

WIDEBAND SPEECH CODING WITH HYBRID DIGITAL-ANALOG TRANSMISSION

Matthias Rüngeler, Fabian Kleifgen and Peter Vary

Institute of Communication Systems and Data Processing (**ind**)
RWTH Aachen University, Germany
{ruengeler|vary}@ind.rwth-aachen.de

ABSTRACT

Efficient digital transmission of speech requires source coding which comes at the price of unavoidable quantization errors. Thus, even in clear channel conditions, the quality of the decoded speech signal is limited due to the quantization errors. Hybrid Digital-Analog (HDA) codes circumvent this limitation by additionally transmitting the quantization error with quasi-analog methods (discrete-time, quasi-continuous-amplitude) with neither increasing the total transmission power, nor the occupied frequency bandwidth on the radio channel. So far, the HDA concept has mainly been applied to random parameters. In this paper, the HDA concept is adapted to the transmission of wideband speech signals using PCM and ADPCM coding. By experimental verification it is shown that the HDA concept may outperform conventional purely digital transmission systems at all channel qualities while additionally eliminating the quality saturation effect.

Index Terms— Hybrid Digital-Analog (HDA) transmission, lattice ADPCM speech coding

1. INTRODUCTION

Conventional digital speech transmission systems are based on the source-channel separation theorem. The digitized analog speech signal is processed by a speech encoder which delivers a sequence of quantized source parameters (e.g., coefficients, gain factors, filtered samples). Thereafter, the bit stream of the source parameters is protected by a channel encoder. These systems are usually optimized for the expected worst-case channel condition, e.g., at the cell edge of a cellular radio system. However, if the receiver experiences a better channel quality (cSNR), the transmission fidelity does not exceed the design transmission fidelity and, thus, the signal to noise ratio (pSNR) of the decoded source parameters saturates.

This saturation is caused by the inherent errors introduced by the digital speech encoder which, even in the case of error-free transmission of the coded bits, can never be compensated for. To combat this saturation with increasing channel qualities, one common solution is using adaptive multi rate (AMR) systems which require a feedback channel. But they may not be used, e.g., in broadcast systems with different channel qualities for different receivers and unicast systems with

rapidly changing channel quality. Also systems with inexpensive transmitters not designed to adapt to changing channels, e.g., transmitters used for wireless microphones or headsets, may not benefit from the AMR principle.

The saturation of the transmission fidelity can also be circumvented by adding continuous-amplitude processing and transmission. With these Hybrid Digital-Analog (HDA) transmission systems, the transmission fidelity improves for increasing channel qualities without the need for adapting the configuration of the source and channel encoding.

In the literature, some HDA transmission designs are compared to purely digital transmission. [1, 2, 3] focus on speech transmission but show a superior quality just for some specific channel qualities. For random variables which are either independent and identically distributed (i.i.d.) or follow a Gauss-Markoff correlation, Skoglund et al. [4] have proposed a system heavily relying on numerical optimizations and exhaustive search in the decoding algorithms. In [5, 6] an approach for HDA transmission benefiting from the power and flexibility of conventional digital channel codes is presented. Additionally, it is shown that to any purely digital transmission system, a superior HDA transmission system can be designed which uses the same transmission resources.

In this paper, first the key results from [5], how to design superior HDA transmission systems for i.i.d. random variables are reviewed in Sec. 2. The system is adapted to speech signals in Sec. 3 using block adaptive PCM. An even more powerful ADPCM approach and different options to benefit from HDA transmission are shown in Sec. 4. Section 5 presents simulation results.

2. HYBRID DIGITAL-ANALOG TRANSMISSION AND PURELY DIGITAL TRANSMISSION

Fig. 1 shows a conventional digital transmission system. Here, the source signal¹ s consists of realizations of an i.i.d. continuous-amplitude and discrete-time random variable with dimension $1 \times M$ and probability density function (pdf) $p(s)$. The entries of the vector s are quantized (Q) and a bit-mapper (BM) generates in total ℓ_{v_D} source bits, yielding the bit vector v_D . Subsequently, a digital channel code followed by digital modulation transforms the source bits into N real-valued symbols forming the vector y_D . Since channel coding

¹In [5] u is used instead of s .

