SOURCE CODING ACTIVITIES AT THOMSON-CSF COMMUNICATIONS

Cédric DEMEURE, Philippe GOURNAY, and Frédéric CHARTIER

Thomson-CSF Communications
66, rue du fossé blanc – B.P. 156
92231 Gennevilliers CEDEX – France
e_mail: cedric.demeure@tcc.thomson-csf.com

ABSTRACT

This paper presents a summary of the activities in research and development in the area of source coding (voice, audio, image and video) at Thomson-CSF Communications (TCC) and the associated applications and products. The main areas of research at TCC on these subjects are in particular very low bit rate speech coding and watermarking.
Keywords: speech, vocoder, LPC, CELP, HSX, image, video, MPEG, JPEG, intelligibility, watermarking.

1. INTRODUCTION

Thomson-CSF Communications (TCC) has long been active in source coding activities first on the low bit rate compression of the speech signal for secure transmission on radio links then on audio coding for broadcasting and finally coming from security applications on watermarking of image and video streams, and so on JPEG 2000 image coding.

These activities were first focused only on military applications. But the quality of low bit rate coding of speech for example has reached such a good quality that the door was opened to commercial applications. Since the last two or three years, several products were developed in association with other companies such as paging receivers with speech message synthesis, answering machines with augmented recording time and satellite telephony. Section 2 contains the describtion of these activities and the associated technology.

TCC large commitment to security issues was the basis of our activity in image and video coding. Watermarking of image and video sources has many applications in the commercial world, starting from copyright assessment, channel monitoring in TV market, author authentication, property and integrity verification, to name a few.

2. SPEECH CODING

2.1 Introduction

Starting from the original sampling of the telephone line bandwidth, several algorithms have been proposed to reduce the bit rate while at the same time maintaining a satisfying subjective quality.

Typically three classes of bit rates are considered:

- high bit rates, higher than 16 kbps, corresponding to coding algorithms taking only into account the temporal variation of the input signal (or waveform), independently of the fact that speech is processed. In this class are found PCM and Delta coders.
- medium bit rates, from 4 to 16 kbps, corresponding to hybrid techniques using wave form methods and taking into account various properties of speech or of the acoustics perception of the human ear. The main component of this class is the CELP coder [1].
- low and very bit rates, from some tenths of bits per second to 4 kbps, correspond to vocoders (VOice CODER) highly specific to speech coding.

A comparison of quality and bit rate reduction is show in the following figure.

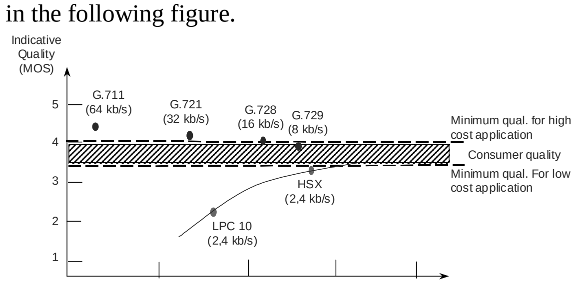

Figure 1: *Description of speech coder quality evolution*

This figure shows:

- first the fact that every 5 to 10 years the bit rate of speech coders was divided by two, keeping the quality constant (G.711 at 64 kb/s in 1972, G.721 at 32 kb/s in 1984, G.728 at 16 kb/s in 1992 and G.729 at 8 kb/s in 1996).

- then the quality evolution obtained at the fixed bit rate of 2.4 kb/s. A significant improvement has been achieved from the LPC 10 coder of 1984 to coders such as the HSX described below.

TCC has from the beginning (around 20 years ago) always focused research and development activities on this last type of algorithm in order to achieve very low bit rate coding. The application for these vocoders comes from the ability to better protect (cypher) the content of a digital speech stream than using the original analog signal. Thus a natural application was in the security and military domains. The need for low bit rate being imposed by several digital radio-communications exhibiting low capacity, like HF and VHF modems.

2.2 Traditional low bit rate vocoders

2.2.1 LPC coders

The classical, binary, voiced/unvoiced LPC10 vocoder [2,3] has long dominated the field of low bit rate speech coding. It offered a service at 2400 bps and was standardized by NATO (STANAG 4198). Though quite intelligible, the output speech was of poor quality; its usage was thus limited to very specific applications, mainly professional and military. The architecture of the LPC10 coder is based on a simple exciter (impulse train or white noise) and a linear predictive filter (hence the name).

