

---

---

**Simultaneous interpreting — Quality  
and transmission of sound and image  
input — Requirements**

*Interprétation simultanée — Qualité et transmission des signaux  
audio-vidéo — Exigences*

STANDARDSISO.COM : Click to view the full PDF of ISO 20108:2017



STANDARDSISO.COM : Click to view the full PDF of ISO 20108:2017



**COPYRIGHT PROTECTED DOCUMENT**

© ISO 2017, Published in Switzerland

All rights reserved. Unless otherwise specified, no part of this publication may be reproduced or utilized otherwise in any form or by any means, electronic or mechanical, including photocopying, or posting on the internet or an intranet, without prior written permission. Permission can be requested from either ISO at the address below or ISO's member body in the country of the requester.

ISO copyright office  
Ch. de Blandonnet 8 • CP 401  
CH-1214 Vernier, Geneva, Switzerland  
Tel. +41 22 749 01 11  
Fax +41 22 749 09 47  
copyright@iso.org  
www.iso.org

# Contents

	Page
<b>Foreword</b> .....	<b>iv</b>
<b>Introduction</b> .....	<b>v</b>
<b>1 Scope</b> .....	<b>1</b>
<b>2 Normative references</b> .....	<b>1</b>
<b>3 Terms and definitions</b> .....	<b>1</b>
<b>4 Sound input to interpreters</b> .....	<b>2</b>
4.1 General.....	2
4.2 Microphones.....	2
4.2.1 Conference system microphones.....	2
4.2.2 Auxiliary sources.....	3
4.2.3 Microphone management.....	3
4.3 Conference system.....	3
4.3.1 General.....	3
4.3.2 Sound input and output.....	3
4.4 Intelligibility.....	7
<b>5 Image input to interpreters</b> .....	<b>7</b>
5.1 General.....	7
5.2 Image quality.....	7
5.3 Image projection.....	7
<b>6 Lip sync</b> .....	<b>7</b>
<b>7 Transmission of audio and video from a distant site</b> .....	<b>8</b>
7.1 Quality.....	8
7.2 Latency.....	8
<b>Annex A (informative) Room acoustics</b> .....	<b>9</b>
<b>Annex B (informative) Image content</b> .....	<b>10</b>
<b>Annex C (informative) Measurement of audio and video output quality</b> .....	<b>11</b>
<b>Bibliography</b> .....	<b>12</b>

## Foreword

ISO (the International Organization for Standardization) is a worldwide federation of national standards bodies (ISO member bodies). The work of preparing International Standards is normally carried out through ISO technical committees. Each member body interested in a subject for which a technical committee has been established has the right to be represented on that committee. International organizations, governmental and non-governmental, in liaison with ISO, also take part in the work. ISO collaborates closely with the International Electrotechnical Commission (IEC) on all matters of electrotechnical standardization.

The procedures used to develop this document and those intended for its further maintenance are described in the ISO/IEC Directives, Part 1. In particular the different approval criteria needed for the different types of ISO documents should be noted. This document was drafted in accordance with the editorial rules of the ISO/IEC Directives, Part 2 (see [www.iso.org/directives](http://www.iso.org/directives)).

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. ISO shall not be held responsible for identifying any or all such patent rights. Details of any patent rights identified during the development of the document will be in the Introduction and/or on the ISO list of patent declarations received (see [www.iso.org/patents](http://www.iso.org/patents)).

Any trade name used in this document is information given for the convenience of users and does not constitute an endorsement.

For an explanation on the voluntary nature of standards, the meaning of ISO specific terms and expressions related to conformity assessment, as well as information about ISO's adherence to the World Trade Organization (WTO) principles in the Technical Barriers to Trade (TBT) see the following URL: [www.iso.org/iso/foreword.html](http://www.iso.org/iso/foreword.html).

This document was prepared by Technical Committee ISO/TC 37, *Terminology and other language and content resources*, Subcommittee SC 5, *Translation, interpreting and related technology*.

## Introduction

In simultaneous interpreting, sound input is the interpreter's most important source of information. The image of the speaker and of all visual materials used in the conference is almost as important.

