The Choice of VoIP Codec for Mobile Devices

Conference Paper · February 2014

3 authors:

Kamiar Radnosrati
Linköping University

Dmitri Moltchanov
Tampere University of Technology

Yevgeni Koucheryavy
Tampere University of Technology

The Choice of VoIP Codec for Mobile Devices

K. Radnosrati
Converging Networks Laboratory
VTT Technical Research Centre of Finland
Oulu, Finland FI-90570
e-mail: kamiar.radnosrati@vtt.fi

D. Moltchanov and Y. Koucheryavy
Department of Electronics and Communications Engineering
Tampere University of Technology
Tampere, Finland FI-33101
e-mail: {moltchan,yk}@cs.tut.fi

***Abstract*—As modern wireless access networks are moving towards packet based wireless access one may expect mobile cellular telephony to be eventually replaced by voice-over-IP (VoIP) applications. The choice of the codec in these applications is not straightforward as packet-based power-aware wireless communications bring new factors into the play. We study interdependencies between the bitrate, energy consumption, and the perceived quality provided by the voice codecs. We show that it is sufficient to equip a software with three codecs only. These are G.729.E, G.711.1 and G.723.1 codecs. Among those, G.723.E provides the best trade-off between the involved factors. When the system is overloaded and/or the power consumption is the most important metric (i.e., a mobile is running out of power) G.723.1 provides the best possible capacity and energy savings at the expense of significant quality degradation. Finally, when the system is underloaded while the amount of power spent for running the service is not important G.711.1 provides the best possible heard quality at exceptionally high power consumption.**

Keywords-VoIP, energy conservation, codec, perceived quality.

I. INTRODUCTION

With the current generation of cellular wireless access technologies offering a wideband packet-based access over the air interface it is expected that voice-over-IP (VoIP) applications will be responsible for most part of the voice traffic. Although the full transition has yet to be made due to slow uptake of IP multimedia subsystem (IMS), forcing operators to use intermediate solutions such as circuit-switching fallback (CSFB), it will happen sooner than later or the customers may start moving towards third-party VoIP applications. the topic of perceived quality evaluation of VoIP codecs has been the subject of the study in [1], [2], and [3] among others. However, the main aim of this investigation is to describe the effect of loss correlation etc.

However, in addition to finalizing the convergence towards a unified all-IP multi-service network, this transition brings additional challenges to software developers. One of the choices that we need to make when developing a VoIP software is the type of the codec to use. The choice of the voice codec is more complicated in wireless environment as in addition to perceived quality provided to the user, one needs to take into account additional additional factors, such as power consumption and bitrate. There are a number of reasons for that. First of all, the uptime of mobile devices depends on their battery power that is evidently not growing at the pace of communication technologies implying that the chosen codec must be as energy efficient as possible. Secondly, wireless technologies are more prone to occasional packet losses that may affect the perceived quality provided by codecs differently. Indeed, when a codec with high compression ratio is used the amount of bandwidth required from the network is minimized while the data flow becomes very sensitive to packet losses. Conversely, for low compression ratios the bitstream is less sensitive to packet losses while the amount of required bandwidth becomes significant. Recall that the compression ratio affects the amount of energy required for both compression and transmission. Further, although the bitrate of most codecs are fairly low compared to the available capacity of modern cellular technologies, minimizing it is still an important issue for network operators, especially, in densely population areas.

Finally, in those applications where the type of the codec is allowed to be changed on-the-fly we are interested which codec maximizes a certain characteristic that can be important for current operational regime of a mobile. For example, when the battery of a mobile is running out of power while the voice session is currently ”on” we are interested in maximizing the battery lifetime at the expense of slightly degraded quality. Indeed, for a given wireless access technology, a certain codec is characterized by a certain amount of power consumption required for transmission. At the same time, the amount of energy required for compression depends on the hardware configuration only. Using codecs with different compression ratio affects both components differently. We will see that for WLAN technology these components are comparable making the choice of the optimal codec less obvious. Considering the abovementioned interdependencies, choosing the appropriate codec minimizing energy consumption of a device and maximizing quality provided to the user is a complex task.

