This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the ICC 2007 proceedings.

Performance Evaluation of Quality of VoIP service over UMTS-UTRAN R99

Andrea Barbaresi, Andrea Mantovani
Telecom Italia - Via G. Reiss Romoli, 274 – I-10148 Torino (TO), Italy
andrea.barbaresi@telecomitalia.it
andrea.mantovani@guest.telecomitalia.it

***Abstract* – Release99 of UMTS (Universal Mobile Telecommunication System) network supports conversational services by means of UTRAN (Universal Terrestrial Radio Access Network) Dedicated transport CHannels (DCH). In this paper we evaluate by means of dynamic simulations the impacts of the radio access network on the end-to-end Quality of Service (QoS) of Voice over IP (VoIP). In order to characterize the QoS level, an analytical model devoted to predict the Mean Opinion Score (MOS) has been considered. Finally, an investigation of the most proper set-up of the service is also presented, taking into account inter-layer parameters such as the dimension of the VoIP client buffer as well as the BLock Error Rate at the air interface (BLER).**

Keywords: QoS, VoIP, MOS, UMTS, UTRAN, DCH, client buffer.

I. Introduction

During the last few years, the growth of Internet Protocol (IP) based networks and broadband access offered at low cost has focused a lot of interest towards Voice over IP (VoIP) service. Nowadays, the VoIP technology can replace in most cases traditional circuit switched (CS) telephony, resulting in reduced costs for the end-user. Obviously, in order to make the VoIP service really attractive, the Quality of Service (QoS) must be close to that obtained by traditional telephony.

Furthermore, the current trend in the evolution of cellular networks is moving towards "all-IP" architectures [1]; this convergence toward a packet switched (PS) only domain could be particularly interesting for operators not only as a mean to reduce CAPital EXpenditures (CAPEX) and OPEration EXpenditures (OPEX) but also to simplify network management.

For this reason, the interest for real-time (i.e. conversational and streaming) PS services is very high also in the current releases of third generation (3G) cellular networks. In particular, many studies focused on the optimization of VoIP over Release5 HSDPA (High Speed Downlink Packet Access) and Release6 HSUPA (High Speed Uplink Packet Access) are available from literature [2],[3]. In this context, in order to be able to manage and accommodate VoIP in an integrated way over different air interfaces, the analysis of VoIP QoS over UTRAN Release99 dedicated transport channels represents a good starting point for the studies on other architectures.

As stated above, in order to obtain a large end-user acceptance, the network has to convey a QoS similar to that offered by traditional CS technology. This can be achieved by an accurate tuning of the main parameters that affect QoS; as a matter of fact, while standard CS telephony was designed specifically for voice calls and provides good speech quality and good spectral efficiency, VoIP must cope with several problems, typical of PS networks, related to packet loss, delay, jitter and the header overhead introduced by the protocol stack.

The scope of this paper concerns the characterization and analysis of VoIP QoS over UTRAN Release99 dedicated transport channels; particular attention has been paid to the search of optimum values for some of the main parameters that directly affects QoS. In order to perform this study, a proprietary event-driven simulation tool has been used, named RAMSET™ (Radio Access of Mobile Systems Evaluation Tool).

This paper is organized as follows. Section II presents some general information about VoIP service and its architecture in 3G cellular networks according to Third Generation Partnership Project (3GPP). Section III shows an accurate bi-directional traffic model for VoIP sources used in the simulations, while Section IV presents the model used for the QoS evaluation. In Section V a general description of the simulated scenario is provided, as well as information about simulation hypotheses. Section VI shows the main simulation results obtained and, finally, Section VII summarizes the conclusions coming from this study.

II. Voice over IP in 3G networks

When considering the transmission of voice traffic over PS networks, it is very important to be able to manage a large amount of traffic, preserving as much as possible the quality of every single voice source and solving the bottleneck problem represented by the bandwidth capacity. For this reason, various voice compression schemes have been developed in order to reduce bandwidth requirements for vocal information (typically, 64kbps for a Pulse Code Modulation signal with a sampling frequency of 8kHz), while maintaining a good perceived quality. Among these compression schemes, known as codecs, 3GPP has chosen AMR (Adaptive Multi-Rate) as a standard codec for voice signals in 3G mobile communications.

