A NOVEL APPROACH TO THE SPEECH AND CHANNEL CODING SYSTEM

Anton Čižmár
Technical University of Košice
Department of Radioelectronics
Park Komenského 13, 041 20 Košice
Slovakia

Abstract

In this paper a novel approach to the speech-and-channel coding system is proposed. The main idea of this approach is splitting an input speech signal into its integer and fractional parts and their different coding to fulfil the demand of bit-rate-reduction and unequal error protection, respectively. The experiments with the fractional part of the signal verify this hypothesis and the results of the average SNR of the input and output speech signal support a practical implementation of the system.

Keywords:

1. Introduction

Variable rate speech coding could be very attractive in all that applications where a common resource, in a fixed amount, is shared by many users or when the network impairments reduce the effective channel capacity. For example, in a mobile radio system, where speech coders have to provide good quality and to be robust against transmission errors, variable rate coding allows dynamic control of bit-rates assigned to speech and error protection. There is a strong demand for significantly increasing the capacity of the current mobile radio system. This can be done by converting the existing analog system to a digital system. All subsystems of future digital cellular mobile radio systems such as speech and channel coding, multiple access and digital modulation methods are currently being extensively studied [1],[2].Also different applications, as satellite and long haul connections, voice store and forward, could benefit by coders employing this technique.

Multirate speech coding could take advantage of the use of embedded quantizers that allow bit deletion and insertion if the proper information is sent to the receiver.

The need for unequal error protection arises in many source coding applications for example in pulse code modulation (PCM) where the effect of an error in the most significant bit (MSB) is more damaging than an error in the least significant bit (LSB). In many situations of mobile radio system, the channel is extremely time varying. Depending on the local channel characteristics, one would like to change the allocation of the bits for speech and channel coding. On a good channel, fewer bits should be spent on channel coding than on a noisy channel so that the overall noise due to both quantization and the channel distortion is minimized. Thus,a flexible and adaptive communication system should be built to satisfy the need for unequal error protection.

In this paper,a new joint speech-and- channel coding system is proposed. The main idea of this system is splitting input speech signal into its integer and fractional parts and their different processing to fulfil the demands of bit-rate-reduction of speech signal and unequal error protection channel coding, respectively.

2. Speech and channel coding system

Traditionaly, speech and channel coding have had complementary roles in digital communication. The speech coder has tried to minimize the bits-per-sample for high-quality signal representation, while the channel coder (modulator and error protection system) has attempted to maximize the bits-per-second-per Hertz that can be used on a transmission or storage medium.

If the output of a speech coder can be categorized into parts of varying sensitivity to bit errors in transmission, a given total overhead for error protection can be used in an unequal error protection scheme that will have the final effect of extending the range of channel quality over which a specified quality of signal communication can be maintained. In situations where the transmitter has information about channel quality, a joint speech-and-channel coding algorithm can make a suitable allocation of the total bit rate for speech coding and error protection. This again, has the effect of utilizing transmission media at low levels of channel quality: a slight undercoding of the signal in a quantization-noise sense, together with a stronger focus on error protection, can realize a specified level of total (quantization-plus-channel) noise over a wider range of channel quality. Recent examples of this appear in digital cellular technology for mobile radio telephony and digital transmission proposals for high-definition television.

There are many possible alternatives for the design of speech coders [3]. Speech quality, robustness against channel errors, complexity, and overall delay are some of the important properties.

The design of an error protection scheme usually consists of selecting a fixed channel code with a certain rate, complexity, and correction capability that is uniform for all the data to be transmited. In many cases, the data to be transmitted have different error protection needs and the channel is time varying. An example of the former is PCM word where the effect of an error in the MSB is more damaging than an error in the LSB.

A block diagram of the system considered in this paper is shown in Fig.1.

number of bits to be deleted in either case depends on the local channel characteristics by which one would like to change the allocation of the bits for speech and channel coding. On a good channel, fewer bits should be spent on channel coding than on a noisy channel.

To support the idea of possible bit-rate-reduction of speech signal in this system we have implemented a transform operation on the Fx_n part of the signal. For simplicity we have decided to use Discret Fourier Transform (DFT block), but other transforms can be used for improving the efficiency of the system [4],[5]. The bit-rate-reduction have been achieved in block DELET 3 by deleting some part of the DFT spectrum.

