MULTIPLE-DESCRIPTION CODING OF SPEECH USING FORWARD ERROR CORRECTION CODES

Kai Clüver, Jan Weil, Thomas Sikora

Communication Systems Group, Technische Universität Berlin
Einsteinufer 17, 10587 Berlin, Germany
phone: +49 30 314-24588, fax: +49 30 314-22514, email: {cluever;weil;sikora}@nue.tu-berlin.de
web: www.nue.tu-berlin.de

ABSTRACT

A flexible framework is presented which performs multiple-description coding of speech signals with two or more channels. The use of forward error correction codes together with a layered speech codec permits encoding into more than two descriptions without excessive increase in complexity. Results of a formal MOS listening test reveal considerable improvements in robustness as long as base layer quality and the number of descriptions are chosen appropriately. A modification of the original encoding scheme allows trading off bit rate savings against robustness to extreme channel conditions. Different coding schemes can easily be compared using a real-time demonstrator software.

1. MULTIPLE-DESCRIPTION CODING

Robustness to temporary channel breakdown can be considerably improved by multiple-description (MD) coding [1]. The coded signal is split into two or more descriptions which are transmitted over the same number of different channels. These channels may indeed consist of different physical links, or of different packets transmitted through a network. The principle of a two-channel MD coded transmission is shown in fig. 1. From the input signal, $x(n)$, the encoder generates two descriptions C_1 and C_2 to be sent over two lossy channels. If no loss occurs, both descriptions will be used by the central decoder to reconstruct the signal $y_0(n)$ with high quality. If one of the descriptions is lost, the received part of the code will enable its corresponding side decoder to yield a reduced-quality version of the output signal, $y_1(n)$ or $y_2(n)$. The transmission will be interrupted only when both descriptions are lost.

Many MD designs aim at balanced descriptions, i. e. equal bit rates ($R_2 = R_1$) and equal distortions of the side decoders. With more than two MD channels, balanced descriptions will yield decoding distortions which do not depend on the individual subset of descriptions but only on the number of descriptions received. The decoded quality will then degrade gracefully with increasing channel failure ratio.

Although transmission robustness potentially increases with the number of channels, few publications on MD speech coding have dealt with more than two descriptions. Usually, two-channel approaches closely depend on the speech coder, e. g. PCM or DPCM [2-5], transform coding [6], or CELP [7], and most designs cannot easily be extended to more than two descriptions. General methods for MD coding with many channels are encoding with diversity [8] or using forward error correction (FEC) codes [9]. This latter approach has been investigated for MD coding of images and video [10] [11].

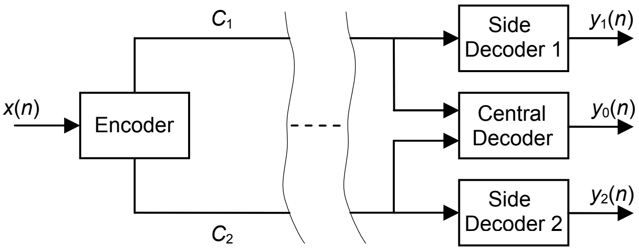


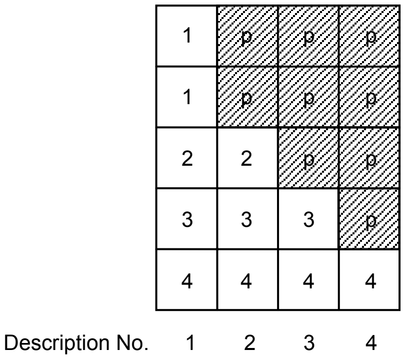
Fig. 1 Structure of a multiple-description coded transmission with two channels

In the following, the structure of a multiple-description speech coder for two or more descriptions using FEC codes as well as a real-time demonstrator software for packetized MD transmission are described. The design of suitable speech coders and the results of both informal and formal listening tests are discussed.