AMR uses ACELP (Algebraic Code Excited Linear Prediction) coding algorithm in order to reduce the overall bitrate of output-encoded signal. There are also other techniques used to further improve the quality/bandwidth ratio, called VAD (Voice Activity Detection) and DTX (Discontinuous Transmission); these techniques are used to localize and characterize vocal inactivity periods that yield to a lower bitrate.

The protocol stack for VoIP service in the User Plane (UP) consists of RTP/UDP/IP protocols; RTP (Real Time Protocol) is designed to carry real-time multi-media traffic, while UDP offers a connectionless transport that minimizes delays. The activation of RTCP (Real Time Control Protocol) in the Control Plane (CP) is optional. Regarding the control procedures, 3GPP has chosen SIP (Session Initiation Protocol) in order to establish, configure and release PS real-time services. SIP protocol can be used either with UDP or TCP transport protocols. The complete protocol stack in the UP can be seen in Figure 1.

III. VOIP TRAFFIC MODEL

When modeling voice traffic for dynamic simulations, it is important to understand the speech process between two users. The research effort [4] in this area found that a general human vocal pattern is mainly characterized by two events: talk spurts (periods in which there is a vocal activity) and silence periods (for example, between words, inside a single word or while listening to the other speaker). In a simplified model only these two events can be considered: in other words, the vocal pattern can be modeled by an alternation between periods of activity and inactivity. This simplified model characterizes single source traffic without considering interaction between callers; a more realistic VoIP model should incorporate at least two users sending voice traffic in both directions. The two-way conversation can therefore be modeled by a 4-state Markov chain [5], as shown in Figure 2, where:

- **State A** indicates that user A is speaking and user B is listening;
- **State B** indicates that user B is speaking while user A is listening;
- **State M**, called *Mutual Silence*, defines a state in which both users are silent;
- **State D**, called *Double Talk*, represents a situation in which both users are speaking simultaneously.

This model can be characterized by the states transition probabilities *p1*, *p2*, *p3*, considering negligible the transition probabilities between states A and B and between D and M, due to their rare occurrence. Furthermore, the sojourn time in each state is modeled by a random variable with exponential distribution, each one respectively defined by parameters λ_A, λ_B, λ_D and λ_M.

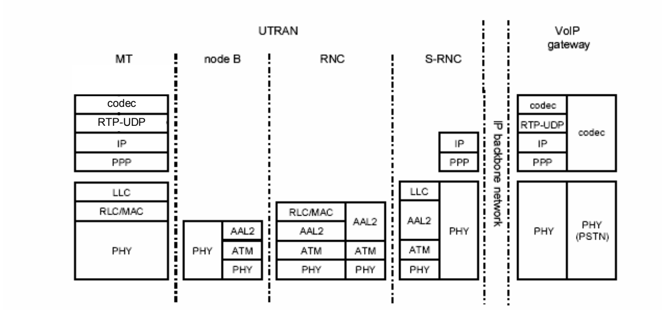

Figure 1 – User plane protocol stack for VoIP service.

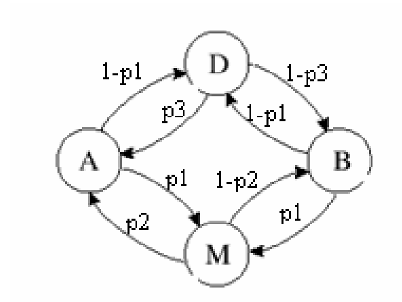

Figure 2 – Bi-directional traffic model for voice conversations.

The typical values taken by the parameters described above are [6]:

$$\begin{cases} p1 = 0.4 \\ p2 = 0.5 \\ p3 = 0.5 \end{cases} \tag{1}$$

$$\begin{cases} \lambda_A = 1/854ms \\ \lambda_B = 1/854ms \\ \lambda_D = 1/226ms \\ \lambda_M = 1/456ms \end{cases} \tag{2}$$

A complete model for a general VoIP traffic source must refer also to the behaviour inside each state, and this aspect depends on the specific voice codec used, which is defined by 3GPP as AMR, as previously stated. Then, during vocal activity and inactivity periods the model must also implement the codec behaviour.