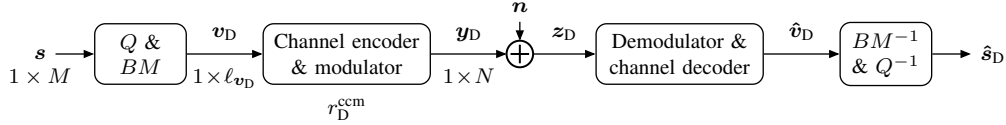


Fig. 1. Purely digital transmission system for a random variable s .

and modulation is combined in one step (ccm), the ratio between the number of bits l_{v_D} and the number of real symbols N is denoted by the coding-modulation rate $r_D^{\text{ccm}} = \frac{l_{v_D}}{N}$. Additive white Gaussian noise n disturbs the channel symbols, thereby yielding the received symbols z_D . The channel signal to noise ratio is

$$\text{cSNR} = \frac{E\{\|y_D\|^2\}}{E\{\|n\|^2\}}. \quad (1)$$

After demodulation, channel decoding, and reconstruction of quantized values, \hat{s}_D gives an estimate of the initial source symbols s .

Fig. 2 illustrates the corresponding HDA transmission system [5]. The general idea is to use a conventional digital transmission system for s and to additionally transmit the quantization error s_H^a by using continuous-amplitude (pseudo-analog) discrete-time processing. The upper branch of the hybrid encoder and decoder is referred to as the *digital* branch and the lower branch as the *analog* branch. All operations, also in the analog branch, are conducted by digital signal processing. The continuous-amplitude symbols are floating or fixed point variables with a precision depending on the digital processor. All variables in the figures are already digitized. The necessary A/D and D/A conversion is assumed to have a high enough fidelity to not affect the results.

The digital branch is a purely digital transmission system; per frame of M source symbols s , D real channel dimensions are used by the digital branch. The analog branch takes $A > 0$ channel uses. Thus, the number of channel uses per HDA frame is $N = D + A$ with $D < N$ and the coding-modulation rate in the digital branch is $r_H^{\text{ccm}} = \frac{l_{v_H}}{N-A} = \frac{l_{v_H}}{D}$. To compare both systems, the respective numbers of channel uses (N) in the digital system and in the HDA system are kept equal.

In the hybrid encoder, quantization Q is applied to the elements of frame s and a bit-mapper (BM) generates the bit vector v_H . The quantized source representation s_H^d is decoded

locally in the transmitter. The quantization error $s_H^a = s - s_H^d$ is processed in the analog branch. The analog mapper uses the continuous-amplitude function $y_H^a = f(s_H^a)$ to map the entries of s_H^a to the entries of y_H^a with length A and average power $\frac{1}{A} E\{\|y_H^a\|^2\} = 1$. The ratio between the input and the output dimensions of the block is $r_H^{\text{mapp}} = \frac{M}{A}$. This mapping $f(\cdot)$ could, e.g., be a linear amplification or a nonlinear function with a rate of $r_H^{\text{mapp}} = 1$ or in case of a mapping yielding one complex symbol for one real input symbol $r_H^{\text{mapp}} = \frac{1}{2}$. In this paper, a scaling with $r_H^{\text{mapp}} = 1$ ($A = M$) is employed.