Since the more sophisticated digital voice coding schemes generate blocks of digital symbols, a block channel coding scheme would seem to be natural. However, no conventional algebraic block coding scheme is known which can accomodate broadly varying unequal error protection needs within one codeword. Here, we suggest the use of rate compatible punctured convolutional

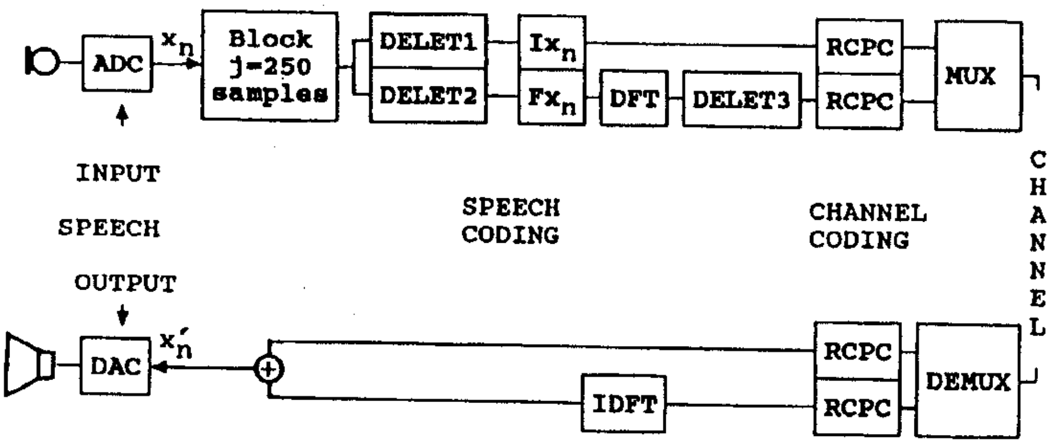

Fig.1
Schematic of speech and channel coding system

Since our intention is to describe the basic principles rather than details of the proposed system, we shall shortly introduce the individual blocks.

Deleting blocks yield an "integer part" denoted by Ix_n (block DELET 1) or a "fractional part" denoted by Fx_n (block DELET 2). The Ix_n part extraction is the same as quantization with a midtread quantizer whereas the Fx_n part is the resulting quantization error. The Fx_n part has its spectrum partially whitened compared to the original spectrum of $\{x\}$. The fractional part of a signal is simply obtained by deleting the MSB's of its PCM binary words. Similarly, the integer part of a signal is obtained by deleting the LSB's of its PCM binary words. Let us suppose that c, $(c = m + k)$ is the total number of bits in PCM word. If we delete k LSB's in the block DELET 1 to obtain Ix_n part of the signal then block DELET 2 deletes m MSB's to make Fx_n part of the signal. The

codes (RCPC) because they match requirements in handling unequal error protection[2]. The main advantage of using the RCPC codes is that optimal decoding of the transmitted signals in fading and noise can be efficiently performed by the Viterbi algorithm.

3. Experimental results

For the numerical results, we have used the word "KNIHA", which means book in Slovak. This word had been pronounced by male speaker and divided into segments each containing 250 samples (corresponding to 25 ms at 10 kHz sampling rate) and SNR was measured over each segment. The average of these segmental measurements over S segments making up an utterance is

p [kHz]	4,8	4,4	4	3,6	3,2	2,8	2,4
k [bits]	SNR_{seg} [dB]						
4	53,33	48,78	46,77	45,82	44,4	42,3	38,15
5	47,82	42,55	39,73	38,3	36,65	35,19	32,3
6	46,25	37,97	35,11	31,88	30,26	29,66	28,63
7	37,14	31,81	28,05	26,86	25,55	24,12	23,68
8	33,62	26,57	23,81	20,49	18,96	18	18,15
9	24,78	19,53	17,47	15,52	14,46	13,79	12,4
10	19,8	16,74	13,58	12,41	11,6	10,48	9,32

Table 1
The resulting SNR_{seg} of the utterance "KNIHA"

$$SNR_{seg} = \frac{1}{S}\sum_{s=0}^{S-1} 10\cdot\log_{10}\left[\frac{\sum_{j=0}^{249} x_{j+250s}^2}{\sum_{j=0}^{249}\left(x_{j+250s} - x'_{j+250s}\right)^2}\right]$$

where x is the original input and x' is the regenerated output speech signal, respectively.

The speech is band limited between 30 and 5000 Hz, sampled at 10 kHz and 12 bits per sample ADC has been used to obtain speech samples for our simulations.

Only speech coding part of the above described system is evaluated in this paper through a mixture of analysis and simulations.