A version using an even lower bit rate (800 bps) was developed by TCC and after a competitive comparison with other solutions was elected in 1993 to be a NATO standard (STANAG 4479). The main improvement over the LPC10 is the reduction by a factor 3 of the bit-rate obtained through a concatenation of three successive frames and an adapted quantisation process (in particular the shape of parameter evolution is quantised between known values). The obtained quality was similar to the original LPC with 3 times less bits. The difference between the two coders is given in table 1.

		2400 bps LPC10	**800 bps**
Framelength		22.5 ms	22.5 ms
Superframe length		none	67.5 ms
Energy		5 bits	10 bits
Pitch / Voicing		7 bits	9 bits
Filter code	**Chosen scheme**		3 bits
	Filter encoding	41 bits	32 bits
Synchronisation		1 bit	none
Total		54 bits	54 bits

Table 1: *LPC10 2400 bps and 800 bps parameters*

This NATO 800 bits/s standard is still now the lowest bit rate existing standard.

2.2.2 CELP coders

Similarly for higher rates (4800 bps) TCC teamed with the University of Sherbrooke and was able to come up with an ACELP coder at 4500 bps that was proposed for standardization at ETSI in the Tetra system. After a world-wide competition the TCC solution was chosen in 1994. TCC is now the unique provider of such technology to Tetra product manufacturers, such as Motorola, Nokia, Simoco/Philips, Frequentis, … Those three coders (800, 2400, and 4800 bps) were incorporated in a single ASIC based upon a DSP core from ST Microelectronics in 1994. This ASIC is currently in use in tactical VHF radio sets. These past few years, there were many improvements in the field of low bit rate speech coding. With the introduction of new models, such as MBE [6], PWI [7], and MELP [8], the quality produced at 2400 bits/s became compatible with a growing number of commercial applications.

However, there is currently a need for an even lower bit rate, typically around 1000 bits/s (see for example [9]). Among the possible applications for such a very low bit rate speech coder are low-cost satellite communications, Internet telephony and vocal paging. The low bit rate HSX speech coder was designed by TCC in order to fill in this gap, while at the same time offering better quality at 2400 bps and intermediate 1200 bps.

2.3 State of the art vocoder: HSX technology

Continuing its collaboration with the University of Sherbrooke, TCC has recently developed a new version of its low bit rate coder based on a the HSX technology. The HSX speech coder is a linear predictive vocoder that uses a simple mixed excitation model, in which the periodic impulse train covers the lower frequencies and the noise is located in the upper frequencies of the input of the LPC synthesis filter. The cutoff frequency between the lower and upper bands is time variable, and the gain between the two excitation signals is adjusted so that the resulting mixed excitation has a flat spectrum. The first part of this section describes the analysis and synthesis part of the HSX speech coder. The second part describes the quantization process for its operation at 1200 bits/s.

2.3.1 Analysis part of the HSX coder

A block diagram of the analysis part of the HSX speech coder is given in Figure 2. The digital speech signal is high-pass filtered and segmented into speech frames that are 22.5 ms long. Two 10th-order LPC analyses are carried out on each frame (one at the middle, the other at the end of the frame). The semi-whitened residual signal (γ is typically equal to 0.80) is then filtered into 4 subbands. The estimate of the pitch is obtained coarsely in the first subband and refined using the other subbands. The cutoff frequency between the lower, voiced band and the upper, unvoiced band is obtained using the 4 subbands. The final value of the pitch and voicing cutoff frequency is given with one frame delay by a robust pitch-and-voicing tracker. Finally, the energy of 4 equal-length subframes is computed pitch-synchronously.