IV. QUALITY ESTIMATION

Delay and packet loss requirements for VoIP lead to accurate design needs for the overall system that support this kind of service: preserving the quality for a VoIP communication means that transmission delay and packet loss rate should be minimized. For this reason, it is important to understand how to measure or estimate the perceived quality in a VoIP session. In this context, two different methods can be used: an objective measurement method or a subjective one [7]. Subjective measurement methods are more reliable, but their implementation is more complex and expensive as they require to collect and analyse information taken from sample groups of people.

Objective measurement methods use some objective parameters (such as network delay and losses, psychoacoustic and elementary parameters) and process these measures to predict how the equivalent subjective perceived quality would be; these methods represent a good trade-off between complexity and reliability. Furthermore, they are particularly suitable for simulations tools because they do not involve real communications.

Among objective metrics, the e-model (defined in ITU-T G.107 standard – [8]) plays an important role. E-model bases its own behaviour on the following consideration: "psychological factors on the psychological scale are additive", which means that a linear relation exists between quality and each impairment factor, and single factors contribute independently to the overall quality value.

The e-model output is a parameter (R) that ranges between 0 and 100, and gives an indication about perceived quality; there is also the possibility [9] to transform R value in the corresponding value in term of MOS (the Mean Opinion Score is the output value of subjective metrics of quality).

R parameter can be calculated as follows:

$$R = R_0 - I_d - I_{e,eff} + A \tag{3}$$

where:

- R_0 represents the perceived quality obtained in the absence of impairment factors;
- The Advantage factor (A) is an additive term that indicates the advantage (in term of quality) observed in a particular communication environment; the recommended value (of 10) for the mobile environment was assumed [8]. In fact, in a mobile communication it can be thought that users accept a greater degradation of quality than in a wired communication, so their opinion will be more "tolerant";
- I_d represents the impairment due to end-to-end delay, and it can be calculated as follows (this is a simplified equation, see [10]):

  $$I_d = 0.024 \cdot d + 0.11 \cdot (d - 177.3) \cdot u(d - 177.3) \tag{4}$$

  where d is the end-to-end delay (measured in ms), and *u(...)* is the Heaviside step function;
- $I_{e,eff}$ indicates the impairment on quality due to packet loss rate, and it can be calculated using:

  $$I_{e,eff} = I_e + (95 - I_e) \cdot \frac{p}{p + Bpl} \tag{5}$$

  where p is the packet loss rate (expressed in %) of transmitted frames. Equipment Impairment Factor (Ie) and Packet-loss Robustness Factor (Bpl) are codec-dependent parameters; for 12.2kbps AMR codec, recommended values for Ie and Bpl are 5 and 10, respectively [11].

Then, R value obtained from e-model can be converted to the corresponding MOS value through the following equation:

$$\begin{cases} MOS = 1 & R < 0 \\ MOS = 1 + 0.035R + 7 \cdot 10^{-6} R(R-60)(100-R) & 0 < R < 100 \\ MOS = 4.5 & R > 100 \end{cases}$$

V. Simulation Methodology

Several dynamic simulations have been conducted in order to evaluate the QoS for VoIP service, taking into account different operational points of UTRAN R99. The specific network layout selected (consisting in 36 macro cells arranged in 12 three-sectorial sites) can be considered representative of a general urban environment, so results described in the next section can be considered valid for this type of scenario. Since the focus of the study is to evaluate the QoS level for VoIP service when different settings are applied, simulation results consist in mean values for all the simulated sessions on each cell under statistical observation. The length of simulations was set up to 3000 seconds, enough to collect results with confidence intervals of 95%, which are statistically significant.

Figure 3 depicts the architecture used for the simulations. Each VoIP call is established between a generic 3G mobile user and the host inside the IP fixed network, therefore all the simulation results reflects this configuration (for example, the measured delay takes into account only one radio link). Other types of VoIP calls, i.e. a call between two mobile users, have not been considered. However, the chosen architecture does not represent a limit for the analysis of other types of VoIP calls since the investigation carried out can also help to collect information for other scenarios; in fact, in a mobile-mobile call, the total delay can be calculated by the sum of uplink and downlink delays (obtained from the mobile-host communication case) and similar considerations can be made for other parameters.

