3rd ETSI Speech Quality Test Event for VoIP Equipment

F. Kettler, F. Rosenberger, H.W. Gierlich, S. Völl, Ph. Cousin,*

HEAD acoustics GmbH, Ebertstr. 30 a, D-52134 Herzogenrath, Frank.Kettler@head-acoustics.de
* European Telecommunications Standards Institute ETSI, *Sophia Antipolis, France*

1. Abstract

ETSI Technical Commitee STQ (Speech Processing, Transmission and Quality Aspects) together with HEAD acoustics as test lab organized the 3rd Speech Quality Test Event (SQTE) for VoIP-Equipment 2004. These single vendor tests continued the success of the 1st and 2nd SQTE, carried out in 2000 and 2002. Due to the high interest of the VoIP community, especially equipment manufacturers and network operators in these kinds of tests, the event was organized as a "twin" event in Europe and the US. Both, VoIP gateway and VoIP terminal manufacturers participated in this test. The results are anonymously compared providing a benchmark of current implementations.

The most important speech quality parameters are summarized in a condensed representation - best describes as a "quality pie" - for each device under test. This presentation discusses some results and gives an overview about some speech quality aspects on this market.

2. Combining "Condensed" Quality Scores and Detailed Parameter Analysis

One focus of the test event was the determination of one-way listening speech quality. This transmission aspect is influenced by the choice of speech coders (G.711, G.729, G.723), the implemented VAD (voice activity detection) and the reaction on network impairments (packet loss concealment (PLC), jitter buffer design). The test setup for IP gateways is shown in figure 1.

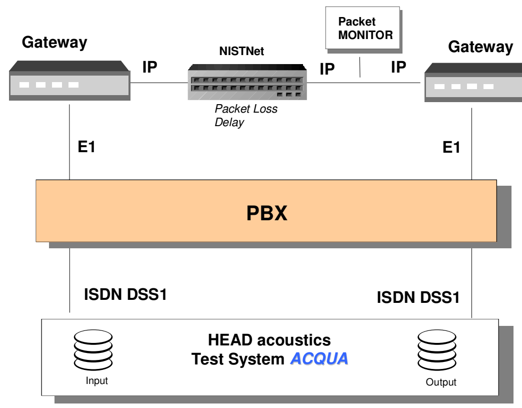

Fig. 1: Test setup for VoIP gateways during the 3rd ETSI SQTE.

Test signals and speech samples were transmitted and analyzed by the test system connected via two ISDN lines. On the packet side IP network impairments are introduced by an IP simulation tool (NISTnet). The tests were carried under signal talk conditions determining the listening speech quality, included echo and double talk performance tests and additionally focused on the quality of background noise transmission.

Figure 2 shows the listening speech quality scores (MOS-LQO acc. to ITU-T P.800.1 [1] using PESQ [2]) and the propagation delay for IP gateways. For both parameters the curves indicate the maximum, minimum and average values over all devices under test. These results are based on the evaluation of 10 gateways using the G.711 speech codec. The test conditions (IP network impairments packet loss [%] and jitter [ms]) are indicated on the x-axis.

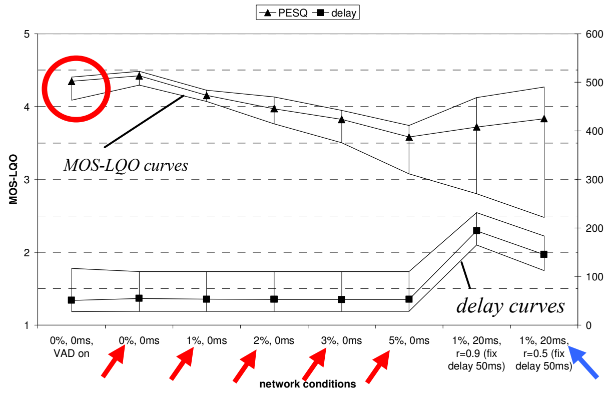

Fig. 2: Minimum, average and maximum MOS-LQO and delay values (10 IP gateways, G.711, 10 ms packet length)

The VAD of all tested implementations does not influence the listening speech quality significantly. A quality difference of 0.3 between the minimum and maximum MOS-LQO can be measured under the VAD test condition (see red circle).

