Bandwidth-Efficient Mixed Pseudo Analogue-Digital Speech and Audio Transmission

Carsten Hoelper, Peter Vary
Institute of Communication Systems and Data Processing (ind), RWTH Aachen University
Muffeter Weg 3a, D-52056, Aachen, Germany
phone: + (49) 241 8026979, fax: + (49) 241 8022186, email: {hoelper,vary}@ind.rwth-aachen.de
web: www.ind.rwth-aachen.de

Abstract— **Today's speech and audio coding and transmission systems are either analogue or digital, with a strong shift from analogue systems to digital systems during the last decades. In this paper, both digital and analogue schemes are combined for the benefit of saving transmission bandwidth, complexity, and of improving the achievable quality at any given signal-to-noise ratio (SNR) on the channel.**

The combination is achieved by transmitting pseudo analogue samples of the unquantized residual signal of a linear predictive digital filter. The new system, Mixed Pseudo Analogue-Digital (MAD) transmission, is applied to narrowband speech as well as to wideband speech and audio. MAD transmission over a channel modeled by additive white Gaussian noise (AWGN) is compared to the GSM Adaptive Multi-Rate speech codec mode 12.2 kbit/s (Enhanced Full-Rate Codec), which uses a comparable transmission bandwidth if channel coding is included and to PCM transmission in the case of audio signals.

I. Introduction

While analogue speech and audio transmission systems suffer badly from high transmission noise, digital systems can completely recover the source signal as long as the channel coding applied is strong enough and the received energy per bit is sufficient according to channel coding.

If the channel SNR increases, the output quality of a digital system will remain constant, even if no errors occur at all. The output quality is limited by the source coder design. For analogue systems, the minimum bandwidth B_a required for the transmission of a speech or audio signal equals the audio bandwidth B_{audio}, see e.g. [6]. Digital systems generally require a higher bandwidth. In this paper we consider digital BPSK (Binary Phase Shift Keying) modulation with a symbol rate $R_d = \frac{1}{T}$ and a Root Raised Cosine (RRC) pulse shaping filter with a roll-off factor $\alpha = 0.5$ requiring a bandwidth of $B_d = R_d(1 + \alpha)$ [4] for bandpass transmission.

In [2] a joined source-channel hybrid digital-analogue (HDA) vector quantization (VQ) scheme is presented. A digital channel is used for transmitting the VQ codebook index of the quantized version of a vector of input samples and an analogue channel (time-discrete, continuous amplitude) is used to transmit the quantization error. Thus, the receiver gets a quantized (digital) representation of the signal and additionally a refinement signal with continuous amplitude.

In contrast to this we follow the basic idea formulated by T. Miki for ADPCM (Adaptive Differential Puls Code Modulation) [1] and propose a hybrid scheme as shown in Figure 1, which requires

- a digital channel for transmitting the following parameters:
  - prediction coefficients a_i
  - gain factors g
- and a pseudo-analogue channel for transmitting discrete-time samples $r_n = r \cdot g$ of the prediction residual r.

It will be shown that this approach is very efficient with respect to the required transmission bandwidth and that it allows to exploit the mechanisms of linear predictive coding (LPC) and noise shaping to produce high quality speech and audio.

While the general scheme of the Mixed Pseudo Analogue-Digital (MAD) transmitter is depicted in Figure 1, the detailed operation is given in Figures 3 (transmitter), 4 (transmission) and 5 (receiver). The objective of MAD transmission is to maximize the subjective quality while minimizing the required transmission bandwidth and coding complexity. It is well known that this objective does generally not yield an MMSE optimum system. The MAD transmission scheme is not adapted to a specific channel SNR and performs well in most channel coditions [3].