2. ENCODING SCHEME USING FORWARD ERROR CORRECTION CODES

A receiver for n descriptions consists of $2^n - 1$ decoders (including the central decoder). Consequently, the MD decoder will be extremely complex for high values of n if explicit side decoders are employed. This problem can be avoided by using a hierarchical (layered) coder together with forward error correction codes for the construction of multiple descriptions [9]. The approach consists of applying unequal erasure protection to n code layers and regrouping the symbols of the resulting codewords into n descriptions. The side decoders are then constructed implicitly by FEC decoding.

Fig. 2 shows the schematic structure of MD coded data for $n = 4$ descriptions. The symbols of the basic $n - 1$ hierarchical layers of a source coder are encoded with different code rates, using a systematic (n,i) error correction code for the i-th layer. The resulting n-symbol codewords form the initial rows of a matrix, and the source code symbols of the highest extension layer ($i = n$), which are not FEC encoded, are grouped into n-symbol words to form the final rows. Then, the columns of the matrix are transmitted separately as n descriptions.



	1	p	p	p
	1	p	p	p
	2	2	p	p
	3	3	3	p
	4	4	4	4
Description No.	1	2	3	4

Fig. 2 Construction of a four-description code (1...4 - source symbols of layers 1 to 4, p - parity symbols)

With $k < n$ descriptions received, only k symbols of each codeword are available at the decoder. Provided that the positions of the lost symbols are known, an (n,i) codeword can be correctly reconstructed as long as the number of missing symbols does not exceed $n - i$. Consequently, the MD decoder will be able to decode the basic k layers of the coded speech equally well with any k of n descriptions.

For FEC encoding, Reed-Solomon (RS) Codes [12] over a Galois field GF(2^r) are applied, the code symbols of which consist of r bits. The construction of multiple descriptions requires a certain minimum amount of source data, which determines the algorithmic MD encoding delay. In order not to cause excessive delays for low-rate speech data, r was set to only 3 bits per symbol. The resulting RS codes over GF(8) are $(7,i)$ codes; consequently, a maximum of 7 descriptions can be encoded. For $n < 7$, the RS codes are shortened appropriately.

At the decoder, codewords with missing source symbols are decoded by constructing a matrix from the RS code generator matrix and the loss pattern, inverting this matrix, and multiplying it with a vector constructed from the received symbols [9]. Matrix inversion, however, is not necessary in two cases: $(n,1)$ RS codes, which reduce to repetition codes, and $(n,n{-}1)$ codes, for which no RS code but a simple addition in GF(8) is employed for both calculation of the parity symbol and decoding.

3. LAYERED SPEECH CODERS

As described above, multiple-description coding with FEC codes requires a layered source code. To this end, two standard speech coders were modified for use within the MD framework.

3.1 PCM

For higher-rate speech coding, standard G.711 logarithmic PCM using the A-law compression characteristic [13] is employed. PCM, like any non-adaptive scalar quantization method, already yields a layered code, as a coarse reconstruction of samples is possible if any number of most significant bits of the codewords are available. The standard codec operates at a full rate of 64 kbit/s, i. e. 8 bit/sample. The only modification necessary to obtain a layered PCM codec was to add decoding rules or decoding tables for bit rates from 7 down to 2 bit/sample.

3.2 CELP

In order to obtain a layered low-rate speech codec, the ACELP codec according to standard G.729 annex A [14] was modified. Instead of the algebraic codebook, a single-pulse excitation is applied while all other functions of the codec (encoding of predictor coefficients, pitch analysis, adaptive codebook, encoding of gains, post-filtering) remain unchanged. This results in a base layer codec which operates at 5.4 kbit/s. Four higher layers are formed by an additional single pulse each, of which positions and relative gains are encoded. For two to five layers, the bit rates amount to 6.6, 7.8, 9.6, and 11.4 kbit/s respectively.

4. PACKETIZED MD CODED SPEECH

MD coding of speech signals was investigated by simulating a packetized transmision link with statistically independent packet losses. Apart from packet loss, no transmission errors (e. g. bit errors) are assumed. Encoding is carried out in frames; from each coded frame n descriptions are derived and transmitted in as many different packets.

