Efficient Evaluation of Voice Quality in GERAN (GSM EDGE Radio Access Network)

N. Paul #, J. Brouet #, C. Thirouard #, J. Antoine °, A. Wautier °, L. Husson °
° Alcatel Research & Innovation Dpt., Route de Nozay, 91 460 Marcoussis, France
# Supélec, Dpt. Radio-Electricité, Plateau de Moulon, 91 192Gif-sur-Yvette, France

Abstract—The transport of multi-media application is now a key feature of the future radio access networks. The evolution of GSM networks is also able in principle to take benefit from end-to-end transfer of such applications. However, careful attention has to be taken regarding the quality perceived by the end user. Indeed, for voice services, mobile subscribers are familiar with a given quality. The introduction of IP services should not decrease this quality. Besides, the introduction of such feature on the radio links should also enable the operators to increase their network capacity. In this context, this paper analyses the perceived voice quality when voice frames are transported on existing packet radio bearers of GSM/EDGE Radio Access Networks. In this situation, the benefit for the operator should be maximal. The quality is evaluated with an objective method that relies on comparison of reference speech sequence with the same sequence that is passed through GERAN radio links. This methodology of evaluation if by far more consistent than simply considering radio link level performance like the bit error rate.

I. INTRODUCTION

In the recent years, second generation cellular systems have been widely deployed worldwide and, in many countries, the number of cellular subscribers is now superceding the number of fixed line subscribers. The most popular cellular standard worldwide is the GSM (Global System for Mobile Communication). It mainly allows users to access to mobile voice services with a good quality. Recently, significant enhancement of the GSM has been achieved for introducing the support of efficient data transmission on the same radio interface. This is the so-called GPRS (General Packet Radio Service) and its enhanced version EGPRS that, in addition, makes use of a modified physical layer (8PSK modulation is used instead of the GMSK modulation when the radio link conditions are favorable, which significantly increases the throughput). In such evolved network, there is, in the network sub-system (NSS), a functional and physical separation between network nodes supporting voice services (that make use of circuit switching via the A interface) and network entities supporting data transmission (that make use of packet switching via the Gn interface). The circuit switched NSS is interfaced with conventional public switched telephone network and the packet switched NSS is interfaced with a packet data network (e.g. the Internet).

In parallel, strong effort is currently spent for unifying circuit switched and packet switched core networks in a so-called all-IP network architecture. In such core networks, services (being real-time or not) will be carried by a unique network infrastructure to the access networks that will serve the end-user. The transport protocols will be largely based on IP protocols with some modifications to selectively guarantee transmission delays across the network according to the service to be provided to the end-user. Indeed, this new core network architecture impacts the radio access networks.

Third generation cellular standards (e.g. UMTS) provide means to interconnect mobile radio networks with this evolved IP network. The GSM standard is also evolving in the same direction. The modifications of the GSM are reflected in the GERAN (GSM EDGE Radio Access Network) standard. In the first releases of this standard, real-time IP applications (e.g. VoIP) are delivered to the mobile user by using conventional circuit switched traffic channels (TCH) on the air-interface. This allows to quickly enable the support of any IP multimedia applications on GSM. Indeed, the signaling protocols for call establishment are rather different, but existing mechanism for the transport of traffic information on the air is unchanged. However, this “straightforward” solution does not bring significant performance gain since the packet nature of the service ends up in the RAN and does not “cross” the air. In future releases of GERAN, end-to-end packet transmission of real-time IP applications is envisaged. In this context, voice and data information of different users will be dynamically multiplexed on the same packet radio channels which will bring overall capacity gain.

This article analyses in this context the transport of voice frames onto packet data channels (PDCH) with the assumption that existing radio bearers defined in GPRS/EGPRS are used. The analysis is focused on the radio link level and does not consider higher layer issues e.g. mechanisms for multiplexing of users. In section II the transport of voice onto packet bearers is presented. In addition, criteria and associated methodology to assess the perceived speech quality by the end user are explained. In section III, the simulation environment is described and reference results on TCH channels are provided. In section IV the performance of voice transport onto EGPRS radio bearers is given along with means to enhance the perceived voice quality. A final conclusion on the methodology and performance improvement is then derived.

