A NEW MODULATION CONCEPT FOR MIXED PSEUDO ANALOGUE-DIGITAL SPEECH AND AUDIO TRANSMISSIOIN

C. Hoelper, P. Vary

(ind), RWTH Aachen University
Muffeter Weg 3a, D-52056, Aachen, Germany
email: {hoelper,vary}@ind.rwth-aachen.de, web: www.ind.rwth-aachen.de

ABSTRACT

Current speech, audio, and video coding and transmission systems are either analogue or digital, with a strong shift from analogue systems to digital systems during the last decades. We have combined both digital and analogue schemes for the benefit of saving transmission bandwidth, complexity, and of improving the achievable quality at any given signal-to-noise ratio on the channel. The combination is achieved by transmitting pseudo analogue samples of the unquantized residual signal of a linear predictive digital filter which is called Mixed Pseudo Analogue-Digital (MAD) transmission. In this paper a new modulation scheme based on QPSK for digital information and an Archimedes spiral for the time discrete, pseudo analogue residual signal is introduced and evaluated.

Index Terms— Linear predictive coding, speech coding

1. INTRODUCTION

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

With increasing channel SNR (signal-to-noise ratio), the output quality of a digital system remains constant due to source coding / quantization, 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. [5]. Digital systems generally require a higher bandwidth.

To combine the advantages of digital and analogue transmission different approaches exist. In [2] 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.

We have proposed a hybrid scheme, Mixed Pseudo Analogue-Digital (MAD) Speech Transmission [3], which requires a digital channel for transmitting the spectral envelope and short term energy plus a pseudo-analogue channel for transmitting discrete-time samples of the prediction residual. T. Miki et.al. have studied a similar scheme for ADPCM (Adaptive Differential Puls Code Modulation) [1].

MAD transmission is very efficient with respect to the required transmission bandwidth and it allows to exploit the mechanisms of linear predictive coding (LPC) and noise shaping to produce high quality speech [3].

In this paper we evaluate the MAD transmission principle with two-dimensional modulation schemes: Binary Phase Shift Keying (BPSK), or Quadrature Phase Shift Keying (QPSK) for the digital information and Amplitude Shift Keying (ASK), or Archimedes Spiral Mapping (ASM) for representation of the pseudo-analogue residual signal.

2. PRINCIPLES OF MAD CODING

The baseband transmission model of the Mixed Pseudo Analogue-Digital transmission system is given in Figures 1 (transmission) and 2 (receiver). We consider speech and audio transmissions. The objective of MAD transmission is to maximize the subjective quality while minimizing the required transmission bandwidth and coding complexity.

2.1. Processing in the Digital Domain

2.1.1. Linear Prediction

Linear Prediction (LP) has proved to be very effective to code speech and audio. The basic idea is to exploit correlation immanent to the input signal. For short-term block adaptive LP, 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.

In our MAD transmission system the conventional prediction filter $H(z) = \frac{1-A(z)}{1-A(z/\gamma)}$ which can be controlled by a factor of $\gamma = 0$ (full prediction) to $\gamma = 1$ (no prediction).

Varying the prediction implies varying the amount of noise shaping [6] at the receiver side, as the channel noise is filtered with the noise shaping filter $\frac{1-A(z/\gamma)}{1-A(z)}$. According to [3] a value of $\gamma \approx 0.5$ yields the best perceptual speech quality for MAD transmission.

The filter coefficients a_i are quantized with standard quantizers. For narrowband input speech (sampling frequency $f_s = 8$ kHz) we use the filter of order 10 from the narrowband Adaptive Multirate speech codec [9] mode 12.2 kbit/s (AMR-NB). The AMR-NB codebook index requires 38 bit per 20 ms frame. Wideband input speech (sampling frequency $f_s = 16$ kHz) and audio (16 kHz, 22 kHz or 32 kHz sampling frequency) are filtered with an filter of order 16. The wideband filter coefficients are quantized with the original wideband Adaptive Multirate (AMR-WB) quantization codebooks [10]. The codebook index from AMR-WB requires 46 bit per 20 ms frame.

