Mixed Pseudo Analogue-Digital Speech Transmission

Carsten Hoelper, Peter Vary
Institute of Communication Systems and Data Processing (ind), RWTH Aachen University, Germany
{hoelper,vary}@ind.rwth-aachen.de

Abstract

Today's speech coding and transmission systems are either analogue or digital, with a strong shift towards digital during the last decades. We combine digital and analogue schemes for the benefit of saving transmission bandwidth, complexity, and of improving the achievable speech quality at any given signal-to-noise ratio (SNR) on the channel. The combination is achieved by transmitting 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 and wideband speech. 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, which uses a comparable transmission bandwidth if channel coding is included. A wideband extension scheme for the low-pass filtered residual is introduced to further reduce the transmission bandwidth.

1. Introduction

While analogue speech 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 the channel coding theorem. However, if the channel SNR increases, the output quality of a digital system will remain constant. Due to the channel coding, digital systems generally require a higher bandwidth than analogue systems. In our paper we consider BPSK (Binary Phase Shift Keying) modulation for transmission of digital information and ASK (Amplitude Shift Keying) for transmission of pseudo analogue samples. The symbol rates are $R_d = \frac{1}{T}$ and $R_a = f_s$ (sampling frequency). A Root Raised Cosine (RRC) pulse shaping filter with roll-off factor $\alpha = 0.5$ is used.
In [1] 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 plus a refinement signal with continuous amplitude. In contrast to this we propose a hybrid scheme as shown in Figure 1, which requires a digital channel for transmitting prediction coefficients a_i and gain factors g and a pseudo-analogue channel for transmitting discrete-time samples $r_n = r \cdot g$ of the prediction residual r.

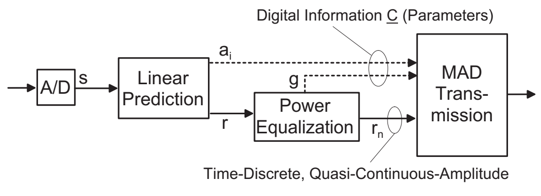

Fig. 1. Mixed Pseudo Analogue-Digital Speech Transmission.

This approach 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. The detailed operation is given in Figures 2 (transmitter), 3 (transmission) and 4 (receiver). The objective is to maximize the speech quality while minimizing the transmission bandwidth and coding complexity. The MAD transmission scheme is not adapted to a specific channel SNR.

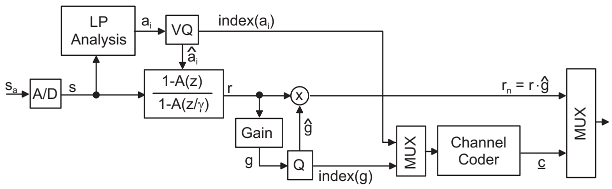

Fig. 2. Mixed Pseudo Analogue-Digital Speech Transmission: Transmitter.

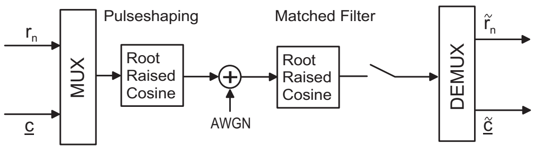

Fig. 3. Mixed Pseudo Analogue-Digital Speech Transmission: Transmission.

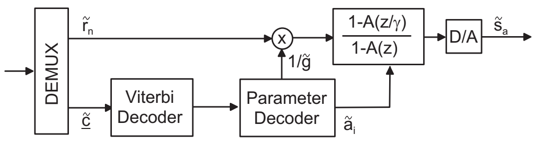

Fig. 4. Mixed Pseudo Analogue-Digital Speech Transmission: Receiver.

2. Digital Linear Prediction

Linear Prediction, e.g. [2], is used in almost all current speech coding standards to exploit correlation immanent to the input signal. 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 segment by segment the energy of the difference between original and predicted signal. In

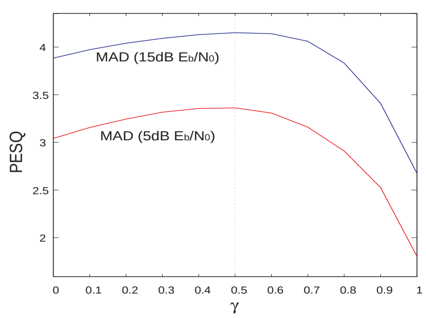

Fig. 5. Influence of prediction strength γ on the subjective speech quality measure PESQ.