New technologies permit different forms of audio and video signals and are now being used increasingly to transmit sound and image to interpreters. Their quality determines the conditions of the interpreters' work. The transmission mode of sound and image is therefore vital and requirements should be standardized.

STANDARDSISO.COM : Click to view the full PDF of ISO 20108:2017

STANDARDSISO.COM : Click to view the full PDF of ISO 20108:2017

# Simultaneous interpreting — Quality and transmission of sound and image input — Requirements

## 1 Scope

This document sets out requirements for the quality and transmission of sound and image input to interpreters and specifies the characteristics of the audio and video signals. The components of typical interpreting systems are specified in ISO 20109. Together with either permanent (see ISO 2603) or mobile (see ISO 4043) booths, these interpreting systems form the interpreters' working environment. In addition to setting out the requirements for on-site interpreting, where participants (speakers and members of the audience) and interpreters are at the same location, this document specifies requirements for different varieties of distance interpreting situations in which the interpreters are not at the same location as one or more of the conference participants.

This document also addresses the work of manufacturers and providers of simultaneous interpreting equipment and technical staff.

In conjunction with either ISO 2603 or ISO 4043, this document and ISO 20109 provide the relevant requirements both for the quality and transmission of sound and image provided to interpreters and for the equipment needed in the booths, the conference room and the distant site(s).

## 2 Normative references

The following documents are referred to in the text in such a way that some or all of their content constitutes requirements of this document. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.

ISO 2603:2016, *Simultaneous interpreting — Permanent booths — Requirements*

ISO 4043, *Simultaneous interpreting — Mobile booths — Requirements*

ISO 18841<sup>1)</sup>, *Interpreting services — General requirements and recommendations*

ISO 20109:2016, *Simultaneous interpreting — Equipment — Requirements*

IEC 60268-16, *Sound system equipment — Part 16: Objective rating of speech intelligibility by speech transmission index*

ITU-R BT.1359, *Relative timing of sound and vision for broadcasting*

## 3 Terms and definitions

For the purposes of this document, the terms and definitions given in ISO 2603, ISO 4043, ISO 18841 and ISO 20109 and the following apply.

ISO and IEC maintain terminological databases for use in standardization at the following addresses:

- IEC Electropedia: available at <http://www.electropedia.org/>
- ISO Online browsing platform: available at <http://www.iso.org/obp>

1) To be published.

**3.1  
interpreting**

rendering of spoken or signed information from a source language to a target language in oral or signed form, conveying both the register and meaning of the source language content

[SOURCE: ISO 18841:2016, 3.1.2]

**3.2  
simultaneous interpreting**

mode of *interpreting* (3.1) performed while a speaker is still speaking or signing

Note 1 to entry: For the purposes of this document, the activity requires specialized equipment.

[SOURCE: ISO 2603:2016, 3.1, modified — Note 1 to entry has been modified.]

**3.3  
booth  
booth for simultaneous interpreting**

self-contained unit enclosing the interpreter's work space

Note 1 to entry: One of the purposes of simultaneous interpreting booths is to provide sound insulation, both from the noise transmitted from the booth's external environment to the booth and vice versa, and from noise passing from one booth to another.

Note 2 to entry: ISO 2603 applies to permanent booths; ISO 4043 applies to mobile booths.

[SOURCE: ISO 2603:2016, 3.2, modified — Note 2 to entry has been added.]

**3.4  
distance interpreting  
remote interpreting**

*interpreting* (3.1) of a speaker in a different location from that of the interpreter, enabled by information and communications technology (ICT)

[SOURCE: ISO 20109:2016, 3.10, modified — The term "simultaneous" has been deleted.]

**3.5  
video display**

electronic device that represents information in a visual form

[SOURCE: ISO 20109:2016, 3.9]

**4 Sound input to interpreters**

**4.1 General**

The overall sound input to interpreters shall comply with ISO 20109:2016, Clause 4.

**4.2 Microphones**

**4.2.1 Conference system microphones**

Conference system microphones shall have the same characteristics as interpreter console microphones set out in ISO 20109:2016, 4.1 and 6.3.

## 4.2.2 Auxiliary sources

### 4.2.2.1 Mobile microphones

Handheld, lapel, head-worn and any other microphone used shall have a dynamic range at least equal to that set out in ISO 20109:2016, 4.1.