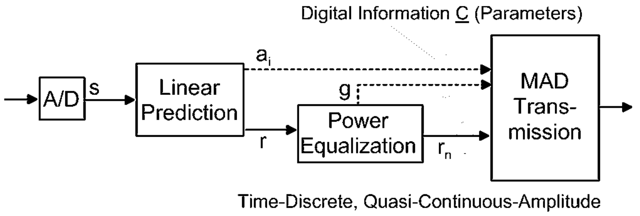
Digital Information C (Parameters)
a_i
g
A/D
s
Linear Prediction
Power Equalization
r
r_n
MAD Trans-mission
Time-Discrete, Quasi-Continuous-Amplitude

Fig. 1. Mixed Pseudo Analogue-Digital Speech and Audio Transmission: Principle.

The remainder of the paper is structured as follows: First the LP filters for narrowband and wideband MAD transmission are introduced in section II-A. Section II-B concentrates on the transmission power, section II-C deals with channel coding, while in section II-D the baseband transmission is reviewed. Section III-A compares the narrowband MAD transmission system to the GSM Enhaced Full-Rate speech codec. In section III-B the wideband version of our MAD transmission

0-7803-9752-5/06/$20.00 ©2006 IEEE

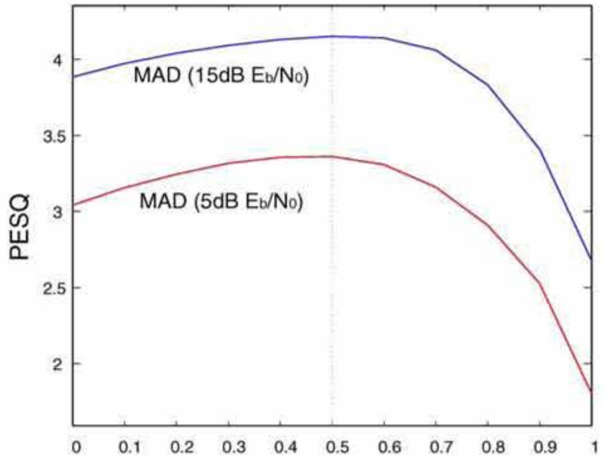

Fig. 2. Influence of prediction strength γ on the subjective speech quality measure PESQ (Perceptual Estimation of Speech Quality [13]).

system is evaluated for speech and audio. Finally section III-C introduces Audio MAD transmission for signals with 22 kHz or 32 kHz sampling frequency. The paper ends with a conclusion.

II. Proposed Method

A. Digital Linear Prediction

Linear Prediction, e.g. [7], [5], has proved to be very effective to code speech and audio signals and is used e.g. in almost all current speech coding standards. The basic idea of linear prediction is to exploit correlation immanent to the input signal. For short-term block adaptive linear prediction, a windowed segment of the input signal is analyzed in order to obtain the filter coefficients $a_1 ... a_N$ (LP filter order N) which minimize the energy of the difference between original and predicted signal. The transfer function of the general linear prediction analysis filter is

$$1 - A(z) = 1 - \sum_{i=1}^{N} a_i \cdot z^{-i}$$

In our MAD transmission system the strength of the prediction filter

$$H(z) = \frac{1 - A(z)}{1 - A(z/\gamma)}$$

can be controlled by a factor of $\gamma = 0$ (full prediction) to $\gamma = 1$ (no prediction), see Figures 3 and 5. Varying the strength of the prediction filter implies varying the amount of colouring of the audible noise at the receiver side, as the white channel noise is filtered with the LP synthesis filter $\frac{1-A(z/\gamma)}{1-A(z)}$. This is called noise shaping in the literature, e.g. [5]. While the audible noise is coloured with the spectral shape of the signal with full prediction ($\gamma = 0$; $\frac{1}{H(z)} = \frac{1}{1-A(z)}$), the audible noise remains white without prediction ($\gamma = 1$; $\frac{1}{H(z)} = 1$). Figure 2 shows the measured perceptual speech quality (Perceptual Estimation of Speech Quality [13]) for different γ. The exemplary simulation results indicate that $\gamma \approx 0.5$ yields the best quality regardless of the channel SNR.