Our experiments have mainly been oriented towards

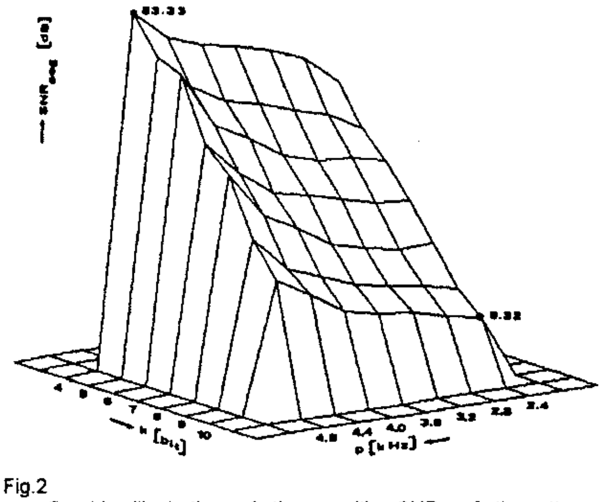

Fig.2
Graphic illustration of the resulting SNR_{seg} of the utterance 'KNIHA'.

blocks DELET 1, DELET 2, DFT and DELET 3, to find out the behaviour of the proposed system, when some limitations to the system have been imposed. These limitations are concerned with the reduction of the Fx_n spectrum, so that some rate of bit-reduction could be achieved. This has been obtained by deleting a certain part of the Fx_n spectral frequencies. Experiments have been done when the fractional part of the signal signal consisted of k=4,5,6,7,8,9 or 10 LSB of the input PCM codeword which corresponded to m=8,7,6,5,4,3 and 2 MSB of the integer part of the signal. After deleting operations, DFT of

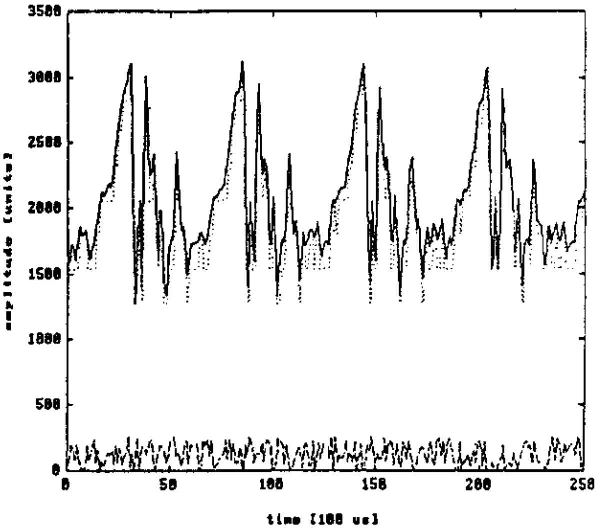

Fig.3
Time waveform of the input speech signal (solid line), integer part Ix_n (dotted line) and fractional part Fx_n (below) for a speech segment of the utterance 'KNIHA'.

the Fx_n signal is imposed and only some part of this spectrum cuts in block DELET 3 at frequency p [kHz] has been transmitted to regenerate signal by IDFT. The remaining part of the spectrum is put to zero. As we have mentioned above the block of encoders and decoders (RPCP) have not been analysed in this paper.

The resulting segmental SNR_{seg} is shown in Tab.1 and Fig.2 ilustrates these results graphically.

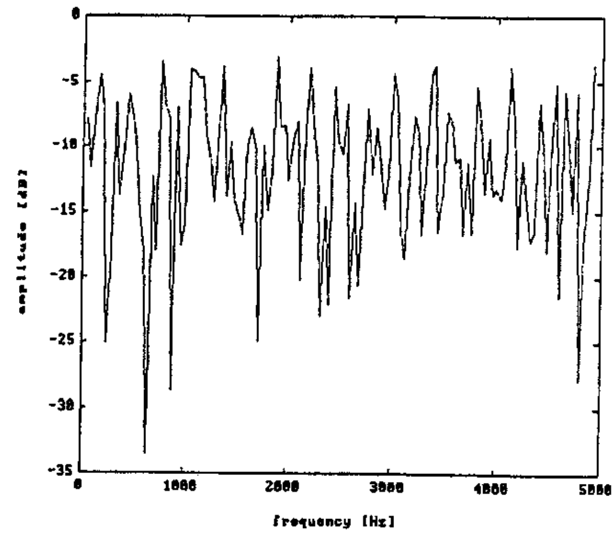

Fig.4
DFT spectrum of the fractional part Fx_n for the speech segment illustrated in Fig.3.