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). Varying the strength of the prediction filter implies varying the amount of colouring of the audible noise at the receiver side (noise shaping) [2]. Figure 5 shows the measured perceptual speech quality (Perceptual Estimation of Speech Quality, PESQ [9]) for different γ. The exemplary simulation results indicate that $\gamma \approx 0.5$ yields the best speech quality regardless of the channel SNR. The LP filter coefficients a_i are quantized with a vector quantizer (VQ). For narrowband input speech (300 Hz ... 3.4 kHz audio bandwidth, 8 kHz sampling rate) we use the VQ from the narrowband Adaptive Multirate (AMR-NB) speech codec [6] mode 12.2 kbit/s. The codebook index from AMR-NB requires 38 bit/20 ms speech frame. For wideband input speech (70 Hz .. 7 kHz audio bandwidth, 16 kHz sampling frequency) we use the original wideband Adaptive Multirate (AMR-WB) VQ [7]. The codebook index from AMR-WB requires 46 bit/20 ms speech frame.

3. Transmission Aspects

To have equal mean output power of the transmitters, a gain $g = \sqrt{1/\sum r(k)^2}$ is calculated in each 5 ms subframe of the residual signal $r(k)$. The gains g are quantized with a scalar 5-bit Lloyd-Max quantizer (Q) and transmitted together with the LP coefficients a_i (Figure 2). Gains g and LP coefficients a_i form the ditital information of the MAD transmission scheme. To combat transmission errors, a rate $1/2$ convolutional channel code [4] with polynominals $G_0 = 1 + D^3 + D^4$ and $G_1 = 1 + D + D^3 + D^4$ is applied. This is the channel code of GSM system (full-rate speech coding) [8]. The GSM full rate system will be taken as a reference in section 4. At the receiver side of both systems a hard-decision Viterbi decoder [4] is used in all cases, which was chosen for reasons of complexity and comparability. The residual signal is not quantized; instead the time-discrete, continuous-amplitude samples are fed to the Root Raised

Cosine filter in addition (time multiplex) to the digital data and transmitted over the AWGN channel. Transmission of analogue and digital parts is investigated in the baseband model of BPSK/ASK. 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) baseband bandwidth [3] for the combined signal equals $B = (1 + \alpha)(R_a + R_d) = 1.5(R_a + R_d)$ with R_a the analogue sample rate and R_d the digital bit rate [4].

4. Comparing MAD To Narrowband AMR

To evaluate the new coding scheme, it was compared to the GSM Adaptive Multirate (AMR-NB) codec [6] mode 12.2 kbit/s (GSM Enhanced Fullrate) operating at 22.8 kbit/s including channel coding [8]. 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 all cases. The required bandwidth of the AMR-NB speech codec is $B_{AMR_{NB}} = 34.2$ kHz. Figure 6 shows the measured wideband PESQ values for different E_b/N_0. Wideband PESQ has been chosen to be able to compare the AMR-NB to both narrowband (MAD-NB) and wideband MAD (MAD-WB) transmission.

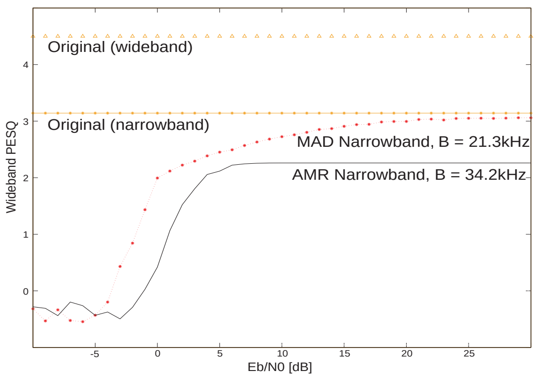

Fig. 6. Comparison of AMR-NB and MAD-NB coding.

Narrowband MAD (MAD-NB) transmission needs 38 bit/20 ms to quantize the LP coefficients of order 10, 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 $R_{d\_NB} = 6.2$ kbit/s after channel coding. With the sampling rate $f_{s_{NB}} = 8000$ Hz the bandwidth used for the residual signal equals $B_{a_{NB}} = 12$ kHz and the bandwidth needed for the digital part equals $B_{d_{NB}} = 9.3$ kHz.Thus, a total bandwisth $B_{MAD_{NB}} = B_{a_{NB}} + B_{d_{NB}} = 21.3$ kHz is used. The line with stars in Figure 6 shows the measured wideband PESQ values of MAD-NB transmission for different E_b/N_0. It may be noted that besides a reduction in bandwidth of about 38%, the MAD transmission scheme also has 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 and codebook searches. 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.

