White Paper

ETSI Speech Quality Test Event Calling Testing Speech Quality of a VoIP Gateway –

A white paper from the ETSI 3rd SQTE
(Speech Quality Test Event)

Version 1
July 2005

AudioCodes

ETSI Speech Quality Test Event

White Paper

Table of Contents

Introduction 4

Voice Quality Testing Methods 4

ETSI Speech Quality Test Event (SQTE) 5

The Test Setup 6

The Test Categories 6

Listening Aspects 7

Test Results 7

Specific Test Results 9

Echo Canceller Double Talk Test 9

Echo Canceller Background Noise Test 10

Summary 11

References 11

About AudioCodes 13

Table of Figures

Figure 1: Test Setup 6

Figure 2: Mediant 2000 Test Results 9

Introduction

Speech Quality is getting increased attention by operators as VoIP services are competing with and replacing traditional PSTN. This process is now being accelerated because VoIP technology has developed in terms of Speech Quality and is being adopted by mainstream telecom carriers.

The purpose of this white paper is to review advanced Speech Quality Testing techniques that were used in the ETSI 3rd Speech Quality Test Event [1] and provide some examples from tests that were done on the AudioCodes Mediant™ 2000 VoIP Gateway.

Voice Quality Testing Methods

Voice quality testing methods can be divided to subjective and objective tests.

In subjective tests, human listeners hear and rank the quality of processed voice files according to a certain scale. The most common scale is called MOS (Mean Opinion Score) and is composed of 5 scores of subjective quality:

1 – Bad
2 – Poor
3 – Fair
4 – Good
5 – Excellent

The MOS score of a certain vocoder is the average of all the ranks voted by the different listeners of the different voice files used in the experiment. The disadvantage of subjective tests is that many listeners are required to produce a statistically stable score. Therefore, objective test algorithms were developed in order to simplify the testing process.

Modern objective test algorithms try to mimic the human listening system using psycho-acoustical models in order to produce a score that will be similar (on average) to the score produced by a human listener. Moreover, the scale of modern objective test algorithms parallels the MOS scale (by using statistical regression analysis), so one can roughly compare results obtained by different objective and subjective tests. However, this comparison should be done carefully because results can vary due to the speech database used, test environment and limitations of the testing methods.

The state-of-the-art in objective test algorithms is the PESQ algorithm (Perceptual Evaluation of Speech Quality) that was published as ITU standard P.862 [2]. This algorithm was used in the ETSI SQTE together with an additional algorithm – TOSQA [9].

Another classification of Speech Quality testing methods involves listening and conversational situations. In listening speech quality tests, the PESQ algorithm mentioned above is used to judge the effect of voice compression (depending on the vocoder) and of Packet Loss and Jitter (depending on the network conditions) on the speech quality. In conversational speech quality tests, additional factors that influence human speech are considered. These factors include echo, double talk, and background noise conditions, as well as VAD (Voice Activity Detection) and CNG (Comfort Noise Generation).

While many testing labs and vendors test only the listening speech quality (due to the simplicity of the test), testing the conversational speech quality, which is of at least equal importance, is much more difficult to test. While attempts were made to test the echo performance, the most critical of conversational aspect [5] [6] [7], Head Acoustics combined all of them during the ETSI event to provide a full picture of the conversational aspects of the tested implementation.

ETSI Speech Quality Test Event (SQTE)

The ETSI SQTE [1], which has been held several times over the past years, gives participating VoIP equipment vendors the opportunity to test their equipment under controlled conditions using common and proven test methods. This allows vendors to receive an objective and comparative analysis of their respective voice quality.

