Audio-Visual Conference through the Ionosphere at 4 kbps

Antonio Navarro, Joao Angeja, Joao Tavares*, Rui Rodrigues, Luis Carvalho
Telecommunications Institute- Aveiro University, 3810 Aveiro, Portugal, navarro@av.it.pt
and
Fernando Perdigao
Telecommunications Institute- Coimbra University, 3030 Coimbra, Portugal, fp@co.it.pt

Abstract

Unlike in data communications, multimedia communications through heterogeneous networks is still far away of being achieved due to lacking of solutions at all protocol layers. The main objective of this research work is to evaluate the performance of a high frequency (HF) wireless network for transporting multimedia services. Beyond of allowing civil/amateur communications, HF bands are also used for long distance wireless military communications. Therefore, our research work is based on NATO Link and Physical layer standards, STANAG 5066 and STANAG 4539 respectively. A typical transmission bandwidth is of about 3 kHz with a throughput bit rate up to 12800 bps resulting in a very low channel capacity and therefore imposing serious challenges for reliable real time multimedia communications. This paper describes a complete video conference system designed to allow an end-to-end communication and discusses the quality of service in terms of packet lost and delay.

1. Introduction

Service access from anywhere by several user at anytime is foreseen to the future. In this context of network convergence, HF links is part of the global network playing an important role connecting long distance users. Therefore, convergence goes through vertical and horizontal integration of services, networks, systems, platforms and terminals. However, an efficient end-to-end content retrieval to any device through heterogeneous networks requires a serious and deep research in every aspect of communication. Besides, an HF link as part of a whole communication network represents a bottleneck. Each HF communication channel bandwidth is typically limited to 3 or 4 kHz. Considering that the Internet Protocol (IP) is the glue technology of the future convergence of networks, this paper presents the performance of IP packet transmission over HF links. The three main objectives are: to provide guidance for HF communication system designers; to propose improvements to STANAG [1,2] standards namely STANAG 5066 and STANAG 4539; to promote new standard technologies such as MPEG-4. The incentive and justification for this research work came from the Portuguese Navy's desire to develop multimedia services, namely video conference over HF channels. The research work presented in this paper makes use of an HF radio channel simulator [3] implemented according to Watterson's HF ionosphere model [4]. On top of this fading model, we implemented the physical and link layers defined by STANAG. IP/UDP/RTP header compressed packets are then transmitted over this layer structure. Our performance measure is based on packet loss and delay. The latter is due to packet filling with media data at the emitter peer.

This paper is divided into 7 sections. In the first two sections, we briefly describe the MPEG-4 speech and Face Animation Parameter (FAP) coding. In the following three sections, we propose an Audio and FAP packet encapsulation scheme and an HF gateway solution. Finally, the last two sections present results and conclusions.

2. Speech coding

At the application layer, the proposed system is based on the MPEG-4 audio [5] and visual [6] standards. The audio codec is the Harmonic Vector eXcitation Coding (HVXC) whereas the human face movements are synthesized using the FAPs. In contrast to general audio coders for natural audio

* Joao Tavares is sponsored by FCT-Portugal.

0-7803-8603-5/04/$20.00 ©2004 IEEE.

signals, speech coders are usually expected to handle human voice as an input signal. The most common and widely used model-based speech analysis system is linear prediction (LP), also known as linear predictive coding (LPC).

There are various types of speech coders but very-low-bit-rate coders are parametric. They can obtain good quality speech at very low bit rates such as 2 to 4 kbps. One particular parametric speech coder is the MPEG-4 HVXC which allows coding of speech signals at very low bit rates with reasonable quality – that is, communication quality to near toll quality. In fact, two bit rates are accepted: 2 and 4 kbps, defining the so-called 2- and 4-kbps modes. This is possible due to the efficient representation of the LPC residual, in which harmonic coding for voiced segments and VXC for unvoiced segments are used. Figure 1 shows an overall block diagram of the HVXC encoder.