PCM coding frames may, within limits, be of virtually any length. Within the MD framework outlined above, PCM code may be split into 2 to 7 descriptions. At the decoder, loss of all packets that represent a frame is dealt with by a frame erasure concealment algorithm based on linear prediction and pitch analysis [15]; for coding frames longer than 15 ms, the algorithm was modified as suggested in [16].

Due to the inherent coding frame of the CELP codec, only multiples of 10 ms are permitted for the length of MD-CELP frames. With five layers, the number of descriptions possible for CELP coded speech ranges from 2 to 5, and frame erasure concealment is carried out according to the original standard [14].

4.1 Bit rate allocation

As the base layer is FEC encoded with the lowest code rate of $1/n$, it contributes a major part of the gross bit rate of the MD code (cf. Fig. 2). On the other hand, the base layer determines the quality of only one description received, which should meet some minimum requirements. This is especially true for PCM, as packet loss causes short noise bursts in the decoded signal which are potentially more annoying than constant noise. Based on preliminary informal listening tests, a compromise bit rate of 4 bit/sample was chosen for the PCM base layer - whenever possible: for 6 and 7 descriptions, the base layer rate has to be reduced to 3 and 2 bit/sample respectively. Contrasted with that, the quality of the 5.4 kbit/s CELP base layer turned out to be sufficient, all the more because here the effects of packet loss on the speech signal set in smoothly. The resulting bit rates for MD encodings of 10 ms frames are given in table 1. In spite of limiting especially the bit rate of the base layer, MD encoding causes a considerable increase in bit rate for higher numbers of descriptions.

no. of descriptions	R_{PCM} / kbit/s	R_{CELP} / kbit/s
1	64	11.4
2	96.6	16.8
3	136.8	23.4
4	172.8	30
5	214.5	39
6	217.8	
7	207.9	

Table 1 Gross bit rates of MD encoded speech

4.2 MOS evaluation

In order to evaluate the subjective quality of the MD codecs, a formal listening test was conducted to obtain the mean opinion scores (MOS). To form the test samples, randomly chosen unique utterances spoken by four (two female, two male) speakers were combined, resulting in an average duration of 10 s. The samples were PCM or CELP encoded into a varying number of descriptions. Additionally, statistically independent random packet losses in the range of 2 to 50 % loss ratio were simulated. The samples were presented in random order via headphones. For each test sample, the subjects had to choose one of five quality levels. A total of 33 subjects participated in the listening test.

For most resulting mean values, half the width of the 95 % confidence interval turned out to be between 0.2 and 0.3; therefore, only differences of more than about 0.5 may be considered statistically significant.

MOS test results for PCM are shown in fig. 3. For comparison, the horizontal dashed line indicates the MOS of lossless transmission. PCM with 2 descriptions scores worst, and at loss ratios below 10 %, the MOS roughly increases with increasing n. With 20 % or more packets lost, MD coding with $n = 5$ scores best, whereas coding with 6 or 7 descriptions suffers from the coarse quantization of the base layer. Single-description PCM with erasure concealment only (dashed curve) scores higher than most MD encodings at loss ratios of up to 10 %, and even at 20 % performs better than MD coding with $n = 2$. Obviously, the noisy bursts caused by MD decoding are regarded as more annoying than the artifacts induced by the erasure concealment algorithm.

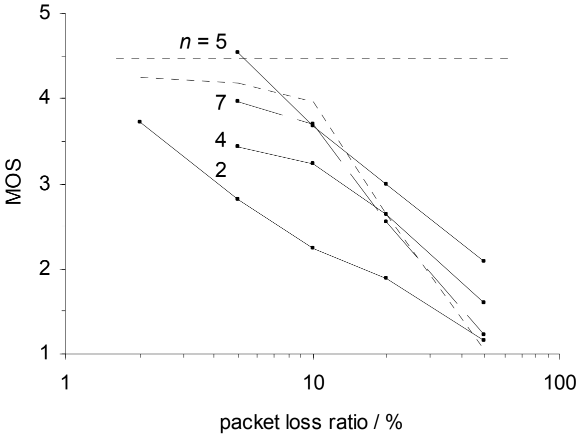


Fig. 3 Mean opinion scores of multiple-description PCM; single-description scores are marked by the dashed curve.

