Microsoft Teams and Skype Audio Test Specification

For personal and conferencing audio devices
built for Microsoft Teams calling.

V4 April 2019

Table of Contents

1 Revision History

Revision	Published	Description
4.0	*4/2019*	Combined specification for Headset/Handset/Speakerphone and Conferencing devices. A single document now covers both RAW and Custom processing devices. Renamed “Group Speakerphone” to be “Center of Room Speakerphone”, and “Front of Table Speakerphone” to “Edge of Room Speakerphone” to better account for new devices that could be mounted at front or side of room or in ceiling. Testing to be conducted over Microsoft Teams call (Skype for Business will be tested in E2E sanity tests). Section 2.5 categorization of devices revised: greater granularity of meeting room devices which will align with how Teams Devices Marketplace recommends devices. 4.1.1 and 4.1.3 test shall be conducted with 3 microphone boom positions in case of headset with boom. 4.1.1 Send path - total quality loss – requirements changed. 4.1.2 Send path and 4.2.3 Receive path E2E latency requirement can be marked “for info only” if the DUT passes the HLK requirements in 3.2.2 and 3.2.3 Requirements relaxed. 4.1.3 - 4.1.7and 4.3.4 testcase requirements determine the Conference device room categorization 4.1.3 Send path – activation level in send direction. A **new test** based on ES 202 740 4.1.4– 4.1.5 Send path - signal level, requirements changed 4.1.6 Send path - idle channel SNR – requirement harder for some device types 4.1.7 Send path - active channel SpNR – requirement harder for some device types 4.1.10 Send path – frequency response, tolerance mask updated for Conferencing speakerphones 4.1.11 Send path – noise level with maximum microphone gain standard requirements added 4.2.1 Receive output level – test signal changed, and requirements modified for speakerphone UI-s 4.3.1 Echo path – terminal coupling loss changed from using TCLw to TCL. Requirements harder for some device types. 4.3.2 Echo path - EQUEST MOS at nominal playback volume – harder for some device types 4.3.3 Echo path – send signal attenuation during doubletalk – harder for some device types. Premium requirements added 4.3.4 Echo path – send signal attenuation during doubletalk – requirements changed for some device types 4.3.5– 4.3.7 **new tests** for RAW mode devices without DUT side audio processing 4.3.8 Echo path - EQUEST MOS at max playback volume – requirements harder for some device types 4.3.9 Echo path - Echo Control Characteristics (ECC) – Max playback volume – **new test** 4.3.10 Echo path – send signal attenuation during doubletalk – Max playback volume – **new test**

		4.4.1 Send path – send quality in the presence of ambient noise – personal devices – requirement harder for some device types. Hoth noise replaced with conference room noise from ETSI 103 224 4.4.2 Send path – send quality in the presence of ambient noise – conferencing devices – level of the background noises increased slightly. 4.4.1 Send path – send quality in the presence of ambient noise – conferencing devices – level of the background noises increased slightly. 4.4.2 Send path - speech quality for alternating near end talkers (conferencing devices only) – a new modification of a spec v3 test 4.4.5 Speech pickup quality and stability 4.4.3 Send path - speech level for alternating near end talkers – conferencing devices – requirement changed. Test separated from spec v3 test 4.4.5 Speech pickup quality and stability 4.4.6 Echo path - EQUEST MOS in reverberant room – requirements harder for some device types 4.4.7 Echo path – AEC convergence at call start – **new test** 4.4.8 – **new tests** for AEC stability in non-static usage scenarios in reverberant room. Minor test procedure corrections in public release vs. April release for this section. 4.4.12 Distractor attenuation for open office headsets – 5 angles used instead of 3, requirement for average and minimum defined. Section 5 restructured 5.3.1 Reverberant room – room reverb definitions and recommendations clarified. Test setups changed for some of the speakerphone UI subcategories in anechoic room to align with Microsoft Speech Platform device test setups. Some conferencing device test setups added.
3.0.1	*12/2016*	Section 3.1 Note added that latest HLK version matching the Windows version on DUT should be used as test tools. 4.3.1 Echo path – terminal coupling loss weighted (TCLw). Nominal speech level value changed from -18 to -24 as there was a dBm0/dBFS conversion error. Section 5. Audio test setup. RME Fireface UCX added as optional REF soundcard to the equipment list. Section 5.6 Reverberant room recommendations amended to increase overlap with Microsoft Speech Platform test room recommendations.
3.0	*11/2016*	Conferencing solutions separated into another specification. Note added about certification program not accepting USB speakerphones to be tested against the personal speakerphone category. Branding changed to Skype for Business certification. Subjective E2E tests are revised. Added the test to measure timestamp error and glitches: Windows 10 HLK – Communication Audio Fidelity. Changed the previously HCK based tests to be based on Windows 10 HLK and the measurement method is revised for latency tests. Also including effect querying. Two additional usage based scenarios are added to headset category: open office and outdoors. Added new requirement for open office headsets: Speech pickup quality for open office headsets. Section 5.1 is updated with test setup related details. Send signal attenuation during doubletalk tests added to Sections 4.3 and 4.4.

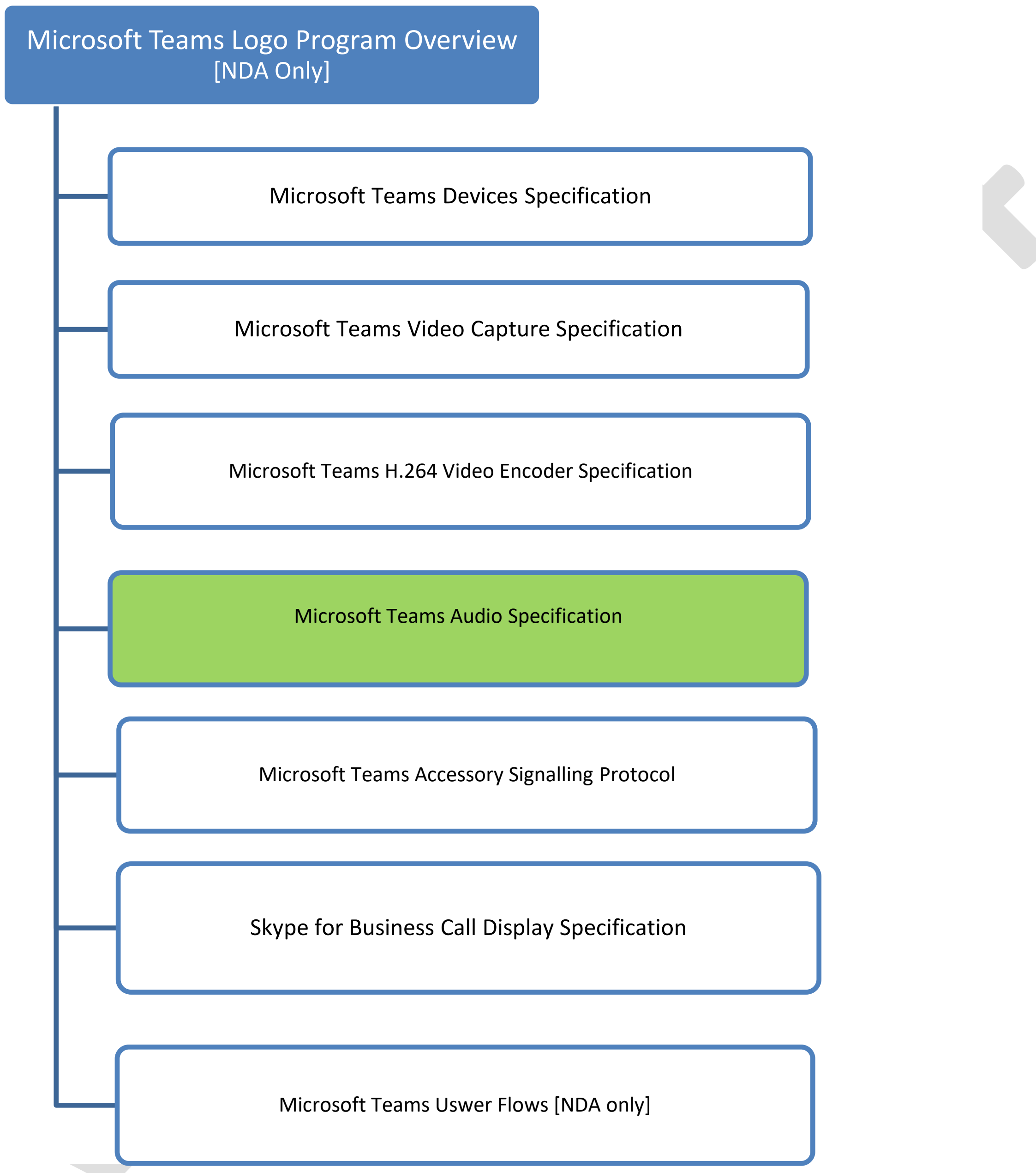
		Note added to Sidetone Masking Rating requirement about headsets with active noise cancelling in receive path.
		Requirement revised:
		• 3.1 E2E Scenarios: Audio Sanity tests. Section 3 is restructured respectively and recommendations about the speech pickup patters are moved to Appendix.
		• 4.1.9 Send path - distortion and noise.
		  ○ 5623-7079Hz band removed
		• 4.2.6 Receive path - distortion and noise
		  ○ -22dBFS requirements removed
		4.4.6 Echo path - ECC requirements modified for echo leakage.
		4.4.12 Distractor attenuation for open office headsets. A new test case targeted especially for headsets built/marketed for Open Office use.
		Section 5.6 is restructured, and details specified:
		• Use of HATS and artificial mouth described.
		• Open office headset test setup added.
		• Reference microphone position defined to be at MRP for headsets.

2.0	*9/2014*	Section 2.5.4 added to explain primary and secondary use cases and respective requirements. Custom processing usage description is amended in Section 2.5. 4.1.11 Send path – noise level with maximum gain Premium test case added. Error corrected in TCLw test description. 5.8 Reverberant room conditions clarified (reverb room limit values are in RT30).

2 Introduction

2.1 Overview

The family of documents supporting the Microsoft Teams Certification Program is shown below and contains detailed requirements that candidate devices being submitted to the *Certification Program* must meet.



The technical requirements listed in this document, the Microsoft Teams *Audio Test Specification*, have been derived solely for the purpose of maximizing interoperability and optimizing the quality of experience of devices used with Microsoft Teams and Skype for Business. Any use of this technical specification for platforms other than optimizing the quality for Microsoft Teams, Skype for Business, or Skype apps is not authorized.

Partners who license, develop, market, and/or sell Microsoft Teams devices that are qualified by Microsoft, are required to adhere to the specifications outlined in this document. Partners seeking changes, modifications and/or additions to this specification will be required to receive written approval from Microsoft before certification of the device. Microsoft 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 Microsoft Teams platform, new device categories, as well as performance improvements in the hardware used in peripheral devices.

2.2 Performance levels

This document provides performance requirements for Microsoft Teams certified devices.

There are two levels of device performance defined by this technical specification:

- **Standard:** Defines a good audio quality level that, when achieved, makes the device eligible for Microsoft Teams Hardware Certification.
  For other entry criteria besides device AV quality, please refer to the Program Overview document mentioned in Section 2.1.
- **Premium:** Provided as a design target. Offered as a guidance to achieve exceptional audio quality. Passing the premium requirements is not a mandatory for the Hardware Certification.

2.3 Additional References

Document Name	Version	Hyperlink
TIA-920.110-B Transmission Requirements for Wideband Digital Wireline Telephones with Handset	10/2015	http://www.tiaonline.org/standards/
TIA-920.120-B Transmission Requirements for Wideband Digital Wireline Speakerphone Telephones	02/2017	http://www.tiaonline.org/standards/
TIA-920.130-B Transmission Requirements for Wideband Digital Wireline Telephones with Headset	4/2018	http://www.tiaonline.org/standards/
ITU-T G.100.1 The use of the decibel and of relative levels in speechband telecommunications	11/2001	http://www.itu.int
ANSI S1.4 Specification for Sound Level Meters	R2006	http://www.ansi.org/
ITU-T P.51 Artificial mouth	08/1996	http://www.itu.int
ITU-T P.56 Objective measurement of active speech level	12/2011	http://www.itu.int
ITU-T P.57 artificial ears	04/2009	http://www.itu.int
ITU-T P.58 Head and torso simulator for telephonometry	08/1996	http://www.itu.int
ITU-T P.341 Transmission characteristics for wideband digital loudspeaking and hands-free telephony terminals	03/2011	http://www.itu.int
ITU-T P.501 Test signals for use in telephonometry	07/2012	http://www.itu.int
ITU-T P.863 Perceptual objective listening quality assessment (POLQA)	01/2011	http://www.itu.int
ITU-T P.502 Objective test methods for speech communication systems using complex test signals	05/2000	http://www.itu.int
ITU-T P.1120 Super-wideband and fullband hands-free communication in motor vehicles	03/2017	http://www.itu.int
ITU-T G.160 Voice enhancement devices	06/2012	http://www.itu.int
ETSI 126 131 / 132 Terminal acoustic characteristics for telephony; Requirements (3GPP TS 26.131 TS 26.132version 11.2.0 Release 11)	04/2013	ETSI
ETSI EG 202 396-1 Speech quality performance in the presence of background noise; Part 1: Background noise simulation technique and background noise database	09/2008	ETSI
ETSI TS 103 106 Speech quality performance in the presence of background noise: Background noise transmission for mobile terminals-objective test methods	03/2013	ETSI
IEEE 269-2010 Standard Method for Measuring Transmission Performance of Analog and Digital Telephone Sets, Handsets, and Headsets	2010	http://ieeexplore.ieee.org
IEEE 1329-2010 Standard Method for Measuring Transmission Performance of Hands-free Telephone Sets	2010	http://ieeexplore.ieee.org
Windows 10 Hardware Lab Kit	12/2011	MSDN
POLQA Application Guide		HA appnote
Echo Quality Evaluation of Speech in Telecommunications (EQUEST)		HA appnote
Microsoft Speech Platform (ie Cortana) performance recommendations		Specs/tools Released via Microsoft Collaborate Contact your microsoft account representative
Windows Audio Processing Object Architecture		Docs Microsoft
Windows Audio Signal Processing Modes		Docs Microsoft

Table 1: Additional references

2.4 Definitions

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.
ACQUA	ACQUA is a dual-channel analysis system developed by HEAD acoustics. The acronym stands for Advanced Communication Quality Analysis. http://www.head-acoustics.de/eng/telecom_acqua.htm
AEC	Acoustic Echo Cancellation
AGC	Automatic Gain Control. (This is not same as OSGC - Microphone gain adjustment visible in OS audio input mixer).
ANC	Active Noise Cancellation – generally means ambient noise cancellation in headset earpieces.
APO	Audio processing objects (APOs), provide software based digital signal processing for Windows audio streams. An APO is a COM host object that contains an algorithm that is written to provide a specific Digital Signal Processing (DSP) effect.
Artificial ear	A device used to measure the acoustic output of earphones, headsets, and handsets. 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.
CNG	Comfort Noise Generator provides a “replacement noise” to send path during periods where the acoustic echo cancellation is actively suppressing far end pickup and also microphone send noise gets attenuated.
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.
Critical distance	The microphone distance from near end talker mouth where the sound pressure level of the direct sound becomes equal to the reverberant sound. Placing microphone further than this limit will influence the call quality in negative way. Using directional microphones or microphone arrays can extend this limit but only if the room background noise is low.
dBm0	A tone that exercises maximum level has a power of 3.14dBm0 for A-law PCM G.711, 3.17dBm0 for μ-law PCM G.711 and L16-256. In this document, we adopt the μ-law PCM G.711 and L16-256 definition of dBm0 in the digital domain. Therefore, the relationship between dBm0 and dBov is as follows: $L_{\mu-law}(dBm0) = L_{ov}(dBov) + 6.18\ dB$
dBov/dBFS	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). 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).

	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.
dBPa	The sound pressure level, in decibels, of a sound is 20 times the logarithm to the base 10 of the ratio of the pressure of this sound to the reference pressure of 1 Pascal (Pa). Note: 1 Pa = 1 N/m2 (0 dBPa corresponds to 94 dBSPL).
dBSPL	The sound pressure level, in decibels, of a sound is 20 times the logarithm to the base 10 of the ratio of the pressure of this sound to the reference pressure of 2 X 10-5 N/m2 (0 dBPa corresponds to 94 dBSPL).
DGC	Digital Gain Control - send path level adjustment after Microsoft Teams pre-processing.
Double-talk	This is a condition when both the far-end participant and near-end participant are talking at the same time
DRP	Drum Reference Point
DSP	Digital Signal Processing
DUT	Device under testing.
DUT Editor REF Editor	Standalone applications to control Microsoft Teams client’s audio preprocessor settings.
ERP	Ear Reference Point
EQUEST	EQUEST is an objective analysis software to evaluate acoustic echo canceler performance. The acronym stands for Echo Quality of Speech in Telecommunications. http://head-acoustics.de/downloads/eng/application_notes/telecom/Appl_note_EQUEST_e0.pdf
Far-end	The far-end participant refers to the user who is talking to the user who is using the device under test (that is, the near-end user).
FFC	Free Field Correction
FFT	Fast Fourier Transform
HATS	Head And Torso Simulator. The HATS is usually equipped with an artificial mouth and two artificial ears.
HLK	Windows Hardware Lab Kit. It is designed to help to deliver systems, software and hardware products that are compatible with Windows and run reliably on Windows 10 for desktop editions (Home, Pro, and Enterprise), and the next version of Windows Server.
Lip Plane	Outer plane of the lip ring.
Lip Ring	The lip ring defines both the reference axis of the mouth and the Lip Plane.
Microphone pod	A table top microphone solution as a part of conferencing device that could house one or more microphone capsules in one enclosure.
MOS-LQO	Mean Opinion Score Listening Quality Objective.
MRP	Mouth Reference Point is a point 25mm in front of the lip plane of the artificial mouth. This is a point where the speech level is calibrated before measurements for normal, loud, and quiet speech.
MTR	Microsoft Teams Rooms
Near-end	The near-end participant refers to the user who is using the device under test (DUT), as compared with the far end participant they are talking with.

NS	Noise Suppression
OEM	Original Equipment Manufacturer
Offload mode	Microsoft Teams pre-processor blocks are disabled and the custom processing modules in hardware DSP or software are used instead.
OSGC	Microphone gain adjustment visible in OS audio input mixer.
POLQA	*Perceptual Objective Listening Quality Assessment*, also known as ITU-T Rec. P.863 is an ITU-T Standard that covers a model to predict speech quality by means of digital speech signal analysis.
RMS	*Root Mean Square*, also known as the **quadratic** mean, is a statistical measure of the magnitude of a varying quantity
SDNR	Signal to Distortion and Noise Ratio
SDR	Signal to Distractor Ratio. Term used for Open office headset tests to describe the level ratio of near end speech compared to nearby distractor speech.
SNR	Signal to Noise Ratio
SpNR	Speech to Noise Ratio
SRS	Skype Room System v2
STMR	Sidetone Masking Rating
TC	Terminal Coupling (dB) loudness of the acoustic echo signal from loudspeaker compared to a near end send signal in send path.
TCL	Terminal Coupling Loss (dB) loudness of the acoustic echo signal from loudspeaker after DUT echo cancellation processing compared to a near end send signal in send path.
TCLw	Weighted Terminal Coupling Loss (dB) - loudness of the acoustic echo signal from loudspeaker after DUT echo cancellation processing compared to a near end send signal in send path. Echo signal is weighted.
TS	Time Stamp
UI	User Interface
VAD	Voice Activity Detection

Table 2: Definitions

2.5 Classification of the products

In this document, context audio requirements fall into one of the following groups, according to the device's acoustic user interface (UI) type:

- Headset audio UI:
  - Monaural headset
  - Binaural headset
  - Open Office Headset (optional enhanced-performance category for headsets, which will receive special recognition in marketing content)
- Handset audio UI
- Speakerphone audio UI:
  - Handheld speakerphone
  - Personal speakerphone
    - Solutions enabling multiple acoustic user interfaces are described in Section 2.5.4
- Conferencing speakerphone audio UI

Audio requirements for each group are based on international standards and Microsoft own technical and use case studies.

If the solution is not described by any of the above categorization and does not fit with test setups in sections 5.2 and 5.3, then the test plan for a DUT is to be agreed with Microsoft prior to the testing.

This specification also applies to Microsoft Teams devices running other operating systems than Windows for PC's. The section 3.2 of the specifications can be skipped in such case.

2.5.1 Headset Audio UI Group

Headset audio UI products consist of two main components – one or two earpieces and a microphone (or microphones). 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.

This document categorizes the Headset audio UI group as follows:

Monaural headset

Binaural headset

Passing the additional requirements for **open office** is optional. Passing these requirements is pre-requisite for a product listing under the open office headset category on Microsoft webpages.

2.5.2 Handset Audio UI Group

Handset audio UI products 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 generally either a desk phone, or a mobile phone.

2.5.3 Personal 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. DSP mixer based solution offer flexibility to use separate microphone and speaker solutions that can be permanently installed in room and are thus modular in nature.

In general, the following applies to all speakerphone type devices:

- The user's speech level decreases the farther the user is from the microphone. The influence of room reflections and reverberation on the speech signal increase the farther the user is from the microphone. It is especially relevant for conferencing solutions as the mouth to microphone distances above 2m start to cause speech quality drop due to room reverb influencing the near end speech pickup more and more.
- 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.

The **Personal solution** use case products are classified as following:

- **Deskphone** - to be considered a desk phone and allowed to be tested at personal distance rather than conferencing distance, the phone must have design factors that make it non-portable and clearly intended for arms-length use, typically this includes non-portable size, power and/or ethernet connections along dialpad and handset. It must not be marketed or otherwise recommended for use in conference rooms. If in doubt, the longer distance requirements for conferencing devices will apply.

Note: as of April 2019, handsets and tablet devices are not an active certification category and PC category is going through changes that are beyond the scope of this specification. Partners should contact Microsoft for the latest status on the following categories prior to submitting for certification.

- **Handheld Speakerphone** – speakerphone UI device where mouth to microphone distance up to 0.3m. This category includes mainly handsets in hands free mode and tablet type devices.
- **Personal (integrated) Speakerphone** – speakerphone UI device where mouth to microphone distance up to 1m. This category includes mainly laptops, all-in-one PCs where the speakerphone is integrated into a personal computing product designed and marketed for personal use, and tablets with a built-in stand. Microphone component of USB webcams may be considered in this category if the camera is not intended for conference room use.

All accessory speakerphones (e.g. USB speakerphones) and other devices which customers will likely use for usage exceeding 1m **are to be tested as conferencing devices.**

2.5.4 Personal solutions that enable multiple UIs

In case a personal solution enables multiple audio UIs then the use cases are prioritized based on how Microsoft Teams is used for calling. The main use case(s) must pass the whole specification whereas all secondary one(s) need to pass a subset of testcases listed below:

Requirements necessary to be tested with the secondary use case:

- 4.1.1 Send path - total quality loss
- 4.1.4 Send path - signal level with normal speech
- 4.1.6 Send path - idle channel SpNR
- 4.1.8 Send path - single frequency interference

- 4.2.1 Receive path - output level
- 4.3.1 Echo path - terminal coupling loss (TCL)
- 4.3.2 Echo path - EQUEST MOS at nominal playback volume
- 4.4.5 Echo path - EQUEST MOS in reverberant room
- 4.4.7 - 4.4.8 Echo path – convergence and variable echo path test

There can be multiple main UIs in following cases:

- If the main scenario is different for audio and video calls, then both must be tested against the whole specification.
- If there are multiple equally valid main scenarios, then all scenarios must be tested unless Microsoft agrees to a testing optimization based on the hardest scenario.

General guidance for choosing the main use case and the respective audio UI(s):

- Desk phones:
  - Handset UI
  - Personal speakerphone UI

  If both modes are supported, then full tests flow I shall be conducted for both modes. On partner request Desk phone type of devices can be tested against the Phone room conferencing device requirements.

- Mobile handsets:
  - Audio only call scenarios
    - Main use case for audio calling scenario: handset audio UI
    - Secondary use case for audio calling scenario*: personal speakerphone audio UI (device flat on table). Video can be disabled for this scenario.*

      Note! Only if the audio processing is customized for this use case.
    - Secondary use case for audio calling scenario*: headset UI.*

      Note! Only in case the product is sold with a headset bundled.
  - Video call scenarios
    - Main use case for video calling scenario: handheld speakerphone UI *(user facing the display and using the user facing camera).*
    - Secondary use case for video calling scenario: handheld speakerphone UI *(user facing the display and using back camera).*

      Note! In subjective evaluations also, a scenario where a second person is talking at the back of the display in addition to person holding the phone is checked.
- Tablet PCs without integrated stand:
  - Main calling scenario: handheld speakerphone UI *(user facing the display and using the user facing camera).*
  - Secondary use case: handheld speakerphone UI *(user facing the display and using back camera).*
  - Secondary use case*: headset UI.*

    Note! Only in case the product is sold with a headset bundled.
  - Secondary use case*: personal speakerphone audio UI (device flat on table). Video can be disabled for this scenario.*

    Note! Only if the audio processing is customized for this use case.
- Tablet PCs with integrated stand:
  - Main calling scenario: personal speakerphone UI.
  - Secondary use case: handheld speakerphone UI.
  - Secondary use case for audio calling scenario*: headset UI*

Note! Only in case the product is sold with a headset bundled.

  - Secondary use case for audio calling scenario*: personal speakerphone audio UI (device flat on table). Video can be disabled for this scenario.*

    Note! Only if the audio processing is customized for this use case.

- Detachable type of PC:
  - Main calling scenario: personal speakerphone UI.
  - Secondary use case: handheld speakerphone UI.
- Convertible type of PC:
  - Main calling scenario: personal speakerphone UI (keyboard in front of the user).
  - Secondary use case: personal speakerphone UI (tent mode - keyboard behind the display).

If vendor has documented additional usage scenarios, then all those must meet the requirements for secondary use cases.

Vendor's recommendations provided before testing are considered for prioritizing the use cases. Without such recommendations a tester can choose the main use cases (including device orientation).

2.5.5 Conferencing solutions category devices

The conferencing category of devices are built as a conferencing solution for a group of people. Generally, the devices under this category are built to be installed in meeting rooms and enable or support the Microsoft Teams Rooms client. Some smaller devices could also allow to be carried along for adhoc meetings or used as a collaboration device in an open neighborhood office area. The indication of group size versus the meeting room size in given in below table.

Room type	Number of people
Phone room	Up to 2 people
Focus room	Up to 5 people
Collaboration: huddle space (edge of room, half circle layout, office ambient noise)	Up to 5 people
Collaboration: lounge space (edge of room + table, half circle layout, office ambient noise)	Up to 6 people
Collaboration: meeting room (edge of room, half circle layout, meeting room ambient noise)	Up to 5 people
Small meeting room	Up to 7 people
Medium meeting room[1]	Up to 11 people
Large meeting room[1]	Up to 18 people
Extra-large meeting room[1]	Up to 30 people

Table 3: Microsoft Teams Rooms target room types

Devices are classified as following:

- **Center of room speakerphone** – a tabletop conference room device where the main unit with microphone(s) and speaker(s) is typically placed on the conference room table. The devices in this group could also have extra extension microphones that help extend the total microphone pickup range of the solution. Sub-category of Center of room speakerphones are linked speakerphones where the main unit like units are daisy chained. As each such unit will have both microphone and speaker these units should follow the multiple mic and multiple speaker distances outlined in Table 4.

[1] DUT devices or their supporting equipment (like network switches etc.) should not use active cooling fans. If they do, an extra focus should be put on creating a very quiet cooling solution. If loud or variable noise is present during the E2E tests (3.1) then that can be a blocking failure even if objective tests pass.

- **Edge of room speakerphone** – main unit with microphone(s) and loudspeaker(s) (and possibly a camera) are co-located with a large screen at one end of the conference room or on a side wall. These devices can be grouped as
  - USB speakerphones to enable AV capabilities for existing big screens. The devices in this group could also have extra extension microphones that help extend the total microphone pickup range of the solution.
  - Audio calling and touch enabled big screen devices. These devices can be in two subgroups
    - The front or side of room big screen with permanent placement on wall with built in microphone(s) and speaker(s). These devices fall under Collaboration : Meeting Room
    - Free moving big screen with built in microphone(s) and speaker(s). These devices allow use in conference rooms, collaboration meeting rooms, huddle spaces or lounge spaces. The latter two use cases need a different configuration for microphone and speaker use as there is a need to focus the send and receive audio to a smaller area in front of the big screen, while attenuating the surrounding disturbances.
  - **Soundbar** type of devices with built in microphone(s) and speaker(s) fall into this category. These are audio accessory devices that are mounted close to the front of room display or projector screen. The devices in this group could also have extra extension microphones that help extend the total microphone pickup range of the solution.
- **Modular conferencing devices** – some conference devices are built with microphones and speakers as separate modules. These devices can allow a mix of Center of room and Edge of room placements. As an example, a solution with several extension microphones on table and speakers wall mounted to the front wall would fall under this category.
- **DSP-mixer based solutions** – these are flexible audio-conferencing solution supporting USB audio interface to Microsoft Teams Rooms PC. DSP mixer devices enable multiple audio inputs and outputs and a customizable audio processing DSP. These devices allow tailoring the solution to the room by picking the best suited microphone and speaker solutions as well as tuning the sound and echo cancellation to the specifics of the conference room in question. Please refer to Section 6.1 for additional details.

Due to the complexity and variability of possible devices for each of the listed room types the actual audio testing uses the following usage scenario-based criteria to define the requirements for each device type:

a) Furthest near end talker to microphone distance
   - Users up to 1.5m radius around the microphone (small portable speakerphones)
   - Users up to 2.3m radius around microphone (extension microphones and center of room speakerphones)
   - Users up to 3.5m in audio/video pickup area (Edge of room microphone array)
   - Users up to 4.5m in audio/video pickup area (Edge of room microphone array)

b) Furthest near end listener to speaker distance

c) Microphone to speaker distance (only for modular devices where the microphone and speaker are not housed in same physical enclosure)

Real conference rooms always have room reverberation and background noise. This does influences the Critical Distance for microphone pick up. This is a distance for a given room where the sound pressure level of direct sound and the reverberant sound become equal. Placing a simple omni microphone further than this limit will influence the call quality in negative way due to worse intelligibility. While it is possible to meet the speech level and speech to noise ratio test for 3.5m and 4.5m distances in anechoic conditions it does not mean this would guarantee a good call quality in real conference rooms. Thus, a single device with 360-degree pickup can cover areas of up to 2.3m radius. If the device allows extension microphones, then each additional microphone can extend the pickup area by extra 2.3m depending on the test result.

The following table outlines the suitable solutions and distances for each type of a meeting room while trying to comply with optimal Critical Distance for the microphone.

Room type[2]	**Distance from furthest user to MICROPHONE**		**Distance from SPEAKER to furthest user**	
	Microphone(s) center of room	Microphone(s) at edge of room	Speaker(s) on table	Speaker(s) at edge of room
Phone room (2m*2m)	**1.5m**	**2.3m**	**1.5m**	**2.3m**
Focus room (3m*3m)	**2.3m**	**3.5m** must have directional microphone pickup	**2.3m**	**3.5m**
Small meeting room (3m*4.5m or 4.5*4.5 if square)	**2.3m**	**3.5m** must have directional microphone pickup	**2.3m**	**3.5m**
Medium meeting room (4.5m*6m)	**On Table - 2.3m** must have a main device/mic with 1 or more extension microphone support or have min two 2.3m mics for long table shapes **Ceiling mount - 3.5m** Directional microphone	**2.3M or 3.5m array microphone (directional) + 1 extension** Or **4.5m array microphone (directional)**	**2.3m** (if min 2 speakers are bundled including cases of linked speakerphones) **3.5m**	**4.5m**
Large meeting room (4.5m*8.5m)	**On Table 2.3m – main unit** must have a main device/mic with 2 or more extension microphone support or have min three 2.3m mics for long table shapes **Ceiling mount** 2 or 3 directional ceiling panel microphones to cover the area with overlap	**3.5m or 4.5m (directional) + 2 extension**	**2.3m or 3.5m** (if min 2 speakers are bundled) **4.5m** (if 1 speaker is bundled)	**2.3m or 3.5m** or 4.5m– for ceiling speakers if ≥2 are bundled **7.5m** - if speaker at front of room
Extra-large meeting room (6m*10m)	**On Table 2.3m** DSP mixer with 4 or more extension microphone support or device with min four 2.3m mics for long table shapes **Ceiling mount** 3 or more directional ceiling panel	**NA**	**2.3m** (if min ≥3 speakers are bundled) **3.5m or 4.5m** (if 2 speakers are bundled)	**2.3m** (if ≥4 speakers are bundled) **3.5m or 4.5m** (if 2 speakers are bundled)

[2] DUT devices or their supporting equipment (like network switches etc.) should not use active cooling fans. If they do, an extra focus should be put on creating a very quiet cooling solution. If loud or variable noise is present during the E2E tests (3.1) then that can be a blocking failure even if objective tests pass.

Room type[2]	Distance from furthest user to MICROPHONE		Distance from SPEAKER to furthest user	
	microphones to cover the area with overlap			
Collaboration: huddle space[3] (2.3m radius w/ office ambient noise)	**N/A**	**2.3m**	**NA**	**2.3m**
Collaboration: lounge space (3.5m radius)	**2.3m**	**3.5m** must have directional microphone pickup	**2.3m**	**3.5m**
Collaboration: meeting room[3] (2.3m radius w/ meeting room noise)	**NA**	**2.3m**	**NA**	**2.3m**

Table 4: Conferencing devices microphone and speaker distances for room types

Note! All the ceiling mounted microphones without in built DSP functionality and speaker shall be tested against the 2.3m usage distance requirements as default. It is assumed that enough such microphones and speakers will be installed in a meeting room, so that the talker to mic and speaker distance does not exceed the 2.3m.
If the ceiling microphone provides DSP capabilities and directional microphone pickup in sectors of the 360 degree to allow greater distances to near end talkers, then the test condition of 3.5m will be used for Medium and Large room categorization.

2.5.6 Determining the test positions for Conferencing devices

Previous section defined the rooms and microphone speaker distances. This section gives guidance on choosing the test positions and level requirements.

Step 1: define hardest usage position for your device. Hardest means the longest distance from microphone to furthest user for Send path testing and longest distance from speaker to furthers user for Receive path testing.

Example device A: "puck style" speakerphone. Microphone and speaker are at same location, center of table.

- For this type of device the microphone and speaker distance to furthest user are equal, thus depending on the target room either Focus/Small room distance of 1.5m will be used or if the device is aimed for Medium sized rooms as well then the 2.3m distance will be used for both send and receive path.

Example device B: speaker bar with built in microphone. Microphone and speaker are at same location, front of room position.

- For this type of device the microphone and speaker are again at same distance from farthest user, so based on the target room, the test distances would be 3.5m for Focus and Small room and 4.5m for the Medium size room. Large room support would not be possible without a bundled extension microphone to be placed on meeting room table.

  Note that this is one of the hardest scenarios as microphone and speaker to user distances are long, so near end speech is quieter while playback needs to be louder. Due to closeness of microphone and

[3] The key difference between huddle and meeting room space for collaboration device is the type of background noise. See testcase 4.4.1 The collaboration devices must also be tested against relevant meeting room scenario.

speaker this creates a very challenging acoustic echo scenario, so passing the Echo path test, especially doubletalk tests, will be very hard.

Example device C: speaker bar with separate table-mounted microphone. Speaker is front of room position, microphone is center of table.

- For this type of device, the microphone and speaker to furthest user distances differ. For microphone the distance will be 1.5m for Focus and Small room target or 2.3m for Medium room target. For speaker the distances would be 3.5m for Focus and Small room and 4.5m for Medium room target.

The "hardest requirements rule": for devices that support multiple deployment options. If a device supports multiple deployment options, then the device must successfully pass requirements at all of the hardest deployment option. In some cases, this means multiple positions must be tested.

- Example B above (mic and speaker in soundbar). If the product offers the option of an extension microphone for table top use, but the extension is optional and not bundled with the sales package, then the hardest distances above are still valid for main unit. The extension microphone can be tested with 2.3m distance and if tests are passed the main unit with extension can extend the target room from Focus/Small to Medium rooms for example. But there will be an according note in the product pages and the listing will indicate the need of the optional microphone extension package.
- Example device D: if deployment options for the device allow front of room fixed installation or the neighborhood collaboration use in open office space then the hardest requirement rule means that front of room distance should be used for a target room size, testing must be performed in both neighborhood and conference room positions. Such a device must perform well in conflicting challenges:
  - neighborhood use: unwanted near end sounds around the device need to be suppressed
  - meeting room use: will need a good pickup from far distances unless an extension microphone is provided with the product.
  - it is recommended to have a “device usage mode” as a setting that can be changed based on the deployment scenario. This will allow tuning the audio performance to a specific scenario. This usage mode should be configurable (and queryable) remotely to enable managed room and support scenarios.

Step 2: record all supported the microphone, speaker and camera usage distances for the device in the “device submission template” and identify the hardest position(s) for testing.

Position Description	Microphone distance	Speaker distance
Front of sound bar speaker	-	3.5M
table mounted microphone	2.3M	-
Front of room mounted microphone	4.5M	-
<other supported usage positions>	…	…
Hardest distance (max)	**4.5M**	**3.5M**

Table 5 Example table for modular speakerphone targeting mid-sized room performance

Step 3: run lab tests (anechoic chamber, reverberant room, camera tests) with the distances defined according to hardest position distance(s) in step 2. Subjective tests will be performed in all supported positions.

Conferencing speakerphones will be assigned a supported room size as additional designation for the certification. The supported room size will be determined based on the combination of subjective tests at the usage ranges defined in the table above and the objective tests listed below. These will have different test levels and or requirements for each room size and/or position.

4.1.4 Send path - signal level with normal speech

4.1.5 Send path - signal level with quiet speech

4.1.6 Send path - idle channel SpNR

4.1.7 Send path - active channel SpNR

4.2.1 Receive path - output level

[multiple sections] Echo path - Echo Control Characteristics (ECC)

4.3.4 Echo path – send signal attenuation during doubletalk

2.6 Audio processing paths

2.6.1 Audio capture and render APO pipelines

In case of in-built audio on PC-s running Windows 10, applications can choose the Audio Signal Processing Mode. For example: Raw, Default, Media, Speech, Communication etc. It is possible to define a different set of available Audio Effects for each of the audio processing modes (SFX – Stream Effect, MFX - Mode Effect, EFX - Endpoint Effect)

Endpoint Effect (EFX) are applied to all streams that use the same endpoint. An endpoint effect is always applied, even to RAW streams. The endpoint effects should be used with care as for applications like Microsoft Teams, Skype for Business, Skype and Cortana these effects will be in the acoustic echo path and thus can severely influence the linearity of echo path. This can cause echo leaks and poor user experience. Some effects that should be placed in the endpoint area are speaker protection and speaker compensation.

	Allowed as an endpoint effect	Comment
Acoustic Echo Cancellation (AEC)	no	
Noise Suppression (NS)	no	
Automatic Gain Control (AGC)	no	
Beam Forming (BF)	no	
Constant Tone Removal	no	
Equalizer	**yes**	only a static type of EQ is allowed
Loudness Equalizer	no	
Bass Boost	no	
Virtual Surround	no	
Virtual Headphones	no	
Speaker Fill	no	
Room Correction	no	
Bass Management	no	
Environmental Effects	no	
Speaker Protection	**yes**	should only engage at loud volume and only for signal peaks to avoid influece on echo path linearity

Speaker Compensation	no	
Dynamic Range Compression	no	

Table 6: Device APO-s that are allowed as always on endpoint effects

As required on the Windows 10 Hardware Guidelines page above the "raw mode" must bypass all of the MFX and SFX type of effects.

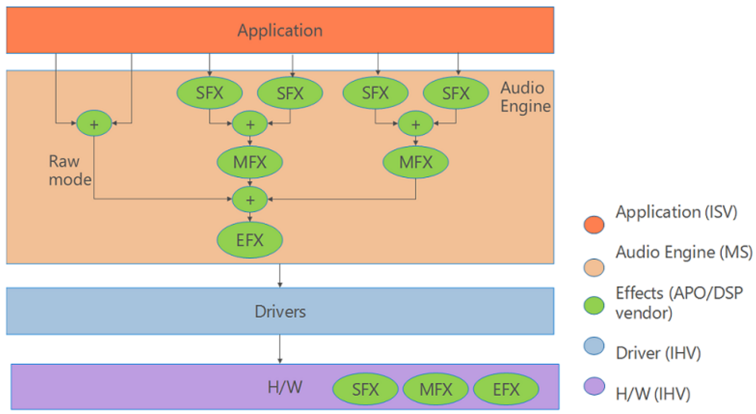


Figure 1: Example of Windows 10 audio processing modes

In case of USB or wireless connected accessories there is no possibility to turn the audio effects on or off from GUI. So, all effects included in device hardware are 'always on'. This means that USB devices will not have an option for testing in Raw mode. It is important to validate the performance of these effects against all use cases of the device. A good and passing performance of an echo cancellation against the Microsoft Teams specification does not warrant a good performance with various Speech recognition solutions like Microsoft Speech Platform and Cortana.

It is recommended (optional) that the USB devices would implement the USB v3.0 specification Audio Function Topology reporting as described Audio Device Class Definition document Section 3.13.9.3 Multi-Function Processing Unit. Correct info in this descriptor can help the operation of future Teams or Skype clients. Teams or Skype will potentially use this info to better tune the audio processing in the client.

The following speakerphone and conference phone relevant descriptors are defined in USB v3.0 specification:

- Algorithm Undefined
- Beam Forming Algorithm
- Acoustic Echo Cancellation Algorithm
- Active Noise Cancellation Algorithm
- Blind Source Separation Algorithm
- Noise Suppression/Reduction

2.6.2 Audio processing path options for DUT devices

In case of in-built audio device in "raw mode", any custom digital signal processing SFX and MFX are bypassed and Skype processing is used. This is the mode Microsoft Teams and Skype for Business clients will use by default.

For this specification, the 'Communications' and 'Default' audio modes are of interest. Via these audio processing modes Skype and Teams can use the device's custom audio processing effects in Skype calls. We will refer to such enabling of the use of 'Communication' and/or 'Default' modes as Offloading.

Device type	Audio modes available for Logo approval	Minimum effects	Identifiers needed as part of test report
Headset without hardware AEC audio processing	Teams processing	NA	USB PID/VID
Headset with hardware AEC and other non-linear processing (VAD/ANC/Beamforming)	Cascaded Teams processing	NA	USB PID/VID[4]
PC-s with embedded audio device without AEC, NS	RAW + Teams processing	NA	Machine Model Machine Type Audio device SSID Vendor ID, Product ID
PC-s with embedded audio device with 3rd party AEC, NS etc.	Hardware/Software offloading[5]	AEC, NS	Machine Model Machine Type Audio device SSID List of certified effects combination
Conferencing devices with Room System capability built in	Hardware/Software offloading	AEC, NS	
Blutooth connected mobile phone docking stations	Pending client support for offload on Android and iOS		
USB connected Handsets or Desk-phones	Hardware offloading (must pass offload in all audio modes, e.g., speakerphone, handset)	AEC, NS	USB PID/VID
Desk-phones with embedded Teams client	Hardware offloading / Cascaded Teams processing	AEC, NS	Custom client build for device
Wireless connected speakerphone devices with USB dongle	Hardware offloading	AEC, NS	USB PID/VID (unique PID required)
Conferencing devices USB connected	Hardware offloading	AEC, NS	USB PID/VID

Table 7: Device APO offloading options

[4] As a standard practice, headsets will not be added to offload processing whitelist since cascaded processing generally doesn't harm headset performance.
[5] As the audio driver can allow user to turn the 3rd party effects on and off the SfB or Teams client will check if the audio effects approved are also enabled at a call start. If not, the client will revert to using client in built processing.

