TRANSMISSION OF VARIABLE-RATE ENCODED SPEECH SAMPLES ON PACKET RADIO NETWORKS *

Fulvio Babich, Sergio Carrato and Francesca Vatta
D.E.E.I., University of Trieste
via A. Valerio, 10, 34127 Trieste, Italy
Tel. +39 40 6763458 - 6767147; Fax: +39 40 6763460
e-mail: babich@univ.trieste.it
e-mail: vatta@univ.trieste.it
e-mail: carrato@imagets.univ.trieste.it

ABSTRACT

This paper presents the performance evaluation of different speech coding techniques in wireless packet switching networks: the goal of our study is to increase network capacity while maintaining a smooth degradation of quality at high loads and heavy interference, in order to make it possible for different kinds of information to coexist in a single network infrastructure. In the paper we propose a variable-rate multimode and embedded encoding technique as effective for handling network congestion and channel impairments that both cause discarding or erasure of frames of information. Therefore this approach is important not only in TDMA packet switched communications with statistical multiplexing (leading to greater efficiency and flexibility than basic TDMA, that assigns a fixed portion of channel resources to each user), but also in a CDMA-based mobile system that is strictly limited by interference.

1 INTRODUCTION

Nowadays there is an evolution towards Personal Communication Systems (PCS), i.e. towards the creation of a single network infrastructure that will make it possible for all people to transfer information between any desired locations, but whose capacity and channel reliability will be strictly limited, as well as, therefore, its transmission quality. Consequently, this paper presents the performance evaluation of different speech coding techniques in wireless packet switching networks, focusing on the problem of delivering a high capacity service for speech, data and supplementary services at a high quality level.

The goal of our study is to increase network capacity while maintaining a smooth degradation of quality at high loads and heavy interference, in order to make it possible for different kinds of information to coexist in a single network infrastructure. In order to achieve the above mentioned goal, embedded and variable bit-rate (VBR) multimode encoding should be combined to fully exploit the capacity of the system: more precisely, such coding schemes solve the difficult problem of having the output bit-rate vary in an optimal sense, in response to source entropy as well as channel quality and network congestion. We want here to emphasise that we can consider channel quality and network congestion as two aspects of the same problem, i.e. packet or frame loss, because in the case of a radio channel with a deep signal fade the effect of channel impairments is equivalent to that of packet losses due to network congestion, even if the access protocol does not involve explicit packetisation. Therefore this problem is important not only in TDMA packet switched communications with statistical multiplexing (leading to greater efficiency and flexibility than basic TDMA, that assigns a fixed portion of channel resources to each user), but also in a CDMA-based mobile system that is strictly limited by interference.

Furthermore, since channel and network conditions play a major role in the design of the coding scheme, there is the need to increase interaction of source coding technology with other disciplines such as channel coding and networking: this is the reason why the coding scheme should include a proper interaction between source and channel coding and between coding and networking (multiple access), taking into account the current channel and network conditions.

In the following, section 2 deals with the advantage of using a variable-rate speech coding technique instead of a fixed-rate one, because it allows a significant gain in capacity. Section 3 deals with the advantage of using, in addition, an embedded scheme, that provides a flexible way to alleviate congestion at any point in the network. Finally section 4 comprises the conclusions and future work arguments.

* This work has been partially supported by MURST "ex quota 40%"

2 VARIABLE-RATE

In a mobile telecommunication system based on TDMA packet switched communications with statistical multiplexing, the use of a variable rate transmission scheme is fundamental in order to completely exploit the capacity of the system.

More precisely we have made a comparison between the Fast Variable Rate CELP coder (FVR-CELP), developed by CSELT S.p.A. in the UMTS Code Division Testbed (CODIT) Project [1], and the fixed-rate ITU-T G.728 16 kbit/s standard [2], which is assumed as a reference in terms of speech quality. The FVR-CELP algorithm has an actual rate selected, on a frame by frame basis (being the frame length in the order of 10 ms) according to the local character of the input signal, among seven rates ranging from 400 bit/s to 16 kbit/s, with an average rate below 7 kbit/s over typical conversational input speech. The lower two rates are devoted to represent background noise, while the higher five ones are intended to be used for speech. These five rates are achieved by including a number of excitation contributions, variable from one up to three.