### 4.2.2.2 External sources

Sound input from external sources like laptops and videoconferencing systems shall have a dynamic range at least equal to that set out in ISO 20109:2016, 4.1.

It shall be possible to switch videoconferencing microphones ON and OFF or to mute them.

Acoustic echo cancelling shall be in place at all sites.

### 4.2.2.3 Ambient microphones

Ambient microphones shall provide interpreters with the ambience of the conference room.

If enabled, one or more microphones shall switch ON when all other sound sources are switched OFF.

The ambient microphone(s) shall switch OFF when any conference system microphone is switched ON or any other sound source is fed into the system. Its level shall be adjustable.

### 4.2.2.4 Chime

A chime may be transmitted to the interpreters' headphones. The sound level of the chime shall be at least 20 dB less than the nominal level of the interpreting system.

## 4.2.3 Microphone management

The number of microphones switched ON shall be configurable with preferably only one microphone switched ON.

If, in addition to the speaker's microphone, the chairperson's microphone is to be independent, it should be possible to switch it ON at any moment.

A button should be available to the chairperson, which when pressed switches OFF the other microphones.

Distant sound sources and distant microphones shall be deemed to be microphones in the microphone management system and may also be controlled by an external management system.

## 4.3 Conference system

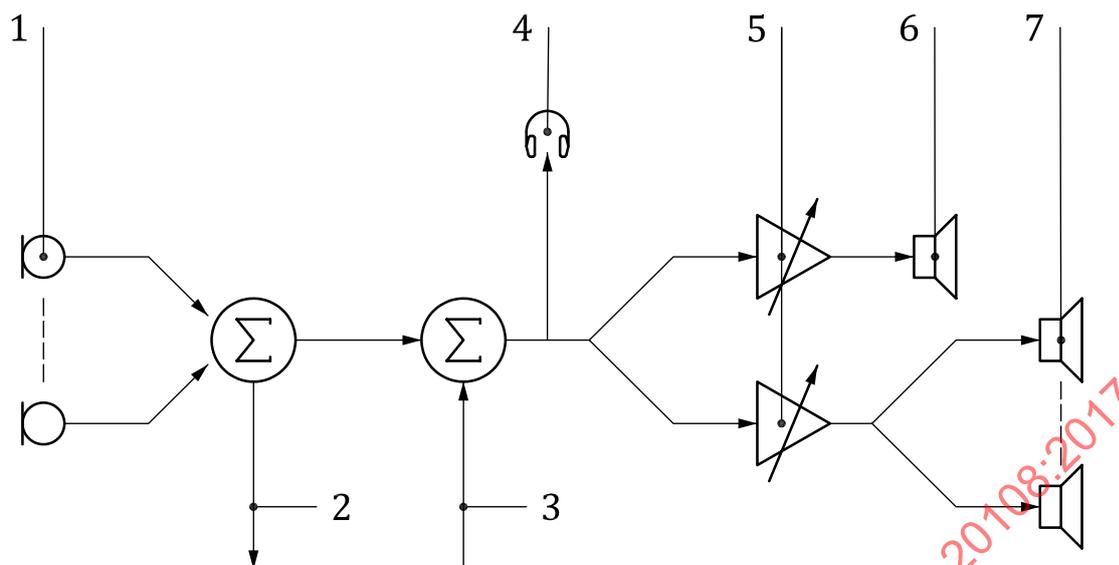
### 4.3.1 General

In a conference room with more than one microphone, the conference system is used to generate the original/floor sound input both for the interpreters and for the speech reinforcement system.

### 4.3.2 Sound input and output

#### 4.3.2.1 Overview

The conference system shall have at least one auxiliary input and one original/floor output (see [Figure 1](#)), which allows a setup at least in the modes set out in [4.3.2.2](#) and [4.3.2.3](#).



