Some Experiences with VoIP Over Converging Networks *

Pekka Perälä, Martín Varela†
VTT Electronics
PL1100, FI-90571, Oulu, Finland
{pekka.perala,ext-martin.varela-rico}@vtt.fi

May 14, 2007

Abstract

Over last few years an increasing number of multi-access devices have emerged, which enable users to stay connected almost everywhere, and in many cases very cheaply. It is likely just a matter of time until VoIP capabilities become ubiquitous in these devices (be it in the form of applications, or in a transparent mode such as the one used in UMA phones), and since they run on heterogeneous networks, with different performance issues, the quality experienced by the users will likely be very variable. In this paper we present some measurements performed with Skype in converging network environments.

1 Introduction

The emergence of multi–access devices, coupled with the wide availability of 3G cellular networks and wireless access networks (some of which are free to use), make the use of VoIP an interesting option for the user on the go. Many newer mobile devices come with SIP-enabled VoIP clients, and some even allow to run popular applications such as Skype. This this could very well create a wide user base for mobile VoIP solutions.

Furthermore, the availability of relatively cheap unlimited-data plans from many cellular operators might build a stronger case for using VoIP instead of normal calls in many situations. Cellular operators and equipment manufacturers are also interested in using the multi–access capabilities of new terminals in order to offload some of their cellular users onto WiFi networks, for instance (e.g. Unlicensed Mobile Access, or UMA [1]), in order to reduce operational costs.

In this paper we will address the quality aspects of these services. That is, how well does for instance Skype, work in a mobile environment. This mobile environment could be a notebook equipped with a UMTS (Universal Mobile Telecommunications System) or HSDPA (High–Speed Downlink Packet Access) card, using Mobile IP to roam between WLANs and a 3G network, or it could be a mobile phone running Skype, either over WiFi or 3G.

We have performed several tests on these scenarios at VTT's Converging Networks Laboratory [2], two public operator's 3G (UMTS) networks, and the city-wide PanOULU WiFi open network, which provides free Internet access in the city of Oulu (the network has over 500 access points scattered over the city and the university campus), as well as internal research LAN and WLANs. In this paper we present our experiences with the perceived quality of VoIP (as provided by Skype) on these mobile contexts. We chose Skype, since its wide availability and huge user base make it a very likely option as VoIP software for people on the go.

The rest of this paper is organized as follows. Section 2 describes the experimental setups that were used. In Section 3 we present and discuss the results we obtained for

*M. Varela's work was carried out during the tenure of an ERCIM post-doctoral fellowship. Part of this work was funded by the ITEA Easy Wireless project.

†Corresponding author.

the QoS. Finally, we conclude the paper with Section 4.

2 Experimental Setups

We performed experiments in two basic scenarios. The first one consisted of using Skype in a Windows based smartphone (HTC TyTN), both over WiFi and over 3G. The second scenario was also based on Skype, but this time running on a notebook, with Mobile-IP [3] support, and roaming between wired, WiFi, and 3G connections.

2.1 Smartphone Scenario

The first scenario we considered was that of a smartphone user, who might use the multi-access capabilities of his device to maintain connectivity on the go. We used a device based on Windows Mobile, since there is a native Skype client for the platform. Other devices might as well be used with Skype or even other widely used VoIP services (such as GTalk, MSN, etc.), for example by using a gateway service like Fring, which has Symbian (S60) clients.

In order to generate the samples, we used an acoustic coupling between the device's headphones and microphone, since an internal mixer was not available. We then played a series of four studio samples in French, German and English in a loop using the device's media player, and used this as an input to the Skype client. This client was in an active call with another Skype client residing in VTT's company network. The degraded samples were recorded at the other end using a sound hijacking program.

The resulting degraded samples were then manually time–aligned to the original ones, since there were some instances of temporal compression, or some pauses in the stream. The alignment was performed by either adding or suppressing silent intervals to the degraded stream, so as to keep both as synchronized as possible. No speech was cut, but there was clipping in some cases, which led to slight synchronization issues.

The samples were then processed with a sliding window (5 or 10 seconds, in most cases), with steps of 1 or 0.5 seconds, and then each segment was assessed with PESQ [4]. The different window and step values were used to compare the PESQ assessments to what the subjective (as per one expert listener's opinion) quality was. We found that the scores provided by PESQ varied significantly with the length of the time window, and especially with the amount of silence that fell within it. The smoothest results, and those we found closest to the actual quality were the ones with a larger window size. We are currently considering extending these tests with PSQA [5], to assert whether we can get more consistent results than with PESQ, and to develop a real–time monitor.

