NEW NATO STANAG NARROW BAND VOICE CODER AT 600 BITS/S

Gwénaël Guilmin, François Capman, Bertrand Ravera, Frédéric Chartier

THALES Communications France
Signal Processing and Multimedia Department
160 boulevard de Valmy - BP82, 92704 COLOMBES CEDEX – France
gwenael.guilmin,francois.capman,bertrand.ravera,frederic.chartier@fr.thalesgroup.com

ABSTRACT

This paper describes a new very low bit rate speech coder at 600 bits/s based on the MELPe algorithm [3]. This coder associated with the other bit rates at 2400 bits/s and 1200 bits/s constitute the new NATO standard STANAG 4591. The STANAG 4591 noise pre-processor is also part of the 600 bits/s coder. At 600 bits/s, four consecutive frames are grouped into a super-frame and jointly quantized to obtain high coding efficiency. The inter-frame redundancy is exploited with distinct quantization schemes for different unvoiced/voiced frame combinations in the super-frame. The paper gives a brief description of the algorithm, presents the quantization process operating at 600 bits/s, and gives some evaluation results together with CPU and memory requirements for its implementation on DSP. We also propose a FEC (Forward Error Corrector) scheme adapted to the new HF NATO communications standard STANAG 4444 [6].

1. INTRODUCTION

In 1997, after the selection of new US MIL-STD 3005 speech coder at 2400bits/s [4], the NATO C3 Agency launched a competition to select a speech coding algorithm addressing narrow band strategic, tactical, SATCOMs and internetworking applications. The goal was firstly to provide a single, high performance speech coding algorithm for seamless interoperability across and between these domains, and secondly to replace the former well-known LPC-10 algorithm at 2400 bits/s and 800bits/s, the STANAG 4198 [2] and the STANAG 4479 [1].

During this competition the MELPe algorithm was selected for the 2400 bit/sec and the 1200 bits/s bit rates [5]. More recently, tactical and strategic end-users expressed their interest for lower bit rate than 1200 bits/s. The NATO Ad Hoc Working Group proposed new competition to selected a narrow band voice coder at 600 bits/s and based on MELPe algorithm. This paper presents the candidate which has been selected to be the new NATO standard at 600 bits/s.

2. CODER OVERVIEW

The coder reuses the same analysis modules as the original 2400 bits/s MELPe coder except the changes described below. A block or "super-frame" consisting of four consecutive frames is formed to quantize the transmitted parameters more efficiently in the 600 bits/s mode. The frame size of the 600 bits/s coder remains the same as for the original MELPe coder: 22.5 ms. In order to avoid big pitch errors, the length of look-ahead is similar to the one used in the 1200 bits/s coder. A pitch smoother was introduced in the 600 bits/s coder. The algorithmic delay for the coder in the 600 bits/s mode is 126.25 ms, and the corresponding buffer structure is shown in Figure 2.

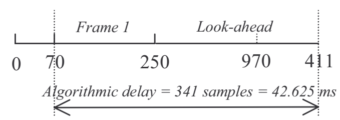

Figure 1 The buffer structure of standard 2400 bits/s MELPe coder.

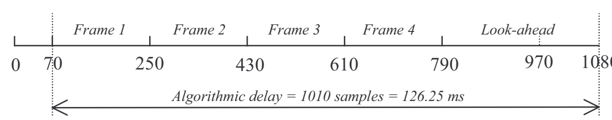

Figure 2 The buffer structure of the 600 bits/s coder.

The transmitted parameters for the 600 bit/sec coder are the same as for the 2400 bit/s MELPe coder. The 600 bit/s coder has several modes that use different quantization schemes. Mode selection is done according to the UV patterns of the super-frame.

2.1. Encoding Mode Determination

The quantized band-pass voicing pattern is used to determine both the encoding and decoding mode of the 600 bits/s coder. This mode depends on the global voicing decision taken on a frame basis. The Table 1 gives the selected mode according to each sequence of voicing decisions for the processing of a super-frame.

Encoding/Decoding mode	Global Voicing Decision for one super-frame
1st mode, purely unvoiced	{U,U,U,U}
2nd mode, single voiced frame	{U,U,U,V},{U,U,V,U}, {U,V,U,U},{V,U,U,U}
3rd mode, more than two voicing transitions	{U,V,U,V},{U,V,V,U}, {V,U,U,V},{V,U,V,U}
4th mode, single voicing transition	{V,V,U,U},{U,U,V,V}
5th mode, single unvoiced frame	{V,V,U,V},{V,V,V,U}, {U,V,V,V},{V,U,V,V}
6th mode, purely voiced	{V,V,V,V}

Table 1. Encoding and Decoding modes of the 600 bit/s.

