Hybrid Digital-Analog Transmission for Wireless Acoustic Sensor Networks

Matthias Rüngeler, Peter Vary

Institute of Communication Systems and Data Processing (IND), RWTH Aachen University, Aachen, Germany

Abstract

A very crucial task of a Wireless Acoustic Sensor Network (WASN) is the wireless transmission of the acoustic signal to a central fusion node. The positioned microphones experience different radio channel qualities, since they may be, e.g., positioned at different distances to the fusion node. The design of the coding and modulation scheme is usually governed by the worst-case radio channel quality. Mostly, limited complexity of the sensors and/or latency constraints do not allow for a feedback channel and adaptation. The conventional solution employs purely digital transmission using a quantizer in the audio encoder with the unavoidable quantization noise which limits the end-to-end SNR even with increasing radio channel quality. No advantage from good radio channels can be taken. In the fusion node, an audio signal processing algorithm (fusion algorithm) combines the received microphone signals and the output quality depends on the SNR of each received microphone signal. Using purely digital transmission, the quality of output is limited due to the quantization noise.

In this work, the use of Hybrid Digital-Analog (HDA) transmission systems for WASNs is proposed and the transmission quality is analyzed. The benefit is that the quality of the output of the fusion node improves with increasing radio channel qualities. It is shown that for fusion algorithms whose output quality is governed by the SNR of the individual received microphone signals, the HDA system supersedes purely digital transmission for all radio channel qualities. Especially for WASN, with many transmitters which may not be able to adapt to the radio channel quality, HDA transmission systems can show their full potential.

Keywords: Hybrid Digital-Analog (HDA), Purely digital transmission, Saturation due to quantization noise, Maximum Ratio Combining (MRC)

1. Introduction

One key task of a Wireless Acoustic Sensor Network (WASN) is the wireless transmission of an audio signal from several microphones to a fusion node for further processing. The microphones may be positioned at different distances to the fusion node and therefore experience different individual radio channel qualities. Cost and complexity constraints and the possible lack of a feedback channel to each of the microphones impede the adaptation of each transmission link to its individual radio channel quality. Therefore, each transmitter is usually designed for the worst case radio channel to ensure reliable transmission without adaptation.

Each of these transmission links can be regarded as a point-to-point communication for which well known techniques may be employed, separate source and channel coding and an intermediate *binary* representation of the source signal lead to asymptotically optimal solutions [1]. For a given, fixed radio channel quality, or for systems which can adapt to the instantaneous radio channel quality, this approach is widely used in practice. For a fixed, non adaptive transmission system, which operates at a channel quality above the design channel quality, all bits are transmitted without error and the end-to-end SNR of the audio signal is limited by the unavoidable quantization noise introduced in the source encoder. A further improvement of the radio channel quality does not further improve the end-to-end SNR.

Hybrid Digital-Analog transmission systems tackle this challenge for point-to-point communication. In [2] it is shown that an additional transmission of the quantization error using continuous-amplitude means leads to an improvement of the end-to-end SNR with increasing channel quality. In [3, 4] the design of HDA systems

Email addresses: ruengeler@ind.rwth-aachen.de (Matthias Rüngeler), vary@ind.rwth-aachen.de (Peter Vary)

URL: www.ind.rwth-aachen.de (Peter Vary)

Preprint submitted to Special Issue on wireless acoustic sensor networks and ad hoc microphone arrays *December 9, 2014*

using well known digital channel codes is elaborated and it is shown how for every purely digital transmission system a superior HDA system can be designed, while the number of channel uses, i.e., the transmission bandwidth on the channel, and the transmission power is kept constant. Especially for WASN with many microphones and fixed transmission systems which have to cope with different radio channel qualities, HDA transmission systems are of special interest. For each microphone the same fixed HDA system design can be used while the end-to-end SNR is improving with increasing radio channel quality.

At the fusion node, an audio signal processing algorithm (fusion algorithm) combines the received microphone signals to recover an improved audio signal or the position of the audio source. In any case, the quality of the output of the fusion algorithm is dependent on the SNR of the received signals. Depending on the type of fusion algorithm, different dependencies may occur, in some cases the worst signal may dictate the output quality, in other cases the best signal.

Using purely digital transmission systems, the input quality is bounded by the quantization noise and hence, the output quality of the processing in the fusion node is limited by the errors introduced at the beginning of the transmission chain. Using Hybrid Digital-Analog transmission, some microphone signals are received with a better SNR, depending on their radio channel quality which may lead to an improved quality of the output signal of the fusion node.

In Section 2 the system model with an arbitrary fusion algorithm is considered whose output quality may depend in different ways from the input SNR of the received microphone signals. In Section 3 the design of the HDA system is revised and its performance for one example is depicted. Section 4 gives simulation results comparing purely digital and HDA transmission systems for different dependencies of the output quality on the input qualities for a WASN. Section 5 concludes this paper.