After multiplexing the symbols from the digital and the analog branch and transmission over the AWGN channel, the received symbols are demultiplexed and conveyed to the digital and analog decoding branches. The analog demapper then gives \hat{s}_H^a as the estimate of the quantization error which can be facilitated using several alternative methods such as maximum likelihood (ML), minimum mean square error (MMSE) and linear minimum mean square error (LMMSE) estimators. The ML estimator applies the inverse function of $f(\cdot)$ whereas the LMMSE estimator additionally weights the received symbols with $\frac{\text{cSNR}}{(1+\text{cSNR})}$ before applying the inverse function of $f(\cdot)$ [7]. The MMSE estimator additionally incorporates the source and noise pdf and calculates the conditional expectation $E\{s_H^a | z_H^a\}$. The outputs of the analog and digital branches are added, whereby \hat{s}_H gives an estimate of the initial source symbols. Finally, the end-to-end parameter signal to noise ratio for both systems is described by

$$\text{pSNR} = \frac{E\{\|s\|^2\}}{E\{\|s - \hat{s}\|^2\}}. \quad (2)$$

For a fair comparison between the purely digital and the HDA transmission systems, the transmission power as well as the number of channel uses (N) should be equal. Thus, the number of channel uses D for the digital branch has to be lowered to $D = N - A$. This can be achieved by lowering the fidelity of the quantizer in the HDA system. Then fewer bits

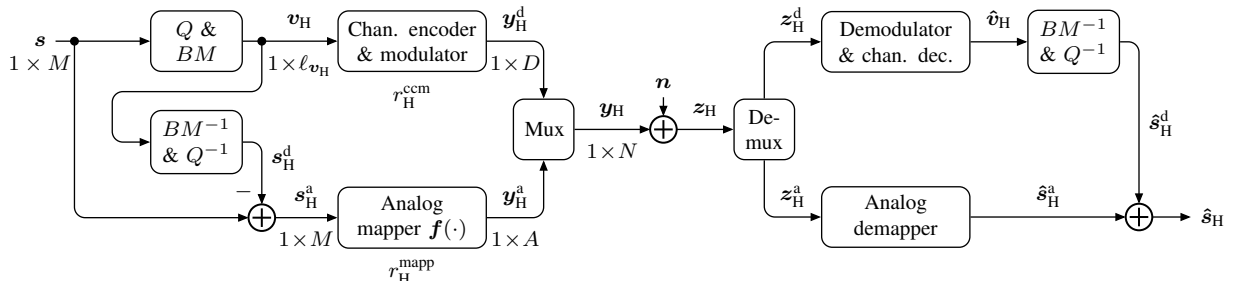


Fig. 2. HDA transmission system for a random variable s .

need to be protected against channel noise and thus the same or even stronger channel coding can be applied ($r_D^{\text{ccm}} \geq r_H^{\text{ccm}}$) compared to the purely digital case. The number of quantization bits of the HDA system cannot be lowered too much, since then the loss in fidelity due to coarser quantization cannot be compensated anymore by the analog branch. In [5], it is stated by how much the fidelity can be lowered to design a superior HDA system for any purely digital system. 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 of increasing the end-to-end pSNR with rising channel qualities.

In [6], this system is extended with block transform coding to exploit the redundancy in Gauss-Markoff sources. It can be shown that also here, the HDA system exhibits a superior performance to purely digital transmission for all channel qualities.

3. HDA BLOCK ADAPTIVE PCM TRANSMISSION

In contrast to the previous section, here, speech signals are used whose short term statistics (e.g., signal level) change rapidly. The analog mapper in Fig. 2 which ensures an average power of 1 of y_H^a cannot be parametrized without knowing the actual samples. Thus, for speech transmission, block wise adaptation is needed. The same holds for the range of the input signal s .

Therefore, in Fig. 3 two normalization blocks are added compared to Fig. 2. A block normalization scales the input vector to use the full input range of the subsequent quantizer Q (linear, logarithmic or vector quantization), yielding adaptive quantization forward (AQF). The power normalization scales the quantization error u_H^a to ensure that for any input signal, the vector y_H^a always has a power of $\frac{1}{A} \|y_H^a\|^2 = 1$.

The scaling factors have to be revealed to the receiver, thus, L_{B1} bits are spent for the block normalization factor and L_{B2} bits are used for the power normalization factor. These bits are multiplexed with the quantization bits v_H , protected

by channel coding and transmitted using the digital branch of the HDA system. The inverse to the power normalization in the receiver additionally weights the received values with $\frac{c\text{SNR}}{(1+c\text{SNR})}$ to achieve LMMSE performance.