In this paper, we carry out an in-depth study of interdependencies between the perceived quality measured by the objective performance metrics, energy consumption spent for encoding and transmission, and bitrate of the codec in wireless environment. Both of these metrics are modulated by two factors that are conventionally assumed to be independent of each other. These are the type of the codec and loss behavior of a channel. Our major finding are as follows (i) for adaptive systems G.729.E, G.711.1 and G.723.1 are sufficient to cover all regimes of a mobile (ii) for conventional regime of a mobile

G.729.E provides the best trade-offs between perceived quality, total energy consumption and the required bitrate (iii) for high-quality service one needs to use G.711.1 at the expense of exceptionally high total power consumption, (iv) in energy saving regime G.723.1 provide some rather insignificant performance gains over G.729.E. Also, it is important to note that all the studied codecs, except for plain a/μ-law G.711, are characterized by similar response to the packet losses implying that the best codec after compression remains the best after any amount packet losses. Finally, for IEEE WLANs, under a certain choice of parameters, the amount of energy spent for encoding is comparable to transmission power implying that the choice of the optimal codec depends on a given technology. In particular, G.729.E is no longer the optimal codec when operating in IEEE WLAN environment.

The paper is organized as follows. In Section II, we introduce the QoE metric we use in this paper. In Section III, we numerically evaluate those trade-offs involved in our study. Discussion on the optimal choice of the codec is provided in Section IV. Conclusions are given in the last section.

II. Perceived quality metric

Quality of VoIP codecs is evaluated at the application layer using specific tests developed for assessing the perceived speech quality. To perform these tests a number of methods have been suggested in the past. We distinguish between subjective and objective tests. Those tests involving surveying humans are called subjective tests. Objective tests are based on deriving applications layer performance metrics based on network performance parameters. These tests try to provide the relationship between network performance and subjective QoE metric.

Subjective metrics assessing quality of voice communications are mostly based on the mean opinion score (MOS) scale. MOS provides numerical indication of the quality of the voice after compression and/or transmission. The value of MOS is a number ranging from 1 to 5 with 5 corresponding the the best possible quality. MOS is estimated by averaging the results of a set of subjective tests, where a number of humans grade the heard audio quality of test sentences.

The widely recognized objective metric for VoIP applications is defined in the so-called E-model standardized by ITU-T [4]. According to E-model the psychoacoustic speech quality is defined as a non-linear additive function of different impairments. The measure of the quality is called an R-factor which is given by

$$R = R_0 - I_s - I_d - I_e + A, \tag{1}$$

where R_0 represents noise and loudness in terms of the signal-to-noise ratio at 0dBr point, I_s accounts for impairments occurring simultaneously with speech, I_d represents impairments that are delayed with respect to speech, I_e is the effect of special equipment, A is the advantage factor. Simply put, I_d is the delay of a packet, encoding impairments are included in I_s, while the compression and network losses are in I_e. The advantage factor accounts for special environments, where a user may sacrifice the quality with respect to availability of the service. The value of R-factor varies in between 0 and 100.

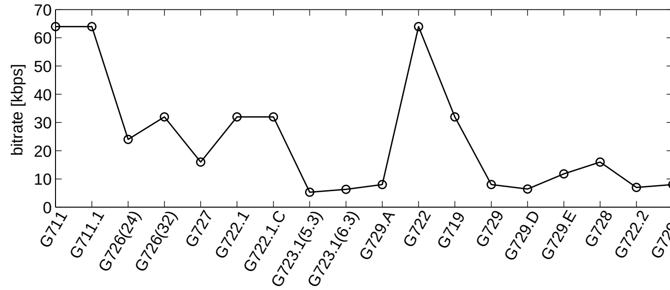

Figure 1: Raw bitrates of codecs.

Parameterizing the model we see that the advantage factor should nowadays be set to 0 as users get accustomed to wireless voice services. Following the work of Clark in [5], the highest possible value of $R_0 - I_s$ is set to 94 resulting in reduced expression $R = 94 - I_d - I_e$ setting the upper bound on the perceived quality. VoIP routes in the Internet are usually provisioned such that the end-to-end delay impairment factor, I_d, is less than the maximum tolerable delay ($150 - 200$ms, [7]). In this case the quality of speech transmission is dominated by I_e, i.e., $R = 94 - I_e$. The effect of I_e has been found using extensive subjective tests.

