LSP Transcoding between G729A and GSM AMR

Helder C. Bertan and Luís G. P. Meloni
Universidade Estadual de Campinas, Campinas, SP, Brazil

***Abstract*—This paper presents a new method for Linear Spectrum Pair transcoding between G.729A and GSM AMR codecs. The transcoding in the bitstream domain gives better quality and lower complexity than the conventional method in the speech domain. Several simulation results are presented using Perceptual Quality Speech Measure showing a gain of about 0.45 in this measure. This scheme is part of a complete transcoding process under development.**

I. Introduction

THE convergence between packets and switching networks is now a reality. The circuit switches were conceived for uncompressed, real-time voice communications. The *Public Switching Telephone Network* (PSTN) involves wired and wireless accesses. For wireless access different voice codec standards have been used. The most common are IS-641 [1], IS-127 [2] and GSM 06.90 [3], for TDMA, CDMA and GSM cellular standards, respectively. The connection between wireless and wired networks is normally accomplished via *mobile telephone switches* (MTS). In the PSTN, the connection or path is exclusive, as it is shared with no other users. While each circuit acts independently in a call operation, they usually act in concert under command and control of a system such as the Signaling System 7 (SS7). Circuits network offers outstanding performance, but it is highly inefficient for any other real-time signals, such as audio and video.

Packets network was conceived for North American defense needs, as the Advanced Research Project Agency Network, later it spawned as the commercial Internet. This network transports data in discrete units known as packets, or datagrams, each of which is of a fixed minimum and maximum size. Packets are individually addressed, allowing the switches to share the physical resources among a huge number of users. This translates into greater efficiency and lower associated costs. The tradeoff is that the transmission performance can be poor as packets suffer important delays, jitters and losses through the network. Most packet protocols involve some sort of error detection and correction, so damage packages can be recovered or discarded. Lost and discarded packages are retransmitted, so the whole received file can be reassembled without errors. This is in short what TCP/IP on Internet does.

Helder C. Bertan and Luís G.P. Meloni are with the Departamento de Comunicações, Faculdade de Engenharia Elétrica e de Computação, Universidade Estadual de Campinas, SP, Brazil, 13.083-970, Phone: +55 19 3788-3818, Fax: +55 19 3289-1395.
Emails: {hcbertan, meloni}@decom.fee.unicamp.br.

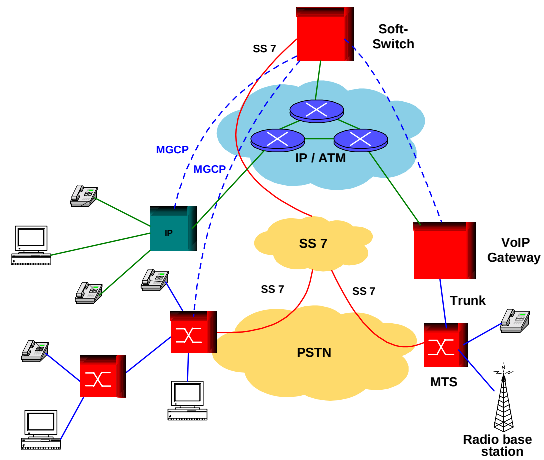

Fig. 1 Public Switching Telephone and Packet Networks Convergence.

The convergence between circuits and packets networks allows using networks resources more efficiently. The capacity to convey real time signals poses other demands over the packet network, more precise control on delays, jitters and bandwidths, in order to guarantee the quality of service. Besides voice, other signals such as audio and video are candidates to this “new” network, driven by new services such as audio cast and videoconference. The ATM (*asynchronous transfer mode*) aims to overcome some of the above difficulties. New protocols, such as SIP or MGCP illustrated in Fig.1, allow the telephony voice service over the Internet (VoIP). Several voice coders have been proposed to convey narrow-band voice packets over these networks, such as G.723.1 [4], G.729 [4] and G.729A [6].

