambda}_k)^{-1}\mathbf{f}_k. \tag{7.34}$$

Note that, unlike the original BLCMV solution, here the optimal weights $\mathbf{w}_k^{\star}$, as well as the PSD matrix $\boldsymbol{\Phi}_k$ are functions of the optimal bit-rates, which will be derived in the following.

As the constraint functions for r_{km}^{L} and r_{km}^{R} are separable, we can independently solve the KKT equations w.r.t. the corresponding rates. We start with the solution for r_{km}^{L}. Solving (7.31b) for v_{km}^{L}, and substituting it into the complementary slackness condition in (7.31j), we have

$$\left(\frac{-2\ln 2|w_{km}^{\mathrm{L}}|^2\, k_{km}^{\mathrm{L}}}{K 2^{2r_{km}^{\mathrm{L}}}} + \lambda_{\mathrm{L}}\right) r_{km}^{\mathrm{L}} = 0. \tag{7.35}$$

Looking at (7.35), there are two cases here: 1) the optimal rate r_{km}^{L} is set to zero, when based on (7.31j), the variable v_{km}^{L} has to be strictly greater than zero, which, by looking at (7.31b), implies $\frac{\lambda_{\mathrm{L}} K}{2\ln 2} \geq |w_{km}^{\mathrm{L}}|^2\, k_{km}^{\mathrm{L}}$. 2) $v_{km}^{\mathrm{L}} = 0$, then solving (7.31b) for r_{km}^{L}, the optimal non-zero valued rates are given by

$$r_{km}^{\mathrm{L}\star} = \frac{1}{2}\log_2\left(\frac{|w_{km}^{\mathrm{L}\star}|^2\, k_{km}^{\mathrm{L}}}{\frac{K\lambda_{\mathrm{L}}^{\star}}{2\ln 2}}\right), \tag{7.36}$$

which implies $\frac{\lambda_{\mathrm{L}} K}{2\ln 2} < |w_{km}^{\mathrm{L}}|^2\, k_{km}^{\mathrm{L}}$. Combining cases 1 and 2, we have

$$r_{km}^{\mathrm{L}\star}(\lambda_{\mathrm{L}}'^{\star}, w_{km}^{\mathrm{L}\star}) = \left[\frac{1}{2}\log_2\left(\frac{|w_{km}^{\mathrm{L}\star}|^2\, k_{km}^{\mathrm{L}}}{\lambda_{\mathrm{L}}'^{\star}}\right)\right]^{+}, \tag{7.37}$$

where $\lambda_{\mathrm{L}}'^{\star} = \frac{K\lambda_{\mathrm{L}}^{\star}}{2\ln 2}$. The operator $[\cdot]^{+}$ assures positive rates and projects all negative values onto zero. The parameter $\lambda_{\mathrm{L}}'^{\star}$ must satisfy the KKT condition (7.31d) with equality, as argued in [27]. Note that the rates are functions of the weights $w_{km}^{\mathrm{L}\star}$ and the water-falling threshold parameter $\lambda_{\mathrm{L}}^{\star}$. Therefore, the alternating optimization is proposed to be used to solve theses equations in (7.37) and (7.34). A similar proof holds for $r_{km}^{\mathrm{R}\star}$.

Finally to find the optimal $\lambda_{\mathrm{L}}^{\star}$ and $\lambda_{\mathrm{R}}^{\star}$, a similar water-filling approach, as proposed in [27] (in the last part of the proof in the appendix), can be used.

REFERENCES

[1] M. Brandstein and D. Ward, *Microphone Arrays: Signal Processing Techniques and Applications*, Berlin, Germany: Springer Science and Business Media, 2001.

[2] P. Vary and R. Martin, *Digital Speech Transmission: Enhancement, Coding And Error Concealment*, John Wiley and Sons, 2006.

[3] R. Sockalingam, M. Holmberg, K. Eneroth, and M. Shulte, "Binaural hearing aid communication shown to improve sound quality and localization," The Hearing Journal, vol. 62, no. 10, pp. 46–47, 2009.

[4] D. Marquardt, "Development and Evaluation of Psychoacoustically Motivated Binaural Noise Reduction and Cue Preservation Techniques," PhD Dessertation, Universiy of Oldenburg, 2015.