In our Mixed Pseudo Analogue-Digital (MAD) transmission system, the LP filter coefficients a_i are quantized with a vector quantizer (VQ), and the gain factor g described in section II-B is quantized with a scalar quantizer (Q). The quantizer codebook indices of these quantizers form the digital part of the transmission. For this part, channel coding as described in section II-C is applied.

We consider narrowband input speech (300 Hz ... 3.4 kHz audio bandwidth, 8 kHz sampling rate) and use routines from the narrowband Adaptive Multirate (AMR-NB) speech codec [10] mode 12.2 kbit/s. Two sets of LP filter coefficients of order 10 are calculated per 20 ms frame, which are jointly quantized using split matrix quantization (SMQ) with the original AMR-NB quantization codebooks [10]. The codebook index from AMR-NB requires 38 bit per 20 ms frame.

Wideband input speech (70 Hz .. 7 kHz audio bandwidth, 16 kHz sampling frequency) and audio (16 kHz, 22 kHz or 32 kHz sampling frequency) require an LP filter of order 16. One set of LP filter coefficients is calculated every 20 ms frame, which is quantized using a combination of split vector quantization (SVQ) and multi-stage vector quantization (MSVQ) with the original wideband Adaptive Multirate (AMR-WB) quantization codebooks [11]. The codebook index from AMR-WB requires 46 bit per 20 ms frame.

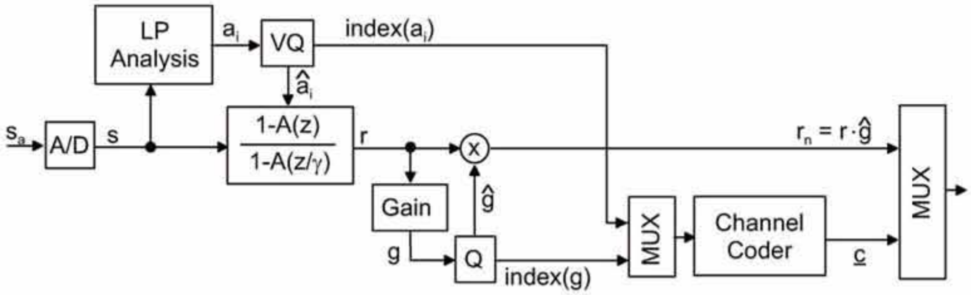

Fig. 3. Mixed Pseudo Analogue-Digital Speech and Audio Transmission: Transmitter.

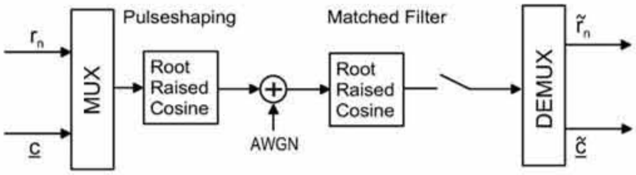

Fig. 4. Mixed Pseudo Analogue-Digital Speech and Audio Transmission: Transmission.

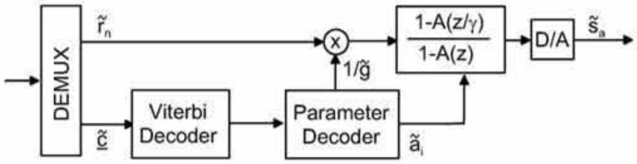

Fig. 5. Mixed Pseudo Analogue-Digital Speech and Audio Transmission: Receiver.

B. *Power Equalization*

If different transmission systems are to be compared, the mean output power of the transmitters must be the same. Thus, for each 5 ms subframe of the residual signal $r(k)$, a gain $g = \sqrt{1/\sum r(k)^2}$ is calculated. Multiplying $r(k)$ by g in each subframe results in continuous-amplitude samples with an average power of 1, which is equivalent to the digital transmission of the symbols 1 and -1, respectively. The gains g are quantized with a scalar 5-bit Lloyd-Max quantizer and transmitted together with the LP coefficients a_i (compare Figure 3). Gains g and LP coefficients a_i form the digital information of the MAD transmission scheme.