2.2. Bit Allocation

The bit allocation schemes for 2400 bits/s is shown in Table 2 for voiced and unvoiced frame.

Parameters	Bits for quantization of one frame (180 samples = 22.5 ms)	
	2400 bits/s Voiced Frame	2400 bits/s Unvoiced Frame
Pitch (Voicing decisions)	7	7
Parity	0	0
LSFs	25	25
Gains	8	8
Band-Pass Voicing	4	0
Fourier Magnitudes	8	0
Jitter	1	0
Synchronization	1	1
Error Protection	0	13
Total	**54**	**54**

Table 2. Bit Allocation for the 2400 bits/s speech coder.

The bit allocation schemes for the 600 bits/s modes are shown in Table 3. Jitter, synchronization and error protection are not any more transmitted in the 54 bits.

Parameters	Bits for quantization of one super-frame (720 samples = 90 ms)					
	Mode 1	Mode 2	Mode 3	Mode 4	Mode 5	Mode 6
Pitch (Voicing decisions)	0	6	8	8	8	8
Parity	0	0	0	0	0	0
LSFs	36	30	30	30	30	32
Gains	13	13	11	11	11	9
Band-Pass Voicing	5	5	5	5	5	5
Total	**54**	**54**	**54**	**54**	**54**	**54**

Table 3. Bit Allocation for the 600 bits/s speech coder.

3. QUANTIZATION SCHEMES

The basic structure of the encoder remains the same as for the 2400 bits/s MELPe coder. The 1200 bits/s pitch smoother module is adapted to the 600 bit/s mode, and Band-Pass voicing constraints are added. The coder extracts the feature parameters of four successive frames using the same algorithm as in the MELPe standard. Then the pitch parameter is smoothed to take benefit of the longer block size. Constraints are then applied to the Band-Pass voicing pattern.

3.1. Pitch Smoother

The pitch smoother takes the pitch estimates from the MELPe analysis module as the starting point. A set of new parameters is computed every 11.25 ms, i.e., each half frame. The eleven computation positions in the current super-frame are illustrated in Figure 3.

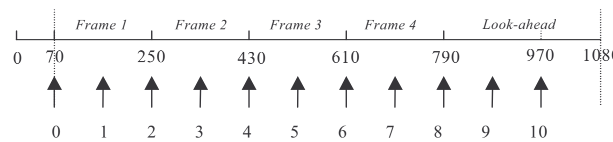

Figure 3. The computation positions of parameters for pitch smoother.

As in the 1200 bits/s model, frame classification into onset and offset frames is introduced in order to guide the pitch and class smoothing process. The parameters introduced for onset/offset classification include: energy, zero-crossing rate, peakiness measurement in speech domain, maximum correlation coefficient in pitch search range.

3.2. Parameter quantization

The bit-allocation is shown in Table 3. New quantization schemes are designed to take advantage of the long block size by using Vector Quantization and classification based on the statistic properties of voiced and unvoiced speech. The Fourier magnitudes are not used in the 600 bits/s coder, and no jitter information is transmitted.

3.2.1. Band-Pass Voicing Quantization

The U/V decisions are vector quantized using 5 bits. The corresponding voicing pattern is used to determine the encoding/decoding mode of the 600 bit/sec coder. A weighted Euclidean distance is used to find the optimal voicing pattern, and the corresponding weight vector is applied to the sub-band voicing information {1.0, 1.0, 0.7, 0.4, 0.1}. Additional constraints are applied as pre-processing of the band-pass voicing pattern: for each frame the voicing vector is replaced by the closest vector among the following set: {(0,0,0,0,0), (1,0,0,0,0), (1,1,1,0,0), (1,1,1,1,1)}.

3.2.1. Pitch Quantization Scheme

A specific quantization scheme is applied to the pitch parameter according to the encoding mode. In the first mode, the super-frame is fully unvoiced and then no pitch information is transmitted. In the second mode, a single frame is voiced, and the corresponding pitch lag is encoded

in the log-domain using a 6 bits scalar quantizer. In the remaining modes (3, 4, 5, and 6), one single pitch value is transmitted using 5 bits, together with the corresponding location within the super-frame using 2 bits, and the selected trajectory type using 1 bit.
The optimal pitch trajectory is searched through a grid bounded by the maximum and minimum of the input pitch pattern. The selection process is performed in the frequency domain, and then corresponding pitch lags are first converted in Hz. Four different types of trajectories are considered: constant path, direct linear path, linear path followed by a constant path, a constant path followed by a linear path. These trajectories are illustrated in the Figure 4.