The performance of E-model was shown to correlate well with MOS grades under assumption of independent packet losses. In wireless networks, there are various mechanisms trying to remove the memory of the channel (e.g., interleaving). In wired Internet the major source of memory is droptail queuing. However, random earlier detection (RED) is gradually replacing droptail in the wired Internet making the packet loss process uncorrelated. Thus, the packet loss process in wireless-cum-wired configuration can be considered memoryless implying that there are no significant grouping of packet losses. It is important to note that this assumption can be relaxed whenever appropriate. For our work taking into account the effect of loss correlation would result in unnecessary increase of complexity and may hide the main message of the study.

III. Interdependencies

A. Rate requirements

Raw bitrates of codecs are shown in Fig. 1. It should be noted that some codecs have variable bit rates. For such codecs, we show only one of their data rates and the corresponding bandwidth. All the metrics we consider in what follows are calculated with respect to IP packet, i.e., energy consumption is expressed as mW per IP packet while rate is in IP packets per second.

The actual amount of data generated by a voice codec per sampling interval can be represented as $S = H + P$, bytes, where H is header size and P is the payload of voice packets. Further, denoting by the R the number of packets emitted per second and by the bandwidth, B, required for transmission is calculated as $B = S*R$ KBps. Thus, to estimate the bandwidth of a codec we need to know how much overhead, H, is added to the payload of a codec. When compressed real-time transport protocol (cRTP) is not used the IP/UDP/RTP headers amount up to 40 bytes due to the following components (i) IP header, 20 bytes, (ii) UDP header, 8 bytes, RTP, 12 bytes. In

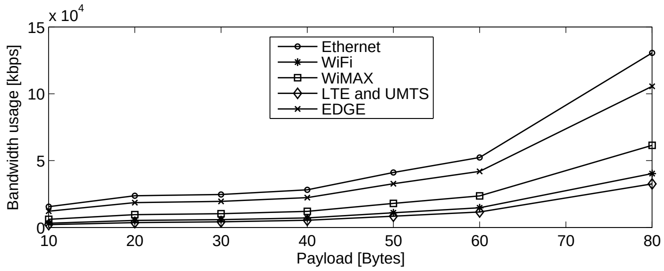

Figure 2: Bandwidth usage of audio codecs.

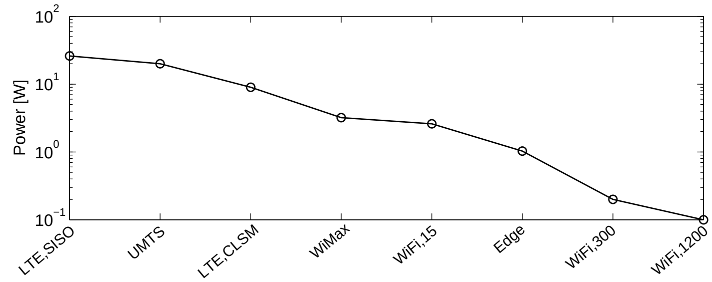

Figure 3: Energy consumption of transmission technologies.

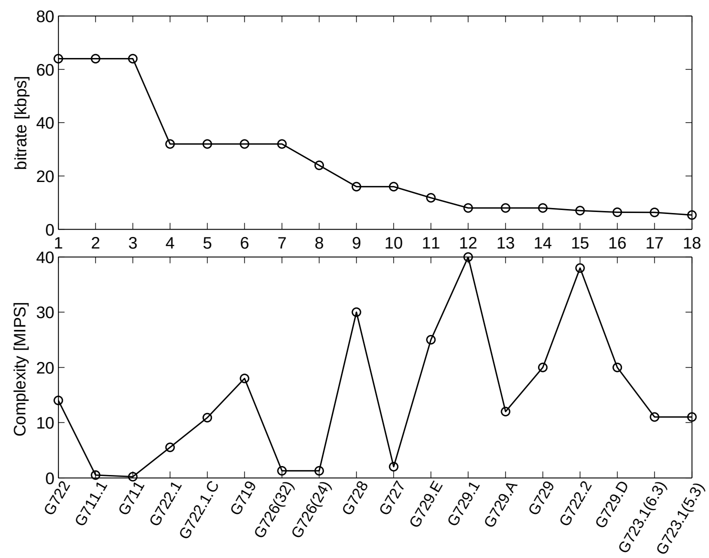

Figure 4: Complexity and bitrate for different audio codecs.

those cases, when cRTP is used, the 40 bytes overhead reduced to just 2 bytes. In case of Ethernet, there will be additional 18 bytes that includes frame check sequence (FCS) and cyclic redundancy check (CRC) headers.