[5] 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, 2012.

[6] B. Cornelis, S. Doclo, T. Van dan Bogaert, M. Moonen, and J. Wouters, "Theoretical analysis of binaural multimicrophone noise reduction techniques," IEEE Transactions on Audio, Speech, and Language Processing, vol. 18, no. 2, pp. 342–355, 2010.

[7] T. J. Klasen, T. Van den Bogaert, M. Moonen, and J. Wouters, "Binaural noise reduction algorithms for hearing aids that preserve interaural time delay cues," IEEE Transactions on Signal Processing, vol. 55, no. 4, pp. 1579–1585, 2007.

[8] S. Doclo, W. Kellermann, S. Makino, and S. E. Nordholm, "Multichannel signal enhancement algorithms for assisted listening devices: Exploiting spatial diversity using multiple microphones," IEEE Signal Processing Magazine, vol. 32, no. 2, pp. 18–30, 2015.

[9] D. Marquardt, E. Hadad, S. Gannot, and S. Doclo, "Optimal binaural lcmv beamformers for combined noise reduction and binaural cue preservation," in 2014 14th International Workshop on Acoustic Signal Enhancement (IWAENC), pp. 288–292, 2014.

[10] A. I. Koutrouvelis, R. C. Hendriks, J. Jensen, and R. Heusdens, "Improved multimicrophone noise reduction preserving binaural cues," in 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 460–464, 2016.

[11] E. Hadad, D. Marquardt, D. Doclo, and S. Gannot, "Theoretical analysis of binaural transfer function mvdr beamformers with interference cue preservation constraints," IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 23, no. 12, pp. 2449–2464, 2015.

[12] A. I. Koutrouvelis, R. C. Hendriks, R. Heusdens, and J. Jensen, "Relaxed binaural LCMV beamforming," IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 25, no. 1, pp. 137–152, 2017.

[13] A. I. Koutrouvelis, R. C. Hendriks, R. Heusdens, and J. Jensen, "A convex approximation of the relaxed binaural beamforming optimization problem," IEEE/ACM Transactions on Audio,Speech, and Language Processing, vol. 27, no. 2, pp. 321–331, 2019.

[14] S. Doclo, A. Spriet, J. Wouters, and M. Moonen, "Frequency-domain criterion for the speech distortion weighted multichannel wiener filter for robust noise reduction," Speech Communication, vol. 49, no. 7-8, pp. 636–656, 2007.

[15] 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, no. 3, pp. 543–558, 2016.

[16] D. Marquardt, E. Hadad, S. Gannot, and S. Doclo, "Theoretical analysis of linearly constrained multi-channel wiener filtering algorithms for combined noise reduction and binaural cue preservation in binaural hearing aids," IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 23, no. 12, pp. 2384–2397, 2015.

[17] S. Doclo, M. Moonen, T. Van den Bogaert, and J. Wouters, "Reducedbandwidth and distributed MWF-Based noise reduction algorithms for binaural hearing aids," IEEE Transactions on Audio, Speech, and Language Processing, vol. 17, no. 1, pp. 38–51, 2009.

[18] S. Doclo, T. C. Lawin-Ore, and T. Rohdenburg, "Rate-constrained binaural MWF-based noise reduction algorithms," in Proc. ITG Conference on Speech Communication, Bochum, Germany, 2010.

[19] O. Roy and M. Vetterli, "Rate-constrained collaborative noise reduction for wireless hearing aids," IEEE Transactions on Signal Processing, vol. 57, no. 2, pp. 645–657, 2009.

[20] S. Srinivasan and A. C. den Brinker, "Analyzing rate-constrained beamforming schemes in wireless binaural hearing aids," in 2009 17th European Signal Processing Conference, pp. 1854–1858, 2009.

[21] S. Srinivasan, "Low-bandwidth binaural beamforming," Electronics Letters, vol. 44, no. 22, pp. 1292–1293, 2008.

[22] S. Srinivasan and A. den Brinker, "Rate-constrained beamforming in binaural hearing aids," EURASIP Journal on Advances in Signal Processing, pp. 1–9, 2009.

[23] O. Roy and M. Vetterli, "Collaborating hearing aids," in Proceedings of MSRI Workshop on Mathematics of Relaying and Cooperation in Communication Networks, 2006.