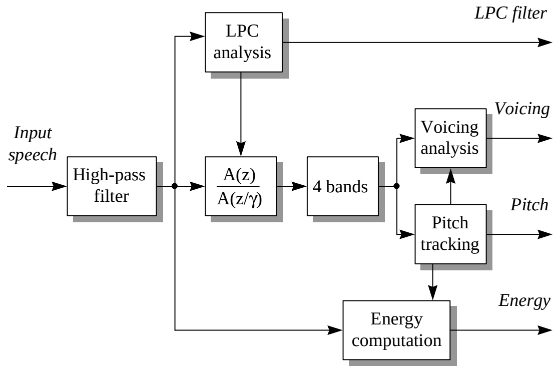

Figure 1 : *Block diagram of the HSX analysis part*

2.3.2 Synthesis part of the HSX coder

A block diagram of the synthesis part of the HSX speech coder is given in Figure 3. The mixed excitation with flat spectrum is the sum of a low-pass, periodic signal and a high-pass, random signal. The harmonic signal with adequate spectral envelope is obtained by passing a periodic impulse train through a bank of low-pass filters. The random signal with associated spectral envelope is obtained by an inverse Fourier transform and an overlap-and-add technique. The LPC synthesis filter is interpolated 4 times per frame. An adaptive spectral postfilter based on the LPC parameters helps providing a better reconstitution of nasal sounds. Finally, an automatic gain control procedure ensures that the energy of the output signal is close to the energy of the input signal.

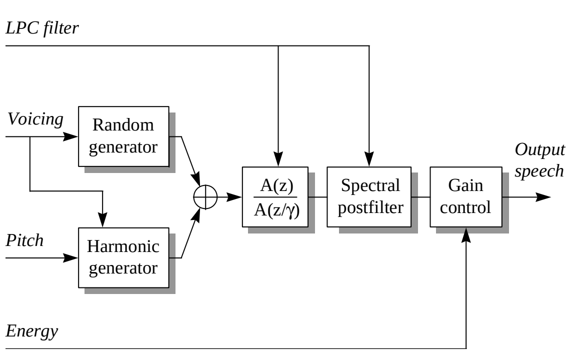

Figure 2: *Block diagram of the HSX synthesis part*

2.3.3 Some HSX evaluation results

We conducted a simplified rhyme test in order to assess the intelligibility performance of the 1200 bits/s version of the HSX speech coder. This test was designed in 1983 by the “Institut de Phonétique de l'Université d'Aix-Marseille” (IPAM) under contract with TCC. It focuses on only 2 elementary phonetic attributes of the initial consonant of the words: graveness and compactness. The average intelligibility score for French is estimated from the intelligibility scores for these two attributes using a linear regression formula. This simplified rhyme test was shown by the IPAM to be statistically consistent with the classical, comprehensive rhyme test for French.

The test was performed with 8 listeners, 4 different test sequences (one test sequence consisting of 56 words), and 2 coders : the 1200 bits/s HSX speech coder and the 2400 bits/s LPC10-E vocoder (version 52). Two different test sequences processed by different vocoders were presented to each listener. The test results presented in Table 2 show that the intelligibility score for the 1200 bits/s HSX speech coder is in average about 2.5 points higher than the one of the 2400 bits/s LPC10-E vocoder. The large standard deviation value in the case of the 2400 bits/s LPC10-E vocoder is due to two high (97.14) intelligibility scores. Discarding the two most extreme measures for each of the coders, the average score and standard deviation become respectively 96.55 and 0.65 for the HSX speech coder, 93.01 and 0.78 for the LPC10-E vocoder.

Vocodeur	Score	StdDev
1200 bits/s HSX vocoder	96.51	1.07
2400 bits/s LPC10-E vocoder	94.04	1.91

Table 2 : *Intelligibility scores*

We also informally evaluated the performance of the 1200 bits/s HSX speech coder in various input and channel conditions. It was found to be robust to background noise, and to perform well in the presence of fast speech. The quality of the synthesized speech was found to be very good, with very few annoying artifacts, for a uniform bit error rate up to 1%. The 1200 bits/s HSX speech coder includes an extrapolation-based parameter recovery procedure that can be put into operation when the channel decoder fails to decode and delivers a "bad frame indicator". It is also extremely tolerant to frame losses.

2.3.4 HSX real-time implementation

Starting from the floating point version simulation, a complete fixed point algorithm description was derived using the low-level operators of the C-ETSI (EFR-GSM, G.729, G.723…). Both coding and decoding are then optimized for the target DSP. The use of a complete intermediate fixed-point reference simulation algorithm allows for very simple implementation on other targets. The table below gives the memory and CPU requirements for the fixed point implementation on TI C54x family of processors (16 bits).