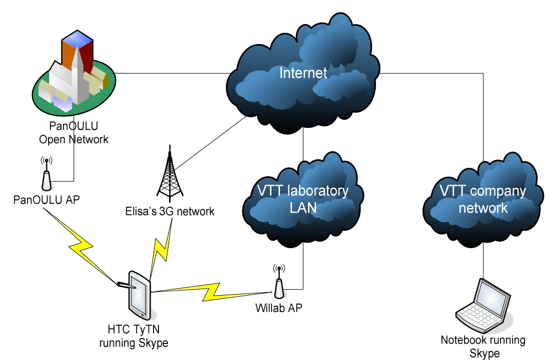

Figure 1: Experimental setup for the smartphone scenario. At any given time, the device was connected to only one of the three available networks.

The tests were performed over three different connections. We used two WLANs, the PanOULU network, which is a city–wide free access WLAN available in the city of Oulu, and the Willab WLAN, which is an internal research network. We also used a public 3G network from Elisa, a Finnish phone operator. In all cases we noticed significant variations in quality during testing. This is consistent with some very reduced scale living–lab testing we have performed, during which the quality of calls ranged from above toll to unusable, both over WLAN and 3G. There were slight delay issues in some cases, but since we didn't measure the actual delays, we didn't take those into account in this study. Figure 1 depicts the architecture used for this scenario.

2.2 Mobile IP Scenario

In the second measurement scenario we performed vertical handovers (HO) over three access networks, while having a Skype call active during the handovers. In addition to Skype, the mobile node (a notebook) client was also running Mobile IP (MIP) client software, which took care of performing the handovers and choosing always the best available network. The mobile node was on a foreign network, as would be a mobile user on the go. There were three access networks used: VTT's Laboratory LAN, the the PanOULU public WLAN and TeliaSonera's 3G/UMTS network, which is a commercial Finnish cellular network. Figure 2 illustrates the experimental setup used.

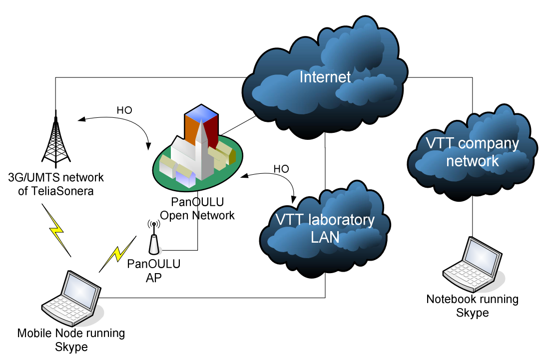

Figure 2: Experimental setup for the Mobile IP scenario. The MIP client chose the optimal connection at any given time.

The tests were performed in a similar way as those of the previous measurement scenario, but in this case, a notebook was used as the mobile node and a wired connection was used instead of a WLAN AP in order to connect to VTT's laboratory LAN. In these tests, the acoustic coupling between microphone and headphones was replaced by an internal sound mixer, in order to avoid any quality degradation due to the former.

During the tests, the LAN and WLAN connections were deliberately made unavailable during speech segments, so the effect of the handovers would be most noticeable. Also, as well as in the previous scenario, the recordings were done on the uplink, which is useful for setting a lower bound to the performance on the 3G network (since there is usually more bandwidth available for the downlink).

3 Quality Results

In this section we present a representative sample of the results we obtained. If one word had to be used to describe the perceived quality (both in terms of objective and subjective experience), that word would be *variable*. Both in the formal testing, as well as in our reduced living–lab tests we found that the perceived quality ranged from fairly good to unusable (sometimes within the same test run), in all scenarios.

3.1 Results from the Smartphone Scenario

The results we obtained with the smartphone were in line with our expectations. While the system worked, the quality tended to be quite low most of the time. On the other hand, there is an *advantage factor* to take into account when assessing these results, since running VoIP applications on multi–access devices drastically reduce costs for the user, while increasing availability. While the extent of this *advantage factor* is out of the scope of this study, it will probably have a big effect on the overall assessment.

Most of the calls we recorded and processed for this scenario were slightly below toll quality. However, the performance wasn't generally an insurmountable obstacle to effective communication. Also, the performance recorded for the WLANs tended to be better than that of the 3G network.