Objective and informal subjective measurements were carried out employing the two above mentioned coding algorithms (FVR-CELP and G.728 LD-CELP) to process a set of sentence pairs over a range of frame lost or erasure conditions. In order to quantify the objective speech degradation inflicted by network conditions (expressed in terms of 10 ms frame loss percentage), we have employed two widely used objective distance measures, namely, the segmental signal-to-noise ratio (SEGSNR) and the log-spectral distance measure (SD). The former evaluates the SNR in dB for typically 15-30 ms long speech segments (chosen: 16 ms) and averages these values in terms of dBs, giving more fair weighting to low-energy unvoiced speech segments during quality evaluation. In contrast, the SD quantifies the spectral degradation due to processing and, hence, it is more appropriate for the quality assessment of the coder. In our graphs we report the absolute SD values as shown in Fig. 1. The graphs also include the performance of the embedded coding schemes that will be explained in the following section (dotted lines).

Informal subjective measurements of the speech degradation inflicted by frame erasures were done by using the ACR (Absolute Category Rating) method, that employs the five-point rating scale reported in Table I [3]. The experiment was designed to evaluate the transmission performance of the variable bit-rate (400 bit/s to 16 kbit/s) coding algorithm, compared to the LD-CELP codec conforming to the ITU-T Recommendation G.728 (16 kbit/s). The two coding algorithms were employed to process a set of sentence pairs, uttered by four different speakers (2 males and 2 females), the four talkers producing 4 sentence-pairs each, before carrying out the informal listening subjective tests. The language was Italian and 16 listeners were employed. In our graphs we report the measure of the codec performance in terms of Mean Opinion Score (Fig. 2) and equivalent Q values (Fig. 3), including four MNRU (Modulated Noise Reference Unit) anchoring points (namely 5, 15, 25 and 35 dB) in the listening tests. The graphs also include the performance of the embedded coding schemes that will be explained in the following section (dotted lines).

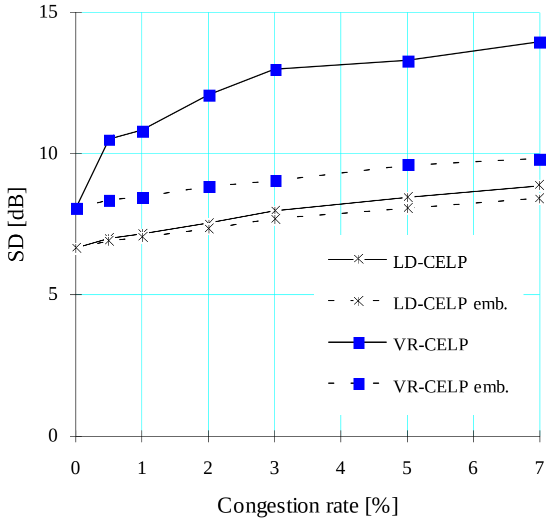

Figure 1 - Speech SD degradation versus congestion rate

Mean Opinion Score	Quality Opinion
5	Excellent
4	Good
3	Fair
2	Poor
1	Bad (or Unsatisfactory)

