



Skype Hardware Certification Specification

Audio Requirements for Computer Accessories and Computers

Version 8.3.1_ACC

Disclaimer: This document is provided "as-is". Information and views expressed in this document, including URL and other Internet Web site references, may change without notice. You bear the risk of using it.

This document does not provide you with any legal rights to any intellectual property in any Microsoft product. You may copy and use this document for your internal, reference purposes.

Feedback: You have no obligation to give Microsoft any suggestions, comments or other feedback ("Feedback") relating to the materials herein ("Materials"). However, any Feedback you voluntarily provide may be used in Microsoft products and related specifications or other documentation (collectively, "Microsoft Offerings") which in turn may be relied upon by other third parties to develop their own products. Accordingly, if you do give Microsoft Feedback on any version of these Materials or the Microsoft Offerings to which they apply, you agree: (a) Microsoft may freely use, reproduce, license, distribute, and otherwise commercialize your Feedback in any Microsoft Offering; (b) you also grant third parties, without charge, only those patent rights necessary to enable other products to use or interface with any specific parts of a Microsoft product that incorporate your Feedback; and (c) you will not give Microsoft any Feedback (i) that you have reason to believe is subject to any patent, copyright or other intellectual property claim or right of any third party; or (ii) subject to license terms which seek to require any Microsoft Offering incorporating or derived from such Feedback, or other Microsoft intellectual property, to be licensed to or otherwise shared with any third party.

© 2013 Microsoft Corporation. All rights reserved.

Document Version

Version: 8.3.1_ACC	Status: Final
Security Classification: Public	
Last Saved: 2013-10-29	
By: Ergo Esken	
Filename: CSpec_ACC_Audio_v8.3.1	

Copyright Notice

©2013 Skype. All rights reserved.

Trademarks

Skype™ is a trademark of Skype.

Windows™ is a registered trademark of Microsoft Corporation in the United States and other countries.

All other product and company names are the property of their respective owners and are hereby acknowledged.

Contact Us



HW Certification Team
Skype Technologies OÜ
Akadeemia tee 15B, Tallinn
12618 ESTONIA

Revision History

Version	Date	Comments	Valid
8.3.1	2013-03-12	DUT Editor snapshots are replaced with description under tests. Note added about the possibility that Disable AGC functionality might not work with some clients in DUT Editor. It must be verified and if not working then adjusted manually. Ref/DUT Editor 3.0.100 description is added to chapter 7.	2013-03-15
8.3	2012-12-14	Diffuse field correction specified to be used for POLQA testing for Headset and Handset UI. Free Field correction to be used for POLQA testing for speakerphone UI. Loud, normal, quiet speech test sequence altered to allow better automation with ACQUA system. Analog Gain Control enabled for some Send path and Echo path tests to allow Skype to tune analog input gain to suitable level and avoid input overloads. Alternate 1/3 rd octave test signal allowed for THD tests, to lessen the impact of acoustic reflections. AEC test signal images updated with the description of new real speech based signal.	2013-01-15
8.1	2012-08-16	Several changes to testcases and introduction of new testcases has been done to more closely align with the Microsoft Lync Logo program requirements. http://technet.microsoft.com/en-us/lync/gg278181 The following is the summary of most important changes: Testcases using Microsoft Device Comformance Tool added to better test the device/driver level performance. IEEE recommended male and female real speech signals are used for majority of the testcases instead of the previous ITU-T P.50 artificial speech samples. The new ITU-T P.863 (ex P.OLQA) objective speech quality analysis tool is used alongside the PESQ tool. Most frequency response masks have changed to more closely align with Lync Logo requirements (and TIA and IEEE standards for Wideband telephony). The headset and handset receive direction testing uses ERP/DRP correction instead of the previous DRP/DF correction. The receive direction levels are measured in dBSPL instead of the A-weighted SPL level. The preferred listening levels have been changed to same with Lync Logo requirements and upcoming IEEE standard. Single frequency interference test is introduced. Weighted terminal coupling loss – TCLw requirement has been added under each Echo Path test sequence. The Echo Path tests use IEEE real speech samples instead of the	2013-01-15

		<p>artificial speech samples.</p> <p>The test conditions have been changed such that the Skype analog gain control (AGC) is disabled during the testing. Prior to test the AGC is set manually to a level that satisfies the send path signal level criteria with normal speech.</p>	
7.0	2011-08-30	<p>Device under test Skype client uses same setting throughout the tests (Skype Out Volume, Noise Suppression and Digital Gain are disabled – AGC control and Skype Echo Cancellation remain active for all tests).</p> <p>Moved the requirements for following under test entry criteria</p> <ul style="list-style-type: none"> • Analog gain adjustment latency • Device – Sampling frequency accurac • No excessive DC-offset in device’s microphone recording <p>Fixed the definition of preferred listening level.</p> <p>Echo path delay requirement changed.</p> <p>Headset:</p> <p>Speech to background noise requirement removed</p> <p>Group Speakerphone:</p> <p>Send path testing changed to 1m physical distance only</p> <p>0,7m speech level tests test cases deleted to optimize test time</p> <p>Send path frequency response mask updated</p> <p>Long-Range Speakerphone:</p> <p>Send path testing changed to 1m physical distance only</p> <p>Some 1m speech level test cases deleted to optimize test time</p> <p>Send path frequency response mask updated</p> <p>Preferred loudness level amended</p> <p>Receive path SpNR lowered to 40dB</p> <p>Real room test specification is a separate document from now on.</p> <p>Linked to this specification through chapter 5.8</p> <p>Reference TV-s that will be used for camera and microphone only Long Range solution testing are defined in section 7.1.13</p>	2011-09-01
6.0.5	2010-09-05	<p>Defined 8N as a default force between HATS right ear and handset for handset measurements.</p> <p>3.4.5 Echo path test case has been moved to 3.2.14 to simplify testing flow in ACQUA Skype Standard</p> <p>4.3.4 Echo path test case has been moved to 4.2.12 to simplify testing flow in ACQUA Skype Standard</p> <p>The correction for artificial ear changed for Speakerphone Acoustic UI group from Diffuse field to Free Field correction.</p> <p>The free field microphone and ITU P.51 compliant mouth simulator listed as acceptable alternative to Head and Torso Simulator for Speakerphone UI group devices testing – see Measurement setup.</p> <p>The Skype-SA option package (Code:60000) from HEAD Acoustics listed under test setups section.</p>	2010-10-01
6.0.4	2010-06-21	<p>Changed the output of 4.3.4.</p>	2010-07-01

		Corrected tables in outputs of: 5.3.10; 5.4.15; 5.5.15 Changed image of real room test environment for long range speakerphone solutions.	
6.0.0	2010-05-31	Send path, loud speech level test signal changed to female artificial speech. Speakerphone AEC test cases better defined Real room test setups and process defined New test case for send path speech quality in real environments. Long range speakerphone section measurement conditions better defined.	2010-07-01
5.0_acc	2009-12-01	Requirements separated into different documents for PC/MAC accessories and devices build on Skype SDK Full rewrite of test case descriptions	2010-04-01

Table of Contents

Document Version	2
Copyright Notice	2
Trademarks	2
Contact Us.....	2
Revision History	3
Table of Contents.....	6
1.0 Introduction	10
1.1 Audio UI Groups.....	11
1.1.1 Headset Audio UI Group	11
1.1.2 Handset Audio UI Group	12
1.1.3 Speakerphone Audio UI Group	12
1.1.4 Other Audio Product Group	14
1.2 Audio Requirements, Priorities, and Quality Ratings.....	15
1.2.1 Audio performance	15
1.2.2 Quality expectations for the audio UI groups.....	15
1.3 Definitions	16
2.0 General Audio Requirements Valid for All Groups	27
2.1 General requirements.....	27
2.1.1 Wind/Puff filter (headset/handset)	27
2.1.2 Stereo audio rendering	27
2.1.3 Microphone arrays (if DUT is built for MS array signal processing)	27
2.2 Requirements to be tested using Microsoft UC Device Logo Conformance Test Tool.....	28
2.2.1 Sampling Frequency Accuracy	29
2.2.2 Bit depth.....	29
2.2.3 Time Stamp	29
2.2.4 Clock Synchronization	30
2.2.5 Latency	31
2.2.6 Glitch	32
2.2.7 Device input gain control	34
2.2.8 Coupling total harmonic distortion and noise (THDN).....	35
2.2.9 On-board audio processing.....	36
2.2.10 DC offset.....	39
2.2.11 Clipping of far-end signal due to microphone boost	40
2.3 All Groups: Audio Performance over Skype call	41
2.3.1 Echo path - round trip delay (Skype end to end test)	41
2.3.2 Send path - total quality loss (Skype end to end test)	42
2.3.3 Receive path - total quality loss (Skype end to end test).....	43
3.0 Headset Audio UI Group	44
3.1 Headset: Audio Test Instructions.....	44
3.1.1 Test environment.....	44
3.1.2 Measurement setup – Head and Torso Simulator (HATS)	44
3.2 Headset: Audio Performance Requirements	47
3.2.1 Send path - signal level with loud speech	47
3.2.2 Send path - signal level with normal speech (send loudness)	47
3.2.3 Send path - signal level with quiet speech.....	48
3.2.4 Send path - frequency response	48
3.2.5 Send path – speech signal to self noise ratio.....	50
3.2.6 Send path – speech signal to self noise ratio during speech	51
3.2.7 Send path – single frequency interference.....	52
3.2.8 Receive path – preferred loudness level in ear.....	52
3.2.9 Receive path – frequency response.....	53
3.2.10 Receive path – speech signal to noise ratio (SpNR).....	54
3.2.11 Receive path – single frequency interference	54

3.2.12	Weighted terminal coupling loss (TCLw)	55
3.2.13	Echo path – acoustic echo cancellation	56
3.2.14	Echo path – send path signal level during two way conversation	57
3.2.15	Echo path – sidetone delay	57
3.3	Headset: Loudness and frequency response stability requirements	58
3.3.1	Send path – frequency response stability	58
3.3.2	Receive path – earpiece frequency response stability	59
3.3.3	Receive path – loudness level adjustment range	60
3.3.4	Receive path – maximum loudness level	60
3.3.5	Weighted terminal coupling loss (TCLw) at max volume	61
3.4	Headset: Requirements for Skype Super Wideband Certification (Optional)	62
3.4.1	Echo path – round trip delay – Super Wideband quality	62
3.4.2	Send path – frequency response – Super Wideband	63
3.4.3	Receive path – frequency response – Super Wideband	64
3.6	Headset: Supporting Audio Documentation Requirements	65
3.6.1	Verifying supporting documentation for Headset Audio UI group	65
4.0	Handset Audio UI Group	66
4.1	Handset: Audio test instructions	66
4.1.1	Test environment	66
4.1.2	Measurement setup – Head and Torso Simulator (HATS)	66
4.2	Handset Mode: Audio Performance Requirements	69
4.2.1	Send path – signal level with loud speech	69
4.2.2	Send path – signal level with normal speech	69
4.2.3	Send path – signal level with quiet speech	70
4.2.4	Send path – frequency response	71
4.2.5	Send path - speech signal to self noise ratio	72
4.2.6	Send path - speech signal to self noise ratio during speech	73
4.2.7	Send path – single frequency interference	73
4.2.8	Receive path - preferred loudness level in ear	74
4.2.9	Receive path - frequency response	74
4.2.10	Receive path - speech signal to noise ratio (SpNR)	76
4.2.11	Receive path – single frequency interference	76
4.2.12	Weighted terminal coupling loss (TCLw)	77
4.2.13	Echo path - acoustic echo cancellation	78
4.2.14	Echo path – send path signal level during two way conversation	78
4.2.15	Echo path – sidetone delay	79
4.3	Handset Mode: Loudness and Frequency Response Stability Requirements	80
4.3.1	Send path - frequency response stability	80
4.3.2	Receive path – earpiece frequency response stability	81
4.3.3	Receive path – loudness level adjustment range	82
4.3.4	Weighted terminal coupling loss (TCLw) at max volume	83
4.3.5	Ring tone loudness	84
4.4	Handset in Speakerphone Mode (if Available): Audio Performance Requirements	85
4.4.1	Send path - signal level with loud speech	85
4.4.2	Send path - signal level with normal speech	86
4.4.3	Send path - signal level with quiet speech	86
4.4.4	Send path - frequency response	87
4.4.5	Send path - speech signal to self noise ratio	88
4.4.6	Receive path - preferred loudness level in ear	88
4.4.7	Receive path - frequency response	89
4.4.8	Weighted terminal coupling loss (TCLw)	90
4.4.9	Echo path - acoustic echo cancellation	91
4.4.10	Echo path – send path signal level during two way conversation	92
4.5	Handset: Supporting Audio Documentation Requirements	93
4.5.1	Verifying supporting documentation for Handset audio	93

5.0	Speakerphone Audio UI Group	94
5.1	Speakerphone: Audio Test Instructions.....	94
5.1.1	Test environment – anechoic room	94
5.1.2	Measurement setup.....	94
5.2	Speakerphone: DUT Usage Distance 35..50cm (Handheld and Portable Speakerphone)	95
	Handheld speakerphone, tablet and laptop PC test position.....	95
5.2.1	Send path - signal level with loud speech	97
5.2.2	Send path - signal level with normal speech.....	97
5.2.3	Send path - signal level with quiet speech.....	98
5.2.4	Send path - frequency response	99
5.2.5	Send path - speech signal to self noise ratio.....	100
5.2.6	Send path - speech signal to self noise ratio during speech	101
5.2.7	Send path – single frequency interference	102
5.2.8	Receive path - preferred loudness level in ear	102
5.2.9	Receive path - frequency response.....	103
5.2.10	Receive path - speech signal to noise ratio (SpNR).....	104
5.2.11	Receive path – single frequency interference	104
5.2.12	Receive path – total harmonic distortion at preferred listening level.....	105
5.3	Speakerphone: DUT Usage Distance Up To 70cm (Personal Speakerphone).....	106
	Personal speakerphone recommended test position.....	106
5.3.1	Send path - signal level with loud speech	108
5.3.2	Send path - signal level with normal speech.....	108
5.3.3	Send path - signal level with quiet speech.....	109
5.3.4	Send path - frequency response	110
5.3.5	Send path - speech signal to self noise ratio.....	111
5.3.6	Send path - speech signal to self noise ratio during speech	112
5.3.7	Send path – single frequency interference	113
5.3.8	Receive path - preferred loudness level in ear	113
5.3.9	Receive path - frequency response.....	114
5.3.10	Receive path - speech signal to noise ratio (SpNR).....	115
5.3.11	Receive path – single frequency interference	115
5.3.12	Receive path – total harmonic distortion at preferred listening level.....	116
5.4	Speakerphone: DUT Usage Distance Up To 150cm (Group Speakerphone).....	117
	Group speakerphone test position.	117
5.4.1	Send path - signal level with loud speech	119
5.4.2	Send path - signal level with normal speech (1,5 m speech level)	119
5.4.3	Send path - signal level with quiet speech (1,5 m speech level)	120
5.4.4	Send path - speech signal to self noise ratio (1,5 m speech level)	120
5.4.5	Send path - speech signal to self noise ratio during speech (1,5 m speech level).....	121
5.4.6	Send path - frequency response	122
5.4.7	Send path – single frequency interference (1,5 m speech level).....	123
5.4.8	Receive path - preferred loudness level in ear	123
5.4.9	Receive path - frequency response.....	124
5.4.10	Receive path - speech signal to noise ratio (SpNR).....	125
5.4.11	Receive path – single frequency interference	125
5.4.12	Receive path – total harmonic distortion at preferred listening level.....	126
5.5	Speakerphone: DUT usage Distance Up To 5m (Long Range Speakerphone)	127
	Long range speakerphone test position in anechoic room.	127
5.5.1	Send path - signal level with loud speech (1m speech level).....	130
5.5.2	Send path - signal level with normal speech (1m speech level)	130
5.5.3	Send path - signal level with normal speech (4m speech level)	131
5.5.4	Send path - signal level with quiet speech (4m speech level)	131
5.5.5	Send path - speech signal to self noise ratio (4m speech level)	131
5.5.6	Send path - speech signal to self noise ratio during speech (4m speech level).....	132
5.5.7	Send path - frequency response	133

5.5.8	Send path - single frequency interference.....	134
5.5.9	Receive path - preferred loudness level in ear at 1m distance.....	134
5.5.10	Receive path - frequency response at 1m distance.....	135
5.5.11	Receive path - speech signal to noise ratio (SpNR) at 1m distance.....	136
5.5.12	Receive path – single frequency interference.....	136
5.5.13	Receive path – total harmonic distortion at preferred listening level.....	137
5.6	Speakerphone: Acoustic Echo Related Requirements – Anechoic Room Testing.....	138
5.6.1	Weighted terminal coupling loss (TCLw).....	138
5.6.2	Echo path - acoustic echo cancellation (with Skype AEC).....	139
5.6.3	Echo path – send path signal level during two way conversation.....	141
5.6.4	Echo path - acoustic echo cancellation (4m speech level) – applicable for Long Range UI only !.....	141
5.6.5	Echo path – send path signal level during two way conversation (4m speech level) - for Long Range UI only !.....	142
5.6.6	Weighted terminal coupling loss (TCLw) – max playback volume.....	143
5.7	Speakerphone: Audio Real Room Test Requirements.....	145
5.7.1	Audio Real Room Testing.....	145
5.8	Speakerphone: Requirements for Skype Super Wideband Certification (Optional).....	146
5.8.1	Echo path - round trip delay – Super Wideband quality.....	146
5.8.2	Send path - frequency response – Super Wideband.....	147
5.8.3	Receive path - frequency response – Super Wideband.....	148
5.9	Speakerphone: Supporting Audio Documentation Requirements.....	149
5.9.1	Verifying supporting documentation for Speakerphone audio.....	149
6.0	Other Audio Product Group.....	150
6.1	Other Audio Product: Audio Test Instructions.....	150
6.1.1	Objective testing measurement setup.....	150
6.2	Other Audio Product: Analog In/Out Interface Product (Soundcard).....	150
6.2.1	Output socket for headphones.....	150
6.2.2	Input socket for headset microphone or professional microphone.....	150
6.2.3	Microphone input socket - Frequency response.....	151
6.2.4	Headphone output socket frequency response.....	152
6.3	Other audio product: Supporting audio documentation requirements.....	153
6.3.1	Verifying supporting documentation for Other audio product.....	153
7.0	Test Setup And Test Environment Details.....	154
7.1.1	Microsoft Device Conformance Test Tool test setup.....	154
7.1.2	Objective test measurement setup.....	155
7.1.3	Test setup hardware/software availability.....	155
7.1.4	Lossless network condition test setup.....	156
7.1.5	List of equipment at Skype AudioLab.....	157
7.1.6	Reference Skype client setup details.....	158
7.1.7	DUT Skype client setup details.....	160
7.1.8	Head and Torso simulator and calibrations.....	161
7.1.9	Test signals used.....	162
7.1.10	Speech to noise ratio calculation example.....	163
7.1.11	Skype AudioLab – anechoic room.....	167
7.1.12	Compatible testing environments.....	168
7.1.13	Reference TV-s for Long Range Speakerphone UI group testing.....	169
8.0	Appendix.....	170
8.1	References.....	170

1.0 Introduction

This specification defines the audio requirements for Accessories built for PC type of computers on Skype Certified level.

This document emphasizes those audio requirements that significantly impact the end user experience—speech qualities, signal noise, echo/delays, and so forth. Test cases based on these audio requirements do not replace other necessary testing and certification(s) that a vendor should and must perform, such as those relating to health and safety regulations, product durability, and so forth.

The technical requirements listed and respective testing methodologies in this document have been derived from earlier versions of Skype and Microsoft Lync video requirements.

This document applies only to solutions that can be used with the public client of Skype for Windows. Other audio capable solutions will be tested against the CSpec_Audio_SDK test specifications.

Passing the requirements is a pre-condition to associate the device with Skype branding.

Skype reserves the right to update the contents of this technical specification at any time without prior notice. Purposes of such updates include the capture of new capabilities in Skype platforms, new device categories, as well as performance improvements in the hardware used in peripheral devices.

In addition to the audio requirements, any product under test must comply with Skype Functionality Specifications and Skype Video Specification (if applicable), which can be downloaded from [Skype Developer Zone](#).

1.1 Audio UI Groups

Audio requirements fall into one of the following groups, according to the device's acoustic or electric user interface (UI) type:

- Headset audio UI
- Handset audio UI
 - Handset mode
 - Handset in speakerphone mode (if available)
- Speakerphone audio UI
 - Mouth to microphone distance up to 0.35m (*Handheld Speakerphone*)
 - Mouth to microphone distance up to 0.7m (*Personal Speakerphone*)
 - Mouth to microphone distance up to 1.5m (*Group Speakerphone*)
 - Mouth to microphone distance up to 5m (*Long Range Speakerphone*)
- Other audio products (without acoustic interface)

Usage distance refers to distance between the user mouth and microphone. Thus for example a TV solution with wireless microphone in remote control placed close to user would be in category of “group speakerphone” if the microphone is instructed to be placed not further than 1,5m maximum.

Each audio requirement group is based on international standards, common use case analysis, Skype's own technical and usability studies, and common sense.

1.1.1 Headset Audio UI Group

Headset audio UI products connect to a device running Skype, and consist of two main components – a mono or two stereo earpieces and a microphone. These components are assembled together so that the headset can be worn on the user's head or attached to the user's ear(s).

The microphone is typically no further than 20 cm from user's mouth or the mouth reference point (MRP) of a head and torso simulator (HATS).

General Skype Certification requirements categorize the Headset audio UI group as follows:

- **Monaural**
- **Binaural**

Examples of Headset audio UI devices are illustrated below:



1.1.2 Handset Audio UI Group

Handset audio UI products connect to Skype network, and consist of a small speaker and a microphone assembled into a single unit that the user holds in his hand and puts next to his/her ear when in a call. The form factor of these devices is similar to that of a landline or mobile phone.

Examples of Handset audio UI devices are illustrated:

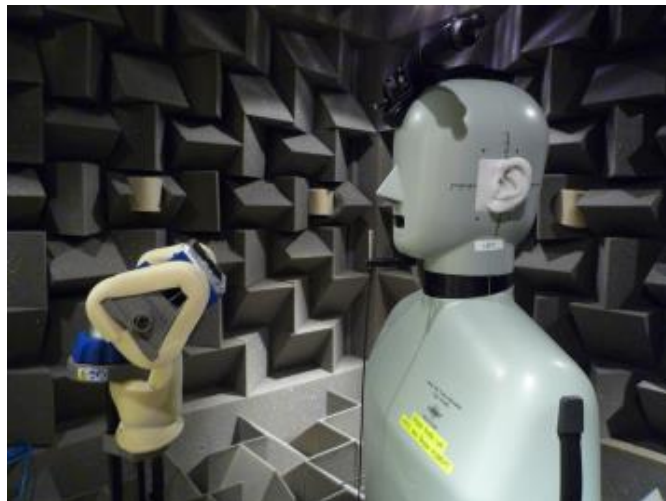


1.1.3 Speakerphone Audio UI Group

Speakerphone audio UI products typically include both a microphone and a loudspeaker in a single unit, making them a full audio solution with both send and receive signal paths. There can be devices that include only one of the two. For example, a webcam with a built-in microphone is categorized as a Speakerphone UI product, but it supports only the send signal path.

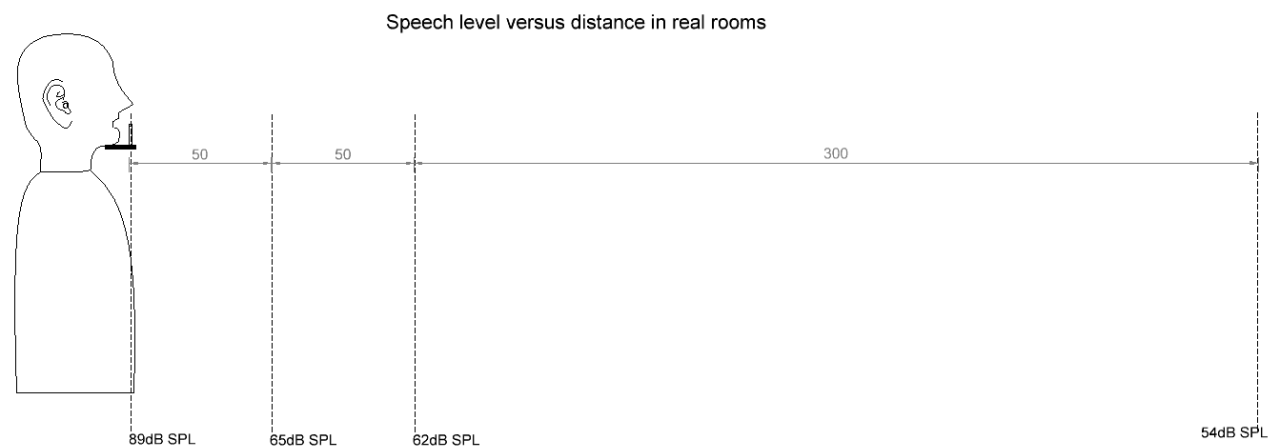
When using a speakerphone for Skype calls there are two main scenarios: **hand held** and **hands free**.

Handheld use cases are those where the device is held by hand in front of user, and the DUT microphone is no further than 35 cm from mouth or the mouth reference point (MRP) of a head and torso simulator (HATS).



Hands free use cases vary over a much wider range, with the DUT being placed at an arbitrarily far distance from user. This introduces three main phenomena:

- The user’s speech level decreases the farther the user is from the microphone.



- The loudspeaker level must be adjusted upward the farther the loudspeaker is from the user, which typically results in louder acoustic echo signals being picked up by the DUT microphone.
- The influence of room reflections and reverberation on the speech signal increase the farther the user is from the microphone.

Thus the hands free use case products are classified by usage distance:

- Mouth to microphone distance up to 0.35 m* (*Handheld Speakerphone*)
 - Mouth to microphone distance up to 0.7 m (*Personal Speakerphone*)
 - Mouth to microphone distance up to 1.5 m (*Group Speakerphone*)
 - Mouth to microphone distance up to 5 m (*Long Range Speakerphone*)
- * For the time being all portable devices running Skype for Windows are classified as Handheld Speakerphones in this test specification context. This considers tablet PCs, ultra-books and laptops running Skype for Windows.
- ** If the tablet is sold with a docking station which contains loudspeakers and/or microphones then those shall be tested as well.

Examples of Speakerphone audio UI devices are:

<p>Handheld and Portable Speakerphone</p> <p>(includes portable devices running Skype for Windows)</p>	
--	--

Personal Speakerphone (includes the All In One stationary type of computers running Skype for Windows)	
Group Speakerphone	
Long Range Speakerphone (accessories for living room solutions based on Skype for Windows)	

1.1.4 Other Audio Product Group

These products comprise various links in the audio signal chain for a Skype environment. While they **do not provide an acoustic user interface**, they can still have a significant impact upon the overall audio quality and the resulting user experience. For example, an interface device that provides a conversion of audio from one format to another might degrade the quality with additional delay, compression loss, bandwidth limitation, noise, distortion, interference problems, and so forth.

The products belonging to this group are for example the Analog Terminal (Telephone) Adapters (ATA).



1.2 Audio Requirements, Priorities, and Quality Ratings

Skype audio requirements ensure that Skype Certified products meet or exceed minimally acceptable performance level, that is, they have good sound quality and offer users an easy, stress-free conversation experience. Requirements related to the usability and functionality of a particular type of audio device, such as a handset, are described in other Skype certification reference documents. While many of the requirements detailed in this document apply equally to all of the audio UI groups listed in [section 2](#), individual audio UI groups can and do have their own unique, specific requirements.

1.2.1 Audio performance

Audio performance relates directly to the audio quality of the product under test. High-level performance parameters are concerned with subjective measures, such as intelligibility, naturalness, and conversational effort. Low-level performance parameters are concerned with objective, technical parameters such as frequency response, sensitivity, distortion, noise, and acoustic echo.

Intelligibility and naturalness are typically measured with subjective listening quality metrics. Since intelligibility can be difficult to quantify, the user's perception of the naturalness of a conversation is typically the same as their perception of its intelligibility. Conversational effort is typically related to how carefully a speaker needs to enunciate their words to be easily understood, and can be greatly affected by line delay and dropouts.

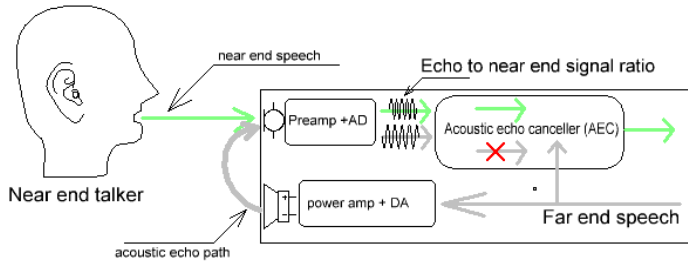
1.2.2 Quality expectations for the audio UI groups

End user expectations of a product's audio quality can vary according to:

- price—the more expensive, the higher the expectation
- brand—reputation for reliability and ease-of-use
- advertising claims—establish minimal expectations
- intended use—primarily concerned with ruggedness and reliability, but some features might be essential for or irrelevant in certain environments
- experience with similar products—should be as good if not better

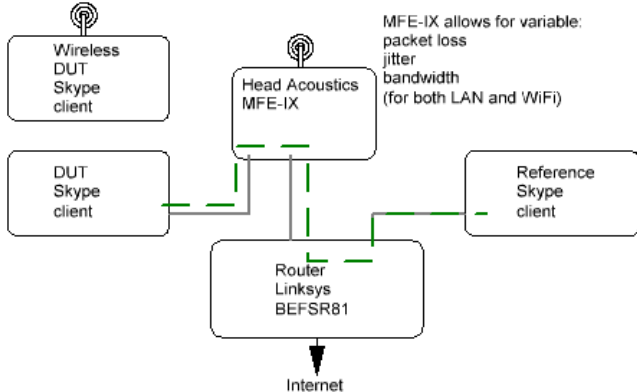
While most of the audio requirements are functionally based according to the audio UI group, several are technology dependent. In no case, however, are any requirements dependent on the MSRP of the product. As an example of a technology dependency, compare a narrowband cordless headset with a wideband plug-in headset. Because of bandwidth limitations inherent in the underlying technologies, a cordless headset cannot benefit from Skype's ability to deliver wideband and super wideband audio.

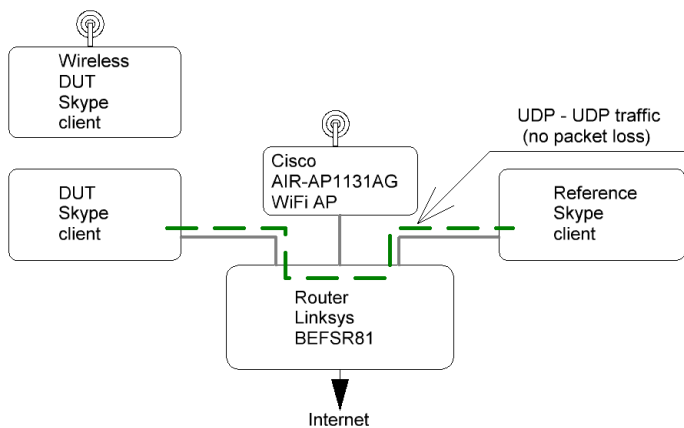
1.3 Definitions


A to D conversion	Device which converts continuous analog signals to discrete digital numbers
A-weighting	A frequency weighting curve defined in IEC179 and various other standards, widely used in sound level meters . A-weighting is an inverse curve for an equal loudness contour of human hearing at quiet levels (based on the 40-phon Fletcher-Munson curves). A-weighted measurements estimate how people perceive the loudness of a sound, allowing for the fact that human hearing has different sensitivities to different frequencies.
Acoustic echo	<p>Capture of the far end signal playback from a loudspeaker/earpiece by the DUT microphone. This signal must be cancelled so the far end speaker is not disturbed by his/her own voice echoing back after some delay period</p> 
Acoustic echo to near end speech ratio	The ratio of the acoustic echo signal to the near end speech signal in the sending path after AD conversion. The stronger is the echo signal, the greater the difficulty in canceling that signal. See Acoustic echo
Acoustic user interface	Enables a user to hear or speak over the communication system. Products providing acoustic UI have microphones, earpieces and/or loudspeakers. Check Section 1.2.
ACQUA (Advanced Communication Quality Analysis)	A dual channel analysis system for acoustic and electric transmission paths.
Active noise gating	A sound processing method that reduces background noise by muting a sound signal when it falls below a certain level, restoring it when the level increases again
Active speech part	Term for analyzing the speech level. Only the active speech parts are selected and not silences. Refer to test signals for examples.
AEC (Acoustic Echo Canceller)	A circuit or algorithm designed to eliminate acoustic echoes and prevent howling due to acoustic feedback from loudspeaker to microphone.
AGC (Analog Gain Control)	Programmatically adjustable analog amplification before the AD converter. Visible as input level slider in Windows / audio / recording settings.

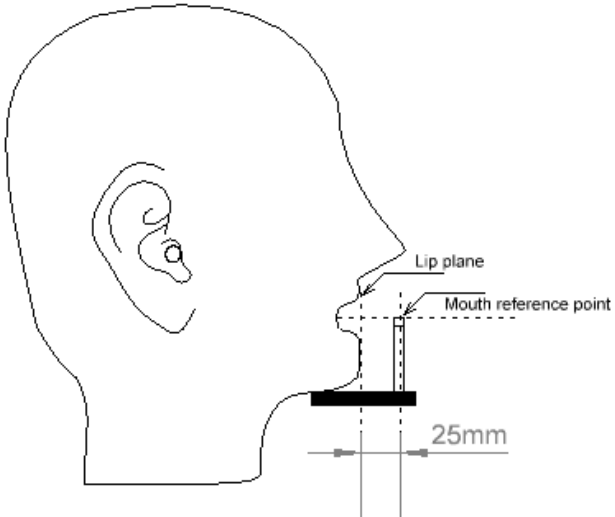
Automatic AGC	Adaptive system that runs inside the audio Codec and keeps the gain to an appropriate recording level for a range of input signals. Automatic AGC is not recommended feature for Skype/Lync devices as it creates a non linear action into the echo path and causes echo leaks during the calls.
Alternate position	Slightly modified position of DUT headset/handset from main test position to determine the stability of frequency responses.
Analog Telephone Adapter (ATA)	A device that enables you to make Skype calls through your computer using a standard PSTN handset.
Anechoic room	A test chamber lined with sound absorbing material to eliminate all internal sound reflections, and effectively isolated from all external noise sources, such as building mechanical systems (air handling, elevators), street noise, and so forth..
Application Programming Interface (API)	A set of routines, data structures, object classes and/or protocols provided by libraries and/or operating system services in order to support the building of applications
Artificial ear	A device used to measure the acoustic output of earphones. This acoustic ear incorporates an acoustic coupler and a calibrated microphone to measure sound pressure. The resulting combination has an overall acoustic impedance similar to that of the average human ear over a given frequency band.
Artificial mouth	A device consisting of a loudspeaker mounted in an enclosure and having a directivity and radiation pattern similar to those of the average human mouth. The frequency response of artificial mouth is compensated such that it provides a flat frequency response at MRP from 80Hz to 11kHz.
Artificial speech	A mathematically defined signal which reproduces human speech characteristics, relevant to the characterization of linear and non-linear telecommunication systems. It is intended to give a satisfactory correlation between objective measurements and tests with real speech. See male artificial speech test signal .
Background noise reference level	Any sound other than the sound being monitored. In this document diffuse Hoth noise at 62 dB SPL(A) level at HATS head position when HATS is not in sound field during level calibration is considered a reference level.
C50	C50, C custom The C50 parameter (Clarity or Klarheitsmass) is the early to late arriving sound energy ratio. It is defined as: $C_{50} = 10 \log \left(\frac{D_{50}}{1 - D_{50}} \right)$ and expressed in dB. C50 is calculate for early time interval up to 50 ms compared to late arriving sound after 50 milliseconds.
Certification	Process of testing a device/or solution against Skype Certification requirements. Please refer to Skype Certified mark.

Clipping	The distortion of an audio signal in which the tops of peaks with high amplitude are cut off, caused by, for example, overloading of amplifier circuits or AD converter.
Conversational quality	Defines how good or how bad are perceived audio parameters that affect conversational quality. Typical parameters are: delay, acoustic echo, noise and continuity of transmission.
Cordless handset	A handset that connects wirelessly to PC or other device through radio frequencies, such as a Bluetooth and DECT handset.
Cordless headset	A headset that connects wirelessly to PC or other device through radio frequencies, such as a Bluetooth headset.
dB / Decibel	A logarithmic unit of measurement that expresses the magnitude of a physical quantity (usually power or intensity) relative to a specified or implied reference level.
dBFS (dBov) (decibels relative to full scale)	<p>The signal level of a digital signal relative to its overload or maximum level is given by dBov. This is also commonly referred to as dBFS (Full Scale).</p> <p>For example, a rectangular function with only the positive or negative maximum number has a level of 0 dBov; For a maximum scale digital sine signal the peak level is 0 dBov and RMS is -3.01 dBov. (ITU-T G.100.1).</p> <p>This specification assumes that dBov and dBFS are equivalent, i.e., dBFS also assumes that the maximum scale digital sine signal the peak level is 0 dBov and RMS is -3.01 dBov.</p>
dBm0	<p>Abbreviation for the power in dBm measured at a zero transmission level point. For measurement setup for Skype devices this is the point between Reference Skype client soundcard and the measurement front end. Please refer to reference Skype setup for details.</p> <p>In practice the conversion from dBFS is as:</p> $Y \text{ dBm0} \approx X \text{ dBFS} + 6 \text{ dB}$
DC	Direct current.
DECT (Digital Enhanced Cordless Telecommunications)	Digital wireless communication standard, which is primarily used in cordless phone systems.

Degraded network condition	 <p>MFE-IX allows for variable: packet loss jitter bandwidth (for both LAN and WiFi)</p> <p>Wireless DUT Skype client</p> <p>DUT Skype client</p> <p>Head Acoustics MFE-IX</p> <p>Router Linksys BEFSR81</p> <p>Reference Skype client</p> <p>Internet</p> <p><u>Network simulator is used between DUT and network to introduce packet loss, jitter and bandwidth limitation IF defined in specific test cases.</u></p>
Delay	<p><u>send path delay</u> (time it takes for signal to get from DUT microphone to reference client soundcard output)</p> <p><u>receive path delay</u> (time it takes for signal to get from reference Skype input to playback in DUT loudspeaker/earpiece)</p>
Diffuse field correction (DF / DFC)	The overall frequency response from undisturbed diffuse-field to the Drum Reference Point of the simulated listener (HATS)
Double talk	Situation where two or more parties on a call are talking at the same time.
Drum Reference Point (DRP)	A point located at the end of the HATS ear canal, corresponding to the eardrum position in human ear.
DUT	Device under test.
Ear Reference Point (ERP):	A virtual point for geometric and acoustic reference located outside the entrance to the ear canal. The exact location is specified for each type of ear simulator.
Electric to electric Skype call	Skype call between two Skype clients that are measured from electric outputs and inputs of good quality sound cards. Acoustic interfaces are not present.
Far end	The DUT Skype client end is defined as near end. Far end is the other end of the Skype call – in measurement setup far end is same as reference Skype client. Refer to measurement setup for details.
Free Field Correction (FF)	The frequency response from undisturbed free-field to the Drum Reference Point. In this document the FF correction always means the 0 degree horizontal and vertical direction – ie sound coming from right in front of HATS.
Good quality reference device	See reference device .
Head and Torso Simulator (HATS)	Head and torso simulator used for sound quality and telecom measurements. Skype AudioLab uses a B&K 4128C.
High level volume setting	For a DUT having adjustable playback volume, “high level volume setting” refers to a level between 75% and 100% of the available range

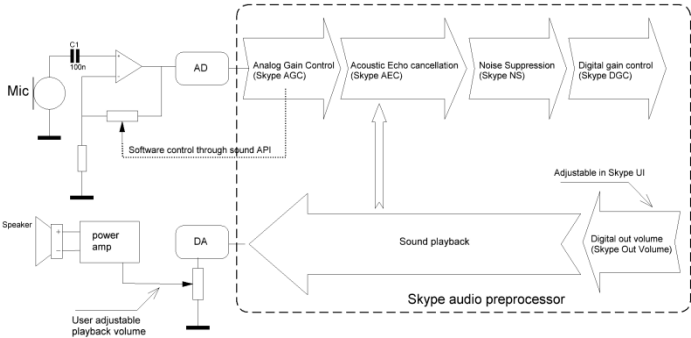
Hoth noise	Models indoor ambient noise when evaluating communications systems such as telephones, according to IEEE standard 269-2001 (revised from 269-1992), "Draft Standard Methods for Measuring Transmission Performance of Analog and Digital Telephone Sets, Handsets and Headsets."
Intelligibility	<p>A sub-parameter of audio performance, describing the ability to:</p> <ul style="list-style-type: none"> Recognize words and their meanings. Convey non-verbal information, such as the speaker's identity and emotional state.
ITU-T (Telecommunication Standardization Sector)	The primary standards organization for speech transmission and quality under the auspices of the International Telecommunication Union .
Listening quality	The subjective speech quality as perceived by the listener, specifically its intelligibility and naturalness .
Local network condition	 <pre> graph TD AP[Cisco AIR-AP1131AG WiFi AP] --- WDC[Wireless DUT Skype client] AP --- DSC[DUT Skype client] AP --- R[Router Linksys BEFSR81] R --- I[Internet] R --- RS[Reference Skype client] RS -.-> UDP - UDP traffic (no packet loss) AP </pre>
Local user	The person who is using a product under test.
Loudness	The subjective loudness of sound, as perceived by the listener.
Loud speech level	<p>The level/volume/loudness of speech in loud conversation between people, typically as a result of being in a noisy environment. In technical terms, this is approximately 10 dB SPL greater than normal speech level or 99 dB SPL when measured at 25 mm (Mouth Reference Point) and 75 dB SPL when measured at 50cm.</p> <p>Compare to quiet speech level and normal speech level.</p>
Loudspeaker	Converts electric audio signal to acoustic signal, used to play back speech to the user.
Lowered speech level	See quiet speech level .
Measurement front end	A device incorporating microphone preamplifiers, mouth amplifier, AES/EBU input and outputs

Measurement turntable	<p>A device allowing rotating the DUT around its center axis in horizontal domain by 5 degree or smaller steps.</p>  <p>For example</p>												
Microphone self noise	Residual noise, or the inherent noise level of a microphone when no signal is present and microphone resides in a very quiet environment.												
Midband speech (MB)	Speech transmission having a sampling frequency of 12kHz and a bandwidth of 100-6000 Hz.												
Mean Opinion Score	Check definition for <i>MOS</i>												
MOS (Mean Opinion Score)	<p>A numerical indication of the perceived quality of speech or audio. Typically an average of several listeners who have performed a specific MOS test in controlled and formal way. MOS scale is defined in ITU-T recommendation P.800 for speech quality as:</p> <table border="1" data-bbox="659 1014 940 1332"> <thead> <tr> <th>MOS</th><th>Quality</th></tr> </thead> <tbody> <tr> <td>5</td><td>Excellent</td></tr> <tr> <td>4</td><td>Good</td></tr> <tr> <td>3</td><td>Fair</td></tr> <tr> <td>2</td><td>Poor</td></tr> <tr> <td>1</td><td>Bad</td></tr> </tbody> </table> <p>In standardized tests, a good quality narrowband call, for example between mobile phones, can reach MOS slightly above 4. Wideband call can reach close to 5 in good conditions. MOS below 3 is generally considered to be too low.</p>	MOS	Quality	5	Excellent	4	Good	3	Fair	2	Poor	1	Bad
MOS	Quality												
5	Excellent												
4	Good												
3	Fair												
2	Poor												
1	Bad												
MOS-LQO	Mean Opinion Score – Listening Quality measured with Objective tools. See Opticom PESQ for example.												