[24] J. Amini, R. C. Hendriks, R. Heusdens, M. Guo, and J. Jensen, "Asymmetric coding for rate-constrained noise reduction in binaural hearing aids," IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 27, no. 1, pp. 154–167, 2019.

[25] J. Amini, R. Hendriks, R. Heusdens, M. Guo, and J. Jensen, "Operational rate-constrained beamforming in binaural hearing aids," in 26th European Signal Processing Conference (EUSIPCO), 2018.

[26] Y. Shoham and A. Gersho, "Efficient bit allocation for an arbitrary set of quantizers," IEEE Transactions on Acoustics, Speech, and Signal Processing, vol. 36, no. 9, pp. 1445–1453, 1988.

[27] J. Amini, R. Hendriks, R. Heusdens, M. Guo, and J. Jensen, "Rate-constrained noise reduction in wireless acoustic sensor networks," IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 28, pp. 1–12, 2020.

[28] 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, no. 5, pp. 442–448, 1977.

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

[30] T. Berger, Z. Zhang, and H. Viswanathan, "The CEO problem [multiterminal source coding]," IEEE Transactions on Information Theory, vol. 42, pp. 887902, 1996.

[31] S. Boyd and L. Vandenberghe, *Convex Optimization,* Cambridge University Press, New York, NY, USA, 2004.

[32] L. Grippo and M. Sciandrone, "On the convergence of the block nonlinear Gauss-Seidel method under convex constraints," Operations Research Letters, vol. 26, no. 3, pp. 127 – 136, 2000.

[33] H. Kayser, S. D. Ewert, J. Anem¨uller, T. Rohdenburg, V. Hohmann, and B. Kollmeier, "Database of multichannel in-ear and behind-the-ear head-related and binaural room impulse responses," EURASIP J. Adv. Signal Process, vol. 2009, pp. 6:1–6:10, 2009.

[34] J. Kominek, A. W. Black, and V. Ver, "CMU arctic databases for speech synthesis," Tech. Rep., 2003.

[35] R. M. Corey, N. Tsuda, and A. C. Singer, "Acoustic impulse responsesvfor wearable audio devices," in ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 216–220, 2019.

8

Conclusion and Future Research Directions

In this chapter, we draw the conclusions of the dissertation. In addition, we mention some future research questions which are raised based on the dissertation.

8.1. Conclusions

Binaural hearing aids are shown to be capable of utilizing advanced multi-microphone noise reduction techniques to enhance the acoustic scene for the hearing aid user. Multi-microphone noise reduction techniques for hearing aids aim at improving the quality and intelligibility of the speech sources. Thanks to wireless technology, hearing aids of the left and the right side ears can potentially collaborate with each other, as well as with other wireless assistive devices in the acoustic scene to improve the amount of noise reduction and allow preservation of the location of the sound sources. One common way to achieve proper noise reduction is to combine the observations from all microphones and output an estimate of the desired sources for the left and the right ears, while reducing the environmental noise. In these approaches, a certain fidelity criterion is optimized to achieve a good estimate of the desired sources, which can be, for example, the MSE or other measures like intelligibility metrics [1, 2]. The observations from the right- and left side hearing aids consist of the binaural information about the sources in the acoustic scene. This binaural information includes the interaural level differences (ILD) and the interaural phase/time differences (IPD/ITD). Preservation of these binaural cues is very important to get a natural impression of the acoustic scene [3].

To perform the multi-microphone noise reduction algorithms the observations from all sensors should be received by the central processor. As the power supply of the devices is limited, the data must be compressed/quantized before transmission. Typically the observations are quantized at certain fixed bit-rates. Therefore, the rate of transmission between the devices is constrained in the noise reduction problem. The goal of this dissertation was to (optimally) incorporate the quantization noise into the multi-

microphone noise reduction problems, taking rate constraints into account.

The goal of rate-constrained noise reduction techniques is similar to the original noise reduction techniques, except that the data to be processed is incomplete (quantized). The performance of the rate-constrained noise reduction methods depends on the amount and distribution of the bit budget across different sensors and frequencies [4]. By using higher bit-rates to represent the microphone signals, the performance will increase. This introduces a trade-off between the transmission rate and the accuracy of the noise reduction algorithm. By measuring the noise reduction performance via distortion measures (e.g. MSE), the problem can be seen as the rate-distortion trade-off, which has been well-studied from an information-theoretic perspective in [5–8]. Looking at the rate-distortion problem from a noise reduction perspective, we propose, in this thesis, different algorithm to incorporate the quantization rate into the noise reduction problem and find different rate-distortion trade-offs aimed at noise reduction.