One important module to link both networks is VoIP gateways, whose main function is voice codec conversions (e.g., G.711 to G.723.1 or G.729 and vice-versa). As the networks convergence takes place, those systems needs to deal with bulky traffic capacity, which is feasible with digital signal processor technology. We expected in this decade a huge traffic of telephony calls to flow over packet network, posing important requirements over voice codecs and transcoders in the VoIP gateways. In this scenario, it is much more likely to coexist both networks for a long time.

The conventional transcoding process uses a sequence of decoder and coder by the other codec. Preliminary studies show that there are many advantages in implementing transcoding in the bitstream domain instead of in the speech domain [7]. There are certain similarities between the coders IS.641, GSM 06.90 and G.729 that could be explored in an efficient transcoding process. This paper deals with the development of a scheme for transcoding the GSM 06.90 and G.729A, focusing in the transcoding of the LSP parameters. The GSM 06.90 is also called *adaptive multi-rate* (AMR) coder. The LSP transcoding method proposed in this paper applies to all GSM AMR bit rates, excepting the 12.2 kbit/s rate that uses two LSP analysis per frame. As shown in [7] for the coders the IS.641 and G.729, a complete transcoding process may be conducted in the bitstream domain, where the results show a reduction in the computation complexity of around 85%. These transcoding techniques are very promising for VoIP gateways in an MTS.

II. GSM AMR AND G.729A CODECS

The linear prediction analysis used in both coders GSM 06.90 and G.729(A) is similar. The G729(A) parameters are computed in a rate twice as high as the GSM 06.90 rate. Figure 2 shows the window positioning for both coders.

There is a lookahead of 5 ms in the linear prediction analysis; this means that 40 samples of future frame are necessary, introducing an additional delay of 5 ms in coding. The analysis window involves 120 samples of past frame, 80 samples of the current frame and 40 of future frame.

The windows are shifted at every 80 samples in the G.729(A) and at every 160 samples in the GSM AMR. For transcoding, a process of buffering, regression, and interpolation of parameters is necessary. Both coders use an asymmetrical window formed in partly by a Hamming window and partly by a cosine window, defined by

$$w_{lp}(n) = \begin{cases} 0.54 - 0.46\cos(\frac{2\pi n}{399}), & n = 0,...,199 \\ \cos(\frac{2\pi(n-200)}{159}), & n = 200,...,239. \end{cases} \quad (1)$$

After windowing, the main processing tasks are autocorrelation computation, format bandwidth expansion

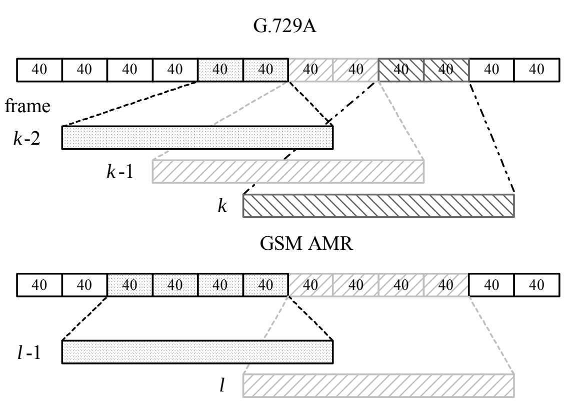

Fig. 2. Window positioning for G.729A and GSM AMR codecs.

by an exponential window, and Levinson-Durbin linear prediction [8]. The same procedure is used in both coders for conversion from LP parameters to Line Spectral Pair (LSP) parameters.

III. LSP TRANSCODING

The conventional transcoding process is to decode the signal to linear PCM and then apply the other coder, as shown in Fig. 3, as meant by the standard process for transcoding. This method presents a higher complexity computation than conversion in the bitstream domain. It will be shown by simulations that bitstream transcoding also offers better speech quality.