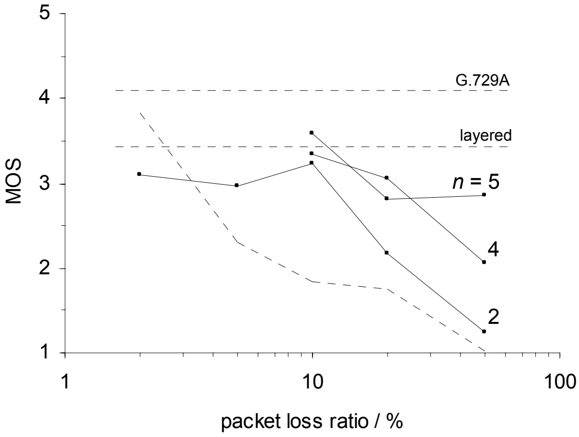


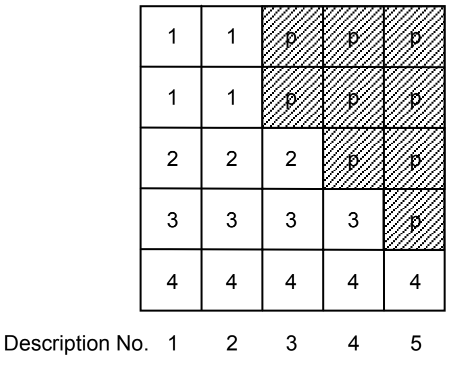
Fig. 4 Mean opinion scores of multiple-description CELP; scores of G.729A CELP are marked by the dashed curve.

Fig. 4 depicts the results for MD-CELP coding. For both standard and modified codecs, the scores of lossless transmission are given by horizontal dashed lines. The modified layered CELP codec scores almost 0.7 points less than the standard codec. At loss ratios of up to 10 %, any MD coding yields scores comparable to undistorted decoding, and at higher loss ratios, the number of descriptions roughly determines the achievable quality. CELP coding with five descriptions results in fair quality even at 50 % loss. Reports of subjects led to the assumption that quality impairments are mainly due to signal level variations caused by the erasure concealment algorithm. This was confirmed both in informal listening tests and by the MOS results of the single-description standard CELP codec (dashed curve in fig. 4), which suffers from heavier degradation under frame erasures than PCM with concealment.

4.3 Modifications of the original configuration

With increasing packet loss, the MD-PCM coding configuration which was evaluated in the formal listening test shows an unexpectedly severe decrease in mean opinion scores. One reason for this was found to be the choice of 4 bit/sample for the base layer bit rate, which is obviously too low for sufficient speech quality. Setting the base layer rate to 5 bit/sample, which is possible for 2, 3, or 4 descriptions, would further increase the respective bit rates to 104.4, 149.4, and 196.8 kbit/s (cf. table 1). On the other hand, informal listening comparisons back the assumption that now e. g. four-description encoding yields results comparable to, if not better than, five-description coding with 4 bit/sample base layer.

Another drawback of the coding configurations investigated above are extremely high gross bit rates which render MD coding with many descriptions inefficient and, for many applications, unacceptable. Therefore, modifications of the MD code structure were considered which permit trading robustness for bit rate savings. In the example shown in fig. 5, the base layer is FEC encoded with a code rate of $2/n$, which, on one hand, means that a minimum of two descriptions have to be received to enable the decoder to yield a coarse reconstruction of the speech signal. On the other hand, the increase in bit rates is mitigated, as shown in table 2. (The PCM bit rates given are those for a 4 bit/sample base layer, as in table 1. With a base layer rate of 5 bit/sample, bit rates in table 2 change to 84.6, 108, and 133.5 kbit/s for 3, 4, and 5 descriptions respectively.) Informal listening tests showed that, for low to medium packet loss ratios, the decoded quality does not decrease compared to unmodified MD as long as the speech signal is encoded with a sufficient number of descriptions.