The purpose of the event as described by ETSI:
The tests will take account of all conversational quality aspects, i.e., during single talk as well as during the interaction of both subscribers. The test results will allow a state-of-the-art benchmark comparison of the VoIP equipment under test and will underline the importance of considering both end-to-end and mouth-to-ear scenarios. In addition, the results will help the participating manufacturers in their efforts to further optimize their products with regard to overall speech quality. The tests are designed and implemented by HEAD acoustics in order to

- Determine the current speech transmission quality of the tested equipment
- Further analyze parameters responsible for the results
- Discuss and optimize parameter settings for the equipment

The tests will include all conversational aspects, including;

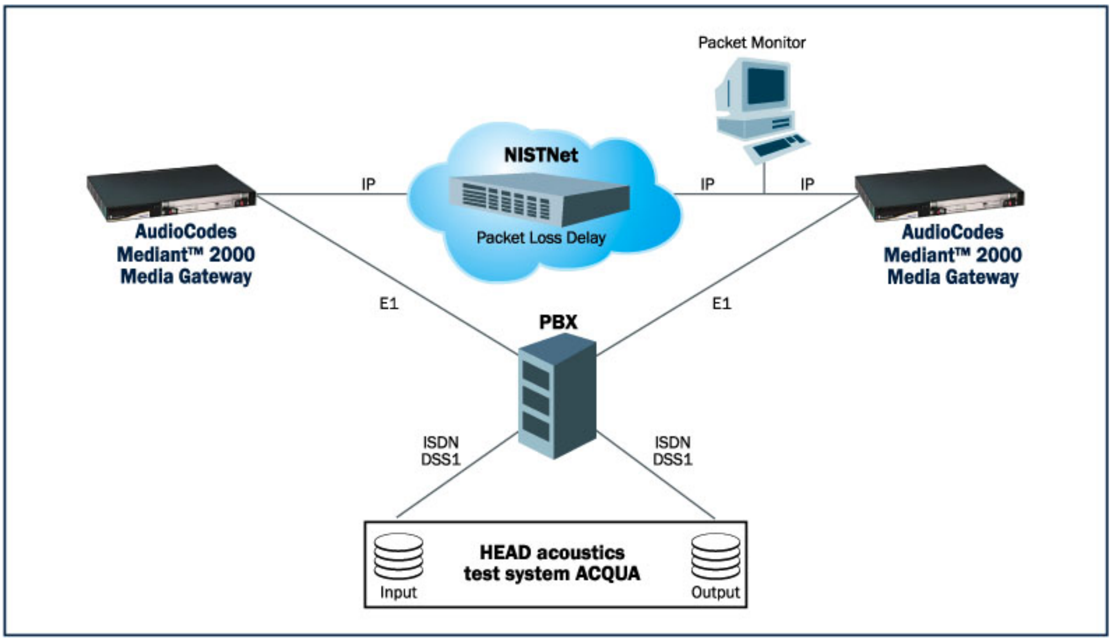
- Speech sound quality
- Echo measurements
- Double talk performance
- Transmission quality in the presence of background noise

The 3rd ETSI SQTE was held in 2 sessions in Europe and USA in 2004. Ten different organizations participated in the Gateway test sessions and an additional session was devoted to IP Phones [3]. The testing was done by HEAD Acoustics and the combined anonymous test results were published by ETSI in [4]. Each participant also received their own report with more details of each specific test.

The Test Setup

The Test Setup was composed of Head Acoustic ACQUA Test System that injected test signals to two AudioCodes Mediant 2000 gateways via their E1 interfaces. The two Mediant 2000 units were connected to a NIST Net IP network simulator that was used for injecting network impairments like packet loss, delay and jitter.

Figure 1: Test Setup



The Test Categories

Conversational Aspects

- ***Echo DT – Echo performance during double talk***

Tests the effect of speech clipping and discontinuity when 2 parties are talking together

- ***TCLw – Echo performance during single talk (Terminal Coupling Loss) according to ITU P.340 [5]***

Tests the level of the residual echo at different conditions of Hybrid ERL (Echo Return Loss)

- ***DT – Double Talk characterization according to ITU P.340 [5]***

Tests the degree of which the conversation is full duplex (i.e. 2 parties can talk simultaneously without cutting/clipping each other's speech)

- ***BGNT (NLP+CN) – Background Noise Performance of the Echo Canceller Non Linear Processor (NLP) and Comfort Noise Generator (CN)***