To summarize what has been addressed in this thesis, the research questions raised in the introduction chapter will be answered briefly in the following. More detailed conclusions will be explained in the remaining sections, in this chapter.

1- What is the effect of the quantization on the noise reduction performance, and how do quantization related assumptions affect the performance of the quantization aware noise reduction?

   - To answer this research question, we have proposed in Chapter three a rate-dependent BMVDR noise reduction algorithm based on uniform quantization for binaural hearing aids. In this chapter, we showed the trade-off between the amount of the total rate spent to quantize the contralateral observation transmitted to the other HA, and the SNR gain of the BMVDR. The more rates are spent to quantize the data, the more SNR gain is achieved by the method. In addition, we investigated some assumptions made on the form of quantization, which influences the SNR performance, especially in the low-bitrate scenarios. These assumptions are: 1) the quantization noise is uncorrelated to the signal to be quantized. 2) And the quantization noise of the observations are uncorrelated from each other. We showed that by using the dithering technique, we can decorrelate the quantization noise from the signal on the additive signal model, making the assumptions to take the quantization into account valid at all rates.

2- The optimal binaural rate-constrained method in [4] unavoidably requires the joint statistics to be known at both processing nodes. Can we design a coding algorithm from an information-theoretic point of view, which can inherently estimate the joint statistics to be applied to provide an optimal solution at least for one processor?

   - To answer this research question, we have proposed in Chapter four an asymmetric coding scheme to do optimal binaural rate-dependent noise reduction without pre-knowledge on the joint statistics. This approach consists of two communication links. One link from the right HA to the left HA and vice versa. In one of the two links, we proposed a vector case of the probability

preserving rate-distortion trade-off, and with that, we can simply retrieve the (unquantized) joint correlations from the quantized data and use it in the other link to achieve the optimal trade-off.

3- Existing methods for rate-constrained binaural noise reduction typically consider only two processing nodes. Can we generalize the binaural hearing set up with a smart rate allocation technique to enable more assistive devices to cooperate for improved noise reduction performance?

- To answer this research question, in Chapter five we have proposed an operational rate-constrained noise reduction algorithm that enables us to utilize not only the binaural information but also the information from additional assistive devices (e.g. mobile phones) together to construct a (small) WASN. In this approach, a discrete optimization technique is used to optimize rate allocations across frequencies and sensors. The above-mentioned method uses an exhaustive search-based approach to find the best rate allocation across microphones, but still uses an efficient rate allocation optimization across frequencies. However, due to the exhaustive search, this becomes intractable when the size of the WASN grows. To address this issue, in Chapter six we have proposed a rate-constrained noise reduction algorithm that suits arbitrary-size WASN, without this scalability issue. We proposed an optimization problem that aims at minimizing the MSE between the estimated signal at the central fusion center and the desired signal by jointly finding the best estimation weights and quantization rates across all frequencies and sensors. Under certain assumptions, the estimations weights are found to be the rate-dependent Wiener filter coefficients, and the optimal rates are found to be the solution to the reverse water-filling problem. The results have shown that the proposed solution performs almost as good as the optimal exhaustive search-based algorithm, with much lower complexity and an ability to be used in large-size WASNs.

4- Most of the existing rate-constrained problems do not take the preservation of spatial cues into account when designing the optimal rate allocation algorithms. Can we efficiently link the rate-constrained problem to have a spatially correct rate-constrained noise reduction system?

- To answer this research question, in Chapter seven, inspired by the proposed method in Chapter six, we have proposed a spatially correct rate-constrained noise reduction algorithms, which aims similarly (to the previous chapter) at jointly finding the optimal estimation weights and rates, while additionally preserving the (binaural) spatial information (cues) in the binaural HA setup. Assuming two fusion centers (for the right and left HAs), based on MSE criteria and linear constraints on the spatial information, the optimal weights are found to be the LCMV filter coefficients and the rates are to be the solution to two reverse water-filling problems.