Fig. 2 shows the differences in bandwidth usage for different wireless access technologies and Ethernet. The payload size for these codecs changes between 10 and 80 bytes with step of 10 bytes. We see that the Ethernet imposes the highest bandwidth usage. Further, observe that the bandwidth requirements for all wireless access technologies presented here grow exponentially fast as the amount of payload increases. Out of all considered technologies EDGE requires the highest amount of bandwidth. The amount of bandwidth required by Wi-Fi and and Wi-Max technologies is comparable. LTE and UMTS requires the same amount of bandwidth and introduce the minimum overhead.

B. Energy consumption

1) Transmission energy: In wireless communications, the transmission power is different in different modes of operation. Conventionally, we distinguish between idle, sleep, transmission, and reception states. Here, we are interested in two of them, namely, transmission and reception states. Notices that it has been shown that that energy consumption in idle and receive states are almost the same [8]. The only difference between them is the amount of power spent by amplifying the received signal in receiving states. However, usually this energy is significantly smaller compared to that one required for transmission. Thus, we concentrate on the power consumption in transmission mode only.

Fig. 3 presents the power consumption measured in mW per packet for different wireless access technologies in logarithmic scale. As one may observe the difference between power consumption of wireless access technologies available today could be as high as two orders of magnitude proving the importance of choosing an appropriate radio interface and transmission technology one-the-fly. Another observation is the importance of choosing codecs. Indeed, different codecs produce their outputs in a wide range of data rates. Particularly, as we already observed the raw data rate ranges from 5.3Kbps for G.723.1 to 64kbps for G.722 or G.711.

2) Encoding energy: The energy spent for encoding varies with the type of the codec and its special features. Unfortunately, the actual energy depends on the type of a digital signal processor used for encoding. One way to provide hardware independent estimates of the encoding complexity is to calculate the amount of operations required for encoding. Fig. 4 shows raw bitrate and complexity of voice codecs measured in millions operations per second (MIPS), where codecs are sorted in descending order of their bitrates. We see that the general trend is increase of the complexity in response to smaller bitrates. At one extreme there are G.711 and G.711.1 codecs having 64Kbps raw rates and requiring very small processing power (0.01mW for G.711 and 0.025mW for G.711.1 on C55x processors family). There are exceptionally complex codecs such as G.729.1 and G.722.2 requiring low raw rates. However, there are some exceptions, such as G.727, G.729 Annex A, and G.723.1 codecs characterized by rather low rates and moderate encoding power consumptions.

Investigations done in this paper show that the range of encoding energy for these codecs varies from below 1mW (G.711 with C55X) per packet to something around 12mW in some cases. Encoding power consumption for several processors is shown in Fig. 5. Note that for some codecs the amount of power required for encoding is comparable to the amount of power required for wireless access technologies. For example, the most complex codec G.729.1 running at C54x architecture requires a power of approximately 12mW, which is more than the amount of energy required for transmission

 ISBN: 978-1-61208-318-6

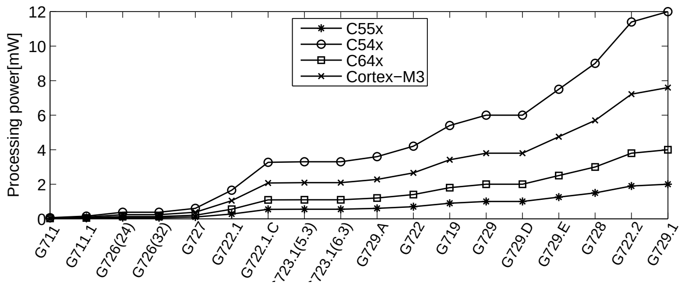

Figure 5: Encoding power consumption of voice codecs.

over LTE in CLSM regime (approximately 9mW). On the other hand, low complexity codecs, such as G.711, G.711.1 requires exceptionally small amount of energy at all platforms. Note that the choice of the processor, plays an extremely important role and the difference in the encoding power could be as high as several mW. This difference becomes bigger as we go from low complexity codecs to the high complexity ones.

