Handovers between Narrowband and Wideband Speech Transmission

Sebastian Möller, Marcel Wältermann, Błażej Lewcio, Niklas Kirschnick, Pablo Vidales

Deutsche Telekom Laboratories, TU Berlin, Ernst-Reuter-Platz 7, 10587 Berlin
E-Mail: {sebastian.moeller, marcel.waeltermann, blazej.lewcio, niklas.kirschnik, pablo.vidales}@telekom.de
Web: http://www.qu.t-labs.tu-berlin.de/

Abstract

Next Generation Networks provide seamless access to speech and multimedia transmission services at different bandwidths. Depending on the audio bandwidth of the speech codecs used in the respective networks, as well as on network degradations resulting from the handover, coding, and packet loss, the quality experienced by the user will differ within a single call. In this paper, we present initial results from a series of subjective listening-only and conversational experiments in which speech quality is quantified as a function of handover, codec change, and network characteristics. The results show when and under which circumstances a handover should be scheduled in order to obtain best speech quality. They are important for the development of high-quality handover strategies.

1 Introduction

In Next Generation Networks (NGNs), a multitude of different network technologies provide wireless and ubiquitous access to speech and multimedia services. The independence of network and service layers enables users to move through geographical areas covered by different wireless network technologies, while the service is preserved. In order to guarantee seamless mobility also for time-critical services such as Voice-over-IP (VoIP), sophisticated mobility-enabling protocols are necessary which ascertain a fast and robust roaming between different wireless networks (so-called vertical handovers).

Depending on the network technology and the available network bandwidth, vertical handovers may also lead to a change in the audio bandwidth. For example, GSM currently mainly transmits the traditional telephone band (300-3400 Hz), whereas HSDPA and WLAN can also be used with codecs offering wideband (50-7000 Hz) speech transmission. As a result of a handover, the codec may change as well, depending on the VoIP service characteristics. Consequently, also the audio bandwidth provided by the codec may switch between narrowband (NB) and wideband (WB) within a single call. So far, the authors are not aware of any investigations which quantify the effect of transmission bandwidth changes *within a single call* on perceived quality. Perceptual effects of bandwidth change have to be put into a relationship to other degradations occurring simultaneously in the network, like coding distortions and packet loss. A quantitative description of such quality degradations is necessary in order to design optimum handover techniques, and to design high-quality networks.

Subjective tests are still the only valid and reliable means for evaluating the quality during handovers. Once the perceptual effects have been quantified, they may later be described by instrumental quality prediction models like the ones recommended by the International Telecommunication Union (ITU-T). This body currently recommends signal-based models which estimate the quality of transmitted speech in a listening-only situation by comparing input and output signals on a perceptual level [1][2], as well as parametric models which estimate conversational speech quality on the basis of network parameters (loudness ratings, noise levels, packet loss rates, etc.) [3]. Unfortunately, none of these models has proven to be applicable to predict the effects of bandwidth changes and NGN handovers.

In this paper, we present the first part of an in-depth study on the effects of vertical handovers on speech quality in NGNs. The aim of the entire study is to quantify and predict the effects of different handover and network characteristics (handover point within a call, networks used before and after the handover, audio bandwidth available in these networks, and packet loss) on the quality perceived by the user. The following research questions guided the design of our experiments:

- Which network and handover characteristics are the most relevant ones from a perceptual point-of-view?
- Is it advantageous to switch from NB to WB whenever possible, or does the bandwidth switching degrade perceived quality? If yes, under which circumstances?
- Is it possible to predict the effects of handovers with instrumental models like the ones cited above?

In order to get analytic insights into the perceptually relevant effects, we focused on the listening-only situation in the first part of our investigation, and tried to partially answer the first two questions. In the near future, we will use the test results for improving the mentioned quality prediction models for NGN handover scenarios, in order to answer also the third research question.