1	1	p	p	p
1	1	p	p	p
2	2	2	p	p
3	3	3	3	p
4	4	4	4	4

Fig. 5 Modified five-description code with four source code layers

4.4 Real-time demonstration software

In order to demonstrate the effect of MD coding beyond the limited number of testing conditions of MOS evaluation or informal experiments reported above, a software application has been developed which makes various coding configurations easily comparable [17]. The software allows up to 20 streams to be requested from a server which codes the speech data in real time and sends it to the requesting client software. Random packet loss may be simulated within the client for robustness comparisons of different coding schemes.

no. of descriptions	R_{PCM} / kbit/s	R_{CELP} / kbit/s
3	81	14.4
4	103.2	18
5	121.5	22.5
6	145.8	
7	149.1	

Table 2 Gross bit rates of modified MD coding

5. DISCUSSION

Using FEC codes, a flexible framework for multiple-description coding of speech signals is obtained. On the condition that the quality of the base layer code is sufficiently high, robustness to packet loss is increased by increasing the number of descriptions. The original code construction may be modified in order to avoid extremely high gross bit rates, with only a small decrease in robustness.

With the base layer rate and the number of descriptions of MD PCM chosen appropriately, an improvement in quality is achieved especially for higher packet loss ratios. MD coding of PCM with only two descriptions, however, does not achieve the performance of other, explicitly two-channel, approaches [4] [5]. For this reason, using FEC codes is advantageous only for generating higher numbers of multiple descriptions, in order to obtain higher robustness to packet loss compared with the two-description case.

The tentative layered CELP codec developed for these experiments yields a quality that is clearly inferior to the standard codec on which it is based. Furthermore, the layers of the modified CELP codec do not represent significantly different quality levels. This partly explains the robustness of MD-CELP coding, since only small variations occur as long as speech frames are not completely erased. Concealment of erasures, on the other hand, induces severe degradation. Further work should therefore include the development of better excitation codebooks for the higher layers and a reoptimization of the erasure concealment algorithm with the aim of reduced signal level variation, both in order to obtain a layered narrowband CELP codec which is better suited to MD coding with more than two descriptions.

All experiments with MD transmission were performed simulating independent single packet losses. The assumption of statistical independence is reasonable for low-rate transmission of single packets through higher-rate networks, as in this case packet loss bursts caused by congestion are usually shorter than the distance between two consecutive packets of the low-rate signal. Variation of transmission paths further reduces the risk of bursty losses. MD coded data, however, are generated in a burst of *n* packets for each frame. In order to maintain statistical independence of packet loss, the descriptions should be transmitted one by one, with the sending times evenly distributed over the coding frame, which would introduce an additional algorithmic delay of approximately one frame.

On account of the possibly bursty effects of packet loss, packetized transmission may be considered to be the worst case scenario for MD coded speech signals. In scenarios with temporary failure of one or more of several transmission links, equal or better quality ratings may be expected.

6. CONCLUSION

A higher number of channels for multiple-description coded narrowband speech signals potentially increases the robustness to temporary channel failure. Combining a layered speech codec with forward error correction coding yields a flexible framework for MD coding with more than two descriptions. An improvement in robustness, however, will only be achieved if the quality of the base layer of the speech codec is sufficient. This is especially true for packetized MD transmission, in which packet losses may cause annoying artifacts in the signal.

A drawback of MD coding is the considerable increase in bit rate that renders the transmission of many descriptions inefficient. Modification of the MD encoding scheme permits trading off robustness to extreme failures against bit rate savings, without loss of quality at low to medium packet loss ratios.