1w=16 bits	**Full-Duplex**	**Coder**	**Decoder**
Program Memory	7 kw	4.5 kw	3 kw
Data (ROM)	9.2 kw	8.7 kw	9.2 kw
Data (RAM)	5.5 kw	4 kw	2.5 kw
Complexity	22 MIPS	18 MIPS	4 MIPS

Table 3: *HSX fixed point implementation requirements*

Table 4 lists the requirements for the same vocoder implementation on a floating point DSP processor (TI C3x). These values have been measured on a real time version based upon C language source code compiled and linked using Texas tools. Only a few routines corresponding to most of the CPU requirements were further optimized in assembly code (FIR filters, correlation, and quantifiers).

1w=32 bits	**Full-Duplex**	**Coder**	**Decoder**
Program Memory	17.7 kw	11.7 kw	11.7 kw
Data Memory	6 kw	4 kw	3 kw
Complexity	25 MIPS	19 MIPS	6 MIPS

Table 4: HSX floating point requirements

2.3.5 Applications

The HSX technology has recently being implemented in several commercial applications such as a radio pager based on MobiDarc® (FM-sub digital carrier) developed with the Info Télécom compagny to offer speech message paging, or an all-electronic answering machine (avoid the use of a tape recorder). In this last application, the use of a 1200 bits/s coder instead of a 2400 bits/s coder was decided. It provides acceptable quality for such an application for the mass consumer market. This in return allows to further reducing the necessary RAM capacity for a given recording time or equivalently increasing the recording time by a factor 2.

2.4 Other activities in speech and audio coding

2.4.1 Noise reduction

Such low bit rate coders are rather sensitive to the speech sensor system and also to the amount of surrounding noise. Classical single-sensor noise pre-processing (such as the one used in hand-free telephony systems) is no longer effective when extremely noisy conditions must be endured (like armored vehicle or fighter aircraft). For these reasons, TCC has launched a thesis on this subject in collaboration with the University of Rennes. The noise reduction system used is a modified Wiener filtering, given by the minimization of the mean-square spectral error. Two hypotheses are considered: presence and absence of speech signal.

It was found that coupling the noise pre-processor with the vocoder parameter estimation algorithm was the most efficient way to obtain satisfactory results. A complete paper on this subject is given in [12].

2.4.2 Audio coding

Thomson-CSF is a traditional actor of the broadcasting industry through its affiliate Thomcast that offers power transmitters in TV and radio operators (one of the world leader in this domain). Since 1996, collaboration between TCC and Thomcast has taken place in the Eureka project Nadib, aiming at proposing a standard to replace traditional AM audio broadcasting by a digital system. Such a standard should allow to introduce digital like quality (and no longer a high sensitivity to propagation) in a traditionally poor quality media and also to propose value added service such as PAD, RDS and other data broadcasting services.

A worldwide consortium has been organized named DRM (for Digitale Radio Mondiale) to propose at the ITU level a system description before the end of 2000, and a further ISO standard. The audio coding scheme under investigation to be included in this system is based upon the MPEG IV AAC audio coder and a CELP coder for speech (in competition with a G 729 coder).

TCC has developed a demonstrator for this system using an MPEG II layer 3 coder. Upgrade to the final scheme is planned. This demonstrator is currently used as a reference receiver for extensive testing throughout the world (9 receivers used with 5 different transmitter locations and numerous receiver locations) [13].

2.5 Very low bit rate coding research

To obtain even lower bit rates (200 to 600 bps), it is necessary to change the underlying model and to work on frames of variable length coming from a segmentation process. Segment indexes are then deduced and the synthesis process may then reproduce the original word (or phonem). A symbolic transcription is done either using a dictionnary of variable length elementary units (phonems, transitions, …) leading to phonetic vocoders, or acoustic units obtained automatically using a learning process.