8.1.1. ON THE EFFECT OF QUANTIZATION ON BINAURAL BEAMFORMING FOR HEARING AIDS

In Chapter three, we worked on the first research question, which addresses the effect of quantization noise on binaural beamforming, mentioned earlier in this chapter. We started the road to the rate-constrained multi-microphone noise reduction algorithms based on the MSE criteria. In Chapter three, we explained how the quantization noise, which occurs when imperfect/quantized data is to be processed by the beamformers, is taken into account in noise reduction problems. As an example, we used the binaural multi-microphone MVDR beamformer. First, the signal model is modified by including additive uniform quantization noise. In addition, the correlation matrix model is modified taking into account the correlation matrix of the quantization noise. With this, the MVDR problem is written using the modified noise correlation matrix. The correlation matrix of the quantization noise is assumed to be diagonal and additive, meaning that : 1) the quantization noise is uncorrelated to the signal to be quantized and 2) the quantization noises of different sensors are uncorrelated. As we assumed uniform quantization, we investigated these two assumptions using correlation measures, defined in Chapter three (Section 3.5). The results show that at lower quantization bitrates the cross power spectral density matrix of the quantization noise is not always diagonal, depending on the position of the sound sources with respect to the microphone array. Therefore, we used the dithering technique to decorrelate the signal form the quantization noise, and also to decorrelate the quantization noise across microphones. We concluded by simulations that using dithering, the assumptions on the modified signal model are always valid at all rates and all source positions. We also showed that the output SNR (as a function of bit-rate) of the beamformer, which takes into account the quantization noise correlation matrix is much higher than the one without taking into account the quantization noise correlation matrix.

8.1.2. INFORMATION-THEORETIC STUDY OF RATE-CONSTRAINED NOISE REDUCTION FOR HEARING AIDS

In Chapter four, we answered the second research question on the limitation of the optimal rate-constrained noise reduction for hearing aids in [4]. This limitation is the inevitable requirement of joint statistics at both processors. The problem can be viewed as a source coding problem with/without side information at the decoder which has been well-studied from an information-theoretic viewpoint. To overcome this limitation, we proposed an asymmetric coding framework for rate-constrained noise reduction for hearing aids. For transmission of the information form one HA to the other HA (Link 1), we extended the so-called probability density preserving coding algorithm for vector sources and showed how to retrieve the joint statistics from the probability density preserved quantized data. Then, we used this information to approach the optimal performance in the other link, without knowing the joint statistics in advance. In this chapter, the noise reduction problem is viewed from an information-theoretic perspective. Therefore, we found the theoretic bounds on the performance of the rate-constrained noise reduction for different coding schemes. It note worthy that to implement these algorithms, one needs a sufficient amount of data and estimators to estimate the sources from the observed and quantized data. In the next chapters, more practical

algorithms are presented in which we use simpler coding schemes (like uniform quantizers).

We can conclude that the proposed probability density preserving coding scheme can help us to preserve the statistics of the data (without pre-knowledge of the statistics), while a direct source coding approach, which uses reserve water filling to allocate the rates to different frequencies, may allocate zero bits to some frequency components, and therefore, this information will be lost and cannot be retrieved perfectly at the decoder. However, the proposed approach is only suitable when there are just two processing nodes. The scenarios with more than two nodes are not taken into account in this algorithm. Therefore, in the next chapters, we propose new algorithms that are able to allocate the rates also to additional assistive agents (microphones) to improve the noise reduction and utilize more sensors in the noise reduction for wireless acoustic sensors.

8.1.3. RATE-CONSTRAINED NOISE REDUCTION FOR GENERALIZED BINAURAL HEARING AID SETUPS (SMALL-SIZE WASNS)

In Chapter five, we extended the binaural setup to enable more sensors to contribute to the noise reduction task. Some additional assistive microphones are now included in the setup in order to send additional information to the central fusion center to improve the noise reduction performance. This work is inspired by operational source coding schemes [9], where the algorithms try to find the best bitrate allocation among a cloud of operating points in the rate-distortion space. The proposed method utilizes discrete optimization to allocate the rates across frequency and uses an exhaustive search to allocate the rates across sensors. This enables us to have an optimized rate-constrained noise reduction framework for more than two nodes. Therefore, we tried to answer the third question, which was about the generalization of the binaural rate-constrained noise reduction to more than two processing nodes, mentioned in the introduction chapter. Although the proposed method is simple and effective for rate allocation across frequencies, the algorithm is not scalable with the size of the network, as we used exhaustive search (a non-polynomial search), which becomes intractable when the size of the network grows. Therefore, in Chapter six, we worked on a scalable solution, as detailed in the next section.