Table I - Rating scale

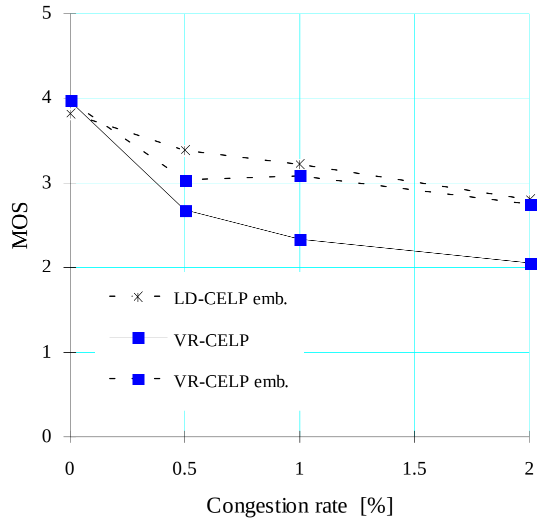

Figure 2 - Speech MOS degradation versus congestion rate

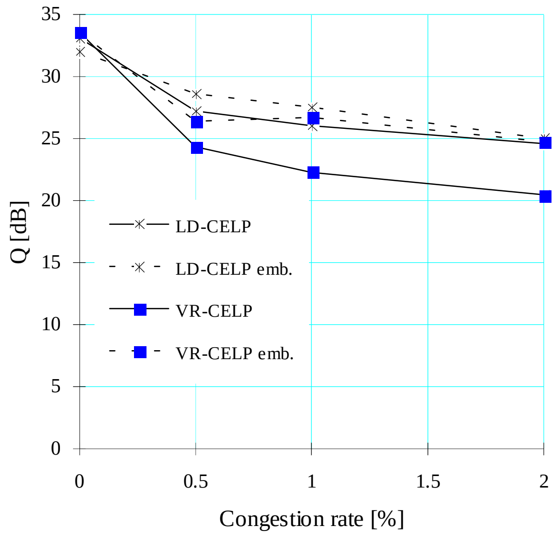

Figure 3 - Equivalent Q rating degradation versus congestion rate

The use of a variable rate transmission scheme [4], being its output bit-rate variable in an optimal sense in response to source entropy, allows a significant gain in capacity because its average bit-rate is much lower than a comparable unimodal fixed-rate coder, but, as one can see from Figs. 1-3, suffers a more evident degradation of quality at high loads and heavy interference (i.e. increasing the frame loss percentage), without introducing any error concealment technique or embedding scheme. Indeed, multimode VR-coding efficiently adapts the coding scheme and parameter bit-allocation to suit the short-term statistical character of the speech, and, therefore, when the frame loss percentage increases, the perceived quality of the VR coder tends to degrade much more for some speech segments than for others, while the perceived quality of the fixed-rate coder degrades in the same way for all speech segments, giving a better overall quality result. This difference is evident looking at the SD and Q values that, as said above, are more appropriate for the quality assessment of the coders.

3 EMBEDDING

Since embedding produces a data stream from which reduced data streams can be extracted by a simple bit dropping procedure [4], the main reason for using an embedded speech coding technique, in addition to a VR multimode encoding, is to provide a flexible way to alleviate congestion at any point in the network (where for network we mean the digital radio access link and the wired network) without the need to exchange control messages among various nodes in the backward path of the connection: frames, in fact, can be dropped in any point of the transmission path, without changing encoder and decoder configuration.

If under heavy traffic we apply cell discarding to fixed-low-bit-rate coding schemes, the cell loss can seriously degrade speech quality. We have to use, therefore, coding schemes that allow us to divide the code word in two or more segments of different importance, and to assemble these segments into separate packets: categorising successively these voice packets as high-priority (more significant) and low-priority (less significant) information, the low-priority packets may be discarded during congestion. These coding schemes (such as the CCITT G.727 embedded ADPCM algorithm [5]) give graceful and smooth degradation of speech quality at high loads by removing code bits beginning with the LSB (Least Significant Bits), resulting in shorter packets that reduce the network load. The two CELP algorithms mentioned in the previous section are amenable to an embedded coding scheme too, such that bits of decreasing significance are assembled into packets of decreasing priorities so that lower priority packets can be dropped. In order to classify bits of decreasing significance in the two coded bit-streams (obtained employing the FVR-CELP and G.728 LD-CELP algorithms) we have adopted two commonly used approaches [6]. One approach of quantifying the sensitivity of various coded bits is to perturb systematically a given bit in every frame of the output bit-stream and evaluate the SEGSNR or SD degradation. The problem with this approach is that it does not take adequate account of the different error propagation properties of different bits. Therefore, we have also used another error sensitivity measure, which takes account of the error propagation effects. For each bit we have found the average SNR degradation due to a single bit error both in the frame in which the error occurs and in the frames following the frame in error. On the basis on these results, the total SNR degradation is then found by integrating the degradation caused in the erroneous frame and in the following frames that are affected by the error. The two sets of degradation figures have been combined to produce an overall sensitivity measure, which has been invoked to assign the speech bits to various priority packets. In Fig. 4 we report the normalised SEGSNR degradation of the FVR-CELP codec due to 100% bit error rate and the total normalised SNR degradation of the FVR-CELP codec due to the propagation of single errors, considering the most significant bits of mode 7.
Objective and subjective measurements were carried out employing the two above mentioned coding algorithms, to process the same set of sentence pairs used in the test described in the previous section, over a range of congestion conditions, but using an embedded coding scheme: namely, bits of decreasing importance have been assembled, according to the hierarchy determined with the above mentioned sensitivity measurements, into packets of decreasing priorities, dropping only the lower priority packets, instead of dropping the whole 10

ms frame as in the previous test. The measurement results are reported in Figs. 1-3: from the figures we derive that the use of an embedded coding scheme makes the variable rate transmission scheme suffer a degradation equal to that suffered by a unimodal fixed-rate coder. It has to be reminded, however, that the variable rate encodings permit one to achieve a significant capacity gain, since, due to lower average bit rate, a similar congestion or interference condition corresponds to a higher number of users in the system, being the actual gain dependent on the user statistical characteristics.