Mouth reference point (MRP)	<p>Mouth reference point is a point 25mm in front of HATS lip plane. This is a point where the speech level is calibrated before measurements for normal, loud and quiet speech.</p> 
Narrowband speech (NB)	Speech transmission having a sampling frequency of 8kHz and a bandwidth of 100-3400 Hz. This is typical of PSTN (landlines) and mobile phones. Compare with wideband and super wideband .
Naturalness	How natural is listening (and speaking) in conversation. Technical parameters are: adequate loudness, natural frequency content, low noise and distortion. This is a sub-parameter of audio performance.
Near end	The end of HATS and where the DUT is connected to. Refer to measurement setup for details.
Non-acoustic user interface	Products that do not provide acoustic interface.
Normal speech level	<p>The level/volume/loudness of speech in normal conversation between people. In technical terms, this is approximately 62 dB SPL (A-weighted) when measured 1 m from the user's mouth. When measured at 25 mm (Mouth Reference Point), the level is defined to be -4.7 dBPa or 89.3 dB SPL according to ITU-T recommendations. In real life this level commonly varies +/-5 dB depending on the person.</p> <p>Compare to quiet speech level and loud speech level.</p>
Objective testing	Measures quality by means of technical measurement tools.
Preferred listening level	Preferred listening level is derived from face to face conversation between two people standing 1m from each other. The preferred level is 62dB SPL(A) in quiet environment and 65dB SPL(A) or even higher in noisier environments, when listening with two ears. When listening with one ear only (a handset use case for example), the preferred level is 70..75dB SPL (A)

PESQ (Perceptual Evaluation of Speech Quality)	<p>An ITU-T P.862.x compliant tool from that provides accurate and repeatable estimates of speech quality degradation that can occur over a telephony network. PESQ compares the audio signal sent over a network with the corresponding (degraded) audio signal received from that network.</p> <p>PESQ also identifies</p> <ul style="list-style-type: none"> • distortion artifacts • additional coding artifacts • temporal artifacts • additional noise and delay changes <p>PESQ does not identify acoustic interface artifacts.</p> <p>Skype emphatically <i>does not</i> use PESQ to measure the quality of an acoustic interface, as PESQ was not designed for nor has it been verified for acoustic interfaces. Skype <i>does not</i> use PESQ as an absolute metric for acoustic interfaces; Skype <i>does</i> use PESQ as a relative metric to compare an acoustic interface to a known reference device.</p>
P.OLQA	POLQA is the next-generation voice quality testing technology for fixed, mobile and IP based networks. POLQA was standardized by the International Telecommunication Union (ITU-T) as new Recommendation P.863 (2011).
Portable device	Lightweight and small devices running Skype for Windows and having webcam, microphone(s) and loudspeaker(s) integrated. For example: laptops, tablet computers, ultra-books and notebooks
Product	Part of a solution provided by submitting vendor, includes for example Handset, Cradle, Base station, dongle, application, and drivers (if available).
Purpose	Statement of the reasoning for the requirement that the test case supports.
Quiet speech level	<p>The level/volume/loudness of speech in relaxed conversations, typically when the caller does not wish to annoy others nearby or be overheard. In technical terms, this is approximately 10 dB SPL (A-weighted) less than normal speech level or 79 dB SPL when measured at 25 mm (Mouth Reference Point) and 55 dB SPL when measured at 50 cm.</p> <p>Compare to normal speech level and loud speech level.</p>
Receiving direction	The receiving direction refers to far end speech playback through the DUT loudspeaker/earpiece.

Reference device	The device selected from same audio-UI category as the device under test. The reference device exhibits good measured and subjective call quality performance and thus gives a reference of quality level achievable in given category with technical solutions available on market.
Reference Skype	Skype client PC to which the DUT Skype calls to during measurements.
Reference Skype input	Reference Skype input is a SPDIF input wired from measurement front end to RME9632 soundcard in reference Skype PC. Please refer to objective measurement setup for reference.
Reference Skype output	Reference Skype output is a SPDIF output wired from RME9632 soundcard in reference Skype PC to measurement front end input. Please refer to objective measurement setup for reference.
Recommended test position.	The most optimal real use like headset or handset test position on HATS which gives the flattest frequency response curves. For speakerphones the diagrams of recommended test position are given before each section.
Ring tone	A digital sound file that a device under test (DUT) plays to indicate an incoming call.
Round trip delay	Round trip delay is a sum of send path delay (time it takes for signal to get from DUT microphone to reference client soundcard output) receive path delay (time it takes for signal to get from reference Skype input to playback in DUT loudspeaker/earpiece) Please refer to objective measurement setup for reference.
RMS	Root Mean Square – a calculation method for average power of signal.
Self noise	Residual noise or the inherent noise level of a microphone when no signal is present and microphone resides in a very quiet environment.
Sending direction	The sending direction refers to near end speech transmission through the DUT sending path.
Signal to noise ratio (SNR)	The dB ratio between the signal level and the average noise level when a signal is not present. When determining this value, the surrounding environment should be very low noise, preferably an anechoic room with good sound isolation.
Simulated 4m test condition	Speakerphone UI products can have a long usage range. In Skype AudioLab the testing for maximum distance send path test cases are done using a simulated 4m test condition. This means that during anechoic room testing the TV stays physically at 1m distance, but the speech level in HATS mouth is decreased by 8dB respectively for each normal, quiet and loud speech for the maximum usage distance test cases. 8dB represents the speech level difference from 1m to 4m distance in average living rooms.
Skype Acoustic Echo Cancellor (Skype AEC)	Software algorithm for acoustic echo removal from send path input signal. Refer to Skype audio preprocessor.

Skype Analog Gain Control (Skype AGC)	Software algorithm for adjusting the analog microphone gain to yield an optimal level through the soundcard API. See Skype audio preprocessor .
Skype Digital Gain Control (Skype DGC)	Software algorithm for digitally adjusting the send path signal to yield an optimal level for a far end client. See Skype audio preprocessor .
Skype Noise Suppression (Skype NS)	Software algorithm for removing microphone self- noise and other stationary and unwanted noises. See Skype audio preprocessor .
Skype audio preprocessor	 <p>Refer to DUT client setup instructions for more info.</p>
Solution	The product + the latest Skype version
Sound Pressure Level (SPL)	<p>A measure of the change in ambient pressure resulting from a sound. SPL (in dB) is defined as:</p> $L_p = 20 \log_{10}(p/p_0)$ <p>where p is a sound pressure and p_0 is the reference level of 20 μPa.</p> <p>A 0 (zero) dB SPL describes a hearing threshold of silence at 1 kHz tone; the SPL of normal conversation varies between 50-75 dB. Unless otherwise indicated, SPL refers to the RMS power of the signal (the <i>effective</i> sound pressure as opposed to the <i>instantaneous</i> sound pressure).</p>
Speech to noise ratio (SpNR)	The dB ratio between the active speech level and the average noise level during the silence gaps in a conversation. When determining this value, the surrounding environment should be very low noise, preferably an anechoic room with good sound isolation.
SpNR speech sample modified SpNR speech sample	<p>Input signal used in speech to self noise ratio tests: IEEE 269-2010 uncompressed male speech sample with 1 second of silent gap inserted</p> <p>The modified SpNR speech has sine wave components inserted to the silent gap to keep the potential noise gating function deactivated.</p>
Subjective testing	Quality rating based on judgments of test subjects. This requires people to talk or/and listen and rate the quality.

Super wideband speech (SWB)	Speech transmission having a sampling frequency of at least 24kHz and a bandwidth of 50-11000 Hz. Compare with narrowband, midband and wideband.
TCLw	Weighted Terminal coupling Loss (dB)
Total Harmonic Distortion (THD)	The total harmonic distortion, or THD, of a signal is a measurement of the harmonic distortion present and is defined as the ratio of the sum of the powers of n (where n is variable) harmonic components to the power of the fundamental frequency. Lesser THD allows the components in a loudspeaker, amplifier or microphone or other equipment to produce a more accurate reproduction by reducing harmonics added by electronics and audio media.
THD+N THDN	Total Harmonic Distortion plus Noise.
Two-way delay	See round trip delay
Objective testing	Measures quality by means of technical measurement tools.
User Interface (UI)	Aggregate of means by which people - the users - interact with the device or computer program
Vendor	Manufacturer who is submitting a solution for Skype Certification.
Wideband speech (WB)	Speech transmission having a sampling frequency of 16 kHz and a bandwidth of 50-7500 Hz. The majority of Skype calls use wideband. Compare with narrowband and super wideband .

2.0 General Audio Requirements Valid for All Groups

2.1 General requirements

2.1.1 Wind/Puff filter (headset/handset)

- Purpose:** A wind filter is used to prevent breathing sound from being picked up by the microphone in a handset or headset. Without it, far-end participants may hear excessive breathing sound.
- Input:** Presence of hardware wind/puff filter should be verified visually or by blowing to microphone from the side.
- Required:** All handsets and headsets that position the microphone within a radius of up to 60cm from artificial mouth MRP or mobile headsets that are to be used outside buildings must have a wind filter implemented in hardware.
- Note:** Not applicable for speakerphones.

2.1.2 Stereo audio rendering

- Purpose:** Lync W15 2013 will support stereo audio capture for conference room systems which will be equipped with special stereo capture devices to allow transmission of stereo signals from such room systems. Note that the regular Lync 2013 clients will not support stereo capture and will only transmit monaural signal. Sending stereo signals from conference rooms. This will allow a better user experience for remote users due to the spatial audio perception which makes it easier to distinguish between the multiple users sitting in the conference room. To allow the user to listen to such stereo audio signals using the Lync 2013 client is scenario all Lync logo devices which have two loudspeakers must be able to render stereo audio.
- Input:** Play back a stereo test signal 7.1.9 and verify stereo playback and left and right channel correctness.
- Required:** A device that has two loudspeakers (e.g. binaural headset, PC) must be able to render a stereo audio signal.
- Some headsets can rotate the microphone boom such that they can be worn also with the left loudspeaker sitting on the right ear. It is recommended that such devices have a clear marking on the Left and Right ear, so that the users know which way the headset should be worn to get a correct stereo representation.
- Note:** Cordless devices using Bluetooth and DECT are exempt from this requirement. For Bluetooth devices the reason is that the headset and hands-free profile do not support wideband stereo signals. DECT systems would degrade in terms of wireless density due to the additional use of frequency slots for stereo. This exemption may be re-examined in future revisions of this specification.

2.1.3 Microphone arrays (if DUT is built for MS array signal processing)

- Purpose:** Microphone array signal processing depends on the accuracy of the API information given by the device under testing. This test is to verify whether the driver accurately reports

microphone-array geometry information and whether the Voice Capture DMO operating in microphone-array mode processes the captured data accurately.

Input: Use the tool from:

<http://msdn.microsoft.com/en-us/windows/hardware/gg462998>

This simple command-line tool verifies whether your driver accurately reports microphone-array geometry information and whether the Voice Capture DMO operating in microphone-array mode processes the captured data accurately. The tool should be used only to verify the following characteristics, which are required for proper functioning and integration of the microphone array processing algorithm.

- Audio client API - Verifies that the Voice Capture DMO operating in microphone array mode can read the Audio driver's property set that contains the microphone array geometry information.
- Microphone array format descriptor and geometry - Verifies that the microphone array descriptor format is accurate. The test also verifies and reports whether the microphone array geometry reported by the driver falls within the range specified by Microsoft.
- DMO source mode capture and processing - Verifies that this process is working correctly. This test should pass if the two previous tests are successful.

Required: If the device under testing contains a microphone array which is made available to Lync as a recording audio device, we recommend following the guidelines in the following two white papers:

- <http://msdn.microsoft.com/en-us/windows/hardware/gg462985>

- <http://msdn.microsoft.com/en-us/windows/hardware/gg462992>

It is important to note that the microphones need to be uni-directional and have a good signal to noise ratio of at least 60dB. Additionally the microphone array verification tool should be used to validate the correct implementation. We recommend that the spacing between the microphones reported by the tool should be identical to the actual microphone spacing. The tool and its description can be found here:

<http://msdn.microsoft.com/en-us/windows/hardware/gg462998>

Note: This test case is applicable only to speakerphone type devices built to support the Microsoft array signal processing.

2.2 Requirements to be tested using Microsoft UC Device Logo Conformance Test Tool

The requirements in this section will be tested with the device conformance tool. The tool is called Microsoft UC Device Logo Conformance Tool (UCDC tool). Currently the test tool is available only under

limited distribution to official Skype test labs and OEMs that are participating in the Skype or Lync Logo program.

Please refer to '7.1.1 Microsoft Device Conformance Test Tool test setup' for details on the test setup.

This test tool is not part of the HEAD Acoustics Skype AS test standard.

2.2.1 Sampling Frequency Accuracy

Purpose: If the device sampling rate deviates too much from the claimed sampling rate, the device will consume data either too fast or too slow for the render path which will lead to an audio buffer under run or overflow. When this happens, users may hear speech cut-outs or glitches in the loudspeaker signal. The same may happen with the microphone signal where again the speech signal will have cut-outs. In addition, such glitches may also cause echo leak to the far end. Another possible impact of high sampling frequency error is that the device clock synchronization algorithm may have problems synchronizing its sampling frequency with the PC accurately and again this may lead to echo.

Input: For USB devices first run the "Test PC Clock Frequency Accuracy Test" which will compare the PC clock of the reference PC with the network clock.

Choose the timestamp test from UCDC tool which will also evaluate the sampling frequency.

The tool will play and record a signal and will estimate the sampling frequency deviation based on the timestamps attached to the data sent to the device loudspeaker and the data captured by the device microphone.

An anechoic chamber is not required because the audio signal itself is not analyzed.

Required: Sampling rates for capture or render shall be one of the following: 48 kHz, 44.1 kHz or 16 kHz. The maximum deviation from claimed sampling rate for both, capture and render must be <0.04% when measured by a reference PC. The same requirement applies to PCs where the sampling rate will be measured without a reference PC.

2.2.2 Bit depth

Purpose: The bit depth requirement is to ensure that the digital signal has a sufficient dynamic range, so that the quantization noise is negligible.

Input: Choose the timestamp test from UCDC tool which will also evaluate the bit depth.

Anechoic chamber is not required because the audio signal itself is not analyzed.

Required: The A/D resolution shall be at least 16 bits.

2.2.3 Time Stamp

Purpose: The time stamp indicates the position of the render and capture stream with respect to the CPU clock. If the time stamp error is too high, then the acoustic echo canceller cannot align the loudspeaker and microphone signal stream very accurately and this can cause echo to appear in the call. Additionally it may be difficult to interrupt the far-end

participants. Also if the render time stamp has high errors, then audio may glitch and result in pops in playback.

Input: Select the DUT as the audio device in the UCDC tool.
Start the timestamp test. A signal will be rendered and captured at the same time.

Anechoic chamber is not required for this test because only the time stamp will be analyzed, not the actual audio signal itself.

Required: Time stamp requirements exist for three different signal streams: capture, loopback, and render stream. The time stamp error is defined as the difference between the actual time stamp and the time stamp computed based on claimed sampling rate. The maximum time stamp error is defined as the average value of the 1% highest absolute time stamp error. The requirements are:

	Capture stream	Loopback stream	Render stream
Maximum timestamp error	< 0.5ms	<0.5ms	<2ms
Timestamp error standard deviation	<0.04ms	<0.04ms	N/A

Time stamp is determined by device streaming positions (DevPos), application streaming position (AppPos) and system performance counter (QPC: query performance counter):

- DevPos: Device streaming position is the count of samples that device has captured/rendered. For USB audio devices, this position is reported by USB audio driver. For other types of devices, it is reported by device or driver.
- AppPos: Application streaming position is the sample counts that application has received from audio APIs (for capture) or sent to audio APIs (for loopback and render).
- QPC: Performance counter is used as a high resolution timer. QPC time should be queried at the same time that the device stream position is queried. In practice it is done one after the other.

For any capture or render frame in an application, there is an associated application stream position. Assuming that the audio stream sampling rate is FS, the time stamp is calculated as: $TS = QPC + (AppPos - DevPos) / FS$.

Note: Note that if the time stamp has a constant offset error it will not be reflected in time stamp error measurement. Instead it will affect the latency test. So the requirement for time stamp constant offset error is incorporated into the latency test.
The echo cancellation in Skype client is less sensitive to the timestamp errors, thus in case of failure against this test is reviewed and decision is made if the performance is allowed. Failing result against this test still does indicate some shortcoming in the device performance and vendor should try to fix the underlying issue.

2.2.4 Clock Synchronization

Purpose: It is usually necessary to synchronize the clock on a device with both capture and render with the clock on the PC. If the two clocks are too different, audio glitches may occur and the far end may experience echoes.

Input: This test is required only for devices that support both capture and render. It is not required for Skype PC.

The test shall be performed in an anechoic chamber:

1. Connect a reference microphone (for example, an analog table-top microphone with a 3.5-mm jack) to the PC.
2. Select the DUT to play back the test tone, and select the reference microphone for capture.
3. Disable all audio enhancements processing on the microphone if any.
4. To run the test, put the reference microphone next to the loudspeaker of the DUT. The distance should be small to have high SNR. Adjust speaker and microphone volume to have sufficient playback capture volume. To maintain high SNR, if the capture signal clips, always lower the reference microphone volume first.

The test will run for the default duration of 5 minutes.

Required:

The expected performance is the following:

1. The clock synchronization shall be stabilized two seconds after the device is plugged into the test computer.
2. After clock synchronization is stabilized, the device clock rate shall stay at one frequency or oscillate between two frequencies within certain distances. In the latter case the distance between the maximum frequency and the minimum frequency shall be less than 0.3 Hz
3. The clock adjustment shall occur only if the cumulative error is higher than one sample.

Capture-only devices such as webcams with an integrated microphone are exempt from this requirement.

Note:

The echo cancellation in Skype client is less sensitive to the clock synchronization errors, thus in case of failure against this test is reviewed and decision is made if the performance is allowed. Failing result against this test still does indicate some shortcoming in the device performance and vendor should try to fix the underlying issue.

2.2.5 Latency

Purpose:

In two-way communication applications, it is important to limit the end-to-end latency to ensure natural conversation. When the latency is too long, users are more likely to talk over each other and find it difficult to interrupt each other. The requirement makes sure that together with Skype, the end-to-end latency is appropriate for natural conversation.

Input:

This test needs to be performed in an anechoic chamber to ensure there is no background noise introducing errors to the test. The test procedure is slightly different for Skype devices and PC.

Skype devices (PC accessories):

1. Connect a reference microphone and a reference speaker to the PC. The reference speaker and microphone shall be a 3.5-mm jack device without any additional audio processing so that the device latency of the reference devices is very short, and can be estimated and subtracted from the overall latency.

2. Start the UCDC tool, select the latency test, and select “Reference Mic” and “Reference Speaker”.
3. Place the DUT and the reference speaker and microphone in the anechoic chamber, as close to each other as possible to minimize latency due to acoustic transmission. Make sure the reference microphone faces the reference speaker and the DUT speaker, and the DUT microphone faces the reference speaker. The distance from any pair of microphone to speaker shall be within 34 cm (latency error due to acoustic transmission <1 ms).
4. Adjust speaker and microphone volumes to ensure microphone capture does not clip and has sufficient SNR.

PC:

1. Start the UCDC tool on the PC.
2. Select latency test and select “PC” from the drop-down menu.
3. Place the DUT in the anechoic chamber and start the latency test.

Required:

The latency is defined as capture and render round-trip latency due to the device which is not reported to the OS. This does not include delays introduced by the operating system such as latency by the USB driver.

	Corded devices	Corded devices with AEC	Cordless devices without AEC	Cordless devices with AEC	Capture-only devices (e.g. webcams)	PC
Max latency	<50 ms	<70 ms	<50ms	<105 ms	<25 ms	<50 ms

It should be noted that a device is classified as “with AEC” only if it meets the TCLw requirements outlined in Sections 3.2.12 , 4.2.12 , 5.6.1 respectively for devices with AEC. In the case that there is built-in AEC which fails the TCLw requirements, the stricter latency requirement for devices without AEC will apply.

Note that the latency test depends on the time stamp accuracy as given by the requirements in Section 2.2.3. A constant time stamp error offset can cause a negative latency result. In this case the DUT is considered failing the requirements as this offset may affect the capability of the AEC to align the capture and render stream and therefore echo leaks may appear.

The latency is computed as following:

Latency = Acoustic Latency Measured – Time stamp Latency Measured – Reference Device Latency

Time stamp latency is computed as following:

Time stamp latency = loop back time stamp – capture time stamp

Time stamp latency includes the OS and driver latency.

2.2.6 Glitch

Purpose: Capture or render glitches affect audio quality where users will hear speech cut-outs. This should be minimized.

Input: This test shall be performed in an anechoic chamber:

1. In the UCDC tool, select the DUT microphone and speaker.
2. Select the reference microphone and speaker.
3. Place the reference microphone close to the DUT speaker.
4. Place the DUT microphone close to the reference speaker.
5. Start the test.

The tool will play back and capture a signal at the same time. It will then compute automatically the number of glitches observed during the test.

Required:

Under normal CPU conditions, the time interval between two glitches shall be 5 minutes or more. An audio glitch is defined as missing, or insertion of one sample or more of audio data. Glitches caused by the operating system are excluded from this measurement. (An operating system glitch is reflected in time stamp glitch/discontinuity. We ignore glitches caused from time stamp glitch.)

For cordless devices (e.g. Bluetooth or DECT) the adjustment to the clock rate mismatch is often implemented by removing or inserting several samples. For such devices up to one single- or double-sample insertion/removal glitch per second is allowed. Microsoft recommends that the sample insertion/removal is performed during a zero crossing so that perceptual impact of this operation is minimized. For glitches with insertion/removal of more than two samples the time interval shall be 5 minutes or more. Additional glitches caused by radio communication errors shall meet published standards from the appropriate industry groups.

Note:

The current version of the UCDC tool will still flag single- and double-sample removal/insertion as a regular glitch for cordless devices. An updated version of the UCDC tool which will allow such insertion/removal according to the requirement above will be released at a later date. In the interim cordless devices failing the glitch test will need to be tested subjectively.

2.2.7 Device input gain control

Purpose: Automatic gain control on devices interferes with Skype/Lync audio processing, and may cause echo leak or speech level fluctuations, therefore is not recommended. Instead it is recommended that the device allows applications to control the analog microphone capture gain by exposing it to the OS.

We recommend that the device exposes analog gain to the application as opposed to digital gain. This allows Lync to adjust the gain to avoid clipping when the microphone input from loudspeaker or local human speaker is clipping. Exposing digital gain will not help handling this situation.

Input: This test shall be performed in an anechoic chamber:

1. In the UCDC tool, select the DUT microphone.
2. Select the reference speaker.
3. Start the test.

Required: If the device or Skype PC supports analog gain changes controllable by the application, the accuracy of the gain adjustment shall be within +/-1dB. That is, if the application requires a gain increase of 3 dB, the actual gain change shall be within 2 to 4 dB. Also the gain change shall be applied within 200 ms of sending the gain change request message.

If the DUT is classified as a device with AEC (i.e. if it meets the TCLw requirements for devices with AEC as stated in Sections 3.2.12 , 4.2.12 , 5.6.1 respectively) then an automatic gain control is allowed.

If the DUT is classified as a device without AEC then the DUT must not have automatic gain control for either microphone or speaker.

2.2.8 Coupling total harmonic distortion and noise (THDN)

Purpose: To ensure appropriate echo cancellation and duplex performance in AEC, it is required that the device have sufficient linearity. Coupling total harmonic distortion and noise is chosen as a measure because it takes into account both harmonic and non-harmonic distortions.

Input: This test shall be performed in the anechoic chamber:

1. Select the coupling THDN test.
2. Maximize the playback volume (this includes PC volume and device volume) and adjust microphone gain to make sure that the microphone does not clip.
3. The test tool will inform the user if there is clipping in the microphone capture.

Required: If the DUT is classified as a device with AEC (i.e. if it meets the TCLw requirements for devices with AEC as stated in Sections 3.2.12 , 4.2.12 , 5.6.1 respectively) then there are no requirements for the THDN.

If the DUT is classified as a device without AEC and it has both audio capture and render then the coupling THDN within 100 to 7100 Hz must meet the following requirement:

$THDN(k_f) > 32\text{dB} - \text{Coupling loss}(k_f) + (\text{Send loudness}_{nominal} - \text{Send loudness}_{measured})$
The nominal send loudness is defined as -18dBm0 (-24dBFS)

The THDN is defined as following:

$$THDN(k_f) = 10 * \log_{10} \frac{\sum_{k=0}^K |H(k)|^2}{\sum_{k=0, k \neq k_f}^K |H(k)|^2}$$

k_f is the frequency the THDN is computed for.

The test is performed using 1/12th Octave steps sine waves as input. The power summation in the above equation is computed in linear frequency domain using DFT.

The normalization with respect to the coupling loss, that is, the attenuation of the fundamental frequency, ensures that devices with high coupling loss (for example, headsets) can achieve the required coupling THDN. In some cases however, the coupling loss may be artificially high because of low analog gain at the device microphone. This will be compensated for by the digital AGC in Skype. To account for this compensation a normalization of the coupling THDN with respect to the target send loudness is included. A sine wave with peak amplitude of -3dBFS is used for the measurement.

2.2.9 On-board audio processing

Purpose: If a device or PC relies on the Skype or Lync AEC, any time-variant processing in the speaker or microphone path will be perceived as an echo path change by the AEC and may lead to echo leak. Additionally, sound effects can be perceived as added reverberation which makes the task for the AEC much more difficult and again may lead to echo. If a device or PC has built-in AEC, the echo attenuation by the device alone must meet the requirement. Also, the end to end audio quality when used with Lync audio processing module must be ensured.

Input: Each device shall provide documentation of the types of included on-board audio processing when submitting the device for qualification.

The following linear and time-invariant on-board audio processing is allowed for all devices:

- Fixed beamforming that meets the directivity requirements.
Note that beam switching between several fixed beamformers is not allowed as it will be perceived as an echo path change by the AEC.
- Microphone or speaker fixed frequency equalization

Following is a non-exhaustive list of on-board nonlinear/time-variant digital signal processing that is not recommended for a device without AEC:

- Adaptive beamforming
- Fixed beamforming with beam switching logic
- Noise suppression
- Noise gating
- Acoustic echo cancellation
- Automatic gain control
- Automatic peak limiters for microphone
- Dynamic range compression including compressor or limiters. (A limiter implemented for acoustic safety reasons is exempt.)

For non-linear (e.g. dynamic range compression) or time-variant (e.g. automatic gain control) on-board processing the requirement differs for devices with and without AEC and will be explained in the following. The DUT can only be classified as a device with AEC if it meets the TCLw requirements for devices with AEC.

If on-board processing is included in the DUT verify and document which processing is available and which are enabled by default. Go to Control Panel->Hardware and Sound->Sound and then check the recording device properties as shown in figure below. For PCs running Windows 7 or 8, document which playback processing is available and which are enabled by default.

Required: **Cordless Devices:** A cordless device (for example, Bluetooth or DECT headset, handset or speakerphone) shall have built-in AEC and shall meet the TCLw requirements for devices with AEC. (Due to wireless link between the device and the PC that introduces extra delays and possible signal gaps and interruptions the Skype AEC cannot guarantee the AEC performance needed and thus the device needs an in built hardware AEC.

Corded USB Device/PC with AEC: In case the DUT is classified as a corded USB device or a PC with AEC, then time-variant and/or nonlinear processing is allowed, including automatic AGC. However, it should be pointed out that the Skype audio processing will be turned on as well and the echo cancellation and total quality loss tests must still pass.

Corded USB Device without AEC: When the DUT is classified as a corded USB device without AEC, the device shall not include any time-variant and/or nonlinear processing for both, the send signal and receive signal path. Additionally, any other sound effects such as introducing additional reverberation shall be turned off. It should be noted that this category also applies to webcams.

Following is a non-exhaustive list of on-board nonlinear/time-variant digital signal processing that is not allowed for a device without AEC:

- adaptive beamforming
- fixed beamforming with beam switching logic
- noise suppression
- acoustic echo cancellation
- automatic gain control
- dynamic range compression including compressor or limiters. (A limiter implemented for acoustic safety reasons is exempt).

PC without AEC

When the DUT is classified as a PC without AEC then depending on the operating system (OS) different requirements apply as described in Table below.

OS	Receive side processing	Send side processing
Windows 7	Don't include or turn off by default	Don't include or turn off by default
Windows 8	Allowed	Don't include or turn off by default

Unfortunately the Windows platform does not allow applications such as Skype to programmatically turn on/off on-board processing. Skype provides high quality DSP, including a robust AEC, AGC, and dynamic range compression and thus it is both unnecessary and counterproductive for the PC to perform any non-linear or time-variant processing on calls with Skype as this may lead to echo leaks or to audio quality degradations.

In case the PC manufacturer wants to provide value-added processing for other applications then the requirements above describe that this processing shall be disabled by default. Note this solution being problematic if the user turns on the on-board processing via the control panel and then performs a Skype (or Lync) call. In this case the Skype call quality may be severely affected by the on-board processing. Therefore it is highly recommended that PCs should not include any nonlinear or time-variant processing such as the non-exhaustive list above.

NB! There can be an exception to the above requirements if the requested default software processing modes for receive or send direction fulfill the following :

- The End to End testing results for the Total Quality Loss testcases 2.3.2, 2.3.3 remain same MOS-LQ or higher compared to when tested without all processing
- Echo Cancellation related testcases pass all the requirements when the requested extra processing is enabled.

Note: Additionally it should be pointed out that the Windows 8 Hardware Certification Requirements for devices (<http://msdn.microsoft.com/en-us/library/windows/hardware/hh748200.aspx>) require any non-linear audio processing to be exposed and controllable (Clause Device.Audio.Base.NoUndiscoverableDevice). This will allow the user to turn on/off on-board processing. However, currently the Windows platform does not allow applications such as Skype to programmatically turn on/off such on-board processing and therefore it is crucial to meet the above requirements.

In Lync 2013 stereo capture was introduced for certain Lync conference room systems. If a device with stereo speakers (e.g. a PC) wants to integrate an AEC then it needs to ensure that the AEC can also handle stereo signals and that the AEC can handle rapid changes of the loudspeaker playing back monaural signals and stereo signal as common in a conference call where at least one participant with a monaural capture device and one conference room with a stereo device are joining the conference call.

2.2.10 DC offset

Purpose: To ensure appropriate dynamic range of the audio capture signal and keep quantization noise under control, the DC offset shall be within a specified range.

Input: The testing sequence:

- Start the UCDC tool with the DUT connected to the PC (or in case of the DUT being a PC then start the UCDC tool directly on the PC).
- Start the DC offset test. The tool will automatically step through the microphone gain settings to ensure that the DC offset is within the requirement for all settings.

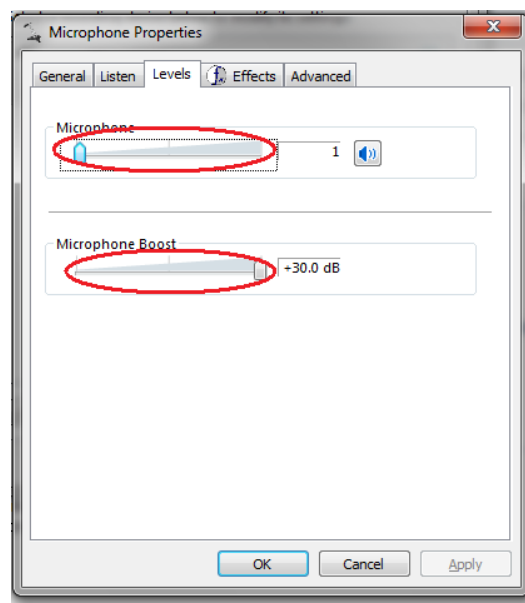
Required: The DC offset of the audio capture signal shall be within $\pm 15\%$ of the peak amplitude. This requirement shall be met for all microphone gain settings.

2.2.11 Clipping of far-end signal due to microphone boost

Purpose: When microphone boost is turned on, the signal from the loudspeaker may clip (saturate) due to the increased sensitivity on the microphone. The nonlinearity caused by microphone saturation may affect the performance of the AEC and result in echo leak to the far end.

Input: Check in the audio control panel whether the DUT has a microphone boost setting. If it does, then set the microphone boost to maximum, the microphone gain to minimum and the loudspeaker gain to maximum (see screenshot below). Make sure to mark down the default values so they can be restored after this test

The current version of the device conformance tool does not yet support this test. Until the tool supports the test, we recommend using a commercially available audio editor (for example, Audacity audio editor) to render the recommended artificial speech and capture the microphone signal at the same time. Check the microphone signal to make sure that the signal is not clipped.



If the DUT supports microphone boost setting, the following settings shall be used for this test:

- set microphone boost to maximum (see screen shot above)
- set microphone gain setting to minimum (see screen shot above)
- set loudspeaker gain setting to maximum

Required: With these settings the microphone capture signal shall not be clipped when rendering an artificial speech with peak level of 0dBFS. The gain due to the microphone boost and the gain due to the microphone gain slider should be implemented as a combined gain in hardware or driver. Otherwise microphone clipping may occur if microphone boost is applied first, or if the minimum microphone gain is applied first and then microphone boost is applied, quantization noise may increase excessively.

2.3 All Groups: Audio Performance over Skype call

Requirements below are valid for all groups: Headset UI, Handset UI, Speakerphone UI, Other-audio products.

Tests in this section are part of the HEAD Acoustics ACQUA Skype AS standard.

2.3.1 Echo path - round trip delay (Skype end to end test)

Purpose: Call interactivity and acoustic echo audibility is dependent of the round trip delay. The purpose of this test is to ensure that the round trip delay during Skype to Skype call using the DUT in lossless local network is below the set maximum limit.

Input: Use recommended test position for DUT.

Make a Skype to Skype audio only call.

Let the Skype calls stabilize for > 3 minutes. As test is done in lossless network jitter buffer in both Skype client should adjust to short length during this time.

Measure the delay in sending direction and then in receiving direction by using the [short delay test signal](#) and [long delay test signal](#). The delay is calculated using cross correlation calculation.

Round trip delay = sending direction delay + receiving direction delay.

Required: The calculated round trip delay must not exceed the figures below:

300ms – for devices with wired connection between Skype client PC and device under test.

340ms – for devices using a wireless link between the computer and device under test.

Note: The average calculated sending direction delay and receiving direction delay is verified from PESQ total quality loss test results (2.3.2 and 2.3.3). PESQ provides delay graph based on the time alignment of degraded signal versus reference signal, thus provides a very precise delay calculation.

2.3.2 Send path - total quality loss (Skype end to end test)

Purpose: To ensure that the call quality in send direction during the Skype to Skype call using DUT in lossless local network condition is not degraded by device or Skype client implementation.

Input: Use recommended test position for DUT.

Make a Skype to Skype audio only call.

Play multiple [real speech samples](#) through artificial mouth at normal speech level (it is recommended to use minimum of 8 samples with total length of 8..12 second each containing 60..80% of active speech). Record the speech samples in reference Skype output. Use the software tool compliant to [ITU-T P862 and ITU-T P862.2 and P.863 \(ex P.OLQA\)](#) to analyze and calculate the average MOS-LQO score.

The source signal files are different for both tools. Both are based on real speech samples but the P.OLQA files contain high frequencies up to 14kHz.

Skype uses PESQ and P.OLQA tool for this analysis. In case of Skype Standard in ACQUA test package the ACQUA in built PESQ and P.OLQA SMD-s are used for this test case.

Required:

	Headset / Handset category	Speakerphone UI categories
PESQ Narrowband MOS-LQO score	> 3.8 MOS-LQO	> 3.5 MOS-LQO
P.OLQA The Super-Wideband MOS-LQ score	> 3.5 MOS-LQ	> 3.1 MOS-LQ

* Requirements for P.OLQA could be readjusted for future version of the specification as Skype is currently gaining knowledge on with this new test tool.

Note: Skype acknowledges the fact that PESQ has not been designed and verified for acoustic interfaces therefore PESQ is not used as a measure of a quality of acoustic interface, but only to identify problems mentioned in the list up.

2.3.3 Receive path - total quality loss (Skype end to end test)

Purpose: To ensure that the call quality in receive direction during the Skype to Skype call using DUT in lossless local network condition is not degraded by device or Skype client implementation.

Input: Use recommended test position for DUT.

Make a Skype to Skype audio only call.

Adjust the earpiece/loudspeaker volume to preferred listening level (note that the preferred listening level is different based on product category!)

Play multiple real speech samples into reference Skype input with normal speech level (it is recommended to use minimum of 8 samples with total length of 8..12 second each containing 60..80% of active speech). Record the speech samples in artificial ear. Use the software tool compliant to ITU-T P862 and ITU-T P862.2 and P.863 (ex P.OLQA) to analyze and calculate the average MOS-LQO score.

The source signal files are different for both tools. Both are based on real speech samples but the P.OLQA files contain high frequencies up to 14kHz.

Skype uses PESQ and P.OLQA tool for this analysis. In case of Skype Standard in ACQUA test package the ACQUA in built PESQ and P.OLQA SMD-s are used for this test case.

Required:

	Headset / Handset category	Speakerphone UI categories
PESQ Narrowband MOS-LQO score	> 3.8 MOS-LQO	> 3.5 MOS-LQO
P.OLQA The Super-Wideband MOS-LQ score	> 3.5 MOS-LQ	> 3.1 MOS-LQ

* Requirements for P.OLQA could be readjusted for future version of the specification as Skype is currently gaining knowledge on with this new test tool.

Note: For headset and handset UI category the Diffuse Field compensation is used for the HATS artificial ear.

For speakerphone UI categories the Free Field compensation is used for the HATS artificial ear.

Skype wants to point out clearly that Skype acknowledges the fact that PESQ has not been designed and verified for acoustic interfaces therefore PESQ is not used as a measure of a quality of acoustic interface, but only to identify problems mentioned in the list up.

3.0 Headset Audio UI Group

3.1 Headset: Audio Test Instructions

Tests in this section are part of the HEAD Acoustics ACQUA Skype AS standard.

3.1.1 Test environment

The test environment should be such that it would not notably influence the measured results compared to results in anechoic environment.

In headset UI case both the earpiece and microphone are very close to HATS, thus the room impact is smaller compared to room requirements needed for testing speakerphone UI products. However:

- There should be no big items closer than 1m from the artificial mouth in any direction.
- The average room noise level should be below 29dB SPL(A).
- The room reverberation time (RT60) should be below 0.3 seconds at frequencies between 150 Hz to 12 kHz.

3.1.2 Measurement setup – Head and Torso Simulator (HATS)

3.1.2.1 Mouth simulator

Head and Torso Simulator (HATS) compliant with ITU-T [P.58](#) is used as a mouth simulator.

The frequency spectrum of mouth simulator is calibrated and frequency compensated at Mouth Reference Point (MRP) to be flat between 100 Hz to 11000 Hz.

The normal speech level for active speech part of male and female artificial speech is calibrated to be 89 dB SPL at MRP. The analysis is done with no frequency weighting and in frequency range from 50Hz to 20'000Hz

Quiet and loud speech signals are respectively 10 dB quieter and 10 dB louder compared to normal speech level.

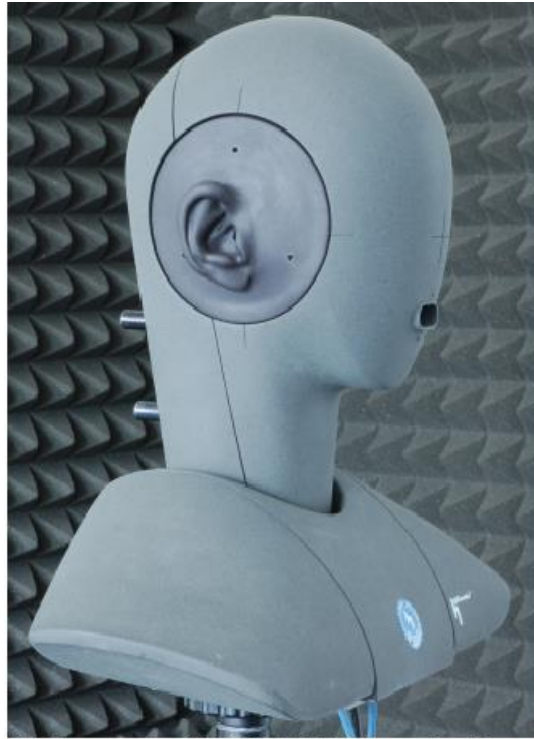


B&K Type 4128C HATS (alternative 1)

3.1.2.2 *Ear simulator*

Head and Torso Simulator (HATS) compliant with ITU-T [P.58](#) and artificial ear compliant with ITU-T P.57 Type 3.3 is used as an ear simulator.

The frequency response measured by the ear simulator at the drum reference point (DRP) needs to be transformed to the ear reference point (ERP) according to IEEE Std. 269 before comparing it to the mask. Note that even a handset/headset receiver with a flat frequency response in the free field will exhibit a non-flat frequency response at ERP (see IEEE Std. 1652 for a more detailed discussion). If manufacturer request some of the other standard corrections such as DRP to ERP (Ear Reference Point) or DRP to free field or DRP to diffuse field as Skype requirements used before then these can be used and/or taken into account when interpreting the results. Skype will make a pass/fail decision case by case in such case. The measurement microphone at DRP is calibrated using microphone calibrator. Skype uses G.R.A.S Type 42AB microphone calibrator for this purpose.



HEAD acoustics HMS II.3 HATS (alternative 2)

3.1.2.3 Headset positioning on HATS

If the manufacturer/vendor/user documentation provides guidelines how the headset should be used/worn, Skype audio engineers will take such recommendations into consideration.

The headset will be placed on HATS and adjusted to resemble the positioning on real human as closely as possible. Frequency response measurements will be done several times while repositioning the headset slightly – the position with flattest (best) results will be used as a recommended test position. This position will be photographed and documented in audio test report.



For earpiece and microphone frequency response stability test cases the headset is removed and replaced on HATS several times and the positioning altered slightly. Depending on headset type (on ear, in ear etc) the possible position alterations of the microphone/earpiece will be different, thus the actual position changes will be documented case by case.

3.1.2.4 Headset tone control, EQ, music mode, etc. settings (if available)

If the DUT has multiple settings for earpiece tone control, then the DUT is tested with all available receive path tone control settings and the requirements 3.2.9 must pass with all settings.

3.2 Headset: Audio Performance Requirements

Please see the sections **7.1.6 Reference Skype client setup details** and **0 DUT client setup details** for instructions how to enable the testing mode used during all of the below measurements

3.2.1 Send path - signal level with loud speech

Purpose: To ensure that DUT send path provides optimal signal level for far end Skype client with loud speech input.

Input: Use recommended test position for DUT. Use following settings for the DUT Editor:

- Enable DUT Client Mode
- Disable AEC

Skype is allowed to automatically adjust the input gain setting for this test case.

Play the IEEE 269-2010 compressed male speech sample at loud speech level for 30 seconds to allow Skype to find optimal input gain level.

Play the IEEE 269-2010 compressed male speech sample at loud speech level and record the signal at reference Skype output.

Required: Calculate the average RMS level in reference Skype output for the active part of compressed male speech sample. **The level must be more than -24dBm0 RMS (equals -30dBFS RMS).** The peaks of the speech signals must not overload the input causing clipping / overload.

Note: The compressed male speech is used for loud speech as it has a lower crest factor and thus better simulates a real world conditions as the crest factor of human voice is lower when people speak louder.

3.2.2 Send path - signal level with normal speech (send loudness)

Purpose: To ensure that DUT send path provides optimal signal level for far end Skype client with normal speech input.

Input: Use recommended test position for DUT. Use following settings for the DUT Editor:

- Enable DUT Client Mode
- Disable AEC

Skype is allowed to automatically adjust the input gain setting for this test case.

Play the IEEE 269-2010 uncompressed male speech sample at normal speech level for 30 seconds to allow Skype to find optimal input gain level.

Play the IEEE 269-2010 uncompressed male speech sample at normal speech level and record the signal at reference Skype output.

Required: Calculate the average RMS level in reference Skype output for the active part of male speech sample. **The level must be more than -24dBm0 RMS (equals -30dBFS RMS).** The peaks of the speech signals must not overload the input causing clipping.

3.2.3 Send path - signal level with quiet speech

Purpose: To ensure that DUT send path provides optimal signal level for far end Skype client with quiet speech input.

Input: Use recommended test position for DUT. Use following settings for the DUT Editor:

- Enable DUT Client Mode
- Disable AGC
- Disable AEC

The same input AGC setting as for normal speech level test is used for this test case. Skype automatic gain adjustment is disabled.

Play the IEEE 269-2010 uncompressed speech sample at quiet speech level and record the signal at reference Skype output.

Required: Calculate the average RMS level in reference Skype output for the active part of male speech sample. **The level must be more than -34dBm0 RMS (equals -40dBFS RMS).** The peaks of the speech signals must not overload the input causing clipping.