C. *Channel Coding For LPC And Gains*

To secure the LP coefficients a_i and gains g, a rate $1/2$ convolutional channel code [8] is applied. The polynomials have been chosen to

$$\begin{aligned} G_0 &= 1 + D^3 + D^4 \text{ and} \\ G_1 &= 1 + D + D^3 + D^4 \end{aligned}$$

to use the same channel code as the GSM system with full-rate speech coding [12], to which the new system will be compared in section III-A. At the receiver side of both systems a hard-decision Viterbi decoder [8] is used in all cases. This decoder was chosen for reasons of complexity and comparability.

D. *Baseband Transmission Model*

The residual signal is not quantized; instead the time-discrete, continuous-amplitude samples are directly fed to the Root Raised Cosine filter in addition (time multiplex) to the digital data and transmitted over the AWGN channel. Thus, instead of quantization noise there is channel noise.

Transmission of analogue and digital parts is investigated in the baseband. To prevent inter-symbol interference, analogue and digital pulses are shaped with a Root Raised Cosine filter (roll-off factor $\alpha = 0.5$). The required (two-sided) bandwidth [6] for the combined signal equals

$$B = B_a + B_d = (1 + \alpha) \cdot (R_a + R_d) = 1.5(R_a + R_d)$$

with R_a the analogue sample rate and R_d the digital bit rate.

III. Experimental Results and Discussion

A. *Comparison To Narrowband AMR*

To evaluate the new coding scheme for speech applications, it was compared to the GSM Adaptive Multirate (AMR-NB) codec [10] mode 12.2 kbit/s (GSM Enhanced Fullrate) operating at 22.8 kbit/s including channel coding [12]. The AMR-NB bitstream was fed to the same pulse shaping filter and AWGN channel as described above. In addition to the convolutional code, no further error concealment was used in both cases. The required bandwidth of the AMR-NB speech codec is

$$B_{AMR_{NB}} = 1.5 \cdot 22.8\,\text{kbit/s} = 34.2\,\text{kHz}.$$

The black solid line in Figure 6 shows the measured wideband PESQ values (Perceptual Estimation of Speech Quality [13]) for different E_b/N_0. Wideband PESQ has been chosen to be able to compare the AMR-NB to both, narrowband and wideband MAD transmission. The original wideband speech signal scores 4.5 on that scale; the narrowband reference signal is a low-pass filtered version of the wideband signal, scoring 3.14. With error free transmission, the digital narrowband AMR-NB speech codec scores 2.26.

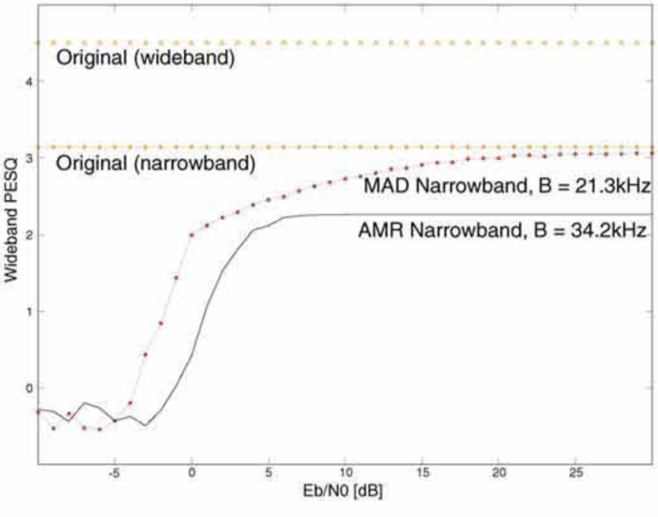

Fig. 6. Comparison of AMR-NB coding and narrowband MAD speech coding.