2.1.2. Power Equalization

If different transmission systems are to be compared, the mean output power of the transmitters must be the same. To ensure equal average energy per transmitted symbol, for each 5 ms subframe (N_s samples) of the residual signal $r(k)$, a gain $g = \sqrt{N_s / \sum r(k)^2}$ is calculated. Multiplying $r(k)$ by the quantized gain factor $\hat{g}$ in each subframe results in continuous-amplitude samples $r_n(k) = r(k) \cdot \hat{g}$ 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 coefficients a_i (compare Figure 1). Gains g and coefficients a_i form the digital information of the MAD transmission scheme.

2.1.3. Channel Coding For LPC And Gains

To protect the LP coefficients a_i and gains g, a rate $1/2$ convolutional channel code [8] is applied. The polynomials are $G_0 = 1 + D^3 + D^4$ and $G_1 = 1 + D + D^3 + D^4$ i.e. the same channel code as used for the GSM system with full-rate speech coding [11]. The output $\underline{c}$ of the channel coder has the rate R_d. At the receiver side of both systems a hard-decision Viterbi decoder [8] is used in all cases. This scheme was chosen for reasons of complexity and comparability.

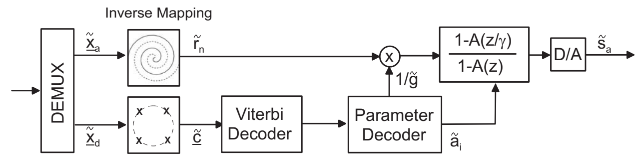

Fig. 2. MAD Transmission: Receiver.

2.2. Signal Space Mapping

In the signal space we consider two different mappings. For the digital information these are Binary Phase Shift Keying (BPSK) and Quadrature Phase Shift Keying (QPSK), respectively. The quasi-continuous residual r_n is transmitted with Amplitude Shift keying (ASK) or Archimedes Spiral Mapping (ASM). Transmission of analogue and digital parts is investigated in the baseband. To prevent inter-symbol interference, the multiplexed analogue and digital pulses are shaped with the same Root Raised Cosine filter (roll-off factor $\alpha = 0.5$).

2.2.1. Baseband Transmission Model for BPSK/ASK

The residual signal is not quantized; instead the normalized, time-discrete, continuous-amplitude samples $r_n = \hat{g} \cdot r$ 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. The required (two-sided) bandwidth [5] for the combined signal equals
$B = B_a + B_{d_{BPSK}} = (1+\alpha) \cdot (R_a + R_d) = 1.5(R_a + R_d)$
with R_a the analogue sample rate and R_d the digital bit rate.

2.2.2. Baseband Transmission Model for QPSK/ASM

The normalized, time-discrete, continuous-amplitude samples $r_n = \hat{g} \cdot r$ are mapped to an Archimedes Spiral (see Figure 3). The Archimedes Spiral is defined in polar coordinates as

$$\varphi_{Ar}(r_n) = \begin{cases} \frac{r_n}{c} & \text{for } r_n \geq 0 \\ \frac{|r_n|}{c} + \pi & \text{for } r_n < 0 \end{cases}$$

with angle φ_{Ar}, radius $|r_n|$ and a real constant c that defines the tightness of the spiral [7]. Thus the complex signal samples transmitted equal