Tests the ability of the echo canceller to suppress the residual echo to the background noise level so it will be perceived by the user as the regular background noise rather than disturbing echo

- ***BGNT (VAD+CN) – Background Noise Performance of the Voice Activity Detector (VAD) and Comfort Noise Generator (CNG)***

Tests the ability of the VAD to stop voice packet transmission without causing the listener to notice any difference in the background noise level and spectral characteristics

- ***VAD – Performance of the Voice Activity Detector (VAD) and Automatic gain Control (AGC)***

Tests the ability of the VAD and AGC not to cause any noticeable fluctuations in the signal level and continuity

Listening Aspects

- ***G.711 5% PL – PESQ results of G.711 with 5% Packet Loss***

Tests the ability of G.711 PLC algorithm to conceal the high amount of 5% packet loss

- ***G.711 5% PL 20 ms J – PESQ results of G.711 with 5% Packet Loss and 20 ms Jitter***

Tests the ability of G.711 PLC algorithm and Jitter Buffer to conceal the high amount of 5% packet loss and Network Jitter of 20 ms

- ***G.729 5% PL – PESQ results of G.729 with 5% Packet Loss***

Tests the ability of G.729 PLC algorithm to conceal the high amount of 5% packet loss

- ***G.729 5% PL 20 ms J – PESQ results of G.729 with 5% Packet Loss and 20 ms Jitter***

Tests the ability of G.729 PLC algorithm and Jitter Buffer to conceal the high amount of 5% packet loss and Network Jitter of 20 ms

- ***G.723 5% PL – PESQ results of G.723 with 5% Packet Loss***

Tests the ability of G.723 PLC algorithm to conceal the high amount of 5% packet loss

- ***G.723 5% PL 20 ms J – PESQ results of G.723 with 5% Packet Loss and 20 ms Jitter***

Tests the ability of G.723 PLC algorithm and Jitter Buffer to conceal the high amount of 5% packet loss and Network Jitter of 20 ms

Test Results

ETSI summarized the results of each participant in a pie chart where each test category is represented by a slice. The radius of each slice is proportional to the performance in that category. Additionally, a color code is used as follows:

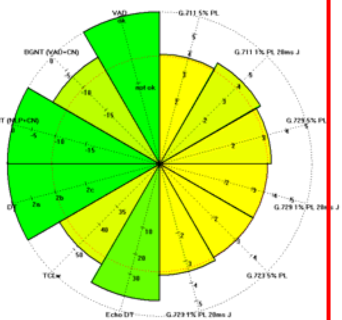
Yellow – the performance is OK or below the requirement (by a deviation represented by Red)
Red – represents the amount of deviation from the requirements
Light Green – meet or exceeds the requirement
Dark Green – much higher than the requirement

The results of all 10 gateways are presented on the next page where AudioCodes Mediant 2000 is marked as P2:

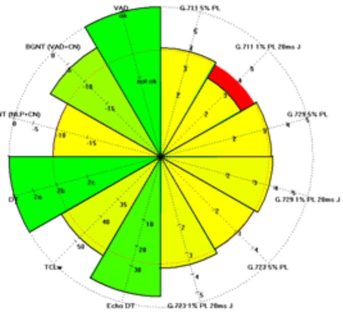
This figure, taken from the ***Anonymized Test Report "Gateways"***, provides an overview over all "Gateway Pie Charts". Note that the gateways are analyzed in random order.

AudioCodes Mediant 2000 gateway is represented by P2.