Skype will enable the offloading only on products verified to provide sufficient quality. The logo specification testing is the most important input to this decision, but other checks can be included depending on the product group and complexity of the product. The enabling of 3rd party APO processing can be rejected or revoked later in some rare cases. Some examples listed below:

- Custom processing gives lower score in user Call Quality Feedback ratings calculated over significant population.
- Custom processing enables only wideband audio (16kHz sampling) whereas in the future MS Teams processing can may enable super wideband audio (24kHz or 36kHz sampling) on the same device.
- Custom processing causes echo leakage or functions in an unreliable manner in some usage scenarios. For example, in noisy or reverberant environments or scenarios where user changes from one acoustic UI scenario to another during the call (example: switching from docked to handheld or from front camera to back camera).
- Future version of the custom processing APO-s (driver or firmware updates) distributed to users show a degraded performance in Call Quality Feedback ratings calculated over significant population.

3 Entry criteria for audio testing

A solution submitted to testing must meet some essential requirements to enable the testing.

If any of the tested items in this chapter fails, then the tester has right to stop testing any further and the test will be considered completed (and the test fees will not be refunded).

If device is a wireless device which supports using while charging, the tests in this section should be performed both in normal (wireless) mode, and in charging mode (device should have the battery discharged prior to conducting the tests in charging mode so that active charging is occurring during the sanity tests).

3.1 E2E Scenarios: Audio Sanity tests

Besides objective tests a subjective performance evaluation is also necessary. The tests described in below section are targeted mainly to be conducted in an intended usage scenarios and environments and in conditions which are not covered by the objective test specification but are expected in everyday use. Many of the subjective test descriptions are similar to ITU-T P.1120 specification Annex C. This specification is a recommended reading for subjective tests and includes a sample questionnaire.

All devices shall be evaluated over Microsoft Teams and Skype for Business call at each engineering stage (e.g. by the OEM at major internal milestones, and again by the test lab prior to beginning the quantitative testing). The goal is to identify any obvious problems before beginning the quantitative testing. The DUT client is to be configured as defined in Section 5.1.4.

3.1.1 Tests environment recommendations for end to end audio test

The evaluation should be performed under normal usage condition and some mild stress conditions. All tests below require two people, and few tests require an additional assisting tester.

Normal usage conditions for personal devices:

- Moderately sized room, typically at least 10 square meters with reverberation time RT60 of 0.4 to 0.8 second The 'Reverberant Room' test room used for Section 4.4 tests is equally suitable for use as a test room for this test.
- Room ambient noise level: 30 to 34dBSPL(A).
    - If the noise level in room is below the above suggestion, then additional noise can be generated by an active speaker pointed to corner of the room playing ETSI 103 224 "Conference1_handsfree.wav" noise test file at a low level. The active speakers used should be able to produce a frequency range of 150Hz to 10kHz as a minimum.
- Receive volume control is set to a comfortable listening level.
- Near end user speech at normal level, typically 62-65 dBSPL(C) at 50 cm.

Noisy usage conditions for personal solutions:

- Room ambient noise: 55 to 58d8SPL(A) at DUT microphone position. The noise can be generated by an active speaker pointed to corner of the room playing ETSI 103 224 "Callcenter2 handsfree.wav" noise test file. The active speakers used should be able to produce a frequency range of 150Hz to 10kHz as a minimum. Receive volume control is set to a comfortable listening level.
- Near end user speech at slightly raised level, typically 65-68 dBSPL(C) at 50 cm.

Normal usage conditions for conferencing solutions:

- Similar conference room as intended for the DUT usage scenario (Small/Medium/Large) with reverberation time RT60 of 0.4 to 1 seconds. Room ambient noise level: 30 to 34dBSPL(A).

  - If the noise level in room is below the above suggestion, then the same two speaker setup suggested for noisy usage conditions can be used to generate a low level of Hoth noise into the room to meet the criteria
- Receive volume control is set to a comfortable listening level.
- Near end user speech at normal level, typically 62-65 dBSPL(A) at 50cm.

Noisy usage conditions for conferencing solutions:

- Room ambient noise: 38 to 42dBSPL(A) at DUT microphone position.
  - The noise can be generated by an active speakers pointed to the corner of the room playing ETSI 103 224 "Conference3_handsfree.wav" noise test file. The active speakers used should be able to produce a frequency range of 150Hz to 10kHz as a minimum.
- Receive volume control is set to a comfortable listening level.
- Near end user speech at slightly raised level, typically 65-68 dBSPL(C) at 50 cm.
- As part of the test an additional single noise source shall be tested as well. This could be a real projector or a laptop fan close to the DUT or another active speaker simulating the projector noise. The single source noise should be 44 to 48dBSPL(A) at DUT microphone

Far-end user(s) setup conditions:

- The far-end talker is in a quiet office environment with a high-quality reference corded binaural headset. List of suitable corded headsets can be found here.

DUT setup conditions:

- Review the DUT user manual for the usage scenarios that are recommended by the manufacturer. The near-end talker uses the product in a real meeting room/office as instructed in the user manual. If there isn't any manufacturer's specification, then similar positioning should be used as described in Section 5.3. If the DUT is a capture-only device (e.g. webcam), then desktop speakers shall be used as the render device.
- Ensure that the default settings of the DUT are used with exception of playback loudness that is adjusted by a tester to a comfortable/preferred listening level (if possible then the used playback level should be noted down for later reference).
- In case a DUT includes a camera or comes bundled with a camera, then a two-way video call shall be active during tests.
- All embedded wireless technologies shall be enabled during the test (e.g. WiFi, Bluetooth).

Test Score definitions scale:

Test scoring will be done on a scale of 1 to 5, with a score of 1 being BAD and 5 being GOOD
5: No detectable flaws in the observed metric (Pass)
4: Some minor flaws detectable by an observant user (Pass)
3: Flaws detectable to the casual user (Possible Fail further review required)
2: Serious flaws making the call difficult to continue (Fail)
1: Very serious flaws preventing the completion of the call (Fail)

3.1.2 Tests on all audio UI categories

This section contains the requirements and end to end assessments for all audio UI categories. If there are category specific requirements, then these are defined in other sections.

3.1.2.1 *Verifications*

All items defined below must be met (and the test score is not required):

- If the DUT enables changing acoustic UI then AGC and AEC must adapt within 5 seconds. For example: connecting/removing an analog headset or extension unit, switching between handset/speakerphone modes and switching between front/rear camera, switching a convertible laptop from primary to secondary position.
- If device supports multi-connectivity (e.g., corded + Bluetooth to cell phone) device must be in multi-connectivity mode. Occasionally throughout the tests below user should trigger audio from the cell phone or secondary paired device (e.g., play multi-media or system sounds). Device shall prioritize call audio and not suffer from glitches, gaps or bandwidth/volume shifts for the voice call.

3.1.2.2 *Assessments in quiet environment*

With the DUT in normal usage environment, place an E2E Microsoft Teams call between the two users. Enable video if DUT solution includes a camera.

DUT user assesses the following areas on the scale from 1 to 5:

- DUT Receive speech quality
  - Rate the quality of DUT receive signal at preferred listening level: is there any low level distortion, rattles and vibrations and/or moderate band limitation effects
  - Is the speech/loud speech distorted at DUT maximum volume settings?
  - Do you hear electric interferences (such as buzzing sound)?
  - Do you hear annoyingly high noise floor in speaker signal or noise fluctuations?
  - Speak simultaneously. Is the communication fluent? Can the DUT user interrupt the far-end speaker when the latter is speaking?
  - Set the DUT volume at maximum. Speak simultaneously. Can the DUT user interrupt the far-end speaker when the latter is speaking?
  - Rate the overall impression.

Far end user assesses the following areas on the scale from 1 to 5:

- DUT send speech quality
  - Speech level fluctuations - were there drop outs, cut offs, missing words or syllables, heavily chopped voice. What if DUT user speaks quietly?
  - Speech quality/speech naturalness - was there synthetic voice and/or heavy distortion and/or higher degree of band-limitation
  - Intelligibility/listening-effort - were there words that were hard to understand, considerable effort required.
  - Signal-to-noise ratio – were there times with high noise level, almost same level as speech, clearly disturbing, but call would be continued
  - Signal-to-noise ratio – were there clearly audible send noise fluctuations?
  - Background noise quality – was the background noise in call clearly unnatural/distorted/synthetic sounding, or many artefacts/clicks/plops, or often variation in sound and level
  - Adaptation to background noise – did the adaptation take 3 … 10 seconds or >10 seconds
  - Rate quality of speech: is it too dark, too bright, metallic? Is the dynamics healthy? - If necessary, re-evaluate in the end of the call.
  - Do you hear electric interferences (such as buzzing sound)? How annoying is it?
  - Does the DUT mute function cause interferences?
  - Do you hear disturbing handling noise when DUT user touches the devices (to change volume etc.)?

  - In case of DUT with keyboard or touchscreen have DUT user type in some notes. Evaluate if and how disturbing is the microphone pickup of the keyboard or touch sounds.

  Design note: devices should avoid placing microphone near sources of noise such as fans, keyboard, touchpad etc.

- When the DUT is used in various positions and holding manners: Do you find some additional audible problems not covered by the previous test cases (for example microphone gets blocked in some usage condition)?
- DUT echo cancellation performance:
  - When the call starts, can you hear the echo of your own voice? Does the echo disappear after <5 seconds (preferably <2sec)?
  - Frequency of echo occurrence – infrequently several times/echo occurs more often than not / Permanent
  - Echo intelligibility - hardly recognizable as voice/distorted voice/slightly distorted voice/clear voice
  - Comfort noise quality
    - Clear difference between comfort noise and natural background noise perceivable; Clearly unnatural/distorted/synthetic sounding, or many artefacts/clicks/plops, or often variation in sound and level
    - Comfort noise does not sound like the natural background noise at all; Very unnatural/distorted/synthetic sound, or permanent artefacts/clicks/plops, or permanent variations in sound and level, very uncomfortable to listen to.
  - Speech level variation during double talk
    - Many drop outs, cut-offs, missing words or syllables, heavily chopped voice, or high constant attenuation being switched during the double talk phases
    - Not possible to hear the other side at all during double talk
  - Intelligibility/listening-effort during double talk
    - Some words were hard to understand during double talk, moderate listening-effort is required
    - Many words were hard to understand during double talk, considerable listening-effort is required
    - No meaning understood with any feasible effort during double talk
  - Echo canceller stability
    - When DUT is used so that echo path changes. For example: DUT user is adjusting the playback volume, bending over the device, moving hands on keyboard etc.
      - Some echo can be heard, but disappears very quickly
      - The echo disappears slowly, the recurrences are audible for a few seconds
      - The echo disappears very slowly, the recurrences are audible for more than 10 seconds
      - The echo builds up like in an unstable feedback system
    - When DUT volume is at maximum, can you hear the echo of your own voice when you speak in a loud voice (e.g. numbers)?
      - Some echo can be heard, but disappears very quickly
      - The echo disappears slowly, the recurrences are audible for a few seconds
      - The echo disappears very slowly, the recurrences are audible for more than 10 seconds
      - The echo builds up like in an unstable feedback system
- Rate the overall impression.

3.1.2.3 *Assessments in noisy environment*

With the DUT in noisy usage environment, place an E2E Microsoft Teams video call between the two users. Choose the appropriate noisy environment for personal or conferencing solution according to the descriptions provided in Section 3.1.1.

Far end user assesses the following areas on the scale from 1 to 5:

- DUT send speech quality
  - Speech level fluctuations - were there drop outs, cut offs, missing words or syllables, heavily chopped voice. What if DUT user speaks quietly?
  - Speech quality/speech naturalness - was there synthetic voice and/or heavy distortion and/or higher degree of band-limitation
  - Intelligibility/listening-effort - were there words that were hard to understand, considerable effort required.
  - Signal-to-noise ratio – were there times with high noise level, almost same level as speech, clearly disturbing, but call would be continued
  - Signal-to-noise ratio – were there clearly audible send noise fluctuations?
  - Background noise quality – was the background noise in call clearly unnatural/distorted/synthetic sounding, or many artefacts/clicks/plops, or often variation in sound and level
- Comfort noise quality
  - Clear difference between comfort noise and natural background noise perceivable; Clearly unnatural/distorted/synthetic sounding, or many artefacts/clicks/plops, or often variation in sound and level
  - Comfort noise does not sound like the natural background noise at all; Very unnatural/distorted/synthetic sound, or permanent artefacts/clicks/plops, or permanent variations in sound and level, very uncomfortable to listen to.

3.1.2.4 *Assessments in a conference call*

With the DUT in normal usage environment, create a Teams conference call between minimum of three users.
In case of room systems testing the Skype Room Systems client will be used as one leg of the call, for personal devices all 3 call legs may be using desktop client.
In addition to the far end there must be at least one additional far tester using a headset in the call.

Far end user assesses the following areas on the scale from 1 to 5:

- DUT send speech quality
  - Speech level fluctuations - were there drop outs, cut offs, missing words or syllables, heavily chopped voice. What if DUT user speaks quietly?
  - Speech quality/speech naturalness - was there synthetic voice and/or heavy distortion and/or higher degree of band-limitation
  - Intelligibility/listening-effort - were there words that were hard to understand, considerable effort required.

3.1.3 Headset and handset audio UI specific tests

With the DUT in a normal usage environment, place an E2E Microsoft Teams video call between the two users. One more tester should be present as assistant next to the DUT user.

DUT user assesses the following areas on the scale from 1 to 5:

- How comfortable is the sidetone level? Cover normal, low, and maximum playback levels.
- Is there some delay or distortion in the sidetone?
  - If the headset has active noise cancellation mode in playback, then it is critical to do these assessments in both mode – active noise cancellation ON and OFF.

Far end user assesses the following areas on the scale from 1 to 5:

- Does the exhale air flow from user's nose cause the headset microphone to pick up breathing noises?
  - DUT user should move the microphone boom up and down in small steps to verify the typical positional range of the boom. If boom is length adjustable then also different steps length wise should be checked. If possible both male and female DUT users should try the device and repeat this test.
- If DUT claims distractor attenuation (Open Office Headsets) are there some directions from where the attenuation is much worse? (Assisting tester should be talking and changing positions outside the claimed pick-up area.)
  - This test should also be repeated in open-office area with typical office activity and speech distractions.

3.1.4 Speakerphone audio UI specific tests

With the DUT in a typical usage environment for a product, start a video call between the two endpoints. Minimum of two near end talkers should be present on the DUT side.

Far end user assesses the following areas on the scale from 1 to 5:

- AGC must adapt in different conditions within 5 seconds. Conditions to be verified:
  - DUT user sets the OSGC to maximum and starts the call speaking normally.
  - DUT user sets the OSGC to 0 and starts the call speaking normally.
  - During the call DUT user blows into the microphone and then continues speaking normally.
- DUT user the whole field of view of the video?
  - If DUT claims directionality: How noticeable is the separation? (Assisting tester should be talking and changing positions outside the claimed pick-up area.)

3.1.5 Conferencing device audio UI specific tests

With the DUT in a similar conference room as the end product classification states (small/medium/large room), start a end to end video call between the two endpoints. Minimum of two testers (preferably 3..4 persons) should be present on the DUT side. Use a certified conferencing camera if the product bundle does not include a camera.

Following scenarios are to be tried:

- Each 4 near end talker says 2..3 sentences and then the turn goes to next near end talker. Far end participant will note how smooth and seamless the transition is from talker to talker. There should be no evident period of low speech level or period where words were hard to understand and considerable effort was required.
  - In case of table microphones this test shall be repeated by placing laptops with lid open between the near end talker and the table microphone
- The far end participant talks at normal and then at loud level.
  - Near end users note the clarity of playback speech. Also, any sign of distortion or noise boost during far end being silent should be noted.
  - Far end user is listening for echoes of his/her own voice.
- The far end talks at rapid pace with minimal silent gaps in between words and sentences. Near end talkers will try to interrupt the far end talker with their own sentences. Far end will register if the break in attempts

succeed (meaning far end user is able to hear the others trying to break into the conversation). This will validate the double talk capability of the device. During this test far end user is also listening for echoes of his/her own voice.

- System stability tests:
    - Maximize the playback volume on DUT. Far end user should talk loudly and listen for his/her own echo's leaking back. Same shall be repeated with near end talkers interrupting the far end user occasionally
    - While in call a near end user picks up a laptop and places in between speaker and one of the microphones while the far end is talking. The echo canceller leaks should disappear in max 2 seconds time.
    - One of the near end user will cough loudly while close to the microphone (50cm distance). Far end user will observe if automated gain controls will react reducing the signal overload. After the coughing has ended near end users start to talk again. Far end user should observe if the normal send level is retuned in max 10 seconds time.
- If the camera pan/tilt/zoom can be controlled: Does the far end user hear any clearly audible and disturbing mechanical noise picked up when these are adjusted?
- When near end mutes the microphone from Rooms system GUI, does the send signal get muted? Is there a clear indication of microphone being muted?
- When the near end user mutes the microphone from the buttons on device (if available), does the send signal get muted? Is there a clear indication of microphone being muted?
- If DUT claims directionality: How noticeable is the separation? (Assisting tester should be talking and changing positions outside the claimed pick-up area.)
- For big screen devices with touch support
    - Evaluate the send speech quality while the screen is tapped or drawn on using the pen/stylus. How audible are these touches on screen? Does the tapping to select and open apps on screen create uncomfortably loud sound into DUT send signal?
    - If the DUT is of a collaboration use type (movable on wheels; touch screen)
        - Test with a group of people (2 or more) front of the device. Does the DUT pick up all talkers equally, even if they are at the edges of camera field of view.
        - Does the echo cancellation pass all stability clauses in all devices Section 3.1.2
            - When users move places in front of the device
            - When the DUT is rotated portrait/landscape (if supported) or moved on wheels slightly
    - If the DUT screen enables rotation
        - Tests the rotation during a Teams call. Does it cause uncomfortably loud sound into DUT send signal? Is there a detectable difference in send speech quality and intelligibility?
        - Does the echo cancellation pass all stability clauses in all devices Section 3.1.2
        - If there are noticeable differences resulting from the rotation, how long does it take to converge.
- If the DUT has a cooling fan for CPU, does the fan noise increase when in video call.
    - In case of fan noise increase, is the noise noticeable but not disturbing / disturbing / very annoying for DUT user.

- How does the fan noise influence the far end user experience, is the noise noticeable but not disturbing / disturbing / very annoying?

3.1.5.1 *Note*

There are Excel spreadsheets available to help conduct the above audio sanity tests. Please contact your Microsoft partner manager to get it.

3.2 Tests without E2E call

The tests in this section are using the Windows HLK components from https://developer.microsoft.com/en-us/windows/hardware/windows-hardware-lab-kit

The below instructions are written for test tools included in latest official HLK at the time of writing (HLK for Windows 10 Version 1809). The actual HLK tools used can change as the new versions of the HLK are released.

The HLK tools to use should be chosen based on the Windows version of the DUT PC.

For devices running operating systems where the Windows HLK is not available this section of tests can be skipped and marked as N/A (not applicable)

The tests in this Section can be conducted in quiet office room or meeting room instead of the anechoic test room.

3.2.1 Windows HLK - Communications Audio Fidelity

3.2.1.1 *Test Purpose*

The Communication Audio Fidelity (CAF) test is verifying several essential audio attributes that are necessary for the audio tests defined in this document. It also covers certain areas that are left out from the portion of this specification that are tested by Head Acoustics test harness.

3.2.1.2 *Requirements*

The solution submitted to testing must have passed the CAF audio test applicable. The passed report must be provided before the testing is started.

Especially the following metrics must pass CAF required values:

- Mic TS Glitch (<12.5)
- Mic TS Drift (<0.08)
- Mic TS Error (< 0.125)
- Spk TS Glitch (<12.5)
- Spk TS Drift (<0.08)
- Spk TS Error (< 0.125)
- Aec Quality Test (<2%)

3.2.1.3 *Test process*

Follow the guidelines here: http://msdn.microsoft.com/en-us/library/windows/hardware/dn390880.aspx

3.2.2 Windows HLK - send path audio latency

3.2.2.1 *Purpose*

The Windows Hardware Lab Kit (HLK) includes a test for measuring the latency induced by the APO effects, device drivers and hardware. 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. Longer latency also causes the possible acoustic echo leakage to be more audible and annoying, thus minimizing the latency helps reduce the perceived echo leakage.

3.2.2.2 *Requirements*

The DUT send latency is measured and the latency of the reference device (Polycom CX100) is deducted.

$$Latency\_SND_{DUT} - Latency_{REF} = DUT\_rel\_SND\_latency$$

Equation 1: DUT relative send latency

	DUT_rel_SND_latency send direction	
	Standard (ms)	Premium (ms)
USB or embedded devices	< 120	< 90
Cordless devices	< 140	< 110

Table 8: DUT relative send latency requirements

3.2.2.3 *Test process*

- Configure the CX100 as both the default render and capture device. Run the *te.exe LatencyTest.dll* and measure the loop latency of the reference device ($Latency_{REF}$)

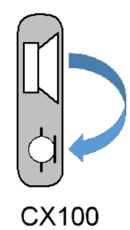


Figure 2: Audio path for measuring the loop latency of the reference device

- Configure the CX100 as a 'Default Device' and 'Default Communication Device' under the Windows 'Playback devices'.
  Configure the DUT device as a 'Default Device' and 'Default Communication Device' under the Windows 'Recording devices'.

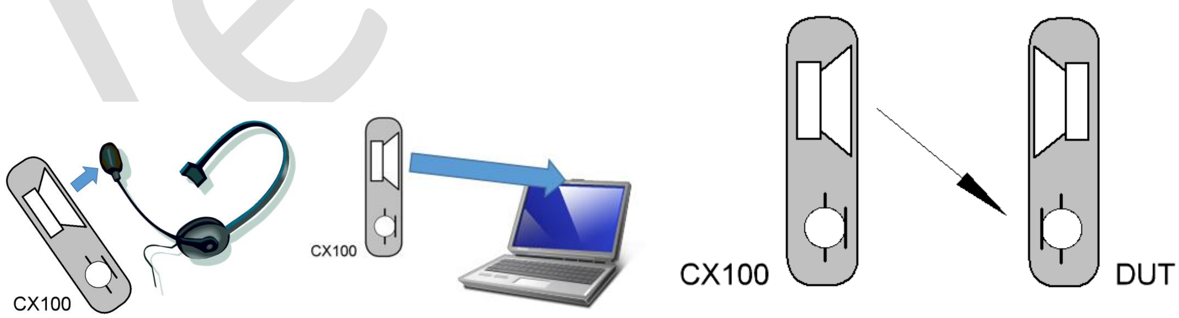


Figure 3: Audio path for measuring DUT send latency

- Run the *te.exe LatencyTest.dll* on the DUT PC Command Prompt, or on test PC in case of plug in devices. The PC used for testing shall run Windows 10 64bit.
- Read the 'LatencyTest::Custom', 'LatencyTest::Raw', and 'LatencyTest::Communications' from the text output on the screen as a result.

- Calculate the DUT relative send latencies by deducting the reference device latency for each of the modes. All must meet the requirement defined above.

3.2.3 Windows HLK – receive path audio latency

3.2.3.1 *Purpose*

The Windows Hardware Lab Kit (HLK) includes a test for measuring the latency induced by the device drivers and hardware. 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. Longer latency also causes the possible acoustic echo leakage to be more audible and annoying, thus minimizing the latency helps reduce the perceived echo leakage.

3.2.3.2 *Requirements*

The DUT receive latency is measured and the latency of the reference device (Polycom CX100) is deducted.

$$Latency\_RCV_{DUT} - Latency_{REF} = DUT\_rel\_RCV\_latency$$

Equation 2: DUT relative receive latency

	DUT_rel_RCV_latency receive direction	
	Standard (ms)	Premium (ms)
USB or embedded devices	< 100	< 70
Cordless devices	< 120	< 100

Table 9: DUT relative receive latency requirements

3.2.3.3 *Test process*

- Use the $Latency_{REF}$ measured in Section 3.2.2.
- Configure the DUT device as a 'Default Device' and 'Default Communication Device' under the Windows 'Playback devices'.
  Configure the CX100 as a 'Default Device' and 'Default Communication Device' under the Windows 'Recording devices'.

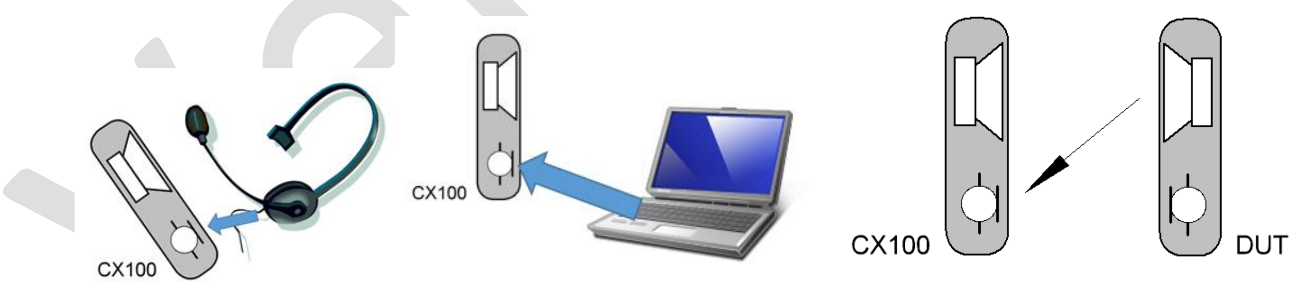


Figure 4: Audio path for measuring DUT receive latency

- Run the *te.exe LatencyTest.dll* on the DUT PC Command Prompt, or on test PC in case of plug in devices. The PC used for testing shall run Windows 10 64bit.
- Read the 'LatencyTest::Custom', 'LatencyTest::Raw', and 'LatencyTest::Communications' from the text output on the screen as a result.
- Calculate the DUT relative receive latencies by deducting the reference device latency for each of the modes. All must meet the requirement defined above.

3.2.4 Audio processing path for PC based solutions using APO offload mode

3.2.4.1 *Purpose*

It is important for an APO or driver to correctly expose all the effects that are provided in the default and communication render and capture streams as effects that are discoverable by software client.

For devices that apply for "RAW + Skype processing" should not have any nonlinear processing APO effects in echo path (such as dynamic range compression for render or AEC in capture pipeline). Please refer to section 2.6.1 for details.

3.2.4.2 *Requirements*

- The solution submitted to testing for "Hardware / Software offloading" must support the following custom effects as a minimum under Other Default: {}, Communications Default: {}
    - AcousticEchoCancellation (including comfort noise generator)
    - NoiseSuppression (stationary noise suppression)
- The solution submitted to testing for "Hardware / Software offloading" can support the following custom effects as a maximum under Other Default: {}, Communications Default: {}
    - AcousticEchoCancellation (including comfort noise generator)
    - NoiseSuppression (stationary noise suppression)
    - BeamForming (multi microphone processing for increased mic directivity

  Note that the AGC (AutomaticGainControl) offloading is not supported/allowed.

- The list of effects visible in Effect Discovery query result must match with the Playback devices / Microphone Devices -> Device Properties -> Enhancements UI. Each effect should have identifiable UI selection that can be enabled/disabled.
  The effects should not expose to user under naming which is not matching the operation of the effects or via single Effects On / Effects Off toggle selection.

3.2.4.3 *Test process*

Verify the list of APO effects by entering *'te.exe EffectsDiscoveryCPPTest_mc.dll'* to Command Prompt (for other platforms a similar query is made with custom method for each platform).

3.2.4.4 *Note*

This testcase is not applicable for external devices connecting to USB or via Cordless connection

3.2.5 Audio processing path for PC based solutions when using a "raw mode"

3.2.5.1 *Purpose*
PC based devices opened in "raw mode" should not have any nonlinear APO effects in echo path. For example, a dynamic range compression for render path or AEC in capture path is not allowed in "raw mode". Please refer to section 2.6.1 for details.

3.2.5.2 *Requirements*
- Verify the list of APO effects in raw capture and render. The list should not include any nonlinear APO's.
- Validate that there is full uninterrupted echo leaking back when MS Teams AEC is bypassed. If echo is cancelled it is an indication that there is a hidden AEC APO in the capture pipeline. In such case the testcase should be marked as "Failed"

3.2.5.3 *Test process*
- Verify the list of APO effects in raw capture and render by entering *'te.exe EffectsDiscoveryCPPTest_mc.dll'* to Command Prompt. The effect list should not expose any non-linear processing APO's for Other Raw: {}
- Use the following steps to validate that there are no render or capture APO's in the echo path:

Use the REF/DUT editor to set the DUT client to "Reference test mode" (this will bypass all MS Teams audio processing).

- Make an end to end test call from a far end client using a headset (same process as for section 3.1).
- Validate that there is full undistorted echo leaking back as MS Teams AEC is bypassed.
  If acoustic echo from DUT speaker is cancelled it is an indication that there is a hidden AEC APO in the capture pipeline. In such case the testcase should be marked as "Failed".

3.2.5.4 *Note*
This testcase is not applicable for external devices connecting to USB or via Cordless connection.

4 Audio quality requirements

All tests described under this chapter are conducted over an End to End Skype call and using the HEAD acoustics ACQUA based measurement system as test management and analysis tool. Section 5.2 provides the details about positioning and measurement equipment used.

REF PC client setup details in Section 5.1.3.

DUT client setup details in Section 5.1.4.

In case DUT device does not have a camera embedded or bundled then use one of the recommended cameras in Section 5.1.5

4.1 Send path tests

4.1.1 Send path - total quality loss

4.1.1.1 *Purpose*

To ensure that the call quality in send direction during the Microsoft Teams call using DUT in lossless local network condition is not degraded by device or interaction with Microsoft Teams client.

Evaluating the following speech quality criteria is in focus for this test:

- Noise Reduction – example: processing artifacts in form of noise gating etc.
- Voice Activity Detection – example: clipping the beginnings of speech
- Voice Enhancements – example: alteration in speech pitch
- Discontinuous Transmission – example: gaps, crackling sounds, clicks, pops etc.

4.1.1.2 *Requirement*

	POLQA, Super-Wideband mode	
	Standard (MOS-LQO)	Premium (MOS-LQO)
Headset Wired[6]	≥ 3.9	≥ 4.1
Headset Wireless[6]	≥ 3.7	≥ 3.9
Handset	≥ 3.9	≥ 4.1
Handheld speakerphone Personal speakerphone	≥ 3.5	≥ 3.7
Conferencing speakerphones	≥ 3.6	≥ 3.8

Table 10: Send total quality loss requirements

4.1.1.3 *Test procedure*

- Use the recommended test position for the DUT as described in Section 5.2.4
- Real speech male and female samples are used for the test. The source speech samples have to be prepared according to the guidelines in ITU-T P.863. The sample files should use 48 kHz sampling rate and be pre-filtered to 50Hz … 14 kHz pass band according to ITU-T G.191.

[6] For headsets and Open Office headsets with microphone boom this test is run at 3 different microphone boom positions as described in Section 5.2.4.1

- Play multiple real speech samples through artificial mouth at normal speech level. Record the speech samples in reference Microsoft Teams output. Use the software tool compliant to ITU-T P.863 (POLQA) to analyze and calculate the average MOS-LQO score.

4.1.2 Send path - latency

4.1.2.1 *Purpose:*

Call interactivity and acoustic echo audibility is dependent of the round-trip delay. The purpose of this test is to ensure that the sending delay portion of the round-trip delay during a Microsoft Teams call using the DUT in lossless local network is below the set maximum limit.

4.1.2.2 *Requirement*

Send path latency is defined as a time it takes for a speech signal to travel from HATS or artificial mouth to DUT microphone to Far End (Reference Microsoft Teams) output over Microsoft Teams call in lossless network condition.

	Send End to End latency	
	Standard (ms)	Premium (ms)
All categories Custom processing	< 230	< 210
All categories RAW + Teams processing	<190	<180

Table 11: Send latency requirements

For devices that have been tested with “Windows HLK – send path audio latency” and pass the requirement in Section 3.2.2.2, then the requirement above can be marked as “NA – latency result for information only”.

4.1.2.3 *Test Procedure*

Calculate the latency from the recorded speech samples used for POLQA in Section 4.1.1 and compare the calculated average send path latency with required limit.

4.1.3 Send path – activation level in send direction

4.1.3.1 *Purpose*

During the silence periods of a near end the DUT device might go into idle mode with extra noise suppression or send signal attenuation. This test determines what is the near end acoustic signal level needed to recover from the idle mode and what is the build up time for the signal.

4.1.3.2 *Requirement*

The computed average RMS level in reference Microsoft Teams output for the active part of male speech sample must satisfy the requirement in the table below.

	Activation level in send direction Analysis band 100Hz – 12kHz
	Standard
Headset[7]	$L_{S,min} \leq -20dBPa$ $T_{r,S,min} \leq 15ms$
Open office headset	$L_{S,min} \leq -23dBPa$ $T_{r,S,min} \leq 15ms$
Handset	$L_{S,min} \leq -20dBPa$ $T_{r,S,min} \leq 15ms$
Personal speakerphone	$L_{S,min} \leq -20dBPa$ $T_{r,S,min} \leq 15ms$
Conferencing devices	$L_{S,min} \leq -23dBPa$ $T_{r,S,min} \leq 15ms$

Table 12: Activation level in send direction

$L_{S,\mathrm{min}}$ - send direction minimum activation level (at mouth MRP)
$T_{r,S,min}$ - send direction build up time

4.1.3.3 *Test Procedure*

- DUT client setup details in Section 5.1.4. **(note the disable Teams digital AGC test mode)**
- Use the recommended test position for the DUT as described for respective device type in Section 5.2.4
- For the test stimulus use the test signal from ITU-T P.501 clause 7.3.4 - the test sequence is a series of single words with a level increase for each next word.
- Render the test stimulus at the artificial mouth.
- Capture and analyze the send signal in reference Microsoft Teams output. The source signal is filtered by the transfer function of the DUT and then the captured DUT send signal is compared to the filtered source signal and level versus time graph is displayed.
- The minimum activation level is determined from the first single word that passes without attenuation. That level indicates the activation level of the test object. This result should then be compared to the requirement above.
- Open Office Headset will be tested with all 3 microphone boom positions as shown in Section 5.3.2.3 if the DUT has an adjustable microphone boom.

[7] For headsets and Open Office headsets with microphone boom this test is run at 3 different microphone boom positions as described in Section 5.2.4.1

4.1.4 Send path - signal level with normal speech

4.1.4.1 *Purpose*

If the send output level is too low, then the far-end participants may have difficulties hearing the near end. At same time too loud a send signal can cause the peaks of the speech signals to overload the input causing clipping and other undesired artifacts.

4.1.4.2 *Requirement*

The computed average RMS level in reference Microsoft Teams output for the speech sample must satisfy the requirement in the table below.

Note: this test is provided as an alternate test method for devices that are incapable of disabling Teams DAGC as required for test 4.1.12 Send path – signal level with normal speech from device. That test is considered more accurate due to the more controlled test method.

		Signal level with normal speech Analysis band 100Hz – 12kHz	
		Standard (dBFS[8])	Premium (dBFS)
Headset		[-24, -14]	[-19, -14], no overload
Handset		[-24, -14][9]	[-19, -14], no overload
Personal speakerphone		[-24, -14]	[-21, -14]
Conferencing devices	up to 1.5m mic usage	[-24, -14]	[-21, -14]
	up to 2.3m mic usage	[-24, -14]	[-21, -14]
	up to 3.5m mic usage	[-24, -14]	[-21, -14]
	up to 4.5m mic usage	[-24, -14]	[-21, -14]

Note: For "raw mode" device 6dB can be deducted from each minimum requirement

Table 13: signal level with normal speech requirements

4.1.4.3 *Test Procedure*

- Use the recommended test position for the DUT as described in Section 5.2.4
- For the test stimulus use the IEEE 269-2010 uncompressed male speech signal.
- Leave >15 second stabilization time by playing the same source signal before starting the measurement.
- Render the test stimulus at the artificial mouth with speech signal level as described in Section 5.2.3.
- Compute the average RMS level in reference Microsoft Teams output for the full speech sample and check for clipping.
  For Premium requirement a check for send path overload/clipping. Max 2ms long section of audio samples with 0dBFS level are allowed during the test sample.

[8] ACQUA test system uses dBm0. dBm0 ≈ dBFS + 6dB

[9] For handsets, a send loudness rating (SLR) can be measured in addition to the send level in dB. For handset devices passing a wideband SLR criteria of 8 +3/-5 according to ITU-T Rec. P.79 can be considered a pass in test report overruling the dB value.

4.1.5 Send path - signal level with quiet speech

4.1.5.1 *Purpose*

If the send output level is too low, then the far-end participants may have difficulties hearing the near end. This could be particularly problematic in the beginning of the call before the automatic gain control stabilizes. The DUT processing should not have a wrongly tuned voice activity detector that suppresses the quiet speech.

4.1.5.2 *Requirement*

The computed average RMS level in reference Microsoft Teams output for the speech sample must satisfy the requirement in the table below.

<table>
<tr><td colspan="2" rowspan="2"></td><td colspan="2">Signal level with quiet speech
Analysis band 100Hz – 12kHz</td></tr>
<tr><td>Standard
(dBFS)</td><td>Premium
(dBFS)</td></tr>
<tr><td colspan="2">Headset</td><td>≥-32</td><td>≥-28</td></tr>
<tr><td colspan="2">Handset</td><td>≥-32</td><td>≥-28</td></tr>
<tr><td colspan="2">Personal speakerphone</td><td>≥-34</td><td>≥-31</td></tr>
<tr><td rowspan="4">Conferencing devices</td><td>up to 1.5m mic usage</td><td>≥-34</td><td>≥-31</td></tr>
<tr><td>up to 2.3m mic usage</td><td>≥-34</td><td>≥-31</td></tr>
<tr><td>up to 3.5m mic usage</td><td>≥-37</td><td>≥-31</td></tr>
<tr><td>up to 4.5m mic usage</td><td>≥-37</td><td>≥-31</td></tr>
</table>

Note: For "raw mode" testing 6dB can be deducted from each minimum requirement

Table 14: Signal level with quiet speech requirements

4.1.5.3 *Test Procedure*

- Use the recommended test position for the DUT as described in Section 5.2.
- For the test stimulus use the IEEE 269-2010 uncompressed male speech signal.
- Render the test stimulus at the artificial mouth with quiet speech signal level as described in Section 5.2.3.
- Leave >15 second stabilization time by playing the same source signal before starting the measurement.
- Compute the average RMS level in reference Microsoft Teams output for the full speech sample.

4.1.6 Send path - idle channel SpNR

4.1.6.1 *Purpose*

If the self-noise is high, the far-end participants may find the noise disruptive. Also, it may affect speech intelligibility as the speech intelligibility decreases when noise is present. Additionally, this can also lead to low send signal levels because the digital gain control may limit the amount of noise, which can be added to a conference from one endpoint and, therefore, might not allow the maximum amplification of the send signal. This test verifies that the device has a good Speech to Noise ratio during longer silences between speech activity.

4.1.6.2 *Requirement*

		Speech to Noise Ratio during idle channel ($SpNR_{idle}$) Analysis band 100Hz – 12kHz	
		Standard (dB)	Premium (dB)
Headset		≥ 43	≥ 46
Open office headset		≥ 46	≥ 49
Handset		≥ 43	≥ 46
Personal speakerphone		≥ 40	≥ 43
Conferencing devices	up to 1.5m mic usage	≥ 40	≥ 43
	up to 2.3m mic usage	≥ 40	≥ 46
	up to 3.5m mic usage	≥ 40	≥ 46
	up to 4.5m mic usage	≥ 37	≥ 43

Note: For “raw mode” testing 3dB can be deducted from each minimum requirement

Table 15: Speech signal to self-noise ratio during idle channel requirements

4.1.6.3 *Test Procedure*

- Use the recommended test position for the DUT as described in Section 5.2.4
- Play the IEEE Male speech + silence test signal at normal speech level and record the signal at reference Microsoft Teams output.
- Calculate the non-weighted level of active speech signal.
- Calculate the A-weighted RMS level of noise in reference Microsoft Teams output for the 1 second silence part at end of the test signal for $SpNR_{idle}$.
- Calculate the $SpNR_{idle}$ (dB) = active speech level – A-weighted noise level during longer silence

4.1.7 Send path - active channel SpNR

4.1.7.1 *Purpose*

The self-noise measurement methodology applied in Section 4.1.6 is not indicative of end quality in case noise gating or aggressive noise suppression processing is applied in the send path. This test verifies that the device has a good Speech to Noise ratio during the speech activity (SpNR – speech to noise ratio during speech).

4.1.7.2 *Requirement*

		Speech to Noise Ratio for active channel ($SpNR_{active}$) Analysis band 100Hz – 12kHz	
		Standard (dB)	Premium (dB)
Headset		≥ 40	≥ 43
Handset		≥ 40	≥ 43
Personal speakerphone		≥ 37	≥ 40
Conferencing devices	up to 1.5m mic usage	≥ 37	≥ 40
	up to 2.3m mic usage	≥ 37	≥ 40
	up to 3.5m mic usage	≥ 37	≥ 40
	up to 4.5m mic usage	≥ 34	≥ 37

Note: For "raw mode" testing 3dB can be deducted from each minimum requirement

Table 16: Speech signal to self-noise ratio during active channel requirements

4.1.7.3 *Test Procedure*

- Use the recommended test position for the DUT as described in Section 5.2.4
- Play the IEEE Male speech for noise during speech test signal at normal speech level and record the signal at reference client output.
- Calculate the non-weighted level of active speech signal.
- Calculate the A-weighted RMS level of noise in reference Microsoft Teams output for the modified 1 second silence part at end of the test signal for $SpNR_{active}$. (The sine signal components are filtered out by band stop filters prior to noise level calculation.)
- Calculate the $SpNR_{active}$ (dB) = active speech level – A-weighted noise level.

4.1.7.4 *Note*

Microsoft reserves the right 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.

4.1.8 Send path - single frequency interference

4.1.8.1 *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.

4.1.8.2 *Requirement*

The A-weighted peak level of highest single frequency interference noise shall be less than 10 dB over the wide band send noise level measured in Section 4.1.6.

If the peak interference noise level measured is below -80dBV then the result of send interference noise test can be marked as Pass over ruling the calculated result.

4.1.8.3 *Test Procedure*

- Use the recommended test position for the DUT as described in Section 5.2.4
- Play the IEEE Male speech + silence test signal at normal speech level and record the signal at reference Microsoft Teams output.
- Calculate the interference noise spectrum analysis using a constant bandwidth 30Hz, Flat Top filter, FFT with 4096 sample size.
- Create a tolerance mask by 1/1 octave intensity smoothing the noise curve and applying +10dB to the curve throughout the analysis frequency range.
- The test result is considered pass if the unsmoothed interference peak remains below the tolerance mask.