$$\underline{x}_a = r_n exp(j \frac{|r_n|}{c})$$

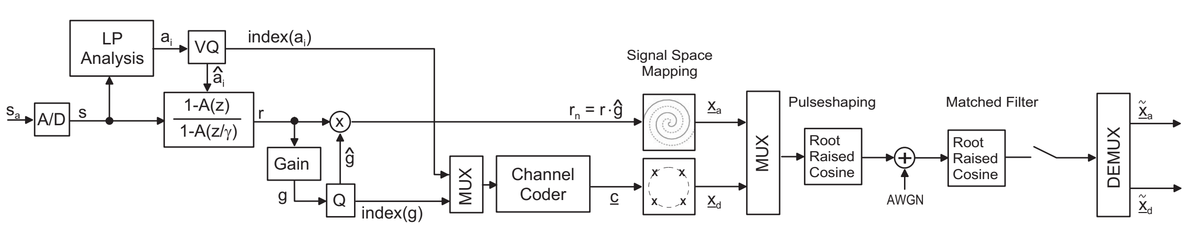

Fig. 1. Mixed Pseudo Analogue-Digital Speech and Audio Transmission: Transmitter and Channel.

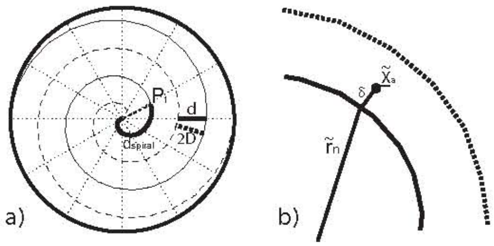

Fig. 3. a) Archimedes Spiral. $r_n \geq 0$ solid, $r_n < 0$ dashed. b) Decoding of a received symbol $\underline{\tilde{x}}_a$ into $\tilde{r_n}$

Archimedes sprial mapping (ASM) uses exactly the same transmit power as the ASK described above, regardless of the constant c, as the amplitudes r_n of ASK and ASM are equal. The complex signal is fed to the Root Raised Cosine filter in addition (time multiplex) to the digital QPSK signal $\underline{x}_d$ and transmitted over the AWGN channel. Again, instead of quantization noise there is channel noise. The required (two-sided) bandwidth for the complex signal equals $B = B_a + B_{d_{QPSK}} = (1+\alpha)\cdot(R_a + 0.5R_d) = 1.5(R_a + 0.5R_d)$ with R_a the analogue sample rate and R_d the digital bit rate.

3. PROPERTIES OF THE ARCHIMEDES SPIRAL

The radial distance d between the two parts of the spiral representing positive and negative values r_n (solid and dashed line in Figure 3) is constant (except for the center of the spiral). It can be calculated as $d = \pi c$. Thus the smaller c the tighter the spiral.

The length of the spiral from the center to a point P_1 at the angle φ_1 can be calculated as [7]

$$d_{spiral} = c\left[\frac{\varphi_1}{2}\sqrt{1+\varphi_1^2} + \frac{1}{2}\ln(\varphi_1 + \sqrt{1+\varphi_1^2})\right] \geq |r_n(P_1)|.$$

The angle γ_{Ar} between a radius and the tangent of the spiral at this radius can be found to be $\gamma_{Ar} = \arctan(\varphi_{Ar}) = \arctan(r_{Ar}/c)$ (see Figure 3b).

To decode the received complex signal samples $\underline{\tilde{x}}_a$, the receiver finds the closest distance δ to the Archimedes Spiral and decodes the corresponding radius $\tilde{r}_n$ (Figure 3b).

Decoding of a distorted signal results in two different possible errors:

I A smaller displacement of $\underline{\tilde{x}}_a$ on the correct branch of the spiral or

II a much larger displacement of $\underline{\tilde{x}}_a$ due to selection of a wrong branch.

If only errors of type I occur the mean displacement of ASM $|r_n - \tilde{r}_{n,ASM}|$ due to the added noise will be lower than that of ASK $|r_n - \tilde{r}_{n,ASK}|$, as the length d_{spiral} of the spiral is greater than the ASK amplitude r_n and thus the noise is compressed by the mapping from the Archimedes Spiral to $\tilde{r}_n$.