The purely digital system in Fig. 1 is also adapted to speech signals by adding the block normalization (system is not depicted here). To ensure a fair comparison, the encoded bits are punctured to yield the same vector length N for all systems.

4. HDA ADPCM CODING

In this section, more sophisticated speech coding than PCM is combined with HDA transmission. A lattice ADPCM speech coder [8] is used as the basis and is depicted in the gray boxes in Fig. 4. A special feature of this codec is the use of an allpass prefilter with a delay of L_{AP} samples as a smearing filter to improve the coding performance for transient sounds. Also adaptive quantization backwards (AQB) is used for quantizing the prediction error. When designing HDA systems by extending purely digital transmission, the question arises which signals are transmitted via the digital and analog branch. Here, via the digital branch, the bits produced by the original codec, i.e., the quantization bits v of the signal u , are transmitted. For the analog branch, at least two possibilities arise. The simplest and always applicable solution is to regard the purely digital version of the speech coder as a black box. The speech signal is encoded and the resulting bit stream is decoded at transmitter side while the difference between the decoded and original signal is transmitted via the analog branch. This solution is depicted in Fig. 4 with a) when positioning the switches to Δs and $\Delta \hat{s}$, respectively. The decoded signal is already available as s'_{AP} in the encoder. The effect of the allpass has to be inverted and its delay has to be compensated before calculating the signal for the analog branch.

Potentially more promising are some approaches which derive the signal for the analog branch from variables inside the speech coder. Here, the component which introduces the distortion is the quantizer of the signal u . Consequently, this distortion should directly be transmitted via the analog branch. However, since the speech codec uses closed loop prediction, the encoder needs to know how the prediction co-

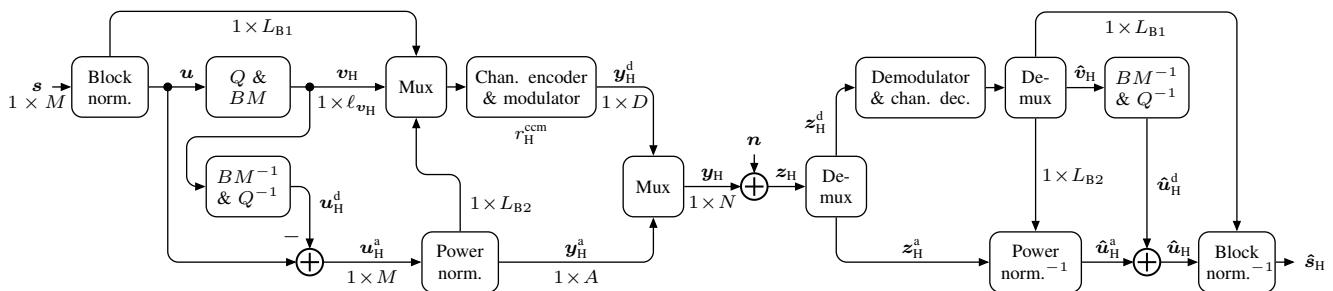


Fig. 3. HDA block adaptive PCM transmission system.

efficients will be estimated in the decoder. Thus, the same prediction error has to be used for the estimation of the prediction coefficients in the encoder as well as in the decoder. Therefore, adding the received quantization error $\Delta\hat{u}$ to the received prediction error would generate different estimates of the prediction coefficients in the decoder than in the encoder. This could lead to annoying artifacts or even unstable behavior. The solution is depicted in Fig. 4 with b). Only the received bits are used for generating the prediction coefficients without considering the analog branch. If the bits are received without error, the same prediction coefficients are estimated in the decoder as well as in the encoder. The same holds for the intensity block. The received quantization error $\Delta\hat{u}$ is not discarded, but weighted with \hat{i} and added to the result from the purely digital branch of the decoder. Filtering is not necessary, since in closed loop prediction, the quantization error $\Delta\mathbf{u} \cdot \mathbf{i}$ distorts the output signal of the decoder without further filtering. This way, even in a closed loop prediction scenario, the analog branch improves the performance.