The Fig. 4 shows the bitstream conversion used in

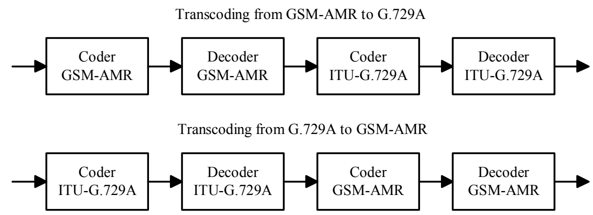

Fig. 3. Conventional transcoding.

transcoding. The task here is to develop a mapping process for conversion from one bitstream to another while maintaining high speech quality. Even a hybrid

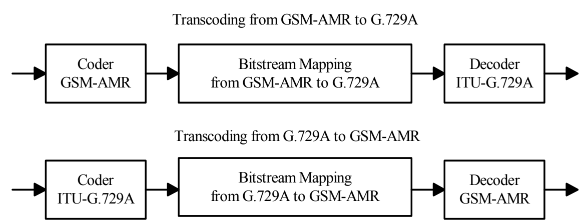

Fig. 4. Bitstream transcoding.

process combining this method and the previous one is attractive for purposes of complexity reduction.

The paper focuses only in the LSP transcoding. From the G.929(A) to the GSM AMR, it is necessary to buffer one frame to compute each LSP vector of the GSM AMR. In the reverse sense, a process of interpolation is necessary

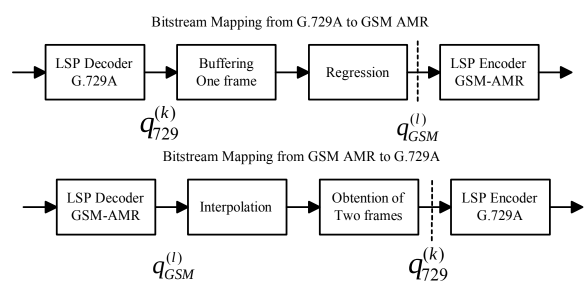

Fig. 5 LSP transcoding in the bitstream domain

to obtain the parameter rate of the G.729(A) (Fig. 5).

In both coders, the LSP parameters are computed frame by frame, so an interpolation scheme is used for computation of subframe LSP vectors. For the G.729 coder, the interpolation equations for frame k are:

$$\mathbf{q}_{1,729}^{(k)} = 0.5\mathbf{q}_{729}^{(k-1)} + 0.5\mathbf{q}_{729}^{(k)} \qquad (2)$$
$$\mathbf{q}_{2,729}^{(k)} = \mathbf{q}_{729}^{(k)}$$

and for the GSM AMR, frame l, where l=k/2:

$$\mathbf{q}_{1,\mathrm{GSM}}^{(l)} = 0.75\mathbf{q}_{\mathrm{GSM}}^{(l-1)} + 0.25\mathbf{q}_{\mathrm{GSM}}^{(l)} \qquad (3)$$
$$\mathbf{q}_{2,\mathrm{GSM}}^{(l)} = 0.5\mathbf{q}_{\mathrm{GSM}}^{(l-1)} + 0.5\mathbf{q}_{\mathrm{GSM}}^{(l)}$$
$$\mathbf{q}_{3,\mathrm{GSM}}^{(l)} = 0.75\mathbf{q}_{\mathrm{GSM}}^{(l-1)} + 0.25\mathbf{q}_{\mathrm{GSM}}^{(l)}$$
$$\mathbf{q}_{4,\mathrm{GSM}}^{(l)} = \mathbf{q}_{\mathrm{GSM}}^{(l)}$$

The following error criteria has been proposed for conversion from G.729 to IS.641 [7]:

$$\varepsilon^{(l)} = \sum_{s=1}^{4} W_s (\mathbf{q}_{s,729}^{(2l)} - \mathbf{q}_{s,\mathrm{GSM}}^{(l)})^{\mathrm{T}} (\mathbf{q}_{s,729}^{(2l)} - \mathbf{q}_{s,\mathrm{GSM}}^{(l)}) \qquad (4)$$

where W_s is a weighting factor for the subframes, and the superscript T indicates the transposed vector. We observe that in the above equation, besides the transmitted parameters, the linear interpolated parameters are also necessary. In this paper, we propose a different error criteria that gives similar results when using (4), but it conducts to a less complex scheme:

$$\varepsilon^{(l)} = \sum_{s=2,4} W_s (\mathbf{q}_{s,729}^{(2l)} - \mathbf{q}_{s,\mathrm{GSM}}^{(l)})^{\mathrm{T}} (\mathbf{q}_{s,729}^{(2l)} - \mathbf{q}_{s,\mathrm{GSM}}^{(l)}) \qquad (5)$$