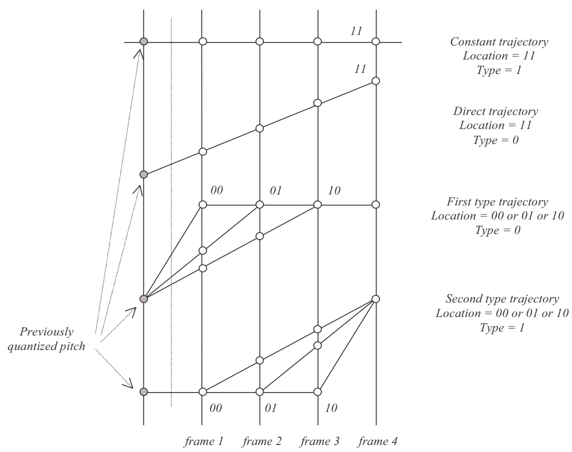

Figure 4. Quantized pitch trajectories in frequency domain

3.2.3. LSF Quantization

For each super-frame, two sub-vectors are constructed and quantized: the first one is the concatenation of the two first LSF vectors, and the second one the concatenation of the two last LSF vectors. As in the 2400 bits/s mode the quantization process is based on Multi-Stage Vector Quantization (MSVQ) (Figure 5). The same weighting coefficients are used for the distance computation.

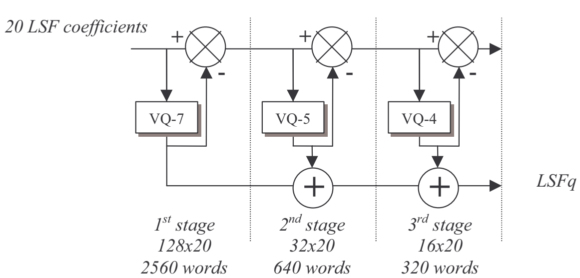

Figure 5. Example of LSF quantization using MSVQ(7,5,4).

The bit allocation and the MSVQ configuration are optimized according to the encoding mode. Different codebooks are used according to the voicing information. To this end, six voicing classes are determined as described in the following table. If the two frames of one sub-vector are voiced, the band-pass voicing pattern is used to define three sub-classes. In the first sub-class (VV type 1) at least one of the two frames is declared unvoiced for the four highest sub-bands. In the second sub-class (VV type 2) at least one of the two frames is declared unvoiced in the highest sub-band. And in the third sub-class (VV type 3) both frames are fully voiced.

3.2.4. Gain Quantization

In the 600 bit/sec coder, two gain parameters are calculated per frame, i.e., 8 gains are obtained for each super-frame. The 8 gain parameters are vector-quantized using either a 9 bits vector quantizer, a (6,5) MSVQ quantizer or a (7,6) MSVQ quantizer, with a MSE criterion defined in the logarithmic domain (Figure 6).

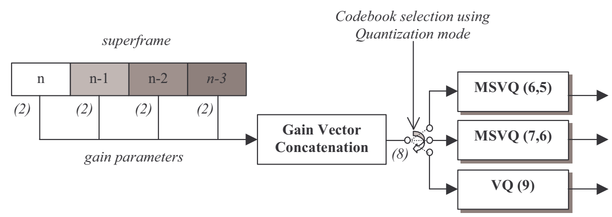

Figure 6. Quantization of gain values.

4. DECODER OVERVIEW

4.1. Parameters Decoding

The 5 bits allocated to the quantization of voicing patterns are first decoded, and the corresponding decoding mode is then determined. According to this decoding mode, the pitch trajectory is obtained either by (i) setting an unvoiced default value if the super-frame is declared unvoiced, (ii) decoding a single pitch value if a single frame is declared voiced, its location being given by the voicing pattern, (iii) linear interpolation using the trajectory type and frame location indicators.
Similarly to the encoding phase, the LSF coefficients are retrieved using the selected MSVQ codebook according to the decoding mode. Then, the restored LSF coefficients are checked for ascending order and minimum separation.
Similarly to the encoding phase, the gain coefficients are retrieved using the selected VQ or MSVQ codebook according to the decoding mode.

4.2. MELPe Synthesis

The basic structure of the decoder is the same as in the 2400 bits/s MELPe coder. The synthesis function is called for each of the four elementary frames of the super-frame. Since no aperiodic flag is transmitted, a jitter parameter is regenerated using the voicing information only. If a frame is unvoiced, the related jitter is set to 25 %, and if voiced, it is set to 0 %.