4.1.9 Send path - distortion and noise

4.1.9.1 *Purpose*

This requirement ensures correct send speech quality so that the far end does not hear distorted or noisy speech. It is also required for optimal performance of AEC so no echo leak results from nonlinear distortions.

4.1.9.2 *Requirement*

		SDNR (pulsed noise signal-to-distortion-and-noise ratio)	
		Standard (dB)	Premium (dB)
	Frequency band	Signal level: 89dBSPL @MRP	
ALL CATEGORIES	224-282	≥30	≥32
	282-355	≥30	≥32
	355-447	≥30	≥32
	447-562	≥30	≥32
	562-708	≥30	≥32
	708-891	≥30	≥32
	891-1122	≥30	≥32
	1122-1413	≥30	≥32
	1413-1778	≥30	≥32
	1778-2239	≥30	≥32
	2239-2818	≥30	≥32
	2818-3548	≥30	≥32
	3548-4467	≥30	≥32
	4467-5623	≥28	≥28

Note: 24dB=6.3%, 26dB=5%, 28dB=4%, 30dB=3.2%, 32dB=2.5%, 34dB=2%

Table 17: Send SDNR requirements

4.1.9.3 *Test Procedure*

- Use the recommended test position for the DUT as described in Section 5.2.4
- Leave >15 second stabilization time by playing the same source signal before starting the measurement at 89dBSPL level. Use the IEEE 269-2010 uncompressed male speech signal as a stabilization signal.
- The test stimulus is white noise band limited to 1/3rd octave and pulsed 250 ms ON (active period), 150 ms OFF (conditioning period). This sequence should be repeated ten times for a total stimulus of four seconds. This test signal should place the DUT in a well-defined, reproducible state for the period of the measurement.
- Apply the test signal at the MRP, at the defined level and frequencies specified in Table 17.
- Calculate the ratio of the signal power to the total A-weighted distortion and noise power of the signal output.

Process the last 200 ms of each measured noise burst using Hanning windowing with 5 Hz resolution FFT from 100 Hz to 8000 Hz. After time averaging, remove the data within the notch frequency band.

4.1.9.4 *Notes*

More information about the SDNR test method can be found from IEEE 269-2010 Annex L.

4.1.10 Send path - frequency response

4.1.10.1 *Purpose*

The frequency response requirement is to make sure that the speech sounds natural and sufficiently intelligible from tonality perspective.

4.1.10.2 *Requirements*

The frequency response requirements are given for each device category.

		Send path frequency response			
		Standard		Premium	
	F(Hz)	Lower limit (dB)	Upper limit (dB)	Lower limit (dB)	Upper limit (dB)
Headset	100		6		5
	150		6	-5	5
	200	-6	6	-5	5
	5000	-6	6	-5	5
	6300	-11	6	-5	5
	8000		6	-5	5
	10500		6	-10	5
Handset	100		6.5		5
	150		6.5		5
	200	-6.5	6.5	-5	5
	5000	-6.5	6.5	-5	5
	6300	-11.5	6.5	-10	5
	8000		6.5		5
	10500		6.5		5
Handheld Speakerphone	100		7		3
	150		7		6
	200	-7	7	-6	6
	5000	-7	7	-6	6
	6300	-13	7	-6	6
	8000		7	-6	6
Personal Speakerphone	100		7		6
	180		7	-6	6
	200	-7.5	7	-6	6
	5000	-7.5	7	-6	6
	6300	-12.5	7	-6	6
	7500		7	-11	6
	10500		7		6

Table 18: Send frequency response requirements for personal solutions

	F(Hz)	Send path frequency response: Standard Lower limit (dB)	Standard Upper limit (dB)	Premium Lower limit (dB)	Premium Upper limit (dB)
Conferencing solutions	100	-30	-2	-30	-3
	200	-7.5	2	-6.5	1
	2000	-3.5	6	-2.5	5
	5000	-3.5	6	-2.5	5
	6300	-6	6	-6	5
	7500	-11	6	-11	5
	10500		6		5

Table 19: Send frequency response requirements for conferencing solutions

The frequency response masks are 'floating' meaning that if the limit is +/-7dB then the mask moves up/down to find the best fit for the measured response.

Figure 5: Headset send frequency response mask

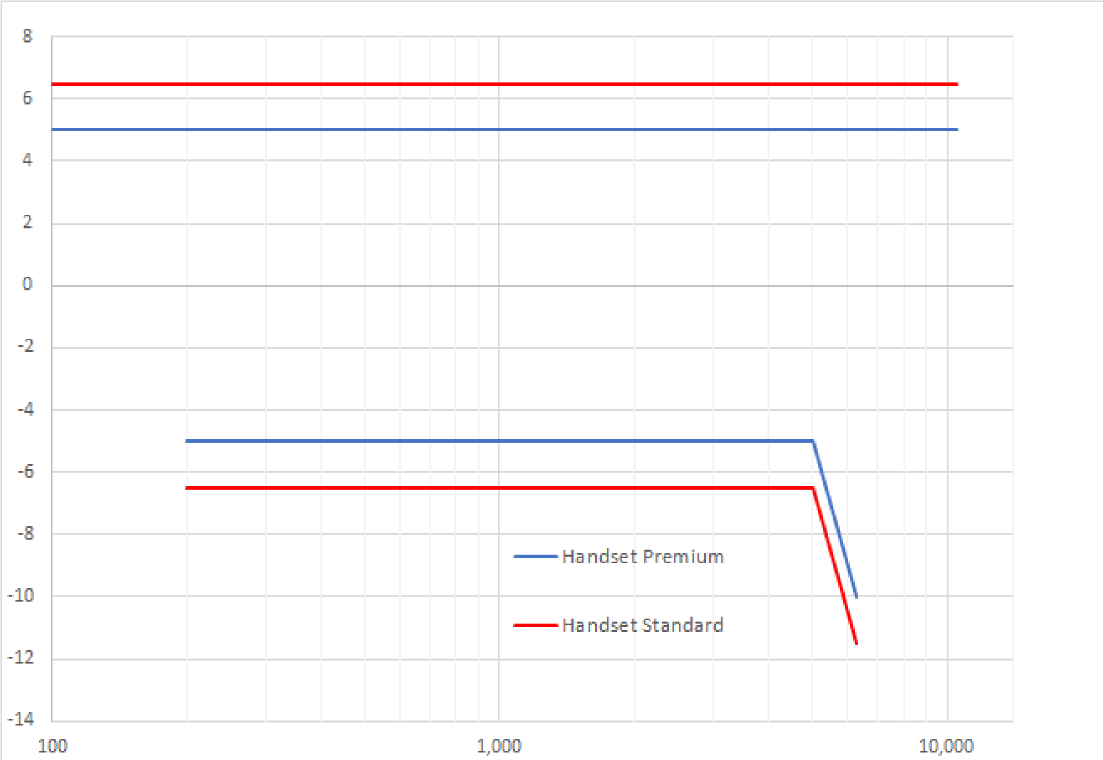


Figure 6: Handset send frequency response mask

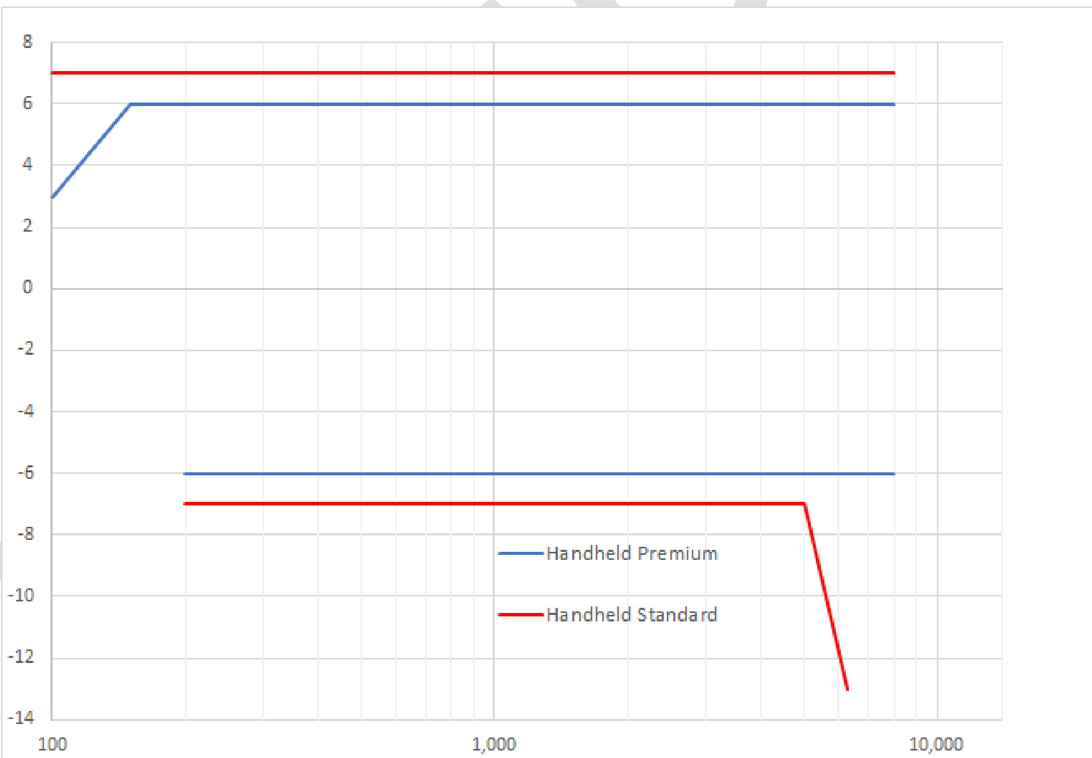


Figure 7: Handheld Speakerphone send frequency response mask

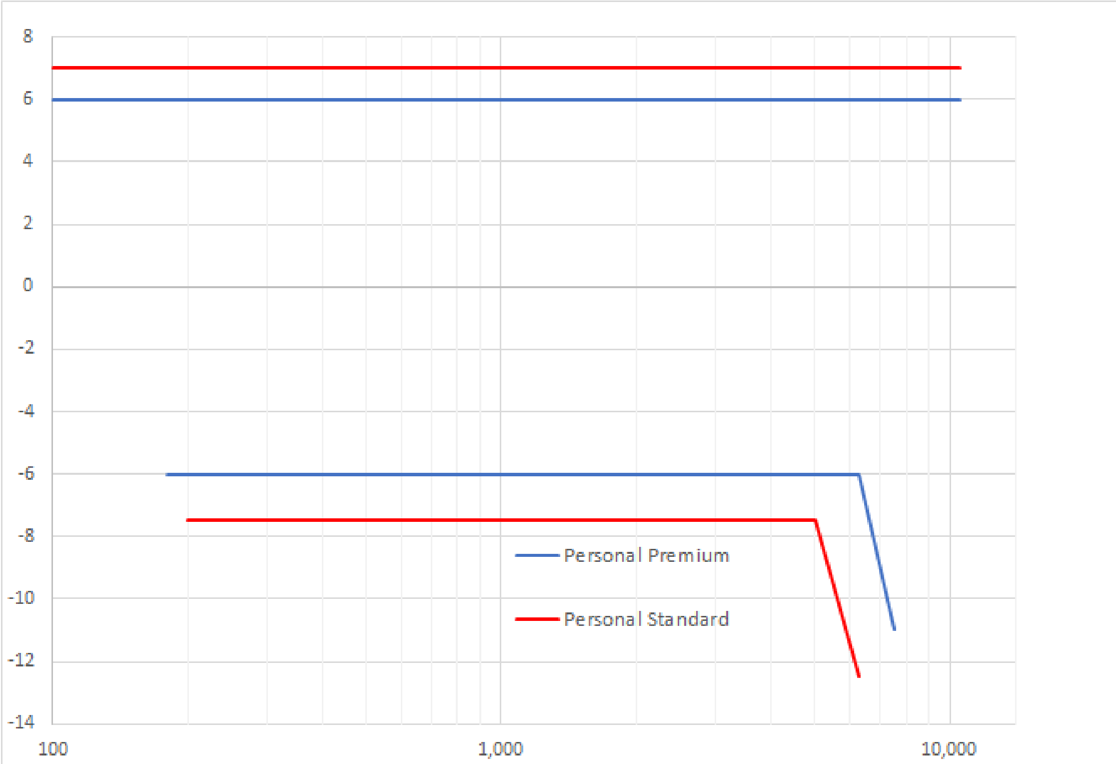


Figure 8: Personal Speakerphone send frequency response mask

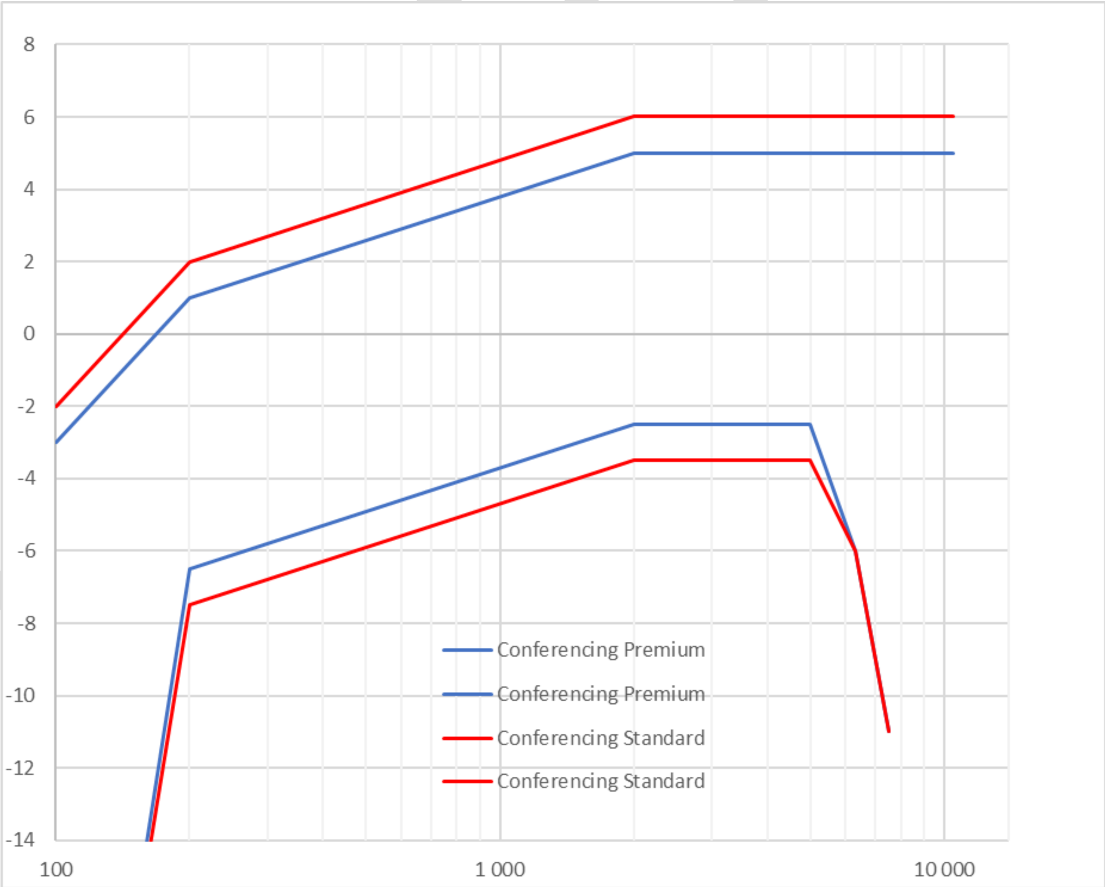


Figure 9: Conferencing device send frequency response mask

4.1.10.3 *Test Procedure*

- Use the recommended test position for the DUT as described in Section 5.2.4
- Run the test signal for >15 seconds for stabilization by playing the IEEE 269-2010 uncompressed male speech signal.
- Measurements shall be done over a range of 100 Hz through 10500 Hz:
    - 1/12th octave bands for handset and headset requirements.
    - 1/3rd octave bands for portable device and speakerphone requirements to account for variability due to sound reflections.
- The test stimulus is the IEEE Male-Female signal for frequency response.
- For headset the headset shall be removed and put back on HATS for 3 times.
- For handset the frequency response shall be measured at 3 positions as described in Section 0 using
    - Angle A 21 deg, angle C – default or manufacturer recommendation.
    - Angle A 35 deg, and angle C in step 1 minus 5 degrees.
    - Angle A 15 deg, and angle C in step 1 plus 5 degrees.
- For speakerphone UI (personal and conferencing devices) the frequency response shall be measured at 3 positions, the Free Field microphone and artificial mouth shall be placed at 0deg, +15deg, -7.5deg as described in Section 5.1.7.
- The final resulting frequency response compared against tolerance mask is a dB average in each band of the 3 measured responses.

4.1.11 Send path – noise level with maximum microphone gain

4.1.11.1 *Purpose*

The send path in device under test should be optimized so that the send path gains available allow Microsoft Teams to achieve an optimal send signal level in the main usage scenario, but DUT should not have excessive headroom of gains where the Speech to Noise ratio decreases and causes a severe degradation in call quality.

Even when the DUT is multi-function device such as a tablet or laptop it is expected that the quiet speech (with level around 55dBSPL at 50cm) is the quietest capture signal to optimize for. Recordings of events, music etc. are all louder scenarios than speech recordings.

Limiting the available maximum gain and/or boost is especially important if product uses digital MEMS microphones. In such scenario the OSGC is implemented via digital gain only and increasing the gain equals increasing the input noise equally as much causing very poor signal to noise ratios at higher gains.

The boost feature under microphone input settings should not be used whenever possible. A single OSGC slider is much preferred by Microsoft Teams as the combination on Boost and OSGC slider can create a very complex gains structure that severely degrades the automated gain adjustment process in Microsoft Teams.

4.1.11.2 *Requirement*

The average A-weighted noise level at normal speech (0dB) use case is compared to the averaged noise of the -15, -18 and -21dB samples. The noise should not increase more than the requirement below.

	Noise level increase with quiet speech	
	Standard (dBFS)	Premium (dBFS)
Headset Handset	≤ 5dB	≤ 3dB
Speakerphone UI categories	≤ 7dB	≤ 4dB

Table 20: Send path noise level with maximum gains

4.1.11.3 *Test Procedure*

- Use the recommended test position for the DUT as described in Section 5.2.4
- The test stimulus is the IEEE Male-Female signal for frequency response (8 sentences of male and female speech).
- Play the test signal with following attenuations
  - 0dB (89dBSPL @ Mouth MRP – normal speech level)
  - -3dB
  - -6dB
  - -9dB (close to the quiet speech level)
  - -12dB
  - -15dB
  - -18dB
  - -21dB
- Analyze the rms level of the last sentence in sequence and analyze the A-weighted noise level during 1 second period after the last sentence for each of the tested levels above.
- Calculate the noise levels difference of first sample versus the average of the last 3 samples

4.1.12 Send path – signal level with normal speech from device

4.1.12.1 *Purpose*

This test checks for the send signal level when the Teams client digital AGC is not used. Meeting the requirements below will ensure that the client does not have to use excessive digital gain to reach the optimal signal levels for group calls.

4.1.12.2 *Requirement*

This requirement applies only to devices using Windows OS or accessories connected to Windows PC or a PC running Microsoft Teams Rooms.

The computed average RMS level in reference Microsoft Teams output for the speech sample must satisfy the requirement in the table below.

		Signal level with normal speech Analysis band 100Hz – 12kHz	
		Standard (dBFS[10])	Premium (dBFS)
Headset		[-31, -14]	[-27, -14], no overload
Handset		[-31, -14][11]	[-27, -14], no overload
Personal speakerphone		[-34, -14]	[-30, -14]
Conferencing devices	up to 1.5m mic usage	[-34, -14]	[-30, -14]
	up to 2.3m mic usage	[-34, -14]	[-30, -14]
	up to 3.5m mic usage	[-34, -14]	[-30, -14]
	up to 4.5m mic usage	[-34, -14]	[-30, -14]

Table 21: signal level from device with normal speech requirements

4.1.12.3 *Test Procedure*

- Use the recommended test position for the DUT as described in Section 5.2.4
- DUT client setup details in Section 5.1.4. **(note the disable Teams digital AGC test mode)**
- For the test stimulus use the TCL test signal.
- Render the test stimulus at the artificial mouth with speech signal level as described in Section 5.2.3.
- Compute the average RMS level in reference Microsoft Teams output for the full speech sample.
- For Premium requirement, a check for send path overload/clipping. Max 2ms long section of audio samples with 0dBFS level are allowed during the test sample.

[10] ACQUA test system uses dBm0. dBm0 ≈ dBFS + 6dB

[11] For handsets, a send loudness rating (SLR) can be measured in addition to the send level in dB. For handset devices passing a wideband SLR criteria of 8 +3/-5 according to ITU-T Rec. P.79 can be considered a pass in test report overruling the dB value.

4.2 Receive path

For headset and handset UI category the DRP to Diffuse Field compensation is used for the HATS artificial ear.
For speakerphone UI categories the Free Field microphone is used.

4.2.1 Receive path - output level

4.2.1.1 *Purpose*

To ensure that the receive path volume control allows setting the volume to the preferred listening level.

4.2.1.2 *Requirement*

The receive output level is being measured as the sound pressure level at the measurement microphone or artificial ear in dBSPL while the DUT is playing back a speech signal.

		Receive output level target Standard (dBSPL)
Headset monaural Headset binaural		76 ± 2 70 ± 2
Handset		76 ± 2[12]
Handheld speakerphone Personal speakerphone		65 ± 2[13]
Conferencing devices	up to 1.5m speaker usage	67 -1/+2
	up to 2.3m speaker usage	69 ± 2
	up to 3.5m speaker usage	71 ± 2
	up to 4.5m speaker usage	73 ± 2
	Up to 7.5m speaker usage	76 -1/+2

Table 22: Receive output level requirements

Some requirements in this document may require a nominal receive output level during test.

The device render volume setting which results in a receive output level closest to the requirement given in Table 22 is defined as nominal level.

For conferencing devices the playback level target for the largest speaker to listener distance supported is to be selected.

4.2.1.3 *Test Procedure*

- Use the recommended test position for the DUT as described in Section 5.2.4
- The test stimulus is the IEEE 269-2010 male speech signal with active speech level of -18dBFS.
- Measure the receive RMS output level in dBSPL by the artificial ear/measurement microphone in range of 100Hz to 12000Hz.
- To meet this requirement, adjust the playback volume as follows:
  - Adjust the device volume control to achieve the nominal receive level.
  - If the nominal receive level cannot be achieved, then the maximum volume the device allows shall be used.

[12] For handsets a receive loudness rating (RLR) can be measured in addition to the receive level in dB. For handsets passing a wideband RLR criteria of 2 +/-3 measured according to ITU-T Rec. P.79 can be considered a pass in test report, over-ruling the dBSPL value. DRP-ERP correction is used for the RLR test!

[13] For laptop in tent mode or a tablet in kickstand mode, the requirement can be relaxed by upto 3dB due to level drop caused by increased usage distance.

4.2.2 Receive path - total quality loss

4.2.2.1 *Purpose*

Evaluating the following speech quality criteria is in focus for this test:

- Receive side Noise Reduction – example: causing fluctuating the noise level during playback.
- Voice Enhancements – example: excessive dynamic range compression
- Discontinuous Transmission – example: gaps, crackling sounds, clicks, pops etc.

4.2.2.2 *Requirement*

	POLQA, Super-Wideband mode	
	Standard (MOS-LQO)	Premium (MOS-LQO)
Headset	≥ 3.7	≥ 4.1
Handset	≥ 3.7	≥ 3.9
Handheld speakerphone Personal speakerphone	≥ 3.3	≥ 3.7
Conferencing speakerphones	≥ 3.7	≥ 3.9

Table 23: Receive total quality loss requirements

4.2.2.3 *Test procedure*

- Use the recommended test position for the DUT as described in Section 5.2.4
- Use the same nominal receive output level as in Section 4.2.1.
- Real speech male and female samples are used for the test. The source speech samples have to be prepared according to the guidelines in ITU-T P.863. The sample files should use 48kHz sampling rate and be pre-filtered to 50Hz … 14kHz pass band according to ITU-T G191.
- Record the speech samples at DUT output. Use the software tool compliant to ITU-T P.863 (POLQA) to analyze and calculate the average MOS-LQO score.

4.2.3 Receive path - latency

4.2.3.1 *Purpose:*

Call interactivity and acoustic echo audibility is dependent of the round-trip delay. The purpose of this test is to ensure that the receiving delay portion of the round-trip delay during a Microsoft Teams call using the DUT in lossless local network is below the set maximum limit.

4.2.3.2 *Requirement*

	Receive End to End latency	
	Standard (ms)	Premium (ms)
All categories Custom processing	<200	<180
All categories RAW + Teams processing	<180	<170

Table 24: Receive latency requirements

For devices that have been tested with “Windows HLK – receive path audio latency” and pass the requirement in 3.2.3.2, then the requirement above can be marked as “NA – latency result for information only”.

4.2.3.3 *Test Procedure*

Calculate the latency from the recorded speech samples used for POLQA in Section 4.2.2 and compare the calculated average receive path latency with required limit.

4.2.4 Receive path - idle channel noise

4.2.4.1 *Purpose*

If the noise level is high, the participants may find the noise disruptive. Also, it may affect speech intelligibility as the speech intelligibility decreases when noise is present.

4.2.4.2 *Requirement*

	Receive idle channel noise Analysis band 100Hz – 20kHz	
	Standard (dBSPL(A))	Premium (dBSPL(A))
Headset	≤ 32	≤ 29
Handset	≤ 32	≤ 29
Handheld speakerphone Personal speakerphone	≤ 32 ≤ 36 if DUT has fan	≤ 27 ≤ 32 if DUT has fan
Conferencing speakerphones[14]	≤ 32	≤ 27

Table 25: Receive idle channel noise requirements

4.2.4.3 *Test Procedure*

- Use the recommended test position for the DUT as described in Section 5.2.4
- Use the same nominal receive output level as in Section 4.2.1.
- Play the IEEE Male speech + silence test signal to reference Microsoft Teams input and record the signal in HATS ear or FF microphone.
- Calculate the A-weighted RMS level of noise in reference Microsoft Teams output for the 1 second silence part at end of the test signal for SpNR.

[14] DUT solutions that include a compute unit for Microsoft Teams Rooms or other supporting equipment such as PoE network switches etc. have to meet the same dBSPL(A) noise emission requirement at 1m distance.

4.2.5 Receive path - single frequency interference

4.2.5.1 *Purpose*

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

4.2.5.2 *Requirement*

The receive A-weighted single frequency interference noise level shall be at least 10 dB quieter than the wideband receive noise level.

If the peak interference noise level measured is below 24dBSPL then the result of receive interference noise test can be marked as Pass over ruling the calculated result.

4.2.5.3 *Test Procedure*

- Use the recommended test position for the DUT as described in Section 5.2.4
- Use the same nominal receive output level as in Section 4.2.1.
- Play a test signal containing digital silence to reference Microsoft Teams input and record the signal in HATS ear or FF microphone.
- Calculate the interference noise spectrum analysis using a constant bandwidth 30Hz, Flat Top filter, FFT with 4096 sample size.
- Create a tolerance mask by 1/1 octave intensity smoothing the noise curve and applying +10dB to the curve throughout the analysis frequency range.
  The test result is considered pass if the unsmoothed interference peak remains below the tolerance mask.

4.2.6 Receive path - distortion and noise

4.2.6.1 *Purpose*

Distortion or noise in the receiving path affects listening quality for the near end user. For example, if the near end user uses a small loudspeaker driven at a very high output level, buzzing and other non-linear distortion might occur. Receive distortion may also cause echo leak due to nonlinearity of the echo.

4.2.6.2 *Requirement*

Note:

	Frequency band	SDNR (pulsed noise signal-to-distortion-and-noise ratio) Standard (dB) Level: -16dBFS	Premium (dB) Level: -16dBFS
Headset Handset	224-282	≥30	≥32
	282-355	≥32	≥34
	355-447	≥32	≥34
	447-562	≥32	≥34
	562-708	≥32	≥34
	708-891	≥32	≥34
	891-1122	≥32	≥34
	1122-1413	≥32	≥34
	1413-1778	≥32	≥34
	1778-2239	≥32	≥34
	2239-2818	≥32	≥34
	2818-3548	≥32	≥34
	3548-4467	≥32	≥34
	4467-5623	≥30	≥30
Handheld speakerphone Personal speakerphone	224-282	NA	NA
	282-355	NA	≥24
	355-447	NA	≥24
	447-562	NA	≥24
	562-708	≥20	≥26
	708-891	≥22	≥26
	891-1122	≥24	≥26
	1122-1413	≥24	≥26
	1413-1778	≥24	≥26
	1778-2239	≥24	≥26
	2239-2818	≥24	≥26
	2818-3548	≥24	≥26
	3548-4467	≥24	≥26
	4467-5623	≥24	≥26

\- 18dB=12.6%, -20dB=10%, -22dB=8%, -24dB=6.3%, -26dB=5%, -28dB=4%, -30dB=3.2%

Table 26: Receive SDNR requirements for personal solutions

Note: -

		SDNR (pulsed noise signal-to-distortion-and-noise ratio)	
		Standard (dB)	Premium (dB)
	Frequency band	Level: -16dBFS	Level: -16dBFS
Conferencing solutions	224-282	≥20	≥24
	282-355	≥20	≥24
	355-447	≥22	≥24
	447-562	≥24	≥24
	562-708	≥24	≥26
	708-891	≥24	≥26
	891-1122	≥24	≥26
	1122-1413	≥24	≥26
	1413-1778	≥24	≥26
	1778-2239	≥24	≥26
	2239-2818	≥24	≥26
	2818-3548	≥24	≥26
	3548-4467	≥24	≥26
	4467-5623	≥24	≥26

18dB=12.6%, -20dB=10%, -22dB=8%, -24dB=6.3%, -26dB=5%, -28dB=4%, -30dB=3.2%

Table 27: Receive SDNR requirements for conferencing devices

4.2.6.3 *Test Procedure*

- Use the recommended test position for the DUT as described in Section 5.2.4
- Use the loudspeaker gain control setting determined in Section 4.2.1.
- The test stimulus is white noise band limited to 1/3rd octave and pulsed 250 ms ON (active period), 150 ms OFF (conditioning period). This sequence should be repeated ten times for a total stimulus of four seconds. This test signal should place the device in a well-defined, reproducible state for the period of the measurement.
- Apply the test signal at the defined level and frequencies as specified in Table 26.

Process the last 200 ms of each measured noise burst using Hanning window with 5 Hz resolution FFT from 100 Hz to 8000 Hz. After time averaging, remove the data within the notch frequency band.

4.2.6.4 *Note*

More information about the SDNR test method can be found from IEEE 269-2010 Annex L.

4.2.7 Receive path - frequency response

4.2.7.1 *Purpose*

The frequency response requirement is required to make sure that the speech sounds natural and sufficiently intelligible from tonality perspective when using Microsoft Teams.

4.2.7.2 *Requirements*

The frequency response requirements are given for each device category. If the DUT has multiple settings for earpiece or loudspeaker tone control, then the DUT is tested with all available receive path tone control settings.

		Receive path frequency response			
		Standard		Premium	
	F(Hz)	Lower limit (dB)	Upper limit (dB)	Lower limit (dB)	Upper limit (dB)
Headset	100		7		6
	150		7	-10	6
	200	-11	7	-6	6
	300	-7	7	-6	6
	1000	-7	7	-6	6
	2000	-7	9	-6	8
	5000	-7	9	-6	8
	6300	-13	9	-6	8
	8000		9	-6	8
	10500		9	-12	8
Handset	100		7.5		6
	150		7.5		6
	200	-11.5	7.5	-10	6
	300	-7.5	7.5	-6	6
	1000	-7.5	7.5	-6	6
	2000	-7.5	9.5	-6	8
	5000	-7.5	9.5	-6	8
	6300	-13.5	9.5	-12	8
	8000		9.5		8
	10500		9.5		8
Handheld Speakerphone	100		7		6
	400		7	-12	6
	500		7	-6	6
	630	-13	7	-6	6
	800	-7	7	-6	6
	1000	-7	7	-6	6
	5000	-7	7	-6	6
	6300	-13	7	-6	6
	8000		7	-6	6
	10500		7	-12	6
Personal Speakerphone	100		7		6
	200		7	-12	6

		Receive path frequency response			
		Standard		Premium	
	250		7	-9	6
	355	-12	7	-6	6
	485	-7	7	-6	6
	5600	-7	7	-6	6
	7100	-10	7	-6	6
	10500		7	-10	6
Conferencing solutions	100		7		6
	180		7	-12	6
	200	-13	7	-10	6
	250	-10	7	-6	6
	315	-7	7	-6	6
	5000	-7	7	-6	6
	6300	-10	7	-6	6
	7500		7	-11	6
	10500		7		6

Table 28: Receive frequency response requirements

The frequency response masks are 'floating' meaning that if the limit is +/-7dB then the mask moves up/down to find the best fit for the measured response.

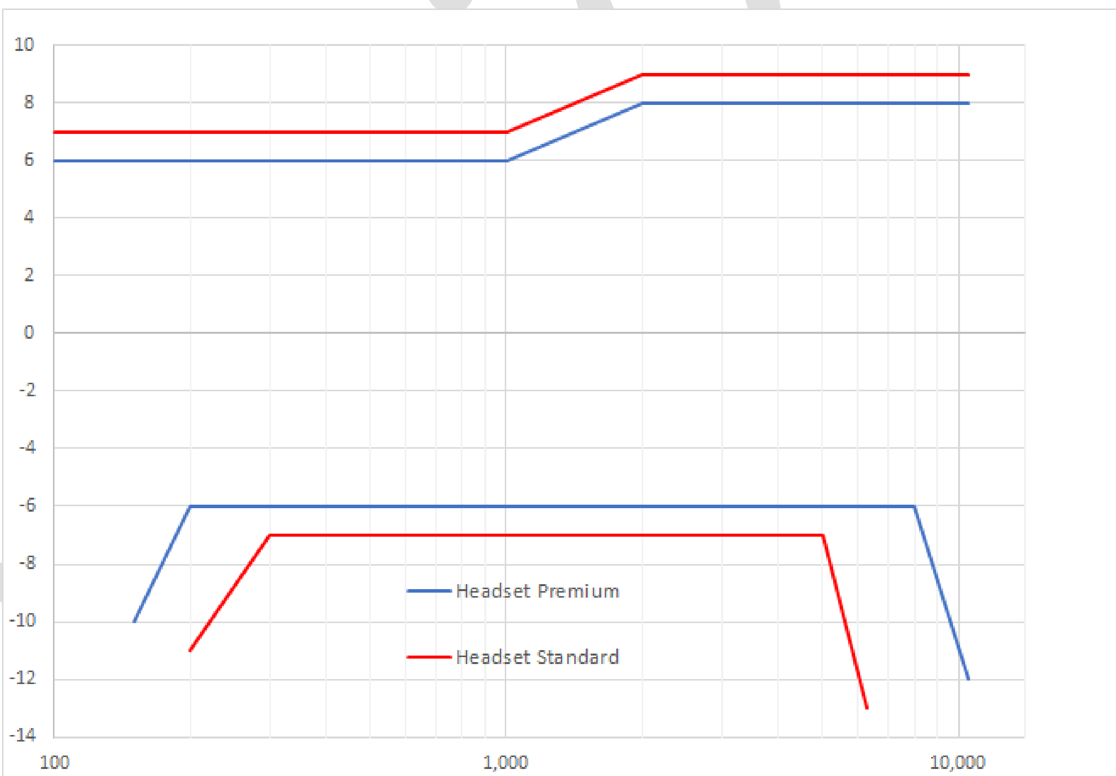


Figure 10: Headset receive frequency response mask

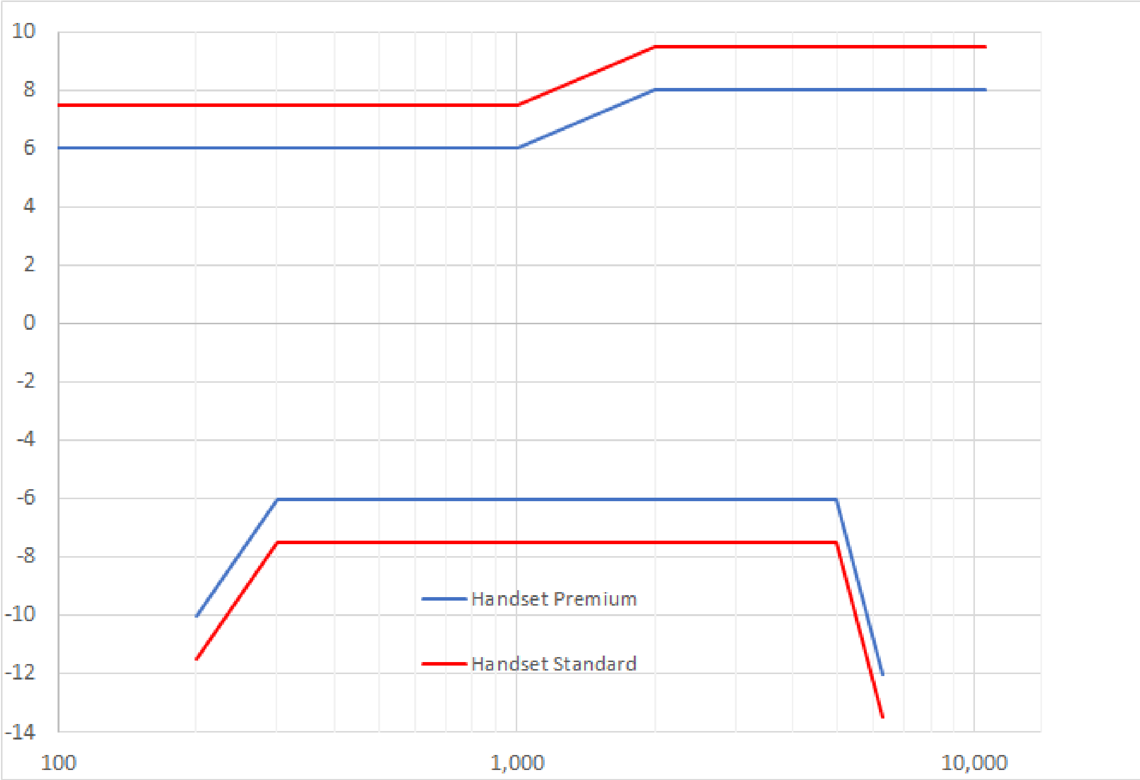


Figure 11: Handset receive frequency response mask

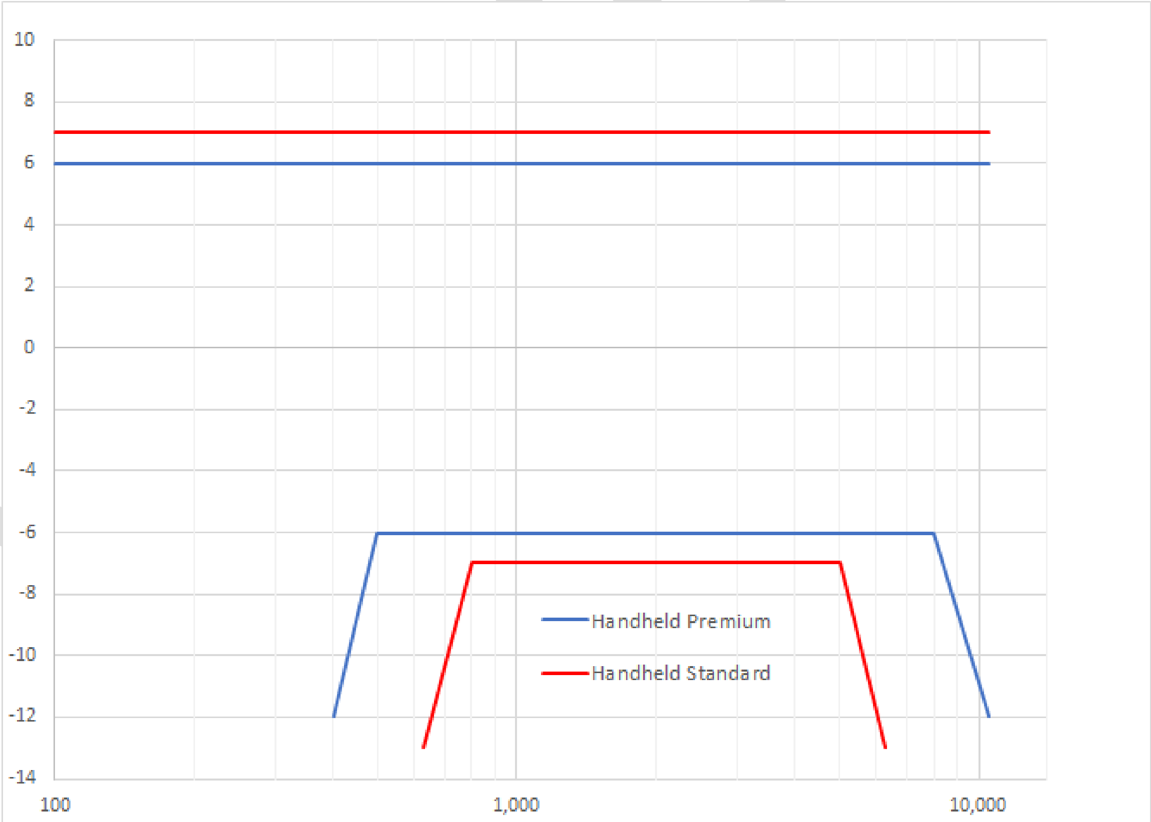


Figure 12: Handheld Speakerphone receive frequency response mask

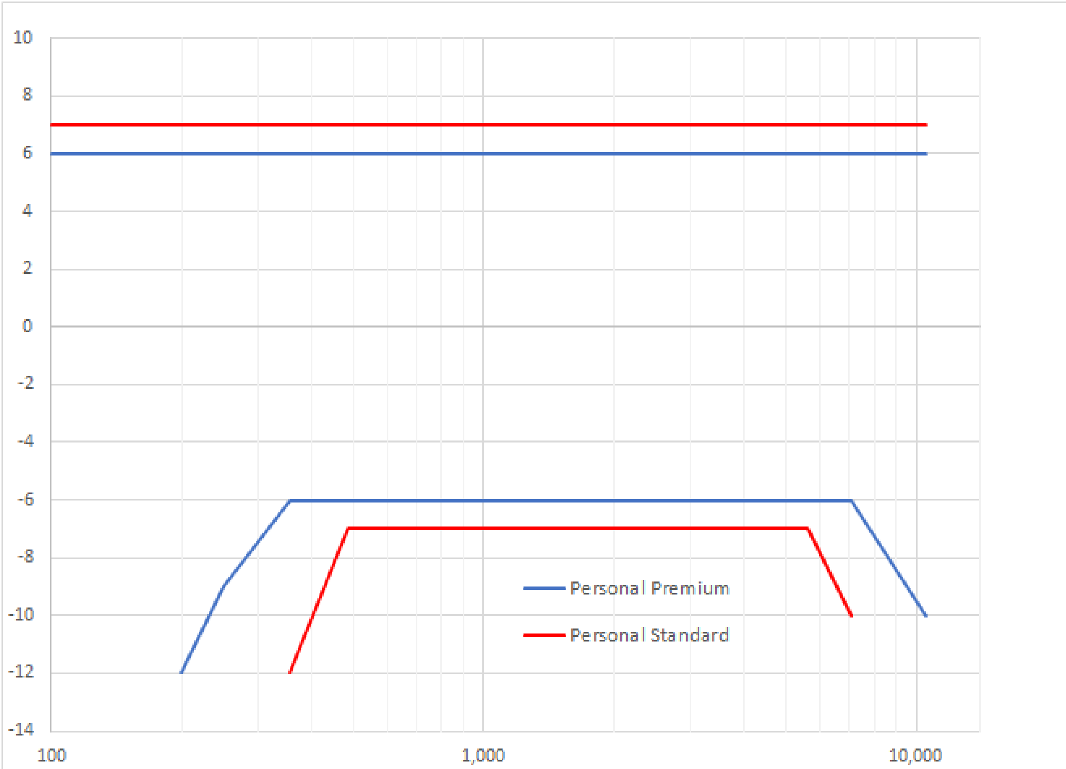


Figure 13: Personal speakerphone receive frequency response mask

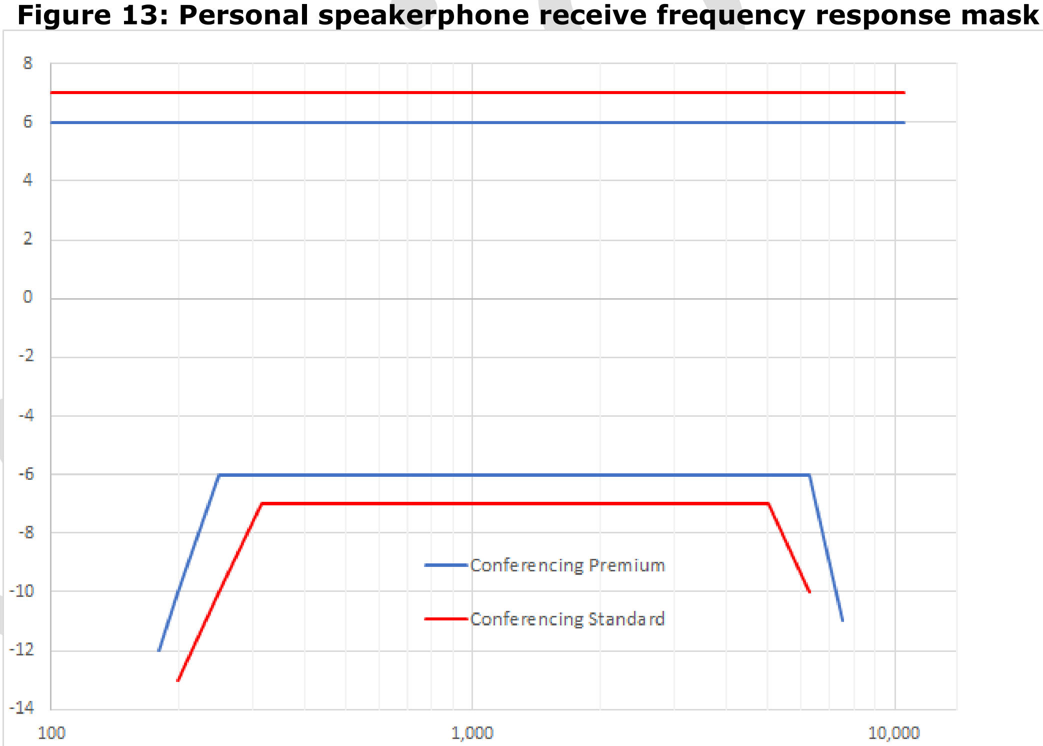


Figure 14: Conferencing device receive frequency response mask

4.2.7.3 *Test Procedure*

- Use the recommended test position for the DUT as described in Section 5.2.4
- Use the same nominal receive output level as in Section 4.2.1.