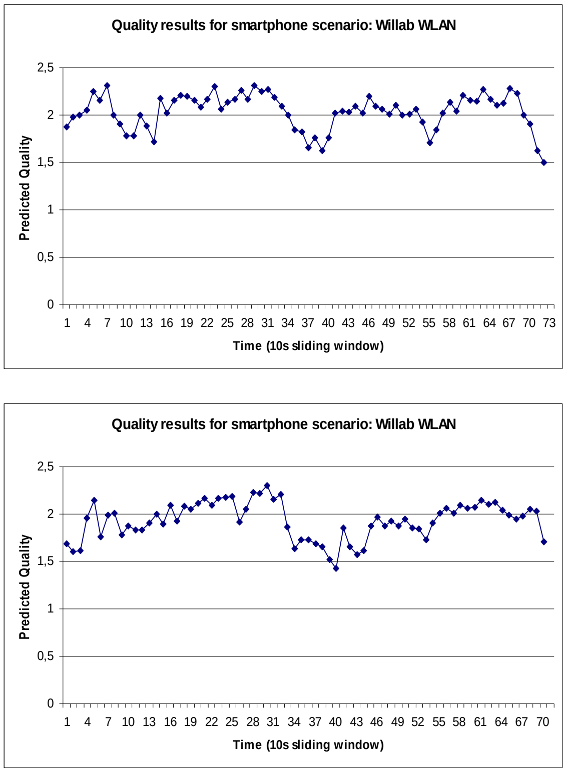

Figure 3: Quality results for the smartphone scenario. These calls were made over the Willab WLAN, which is used for research purposes, and quite active.

Figures 3 and 4 show typical results for the experiments running over WLANs. As mentioned above, the overall quality was slightly below toll quality in most cases, though some noticeable glitches appeared from time to time.

As for 3G, the quality in these cases was mostly lower than when using a WLAN access. The speech was still mostly intelligible, although there were some points at which it became garbled, or parts went missing. Figure 5 shows results for measurements done several months apart over the Elisa UMTS network. While these results do not look very encouraging, there are two issues to take into account. Firstly, flat rate 3G data plans are

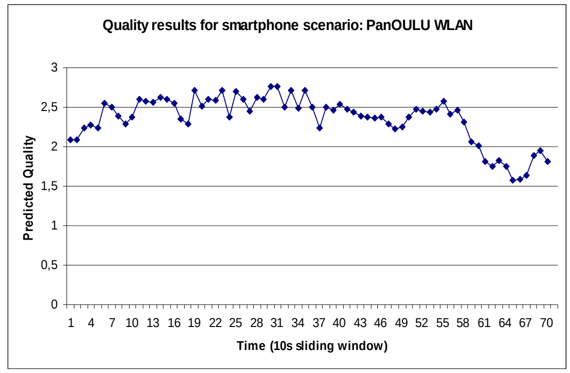

Figure 4: Quality results for the smartphone scenario. This call was made over the PanOULU city–wide open network.

becoming cheaper, and many people are adopting them. Secondly, commercial HSDPA networks are already operational, and HSDPA devices are becoming increasingly available. It seems reasonable to expect that within a few years' time, the results obtained will be satisfactory for most users.

3.2 Results from the Mobile IP Scenario

An important part of this work was to measure the effect of MIP–originated handovers on the perceived listening quality of the VoIP stream. While the MIP client allows for selection of the best available connection at any time, the handovers are not seamless, and there is a small connection break each time the access network is changed. These breaks are generally below 2s in length [6], but they are definitely noticeable.

Another interesting finding, is that Skype seems to have sometimes problems when roaming back into the LAN. The issue was not always present, and indeed in some cases connecting back to the LAN was barely, if even noticeable, as the connection break tends to be shorter in this case. There were, however, some instances in which roaming back into the LAN produced huge "oscillations" in the perceived quality, instead of settling in a (expectedly good) quality level. At the time of

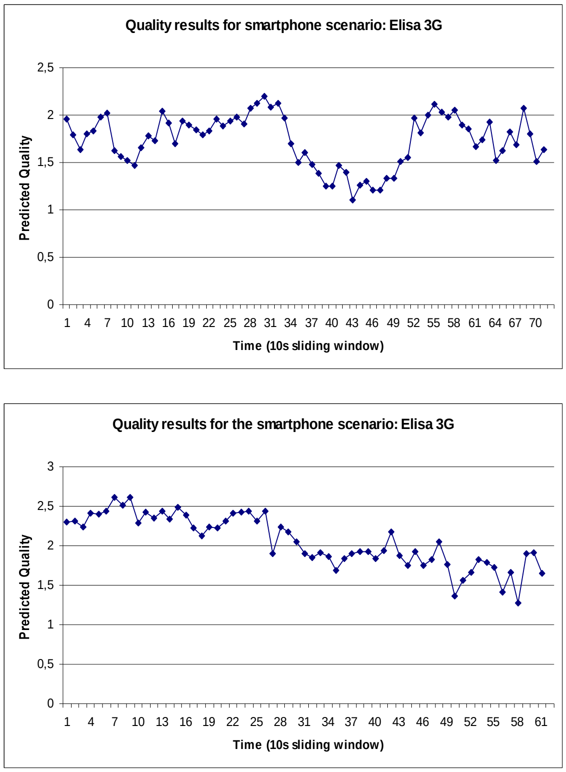

Figure 5: Quality results for the smartphone scenario. These calls were made over the Elisa public 3G (UMTS) network.

writing we do not know the exact cause for this behavior, but it seems likely that it might be related to the way in which Skype adapts to variations in available bandwidth (said variations in these cases may be, in practice, of one order of magnitude or more).