2. System Model

2.1. Fusion Node and Noisy Transmission

In the acoustic scene shown in Fig. 1, there may be several target sound sources $(\boldsymbol{s}_a, \boldsymbol{s}_b, \dots)$ and undesired noise sources $(\boldsymbol{n}_a, \boldsymbol{n}_b, \dots)$. For each of the L microphones, there is an individual acoustic transfer function $\mathbf{T}_i$ with $1 \leq i \leq L$ which models the acoustic influence of the scene on the signals before being captured as one

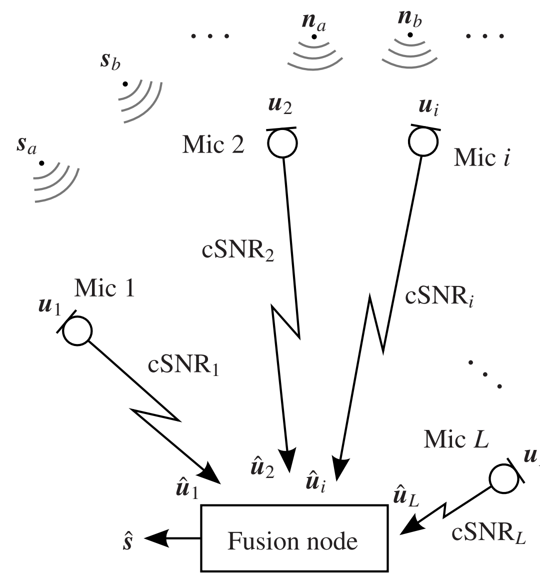

Figure 1: Target sound sources $(\boldsymbol{s}_a, \boldsymbol{s}_b, \dots)$ and undesired noise sources $(\boldsymbol{n}_a, \boldsymbol{n}_b, \dots)$ form an acoustic scene. Several microphones capture audio signals $\boldsymbol{u}_i$ and transmit them to a fusion node where the signals are received as $\hat{\boldsymbol{u}}_i$. Each transmission link $1 \leq i \leq L$ experiences a different radio channel quality cSNR$_i$ and the output of the fusion node is $\hat{\boldsymbol{s}}$.

of the microphone signals $\boldsymbol{u}_i$:

$$\boldsymbol{u}_i = \mathbf{T}_i \left(\boldsymbol{s}_a, \boldsymbol{s}_b, \dots, \boldsymbol{n}_a, \boldsymbol{n}_b, \dots \right) \tag{1}$$

The microphone signal $\boldsymbol{u}_i$ is a vector formed by samples of the waveform captured by microphone i. Each microphone signal is transmitted over the radio link using the same type of transmission system. At the fusion node, for each microphone the audio signal $\hat{\boldsymbol{u}}_i$ is received.

The distortion which is introduced by the transmission depends on the transmission system (e.g., purely digital or Hybrid Digital-Analog) and the channel quality (cSNR$_i$) which may be different on each individual radio link.

The fusion algorithm $\mathbf{F}(\cdot)$ combines the received signals $(\hat{\boldsymbol{u}}_i)$ and obtains an estimate $\hat{\boldsymbol{s}}$ of the desired target signal $\boldsymbol{s}$:

$$\begin{aligned} \hat{\boldsymbol{s}} &= \mathbf{F} \left(\hat{\boldsymbol{u}}_1, \hat{\boldsymbol{u}}_2, \dots, \hat{\boldsymbol{u}}_L \right) \\ &= \tilde{\boldsymbol{s}} + \boldsymbol{e} + \boldsymbol{d} \\ &= \boldsymbol{s} + \boldsymbol{d} \end{aligned} \tag{2}$$

The fusion algorithm may perform beamforming, dereverberation, noise reduction, source separation, etc. The target (clean) signal of the fusion algorithm is $\tilde{\boldsymbol{s}}$ while the influence of the undesired noise sources $(\boldsymbol{n}_a, \dots)$ is modeled by the error $\boldsymbol{e}$. The scope of this paper is the

Published in Special Issue on wireless acoustic sensor networks and ad hoc microphone arrays, Signal Processing Feb. 2015, DOI: 10.1016/j.sigpro.2014.07.005 2/9

analysis of the distortion $\boldsymbol{d}$ in the output signal of the fusion node which is introduced by the transmission. The actual performance of the particular fusion algorithm or the influence of captured noise is not considered here, thus, the signals $\tilde{\boldsymbol{s}}$ and $\boldsymbol{e}$ are combined to $\boldsymbol{s}$. This way, the output of the fusion node is modeled by $\boldsymbol{s}$ and a term $\boldsymbol{d}$ which is dependent on the transmission system and the radio channel quality .

2.2. Generic Purely Digital Transmission

A generic purely digital transmission system is depicted in Figure 2. A continuous-amplitude source vector $\boldsymbol{u}_i$ from microphone i forming one frame with M symbols which could be the captured discrete-time, continuous-amplitude samples or parameters of an audio or speech codec. There are several A/D and D/A converters depicted which use such a high resolution that the influence of the conversion will be neglected in the following. The digital encoder, which includes the source and channel coding as well as the modulator, produces an N dimensional channel vector $\boldsymbol{y}_{\mathrm{D},i}$. In contrast to the A/D convertors in Fig. 2, the distortion introduced by the quantizer in the digital source encoder is taken into account in the following. An Additive White Gaussian Noise (AWGN) channel adds a Gaussian noise vector with the variance $\sigma^2_{n,i}$ per dimension which leads to the following radio channel SNR cSNR_i for microphone i:

$$\mathrm{cSNR}_i = \frac{E\left\{\left\|\boldsymbol{y}_{\mathrm{D},i}\right\|^2\right\}}{\sigma^2_{\boldsymbol{n},i}}. \tag{3}$$

The digital decoder decodes the received vector $\boldsymbol{z}_{\mathrm{D},i}$ for the current frame to the estimate $\hat{\boldsymbol{u}}_{\mathrm{D},i}$ of the initially captured symbols.