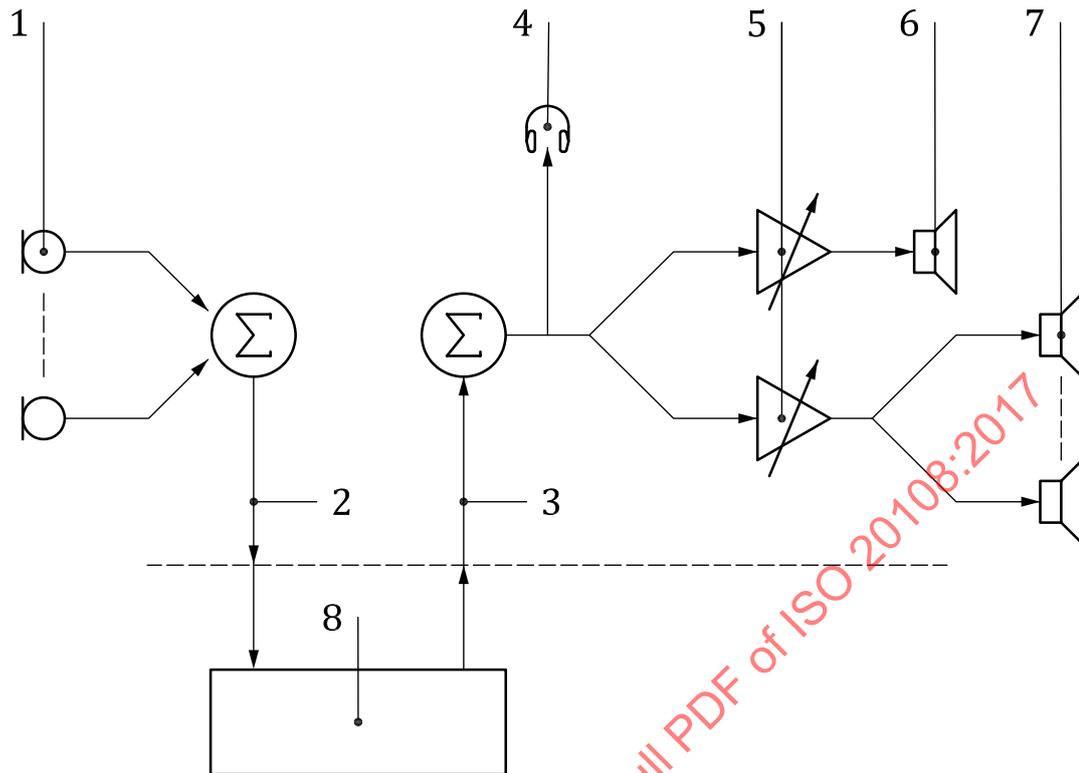
**Key**

- 1 conference system microphones
- 2 dry floor output
- 3 auxiliary input
- 4 processed floor output
- 5 volume controls
- 6 speech reinforcement system
- 7 discussion unit speakers

**Figure 1 — Standard mode**

**4.3.2.2 Insertion mode**

In insertion mode, the normal audio path of the summed output of the conference system is interrupted to allow an audio processing device to be connected. The external audio device is connected between the original/floor output and the auxiliary input (see [Figure 2](#)).