The simulation tool performs a detailed analysis of UTRAN behavior, taking into account all the details concerning radio protocols involved in the service delivery, while considering some simplifications on the other network devices involved. In particular, delay results reported in the next section refer to the contribution of the air interface within UTRAN, without considering the delay contributions introduced by the Core Network (CN) and IP fixed network. Consequently, typical mean values of delay for the transport part of the network [12] should be added to the reported results in order to estimate the end-to-end delay.

In order to perform simulations, some specific modules for VoIP service have been used:

- VoIP traffic generator: it produces VoIP traffic for the peer entities involved in the voice call. It follows the traffic model described in Section III;
- VoIP client: it uses RTP protocol to send VoIP frames over the network (RTCP protocol activation is optional). When receiving VoIP frames, it implements a static buffer technique to compensate jitter effects;
- VoIP sessions control module: it uses SIP protocol to establish and release VoIP sessions.

The simulation work has been focused on the evaluation of QoS experimented by VoIP users with respect to variations of some of the most relevant and variable factors; particularly, we focused on the radio block error rate (BLER) and the de-jittering buffer length.

The range used for the above-mentioned variables has been defined as follows:

- BLER ∈ {2, 5, 10} %
- De-jittering buffer length ∈ {0, 20, 60, 180} ms

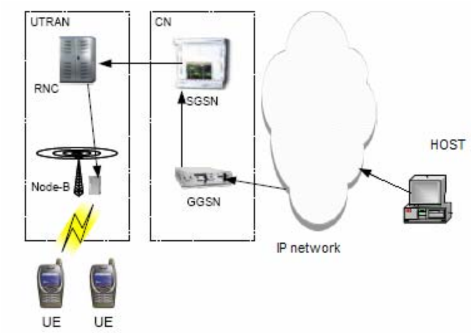

Figure 3 – Simulated architecture.

Simulations have been conducted with the following parameters for VoIP service:

- Mean VoIP call duration = 120 s;
- AMR codec at 12.2kbps bitrate and Voice Activity Detection mechanism activated;
- SIP protocol used above the TCP stack;
- IPv4 protocol stack with ideal RObust Header Compression mechanism (ROHC) (i.e. it is assumed that the algorithm is always in the steady state);
- 16.8kbps UL/16.8kbps DL PS Conversational RAB for user data, specifically optimized for VoIP traffic [13];
- 8kbps UL/8kbps DL PS Interactive RAB for SIP traffic as defined by 3GPP specifications [13];

Regarding the number of users assigned to each cell, it has been supposed to perform simulations in medium traffic load conditions (i.e. 15 active users per cell). This hypothesis has been chosen in order to evaluate the QoS without considering congestion limitations due to cell over-loaded situations.

VI. Simulation Results

This section shows the main results obtained by mean of simulations, with a focus on the following application-layer parameters:

- Transmission delay: end-to-end delay measured between mobile device and VoIP host client (with the assumptions already described in Section V);
- Number of playout interruptions: number of times per session in which the de-jittering buffer becomes empty when the conversation is active and the VoIP decoder should play a voice frame (i.e. "buffer underflow" condition);
- MOS: value of Mean Opinion Score estimated by the e-model described in Section IV;

Table 1 shows the values of mean transmission delay with respect to de-jittering buffer length and BLER target values. It is worth noting that values increase when increasing de-jittering buffer length since the end-to-end delay is composed, in the considered scenario, by two factors: the radio interface transmission delay and the delay introduced by the VoIP client buffer.

Table 1 – Transmission delay vs buffer and BLER.

Transmission Delay [ms]	Buffer 0ms	Buffer 20ms	Buffer 60ms	Buffer 180ms
BLER = 2%	20	41.3	71.3	132.2
BLER = 5%	20	41.2	71.3	131.9
BLER =10%	20	40.2	70.3	131.7

Fixing a value for the BLER and in the absence of de-jittering buffer (0ms buffer length), the measured transmission delay corresponds to one TTI (Transmission Time Interval); when the buffer is present, the total delay is also composed by the mean delay introduced by the buffer, which is proportional to its length. More in detail, the delay introduced by the de-jittering buffer depends on its mean filling level during the VoIP session; for example, when a buffer of 180 ms is considered, results show a mean transmission delay of about 132 ms (i.e. lower than the maximum buffer length). Figure 4 shows in detail the trend of transmission delay versus buffer length in the 2% BLER case (values are expressed relatively to the minimum transmission delay). Fixing a value for the buffer length, the target BLER has low influence on the mean transmission delay: increasing the BLER value leads to a greater number of lost packets that the buffer does not receive (the considered RAB combination assumes the RLC protocol to operate in Unacknowledged Mode, i.e. without retransmissions), so the mean delay introduced by the buffer is slightly different.