3.2.4 Send path - frequency response

Purpose: The test checks that the frequency response of the DUT send signal path is flat enough to meet Skype requirement and take full advantage of Skype voice codec bandwidth.

Input: Use recommended test position for DUT. Use following settings for the DUT Editor:

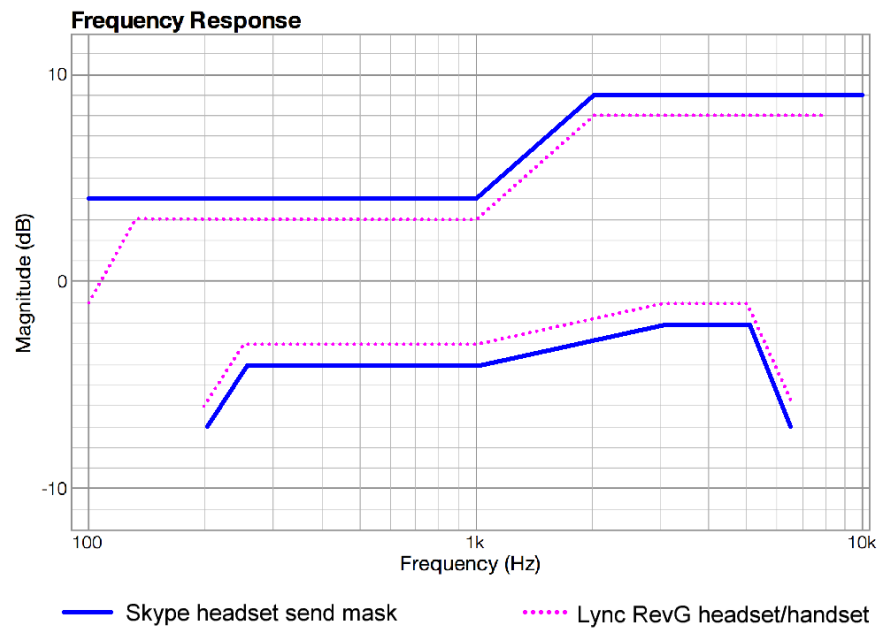
- Enable DUT Client Mode
- Disable AGC
- Disable AEC

The same input AGC setting as for normal speech level test is used for this test case. Skype automatic gain adjustment is disabled.

Play the [male artificial speech for frequency response](#) at normal speech level and record the signal at reference Skype output.

Required: The frequency response is calculated by comparing the 1/12 octave spectrum of the prior recorded reference result at MRP to the DUT send signal recorded at reference Skype output.

The resulting frequency response graph fits into the tolerance mask on the next page:



Frequency	Lower limit	Upper limit
99 Hz	-80,0 dB	80,0 dB
100 Hz	-80,0 dB	4,0 dB
199 Hz	-80,0 dB	4,0 dB
200 Hz	-7,0 dB	4,0 dB
250 Hz	-4,0 dB	4,0 dB
1000 Hz	-4,0 dB	4,0 dB
2000 Hz	-3,0 dB	9,0 dB
3000 Hz	-2,0 dB	9,0 dB
5000 Hz	-2,0 dB	9,0 dB
6500 Hz	-7,0 dB	9,0 dB
6501 Hz	-80 dB	9,0 dB
10000 Hz	-80 dB	9,0 dB
10001 Hz	-80 dB	80 dB

3.2.5 Send path – speech signal to self noise ratio

Purpose: Too high self noise in microphone signal decreases the intelligibility of the speech and influences the total call quality in negative way. This test tests for speech to self noise ratio in DUT send path.

Input: Use recommended test position for DUT. Use following settings for the DUT Editor:

- Enable DUT Client Mode
- Disable AGC
- Disable AEC

The same input AGC setting as for normal speech level test is used for this test case. Skype automatic gain adjustment is disabled.

Play the [SpNR speech sample](#) at normal speech level and record the signal at reference Skype output.

Required: Calculate the RMS level in reference Skype output for the active speech part (the active speech does not include pauses or silences) → this is defined as Speech level.

Calculate the A-weighted RMS level of noise in reference Skype output for the 1 second silence part at end of the test signal for SpNR → this is defined as Noise level.

The calculated speech to noise ratio (SpNR ratio) is more than 35 dB

Speech level – Noise level > 35

Note: Please note that Skype specifies speech to noise ratio (SpNR), this is not the same as signal to noise ratio (SNR) specification often found in datasheet that specifies a sine signal to noise ratio.

3.2.6 Send path – speech signal to self noise ratio during speech

Purpose: Too high self noise in microphone signal decreases the intelligibility of the speech and influences the total call quality in negative way. Thus this test tests for speech to self noise level in DUT send path. This test tests for speech to self noise level in DUT send path during speech.

Input: Use recommended test position for DUT. Use following settings for the DUT Editor:

- Enable DUT Client Mode
- Disable AGC
- Disable AEC

The same input AGC setting as for normal speech level test is used for this test case. Skype automatic gain adjustment is disabled.

Play the [modified SpNR speech sample](#) at normal speech level and record the signal at reference Skype output.

Required: Calculate the RMS level in reference Skype output for the active speech part (the active speech does not include pauses or silences) → this is defined as Speech level.

Calculate the A-weighted RMS level of noise in reference Skype output for the 1 second silence part in middle of the speech signal for SpNR during speech → this is defined as Noise level during speech. (The sine signal components are filtered out by band stop filters prior to noise level calculation.)

The calculated speech level minus noise level (SpNR ratio) is more than 33 dB

Speech level – Noise level during speech > 33 dB

Note: Skype leaves a freedom to alter or improve the above test signal or post-processing of result without prior notice. The speech part of the signal will always remain the same.

3.2.7 Send path – single frequency interference

Purpose: Narrow-band noise, including single frequency interference, is an impairment that can be perceived as a tone, depending on its level relative to the overall weighted noise level. This can be caused by electrical noise in soundcards or by fan or hard disk drive noise on laptops. This requirement makes sure that no tonal noise is present in the send signal.

Input: Use recommended test position for DUT. Use following settings for the DUT Editor:

- Enable DUT Client Mode
- Disable AGC
- Disable AEC

The same input AGC setting as for normal speech level test is used for this test case. Skype automatic gain adjustment is disabled.

Play the [SpNR speech sample](#) at normal speech level and record the signal at reference Skype output.

Required: Calculate the A-weighted peak noise level for the 1 second silence part at end of the test signal for SpNR with an effective bandwidth of not more than 31 Hz. When using a 48 kHz sampling rate recording this means a minimum FFT size of 4096. Frequency range analyzed is 100Hz to 12000Hz. For FFT analysis the “Flat Top” windowing is employed.

The measured peak noise level is $\leq -74\text{dBm0}$ ($\leq -80\text{dBFS}$)

3.2.8 Receive path – preferred loudness level in ear

Purpose: To ensure that the earpiece/loudspeaker is able to provide optimal receiving signal loudness in ear.

Skype uses HATS equipped with ITU-T P.57 specified Type 3.3 artificial ear for receive direction.

Input: Use recommended test position for DUT.

Play the IEEE 269-2010 uncompressed male speech to the reference Skype input and record the acoustic signal in artificial ear.

Required: Calculate the RMS SPL level in artificial ear for the active part of male artificial speech signal (active speech does not include pauses or silences).

The level must be set to:

78 dB .. 82 dB SPL – for monaural

73 dB .. 77 dB SPL – for binaural

The level set in this test case will be used for other receiving direction tests that refer to preferred loudness level for playback loudness.

3.2.9 Receive path – frequency response

Purpose: The test checks that the frequency response of the DUT receiving path is flat enough to meet minimum requirement taking into consideration the technical limitations of wireless technologies.

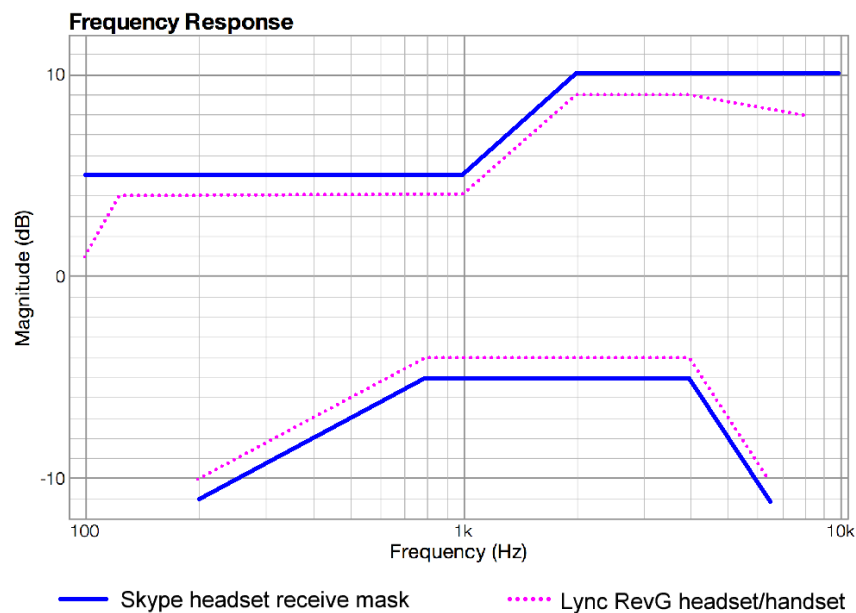
Input: Use recommended test position for DUT.

Use same receive loudness setting as set in 3.2.8.

Play the [male artificial speech for frequency response](#) to the reference Skype input.
Record the acoustic receive signal in artificial ear.

Required: The frequency response is calculated by comparing the 1/12 octave spectrum of the prior recorded reference result to the DUT receive signal in artificial ear with the DRP/ERP correction applied.

The resulting frequency response graph fits into the tolerance mask below:



Frequency	Lower limit	Upper limit
99 Hz	-80,0 dB	80,0 dB
100 Hz	-80,0 dB	5,0 dB
199 Hz	-80,0 dB	5,0 dB
200 Hz	-11,0 dB	5,0 dB
800 Hz	-5,0 dB	5,0 dB
1000 Hz	-5,0 dB	5,0 dB
2000 Hz	-5,0 dB	10,0 dB
4000 Hz	-5,0 dB	10,0 dB
6500 Hz	-11,0 dB	10,0 dB
6501 Hz	-80,0 dB	10,0 dB
10000 Hz	-80,0 dB	10,0 dB
10001 Hz	-80,0 dB	80,0 dB

3.2.10 Receive path – speech signal to noise ratio (SpNR)

Purpose: Too high self noise in receive path decreases the intelligibility of the speech and influences the total call quality in negative way. This test tests for speech to noise ratio in DUT receive path.

Input: Use recommended test position for DUT.

Use same receive loudness setting as set in 3.2.8.

Play the IEEE 269-2010 uncompressed male speech to the reference Skype input and record the acoustic signal in artificial ear.

Required: Calculate the RMS SPL level in artificial ear for the active part of male artificial speech (active speech does not include pauses or silences) → Speech level.

Calculate the A-weighted RMS SPL level in artificial ear for the 1 second silence part in middle of the male artificial speech signal for SpNR → Noise level.

The calculated speech to noise ratio (SpNR ratio) is more than 40 dB

Speech level – Noise level > 40 dB

Note: Please note that Skype requires speech to noise ratio (SpNR), this is not the same as signal to noise ratio (SNR) specification often found in datasheet that specifies a sine signal to noise ratio!

3.2.11 Receive path – single frequency interference

Purpose: Tonal noise may be perceived in the loudspeaker signal if receive single frequency interference is too high.

Input: Use recommended test position for DUT.

Play the signal containing silence into the REF Skype input

Required: Record a 5 second sound sample from HATS ear microphone. Measure the A-weighted peak noise level over the frequency range of 50 to 20000 Hz with an effective bandwidth of not more than 31 Hz. For 48 kHz sampling rate recording this means a minimum FFT size of 4096. For FFT analysis the “Flat Top” windowing is employed.

Compare the analyzed level to the noise level calculated in test case 3.2.10. Receive A-weighted single frequency interference noise peak level shall be at least 10 dB quieter than the averaged receive noise level.

3.2.12 Weighted terminal coupling loss (TCLw)

Purpose: The amount of acoustic echo in the microphone signal is measured by the TCLw and the acoustic echo should be minimized by maximizing the physical distance between the loudspeaker and the microphone. For devices relying on the AEC in Lync, not meeting this requirement will result in echo leak, or distortion and attenuation of speech during double-talk (that is, near-end user and far-end participant talking simultaneously).

For devices with on-board AEC, a failure of this test will lead to echo leaks that are disruptive to the far-end participants.

The TCLw shall be normalized with respect to the nominal send loudness to account for any analog gain difference which would be compensated for by the digital AGC integrated in Skype. The nominal send loudness is defined as -18dBm0 (-24dBFS) The formula for the normalized TCLw is

$$\begin{aligned} TCLw &= TCLw_{measured} + (Send\ loudness_{measured} - Send\ loudness_{nominal}) \\ TCLw &= TCLw_{measured} + (Send\ loudness_{measured} + 24dBov) \end{aligned}$$

The TCLw shall be measured at preferred receive loudness using the IEEE Std. 269 male uncompressed speech signal.

Input: Use recommended test position for DUT. Use following settings for DUT editor **during preparation!**

- Enable DUT Client Mode
- Disable AEC

Play the IEEE 269-2010 uncompressed male speech to the reference Skype input. Skype is allowed to automatically adjust the input gain setting during the preparation period. After Skype has adjusted the input gain to optimal level the AGC is disabled.

Use following settings for DUT editor **for the actual test case run!**

- Enable DUT Client Mode
- Disable AGC
- Disable AEC

Use same receive loudness setting as set in 3.2.8.

Play the IEEE 269-2010 uncompressed male speech sample at normal speech level to reference Skype input and record the signal at reference Skype output.

Required: **TCLw for devices without in built AEC >30dB**
TCLw for devices with in built AEC >52dB

Note: The measurement shall be performed after system stability is reached (including convergence of any echo algorithms): this shall be accomplished by invoking the test signal for at least 2 seconds before the actual measurement occurs.

Measurements shall be done in 1/12th octave bands over a range of 100 Hz through 8000 Hz. The weighted terminal coupling loss is calculated according to ITU-T G.122 Annex B.4 (trapezoidal rule) using the frequency range of 100 to 8000 Hz rather than 300 to 3400 Hz.

Note that TCLw is also tested at maximum volume [here](#).

3.2.13 Echo path – acoustic echo cancellation

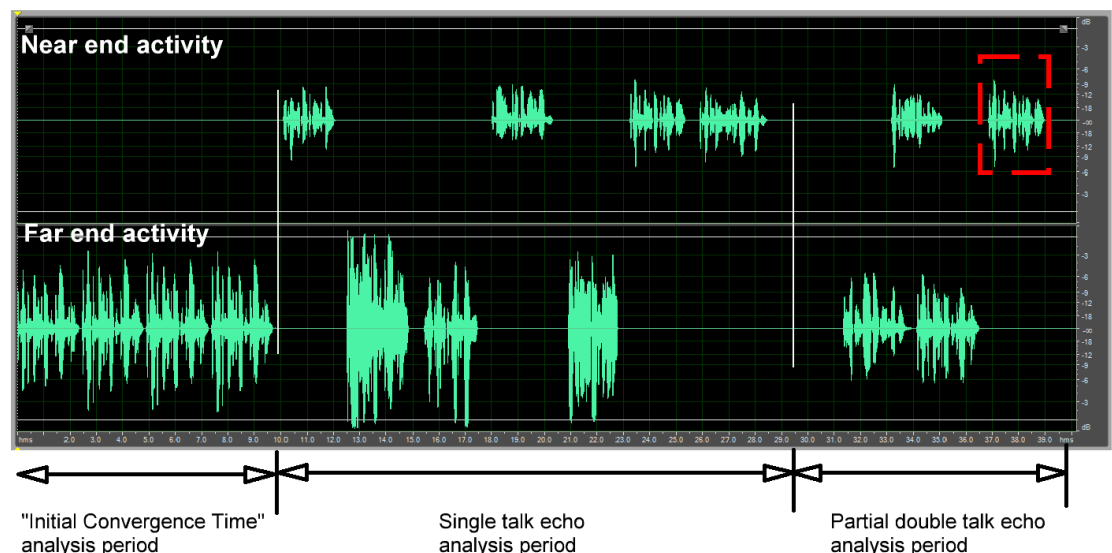
Purpose: The test checks the level of loudspeaker acoustic echo leaking back to far end output. The test signal includes near and far ends speech, alternating at different times. The talkers are occasionally overlapping to simulate use case where users take turns while speaking, but interrupt each other from time to time.

Input: Use recommended test position for DUT. Use following setting for the DUT editor:

- Enable DUT Client Mode

Play the preparation test signals to the reference Skype input and artificial mouth simultaneously. Use same receive loudness setting as set in 3.2.8.

Play the Echo test signal to the reference Skype input and artificial mouth simultaneously.



Required: The recording must comply with the following:

- **Initial Convergence Time** – the echo suppression should reach full cancellation at latest after 4 seconds from the start of far end activity
- **Residual echo / loss of convergence**
Calculate the **level versus time** and **spectrum versus time** graph of reference Skype output. Analyze and listen to the recording.
There should be no echo leaks or bursts higher than +5 dB in level compared to the send path noise floor level during far end speech activity in any frequency range between 50 Hz to 20 kHz.
- During the partial double talk there should be minimal loss of near end speech, especially the beginnings of the near end speech. The partial attenuation of near end speech is allowed, but the far end user should be able to recognize that the near end attempts to speak.
- **Send path noise floor stability / similarity.** It is beneficial to keep low level of comfort noise in the send signal path also during far end speech activity periods. Fully muting or attenuating the send signal during far end activity will make possible echo residuals more audible and could lead to failing result for this test case. Also the spectrum of the generated comfort noise should match that of the send path noise during silent periods of near end speech activity. If not – the noise floor changes will be very audible every time the far end is speaking.

3.2.14 Echo path – send path signal level during two way conversation

Purpose: The test checks the transmitted near end speech level during two way conversation. The send level might be low in such use case, especially on devices where the acoustic echo in microphone is very loud. This usually is due to physical distance between microphone(s) and speaker(s). The loud acoustic echo will force the analog gain control to adjust to lower gain, thus a digital amplification with faster adjustment speed will be needed to compensate for the level loss.

Input: Use the resulting recording from test case 3.2.13

The recording of send path signal in reference Skype output is analyzed. The section marked with red dotted line in above sequence is used for calculation of send path level during conversation

Required: Calculate the RMS level in reference Skype output for the active part of near end speech (time selection marked with red dotted line in above sequence). **The level must be more than -24 dBm0 RMS (equals -30dBFS RMS).**

3.2.15 Echo path – sidetone delay

Purpose: Ideally the sidetone should be a real-time signal. Sidetone delay less than 5 ms is generally perceived as a normal sidetone. Sidetone delay between 5 and 10 ms is generally perceived as unnatural sidetone, with an uncomfortable hollow characteristic. Sidetone delay greater than 10 ms is generally perceived as a distinct talker echo signal.

Input: Remove DUT from HATS. Play the short delay test signal into HATS mouth

Record both HATS ear and Reference Skype output signals simultaneously. Using a cross correlation method measure the HATS mouth to ear delay.

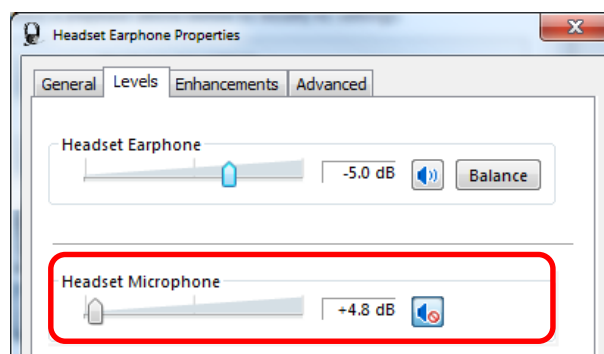
Attach DUT to HATS and enable the side tone path. Play the short delay test signal into HATS mouth

Record both HATS ear and Reference Skype output signals simultaneously. Using a cross correlation method measure the delay of the side tone.

Required: The sidetone delay for both handset and headset shall occur no more than 5ms after the mouth to ear delay signal.

It is desirable for the sidetone delay to be less than 1ms

Note: This test case is only applicable for devices that have a side tone path function.



3.3 Headset: Loudness and frequency response stability requirements

3.3.1 Send path – frequency response stability

Purpose: The test checks that the frequency response of the DUT send path does not change in major way when microphone position is slightly moved around the artificial mouth.

Input: Use recommended test position for DUT. Use following setting for the DUT editor:

- Enable DUT Client Mode
- Disable AGC
- Disable AEC

Play the full [male artificial speech for frequency response](#) sample at normal speech level and record the signal at reference Skype output → [send reference result](#).

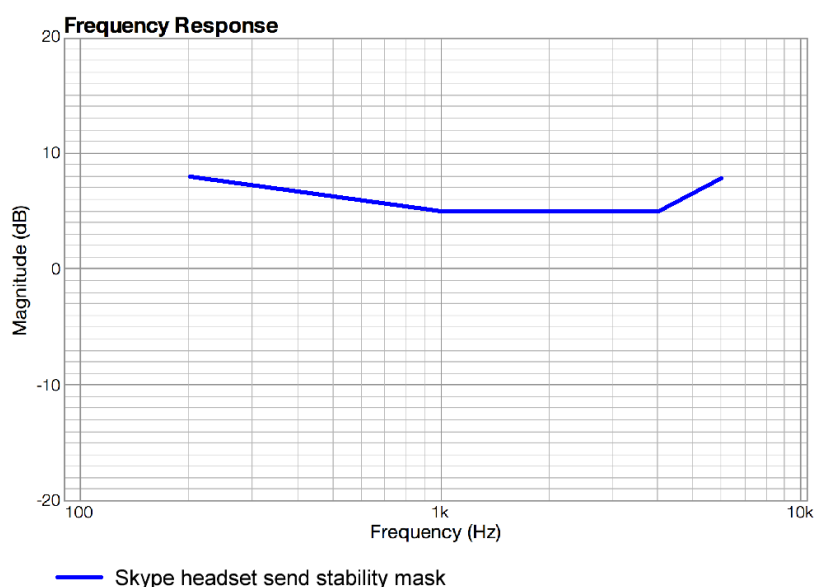
Alternate the DUT position 3 times every time recording the [male artificial speech for frequency response](#).

Required: The frequency response is calculated by comparing the 1/3 octave spectrum of the prior recorded reference result at MRP to the DUT send signal recorded at reference Skype output.

The resulting frequency responses are compared.

Maximum and minimum of each 1/3 octave band is calculated for all recordings.

The difference between maximum and minimum curves is calculated by subtracting the minimum value from the maximum value in every 1/3 octave band. (Note this difference is always positive) **This difference in each octave band must be less than the tolerance mask below.**



3.3.2 Receive path – earpiece frequency response stability

Purpose: The test checks that the frequency response of the DUT receiving path does not change in major way when earpiece position is slightly moved around on artificial ear.

Skype uses ITU-T P.57 specified Type 3.3 artificial ear for receive direction testing with DRP/ERP correction applied.

Input: Use recommended test position for DUT. Make a Skype to Skype call from DUT to reference Skype client in lossless local network condition.

Play the [male artificial speech for frequency response](#) to the reference Skype input. Record the acoustic receive signal in artificial ear → receive reference result.

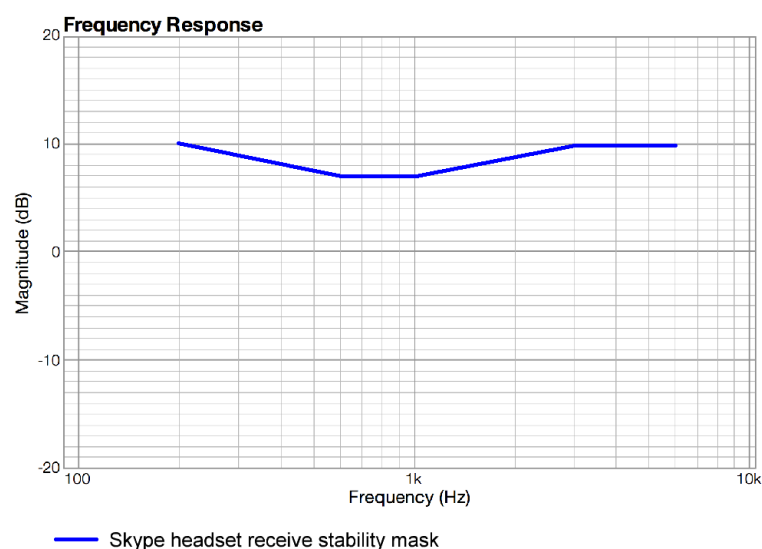
Alternate the DUT position 3 times every time recording the [male artificial speech for frequency response](#).

Required: The frequency response is calculated by comparing the 1/3 octave spectrum of the prior recorded reference result to the DUT receive signal in artificial ear with the DRP/ERP correction applied.

The resulting frequency responses are compared.

Maximum and minimum of each 1/3 octave band is calculated for all recordings.

The difference between maximum and minimum curves is calculated by subtracting the minimum value from the maximum value in every 1/3 octave band. (Note this difference is always positive) **This difference in each octave band must be less than the tolerance mask below.**



3.3.3 Receive path – loudness level adjustment range

- Purpose:** To ensure that the receive path volume adjustment has enough adjustment range to allow volume settings for users personal preference.
- Input:** Adjust the DUT playback volume to 5...25% of full available adjustment range → low level volume setting.
- Play the [male artificial speech](#) to the reference Skype input. Record the acoustic receive signal in artificial ear.
- Adjust the DUT playback volume to 75...100% of full available adjustment range → high level volume setting.
- Play the male artificial speech to the reference Skype input. Record the acoustic receive signal in artificial ear.
- Required:** Calculate the RMS SPL level in artificial ear for the active part of male artificial speech signal for low level volume setting.
- Calculate the RMS SPL level in artificial ear for the active part of male artificial speech signal for high level volume setting.
- The high level minus low level must be > 20 dB.**

3.3.4 Receive path – maximum loudness level

- Purpose:** To ensure that the receive path maximum setting does not provide uncomfortably loud acoustic signal.
- Under some operating systems the DUT defaults to maximum playback volume setting, thus at first use user might hear uncomfortably loud levels.
- Input:** Adjust the DUT playback volume to 100% of full available adjustment range → max level volume setting.
- Play the [male artificial speech](#) to the reference Skype input. Record the acoustic receive signal in artificial ear.
- Required:** Calculate the SPL RMS level in artificial ear for the active part of male artificial speech signal for max level volume setting.
- The maximum loudness level is less than 94dB.**

3.3.5 Weighted terminal coupling loss (TCLw) at max volume

Purpose: The amount of acoustic echo in the microphone signal is measured by the TCLw and the acoustic echo should be minimized by maximizing the physical distance between the loudspeaker and the microphone. For devices relying on the AEC in Lync, not meeting this requirement will result in echo leak, or distortion and attenuation of speech during double-talk (that is, near-end user and far-end participant talking simultaneously).

For devices with on-board AEC, a failure of this test will lead to echo leaks that are disruptive to the far-end participants.

The TCLw shall be normalized with respect to the nominal send loudness to account for any analog gain difference which would be compensated for by the digital AGC integrated in Skype. The nominal send loudness is defined as -18dBm0 (-24dBFS) The formula for the normalized TCLw is

$$\begin{aligned} TCLw &= TCLw_{measured} + (Send\ loudness_{measured} - Send\ loudness_{nominal}) \\ TCLw &= TCLw_{measured} + (Send\ loudness_{measured} + 24dBov) \end{aligned}$$

The TCLw shall be measured at preferred receive loudness using the IEEE Std. 269 male uncompressed speech signal.

Input: Use recommended test position for DUT. Use following settings for DUT editor **during preparation!**

- Enable DUT Client Mode
- Disable AEC

Play the IEEE 269-2010 uncompressed male speech to the reference Skype input. Skype is allowed to automatically adjust the input gain setting during the preparation period. After Skype has adjusted the input gain to optimal level the AGC is disabled.

Use following settings for DUT editor **for the actual test case run!**

- Enable DUT Client Mode
- Disable AGC
- Disable AEC

Use same maximum receive loudness setting as set in 3.3.4

Play the IEEE 269-2010 uncompressed male speech sample at normal speech level to reference Skype input and record the signal at reference Skype output.

Required: **TCLw for devices without in built AEC >30dB**

TCLw for devices with in built AEC >45dB

Note: The measurement shall be performed after system stability is reached (including convergence of any echo algorithms): this shall be accomplished by invoking the test signal for at least 2 seconds before the actual measurement occurs.

Measurements shall be done in 1/12th octave bands over a range of 100 Hz through 8000 Hz. The weighted terminal coupling loss is calculated according to ITU-T G.122 Annex B.4 (trapezoidal rule) using the frequency range of 100 to 8000 Hz rather than 300 to 3400 Hz.

3.4 Headset: Requirements for Skype Super Wideband Certification (Optional)

3.4.1 Echo path – round trip delay – Super Wideband quality

Purpose: Call interactivity and acoustic echo audibility is dependent of the round trip delay. The purpose of this test is to ensure that the round trip delay during Skype to Skype call using the DUT in lossless local network is below the set maximum limit.

Input: Use recommended test position for DUT.

Make a Skype to Skype audio only call.

Let the Skype calls stabilize for > 3 minutes. As test is done in lossless network jitter buffer in both Skype client will adjust to low length.

Measure the delay in sending direction and then in receiving direction by using the [short delay test signal](#) and [long delay test signal](#). The delay is calculated using cross correlation calculation.

Round trip delay = sending direction delay + receiving direction delay.

Required: The calculated round trip delay must be below:

250ms – for devices with wired connection between Skype client PC and device under test.

290ms – for battery powered devices using a wireless link between the computer and device under test.

Note: The average calculated sending direction delay and receiving direction delay is verified from PESQ total quality loss test results. PESQ provides delay graph based on the time alignment of degraded signal versus reference signal, thus provides a very precise delay calculation.

3.4.2 Send path – frequency response – Super Wideband.

Purpose: The test checks that the frequency response of the DUT send signal path is flat enough to meet Skype requirement and take full advantage of Skype Super Wideband voice codec (SILK SWB) bandwidth.

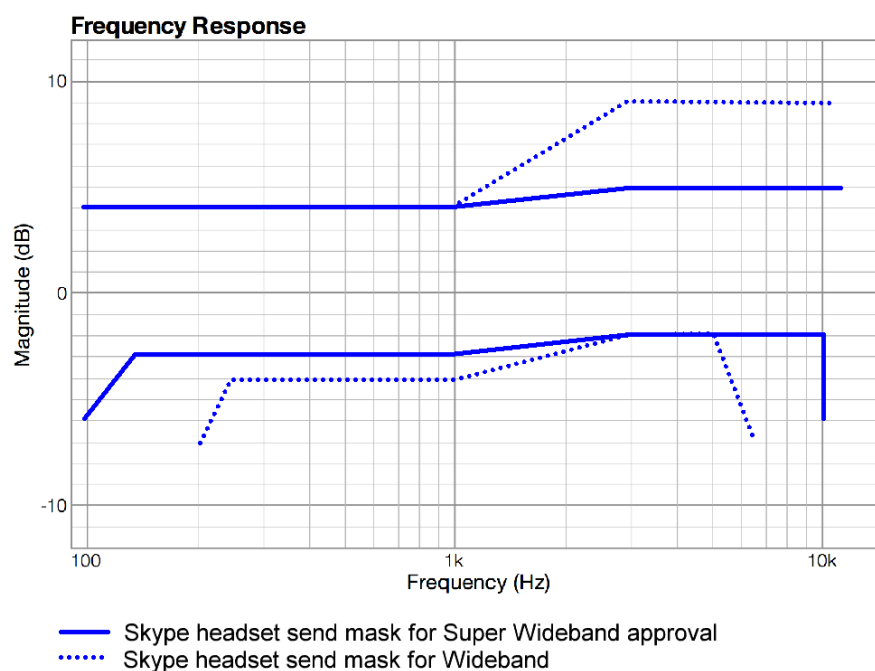
Input: Use recommended test position for DUT. Use following settings for the DUT Editor:

- Enable DUT Client Mode
- Disable AGC
- Disable AEC

Play the [male artificial speech for frequency response](#) at normal speech level and record the signal at reference Skype output.

Required: The frequency response is calculated by comparing the 1/3 octave spectrum of the prior recorded reference result at MRP to the DUT send signal recorded at reference Skype output.

The resulting frequency response graph fits into below tolerance mask



Frequency	Lower limit	Upper limit
99 Hz	-80,0 dB	80,0 dB
100 Hz	-6,0 dB	4,0 dB
150 Hz	-3,0 dB	4,0 dB
1000 Hz	-3,0 dB	4,0 dB
3000 Hz	-2,0 dB	5,0 dB
10000 Hz	-2,0 dB	5,0 dB
10001 Hz	-80,0 dB	5,0 dB
12000 Hz	-80,0 dB	5,0 dB
12001 Hz	-80,0 dB	80,0 dB

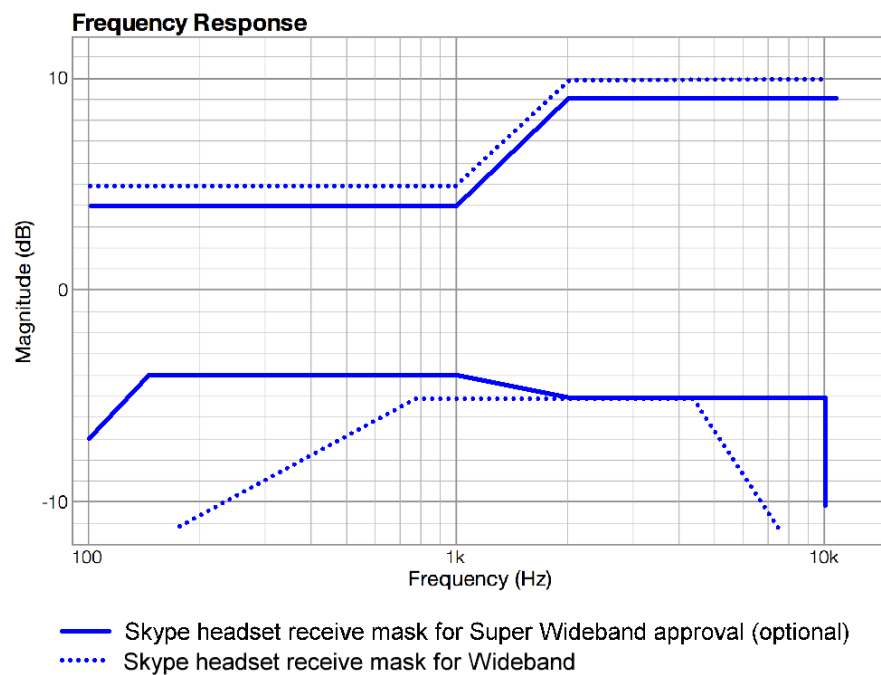
3.4.3 Receive path – frequency response – Super Wideband

Purpose: The test checks that the frequency response of the DUT receiving path is flat enough to meet minimum requirement and take full advantage of Skype Super Wideband voice codec (SILK SWB) bandwidth.

Input: Use recommended test position for DUT.
Use same receive loudness setting as set in 3.2.8.
Play the [male artificial speech for frequency response](#) to the reference Skype input.
Record the acoustic receive signal in artificial ear.

Required: The frequency response is calculated by comparing the 1/3 octave spectrum of the prior recorded reference result to the DUT receive signal in artificial ear with the DRP/ERP correction applied.

The resulting frequency response graph fits into the tolerance mask below:



Frequency	Lower limit	Upper limit
99 Hz	-80,0 dB	80,0 dB
100 Hz	-7,0 dB	4,0 dB
150 Hz	-4,0 dB	4,0 dB
1000 Hz	-4,0 dB	4,0 dB
2000 Hz	-5,0 dB	9,0 dB
10000 Hz	-5,0 dB	9,0 dB
10001 Hz	-80,0 dB	9,0 dB
12000 Hz	-80,0 dB	9,0 dB
12001 Hz	-80,0 dB	80,0 dB

3.6 Headset: Supporting Audio Documentation Requirements

In addition to the user manual (the one that comes with the product) we also ask for supporting audio documentation (for certification testing purposes). Such documentation contains engineering data and engineering test data for the product. Earpiece below means the acoustic output component for sound playback to the user's ear, for example the small loudspeaker.

3.6.1 Verifying supporting documentation for Headset Audio UI group

Purpose: Solution must come with a supporting audio documentation (only for certification testing purposes).

Required: DUT arrives with supporting audio documentation that contains information about:

- Active signal processing: yes/no, if yes then:
 - DUT has in built Acoustic Echo Cancellation: yes/no
 - DUT has in built Active Noise Suppression: yes/no
 - DUT has in built Active Noise Gating: yes/no
 - DUT has in built Automated Gain Control in sending or/and in receiving directions: yes/no
- Microphone: Directionality/design principle (omni-, uni-, noise canceling etc.)
- Microphone: Frequency range (lowest and highest audible frequencies)
- Earpiece: Lowest and highest designed audible frequencies.
- Tone control modes / usage scenarios.
- Supported sample rates (for USB connected headsets)
- Send path and receive path audio signal bandwidth (in case of Wireless headsets)

4.0 Handset Audio UI Group

Tests in this section are part of the HEAD Acoustics ACQUA Skype AS standard.

4.1 Handset: Audio test instructions

4.1.1 Test environment

The test environment should be such that it would not notably influence the measured results compared to results in anechoic environment.

As in headset UI case both the earpiece and microphone are very close to HATS, the room impact is lessened compared to room requirements needed for testing speakerphone UI products. There should be no big items closer than 1m to MRP in any direction.

The average room noise level should be below 29 dB SPL(A).

The room reverberation time (RT60) should be below 0.3 seconds at frequencies between 150 Hz to 12 kHz.

4.1.2 Measurement setup – Head and Torso Simulator (HATS)

4.1.2.1 Mouth simulator

Head and Torso Simulator (HATS) compliant with ITU-T P.58 is used as a mouth simulator.

The frequency spectrum of mouth simulator is calibrated and frequency compensated at Mouth Reference Point (MRP) to be flat between 100 Hz to 11000 Hz.

The normal speech level for active speech part of male and female artificial speech is calibrated to be 89 dB SPL at MRP. The analysis is done with no frequency weighting and in frequency range from 50Hz to 20'000Hz

Quiet and loud speech signals are respectively 10 dB quieter and 10 dB louder compared to normal speech level.



B&K Type 4128C HATS (alternative 1)

4.1.2.2 *Ear simulator*

Head and Torso Simulator (HATS) compliant with ITU-T P.58 and artificial ear compliant with ITU-T P.57 Type 3.3 is used as a ear simulator.

The frequency response measured by the ear simulator at the drum reference point (DRP) needs to be transformed to the ear reference point (ERP) according to IEEE Std. 269 before comparing it to the mask. Note that even a handset/headset receiver with a flat frequency response in the free field will exhibit a non-flat frequency response at ERP (see IEEE Std. 1652 for a more detailed discussion). If manufacturer request some of the other standard corrections such as DRP to ERP (Ear Reference Point) or DRP to free field or DRP to diffuse field as Skype requirements used before then these can be used and/or taken into account when interpreting the results. Skype will make a pass/fail decision case by case in such case. The measurement microphone at DRP is calibrated using microphone calibrator. Skype uses G.R.A.S Type 42AB microphone calibrator for this purpose.



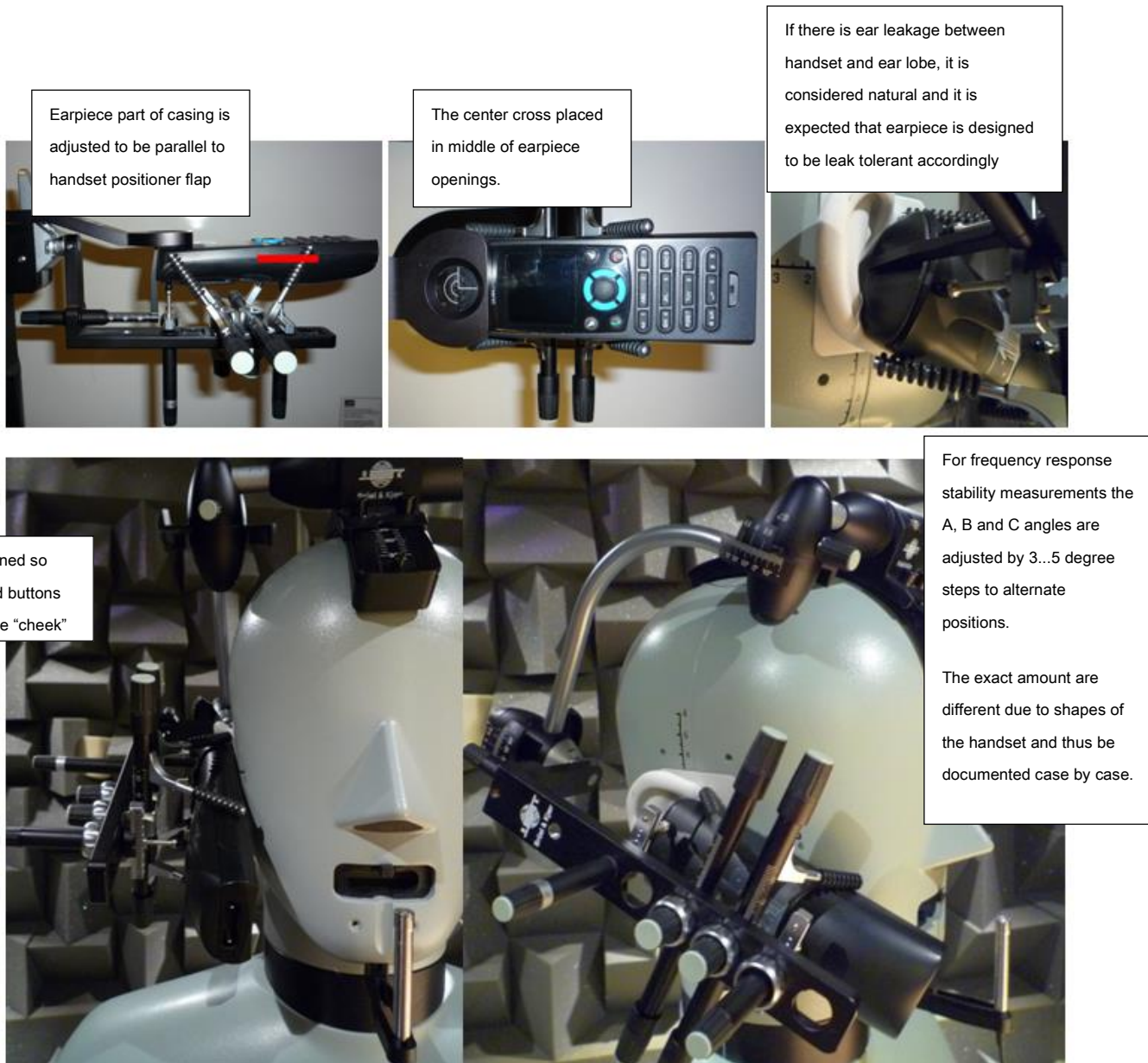
HEAD acoustics HMS II.3 HATS (alternative 2)

4.1.2.3 *Handset positioning on HATS*

If the manufacturer/vendor provides guidelines how the handset should be tested by providing standard handset positioning coordinates as defined in ITU-T [P.64](#), Annex E. Skype audio engineers will take such recommendations into consideration.

Otherwise following guidelines are followed during positioning of the handset device for handset mode testing.

If not instructed otherwise, the measurements shall be conducted with an **application force of 8N** between phone and the right ear.



For frequency response stability measurements the A, B and C angles are adjusted by 3...5 degree steps and pressure towards ear and movement towards back are modified by 2-10 mm to alternate positions. The exact amounts are different due to shapes of the handset and thus will be documented case by case.

4.2 Handset Mode: Audio Performance Requirements

Please see the sections **7.1.6 Reference Skype client setup details** and **0 DUT client setup details** for instructions how to enable the testing mode used during all of the below measurements

4.2.1 Send path – signal level with loud speech

Purpose: To ensure that DUT send path provides optimal signal level for far end Skype client with loud speech input.