2 NGN Testbed

A detailed analysis of the perceptual effects requires controllable network conditions. For this purpose, an NGN testbed has been developed which allows for a manipulation of the mentioned network characteristics. It uses Mobile IPv4 to enable seamless handovers between different radio access technologies, and consists of a Home Agent (HA), a Mobile Node (MN), and a Correspondent Node (CN) which communicate with each other to establish a connection. Details on the hardware and software components as well as on their communication are described in [4].

The VoIP framework is implemented with the help of the PJSIP software [5], to which extensive modifications have been made. It makes use of the SIP/SDP parameter negotiation, parallel media stream establishment, and RTP packet filtering to enable codec changeover during a call. Different changeover techniques have been developed and compared in [6], showing that a soft codec changeover with a shared media transport seems to be the best solution. We coded speech using the G.722.2 (AMR-WB) codec at 23.05 and 12.65 kbit/s in the WB domain, and G.711 log. PCM at 64 kbit/s in the NB domain, both with their native packet loss concealment (PLC). As parallel media streams are allocated during the transition phase between networks, codec changeover can be scheduled before or after network handover. The testbed can additionally introduce random packet loss of a defined percentage *Ppl* in order to degrade network conditions in a controlled way.

3 Test Procedure

We can expect speech quality to vary during a call. Thus, in order to quantify quality, the entire length of a call has to be considered. Standard listening-only tests according to ITU-T Rec. P.800 [7] make use of speech samples of 4-8 s length and are therefore not suitable for this purpose. On the other hand, conversational tests are tedious and they place the listener's focus of attention on the content of the speech signal – and not on the form. While this is ecologically valid, subtle perceptual differences may get blurred as a result of the test scenario.

As a compromise, we opted for a three-fold test protocol:

1. Conversations of 60 s length were simulated by concatenating 5 meaningful speech segments with alternating pauses, playing them back to the test participants, asking them to answer content-related questions during the pauses, and asking for an overall quality judgment at the end of the simulated conversation. This approach has been developed in [8][9] and is now recommended for call-quality measurement in [10].
2. The composing segments of the simulated conversations and some additional segments of approx. 6 s length were presented to the participants in a standard listening-only context, asking for an overall quality rating after each sample.
3. Several anchor conditions were assessed in a standard Short Conversation Test according to [7]. This test serves to counter-check the validity of the listening-only test results, i.e. their transferability to a realistic conversational situation.

Up to now, we carried out two tests of the first type (Tests 1a and 1b), two corresponding tests of the second type (Tests 2a and 2b), and one test of the third type (Test 3). In this paper, we focus on the results of Tests 1a and 2a, and derive initial rules for a quality-driven codec switching logic. In future work, we will concentrate on validating the results with the second pair of listening-only tests (1b and 2b) and the conversational test (3), and on modeling the integration of instantaneous quality judgments to an estimate of the overall call quality.

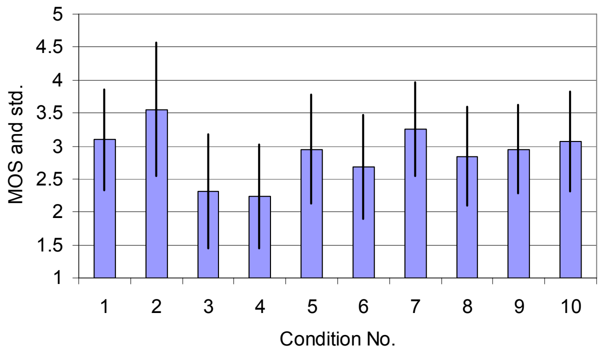

Figure 1: Results of Test 1a.

3.1 Test Conditions

Test 1a concentrates on WB/NB transitions and the effects of packet loss. It contains two conditions with pure NB and WB calls, 4 conditions where packet loss continuously increases until the middle (3rd segment) of the call, and then codec switching occurs to a loss-free network with a different codec (or not), and 4 conditions where NB → WB or WB → NB transitions occur at the beginning (2nd segment), or at the end (4th segment) of a call. Table 1 summarizes the conditions.