TCC has recently started a collaboration project on this subject with Elan Informatique, Info-Réalité, ESIEE and ENST Paris. The aim is to develop a new coder starting from the thesis work of Jan Cernocký [10] (coder family named ALISP for Automatic, Language Independent Speech Processing) using a phonetic vocoder [11]. The center of this coder is a recognition system that after a segmentation process derives a coding unit (CU) for each segment. These CU are derived according to a model that corresponds to a synthesis unit (SU) and associated samples used for the reconstruction phase. The main issue in such a scheme comes from the importance of the learning process, and a lot of energy is spent to make this phase automatic and to reduce its length. Furthermore, some work is dedicated to make this process incremental: the longer the vocoder is used, the better the quality its gives.

3. IMAGE AND VIDEO CODING

TCC has been a long standing actor in the field of security: complete line of equipment for network security, compatible products for the security of telephone and radiotelephone links, protection of satellite links, secure office terminals (fax, videotex, PC), complete systems for secure applications, etc.

Starting from this activity TCC has started several years ago to look at other commercial applications of security technologies such as on image and video watermarking. Watermarking can be applied on any digital object (image, sound, graphics, and software). Image watermarking consists in burying in an invisible but robust manner identifying information (for example 64 bits) inside the image. This image, fixed or animated can be represented in two different domains:

- its original form (pixels)
- a compressed form (transport stream for example).

Watermarking does not consist in adding information in the image side fields, but in the image itself by modifying the pixels. This insertion may be done in both domains, corresponding to different applications schemes.

Invisibility is a fundamental issue here as watermarking cannot be allowed to deteriorate the image quality. It is also a key issue for security, as if made visible, watermarking may be detected and so removed.

Robustness is defined by the watermarking capacity to survive to various manipulations on the image, such as compression in different standards (JPEG, MPEG, wavelets, ..), to the transformation in analog format, as well as a lot of digital manipulations or transformations. Resistance to attacks is characterized by the capacity to be able to read the original watermark after a series of transformations, including natural degradation of the image when transmitted for example on a given imperfect channel (broadcasting for example). Other classical attacks are noise addition (gaussian, impulsive), filtering, cropping, zooming, geometrical modification, resampling and applying several watermarking, to name a few.

In practice, the process may be seen as a transmission channel with extremely high error probability (0.4 typical), which implies a very protected scheme with extremely low false alarm rate (false detection) and very good performance (recovery probability).

Research in robustness of watermarking at TCC is currently going on in collaboration with ENST in Paris [14].

3.1 Video watermarking

In this context, TCC has launched and has been co-ordinating, from 1995 onwards, the ACTS TALISMAN project developing hardware and software systems for protecting still pictures and video sequences and complying with MPEG-2 compression. Through its collaboration with research centers such as ENST, UCL, INPG and others, TCC has issued the first operational system for watermarking motion pictures complying with MPEG-2 compression and tested in real broadcast conditions. The system has the following properties:

- total invisibility of the watermarks on the original high quality work
- high robustness against MPEG-2 compression, broadcast errors, etc
- survivability to the whole broadcast processing chain.

These works, completed in the ACTS OCTALIS project, have been demonstrated at IBC'98, NAB'99 and Montreux'99. The resulting developments are now in collaboration with Thomcast. Since a number of years, TCC has got a huge experience in watermarking for video high bit rate and is now positioned as one of the acknowledged watermarking technology supplier for MPEG-2. The technology has been assessed during several showcases on the EBU network, particularly during the Football Cup in 1998.

The watermarking demonstrator built by TCC is based on the use of a TI C80 DSP. The input and output interfaces are compliant with the CCIR 656 format with a 720 pixels by 576 lines definition at 25 images per second. A new version based on the DSP C62 is under development for use in comparative testing programmed for June 2000 for evaluation tests at the EBU.

3.2 On going watermarking work

TCC is currently involved in two IST programs, Certimark and Migrator 2000.

Certimark is a project aiming at designing, developing and publishing a complete benchmark suite for watermarking technologies. It also aims at making this suite a reference tool both for technology suppliers, setting up a certification process for watermarking algorithms, to definitely bring the European technological offer in a dominant position for protecting pictures in the context of E-Commerce over the Internet. Consequently, Certimark is going to concentrate research on pending key issues in watermarking for protection of still pictures and low bit rate video over the Internet.