P1

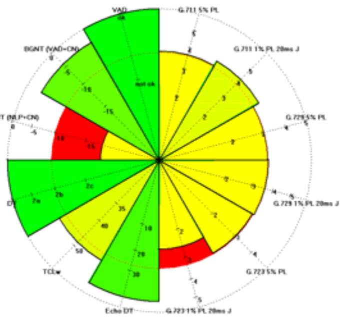


P2



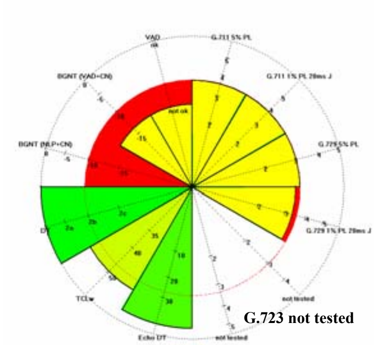
P3

P4

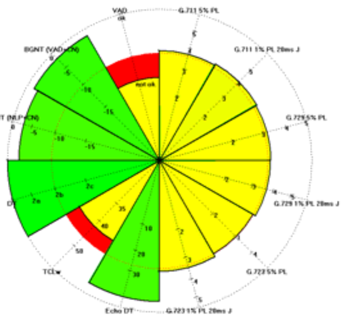


P5

P6

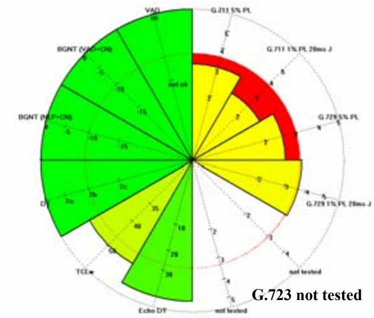


P7



P8

P9

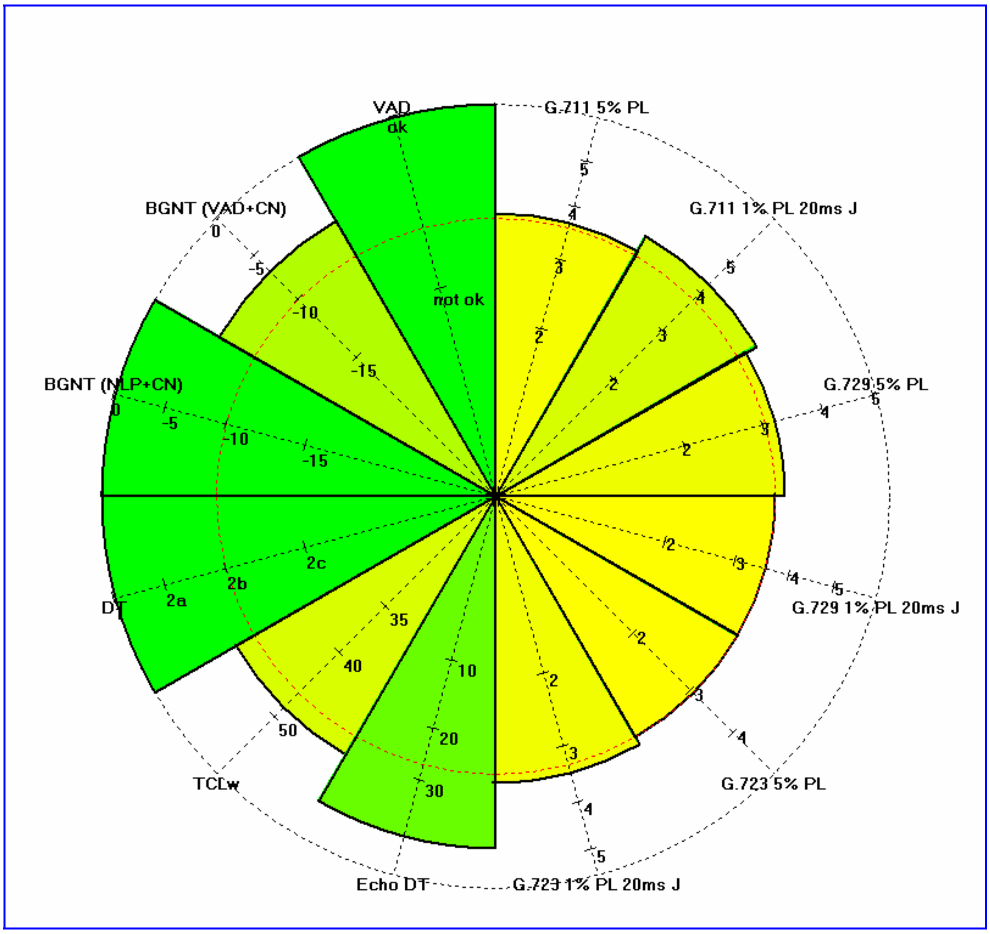


P10

As can be seen from the results above, the AudioCodes Mediant™ 2000 was the only gateway which passed and exceeded all the requirements.