5. PERFORMANCE RESULTS

Several subjective tests were run by US lab with noise pre-processor in different noisy conditions to select the more suitable low bit speech coder. Results, given in the Table 4, show that performance of the STANAG 4591 600 bits/s speech coder is below the 1200 bits/s speech coder performance, but it outperforms the former STANAG 4198 LPC10 at 2400 bits/s.

Condition / Coder	Quiet	Office	Car
ST4591 at 600 bits/s	88.24	86.59	86.31
ST4591 at 1200 bits/s	91.19	90.06	87.22
ST4198 LPC10-e at 2400 bits/s	86.39	83.51	80.84

Table 4. Subjective intelligibility test results for some acoustic conditions.

6. ST4444 FEC PROPOSAL

The proposed FEC scheme related to the 600 bits/s MELP coder is adapted to a specific Wave Form (WF), of the HF NATO communications standard STANAG 4444, dedicated to voice traffic (EWWF4) which provides a 2400 bits/s user data rate. We already integrated this particular WF with NATO standard vocoders such as ST4479 (800 bits/s) and ST4591 (1200 bits/s).

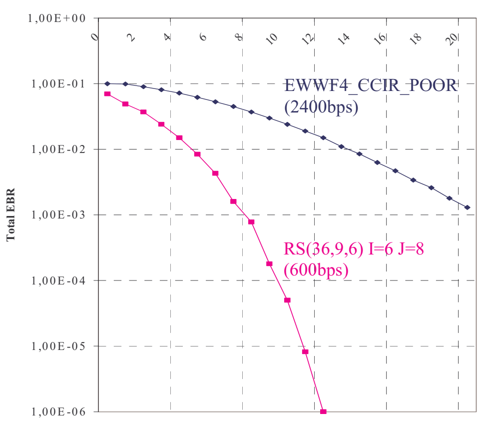

Figure 7. Proposed FEC scheme (CCIR poor channel) with EWWF4 waveform.

The proposed FEC solution includes the following features. A Reed Solomon Block Forward Error Corrector: RS(36,9,6), RS symbols Interleaver (I=6 J=8) (delay 2.4 s), and an End Of Message Detection (120 RS symbols), like ST4444/4479 mode. The obtained performances in case of CCIR poor channel are given below (Figure 7).

The CCIR poor channel includes 2 additional paths with 0 dB relative level, 1Hz shift Doppler and with respectively 0 ms delay and 2 ms delay. A 1.5% Frame Error Rate (FER) is obtained with a perturbed channel characterized by 7 dB SNR.

7. CONCLUSIONS

This paper described a new very low bit rate speech coder at 600 bits/s which was selected for the STANAG 4591. This coder was shown to be highly intelligible, to offer an attractive quality, and to be reasonably complex (Table 5). It would thus also be ideally suited for any other applications that require speech coding at a very low bit rate.

Requirements	ST4591 at 2400 bits/s	ST4591 at 1200 bits/s	ST4591 at 600 bits/s
Program memory (kWord)	20	30	30*
Data memory (kWord)	20	53	60*
Complexity (MIPS)	54	66	70*
* Estimation with 5% confidence, on TI C5416 platform at 140Mhz (embedded memory 128 kW), with Word = 16 bit.			

Table 5. Memory and complexity footprint for the three STANAG 4591 speech coder bit rates.

8. REFERENCES

[1] B. Mouy, P. de La Noue, and G. Goudezeune, "NATO STANAG 4479: A standard for an 800 bps vocoder and channel coding in HF-ECCM system", IEEE International Conference on Acoustics, Speech, and Signal Processing., Detroit, pp. 480-483, May 1995.

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

[3] A. V. McCree, and T.P. Barnwell III, " A mixed excitation LPC vocoder model for low bit rate speech coding", IEEE Transactions on Speech and Audio Processing, Vol. 3, No. 4, pp. 242-250, July 1995.

[4] L.M. Supplee, R.P. Cohn, J.S. Collura, and A.V. McCree, "MELP : The new federal standard at 2400 bits/s", IEEE International Conference on Acoustics, Speech, and Signal Processing, Munich, pp. 1591-1594, April 1997.

[5] T. Wang, et *all.*, "A1200/2400 bps coding suite based on MELP", IEEE Speech Coding Workshop, pp. 90-92 ,2002.

[6] "Draft Annex A to the slow-hop HF EPM STANAG-4444", NATO Standard STANAG-4444-Ed2, January 1999.