Narrowband MAD transmission, suitable for telephone quality speech, needs 38 bit/20 ms to quantize the LP coefficients of order 10 with modules from the AMR-NB narrowband speech codec, 20 bit/20 ms for the gains (4 subframes times 5 bit), and 4 bit/20 ms for termination of the convolutional code, adding up to

$$R_{d\_NB} = (38 + 20 + 4)\frac{\text{bit}}{\text{frame}} \cdot 50\frac{\text{frames}}{\text{s}} \cdot 2 = 6.2\,\text{kbit/s}$$

after channel coding. With the sampling rate $f_{s_{NB}} = 8000\,\text{Hz}$ the bandwidth used for the residual signal equals

$$B_{a_{NB}} = 1.5 \cdot f_{s_{NB}} = 12\,\text{kHz}$$

and the bandwidth needed for the digital part equals

$$B_{d_{NB}} = 1.5 \cdot R_{d\_NB} = 9.3\,\text{kHz}.$$

Thus, a total bandwith

$$B_{MAD_{NB}} = B_{a_{NB}} + B_{d_{NB}} = 12\,\text{kHz} + 9.3\,\text{kHz} = 21.3\,\text{kHz}$$

is used. The line with stars in Figure 6 shows the measured wideband PESQ values of narrowband MAD transmission for different E_b/N_0. It may be noted that besides a reduction in bandwidth of about 38%, the MAD transmission scheme has also significantly reduced requirements for computational power compared to a Code Excited Linear Prediction (CELP) scheme as used in the AMR-NB speech codec, due to the complete absense of open loop pitch, adaptive, and stochastic codebook search. Using MAD transmission, the speech quality rises with improving channel conditions until truly transparent speech transmission is reached. With falling E_b/N_0, MAD degrades gracefully up to the point when the digital information is corrupted and wrong LP indices are decoded. This threshold effect, however, starts at lower E_b/N_0 than with the digital system.

B. Wideband Coding and MAD Audio Coding

If wideband speech (7 kHz audio bandwidth, 16 kHz sampling frequency) or audio (also 16 kHz sampling frequency) is available at the transmitter side, the MAD transmission scheme allows for wideband coding with no additional computational requirements compared to narrowband MAD transmission, despite those caused by the increased sampling rate and LP filter order. The transmission bandwidth also remains well in the same region as the bandwidth required for narrowband AMR-NB transmission.

Wideband MAD transmission differs from narrowband MAD transmission only with respect to the sample rate of the input signal and the linear prediction order. Quantization of the LP coefficients of order 16 is carried out with modules from the AMR-WB wideband speech codec [11]. Wideband transmission requires 46 bit/20 ms for the LP coefficients. Thus, we get

$$R_{d\_WB} = (46 + 20 + 4)\frac{\text{bit}}{\text{frame}} \cdot 50\frac{\text{frames}}{\text{s}} \cdot 2 = 7\,\text{kbit/s}$$

after channel coding. With this and a sampling frequency $f_{s_{WB}} = 16000\,\text{Hz}$, the bandwidth used for the analogue residual becomes

$$B_{a_{WB}} = 1.5 \cdot f_{s_{WB}} = 24\,\text{kHz}$$

and the bandwidth needed for the digital part is

$$B_{d_{WB}} = 1.5 \cdot R_{d\_WB} = 10.5\,\text{kHz}.$$

We can finally obtain the total bandwidth

$$B_{MAD_{WB}} = B_{a_{WB}} + B_{d_{WB}} = (24 + 10.5)\,\text{kHz} = 34.5\,\text{kHz}.$$