The results of the experiments show the potential of multiple description coding for highly robust packetized transmission of speech signals. Future work should include more formal comparisons using more elaborate layered coders in order to determine the achievable quality in dependence of the bit rate. Furthermore, the MD framework with FEC coding outlined here is suitable for any layered speech or audio source codec for robust encoding with graceful degradation under transmission failures.

ACKNOWLEDGEMENTS

The work reported here was funded by the German Research Foundation (DFG) and carried out within VISNET II, a European Network of Excellence funded under the European Commission IST FP6 programme.

REFERENCES

[1] V. K. Goyal, "Multiple Description Coding: Compression Meets the Network", *IEEE Signal Processing Magazine*, vol. 8, no. 5, pp. 74-93, Sept. 2001.

[2] N. S. Jayant, S. W. Christensen, "Effects of Packet Losses in Waveform Coded Speech and Improvements due to an Odd-Even Sample-Interpolation Procedure", *IEEE Trans. on Communications*, vol. COM-29, no. 2, pp. 101-109, Feb. 1981.

[3] A. Ingle, V. A. Vaishampayan, "DPCM System Design for Diversity Systems with Applications to Packetized Speech", *IEEE Trans. on Speech and Audio Processing*, vol. 3, no. 1, pp. 48-58, Jan. 1995.

[4] S. D. Voran, "A Multiple-Description PCM Speech Coder using Structured Dual Vector Quantizers", in *Proc. ICASSP 2005*, pp. I-129 - I-132.

[5] K. Clüver, T. Sikora, "Multiple-Description Coding of Logarithmic PCM", in *Proc. EUSIPCO 2005*.

[6] J.-C. Batllo, V. A. Vaishampayan, "Multiple Description Transform Codes with an Application to Packetized Speech", in *Proc. Int. Symp. on Information Theory*, Trondheim, Norway, June/July 1994, p. 458.

[7] J. Balam, J. D. Gibson, "Multiple Description Coding and Path Diversity for Voice Communication over MANETs", in *Proc. 39th Asilomar Conf. on Signals, Systems, and Computers*, Oct./Nov. 2005, pp. 310-314.

[8] X. Zhong, B.-H. Juang, "Multiple Description Speech Coding with Diversities", in *Proc. ICASSP 2002*, pp. I-177 - I-180.

[9] L. Rizzo, "Effective Erasure Codes for Reliable Computer Communication Protocols", *ACM Computer Communication Review*, vol. 27, no. 2, April 1997

[10] A. E. Mohr, E. A. Riskin, R. E. Ladner, "Unequal Loss Protection: Graceful Degradation of Image Quality over Packet Erasure Channels through Forward Error Correction", *IEEE Journal on Selected Areas in Communications*, vol. 18, no. 6, pp.819-828, June 2006.

[11] T. Stockhammer, "Progressive Video Transmission for Packet Lossy Channels Exploiting Feedback and Unequal Erasure Protection", in *Proc. ICIP 2002*, pp. II-169 - II-172.

[12] R. E. Blahut, *Theory and Practice of Error Control Codes*. Reading MA: Addison-Wesley, 1983.

[13] N. S. Jayant, P. Noll, *Digital Coding of Waveforms*. Englewood Cliffs NJ: Prentice-Hall, 1984.

[14] "Reduced Complexity 8 kbit/s CS-ACELP Speech Codec", *ITU-T Recommendation G.729 Annex A*, Nov. 1996.

[15] K. Clüver, "An ATM Speech Codec with Improved Reconstruction of Lost Cells", in *Proc. EUSIPCO 1996*, pp. 1641-1643.

[16] E. Gündüzhan, K. Momtahan, "A Linear Prediction Based Loss Concealment Algorithm for PCM Coded Speech", *IEEE Trans. on Speech and Audio Processing*, vol.9, no. 8, pp. 778-785, Nov. 2001.

[17] J. Weil, K. Clüver, T. Sikora, "Real-Time Multiple-Description Coding of Speech Signals", in *Proc. 5th Int. Linux Audio Conf., LAC 2007*, Berlin.