For the different packet loss rates but without jitter (see red arrows) the results decrease. 5% packet loss leads to 3.6 average MOS-LQO. The quality range is approximately 0.7 MQS-LQO under this test condition. The quality differences for the implemented PLC strategies are significant. Further tests applying more specific analyses can be used as optimization criteria (see below).

Very high quality differences occur under the jitter conditions (see blue arrow). They are caused by the different jitter buffer implementations and the related control strategies. Impairments are mainly caused by non-optimized phase interpolation during jitter buffer adjustments (see below).

The average end to end delay significantly increases up to 145 ms for the jitter condition (correlation coefficient r=0.5, see blue arrow). A fixed delay of 50 ms was introduced by the test setup, the average equipment delay can therefore be calculated to 95 ms. The difference between the maximum and minimum delay is constant for all test conditions. Jitter does not influence these delay differences between the devices under test.

These "condensed" results provide an overview about the quality range of different implementations. Moreover, an additional focus of this event was to provide ideas for manufacturers how to improve the performance - if necessary. Two examples concerning the implemented packet loss concealment algorithms and the jitter buffer performance are given in the following figures. Figure 3 and 4 compare two gateways measured under the influence of 5% packet loss. The upper window shows the transmitted test signal (a periodically repeated voiced sound). The lower window represents a cross correlation analysis between the transmitted signal and the original test signal. Note that the periodical pattern is caused by the periodical repetition of the voiced sound. The middle window represents the Relative Approach analysis [3], a hearing model based method to detect and optimize audible disturbances (see also [4]). The implemented PLC analyzed in figure 3 does not indicate significant disturbances under the influence of packet loss in the simulated network. The MOS-LQO was determined to 3.7. The delay is 41.1 ms.

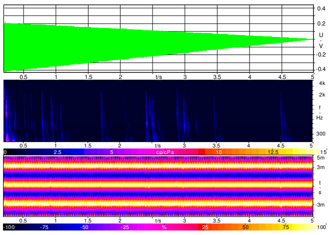

Fig. 3: PLC 1 (MOS-LQO 3.7, delay 41.1 ms)

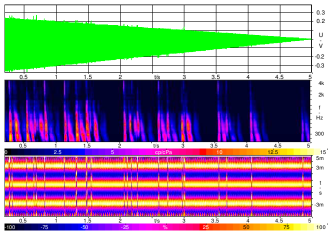

Fig. 4: PLC 2 (MOS-LQO 3.3, delay 34.6 ms)

The gateway analyzed in figure 4 performs with a MOS-LQO of only 3.3 under this test condition. The cross correlation analysis indicates phase shifts caused by the non-optimized phase interpolation. The Relative Approach indicates significant disturbances mainly in the lower frequency range. The delay was even lower for this implementation (34.6 ms). These two examples provide useful information on how to optimize the performance of PLC.

The same tests were carried out introducing 20 ms jitter. Two implementations are analyzed in figure 5 and 6. Jitter is completely considered by the jitter buffer as shown in figure 6. This implementation does not lead to audible disturbances. Signal phase adjustments can only be seen for very low signal amplitudes. The corresponding MOS-LQO was determined to 4.2, the delay amounts to 90 ms. The implementation analyzed in figure 5 shows phase discontinuities introduced by the jitter buffer adjustment and a complete re-synchronization of the signal phase (see red arrow). The MOS-LQO was determined to 3.1. For this implementation the delay was 90 ms.

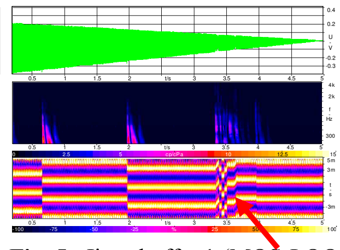

Fig. 5: Jitter buffer 1 (MOS-LQO 3.1, delay 90 ms)

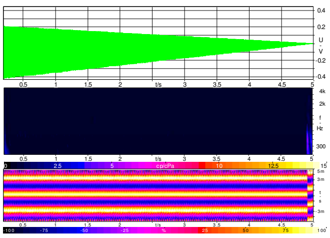

Fig. 6: Jitter buffer 2 (MOS-LQO 4.2, delay 90 ms)

The combination of "condensed" and detailed analyses point out one focus of the ETSI event: comparison of speech quality aspects and optimization criteria.