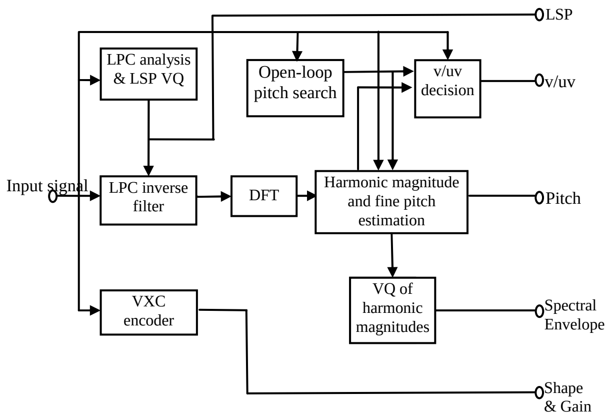

Figure 1. HVXC encoder

In each speech frame corresponding to 20 ms, the encoder generates 40 and 80 bits in the 2 and 4 kbps modes, respectively. Some encoded frames are grouped and encapsulated in packets as explained in Audio and FAP Encapsulation section.

3. Face animation

The MPEG-4 standard [6] specifies a face model in its neutral state, a number of feature points (FDPs) on this neutral face as reference points, and a set of face animation parameters (FAPs), each corresponding to a particular facial action deforming a face model. Deforming a neutral face model according to some specified FAP values at each time instant generates a facial animation sequence. The FAP value indicates the magnitude of the corresponding action, e.g., a big versus a small smile or deformation of a mouth corner. We developed face analysis system to calculate FAPs.

Since the FAPs are required to animate faces of different sizes and proportions, the FAP values are defined in face animation parameter units (FAPU). The FAPUs are computed from spatial distances between major facial features on the model in its neutral state. Figure 2 illustrates a face texture “Tavares” mapped onto a face model. MPEG-4 standard defines 84 FDPs and 68 FAPs.

Figure 2. Face texture mapped onto FAPs

Arithmetic coding is applied to the quantized and temporally predicted FAPs leading to a low-delay FAP coding. At an I-frame, FAPs are coded without prediction. At P-frames following I-frame, the values of FAPs at frame k (FAP_k) are predicted using the previously decoded value FAP_{k-1}. The prediction error is quantized using a different quantization step size (QP) for each FAP multiplied by a global quantization parameter (FQU).

In order to implement an error resilient communication system, we use an alternate frame coding sequence (I P I P….). Besides, as the packet size is small, we considered sending all FAPs from groups 7 (Head rotation) and 8 (Outer lip positions) and the first FAP from group 2 (Jaw, Chin, Inner Lowerlip, Cornerlips, Midlip) corresponding to the open_jaw. The remaining FAPs from group 2 are interpolated from group 8 FAPs. An I- and P-frames together do not exceed an RTP payload size of 50 bytes. Likewise for the same bit rate used in speech case, we obtain the frame (face) rates shown in Table 1.

Table 1. Frame rates

Bit Rate	Frame Rate
2 kbps	10 Hz
4 kbps	20 Hz

4. Audio and FAP encapsulation

Above the transport layer, Real Time Protocol (RTP) [7,8] defined by IETF has been used to encapsulate speech and FAP streams. The speech and FAP media are multiplexed and encapsulated into the same RTP packet. This encapsulation scheme results in four different RTP payload sizes, for delays not greater than 200 ms. Table 2 presents for two different delays and different combinations of speech and FAPs encapsulated into the same RTP packet, six possible cases of media transmission. Nevertheless, the overall transmission bit rate is higher than shown in the most right column of Table 2 as explained in the following section.

Table 2. Communication parameters

Delay (ms)	FAPs 2 kbps	FAPs 4 kbps	Audio 2 kbps	Audio 4 kbps	RTP pyld size (bytes)	Source media rate (bps)
100		50	25		75	6k
		50		50	100	8k
200	50		50		100	4k
	50			100	150	6k
		100	50		150	6k
		100		100	200	8k

5. HF gateway and header compression

In this section we will present a proposal of an HF gateway for real time packet routing. This gateway performs the interconnection between an Ethernet Network (LAN side) and a Stanag 5066 Network (HF side). Figure 3 shows the physical position of this gateway on a real scenario network topology. A particular IP/UDP/RTP header compressor was designed on the HF gateway in order to minimize the network traffic. The compressed packet unit is then encapsulated into the packet structure defined in the NATO standards. The lowest link layer packet is named D_PDU. Once generated, this packet is afterwards passed to the physical layer.