Table 1: Test 1a conditions. H: HSDPA; W: WLAN; *Ppl*: packet loss in %; switching at the beginning (beg.), middle (mid.) or end of a simulated call.

No.	Network(s)	Codec(s)	*Ppl* per segment
1	H	G.711	0
2	W	G.722.2	0
3	H	G.711	0,10,20,10,10
4	W	G.722.2	0,10,20,10,10
5	H→W mid.	G.711→G.722.2	0,10,20,0,0
6	W→H mid.	G.722.2→G.711	0,10,20,0,0
7	H→W beg.	G.711→G.722.2	0
8	H→W end	G.711→G.722.2	0
9	W→H beg.	G.722.2→G.711	0
10	W→H end	G.722.2→G.711	0

The corresponding Test 2a contains all segments of the simulated conversations of Test 1a, plus additional samples with and without network handovers, addressing also Flash-OFDM networks. The resulting list of 25 segments for this test is not reproduced here to save space; instead, the x-axis labels in Fig. 2 provide a short description of the respective test conditions.

3.2 Test Set-Up

Test participants were invited to a sound-insulated laboratory, were instructed about the purpose of the test, and listened to the samples in three sessions of approx. 25 min. each (2 sessions for Tests 1, 1 session for Tests 2). Speech samples were presented over a Sennheiser HMD

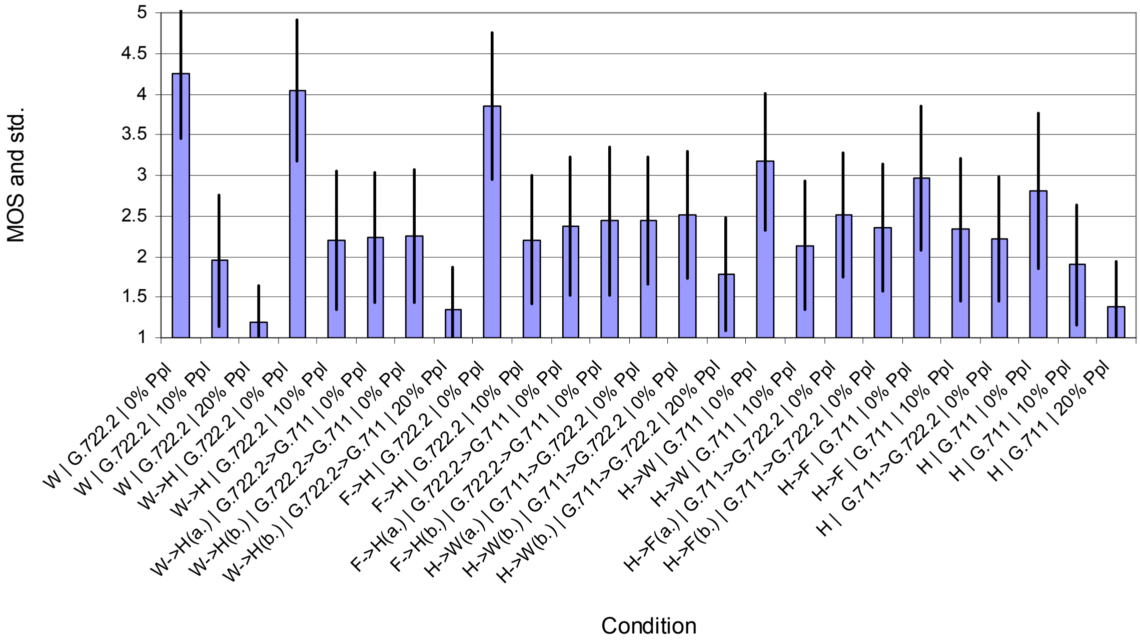

Figure 2: Results of Test 2a. First field: network conditions: W: WLAN; H: HSDPA; F: Flash-OFDM; →: network handover; (b.): codec changeover before network handover; (a.): codec changeover after network handover. Second field: codec conditions: G..711: G.711 at 64 kbit/s; G.722.2: G.722.2 at 23.05 kbit/s; →: codec changeover. Third field: packet loss conditions.