For better clarity, the AudioCodes Mediant 2000 Test Results are shown enlarged below.

Figure 2: Mediant™ 2000 Test Results



Specific Test Results

Examples of specific test results from the ETSI test report follow.

Echo Canceller Double Talk Test

In this test, a near end test signal (green) and a far end test signal (red) are injected simultaneously into the echo canceller. The levels of the signals are changed to create different scenarios for the tested echo canceller. The output of the echo canceller (pink) is expected to be as close as possible to the near end test signal such that no echo and no near end speech clipping will occur. The most difficult part is at the beginning of the test signal where the level of the near end speech is low compared to the far end speech. In these cases it is common for echo cancellers to clip the near end speech but as can be seen in Fig 5.146, the echo canceller tracks the near end speech without clipping.

The following quote is taken from the ETSI report:

“The different level distribution on both sides leads to the result shown in Figure 5.146 and Figure 5.147 for the 14 dB ERL echo path.

*Under this test condition no signal level variation occurs in the analysis given in **Figure 5.146** for the low near end signal levels, i.e. at the beginning of the test signal. All bursts are completely transmitted.*

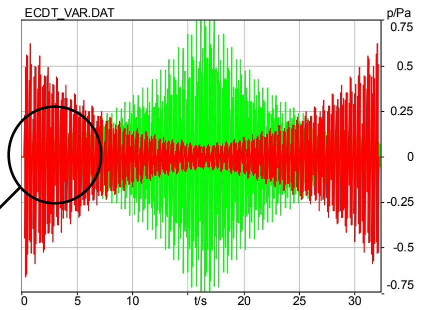


Fig. 5.145: Simulated double talk (II)