Input: Use recommended test position for DUT. Use following settings for the DUT Editor:

- Enable DUT Client Mode
- Disable AEC

Skype is allowed to automatically adjust the input gain setting for this test case.

Play the IEEE 269-2010 compressed male speech sample at loud speech level for 30 seconds to allow Skype to find optimal input gain level.

Play the IEEE 269-2010 compressed male speech sample at loud speech level and record the signal at reference Skype output.

Required: Calculate the average RMS level in reference Skype output for the active part of compressed male speech sample. **The level must be more than -24dBm0 RMS (equals -30dBFS RMS).** The peaks of the speech signals must not overload the input causing clipping.

Note: The compressed male speech is used for loud speech as it has a lower crest factor and thus better simulates a real world conditions as the crest factor of human voice is lower when people speak louder.

4.2.2 Send path – signal level with normal speech

Purpose: To ensure that DUT send path provides optimal signal level for far end Skype client with normal speech input.

Input: Use recommended test position for DUT. Use following settings for the DUT Editor:

- Enable DUT Client Mode
- Disable AEC

Skype is allowed to automatically adjust the input gain setting for this test case.

Play the IEEE 269-2010 uncompressed male speech sample at normal speech level for 30 seconds to allow Skype to find optimal input gain level.

Play the IEEE 269-2010 uncompressed male speech sample at normal speech level and record the signal at reference Skype output.

Required: Calculate the average RMS level in reference Skype output for the active part of male speech sample. **The level must be more than -24dBm0 RMS (equals -30dBFS RMS).** The peaks of the speech signals must not overload the input causing clipping.

4.2.3 Send path – signal level with quiet speech

Purpose: To ensure that DUT send path provides optimal signal level for far end Skype client with quiet speech input.

Input: Use recommended test position for DUT. Use following settings for the DUT Editor:

- Enable DUT Client Mode
- Disable AGC
- Disable AEC

The same input AGC setting as for normal speech level test is used for this test case. Skype automatic gain adjustment is disabled.

Play the IEEE 269-2010 uncompressed speech sample at quiet speech level and record the signal at reference Skype output.

Required: Calculate the average RMS level in reference Skype output for the active part of male speech sample. **The level must be more than -34 dBm0 RMS (equals -40 dBFS RMS).** The peaks of the speech signals must not overload the input causing clipping.

4.2.4 Send path – frequency response

Purpose: The test checks that the frequency response of the DUT send signal path is flat enough to meet Skype requirement considering technical limitations from wireless connection.

Input: Use recommended test position for DUT. Use following settings for the DUT Editor:

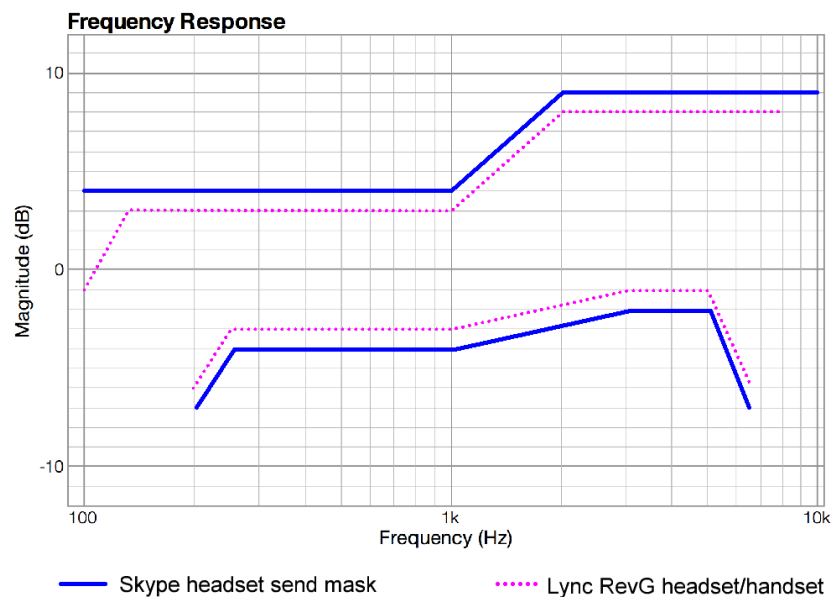
- Enable DUT Client Mode
- Disable AGC
- Disable AEC

The same input AGC setting as for normal speech level test is used for this test case. Skype automatic gain adjustment is disabled.

Play the [male artificial speech for frequency response](#) at normal speech level and record the signal at reference Skype output.

Required: The frequency response is calculated by comparing the 1/12 octave spectrum of the prior recorded reference result at MRP to the DUT send signal recorded at reference Skype output.

The resulting frequency response graph fits into the tolerance mask below:



Frequency	Lower limit	Upper limit
99 Hz	-80,0 dB	80,0 dB
100 Hz	-80,0 dB	4,0 dB
199 Hz	-80,0 dB	4,0 dB
200 Hz	-7,0 dB	4,0 dB
250 Hz	-4,0 dB	4,0 dB
1000 Hz	-4,0 dB	4,0 dB
2000 Hz	-3,0 dB	9,0 dB
3000 Hz	-2,0 dB	9,0 dB
5000 Hz	-2,0 dB	9,0 dB
6500 Hz	-7,0 dB	9,0 dB
6501 Hz	-80 dB	9,0 dB
10000 Hz	-80 dB	9,0 dB
10001 Hz	-80 dB	80 dB

4.2.5 Send path - speech signal to self noise ratio

Purpose: Too high self noise in microphone signal decreases the intelligibility of the speech and influences the total call quality in negative way. This test tests for speech to self noise ratio in DUT send path.

Input: Use recommended test position for DUT. Use following settings for the DUT Editor:

- Enable DUT Client Mode
- Disable AGC
- Disable AEC

The same input AGC setting as for normal speech level test is used for this test case. Skype automatic gain adjustment is disabled.

Play the [SpNR speech sample](#) at normal speech level and record the signal at reference Skype output.

Required: Calculate the RMS level in reference Skype output for the active speech part (the active speech does not include pauses or silences) → this is defined as Speech level

Calculate the A-weighted RMS level of noise in reference Skype output for the 1 second silence part at end of the test signal for SpNR → this is defined as Noise level.

The calculated speech to noise ratio (SpNR ratio) is more than 40 dB

(Speech level – Noise level > 40 dB)

Note: Please note that Skype specifies speech to noise ratio (SpNR), this is not the same as signal to noise ratio (SNR) specification often found in datasheet that specifies a sine signal to noise ratio.

4.2.6 Send path - speech signal to self noise ratio during speech

Purpose: Too high self noise in microphone signal decreases the intelligibility of the speech and influences the total call quality in negative way. Thus this test tests for speech to self noise level in DUT send path. this test tests for speech to self noise level in DUT send path during speech.

Input: Use recommended test position for DUT. Use following settings for the DUT Editor:

- Enable DUT Client Mode
- Disable AGC
- Disable AEC

The same input AGC setting as for normal speech level test is used for this test case. Skype automatic gain adjustment is disabled.

Play the [modified SpNR speech sample](#) at normal speech level and record the signal at reference Skype output.

Required: Calculate the RMS level in reference Skype output for the active speech part (the active speech does not include pauses or silences) → this is defined as Speech level.

Calculate the A-weighted RMS level of noise in reference Skype output for the 1 second silence part in middle of the speech signal for SpNR during speech → this is defined as Noise level during speech. (The sine signal components are filtered out by band stop filters prior to noise level calculation.)

The calculated speech level minus noise level (SpNR ratio) is more than 35 dB

(Speech level – Noise level during speech > 35 dB)

Note: Skype leaves a freedom to alter or improve the above test signal or post-processing of result without prior notice. The speech part of the signal will always remain the same.

4.2.7 Send path – single frequency interference

Purpose: Narrow-band noise, including single frequency interference, is an impairment that can be perceived as a tone , depending on its level relative to the overall weighted noise level. This can be caused by electrical noise in soundcards or by fan or hard disk drive noise on laptops. This requirement makes sure that no tonal noise is present in the send signal.

Input: Use recommended test position for DUT. Use following settings for the DUT Editor:

- Enable DUT Client Mode
- Disable AGC
- Disable AEC

The same input AGC setting as for normal speech level test is used for this test case. Skype automatic gain adjustment is disabled.

Play the [SpNR speech sample](#) at normal speech level and record the signal at reference Skype output.

Required: Calculate the A-weighted peak noise level for the 1 second silence part at end of the test signal for SpNR with an effective bandwidth of not more than 31 Hz. When using a 48 kHz sampling rate recording this means a minimum FFT size of 4096. Frequency range analyzed is 100Hz to 12000Hz. For FFT analysis the “Flat Top” windowing is employed.

The measured peak noise level is $\leq -74\text{dBm0}$ ($\leq -80\text{dBFS}$)

4.2.8 Receive path - preferred loudness level in ear

Purpose: To ensure that the receive path volume adjustment has enough adjustment range to allow volume settings for users personal preference.

Input: Use recommended test position for DUT.

Play the IEEE 269-2010 uncompressed male speech to the reference Skype input and record the acoustic signal in artificial ear.

Required: Calculate the RMS SPL level in artificial ear for the active part of male artificial speech signal (active speech does not include pauses or silences). **The level must be set to 78 dB ..82 dB SPL.**

The level set in this test case will be used for other receiving direction tests that refer to preferred loudness level for playback loudness.

4.2.9 Receive path - frequency response

Purpose: The test checks that the frequency response of the DUT receiving path is flat enough to meet minimum requirement taking into consideration the technical limitations of wireless technologies.

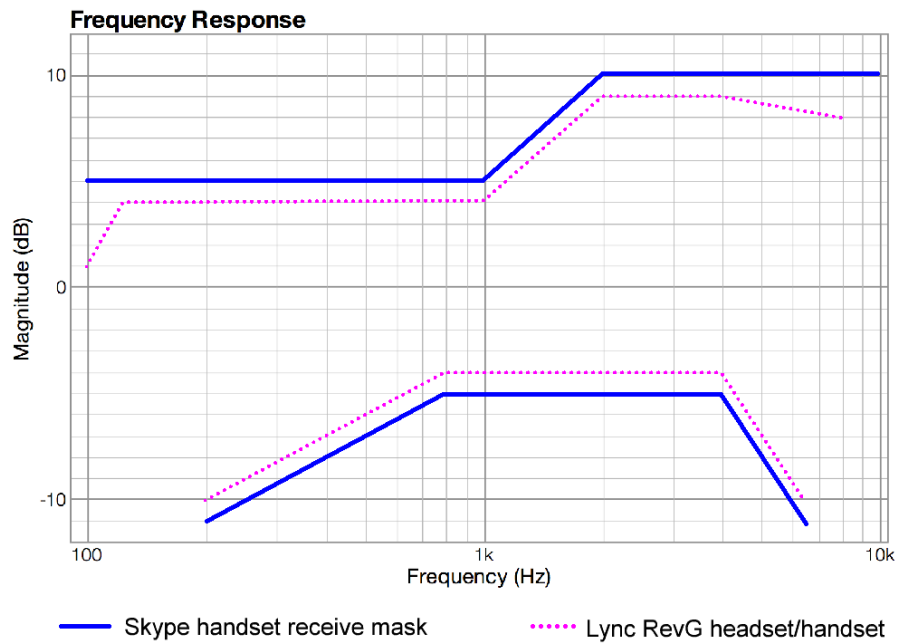
Input: Use recommended test position for DUT.

Use same receive loudness setting as set in 4.2.8.

Play the [male artificial speech for frequency response](#) to the reference Skype input. Record the acoustic receive signal in artificial ear.

Required: The frequency response is calculated by comparing the 1/12 octave spectrum of the prior recorded reference result to the DUT receive signal in artificial ear with the DRP/ERP correction applied.

The resulting frequency response graph fits into the tolerance mask below:



Frequency	Lower limit	Upper limit
99 Hz	-80,0 dB	80,0 dB
100 Hz	-80,0 dB	5,0 dB
199 Hz	-80,0 dB	5,0 dB
200 Hz	-11,0 dB	5,0 dB
800 Hz	-5,0 dB	5,0 dB
1000 Hz	-5,0 dB	5,0 dB
2000 Hz	-5,0 dB	10,0 dB
4000 Hz	-5,0 dB	10,0 dB
6500 Hz	-11,0 dB	10,0 dB
6501 Hz	-80,0 dB	10,0 dB
10000 Hz	-80,0 dB	10,0 dB
10001 Hz	-80,0 dB	-80,0 dB

4.2.10 Receive path - speech signal to noise ratio (SpNR)

- Purpose:** Too high self noise in receive path decreases the intelligibility of the speech and influences the total call quality in negative way. This test tests for speech to noise ratio in DUT receive path.
- Input:** Use recommended test position for DUT.
- Use same receive loudness setting as set in 4.2.8.
- Play the IEEE 269-2010 uncompressed male speech to the reference Skype input and record the acoustic signal in artificial ear.
- Required:** Calculate the RMS SPL level in artificial ear for the active part of male artificial speech (active speech does not include pauses or silences) → Speech level
- Calculate the A-weighted RMS SPL level in artificial ear for the 1 second silence part in middle of the male artificial speech signal for SpNR → Noise level.
- The calculated speech to noise ratio (SpNR ratio) is more than 40 dB.**
(*Speech level – Noise level > 40 dB*)
- Note:** Please note that Skype requires speech to noise ratio (SpNR), this is not the same as signal to noise ratio (SNR) specification often found in datasheet that specifies a sine signal to noise ratio.

4.2.11 Receive path – single frequency interference

- Purpose:** Tonal noise may be perceived in the loudspeaker signal if receive single frequency interference is too high.
- Input:** Use recommended test position for DUT.
- Play the signal containing silence into the REF Skype input
- Required:** Record a 5 second sound sample from HATS ear microphone. Measure the A-weighted peak noise level over the frequency range of 50 to 20000 Hz with an effective bandwidth of not more than 31 Hz. For 48 kHz sampling rate recording this means a minimum FFT size of 4096. For FFT analysis the “Flat Top” windowing is employed.
- Compare the analyzed level to the noise level calculated in test case** Error! Reference source not found. **Receive A-weighted single frequency interference noise peak level shall be at least 10 dB quieter than the averaged receive noise level.**

4.2.12 Weighted terminal coupling loss (TCLw)

Purpose: The amount of acoustic echo in the microphone signal is measured by the TCLw and the acoustic echo should be minimized by maximizing the physical distance between the loudspeaker and the microphone. For devices relying on the AEC in Lync, not meeting this requirement will result in echo leak, or distortion and attenuation of speech during double-talk (that is, near-end user and far-end participant talking simultaneously).

For devices with on-board AEC, a failure of this test will lead to echo leaks that are disruptive to the far-end participants.

The TCLw shall be normalized with respect to the nominal send loudness to account for any analog gain difference which would be compensated for by the digital AGC integrated in Skype. The nominal send loudness is defined as -18dBm0 (-24dBFS) The formula for the normalized TCLw is

$$\begin{aligned} TCLw &= TCLw_{measured} + (Send\ loudness_{measured} - Send\ loudness_{nominal}) \\ TCLw &= TCLw_{measured} + (Send\ loudness_{measured} + 24dBov) \end{aligned}$$

The TCLw shall be measured at preferred receive loudness using the IEEE Std. 269 male uncompressed speech signal.

Input: Use recommended test position for DUT. Use following settings for DUT editor **during preparation!**

- Enable DUT Client Mode
- Disable AEC

Play the IEEE 269-2010 uncompressed male speech to the reference Skype input. Skype is allowed to automatically adjust the input gain setting during the preparation period. After Skype has adjusted the input gain to optimal level the AGC is disabled.

Use following settings for DUT editor **for the actual test case run!**

- Enable DUT Client Mode
- Disable AGC
- Disable AEC

Use same receive loudness setting as set in 4.2.8.

Play the IEEE 269-2010 uncompressed male speech sample at normal speech level to reference Skype input and record the signal at reference Skype output.

Required: **TCLw for devices without in built AEC >30dB**
TCLw for devices with in built AEC >52dB

Note: The measurement shall be performed after system stability is reached (including convergence of any echo algorithms): this shall be accomplished by invoking the test signal for at least 2 seconds before the actual measurement occurs.

Measurements shall be done in 1/12th octave bands over a range of 100 Hz through 8000 Hz. The weighted terminal coupling loss is calculated according to ITU-T G.122 Annex B.4 (trapezoidal rule) using the frequency range of 100 to 8000 Hz rather than 300 to 3400 Hz.

Note that TCLw is also tested at maximum volume [here](#).

4.2.13 Echo path - acoustic echo cancellation

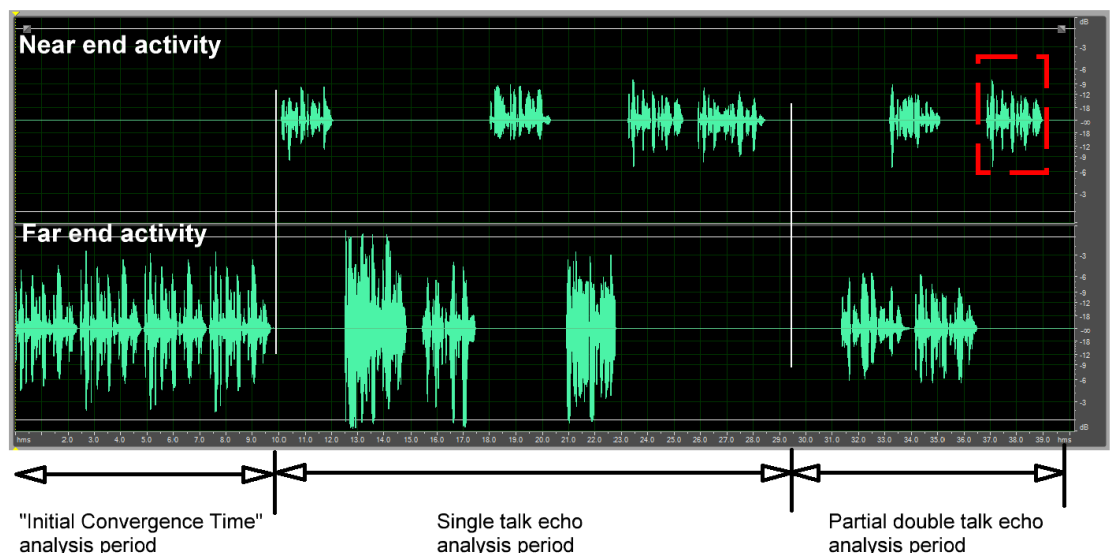
Purpose: The test checks the level of loudspeaker acoustic echo leaking back to far end output. The test signal includes near and far ends speech, alternating at different times. The talkers are occasionally overlapping to simulate use case where users take turns while speaking, but interrupt each other from time to time.

Input: Use recommended test position for DUT. Use following setting for the DUT editor:

- Enable DUT Client Mode

Play the preparation test signals to the reference Skype input and artificial mouth simultaneously. Use same receive loudness setting as set in 3.2.8.

Play the Echo test signal to the reference Skype input and artificial mouth simultaneously.



Required: The recording must comply with the following:

- **Initial Convergence Time** – the echo suppression should reach full cancellation at latest after 4 seconds from the start of far end activity
- **Residual echo / loss of convergence**
Calculate the **level versus time** and **spectrum versus time** graph of reference Skype output. Analyze and listen to the recording.
There should be no echo leaks or bursts higher than +5 dB in level compared to the send path noise floor level during far end speech activity in any frequency range between 50 Hz to 20 kHz.
- During the partial double talk there should be minimal loss of near end speech, especially the beginnings of the near end speech. The partial attenuation of near end speech is allowed, but the far end user should be able to recognize that the near end attempts to speak.
- **Send path noise floor stability / similarity.** It is beneficial to keep low level of comfort noise in the send signal path also during far end speech activity periods. Fully muting or attenuating the send signal during far end activity will make possible echo residuals more audible and could lead to failing result for this test case. Also the spectrum of the generated comfort noise should match that of the send path noise during silent periods of near end speech activity. If not – the noise floor changes will be very audible every time the far end is speaking.

4.2.14 Echo path – send path signal level during two way conversation

Purpose: The test checks the transmitted near end speech level during two way conversation. The send level might be low in such use case, especially on devices where the acoustic echo in microphone is very loud. This usually is due to physical distance between microphone(s) and speaker(s). The loud acoustic echo will force the analog gain control to adjust to lower gain, thus a digital amplification with faster adjustment speed will be needed to compensate for the level loss.

Input: Use the resulting recording from test case 4.2.13

The recording of send path signal in reference Skype output is analyzed. The section marked with red dotted line in above sequence is used for calculation of send path level during conversation

Required: Calculate the RMS level in reference Skype output for the active part of near end speech (time selection marked with red dotted line in above sequence). **The level must be more than -24 dBm0 RMS (equals -30dBFS RMS).**

4.2.15 Echo path – sidetone delay

Purpose: Ideally the sidetone should be a real-time signal. Sidetone delay less than 5 ms is generally perceived as a normal sidetone. Sidetone delay between 5 and 10 ms is generally perceived as unnatural sidetone, with an uncomfortable hollow characteristic. Sidetone delay greater than 10 ms is generally perceived as a distinct talker echo signal.

Input: Remove DUT from HATS. Play the short delay test signal into HATS mouth

Record both HATS ear and Reference Skype output signals simultaneously. Using a cross correlation method measure the HATS mouth to ear delay.

Attach DUT to HATS and enable the side tone path. Play the short delay test signal into HATS mouth

Record both HATS ear and Reference Skype output signals simultaneously. Using a cross correlation method measure the delay of the side tone.

Required: The sidetone delay for both handset and headset shall occur no more than 5ms after the mouth to ear delay signal.

It is desirable for the sidetone delay to be less than 1ms

Note: This test case is only applicable for devices that have a side tone path function.

4.3 Handset Mode: Loudness and Frequency Response Stability Requirements

4.3.1 Send path - frequency response stability.

Purpose: The test checks that the frequency response of the DUT send path does not change in major way when microphone position is slightly moved around the artificial mouth.

Input: Use recommended test position for DUT. Use following settings for the DUT Editor:

- Enable DUT Client Mode
- Disable AGC
- Disable AEC

Play the full [male artificial speech for frequency response](#) sample at normal speech level and record the signal at reference Skype output → send reference result.

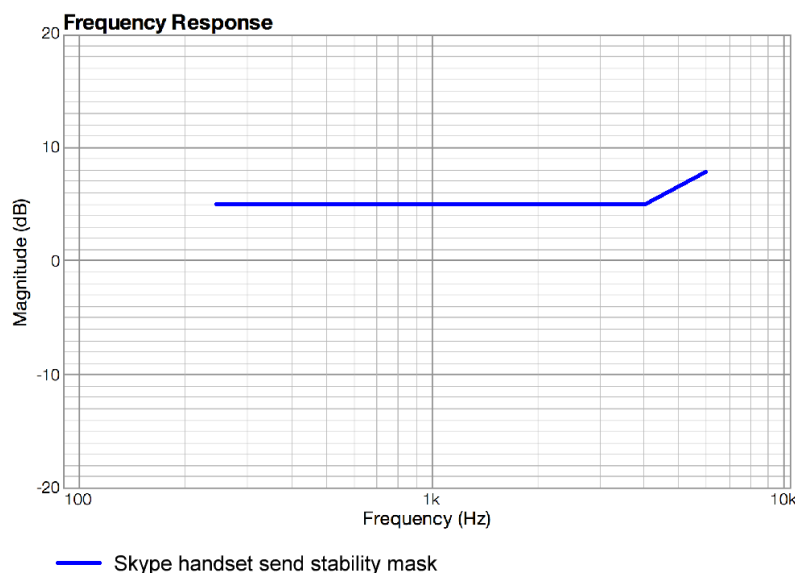
Alternate the DUT position 3 times every time recording the [male artificial speech for frequency response](#).

Required: The frequency response is calculated by comparing the 1/3 octave spectrum of the prior recorded reference result at MRP to the DUT send signal recorded at reference Skype output.

The resulting frequency responses are compared:

Maximum and minimum of each 1/3 octave band is calculated for all recordings

The difference between maximum and minimum curves is calculated by subtracting the minimum value from the maximum value in every 1/3 octave band. (Note this difference is always positive) **This difference in each octave band must be less than the tolerance mask below:**



4.3.2 Receive path – earpiece frequency response stability

Purpose: The test checks that the frequency response of the DUT receiving path does not change in major way when earpiece position is slightly moved around on artificial ear.

Input: Use recommended test position for DUT. Make a Skype to Skype call from DUT to reference Skype client in lossless local network condition.

Play the [male artificial speech for frequency response](#) to the reference Skype input.
Record the acoustic receive signal in artificial ear → [receive reference result](#).

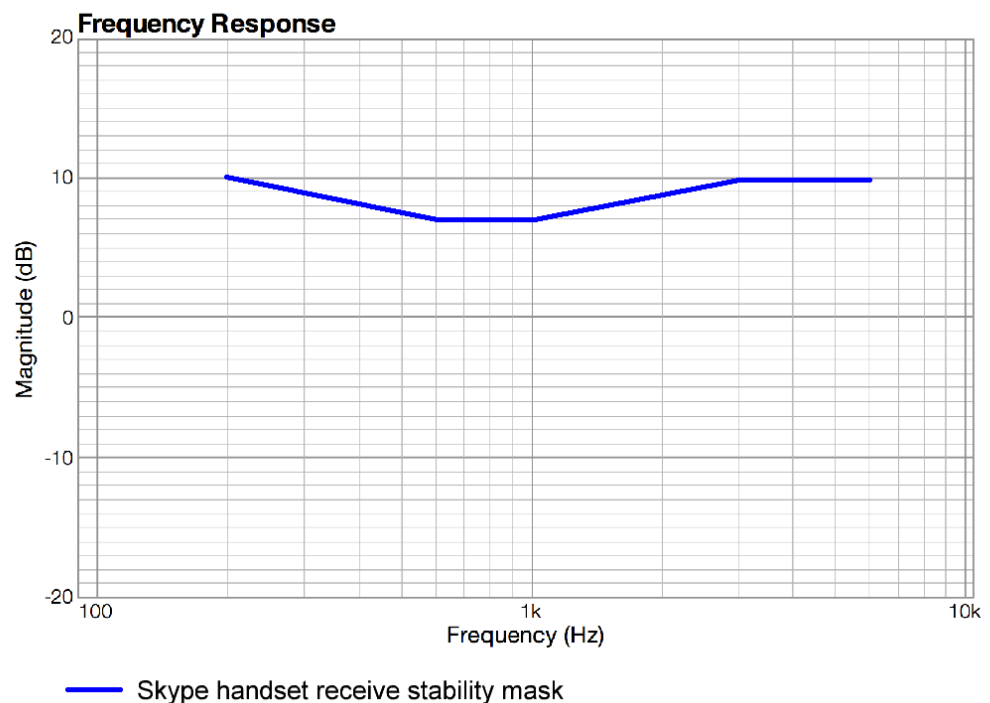
Alternate the DUT position 3 times every time recording the [male artificial speech for frequency response](#).

Required: The frequency response is calculated by comparing the 1/3 octave spectrum of the prior recorded reference result to the DUT receive signal in artificial ear with the DRP/ERP correction applied.

The resulting frequency responses are compared:

Maximum and minimum of each 1/3 octave band is calculated for all recordings

The difference between maximum and minimum curves is calculated by subtracting the minimum value from the maximum value in every 1/3 octave band. (Note this difference is always positive) **This difference in each octave band must be less than the tolerance mask below.**



4.3.3 Receive path – loudness level adjustment range

Purpose: To ensure that the earpiece/loudspeaker is able to provide possibility to adjust receiving signal loudness in ear for personal preference.

Skype uses ITU-T P.57 specified Type 3.3 artificial ear for receive direction.

Input: Adjust the DUT playback volume to 5...25% of full available adjustment range → low level volume setting.

Play the [male artificial speech](#) to the reference Skype input. Record the acoustic receive signal in artificial ear.

Adjust the DUT playback volume to 75...100% of full available adjustment range → high level volume setting.

Play the male artificial speech to the reference Skype input. Record the acoustic receive signal in artificial ear.

Required: Calculate the RMS SPL level in artificial ear for the active part of male artificial speech signal for low level volume setting.

Calculate the RMS SPL level in artificial ear for the active part of male artificial speech signal for high level volume setting.

The high level minus low level must be > 20 dB.

4.3.4 Weighted terminal coupling loss (TCLw) at max volume

Purpose: The amount of acoustic echo in the microphone signal is measured by the TCLw and the acoustic echo should be minimized by maximizing the physical distance between the loudspeaker and the microphone. For devices relying on the AEC in Lync, not meeting this requirement will result in echo leak, or distortion and attenuation of speech during double-talk (that is, near-end user and far-end participant talking simultaneously).

For devices with on-board AEC, a failure of this test will lead to echo leaks that are disruptive to the far-end participants.

The TCLw shall be normalized with respect to the nominal send loudness to account for any analog gain difference which would be compensated for by the digital AGC integrated in Skype. The nominal send loudness is defined as -18dBm0 (-24dBFS) The formula for the normalized TCLw is

$$\begin{aligned} TCLw &= TCLw_{measured} + (Send\ loudness_{measured} - Send\ loudness_{nominal}) \\ TCLw &= TCLw_{measured} + (Send\ loudness_{measured} + 24dBov) \end{aligned}$$

The TCLw shall be measured at preferred receive loudness using the IEEE Std. 269 male uncompressed speech signal.

Input: Use recommended test position for DUT. Use following settings for DUT editor **during preparation!**

- Enable DUT Client Mode
- Disable AEC

Play the IEEE 269-2010 uncompressed male speech to the reference Skype input. Skype is allowed to automatically adjust the input gain setting during the preparation period. After Skype has adjusted the input gain to optimal level the AGC is disabled.

Use following settings for DUT editor **for the actual test case run!**

- Enable DUT Client Mode
- Disable AGC
- Disable AEC

Use the max receive loudness level (adjust handset volume to max).

Play the IEEE 269-2010 uncompressed male speech sample at normal speech level to reference Skype input and record the signal at reference Skype output.

Required: **TCLw for devices without in built AEC >30dB**
TCLw for devices with in built AEC >45dB

Note: The measurement shall be performed after system stability is reached (including convergence of any echo algorithms): this shall be accomplished by invoking the test signal for at least 2 seconds before the actual measurement occurs.

Measurements shall be done in 1/12th octave bands over a range of 100 Hz through 8000 Hz. The weighted terminal coupling loss is calculated according to ITU-T G.122 Annex B.4 (trapezoidal rule) using the frequency range of 100 to 8000 Hz rather than 300 to 3400 Hz.

4.3.5 Ring tone loudness

Purpose: To verify that user can hear the ringing of incoming call in normal and noisy every-day-life environment.

Input: Place the DUT into HATS handset positioner. Set the measurement microphone to 10 cm distance from the handset.

Set ring tone volume of the DUT handset to maximum level. Use a level calibrated measurement microphone and recording setup or a calibrated SPL meter. Play ring tones of handset one by one.

Required: Record the ring tones with a calibrated microphone and check SPL levels offline or check SPL levels with SPL meter. For each tone measure the maximum hold value of SPLfast RMS (125 ms exponential time weighting).

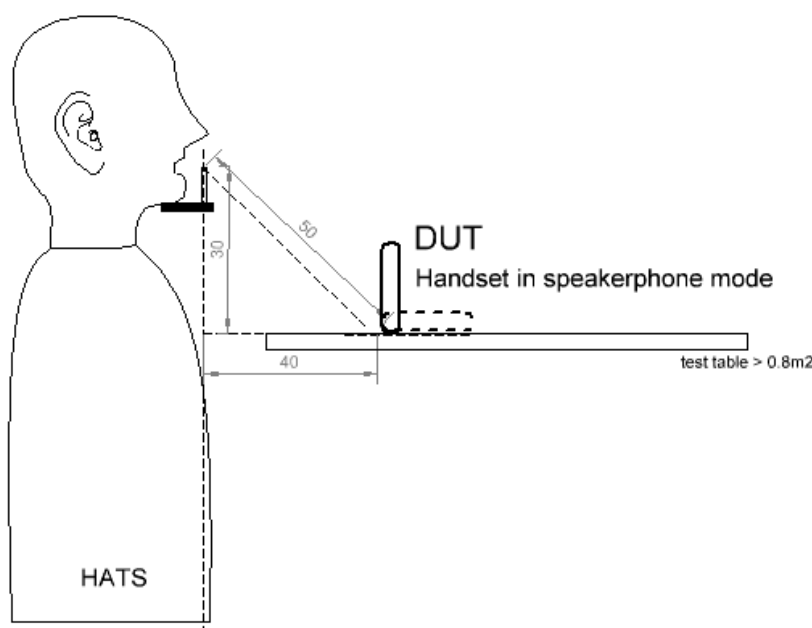
The DUT must provide minimum two ring tones where maximum hold value of SPLfast RMS is louder than 85 dB SPL

Exception: Lower ring tone levels can be accepted if manufacturer provides a detailed written explanation why the above required level cannot be provided. Skype will validate the request and decide case by case weather to accept the lower ring tone level or not.

For example such case can happen if the safety regulations of specific market regions require the product to be limited in ring tone loudness to lower level than required above.

4.4 Handset in Speakerphone Mode (if Available): Audio Performance Requirements

Handset in speakerphone mode test position.



4.4.1 Send path - signal level with loud speech

Purpose: To ensure that DUT send path provides optimal signal level for far end Skype client with loud speech input.

Input: Use recommended test position for DUT. Use following settings for the DUT Editor:

- Enable DUT Client Mode
- Disable AEC

Skype is allowed to automatically adjust the input gain setting for this test case.

Play the IEEE 269-2010 compressed male speech sample at loud speech level for 30 seconds to allow Skype to find optimal input gain level.

Play the IEEE 269-2010 compressed male speech sample at loud speech level and record the signal at reference Skype output.

Required: Calculate the average RMS level in reference Skype output for the active part of compressed male speech sample. **The level must be more than -24dBm0 RMS (equals -30dBFS RMS).** The peaks of the speech signals must not overload the input causing clipping.

Note: The compressed male speech is used for loud speech as it has a lower crest factor and thus better simulates a real world conditions as the crest factor of human voice is lower when people speak louder.

4.4.2 Send path - signal level with normal speech

Purpose: To ensure that DUT send path provides optimal signal level for far end Skype client with normal speech input.

Input: Use recommended test position for DUT. Use following settings for the DUT Editor:

- Enable DUT Client Mode
- Disable AEC

Skype is allowed to automatically adjust the input gain setting for this test case.

Play the IEEE 269-2010 uncompressed male speech sample at normal speech level for 30 seconds to allow Skype to find optimal input gain level.

Play the IEEE 269-2010 uncompressed male speech sample at normal speech level and record the signal at reference Skype output.

Required: Calculate the average RMS level in reference Skype output for the active part of male speech sample. **The level must be more than -28dBm0 RMS (equals -34dBFS RMS).** The peaks of the speech signals must not overload the input causing clipping.

4.4.3 Send path - signal level with quiet speech

Purpose: To ensure that DUT send path provides optimal signal level for far end Skype client with quiet speech input.

Input: Use handset in speakerphone mode test position for DUT. Use following settings for the DUT Editor:

- Enable DUT Client Mode
- Disable AGC
- Disable AEC

The same input AGC setting as for normal speech level test is used for this test case. Skype automatic gain adjustment is disabled.

Play the IEEE 269-2010 uncompressed speech at quiet speech level and record the signal at reference Skype output.

Required: Calculate the average RMS level in reference Skype output for the active part of male speech sample. **The level must be more than -40dBm0 RMS (equals -46dBFS RMS).** The peaks of the speech signals must not overload the input causing clipping.

4.4.4 Send path - frequency response

Purpose: The test checks that the frequency response of the DUT send signal path is flat enough to meet Skype requirement and take full advantage of Skype voice codec bandwidth.

Input: Use recommended test position for DUT. Use following settings for the DUT Editor:

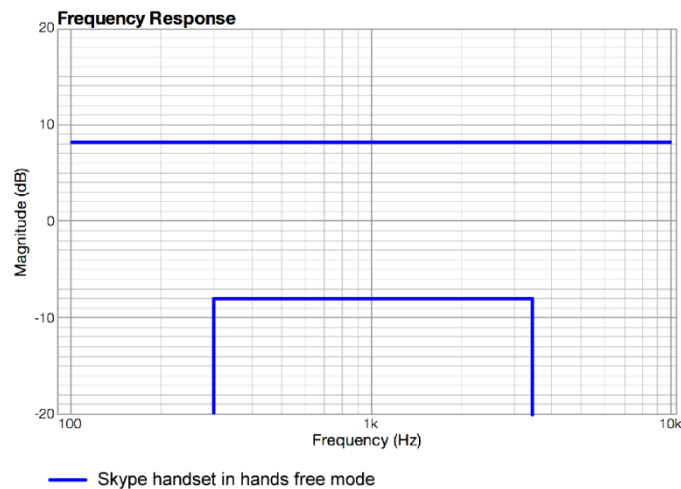
- Enable DUT Client Mode
- Disable AGC
- Disable AEC

The same input AGC setting as for normal speech level test is used for this test case.

Play the [male artificial speech for frequency response](#) at normal speech level and record the signal at reference Skype output.

Required: The frequency response is calculated by comparing the 1/3 octave spectrum of the prior recorded reference result at MRP to the DUT send signal recorded at reference Skype output.

The resulting frequency response graph fits into the tolerance mask below:



Frequency	Lower limit	Upper limit
99 Hz	-80,0 dB	80,0 dB
100 Hz	-80,0 dB	8,0 dB
299 Hz	-80,0 dB	8,0 dB
300 Hz	-8,0 dB	8,0 dB
3400 Hz	-8,0 dB	8,0 dB
3401 Hz	-80,0 dB	8,0 dB
10000 Hz	-80,0 dB	8,0 dB
10001 Hz	-80,0 dB	80,0 dB

4.4.5 Send path - speech signal to self noise ratio

Purpose: Too high self noise in microphone signal decreases the intelligibility of the speech and influences the total call quality in negative way. This test tests for speech to self noise ratio in DUT send path.

Input: Use recommended test position for DUT. Use following settings for the DUT Editor:

- Enable DUT Client Mode
- Disable AGC
- Disable AEC

The same input AGC setting as for normal speech level test is used for this test case.

Play the [SpNR speech sample](#) at normal speech level and record the signal at reference Skype output.

Required: Calculate the RMS level in reference Skype output for the active speech part (the active speech does not include pauses or silences) → this is defined as Speech level

Calculate the A-weighted RMS level of noise in reference Skype output for the 1 second silence part at end of the test signal for SpNR → this is defined as Noise level.

The calculated speech to noise ratio (SpNR ratio) is more than 25 dB.

(Speech level – Noise level > 25 dB)

Note: Please note that Skype specifies speech to noise ratio (SpNR), this is not the same as signal to noise ratio (SNR) specification often found in datasheet that specifies a sine signal to noise ratio.

4.4.6 Receive path - preferred loudness level in ear

Purpose: To ensure that the receive path volume adjustment has enough adjustment range to allow setting volume to preferred listening level.

Input: Use handset in speakerphone mode test position for DUT. Use following settings for the DUT Editor:

- Enable DUT Client Mode
- Disable AGC
- Disable AEC

The same input AGC setting as for normal speech level test is used for this test case.

Play the IEEE 269-2010 uncompressed male speech to the reference Skype input and record the acoustic signal in artificial ear.

Required: Calculate the RMS SPL level in artificial ear for the active part of male artificial speech signal (active speech does not include pauses or silences). **The level must be set to 58..62 dB SPL.**

4.4.7 Receive path - frequency response

Purpose: The test checks that the frequency response of the DUT receiving path is flat enough to meet minimum requirement and take full advantage of Skype voice codec bandwidth.

Skype uses ITU-T P.57 specified Type 3.3 artificial ear for receive direction testing with free field correction applied.

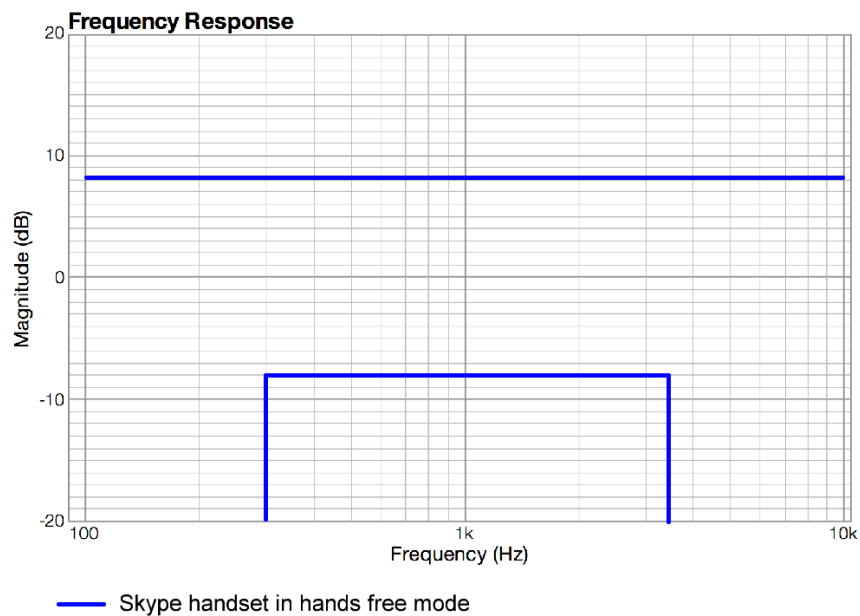
Input: Use handset in speakerphone mode test position for DUT.

Use same receive loudness setting as set in 4.4.6.

Play the [male artificial speech for frequency response](#) to the reference Skype input.
Record the acoustic receive signal in artificial ear.

Required: The frequency response is calculated by comparing the 1/3 octave spectrum of the prior recorded reference result to the DUT receive signal in artificial ear with the free field correction applied.

The resulting frequency response graph fits into the tolerance mask below:



Frequency	Lower limit	Upper limit
99 Hz	-80,0 dB	80,0 dB
100 Hz	-80,0 dB	8,0 dB
299 Hz	-80,0 dB	8,0 dB
300 Hz	-8,0 dB	8,0 dB
3400 Hz	-8,0 dB	8,0 dB
3401 Hz	-80,0 dB	8,0 dB
10000 Hz	-80,0 dB	8,0 dB
10001 Hz	-80,0 dB	80,0 dB

4.4.8 Weighted terminal coupling loss (TCLw)

Purpose: The amount of acoustic echo in the microphone signal is measured by the TCLw and the acoustic echo should be minimized by maximizing the physical distance between the loudspeaker and the microphone. For devices relying on the AEC in Lync, not meeting this requirement will result in echo leak, or distortion and attenuation of speech during double-talk (that is, near-end user and far-end participant talking simultaneously).

For devices with on-board AEC, a failure of this test will lead to echo leaks that are disruptive to the far-end participants.

The TCLw shall be normalized with respect to the nominal send loudness to account for any analog gain difference which would be compensated for by the digital AGC integrated in Skype. The nominal send loudness is defined as -18dBm0 (-24dBFS).

The formula for the normalized TCLw is

$$\begin{aligned} TCLw &= TCLw_{measured} + (Send\ loudness_{measured} - Send\ loudness_{nominal}) \\ TCLw &= TCLw_{measured} + (Send\ loudness_{measured} + 24dBov) \end{aligned}$$

The TCLw shall be measured at preferred receive loudness using the IEEE Std. 269 male uncompressed speech signal.

Input: Use handset in speakerphone mode test position for DUT. Use following settings for DUT editor **during preparation!**

- Enable DUT Client Mode
- Disable AEC

Play the IEEE 269-2010 uncompressed male speech to the reference Skype input. Skype is allowed to automatically adjust the input gain setting during the preparation period. After Skype has adjusted the input gain to optimal level the AGC is disabled.

Use following settings for DUT editor **for the actual test case run!**

- Enable DUT Client Mode
- Disable AGC
- Disable AEC

Use same receive loudness setting as set in 4.4.6

Play the IEEE 269-2010 uncompressed male speech sample at normal speech level to reference Skype input and record the signal at reference Skype output.

Required: **TCLw for devices without in built AEC >-15dB**
TCLw for devices with in built AEC >45dB

Note: The measurement shall be performed after system stability is reached (including convergence of any echo algorithms): this shall be accomplished by invoking the test signal for at least 2 seconds before the actual measurement occurs.