5. SIMULATION RESULTS

For all simulations, samples from the NTT database are employed. 96 sentences in “English” are selected which sum up to a duration of 12 minutes and 48 seconds. They are partitioned to blocks of $M = 80$ samples (5ms @ 16kHz sampling rate). The output bits of the source encoder and the bits from the normalization are channel encoded using convolutional coding with a rate- $\frac{1}{2}$ recursive systematic convolutional code with the generator polynomial $\{1, 15/13\}_8$ — the same code which is used as a component code in the UMTS-LTE Turbo code. To guarantee a fair comparison between all systems, puncturing is applied to the codewords to always yield $N = 560$ channel uses per 5ms. In all simulations BPSK modulation is employed. In the HDA case, $D = 480$ chan-

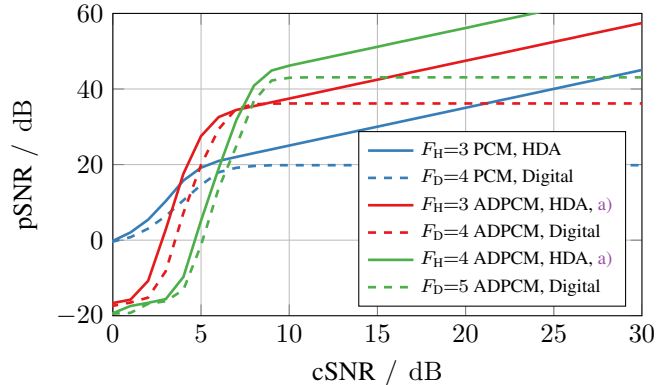


Fig. 5. pSNR for PCM and ADPCM source coding with purely digital and HDA transmission. $M=80$. HDA: $A=80$, $D=480$, $N=560$. Digital: $N=560$. ADPCM HDA case for a) and $L_{B2}=16$, PCM HDA case with $L_{B1}=L_{B2}=10$, PCM digital case with $L_{B1}=10$.

nel uses are allocated to the digital branch and $A = M = 80$ channel uses to the analog branch. LMMSE estimation is employed in the analog branch at the receiver. Fig. 5 shows the performance in pSNR for PCM and ADPCM source coding with purely digital and HDA transmission. The design rules in [5] for HDA systems recommend lowering the fidelity of the quantizer in the HDA case (F_H) by $r_D^{cm} = 0.71$ bit per source symbol for $F_D = 5$ to $F_H = 4.29$ and by $r_D^{cm} = 0.57$ for $F_D = 4$ (i.e., $F_H = 3.43$). Since the ADPCM speech encoder just supports integer numbers of bits per symbol, $F_H = 4$ and $F_H = 3$ are used. It can be seen that PCM shows a lower performance than ADPCM, since it cannot exploit the correlation in the signal. In the HDA ADPCM case, $\Delta\mathbf{s}$ is transmitted. All Hybrid Digital-Analog systems meet or outperform the corresponding purely digital systems for all channel qualities. Just the HDA curve with $F_H = 3$ overlaps with the curve for $F_D = 4$ at $cSNR|_{dB} = 7$ dB due to not completely following