The end-to-end parameter SNR for each radio link transmission is described by

$$\mathrm{pSNR}_i = \frac{E\left\{\|\boldsymbol{u}_i\|^2\right\}}{\mathrm{MSE}_i} = \frac{E\left\{\|\boldsymbol{u}_i\|^2\right\}}{E\left\{\left\|\boldsymbol{u}_i - \hat{\boldsymbol{u}}_{\mathrm{D},i}\right\|^2\right\}} \tag{4}$$

where MSE is the mean square error between the source symbols and the decoded symbols.

2.3. Generic Hybrid Digital-Analog Transmission

Figure 3 shows the Hybrid Digital-Analog transmission system. Here an HDA encoder transforms the source vector $\boldsymbol{u}_i$ into two output vectors. The first vector $\boldsymbol{y}^{\mathrm{d}}_{\mathrm{H},i}$ is generated using the same methods as in a purely digital system. This vector is transmitted in the *Digital* branch. The second vector $\boldsymbol{y}^{\mathrm{a}}_{\mathrm{H},i}$ contains a continuous-amplitude refinement to transmit the source vector with a higher fidelity than $\boldsymbol{y}^{\mathrm{d}}_{\mathrm{H},i}$ permits. This vector contains analog symbols which are generated using digital signal processing and subsequent D/A conversion. These symbols are transmitted as analog samples in the *pseudo analog* branch with a certain computational precision (e.g., 16 bit fixed point or even floating point), therefore the word "pseudo" which is omitted in the following. In the *digital* branch, the channel symbols are formed of discrete sets as generated by modulations such as BPSK or QPSK. In the *pseudo analog* branch, the channel symbols may take all values which the D/A convertor is able to generate.

The dimensions of the channel vectors in the digital (D) and the analog (A) branch add up to $N = D + A$, hence the HDA system relies on the same number of channel uses (same bandwidth) as the purely digital system. The vectors are transmitted via an AWGN channel and the HDA decoder estimates the purely digital source representation $\hat{\boldsymbol{u}}^{\mathrm{d}}_{\mathrm{H},i}$ and the pseudo analog refinement $\hat{\boldsymbol{u}}^{\mathrm{a}}_{\mathrm{H},i}$. Adding up these two vectors yields the overall estimate $\hat{\boldsymbol{u}}_{\mathrm{H},i}$ of the initially captured symbols. Again, all depicted A/D and D/A convertors introduce only negligible distortions, yet the effects of the quantizer in the digital part of the HDA encoder are not negligible. The cSNR_i and the pSNR_i of the HDA system are defined as in the purely digital case.

2.4. Combining the Microphone Signals in Fusion Node

A set of possibly different radio channel qualities cSNR_i ($1 \leq i \leq L$) for the microphone transmission links is called "scenario" in the following. For further assessment, not only one scenario is considered. The scenarios are modeled using a statistical description of the channel qualities while the quality of the radio links is time invariant in each scenario.

At the fusion node all microphone signals are received with an individual pSNR_i which depends on the channel quality cSNR_i of the corresponding radio link i and the MSE_i of the employed transmission system. The pSNR_i depends via (4) on the MSE_i, which may be higher or lower, depending on the effectiveness of the transmission system and the individual radio channel quality. As stated in (2) a fusion algorithm uses the received microphone signals to generate the estimate $\hat{\boldsymbol{s}}$ of the target signal which is the sum of the target signal $\boldsymbol{s}$ and undesired components $\boldsymbol{d}$. Depending on the application, there may be more and different target signals, but the considerations taken here can easily be extended to the more general case.

Published in Special Issue on wireless acoustic sensor networks and ad hoc microphone arrays, Signal Processing Feb. 2015, DOI: 10.1016/j.sigpro.2014.07.005 3/9

Figure 2: Purely digital transmission system with a source vector captured by microphone i.

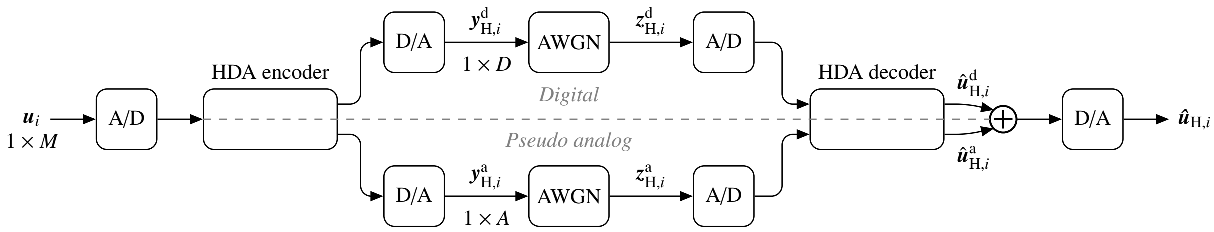

Figure 3: Hybrid Digital-Analog (HDA) transmission system with a source vector captured by microphone i.

The quality fSNR ("fusion" SNR) of the output signal of the fusion node can be defined as follows:

$$\text{fSNR} = \frac{E\left\{||\boldsymbol{s}||^2\right\}}{\text{fMSE}} = \frac{E\left\{||\boldsymbol{s}||^2\right\}}{E\left\{||\boldsymbol{s}-\hat{\boldsymbol{s}}||^2\right\}} = \frac{E\left\{||\boldsymbol{s}||^2\right\}}{E\left\{||\boldsymbol{d}||^2\right\}} \tag{5}$$

The quality in fSNR and hence the distortion fMSE of the output signal of the fusion node depends on the distortion (MSE_i) of the received microphone signals and the properties of the fusion algorithm. Table 1 lists several types of dependencies the fusion algorithms may exhibit.