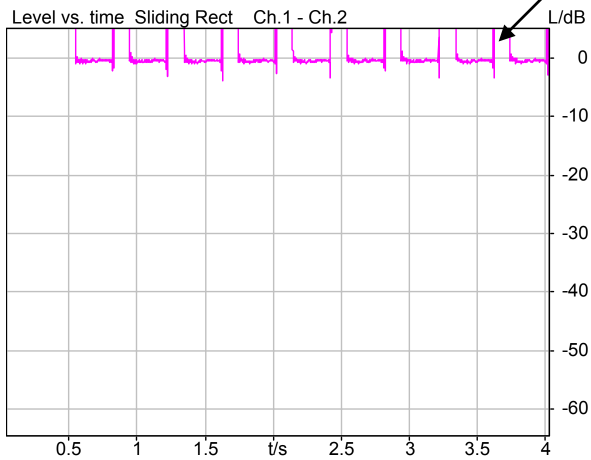


Fig. 5.146: Transmission of near end signal (beginning), 14 dB ERL

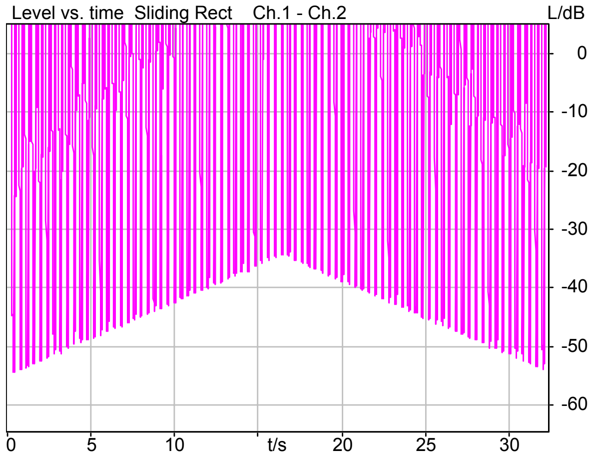


Fig. 5.147: Echo attenuation during double talk (whole sequence), 14 dB ERL

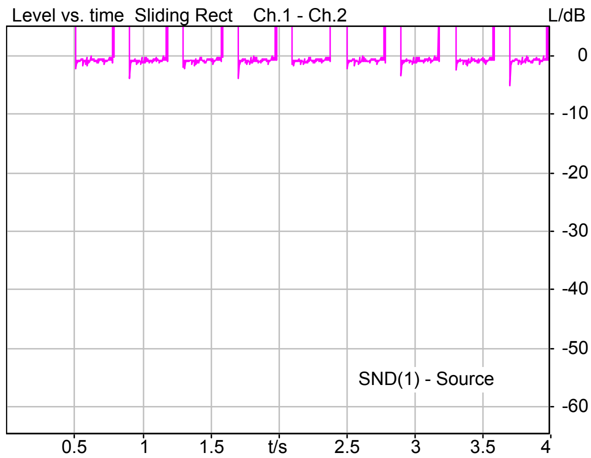


Fig. 5.148: Reference measurement

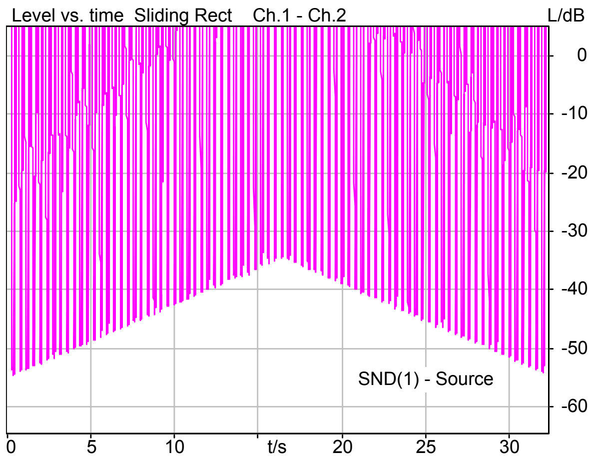


Fig. 5.149: Reference measurement

The echo attenuation is high and the result shown in Figure 5.147 is comparable to the reference measurement in Figure 5.149.
The listening example on CD also demonstrates that the near end signal (male voice) is transmitted without audible degradation, no syllable clipping occurs". [ETSI]

Echo Canceller Background Noise Test

In this test, a noise signal is injected to the near end speech input of the echo canceller and in the same time a speech signal is injected to the far end input (gray) in order to produce an echo that will disturb the adaptation of the echo canceller Comfort Noise Generator. The level of the output of the echo canceller is measured while the echo is present (green) and while the echo is absent (black). By comparing the black and the green contours, the level tracking of the NLP and CNG modules of the echo canceller is tested. As can be seen from Figure 5.152 below, the Echo Canceller Comfort Noise Generator tracks very well the level of the background noise in the presence of the echo signal.

The following part is taken from the ETSI report:

"The background noise transmission tests measured with the application of far end signals with this 14 dB ERL echo path are analyzed in the Figures 5.152 and 5.153. The level versus time analyses of the transmitted background noise signals with and without far end CS signals again demonstrate that the NLP gets active during the application of the far end CSS. The injected comfort noise level is very similar to the original background noise level. The noise contrast is very low, the implementation and its performance can be characterized as smooth and pleasant.
The corresponding listening example can be found on CD ("14 dB ERL - pub with far end speech", "14 dB ERL - cafe with far end speech")" [ETSI]