410 headset at a comfortable listening level, with a background level below 35 dB(A) [8]. After each segment of Test 1a, the participants were asked to answer a short content-related question by selecting one of 3 options presented on a computer screen. At the end of each simulated conversation of Test 1a, as well as after each sample of Test 2a, participants had to rate the overall quality on a 5-point absolute category scale. The test set-up and scale followed mainly the requirements given in [7] and [10].

13 participants took part in Test 1a, and 24 in Test 2a. They were recruited from the normal telephone-user population, did not report any hearing impairment, and received a voucher in return for their effort.

4 Test Results

4.1 Test 1a

Fig. 1 shows the auditory judgments of the simulated conversations of Test 1a, averaged over all test participants and samples used in each condition (mean opinion scores, MOS, and standard deviations, std.). As expected, the pure WB condition (#2) is rated best, and considerably better than the pure NB condition (#1). The packet loss of conditions #3 and #4 impacts quality significantly; these conditions have the lowest quality of the entire test, showing that packet loss seems to be the most dominant quality degradation. Conditions #5 and #6 differ from #3 and #4 in that a network handover occurs, leading to a different audio bandwidth and no packet loss at the end of the call. In both cases, the handover results in a considerable improvement of perceived quality. Apparently, network handover can be an efficient tool for quality improvement in case that packet loss degrades quality significantly. This also holds when a transition from WB to NB is necessary, as the comparison between conditions #4 and #6 shows.

Even when no packets are lost, switching from NB to WB may be advantageous in case that a significant time period remains in order to take profit of the improved quality. A comparison between conditions #7, #8 and #1 shows that quality improves when switching from NB to WB at the beginning of a call; however, when switching occurs at the end of a call, the quality degrades compared to the pure NB case. For the opposite (WB → NB) direction, switching definitely degrades quality; the longer the WB connection remains established, the better the perceived quality.

4.2 Test 2a

The results of Test 2a help to quantify the trade-off between audio bandwidth and packet-loss degradations, see Fig. 2. As in Test 1a, packet loss is the most important quality degradation. In case of zero packet loss, a network handover without changing the codec degrades quality only slightly. However, if the codec has to be changed as well, this has a significant effect on perceived quality. For WB→NB transitions, the impact of bandwidth switching is roughly equivalent to a 5-10% packet loss degradation. For the NB→WB transition, quality improves in all cases, and the improvement is again equivalent to a 5-10% packet loss in the improved condition. These findings are valid both for the transitions between WLAN/HSDPA and Flash-OFDM/HSDPA. No difference was found between codec switching before and after the handover.

4.3 Further Results

The other test results we have analyzed so far confirm most of these findings. The impact of bandwidth switching seems to depend on the basic quality level provided by the network: If the quality is high, then the impact of bandwidth switching is remarkably high as well; if the basic quality level is already low due to high packet loss in the network, then switching codecs does not have a strong effect. In other words: packet-loss degradations are able to mask the positive effect of switching to a higher audio bandwidth.

5 Summary and Conclusions

On the basis of the obtained test results, it is possible to partially answer the first two research questions raised in the introduction:

- Packet loss seems to be the most relevant degradation in the NGN scenarios addressed in our study; it may also mask the perceptual effects of bandwidth switching. Switching from NB to WB and vice versa is a secondary degradation, which is roughly equivalent to the degradation caused by 5-10 % packet loss. Network switching – without switching the codec bandwidth – is far less important from a perceptual point-of-view.
- Switching from a WB to a NB codec is advantageous in case that the packet loss rate is high, and if the switching helps to reduce the packet loss rate, by requiring less throughput. On the other hand, switching from NB to WB may be advantageous only in case that a sufficiently long period of WB speech remains before the end of the conversation. From our limited results, this minimum length seems to be in the range of 30 s.