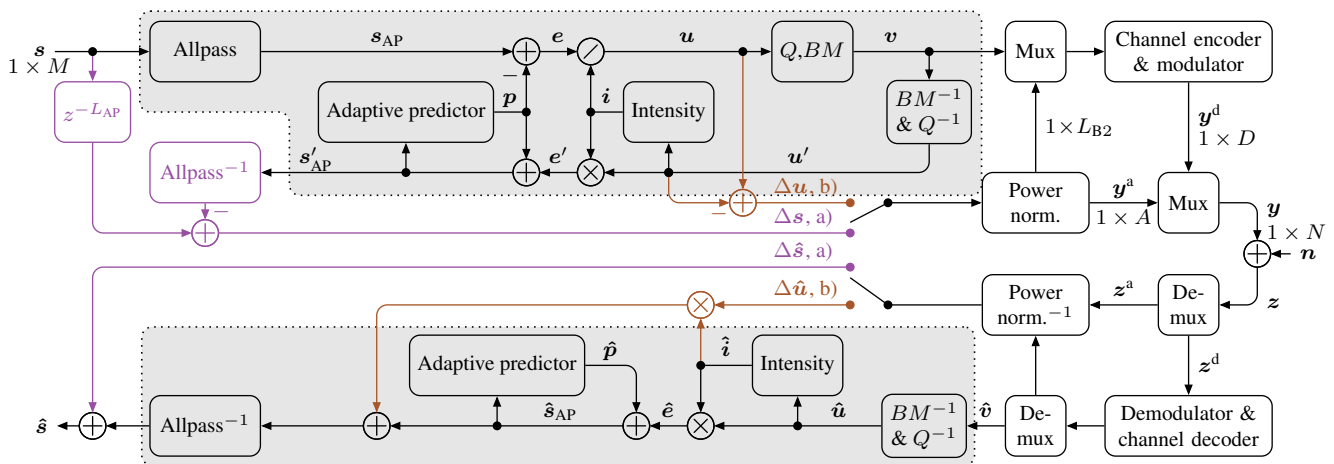


Fig. 4. HDA ADPCM speech coding system. Transmission of a) speech reconstruction error or b) quantization error of prediction error in the analog branch.

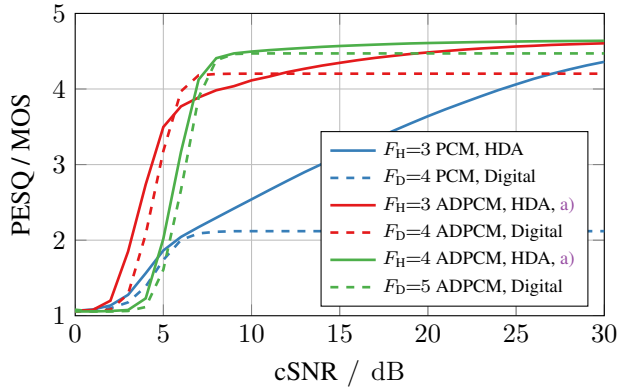


Fig. 6. PESQ for PCM and ADPCM source coding with purely digital and HDA transmission with parameters as in Fig. 5.

the design rules. Similar observations can be made in Fig. 6 for the PESQ measure [9].

In Fig. 7 the performance of the two options a) and b) in ADPCM HDA transmission is compared. Both options benefit from an increased channel quality and improve the transmission fidelity. Using a block length of 5ms ($M = 80$), the performance in pSNR is better for a) than for b). For longer block lengths (30ms, $M = 480$, $N = 3360$), the performance of a) remains the same, while for b) it improves a little but stays below the performance of a).

For the PESQ measure, the effect is reversed (Fig. 8). For bigger blocks, the performance of a) decreases significantly while the performance of b) increases a little and supersedes a). Thus, for bigger blocks, using b) is the better choice. The reason for this is the higher stationarity of the signal Δu (b) compared to Δs (a). This leads to a better perceptual quality when applying the power normalization to longer blocks.

6. CONCLUSION

In this paper the Hybrid Digital-Analog transmission systems for random variables are successfully adapted to speech transmission. PCM coding is enhanced with block wise normalization while the quantization error is transmitted via the analog branch. The more powerful ADPCM speech coding is

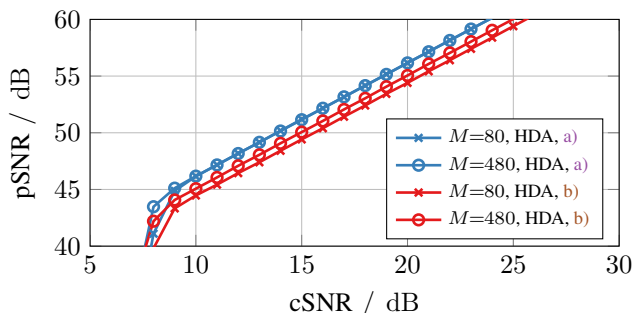


Fig. 7. pSNR for ADPCM source coding with different block lengths M and different options to obtain the signal for the analog branch. $F_H=4$.