Type	fMSE =
mean	$E\left\{\frac{1}{L}\sum_{i=1}^{L}\text{MSE}_i\right\}$
minSNR	$E\left\{\max_{1\le i\le L}(\text{MSE}_i)\right\}$
maxSNR	$E\left\{\min_{1\le i\le L}(\text{MSE}_i)\right\}$
Cohadd	$E\left\{\frac{1}{L^2}\sum_{i=1}^{L}\text{MSE}_i\right\}$
MRC	$E\left\{\left(\sum_{i=1}^{L}\frac{1}{\text{MSE}_i}\right)^{-1}\right\}$
add	$E\left\{\sum_{i=1}^{L}\text{MSE}_i\right\}$

Table 1: Typical types of characteristics of the distortion fMSE at the output of the fusion node. Different fusion algorithms exhibit different types of dependencies of the output quality on the quality of the input signals. The expectation is taken over the considered scenarios. The MSE_i refers to the mean square error of microphone signal i.

All types correspond to the behavior of existing fusion algorithms or represent behaviors which complete the table. The "mean" type is motivated by its simplicity while the output distortion is the average of the distortions of the received microphone signals. The algorithm of the "minSNR" type produces a fusion signal which is dominated by the quality of the worst input signal to the fusion node (worst input has the minimum pSNR_i and maximum mean square error (MSE_i)) while for the "maxSNR" type the quality of the best input dominates. The "Cohadd" type corresponds to the behavior of a coherent addition of noisy signals which corresponds to the behavior of a delay and sum beamformer. Algorithms such as the Multichannel Wiener Filter [5] achieve, under certain circumstances, a combination following Maximum Ratio Combining ("MRC" type). One example of the "add" type is the Elko beamformer [6]. Here, the uncorrelated quantization and channel noise of all input signals is added up.

3. Realization of the Purely Digital and Hybrid Digital-Analog Transmission system

Figure 4 shows a conventional purely digital transmission system. In the following, the A/D and D/A conversion blocks and the indices $_i$ to indicate the microphone number are omitted for simplicity. The source emits continuous-amplitude and discrete-time source symbols $\boldsymbol{u}$ forming one frame with dimension $1\times M$ following the probability density function (pdf) $p(u)$. The symbols could be the captured discrete-time, continuous-amplitude samples of an audio signal or even parameters of any source encoder. The entries of the vector $\boldsymbol{u}$ are quantized (Q) and a bit-mapper (BM) generates ℓ_{v_D} source bits, yielding the vector $\boldsymbol{v}_D$. Subsequently, a digital channel encoder followed by a digital modulator transforms the source bits into N real-valued symbols forming the vector $\boldsymbol{y}_D$ with a symbol power of $E\{y_D^2\} = 1$ averaged over all vector entries y_D. Modulation schemes using complex-valued symbols (e.g., QPSK, 8PSK) may also be considered by noting the equivalence between one complex symbol and two real symbols. Since channel coding and modulation (ccm) is combined as one step, the ratio between the number

Published in Special Issue on wireless acoustic sensor networks and ad hoc microphone arrays, Signal Processing Feb. 2015, DOI: 10.1016/j.sigpro.2014.07.005 4 / 9

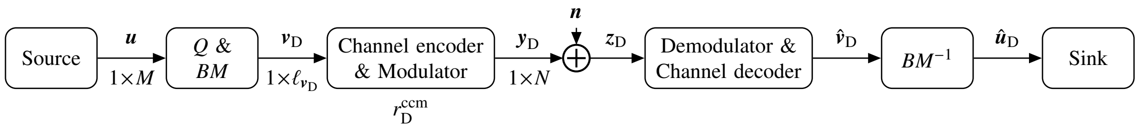

Figure 4: Purely digital transmission system with code rates $r_\mathrm{D}^\mathrm{ccm} = \frac{\ell_{\boldsymbol{v}_\mathrm{D}}}{N}$.

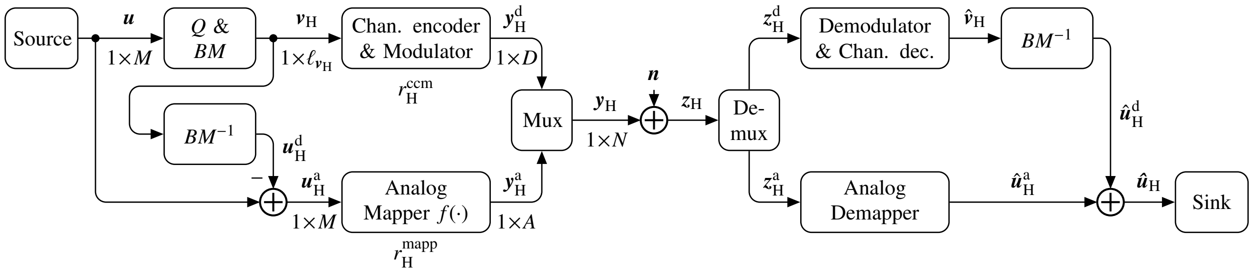

Figure 5: HDA transmission system with code rates $r_\mathrm{H}^\mathrm{ccm} = \frac{\ell_{\boldsymbol{v}_\mathrm{H}}}{N-A} = \frac{\ell_{\boldsymbol{v}_\mathrm{H}}}{D}$ in the digital branch and $r_\mathrm{H}^\mathrm{mapp} = \frac{M}{N-D} = \frac{M}{A}$ in the analog branch.

of bits $\ell_{\boldsymbol{v}_\mathrm{D}}$ and the number of real symbols N is denoted by the rate

$$r_\mathrm{D}^\mathrm{ccm} = \frac{\ell_{\boldsymbol{v}_\mathrm{D}}}{N}. \tag{6}$$

Moreover, the ratio between the number of real source symbols and real channel symbols is given by $r_\mathrm{D} = M/N$.