8.1.4. RATE-CONSTRAINED NOISE REDUCTION FOR WASNS

In Chapter six, we intended to find a solution to the rate-constrained noise reduction problem for wireless acoustic sensor networks with arbitrary size. Based on the linear estimation concept, we proposed an optimization problem that aims at the estimation of a source, assuming that the remotely-observed microphone signals are quantized and transmitted to a fusion center. We proposed to jointly find the estimation coefficients and bit-rate allocation across both frequency and microphones. The objective function consists of the averaged estimation error between the target signal and its estimate in the frequency domain (MSE). The MSE is a function of the estimation weights and allocation bitrates for all frequencies and microphones. The constraints of the optimization problem are assumed to be linear functions of the rates to limit the total bit budget at which information can be coded and transmitted. We proposed a solution to this non-convex optimization problem (details on why the problem is non-convex can be found

in Chapter six) based on the alternating optimization approach. Under certain assumptions, the estimation coefficients are found to be the rate-dependent Wiener coefficients and the rates are found to be the solution to a weighted reverse water filling problem.

Unlike the previous method in [10], the proposed method does not suffer from the scalability issue, as the complexity order of the proposed solution (which is polynomial) is much less than that of the exhaustive search in [10]. However, one of the important aspects of multi-microphone noise reduction is to preserve the spatial information of the sources while suppressing the undesired part of the acoustic scene. This is not considered in the current rate-constrained problem. In the next chapter, we worked on how to incorporate spatial information preservation in the rate-constrained noise reduction problem.

8.1.5. SPATIALLY CORRECT RATE-CONSTRAINED NOISE REDUCTION FOR WASNS

In Chapter seven, we extended our previous work in Chapter six to bring spatial information preservation to the binaural rate-constrained noise reduction problem. To achieve this, based on the linear estimation philosophy, we proposed a multi-fusion center-based optimization framework, in which the goal is to find the best estimation weights and quantization rates while preserving the spatial cues of the sources. The objective function is the summation of the averaged estimation errors between the target signal and its estimate in the frequency domain (MSE) for both the left and the right side HA processors, which is a function of estimation weights and allocation rates for all frequencies and microphones. The constraints of the optimization problem can be categorized into two groups. 1) The first group is assumed to be linear functions of the rates for both links (transmitting the information from the left side to the right side and vice versa) to limit the total bit budget at which information can be coded and transmitted. 2) The second group of constraints are meant for spatial cue preservation and are linear functions of the estimation weights. Compared to the previous chapter, we proposed an extended solution to this non-convex optimization problem based on the alternating optimization approach. Under certain assumptions, the estimation coefficients are found to be the rate-dependent LCMV coefficients and the rates are found to be the solution to a weighted reverse water filling problem for both transmission links. With the proposed method, we presented a spatially correct rate-constrained noise reduction algorithm for binaural hearing aids in arbitrary size WASNs. Therefore, similar to Chapter six, here we do not have the scalability issue as the size of the network grows. In Chapter seven, we tried to answer the last research question from this dissertation.

Based on the generalized binaural HA setup, where additional assistive microphones are collaborating with HAs, the performance of the proposed method is evaluated using SNR gains and ILD and ITD errors. The results showed that the proposed method outperforms the methods with naive/random choices of rates.

8.2. SUGGESTIONS FOR POSSIBLE FUTURE RESEARCH DIRECTIONS

In this dissertation, our goal was to study different rate-constrained multi-microphone noise reduction problems, given the imperfect (compressed) data. We showed, from information-theoretic optimal quantizers to the simple uniform quantizers, how to link the noise reduction problem to the data compression. In this chapter, we give some suggestions on how to continue this research from different perspectives.