Figure 6 provides a baseline for the performance we observed while connected to our lab's LAN. In general the perceived quality was good, although there were some slight fluctuations, and some below–toll quality at the end of the trace.

Figure 7 shows two quality traces in which two handovers were performed, first LAN-WLAN, and then

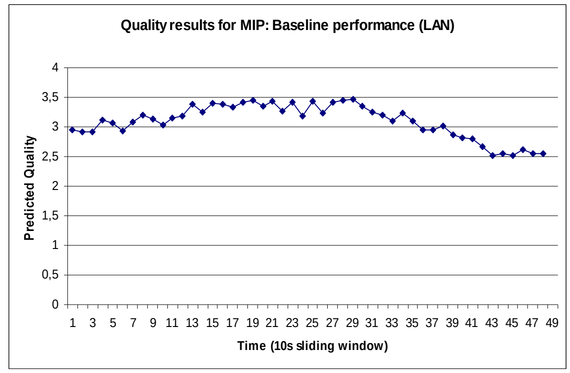

Figure 6: Quality results for the Mobile IP scenario. In this example, no handovers were performed, and the notebook was connected to a LAN.

WLAN-LAN. In both cases, the variation in the quality was not significant. In the recorded audio files the handovers are evident, but otherwise the quality remains stable on both networks.

As in the smartphone scenario, the voice quality obtained over 3G was generally worse than when using either of the other access media. For this scenario we used TeliaSonera's UMTS network, and the performance levels were similar to those obtained with Elisa's network and the smartphone. In Figure 8, this difference can be appreciated. During the 3G period, the quality drops about 1 MOS point, to a level where it is almost unusable.

In Figure 9, the "oscillations" in quality mentioned above, that sometimes happen when roaming back into the LAN can be observed. During approximately 20s, the quality increases and decreases erratically within 1 MOS point. In the recorded trace this effect is particularly annoying. An even more drastic variability in the recorded quality is shown in Figure 10. During that test, the quality was consistently bad across all access networks, which suggests that some intermediate node was being used at the time, and the network was congested somewhere along the path to it. In this case, given the low overall quality and the high variability, the system was practically unusable. This was the worst example we got

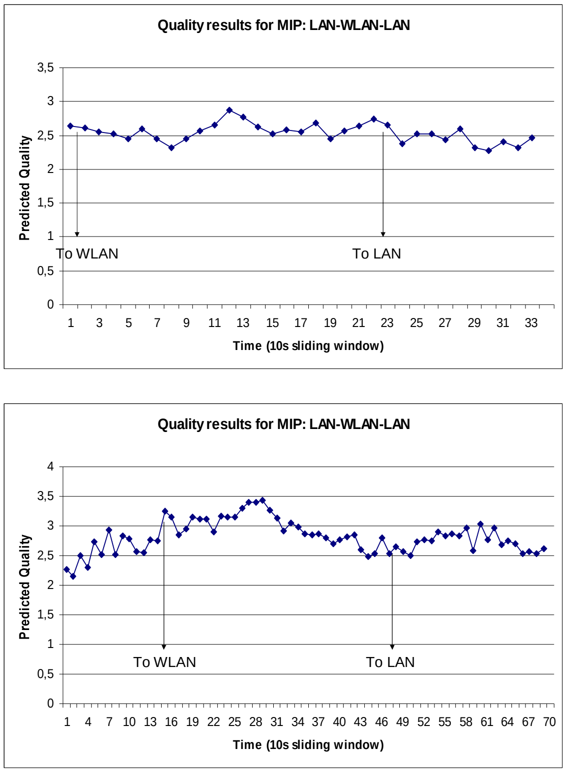

Figure 7: Quality results for the Mobile IP scenario. In these examples, two handovers (LAN-WLAN-LAN) were performed. The quality variations in these cases were relatively small.