C. Perceived quality

Recall that the reduced E-model is given by $R = 94 - I_e$. The only unknown we have is I_e, which is the effective equipment impairment factor taking into account the effect of voice compression and network losses. I_e values for a number of considered codecs are summarized in Fig. 6 [6]. These values represent the perceived quality after compression and do not take into account the effect of packet losses introduced by the transmission medium. Recalling that R factor 94 is the maximum possible value achieved with G.711 while 70 is the minimum acceptable one the set of codecs available today provides the perceived quality across the whole range of acceptable quality.

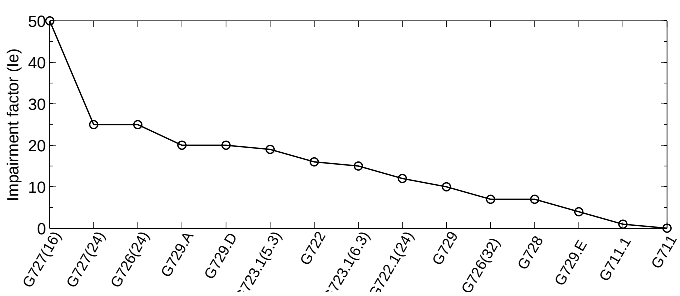

Figure 6: Performance degradation introduced by encoding.

Fig. 7 shows the values of impairment factors for a number of codecs for different values of the packet loss ratio (PLR, measured in percents). As one may observe almost all impairments factors I_e are linear functions of PLR. One of consequence of this behavior is that if the perceived quality is the only metric of interest then the choice of the best codec is independent of the packet loss ratio. In other words, the codec providing the best performance in absence of losses will remain the best one for any value of PLR. The only exception is G.711 codec without packet loss concealment feature whose performance severely degrade when PLR increases.

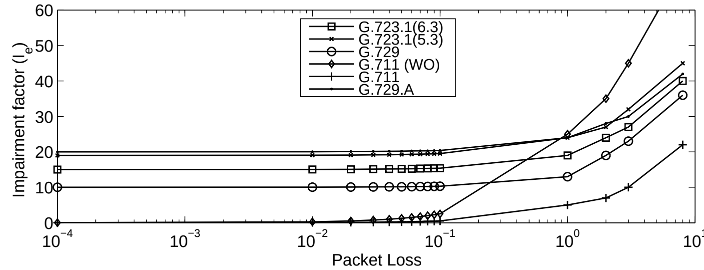

Figure 7: Values of I_e factor for different packet loss ratio.

IV. OPTIMAL CHOICE OF THE CODEC

Consider how the amount of energy spent for compression is related to the power required for transmission. We would like to check whether there are special codecs minimizing the overall power consumption. The amount of power for different processor families required for encoding and transmission is shown in Fig. 8 and Fig. 9. As one may observe the transmission power for codecs is generally way larger than the energy required for compression. One special exception is Wi-Fi access technology operating with high TTI values, where these two sources of energy consumption are comparable for some codecs. Thus, in VoIP applications, for most wireless technologies transmission power dominates the overall power consumption.

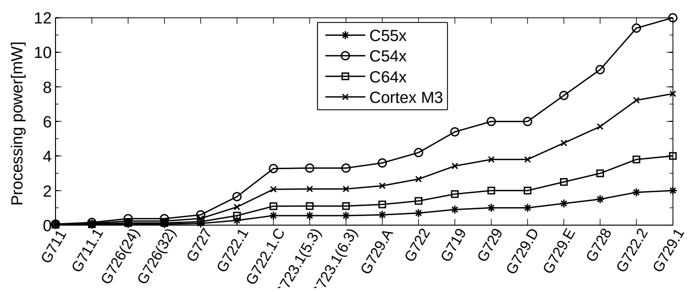

Figure 8: Encoding power consumption.

