SPLIT-BAND LD-CELP WIDEBAND SPEECH CODING AT 24 KBIT/S

Andrea Santilli(*), Aurelio Uncini(**), Francesco Piazza(**)

(*) AETHRA S.r.L. 60020 Palombina (AN) - Italy
(**) Dip. Elettronica ed Automatica, Univ. of Ancona, 60131 Ancona, Italy
Phone: +39 71 220 4453 Fax: +39 71 220 4464
e-mail: upfm@eealab.unian.it

ABSTRACT

Nowaday 7 kHz wideband speech coding requires at least 48 kbit/s as it still depends on the ITU standard G.722. CELP coders have been developed for wideband systems achieving high quality speech coding at rates from 16 kbit/s to 32 kbit/s as the wideband LD-CELP at 32 kbit/s.

In this paper, a new split-band LD-CELP wideband coder at 24 kbit/s is proposed and its performance and complexity are compared with those of the already known wideband LD-CELP.

1. INTRODUCTION

The technologies of ISDN tele-conferencing, have raised a lot of interest in advanced coding algorithm for 7 kHz wideband speech. In fact the increased audio bandwidth gives a significant improvement in overall perceived quality.

Narrow band speech is currently coded by the ITU 16 kbit/s standard G.728 (LD-CELP) [1,2] while, at the moment, wideband speech coding requires at least 48 kbit/s as it still depends on the ITU standard G.722 (SB-ADPCM) [7]. With the introduction of 32 kbit/s, or less, wideband coders, the remaining bit-rate could be used to increase the video quality as the overall rate is fixed (128 kbit/s on the most used 2B ISDN service).

In recent years, CELP coders have been developed for wideband systems [3-6], and have achieved high quality speech coding at rates from 16 kbit/s to 32 kbit/s. For istance the well known LD-CELP has been modified [3] to suit the wideband voice signal using 32 kbit/s.

As pointed out in [3,7], the main problem associated with wideband speech is the spectral noise weighting because the voice signal is very weak in the added high frequencies while noise, with CELP, tends to be white; so an enhanced weighting filter has been adopted to achieve an higher spectral tilt with limited formants.

Another drawback typical of wideband CELP is its gross computational load. For istance in [3], in order to control it, they are compelled to reduce the synthesis filter order from 100 (the proper transposition of the G.728 value 50) to 32; so the coder is not able to exploit the pitch periodicity in voiced speech.

In this paper to overcome the previous problems, a new split-band LD-CELP wideband coder at 24 kbit/s is proposed and its performance and complexity are compared with those of the already known wideband LD-CELP.

2. THE SPLIT-BAND LD-CELP

The main idea of this work is to employ the standard LD-CELP in a split-band scheme (SBLD-CELP), as shown in Fig. 1, where the strict similarity with the G.722 (SB-ADPCM) is apparent. The 16 kHz sampling signal is decomposed in two 8 kHz components by a QMF bank [8]. The lower one is comparable to narrow band speech and so it can be efficiently coded by LD-CELP G.728.

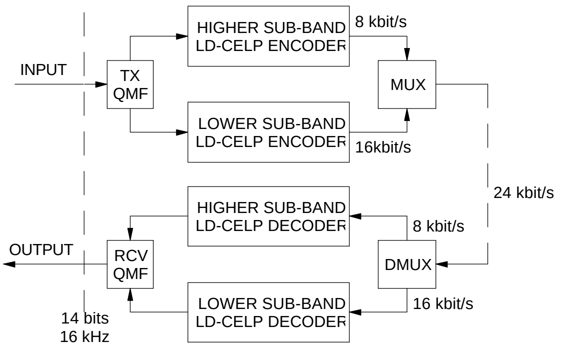

Fig. 1 The overal scheme of the proposed wideband coder.

The higher one is highly unstructured and so it can be coded by a reduced version of the LD-CELP: weighting filter and postfilter are eliminated while the synthesis filter and log-gain predictor order are reduced to 10 and 6 respectively (Fig. 2). Furthermore vector size is doubled accomplishing a 1 bit/sample coding (8 kbit/s). This way an overall 24 kbit/s wideband coding is achieved, ensuring the proper unequal bit and complexity allocation between the two bands.

Such solution exhibits computational saving, as each sub-band scheme does, when filtering and/or autocorrelation are involved; this makes possible to exploit long-term correlations.

Furthermore it avoids the spectral noise spreading that reduces the SNR at high frequencies in a full-band CELP, because each sub-band is treated separately.