The upper line with stars in Figure 7 corresponds to wideband MAD transmission and shows the impressive gain in quality using true wideband coding. The computational complexity of wideband MAD transmission differs only with respect to the sampling frequency from that of narrowband MAD. It is still substantially below that of a narrowband CELP codec. To give another reference, the maximum quality of the AMR-WB wideband speech codec mode 23.05kbit/s is shown. This codec scores 3.59 for the test signals. It was not compared to MAD transmission as the required transmission bandwidth without any channel coding at all would be

$$B_{AMR_{WB}} = 1.5 \cdot 23.05\,\text{kbit/s} = 34.57\,\text{kHz}.$$

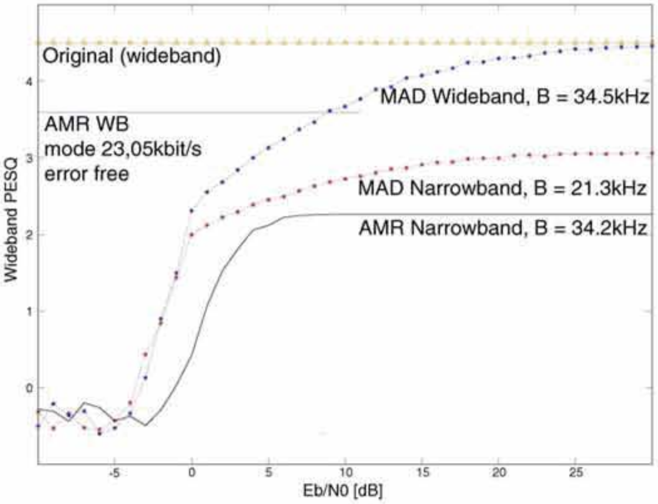

Fig. 7. Comparison of AMR-NB and MAD speech coding.

As our MAD transmission scheme does not make use of any model of speech production, it is suitable not only for speech but also for audio signals. Figure 8 shows an exemplary comparison of the subjective quality of a music signal (Mozart: Kleine Nachtmusik) coded with the GSM Enhanced Full Rate codec, with the wideband MAD codec, and transmitted as PCM samples. While the PESQ measure strictly speaking measures the perceptual quality of speech signals, it correlates well with the subjective impression found in an informal listening test with different kinds of music (Classic, March, Pop, Rock, Instrumental).

C. Audio Coding (f_s =22 kHz or 32 kHz)

Audio signals with 22 kHz or 32 kHz sampling frequency can be handled by our Audio MAD transmission system. Audio MAD transmission differs from 16 kHz wideband MAD transmission only with respect to the sample rate of the input

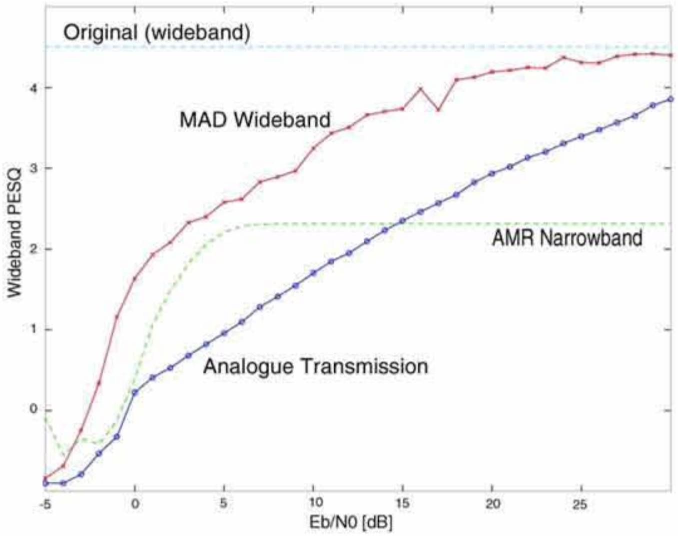

Fig. 8. Comparison of AMR-NB, PCM and MAD audio coding.

signal. Linear prediction of order 16 and quantization of the LP coefficients with modules from the AMR-WB wideband speech codec [11] remain unchanged. Thus, the digital rate is still

$$R_{d\_WB} = (46 + 20 + 4)\frac{\text{bit}}{\text{frame}} \cdot 50\frac{\text{frames}}{\text{s}} \cdot 2 = 7\,\text{kbit/s}$$

after channel coding.