Measurements shall be done in 1/3rd octave bands over a range of 100 Hz through 8000 Hz. The weighted terminal coupling loss is calculated according to ITU-T G.122 Annex B.4 (trapezoidal rule) using the frequency range of 100 to 8000 Hz rather than 300 to 3400 Hz.

4.4.9 Echo path - acoustic echo cancellation

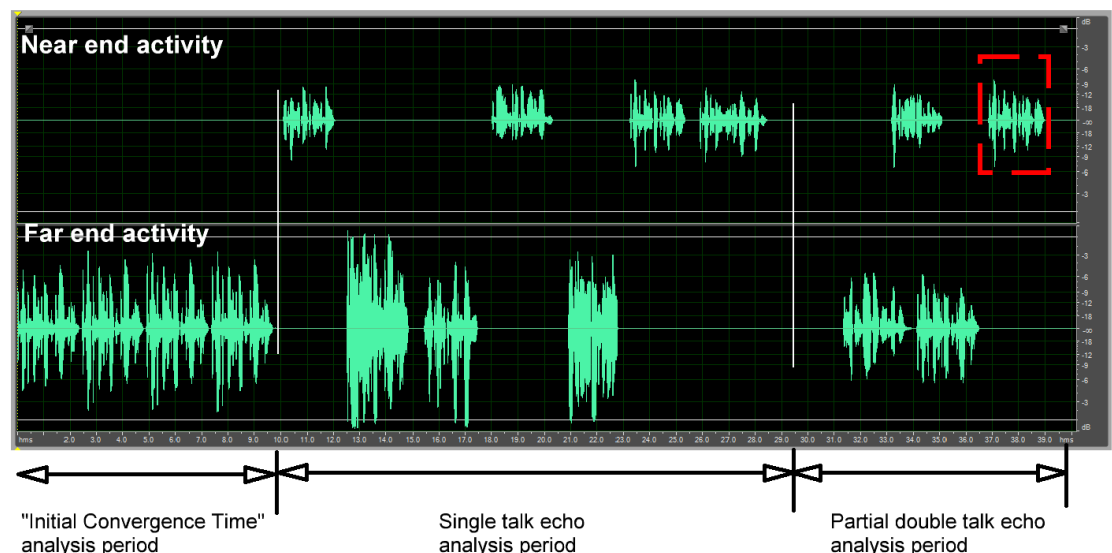
Purpose: The test checks the level of loudspeaker acoustic echo leaking back to far end output. The test signal includes near and far ends speech, alternating at different times. The talkers are occasionally overlapping to simulate use case where users take turns while speaking, but interrupt each other from time to time.

Input: Use recommended test position for DUT. Use following setting for the DUT editor:

- Enable DUT Client Mode

Play the preparation test signals to the reference Skype input and artificial mouth simultaneously. Use same receive loudness setting as set in 3.2.8.

Play the Echo test signal to the reference Skype input and artificial mouth simultaneously.



Required: The recording must comply with the following:

- **Initial Convergence Time** – the echo suppression should reach full cancellation at latest after 4 seconds from the start of far end activity
- **Residual echo / loss of convergence**
Calculate the **level versus time** and **spectrum versus time** graph of reference Skype output. Analyze and listen to the recording.
There should be no echo leaks or bursts higher than +5 dB in level compared to the send path noise floor level during far end speech activity in any frequency range between 50 Hz to 20 kHz.
- During the partial double talk there should be minimal loss of near end speech, especially the beginnings of the near end speech. The partial attenuation of near end speech is allowed, but the far end user should be able to recognize that the near end attempts to speak.
- **Send path noise floor stability / similarity.** It is beneficial to keep low level of comfort noise in the send signal path also during far end speech activity periods. Fully muting or attenuating the send signal during far end activity will make possible echo residuals more audible and could lead to failing result for this test case. Also the spectrum of the generated comfort noise should match that of the send path noise during silent periods of near end speech activity. If not – the noise floor changes will be very audible every time the far end is speaking.

4.4.10 Echo path – send path signal level during two way conversation

Purpose: The test checks the transmitted near end speech level during two way conversation. The send level might be low in such use case, especially on devices where the acoustic echo in microphone is very loud. This usually is due to physical distance between microphone(s) and speaker(s). The loud acoustic echo will force the analog gain control to adjust to lower gain, thus a digital amplification with faster adjustment speed will be needed to compensate for the level loss.

Input: Use the resulting recording from test case 4.4.9

The recording of send path signal in reference Skype output is analyzed. The section marked with red dotted line in above sequence is used for calculation of send path level during conversation

Required: Calculate the RMS level in reference Skype output for the active part of near end speech (time selection marked with red dotted line in above sequence). **The level must be more than -24 dBm0 RMS (equals -30 dBFS RMS).**

4.5 Handset: Supporting Audio Documentation Requirements

In addition to the user manual (the one that comes with the product) we also ask for supporting audio documentation (only for certification testing purpose). Such documentation contains engineering data and engineering test data for the product.

Earpiece below means the acoustic output device, for example a small loudspeaker. Ring tone loudspeaker means the component that reproduces ring tones. In some devices it is the same component that reproduces speech, in others it is a separate element.

4.5.1 Verifying supporting documentation for Handset audio

Purpose: Solution must come with a supporting audio documentation (only for certification testing purposes).

Required: DUT arrives with supporting audio documentation that contains information about:

- Active signal processing: yes/no, if yes then:
 - Acoustic echo cancellation: yes/no, in sending (i.e., microphone) or/and receiving (i.e. earpiece) directions.
 - Noise suppression: yes/no, in sending or/and receiving directions
 - Automated Gain Control: yes/no, in sending or/and receiving directions
 - Other: describe what, sending or/and receiving directions
- Microphone: Directionality/design principle
- Microphone: Frequency range (lowest and highest audible frequencies)
- Earpiece: Lowest and highest designed audible frequencies
- Earpiece: Designed acoustic SPL for receiving speech signal at the user's ear
- Ring tones: Number of tones
- Ring tones: Types of tones: MP3, Midi, wav/PCM etc...
- Ring tones from the loudspeaker: Maximum level of the loudest ring tone at 10 cm distance from the handset in the free field conditions (SPL, fast time weighting max hold)
- Ring tones from the loudspeaker: Level that half of the ring tones exceed when volume settings are on maximum. Measured at 10 cm distance from the handset in the free field conditions (SPL, fast time weighting max hold)

5.0 Speakerphone Audio UI Group

Tests in this section are part of the HEAD Acoustics ACQUA Skype AS standard.

Note: If the **Device Under Test supports Skype Video calls**, then all of the send, receive and echo path tests below have to be done by enabling a two way video call. Skype clients on both sides will determine the resolution, bitrate and frame rate during the call.

5.1 Speakerphone: Audio Test Instructions

5.1.1 Test environment – anechoic room

Please refer to [Test Setup And Test Environment Details](#) for details on [Skype anechoic room](#) or alternative test environments.

5.1.2 Measurement setup

5.1.2.1 *Mouth simulator*

For Speakerphone acoustic UI group there are two acceptable mouth simulators

- Head and Torso Simulator (HATS) compliant with ITU-T P.58 is used as a mouth simulator.
- ITU-T P.51 compliant mouth simulator such as B&K type 4227 or G.R.A.S Type 44AA or 44AB

The frequency spectrum of mouth simulator is calibrated and frequency compensated at Mouth Reference Point (MRP) to be flat between 100 Hz to 11000 Hz.

The normal speech level for active speech part of male artificial speech is calibrated to be 89 dB SPL at MRP.

Quiet and loud speech signals are respectively 10 dB quieter and 10 dB louder compared to normal speech level.

5.1.2.2 *Ear simulator / measurement microphone*

For Speakerphone acoustic UI group there are two acceptable ways to measure receive direction parameters

- Head and Torso Simulator (HATS) compliant with ITU-T P.58 and artificial ear compliant with ITU-T P.57 Type 3.3 is used as a ear simulator. (Skype uses Drum Reference Point (DRP) to **Free Field Correction (FF)** for all receiving path frequency response measurements of devices in Speakerphone UI category)
- Free Field measurement microphone fulfilling the IEC 60651 type 0 or type 1 requirement.

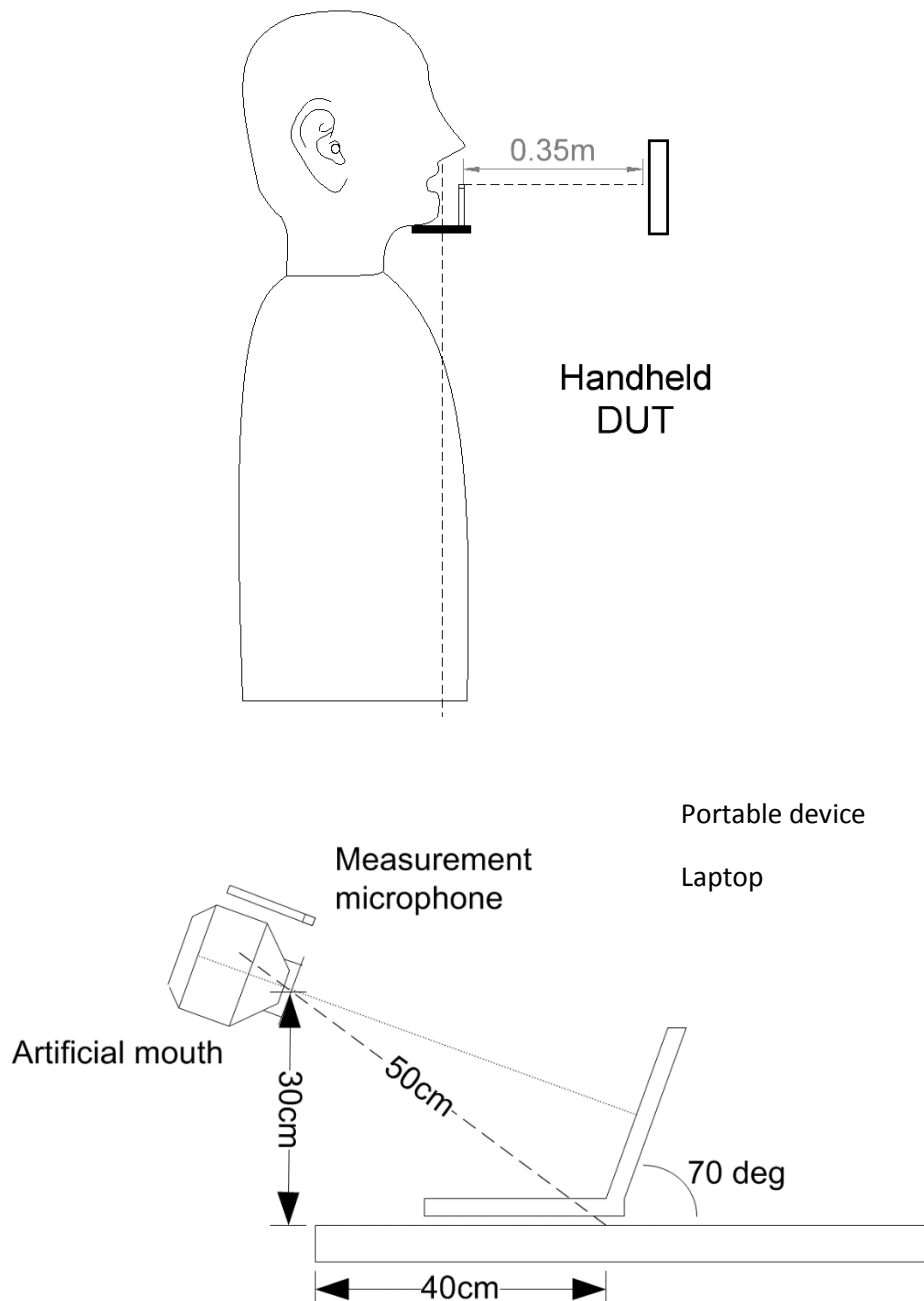
The measurement microphone is calibrated using microphone calibrator. Skype uses G.R.A.S Type 42AB microphone calibrator for this purpose.

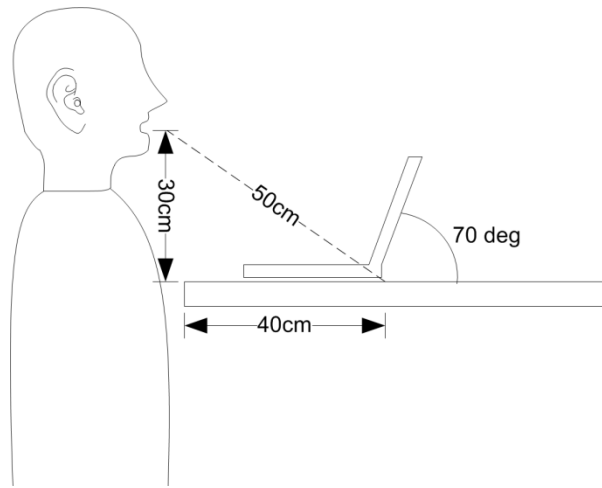
5.2 Speakerphone: DUT Usage Distance 35..50cm (Handheld and Portable Speakerphone)

Handheld speakerphone, tablet and laptop PC test position.

The DUT is placed in front of HATS as drawn below and test table is not used. The physical fixture holding the device should not block the microphone inlet / speaker(s) outlet.

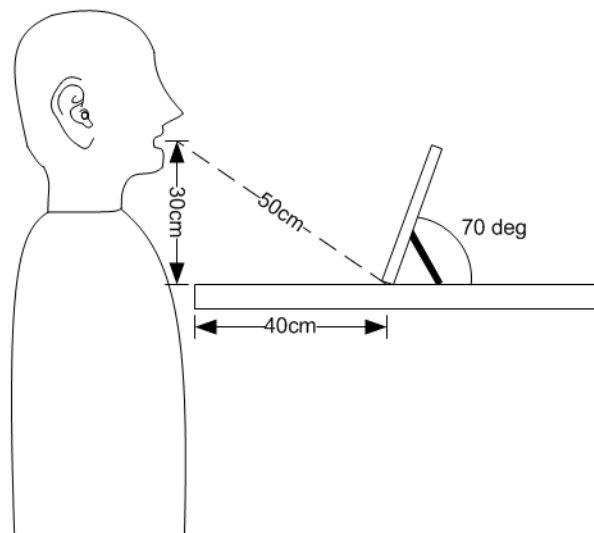
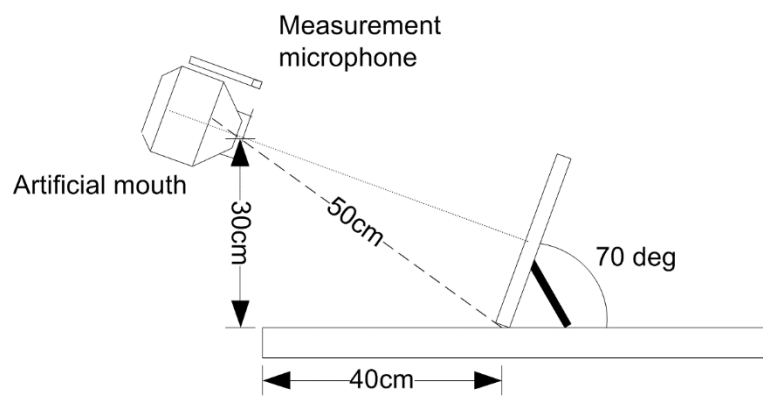
If manufacturer of device or user manual advises other usage scenario, then that is taken into account and a measurement position is agreed between Skype audio engineers and manufacturer/vendor.





Portable device

Tablet PC



5.2.1 Send path - signal level with loud speech

Purpose: To ensure that DUT send path provides optimal signal level for far end Skype client with loud speech input.

Input: Use recommended test position for DUT. Use following settings for the DUT Editor:

- Enable DUT Client Mode
- Disable AEC

Skype is allowed to automatically adjust the input gain setting for this test case.

Play the IEEE 269-2010 compressed male speech sample at loud speech level for 30 seconds to allow Skype to find optimal input gain level.

Play the IEEE 269-2010 compressed male speech sample at loud speech level and record the signal at reference Skype output.

Required: Calculate the average RMS level in reference Skype output for the active part of compressed male speech sample. **The level must be more than -24dBm0 RMS (equals -30dBFS RMS).** The peaks of the speech signals must not overload the input causing clipping.

Note: The compressed male speech is used for loud speech as it has a lower crest factor and thus better simulates a real world conditions as the crest factor of human voice is lower when people speak louder.

5.2.2 Send path - signal level with normal speech

Purpose: To ensure that DUT send path provides optimal signal level for far end Skype client with normal speech input.

Input: Use recommended test position for DUT. Use following settings for the DUT Editor:

- Enable DUT Client Mode
- Disable AEC

Skype is allowed to automatically adjust the input gain setting for this test case.

Play the IEEE 269-2010 uncompressed male speech sample at normal speech level for 30 seconds to allow Skype to find optimal input gain level.

Play the IEEE 269-2010 uncompressed male speech sample at normal speech level and record the signal at reference Skype output.

Required: Calculate the average RMS level in reference Skype output for the active part of male speech sample. **The level must be more than -24dBm0 RMS (equals -30dBFS RMS).** The peaks of the speech signals must not overload the input causing clipping.

5.2.3 Send path - signal level with quiet speech

Purpose: To ensure that DUT send path provides optimal signal level for far end Skype client with quiet speech input.

Input: Use recommended test position for DUT. Use following settings for the DUT Editor:

- Enable DUT Client Mode
- Disable AGC
- Disable AEC

The same input AGC setting as for normal speech level test is used for this test case. Skype automatic gain adjustment is disabled.

Play the IEEE 269-2010 uncompressed speech at quiet speech level and record the signal at reference Skype output.

Required: Calculate the average RMS level in reference Skype output for the active part of male speech sample. **The level must be more than -34dBm0 RMS (equals -40dBFS RMS).** The peaks of the speech signals must not overload the input causing clipping.

5.2.4 Send path - frequency response

Purpose: The test checks that the frequency response of the DUT send signal path is flat enough to meet Skype requirement and take full advantage of Skype voice codec bandwidth.

Input: Use recommended test position for DUT. Use following settings for the DUT Editor:

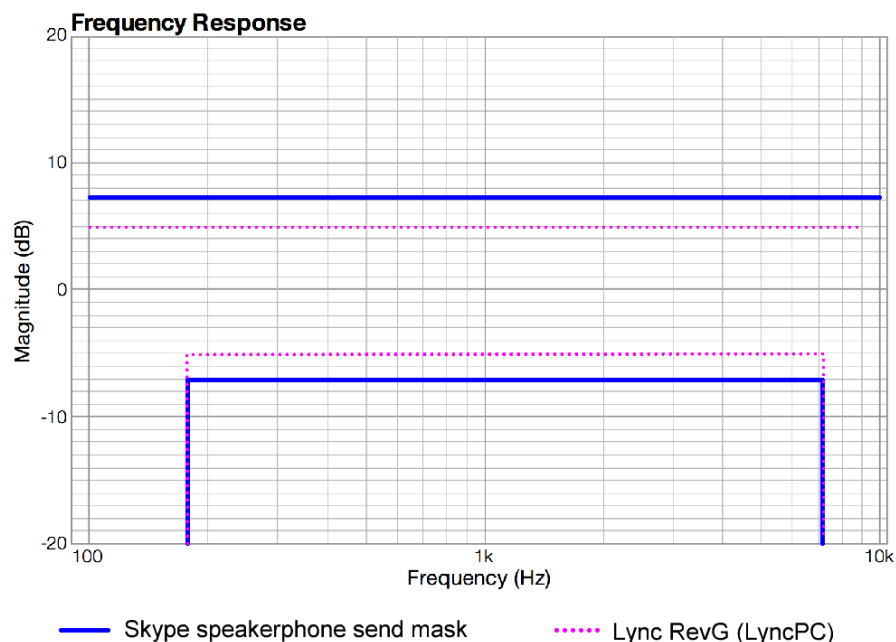
- Enable DUT Client Mode
- Disable AGC
- Disable AEC

The same input AGC setting as for normal speech level test is used for this test case. Skype automatic gain adjustment is disabled.

Play the [male artificial speech for frequency response](#) at normal speech level and record the signal at reference Skype output.

Required: The frequency response is calculated by comparing the 1/3 octave spectrum of the prior recorded reference result to the DUT receive signal in artificial ear with the free field correction applied.

The resulting frequency response graph fits into the tolerance mask below:



Frequency	Lower limit	Upper limit
99 Hz	-80,0 dB	80,0 dB
100 Hz	-80,0 dB	7,0 dB
179 Hz	-80,0 dB	7,0 dB
180 Hz	-7,0 dB	7,0 dB
7100 Hz	-7,0 dB	7,0 dB
7101 Hz	-80,0 dB	7,0 dB
10000 Hz	-80,0 dB	7,0 dB
10001 Hz	-80,0 dB	80,0 dB

NB! The tolerance mask is floating (not fixed) and will move up/down trying to center the measured response between tolerance limits.

5.2.5 Send path - speech signal to self noise ratio

Purpose: Too high self noise in microphone signal decreases the intelligibility of the speech and influences the total call quality in negative way. This test tests for speech to self noise ratio in DUT send path.

Input: Use recommended test position for DUT. Use following settings for the DUT Editor:

- Enable DUT Client Mode
- Disable AGC
- Disable AEC

The same input AGC setting as for normal speech level test is used for this test case. Skype automatic gain adjustment is disabled.

Play the [SpNR speech sample](#) at normal speech level and record the signal at reference Skype output.

Required: Calculate the RMS level in reference Skype output for the active speech part (the active speech does not include pauses or silences) → this is defined as Speech level

Calculate the A-weighted RMS level of noise in reference Skype output for the 1 second silence part at end of the test signal for SpNR → this is defined as Noise level.

Solutions without cooling fan ≥ 30

Solutions with active cooling ≥ 25

Note: Please note that Skype specifies speech to noise ratio (SpNR), this is not the same as signal to noise ratio (SNR) specification often found in datasheet that specifies a sine signal to noise ratio.

5.2.6 Send path - speech signal to self noise ratio during speech

Purpose: Too high self noise in microphone signal decreases the intelligibility of the speech and influences the total call quality in negative way. Thus this test tests for speech to self noise level in DUT send path. This test tests for speech to self noise level in DUT send path during speech.

Input: Use recommended test position for DUT. Use following settings for the DUT Editor:

- Enable DUT Client Mode
- Disable AGC
- Disable AEC

The same input AGC setting as for normal speech level test is used for this test case. Skype automatic gain adjustment is disabled.

Play the [modified SpNR speech sample](#) at normal speech level and record the signal at reference Skype output.

Required: Calculate the RMS level in reference Skype output for the active speech part (the active speech does not include pauses or silences) → this is defined as Speech level.

Calculate the A-weighted RMS level of noise in reference Skype output for the 1 second silence part in middle of the speech signal for SpNR during speech → this is defined as Noise level during speech. (The sine signal components are filtered out by band stop filters prior to noise level calculation.)

Solutions without cooling fan ≥ 27

Solutions with active cooling ≥ 22

Note: Skype leaves a freedom to alter or improve the above test signal or post-processing of result without prior notice. The speech part of the signal will always remain the same.

5.2.7 Send path – single frequency interference

Purpose: Narrow-band noise, including single frequency interference, is an impairment that can be perceived as a tone, depending on its level relative to the overall weighted noise level. This can be caused by electrical noise in soundcards or by fan or hard disk drive noise on laptops. This requirement makes sure that no tonal noise is present in the send signal.

Input: Use recommended test position for DUT. Use following settings for the DUT Editor:

- Enable DUT Client Mode
- Disable AGC
- Disable AEC

The same input AGC setting as for normal speech level test is used for this test case. Skype automatic gain adjustment is disabled.

Play the [SpNR speech sample](#) at normal speech level and record the signal at reference Skype output.

Required: Calculate the A-weighted peak noise level for the 1 second silence part at end of the test signal for SpNR with an effective bandwidth of not more than 31 Hz. When using a 48 kHz sampling rate recording this means a minimum FFT size of 4096. Frequency range analyzed is 100Hz to 12000Hz. For FFT analysis the “Flat Top” windowing is employed.

The measured peak noise level is:

Solutions without cooling fan $\leq -64\text{dBm0}$ ($\leq -70\text{dBFS}$)

Solutions with active cooling $\leq -60\text{dBm0}$ ($\leq -66\text{dBFS}$)

5.2.8 Receive path - preferred loudness level in ear

Purpose: To ensure that the receive path volume adjustment has enough adjustment range to allow setting volume to preferred listening level.

Input: Use recommended test position for DUT.

Play the IEEE 269-2010 uncompressed male speech to the reference Skype input and record the acoustic signal in artificial ear.

Required: Calculate the RMS SPL level in artificial ear for the active part of male artificial speech signal (active speech does not include pauses or silences). **The level must be set to 58..62dB SPL.**

The level set in this test case will be used for other receiving direction tests that refer to preferred loudness level for playback loudness.

5.2.9 Receive path - frequency response

Purpose: The test checks that the frequency response of the DUT receiving path is flat enough to meet minimum requirement and take full advantage of Skype voice codec bandwidth.

Input: Use recommended test position for DUT.

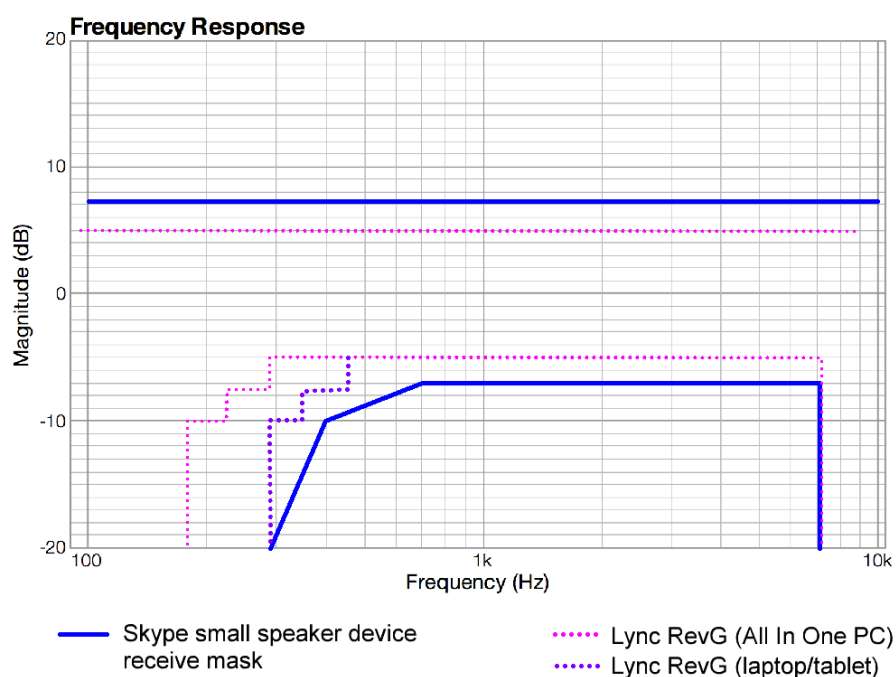
Use same receive loudness setting as set in 5.2.8.

Play the [male artificial speech for frequency response](#) to the reference Skype input.

Record the acoustic receive signal in artificial ear.

Required: The frequency response is calculated by comparing the 1/3 octave spectrum of the prior recorded reference result to the DUT receive signal in artificial ear with the free field correction applied.

The resulting frequency response graph fits into the tolerance mask below:



Frequency	Lower limit	Upper limit
99 Hz	-80,0 dB	80,0 dB
100 Hz	-80,0 dB	7,0 dB
279 Hz	-80,0 dB	7,0 dB
280 Hz	-20,0 dB	7,0 dB
400 Hz	-10,0 dB	7,0 dB
700 Hz	-7,0 dB	7,0 dB
7100 Hz	-7,0 dB	7,0 dB
7101 Hz	-80,0 dB	7,0 dB
10000 Hz	-80,0 dB	7,0 dB
10001 Hz	-80,0 dB	80,0 dB

5.2.10 Receive path - speech signal to noise ratio (SpNR).

- Purpose:** Too high self noise in receive path decreases the intelligibility of the speech and influences the total call quality in negative way. This test tests for speech to noise ratio in DUT receive path.
- Input:** Use recommended test position for DUT.
- Use same receive loudness setting as set in 5.2.8.
- Play the IEEE 269-2010 uncompressed male speech to the reference Skype input and record the acoustic signal in artificial ear.
- Required:** Calculate the RMS SPL level in artificial ear for the active part of male artificial speech (active speech does not include pauses or silences) → Speech level
- Calculate the A-weighted RMS SPL level in artificial ear for the 1 second silence part in middle of the male artificial speech signal for SpNR → Noise level.
- The calculated speech to noise ratio (SpNR ratio) is more than 35 dB**
Speech level – Noise level > 35 dB
- Note:** Please note that Skype requires speech to noise ratio (SpNR), this is not the same as signal to noise ratio (SNR) specification often found in datasheet that specifies a sine signal to noise ratio!

5.2.11 Receive path – single frequency interference

- Purpose:** Tonal noise may be perceived in the loudspeaker signal if receive single frequency interference is too high.
- Input:** Use recommended test position for DUT.
- Play the signal containing silence into the REF Skype input
- Required:** Record a 5 second sound sample from HATS ear microphone. Measure the A-weighted peak noise level over the frequency range of 50 to 20000 Hz with an effective bandwidth of not more than 31 Hz. For 48 kHz sampling rate recording this means a minimum FFT size of 4096. For FFT analysis the “Flat Top” windowing is employed.
- Compare the analyzed level to the noise level calculated in test case** Error! Reference source not found. **Receive A-weighted single frequency interference noise peak level shall be at least 10 dB quieter than the averaged receive noise level.**

5.2.12 Receive path – total harmonic distortion at preferred listening level.

Purpose: To verify that DUT loudspeaker has low enough total harmonic distortion. This is important for subjective call quality, but also for acoustic echo cancellation.

Input: Use recommended test position for DUT.

Use same receive loudness setting as set in 5.2.8.

Play the stepped sine signal with -3dBm0 RMS (-9dBFS RMS) or a 1/3rd octave band limited pink noise with rms signal level of -3dBm0 RMS (-9dBFS RMS) to the reference Skype input. Record the acoustic receive signal in artificial ear.

Required: The receive signal is analyzed for total harmonic distortion + noise

The resulting total harmonic distortion + noise must not exceed

< 10% (20dB) for frequency 315Hz.

< 5% (26dB) for frequencies 400Hz, 800Hz, 1600Hz, 2500Hz.

Note: Skype test will measure THD also at frequencies below the 315Hz and plot the result for the informational purpose. Vendor should make every attempt to keep the THD below 10% also at lower frequencies as the distortions there are equally problematic for Skype's acoustic echo cancellation.

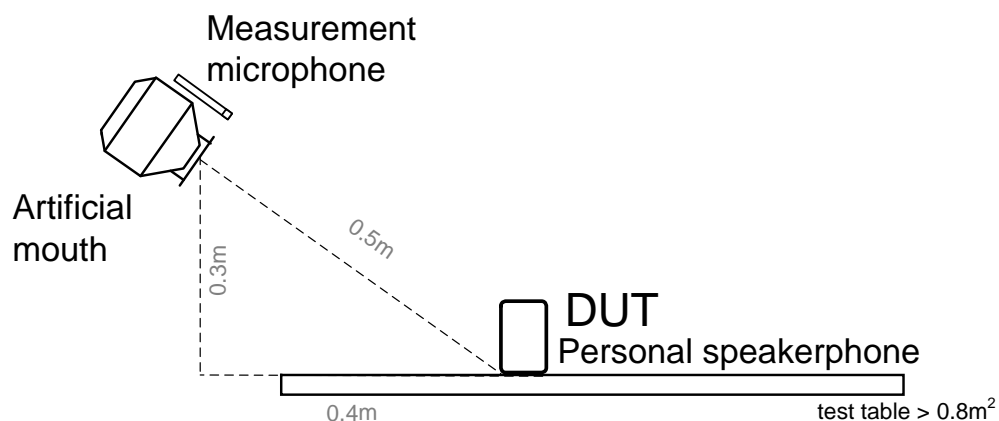
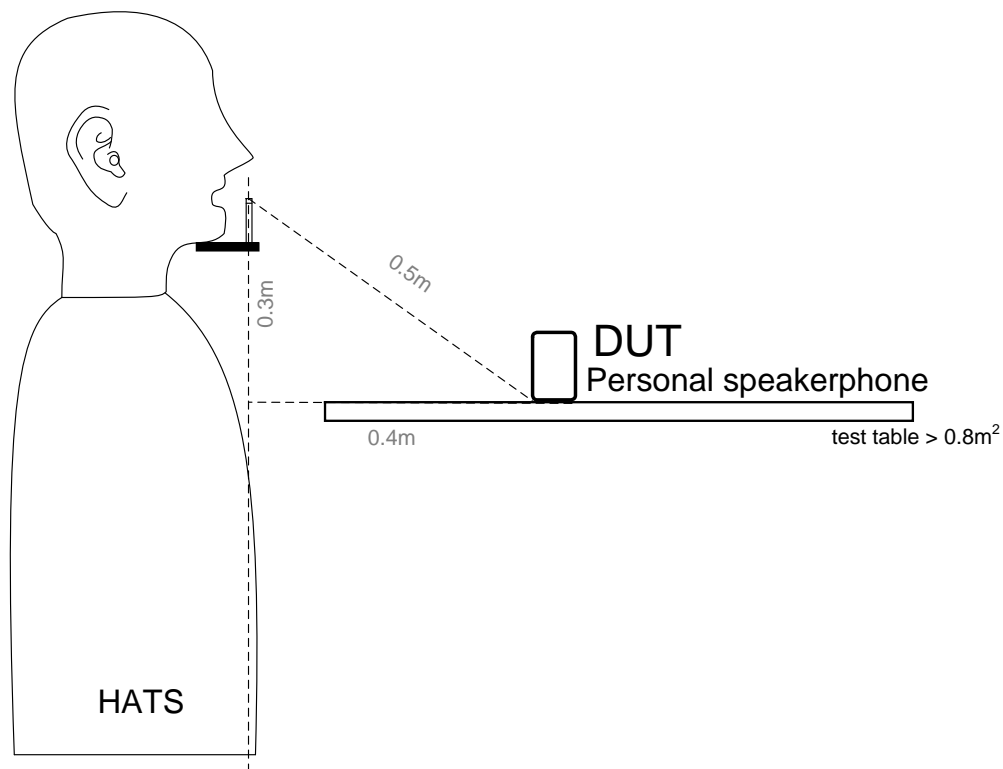
Correct high pass filtering and EQ should be used in low frequencies to avoid speaker to have a speaker cone over excursion below the speaker resonant frequency.

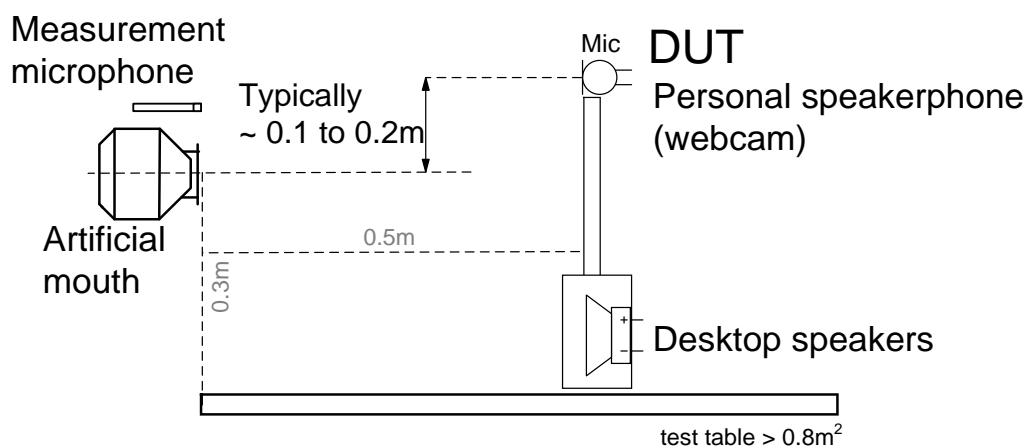
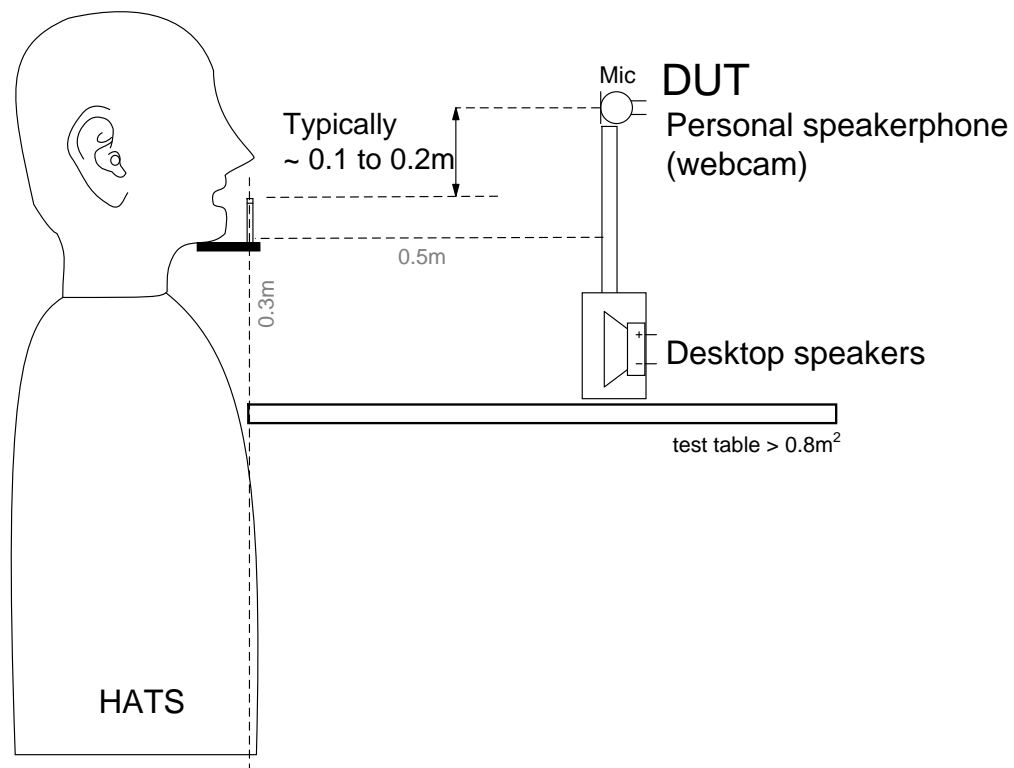
5.3 Speakerphone: DUT Usage Distance Up To 70cm (Personal Speakerphone)

Personal speakerphone recommended test position.

The recommended test position for personal speakerphone is shown below (webcams containing microphones are placed as they are intended to be used typically).

If manufacturer of device or user manual advises other usage scenario, then that is taken into account and a measurement position is agreed between Skype audio engineers and manufacturer/vendor.





5.3.1 Send path - signal level with loud speech

Purpose: To ensure that DUT send path provides optimal signal level for far end Skype client with loud speech input.

Input: Use recommended test position for DUT. Use following settings for the DUT Editor:

- Enable DUT Client Mode
- Disable AEC

Skype is allowed to automatically adjust the input gain setting for this test case.

Play the IEEE 269-2010 compressed male speech sample at loud speech level for 30 seconds to allow Skype to find optimal input gain level.

Play the IEEE 269-2010 compressed male speech sample at loud speech level and record the signal at reference Skype output.

Required: Calculate the average RMS level in reference Skype output for the active part of compressed male speech sample. **The level must be more than -24dBm0 RMS (equals -30dBFS RMS).** The peaks of the speech signals must not overload the input causing clipping.

Note: The compressed male speech is used for loud speech as it has a lower crest factor and thus better simulates a real world conditions as the crest factor of human voice is lower when people speak louder.

5.3.2 Send path - signal level with normal speech

Purpose: To ensure that DUT send path provides optimal signal level for far end Skype client with normal speech input.

Input: Use recommended test position for DUT. Use following settings for the DUT Editor:

- Enable DUT Client Mode
- Disable AEC

Skype is allowed to automatically adjust the input gain setting for this test case.

Play the IEEE 269-2010 uncompressed male speech sample at normal speech level for 30 seconds to allow Skype to find optimal input gain level.

Play the IEEE 269-2010 uncompressed male speech sample at normal speech level and record the signal at reference Skype output.

Required: Calculate the average RMS level in reference Skype output for the active part of male speech sample. **The level must be more than -24dBm0 RMS (equals -30dBFS RMS).** The peaks of the speech signals must not overload the input causing clipping.

5.3.3 Send path - signal level with quiet speech

Purpose: To ensure that DUT send path provides optimal signal level for far end Skype client with quiet speech input.

Input: Use recommended test position for DUT. Use following settings for the DUT Editor:

- Enable DUT Client Mode
- Disable AGC
- Disable AEC

The same input AGC setting as for normal speech level test is used for this test case. Skype automatic gain adjustment is disabled.

Play the IEEE 269-2010 uncompressed speech at quiet speech level and record the signal at reference Skype output.

Required: Calculate the average RMS level in reference Skype output for the active part of male speech sample. **The level must be more than -34dBm0 RMS (equals -40dBFS RMS).** The peaks of the speech signals must not overload the input causing clipping.

5.3.4 Send path - frequency response

Purpose: The test checks that the frequency response of the DUT send signal path is flat enough to meet Skype requirement and take full advantage of Skype voice codec bandwidth.

Input: Use recommended test position for DUT. Use following settings for the DUT Editor:

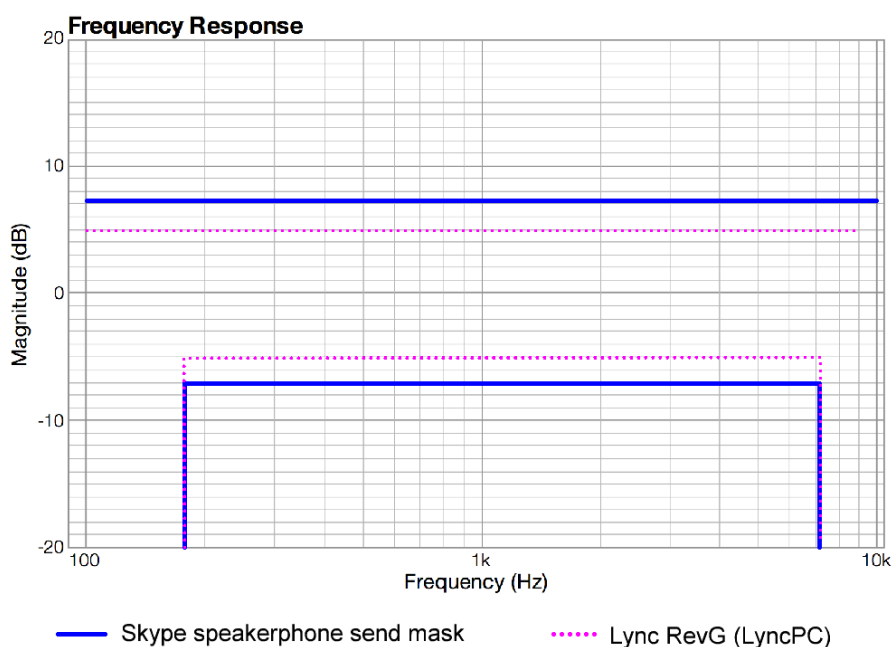
- Enable DUT Client Mode
- Disable AGC
- Disable AEC

The same input AGC setting as for normal speech level test is used for this test case. Skype automatic gain adjustment is disabled.

Play the [male artificial speech for frequency response](#) at normal speech level and record the signal at reference Skype output.

Required: The frequency response is calculated by comparing the 1/3 octave spectrum of the prior recorded reference result to the DUT receive signal in artificial ear with the free field correction applied.

The resulting frequency response graph fits into the tolerance mask below:



Frequency	Lower limit	Upper limit
99 Hz	-80,0 dB	80,0 dB
100 Hz	-80,0 dB	7,0 dB
179 Hz	-80,0 dB	7,0 dB
180 Hz	-7,0 dB	7,0 dB
7100 Hz	-7,0 dB	7,0 dB
7101 Hz	-80,0 dB	7,0 dB
10000 Hz	-80,0 dB	7,0 dB
10001 Hz	-80,0 dB	80,0 dB

5.3.5 Send path - speech signal to self noise ratio

Purpose: Too high self noise in microphone signal decreases the intelligibility of the speech and influences the total call quality in negative way. This test tests for speech to self noise ratio in DUT send path.