Substituting (2) and (3) in (5) and optimizing in relation to $\mathbf{q}_{\mathrm{GSM}}^{(l)}$, it conducts to the following equation

$$\mathbf{q}_{\mathrm{GSM}}^{(l)} = \frac{2W_2\, \mathbf{q}_{729}^{(2l-1)} + 4W_4\, \mathbf{q}_{729}^{(2l)} - W_2 \mathbf{q}_{\mathrm{GSM}}^{(L-1)}}{W_2 + 4W_4} \qquad (6)$$

The weighing factors are determined experimentally by using speech data and a perceptual measure as PSQM [9].

For conversion in the inverse sense, from GSM AMR to G.729(A), a linear interpolation of the GSM AMR parameters gives a parameter vector rate twice as high, as needed by the G.729(A) coder

$$\mathbf{q}_{729}^{(2l-1)} = 0.5\mathbf{q}_{\mathrm{GSM}}^{(l-1)} + 0.5\mathbf{q}_{\mathrm{GSM}}^{(l)} \qquad (7)$$
$$\mathbf{q}_{729}^{(2l)} = \mathbf{q}_{\mathrm{GSM}}^{(l)}$$

IV. Speech Data Preparation and Performance Measures

The speech signals used in simulation are from the "Telephone Network Acoustic-Phonetic Continuous Speech Corpus"- NTIMIT [10]. The corpus use a sample frequency of 16 kHz. In simulations, the signals were filtered and decimated to 8 kHz of sample frequency by the following low-pass FIR filter designed in MATLAB:

```
f = [0 0.425 0.5 1];
m = [1 1 0 0];
b = remez(96,f,m);
```

This filter has 97 taps, a cutoff frequency of about 3400 Hz and offers 64 dB of attenuation at 4000 Hz. The SNR of the original and filtered signals is at minimum 30 dB, showing that actually almost all energy of the signal is inside the frequency band of the telephone channel. The signals were left justified as demanding the coder's schemes, by a gain of 2.

The perceptual subjective quality measure (PSQM) [9] was used in simulations for performance analysis. The PSQM computation uses a sampling frequency of 8 kHz. Time alignment was achieved by using the known coder delay and by a procedure of finding the best PSQM around $\pm M$ samples of this delay. The signals for PSQM computation were also scaled to –26 dBov as specified by [9] and [11]. As the active speech level depends on the scale factor, -26 dBov was achieved after a 6 times iteration.

We have also used an objective measure to analyze the different schemes of LSP transcoding based on Spectral Distortion (SD) [12]:

$$\mathrm{SD}^2 = \frac{100}{f_2 - f_1} \int_{f_1}^{f_2} \left(\log_{10} \frac{P(e^{j2\pi f/f_s})}{\hat{P}(e^{j2\pi f/f_s})}\right)^2 df\,, \qquad (8)$$

where $P(e^{j2\pi f/f_s}) = 1/\left|A(e^{j2\pi f/f_s})\right|^2$ is the power spectrum response of LPC filter of the coder and $\hat{P}(e^{j2\pi f/f_s})$ is the analogous for the regressed or interpolated parameters used in transcoding process. In computation of (8) we have used a discrete version of this equation; the computation of the power spectrum uses a FFT of 512 points. The lower and upper frequencies in (8) are 125 Hz and 3400 Hz, respectively [7].

V. Simulation Results

In the first simulation, we have measured the speech quality of the standard coders (Table I). We have used the test corpus of NTIMIT for the region DR1. It is important to notice that when using signal scaling to -26 dBov, the PSQM measures are slightly higher. Both coders present a similar PSQM around 2.

In the second simulation we performed the standard transcoding process as illustrated in Fig. 3. Observe a decrease in speech quality of about 0.7 PSQM. This quality loss is mainly attributed to the post filtering.

Table III shows the results of G.729A to GSM AMR transcoding using (6). We have conducted three simulations using different values for W_2 and W_4. The best results were obtained for W_2 = 0.2 and W_4 = 0.8. Table IV shows results for comparison with the regression equation proposed in [7]. We observe that both equations give very similar results, but the proposed one is less complex.