The results may help to design efficient network handover strategies, cf. e.g. [11]. In bad network conditions (high packet loss rate), a handover should be made if the packet loss rate can be reduced by this step. In this situation, it is not important whether the audio bandwidth can be maintained or not; the reduction of packet loss should be the ultimate goal. In contrast to this, a handover can also be fruitful in good network conditions (low packet loss rate). In this case, switching to a higher audio bandwidth can help to significantly improve quality. The improvement is most effective if it occurs early in the call; the remaining call duration should be more than 30 s. Switching from WB to NB is always linked to a loss in quality; network handover without codec switching, however, does not significantly impact the perceived quality.

6 Acknowledgment

The study was carried out in the frame of the "Mobisense" project funded by Deutsche Telekom AG, and the project "Attribute-based Speech Quality Measures" funded by the DFG (MO 1038/5-2). The funding and fruitful discussions with team members from DAI-Labor and T-Labs are gratefully acknowledged. Pablo Vidales is member of the National System of Researchers (SNI) in Mexico since January 2007.

Literature

[1] ITU-T Rec. P.862. *Perceptual Evaluation of Speech Quality (PESQ): An Objective Method for End-to-end Speech Quality Assessment of Narrow-band Telephone Networks and Speech Codecs*. International Telecommunication Union, Geneva, 2001.

[2] ITU-T Rec. P.862.2. *Wideband Extension to Recommendation P.862 for the Assessment of Wideband Telephone Networks and Speech Codecs*. International Telecommunication Union, Geneva, 2007.

[3] ITU-T Rec. G.107. *The E-Model, and Computational Model for Use in Transmission Planning*. International Telecommunication Union, Geneva, 2005, and Amendment 1, 2006.

[4] Pablo Vidales, Niklas Kirschnick, Blazej Lewcio, Frank Steuer, Marcel Wältermann and Sebastian Möller. Mobisense Testbed: Merging User Perception and Network Performance. In *Proc. 4th Int. Conf. on Testbeds and Research Infrastructures for the Development of Networks & Communities*, Innsbruck, March 18-20, 2008.

[5] PJSIP – Open Source SIP Stack and Media Stack for Presence, Im/instant Messaging, and Multimedia Communication. http://www.pjsip.org, 2008.

[6] Marcel Wältermann, Blazej Lewcio, Pablo Vidales and Sebastian Möller. A Technique for Seamless VoIP-Codec Switching in Next Generation Networks. In *IEEE International Conference on Communications (ICC 2008)*, Beijing, 2008.

[7] ITU-T Rec. P.800. *Methods for Subjective Determination of Transmission Quality*. International Telecommunication Union, Geneva, 1996.

[8] J. Berger, A. Hellenbart, R. Ullmann, B. Weiss, S. Möller, J. Gustafsson and G. Heikkilä. Estimation of 'Quality per Call' in Modelled Telephone Conversations. In *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP 2008)*, Las Vegas, pages 4809-4812, 2008.

[9] Benjamin Weiss, Sebastian Möller and Jens Berger. Wahrgenommene Sprachqualität in Telefongesprächen bei zeitlich variierenden Übertragungseigenschaften. In *Elektronische Sprachsignalverarbeitung. Tagungsband der 18. Konferenz*, Cottbus, K. Fellbaum (ed.), TUDpress, Dresden, pages 210-217, 2007.

[10] ETSI TR 102 506. *Speech Processing, Transmission and Quality Aspects (STQ); Estimating Speech Quality per Call*. European Telecommunications Standards Institute, Sophia Antipolis, 2007.

[11] Niklas Kirschnick. *Mobility Management in Next Generation Networks Based on User Perception*. Diploma thesis, TU Berlin, 2008.