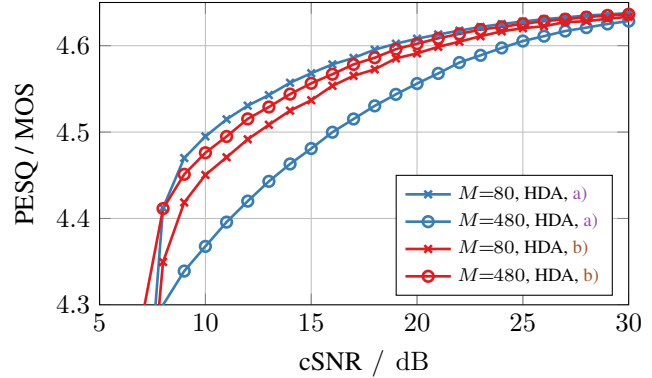


Fig. 8. PESQ for ADPCM source coding with different block lengths M and different options to obtain the signal for the analog branch with parameters as in Fig. 7.

adapted to transmit either the error resulting when quantizing the prediction error, or the speech reconstruction error as the analog signal. Simulations show the superiority of block adaptive HDA PCM and HDA ADPCM transmission over conventional PCM and ADPCM transmission of speech signals. This holds for all channel conditions for pSNR and for the majority of channel conditions for PESQ.

Acknowledgement

Thanks to Michael Czarnetzki for programming support and fruitful discussions.

7. REFERENCES

- [1] T. Miki, C.-E. Sundberg, and N. Seshadri, "Pseudo-analog speech transmission in mobile radio communication systems," *Vehicular Technology, IEEE Transactions on*, vol. 42, no. 1, Feb. 1993.
- [2] N. Phamdo and U. Mittal, "A joint source-channel speech coder using hybrid digital-analog (hda) modulation," *IEEE Transactions on Speech and Audio Processing*, vol. 10, no. 4, 2002.
- [3] C. Hoelper and P. Vary, "Bandwidth Efficient Mixed Pseudo Analogue - Digital Speech Transmission," in *Proceedings of European Signal Processing Conference (EUSIPCO)*. EURASIP, Sep. 2006.
- [4] M. Skoglund, N. Phamdo, and F. Alajaji, "Design and Performance of VQ-based Hybrid Digital-Analog Joint Source-Channel Codes," *IEEE Transactions on Information Theory*, vol. 48, no. 3, Mar. 2002.
- [5] M. Rüngeler, J. Bunte, and P. Vary, "Design and Evaluation of Hybrid Digital-Analog Transmission Outperforming Purely Digital Concepts," *Communications, IEEE Transactions on*, vol. 62, no. 11, Nov 2014.
- [6] M. Rüngeler and P. Vary, "Hybrid Digital Analog Transform Coding," in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. Vancouver: IEEE, May 2013.
- [7] O. Fresnedo, F. Vazquez-Araujo, L. Castedo, and J. Garcia-Frias, "Low-Complexity Near-Optimal Decoding for Analog Joint Source Channel Coding Using Space-Filling Curves," *IEEE Communications Letters*, vol. 17, no. 4, 2013.
- [8] M. Dietrich, "Performance and implementation of a robust ADPCM algorithm for wideband speech coding with 64 kbit/s," in *Proc. of Intl. Zürich Seminar on Digital Communications*, Switzerland, 1984.
- [9] A. Rix, J. Beerends, M. Hollier, and A. Hekstra, "Perceptual evaluation of speech quality (PESQ) - A new method for speech quality assessment of telephone networks and codecs," in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Salt Lake City, Utah, USA, 2001.