For errors of type II we find that as the branches of the spiral are equally spaced at the distance $2D$ (see Figure 3a) the propability of choosing an incorrect branch in the outer branches, i.e. $\varphi_{Ar} > \pi$, can be calculated to be $p_e(D) = 1 + erf(\frac{D}{\sqrt{2}\sigma_n})$ with noise power σ_n^2.

Figure 4 shows the effect of AWGN with 10dB SNR on the transmitted spiral for different c.

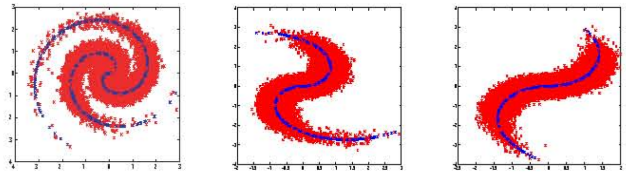

Fig. 4. ASM with SNR = 10dB and $c = 0.5$, $c = 1.5$, and $c = 2.5$, respectively.

4. EXPERIMENTAL RESULTS AND DISCUSSION

4.1. MAD Transmission with BPSK/ASK

The MAD transmission scheme has been evaluated in detail for narrowband and wideband speech and it was compared to the GSM Adaptive Multirate codec mode 12.2 kbit/s (GSM Enhanced Fullrate) operating at 22.8 kbit/s including channel coding in [3]. The resulting speech quality and bit rate can be seen in Figure 5, which shows the measured wideband PESQ values (Perceptual Estimation of Speech Quality [12]) for different E_b/N_0 with E_b the energy per coded AMR bit.

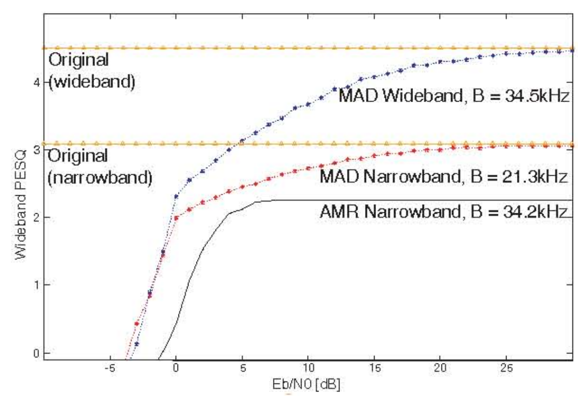

Fig. 5. AMR-NB and MAD speech coding with BPSK/ASK.

It may be noted that besides a possible reduction in bandwidth of about 38% in the narrowband case, the MAD transmission scheme (wideband as well as narrowband) 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 E_b/N_0 decreasing, MAD degrades gracefully up to the point when the digital information is corrupted and wrong LP indices are decoded.

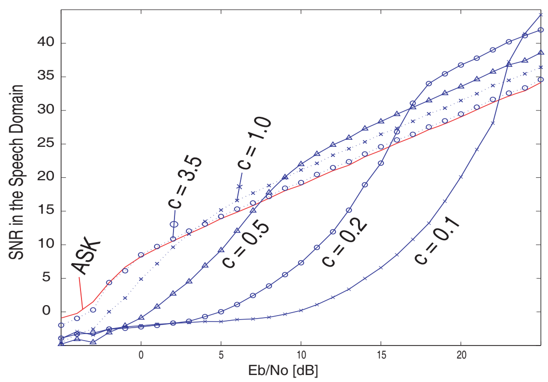

Fig. 6. Effect of ASM.

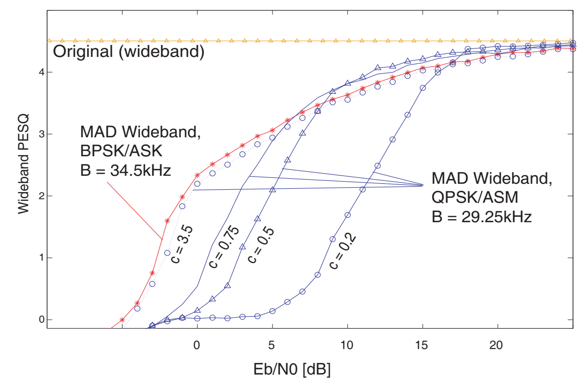

Fig. 7. Evaluation MAD speech coding with QPSK/ASM.