Additive white Gaussian noise (AWGN) $\boldsymbol{n}$ with variance σ_n^2 per dimension disturbs the channel symbols, thereby yielding the received symbols z_D. After demodulation, channel decoding and reconstruction of quantization levels, $\hat{\boldsymbol{u}}_\mathrm{D}$ gives an estimate of the initial source symbols $\boldsymbol{u}$. In case of hard decision demodulation and channel decoding the reconstruction of quantization levels is a maximum likelihood (ML) estimator. If soft decision demodulation and decoding is employed, also maximum a-posteriori (MAP) or minimum mean square error (MMSE) estimation can be used. In this paper an ML estimator is employed.

Figure 5 illustrates the HDA transmission system [3]. The general idea is to use a conventional digital transmission system for $\boldsymbol{u}$ which is differently designed, though, and additionally transmit the quantization error $\boldsymbol{u}_\mathrm{H}^\mathrm{a}$ by using continuous-amplitude (pseudo analog) discrete-time processing. The upper branch of the hybrid encoder and decoder is the *digital* branch and the lower branch the *pseudo analog*, or in short *analog* branch. All operations, also in the analog branch, are conducted using digital signal processing. The discrete-time continuous-amplitude symbols are represented with a precision depending on the digital processor.

The digital branch is a purely digital transmission system; per frame of M source symbols $\boldsymbol{u}$, the number of real channel dimensions used by the digital branch is D. The analog branch utilizes $A > 0$ channel uses. Thus, the number of channel uses per HDA frame is

$$N = D + A \tag{7}$$

with $D < N$ and the code rate of the channel coding and the modulation in the digital branch is

$$r_\mathrm{H}^\mathrm{ccm} = \frac{\ell_{\boldsymbol{v}_\mathrm{H}}}{N - A} = \frac{\ell_{\boldsymbol{v}_\mathrm{H}}}{D}. \tag{8}$$

In order to compare both systems, the respective numbers of channel uses (N) in the purely digital system and in the HDA system are kept equal.

In the hybrid encoder, scalar quantization Q may be applied to the elements of frame $\boldsymbol{u}$. Alternatively a vector quantizer might be applied to the complete frame or parts of the frame. Then a bit-mapper (BM) generates the source bits $\boldsymbol{v}_\mathrm{H}$. Subsequently, the quantized source representation $\boldsymbol{u}_\mathrm{H}^\mathrm{d}$ is decoded in the transmitter. The distortion $\boldsymbol{u}_\mathrm{H}^\mathrm{a} = \boldsymbol{u} - \boldsymbol{u}_\mathrm{H}^\mathrm{d}$ introduced by the quantizer is processed in the analog branch. The analog mapper uses the continuous-amplitude function $f(\cdot)$ to map the entries of the $\boldsymbol{u}_\mathrm{H}^\mathrm{a}$ to the entries of $\boldsymbol{y}_\mathrm{H}^\mathrm{a}$ with length A and average power $E\{(y_\mathrm{H}^\mathrm{a})^2\} = 1$:

$$\boldsymbol{y}_\mathrm{H}^\mathrm{a} = f(\boldsymbol{u}_\mathrm{H}^\mathrm{a}). \tag{9}$$

The function $f(\cdot)$ can also be defined to work on several entries of $\boldsymbol{u}_\mathrm{H}^\mathrm{a}$ in one step and also output multiple entries of $\boldsymbol{y}_\mathrm{H}^\mathrm{a}$. The ratio between the input and the output

Published in Special Issue on wireless acoustic sensor networks and ad hoc microphone arrays, Signal Processing Feb. 2015, DOI: 10.1016/j.sigpro.2014.07.005 5/9

dimensions of the block is

$$r_{\mathrm{H}}^{\mathrm{mapp}} = \frac{M}{N-D} = \frac{M}{A}. \tag{10}$$

This mapping $f(\cdot)$ could, e.g., be a linear amplification or a nonlinear function with a rate of $r_{\mathrm{H}}^{\mathrm{mapp}} = 1$ or in case of a mapping yielding one complex symbol for one real input symbol $r_{\mathrm{H}}^{\mathrm{mapp}} = 1/2$. In this paper a scaling with $r_{\mathrm{H}}^{\mathrm{mapp}} = 1$ is employed.

After multiplexing the modulated symbols from the digital and the analog branch and transmitting over the AWGN channel, the received symbols are demultiplexed and conveyed to the digital and analog decoding branches.