With respect to the number of lost VoIP frames measured by VoIP client, it has been observed that this number matches the target BLER at radio interface. This seems to be quite obvious if it is considered that, according to the RAB combination, every single VoIP frame is encapsulated in one RTP/UDP/IP packet, which corresponds exactly to one radio frame, so the packet loss at application layer assumes the same value of the target BLER at the radio interface.

Then, an analysis of QoS has been made observing the value of MOS experienced during simulations; Table 2 shows the values of predicted MOS versus buffer length and target BLER. As stated in the previous section, the reported MOS values reflect only the contributions introduced by the UTRAN, without considering any additional delays due to the CN and the IP fixed network. Hence, the MOS estimation can be considered representative of the end-to-end QoS for mobile users which do not experience any delay limitations introduced by the transport part of the network. It can be observed that the value of predicted MOS decreases, for a fixed BLER, when increasing the buffer length. This behaviour can be explained considering that, as stated above, the total MOS is calculated with the model described in Section IV, which considers as impairment factors on the service quality both the end-to-end delay and the packet loss rate: first term remains constant thanks to the power control that adjusts the BLER values to the target, while the second term increases when increasing buffer length, as shown in Table 1. Then, the overall predicted quality slightly decreases when buffer length increases. Moreover, higher values of target BLER, with a fixed value for buffer length, lead to a decrease on quality (Figure 5).

From the above considerations, it should be concluded that, fixing a target BLER, the preferable solution for the VoIP client architecture is a client without de-jitter buffer (in fact, the maximum MOS value is reached for 0ms buffer length). However, from a practical point of view, this conclusion should be emended taking into account also the values assumed by another parameter affecting the VoIP QoS: the mean number of playout interruptions.

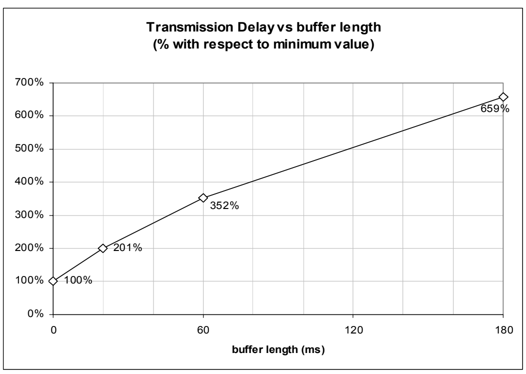

Figure 4 – Transmission delay vs buffer (BLER=2%).

As a matter of fact, we can see that this parameter takes very high values for de-jitter buffer lengths of 0ms and 20ms (more than 3000 interruptions have been observed) , while its value rapidly decreases for greater buffer lengths (see Table 3 and Figure 6), indicating an higher QoS level (it can be assumed that playout interruptions could cause a degrade on the perceived quality because they produce "silence spikes" during the conversation).

Therefore, considering the number of playout interruptions as an impairment factor for the overall perceived quality in the communication, it can be stated that the optimum value for the de-jittering buffer length in a VoIP system is 60ms for every considered target BLER. The graphical comparison between predicted MOS and number of playout interruptions, showing the trade-off in terms of quality can be seen in Figure 7.

Table 2 – Predicted MOS vs buffer, BLER.

MOS	Buffer 0ms	Buffer 20ms	Buffer 60ms	Buffer 180ms
BLER = 2%	4.04	4.03	3.99	3.91
BLER = 5%	3.35	3.32	3.28	3.21
BLER =10%	2.57	2.55	2.51	2.43

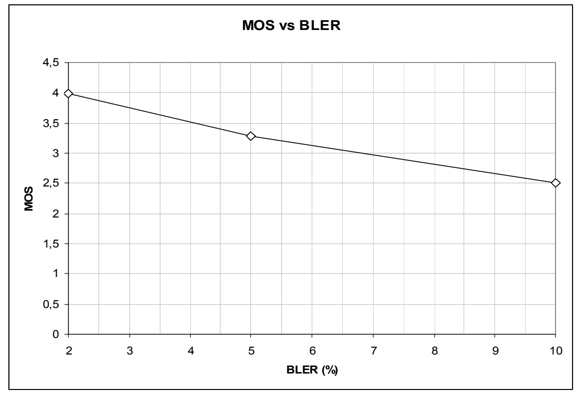

Figure 5 – MOS vs BLER (buffer = 60ms).