Input: Use recommended test position for DUT. Use following settings for the DUT Editor:

- Enable DUT Client Mode
- Disable AGC
- Disable AEC

The same input AGC setting as for normal speech level test is used for this test case. Skype automatic gain adjustment is disabled.

Play the [SpNR speech sample](#) at normal speech level and record the signal at reference Skype output.

Required: Calculate the RMS level in reference Skype output for the active speech part (the active speech does not include pauses or silences) → this is defined as Speech level

Calculate the A-weighted RMS level of noise in reference Skype output for the 1 second silence part at end of the test signal for SpNR → this is defined as Noise level.

The calculated speech level minus noise level (SpNR ratio during speech) is:

Webcams / Speakerphone ≥33

“All In One” without cooling fan ≥30

“All In One” with active cooling ≥25

Note: Please note that Skype specifies speech to noise ratio (SpNR), this is not the same as signal to noise ratio (SNR) specification often found in datasheet that specifies a sine signal to noise ratio.

5.3.6 Send path - speech signal to self noise ratio during speech

Purpose: Too high self noise in microphone signal decreases the intelligibility of the speech and influences the total call quality in negative way. Thus this test tests for speech to self noise level in DUT send path. This test tests for speech to self noise level in DUT send path during speech.

Input: Use recommended test position for DUT. Use following settings for the DUT Editor:

- Enable DUT Client Mode
- Disable AGC
- Disable AEC

The same input AGC setting as for normal speech level test is used for this test case. Skype automatic gain adjustment is disabled.

Play the [modified SpNR speech sample](#) at normal speech level and record the signal at reference Skype output.

Required: Calculate the RMS level in reference Skype output for the active speech part (the active speech does not include pauses or silences) → this is defined as Speech level.

Calculate the A-weighted RMS level of noise in reference Skype output for the 1 second silence part in middle of the speech signal for SpNR during speech → this is defined as Noise level during speech. (The sine signal components are filtered out by band stop filters prior to noise level calculation.)

The calculated speech level minus noise level (SpNR ratio during speech) is:

Webcams / Speakerphone ≥30

“All In One” without cooling fan ≥27

“All In One” with active cooling ≥22

Note: Skype leaves a freedom to alter or improve the above test signal or post-processing of result without prior notice. The speech part of the signal will always remain the same.

5.3.7 Send path – single frequency interference

Purpose: Narrow-band noise, including single frequency interference, is an impairment that can be perceived as a tone, depending on its level relative to the overall weighted noise level. This can be caused by electrical noise in soundcards or by fan or hard disk drive noise on laptops. This requirement makes sure that no tonal noise is present in the send signal.

Input: Use recommended test position for DUT. Use following settings for the DUT Editor:

- Enable DUT Client Mode
- Disable AGC
- Disable AEC

The same input AGC setting as for normal speech level test is used for this test case. Skype automatic gain adjustment is disabled.

Play the [SpNR speech sample](#) at normal speech level and record the signal at reference Skype output.

Required: Calculate the A-weighted peak noise level for the 1 second silence part at end of the test signal for SpNR with an effective bandwidth of not more than 31 Hz. When using a 48 kHz sampling rate recording this means a minimum FFT size of 4096. Frequency range analyzed is 100Hz to 12000Hz. For FFT analysis the “Flat Top” windowing is employed.

The measured peak noise level is:

Webcams / Speakerphone $\leq -70\text{dBm0}$ ($\leq -76\text{dBFS}$)

“All In One” without cooling fan $\leq -64\text{dBm0}$ ($\leq -70\text{dBFS}$)

“All In One” with active cooling $\leq -60\text{dBm0}$ ($\leq -66\text{dBFS}$)

5.3.8 Receive path - preferred loudness level in ear

Purpose: To ensure that the receive path volume adjustment has enough adjustment range to allow setting volume to preferred listening level.

Input: Use recommended test position for DUT.

Play the IEEE 269-2010 uncompressed male speech to the reference Skype input and record the acoustic signal in artificial ear.

Required: Calculate the RMS SPL level in artificial ear for the active part of male artificial speech signal (active speech does not include pauses or silences).

The level must be set to:

Webcams (for the reference speaker) / Speakerphone: 63..67dB SPL

“All In One” PC: 68..72dB SPL

The level set in this test case will be used for other receiving direction tests that refer to preferred loudness level for playback loudness.

5.3.9 Receive path - frequency response

Purpose: The test checks that the frequency response of the DUT receiving path is flat enough to meet minimum requirement and take full advantage of Skype voice codec bandwidth.

Input: Use recommended test position for DUT.

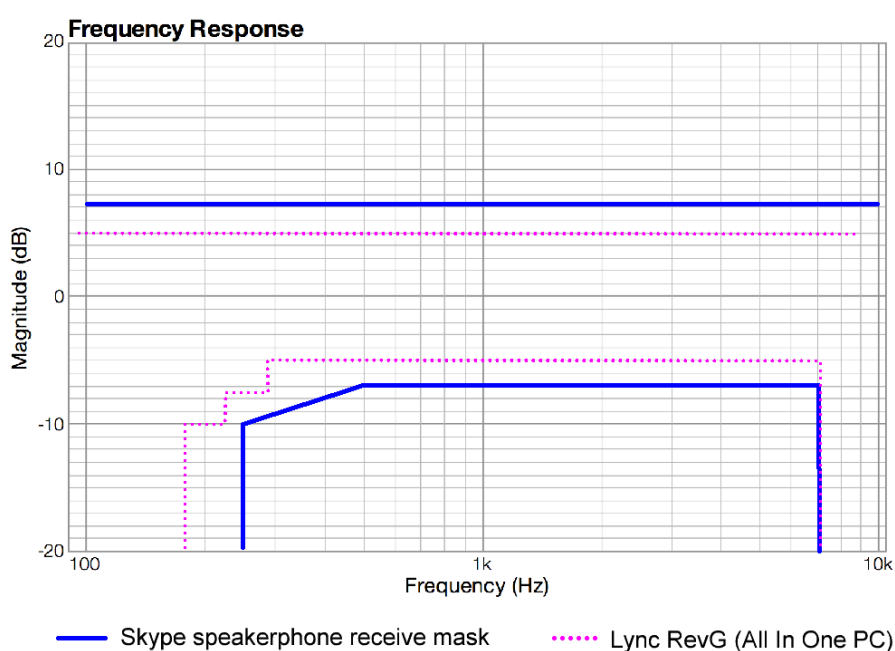
Use same receive loudness setting as set in 5.3.8

Play the [male artificial speech for frequency response](#) to the reference Skype input.

Record the acoustic receive signal in artificial ear.

Required: The frequency response is calculated by comparing the 1/3 octave spectrum of the prior recorded reference result to the DUT receive signal in artificial ear with the free field correction applied.

The resulting frequency response graph fits into the tolerance mask below:



Frequency	Lower limit	Upper limit
99 Hz	-80,0 dB	80,0 dB
100 Hz	-80,0 dB	7,0 dB
249 Hz	-80,0 dB	7,0 dB
250 Hz	-10,0 dB	7,0 dB
500 Hz	-7,0 dB	7,0 dB
7100 Hz	-7,0 dB	7,0 dB
7101 Hz	-80,0 dB	7,0 dB
10000 Hz	-80,0 dB	7,0 dB
10001 Hz	-80,0 dB	80,0 dB

NB! The tolerance mask is floating (not fixed) and will move up/down trying to center the measured response between tolerance limits.

5.3.10 Receive path - speech signal to noise ratio (SpNR).

- Purpose:** Too high self noise in receive path decreases the intelligibility of the speech and influences the total call quality in negative way. This test tests for speech to noise ratio in DUT receive path.
- Input:** Use recommended test position for DUT.
- Use same receive loudness setting as set in 5.3.8
- Play the IEEE 269-2010 uncompressed male speech to the reference Skype input and record the acoustic signal in artificial ear.
- Required:** Calculate the RMS SPL level in artificial ear for the active part of male artificial speech (active speech does not include pauses or silences) → Speech level
- Calculate the A-weighted RMS SPL level in artificial ear for the 1 second silence part in middle of the male artificial speech signal for SpNR → Noise level.
- The calculated speech to noise ratio (SpNR ratio) is more than 40dB.**
Speech level – Noise level > 40dB
- Note:** Please note that Skype requires speech to noise ratio (SpNR), this is not the same as signal to noise ratio (SNR) specification often found in datasheet that specifies a sine signal to noise ratio!

5.3.11 Receive path – single frequency interference

- Purpose:** Tonal noise may be perceived in the loudspeaker signal if receive single frequency interference is too high.
- Input:** Use recommended test position for DUT.
- Play the signal containing silence into the REF Skype input
- Required:** Record a 5 second sound sample from HATS ear microphone. Measure the A-weighted peak noise level over the frequency range of 50 to 20000 Hz with an effective bandwidth of not more than 31 Hz. For 48 kHz sampling rate recording this means a minimum FFT size of 4096. For FFT analysis the “Flat Top” windowing is employed.
- Compare the analyzed level to the noise level calculated in test case** Error! Reference source not found. **Receive A-weighted single frequency interference noise peak level shall be at least 10 dB quieter than the averaged receive noise level.**

5.3.12 Receive path – total harmonic distortion at preferred listening level.

Purpose: To verify that DUT loudspeaker has low enough total harmonic distortion. This is important for subjective call quality, but also for acoustic echo cancellation.

Input: Use recommended test position for DUT.

Use same receive loudness setting as set in 5.3.8.

Play the stepped sine signal with -3dBm0 RMS (-9dBFS RMS) or a 1/3rd octave band limited pink noise with rms signal level of -3dBm0 RMS (-9dBFS RMS) to the reference Skype input. Record the acoustic receive signal in artificial ear.

Required: The receive signal is analyzed for total harmonic distortion + noise

The resulting total harmonic distortion + noise must not exceed

< 6.3% (24dB) for frequency 315Hz.

< 5% (26dB) for frequencies 400Hz, 800Hz, 1600Hz, 2500Hz.

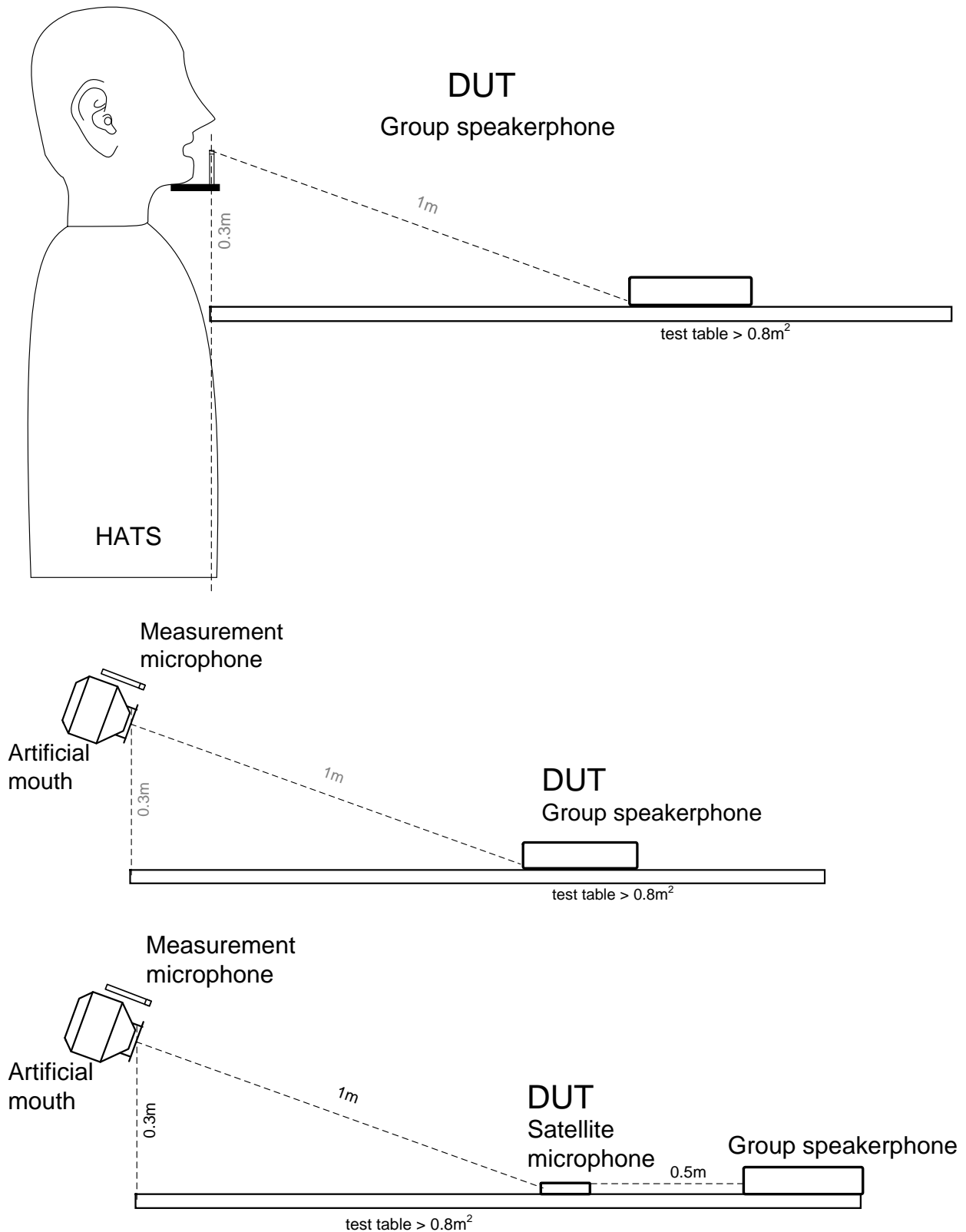
Note: Skype test will measure THD also at frequencies below the 315Hz and plot the result for the informational purpose. Vendor should make every attempt to keep the THD below 10% also at lower frequencies as the distortions there are equally problematic for Skype's acoustic echo cancellation.

Correct high pass filtering and EQ should be used at low frequencies to avoid speaker to have a speaker cone over excursion below the speaker resonant frequency.

5.4 Speakerphone: DUT Usage Distance Up To 150cm (Group Speakerphone)

Group speakerphone test position.

The recommended test position for group speakerphone is shown below. The group speakerphone is measured with microphone(s) at 1m distance from HATS MRP.



In anechoic room the testing the 1meter distance is used for practical reasons. This means that during anechoic room testing with the “**1.5 meter speech level**” the DUT stays physically at 1m distance, but **the speech level in HATS mouth is decreased by 3dB** respectively for each normal, quiet and loud speech compared to normal 1meter levels. 3dB represents the estimated average speech level drop in microphone when near end talker is moving from 1m to 1.5m distance in average meeting rooms.

Speech levels at HATS MRP (mouth reference point) for Group Speakerphones

	1 meter level test case	1.5 meter speech level
Normal speech level	89 dB SPL	86 dB SPL (test case 5.4.4)
Loud speech level	99 dB SPL	96 dB SPL (test case 5.4.3)
Quiet speech level	79 dB SPL	76 dB SPL (test case 5.4.2)

Note! The nominal difference between the MRP (Mouth Reference Point) and HFRP (Hands-Free Reference Point at 0.5 meter distance) is 24 dB SPL. ACQUA calibrates the level of the artificial mouth at the MRP (Mouth equalization is active during calibrations).

If manufacturer/vendor of device advises other measurement scenario, then that is taken into account and a measurement positions are agreed between Skype audio engineers and manufacturer/vendor.

In case a solution provides a partial acoustic audio interface (input only) and is intended to work with a TV then we complete the acoustic UI and running the tests with a TV. This is necessary to understand the echo path performance and depending on the intended usage position the TV panel might affect the send path.

A 32” or 42” TV from the product lineup of Samsung, LG, Sony and Vizio is used as a playback / display device. Please refer to section 7.1.13 for the exact models of TV-s used. The TV in-built loudspeakers are used.

5.4.1 Send path - signal level with loud speech

Purpose: To ensure that DUT send path provides optimal signal level for far end Skype client with loud speech level at maximum recommended usage distance from DUT.

Input: Use the maximum recommended usage distance test position for DUT. Use following settings for the DUT Editor:

- Enable DUT Client Mode
- Disable AEC

Skype is allowed to automatically adjust the input gain setting for this test case.

Play the IEEE 269-2010 compressed male speech sample at loud speech level for 30 seconds to allow Skype to find optimal input gain level.

Play the IEEE 269-2010 compressed male speech sample at loud speech level and record the signal at reference Skype output.

Required: Calculate the average RMS level in reference Skype output for the active part of male speech sample. **The level must be more than -24dBm0 RMS (equals -30dBFS RMS).**

The peaks of the speech signals must not overload the input causing clipping.

The crest factor of analyzed signal must not be compressed more than 5dB compared to source signal!

Note: The compressed male speech is used for loud speech as it has a lower crest factor and thus better simulates a real world conditions as the crest factor of human voice is lower when people speak louder.

5.4.2 Send path - signal level with normal speech (1,5 m speech level)

Purpose: To ensure that DUT send path provides optimal signal level for far end Skype client with normal speech input at maximum recommended usage distance from DUT.

Input: Use the maximum recommended usage distance test position for DUT. Use following settings for the DUT Editor:

- Enable DUT Client Mode
- Disable AEC

Skype is allowed to automatically adjust the input gain setting for this test case.

Play the IEEE 269-2010 uncompressed male speech sample at 1.5m simulated normal speech level for 30 seconds to allow Skype to find optimal input gain level.

Play the IEEE 269-2010 uncompressed male speech sample at 1.5m simulated normal speech level and record the signal at reference Skype output.

Required: Calculate the average RMS level in reference Skype output for the active part of male speech sample. **The level must be more than -24dBm0 RMS (equals -30dBFS RMS).** The peaks of the speech signals must not overload the input causing clipping.

5.4.3 Send path - signal level with quiet speech (1,5 m speech level)

Purpose: To ensure that DUT send path provides optimal signal level for far end Skype client with quiet speech level at maximum recommended usage distance from DUT.

Input: Use the maximum recommended usage distance test position for DUT. Use following settings for the DUT Editor:

- Enable DUT Client Mode
- Disable AGC
- Disable AEC

The same input AGC setting as for normal speech level test is used for this test case. Skype automatic gain adjustment is disabled.

Play the IEEE 269-2010 uncompressed speech at 1.5 meter simulated quiet speech level and record the signal at reference Skype output.

Required: Calculate the average RMS level in reference Skype output for the active part of male speech sample. **The level must be more than -34dBm0 RMS (equals -40dBFS RMS).** The peaks of the speech signals must not overload the input causing clipping.

5.4.4 Send path - speech signal to self noise ratio (1,5 m speech level)

Purpose: Too high self noise in microphone signal decreases the intelligibility of the speech and influences the total call quality in negative way. This test tests for speech to self noise ratio in DUT send path.

Input: Use the maximum recommended usage distance test position for DUT. Use following settings for the DUT Editor:

- Enable DUT Client Mode
- Disable AGC
- Disable AEC

The same input AGC setting as for normal speech level test is used for this test case. Skype automatic gain adjustment is disabled.

Play the [SpNR speech sample](#) at 1,5m simulated normal speech level and record the signal at reference Skype output.

Required: Calculate the RMS level in reference Skype output for the active speech part (the active speech does not include pauses or silences) → this is defined as Speech level

Calculate the A-weighted RMS level of noise in reference Skype output for the 1 second silence part at end of the test signal for SpNR → this is defined as Noise level.

The calculated speech to noise ratio (SpNR ratio) is more than 25dB.
(Speech level – Noise level > 25 dB)

Note: Please note that Skype specifies speech to noise ratio (SpNR), this is not the same as signal to noise ratio (SNR) specification often found in datasheet that specifies a sine signal to noise ratio.

5.4.5 Send path - speech signal to self noise ratio during speech (1,5 m speech level)

Purpose: Too high self noise in microphone signal decreases the intelligibility of the speech and influences the total call quality in negative way. Thus this test tests for speech to self noise level in DUT send path. This test tests for speech to self noise level in DUT send path during speech.

Input: Use recommended test position for DUT. Use following settings for the DUT Editor:

- Enable DUT Client Mode
- Disable AGC
- Disable AEC

The same input AGC setting as for normal speech level test is used for this test case. Skype automatic gain adjustment is disabled.

Play the [modified SpNR speech sample](#) at normal speech level and record the signal at reference Skype output.

Required: Calculate the RMS level in reference Skype output for the active speech part (the active speech does not include pauses or silences) → this is defined as Speech level.

Calculate the A-weighted RMS level of noise in reference Skype output for the 1 second silence part in middle of the speech signal for SpNR during speech → this is defined as Noise level during speech. (The sine signal components are filtered out by band stop filters prior to noise level calculation.)

The calculated speech level minus noise level (SpNR ratio) is more than 20dB.
(Speech level – Noise level during speech > 20 dB)

Note: Skype leaves a freedom to alter or improve the above test signal or post-processing of result without prior notice. The speech part of the signal will always remain the same.

5.4.6 Send path - frequency response

Purpose: The test checks that the frequency response of the DUT send signal path is flat enough to meet Skype requirement and take full advantage of Skype voice codec bandwidth.

Input: Use recommended test position for DUT. Use following settings for the DUT Editor:

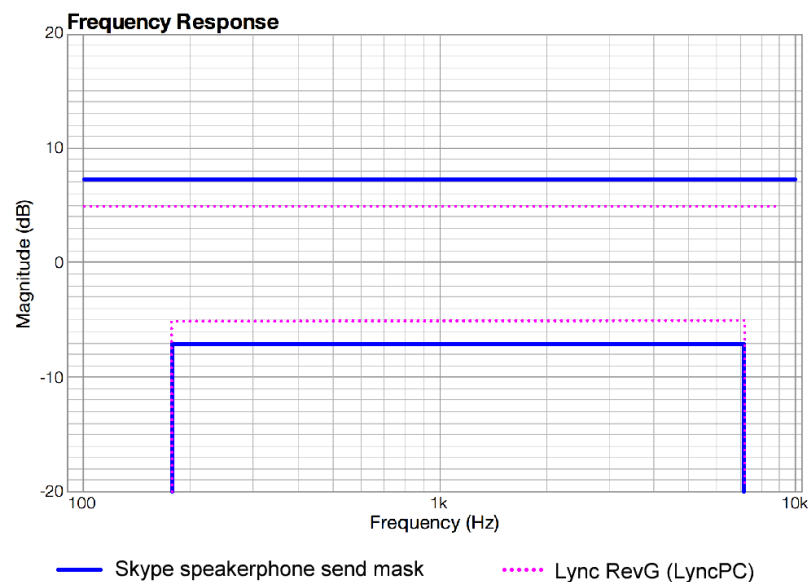
- Enable DUT Client Mode
- Disable AGC
- Disable AEC

The same input AGC setting as for normal speech level test is used for this test case. Skype automatic gain adjustment is disabled.

Play the [male artificial speech for frequency response](#) at normal speech level and record the signal at reference Skype output.

Required: The frequency response is calculated by comparing the 1/3 octave spectrum of the prior recorded reference result at MRP to the DUT send signal recorded at reference Skype output.

The resulting frequency response graph fits into the tolerance mask below:



Frequency	Lower limit	Upper limit
99 Hz	-80,0 dB	80,0 dB
100 Hz	-80,0 dB	7,0 dB
179 Hz	-80,0 dB	7,0 dB
180 Hz	-7,0 dB	7,0 dB
7100 Hz	-7,0 dB	7,0 dB
7101 Hz	-80,0 dB	7,0 dB
10000 Hz	-80,0 dB	7,0 dB
10001 Hz	-80,0 dB	80,0 dB

5.4.7 Send path – single frequency interference (1,5 m speech level)

Purpose: Narrow-band noise, including single frequency interference, is an impairment that can be perceived as a tone, depending on its level relative to the overall weighted noise level. This can be caused by electrical noise in soundcards or by fan or hard disk drive noise on laptops. This requirement makes sure that no tonal noise is present in the send signal.

Input: Use recommended test position for DUT. Use following settings for the DUT Editor:

- Enable DUT Client Mode
- Disable AGC
- Disable AEC

The same input AGC setting as for normal speech level test is used for this test case. Skype automatic gain adjustment is disabled.

Play the [SpNR speech sample](#) at 1,5m simulated normal speech level and record the signal at reference Skype output.

Required: Calculate the A-weighted peak noise level for the 1 second silence part at end of the test signal for SpNR with an effective bandwidth of not more than 31 Hz. When using a 48 kHz sampling rate recording this means a minimum FFT size of 4096. Frequency range analyzed is 100Hz to 12000Hz. For FFT analysis the “Flat Top” windowing is employed.

The measured peak noise level is: $\leq -70\text{dBm0}$ ($\leq -76\text{dBFS}$).

5.4.8 Receive path - preferred loudness level in ear

Purpose: To ensure that the receive path volume adjustment has enough adjustment range to allow setting volume to preferred listening level.

Input: Use recommended test position for DUT.

Play the IEEE 269-2010 uncompressed male speech to the reference Skype input and record the acoustic signal in artificial ear.

Required: Calculate the RMS SPL level in artificial ear for the active part of male speech sample (active speech does not include pauses or silences). **The level must be set to 68 dB ..72 dB SPL.**

The level set in this test case will be used for other receiving direction tests that refer to preferred loudness level for playback loudness.

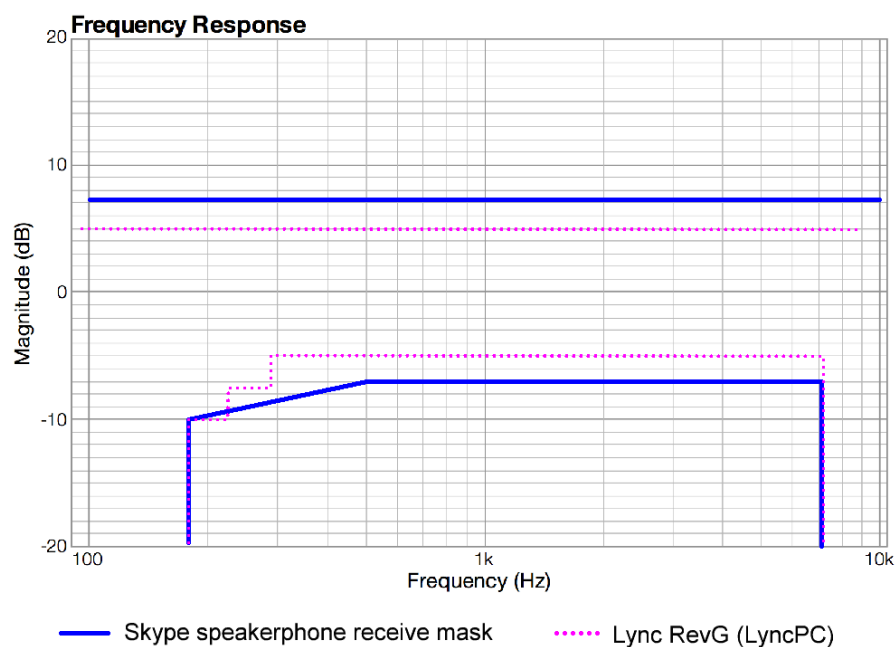
5.4.9 Receive path - frequency response

Purpose: The test checks that the frequency response of the DUT receiving path is flat enough to meet minimum requirement and take full advantage of Skype voice codec bandwidth.

Input: Use recommended test position for DUT.
Use same receive loudness setting as set in 5.4.8
Play the [male artificial speech for frequency response](#) to the reference Skype input.
Record the acoustic receive signal in artificial ear.

Required: The frequency response is calculated by comparing the 1/3 octave spectrum of the prior recorded reference result to the DUT receive signal in artificial ear with the free field correction applied.

The resulting frequency response graph fits into the tolerance mask below:



Frequency	Lower limit	Upper limit
99 Hz	-80,0 dB	80,0 dB
100 Hz	-80,0 dB	7,0 dB
179 Hz	-80,0 dB	7,0 dB
180 Hz	-10,0 dB	7,0 dB
500 Hz	-7,0 dB	7,0 dB
700 Hz	-7,0 dB	7,0 dB
7100 Hz	-7,0 dB	7,0 dB
7101 Hz	-80,0 dB	7,0 dB
10000 Hz	-80,0 dB	7,0 dB
10001 Hz	-80,0 dB	80,0 dB

NB! The tolerance mask is floating (not fixed) and will move up/down trying to center the measured response between tolerance limits.

5.4.10 Receive path - speech signal to noise ratio (SpNR)

- Purpose:** Too high self noise in receive path decreases the intelligibility of the speech and influences the total call quality in negative way. This test tests for speech to noise ratio in DUT receive path.
- Input:** Use recommended test position for DUT.
- Use same receive loudness setting as set in 5.4.8
- Play the IEEE 269-2010 uncompressed male speech to the reference Skype input and record the acoustic signal in artificial ear.
- Required:** Calculate the RMS SPL level in artificial ear for the active part of male artificial speech (active speech does not include pauses or silences) → Speech level.
- Calculate the A-weighted RMS SPL level in artificial ear for the 1 second silence part in middle of the male artificial speech signal for SpNR → Noise level.
- The calculated speech to noise ratio (SpNR ratio) is more than 40 dB**
Speech level – Noise level > 40 dB
- Note:** Please note that Skype requires speech to noise ratio (SpNR), this is not the same as signal to noise ratio (SNR) specification often found in datasheet that specifies a sine signal to noise ratio.

5.4.11 Receive path – single frequency interference

- Purpose:** Tonal noise may be perceived in the loudspeaker signal if receive single frequency interference is too high.
- Input:** Use recommended test position for DUT.
- Play the signal containing silence into the REF Skype input.
- Required:** Record a 5 second sound sample from HATS ear microphone. Measure the A-weighted peak noise level over the frequency range of 50 to 20000 Hz with an effective bandwidth of not more than 31 Hz. For 48 kHz sampling rate recording this means a minimum FFT size of 4096. For FFT analysis the “Flat Top” windowing is employed.
- Compare the analyzed level to the noise level calculated in test case** Error! Reference source not found. **Receive A-weighted single frequency interference noise peak level shall be at least 10 dB quieter than the averaged receive noise level.**

5.4.12 Receive path – total harmonic distortion at preferred listening level

Purpose: To verify that DUT loudspeaker has low enough total harmonic distortion. This is important for subjective call quality, but also for acoustic echo cancellation.

Input: Use recommended test position for DUT.

Use same receive loudness setting as set in 5.4.8.

Play the stepped sine signal with -3dBm0 RMS (-9dBFS RMS) or a 1/3rd octave band limited pink noise with rms signal level of -3dBm0 RMS (-9dBFS RMS) to the reference Skype input. Record the acoustic receive signal in artificial ear.

Required: The receive signal is analyzed for total harmonic distortion + noise

The resulting total harmonic distortion + noise must not exceed

< 6.3% (24dB) for frequency 315Hz.

< 5% (26dB) for frequencies 400Hz, 800Hz, 1600Hz, 2500Hz.

Note: Skype test will measure THD also at frequencies below the 315Hz and plot the result for the informational purpose. Vendor should make every attempt to keep the THD below 10% also at lower frequencies as the distortions there are equally problematic for Skype's acoustic echo cancellation.

Correct high pass filtering and EQ should be used at low frequencies to avoid speaker to have a speaker cone over excursion below the speaker resonant frequency.

5.5 Speakerphone: DUT usage Distance Up To 5m (Long Range Speakerphone)

Long range speakerphone test position in anechoic room.

The recommended test position for long range speakerphone is shown below. The long range speakerphone is measured with microphones at 1m vertical plane from artificial mouth MRP.

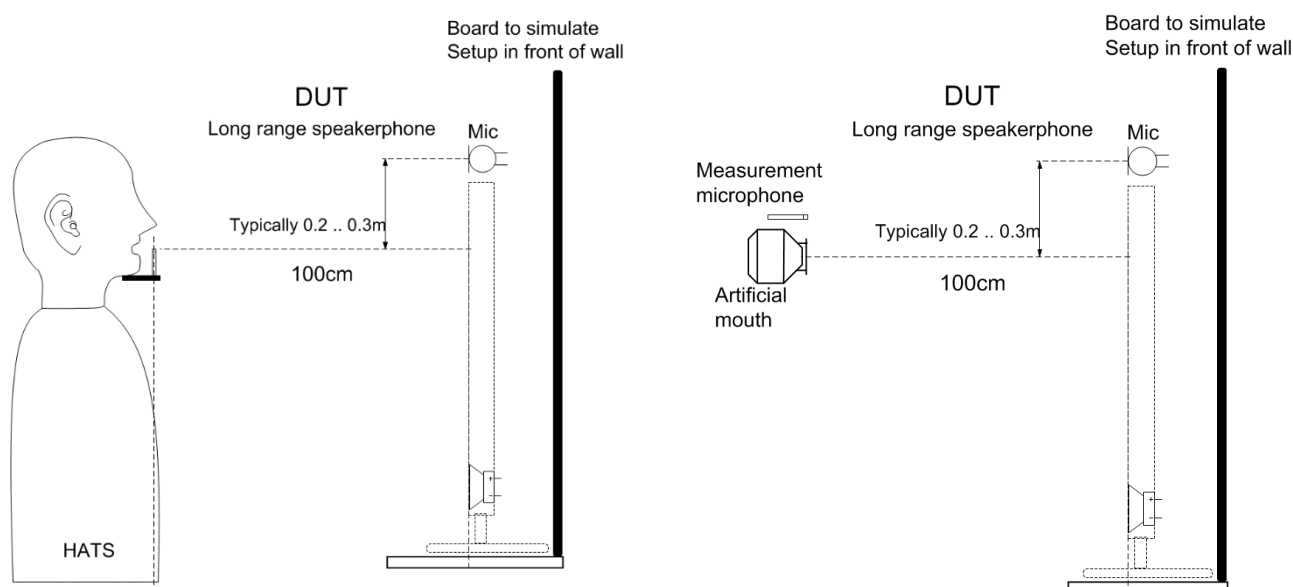
In anechoic room the testing for maximum distance send path test cases is done using a simulated 4m test condition. This means that during anechoic room testing with the “**4 meter speech level**” the DUT stays physically at 1m distance. The speech level in HATS mouth is decreased by 8dB respectively for each normal, quiet and loud speech compared to normal 1meter levels. **8dB represents the estimated average speech level drop in microphone when near end talker is moving from 1m to 4m distance in average living rooms.**

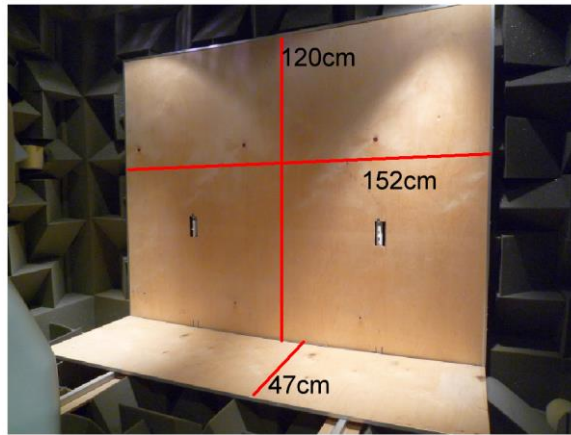
Speech levels at HATS MRP (mouth reference point) for Long Range Speakerphones

	1 meter level test case	4 meter level test case
Normal speech level	89 dB SPL (test case 5.5.2)	81 dB SPL (test case 5.5.3)
Loud speech level	99 dB SPL (test case 0)	91 dB SPL
Quiet speech level	79 dB SPL	71 dB SPL (test case 5.5.4)

Note! The nominal difference between the MRP (Mouth Reference Point) and HFRP (Hands-Free Reference Point at 0.5 meter distance) is 24 dB SPL. ACQUA calibrates the level of the artificial mouth at the MRP (Mouth equalization is active during calibrations).

If manufacturer/vendor of device advises other measurement scenario, then that is taken into account and a measurement positions are agreed between Skype audio engineers and manufacturer/vendor.

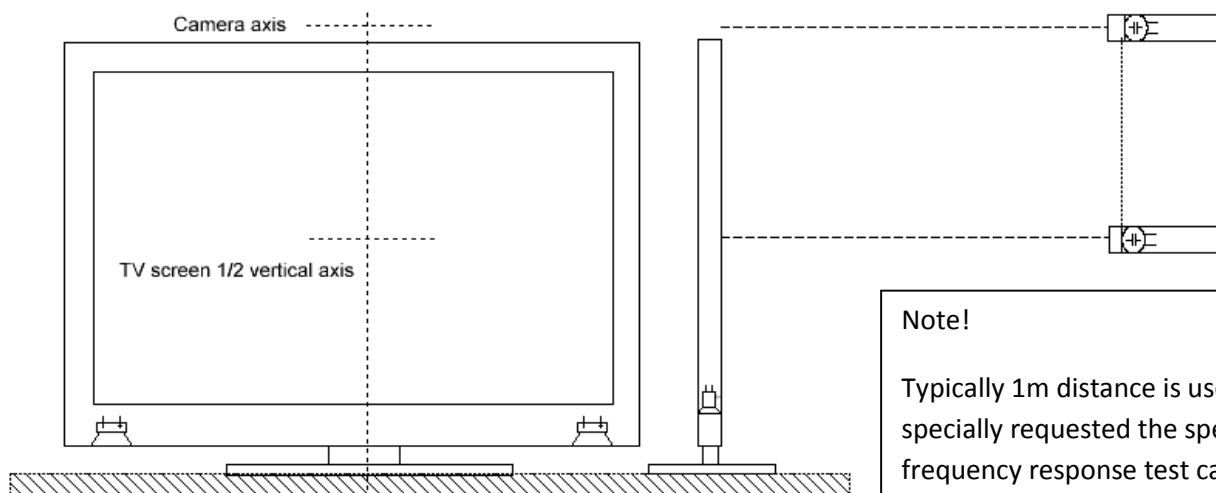




The sample of physical back-wall setup in Skype Tallinn anechoic room. Material is 8mm thick plywood. Same or thicker MDF board is equally ok to use.

Long range speakerphone test position if DUT is a TV product.

If the vendor requests then the receive path loudspeaker tests can be conducted by removing the HATS and replacing the artificial ear with the free field microphone. The microphone will be positioned according to below instructions.



As a default measurement setup in Skype AudioLab testing a table is used and TV is placed on table.

Note!

Typically 1m distance is used, when specially requested the speaker frequency response test can be also conducted at 2m or 2.5m for the TV-s with screen size of >46inch.

Long range speakerphone test position if DUT is a Set Top Box or similar device that connects to TV in order to enable Skype calls as a product falling in Long Range Speakerphone UI.

In case a solution provides a partial acoustic audio interface (input only) and is intended to work with a TV then we complete the acoustic UI and running the tests with a TV. This is necessary to understand the echo path performance and depending on the intended usage position the TV panel might affect the send path.

- A 32" or 42" TV from the product lineup of Samsung, LG, Sony and Vizio is used as a playback / display device. **Please refer to section 7.1.13 for the exact models of TV-s used.** The TV in built loudspeakers are used.
- Whenever available, the HDMI output is used as a default for both the receive audio and for UI and remote video playback. If only analog output is available for audio, the analog signal is used instead.
- Camera is placed on top of the TV bezel in middle, unless user instructions are provided either on paper or as UI guidelines and the instructions give clear guidelines how to position the camera / microphones.
- TV playback audio is adjusted to "preferred listening level" unless the device provides active UI notifications during Skype call that instruct the user to lower the playback volume to avoid acoustic echo leaking. If such message instruct using a lower level than the specified lower limits in test case 5.5.8 then the lowest limit is used.

5.5.1 Send path - signal level with loud speech (1m speech level)

Purpose: To ensure that DUT send path provides optimal signal level for far end Skype client with loud speech level when user is 100cm from DUT.

Input: Use the 100cm test position for DUT. Use recommended test position for DUT. Use following settings for the DUT Editor:

- Enable DUT Client Mode
- Disable AEC

Skype is allowed to automatically adjust the input gain setting for this test case.

Play the IEEE 269-2010 compressed male speech sample at loud speech level for 30 seconds to allow Skype to find optimal input gain level.

Play the IEEE 269-2010 compressed male speech sample at loud speech level and record the signal at reference Skype output.

Required: Calculate the average RMS level in reference Skype output for the active part of compressed male speech sample. **The level must be more than -24dBm0 RMS (equals -30dBFS RMS).**

The peaks of the speech signals must not overload the input causing clipping.

The crest factor of analyzed signal must not be compressed more than 5dB compared to source signal!

Note: The compressed male speech is used for loud speech as it has a lower crest factor and thus better simulates a real world conditions as the crest factor of human voice is lower when people speak louder.

5.5.2 Send path - signal level with normal speech (1m speech level)

Purpose: To ensure that DUT send path provides optimal signal level for far end Skype client with normal speech input when user is 100cm from DUT.

Input: Use the 100cm test position for DUT. Use recommended test position for DUT. Use following settings for the DUT Editor:

- Enable DUT Client Mode
- Disable AEC

Skype is allowed to automatically adjust the input gain setting for this test case.

Play the IEEE 269-2010 uncompressed male speech sample at normal speech level for 30 seconds to allow Skype to find optimal input gain level.

Play the IEEE 269-2010 uncompressed male speech sample at normal speech level and record the signal at reference Skype output.

Required: Calculate the average RMS level in reference Skype output for the active part of male speech sample. **The level must be more than -24dBm0 RMS (equals -30dBFS RMS).** The peaks of the speech signals must not overload the input causing clipping.

5.5.3 Send path - signal level with normal speech (4m speech level)

Purpose: To ensure that DUT send path provides optimal signal level for far end Skype client with normal speech input at maximum recommended usage distance from DUT.

Input: Use the maximum recommended usage distance test position for DUT. Use recommended test position for DUT. Use following settings for the DUT Editor:

- Enable DUT Client Mode
- Disable AEC

Skype is allowed to automatically adjust the input gain setting for this test case.

Play the IEEE 269-2010 uncompressed male speech sample at 4m simulated normal speech level for 30 seconds to allow Skype to find optimal input gain level.

Play the IEEE 269-2010 uncompressed male speech sample at 4m simulated normal speech level and record the signal at reference Skype output.

Required: Calculate the average RMS level in reference Skype output for the active part of male speech sample. **The level must be more than -24dBm0 RMS (equals -30dBFS RMS).** The peaks of the speech signals must not overload the input causing clipping.

5.5.4 Send path - signal level with quiet speech (4m speech level)

Purpose: To ensure that DUT send path provides optimal signal level for far end Skype client with quiet speech level at maximum recommended usage distance from DUT.

Input: Use the maximum recommended usage distance test position for DUT. Use following settings for the DUT Editor:

- Enable DUT Client Mode
- Disable AGC
- Disable AEC

The same input AGC setting as for normal speech level test is used for this test case. Skype automatic gain adjustment is disabled.

Play the IEEE 269-2010 uncompressed speech at 4m simulated quiet speech level and record the signal at reference Skype output.

Required: Calculate the average RMS level in reference Skype output for the active part of male speech sample. **The level must be more than -34dBm0 RMS (equals -40dBFS RMS).**

5.5.5 Send path - speech signal to self noise ratio (4m speech level)

Purpose: Too high self noise in microphone signal decreases the intelligibility of the speech and influences the total call quality in negative way. Thus this test tests for speech to self noise level in DUT send path.

Input: Use recommended test position for DUT. Use following settings for the DUT Editor:

- Enable DUT Client Mode
- Disable AGC
- Disable AEC

The same input AGC setting as for normal speech level test is used for this test case. Skype automatic gain adjustment is disabled.

Play the [SpNR speech sample](#) at 4m simulated normal speech level and record the signal at reference Skype output.

Required: Calculate the RMS level in reference Skype output for the active speech part (the active speech does not include pauses or silences) → this is defined as Speech level

Calculate the A-weighted RMS level of noise in reference Skype output for the 1 second silence part at end of the test signal for SpNR → this is defined as Noise level.

The calculated speech to noise ratio (SpNR ratio) is more than 25 dB.

(Speech level – Noise level > 25 dB)

Note: Please note that Skype specifies speech to noise ratio (SpNR), this is not the same as signal to noise ratio (SNR) specification often found in datasheet that specifies a sine signal to noise ratio.