If wideband speech 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 of 34.2 kHz required for narrowband AMR-NB transmission. Quantization of the LP coefficients of order 16 is carried out with modules from the AMR-WB wideband speech codec [7] using 46 bit/20 ms. Thus, we get $R_{d\_WB} = 7$ kbit/s after channel coding. With this and a sampling frequency $f_{s_{WB}} = 16000$ Hz, the bandwidth used for the residual becomes $B_{a_{WB}} = 24$ kHz and the bandwidth needed for the digital part is $B_{d_{WB}} == 10.5$ kHz. We can finally obtain the total bandwidth $B_{MAD_{WB}} = B_{a_{WB}} + B_{d_{WB}} = 34.5$ kHz.

Figure 7 shows the impressive gain in speech quality using true wideband coding.

5. Wideband Extension

With most of the additional bandwidth required for wideband coding being taken by the analogue residual, the next step is to reduce this part again, which is done by low-pass filtering and downsampling of the residual on the transmitter side. Due to the spectral flatness of the residual, the receiver can estimate the omitted frequency band. This principle is used e.g. in Residual Excited Linear Predictive (RELP) speech codecs like the GSM full-rate codec [5].

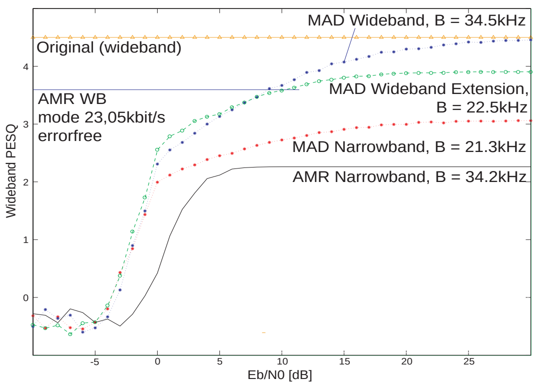

Fig. 7. Comparison of AMR-NB and MAD coding.

Due to the requirement of spectral flatness in the residual, for MAD transmission with wideband extension (BWE, bandwidth extension) the LP filter must carry out a full prediction, that is $\gamma = 0$. The residual signal r_{16kHz} is low-pass filtered with a cut-off frequency of $f_{lp} = 4000\,\text{Hz}$ and subsampled to $f_{s_{NB}} = 8000\,\text{Hz}$. The subsampled residual $r_{lp,8kHz}$ is then transmitted over the AWGN channel as described above for the narrowband version. At the receiver side, the signal is upsampled to $f_{s_{WB}} = 16000\,\text{Hz}$ and low-pass filtered again. All operations (except for the additive channel noise) are transparent for the lower band. The upper band is estimated by spectral folding which can be described as modulation with the modulation frequency equal to the Nyquist frequency $f_N = 8\,\text{kHz}$ of the interpolated residual $r'_{lp,16kHz}$ [10]. This way the modulation is simplified to a sequence of alternating signs: $r'_{wbx,16kHz}(k) = r'_{lp,16kHz}(k)(1 + (-1)^k)$.

Figure 7 also shows the results for MAD-BWE. This transmission scheme only uses a bandwith of $B_{a_{BWE}} = 12\,\text{kHz}$ and $B_{d_{BWE}} = 10.5\,\text{kHz}$ which in the sum equals $B_{MAD_{BWE}} = B_{a_{BWE}} + B_{d_{BWE}} = 22.5\,\text{kHz}$.

While true wideband coding naturally reaches a higher quality than the wideband extension scheme, the latter still outperforms both narrowband systems operating at a similar or lower channel bandwidth.

6. Conclusion

A new principle, Mixed Pseudo Analogue-Digital Speech 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 signals, yielding almost transparent quality for good channels. With weaker channels the speech 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 AGN channels and bandwidth critical applications. The general MAD scheme does not require any prior knowledge of the channel.

References

[1] 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.

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

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

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

[5] ETSI Rec. GSM 06.10, *GSM Full Rate Sp. Transc.*, 1988.

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

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

[8] ETS 300575, *GSM (Phase 2); Channel Coding GSM*.

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

[10] P. Jax, *Enhancement of Bandlimited Speech Signals: Algorithms and Theoretical Bounds*, Mainz, 2002.