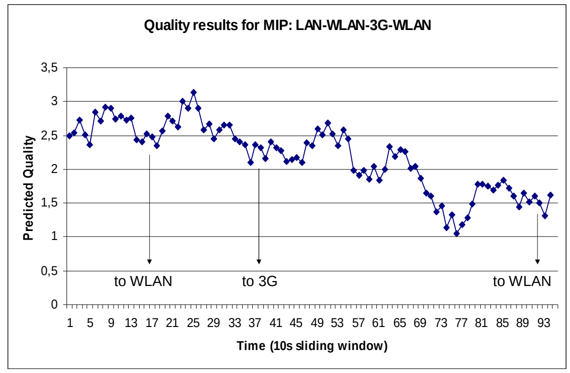

Figure 8: Quality results for the Mobile IP scenario. Three handovers were performed in this example. Notice how the 3G performance is significantly worse than that of any of the other two access media.

while performing these tests.

4 Conclusions

In this paper we have studied the performance, in terms of listening quality, of Skype when used in a converging networks context. To this end, we used multi–access devices, both with and without mobility support, to assess how well they can be used with a (popular) VoIP application. The advances in current wireless technologies, as well as the decrease in data plan prices from cellular operators and the wide availability of WLANs currently make possible to make VoIP access ubiquitous, and to reduce usage costs. The availability of HSDPA networks and devices is increasing daily, and this will significantly increase the performance of VoIP applications. Furthermore, operators are also moving to an all-IP architecture with the deployment of the IP Multimedia Subsystem (IMS), and thus VoIP (and other multimedia) over cellular networks will be better supported.

As expected, the results obtained show that VoIP quality in these contexts is still not on par with it's circuit-switched counterpart. Lower–than–toll MOS were the norm rather than the exception, and the perceived quality was very variable. Still, the possibility to use VoIP in multi-access devices enlarges the reach of VoIP applications, since they can now be used by mobile users, often at a significant cost advantage. Thus, an *advantage factor* should be considered when assessing these systems and comparing them to the traditional ones.

References

[1] Unlicensed Mobile Technology (UMA) website. http://www.umatechnology.org/.

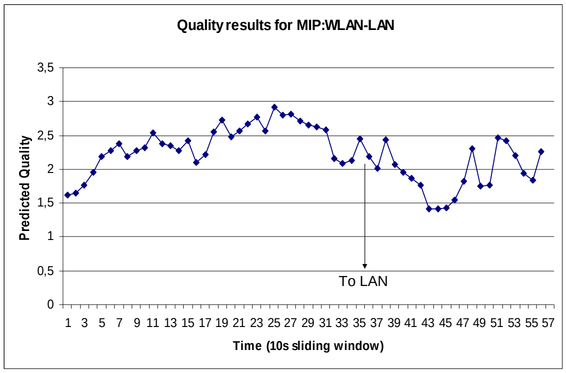

Figure 9: Quality results for the Mobile IP scenario. Three handovers were performed in this example. Notice how the 3G performance is significantly worse than that of any of the other two access media.

[2] VTT. Converging Networks Laboratory. http://cnl.willab.fi/.

[3] IETF Network Working Group. IP Mobility Support for IPv4 (RFC 3220), August 2002.

[4] ITU-T Recommendation P.862. Perceptual Evaluation of Speech Quality (PESQ), an Objective Method for End-To-End Speech Quality Assessment of Narrowband Telephone Networks and Speech Codecs, 2001.

[5] Samir Mohamed Gerardo Rubino and Martín Varela. Performance Evaluation of Real-time Speech through a Packet Network: a Random Neural Networks-based Approach. *Performance Evaluation*, 57(2):141–162, may 2004.

[6] Mirjami Taramaa Mikko Koivula and Pekka Ruuska. Differentiated Services and Vertical Handovers Supporting Multimedia in Heterogeneous Networks. In *18th Annual IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC'07), to appear*, Athens, Greece, September 2007.

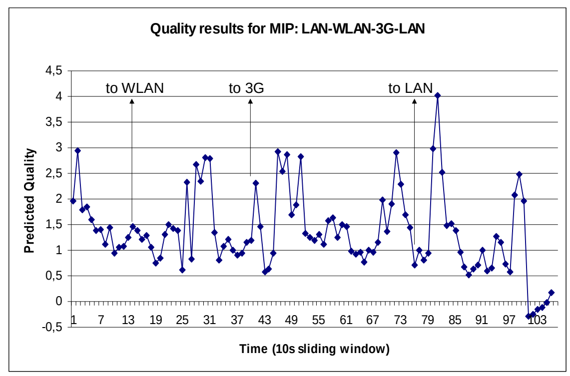

Figure 10: Quality results for the Mobile IP scenario. Three handovers were performed in this example. The quality was particularly bad, and noticeably variable during the test.