Table 3 – Mean number of playout interruptions vs buffer, BLER.

Playout interruptions	Buffer 0ms	Buffer 20ms	Buffer 60ms	Buffer 180ms
BLER = 2%	>3000	2688	73.5	19.2
BLER = 5%	>3000	2920	167.6	44.4
BLER =10%	>3000	>3000	301.3	83.5

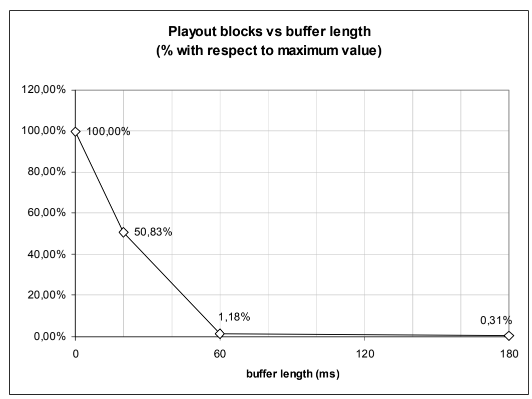

Figure 6 – Playout interruptions vs buffer (values expressed relatively to the maximum).

An example of cumulative distribution function for the MOS score, obtained with a buffer length of 60ms and a target BLER of 2% is shown in Figure 8.

Another effect capable of defining the overall QoS level of VoIP service is the time needed to setup a call. As far as the performances of SIP protocol are concerned, it has been observed that the performances are quite constant when varying target BLER and de-jittering buffer length.

Finally, the observed mean value of delay needed to establish the VoIP session (remember that SIP protocol is carried by a different RAB than the RTP traffic) is about 6.2 seconds; in Figure 9 the cdf (cumulative distribution function) related to this delay is shown.

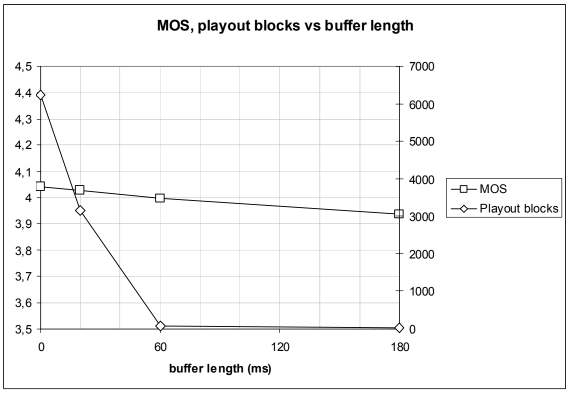

Figure 7 – MOS, playout interruptions vs buffer length (BLER=2%).

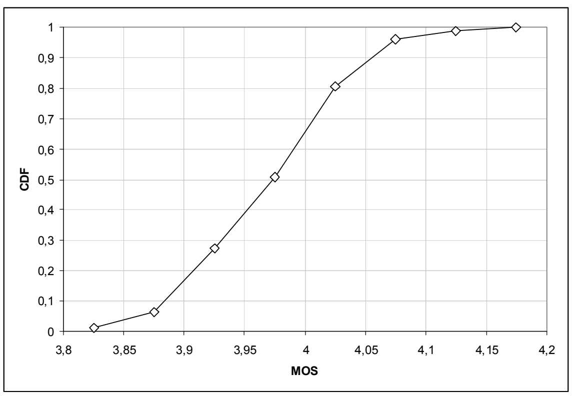

Figure 8 – CDF for MOS score (buffer=60ms, BLER=2%).