II. TRANSMISSION OF VOICE OVER PACKET RADIO BEARERS

A. Voice over EGPRS

For taking maximum benefits from end-to-end packet voice transmission using a cellular system as the access network, the packet voice frames have to be carried on the air interface using packet radio channels. As a consequence, provided that the radio resources are appropriately managed, several users can be multiplexed on the same radio channels. Indeed, voice is a bursty stream (i.e. a succession of talk-spurts and silences). If the radio resource scheduler has, means to quickly detect transitions from talk-spurts to silences (and vice versa), and also the capability to ensure that voice packets are transmitted on the air within a given delay bound with limited jitter, the non-speech activity intervals can be used to convey information streams (either real-time or not) from the same or other users. Besides, the possibility to use higher rate radio channels when the 8PSK modulation is used
allows to increase the number of sub-channels per time slot. All that enables in principle to significantly increase the overall network capacity. However, there are several issues that arise from the characteristics of the packet radio bearers as they are defined in the GERAN standard but also from the characteristics of the voice codecs that are being used on radio links [1].

The cellular voice codecs deliver for each sampled speech frame period (of 20 ms in GSM) a stream of compressed bits. These bits are classified according to their degree of relevance into 2 classes, namely class I bits and class II bits. Class I bits are further divided into two sub-categories: class Ia and class Ib. The correct reception of class Ia bits is essential for reconstructing the original speech frames. Class Ib bits tolerate some residual errors at the receiver side. Finally, class II bits tolerate a higher error rate. In the GSM standard, these characteristics are exploited to optimize the trade-off between the available bandwidth and the forward error correction (FEC) schemes. A convolutional code is used to protect class I bits. In addition, class Ia bits are protected by a cyclic redundancy code (CRC) that then checks the validity of the sequence at the receiver. Class II bits are transmitted without any FEC. This way of coding is called unequal error protection (UEP). In addition to UEP, the encoded bit streams are interleaved on several TDMA frames using a diagonal interleaving of depth 8. Figure 1 gives a schematic of conventional UEP scheme for the GSM full rate (FR) codec. Reference performance for different codecs used in GERAN can be found in [2].

Figure 1: Transmission of GSM FR frames on TCH.

On the opposite, the packet radio bearers of GPRS/EGPRS have been designed for transporting data streams. In this context, any bit to be conveyed on the air has equal importance. However, all the modulations and coding schemes (MCS) that have been defined for GPRS/EGPRS use equal error protection (EEP) schemes. In this situation, class I bits are not enough protected (this is especially true for class Ia bits) and class II bits are over-protected. For instance, if GSM FR frames are transported MCS3 EGPRS bearer, this provides 0.85 code rate in GMSK mode for bits (cf. figure 2). Besides, the transmission on the radio interface in GPRS/EGPRS is based on radio blocks that consist of 4 bursts in 4 consecutive frames. The interleaving scheme used is then rectangular and limited to a depth of 4. A rough evaluation of the degradation of the performance when using the packet links for transmitting voice instead of the optimized voice bearers is presented in [3]. The following sections further analyze this degradation.

Another critical issue is related to the size of the headers of VoIP frames. Indeed, those headers (RTP/UDP/IP) are much more larger (40 to 60 bytes) than the voice frames to be transported (32.5 bytes if GSM FR codec is used). For spectral efficiency sake, the header can be easily compressed down to 1 to 2 bytes, making then relevant the transport of VoIP frames on the air. Example of such compression scheme and associated performance can be found in [4-5]. In the present paper, perfect reception of compressed headers is assumed.

B. Evaluation of voice quality

There is an interest to transmit voice frames onto packet radio bearers only if the quality perceived by the end user is similar to the one that is provided by the TCH bearers, or alternatively, the perceived quality is acceptable and the system capacity is increased. It is then essential to assess the quality perceived by the end user. Commonly, the quality is assumed to be acceptable if average link level performance metrics are below predefined thresholds. In GSM, the metrics are the frame error rate (FER), the residual bit error rate (RBER) on class Ib bits and the BER of class II bits [2]. Those performance thresholds are obviously linked to the coding and the interleaving schemes that are being used. Taking the same target values when voice frames are transmitted on radio bearers is then not really appropriate. Besides, the targets do not take into account the error frame distribution into consideration, which is also a potential source of different perception of the voice quality for the same average link performance. As a consequence, other means to assess the voice quality should be employed.