4.2. MAD Transmission with QPSK/ASM

Using QPSK instead of BPSK for transmission of the digital information does not change the output SNR. Yet with the symbol rate reduced by a factor of two the required bandwidth is cut in half. This means that in the case of wideband MAD Transmission the channel bandwidth required for digital transmission of LP coefficients and gains can be reduced from $B_{d_{BPSK}} = 10.5\,\text{kHz}$ to $B_{d_{QPSK}} = 5.25\,\text{kHz}$. Mapping the time-discrete quasi-continuous-amplitude samples of the residual signal to the Archimedes spiral does not affect the required bandwidth as only the phase is changed and not the transmission rate. Thus the complete channel bandwidth required for MAD wideband speech transmission thus is reduced from $B_{BPSK} = (24 + 10.5)\,\text{kHz} = 34.5\,\text{kHz}$ to $B_{QPSK} = (24 + 5.25)\,\text{kHz} = 29.25\,\text{kHz}$. with $B_a = 1.5 \cdot 16\,\text{kHz} = 24\,\text{kHz}$.

The effect of mapping the residual to an Archimedes Spiral can best be studied in Figure 6 which shows the output SNR of the speech signal over the channel E_b/N_0 for different constants c. While for good channels c may be smaller (and thus the spiral tighter) to have a higher gain in SNR, it must be increased with the noise power. Figure 7 shows the speech quality achieved with the new modulation scheme in comparison to MAD coding with BPSK/ASK.

5. CONCLUSION

A new modulation concept, Archimedes Spiral Mapping, has been introduced and studied in application with a Mixed Pseudo Analogue-Digital Transmission system, that combines the advantages of robust digital transmission of parameters and bandwidth-efficient transmission of pseudo analogue samples of a prediction residual. This scheme allows high quality transmission of speech and audio signals, yielding almost transparent quality for good channels. With weaker channels, the quality degrades gracefully. With QPSK for the digital information, 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, Archimedes Spiral Mapping, however, does benefit from a channel estimation with regard to the noise power. Compared to [3] a benefit of 5dB up to 15dB SNR in the speech domain is possible for good channels (Eb/N0 of 10dB to 25dB, see Figure 6) .

6. REFERENCES

[1] T. Miki, C.-E.W. Sundberg, N. Seshadri, "Pseudo-Analog Speech Transmission in Mobile Radio Communication Systems, *IEEE Trans. on Vehicular Technology*, 2/93.

[2] M.Skoglund, N.Phamdo, F.Alajaji, "Design & Performance of VQ-Based Hybrid Digital-Analog Joint Source-Channel Codes, *IEEE Trans.on Inf.Th.*, v.48, no.3, 03/02.

[3] C. Hoelper, P.Vary, "Bandwidth-Efficient Mixed Pseudo Analogue-Digital Speech Transmission, *European Signal Processing Conference EUSIPCO*, Florence, 9/06.

[4] C. Hoelper, P.Vary, "Bandwidth-Efficient Mixed Pseudo Analogue-Digital Speech and Audio Transmission, *Int. Workshop On Multimedia Sig. Proc.*, Victoria, 10/06.

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

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

[7] I.N. Bronstein, K.A. Semendjajew, *Taschenbuch der Mathematik*, Teubner, 1991.

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

[9] ITU-T 26.090, *Adaptive Multi-Rate speech Codec*.

[10] ITU-T 26.190, *AMR - Wideband speech codec*.

[11] ETS 300575, *GSM (Ph.2); Channel Coding GSM 05.03*.

[12] ITU-T P.862.2, *Perceptual Evaluation of Speech Qual.*.