Title
Summary Report - Listening Test on Teleconference Devices

Journal no. 
SenseLab 009-13  
Project no. 
1100245  
Our ref. 
JRA, SVL, TOS

Client 
Cisco  
50 W Tasman Dr  
San Jose, CA 95134

Summary
A two-part listening test was performed on recordings of the Cisco® Unified IP Conference Phone 8831 (CP-8831) and Polycom SoundStation IP 7000 (IP 7000).

Industry-standard listening test methodologies were applied in assessing different aspects of the perceived audio quality on these recordings.

The results from the listening tests overall showed statistically significant better performance of the CP-8831 over the IP 7000.

The largest difference between the two products was found in the uplink situation, where the evaluation of the background noise showed strong performance of the CP-8831 over the IP 7000.

Figure 1 shows the results (error bars on the figure include 95-percent confidence intervals), and Figure 2 shows the uplink mean score ratings for the three test questions related to: distortion of speech signal, level of background noise, and overall quality as averaged across the three test locations and languages and experiment conditions.

Figure 1 Receive-End Mean Score Ratings Averaged Between Test Locations and Experimental Conditions
Figure 2 Uplink Mean Score Ratings (including 95-percent confidence intervals) for the Three Test Questions Related to Distortion of Speech Signal, Level of Background Noise, and Overall Quality Averaged Across the Three Test Locations and Languages and Experiment Conditions

*This document is a summary report. For full description of listening test, please refer to the document “SenseLab 005-13”.

The report must not be reproduced, except in full, without the written approval of DELTA.
Test Procedure

The test was divided into two separate parts, evaluating receive-end and uplink performance separately. The recordings captured the performance of the devices in a range of conditions, including narrowband (NB, 300–3400 Hz) and wideband (WB, 50–7000 Hz) settings for three languages.

All listeners were presented with written instructions, and completed a small familiarization module prior to starting each of the separate parts. The familiarization module introduced the listener to the variety of test material as well as test methodology and scales.

The test was performed on individual listening stations or booths, where the listeners were seated and given written and verbal instructions. They then proceeded to listen to recordings of the products and give ratings according to the methods suggested in the ITU-T P.800 ACR and ITU-T P.835. To avoid any order-of-presentation effect, each listener was presented with a unique test randomization in a double-blind paradigm. None of the listeners was informed about which products were under test.

A total of 96 listeners participated in the ITU-T P.800 ACR test, and 85 listeners in the ITU-T P.835 test. The tests were performed in three countries using native speaking listeners.

Results: Receive End

Receive-end performance captures how the device reproduces the speech signal from the incoming call. It is the audio quality that is experienced when listening to the device in real use cases.

Differences in performance were found for both narrowband and wideband settings. The magnitude of difference between the performances is in the range of 0.2 Listening Quality Mean Opinion Scores (LQ-MOS). All systems tested significantly different (P<0.001). Refer to Error! Reference source not found. for details.
Results: Uplink

Uplink performance describes the quality of the signal sent from the device. It includes the speech signal and some level of background noise. Uplink performance is typically not experienced directly by the device user, but instead influences how the receivers of the call experience the audio quality.

Three test questions are used to assess the following parameters:

- **Background noise**: The CP-8831 performed significantly better than the IP 7000 in suppressing the background noise (simulated ventilation noise). The magnitude of difference is between 0.8 LQ-MOS for narrowband and 1.0 LQ-MOS for wideband settings.

- **Distortion of speech**: The two systems performed statistically similar in wideband settings. For narrowband the CP-8831 performed significantly better in the range of ~0.1 LQ-MOS.

- **Overall quality**: The CP-8831 performed significantly better than the IP 7000 in the overall quality test question. The magnitude of difference is in the range of ~0.3 LQ-MOS. Refer to Error! Reference source not found. for details.

DELTA SenseLab

DELTA is a Danish technology company that is a part of GTS – Advanced Technology Group, approved by the Danish Ministry of Science, Innovation, and Higher Education. DELTA SenseLab is an independent test laboratory specializing in performing listening and perceptual tests of audio codecs, audio hardware, hearing aids, and telecommunication devices.

DELTA, 02 July 2013

Jesper Ramsgaard  
Specialist, SenseLab

Søren Vase Legarth  
Head of Dept. SenseLab

*This document is a summary report. For full description of listening test, please refer to the document “SenseLab 005-13”.

The report must not be reproduced, except in full, without the written approval of DELTA.