5.5.6 Send path - speech signal to self noise ratio during speech (4m speech level)

Purpose: To ensure that the DUT internally generated self noise of microphone capsule, microphone amplifiers and A to D conversion is below the required limit also when active noise gating is used in microphone signal path. Too high self noise in microphone signal during speech decreases the intelligibility of the speech and influences the total call quality in negative way.

Input: Use recommended test position for DUT. Use following settings for the DUT Editor:

- Enable DUT Client Mode
- Disable AGC
- Disable AEC

The same input AGC setting as for normal speech level test is used for this test case. Skype automatic gain adjustment is disabled.

Play the [modified SpNR speech sample](#) at normal speech level and record the signal at reference Skype output.

Required: Calculate the RMS level in reference Skype output for the active speech part (the active speech does not include pauses or silences) → this is defined as Speech level.

Calculate the A-weighted RMS level of noise in reference Skype output for the 1 second silence part in middle of the speech signal for SpNR during speech → this is defined as Noise level during speech. (The sine signal components are filtered out by band stop filters prior to noise level calculation.)

The calculated speech level minus noise level (SpNR ratio) is more than 20 dB.

(Speech level – Noise level during speech > 20 dB)

Note: Skype leaves a freedom to alter or improve the above test signal or post-processing of result without prior notice. The speech part of the signal will always remain the same.

5.5.7 Send path - frequency response

Purpose: The test checks that the frequency response of the DUT send signal path is flat enough to meet Skype requirement and take full advantage of Skype voice codec bandwidth.

Input: Use recommended test position for DUT. Use following settings for the DUT Editor:

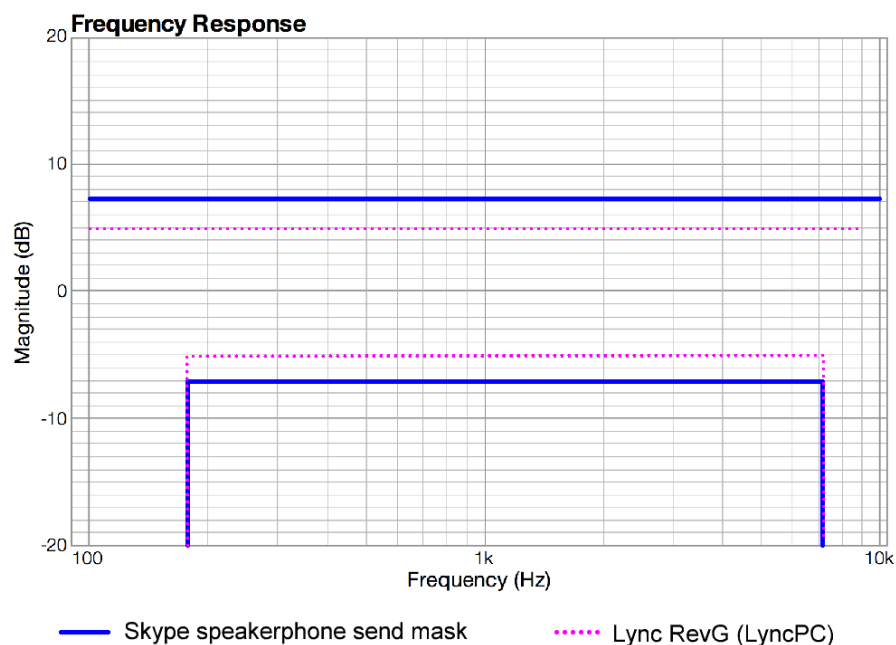
- Enable DUT Client Mode
- Disable AGC
- Disable AEC

The same input AGC setting as for normal speech level test is used for this test case. Skype automatic gain adjustment is disabled.

Play the [male artificial speech for frequency response](#) at normal speech level and record the signal at reference Skype output.

Required: The frequency response is calculated by comparing the 1/3 octave spectrum of the prior recorded reference result at MRP to the DUT send signal recorded at reference Skype output.

The resulting frequency response graph fits into the tolerance mask below:



Frequency	Lower limit	Upper limit
99 Hz	-80,0 dB	80,0 dB
100 Hz	-80,0 dB	7,0 dB
179 Hz	-80,0 dB	7,0 dB
180 Hz	-7,0 dB	7,0 dB
7100 Hz	-7,0 dB	7,0 dB
7101 Hz	-80,0 dB	7,0 dB
10000 Hz	-80,0 dB	7,0 dB
10001 Hz	-80,0 dB	80,0 dB

NB! The tolerance masks above are floating (not fixed) and will move up/down trying to center the measured response between tolerance limits.

5.5.8 Send path - single frequency interference

Purpose: Narrow-band noise, including single frequency interference, is an impairment that can be perceived as a tone, depending on its level relative to the overall weighted noise level. This can be caused by electrical noise in soundcards or by fan or hard disk drive noise on laptops. This requirement makes sure that no tonal noise is present in the send signal.

Input: Use recommended test position for DUT. Use following settings for the DUT Editor:

- Enable DUT Client Mode
- Disable AGC
- Disable AEC

The same input AGC setting as for normal speech level test is used for this test case. Skype automatic gain adjustment is disabled.

Play the [SpNR speech sample](#) at normal speech level and record the signal at reference Skype output.

Required: Calculate the A-weighted peak noise level for the 1 second silence part at end of the test signal for SpNR with an effective bandwidth of not more than 31 Hz. When using a 48kHz sampling rate recording this means a minimum FFT size of 4096. Frequency range analyzed is 100Hz to 12000Hz. For FFT analysis the “Flat Top” windowing is employed.

The measured peak noise level is: $\leq -70\text{dBm0}$ ($\leq -76\text{dBFS}$)

5.5.9 Receive path - preferred loudness level in ear at 1m distance

Purpose: To ensure that the receive path volume adjustment has enough adjustment range to allow setting volume to preferred listening level.

Input: Use recommended test position for DUT.

Play the IEEE 269-2010 uncompressed male speech to the reference Skype input and record the acoustic signal in artificial ear.

Required: Calculate the RMS SPL in artificial ear for the active part of male artificial speech signal (active speech does not include pauses or silences). **The level must be set to according to table below**

Screen diagonal ≤ 45 inch	68..72 dBSPL
Screen diagonal > 45 inch	70..74 dBSPL

If the DUT does not have playback speakers a TV set is selected by Certification team to be used as a playback device. The volume is then adjusted according the above table. Typically a 32 inch or 40 inch TV is used for this purpose.

Note: The required level has been changed lower as the testing takes place in anechoic environment. The levels in real rooms will typically be 4..5dB louder with same volume setting due to boundary boost by nearby walls, floor etc.

The level set in this test case will be used for other receiving direction tests that refer to preferred loudness level for playback loudness.

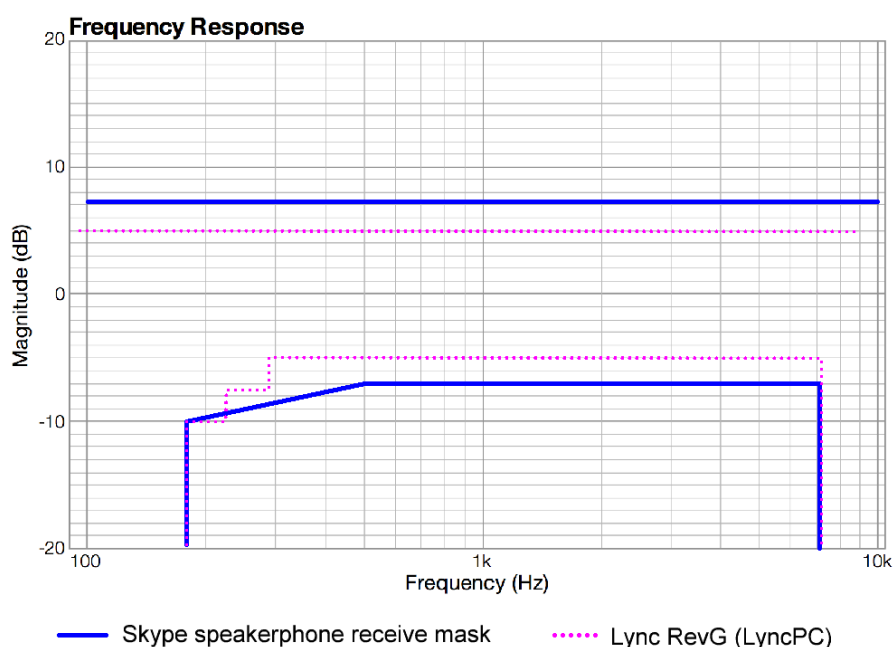
5.5.10 Receive path - frequency response at 1m distance

Purpose: The test checks that the frequency response of the DUT receiving path is flat enough to meet minimum requirement and take full advantage of Skype voice codec bandwidth.

Input: Use recommended test position for DUT.
Use same receive loudness setting as set in 5.5.9
Play the [male artificial speech for frequency response](#) to the reference Skype input.
Record the acoustic receive signal in artificial ear.

Required: The frequency response is calculated by comparing the 1/3 octave spectrum of the prior recorded reference result to the DUT receive signal in artificial ear with the free field correction applied.

The resulting frequency response graph fits into the tolerance mask below:



Frequency	Lower limit	Upper limit
99 Hz	-80,0 dB	80,0 dB
100 Hz	-80,0 dB	7,0 dB
179 Hz	-80,0 dB	7,0 dB
180 Hz	-10,0 dB	7,0 dB
500 Hz	-7,0 dB	7,0 dB
7100 Hz	-7,0 dB	7,0 dB
7101 Hz	-80,0 dB	7,0 dB
10000 Hz	-80,0 dB	7,0 dB
10001 Hz	-80,0 dB	80,0 dB

NB! The tolerance mask is floating (not fixed) and will move up/down trying to center the measured response between tolerance limits.

5.5.11 Receive path - speech signal to noise ratio (SpNR) at 1m distance

Purpose: Too high self noise in receive path decreases the intelligibility of the speech and influences the total call quality in negative way. This test tests for speech to noise ratio in DUT receive path.

Input: Use recommended test position for DUT.

Use same receive loudness setting as set in 5.5.9

Play the IEEE 269-2010 uncompressed male speech to the reference Skype input and record the acoustic signal in artificial ear.

Required: Calculate the RMS SPL level in artificial ear for the active part of male artificial speech (active speech does not include pauses or silences) → Speech level

Calculate the A-weighted RMS SPL level in artificial ear for the 1 second silence part in middle of the male artificial speech signal for SpNR → Noise level.

The calculated speech to noise ratio (SpNR ratio) is more than 40 dB.

Note: Please note that Skype requires speech to noise ratio (SpNR), this is not the same as signal to noise ratio (SNR) specification often found in datasheet that specifies a sine signal to noise ratio.

5.5.12 Receive path – single frequency interference

Purpose: Tonal noise may be perceived in the loudspeaker signal if receive single frequency interference is too high.

Input: Use recommended test position for DUT.

Play the signal containing silence into the REF Skype input

Required: Record a 5 second sound sample from HATS ear microphone. Measure the A-weighted peak noise level over the frequency range of 50 to 20000 Hz with an effective bandwidth of not more than 31 Hz. For 48kHz sampling rate recording this means a minimum FFT size of 4096. For FFT analysis the “Flat Top” windowing is employed.

Compare the analyzed level to the noise level calculated in test case Error! Reference source not found. **Receive A-weighted single frequency interference noise peak level shall be at least 10 dB quieter than the averaged receive noise level.**

5.5.13 Receive path – total harmonic distortion at preferred listening level

Purpose: To verify that DUT loudspeaker has low enough total harmonic distortion. This is important for subjective call quality, but also for acoustic echo cancellation.

Input: Use recommended test position for DUT.

Use same receive loudness setting as set in 5.5.9.

Play the stepped sine signal with -3dBm0 RMS (-9dBFS RMS) or a 1/3rd octave band limited pink noise with rms signal level of -3dBm0 RMS (-9dBFS RMS) to the reference Skype input. Record the acoustic receive signal in artificial ear.

Required: The receive signal is analyzed for total harmonic distortion + noise

The resulting total harmonic distortion + noise must not exceed

< 6.3% (24dB) for frequency 315Hz.

< 5% (26dB) for frequencies 400Hz, 800Hz, 1600Hz, 2500Hz.

Note: Skype test will measure THD also at frequencies below the 315Hz and plot the result for the informational purpose. Vendor should make every attempt to keep the THD below 10% also at lower frequencies as the distortions there are equally problematic for Skype's acoustic echo cancellation.

Correct high pass filtering and EQ should be used at low frequencies to avoid speaker to have a speaker cone over excursion below the speaker resonant frequency.

5.6 Speakerphone: Acoustic Echo Related Requirements – Anechoic Room Testing

NB! The test cases below are valid for all devices of Speakerphone category.

5.6.1 Weighted terminal coupling loss (TCLw)

Purpose: The amount of acoustic echo in the microphone signal is measured by the TCLw and the acoustic echo should be minimized by maximizing the physical distance between the loudspeaker and the microphone. For devices relying on the AEC in Lync, not meeting this requirement will result in echo leak, or distortion and attenuation of speech during double-talk (that is, near-end user and far-end participant talking simultaneously).

For devices with on-board AEC, a failure of this test will lead to echo leaks that are disruptive to the far-end participants.

The TCLw shall be normalized with respect to the nominal send loudness to account for any analog gain difference which would be compensated for by the digital AGC integrated in Skype. The nominal send loudness is defined as -18dBm0 (-24dBFS). The formula for the normalized TCLw is

$$\begin{aligned} TCLw &= TCLw_{measured} + (Send\ loudness_{measured} - Send\ loudness_{nominal}) \\ TCLw &= TCLw_{measured} + (Send\ loudness_{measured} - (-24dBov)) \end{aligned}$$

The TCLw shall be measured at preferred receive loudness using the IEEE Std. 269 male uncompressed speech signal.

Input: Use recommended test position for DUT. Use following settings for DUT editor **during preparation!**

- Enable DUT Client Mode
- Disable AEC

Play the IEEE 269-2010 uncompressed male speech to the reference Skype input. Skype is allowed to automatically adjust the input gain setting during the preparation period. After Skype has adjusted the input gain to optimal level the AGC is disabled.

Use following settings for DUT editor **for the actual test case run!**

- Enable DUT Client Mode
- Disable AGC
- Disable AEC

Use same Preferred loudness level setting from the respective speakerphone sub category (5.2.8 , 5.3.8, 5.4.8 or 5.5.9).

Play the IEEE 269-2010 uncompressed male speech sample at normal speech level to reference Skype input and record the signal at reference Skype output.

Required: TCLw for devices without in built AEC:

Device type	TCLw at nominal receive loudness
Personal speakerphone	>-10dB
PC, all-in-one	>-10dB
PC, laptop	>-10dB
PC, tablet	>-15dB
Long-range speakerphone	>-15dB

TCLw for devices with in built AEC:

Device type	TCLw at nominal receive loudness
Speakerphone	>45dB
PC of any type	>45dB

Note: The measurement shall be performed after system stability is reached (including convergence of any echo algorithms): this shall be accomplished by invoking the test signal for at least 2 seconds before the actual measurement occurs.

Measurements shall be done in 1/3rd octave bands over a range of 100 Hz through 8000 Hz. The weighted terminal coupling loss is calculated according to ITU-T G.122 Annex B.4 (trapezoidal rule) using the frequency range of 100 to 8000 Hz rather than 300 to 3400 Hz.

Note that TCLw is also tested at maximum volume [here](#).

5.6.2 Echo path - acoustic echo cancellation (with Skype AEC)

Purpose: The test checks the level of loudspeaker acoustic echo leaking back to far end output. The test signal includes near and far ends speech, alternating at different times. The talkers are occasionally overlapping to simulate use case where users take turns while speaking, but interrupt each other from time to time.

Input: Use recommended test position for DUT or the DUT placement guidelines in user documentation or “on screen” guidelines.

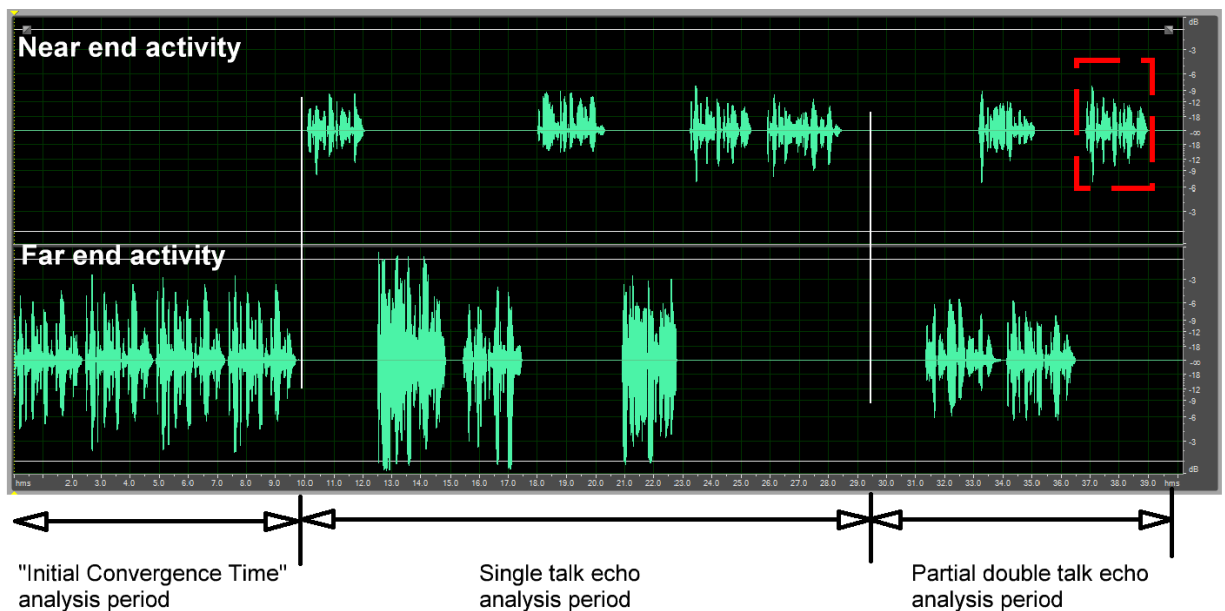
Use following setting for the DUT editor:

- Enable DUT Client Mode

Play the preparation test signals to the reference Skype input and artificial mouth simultaneously.

Use same receive loudness setting as set in respective category’s preferred listening level test case.

Play the Echo test signal to the reference Skype input and artificial mouth simultaneously.



Record the above sequence starting from far end activity start at reference Skype output.

Required: The recording must comply with the following:

- **Initial Convergence Time** – the echo suppression should reach full cancellation at latest after 4 seconds from the start of far end activity
- **Residual echo / loss of convergence**

Calculate the **level versus time** and **spectrum versus time** graph of reference Skype output. Analyze and listen to the recording.

There should be no echo leaks or bursts higher than +5 dB in level compared to the send path noise floor level during far end speech activity in any frequency range between 50 Hz to 20 kHz.

- During the partial double talk there should be minimal loss of near end speech, especially the beginnings of the near end speech. The partial attenuation of near end speech is allowed, but the far end user should be able to recognize that the near end attempts to speak.
- **Send path noise floor stability / similarity.** It is beneficial to keep low level of comfort noise in the send signal path also during far end speech activity periods. Fully muting or attenuating the send signal during far end activity will make possible echo residuals more audible and could lead to failing result for this test case. Also the spectrum of the generated comfort noise should match that of the send path noise during silent periods of near end speech activity. If not – the noise floor changes will be very audible every time the far end is speaking.

Note! As the anechoic room testing takes place in environment with no reverberation and room reflection the echo tail length is not tested properly. Thus the speakerphone devices should always be used and evaluated also in real rooms. For Group Speakerphone and Long Range Speakerphone Skype has a separate specification - Real Room Audio Requirements.

5.6.3 Echo path – send path signal level during two way conversation

Purpose: The test checks the transmitted near end speech level during two way conversation. The send level might be low in such use case, especially on devices where the acoustic echo in microphone is very loud. This usually is due to physical distance between microphone(s) and speaker(s). The loud acoustic echo will force the analog gain control to adjust to lower gain, thus a digital amplification with faster adjustment speed will be needed to compensate for the level loss.

Input: Use the resulting recording from test case 5.6.2

The recording of send path signal in reference Skype output is analyzed. The section marked with red dotted line in above sequence is used for calculation of send path level during conversation

Required: Calculate the RMS level in reference Skype output for the active part of near end speech (time selection marked with red dotted line in above sequence). **The level must be more than -28 dBm0 RMS (equals -34 dBFS RMS).**

5.6.4 Echo path - acoustic echo cancellation (4m speech level) – applicable for Long Range UI only !

Purpose: The test checks the level of loudspeaker acoustic echo leaking back to far end output. The test signal includes near and far ends speech, alternating at different times. The talkers are occasionally overlapping to simulate use case where users take turns while speaking, but interrupt each other from time to time.

Input: Use recommended test position for DUT or the DUT placement guidelines in user documentation or “on screen” guidelines.

Use following setting for the DUT editor:

- Enable DUT Client Mode

Play the preparation test signals to the reference Skype input and artificial mouth simultaneously.

Use same receive loudness setting as set in respective category’s preferred listening level test case.

Play the Echo test signal to the reference Skype input and artificial mouth simultaneously. The 4m speech level is used for near end talker (HATS mouth)

Required: The recording must comply with the following:

- **Initial Convergence Time** – the echo suppression should reach full cancellation at latest after 4 seconds from the start of far end activity
- **Residual echo / loss of convergence**
Calculate the **level versus time** and **spectrum versus time** graph of reference Skype output. Analyze and listen to the recording.

There should be no echo leaks or bursts higher than +5 dB in level compared to the send path noise floor level during far end speech activity in any frequency range between 50 Hz to 20 kHz.

- During the partial double talk there should be minimal loss of near end speech, especially the beginnings of the near end speech. The partial attenuation of near end speech is allowed, but the far end user should be able to recognize that the near end attempts to speak.
- **Send path noise floor stability / similarity.** It is beneficial to keep low level of comfort noise in the send signal path also during far end speech activity periods. Fully muting or attenuating the send signal during far end activity will make possible echo residuals more audible and could lead to failing result for this test case. Also the spectrum of the generated comfort noise should match that of the send path noise during silent periods of near end speech activity. If not – the noise floor changes will be very audible every time the far end is speaking.

Note! As the anechoic room testing takes place in environment with no reverberation and room reflection the echo tail length is not tested properly. Thus the speakerphone devices should always be used and evaluated also in real rooms. For Group Speakerphone and Long Range Speakerphone Skype has a separate specification - Real Room Audio Requirements.

5.6.5 Echo path – send path signal level during two way conversation (4m speech level) - for Long Range UI only !

Purpose: The test checks the transmitted near end speech level during two way conversation. The send level might be low in such use case, especially on devices where the acoustic echo in microphone is very loud. This usually is due to physical distance between microphone(s) and speaker(s). The loud acoustic echo will force the analog gain control to adjust to lower gain, thus a digital amplification with faster adjustment speed will be needed to compensate for the level loss.

Input: Use the resulting recording from test case 5.6.4
The recording of send path signal in reference Skype output is analyzed. The section marked with red dotted line in above sequence is used for calculation of send path level during conversation

Required: Calculate the RMS level in reference Skype output for the active part of near end speech (time selection marked with red dotted line in above sequence). **The level must be more than -34dBm0 RMS (equals -40dBFS RMS).**

5.6.6 Weighted terminal coupling loss (TCLw) – max playback volume

Purpose: The amount of acoustic echo in the microphone signal is measured by the TCLw and the acoustic echo should be minimized by maximizing the physical distance between the loudspeaker and the microphone. For devices relying on the AEC in Lync, not meeting this requirement will result in echo leak, or distortion and attenuation of speech during double-talk (that is, near-end user and far-end participant talking simultaneously).

For devices with on-board AEC, a failure of this test will lead to echo leaks that are disruptive to the far-end participants.

The TCLw shall be normalized with respect to the nominal send loudness to account for any analog gain difference which would be compensated for by the digital AGC integrated in Skype. The nominal send loudness is defined as -18dBm0 (-24dBFS). The formula for the normalized TCLw is

$$\begin{aligned} TCLw &= TCLw_{measured} + (Send\ loudness_{measured} - Send\ loudness_{nominal}) \\ TCLw &= TCLw_{measured} + (Send\ loudness_{measured} - (-24dBov)) \end{aligned}$$

The TCLw shall be measured at preferred receive loudness using the IEEE Std. 269 male uncompressed speech signal.

Input: Use recommended test position for DUT. Use following settings for DUT editor **during preparation!**

- Enable DUT Client Mode
- Disable AEC

Play the IEEE 269-2010 uncompressed male speech to the reference Skype input. Skype is allowed to automatically adjust the input gain setting during the preparation period. After Skype has adjusted the input gain to optimal level the AGC is disabled.

Use following settings for DUT editor **for the actual test case run!**

- Enable DUT Client Mode
- Disable AGC
- Disable AEC

Use same Maximum loudness setting.

Play the IEEE 269-2010 uncompressed male speech sample at normal speech level to reference Skype input and record the signal at reference Skype output.

Required: **TCLw for devices without in built AEC:**

Device type	TCLw at maximum receive loudness
Personal speakerphone	>-15dB
Group speakerphone	>-15dB
PC, all-in-one	>-15dB
PC, laptop	>-15dB
PC, tablet	>-20dB
Long-range speakerphone	>-20dB

TCLw for devices with in built AEC:

Device type	TCLw at maximum receive loudness
Speakerphone	>40dB
PC of any type	>40dB

Note: The measurement shall be performed after system stability is reached (including convergence of any echo algorithms): this shall be accomplished by invoking the test signal for at least 2 seconds before the actual measurement occurs.

Measurements shall be done in 1/3rd octave bands over a range of 100 Hz through 8000 Hz. The weighted terminal coupling loss is calculated according to ITU-T G.122 Annex B.4 (trapezoidal rule) using the frequency range of 100 to 8000 Hz rather than 300 to 3400 Hz.

5.7 Speakerphone: Audio Real Room Test Requirements

5.7.1 Audio Real Room Testing

Purpose: As the anechoic room testing takes place in environment with no reverberation and room reflection the echo tail length is not tested properly. Thus the speakerphone devices should always be used and evaluated also in real rooms. For Group Speakerphone and Long Range Speakerphone Skype has a separate specification - Real Room Audio Requirements.

Also the speech pickup quality might degrade when background noise is present in the room the DUT is used in. Anechoic room has a very low background noise level and thus does not test for this. Thus a Real Room Testing is done with a pre-defined background noise presence among other tests.

Input: Refer to the Audio Real Room Requirements document

Required: A DUT must meet all the Audio Real Room Requirements applicable for the respective category (categories are either a Group Speakerphone or Long Range Speakerphone)

5.8 Speakerphone: Requirements for Skype Super Wideband Certification (Optional)

5.8.1 Echo path - round trip delay – Super Wideband quality

- Purpose:** Call interactivity and acoustic echo audibility is dependent of the round trip delay. The purpose of this test is to ensure that the round trip delay during Skype to Skype call using the DUT in lossless local network is below the set maximum limit.
- Input:** Use recommended test position for DUT.
- Make a Skype to Skype audio only call.
- Let the Skype calls stabilize for > 3 minutes. As test is done in lossless network jitter buffer in both Skype client will adjust to low length.
- Measure the delay in sending direction and then in receiving direction by using the [short delay test signal](#) and [long delay test signal](#). The delay is calculated using cross correlation calculation.
- Round trip delay = sending direction delay + receiving direction delay.
- Required:** The calculated round trip delay must be below:
- 300ms** – for devices with wired connection between Skype client PC and device under test.
- 350ms** – for battery powered devices using a wireless link between the computer and device under test.
- Note:** The average calculated sending direction delay and receiving direction delay is verified from PESQ total quality loss test results. PESQ provides delay graph based on the time alignment of degraded signal versus reference signal, thus provides a very precise delay calculation.

5.8.2 Send path - frequency response – Super Wideband.

Purpose: The test checks that the frequency response of the DUT send signal path is flat enough to meet Skype requirement and take full advantage of Skype Super Wideband voice codec (SILK SWB) bandwidth.

Input: Use recommended test position for DUT. Use following settings for the DUT Editor:

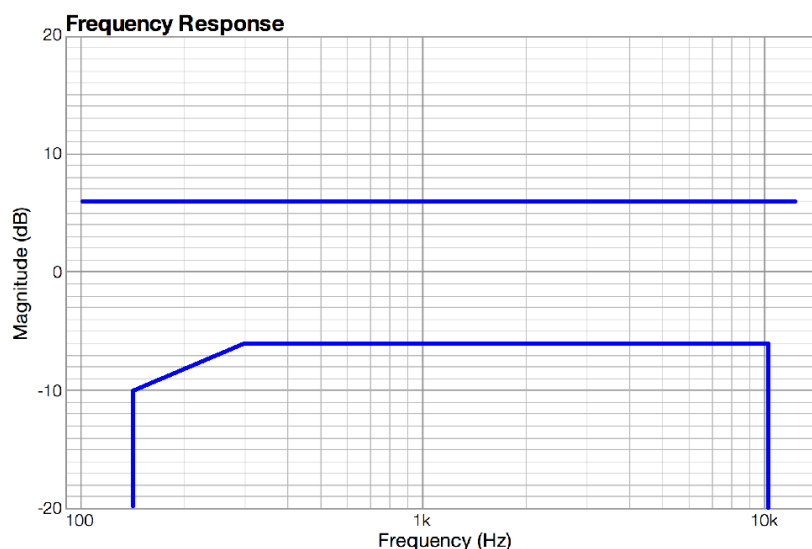
- Enable DUT Client Mode
- Disable AGC
- Disable AEC

The same input AGC setting as for normal speech level test is used for this test case.
Skype automatic gain adjustment is disabled

Play the [male artificial speech for frequency response](#) at normal speech level and record the signal at reference Skype output.

Required: The frequency response is calculated by comparing the 1/3 octave spectrum of the prior recorded reference result at MRP to the DUT send signal recorded at reference Skype output.

The resulting frequency response graph fits into below tolerance mask:



— Skype speakerphone send mask for Super Wideband

Frequency	Lower limit	Upper limit
99 Hz	-80,0 dB	80,0 dB
100 Hz	-80,0 dB	6,0 dB
179 Hz	-80,0 dB	6,0 dB
180 Hz	-10,0 dB	6,0 dB
400 Hz	-6,0 dB	6,0 dB
10000 Hz	-6,0 dB	6,0 dB
10001 Hz	-80,0 dB	6,0 dB
12000 Hz	-80,0 dB	6,0 dB
12001 Hz	-80,0 dB	80,0 dB

NB! The tolerance masks above are floating (not fixed) and will move up/down trying to center the measured response between tolerance limits.

5.8.3 Receive path - frequency response – Super Wideband

Purpose: The test checks that the frequency response of the DUT receiving path is flat enough to meet minimum requirement and take full advantage of Skype Super Wideband voice codec (SILK SWB) bandwidth.

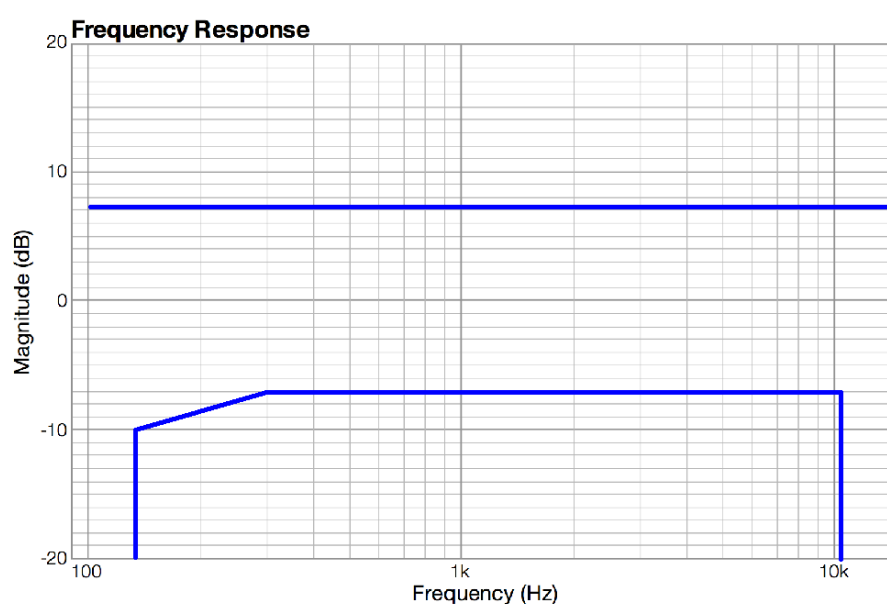
Input: Use recommended test position for DUT.

Use receive loudness setting as set in respective category's preferred listening level test case.

Play the [male artificial speech for frequency response](#) to the reference Skype input.
Record the acoustic receive signal in artificial ear.

Required: The frequency response is calculated by comparing the 1/3 octave spectrum of the prior recorded reference result to the DUT receive signal in artificial ear with the free field correction applied.

The resulting frequency response graph fits into the tolerance mask below:



— Skype speakerphone receive mask for Super Wideband

Frequency	Lower limit	Upper limit
99 Hz	-80,0 dB	80,0 dB
100 Hz	-80,0 dB	7,0 dB
139 Hz	-80,0 dB	7,0 dB
140 Hz	-10,0 dB	7,0 dB
300 Hz	-7,0 dB	7,0 dB
10500 Hz	-7,0 dB	7,0 dB
10501 Hz	-80,0 dB	7,0 dB
12000 Hz	-80,0 dB	7,0 dB
12001 Hz	-80,0 dB	80,0 dB

NB! The tolerance masks above are floating (not fixed) and will move up/down trying to center the measured response between tolerance limits.

5.9 Speakerphone: Supporting Audio Documentation Requirements

In addition to the user manual (the one that comes with the product) in Certification testing we also ask for supporting audio documentation. Such documentation contains engineering data and engineering test data of the product.

5.9.1 Verifying supporting documentation for Speakerphone audio

Purpose: Solution must come with a supporting audio documentation (only for certification testing purposes).

Required: DUT arrives with supporting audio documentation that contains the following information:

- Usage related info:
 - Recommended operating distance
 - Maximum operating distance
- Active signal processing: yes/no, if yes then:
 - Active beam forming microphone and/or loudspeaker: yes/no
 - In built acoustic echo cancellation: yes/no
 - Echo cancellation operating bandwidth (narrowband, wideband, super wideband)
 - Noise suppression: yes/no, in sending or/and receiving directions
 - Automated Gain Control: yes/no, in sending or/and receiving directions
 - Other: describe what, sending or/and receiving directions
- Microphone/s:
 - Frequency range (lowest and highest audible frequencies)
 - Directionality/design principle of a microphone
 - Number of microphones / microphone inputs
 - Microphone phantom power yes/no, supply voltage (if applicable)
 - Microphone input connector type (balanced, unbalanced) (if applicable)
- Loudspeaker/s:
 - Frequency range (lowest and highest audible frequencies)
 - Number of loudspeaker / line outputs (if applicable)
 - Loudspeaker design principle (one/multi way, open/closed box/bass reflex)

6.0 Other Audio Product Group

6.1 Other Audio Product: Audio Test Instructions

6.1.1 Objective testing measurement setup

Objective testing arrangement depends on if the testing is performed with or without acoustic interface. Example of the earlier is a sound card that can be tested together with headset. The example of latter is an audio processing algorithm that does not give direct signal to acoustic interface device. In a case the device is tested together with acoustic interface, the testing setup can be picked from headset, handset or speakerphone test instructions in the previous chapters. In another case when acoustic interface is not used, the electric to electric tests between two Skype clients and the DUT can be performed.

The measurements will be performed mainly in Skype call having all speech enhancement algorithms as they are by default in Skype and potential device audio drivers.

Frequency response results are averaged to 1/3 octave frequency resolution.

6.2 Other Audio Product: Analog In/Out Interface Product (Soundcard)

6.2.1 Output socket for headphones

Purpose: To allow connecting any standard headphone as an external device using a 3.5mm (1/8-inch) plug or alternatively 6,5mm TRS socket.

Required: Device has a stereo headphone output accepting an industry standard 1/8-inch (3.5mm) stereo plug or alternatively 6,5mm TRS socket.

6.2.2 Input socket for headset microphone or professional microphone

Purpose: To allow connecting standard headset microphone or professional microphone.

Required: Microphone input accepts industry standard 3.5 mm (1/8-inch) tip/ring/sleeve microphone sockets where the microphones signal is on the tip, bias is on the ring, and the sleeve is grounded. Bias voltage must be +2.5 ...5.5V DC.

Or alternatively product accepts XLR, mini XLR or 6,5mm TRS connector and provides +24...48V DC phantom power

Note: If headset connector uses other non-standard configuration the pass or fail of this test case will be decided case by case.

6.2.3 Microphone input socket - Frequency response

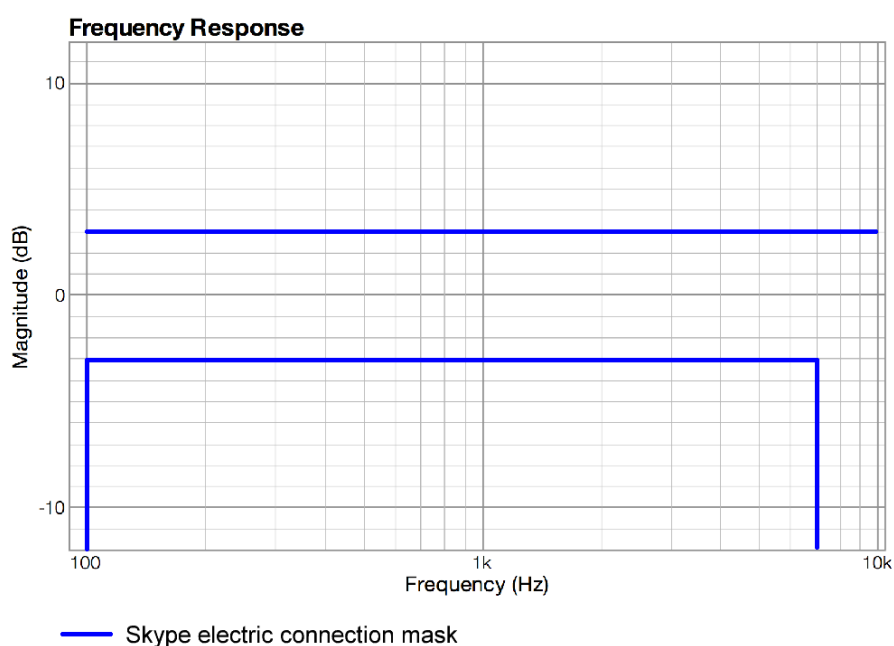
Purpose: To verify that the frequency response of microphone input passes the minimum requirement.

Input: Play back a measurement signal to microphone input socket. Make sure the signal level is appropriate and does not overload / clip the input.

Required: Measure frequency responses of microphone input by comparing the recorded test signals to the original test signal. The resulting frequency responses fit into a wideband tolerance window:

Note 1: This test case is not measured over Skype to Skype call. The [ARTA](#) measurement software or other suitable software and loop back method is used.

Note 2: This test only needs to be passed if there is a microphone input socket available.



Frequency	Lower limit	Upper limit
99Hz	-80,0 dB	20,0 dB
100 Hz	-3,0 dB	3,0 dB
7000 Hz	-3,0 dB	3,0 dB
7001 Hz	-80,0 dB	3,0 dB
10500 Hz	-80,0 dB	3,0 dB
10501Hz	-80,0 dB	20,0 dB

NB! The tolerance masks above are floating (not fixed) and will move up/down trying to center the measured response between tolerance limits.

6.2.4 Headphone output socket frequency response

Purpose: To verify that the headset output frequency response curves pass the minimum requirement.

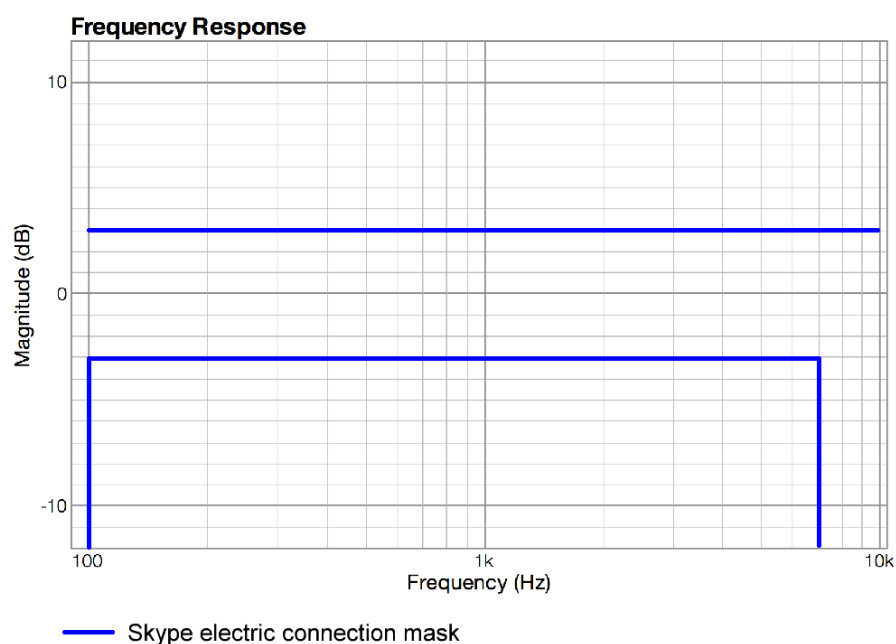
Input: Play back a measurement signal to headphone / headset output.

Required: Measure frequency responses of headphone output by comparing the output signal to the original signal. The resulting frequency responses fit into a super wideband tolerance window:

The test case must pass the requirement with a 50 ohm to 600 ohm resistive load impedances.

Note 1: This test case is not measured over Skype to Skype call. The [ARTA](#) measurement software or other suitable software and loop back method is used.

Note 2: The test case must pass the requirement with a 50 ohm to 600 ohm resistive load impedances.



Frequency	Lower limit	Upper limit
99Hz	-80,0 dB	20,0 dB
100 Hz	-3,0 dB	3,0 dB
7000 Hz	-3,0 dB	3,0 dB
7001 Hz	-80,0 dB	3,0 dB
10500 Hz	-80,0 dB	3,0 dB
10501Hz	-80,0 dB	20,0 dB

NB! The tolerance masks above are floating (not fixed) and will move up/down trying to center the measured response between tolerance limits.

6.3 Other audio product: Supporting audio documentation requirements

In addition to the user manual (the one that comes with the product) we also ask for supporting audio documentation (for certification testing purposes). Such documentation contains engineering data and engineering test data for the product.

6.3.1 Verifying supporting documentation for Other audio product

Purpose: Solution must come with a supporting audio documentation (only for certification testing purposes).