In Chapter four, we studied the information-theoretic source coding algorithms for the binaural noise reduction problem, with only two processing nodes. For example, the well-known WZ source coding was linked to the binaural noise reduction problem. However, the problem of having more than two nodes was not considered in the proposed method. This work can be continued to study the possibility of extending the WZ source coding problems for more than two processing nodes. The question would be how to reduce the redundancy of information contained in each node to be used in other nodes. Or in other words, how to link WZ-based problems for the arbitrary-size WASNs to the noise reduction problem.

In Chapters four, five, six, and seven we proposed different rate-constrained problems based on different scenarios and assumptions. In all the proposed methods, the DFT or KLT transformation was used to process the spatial and temporal information before transmission to the other nodes. We suggest studying more transformation techniques, such as generalized eigenvalue decomposition (G-EVD), to see the effect of different decomposition techniques on the performance of the noise reduction problems. Using different transformation techniques at the encoder side will have an impact on the efficiency of the rate allocation algorithms, therefore, it worth studying it.

In Chapters five, six, and seven we tried to introduce different rate-constrained solutions to the multi-microphone noise reduction in both small-size and arbitrary-size WASNs. One of the underlying assumptions in these proposed methods was that the statistics of the signals does not change rapidly in consecutive speech frames. To deal with the time-varying statistics, we think that we could use the idea of progressive source coding to update our estimated/computed bit-rates based on the updated statistics. This will lead to a notion of time-varying rate-constrained noise reduction, which would of great interest in practical situations, for example in highly non-stationary acoustic scenes.

REFERENCES

[1] D. Marquardt, E. Hadad, S. Gannot, and S. Doclo, “Optimal binaural LCMV beamformers for combined noise reduction and binaural cue preservation,” in 2014 14th International Workshop on Acoustic Signal Enhancement (IWAENC), pp. 288–292, 2014.

[2] R. C. Hendriks, T. Gerkmann; J. Jensen, “DFT-Domain Based Single-Microphone Noise Reduction for Speech Enhancement: A Survey of the State of the Art,” in DFT-Domain Based Single-Microphone Noise Reduction for Speech Enhancement: A Survey of the State of the Art , , Morgan and Claypool, 2013.

[3] T. J. Klasen, T. Van den Bogaert, M. Moonen, and J. Wouters, "Binaural noise reduction algorithms for hearing aids that preserve interaural time delay cues," IEEE Transactions on Signal Processing, vol. 55, no. 4, pp. 1579–1585, 2007.

[4] O. Roy and M. Vetterli, "Rate-constrained collaborative noise reduction for wireless hearing aids," IEEE Transactions on Signal Processing, vol. 57, no. 2, pp. 645–657, 2009.

[5] T. Flynn and R. Gray, "Encoding of correlated observations," IEEE Transactions on Information Theory, vol. 33, no. 6, pp. 773–787, 1987.

[6] T. Berger, *Rate-distortion theory: A mathematical basis for data compression,* Englewood Cliffs, N.J.: Prentice-Hall, 1971.

[7] J. K. Wolf and J. Ziv, "Transmission of noisy information to a noisy receiver with minimum distortion," IEEE Transactions on Information Theory, vol. 16, no. 4, pp. 406–411, 1970.

[8] A. D. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," IEEE Transactions on Information Theory, pp. 1–10, 1976.

[9] Y. Shoham and A. Gersho, "Efficient bit allocation for an arbitrary set of quantizers," IEEE Transactions on Acoustics, Speech, and Signal Processing, vol. 36, no. 9, pp. 1445–1453, 1988.

[10] J. Amini, R. C. Hendriks, R. Heusdens, M. Guo, and J. Jensen, "Operational rate-constrained beamforming in binaural hearing aids," in 26th European Signal Processing Conference (EUSIPCO), 2018.

CURRICULUM VITÆ

Jamal AMINI

19-09-1986 Born in Tehran, Iran. received the B.Sc. degree in computer engineering from Shiraz University, Shiraz, Iran, in 2009, and the M.Sc. degree in electrical engineering from Amirkabir University of Technology (Tehran Polytechnic), Tehran, Iran, in 2011. He is currently a Ph.D. student in the Circuits and Systems (CAS) Group, Faculty of Electrical Engineering, Mathematics and Computer Science, Delft University of Technology. His research interests are on speech enhancement, speech analysis and synthesis, source coding, and voice conversion.