- Measurements shall be done over a range of 100 Hz through 10500 Hz:
    - $1/12^{th}$ octave bands for handset and headset requirements.
    - $1/3^{rd}$ octave bands for portable device and speakerphone requirements to account for variability due to sound reflections.
- The test stimulus is the IEEE Male-Female signal for frequency response.
- For headset the headset shall be removed and put back on HATS for 3 times.
- For handset the frequency response shall be measured at 3 positions as described in 0 using
    - Angle A 21 deg, angle C – default or manufacturer recommendation.
    - Angle A 35 deg, and angle C in step 1 minus 5 degrees.
    - Angle A 15 deg, and angle C in step 1 plus 5 degrees.
- **For speakerphone devices frequency response shall be measured at 3 positions** as described in Section 5.1.8
- The final resulting frequency response compared against tolerance mask is a dB average in each band of the 3 measured responses.

4.2.8 Receive path – dynamic volume compression for quiet signals

4.2.8.1 *Purpose*

The DUT should not apply dynamic volume compensations to low level playback signals. This can cause an unwanted boosting of distractor speech or other noises from other call participants.

4.2.8.2 *Requirement*

Gain change of low level signals	
Standard (dB)	Premium (dB)
≤ 3	≤ 1

Table 29: Receive idle channel noise requirements

4.2.8.3 *Test Procedure*

- Use the recommended test position for the DUT as described in Section 5.2.4.
- Skype for Business DUT mode: Audio offloading - checked.
- To test this requirement, adjust the playback volume to the maximum playback level.
- Play the IEEE Male speech + low level sinus signals test signal to reference Microsoft Teams input and record the DUT output.
- Calculate the A-weighted RMS level of sine signals from 2.6 .. 2.9 seconds.
- Calculate the A-weighted RMS level of sine signals from 10.0 .. 10.5 seconds.
- Compare the level (gain) change between the first and second result to the required limit.

4.3 Echo path

4.3.1 Echo path - terminal coupling loss (TCL)

4.3.1.1 *Purpose*

The level of acoustic echo in the send path signal is measured by the terminal coupling loss – TCL. For devices with on-board AEC, a failure of this test indicates echo leaks that can be very disruptive to the far-end participants.

4.3.1.2 *Requirement*

This requirement applies to devices with both capture and render capabilities.

The TCL shall be normalized with respect to the nominal send output level to account for any analog gain difference. The nominal send output level is defined as -24dBFS. The formula for the normalized TCL is

$$\begin{aligned} TCL &= TCL_{measured} + (Send\ output\ level_{measured} - Send\ output\ level_{nominal}) \\ &= TCL_{measured} + (Send\ output\ level_{measured} - (-24dBFS)) \end{aligned}$$

Equation 3: TCL calculation

The TCL shall be measured at nominal receive output level as defined in Section 4.2.1 and the requirements in Table 30.

	TCL at nominal receive output level	
	Standard (dB)	Premium (dB)
Headset	>50	>56
Handset	>50	>56
Speakerphone UI categories	>50	>56

Table 30: TCL requirements for devices with on-board AEC

4.3.1.3 *Test Procedure*

- Use the recommended test position for the DUT as described in Section 5.2.4
- Use the same nominal receive output level as in Section 4.2.1.
- Run the test signal for AGC and AEC stabilization with the preparation signals specified in the table below:

	Preparation	Analysis
Near end input	speech	silence
Far end input	TCL test signal	TCL test signal

Table 31: Test vectors used in TCL test

- Play the Echo test signals as listed in Table 31.
- First 6 sentences are ignored, and the calculation is done over time of the remaining 6 sentences.

4.3.2 Echo path - EQUEST MOS at nominal playback volume

4.3.2.1 *Purpose*

Call interactivity and acoustic echo audibility is dependent on the round-trip delay. The TCL is a one-dimensional metric and does not take into account the time nor spectral echo characteristics. The EQUEST analysis tool uses a Relative Approach method by HEAD acoustics. This method allows better detection of temporal and spectral components and thus yields a better aurally adequate analysis of echo disturbance.

4.3.2.2 *Requirement*

	EQUEST MOS	
	Standard	Premium
Headset	≥ 4.0	≥ 4.2
Handset	≥ 4.0	≥ 4.2
Speakerphone UI categories	≥4.0	≥ 4.2

Table 32: EQUEST MOS requirements at nominal playback volume

4.3.2.3 *Test Procedure*

- Use the recommended test position for the DUT as described in Section 5.2.4
- Use the same nominal receive output level as in Section 4.2.1.
- Play the Echo test signals as listed in table below:

	Analysis
Near end input (Lch) Far end input (Rch)	EQUEST_ST_echo.wav

Table 33: Test vectors used in EQUEST WB MOS test

- Compare each calculated EQUEST WB MOS against the requirement

4.3.2.4 *Note*

More data about the EQUEST test method can be found at:
http://head-acoustics.de/downloads/eng/application_notes/telecom/Appl_note_EQUEST_e0.pdf
http://www.iwaenc.org/proceedings/2008/contents/papers/9067.pdf

4.3.3 Echo path - Echo Control Characteristics (ECC)

4.3.3.1 *Purpose*

The TCLw metric shows the terminal coupling loss of the AEC in far end single talk usage. EQUEST analysis also shows the AEC performance during single talk situations. In real world calls the echo cancellation performance is much more complex as achieving the good echo suppression could cause other impairments such as distortion or clipping in near end speech or echo leaks when talking turn changes. This test helps to identify the possible echo residuals during and active doubletalk scenarios with alternating talking turns. The test methods used for this test case are based on the ETSI/3GPP documents TS 126.131 and TS 126.132 (Section 6.2 – Echo control characteristics).

The effect of the echo leakage or impact to double-talk can be verified subjectively as outlined in ITU-T P.831 and P832.

4.3.3.2 *Requirement*

	Echo Control Characteristics
	Standard (%)

Category	ST class Segment 1&2	DT class Segment 1&2
A1	N/A	N/A
A2		
B	N/A	N/A
C	N/A	N/A
D	N/A	N/A
E	N/A	N/A
F	< 5	< 5
G	< 5	< 5

Table 34: Echo control characteristics requirements in anechoic room

4.3.3.3 *Test procedure*

- Use the recommended test position for the DUT as described in Section 5.2.4
- DUT client setup details in Section 5.1.4. **(note the disable Teams digital AGC test mode)**
- Use the same playback loudness setting as used in Section 4.2.1 in anechoic room testing.
- Play the Echo test signals as listed in the table below with a near end signal level as defined in Section 5.2.3:

	Single talk test run (for level reference)
Near end input (Lch) &Far end input (Rch)	ECC_ST.wav
	Double talk test run (for final analysis)
Near end input (Lch) & Far end input (Rch)	ECC_DT-wav

Table 35: Test vectors used in AEC performance test

- Recorded signal is Highpass filtered with 2nd order Butterworth filter at 180Hz to avoid low frequency room or microphone noise influence on test results.
- The 13.5-43.5sec as a segment 1 is analyzed and categorized (segment 1 includes also part of the pre-conditioning sequence with far end activity period).
- The 43.5-58.5sec as segment 2 is analyzed and categorized.
- Compare the resulting analysis data against the requirement

4.3.3.4 *Note*

The near end signal level used for Conferencing devices is the simulated speech level as defined in Section 5.2.3. The double talk categories A1+A2 are marked as N/A as the doubletalk attenuation is now tested with the test case “Echo path - send signal attenuation during doubletalk”.

4.3.4 Echo path – send signal attenuation during doubletalk

4.3.4.1 *Purpose*

The ability of a near end user to interrupt the far end user is an important aspect of a voice communication and especially so on meeting room oriented conferencing devices. The test method used follows the process defined in appendix of the ITU-T P.340 and P.502. The test result categorizes the devices by measuring the attenuation of the send signal. The resulting dB values can be mapped to the ITU-T P.340 doubletalk behaviors mapping.

This method does not provide directly linking results compared to the subjective evaluation but does provide an objective characterization of the DUT performance during double-talk.

4.3.4.2 *Requirement*

		ITU-T P.502 based analysis	
		Standard (dB)	Premium (dB)
		Segment 1 and Segment 2	Segment 1 and Segment 2
Headset		≤ 9	≤ 6
Open office headset		≤ 6	≤ 3
Handset		≤ 9	≤ 6
Personal speakerphone		≤ 15	≤ 12
Conferencing devices	up to 1.5m mic usage	≤ 15	≤ 12
	up to 2.3m mic usage	≤ 12	≤ 9
	up to 3.5m mic usage	≤ 15	≤ 12
	up to 4.5m mic usage	≤ 15	≤ 12

Table 36: Send signal attenuation during doubletalk

4.3.4.3 *Test procedure*

- Use the recommended test position for the DUT as described in Section 5.2.4
- DUT client setup details in Section 5.1.4. **(note the disable Teams digital AGC test mode)**
- The 23.5-43.5sec as a segment 1 is analyzed and categorized.
- The 43.5-58.5sec as segment 2 is analyzed and categorized.
- Compare the resulting analysis data against the requirement (both segments must pass the criteria).

4.3.4.4 *Note*

The near end signal level used for Conferencing devices is the simulated speech level as defined in Section 5.2.3

More data about the Automated Double Talk test method can be found at:
http://www.head-acoustics.de/downloads/eng/application_notes/telecom/Appl_note_AutDT_e0.pdf
For info - ETSI ES 202 740 classification:

≤ 3 - Full Duplex Capability
≤ 6 – Partial Duplex Capability
≤ 9 – Partial Duplex Capability
≤12 - Partial Duplex Capability
>12 – No Duplex Capability

4.3.5 Echo path - terminal coupling at max playback volume ("raw mode" only)

4.3.5.1 *Purpose*

The level difference of acoustic echo from loudspeaker compared to the near end signal in send path is measured by the terminal coupling (TC).

4.3.5.2 *Requirement*

This requirement applies only to devices with both capture and render capabilities which support the raw mode and do not have AEC or NS in DUT send path processing.

The TC shall be normalized with respect to the nominal send output level to account for any analog gain difference. The nominal send output level is defined as -24dBFS. The formula for the normalized TCL is

$$\begin{aligned} TC &= TC_{measured} + (Send\ output\ level_{measured} - Send\ output\ level_{nominal}) \\ &= TC_{measured} + (Send\ output\ level_{measured} - (-24dBFS)) \end{aligned}$$

Equation 4: TCL calculation

The TC shall be measured at nominal receive output level as defined in Section 4.2.1 and the requirements in Table 30.

	TC at nominal receive output level	
	Standard (dB)	Premium (dB)
Headset	>30	>36
Speakerphone UI categories	>-10	>-3

Table 37: TC requirements for raw mode devices

4.3.5.3 *Test Procedure*

- Use the recommended test position for the DUT as described in Section 5.2.4
- Use the same nominal receive output level as in Section 4.2.1.
- Run the test signal for AGC and AEC stabilization with the preparation signals specified in the table below:

	Preparation	Analysis
Near end input	silence	silence
Far end (input	TCL test signal	TCL test signal

Table 38: Test vectors used in TCL test

- Play the Echo test signals as listed in Table 31.
- First 6 sentences are ignored, and the calculation is done over time of the remaining 6 sentences.

4.3.6 Echo path – send level during conversation at max playback volume ("raw mode" only)

4.3.6.1 *Purpose*

The DUT device should still be able to provide adequate send speech level even when playback volume is maximized, and microphone gain adjusted to avoid echo signal overloading the input. This can be achieved via better case design, increasing the mic/speaker separation distance (physically bigger case), using directive microphones or other means.

4.3.6.2 *Requirement*

This requirement applies only to devices with both capture and render capabilities which support the "raw mode" and do not have AEC or NS in DUT send path processing.

The computed average RMS level in reference Microsoft Teams output for the speech sample must satisfy the requirement in the table below.

	Signal level with normal speech Analysis band 100Hz – 12kHz	
	Standard (dBFS)	Premium (dBFS)
Personal Speakerphone UI categories	[-38, -14]	[-28, -14], no overload

Note: dBm0 ≈ dBFS + 6dB!

Table 39: signal level with normal speech requirements for raw mode devices

4.3.6.3 *Test Procedure*

- Use the recommended test position for the DUT as described in Section 5.2.
- Use the same nominal receive output level as in Section 4.2.1.
- Play the Echo test signals as listed in table below:

	Analysis
Near end input (Lch) Far end input (Rch)	EQUEST_ST_echo.wav

Table 40: Test vectors used in EQUEST WB MOS test

- The near end test signal is played at normal speech level from the artificial mouth.
- Analyze the near end speech level between 23 …28 sec.

4.3.7 Echo path – echo signal overloading input at max playback level ("raw mode" only)

4.3.7.1 *Purpose*

The DUT microphone input level slider should allow enough adjustment range to avoid the echo signal overloading the input. Overloaded input creates a lot of extra non linearities that do not allow an optimal operation of the echo cancellation algorithm in client.

4.3.7.2 *Requirement*

This requirement applies only to devices with both capture and render capabilities which support the "raw mode" and do not have AEC or NS in DUT send path processing.

Compute the average RMS level in reference Microsoft Teams output for the full speech sample and check for send path overload/clipping. Max 2ms long section of audio samples with 0dBFS level are allowed during the test sample.

4.3.7.3 *Test Procedure*

- Use the recommended test position for the DUT as described in Section 5.2.
- Use the same nominal receive output level as in Section 4.2.1.
- Play the Echo test signals as listed in table below:

	Analysis
Near end input (Lch) Far end input (Rch)	EQUEST_ST_echo.wav

Table 41: Test vectors used in EQUEST WB MOS test

- Compare each calculated EQUEST WB MOS against the requirement

4.3.8 Echo path - EQUEST MOS - Max playback volume

4.3.8.1 *Purpose*

The maximum playback level and minimum gain in OS gain control (OSGC) should be tuned for every individual device in such a way that even when playback level is maximized the loudest peaks in the playback signals would not saturate/overload the microphone path. Also, the loudest allowed playback volume should not cause severe raise of distortion or over excursion of loudspeakers. This could cause AEC to leak.

4.3.8.2 *Requirement*

	EQUEST MOS	
	Standard	Premium
Headset	≥ 3.8	≥ 4.0
Handset	≥ 3.8	≥ 4.0
Speakerphone UI categories	≥ 3.5	≥ 3.7

Table 42: EQUEST MOS requirements at maximum playback volume

4.3.8.3 *Test Procedure*

- Use the recommended test position for the DUT as described in Section 5.2.4
- Adjust the playback volume control such that the maximal receive output level is achieved.
  - In case a UI warning appears then lower the volume until the warning message disappears.
- Run the test signal as described in Table 43 to allow stabilization of AGC and AEC.
- Play the Echo test signals as listed in the table below:

	Analysis
Near end input (Lch) Far end input (Rch)	EQUEST_ST_echo.wav

Table 43: Test vectors used in EQUEST WB MOS test at maximum playback loudness

- Compare each calculated EQUEST WB MOS against the requirement

4.3.9 Echo path - Echo Control Characteristics (ECC) – Max playback volume

4.3.9.1 *Purpose*

Identification of possible echo residuals during and active doubletalk scenarios with alternating talking turns.

4.3.9.2 *Requirement*

	Echo Control Characteristics Standard (%)	
Category	ST class Segment 1&2	DT class Segment 1&2
A1	N/A	N/A
A2		
B	N/A	N/A
C	N/A	N/A
D	N/A	N/A
E	N/A	N/A
F	< 8	< 8
G	< 8	< 8

Table 44: Echo control characteristics requirements in anechoic room

4.3.9.3 *Test procedure*

- Use the recommended test position for the DUT as described in Section 5.2.4
- DUT client setup details in Section 5.1.4. **(note the disable Teams digital AGC test mode)**
- Adjust the playback volume control such that the maximal receive output level is achieved.
  - In case a UI warning appears then lower the volume until the warning message disappears.
- Play the Echo test signals as listed in the table below:

	Single talk test run (for level reference)
Near end input (Lch) Far end input (Rch)	ECC_ST.wav
	Double talk test run (for final analysis)
Near end input (Lch) Far end input (Rch)	ECC_DT-wav

Table 45: Test vectors used in AEC performance test

- Analysis file includes a 23.5 second pre-conditioning sequence
- The 23.5-43.5sec as a segment 1 is analyzed and categorized.
- The 43.5-58.5sec as segment 2 is analyzed and categorized.
- Compare the resulting analysis data against the requirement

4.3.10 Echo path – send signal attenuation during doubletalk – Max playback volume

4.3.10.1 *Purpose*

The ability of a near end user to interrupt the far end user is an important aspect of a voice communication and especially so on meeting room oriented conferencing devices. The test method used follows the process defined in appendix of the ITU-T P.340 and P.502. The test result categorizes the devices by measuring the attenuation of the send signal. The resulting dB values can be mapped to the ITU-T P.340 doubletalk behaviors mapping.

This method does not provide directly linking results compared to the subjective evaluation but does provide an objective characterization of the DUT performance during double-talk.

4.3.10.2 *Requirement*

	ITU-T P.502 based analysis Standard (dB)
	Segment 1 and Segment 2
Headset	≤ 12
Open office headset	≤ 9
Handset	≤ 12
Personal speakerphone	≤ 30
Conferencing devices	≤ 28

Table 46: Send signal attenuation during doubletalk

4.3.10.3 *Test procedure*

- Use the recommended test position for the DUT as described in Section 5.2.4
- DUT client setup details in Section 5.1.4. **(note the disable Teams digital AGC test mode)**
- Analysis file includes a 23.5 second pre-conditioning sequence
- The 23.5-43.5sec as a segment 1 is analyzed and categorized.
- The 43.5-58.5sec as segment 2 is analyzed and categorized.
- Compare the resulting analysis data against the requirement.

4.3.10.4 *Note*

More data about the Automated Double Talk test method can be found at:
http://www.head-acoustics.de/downloads/eng/application_notes/telecom/Appl_note_AutDT_e0.pdf

For info - ETSI ES 202 740 classification:

≤ 3 - Full Duplex Capability
≤ 6 – Partial Duplex Capability
≤ 9 – Partial Duplex Capability
≤12 - Partial Duplex Capability
>12 – No Duplex Capability

4.3.11 Sidetone Masking Rating for headsets and handsets

4.3.11.1 *Purpose*

Sidetone is used to assure the user that the phone is working. It also provides a feedback mechanism to the users, so they are aware of any abnormal noise they may be sending to the far end, such as breathing noise.
The slightly louder sidetone is recommended for sealed type of headsets because the acoustic path from mouth to ear is heavily suppressed. Solutions with speakerphone UI shall not implement a sidetone.

4.3.11.2 *Requirement*

The Sidetone Masking Rating (STMR) is the loudness loss in dB of the path from the mouth to the ear when using the tested handset/headset. STMR is calculated from the ratio of the acoustic output signal from the receiver at the ear reference point (ERP) to the acoustic input signal at the mouth reference point (MRP) over the specified frequency band. Handsets/headsets with adjustable receive levels shall be tested at the minimum, nominal and maximum volume control settings. The STMR requirements are given in Table 47. When sidetone volume is exposed in the control panel, the default sidetone volume shall meet these requirements. No clipping shall be perceived for speech level of 99 dBSPL at MRP

	Side Tone Masking Rating (STMR)
	Standard (dB)
Headset (binaural) Headset (monaural)	[18, 28][15] [12, 24]
Handset	[12, 22]
Speakerphone UI categories	N/A

Table 47: STMR requirements

Notes:

- Handsets/headsets with smaller form factor i.e., with a larger distance between microphone and mouth usually require a larger gain in the send direction to compensate for the lower speech capture level. This also increases the background noise picked up by the microphone and, thus, also the background noise reproduced by the sidetone. For such devices the sidetone level should be reduced appropriately within the requirements given in Table 47 to prevent high background noise in the receiver due to the sidetone.
- Headsets with tight or sealed ear coupling (e.g. hard-cap or sealed insert headphones) or active noise cancellation will reduce the natural acoustic sidetone both during the call and when not in call but when render audio is not active. Therefore, for such devices it is recommended to implement a context aware sidetone which is up to 3dB louder than loosely coupled headphones but shall still fall within the STMR range given in Table 47.
- Measurement of a low STMR for loosely coupled handsets/headsets is difficult with the HATS due to the acoustic coupling between HATS mouth and ear. A reference of the maximum measurable STMR for a certain handset/headset can be obtained by measuring the handset/headset without plugging it into the PC and thus disabling the sidetone of the handset/headset.
- The receive signal is transformed from DRP to ERP for STMR measurements.

[15] If DUT has active noise cancelling mode for receive signal (earpieces) then additional 7dB STMR loss is allowed.

4.3.12 Sidetone Latency for headsets and handsets

4.3.12.1 *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.

If sidetone is supported, then the requirements below shall be met. The sidetone is strongly recommended for sealed type of headsets because the acoustic path from mouth to ear is heavily suppressed. Solutions with speakerphone UI shall not implement a sidetone.

4.3.12.2 *Requirement*

The sidetone delay for both handset and headset shall be less than 5 ms. It is desirable for the sidetone delay to be less than 1ms. Solutions with speakerphone UI shall not implement a sidetone.

	Sidetone latency
	Standard (ms)
Headset	< 5
Handset	< 5
Speakerphone UI categories	N/A

Table 48: Sidetone latency requirements

4.3.12.3 *Test Procedure*

Refer to IEEE Std. 269 for the test procedure.

4.3.12.4 *Note*

It is allowed to deduct the HEAD front end loopback latency from the measured result.

4.4 Tests in reverberant room

The device to be tested might be used in practice in very different environments. These could range from a quiet but reverberant all the way to open space with a loud background noise.

For personal devices simulated background noise environment based on the ETSI EG 202 396-1 is used to estimate the devices ability to keep a good send signal speech quality.

For conferencing devices a single sound source is used for background noise playback as that is more realistic for meeting room scenarios.

4.4.1 Send path – send quality in the presence of ambient noise

4.4.1.1 *Purpose*

For personal devices the quality of the speech signal transmission in noisy environments is becoming more and more important, especially with tablets and laptops that people carry with them. Users can easily make voice calls from a café or a car due to wireless broadband becoming mainstream. For personal devices, the speech quality is objectively tested in such scenarios using a test setup outlined in ETSI EG 202 396-1. The analysis is done based on the ETSI TS 103 106 specification.

Similarly, conferencing devices need to be able to operate in presence of common meeting room noises such as noise from projector or a HVAC systems in the room.

4.4.1.2 *Requirement*

		3QUEST results	
		Standard	Premium
Headset Handset[16]	S-MOS (average score of tested noise cases)	≥ 3.5	≥ 3.8
	N-MOS (average score of tested noise cases)	≥ 3.0	≥ 3.2
	G-MOS (average score of tested noise cases)	-	-
	S-MOS min (lowest score of tested noise cases)	≥ 3.0	≥ 3.3
	N-MOS min (lowest score of tested noise cases)	≥ 2.5	≥ 2.8
	G-MOS min (lowest score of tested noise cases)	-	-
Open Office Headset[16]	S-MOS (average score of tested noise cases)	≥ 3.5	≥ 3.8
	N-MOS (average score of tested noise cases)	≥ 3.5	≥ 3.8
	G-MOS (average score of tested noise cases)	-	-
	S-MOS min (lowest score of tested noise cases)	≥ 3.2	≥ 3.5
	N-MOS min (lowest score of tested noise cases)	≥ 3.2	≥ 3.5
	G-MOS min (lowest score of tested noise cases)	-	-
Personal Speakerphone	S-MOS (average score of tested noise cases)	≥ 3.5	≥ 3.8
	N-MOS (average score of tested noise cases)	≥ 1.8	≥ 2.5
	G-MOS (average score of tested noise cases)	-	-
	S-MOS min (lowest score of tested noise cases)	≥ 3.2	≥ 3.5
	N-MOS min (lowest score of tested noise cases)	-	≥ 2.1
	G-MOS min (lowest score of tested noise cases)	-	-

[16] For Headsets and Open Office Headsets with microphone boom this test is run at 3 different microphone boom positions as described in 5.2.4.1

		3QUEST results	
		Standard	Premium
Conference speakerphone	S-MOS (average score of tested noise cases)	≥ 3.5	≥ 3.8
	N-MOS (average score of tested noise cases)	≥ 2.9	≥ 3.2
	G-MOS (average score of tested noise cases)	-	-

Table 49: Speech quality requirements in presence of background noise

Individual test run S-MOS, N-MOS and G-MOS are calculated over 16 sentences for each background noise type.

The average final MOS score is reported with deviation for each of the noise types applicable.

Noise type	Pub Noise Binaural	Outside Traffic Road Binaural	Outside Traffic Crossroads Binaural	Train Station Binaural	Fullsize Car1 130Km/h Binaural	Cafeteria Noise Binaural	Mensa Binaural	Work Noise Office Callcenter Binaural	Male Single Voice Distractor Binaural	Conference room noise	Projector Noise	Speech level at HATS MRP (dBSPL)
Headset							x	x	x			89
Open Office Headset							x	x	x			89
Handset – desk phone							x	x	x			89
Handheld speakerphone				x		x	x	x	x			89
Handset – mobile	x	x	x	x	x	x	x	x				89+3
Personal speakerphone [17]				x		x	x	x	x			89
Conferencing Speakerphone										x	x	89
Collaboration: meeting room										x	x	89
Collaboration: huddle / lounge space[18]							x	x	x	x	x	89

Table 50: HAE-BGN Background noises used for different device categories

4.4.1.3 *Test Procedure Personal Devices*

- Follow the background noise system setup guidelines for the respective device category.
- Use the recommended test position for the DUT as described in Section 5.3.2
    - Open Office Headset will be tested with 3 different microphone boom positions as shown in 5.3.2.3
- Start the background noise is sync with the measurement.
- Play the preparation test signals (minimum of 4 sentences).
- Play the 16-sentence test signal.

---

[17] If the DUT device is clearly meant to be used indoors of an office or at home then the same noise types as for Headset – open office are used

[18] The HAE-BGN test setup is configured so that the back speakers are muted. Playback levels used are not altered in BGN system. See Section 5.3.3.8 for setup details

- Record and analyze the resulting average MOS against the requirement.

The background noise files can be found at

- http://docbox.etsi.org/stq/Open/EG%20202%20396-1%20Background%20noise%20database/Binaural_Signals/

4.4.1.4 *Test Procedure Conferencing Devices*

- Use the recommended test position for the DUT as described in Section 5.3.3.2.
    - Talker1/Position1 is used for near end signal playback
    - Talker2/Position2 is used playback source for projector and Hoth noise
- MUTE DUT playback for this test.
- Play the Near end + projector noise test signal.
    - Noise level at DUT position shall be 46 ± 1 dBSPL(A)
- Play the Near end + ETSI 103 224 "Conference3_handsfree.wav" noise test signal.
    - Noise level at DUT position shall be 42 ± 1 dBSPL(A)
- Record and analyze the resulting 3QUEST MOS against the requirement.

4.4.2 Send path - speech quality for alternating near end talkers (conferencing devices only)

4.4.2.1 *Purpose*

The speakerphone product built for conferencing could use special processing for microphone signal to create a directional pickup pattern or use special processing to reduce the effect of room reverberations, suppress background noise etc. Such processing is allowed but should perform well in real world use cases and not only in anechoic environment.

The test uses two near end talkers with different mouth to microphone distances.

HEAD acoustics 3QUEST objective speech quality test tool is used to measure the Speech MOS metric for this test case.

4.4.2.2 *Requirement*

	3QUEST S-MOS target (position1 and position2 results must both pass)	
	3QUEST S-MOS [19]	
Condition	Standard	Premium
Talker 1 at 1m (normal speech level)	≥4.0	≥4.2
Talker 2 at 1.5m (normal speech level)	≥3.8	≥4.0

Table 51: Speech pickup requirements

4.4.2.3 *Test Procedure Conferencing Devices*

- Use the recommended test position for the DUT as described in Section 5.3.3.2.
- Use a test signal with altering talking turns between the two near end talkers and a far end signal.
- Analyze the S-MOS for consecutive Talker 1 and then Talker 2 test signal recordings.
- Alter the position of artificial mouth and HATS as described in Section 5.3 and rerun the test until all alternatives are covered for the specific type of device.
- Compare each calculated EQUEST WB MOS against the requirement

4.4.3 Send path – speech level for alternating near end talkers (conferencing devices only)

4.4.3.1 *Purpose*

The speakerphone product built for conferencing could use special processing for microphone signal to create a directional pickup pattern or use special processing to reduce the effect of room reverberations, suppress background

[19] In prior test specification the 3Quest tests were performed at 1, 2, and 4M distances, but results for the longer distances didn't correlate well to real world performance. Longer range performance is now covered by tests in anechoic chamber.

noise etc. Such processing is allowed but should perform well in real world use cases and not only in anechoic environment.

The test uses two near end talkers with different mouth to microphone distances. The speech signal send level is measured during activity of the near end talkers.

4.4.3.2 *Requirement*

The computed average RMS level in reference Microsoft Teams output for the speech sample must satisfy the requirement in the table below.

	Send signal level requirement with normal speech level at mouth MRP	
Talker	Standard (dBFS)	Premium (dBFS)
Talker 1	[-28, -14]	[-26, -14]
Talker 2	[-31, -14]	[-26, -14]

Note: dBm0 ≈ dBFS + 6dB!

Table 52: signal level with normal speech requirements

4.4.3.3 *Test Procedure Conferencing Devices*

- Use the recommended test position for the DUT as described in Section 5.3.3.2.
- Use a test signal with altering talking turns between the two near end talkers and a far end signal.
- Alter the position of artificial mouth and HATS as described in Section 5.3 and rerun the test until all alternatives are covered for the specific type of device.

4.4.4 Receive path - output level in reverberant room

4.4.4.1 *Purpose*

The loudspeaker playback level for DUT is user adjustable. Here the tester must set the playback to required listening level at 1m in a reverberant test room for the following Echo path test.

4.4.4.2 *Requirement*

Document the nominal loudness playback level in HATS ear when the same playback loudness setting is used as set in Section 4.2.1 in anechoic room testing. In case the level is louder than the requirement targets in reverberant room, then decrease the playback level on DUT enough to fit the target range. Document the new setting in final test report.

		Receive output level target Standard (dBSPL)
Headset monaural Headset binaural		76 ± 2 70 ± 2
Handset		76 ± 2[20]
Handheld speakerphone Personal speakerphone		65 ± 2[21]
Conferencing devices*	up to 1.5m speaker usage	67 -1/+2
	up to 2.3m speaker usage	69 ± 2
	up to 3.5m speaker usage	71 ± 2
	up to 4.5m speaker usage	73 ± 2
	up to 7.5m speaker usage[22]	76 -1/+2
Note * for modular solutions where speaker is not at main DUT position the measurement microphone is placed at 1m in front of a speaker for this measurement. In case of 2 or more separate speaker units each speaker can be set to 3dB lower level than above requirement.		

Table 53: Receive output level requirements – reverberant room

4.4.4.3 *Test Procedure Personal Devices*

- Use the recommended test position for the DUT as described in Section 5.3.2
- The test stimulus is the IEEE 269-2010 uncompressed male speech signal pre-processed for active speech level of -18dBFS.
- Measure the receive RMS output level in dBSPL(C) by the artificial ear/measurement microphone.
- If needed adjust the level and rerun the test to fit into the requirement range.

4.4.4.4 *Test Procedure Conferencing Devices*

- Use the test position Talker 1 / Position 1 for the DUT as described in Section 5.3.3.2.
- The test stimulus is the IEEE 269-2010 uncompressed male speech signal pre-processed for active speech level of -18dBFS.

[20] For handsets a receive loudness rating (RLR) can be measured in addition to the receive level in dB. For handsets passing a wideband RLR criteria of 2 +/-3 measured according to ITU-T Rec. P.79 can be considered a pass in test report, over-ruling the dBSPL value. DRP-ERP correction is used for the RLR test!

[21] For laptop in tent mode or a tablet in kickstand mode, the requirement can be relaxed by upto 3dB due to level drop caused by increased usage distance.

[22] Waivers for doubletalk/echo performance will not be allowed for devices that attempt 7.5m range speaker loudness. Devices that struggle should lower loudness levels and will be considered only for smaller room size based on 4.5m range.

- Measure the receive RMS output level in dBSPL(C) by the artificial ear/measurement microphone. If needed adjust the level and rerun the test to fit into the requirement range.

4.4.5 Echo path - EQUEST MOS in reverberant room

4.4.5.1 *Purpose*

Acoustic echo audibility can influence a call quality in significant way. A real environment poses higher requirements for AEC algorithms due to room reverberation. To verify a good performance in real world like environment the EQUEST analysis tool is used. EQUEST uses a Relative Approach method by HEAD acoustics. This method allows better detection of temporal and spectral components and thus yields a better aurally adequate analysis of echo disturbance.

4.4.5.2 *Requirement*

	EQUEST MOS	
	Standard	Premium
Headset	≥ 4.0	≥ 4.2
Handset	≥ 4.0	≥ 4.2
Speakerphone UI categories	≥ 3.9	≥ 4.1
Conferencing devices	≥ 3.9	≥ 4.1

Table 54: EQUEST MOS requirements

4.4.5.3 *Test Procedure Personal Devices*

- Use the recommended test position for the DUT as described in Section Section 5.3.2
- Use the nominal receive output level as in Section 4.4.1
- Play the Echo test signals as listed in table below:

	Analysis
Near end input (Lch) & Far end input (Rch)	EQUEST_ST_echo.wav

Table 55: Test vectors used in EQUEST MOS test

- Compare each calculated EQUEST WB MOS against the requirement. If any one run fails to meet the criteria, then the end result is marked “Fail”.

4.4.5.4 *Test Procedure Conferencing Devices*

- Use the recommended test position for the DUT as described in Section 5.3.3.2.
  Note that the test is to be run multiple times to cover all the given Talker 1 and Talker 2 positions for the respective device type.
- Use the nominal receive output level as in Section 4.4.1
- Play the Echo test signals as listed in table.

	Analysis 1
Near end Talker 1 input	Equest_M_F_test.wav
Near end Talker 2 input	
Far end input (played via HAE-BGN system)	Equest_FE_input.dat
	Analysis 2
Near end Talker 1 input	Equest_F_M_test.wav
Near end Talker 2 input	
Far end input (played via HAE-BGN system)	Equest_FE_input.dat

Table 56: Test vectors used in EQUEST MOS test

- Recorded signal is High pass filtered with 2nd order Butterworth filter at 80Hz to avoid low frequency room or microphone noise influence on test results.

- Compare each calculated EQUEST WB MOS against the requirement. If any one run fails to meet the criteria, then the end result is marked “Fail”.

4.4.6 Echo path - Echo Control Characteristics (ECC)

4.4.6.1 *Purpose*

EQUEST analysis shows the AEC performance during single talk situations. In real world calls the echo cancellation performance is much more complex as achieving the good echo suppression could cause other impairments such as distortion or clipping in near end speech or echo leaks when talking turn changes. This test helps to identify the possible echo residuals during and active doubletalk scenarios with alternating talking turns. The test methods used for this test case are based on the ETSI/3GPP documents TS 126.131 and TS 126.132 (Section 6.2 – Echo control characteristics).

The effect of the echo leakage or impact to double-talk can be verified subjectively as outlined in ITU-T P.831 and P832.

4.4.6.2 *Requirement*

	Echo Control Characteristics	
	Category	Standard (%)
Headset and Handset	A1	N/A
	A2	
	B	N/A
	C	N/A
	D	N/A
	E	N/A
	F	<2
	G	<2
Speakerphone UI devices	A1	N/A
	A2	
	B	N/A
	C	N/A
	D	N/A
	E	N/A
	F	<5
	G	<5

Table 57: Echo control characteristics in reverberant room

4.4.6.3 *Test Procedure Personal Devices*

- Use the recommended test position for the DUT as described in Section Section 5.3.2
- DUT client setup details in Section 5.1.4. **(note the disable Teams digital AGC test mode)**
- Use the nominal receive output level as in Section 4.4.1
- Play the Echo test signals as listed in the table below:

	Single talk test run (for level reference)
Near end input (Lch) Far end input (Rch)	ECC_ST.wav
	Double talk test run (for final analysis)
Near end input (Lch) Far end input (Rch)	ECC_DT-wav

Table 58: Test vectors used in AEC performance test

- Recorded signal is Highpass filtered with 2nd order Butterworth filter at 180Hz to avoid low frequency room or microphone noise influence on test results.
- Analysis file includes a 23.5 second pre-conditioning sequence.
- The 23.5-43.5sec as a segment 1 is analyzed and categorized.
- The 43.5-58.5sec as segment 2 is analyzed and categorized.
- Compare the resulting analysis data against the requirement

4.4.6.4 *Test Procedure Conferencing Devices*

- Use the recommended test position for the DUT as described in Section 5.3.3.2.
- Use the nominal receive output level as in Section 4.4.1
- Play the Echo test signals as listed in the table below:

	Single talk test run (for level reference)
Near end input (Lch) Far end input (Rch)	ECC_ST.wav
	Double talk test run (for final analysis)
Near end input (Lch) Far end input (Rch)	ECC_DT-wav

Table 59: Test vectors used in AEC performance test

- Recorded signal is high pass filtered with 2nd order Butterworth filter at 80Hz to avoid low frequency room or microphone noise influence on test results.
- Analysis file includes a 23.5 second pre-conditioning sequence.
- The 23.5-43.5sec as a segment 1 is analyzed and categorized.
- The 43.5-58.5sec as segment 2 is analyzed and categorized.
- Compare the resulting analysis data against the requirement.

4.4.7 Echo path – AEC convergence time at call start

4.4.7.1 *Test Purpose*

When a call is started it is natural that echo cancellation needs a short time to converge to the echo path. This convergence should happen is a short time not to cause a poor user experience. This test gives a level versus time analysis output that can be used to objectively evaluate the AEC convergence time.

4.4.7.2 *Requirements*

- Analyze the level versus time plot. Does the send path signal level fall under -60dBFS (-66dBm0) in less than 5 seconds after the send audio path opens in a call.

4.4.7.3 *Test process*

- Play the Echo test signals as listed in Table 31.
- As fast as possible set up a call between REF and DUT clients. Call should be started in <15seconds to get a valid analysis plot. If needed the test can be repeated multiple times using two testers or a remote control of the DUT to answer the call.

4.4.8 Echo path – stability loss with variable echo path

4.4.8.1 *Test Purpose*

In real life calls people might adjust the laptop screen or move in front of the DUT. This causes a variable echo path that the DUT must be able to adapt to fast. This test will evaluate the DUT performance in handling small changes in echo path

4.4.8.2 *Requirements*

This test should be run on all Speakerphone UI devices.

	TCL delta to non-variable echo path	
	Standard	Premium
Headset	NA	NA
Handset	NA	NA
Personal Speakerphone UI categories	< 9dB	< 6dB
Conferencing device categories	< 6dB	< 3dB

Table 60: TCL requirement with variable echo path in reverberant room

4.4.8.3 *Test Procedure Personal Devices*

- Use the test setup 5.3.2.8
- Use the nominal receive output level as in Section 4.4.1
- Run the first TCL test with static echo path for echo attenuation level reference
- Run the second test with the rotating reflecting surface moving 90°/sec (fast position change followed by a stable period)
- Check the highest delta from the level versus time curves and compare to the requirement

4.4.8.4 *Test Procedure Conferencing Devices*

- Use the test setup 5.3.3.11
- Use the nominal receive output level as in Section 4.4.1
- Run the first TCL test with static echo path for echo attenuation level reference
- Run the second test with the rotating reflecting surface moving 90°/sec (fast position change followed by a stable period)
- Check the highest delta from the level versus time curves and compare to the requirement

4.4.9 Receive path – maximum output level in reverberant room

4.4.9.1 *Purpose*

Different DUT devices might have a different headroom for maximum loudness. This testcase

4.4.9.2 *Requirement*

		Maximum Receive output level Standard (dBSPL(C))
Headset monaural Headset binaural		N/A
Handset		N/A
Handheld speakerphone Personal speakerphone		[65, 75]
Conferencing devices*	up to 1.5m speaker usage	[68, 73]
	up to 2.3m speaker usage	[69, 74]
	up to 3.5m speaker usage	[71, 76]
	up to 4.5m speaker usage	[73, 78]
	up to 7.5m speaker usage	[76, 81]
Note * for modular solutions where speaker is not at main DUT position the measurement microphone is placed at 1m in front of a speaker for this measurement.		