Required: DUT arrives with supporting audio documentation that contains the following information:

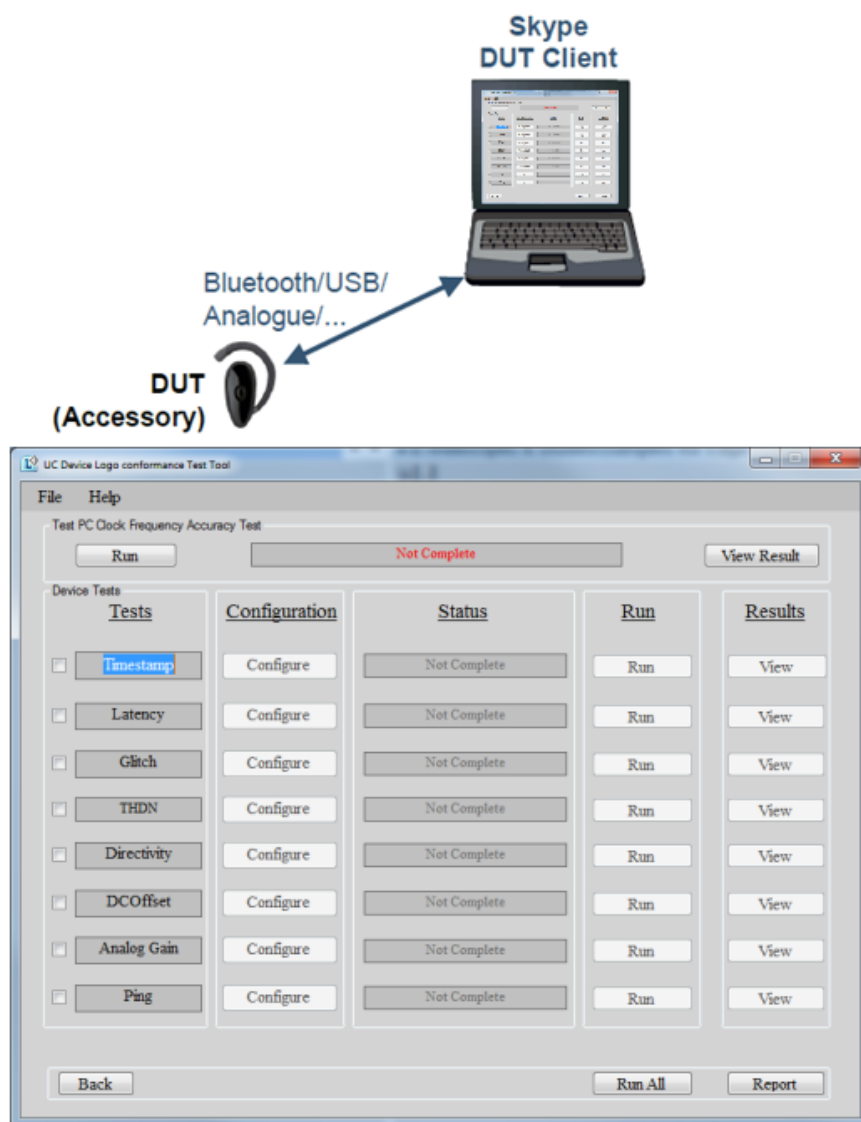
- Sending:
 - Speech signal delay from input to output (if above 5 ms)
 - Usable frequency bandwidth
- Receiving:
 - Speech signal delay from input to output (if above 5 ms)
 - Usable frequency bandwidth
- Connectors (if applicable):
 - Type/s & electric connections (ground, signals, bias voltages...)
 - Target input and output levels/voltages
 - Maximum input and output levels/voltages
 - Maximum and minimum impedances for external connection
- Volume control (if applicable):
 - Range in dB
 - Minimum volume (dBV, dBSPL or similar RMS)
 - Maximum volume (dBV, dBSPL or similar RMS)
- Active signal processing: yes/no
 - if yes then what?

7.0 Test Setup And Test Environment Details

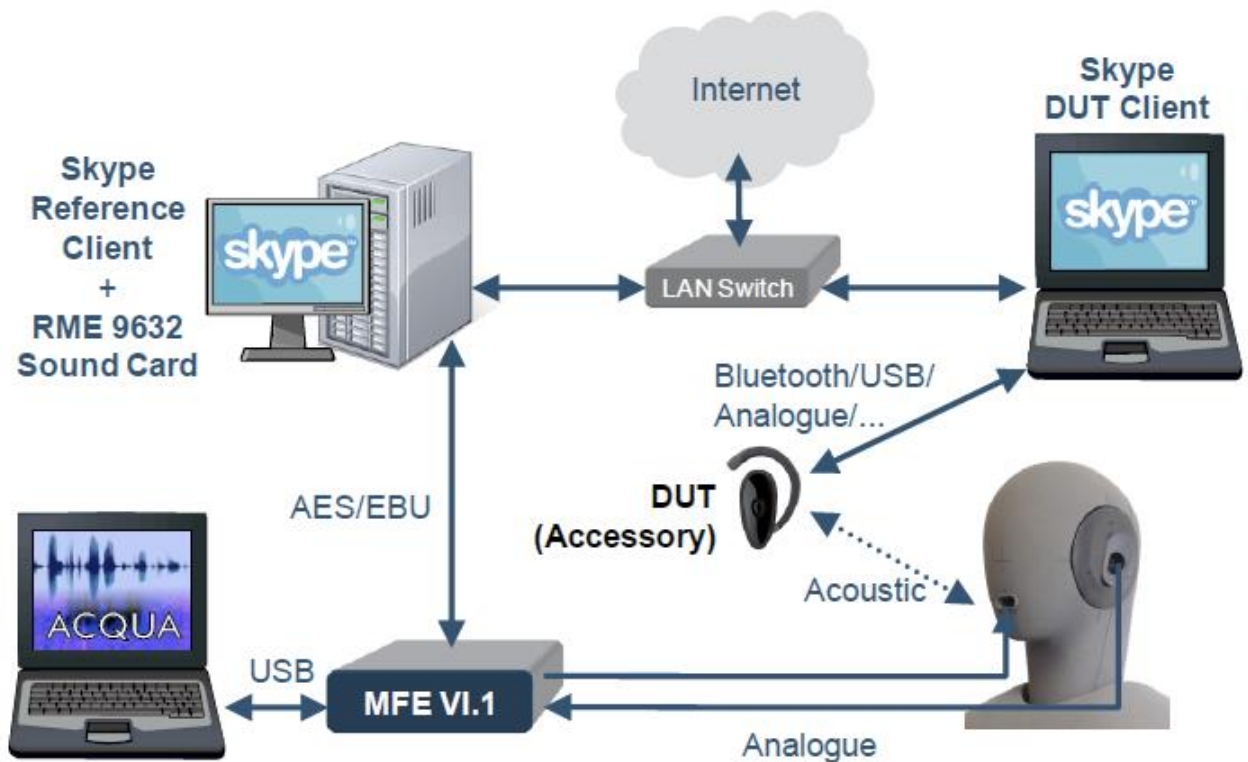
7.1.1 Microsoft Device Conformance Test Tool test setup

This test tool is not part of the HEAD Acoustics Skype AS test standard.

To test Skype devices for the tests in section 2.2 “Requirements to be tested using Microsoft UC Device Conformance Test Tool”, the tool shall be running on a DUT PC and the device under test (DUT) shall be connected to the same reference PC. The reference PC shall have Windows 7 as the operating system and shall pass all the glitch and timestamp requirements. This is to ensure that a DUT submitted for certification will not fail due to failures of the reference PC. When testing PCs, the tool shall be running on the PC itself (and no reference PC is required in that case). The PC under test shall also have Windows 7 or Windows 8 as OS.



7.1.2 Objective test measurement setup



- Skype to Skype call is created between two Skype clients in lossless local network condition, unless specified otherwise in specific test case.
- Reference Skype client runs on PC with Windows 7 32 bit operating system.
- A third computer runs ACQUA audio measurement system and MFE measurement front end connected to HATS and the Reference Skype PC.

7.1.3 Test setup hardware/software availability.

Skype has partnered with Head-Acoustics so that it is possible to purchase the needed hardware and software tools to carry out anechoic room testing of Skype devices against the above Skype requirements.

The test packages consist of the Measurement front-end(s), other necessary hardware tools and ACQUA software. Skype test sequences are offered as additional "Skype measurement standard" package for ACQUA software.

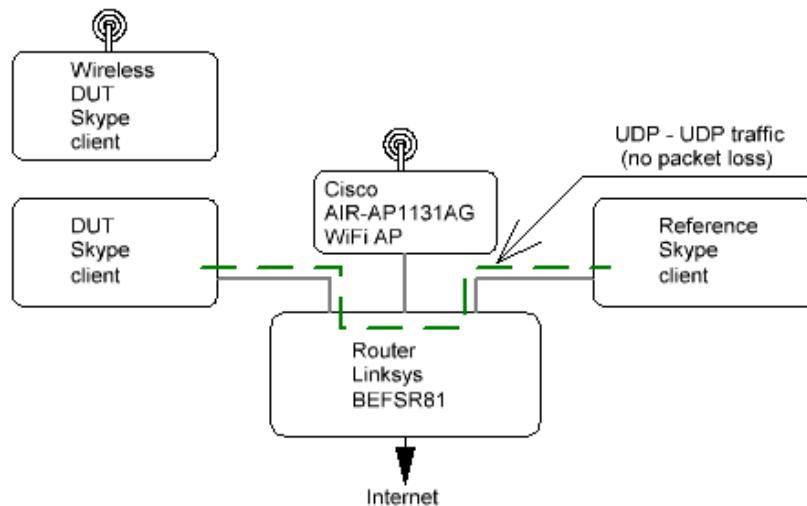
In order to get a quote from Head-Acoustics please contact Mr. Ossi Raivio who will forward the request to respective sales representative in Your country/region:

Ossi RAIVIO, M.Sc.
Sales Telecom
Tel +49 2407 577 114
Fax +49 2407 577 99
Email ossi.raivio@head-acoustics.de
Web <http://www.head-acoustics.de>

HEAD acoustics GmbH · Ebertstr. 30a · 52134 Herzogenrath · Germany
Vertretungsberechtigte Geschäftsführer Prof. Dr.-Ing. Klaus Genuit, Reinhard Scholz · Registergericht Amtsgericht Aachen · Registernummer HRB 3468

7.1.4 Lossless network condition test setup

By “lossless network condition” Skype defines the network setup that uses a local router with DHCP server. This guarantees a lossless high bandwidth and low latency network transmission as Skype peer to peer call is routed locally and is not influenced by the outbound internet connection speed.



After starting the Skype to Skype call ensure in Reference Skype client that the network statistics show a UDP-UDP call with minimal packet loss and no relays.

Example of lossless network condition!

```
Packet loss 0.0% (0)
Send packet loss 0.0%/0.0%
Recv packet loss 0.0%/0.0%
Roundtrip 0ms
NBM audio 3000 / 40 ms video 0 corr 0%
SessionOut UDP (718 packets)
SessionIn UDP (678 packets)
Relays 0
UDP status local Good remote:Good
```

7.1.5 List of equipment at Skype AudioLab

- **Head Acoustics ACQUA software – automated audio testing system with Skype-SA option package (Code: 60000)**
http://www.head-acoustics.de/eng/telecom_acqua.htm
http://www.head-acoustics.de/downloads/eng/acqua/acqua18e_mail.pdf
- **Head Acoustics MFE VI.I Measurement Front End**
http://www.head-acoustics.de/eng/telecom_acqua_mfe_VI_1.htm
http://www.head-acoustics.de/downloads/eng/mfe/D6462e1_MFE_VI_1.pdf
- **Bruel and Kjaer Head and Torso Simulator – model 4128C**
<http://www.bksv.com/1650.asp>
<http://www.bksv.com/pdf/Bp0521.pdf>

NB !
The Head Acoustics HMS II.3 with 3.3 artificial ear is equally valid as HATS
http://www.head-acoustics.de/eng/telecom_hms_II_3.htm
- **Bruel and Kjaer Head and Torso Simulator – handset positioner 4606**
<http://www.bksv.com/pdf/Bp0521.pdf>
- **Soundcard in Reference Skype Client PC – RME HDSP9632**
http://rme-audio.de/en_products_hdsp_9632.php
- **DUT Skype Client and Reference Skype Client PC specification**
Intel DG965SS motherboard with BIOS version MQ96510J.86A.1666.2007.0327.2349
The processor in all PC-s: Intel 630 P4 FSB800 2MB 3.0GHz
2GB (4x512Mb 533MHz DDR2 NON-ECC CL4 Kingston DIMM modules)
Samsung 40Gb SATAII NCQ 7200 RPM 8Mb Hard drive
Samsung DVD ROM
- **Local LAN router with DHCP server** (Linksys BEFSR81 ver 3.1 is used in Skype lab)
- **WiFi access point for wireless devices** (Cisco Aironet AIR-AP1131AG or HEAD Acoustics MFE IX is used in Skype lab)
- **ARTA software – for Clarity measurements in real rooms** <http://www.fesb.hr/~mateljan/arta/>
- **Automated turntable (optional)**
http://www.linearx.com/products/accessories/LT360/LT360_01.htm

7.1.6 Reference Skype client setup details

- **Soundcard in Reference Skype Client PC – RME HDSP9632**

http://rme-audio.de/en_products_hdsp_9632.php

- **Reference Skype Client PC specification**

OS Name Microsoft Windows 7 Enterprise

Version 6.1.7600 Build 7600

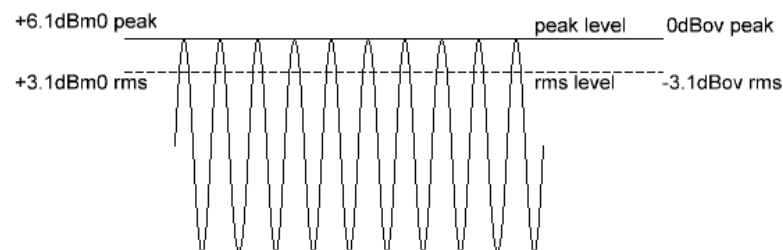
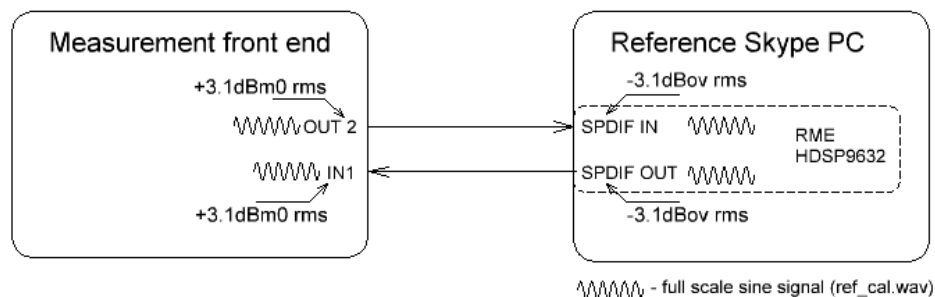
System Type X86-based PC

Processor Intel(R) Core(TM)2 Duo CPU E8400 @ 3.00GHz, 2997 Mhz, 2 Core(s), 2 Logical Processor(s)

Installed Physical Memory (RAM) 4.00 GB

Calibration procedure for Reference Skype client <-> measurement system

- 1) Play back a full scale 1019Hz sine signal to Reference Skype soundcard SPDIF output (ref_cal.wav) – calibrate the measurement system so that it shows +3.1 dBm0 as signal RMS level.
- 2) Generate a 1019 Hz sine signal on measurement system and calibrate the output so that it provides a -3.1 dBFS RMS level (equals 0 dBFS peak) to RME HDSP9632 SPDIF input.



There is 6dB difference between the dBov levels in PC and the dBm0 levels used by Acqua measurement system. The PC level reading is always 6dB less than the Acqua reading

NB! - In order to minimize the influence of Skype client pre-processing, AEC etc. the Skype client audio pre-processor is disabled in reference Skype PC. The details how this can be achieved are available on SkypeKit support website.

7.1.7 DUT Editor and REF Editor setup details

Ensure the DUT is fully charged in case it is a battery powered device.

Ensure that correct wireless connection is selected and shows strong signal.

Do not start any other applications beside the Skype client on DUT.

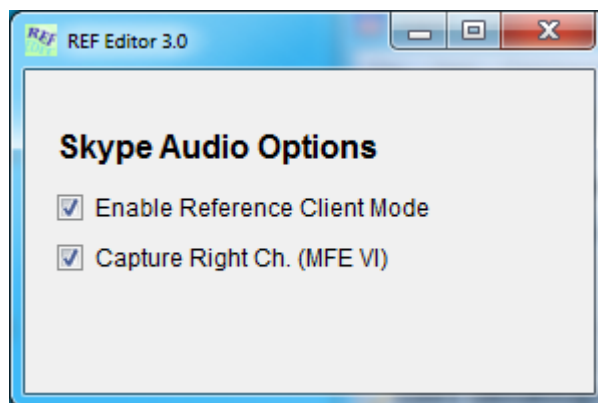
After starting the Skype to Skype call **ensure in Reference Skype client** that the network statistics show a UDP-UDP call with minimal packet loss and no relays.

Example of lossless network condition:

```
Packet loss 0.0% (0)
Send packet loss 0.0%/0.0%
Recv packet loss 0.0%/0.0%
Roundtrip 0ms
NBM audio 3000 / 40 ms video 0 corr 0%
SessionOut UDP (718 packets)
SessionIn UDP (678 packets)
Relays 0
UDP status local Good remote:Good
```

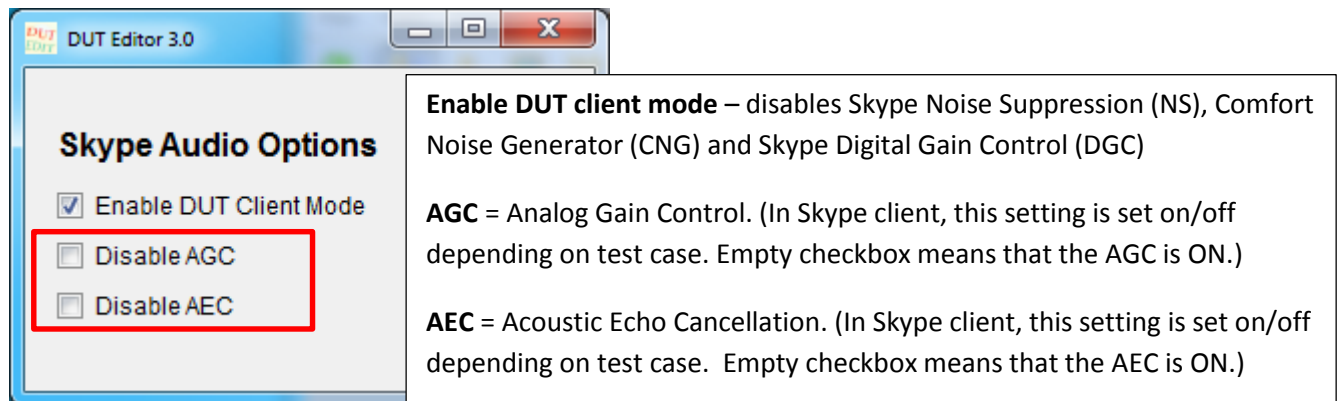
REF Editor 3.0 setup details

Use the REF Editor tool to select a Reference PC mode. Fully Quit Skype client and reopen it to make sure the settings are applied correctly.

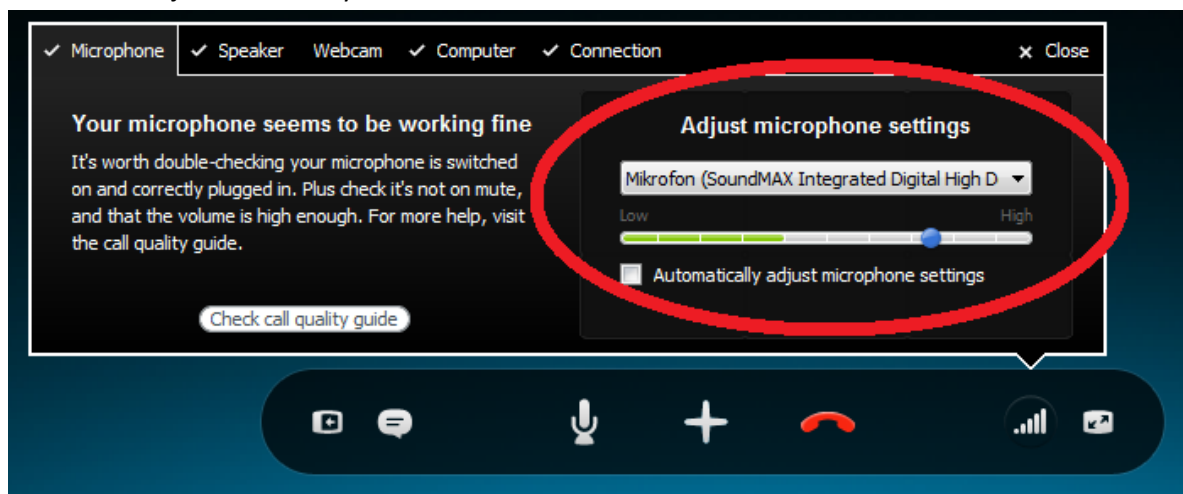


DUT Editor 3.0 setup details

Use the DUT Editor tool to select a DUT PC mode. Settings set by DUT editor activate “on the fly” during the Skype to Skype calls in 1..3 second activation time after change of setting.

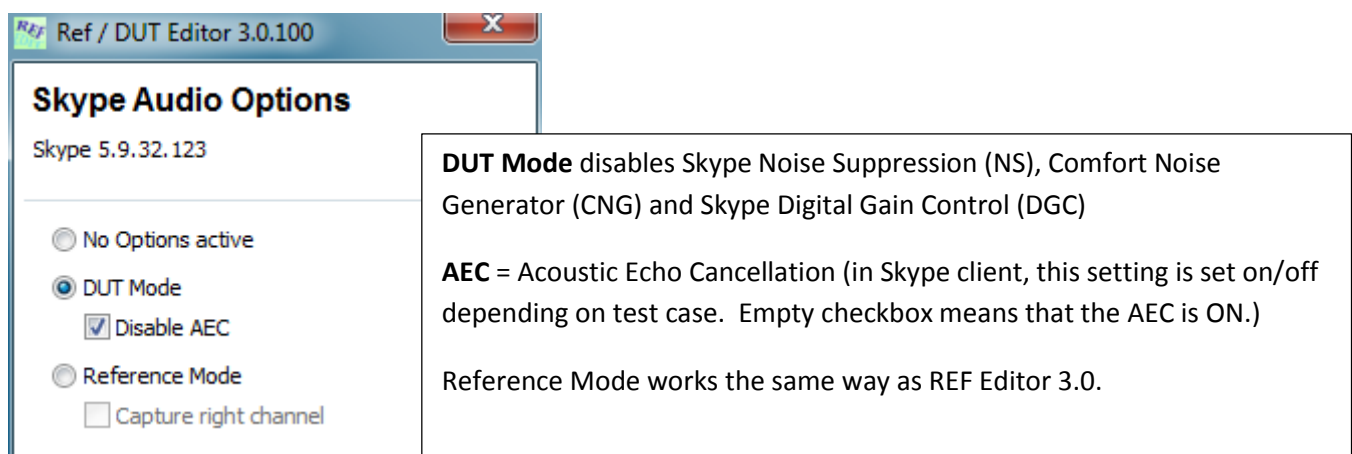


NB! The Disable AGC functionality in DUT Editor 3.0 might not work with some versions of Skype. Make sure that the microphone settings "automatically adjust microphone settings" are turned **OFF (unchecked)**! If it is not then adjust it manually as needed.



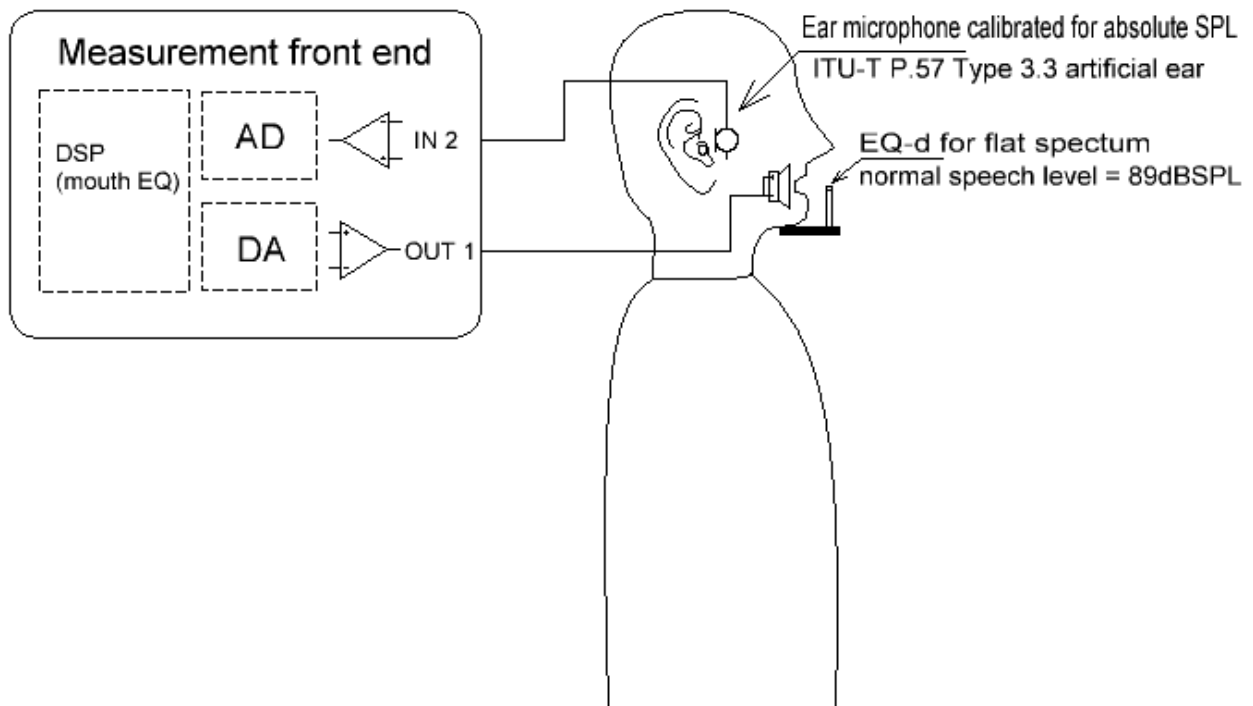
Ref / DUT Editor 3.0.100

The Ref / DUT Editor version 3.0.100 does not provide automatic means to control the AGC. That must be done manually as indicated on the snapshot above. This version comes with ACQUA package.



7.1.8 Head and Torso simulator and calibrations

Skype uses Head and Torso Simulator (HATS) compliant with ITU-T [P.58](#) and artificial ear compliant with ITU-T P.57 Type 3.3.



The frequency spectrum of mouth simulator is calibrated and frequency compensated at Mouth Reference Point (MRP) to be flat between 100 Hz to 11000 Hz.

The normal speech level for active speech part of male and female artificial speech is calibrated to be -5 dBPa (89 dB SPL) at MRP. The analysis is done with no frequency weighting and in frequency range from 50Hz to 20'000Hz

Quiet and loud speech signals are respectively 10 dB quieter and 10 dB louder compared to normal speech level.

Skype uses Drum Reference Point DRP to ERP for all Headset and Handset category receiving path frequency response measurements.

Skype uses Drum Reference Point (DRP) to Free Field Correction (FFC) for all Speakerphone category receiving path frequency response measurements.

7.1.9 Test signals used

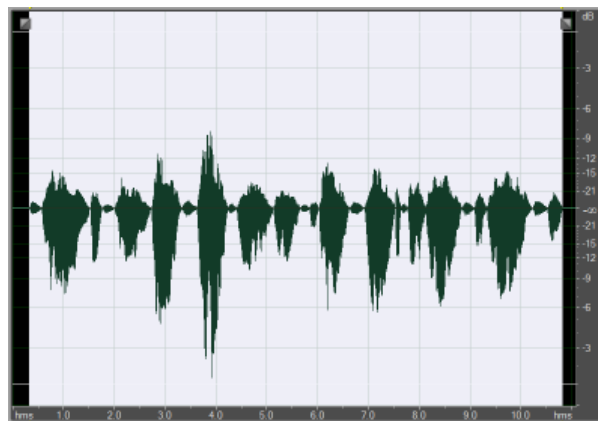
- NB! The test vectors package is available only for vendors that are participating in pre-certification.
- All samples are played to Reference Skype input with exactly the same level as in the source file. For example the IEEE_Male_dual_mono.wav has a -24dBFS average level for the speech. This level is used and played into the Reference Skype input and thus represents also the signal level that will be played back during receive path tests on the DUT.

Short delay test signal	short_delay.wav
Long delay test signal	long_delay.wav
IEEE Male speech	IEEE_Male_dual_mono.wav
IEEE compressed male speech (for loud speech test cases)	IEEE__Male_dual_mono_compressed.wav
IEEE Male speech + silence	IEEE__Male_dual_mono_SPNR.wav
IEEE Male speech for noise during speech test	IEEE__Male_dual_mono_SPNR_speech.wav
Male artificial speech for SpNR during speech	P50_for_SpNR_speech.wav
Modified artificial speech for frequency response	P.50_for_SWB.wav
Echo test signal	IEEE_AEC_single_talk_double_talk_Slevel.wav
Reference Skype output calibration file (Full scale peak to peak Sine signal of 1019Hz)	Ref_cal.wav
Stereo playback test signal	Smooth_as_SILK_stereo_test.wav

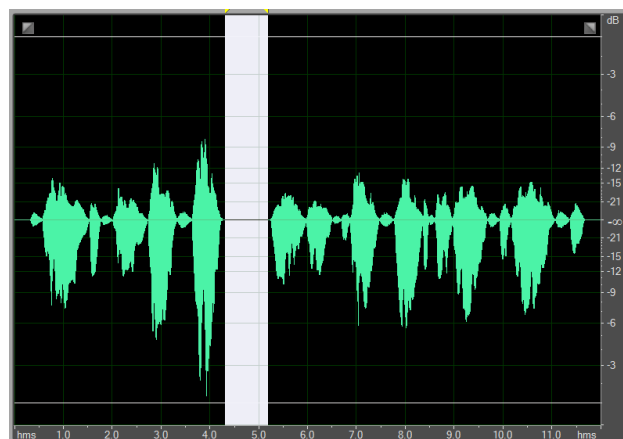
NB! The test vectors Skype uses for POLQA MOS-LQO measurements are Skype original recordings. These have been selected and pre-processed to give a maximum 4.75 MOS score when signal is compared to itself (mode in POLQA to test the signal quality as the test signal).

7.1.10 Speech to noise ratio calculation example

Sample of male artificial speech (active speech part selected– on white background)



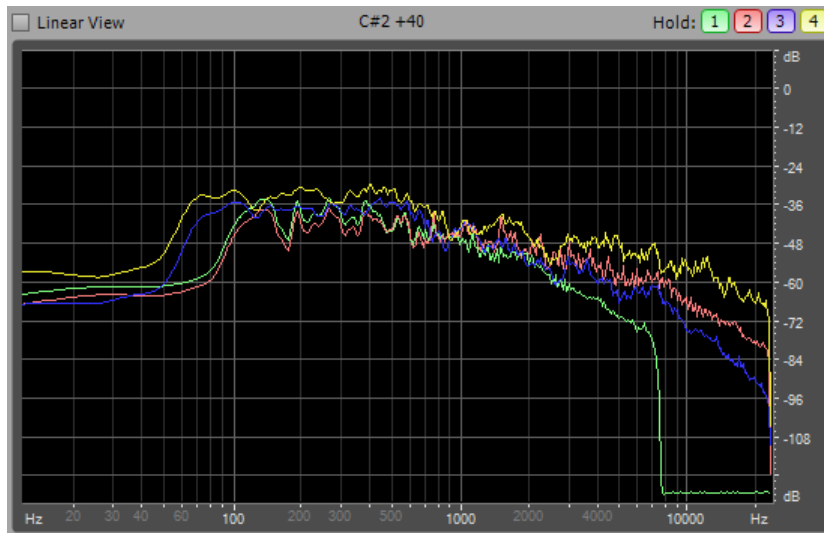
Sample of male artificial speech for SpNR – (noise part selected)



7.1.10.1 Frequency response test signal comparison

Male artificial speech (green) versus Male artificial speech for frequency response (blue) spectrum comparison.

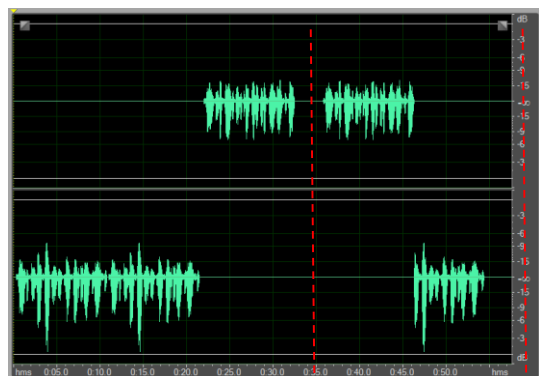
Note the SNR of blue test signal at 5.8 kHz and also high frequency content above 8kHz which makes the signal suitable for Super Wideband measurements.



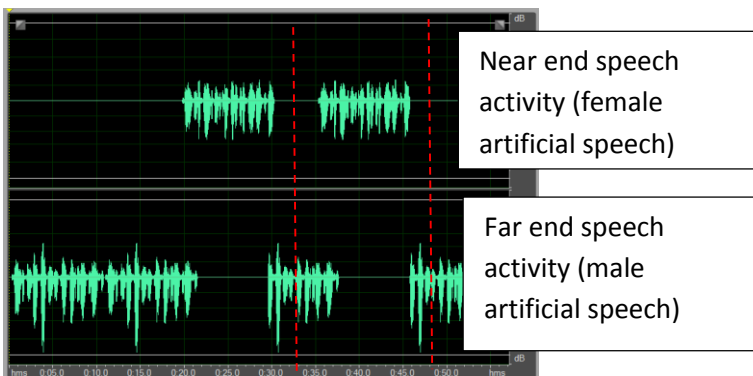
Green = original artificial P.50
Red = Skype modified artificial P.50
Blue = IEEE male speech
Yellow = IEEE compressed male + female samples

7.1.10.2 Single talk and double talk echo test sample analysis

Echo single talk test signal



Echo double talk test signal

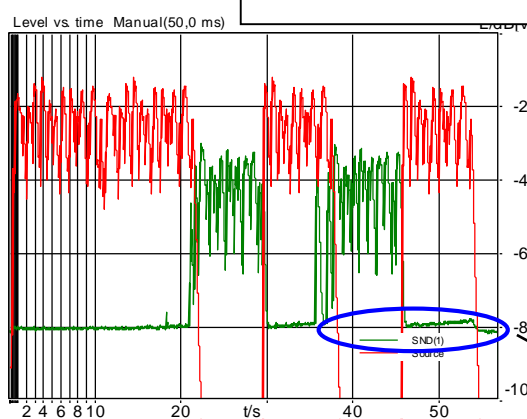
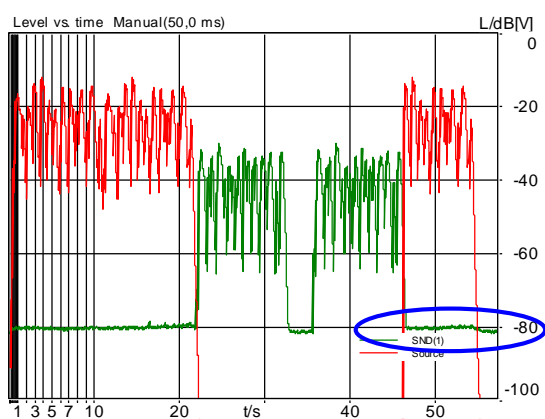


Near end speech activity (female artificial speech)

Far end speech activity (male artificial speech)

Sample result when no echo is leaking – passing result

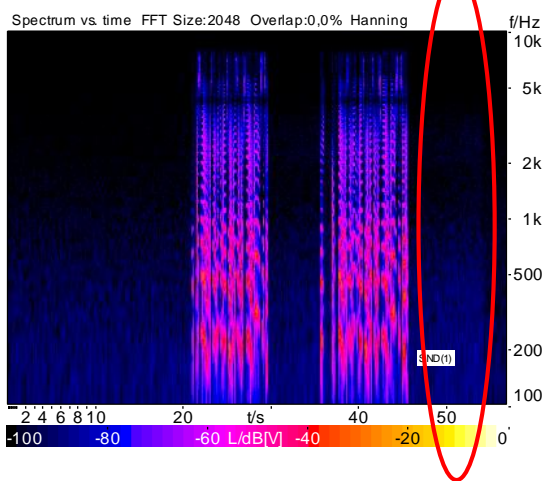
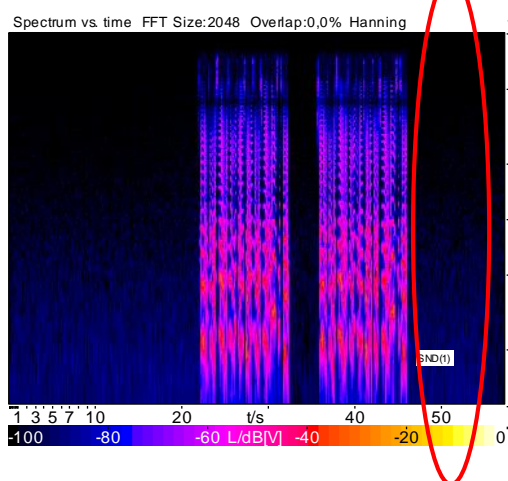
Level versus time



red = far end activity,
green = DUT send signal recorded in far end

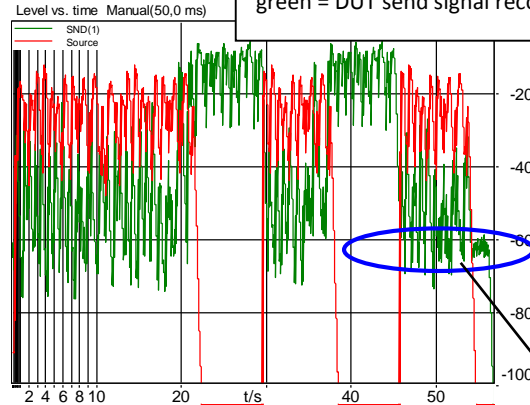
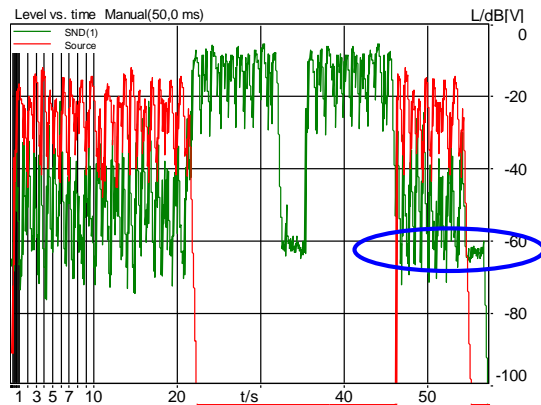
No leaking echo of far end speech visible.

Spectrum versus time versus signal level



Sample result when echo is leaking – failing result

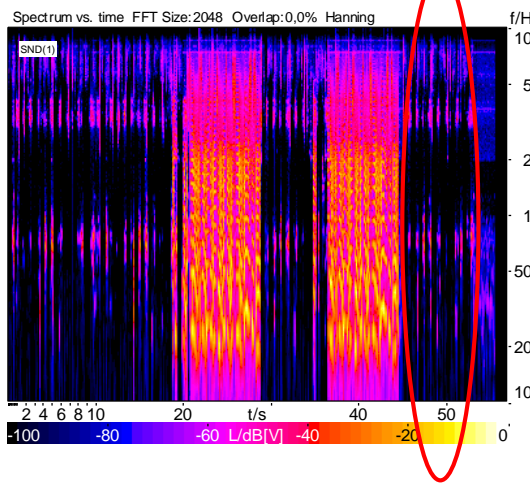
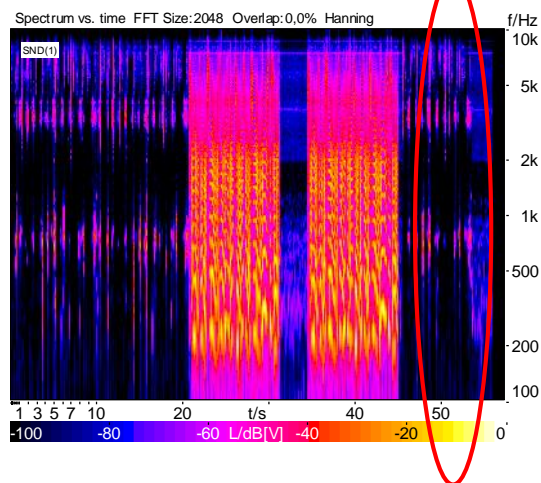
Level versus time



red = far end activity,

green = DUT send signal recorded in far end

Spectrum versus time versus signal level



Echos are leaking, especially around 700Hz and 4000Hz frequency ranges, showing potential problems with loudspeaker performance.

7.1.11 Skype AudioLab – anechoic room

An anechoic chamber for audio measurements has been built in the Skype office building in Tallinn. The chamber has been designed to fulfill the acoustical requirements for high quality wideband VoIP and telecom product testing, ie. free field conditions at a frequency range of 160 Hz – 20 kHz.

The free field performance of the chamber fulfills the ITU-T P.341 recommendation requirement for Test rooms (A.3.1.1) at one third octave bands down to 160 Hz. In hands free and loudspeaker measurements the room is anechoic down to 160 Hz.

The background noise level inside the chamber is lower than the self-noise of typical measurement microphones, and does not pose a limit to measurements in the design frequency range. The measured A-weighted background noise level was below 20 dB.

Airborne noise from activities in the open office or the control room does not interfere with measurements inside the chamber. The measured airborne sound insulation is above 70 dB between the open office and the chamber, and above 55 dB between the control room and the chamber.



7.1.12 Compatible testing environments

Majority of the tests in Skype requirement documents have to be conducted in dedicated test room. Only a speakerphone category has some specific tests that have to be run in real rooms. While the headset or handset can be tested in quiet room with low reverberation, the speakerphone category testing does need an anechoic room as otherwise any room reflections would appear in measured result.

Background noise level in test environment.

Headset and Handset category			Speakerphone category		
	Octave Band Center Frequency (Hz)	Octave Band Level (dB SPL)		Octave Band Center Frequency (Hz)	Octave Band Level (dB SPL)
	63	49		63	40
	125	34		125	25
	250	29		250	20
	500	29		500	20
	1000	29		1000	20
	2000	29		2000	20
	4000	29		4000	20
	8000	29		8000	20
Equals average A-weighted noise level of approximately 29dB SPL			Equals average A-weighted noise level of approximately 20dB SPL		

Reflection-free conditions in test environment

Headset and Handset category	Speakerphone category
<p>Quiet test room with RT60 < 0.2 seconds.</p> <p>No big objects in 1m radius around HATS MRP (mouth reference point)</p>	<p>Free-field condition (anechoic condition) is must for all test beside real room test cases. The room must comply with free field conditions above 160 Hz as outlined in ITU-T P.341 recommendation A.3.1.1</p>
<p>Small objects (test fixtures etc.) must not influence the frequency response measurements more than +/-1 dB</p>	<p>Room has to be big enough to accommodate the device under test, test table and measurement transducers (or HATS). For Long range speakerphones this might mean that the device under test is a 50inch diagonal TV which is placed at 1m distance for testing.</p>

7.1.13 Reference TV-s for Long Range Speakerphone UI group testing.

Skype has selected the following TV-s as representative samples based on the sales volumes and price point.

During testing the default settings are used (as they are “out of box” or after factory settings reset).

Ref TV number	TV model	Screen Size
1	Sony KDL-32BX320	32’’
2	Samsung LN32D450G1D	32’’
3	Vizio M420NV	42’’
4	LG 42LK520	42’’

8.0 Appendix

8.1 References

- [1] ITU-T Recommendation [G.100.1](#): The use of the decibel and of relative levels in speech band telecommunications
- [2] ITU-T Recommendation [G.122](#): Influence of national systems on stability and talker echo in international connections
- [3] ITU-T Recommendation [G.131](#): Talker echo and its control
- [4] ITU-T Recommendations [P-sector](#).
- [5] ITU-T Recommendation [P.50](#): Artificial Voices
- [6] ITU-T Recommendation [P.51](#): Artificial Mouth
- [7] ITU-T Recommendation [P.57](#): Artificial Ears
- [8] ITU-T Recommendation [P.58](#): Head And Torso Simulator (HATS)
- [9] ITU-T Recommendation [P.64](#): Determination of sensitivity frequency characteristics of local telephone systems
- [10] ITU-T Recommendation [P.800](#): Methods for subjective determination of transmission quality
- [11] ITU-T Recommendation [P.862](#): Perceptual Evaluation of Speech Quality
- [12] ITU-T Recommendation [P.862.1](#): Mapping Function for Transforming P.862 raw result scores to MOS-LQO
- [13] ITU-T Recommendation [P.862.2](#): Wideband extension to Recommendation P.862
- [14] [IEEE Standard 269](#): Methods for measuring transmission performance of analog and digital telephone sets, handsets, and headsets
- [15] [IEEE Standard 1329](#): Method for measuring transmission performance of hands-free telephone sets
- [16] [Perceptual Evaluation of Speech Quality tool](#) (PESQ) that complies with ITU-T P.862 recommendation. Skype uses Opticom's version of PESQ that has been integrated into the HeadAcoustic ACQUA system
- [17] [ACQUA Advanced Communication Quality Analysis](#) system by HeadAcoustics
- [18] [Lync Logo program requirements](#)
- [19] [TIA standards for Wideband Telephony](#)
- [20] [HEAD Acoustics](#) – provider of ACQUA test software and measurement front-ends
- [21] [G.R.A.S](#) – provider of measurement microphones and artificial mouth