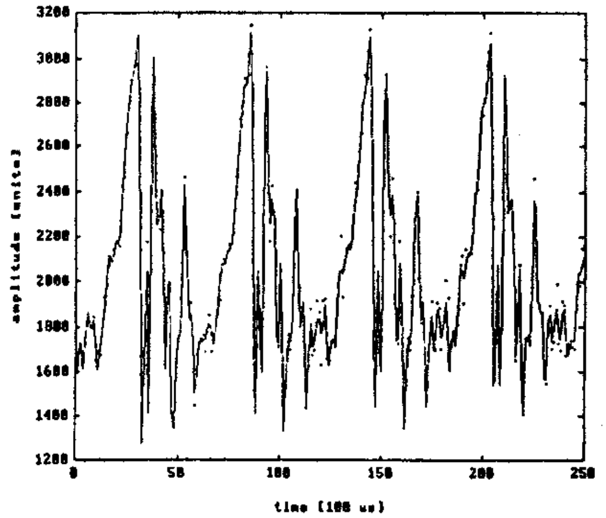

Fig.5
Time waveform of the input speech signal (solid line) and regenerated speech signal (dotted line) for a segment of the utterance 'KNIHA'.

Figures 3, 4, 5, and 6 show the results of some individual blocks for k=8 (number of deleted LSB's) and for cuting frequency p=2.4 kHz, behind which all the frequencies of Fx_n spectrum are zeros (not transmitted). For this case we can see from Tab.1 the resulting SNR_{seg} = 18.15 [dB]. As one can see from Fig.5 and Fig.6, the timewaveform and spectrum of the output signal, are very good aproximations of the input signal.

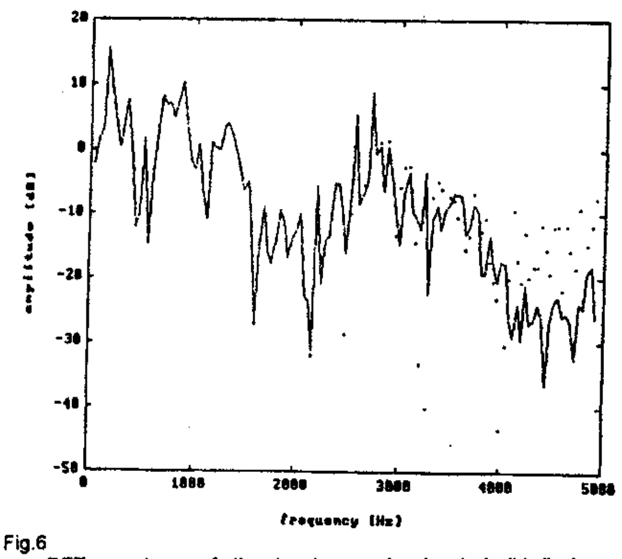

Fig.6
DFT spectrum of the input speech signal (solid line) and regenerated speech signal (dotted line) for the speech segment illustrated in Fig.5.

4. Conclusion

This paper has presented opportunities for integrating speech and channel coding system. Perceptual tests of this system have shown very good quality of regenerated speech when compared to the original speech.

5. References

[1] R.V.Cox, J.Hagenauer, N.Seshadri, and C-E.W.Sundberg, "Subband speech coding and matched convolutional channel coding for mobile radio channels," IEEE Trans. Signal Proces., vol.39, No.8, pp.1717-1731, Aug.1991.

[2] J.Hagenauer, N.Seshadri, C-E.W.Sundberg, "The Performan ce of Rate-Compatible Punctured Convolutional Codes for Digital Mobile Radio," IEEE Trans. Commun., vol.38,No.7, pp.966-980, July 1990.

[3] N.Jayant, "Signal Compression: Technology Targets and Research Directions," IEEE Journal on Selectes Areas in Communication," vol.10, No.5, pp.796-818, June 1992.

[4] I.Franková, A. Čižmár, "Bit-Rate Reduction of Speech Signal Using the Rapid Transform," Slaboproudy obzor (Electronic Horizon), No.9-10, pp.160-164, 1992.

[5] A.Čižmár, "Carpenter-Grossberg Neural Network Model for Speech Wavewform Processing," Procedings of the VI.th conference of Technical University in Kosice, section Radioelectronics, pp.84-89, Sept.1992.

About author, ...

Anton Čižmár was born in Michalovce, Czechoslovakia, in 1956. He graduated from the Slovak Technical University in Bratislava in 1980. He received Ph.D. degree in Radioelectronics from Technical University of Košice, in 1986. Now he is Associate Professor at Department of Radioelectronics Faculty of Electrical Engineering Technical University of Košice. His research interests include speech processing, neural networks, data compresion and digital communications.