Table 61: Receive output level requirements

Exceeding the upper output level requirement can be waivered if all Echo path maximum level tests pass the requirements with a clear margin.

4.4.9.3 *Test Procedure Personal Devices*

- Use the recommended test position for the DUT as described in Section 5.3.2
- Set the DUT device to maximum playback level.
- To test this requirement, adjust the playback volume to the maximal playback level.
- The test stimulus is the IEEE 269-2010 male speech signal <u>with active speech level of -18dBFS</u>.
- Measure the receive RMS output level in dBSPL by the artificial ear/measurement microphone.

4.4.9.4 *Test Procedure Conferencing Devices*

- Use the test position Talker 1 / Position 1 for the DUT as described in Section 5.3.3.2.
- To test this requirement, adjust the playback volume to the maximal playback level.
- The test stimulus is the IEEE 269-2010 male speech signal <u>with active speech level of -18dBFS</u>.

- Measure the receive RMS output level in dBSPL by the artificial ear/measurement microphone.

4.4.10 Echo path - EQUEST MOS - Max playback volume

4.4.10.1 *Purpose*

The acceptable echo cancellation quality should be achieved even in worst case scenario of maximum playback loudness. A good device should have a limit for maximum loudness to avoid echo leakage or to show an UI warning on device or on screen to inform the user in case the playback level is too loud and could impact the call quality.

4.4.10.2 *Requirement*

	EQUEST MOS	
	Standard	Premium
Headset	≥ 4.0	≥ 4.2
Handset	≥ 4.0	≥ 4.2
Speakerphone UI categories	≥ 3.5	≥ 3.9

Table 62: EQUEST MOS requirements at maximum playback loudness in reverberant room

4.4.10.3 *Test Procedure Personal Devices*

- Use the recommended test position for the DUT as described in Section 5.3.2
  Use the playback volume level set in Section 4.4.8.2.
- Run the test signal to allow stabilization of AGC and AEC with the same test stimulus as below.
- Play the Echo test signals as listed in table.

	Analysis
Near end input (Lch) Far end input (Rch)	EQUEST_ST_echo.wav

Table 63: Test vectors used in EQUEST MOS test at maximum playback loudness

- Compare each calculated EQUEST WB MOS against the requirement

4.4.10.4 *Test Procedure Conferencing Devices*

- Use the test position Talker 1 / Position 1 for the DUT as described in Section 5.3.3.2.
- Adjust the playback volume control such that the maximal receive output level is achieved.
    - In case a UI warning appears then lower the volume until the warning message disappears (provided volume level is still above nominal render volume).
- Run the test signal to allow stabilization of AGC and AEC with the same test stimulus as below.
- Play the Echo test signals as listed in table.

•	Analysis
Near end Talker 1 input	Equest_M_F_test.wav
Near end Talker 2 input	
Far end input (played via HAE-BGN system)	Equest_FE_input.dat

Table 64: Test vectors used in EQUEST MOS test

- Compare each calculated EQUEST WB MOS against the requirement

4.4.11 Echo path - Echo Control Characteristics (ECC) – Max playback volume

4.4.11.1 *Purpose*

EQUEST analysis shows the AEC performance during single talk situations. In real world calls the echo cancellation performance is much more complex as achieving the good echo suppression could cause other impairments such as distortion or clipping in near end speech or echo leaks when talking turn changes. This test helps to identify the possible echo residuals during and active doubletalk scenarios with alternating talking turns. The test methods used for this test case are based on the ETSI/3GPP documents TS 126.131 and TS 126.132 (Section 6.2 – Echo control characteristics).

The effect of the echo leakage or impact to double-talk can be verified subjectively as outlined in ITU-T P.831 and P832.

4.4.11.2 *Requirement*

	Echo Control Characteristics	
	Category	Standard (%)
Headset and Handset	A1	N/A
	A2	
	B	N/A
	C	N/A
	D	N/A
	E	N/A
	F	<4
	G	<4
Speakerphone UI devices	A1	N/A
	A2	
	B	N/A
	C	N/A
	D	N/A
	E	N/A
	F	<10
	G	<10

Table 65: Echo control characteristics in reverberant room

4.4.11.3 *Test Procedure Personal Devices*

- Use the recommended test position for the DUT as described in Section Section 5.3.2
- Adjust the playback volume control such that the maximal receive output level is achieved.
    - In case a UI warning appears then lower the volume until the warning message disappears.
- Play the Echo test signals as listed in the table below:

	Single talk test run (for level reference)
Near end input (Lch) Far end input (Rch)	ECC_ST.wav
	Double talk test run (for final analysis)
Near end input (Lch) Far end input (Rch)	ECC_DT-wav

Table 66: Test vectors used in AEC performance test

- Recorded signal is Highpass filtered with 2nd order Butterworth filter at 80Hz to avoid low frequency room or microphone noise influence on test results.

- Analysis file includes a 23.5 second pre-conditioning sequence
- The 23.5-43.5sec as a segment 1 is analyzed and categorized.
- The 43.5-58.5sec as segment 2 is analyzed and categorized.
- Compare the resulting analysis data against the requirement

4.4.11.4 *Test Procedure Conferencing Devices*

- Use the recommended test position for the DUT as described in Section 5.3.3.2.
- Adjust the playback volume control such that the maximal receive output level is achieved.
    - In case a UI warning appears then lower the volume until the warning message disappears.
- Play the Echo test signals as listed in the table below:

	Single talk test run (for level reference)
Near end input (Lch) Far end input (Rch)	ECC_ST.wav
	Double talk test run (for final analysis)
Near end input (Lch) Far end input (Rch)	ECC_DT-wav

Table 67: Test vectors used in AEC performance test

- Recorded signal is high pass filtered with 2nd order Butterworth filter at 80Hz to avoid low frequency room or microphone noise influence on test results.
- Analysis file includes a 23.5 second pre-conditioning sequence.
- The 23.5-43.5sec as a segment 1 is analyzed and categorized.
- The 43.5-58.5sec as segment 2 is analyzed and categorized.
- Compare the resulting analysis data against the requirement.

4.4.12 Distractor attenuation for open office headsets

4.4.12.1 *Purpose*

The purpose of this test is to verify the send path quality in an open office environment and ability to suppress nearby talkers. A good open office headset will provide strong attenuation to all sound beside the DUT user speech. For best performance, a long microphone boom can ease the engineering of suitable headsets as it brings the microphone(s) closer to talker's mouth.

4.4.12.2 *Requirement*

The distractor speech source is placed at 60cm distance from HATS MRP. The levels from HATS mouth are set to the normal speech level. The level of the secondary distractor mouth is adjusted so that the near end speech to distractor level ratio is 16dB at HATS MRP (distractor 16dB quieter).

$$SDR_{send} = SND_{nearend\ speech\ level\ at\ MRP} - SND_{distractor\ only\ level\ at\ MRP}$$

$$SDR_{reference} = REF_{nearend\ speech\ level\ at\ MRP} - REF_{distractor\ speec\ level\ at\ MRP}$$

$$SDR_{DUT} = SDR_{send} - SDR_{reference}$$

Equation 5: Distraction attenuation compared to single omnidirectional microphone at MRP

			Speech to distractor attenuation SDR (dB)	
			For Open Office designation (dB)	Premium (dB)
Headset type	Distractor type		NA	NA
Open office headset	Single Speech Distractor	Average of all angles	≥ 17	≥ 23
		Minimum of all angles	≥ 14	≥ 20

Table 68: Distractor attenuation requirements for open office headset

The numeric requirements are based on user study that showed 75% and 90% of listeners tolerate the distractor at the standard and premium attenuation level respectively.

4.4.12.3 *Test Procedure*

- DUT client setup details in Section 5.1.4. **(note the disable Teams digital AGC test mode)**
- Use the recommended test position for the DUT as described in Section 5.3.2.4.
    - A reference boom position where boom is directed towards mouth is used for this test.
- Leave >30 second stabilization time by playing the same source signal before starting the measurement.
- Alter the position of artificial mouth as described in Section 5.3.2.4 and rerun the test until all alternative five angles for the distractor source are done.
- Record signals from Reference client output and from Reference microphone
- Calculate the SDR_{send} by comparing the signal level during near end speech to the signal level when only distractor is active.
- Calculate the $SDR_{reference}$ by comparing the Reference microphone signal level during near end speech to the signal level when only distractor is active.
- Calculate the SDR_{DUT} for each angle
    - Calculate the average and minimum SDR for all angles

5 Audio test setup

The test setups are outlined in following paragraphs. In some cases, the setups and setting follow the ITU-T, ETSI, TIA and IEEE standards and recommendations. In other cases, Microsoft Teams testing uses a custom way to set up the device under test and test equipment.

5.1 Test equipment and details

Official test results must be produced by the approved certification test labs. If partners build their own labs for pre-testing, the test environment for Sections 4.1 to 4.3 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 29dBSPL(A).
- The room reverberation time (RT60) should be below 0.3 seconds at frequencies between 150 Hz to 12 kHz.

5.1.1 Test setup hardware/software availability.

5.1.1.1 *Acoustic Test Equipment and automation*

Skype has partnered with HEAD acoustics to offer a bundle of the needed hardware and software tools to carry out anechoic and reverberant room testing of Microsoft Teams devices against the requirements above.

HEAD acoustics GmbH
Sales Telecom
Tel +49 2407 577-0
e-mail: info@head-acoustics.de
Web https://www.head-acoustics.com/eng/contact.htm

5.1.1.2 *Cortana tools and automation*

Partner should contact their Microsoft account representative to get access to Cortana specifications and Tools. These tools are distributed via a Microsoft specification and software distribution tool called Collaborate. You will need to identify a tenant administrator (i.e., at your company) to manage access for people within your company, and for Cortana team to approve access to the specific content. Partner is expected to acquire and test to all Cortana specifications and tools well in advance of submitting for Teams certification in order to avoid delays. Due to the close relationship with Windows, Cortana specification and tools are released on a cadence that is independent of the Microsoft Teams specification updates and partner should test using the latest Cortana specification and tools unless Microsoft indicates otherwise.

5.1.2 List of equipment

- **HEAD acoustics ACQUA software – automated audio testing system with**
  - **Skype Specification option package**
  - **ACOPT 21 – 3QUEST**
  - **ACOPT 29 – EQUEST**
  - **ACOPT 30 – POLQA**
  - **ACOPT 32 – Speech Based DT (TS 26 131 and P.502 real speech methods used)**

  http://www.head-acoustics.de/eng/telecom_acqua.htm
  http://www.head-acoustics.de/eng/telecom_acqua_acopt.htm
  http://www.head-acoustics.de/eng/telecom_measurement_standards.htm
- **HEAD acoustics Measurement Front End (one of the below options needed)**
  - **MFE VI.I -** http://www.head-acoustics.de/eng/telecom_acqua_mfe_VI_1.htm
  - **labCORE (consult HEAD for best set of optional modules)**
- **G.R.A.S. 44AB Mouth simulator (or other P.51 compliant mouth simulator).**
  http://www.gras.dk/products/ear-simulator/mouth-simulators/product/281-44ab
- **G.R.A.S. 40AF 1/2" Ext. Polarized Free-field Microphone, High Sensitivity (or other 1/2" IEC 61094-4 compliant free field microphone)**
  http://www.gras.dk/products/measurement-microphone-cartridge/externally-polarized-cartridges-200-v/product/163-40af
- **HEAD acoustics HAE-BGN background noise test system or 3Pass LAB system.**
  http://www.head-acoustics.de/eng/telecom_hae_bgn.htm
  http://www.head-acoustics.de/eng/telecom_3PASS.htm
- **HEAD acoustics HMS II.3 with type 3.3 artificial ear (alternate 1)**
  http://www.head-acoustics.de/eng/telecom_hms_II_3.htm
- **Bruel and Kjaer HATS - model 4128C with type 3.3 artificial ear (alternate 2)**
  https://www.bksv.com/en/products/transducers/ear-simulators/head-and-torso/hats-type-4128c
  - **Using the modified Type 3.3 ear pinna that has a tapered ear canal is accepted for in ear headphones.**
- **Additional amplifier for the second mouth simulator**
  For example:
  http://www.thomann.de/gb/the_tamp_pm40c_endstufenmodul.htm?ref=search_rslt_pm40c_162094_0
- **Soundcard in Reference PC – RME HDSP9632 or HDSPe AIO or RME Fireface UCX via USB**
  http://www.rme-audio.de/en_products_hdspe_aio.php
  http://www.rme-audio.de/en/products/fireface_ucx.php
- **Polycom CX100 USB speakerphone for HLK tests.**
  https://support.polycom.com/content/support/North_America/USA/en/support/voice/cx/communicator_cx100.html

  Local LAN router with DHCP server to create a local subnet between REF and DUT clients. WiFi access point for wireless devices
- **Testing of DUT devices for Microsoft Teams Rooms will require a Microsoft Teams Rooms compute unit (PC).** Suitable solutions can be found at link
  https://products.office.com/en-us/microsoft-teams/across-devices/devices/category?devicetype=20

5.1.3 Reference Microsoft Teams client setup details

- Soundcard in Reference PC – RME HDSP9632 or HDSPe AIO or RME Fireface UCX via USB
  http://www.rme-audio.de/en_products_hdspe_aio.php
  http://www.rme-audio.de/en/products/fireface_ucx.php

  Refer to Audio Standard Operating Procedures for a correct driver/FW version for RME soundcard!
- Use a PC running the latest Windows 10 operating system.
  - Microsoft Teams client – the latest available Windows client from MS Teams website.
  - Use the REF/DUT editor supplied by HEAD acoustics or refer to Audio Standard Operating Procedures available from Microsoft Teams logo program partner manager for guidelines how to enable a 'Reference mode' for the REF client.
- Recommended webcams to be used for Reference client PC are
  http://www.microsoft.com/hardware/en-gb/p/lifecam-studio/Q2F-00015
  http://www.microsoft.com/hardware/en-gb/p/lifecam-cinema/H5D-00014
  https://www.logitech.com/en-us/product/brio

Calibration procedure for reference Microsoft Teams client <-> measurement system

1) Play back a full scale 1019Hz sine signal to reference Skype for Business 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.

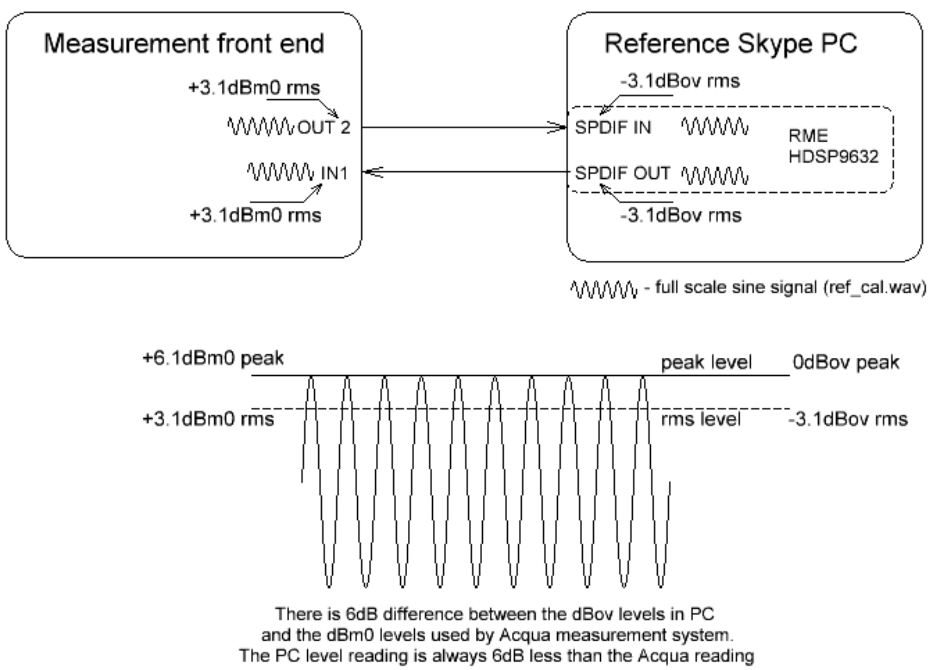


Figure 15: dBm0 conversion to dBov

5.1.4 DUT Skype client setup details

Following table outlines the applicable test modes for each category of devices.

Devices that are seeking certification as an audio offload approved device, the middle section must be run with the client in both offload and non-offload configuration, additionally doubletalk tests (4.3.9 and 4.3.10 must also be run in both modes).

Device type	DUT client configuration / test mode for all tests except the testcases marked on right	Exception Testcases / test mode
USB connected Headsets	Default client settings used	4.1.3/ 4.1.12 / 4.3.3 / 4.3.4 / 4.3.9 / 4.3.10 / 4.4.12 Teams digitalAGC – disabled
Open Office Headsets	Default client settings used	4.1.3/ 4.1.12 / 4.3.3 / 4.3.4 / 4.3.9 / 4.3.10 / 4.4.12 Teams digitalAGC – disabled Teams AEC / NS – disabled
Wireless Headsets with USB dongle	Default client settings used	4.1.3/ 4.1.12 / 4.3.3 / 4.3.4 / 4.3.9 / 4.3.10 Teams digitalAGC – disabled Teams AEC / NS – disabled
USB connected Desk phones	Teams AEC / NS – offloaded	4.1.3/ 4.1.12 / 4.3.3 / 4.3.4 / 4.3.9 / 4.3.10 Teams digitalAGC – disabled Teams AEC / NS – disabled
Desk phones with embedded client	Based on custom build of MS Teams on DUT	4.1.3/ 4.1.12 /4.3.3 / 4.3.4 / 4.3.9 / 4.3.10 Teams AEC / NS / dAGC – disabled (if DUT client allows Teams ADSP configurations)
DUT with embedded audio device without AEC, NS etc. <u>This is setting for the "RAW mode" testing</u>	**Default client settings used**	4.1.3 / 4.1.4 / 4.1.5 / 4.1.6 / 4.1.7 / 4.1.12 / 4.3.5 / 4.3.6 / 4.3.7 RAW capture mode[23] - enabled Teams AEC / NS / dAGC – disabled
DUT with embedded audio device with 3rd party AEC, NS etc.	MS Teams Desktop client used Default/Communications mode[24] - enabled Teams AEC / NS – disabled	4.1.3/ 4.1.12 /4.3.3 / 4.3.4 / 4.3.10 MS Teams Desktop client used Default/Communications mode - enabled Teams AEC / NS – disabled Teams digitalAGC – disabled
Conferencing devices USB connected	Room system client used MTR AEC – disabled MTR NS - disabled	4.1.3/ 4.1.12 /4.3.3 / 4.3.4 / 4.3.10 Room system client used MTR digitalAGC – disabled MTR AEC – disabled MTR NS - disabled
Conferencing devices with Microsoft Teams Rooms capability built in	Room system client used Default/Communications capture mode - enabled MTR AEC – disabled MTR NS - disabled	4.1.3/ 4.1.12 / 4.3.3 / 4.3.4 / 4.3.9 Room system client used Default/Communications capture mode - enabled MTR digitalAGC – disabled MTR AEC – disabled MTR NS - disabled

Table 69: DUT client test modes for Send, Receive and Echo path testing

5.1.5 DUT device setup details - camera

If the DUT solution comes with a camera bundled, then that camera is used during the testing. If there is no camera bundled, then one of the below cameras should be used with the DUT for the calls during testing.

23 Please refer to section 2.6.1 for details on RAW capture mode
24 Please refer to section 2.6.1 for details on Communication capture mode

http://www.microsoft.com/hardware/en-gb/p/lifecam-studio/Q2F-00015

http://www.microsoft.com/hardware/en-gb/p/lifecam-cinema/H5D-00014

https://www.logitech.com/en-us/product/brio

5.1.6 Head and Torso simulator and calibrations

Certification testing uses Head and Torso Simulator (HATS) compliant with ITU-T P.58 and artificial ear compliant with ITU-T P.57 Type 3.3.

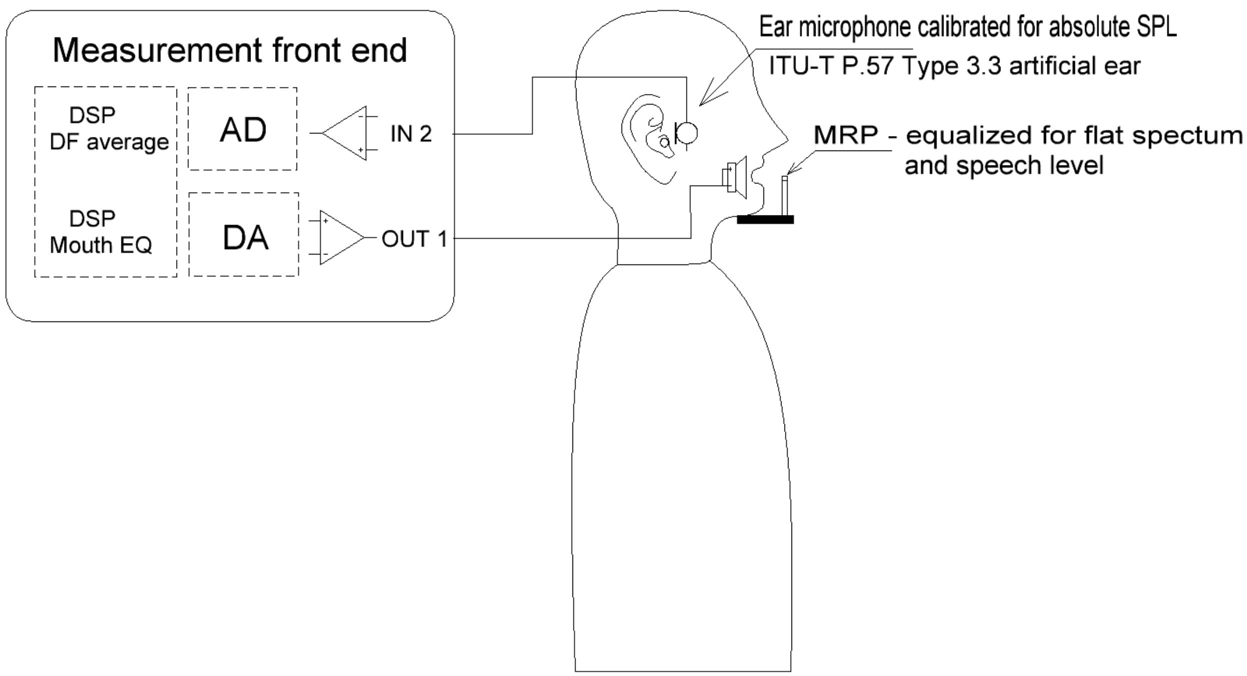


Figure 16: Head and Torso Simulator

The frequency spectrum of mouth simulator is calibrated and frequency compensated with ¼'' pressure field microphone 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 20000Hz.

5.1.7 Mouth simulator and free field microphone

The mouth simulator used for test should be compliant to ITU-T P.51. The recommended model is a G.R.A.S 44AB as this model has a good extension at higher frequencies (above 10kHz) and uses a rounded front interfering less with receive path measurement linearity.

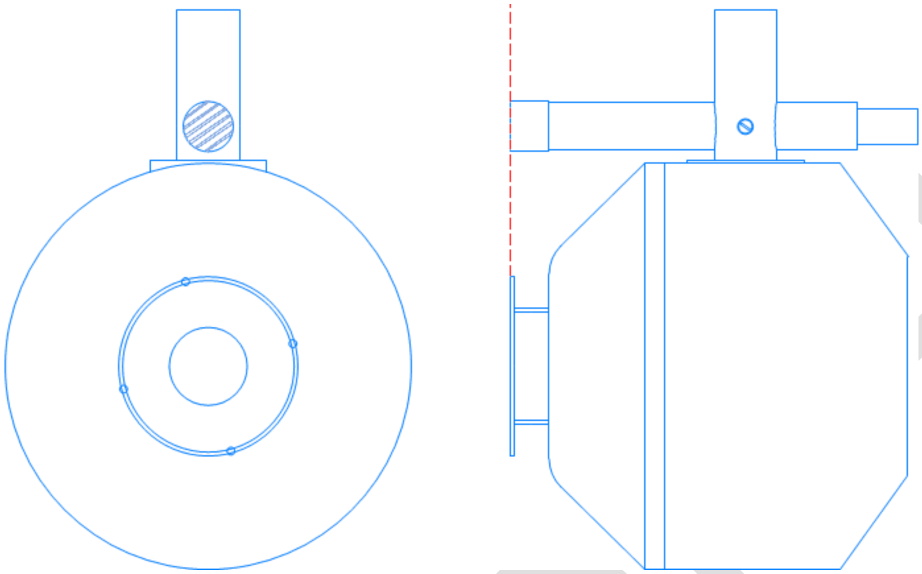


Figure 17: Free field microphone positioning on top of artificial mouth simulator

Note! A special holder for ½’’ microphone is available as an accessory from HEAD acoustics.

5.1.8 Frequency response measurement for Speakerphone UI categories

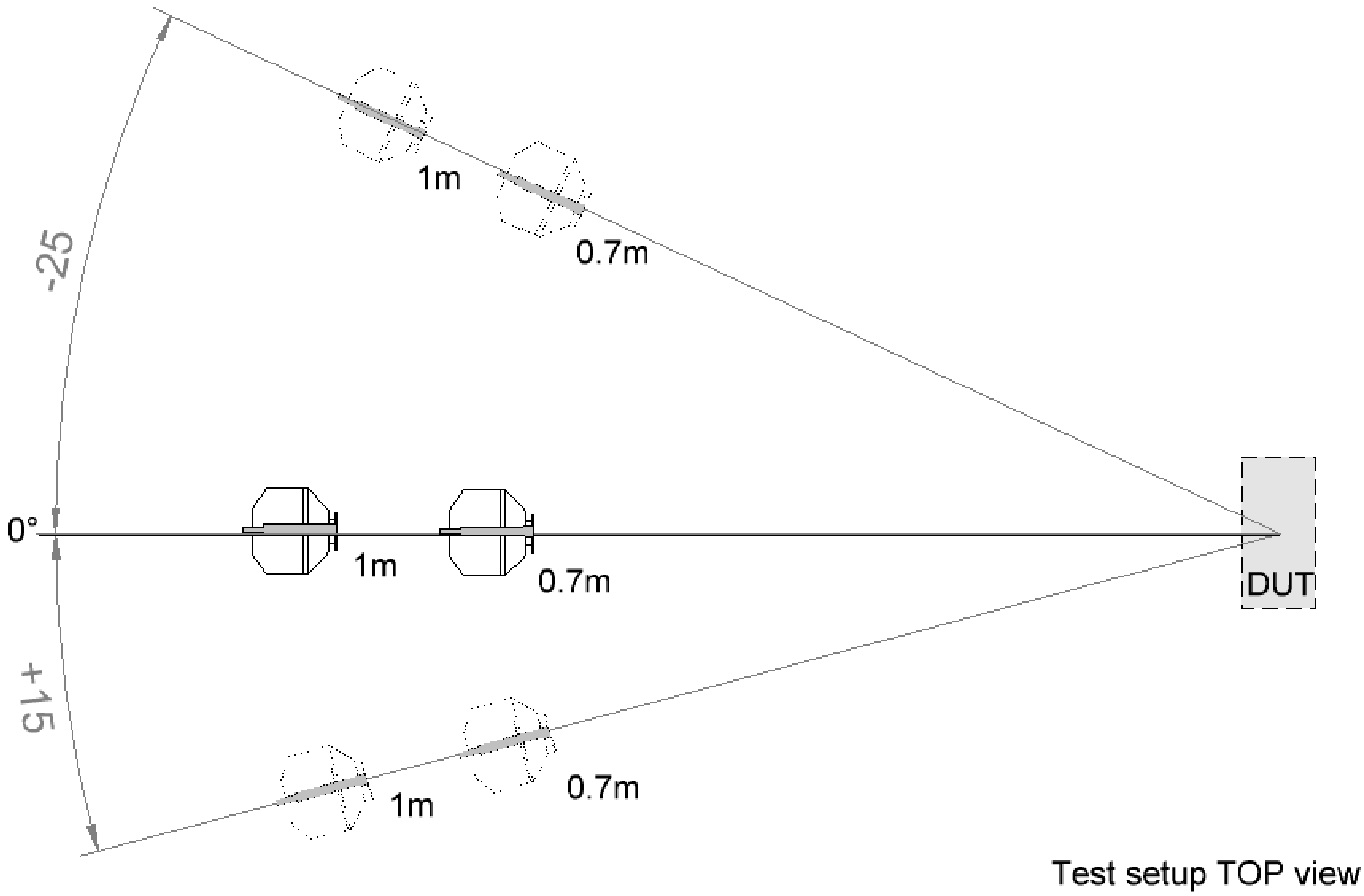


Figure 18: Measurement positions for anechoic frequency response measurements.

Note! It is equally acceptable to rotate the DUT device instead of moving the mouth simulator left and right. HEAD acoustics also offers a Remote-operated Turntable HRT I that can be used to automate the collection of the measurement data. **If the HRT I type turntable is used it should be placed under the test table so it would not raise the device off from table surface!**

5.1.9 Test signals used

All samples are played to reference Microsoft Teams input with the same level as in the source file.

Naming in specification or test sequence	File name
Short delay test signal	short_delay.wav
IEEE 269-2010 uncompressed male speech signal	IEEE_Male_dual_mono.wav
IEEE 269-2010 compressed male speech signal	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
IEEE Male speech + sines	IEEE_Male_speech+sines.wav
Speech activation level	ETSI_speech_activation.wav
TCL test signal	FB_male_female_single-talk_seq_compressed.wav
ECC near end – far end single talk	ECC_ST.wav
ECC near end – far end double talk	ECC_DT.wav
EQUEST near end - far end test signal	EQUEST_ST_echo.wav
Female conditioning signal (short)	FB_female_conditioning_seq_long.wav
Female conditioning signal (long)	FB_female_conditioning_seq_short.wav
Male-Female single talk test signal	FB_male_female_single-talk_seq.wav
Male-Female double talk test signal	FB_male_female_double-talk_seq.wav
Short delay test signal (for HAE-BGN)	short_delay.dat
ECC female-male single talk signal	ECC_F_M_ST.wav
ECC male-female single talk signal	ECC_M_F_ST.wav
ECC female-male double talk signal	ECC_F_M_DT.wav
ECC male-female double talk signal	ECC_M_F_DT.wav
ECC far end single talk signal (for HAE-BGN)	ECC_FE_ST.dat
ECC far end double talk signal (for HAE-BGN)	ECC_FE_DT.dat
EQUEST near end male-female test signal for conference devices	Equest_M_F_test.wav
EQUEST near end female-male test signal for conference devices	Equest_F_M_test.wav
EQUEST far end test signal for conference devices (for HAE-BGN)	Equest_FE_input.dat
Near end + projector noise	RR_NE_projector.wav
Near end + hoth noise	RR_NE_hoth.wav
Reference Microsoft Teams output calibration file (Full scale peak to peak Sine signal of 1019Hz)	Ref_cal.wav
Digital silence	Digital_silence_10s.wav
Stereo playback test signal	Smooth_as_SILK_stereo_test.wav
POLQA test signal set	Skype_POLQA.zip

Table 70: Test vectors used for testing the conferencing solutions

5.2 Anechoic test environment

5.2.1 Anechoic room or semi anechoic room

The anechoic chamber or semi-anechoic chamber used should fulfill the acoustical requirements for high quality wideband VoIP and telecom product testing. Please refer to ITU-T P.341 for the recommended parameters for the test room.

Figure 19: Microsoft Tallinn anechoic room

5.2.1.1 *Background noise level in test environment.*

Headset and Handset category		**Speakerphone category**	
Octave Band Center Frequency (Hz)	Octave Band Level (dBSPL)	Octave Band Center Frequency (Hz)	Octave Band Level (dBSPL)
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 29dBSPL		Equals average A-weighted noise level of approximately 20dBSPL	

Table 71: Background noise in audio test setups

5.2.2 Objective test measurement setup

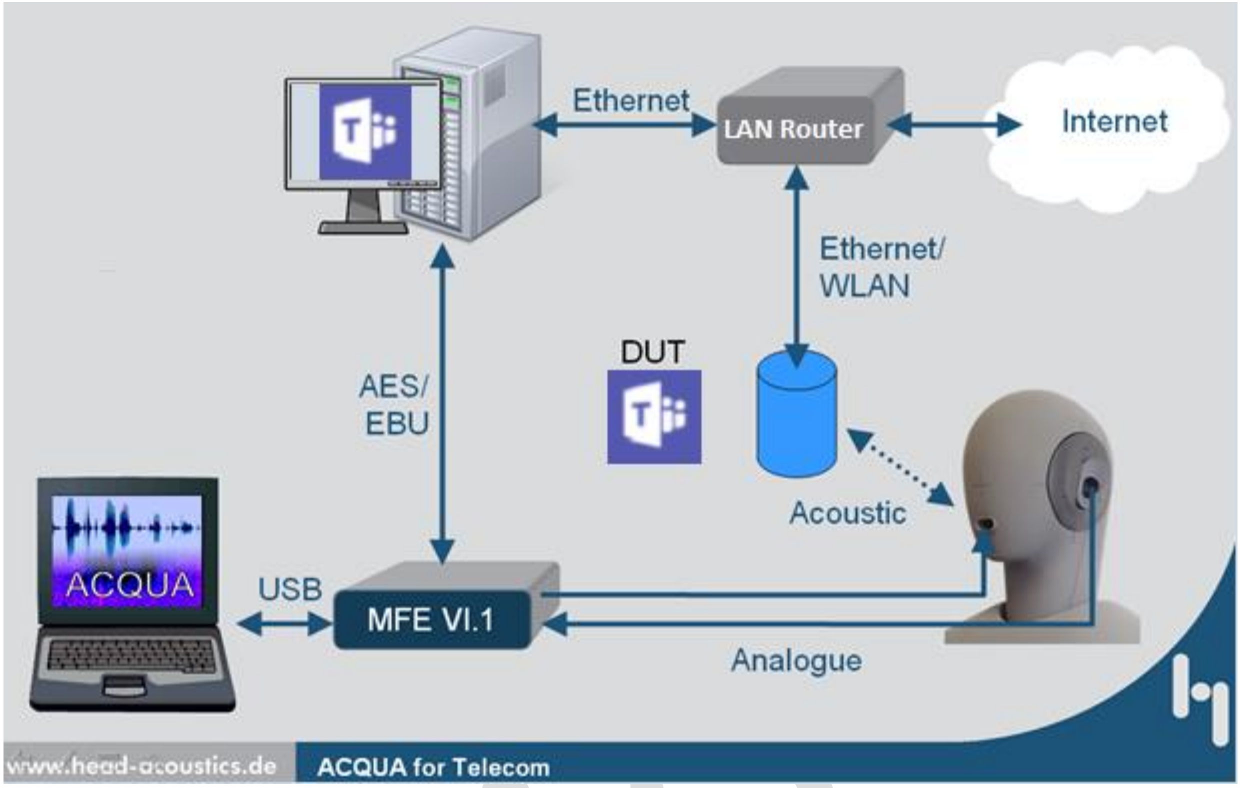


Figure 20: Objective measurement setup

- Microsoft Teams call is created between two clients in lossless local network condition.
- Reference Microsoft Teams client runs on PC with Windows operating system.
- A third computer runs ACQUA audio measurement system and MFE measurement front end connected to HATS and the Reference PC.

5.2.3 Signal levels used for artificial mouth in anechoic room

		Normal speech at artificial mouth MRP	Quiet speech at artificial mouth MRP
Headset		89dBSPL	79dBSPL
Handset		89dBSPL	79dBSPL
Personal speakerphone		89dBSPL	79dBSPL
Conferencing devices	Up to 1.5m mic usage	86dBSPL (-3dB) *	76dBSPL*
	Up to 2.3m mic usage	84dBSPL (-5dB) *	74dBSPL*
	Up to 3.5m mic usage	82.5dBSPL (-6.5dB) *	72.5dBSPL*
	Up to 4.5m mic usage	81dBSPL (-8dB) *	71dBSPL*

* Simulated speech level drop at distance while artificial mouth is physically at 1m distance

Table 72: signal level with normal speech requirements

5.2.4 Anechoic room device setup

5.2.4.1 *Headset positioning on HATS*

If the manufacturer/vendor/user documentation provides guidelines how the headset should be used/worn, the audio test 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.

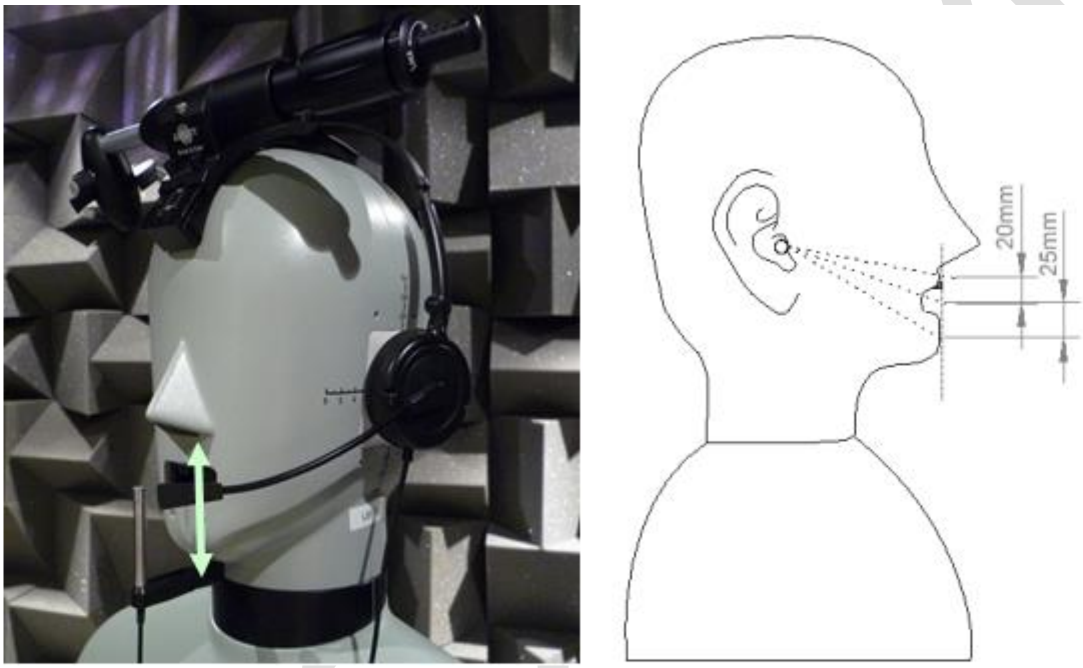


Figure 21: Headset positioning on HATS

For earpiece and microphone frequency response test case the headset is removed and placed again 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.

The microphone is moved in 3 positions. Reference position where boom points to mouth and then 20mm up and 25mm down from that position.

5.2.4.2 *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. the audio test engineer 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 ±2N** between phone and the right ear.

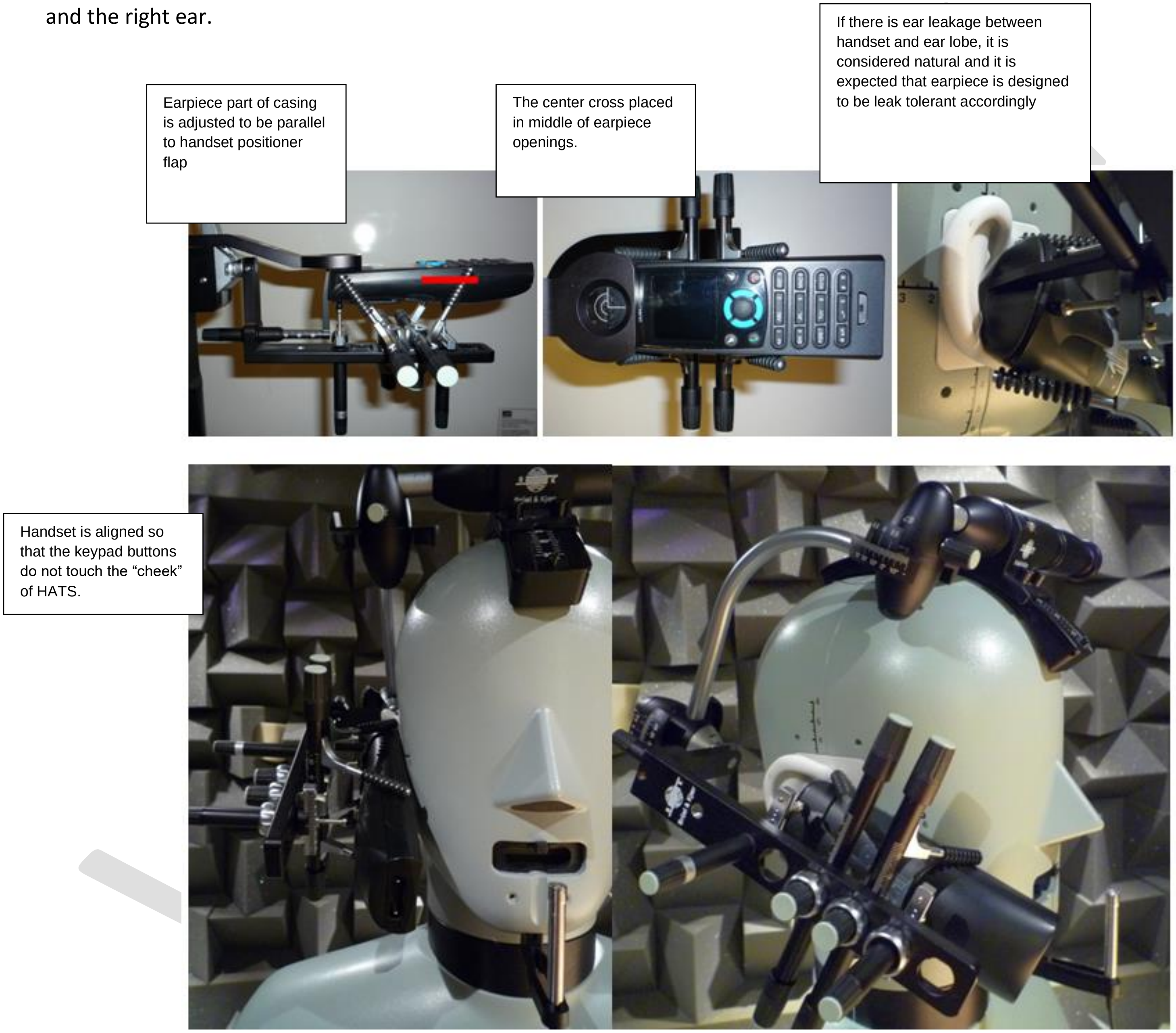


Figure 22: Handset positioning on HATS

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.