The analog demapper then gives $\hat{\boldsymbol{u}}_{\mathrm{H}}^{\mathrm{a}}$ as the estimate of the quantization error which can be facilitated using several methods such as ML, MMSE and linear minimum mean square error (LMMSE) estimators. The ML estimator just inverts the effect of (9) whereas the LMMSE estimator additionally weights the received symbols before the inversion with cSNR/(1 + cSNR) [7, 8]. The MMSE estimator additionally considers the source and noise pdf and calculates the conditional expectation $E\{\boldsymbol{u}_{\mathrm{H}}^{\mathrm{a}}|z_{\mathrm{H}}^{\mathrm{a}}\}$. The outputs of the analog and digital branches are added, whereby $\hat{\boldsymbol{u}}_{\mathrm{H}}$ gives the final estimate of the source symbols.

For a fair comparison between the purely digital and the HDA transmission systems, the transmission power should be equal and also the number of channel uses (N) should be the same. It may be clear that with additional channel uses for the analog branch, the HDA system performs better than a purely digital system. But for a fair comparison with a fixed overall number of channel uses N, the number of channel uses D for the digital branch has to be lowered by $D = N - A$. In [3] it has been shown how for any purely digital a superior HDA system system can be designed. The fidelity of the quantizer of the HDA system and thus the number of bits which need to be transmitted is lowered, such that the channel coding rate of the digital channel encoder in the purely digital system is higher than or the same as the channel coding rate in the digital branch of the HDA system. The number of quantization bits of the HDA system cannot be lowered too much, since then loss in fidelity due to coarser quantization cannot be compensated anymore by the analog branch [3]. If additionally an LMMSE estimator is used as the analog demapper in the analog branch, the performance of the HDA system is superior or equal to the purely digital transmission system at all channel qualities. Especially in the context of unknown radio channel qualities, or with channel qualities which may be higher than expected while designing the system, the HDA system exhibits the very desirable property to increase the end-to-end pSNR with rising channel qualities.

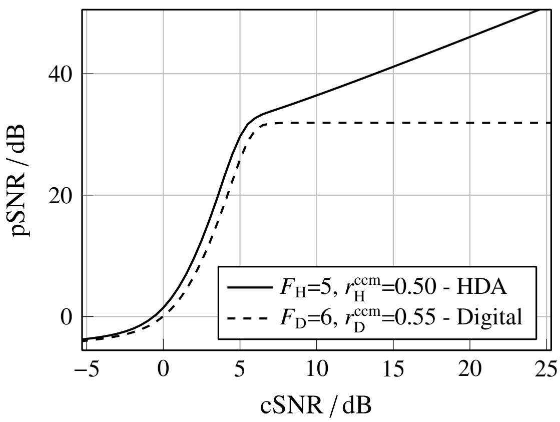

Figure 6: Performance of Hybrid Digital-Analog and purely digital transmission. Gaussian source vector with $M = 80$, Lloyd Max quantization and convolutional coding with puncturing and BPSK modulation. The analog branch uses an LMMSE estimator in the analog demapper. Both simulations employ $N = 880$ channel uses and quantizer with word lengths F_{H} and F_{D} respectively. $\ell_{v_{\mathrm{H}}} = M \cdot F_{\mathrm{H}}$; $\ell_{v_{\mathrm{D}}} = M \cdot F_{\mathrm{D}}$

The LMMSE (and also the MMSE) estimator in the analog branch need the current channel quality. Since, depending on the digital channel coding and modulation, the channel quality is obtained anyways, e.g., using pilots, this channel quality can also be used for the analog branch. In case of non-perfect estimation of the channel quality, the LMMSE estimator exhibits the nice property of graceful degradation; therefore, the loss in performance is not significant.

Figure 6 shows the performance of a purely digital transmission system and an HDA transmission system with an LMMSE estimator in the analog branch. Both systems use the same number of channel uses ($N = 880$), the same transmission power, and encode the same source vectors with length $M = 80$. The source could be an audio signal with a sampling rate of 16 kHz, partitioned to 5 ms blocks. This leads to frames with the size $M = 80$. For simplicity, the pdf of the source is assumed to be Gaussian and in [3] it is shown that also for other pdfs, the HDA systems are superior to purely digital transmission. The digital transmission system employs a scalar Lloyd Max quantizer (LMQ) with $F_{\mathrm{D}} = 6$ bits per symbol leading to a maximum pSNR of 31.9 dB due to the quantization noise. The channel code uses a rate-$\frac{1}{2}$ recursive systematic convolutional code with the generator polynomial $\{1, 15/13\}_8$ —the same code which is used in LTE [9]. Puncturing is used to achieve

Published in Special Issue on wireless acoustic sensor networks and ad hoc microphone arrays, Signal Processing Feb. 2015, DOI: 10.1016/j.sigpro.2014.07.005 6/9

a coding rate of $r_{\mathrm{D}}^{\mathrm{ccm}} = 0.55$. Binary phase shift keying (BPSK) is employed as the modulation. The HDA system employs a quantizer with only $F_{\mathrm{H}} = 5$ bits per symbol which leads to an even more robust channel coding rate of $r_{\mathrm{H}}^{\mathrm{ccm}} = 0.5$. The HDA system exhibits a superior performance for all channel qualities. In contrast to the saturation of the purely digital system, the HDA system increases its performance for improved channel qualities.

For audio and speech coding, usually the correlation in the signal is exploited by linear prediction or by transform coding. In [4] it is shown that HDA systems also show a superior performance in the context of transform coding.