Finally Table V and Table VI show the spectral distortion measure (8) in the transcoding process from G.729A to GSM AMR and vice versa. The 3rd and 4th columns show the percentage of frame outliers in the range 2-4 dB and >4 dB, respectively. Observe that in the first case the spectral distortions are better, due to the fact that the first uses a regression scheme and the latter an interpolation scheme.

TABLE I
PSQM ANALYSIS OF STANDARD CODERS

NTIMIT Data Base	PSQM (Delay=40, tolerance=20)	
	G.729A	GSM AMR
DR1/FAKS0	2,134935	2,109011
DR1/FDAC1	1,919102	1,836625
DR1/FELC0	2,485503	2,446796
DR1/FJEM0	2,021465	1,934423
DR1/MDAB0	2,242581	2,159264
DR1/MJSW0	1,890444	1,862232
DR1/MREB0	1,849667	1,799423
DR1/MRJO0	2,193720	2,136625
DR1/MSJS1	1,966741	1,921458
DR1/MSTK0	2,049034	1,999720
DR1/MWBT0	1,918980	1,892968
Average PSQM	2,061107	2,008959
Standard Deviation	0,189974	0,190391

TABLE II
PSQM ANALYSIS OF STANDARD TRANSCODING

NTIMIT Data Base	PSQM (Delay=40, tolerance=20)	
	G.729A	GSM AMR
DR1/FAKS0	2,834827	2,819268
DR1/FDAC1	2,612706	2,648919
DR1/FELC0	3,201983	3,261654
DR1/FJEM0	2,752629	2,681147
DR1/MDAB0	3,002980	3,003896
DR1/MJSW0	2,605504	2,658813
DR1/MREB0	2,639069	2,658673
DR1/MRJO0	2,978773	2,975842
DR1/MSJS1	2,679824	2,715861
DR1/MSTK0	2,049034	1,999720
DR1/MWBT0	2,696155	2,711384
Average PSQM	2,732135	2,739562
Standard Deviation	0,295893	0,312738

TABLE III
G.729A TO GSM AMR TRANSCODING
PSQM ANALYSIS BY W_1 AND W_2 VARIATION

Data Base DR1/FAKS0	W_2=0,25 W_4=0,75	W_2=0,20 W_4=0,80	W_2=0,15 W_4=0,85
A1	2,234734	2,240010	2,236466
SA2	2,236512	2,236652	2,242796
SI1573	2,400911	2,380989	2,378077
SI2203	2,196746	2,202902	2,198439
SI943	2,208013	2,213118	2,217973
SX133	2,429710	2,421051	2,424245
SX223	2,261894	2,244872	2,244714
SX313	2,120217	2,107216	2,113731
SX403	2,458290	2,447484	2,435931
SX43	2,314404	2,325537	2,337561
Average PSQM	2,286143	2,281983	2.282993
Standard deviation	0,111324	0,108096	0,105742

TABLE IV
G.729A TO GSM AMR TRANSCODING
COMPARISON TO A PREVIOUS METHOD

Data Base	PSQM (delay=40, tolerance=20)	
	Proposed	Kang et al. [7]
DR1/FAKS0	2,281981	2,284423
DR1/FDAC1	1,980144	1,985937
DR1/FELC0	2,583724	2,582135
DR1/FJEM0	2,105820	2,106139
DR1/MDAB0	2,334784	2,336502
DR1/MJSW0	1,996894	1,991878
DR1/MREB0	1,977408	1,971986
DR1/MRJO0	2,282009	2,286308
DR1/MSJS1	2,089671	2,083093
DR1/MSTK0	2,140768	2,140920
DR1/MWBT0	2,028866	2,028856
Average PSQM	2,163824	2,163471
Standard deviation	0,189060	0,189985