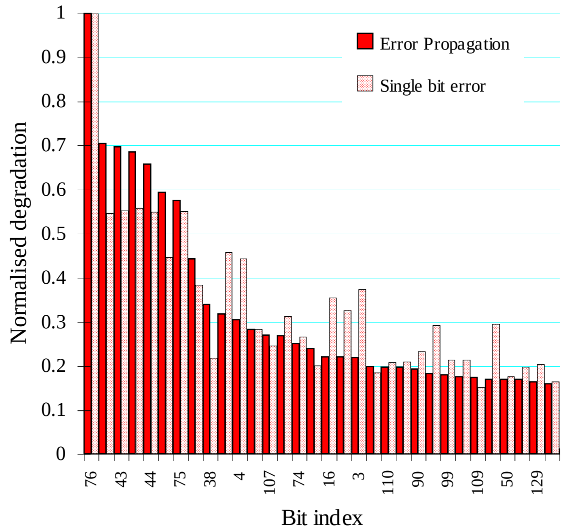

Figure 4 - Normalised SEGSNR degradation of the FVR-CELP codec due to 100% bit error rate and total normalised SNR degradation of the FVR-CELP codec due to the propagation of single errors in various bits.

4 CONCLUSIONS AND FUTURE WORK

In this paper it has been shown that a FVR-CELP coder is suitable to manage network congestion and channel impairments (with respect to the uncontrolled "stealing" of speech frames present in a mobile telecommunication system). By using an embedded coding scheme, the variable rate transmission scheme suffers a degradation equal to that suffered by a unimodal fixed-rate coder, nevertheless the former scheme permits one to achieve a significant capacity gain, since, due to lower average bit rate, a similar congestion or interference condition corresponds to a higher number of users in the system.
Some hints about the features of the FVR-CELP coder and some results about the evaluation of the sensitivity of various bits to errors, in order to introduce an embedded coding scheme, have been presented too.
This analysis can be considered preparatory to a future performance analysis of the FVR-CELP coder in applications involving channels with time-varying error characteristics. For such situations, each mode of the FVR-CELP coder can have a particular allocation of bits between source and channel coding stages, leading to an adaptive and unequal error protection code tailoring the encoding operation to suit both the different sensitivities of various bits to errors (pointed out in the previous section), and also the time-varying channel conditions.
In order to perform the above mentioned allocation of bits between source and channel coding stages, it is necessary to know in detail the channel characteristics and to quantify the objective quality degradation inflicted by channel impairments (i.e. by random and burst errors) to bit groups of different importance, in order to assign the speech bits to different bit protection classes (Unequal Error Protection). This analysis will be argument of a future work.

ACKNOWLEDGEMENTS

The authors want to thank CSELT S.p.A. for the provision of the FVR-CELP coder and Ing. Carlo Fabbiani for his help in the provision of measurement data.

REFERENCES

[1] L. Cellario and D. Sereno, "CELP Coding at Variable Rate", European Transactions on Telecommunications, Vol. 5, No. 5, September-October 1994, pp. 603-613.

[2] J.-H. Chen, R. V. Cox, Y.-C. Lin, N. Jayant and M. J. Melchner, "A Low-Delay CELP Coder for the CCITT 16 kb/s Speech Coding Standard", IEEE Journal on Selected Areas in Communications, Vol. 10, No. 5, June 1992, pp. 830-849.

[3] S. Dimolitsas, "Subjective quality quantification of digital voice communication systems", IEE Proceedings-I, Vol. 138, No. 6, December 1991, pp. 585-595.

[4] A. Das, E. Paksoy and A. Gersho, "Multimode and Variable - Rate Coding of Speech", in "Speech Coding and Synthesis", Chapter 7, pp. 257-288, Elsevier Science B.V., Amsterdam, 1995.

[5] M. H. Sheirf, D. O. Bowker, G. Bertocci, B. A. Orford and G. A. Mariano, " Overview and Performance of CCITT/ANSI Embedded ADPCM Algorithms", IEEE Transactions on Communications, Vol. 41, No. 2, February 1993, pp. 391-399.

[6] L. Hanzo and J. P. Woodard, "An Intelligent Multimode Voice Communications System for Indoor Communications", IEEE Transactions on Vehicular Technology, Vol. 44, N. 4, November 1995, pp. 735 - 748.