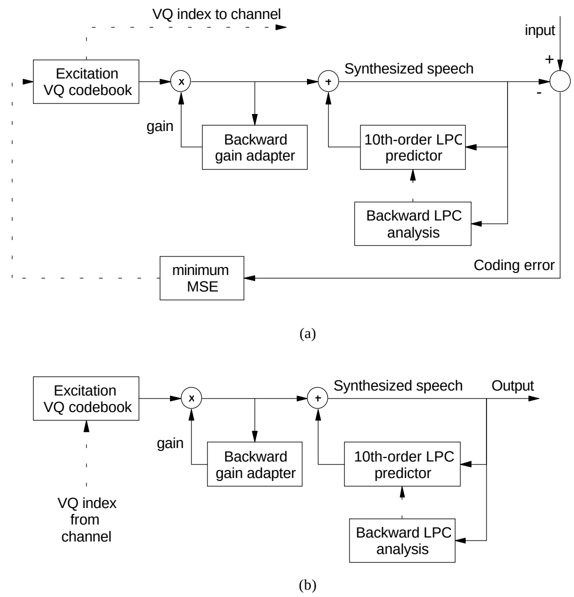

Fig. 2. (a) Highest sub-band Encoder.
(b) Highest sub-band decoder.

It is worthwhile to point out that if postfiltering (optional in G.728) is eliminated also in the lowest sub-band decoder, both channels have a fixed and well-known delay; so it is easy to restore the proper alignment between them, as required by the QMF bank [8] in order to perform aliasing cancellation. However, a full-band postfilter could be added, although it is not necessary. It should be a wideband transposition of the postfilter used in G.728 employing a 20th-order pole-zero short-term postfilter; but in this case the LPC predictor used to compute the filter coefficients a_i , b_i cannot be obtained as by-product of the 50th-order backward LPC analysis, since that is performed into each sub-band separately. So a further 20th-order full-band LPC analysis should be performed, but this might be not worthwhile: a high computational power is required to get a modest quality improvement.

The Split-Band LD-CELP also exhibits another particular advantage over all other wideband coder in audiovisual teleservices [9,10]. As it encodes the lower channel by the G.728 standard, it is possible also for system lacking wideband codec, to properly decode just the lowest sub-band. This feature, a sort of backward-compatibility, is particularly interesting for multi-user systems because it does not compell all the user to choose the G.728 coding in case just one of them did not have the wideband codec.

3. PERFORMANCE

Based on the simulation results, our SBLD-CELP and a WideBand LD-CELP (WLD-CELP) developed according to [3] are compared: the segmental SNR (SNRSEG) has been computed for two sequences of speech, discriminating between female and male utterances. In order to highlight high frequencies quality, sub-band SNRSEG has been also considered. This is straightforward with our scheme, while a further QMF analysis of the reconstructed speech is required for the full-band scheme.

As shown in Table 1, the full-band LD-CELP exhibits a better overall SNRSEG but it has a worse SNRSEG on the highest sub-band. So our solution achieves less quantization noise at higher frequencies, where the auditory system is quite sensitive.

This fact can be highlighted also by the spectrograms of the quantization noise. In Fig. 3, the full-band scheme noise spectrogram is compared to that of our coder. Owing to the noise weighting, both spectrograms resemble to that of the coded voice, particularly at lowest frequencies. But it is apparent that with the split-band coder the reconstruction error energy is restricted in the lowest sub-band. This sort of noise segregation allows to exploit the auditory masking effect and so improves the perceived quality. That has been verified by informal listening tests where the performance of both schemes has been compared using either male and female utterances.

In the last column of Table 1 an estimation of the computational load, expressed as a ratio with that of the narrow band LD-CELP G.728 with no postfilter, is shown. It can be noted that both schemes have about the same complexity and that in the split-band scheme just a third of the computational load is devoted to the upper subband.

Even if low delay is not a primary need in videconference applications, it can be pointed out that our solution exhibits a low delay as the overall one-way delay is less than 4 ms

SNRSEG (dB)	**Male**	**Female**	**C.L.**
WLD-CELP	21.082	23.353	1.49
SBLD-CELP	17.393	20.939	1.48
WLD CH0	21.520	23.767	-
WLD CH1	3.186	3.017	-
SBLD CH0	18.611	21.220	0.92
SBLD CH1	4.575	6.138	0.56

Table 1. Full-Band and Split-Band LD-CELP comparison; CH0 and CH1 are the lowest and highest sub-band respectively.

4. CONCLUSION

In the paper, a new split-band LD-CELP wideband coder has been proposed. This coder is able to get high quality coding at 24 kbit/s with better performance in terms of SNRSEG at the higher frequencies. This results in a perceived speech quality better than that provided by the full-band solution. Both coders require the same computational burden, moreover the new scheme allowing the lowest sub-band to be decoded separately, is also backward-compatible with the standard narrow-band LD-CELP G.728.

REFERENCES

[1] J. H. Chen, R. V. Cox, Y. C. Lin, N. Jayant, M. J. Melchner, "A Low-Delay CELP Coder for the CCITT 16 kb/s Speech Coding Standard" in *IEEE Journal on Selected Areas in Communication*, Vol. 10 No. 5, June 1992, pag. 830-848.