5.2.4.3 *Handheld speakerphone test position*

The DUT is placed as drawn below. The physical fixture holding the device should not block the microphone inlet / speaker(s) outlet.

If the manufacturer/vendor of the device advises another measurement scenario, then that is taken into account and measurement positions are agreed.

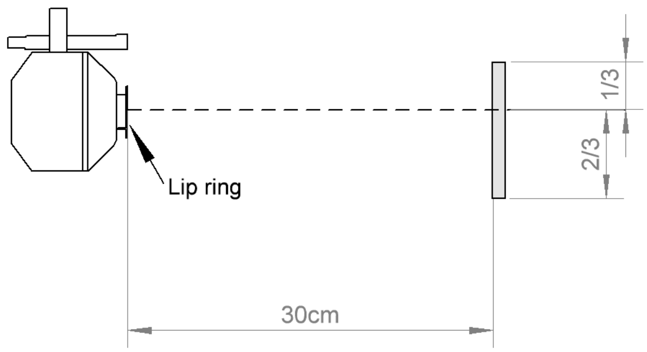


Figure 23: Positioning handheld device

5.2.4.4 *Personal speakerphone recommended test position*

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

If the manufacturer/vendor of the device advises another measurement scenario, then that is taken into account and measurement positions are agreed.

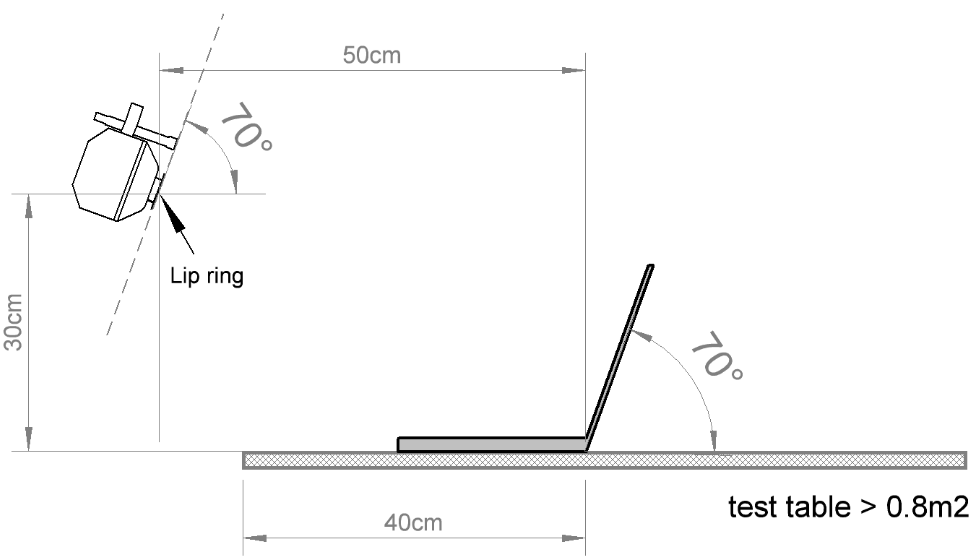


Figure 24: Positioning a laptop using artificial mouth

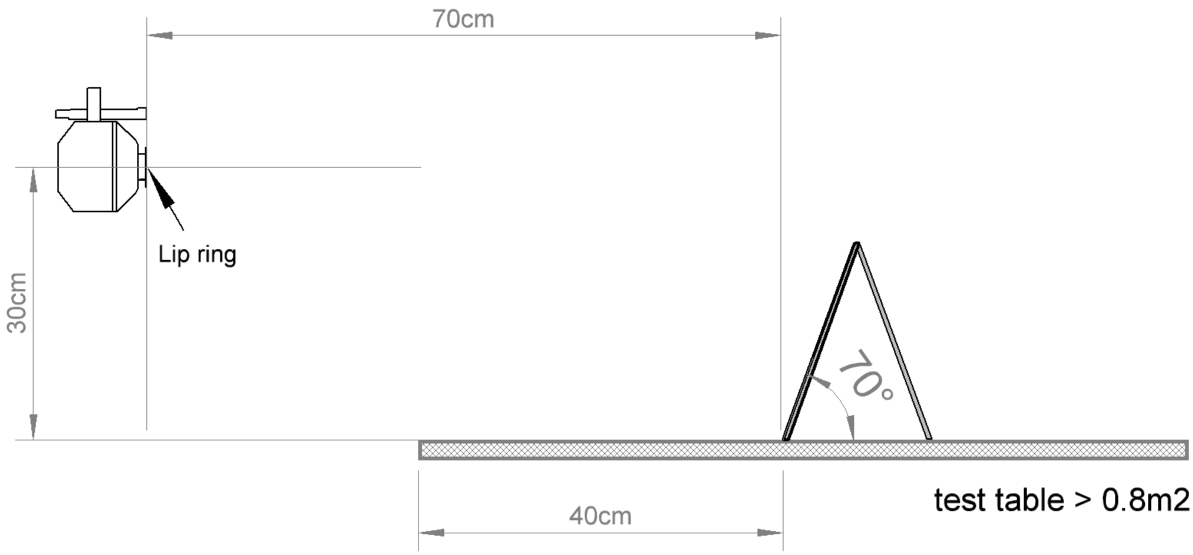


Figure 25: Positioning a laptop in tent mode using artificial mouth

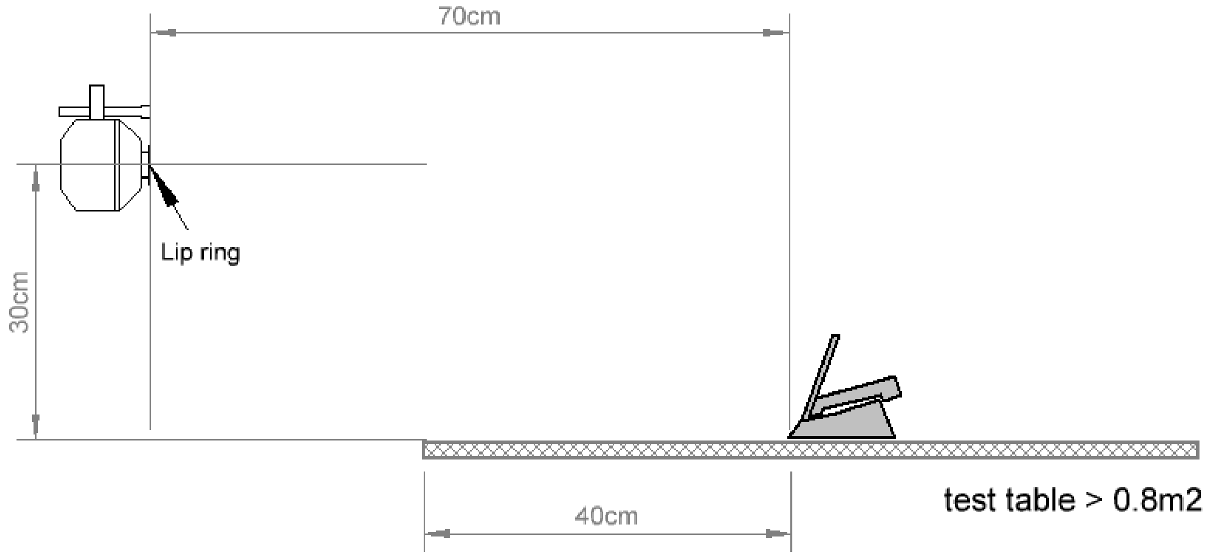


Figure 26: Positioning a desk phone using artificial mouth

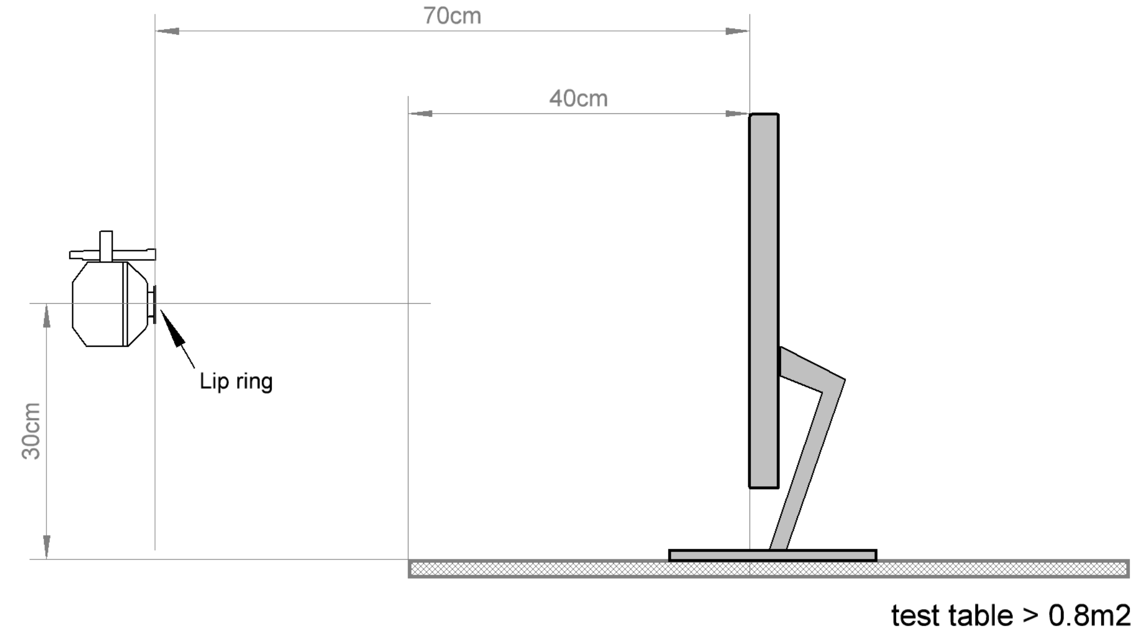


Figure 27: Positioning of an all-in-one (AIO) PC or a multimedia monitor using artificial mouth

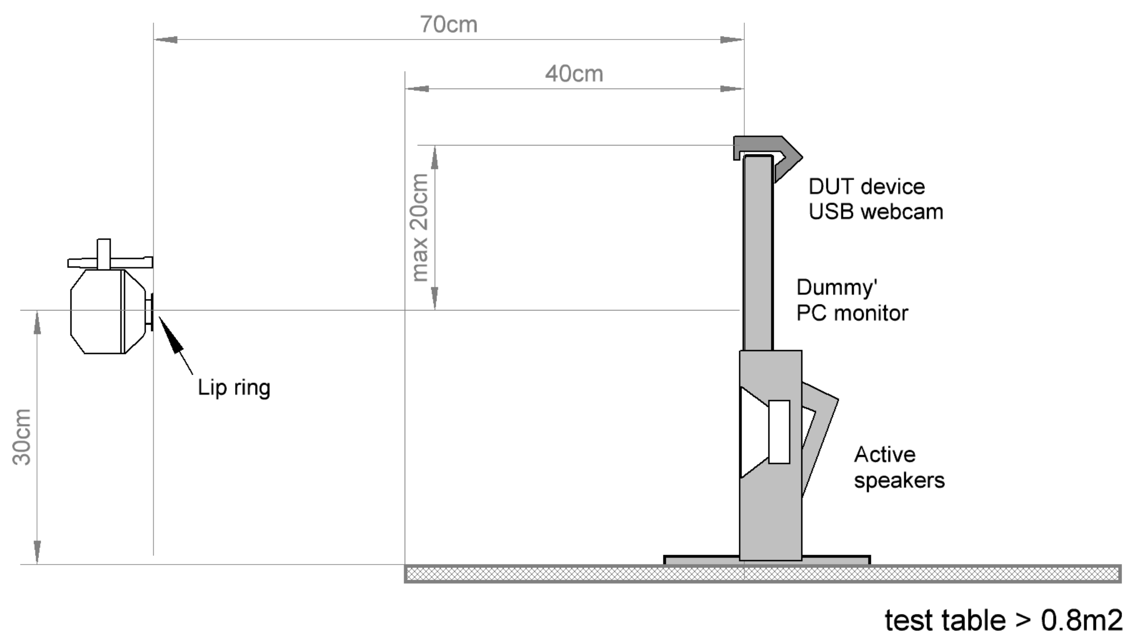


Figure 28: Positioning a standalone webcam using artificial mouth

5.2.4.5 *Table top speakerphone test position*

The recommended test position for center of room speakerphone is shown below. The center of room speakerphone is measured with microphone(s) at 1m distance from Lip Ring.

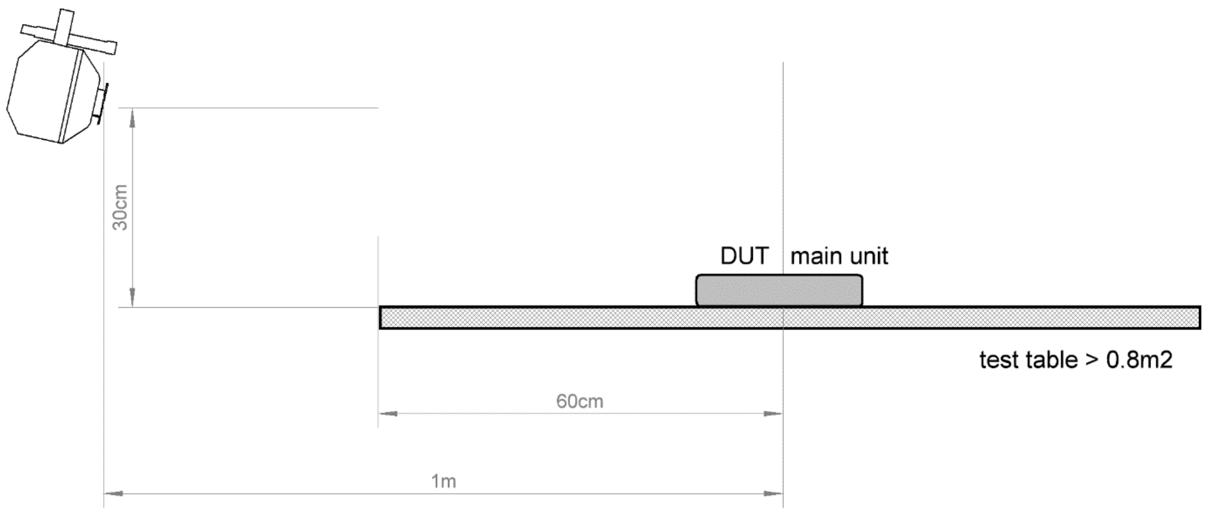


Figure 29: Positioning a center of room speakerphone using artificial mouth[25]

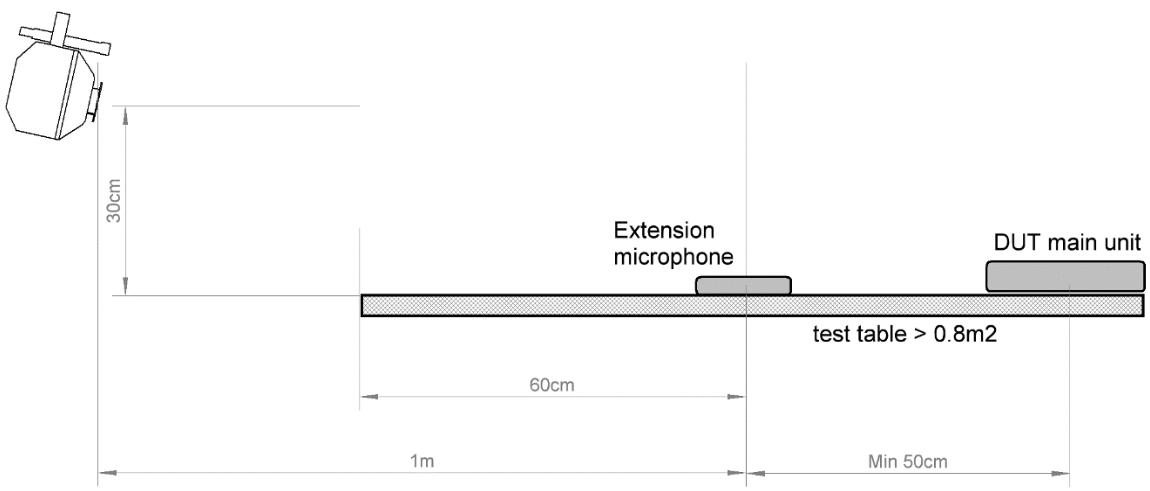


Figure 30: Positioning a center of room speakerphone with extension microphone using artificial mouth

[25] In case the DUT has a tall form factor then the mouth should be pointed towards the top of the DUT device.

5.2.4.6 *Edge of room speakerphone test position in anechoic room.*

The recommended test position for edge of room speakerphone is shown below. If device allows for rotation (landscape/portrait mode) then Frequency response (send 4.1.10 and receive 4.2.7), as well as all E2E tests (3.1) must be tested in both orientations.

If the manufacturer/vendor of the device advises another measurement scenario, then OEM must receive approval from Microsoft for any testing variations prior to submitting for certification.

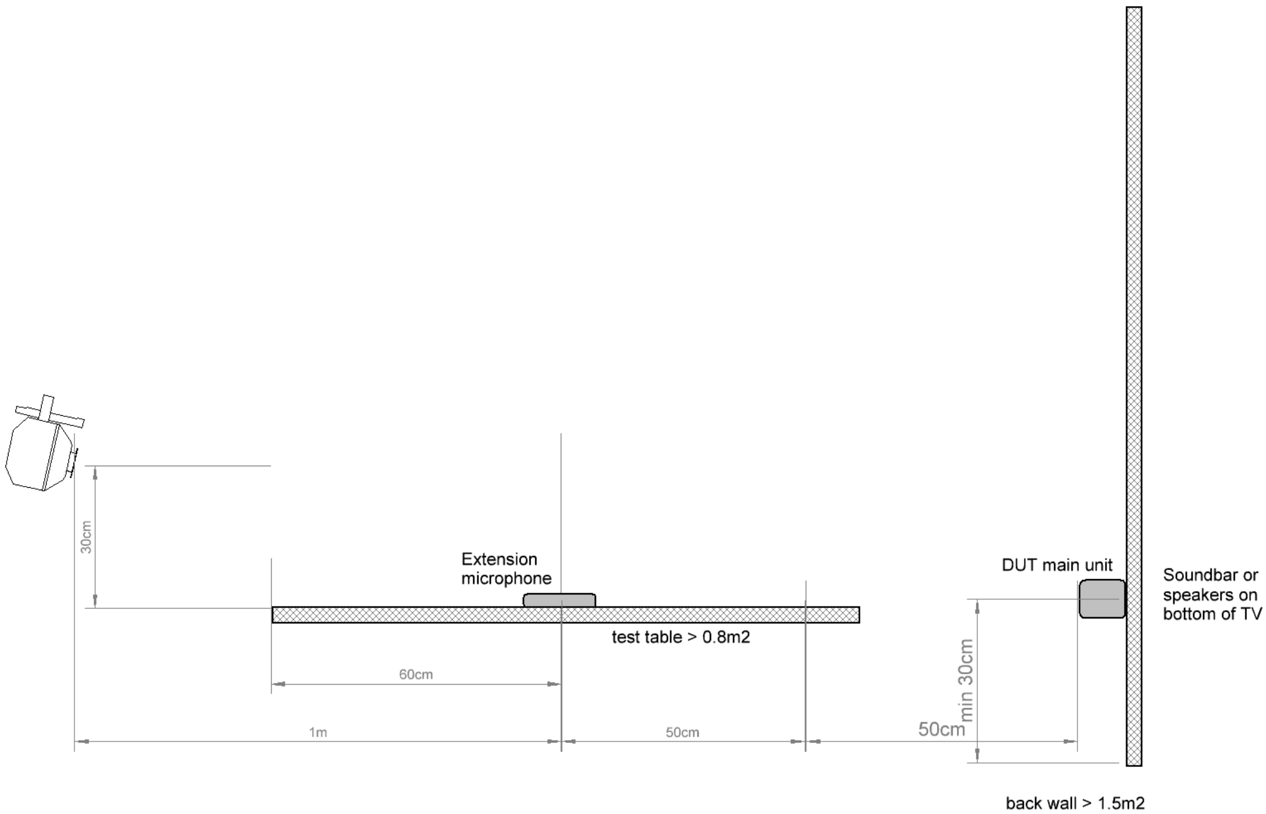


Figure 31: Positioning for microphone on table and soundbar/speaker on wall

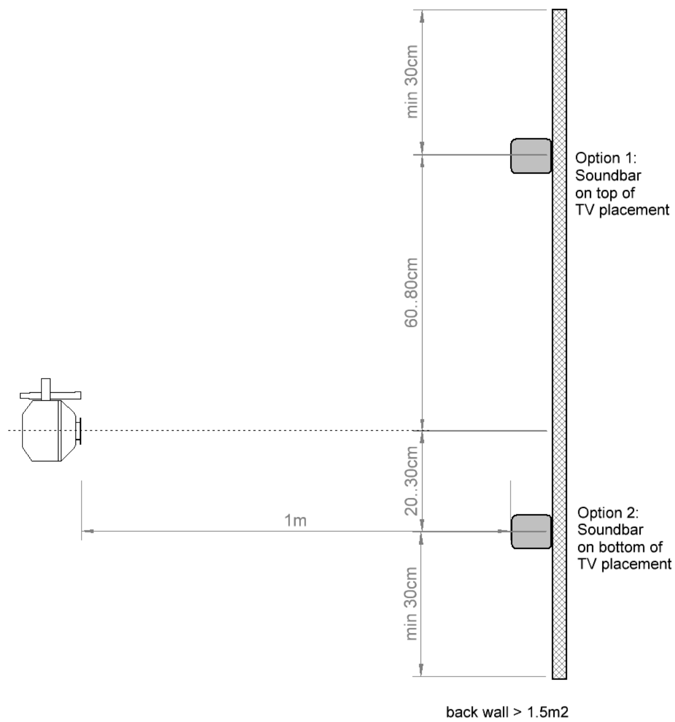


Figure 32: Positioning for soundbar or edge of room speakerphone

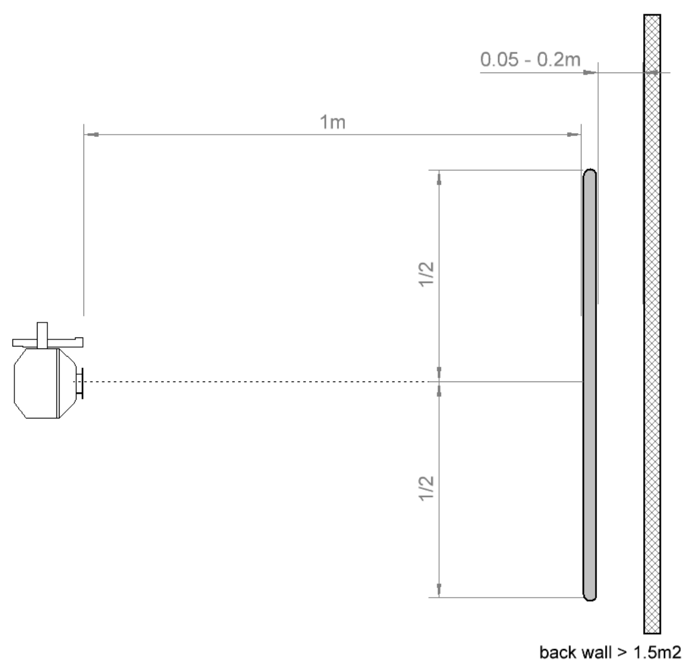


Figure 33: Positioning for big screen type edge of room speakerphone

In case of a free rolling collaboration device the testing should be conducted without the back-wall.

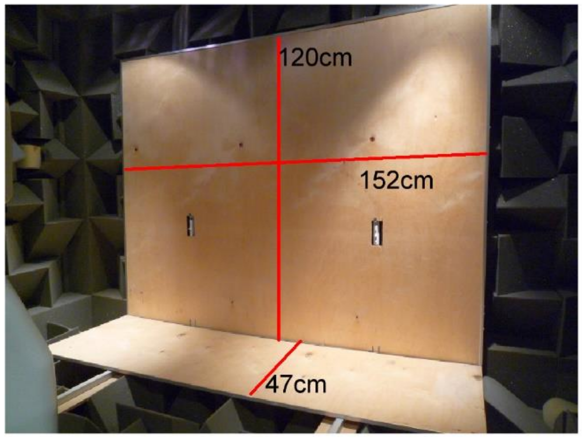


Figure 34: Simulated back-wall

The sample of physical back-wall setup in anechoic room. Material is 8mm thick plywood. Same or thicker MDF board is equally ok to use.

In case the DUT solution has standalone left and right speaker then the speakers should be tested separately with the measurement microphone placed on respective speaker's axis at 100cm distance.

5.3 Reverberant test environment

5.3.1 Reverberant room

- For reverberant room device testing the test room recommendations are following:
  - **Room size** – the room size should be in range between 2.7m X 3.7m and 3.5m X 4.4m. Room height 2.2m to 3.25m.
    - **Treatment of the room –** the reverberation time target for the reverberant test room should be 0.4s < RT60 < 0.7s
      The reverberation time specified is higher compared to ETSI ES202 396-1 recommendation as the goal is to stress the DUT more and make sure the DUT is able to do a good echo cancellation also in more reverberant rooms that are not uncommon in many offices. The aim of the ETSI ES 202 396-1 is to create a suitable environment for reliable background noise simulation scenarios but does not focus on echo canceller performance testing.

Octave Band Center Frequency (Hz)	Lower Limit (RT60, ms)	Upper Limit (RT60, ms)
125	400	950
250	500	850
500	350	700
1000	350	600
2000	350	600
4000	350	600
8000	350	600
Room Average	400	700

Table 73: Reverberant room reverberation time versus octave bands

- The reverb time should be declining toward high frequencies but should not have dips or peaks in some octave bands that deviate more than 0.2 sec compared to the adjacent octave bands on either side.
- **Noise floor –** to avoid room noise influencing the test results the average noise floor in room should be 30 dBSPL(A) or below.

5.3.2 Reverberant room setup for personal solutions

5.3.2.1 *Setup for calibrating the background noise*

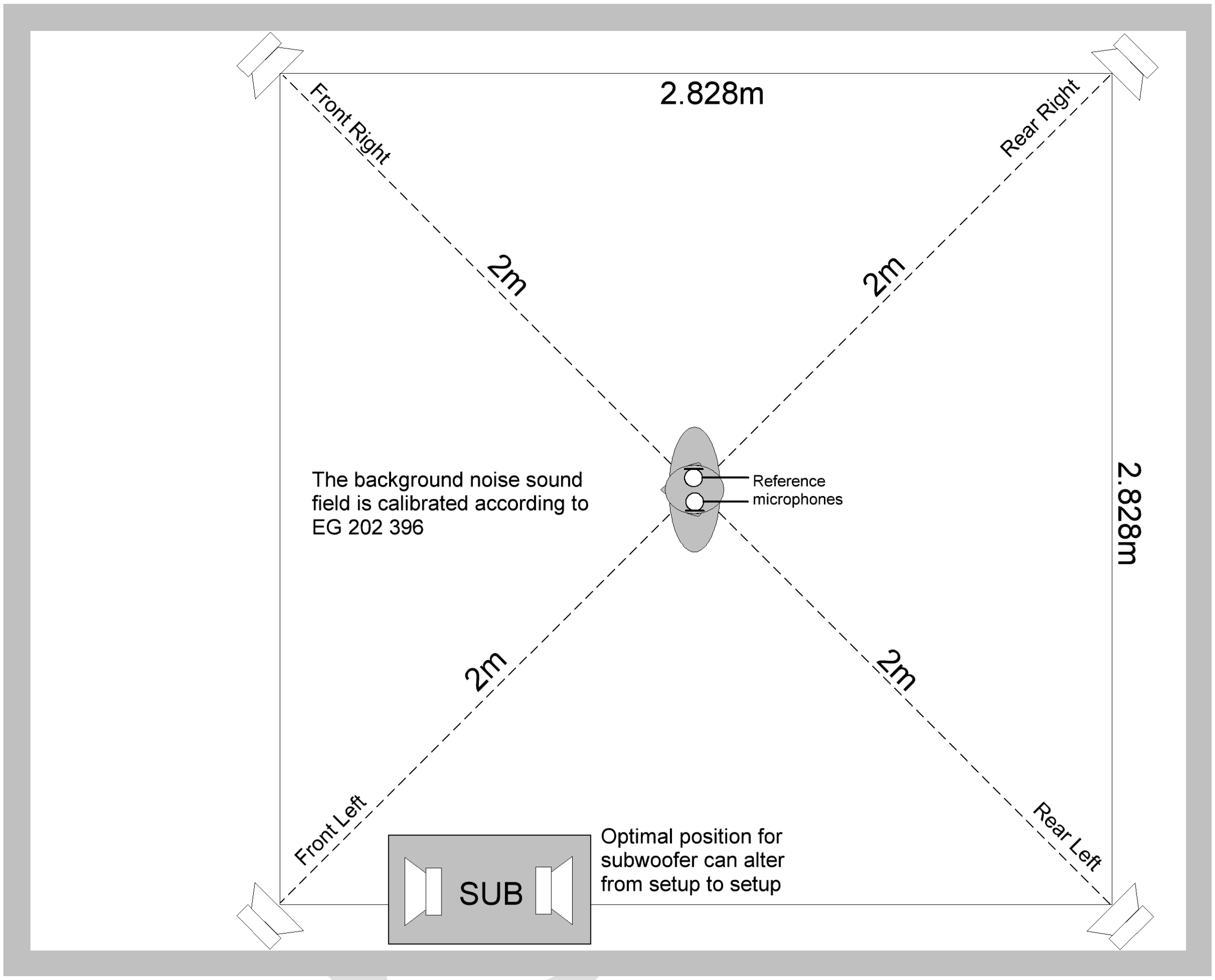


Figure 35: HAE-BGN setup for system calibration (HATS is required)

5.3.2.2 *Background noise test setup for headsets*

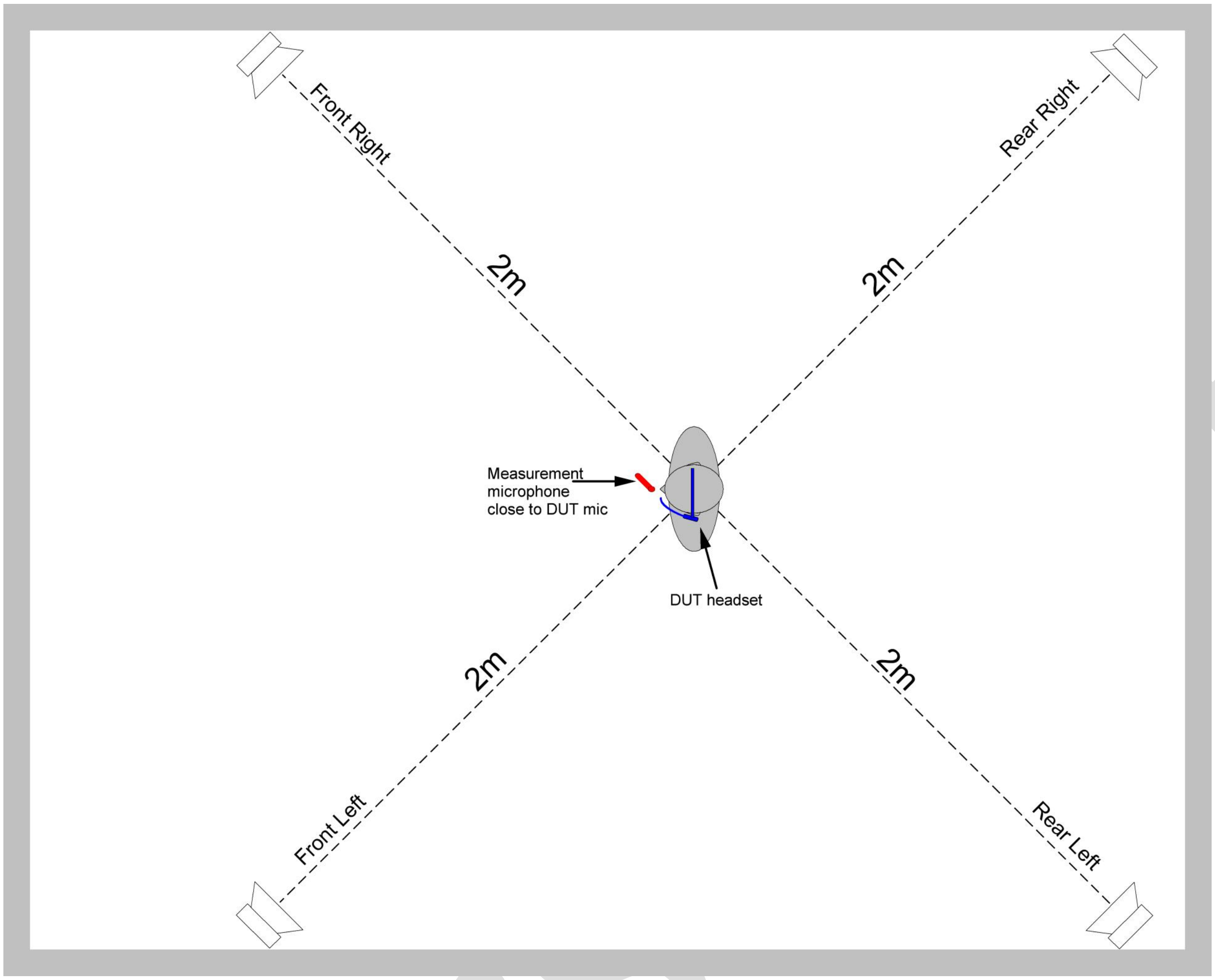


Figure 36: Background noise test setup for headsets

Note: In case of headset the reference microphone is put to next to the DUT microphone. In case of headsets without the boom the Measurement microphone is placed close to the DUT microphone in ear cup. If the DUT has microphones in both earcups, then a manufacturer recommendation for the Measurement microphone placement shall be followed.

5.3.2.3 *Background noise test setup for Open Office headsets*

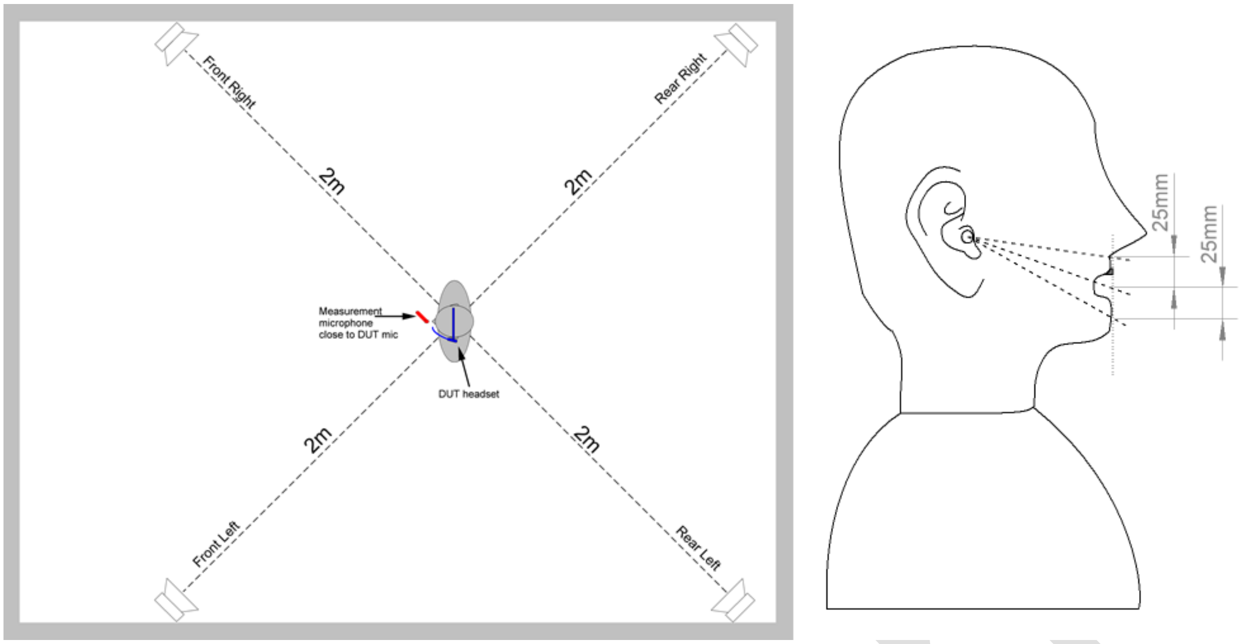


Figure 37: Background noise test setup for Open Office headsets

Note: In addition to testing at a reference boom position (boom pointed to corner of the mouth), the test shall be run also with boom moved up by 20mm and down by 25mm. This is to ensure that the headset microphones allow some individualization of boom position by the user.

If the device has speech feedback on mute, call stop etc. it would be advisable to also detect the boom angle an give the user a warning if headset boom is positioned too high/low for optimal operation.

5.3.2.4 *Speech distractor noise test setup for Open Office headsets*

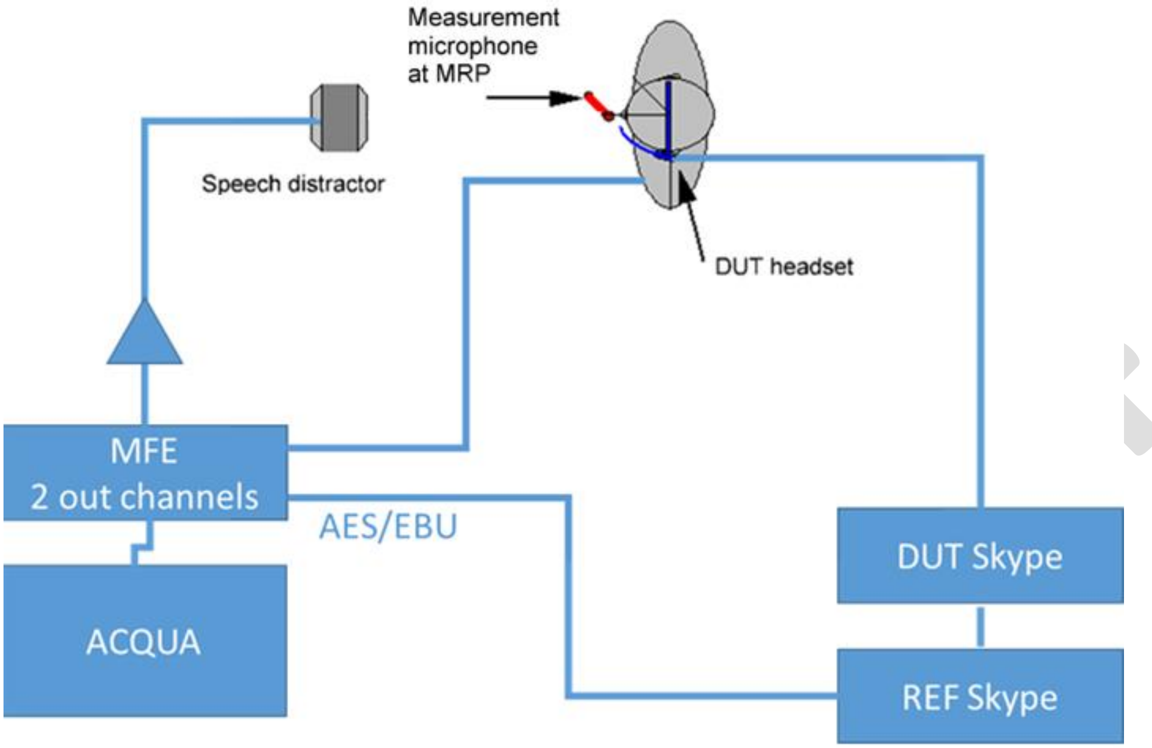


Figure 38: Objective measurement setup for open office headset

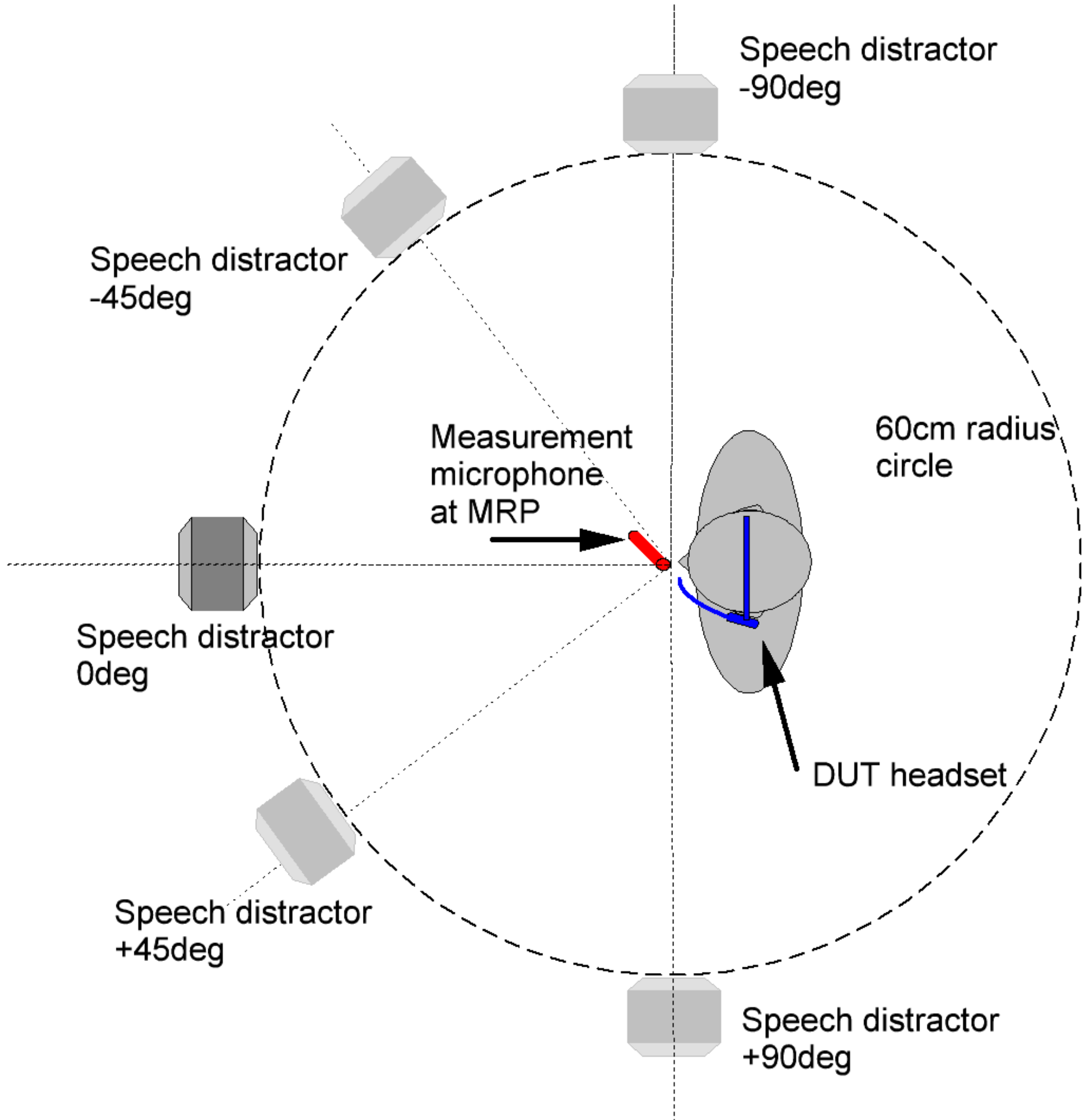


Figure 39: Speech distractor source positions for open office headset test.