So far, we have seen that the trade-offs between perceived quality and energy required for running the service can be complicated. Putting all of these parameters together and viewing the results in one figure would be helpful. Fig. 10 shows total power consumptions of voice codecs per one second interval for C54x processor architecture (recall, that out of all considered platforms C54x requires the most power for encoding). Corresponding I_e values and R factors after compression are shown in Fig. 11. Assume for a moment that the bitrate of the codec is not a concern, i.e., the channel capacity is large enough to accommodate the one with the highest bitrate (G.711 or G.711.1). Even in this case, the choice of the codec optimizing both the total energy consumption and the perceived quality is still non-trivial. If one is targeting the best possible perceived quality G.711.1 is the obvious choice providing the maximum possible value of R-factor after compression. However, the amount of power spent for encoding and transmission is extremely high amounting to approximately 1250mW for UMTS and 1100mW for LTE CLSM. Energy requirements of G.711 is significantly smaller

 ISBN: 978-1-61208-318-6

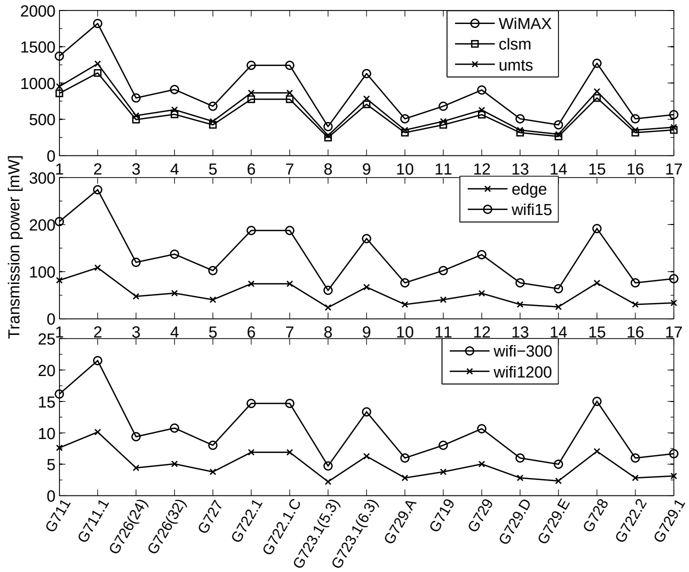

Figure 9: Transmission power consumption.

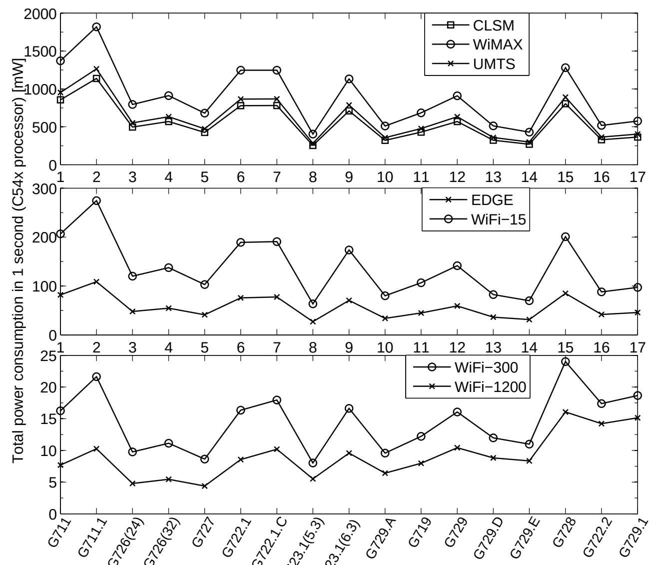

Figure 10: Total power consumption for C54x.

(approximately 750mW for LTE CLSM and 900mW for UMTS) while the perceived quality is kept at the same level. Using G.711 instead of G.711.1 in loss-free environment would allows for save 350mW for both technologies. Assuming the average length of a conversation being equal to 300 seconds (5 minutes) it would result in approximately 100W energy savings. This implies additional 140 seconds of a VoIP call over LTE CLSM or approximately 116 seconds for UMTS VoIP call. However, these are not the best possible energy savings one may achieve. G.729.E codec providing R-factor of 90 while requires around 300mW of energy operating in LTE CLSM and UMTS networks. These significant additional energy savings compared to G.711 (approximately 450mW for LTE CLSM and 600mW for UMTS) comes at just slight decrease of quality (R-factor 90 compared to 94 for G.711). Further, taking into account the the data rate of G.729.E codec is just 12.5Kbps this codec is far superior compared to both G.711 and G.711.1. Energy consumption of of G.729.E codec is comparable to G.726 operating at 32Kbps data rate except for slightly worse R-factor (87 instead of 90). When energy is the most important metric (i.e., power of a mobile is running out) the best possible codec is G.723.1 operating at 5.3 Kbps data rate providing R value 75. This codec is especially useful for those wireless technologies characterized by small transmission energy requirements, i.e., IEEE 802.11 WLANs. However, as one may notice G.726 codec operating at 32Kbps provides significantly better quality (R-factor 87 compared to 79 for G.723.1). For energy saving regime the latter is advisable.