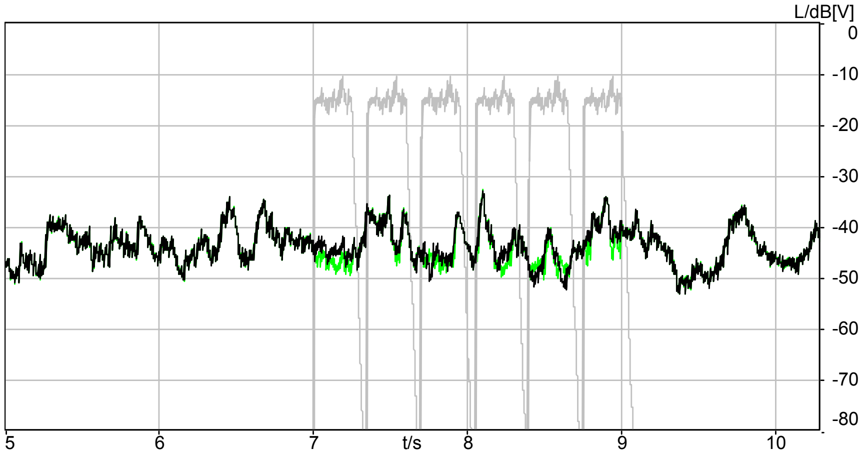


Figure 5.152: Noise from pub, comparison of transmission without far end signal (black), with far end signal (green) and far end signal (grey), **14 dB ERL**

Summary

The Speech Quality testing methodology used in the ETSI SQTE was presented. The summarized AudioCodes Mediant 2000 test results were also presented together with some specific examples from the ETSI test report. The Speech Quality as reflected by the results is outstanding; it meets and exceeds all the requirements.

AudioCodes Mediant™ 2000 was the only gateway which passed and exceeded all the requirements.

For more information, please refer to [8].

References

[1] http://www.etsi.org/plugtests/History/2004SQTE.htm
[2] ITU Recommendation P.862 (02/2001) Perceptual evaluation of speech quality (PESQ), an objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs
[3] http://www.etsi.org/pressroom/Previous/2004/2004_07_sqte.htm
[4] http://www.etsi.org/plugtests/History/DOC/3rd%20SQTEAnonymized%20ReportGatewaysupdate.doc

[5] ITU-T Recommendation P.340: Transmission Characteristics and Speech Quality Parameter of Hands-free Telephones
[6]] ITU-T Recommendation G.168 (2002): Digital Network Echo Cancellers
[7]] ITU-T Recommendation P.502: Objective analysis methods for speech communication systems using complex test signals.
[8] http://www.audiocodes.com/Content.aspx?voip=2189
[9] EG 201 377-1: Speech Processing, Transmission and Quality Aspects (STQ); specification and measurement of speech transmission quality; part 1: Introduction to objective comparison measurement methods for one-way speech quality across networks.

About AudioCodes

AudioCodes Ltd. (NASDAQ: AUDC) enables the new voice infrastructure by providing innovative, reliable and cost-effective Voice over Packet technology and Voice Network products to OEMs, network equipment providers and system integrators. AudioCodes provides its customers and partners with a diverse range of flexible, comprehensive media gateway and media processing technologies, based on VoIPerfect™ – AudioCodes' underlying, best-of-breed, core media gateway architecture. The company is a market leader in voice compression technology and is a key originator of the ITU G.723.1 standard for the emerging Voice over IP market. AudioCodes' voice network products feature media gateway and media server platforms for packet-based applications in the converged, wireline, wireless, broadband access, and enhanced voice services markets. AudioCodes enabling technology products include VoIP and CTI communication boards, VoIP media gateway processors and modules, and CPE devices. AudioCodes' headquarters and R&D facilities are located in Israel with an R&D extension in the U.S. Other AudioCodes' offices are located in Europe, the Far East, and Latin America.

International Headquarters

1 Hayarden Street, Airport City
Lod, Israel 70151
Tel: +972-3-976-4000
Fax: +972-3-976-4040

US Headquarters

2099 Gateway Place, Suite 500
San Jose, CA 95110
Tel: +1-408-441-1175
Fax: +1-408-451-9520

info@audiocodes.com
www.audiocodes.com

Disclaimer

LTRT-00430 Ref. # 07/05 V.1

Notice

The information contained in this document is believed to be accurate and reliable at the time of printing. However, due to ongoing product improvements and changes in the business environment, AudioCodes does not assume any obligation to update or correct any information and does not guarantee or assume liability for the accuracy of the printed material nor can it accept responsibility for errors or omissions. Original Date Published: May 2005