4. Simulation Results

In order to evaluate the performance of WASNs, the radio channel qualities cSNR_i of the individual microphone links have to be chosen. In wireless communication a log-normal distribution of the channel quality is frequently observed which is characterized by a normal distribution of the cSNR in the logarithmic domain. The mean channel quality in dB of all links is set by $\overline{\mathrm{cSNR}|_{\mathrm{dB}}}$ and the variance by Γ in the logarithmic domain across the radio links in the different scenarios. The channel qualities cSNR_i in dB for all L radio links are then modeled by:

$$\mathrm{cSNR}_i|_{\mathrm{dB}} = \mathcal{N}(\overline{\mathrm{cSNR}|_{\mathrm{dB}}}, \Gamma), \qquad 1 \le i \le L. \tag{11}$$

In Figure 7 the performance in fSNR of the output signal of different fusion algorithms with $L = 4$ microphones for a mean radio channel quality $\overline{\mathrm{cSNR}|_{\mathrm{dB}}}$ between -10dB and 50dB and a variance $\Gamma = 5$ is depicted. As a reference, the performance of the transmission with just one microphone with the corresponding cSNR is also shown which is equivalent to the curves in Figure 6.

The curves with the worst performance correspond to the combination type "minSNR". Here the microphone with the lowest pSNR determines the performance of the combined signal. The performance of the "maxSNR" combination type depends on the microphone with the best transmission quality and therefore, in contrast to the "minSNR" type, shows an even better performance than using just one microphone. The "MRC" type combines all received signals in dependence of their individual cSNR_i. This type supersedes all other combination types. E.g., for the digital transmission system with a channel quality for which the

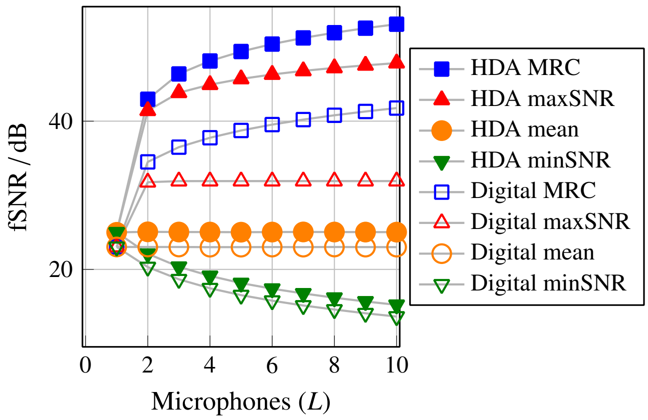

Figure 8: Performance in fSNR of the output signal of the fusion node for varying number L of microphones for a mean radio channel quality $\overline{\mathrm{cSNR}|_{\mathrm{dB}}} = 15$dB and a variance $\Gamma = 5$. For all combination types, the HDA system outperforms purely digital transmission. The gray lines indicate neighboring markers, not interpolated values.

transmission quality has already saturated, the combination types "minSNR" and "maxSNR" show the same performance as using just one microphone, since all individual microphones achieve the same fSNR. But the "MRC" type achieves a gain of 6dB compared to the performance of just one microphone.

The comparison of different combination types for microphone signals is not in the focus of this study. The main focus is the comparison of purely digital and Hybrid Digital-Analog (HDA) transmission in the context of WASNs. Figure 7 clearly shows that the HDA system always supersedes the purely digital transmission systems for the same combination type. Especially for channel qualities in which the performance of the purely digital transmission system saturates due to the unrecoverable noise introduced by the quantizer, the HDA transmission system exhibits its advantages.

In Figure 8 the performance in fSNR of the output signal of different fusion algorithms and a varying number L of microphones for a mean radio channel quality $\overline{\mathrm{cSNR}|_{\mathrm{dB}}} = 15$dB and a channel quality variance $\Gamma = 5$ is depicted. As expected, for just one microphone ($L = 1$), the performance for all combination types is equal. Still, the pSNR in Figure 6 for $\mathrm{cSNR} = 15$dB is not achieved, since in Figure 8 the channel quality is varied according to (11) with $\Gamma = 5$. Two combination types are shown which benefit from an increased number of microphones: the "maxSNR" and the "MRC" type.

The "MRC" type gains most, as well for the HDA as for the digital transmission, when using more micro-

Published in Special Issue on wireless acoustic sensor networks and ad hoc microphone arrays, Signal Processing Feb. 2015, DOI: 10.1016/j.sigpro.2014.07.005 7/9