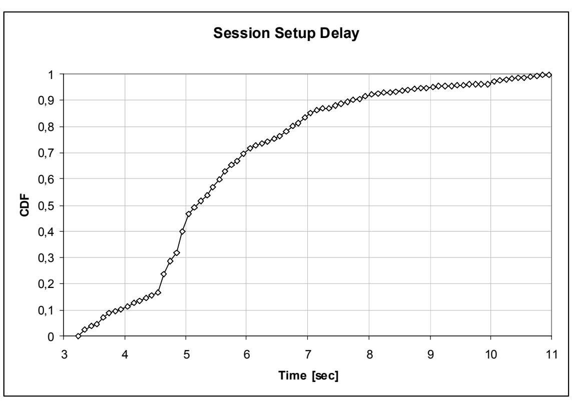

Figure 9 – CDF for session setup delay.

VII. Conclusions

Our analysis on the QoS level offered by dedicated transport channels in UTRAN R99 shows that it is possible to have a quality level of VoIP similar to the one obtained with traditional CS voice service, by tuning some of the main parameters that directly affect QoS. Particularly, considering the architecture described in Section V, it has been found that the most suitable value for target BLER at radio interface should be around 2% since such a value produces a predicted MOS greater than 3.5 (this value is considered as a "threshold" for having an acceptable quality level, see [8] and [14]). With regard to this matter, the carried out radio-link simulations show that, for the considered 16.8kbps conversational PS RAB, this BLER target value can be achieved with Eb/N0 of about 4.9 dB in uplink and Eb/N0 of about 5.2 dB in downlink.

Moreover, the optimum VoIP client architecture should foresee a de-jittering buffer length of 60ms (assuming a fixed buffer technique in order to compensate jitter effects).

Acknowledgements

This work has been performed within the framework of the IST-AROMA project [15] (www.aroma-ist.upc.edu), which is partially founded by the European Community.

References

[1] 3GPP TR 23.882: "3GPP System Architecture Evolution (SAE): Report on technical options and conclusions."

[2] B. Wang, K. I. Pedersen, T. E. Kolding, and P. E. Mogensen, "Performance of VoIP over HSDPA", IEEE VTC-Spring, volume 4, pages, 2335–2339, May 2005.

[3] M. Tao Chen Kuusela, E. Malkamaki, Nokia Research Center, "Uplink Capacity of VoIP on HSUPA", Vehicular Technology Conference, 2006. VTC 2006-Spring. IEEE 63rd

[4] E. Casilari, H. Montes, F. Sandoval, "Modelling of voice traffic over IP networks", 3rd Internacional Symposium on Communication Systems, Networks and Digital Signal Processing (CSNDPS 2002), Stafford (UK), pp. 411-414, July 2002.

[5] ITU-T P.59, "Artificial Conversational Speech".

[6] F. Hammer, P. Reichl, A. Raake, "Elements of interactivity in telephone conversations", *8th International Conference on Spoken Language Processing*, Jeju Island, Korea, October 2004.

[7] S. Rein, F.H.P. Fitzek, M. Reisslein, "Voice quality evaluation in wireless packet communication systems: a tutorial and performance results for ROHC", *IEEE Wireless Comm.*, February 2005.

[8] ITU-T G.107, "The E-model, a computational model for use in transmission planning".

[9] L. Carvalho, E. Mota, R. Aguiar, A.F. Lima, J. Neuman de Souza, A. Barreto, "An E-model implementation for speech quality evaluation in VoIP systems", *IEEE ISCC 2005*, pp. 933-938, Cartagena (Spain), June 2005.

[10] L. Atzori, M.L. Lobina, "Speech playout buffering based on a simplified version of the ITU-T E-model", IEEE Signal Processing Letters, Vol. 11, No. 3, March 2004.

[11] ITU-T G.113 – Appendix I, "Provisional planning values for the equipment impairment factor Ie and packet-loss robustness factor Bpl"

[12] R. Cuny, A. Lakaniem, "VoIP in 3G Networks: an End-to-End Quality of Service Analysis", Nokia Research Center, IEEE VTC May 2003

[13] 3GPP TR 25.993: "Typical examples of RABs and RBs supported by UTRA".

[14] A. Markopoulou, F. Tobagi, M. Karam, "Assessment of VoIP quality over Internet backbones", *INFOCOM 2002*, June 2002.

[15] P. F. Casadevall, R. Lung, A. Vega, A. Barbaresi, N. Nafisi, A. Gomes, L. Correia, "Overview of the AROMA Project", IST Mobile & Wireless Communications Summit 2006.