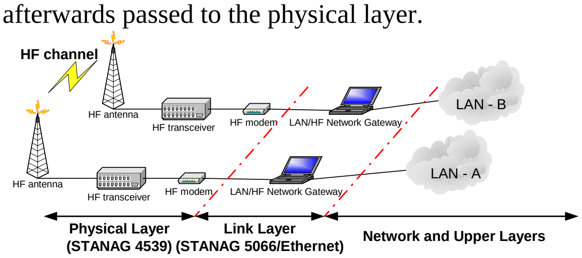

Figure 3. Internetworking with HF link

To minimize the header overheads on the HF channel, we consider that the HF network between the two gateways is a point-to-point link. Therefore, the total size of IP/UDP/RTP headers is compressed from 40 bytes to an average of 4 bytes, as stated on the RFC 2508 [8]. Thus, in the HF link, the summation of all packet headers included at the transport, network and STANAG link layers becomes, on average, equal to 38 bytes (34 bytes at the STANAG link layer). Header compression reduces the total packet overhead from 74 to 38 bytes. The compression/decompression of packet headers is done in the gateway during the routing of packets between networks.

The RTP payload throughput, RTP_PT, is given by,

$$RTP\_PT = \frac{SymbolRate * \log_2 M}{1.125} * CodeRate * (1 - PacketOverheadRatio) \quad (1)$$

where M is the number of constellation points, the symbol rate is equal to 2400 symbol-per-second and the packet overhead ratio is determined as explained in the following example. For the RTP packet payload length of 75 bytes (Case 1):

RTP payload length = 75 bytes;
D_PDU packet length = 75+38=113 bytes;
Packet overhead ratio = 38 / 113 = 0.3363. (2)

The above calculations can be straightforward extended to other cases of packet lengths. The above equation is then calculated for all possible combinations of packet overhead ratios, code rates and modulation schemes. Table 3 shows all obtained values. Let us consider a media data transmission presented in the most right column of Table 2 with a RTP payload generated at the emitter peer with delay time not greater than 200 ms. All possible combinations are indicated by the darker cells in Table 3. With a delay equal to 100 ms and media data transmission of 8 kbps (Table 2), just one combination is possible and is bold marked in Table 3. Once identified all allowed combinations, we should choose the communication parameters (code rate, modulator, packet size) leading to a lower packet loss. Therefore, we simulate different HF channel conditions within a SNR range from 0 to 45 dB [3].

Table 3. Transmission bit rates

	Case 1	75 bytes				**Case 2**	100 bytes			
FEC	**1/2**	**2/3**	**3/4**	**4/5**	**1**	**1/2**	**2/3**	**3/4**	**4/5**	**1**
QPSK	1416	1888	2124	2265	2832	1546	2061	2319	2473	3092
8QAM	2124	2832	3186	3398	4248	2319	3092	3478	3710	4638
16QAM	2832	3776	4248	4531	5664	3092	4122	4638	4947	6184
32QAM	3540	4720	5310	5664	7080	3865	5153	5797	6184	7729
64QAM	4248	5664	6372	6796	8496	4638	6184	6957	7420	**9275**
	Case 3	150 bytes				**Case 4**	200 bytes			
FEC	**1/2**	**2/3**	**3/4**	**4/5**	**1**	**1/2**	**2/3**	**3/4**	**4/5**	**1**
QPSK	1702	2270	2553	2723	3404	1793	2390	2689	2868	3585
8QAM	2553	3404	3830	4085	5106	2689	3585	4034	4303	5378
16QAM	3404	4539	5106	5447	6809	3585	4781	5378	5737	7171
32QAM	4255	5674	6383	6809	8511	4482	5976	6723	7171	8964
64QAM	5106	6809	7660	8170	10213	5378	7171	8067	8605	10756

6. Results

Firstly, we modeled the ionosphere radio channel and the complete network with compressed IP/UDP/RTP on the top of the above mentioned link layer. Secondly, we calculated several channel/network parameters like the SNR, BER, packet overhead efficiency and packet loss. Thirdly, we achieved the best packet size, FEC code rate and modulation. The results are a function of the instantaneous time-varying radio channel characteristics, SNR, delay, a stack of possible modulators (QPSK, 8-QAM, 16-QAM, 32-QAM and 64-QAM) and code rates (1, 1/2, 2/3, 3/4, 4/5) obtained by puncturing a convolutional 1/2 mother code. The transmission baud rate is 2400 sym/s with channel SNRs varying from 0 to 45 dB. Figure 4 shows the curves corresponding to the lowest packet loss percentage ("good" channel condition [4]) for allowed cases shown in Table 3. As expected, the lowest D_PDU packet loss corresponds to the constellations with lower number of levels. In Figure 4, the difference between 8-QAM-FEC 1 and 32-QAM-FEC 1 curves at SNR=15 dB is around 35%. However, the constellations with lower number of levels correspond to a higher packet transmission delay. This interpretation led us to a measure of performance taking into account the delay at the emitter peer, since we are dealing with real time signals.