Migrator 2000 aims at defining and developing a new generation of content creation and personalisation tools for still pictures based on the emerging new JPEG 2000 standard. Much more than just another compression algorithm, JPEG 2000 is a truly innovative standard, both with its new functionalities and with the concept of "open standard". Several new features are included such as a new compression algorithm based on wavelets, scalability in quality and in resolution, error resilience for wireless transmission, region of interest selection at the user's and at the creator's side, new Metadata policy and classification, content identification with the notion of License Plate and a flexible file format to cover wide range of applications.

TCC interest in JPEG 2000 lies in its ability to be used in several environments that covers applications such as surveillance, region of interest image representation and retrieval, robust transmission with incremental definitions, etc. Current work includes quantifier simplification for real time implementation.

4. CONCLUSION

To conclude this paper, we can stress the amazing evolution in speech, audio, image and video compression in the last 10 years, that TCC tries to follow and in some area precede in order to propose interesting new technology advances and to develop new applications and offer business opportunities.

Such niche research activity includes in particular very low speech compression and watermarking.

5. REFERENCES

[1] M.R. Schroeder, B. Atal, *Code-Excited Linear Prediction(CELP): High Quality Speech at Very Low Bit Rates*. Proc. IEEE ICASSP-85, pp. 937-940, Tampa, 1985.

[2] *Parameters and coding characteristics that must be common to assure interoperability of 2400 bps linear predictive encoded speech*, NATO Standard STANAG-4198-Ed1. February 13, 1984.

[3] C. Laflamme, R. Salami, R. Matmti and J-P. Adoul, *Harmonic-Stochastic Excitation (HSX) Speech Coding below 4kbits/s*, IEEE Int. Conf. on ASSP, pp. 204-207, 1996.

[4] G. Guilmin, R. Le Bouquin-Jeannes and P. Gournay, *Study of the Influence of Noise Pre-Processing on the Performance of a Low Bit Rate Parametric Speech Coder*, EUROSPEECH, Budapest, Hungary, pp. 2367-2370, Sept. 1999.

[5] Mouy, B., de La Noue, P., and Goudezeune, G. *NATO STANAG 4479: A standard for an 800 bps vocoder and channel coding in HF-ECCM system*, IEEE Int. Conf. on ASSP, Detroit, pp. 480-483, May 1995.

[6] Griffin, D.W., and Lim, J.S. *Multiband Excitation Vocoders*, IEEE Trans. on Acoustics, Speech, and Signal Processing, vol. 36, no. 8, pp. 1223-1235, 1988.

[7] Kleijn, W.B, and Haagen, J. *Waveform Interpolation for Coding and Synthesis*, in Speech Coding and Synthesis, edited by W.B. Kleijn and K.K. Paliwal, Elsevier 1995.

[8] Supplee, L.M., Cohn R.P., Collura J.S., and McCree, A.V. *MELP : The new federal standard at 2400 bits/s*, IEEE Int. Conf. on ASSP, Munich, pp. 1591-1594, April 1997.

[9] Shoham, Y. *Very low complexity interpolative speech coding at 1.2 to 2.4 Kbps*, IEEE Int. Conf. on ASSP, Munich, pp. 1599-1602, April 1997.

[10] J. Cernocký, *Speech Processing Using Automatically Derived Segmental Units: Applications to Very Low Rate Coding and Speaker Verification*, PhD thesis, Université Paris XI Orsay, 1998.

[11] G. Chollet, J. Cernocký, A. Constantinescu, S. Deligne, F. Bimbot, *Computational models of speech pattern processing, chapter Towards ALISP: a proposal for Automatic Language Independent Speech Processing*, NATO ASI Series. Springer Verlag, pp. 375-388, 1999.

[12] G. Guilmin, R. Le Bouquin, P. Gournay, *on the optimisation of noise pre-processing for a low bit rate parametric speech coder*, Workshop COST 254, Bordeaux, France, March 2000.

[13] C.J. Demeure, P.A. Laurent, B. Le Floch, D.Lacroix, *Optimised COFDM waveform and coding schemes for digital terrestrial radio Broadcasting*, ISCTA conf., July 1999.

[14] S. Baudry, H. Maitre and P. Nguyen, *Channel coding in video watermarking: use of soft decoding to improve the watermark retrieval*, submitted to ICIP 2000.