3. Conversational Aspects

The ETSI SQTE tests covered all conversational aspects including the quality of background noise transmission. Artificial noise scenarios were used as well as realistic signals. Figure 7 and 8 demonstrate two implementations of VAD and comfort noise generation.

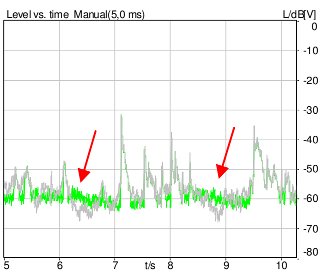

Fig. 7: Noise transmission 1

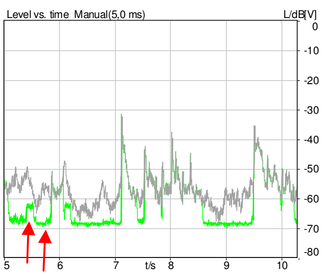

Fig. 8: Noise transmission 2

The curves represent the measured signal level vs. time at the output (green) and the original test signal level vs. time (grey). Comfort noise is continuously adapted in the analysis shown in figure 7 but the level sometimes exceeds the original noise level (see red arrow). The implementation shown in Figure 8 adapts the comfort noise in steps. It does not continuously follow the original signal level. These noise level mismatches are audible and disturbing.

4. "Quality Pies"

In order to provide a quick overview about the results for each gateway under tests for all speech quality aspects a graphical result representation was derived. The focus of this conversational speech quality representation is to provide

- a "quick and easy to read" overview about the implementation including strength and weakness,
- a comparison to recommended values or average results from the event,
- detailed information for development to improve the performance.

The results are summarized in a diagram best described as a "Quality Pie". Three examples of IP gateways are shown in figure 10 [5]. The following assumptions were made: Each parameter is represented by a pie slice. The size of each slice directly correlates to quality. Interaction aspects between single parameters are not considered.

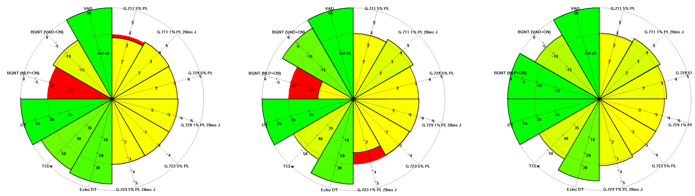

Fig. 10: Overview - speech quality performance (three gateways [5])

The minimum requirement for a parameter or the average results from all manufacturers participating during the event is indicated by an inner red circle. If the measured parameter exceeds the recommended requirement or indicates a quality better than the average performance during the test event, the red circle is not visible and overlapped by the pie slice. Corresponding representations were also derived for IP phones tested during the event [6].

The whole analysis can be found in the anonymized test reports published after the test event. These reports can be found on the ETSI website http://www.etsi.org/plugtests/History/2004SQTE.htm .

5. Conclusions

The ETSI Speech Quality Test Events for VoIP equipment is established as an anonymous international speech quality comparison test. The benefit for manufactures is obvious: On the one hand anonymity guarantees ideal comparison to other implementations on the market. On the other hand the powerful combination of condensed results and detailed parameter analyses provides optimization criteria to improve performance.

6. References

[1] ITU-T Recommendation P.800.1, Mean Opinion Score (MOS) Terminology

[2] ITU-T Rec. P.862, Perceptual Evaluation of Speech Quality (PESQ)

[3] Genuit, K.: Objective Evaluation of Acoustic Quality Based on a Relative Approach, InterNoise '96, Liverpool, UK

[4] Kettler, F.; Gierlich, H.W.; Rosenberger, Application of the Relative Approach to Optimize Packet Loss Concealment Implementations, DAGA 2003, 18.-20.03.2003, Aachen

[5] Anonymized Test Report - Gateways, 3rd ETSI Speech Quality Test Event, ETSI Plugtests Service, HEAD acoustics, Dec. 2004

[6] Anonymized Test Report – IP Phones, 3rd ETSI Speech Quality Test Event, ETSI Plugtests Service, HEAD acoustics, Dec. 2004