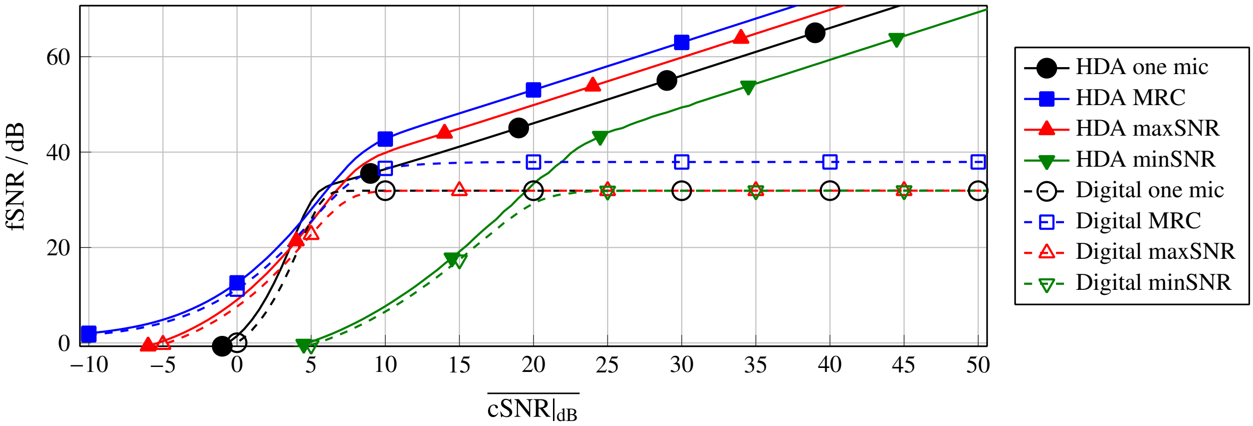

Figure 7: Performance in fSNR of the output signal of the fusion node with of $L = 4$ microphones for a mean radio channel quality $\overline{\mathrm{cSNR}}|_{\mathrm{dB}}$ and a variance $\Gamma = 5$. As a reference, the performance of just one microphone is also depicted. For all combination types, the HDA system outperforms purely digital transmission.

phones. For the HDA transmission system, the gain in fSNR using $L = 10$ microphones instead of just one is remarkable 28 dB. Another notable observation is that the HDA transmission with only 2 microphones supersedes the performance of 10 microphones with purely digital transmission.

The "maxSNR" combination type leads to a gain for the HDA system for additional microphones, since with more microphones the probability of a high cSNR_i is increased and thus also the performance of the best microphone. In case of a purely digital transmission system, an increased channel quality is not rewarded with an improved fSNR, since the performance of the purely digital system has already saturated. Thus, the "maxSNR" type does not benefit from more microphones with a purely digital transmission system.

When averaging the performance of all microphones ("mean"), the performance is independent of the amount of microphones, both for the HDA and for the digital system. With the "minSNR" and the "add" (not depicted) type, the performance of the output signal of the fusion node decreases with more microphones.

Independent of the combination type, it can be observed that again the systems using the HDA transmission show a superior performance to the purely digital transmission. Even with the "maxSNR" type, purely digital transmission does not lead to a gain with additional microphones, but HDA transmission enables an additional gain for an increased number of microphones.

5. Conclusion

Usually, microphones in Wireless Acoustic Sensor Networks (WASNs) use the same type of transmission system to transmit their acoustic signals to a fusion node. The link from each microphone experiences a different radio channel quality due to a different distance to the fusion node, while the transmission system cannot adapt to the radio channel quality because of a missing feedback channel or cost, latency or complexity constraints. Purely digital transmission systems cannot exceed a certain end-to-end SNR due to the inherent quantization noise which is dictated by the overall design for the worst case radio channel quality. One of the strengths of a Hybrid Digital-Analog (HDA) transmission system is the constant improvement of the end-to-end SNR for an increased radio channel quality, also for non-adaptive transmitters.

Independent of the behavior of the fusion algorithm in terms of output quality in dependence of the quality of the received microphone signals, the HDA transmission system supersedes purely digital transmission. It gains most when the fusion algorithm can translate an improved input quality to an increased output quality.

Especially for WASNs with the inherent wide range of radio channel qualities of the wireless link to the individual microphones, HDA transmission leads to an improved performance compared to purely digital transmission.

[1] C. Shannon, A mathematical theory of communication, Bell System Technical Journal 27 (1948) 623–656.

[2] M. Skoglund, N. Phamdo, F. Alajaji, Design and performance

Published in Special Issue on wireless acoustic sensor networks and ad hoc microphone arrays, Signal Processing Feb. 2015, DOI: 10.1016/j.sigpro.2014.07.005 8/9

of VQ-based hybrid digital-analog joint source-channel codes, IEEE Transactions on Information Theory 48 (2002) 708–720.

[3] M. Rüngeler, P. Vary, Surpassing purely digital transmission: A simplified design of hybrid digital analog codes, in: 9th International ITG Conference on Systems, Communications and Coding 2013 (SCC'2013), Munich, Germany, 2013.

[4] M. Rüngeler, P. Vary, Hybrid digital analog transform coding, in: Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), IEEE, 2013, pp. 5109–5113.

[5] S. Doclo, Multi-microphone noise reduction and dereverberation techniques for speech applications, Ph.D. thesis, Katholieke Universiteit Leuven, 2003.

[6] G. Elko, A.-T. N. Pong, A simple adaptive first-order differential microphone, in: Applications of Signal Processing to Audio and Acoustics, 1995., IEEE ASSP Workshop on, 1995, pp. 169–172. doi:10.1109/ASPAA.1995.482983.

[7] O. Fresnedo, F. Vazquez-Araujo, L. Castedo, J. Garcia-Frias, Low-complexity near-optimal decoding for analog joint source channel coding using space-filling curves, Communications Letters, IEEE 17 (2013) 745–748.

[8] S. M. Kay, Fundamentals of Statistical Signal Processing: Estimation Theory, Prentice-Hall Signal Processing Series, 1993.

[9] 3GPP TS 36.212, Evolved Terrestrial Radio Access (E-UTRA); Multiplexing and Channel Coding, 3GPP Technical Specification Group Radio Access Network, 2011.

Published in Special Issue on wireless acoustic sensor networks and ad hoc microphone arrays, Signal Processing Feb. 2015, DOI: 10.1016/j.sigpro.2014.07.005 9 / 9