The most reliable method is to perform subjective tests. A significant range of speech sequences is passed through the system under characterization. These sequences are evaluated and ranked by a representative auditory, statistics are collected and a unique subjective quality note is then derived
for each transmission situation test. But, setting up such experiments is very heavy, especially if the system under test can take a multiplicity of configurations (which is the case when dealing with voice transmission on radio links).

Objective methods have been developed to advantageously replace those testing sessions. The principle is the following one. Reference speech sequences are passed through the system under characterization. Then, the output of the system is compared to the reference speech sequence and a ranking is derived. Basically, the comparison is performed by analyzing the signal difference in several critical bands of the speech sequences making use of complex cognitive models of the human ear perception (figure 3).

![Figure 3: Principle of objective methods.](image)

One of the most often used methods is the Perceptual Speech Quality Measurement (PSQM) method [6]. This method was designed to assess perceived quality of narrow-band voice codecs such as the ones used in cellular systems. The PSQM gives a degradation evaluation between the reference and the output speech sequences. It has been thoroughly validated and gives consistent results when no transmission errors are generated by the system under characterization. However, it is still a good tool for assessing relative information of perceived quality.

For systems with transmission errors, an evolution of the PSQM method, the Perceptual Evaluation of Speech Quality (PESQ) method has recently been proposed in ITU [7]. The principle is similar to the PSQM but in addition, voice frame with errors are accounted for in the evaluation of the resulting voice quality. Besides, it has the advantage to deliver Mean Opinion Score values (MOS) which is a widely understood metric in the telecommunication community. The results presented in the next sections are based on the PESQ method.

III. SIMULATION ENVIRONMENT

A. Simulation platform

The simulation platform consists in several modules that are depicted in figure 4:

- GERAN speech encoder / decoder,
- GERAN channel encoder / decoder,
- GERAN channel model,
- PESQ module.

![Figure 4: Simulation platform.](image)

To speed up the simulations by saving most the computations required for demodulation / equalization, and for generating the fast fading variations of the channel, the radio link level modules (modulator, mobile channel, demodulator) are advantageously replaced by a two-level channel model. It consists in an external and an internal chain. The external chain gives the time variations of the signal to noise ratio (SNR) per burst having a similar auto-correlation and distribution as typical radio-mobile channels (e.g. typical urban 50 km/h –TU50-). The internal chain is based on a Fritchman model (that makes use of the Markov formalism) that describes for each burst the error pattern at the output of the demodulator [8]. Then, one state of the channel is fully described by three components: the initial state probability vector, the state transition matrix, and the observations corresponding to each state. There are internal chains for each SNR state and for each modulation. Here, the values of the Fritchman model are based on optimal receiver structures assuming perfect channel impulse response estimation. Besides, the article is restricted to hard decisions at the output of the demodulator, but soft decisions could also be envisaged [9].

The following results are for TU50 channels without frequency hopping in a noise-limited environment. Obviously, same methodology can be used for any other channel configuration. Besides, only one speech sequence (English language, female) has been used. The final quality is the average quality of the same speech sequence passed through the system with different channel realizations having the same average SNR.

B. Reference performance for GSM FR and GSM EFR:

![Figure 5: Reference performance curves](image)
Reference performance are those of the conventional GSM codecs. Figure 5 plots the voice quality as a function of the SNR for GSM FR, GSM EFR (Enhanced full rate) codecs. For all the codecs, simple error concealment processing is implemented (in case of reception of an erroneous voice frame, previous voice frame is repeated but with a decreased amplitude). Acceptable quality limit is set to a MOS value of 3. Referring to figure 5, the acceptable quality for GSM FR is achieved for a SNR around 9 dB. This value is consistent with the one specified in [2]. Besides, GSM EFR shows better quality for high SNR.