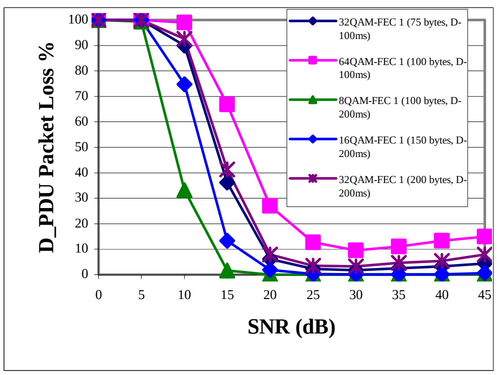

Figure 4. The least D_PDU packet loss curves

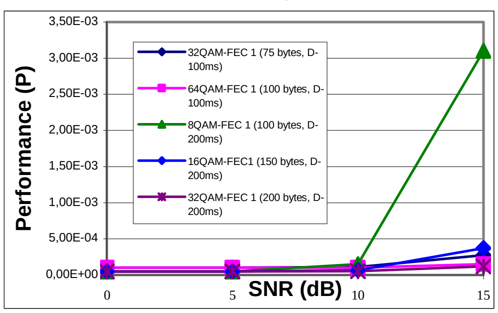

Figure 5. The greatest performance curves at SNR=15 dB

As the delay should also be taken into account, we then defined a performance measure criterion given by,

$$P = \frac{1}{\text{Delay} * \text{Packet loss}} \tag{3}$$

We have applied the above expression to the values obtained from Figure 4 up to SNR=15 dB as shown in Figure 5. As the performance increases significantly above SNR=10 dB, we recommend the following transmission parameters: uncoded, SNR greater than 10 dB and 8-QAM.

7. Conclusions and future work

In this paper, the QoS in real time communications is measured in terms of media packet loss rate and delay. Despite our results show that the FEC codes is not effective (the interleaver was not used due to delay constrains), a careful packet length design is crucial to achieve a high performance real time packet HF communication. Our results show that the most appropriate source packet length (RTP payload length) is equal to 100 bytes (200 ms delay). This RTP packet transports 50 bytes of encoded speech (2 kbps) and 50 bytes of FAP data (2 kbps). In this case, for instance, the packet loss is equal to 1.6 % at SNR=15 dB. Our audio-visual conference solution is straightforward extended to cope with more than two users since the HF network gateway delivers only one user's face movements and speech to the HF link.

8. References

[1] STANAG 5066, "Profile for High Frequency (HF) Radio Data Communications", NATO, Ratification Request, 1999.
[2] STANAG 4539, "Technical Standards for Non-Hopping HF Communications Waveforms", NATO, Ratification Request, 2000.
[3] M. Rostami, J. Angeja, J. Tavares and A. Navarro, "HF Channel Modeling for Real Time Packet Transmission", SPIE - Digital Wireless Communications, Orlando, Florida USA, April 2003.
[4] C.C. Watterson et al, "Experimental Confirmation of an HF Channel Model", IEEE Trans. Commun. Technology., vol. 18, pp. 792-803, 1970.
[5] MPEG-4 Audio Group, "Information Technology – Very Low Bitrate Audio-Visual Coding: Part 3 - Audio", ISO/IEC FCD 14496-3, 1998.
[6] MPEG-4 Video Group, "Information Technology – Coding of Audio-Visual Objects: Part 2 - Visual", ISO/IEC FCD 14496-2, 2001.
[7] H. Schulzrinne et al, 1996, "RTP: A Transport Protocol for Real-Time Applications", Network Working Group Request for Comments, RFC 1889.
[8] S. Casner and V. Jacobson, 1999, "Compressing IP/UDP/RTP Headers for Low-Speed Serial Links", Network Working Group Request for Comments, RFC 2508.