As we already highlighted, Wi-Fi operating with TTI 1200 is a special example of a wireless technology, where the energy consumption for compression is compared to that required for transmission. Observing Fig. 10 we see that the total power consumption for C54x processor family and Wi-Fi with TTI 1200 has a behavior different from other technologies. In fact, G.729.E codec is no longer the one providing one of the best trade-offs between the perceived quality and total power

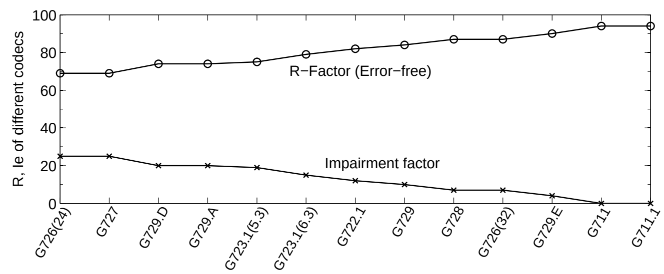

Figure 11: I_e and R-factor after compression.

consumption for zero PLR. G.711 codec outperforms G.729.E in terms of total power consumption providing better perceived quality (7.5mW and R value 94 instead of 10mW and R value 90 for G.729.E). Moreover, G.726 codec operating at 24Kbps is even more (almost twice) energy efficient providing 5.5mW of total power consumption while delivering R value of 90. Note that similar conclusions are true for Cortex-M3 family while C64x family is closer to C55x performance with transmission energy dominating the total power consumption (not shown here). The choice of the optimal codec is more complex when the amount of energy for transmission and compression is comparable.

So far, we considered the trade-off between the amount of energy required for running the service and the perceived quality provided to the user. These factors are often enough when the system is well below its capacity limits. However, when it is about to overflow one also need to take into account the rate requirements of codecs as slight overload may lead to extreme quality degradation for all the users. When choosing the best possible codecs for these conditions one needs to take into account three factors simultaneously: rate requirements, perceived quality, and energy consumption. Taking another look at Fig. 10 and recalling data from Fig. 1 we see that G.729.E is still the best codec optimizing

 ISBN: 978-1-61208-318-6

these three parameters simultaneously (12.5 Kbps, R value 90, 400mW LTE CLSM and UMTS). G.728 operating at 16 Kbps, is characterized by comparable R value (87) but requires significantly more energy (approximately, 800mW and 900mW for LTE CLSM and UMTS for C54x, respectively). Similarly, G.726 operating at 32Kbps has slightly higher power requirements and comparable R value but requires approximately three times more bandwidth. When the system is severely overloaded G.723.1 operating at 5.3 Kbps is the best possible choice from the rate requirements perspective (more than twice better than G.729.E). However, its perceived quality is close to unacceptable (R value 79).

Consider now what happens when the PLR increases. Fig. 12 shows the values of R-factor and corresponding values of MOS for six selected codecs (computed according to the closed-form expression provided in [9]) for a number of codecs for different values of the packet loss ratio. Surprisingly, the choice of the VoIP codec jointly optimizing the considered three factors (rate/energy/quality) is independent of the value of PLR as response of the considered codecs to PLR is qualitatively and quantitatively similar. One exception is the very special behavior of G.711 codec without PLC capabilities whose R value decreases exponentially fast as PLR increases. This codec should never be used in lossy environments such as wireless access. Also, as one may observe, there is an intersection between lines corresponding to G.729.A and G.723.1 (5.3 Kbps) codecs, i.e., up to PLR of approximately 2.5% G.723.1 performs slightly better than G.729.A, while for higher value of PLR G.729.A outperforms G.723.1. This implies that if ones originally uses G.723.1, as PLR increases one needs to change to G.729.A. However, the region where G.729.A outperforms G.723.1 is below MOS 3.5, which is widely accepted as the minimum acceptable quality. Thus, in most cases, the choice of the best codec for non-zero value of PLR (after compression) coincides with the chose made for any non-negligible PLR.