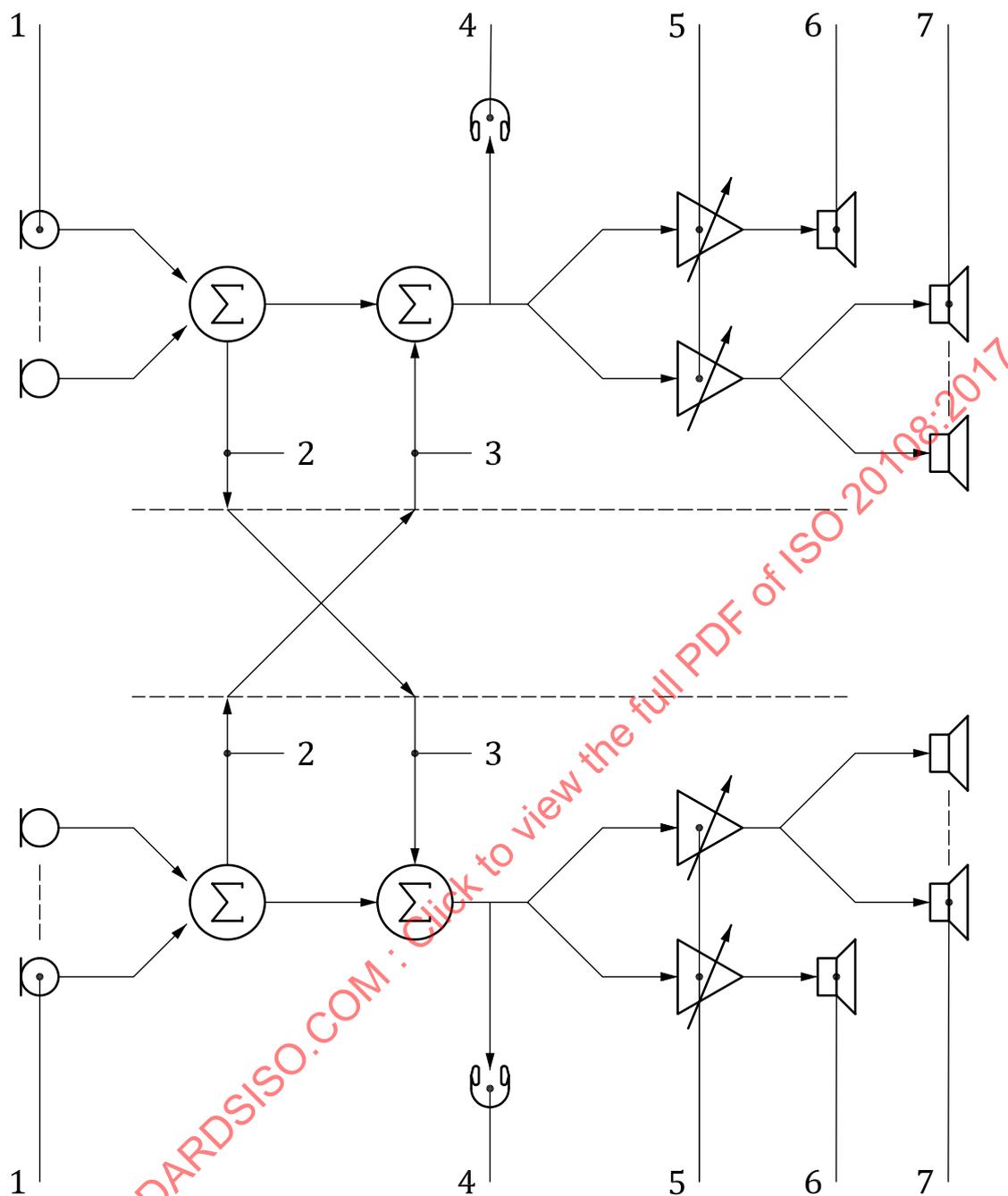
**Key**

- 1 conference system microphones
- 2 dry floor output
- 3 auxiliary input
- 4 processed floor output
- 5 volume controls
- 6 speech reinforcement system
- 7 discussion unit speakers
- 8 audio processor

**Figure 2 — Insertion mode****4.3.2.3 Mix-minus mode**

The main feature of this mode is the fact that sound sources connected to the auxiliary input of the conference system are not routed to the original/floor output of the conference system.

This is desired to avoid echo and feedback in applications where a distant audio connection is made with another conference system in order to connect distant delegates. The result is a considerable improvement in the sound performance of the connection. See [Figure 3](#) for an example of what the audio routing looks like if the two systems (local and distant) are connected.



**Key**

- 1 conference system microphones
- 2 dry floor output
- 3 auxiliary input
- 4 processed floor output
- 5 volume controls
- 6 speech reinforcement system
- 7 discussion unit speakers

**Figure 3 — Mix-minus mode**

#### 4.4 Intelligibility

The number of microphones switched ON in a conference setting with provision of simultaneous interpretation shall always be kept to a minimum. Systems shall include measures to prevent acoustic feedback and acoustic shock and to reduce sound degradation caused by too many microphones switched ON.

Systems shall comply with a speech transmission index (STI) of no less than 0,64 in accordance with IEC 60268-16.

### 5 Image input to interpreters

#### 5.1 General

Requirements regarding the need for video displays inside interpreting booths, and their number, positioning and dimensions are set out in ISO 20109.

Video displays may be placed in front of the booths.

See [Annex B](#) for information regarding image content.