5.3.2.5 *Background noise test setup for handsets*

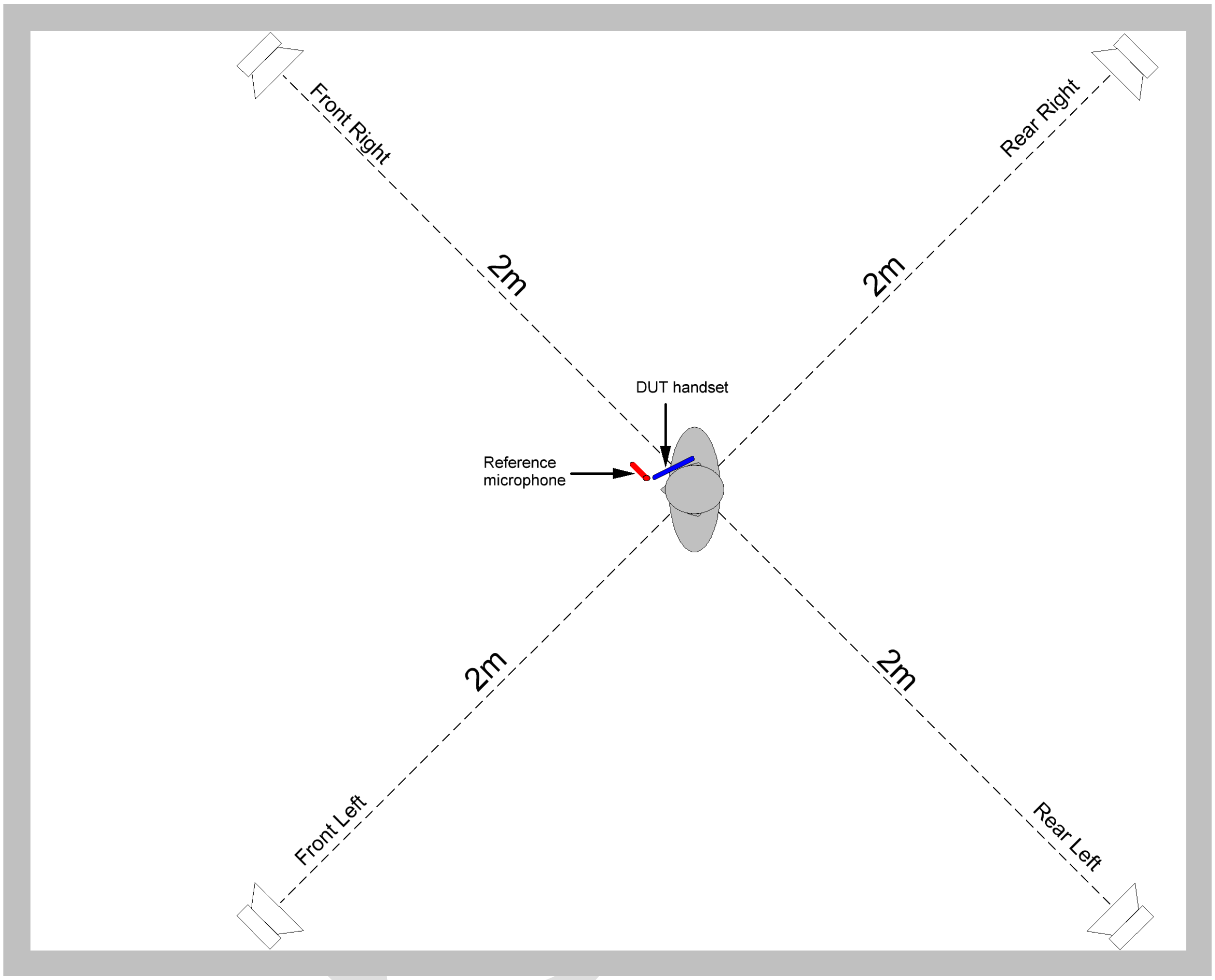


Figure 40: Background noise test setup for handsets

5.3.2.6 *Background noise test setup for handheld speakerphones*

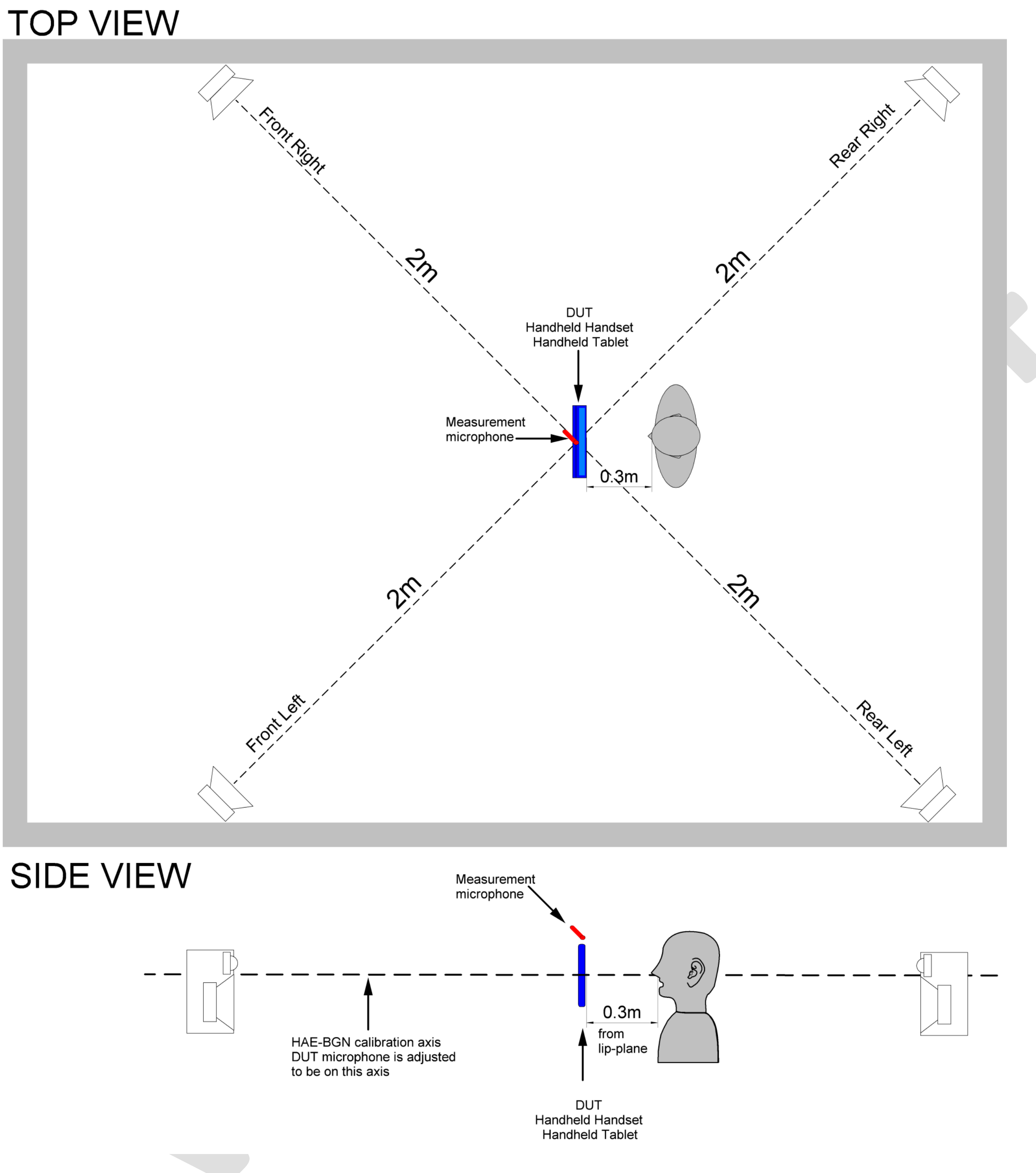


Figure 41: Background noise test setup for handheld speakerphones

5.3.2.7 *Background noise test setup for personal speakerphones*

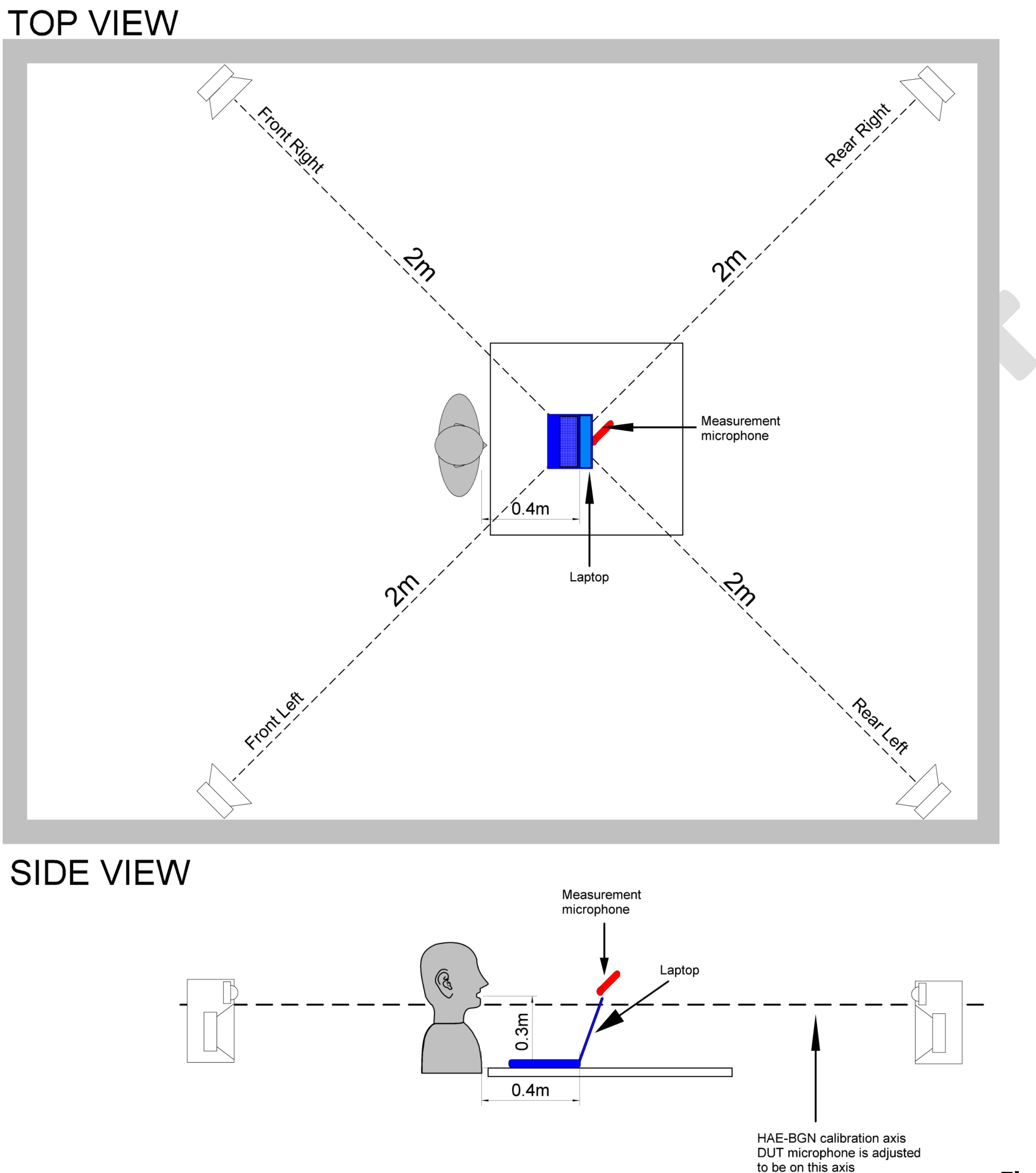


Figure 42: Background noise test setup for personal speakerphones

Note: In case of a narrow and tall devices like Smart Speakers the HAE-BGN axis should be at the same level as microphones in the DUT device. If microphone location is not known, then the top of the device is aligned with calibration axis.

5.3.2.8 *Variable echo path test for personal speakerphones*

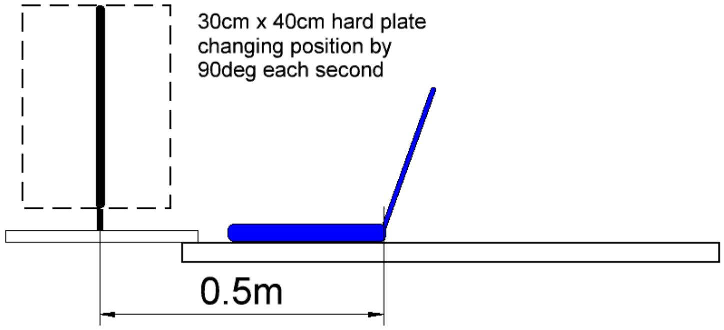


Figure 43: Variable echo path test setup for personal speakerphones

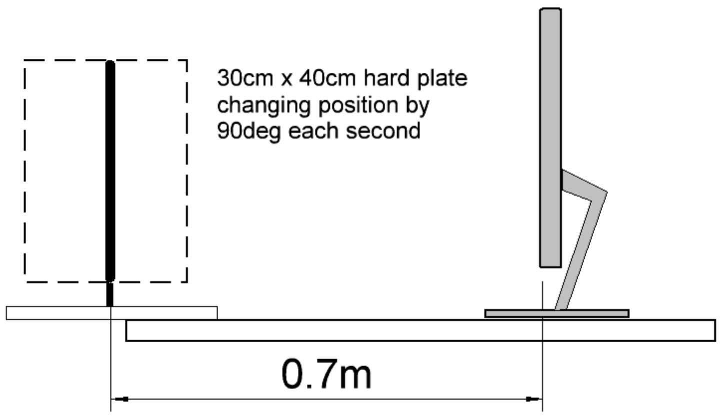


Figure 44: Variable echo path test setup for personal AiO type devices

5.3.3 Reverberant room setup for conferencing solutions

5.3.3.1 *Objective test measurement setups for reverberant room*

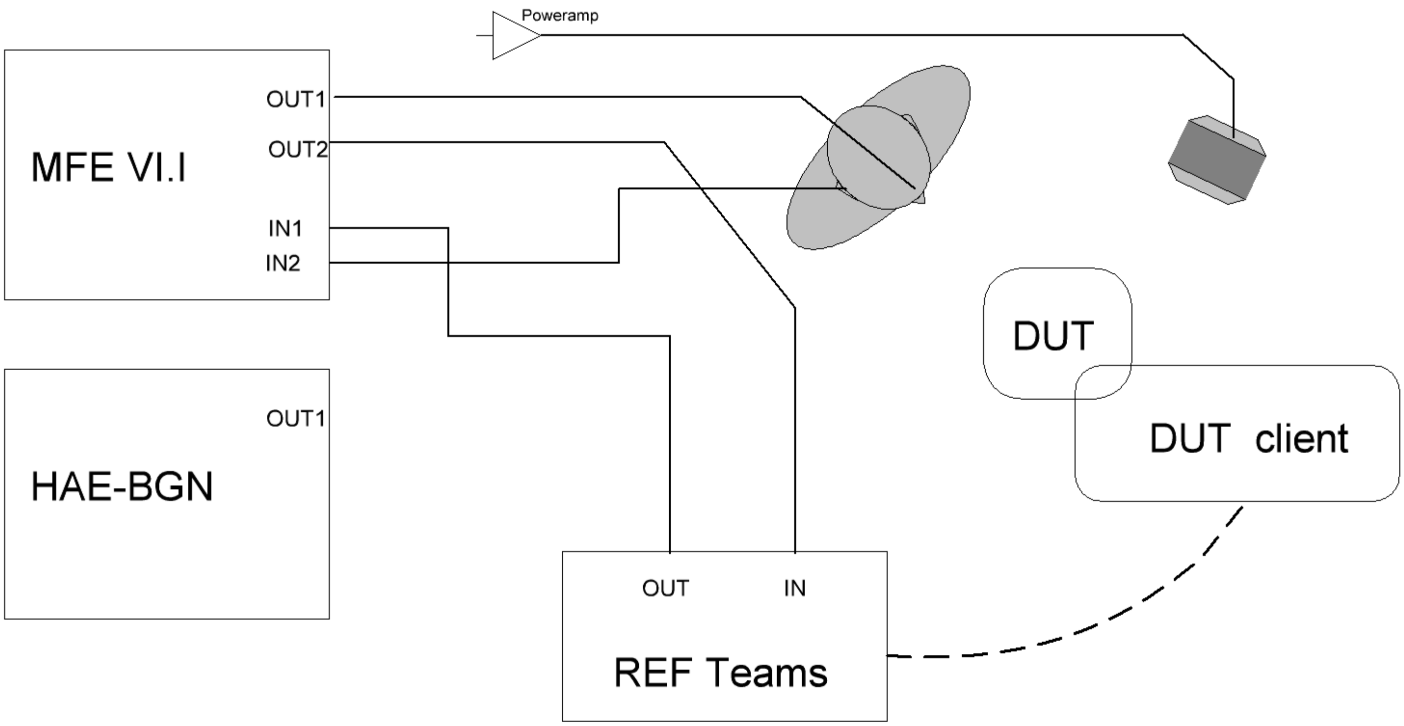


Figure 45: Objective measurement setup for reverb room echo path test

- A call is created between two clients in lossless local network condition.
- DUT uses a Teams Rooms client in offload mode

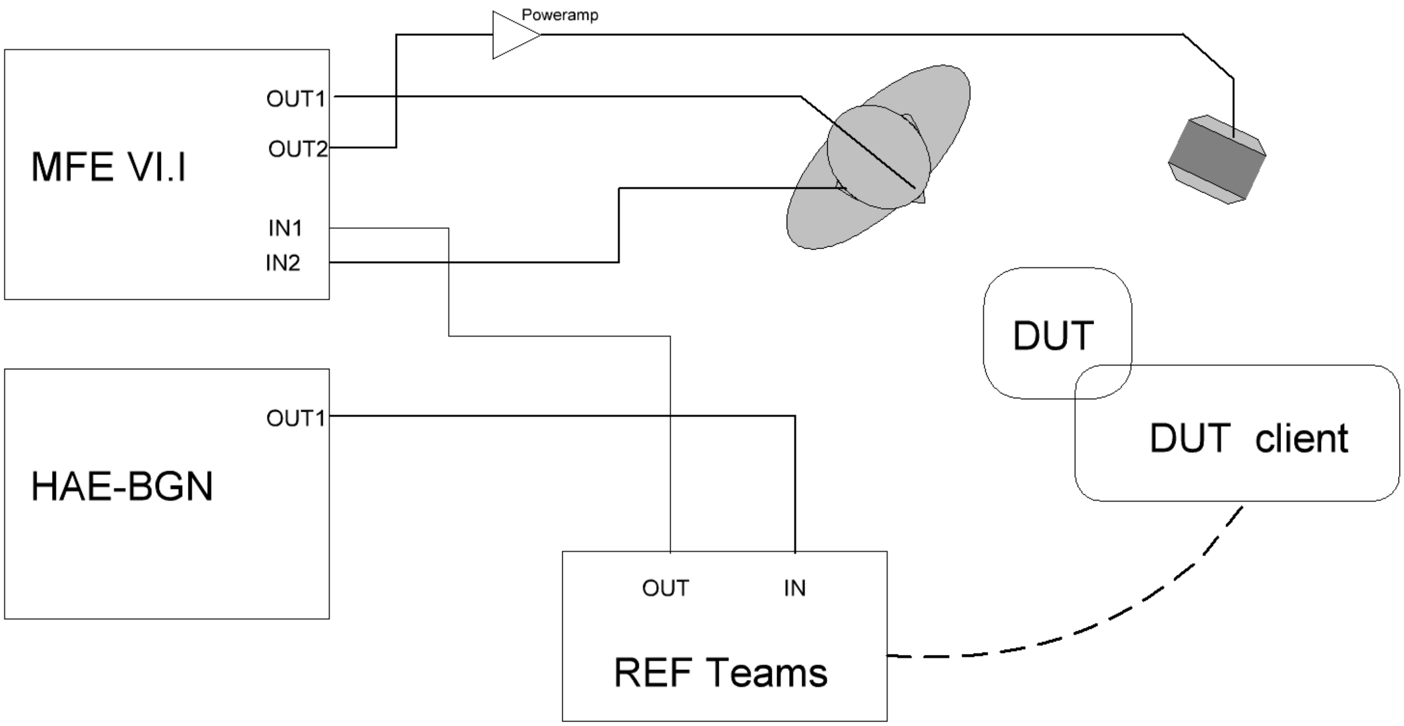


Figure 46: Objective measurement setup for alternating near end talkers

- A call is created between two clients in lossless local network condition.
- DUT uses a Microsoft Teams Rooms (MTR) client in offload mode
- HAE-BGN systems optical output is used to play back the input signal for REF Teams client (this is due to lack of outputs on MFE6.1. The HEAD labCORE front end can do this internally if 4 output channel option is available.
    - Note! The amplifiers / speakers for the background noise are switched off during this testing!!!

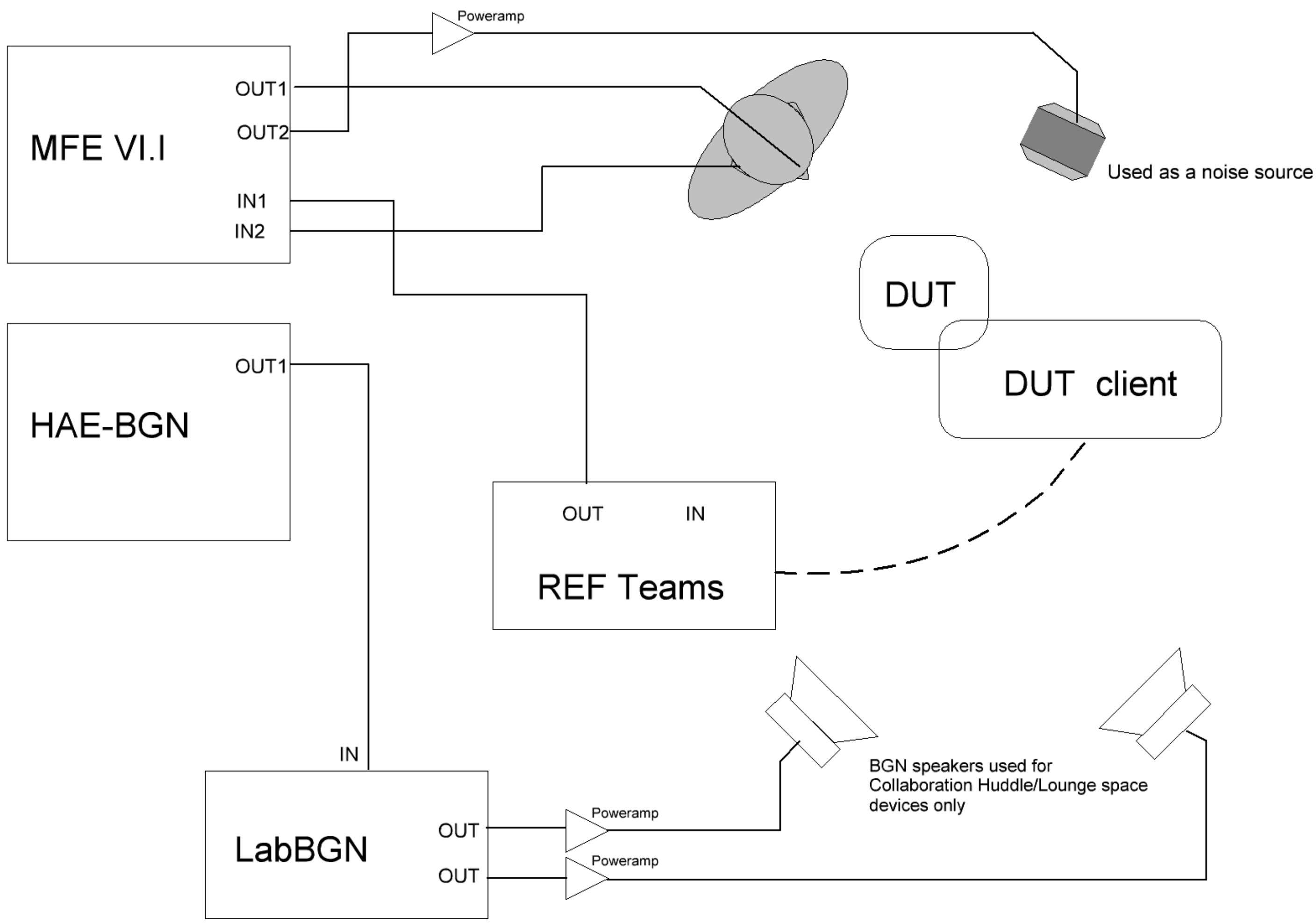


Figure 47: Objective measurement setup for send quality with background noise

5.3.3.2 *Reverberant room device setup*

Center of room and edge of room speakerphones test setup contains a HATS (Talker 1) and an artificial mouth (Talker 2). The following test setup diagrams indicate the positions for the Talker 1 and Talker 2 simulations.

The table used is a rectangular table with each side longer than 800mm (can be a square of 800 x 800mm). In case of dual unit device or extension microphone testing 2 such tables are used.

DUT device is placed in the center of the table in reference to Talker 1 / Position 1. Main device orientation is to be set based on following:

- In case of an audio only device without any extension(s) choose the position where most of buttons are displayed the most convenient way.
- In case of an audio-video device choose the position facing screen (or if no screen then assume that the camera is in the center of a screen).
- In case of an audio only device with extension(s) choose the position on axis from the main using towards the extension.

The optional extension units or extension microphone is placed in middle of the second test table. Orientation should be determined based on user manual or setup guide instructions.

If there are more than one extension units, then those must be connected and switched on but placed under the Talker 1 table on the floor. This will ensure that there is no direct speech pickup from talker 1 or talker 2, but the microphones self noise is still captured in the test results.

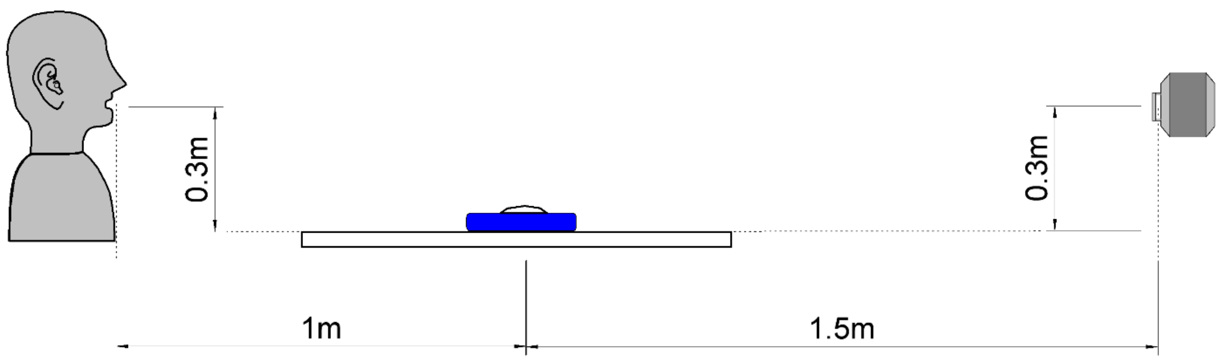


Figure 48: Side view of positioning the DUT and the HATS/artificial mouth

5.3.3.3 *Center of room speakerphones with single unit on table*

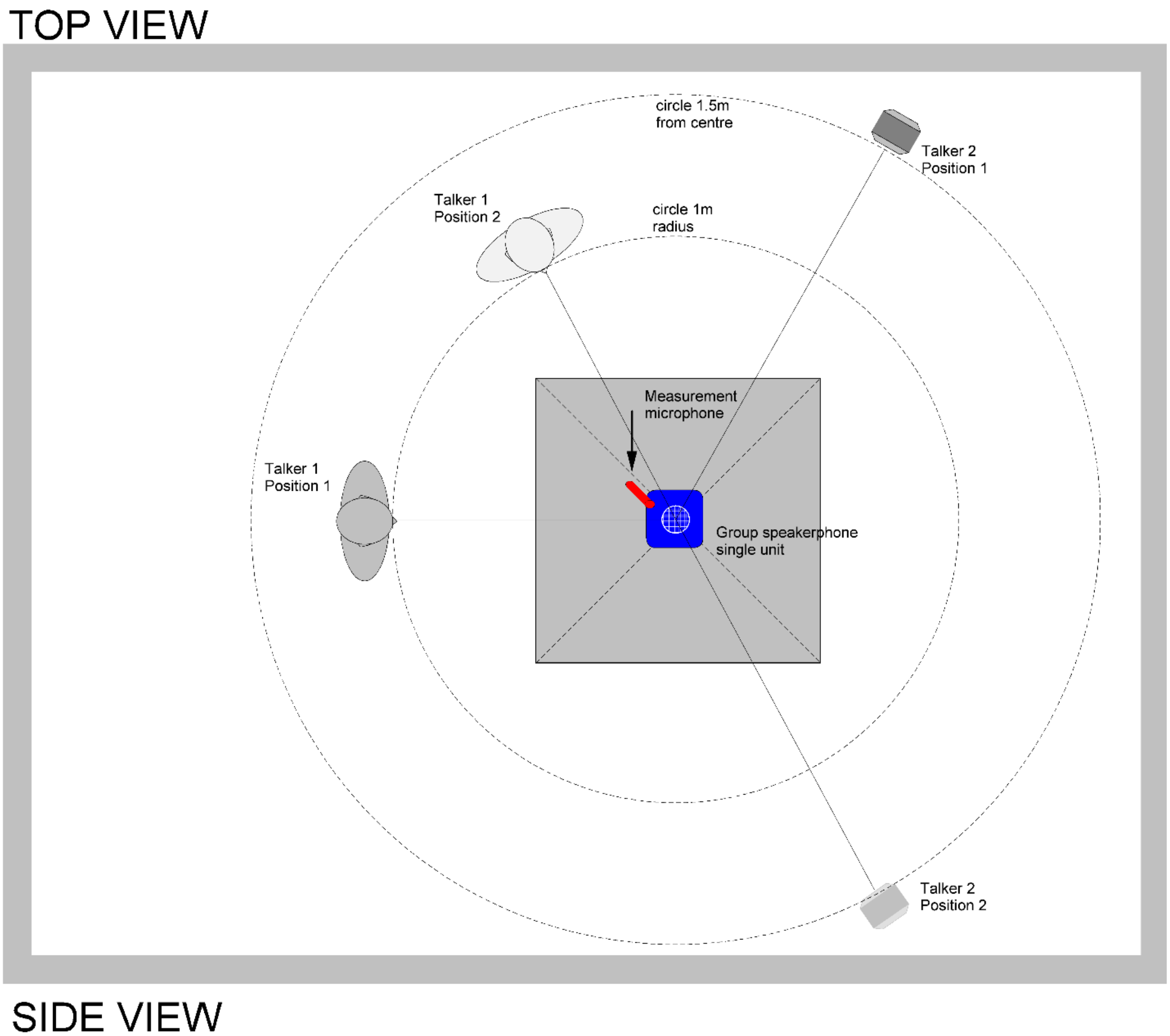


Figure 49: Center of room speakerphones with single unit on table

5.3.3.4 *Center of room speakerphones with master + extension unit or extension microphone(s)*

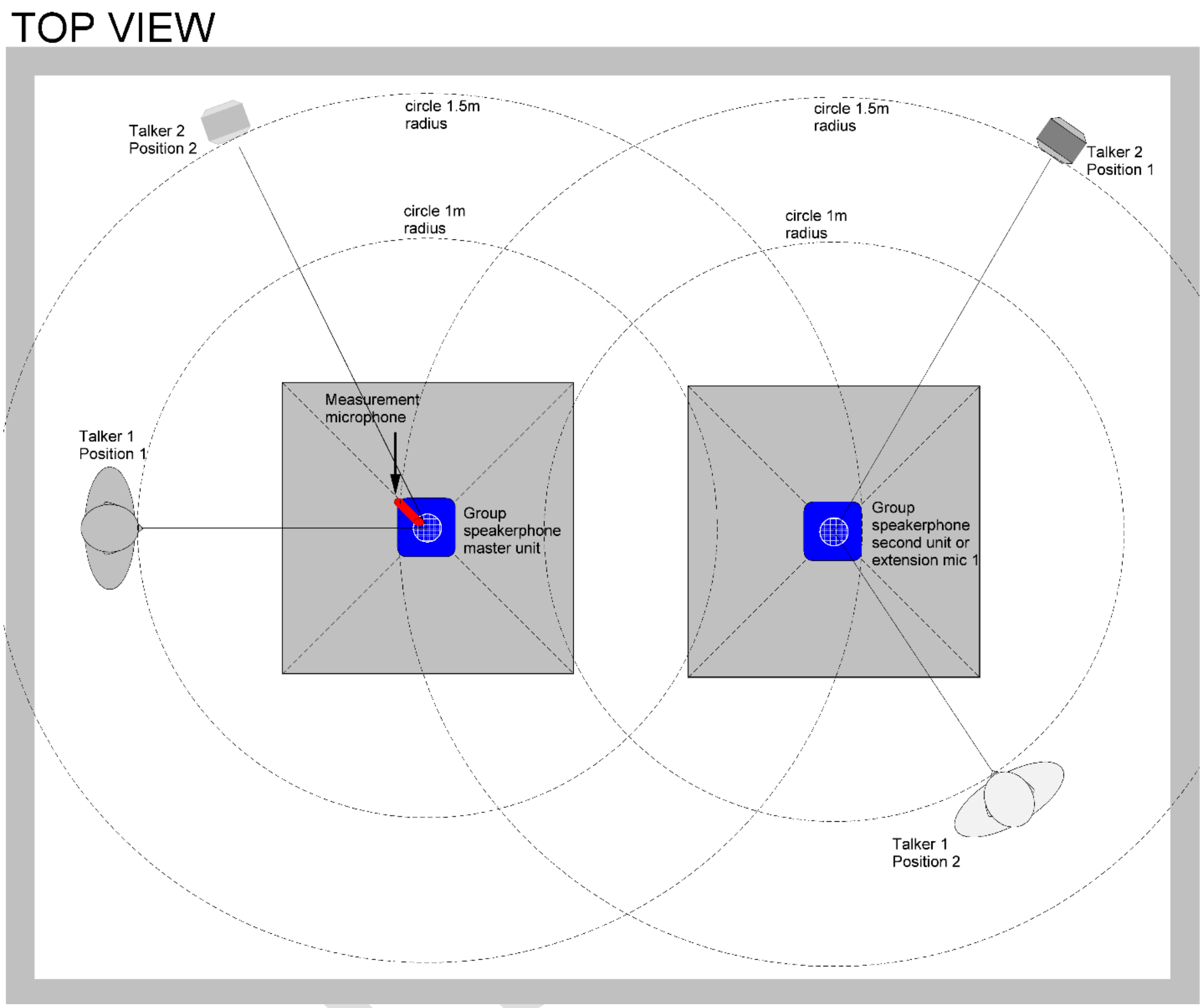


Figure 50: Center of room speakerphones with master + extension unit or extension microphone(s)

5.3.3.5 *Center of room speakerphones with camera, microphone and speaker in same enclosure*

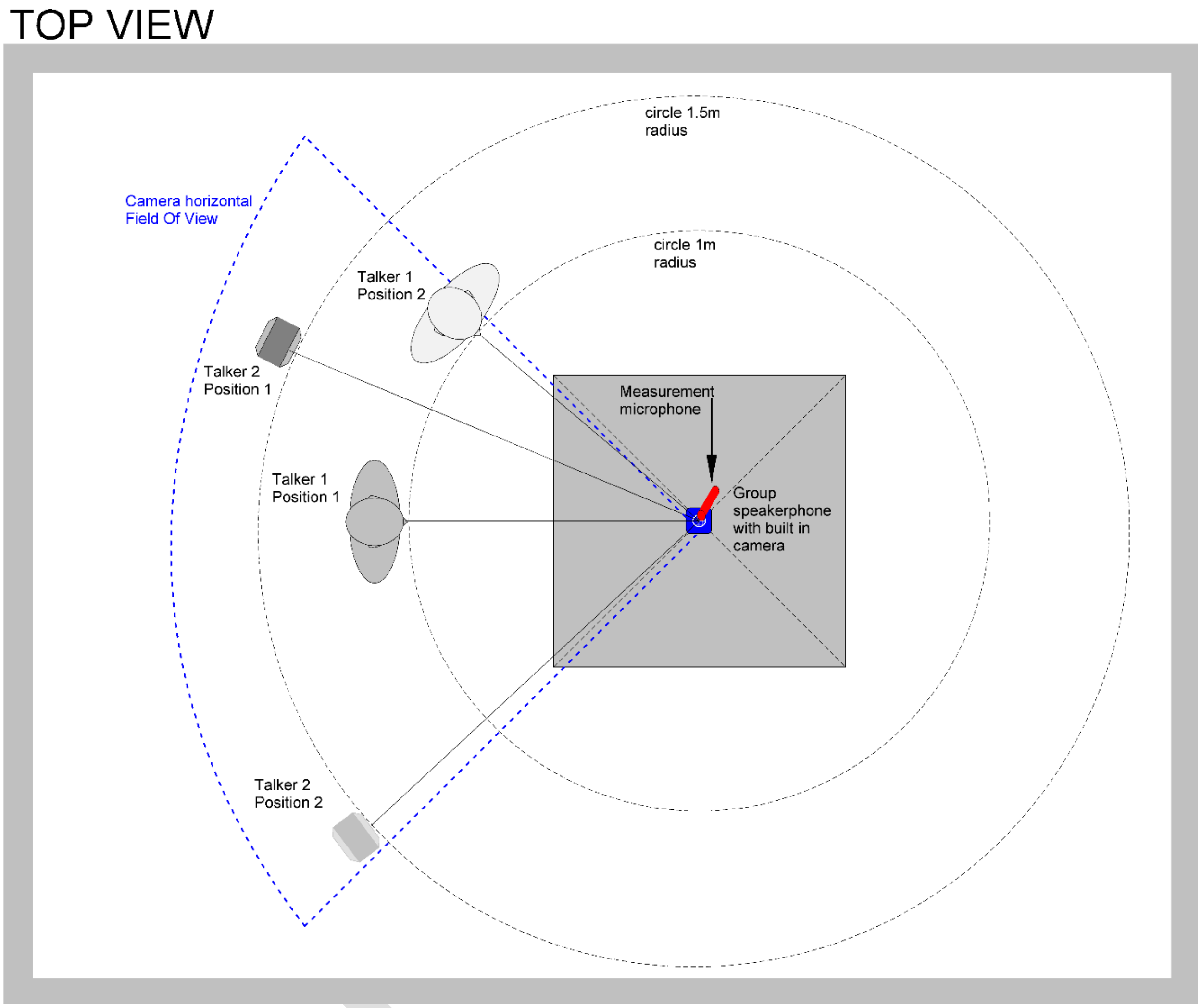


Figure 51: Center of room speakerphones with camera, microphone and speaker in same enclosure

If the DUT provides wider than 90 degree coverage, then Talker 2 positions must be extended so that each next 90 degree is tested at 2 additional positions. For example, for a 360 degree view the talker 2 will be placed in 8 different positions around the device.

Tester should also observe the send video during the test and verify the correct operation of the speech based “active talker detection” is the device provides such functionality.

5.3.3.6 *Center of room speakerphones with camera and microphone as one unit + additional speakerphone or microphones on table.*

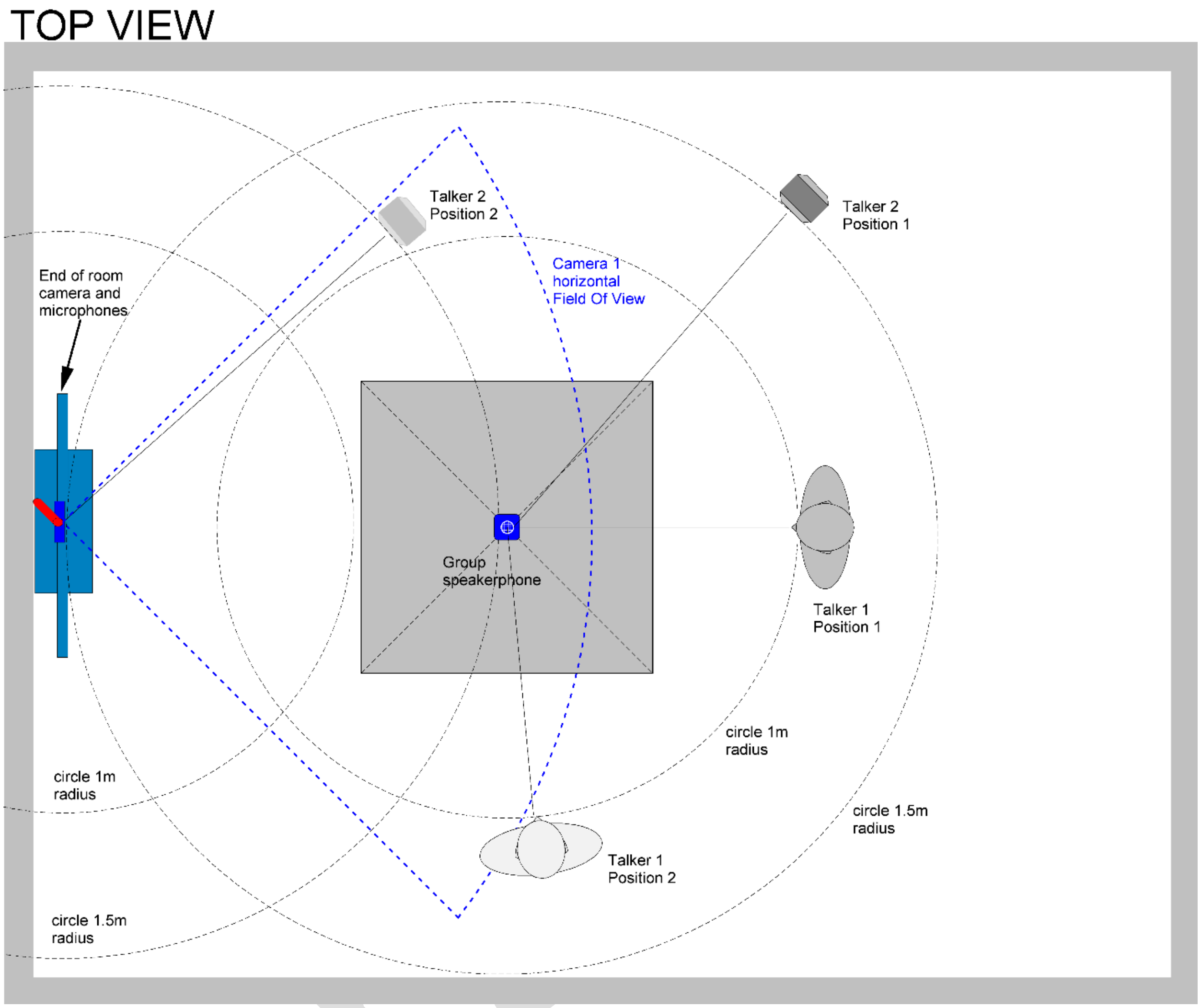


Figure 52: Center of room speakerphones with camera and microphone as one unit + additional speakerphone or microphones on table.

