VoIP QUALITY

State-of-the-art, end-to-end speech quality for voice over IP:

impressions of the first ETSI Voice over IP (VoIP) speech quality test event

Traditional telecommunications experts and data specialists tend to avoid discussions about speech quality expectations. Marketing strategies often play the major role in discussions about the expected speech quality for the users. Moreover, sometimes the detailed understanding of all relevant speech quality aspects and the specific knowledge about appropriate measurement procedures are missing.

In October 2000, the first VoIP speech quality test event was held by ETSI in their headquarters in Sophia Antipolis, France. Over one week, state-of-the-art VoIP equipment from various manufacturers was tested according to measurement methods defined in ITU-T Recommendations as well as in working group 5 of the ETSI VoIP project TIPHON.

This article discusses the measurement methods as well as the general trends that can be analyzed from the results obtained during the test event.

Test Principle
The speech quality test event was organized in order to determine quality parameters for VoIP equipment in both gateways and IP terminals. The tests were performed by the T-Nova Berkom and the HEAD acoustics test labs. The complete end-to-end scenario was considered to reproduce realistic conditions for the user. In order to compare the results obtained for the equipment of different manufacturers, common test procedures were provided to all participants. These procedures were defined and distributed in advance and included:

- a common test set-up providing optional electrical or acoustical interfaces to access the equipment under test
- a common set of network conditions like different rates of packet loss, additional delay and delay jitter
- the use of identical test signals and analysis methods

Common Test Scenarios
The common test set-up provided optional:

- two electrical interfaces (see figure 1),
- one electrical and one acoustical interface (see figure 2) or
- two acoustical interfaces

Manufactures bringing IP gateways could connect their equipment via E1 or T1 interfaces to a PABX and the IP network simulation (NistNet) as shown in figure 1. The testing equipment provided by the two test labs accessed the set-up via one ISDN line on each side to feed and record test signals.

Optionally the set-ups could be accessed via acoustical interfaces necessary to determine end-to-end speech quality parameters (mouth to ear) using a standard ISDN telephone connected to the PABX or IP terminals like handsets, headsets or hands-free phones. IP terminals were directly connected to the IP network simulation.

Figure 2 shows one example of the test set-up using one acoustical interface to access an IP terminal and an electrical interface to access one gateway.

![Figure 2: test set-up for VoIP equipment, combined acoustical/electrical measurements](image)

The artificial head measurement system, HMS II.3 of HEAD acoustics (HATS, head and torso simulator according to ITU-T Recommendation P.58) provided the acoustical interface. The HATS are equipped with an artificial mouth and an artificial ear, type 3.4 according to ITU-T Recommendation P.57. A handset-mounting device guarantees the realistic and reproducible position of the terminal relative to the artificial ear.

Figure 3 shows a head and torso simulator HMS II.3 with mounted handset (note, that the blue cap sponsored by ETSI is not necessarily needed for the acoustical measurements).

Test signals and real speech samples were transmitted and recorded between the two access points. The tests covered not only the single talk conditions to access the speech listening quality, but also double conditions and echo measurements to determine the conversational quality.

**Common Test Conditions**

Table 1 shows the set of test conditions provided by the NistNet IP simulator for measurements of speech listening quality. To ensure a reasonably precise percentage of packet losses, an IP traffic monitoring system, a new development of T-Nova Berkom, was used.

The test conditions using the acoustical interface are given in table 2.

Overall speech quality assessment requires the consideration of additional test conditions. Besides the perceived speech quality in the listening situation parameters like audible echoes in single and double talk situations, the double talk performance (both subscribers talk simultaneously) and the transmission quality of background noise has to be taken into account. The long transmission delays that can be expected in VoIP connections emphasize the importance of these conversational parameters. People...
The injected comfort noise level was significantly higher than the channel noise, which lead to quality degradation in the listening test. TOSQA focuses on the quality assessment during active speech and consequently these conditions were scored a bit too optimistically. However, this auditory reference test was carried out for the test scenario using the electrical interface at both ends.

A second auditory reference test was conducted based on recordings using the acoustical interface. The acoustical interface was provided by the HATS. This interface was provided for manufactures of IP-phones. The reference conditions for the auditory test (single codecs and MNRU) were generated in the same way using a conventionally shaped telephone handset mounted to the HATS. The new TOSQA 2001 ‘terminal extension’ version, providing the quality assessment of acoustically recorded speech material was used. The correlation between the auditory MOS and the objective TOSQA 2001 results reached 0.98.

The test scenario was a complete four-wire connection. Consequently
the test set-up cannot physically cause the echo. But under some test conditions in the double talk situation echo components could be identified which lead to audible and annoying echo. Figure 6a gives one example. High level echo components can be determined during the pauses between the signal bursts. The echo occurred only during double talk, under single talk conditions the connection was echo-free. These echoes will probably lead to customer complaints under real network conditions.

Clipping

The test result of a different implementation is given in figure 6b. The original test signal bursts (red color) are not completely transmitted and the signal is partly clipped. This hampers the naturalness of a conversation because both subscribers have to concentrate on not double talking in order to avoid these disturbances.

Quality of background noise transmission

The environmental conditions during a telephone conversation, for example the background noise in an office or traffic noise particularly in the case of mobile communications, has to be considered. Different types of background noise signals were used during the tests to determine the transmission quality. The following figure 7 demonstrates a typical example using a test signal with an increasing level versus time (the red signal is the original test signal). The green signal was measured after transmission over two gateways.

The signal part applied with the lower level is not completely transmitted. Obviously a detection threshold is implemented in order to save transmission bandwidth during speech pauses. However, this may lead to an annoying noise contrast for the conversational partner, especially if the noise modulation occurs frequently in realistic telephone scenarios.

Summary

The test event was successful and, as indicated by the feedback of all participating companies, useful for the manufacturers who participated. The results of the listening speech quality demonstrate the state-of-the-art performance of equipment under single talk conditions. The test scenarios and the measurements as they are briefly described above determine additional parameters, which highly influence the conversational quality. The tests and the results obtained during the speech quality test event clearly point out the importance of considering the end-to-end scenario. It can be assumed that for all participating manufacturers the results can be used for optimization of their VoIP equipment to improve the overall speech quality.

The event was sponsored by Texas Instruments, Alcatel, T-Nova Berkom and HEAD acoustics. Detailed results from the event will be discussed publicly at the next TIPHON meeting in March 2001.