[2] ITU - Recommendation G.728 (09/92) - "Coding of Speech at 16 kbit/s using Low-DElay Code Excited Linear Prediction".

[3] E. Ordentlich, Y. Shoham, "Low-Delay Code-Excited Linear-Predictive Coding of Wideband Speech at 32 kbps" in *Proc. ICASSP*, Maggio 1991, pag. 9-12.

[4] C. Laflamme, J. P. Adoul, R. Salami, S. Morisette, P. Mabilleau, "16 KBPS Wideband Speech Coding Technique Based on Algebraic CELP" in *Proc. ICASSP*, Maggio 1991, pag. 13-16.

[5] C. Roy, P. Kabal, "Wideband CELP Speech Coding at 16 kbits/sec" in *Proc. ICASSP*, Maggio 1991, pag. 17-20.

[6] A. Fuldseth, E. Harborg, F.T. Johansen, J. E. Knudsen, "Wideband speech coding at 16 kbit/s for a videophone application" in *Speech Communication* 11 (1992) North-Holland, pag. 139-148.

[7] N.S. Jayant, J.D. Johnston, Y. Shoham, "Coding of Wideband Speech" in *Speech Communication* 11 (1992) North-Holland, pag. 127-138.

[8] L. R. Rabiner, R. E. Crochierie, *Multirate Digital Signal Processing* Prentice Hall Cap. 7 pag. 378-386.

[9] ITU - Recommendation H.221 (1990) - "Frame structure for a 64 to 1920 kbit/s channel in Audiovisual Teleservices".

ITU - Recommendation H.242 (1990) - "System for establishing communication between audiovisual terminals using digital channels up to 2 Mbit/s"

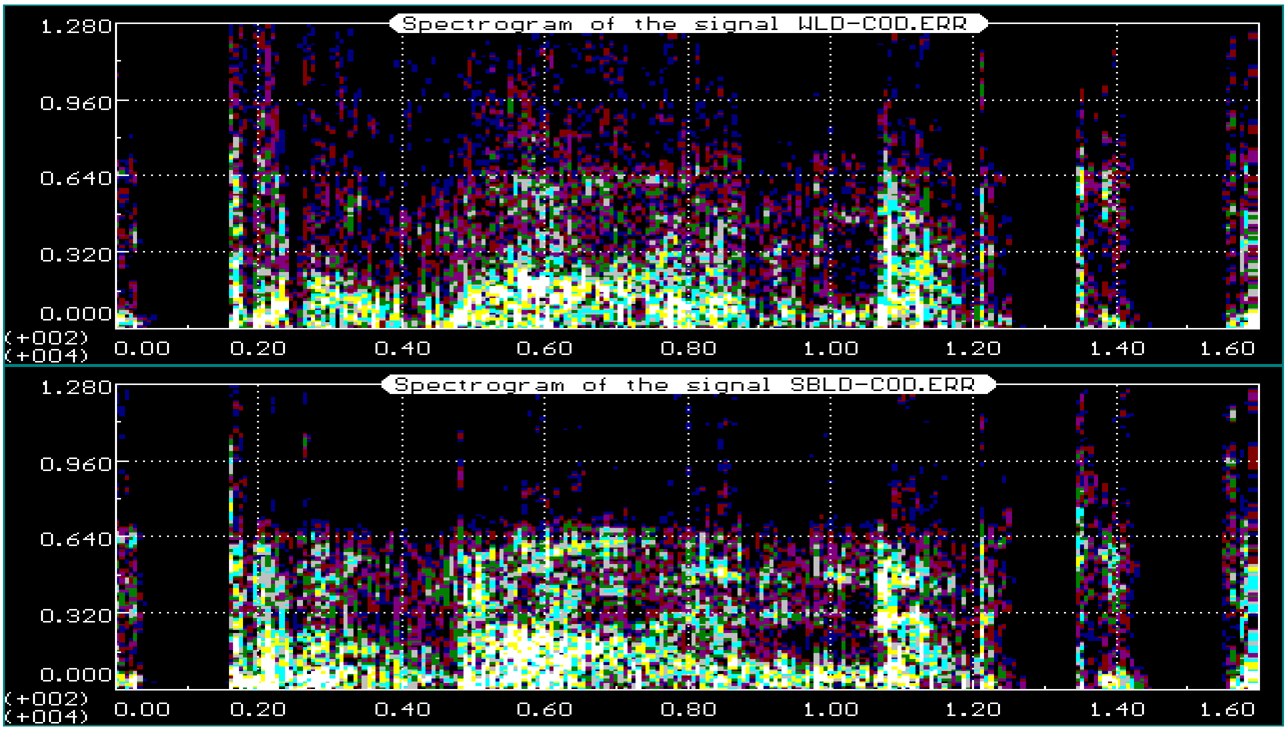

Fig. 3 - Noise Spectrograms: Full Band Coder (top) and Split Band Coder (bottom)