TABLE V
SPECTRAL DISTORTION FROM G.729A TO GSM AMR TRANSCODING

Data Base DR1/FAKS0	Spectral Distortion	Outliers 2-4 dB	Outliers > 4 dB
SA1	1,152454	3,03	0,00
SA2	1,073254	1,66	0,00
SI1573	1,152830	3,23	0,00
SI2203	1,089936	1,71	0,00
SI943	1,064959	3,74	0,00
SX133	1,112824	2,42	0,61
SX223	1,108215	3,90	0,00
SX313	1,184299	3,98	0,00
SX403	1,168900	3,59	0,00
SX43	1,182962	6,56	0,00
Average/Total	1,129063	3,27	0,06

TABLE VI
SPECTRAL DISTORTION FROM GSM AMR TO G.729A TRANSCODING

Data Base DR1/FAKS0	Spectral Distortion	Outliers 2-4 dB	Outliers > 4 dB
SA1	1,432601	9,34	1,52
SA2	1,390052	7,73	1,66
SI1573	1,466650	10,89	1,61
SI2203	1,391907	10,86	1,43
SI943	1,388911	9,89	1,87
SX133	1,423600	10,00	2,12
SX223	1,500180	12,01	2,92
SX313	1,457533	9,66	2,56
SX403	1,537326	16,17	2,69
SX43	1,421159	10,25	2,05
Average/Total	1,440992	10,63	2,00

VI. CONCLUSION

In this paper we have presented a scheme for transcoding the Linear Spectrum Pair (LSP) parameters from G.729A to GSM AMR, and vice versa. The proposed method is simpler than a previous one proposed for the G.729 and IS-641 [7]. The results are similar, but the proposed one is less complex. The simulation results showed a gain of about 0.45 in the PSQM measure when conducting the transcoding in the bitstream domain instead of in the speech domain. The spectral distortion measure is better in the conversion from the G.729A to GSM AMR than in the reverse sense. The proposed method is part of a complete transcoding algorithm under development.

ACKNOWLEDGMENTS

This work has been fully supported by Ericsson Telecomunicações S.A. in the R&D project - record No. 29-P-17007-2000.

REFERENCES

[1] T. Honkanen, J. Vainio, K. Järvinen, P. Haavisto, R. Salami, C. Laflamme and J.-P. Adoul, "Enhanced full rate speech codec for IS-136 digital cellular system," *Proc. ICASSP,* Münich, German, pp. 731-734, Apr. 1997.

[2] M.C. Reccione, "The enhanced variable rate coder: toll quality speech for CDMA," *International Journal of Speech Technology*, vol. 2. no. 4, pp. 305-315, May 1999.

[3] ETSI, GSM 06.90, *Digital cellular telecommunications system (Phase 2+); Adptive Multi-Rate (AMR) speech transcoding*, 1998.

[4] ITU-T Recommendation G.723.1, *Dual rate speech coder for multimedia telecommunication transmitting at 5.3 and 6.3 kbit/s*, Mar. 1996.

[5] ITU-T Recommendation G.729, *Coding of speech at 8 kbit/s using conjugate-structure algebraic-code-excited linear-prediction (CS-ACELP),* Mar. 1996.

[6] ITU-T Recommendation G.729A, *Annex A: Reduced complexity 8 kbit/s CS-ACELP speech codec,* Mar. 1996.

[7] H-G. Kang, H.K. Kim and R.V. Cox, "Improving the transcoding capability of speech coders," *IEEE Trans. on Multimidia,* submitted for publication.

[8] A.M. Kondoz, *Digital Speech, Coding for Low Bit Rate Communications Systems,* West Sussex, UK, 1994.

[9] ITU-T Recommendation P.861, *Objective quality measurement of telephone-band (300-3400 Hz) speech codecs*, Feb. 1998.

[10] C. Jankowski, A. Kalyanswamy, S. Basson and J. Spitz, "NTIMIT: a phonetically balanced, continuous speech, telephone bandwidth speech database," *Proc. ICASSP*, Albuquerque, Apr. 1990.

[11] ITU-T Recommendation P.56, *Objective measurement of active speech level*, Feb. 1993.

[12] K.K. Paliwal and B.S. Atal, "Efficient vector quantization of LPC parameters at 24 bits/frame," *IEEE Trans. Speech Audio Processing,* vo. 1, no. 1, pp. 3-14, Jan. 1993.