With $f_{sWB} = 22000\,\text{Hz}$, the bandwidth used for the analogue residual becomes

$$B_{aWB} = 1.5 \cdot f_{sWB} = 33\,\text{kHz}.$$

The total bandwidth is

$$B_{MADWB} = B_{aWB} + B_{dWB} = (33 + 10.5)\,\text{kHz} = 43.5\,\text{kHz}.$$

With $f_{sWB} = 32000\,\text{Hz}$, the bandwidth used for the analogue residual becomes

$$B_{aWB} = 1.5 \cdot f_{sWB} = 48\,\text{kHz}.$$

The total bandwidth is

$$B_{MADWB} = B_{aWB} + B_{dWB} = (48 + 10.5)\,\text{kHz} = 58.5\,\text{kHz}.$$

Informal listening tests proved the increase in perceptual quality compared to PCM transmission of equal sampling frequency to be comparable to the gain shown in Figure 8 for $f_{sWB} = 16000\,\text{Hz}$.

IV. Conclusion and Future Work

A new principle, Mixed Pseudo Analogue-Digital Speech and Audio Transmission, has been proposed that combines the advantages of robust digital transmission of parameters and bandwidth-efficient transmission of pseudo analogue samples of a prediction residual. The new scheme allows high quality transmission of speech and audio signals, yielding almost transparent quality for good channels. With weaker channels, the quality degrades gracefully. This new scheme uses significantly smaller bandwidth and computational power in comparison to purely digital schemes. Thus it is well suited, e.g., for AWGN channels and bandwidth critical applications. The general MAD scheme does not require any prior knowledge of the channel.

Currently, MAD transmission is extended with a bandwidth extension scheme to estimate the wideband excitation after transmission of a low-pass filtered version of the pseudo analogue residual to save further transmission bandwidth. Also, different modulation schemes are being looked at.

References

[1] Miki, Sundberg, Seshadri, "Pseudo-Analog Speech Transmission in Mobile Radio Communication Systems, *IEEE Trans. on Vehicular Technology*, vol. 42, no. 1, Feb. 1993.

[2] M.Skoglund, N.Phamdo, F.Alajaji, "Design and Performance of VQ-Based Hybrid Digital-Analog Joint Source-Channel Codes, *IEEE Transactions on Information Theory*, vol. 48, no. 3, pp.708-720, March 2002.

[3] C. Hoelper, P.Vary, "Bandwidth-Efficient Mixed Pseudo Analogue-Digital Speech Transmission, submitted to *European Signal Processong Conference EUSIPCO*, Florence, September 2006.

[4] J.G. Proakis, *Digital Communications*, McGraw-Hill, 2001.

[5] P.Vary, R.Martin, *Digital Speech Transmission - Enhancement, Coding and Error Concealment*, Wiley, 2006.

[6] S. Haykin, *An Introduction to Analog and Digital Communications*, Wiley, 1989.

[7] Jayant, Noll, *Digital Coding of Waveforms: Principles and Applications to Speech and Video*, Prentice Hall, 1984.

[8] B. Sklar, *Digital Communications*, Pearson, 2001.

[9] ETSI Rec. GSM 06.10, *GSM Full Rate Speech Transcoding*, 1988.

[10] ITU-T 26.090, *Adaptive Multi-Rate (AMR) speech Codec.*

[11] ITU-T 26.190, *Adaptive Multi-Rate - Wideband (AMR-WB) speech codec.*

[12] ETS 300575, *GSM (Phase 2): Channel Coding GSM 05.03.*

[13] ITU-T P.862, *Perceptual Evaluation of Speech Quality.*