IV. RESULTS

MCS3 (37 bytes payload) is the first EGPRS container that can convey 1 GSM FR frame (32.5 bytes). MCS3 uses GMSK modulation and a code rate of 0.85 (the data bits have few channel error protections). Similarly, MCS6 is the first EGPRS container that can transport 2 GSM frames per container. Hence, two voice users (using MCS6) can share the same physical channel. MCS6 uses 8PSK modulation and a code rate of 0.49 (channel error protection is similar to TCH but the modulation is less robust). The results focused on these MCS since they are the most suited payload format to carry GSM FR frames.

A. Voice over EGPRS: straight transmission

Figure 6: Transmission of GSM FR as such on MCS6.

The first scenario considers that voice frames coming out from the GSM FR codec are transmitted as such on PDCH (like in figure 2). In this case, all the voice frames coming out from the channel decoder are passed to the voice decoder without any indication whether the voice frame is correct or not. Indeed, there is no CRC appended to class Ia bits and it is not appropriate to consider the information of the CRC of the packet payload since it accounts for errors in the whole payload and not only on class Ia bits. As a consequence, performance curve in figure 6 does not include any error concealment method when GSM FR frames are transported on MCS6 bearers. The obtained performance is then significantly below GSM FR on TCH (a MOS equal to 3 is obtained for SNR = 19 dB). However, if we consider a GSM network with a frequency reuse of 9, approximately 50% of the area has a SNR value higher than 19 dB. In this case on 50% of the area 2 users can be multiplexed with good quality on the same physical channel.

B. Voice over EGPRS: performance enhancement

Previous performance can be easily improved. Indeed, an optimal design of a channel code adapted to voice transmission onto packet bearer is one solution for improvement [9]. However, this type of solution requires new packet bearers for any GERAN voice codec and modulations. This would require significant modifications of the standard. The tracks followed here aim at reusing existing packet radio bearers without any modifications of the channel FEC schemes. Improvements are then achieved by additional treatment above the channel encoder / decoder.

The first way of improvement is to append a CRC code to the class Ia bits transmitted in the packet containers. After the channel decoding of the packet payload, the validity of the class Ia bits can be checked. In case an error is detected on these bits, the same simple error concealment mechanism as in section III is applied. Figure 7 shows that 4 dB gain can be achieved with MCS3. For MCS6, a 2 dB gain has been observed.

Figure 7: Performance improvement with simple error concealment scheme (MCS3)

Even if error concealment is used, performances with MCS3 (that does not allow to multiplex several voice users on the same physical channel) are still far from the reference GSM FR performance (6 dB loss). The first cause of degradation is the usage of a very low protection FEC scheme. The second cause of degradation is that only hard decisions are feeding the Viterbi decoder of the FEC in the channel model. Then, MCS3 performance is much more affected than the TCH performance since MCS3 is highly punctured whereas TCH is not punctured at all. However, some improvements are still required.

An additional way of enhancing the performance is to organize differently the transmission of voice frames so that UEP is emulated. The idea is to transport pieces of voice frames that need high level of protection with a highly
protected MCS scheme and other pieces of the same voice frames with a MCS scheme having lowest protection. Obviously, the overall bandwidth to transmit the information should be the same. Figure 8 illustrates the new channel organization for GSM FR frames. Instead of transmitting the frames with MCS3, class Ia bits of 2 consecutive frames are transmitted on a first radio container using MCS1 (GMSK modulation and 0.53 code rate) which has similar characteristics as the TCH encoder. The remaining bits of the 2 frames are transmitted with less protection on a second container, MCS5 (8PSK and 0.37 code rate).

The quality of the transport of the voice frames has been compared with existing transport on TCH link. It obviously resulted in a significant degradation of the performances. However, with some very simple tricks, the obtained voice quality on packet bearers can be significantly improved so that the performances approach the ones of TCH channels. Besides, if high data rates radio containers are used, the loss in quality is balanced by the ability to transport in parallel on the same physical channel several voice communications and then to increase the offered load of voice users (2 in case of MCS6 and GSM FR codecs). Such transport is then promising, since in addition, it exists robust header compression scheme that do not affect the spectral efficiency when carrying RTP/UDP/IP frames on the air and since the use of packet switching on the air enables to multiplex on the same physical channel users with different QoS (silences during speech sequences can be exploited for transmitting e.g. non-real-time data). Therefore, more capacity and more flexibility in the usage of the scarce radio resource can be obtained.

REFERENCES