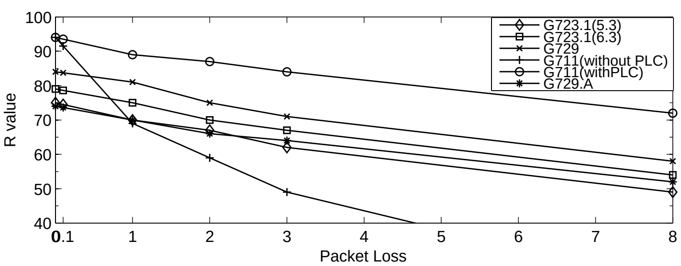

Figure 12: Values of R-factor for different packet loss ratios.

V. CONCLUSIONS

In this work, we examined trade-offs between quality of user experience, compression, and energy consumption for VoIP applications in wireless environment. As opposed to many studies exploring ways to minimize energy consumption of mobile devices we concentrated on the time period when a media application is up and running and studied the way how decrease the amount of energy required for running VoIP service while maintaining the best possible quality provided to the use. This study was motivated by availability of multiple codecs for voice and video information characterized by a wide diversity of compressed data rates and compression algorithms.

Summarizing, we note that the choice of the VoIP codec jointly optimizing the considered three factors (rate/energy/quality) is rather straightforward with only small deviations in the special cases. The reasons are (i) independence of the choice of the codec from packet losses (ii) domination of transmission energy requirements in total power consumption for most considered technologies (iii) one codec being significantly superior than others (G.729.E). From energy/quality/rate joint optimization point of view there are a number of obsolete codecs that are always worse compared to other. These are G.726 (24 Kbps and 32Kbps), G.727, G.729.D, G.729.A, G.723.1 (6.3 Kbps), G.722.1, G.729, G.728, G.711 (without PLC). The only codecs that need to be implemented to optimize the the abovementioned three parameters for any operational regime of a mobile are: G.729.E, G.711.1 (PLC mode), G.723.1 (5.3Kbps). The first codec, G.723.E, provides best possible performance for conventional regime of a wireless ensuring the best possible trade-off between the rate requirements (12.5 Kbps), power consumption (350mW LTE CLSM, UMTS), and the perceived quality (R value 90). When the system is severely overloaded and/or the power consumption is the most important metrics (i.e., a mobile is running out of power) G.723.1 operating at 5.3 Kbps provide the best possible capacity and energy savings (just 300mW for LTE CLSM and UMTS) at the expense of significant quality degradation (R value 75, which is close to the lowest possible quality level). This codec can only operate under zero PLR as any non-zero value of PLR immediately make the heard quality unacceptable. Finally, when the system is underloaded while the amount of power spent for communication is not important G.711.1 with PLC capabilities provides the best possible heard quality at exceptionally high power consumption.

REFERENCES

[1] M. Goudarzi, L. Sun, and E. Ifeachor, "Modelling Speech Quality for NB and WB SILK Codec for VoIP Applications, " Next Generation Mobile Applications, Services and Technologies (NGMAST), Cardiff, UK, 2011, pp. 42-47

[2] A. Marzuki, Y. Chai, H. Zen, L. Wee, K. Lias, and D. Mat, "Performance Analysis of VoIP over 802.11b/e using different codecs, " ISCIT 2010

[3] M. Aamir and S. Zaidi, "QoS analysis of VoIP traffic for different codecs and frame counts per packet in multimedia environment using OPNET," INMIC 2012

[4] G.107, "The E-model, a computational model for use in transmission planning", ITU-T Recommendation, 2003.

[5] A. Clark, "Modeling the Effects of Burst Packet Loss and Recency on Subjective Voice Quality," 2nd IPTel Workshop, 2001, pp.123-127.

[6] G.113, "Transmission impairments due to speech processing", ITU-T Recommendation, 2007.

[7] G.114, "One-way transmission time", ITU-T Recommendation, 1996.

[8] Y. Li, M. Reisslein and C. Chakrabarti, "Energy-Efficient Video Transmission Over a Wireless Link," IEEE Tran. Veh. Tech., vol.58, no.3, 2009, pp. 1229-1244.

[9] H. Assem, D. Malone, J. Dunne, and P. O'Sullivan, "Monitoring VoIP call quality using improved simplified E-model," Proc. ICNC, Jan. 28-31, San-Diego, USA, 2012, pp. 927-931.

 ISBN: 978-1-61208-318-6