5.3.3.7 *Edge of room speakerphone*

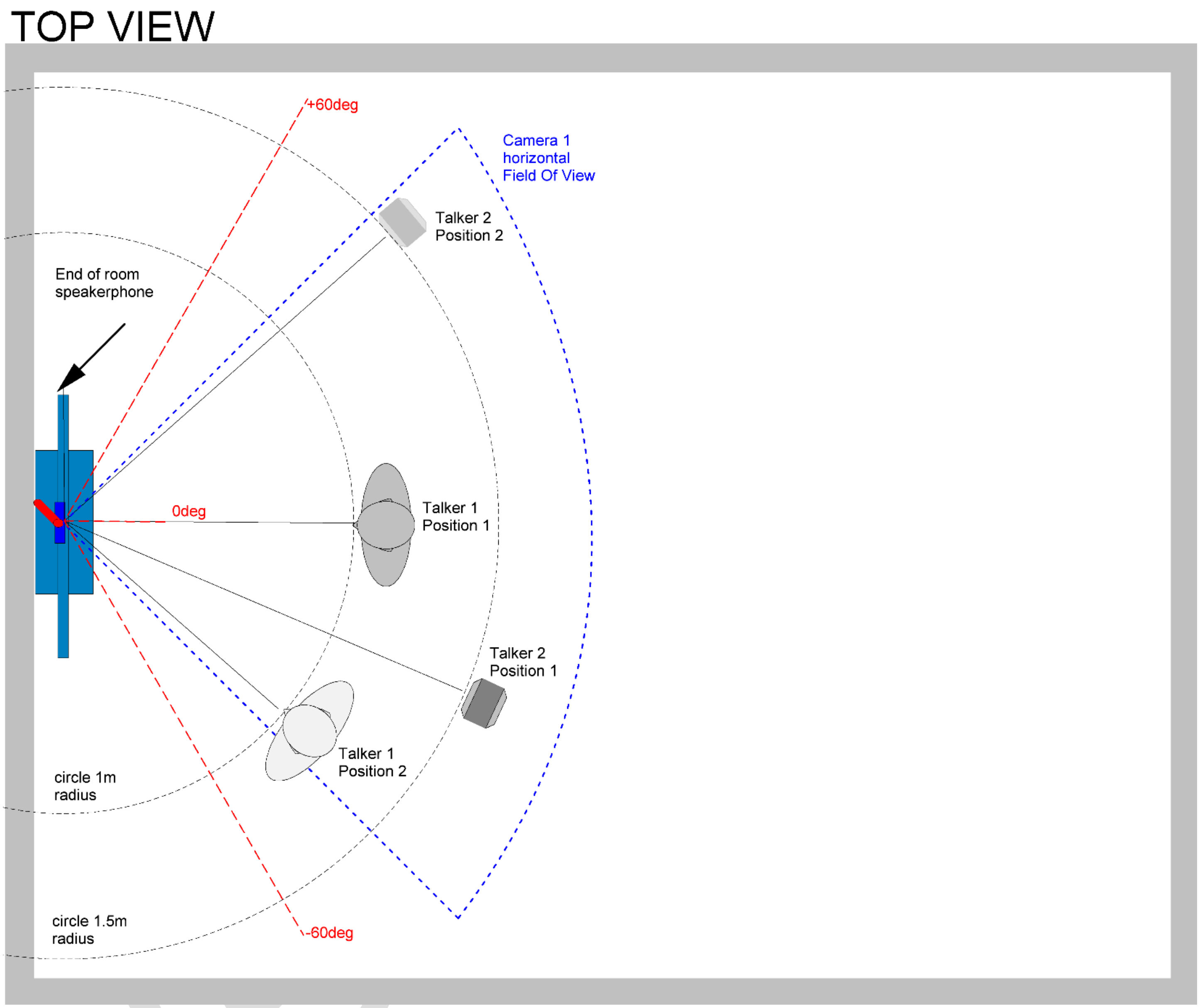


Figure 53: Edge of room speakerphone

HATS and artificial mouth are adjusted vertically to be at the mid height of the TV screen. For DUT devices with camera then for Position 2 the near end talkers are placed on the edge of the camera field of view (blue dash line). If the DUT provides more than 2 cameras, then Talker 2 positions must be extended so that each next camera Field of View area is tested at FOV horizontal edges. Tester should also observe the send video during the test and verify the correct operation of the speech based "active talker detection" works correctly.
If the DUT does not include a camera then the Position 2 is at +60deg and -60 deg lines (red dashed lines).

In case of soundbar type of speakerphone the soundbar is mounted on the wall either on top or bottom of the typical big screen TV. The user guide or a vendor recommendation is taken as a guideline for choosing the mounting position.

5.3.3.8 *Edge of room speakerphone (big screen collaboration device)*

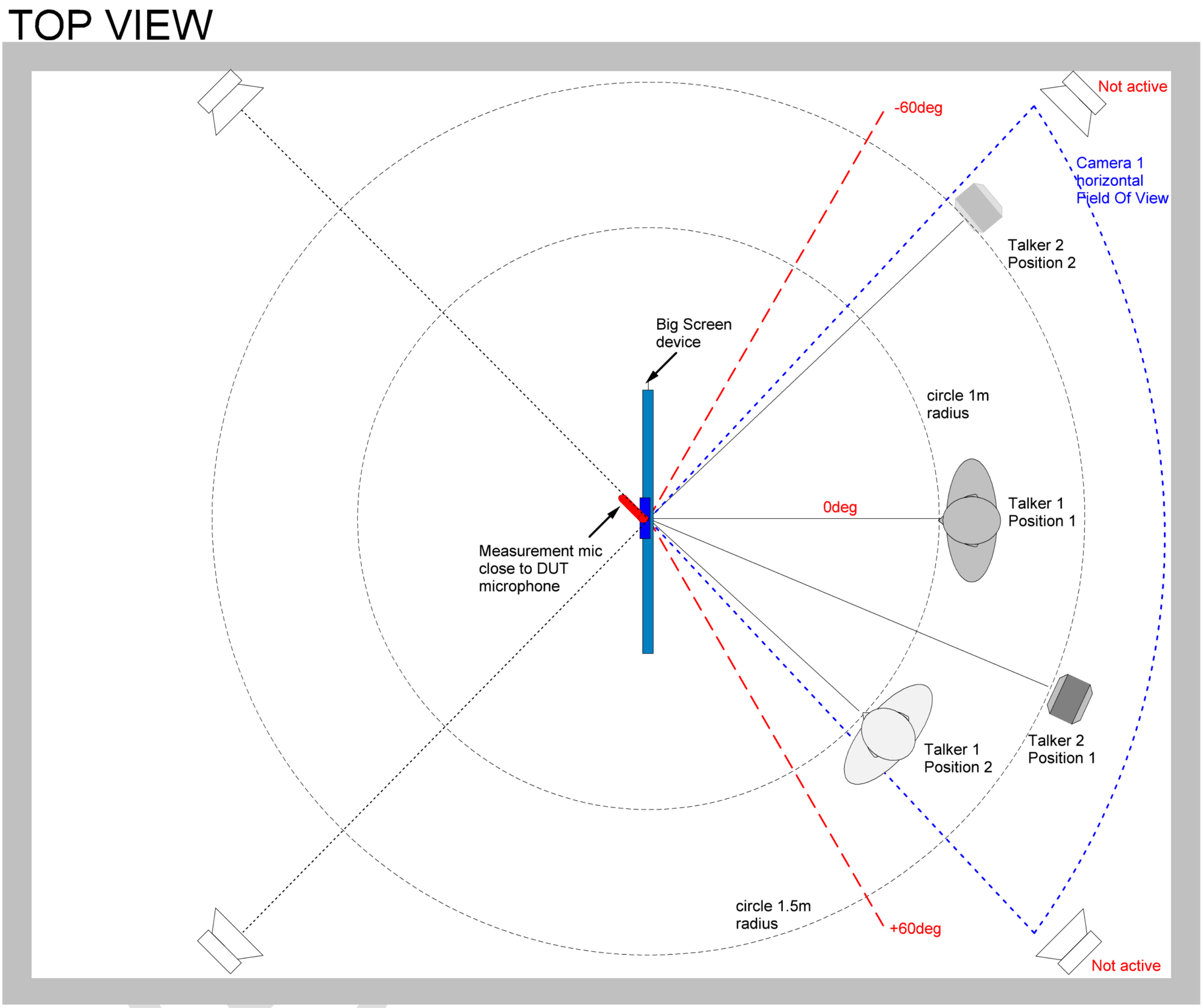


Figure 54: Edge of room speakerphone – big screen collaboration devices

HATS and artificial mouth are adjusted vertically to be at the mid height of the TV screen.

If the DUT does not include a camera then the Position 2 is at +60deg and -60 deg lines (red dashed lines).

5.3.3.9 *DSP mixer with ceiling panel microphone(s) and speaker*

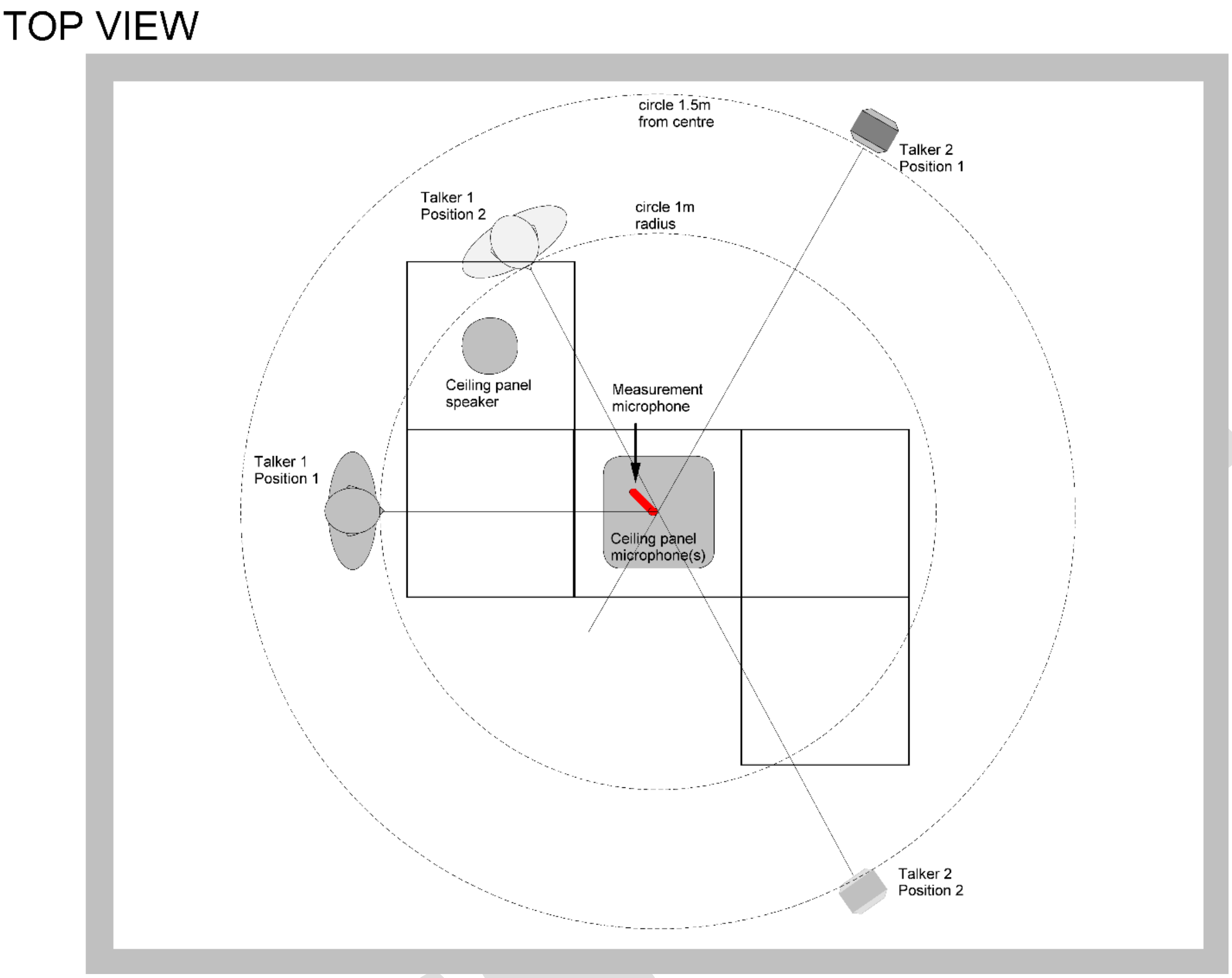


Figure 55: Ceiling panel microphone(s) and speaker – top view

Note: The 60cm*60cm rectangles indicate the minimum amount of ceiling panel framing needed in case the test room does not have a fixed framing installation for the panels.

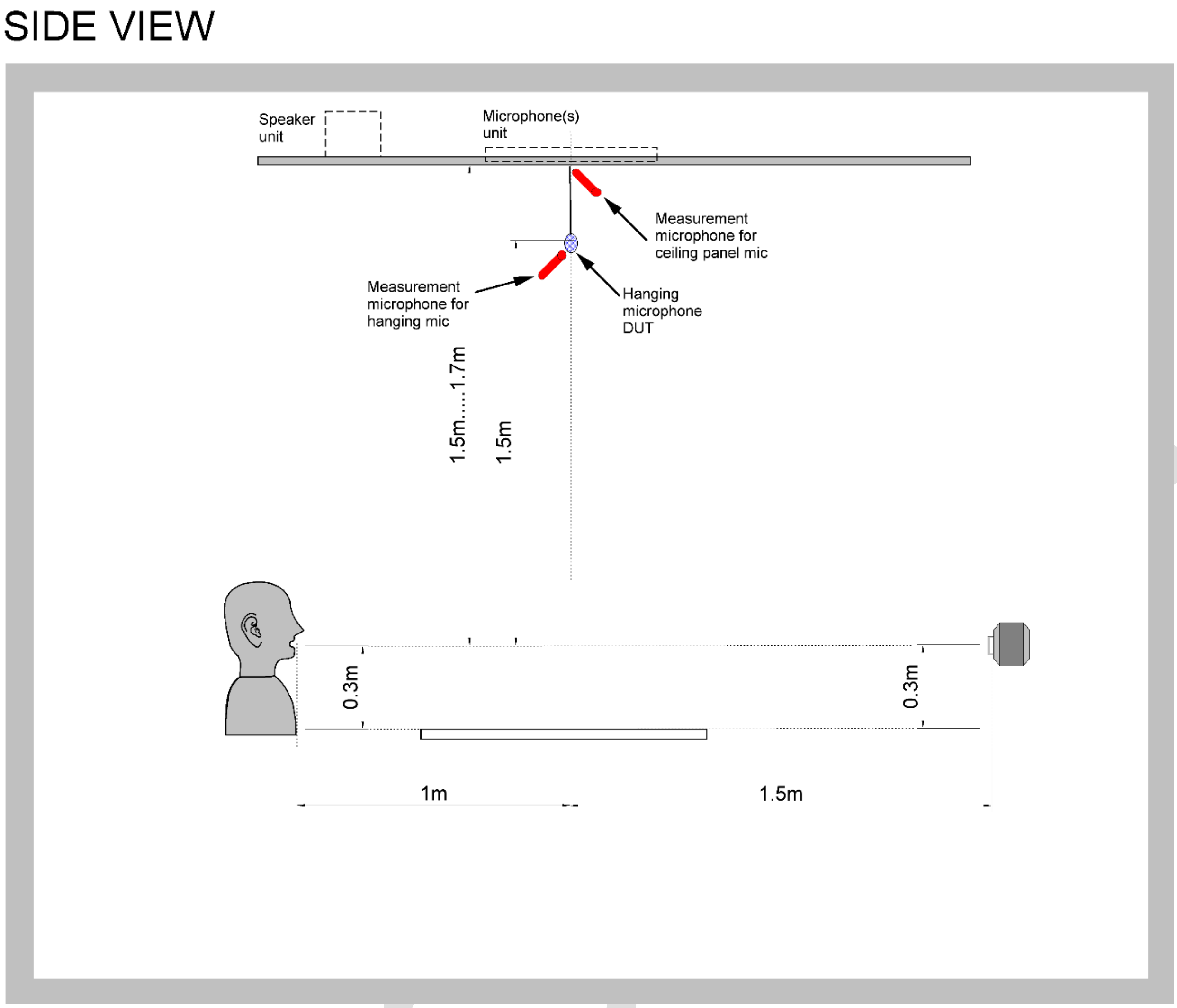


Figure 56: Ceiling panel microphone(s) and speaker – side view

5.3.3.10 *DSP mixer with table microphone(s) and front of room speaker(s)*

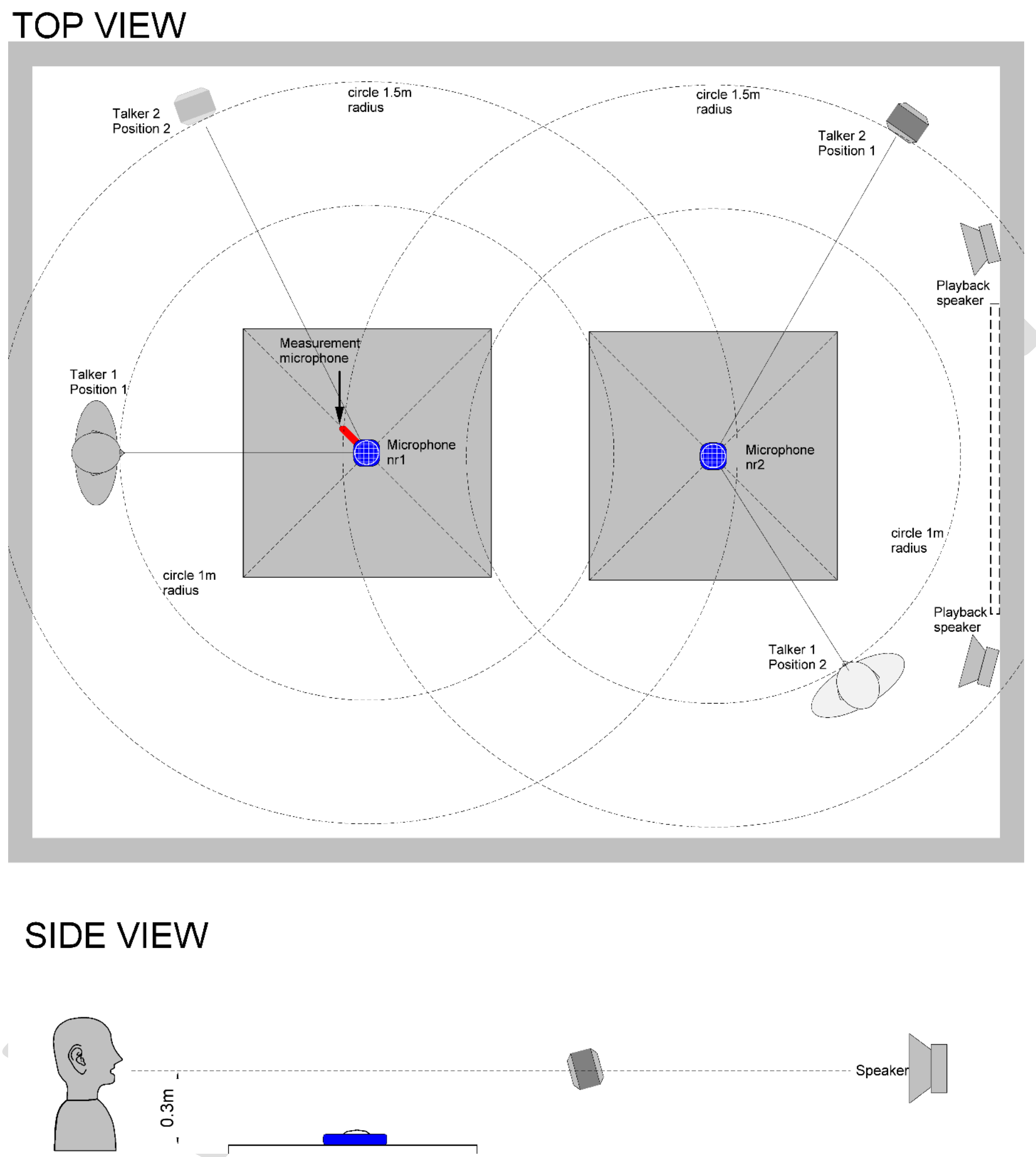


Figure 57: DSP mixer with table microphone(s) and front of room speaker(s)

Note: The above test scenario uses only 2 extension microphones while most Large Meeting Room installations will probably need more. The test setup above is a compromise to allow usage of the objective audio test setup.

5.3.3.11 *Variable echo path test for Center of room speakerphones*

This diagram shows setup for 4.4.8 Echo path – stability loss with variable echo path.

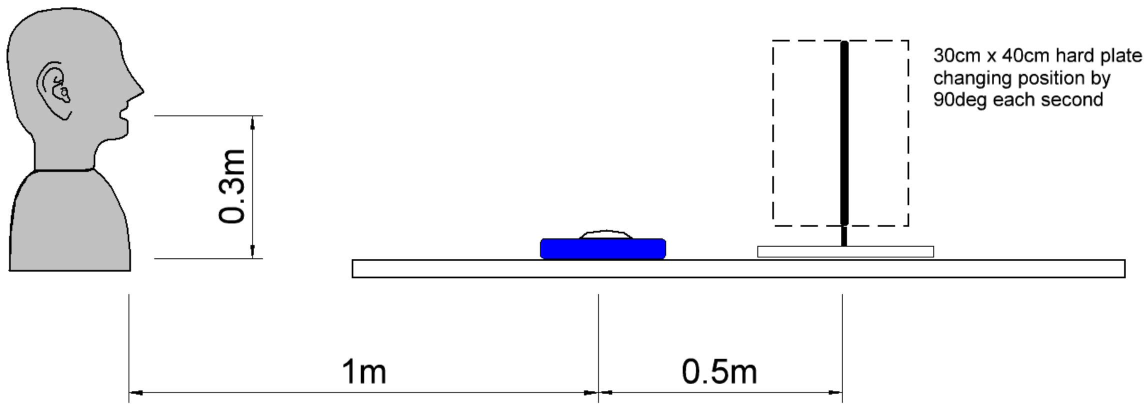


Figure 58: Variable echo path test for Group conferencing devices

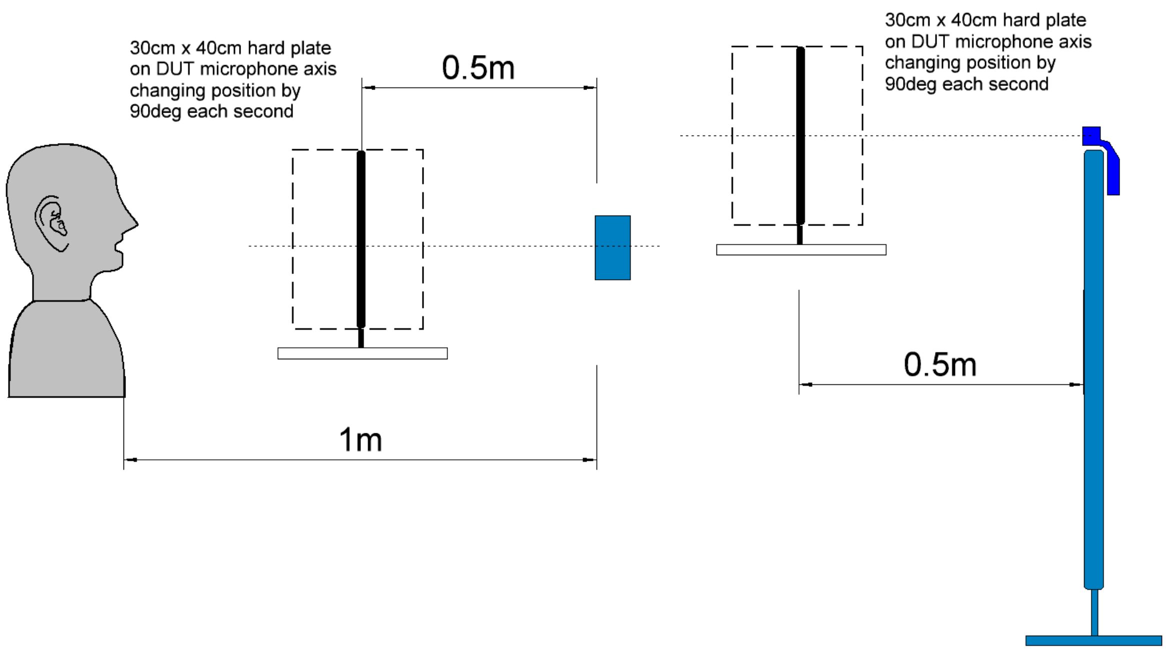


Figure 59: Variable echo path test for Edge of room conferencing devices

6 Appendix

6.1 DSP mixer based conference system solutions

Large meeting rooms require more customizable solutions due to difference in size, layout of the room, size and shape of the table etc. Thus, more flexible solutions such as DSP mixer based solutions are a better option.

This section of specification states the requirements and test setup description for this category of devices. The objective test setups use an anechoic room environment in Section 5.2 and a reverberant room environment specified in Section 5.3. The reverberant room size is smaller than the intended large conference room, but the tradeoff is made to allow the objective tests to be carried out.

Suitable products in this category should provide following technical capabilities:

- Rack mounting option
- USB soundcard interface to Skype Room System computer with 48kHz/16bit or 48kHz/24bit input/output capability
- Unique USB Product ID / Vendor ID required for audio offloading
  - In case the USB audio functionality is an add-on option module, then the PID/VID need to be unique to the specific Logo approved DSP mixer model.
- USB connection must have two-way HID command support to:
  - Allow muting of all microphones from software client.
  - Reflect microphone mute state back to software client.
  - Software client playback volume adjustment affect the system master volume
  - If hardware volume adjustment is available, then adjusting the playback volume must be reflected in software client.
- 4 or more microphone input channels (analog or digital based)
  - In case of analog microphone inputs the device should be able to provide the +48V phantom power to the microphones
- 2 or more analog or digital microphone input channels
- 2 or more analog or digital output channels
- In/out configurations and DSP parameters configurable over USB or IP network.
  - Enables saving and reloading full configuration and parameters file.
- If the device is active cooled by fan then the DSP mixer unit is tested by placing it in the same test position as shown on Figure 29 for center of room speakerphone and the testcases 4.2.4 and 4.2.5 must pass the set requirements.
  - If the device bundle includes power amplifier or a network switch that is active cooled by a fan the same requirements applies to those devices.

Suitable products in this category shall provide following DSP audio processing capabilities:

- Processed audio must be at least Wideband quality (100Hz to 8kHz audio bandwidth)
- Individual Gain/AEC/NS per each microphone channel
- Configurable AEC parameters to be able to handle rooms from low (RT60<0.5sec) to high reverb (RT60<1.2sec) while maintaining best possible doubletalk.
- Configurable NS to allow tuning for a room and expected noise level in the room.
  - Including speech(voice) activity detection to avoid passing non-speech like non-periodic noises.
- Configurable per channel microphone gains to tune the levels per talker to microphone distances.

- Configurable per channel output gains to tune the levels per room and listener to speaker distances.
- Minimum of 5 parametric EQ-s and/or shelving filters per each in/out channel to allow tuning the microphone / speaker response to the room.
- Configurable frequency mof at least 2nd order for all inputs.
- Configurable frequency HP/LP filters of at least 2nd order for all outputs.
- Microphone signal auto-mixer with a possibility to set 1 microphone channel as a priority microphone. In addition to "gating" mode the device must also provide a "gain sharing" mode for auto-mixer option.

Bundle to be provided for testing

In addition to the main DSP mixer a vendor should provide the following:

- Minimum of 2 microphones in case of table or hanging microphones (which can contain several microphone capsules and/or beamforming etc.).
  In case of ceiling panel microphone module replacing a full ceiling tile - 1 unit is sufficient.
- Minimum of 2 speakers for playback. These can be ceiling panel mounted, wall mounted or table mounted versions.
- Any other necessary audio or power modules. For example, power amplifier for speakers in case the main unit does not have that capability. Also in case of IP based microphones/speakers a suitable PoE capable switch must be provided.

After the successful test the Microsoft Technet reference will include details on the accessories used when the DSP mixer device was approved. Thus, the vendor is free to provide 2 or 3 alternate configurations – for example table microphones + wall mounted speakers and another bundle of ceiling panel microphone with ceiling panel speakers.

6.1.1 Testing guidelines

6.1.1.1 *Anechoic room tests guidelines*

- A single physical microphone unit (can contain several microphones) is connected for the testing.
  1. If the unit has more than 1 microphone, then a mode where each microphone provides an individual audio channel is used.
  2. Auto mixer should be configured to allow choosing the correct(active) microphone throughout the tests.
- In case of ceiling panel microphone, it is permitted to reverse the test setup – i.e. place the microphone panel on the "floor" of the anechoic room and raise the mouth simulator to the defined distance. The mouth simulator should be pointing to a horizontal plane and not rotated to point toward the ceiling panel. In the real-world use case of such microphones a talker will not look upwards while talking.
- Loudspeaker is placed 60cm from the microphone behind the microphone when looked at from an artificial mouth direction
- Anechoic room test flow for DSP Mixer based devices:
  1. Run the Receive path - output level test 4.2.1 and set the playback loudness to a preferred playback level
  2. Run Receive path - frequency response test 4.2.7. Tune the HP/LP and EQ filters to pass the requirements
  3. Run the Echo path - terminal coupling loss (TCL) test 4.3.1 and tune the microphone gains to be suitable for best AEC performance
  4. Run the Send path – normal speech preparation tests to adjust the microphone gains to required levels

5. Carry on the full tests flow of Sections 4.1, 4.2, and 4.3
   - If the microphone(s) provide several pickup patterns (cardioid, super cardioid, etc.) then the send path – total quality loss test 4.1.1 and frequency response tests 4.1.10 should be tested for each setting. 4.1.1 and frequency response tests should be tested for each setting.

6.1.1.2 *Reverberant room tests guidelines*

Physical test setups are described in Figure 55, Figure 56, and Figure 57.

- Reverb room test flow for DSP Mixer based devices
  - Run the 4.4.4 Receive output level test. Adjust the playback level to meet the criteria.
  - Verify the microphone signal peak levels during the Echo path – preparation tests.
  - Same tunings for send path should be suitable also for reverberant room tests.
- Run all the conferencing device applicable tests in Section 4.4

6.1.1.3 *End to end test guidelines*

- The DUT solution should be set up in a large meeting room of 35…45m2 with reverberation time of 0.4s < RT60 < 0.9s
- Microphone and playback speaker placements
  - In case of single capsule table microphones, set up at least 4 microphones with at least 2m distances between each microphone (square shape positioning)
    - Single capsule table microphone to closest playback speaker distance should be 1m for at least 1 of the 4 microphones
  - In case of beamforming or directional pattern multi capsule table microphones set up at least 2 microphone units with a minimum 2m distance between the microphones.
    - Multi capsule table microphone to closest playback speaker distance should be 1m for at least 1 of the 2 microphones
  - In case of hanging microphones set up at least 4 microphones with at least 2m distances between each microphone (square shape positioning)
    - Hanging microphone to closest playback speaker distance should be 1m for at least 1 of the 4 microphones
  - In case of ceiling panel microphone set up at least one microphone module
    - Ceiling panel microphone to closest playback speaker distance should be 60cm
- The end to end test calls should be run with the Microsoft Teams Rooms calling a Microsoft Teams client.
- The far end participant should be using a Microsoft Teams certified headset with stereo playback (certified within last 2 years is recommended).
- There should be 4 near end talkers in the room. The goal is to evaluate the multi microphone mixing and its ability to choose the correct microphone channel where an active talker talks.
  - In case of 4 single capsule microphone setup each talker is closest to one of the microphones
  - In case of 2 microphone units 2 talkers are close to first while 2 are closer to second microphone unit
  - In case of ceiling panel microphones people are positioned in circle around the microphone panel's vertical center axis

Following scenarios are to be tried:

- Each 4 near end talker says 2..3 sentences and then the turn goes to next near end talker. Far end participant will note how smooth and seamless the transition is from talker to talker. There should be no evident lag of low speech level or poor clarity of speech while talking turns change.
    - In case of table microphones this test shall be repeated by placing laptops with lid open between the near end talker and the table microphone
- The far end participant talks at normal and then at loud level.
    - Near end users not the clarity of playback speech. Also, any sign of distortion or noise boost should be noted.
    - Far end user is listening for echoes of his/her own voice.
- The far end talks at rapid pace with minimal silent gaps in between words and sentences. Near end talkers will try to interrupt the far end talker with their own sentences. Far end will register if the break in attempts succeed (meaning far end user does detect other trying to break into the conversation). This will validate the double talk capability of the device. During this test far end user is also listening for echoes of his/her own voice.
- System stability tests
    - Maximize the playback volume. Far end user should talk loudly and listen for his/her own echo's leaking back. Same shall be repeated with near end talkers interrupting the far end user occasionally
    - While in call a near end user picks up a laptop and places in between speaker and one of the microphones while the far end is talking. The echo canceller leaks should disappear in max 2 seconds time.
    - One of the near end user will cough loudly while close to the microphone (50cm distance). Far end user will observe if automated gain controls will react reducing the signal overload. After the coughing has ended near end users start to talk again. Far end user should observe if the normal send level is retuned in max 10 seconds time.

6.1.1.4 *Subjective test assessments: far-field voice-command related*

The tests below should be conducted as follows:

- Room size – the room size should be in range from 10m2 to 25m2.
- Treatment of the room – the reverberation time of the room should be
- 0.4s < RT60 < 0.7s If the device is intended for middle of the table placement (conferencing use case) then the near end user should be at 2.5m from the device for following tests.
- If the device is meant to be positioned as End of Table type of device close to the wall of the room, then near end talker should be at 4m distance from the device for following tests.
- If the device is a multi-use case product that can be placed freely in any position of the room, then both the 2.5m and 4m distances for near end user should be tested.

DUT user assesses the following areas on the scale from 1 to 5:

- Give a voice command of "Hey <u>assistant</u>, set (and later unset) **Skype (Teams) to Do Not Disturb mode**" (and possible alternations of same command – like DND) – does the device make an expected action.
- Starting a call with a voice command of "Hey <u>assistant</u>, call <u>Contact</u> **via Skype**" (and possible alternations of same command) – does the device make an expected action.
- Starting a call with a voice command of "Hey <u>assistant</u>, call <u>Contact</u> **on mobile**" (and possible alternations of same command) – does the device make an expected action.

- Starting a call with a voice command of “Hey assistant, call Contact **on landline**” (and possible alternations of same command) – does the device make an expected action.
- Starting a call with a voice command of “Hey assistant, call **dictate the number**” (and possible alternations of same command) – does the device make an expected action.
- Rejecting an incoming call with a voice command of “Hey assistant, **ignore the call**” (and possible alternations of same command) – does the device make an expected action.
- Answer an incoming call “Hey assistant, **answer** the call” (and possible alternations of same command) – does the device make an expected action.
    - For video call enabled devices “Hey assistant**, answer the call without video**” (and possible alternations of same command) – does the device make an expected action.
    - For video call enabled devices “Hey assistant, **answer the call with video**” (and possible alternations of same command) – does the device make an expected action.
- Once the call is ongoing try “Hey assistant, **mute the call**” (and possible alternations of same command) – does the device make an expected action.
- Once the call is ongoing try “Hey assistant, **unmute the call**” (and possible alternations of same command) – does the device make an expected action.
- Once the call is ongoing try “Hey assistant, **increase the volume /decrease the volume**” (and possible alternations of same command) – does the device make an expected action.
- Once the call is ongoing try “Hey assistant, **add a contact the call**” (and possible alternations of same command) – does the device make an expected action.
- Once the call is ongoing try “Hey assistant, **add a reminder to my calendar**” (and possible alternations of same command) – does the device make an expected action.
- Ending a call with a voice command of “Hey assistant, **hang up**” (and possible alternations of same command) – does the device make an expected action.

Far end user assesses the following areas on the scale from 1 to 5

- During an ongoing call - near end user gives a command to an assistant and assistant will respond:
    - IF the design of the DUT is such that far end should hear the command and the assistants answer, then did the assistant voice come through without distortions, gaps, crack sounds?
    - If the design of DUT is such that far end should not hear the assistant voice, then far end should evaluate if no echo’s or other residuals of the assistant speech come through to the far end.
- During an ongoing call: Far end user gives the command “Hey assistant, set max volume” (and possible alternations of same command). Check if DUT device assistant reacts to commands from far end. If it does, this is not an acceptable behavior (error in implementation of echo cancellation).

6.1.1.5 *End to end test assessments in quiet environment*

With the DUT in normal usage environment, place an E2E SfB / Skype or Teams call between the two users.

DUT user assesses the following areas on the scale from 1 to 5:

- Rate the quality of far end speech: is it too dark, too bright, metallic? Is the dynamics healthy?
- Are the highest speech levels distorted with LOUD speech from far end and maximum volume settings at DUT?
- Do you hear electric interferences (such as buzzing sound)? How annoying is it?
- Do you hear annoyingly high noise floor or noise fluctuations?

- Speak simultaneously. Is the communication fluent? Can the DUT user interrupt the far-end speaker when the latter is speaking?
- Set the DUT volume to maximum. Speak simultaneously. Is the communication fluent? Can the DUT user interrupt the far-end speaker when the latter is speaking?
- Rate the overall impression.

Far end user assesses the following areas on the scale from 1 to 5:

- How intelligible is the DUT user speech during single talk?
- Are words or parts of words missing or attenuated during single talk? What if DUT user speaks quietly? What if the laptop is placed between talker and DUT?
- Rate quality of speech: is it too dark, too bright, metallic? Is the dynamics healthy? - If necessary, re-evaluate in the end of the call.
- Do you hear electric interferences (such as buzzing sound)? How annoying is it?
- Does the DUT mute function cause interferences?
- Do you hear annoyingly high noise floor or noise fluctuations?
- Do you hear any other interferences (such as cracks, pops, drop-outs, breaks in static background noise etc.)?
- Do you hear disturbing handling noise? Is the device picking up external mechanical noises from the table surface (for example: external mouse usage)?
- When the call starts, can you hear the echo of your own voice? Does the echo disappear after <5 seconds (preferably <2sec).
- When DUT is used so that echo path changes (for example: let the DUT user adjus the playback volume or bend over the device): can you hear the echo of your own voice? Is the echo canceller adjusting within reasonable time (<3seconds)?
- Speak simultaneously. Is the communication fluent? Can the DUT user interrupt you during speaking?
- Speak simultaneously. Can you hear the echo of your own voice - how badly?
- When DUT volume is at maximum, can you hear the echo of your own voice when you speak in a loud voice (e.g. numbers)?
- When DUT volume is at maximum, can you hear the echo of your own voice, and how badly?
- Rate the overall impression.

6.1.1.6 *End to end test assessments in noisy environment*

With the DUT in noisy usage environment, place an E2E Microsoft Teams video call between the two users.

Far end user assesses the following areas on the scale from 1 to 5:

- Rate quality of DUT user speech in background noise during single talk.
- Are words or parts of words missing or attenuated during single talk? What if DUT user speaks quietly? What if the laptop is placed between talker and DUT?
- Rate quality of background noise. Do you hear noise level variation (pumping, cutting) between speech and silence periods? How annoying is it? Does noise sound otherwise natural and pleasant?
- Speak yourself while the other side is silent. Rate the quality of comfort noise.

6.1.1.7 *End to end test assessments in a conference call*

With the DUT in normal usage environment, place a Microsoft Teams conference call between three users (DUT will use the Skype Room Systems client). In addition to the far end there has to be an assisting tester using a headset and speaking at normal speech level.

Far end user assesses the following areas on the scale from 1 to 5:

- Is the speech level similar to the assisting tester?
- Are short interruptions by assisting tester causing speech degradation?
- Can you hear the echo of the assisting tester voice or your own voice - how badly?

6.2 End to end tests on conferencing devices with camera

This section applies to conferencing devices that have camera integrated or bundled.

With the DUT in a normal usage environment, place an E2E Microsoft Teams video call between the two users. One more tester should be present as assistant next to the DUT user.

Far end user assesses the following areas on the scale from 1 to 5:

- Is the DUT user voice picked up equally well from the DUT-s whole field of view?
- In case the solution provides directional audio then the audio and video directionalities must match.
  - Can be verified during test with near end user moving from far left in video FOV slowly to far right of the video FOV while talking at >1.5m distance from device. Clearly audible changes in level, low/mid/high frequency balance should be reported as issues on test report.
- Active speaker view must not react to the speech reflections from a whiteboard, windows or walls.
- Active speaker view must handle multiple people talking simultaneously. Such talking turn changes must not cause rapid video jumps from one to the other near end user.

If DUT claims directionality: How noticeable is the attenuation of sound sources outside the active pickup area? (Assisting tester should be talking and changing positions inside and outside the claimed pick-up area.)

6.3 Recommended speech pickup in different audio UI categories for personal solutions

This section recommends the speech pickup patterns for devices that support multiple audio UI categories.

Scenarios for DUT that enables multiple audio UIs	Desired behavior
Handheld speakerphone audio UI	Front +/- 45°, suppresses sounds from rear
Personal speakerphone audio UI	Omni-directional or Front +/- 90°
Headset audio UI	Switches to headset processing.
Handheld speakerphone audio UI (audio playback through headphones)	Switches to appropriate handheld or personal speakerphone audio UI mode.
Handset audio UI	Switches to handset processing

Table 74: Recommendations for speech pickup for devices that can be used in different audio UI categories

Scenarios for DUT with camera(s)	Desired behavior
Front-facing camera in use, Handheld speakerphone audio UI	Front +/- 45°, suppresses sounds from rear
Rear-facing camera video call, Handheld speakerphone audio UI	Omni-directional
Camera in use, Personal speakerphone audio UI	Omni-directional or Front +/- 90°
Camera in use, Headset audio UI	Switches to headset processing.
Camera in use, Handheld speakerphone audio UI (audio playback through headphones)	Switches to appropriate speakerphone audio UI mode.
Collaboration: huddle space (edge of room, half circle layout, office ambient noise)	Front pickup similar to camera horizontal FOV, suppresses sounds from sides. ***Nice to have*** auto detection of people up close to device for hands on collaboration causing auto switch to wider angle mic pickup range and mic/speaker gain optimization. ability to detect and attenuate the touch or pencil noises in send signal.
Collaboration: lounge space(edge of room + table, half circle layout, office ambient noise)	Front pickup similar to camera horizontal FOV, suppresses sounds from sides. ***Nice to have*** – extension microphone support to allow placing mic closer to talkers. ability to detect and attenuate the touch or pencil noises in send signal.
Collaboration: meeting room (edge of room, half circle layout, meeting room ambient noise)	Front pickup in half circle around device. ***Nice to have*** – auto detection of people up close to device for hands on collaboration causing auto switch to wider angle mic pickup range and mic/speaker gain optimization. ability to detect and attenuate the touch or pencil noises in send signal.

Table 75: Recommendations for speech pickup for devices with built-in cameras