#### 5.2 Image quality

The quality of image input to the interpreters shall be kept on a level that prevents visible artefacts such as blurring or freezing.

In case of excessive blurring, freezing or artefacts, interpreting may have to be temporarily suspended until the image stabilizes.

#### 5.3 Image projection

The size of the projection screen in the conference room will depend on the viewing distance. Video displays shall be used (see ISO 20109:2016, B.2)

- if the angle of the interpreters' line of vision towards a screen is less than 35°, taking the edge of the booth as a reference,
- if the distance between the booths and the screen is  $\geq 3$  times the screen's diagonal, or
- if the booths are located behind the main speakers or on an upper level (in accordance with ISO 2603:2016, 4.5).

The projection of the direct signal of images of a speaker present in the room shall not lag behind that of the sound by more than 120 ms.

### 6 Lip sync

Sound and image from a distant site shall be synchronized.

This means that sound shall neither lag image by more than 45 ms nor shall it lead it by more than 125 ms according to ITU-R BT.1359.

## 7 Transmission of audio and video from a distant site

### 7.1 Quality

The quality of the transmission of sound and image input to the interpreters shall be kept on a level that prevents audible and visible artefacts due to the effects of both packet level and signal-related impairments caused by coding processes.

Room acoustics facilitating good speech intelligibility are described in [Annex A](#).

The measurement of audio and video output quality is described in [Annex C](#).

### 7.2 Latency

The propagation delay of the image and sound from the source to the interpreters shall not exceed 500 ms.

STANDARDSISO.COM : Click to view the full PDF of ISO 20108:2017

## **Annex A** (informative)

### **Room acoustics**

In order to facilitate good speech intelligibility, any room from which a speaker takes the floor should not cause reverberation or echoes exceeding or falling below the values recommended for the respective type of room according to AS/NZS 2107.

AS/NZS 2107 also specifies methods of measuring the background sound level in unoccupied spaces and the reverberation time in unoccupied spaces in buildings.

STANDARDSISO.COM : Click to view the full PDF of ISO 20108:2017

## Annex B (informative)

### Image content

- If a speaker present in the room is displayed, or where a point-to-point videoconference is in use, that speaker (and any visual materials) should be displayed using the direct signal of the room on the entire screen with full resolution and the correct aspect ratio, avoiding any resizing, resampling or downscaling and for the entire duration of the speech. The multipoint picture should not be used, especially on the video displays in the booths. Visual materials should be displayed to the interpreters on a separate screen as a continuous uninterrupted video feed in both point-to-point and multipoint videoconferences. If separate screens are not available, priority should be given to the visual materials.
- In the case of multipoint videoconferences, the active speaker should be displayed on at least 65 % of the screen with the correct aspect ratio. Other participants may be displayed on the remaining part of the screen.
- Additionally, an overview image of the entire meeting or conference room may be shown in such a way that all participants can be seen clearly.
- The speakers should be viewed from directly in front or at an angle of
  - $\pm 35^\circ$  on a horizontal plane (except for camera positions designed to view the chairperson:  $\pm 5^\circ$ ), and
  - $+13^\circ/-25^\circ$  on a vertical plane (except for camera positions designed to view the chairperson:  $\pm 5^\circ$ ).
- Any available camera may be used to film members of the audience not speaking.
- The camera should zoom in so that the speaker is visible from the waist up and all lip and hand movements are also visible.
- The speaker should be lit sufficiently, using indirect diffused light of one kind, thereby avoiding the creation of hard shadows around him/her. Colour temperature should be between 3 000 K and 3 800 K; if a significant quantity of outdoor light enters the room, colour temperature should not exceed 5 500 K.
- It may be necessary to take measures to prevent outside light from entering the room. Speakers should be advised to wear neutral colours (no stripes or flashy/contrasting colours). To avoid colour casting, white should be avoided. Reflecting surfaces should also be avoided.
- In order to avoid artefacts in the image, speakers should be discouraged from making extravagant gestures.
- If multiple speakers are present at a distant site, the camera(s) should automatically pan, tilt and zoom onto the active speaker. If this is not possible, camera positions should be saved in advance. A camera panning, tilting and zooming into a new speaker should stabilize in no more than 4 s.