INTERPRETED VIDEOCONFERENCE
MINIMUM AUDIOVISUAL TECHNICAL REQUIREMENTS
(updated oct 2010)

The following requirements, relating to various aspects of a videoconferencing link between different locations, are to be considered as the minimum to ensure a level of quality suitable for simultaneous interpretation.

They are based on extensive benchmarking tests carried out by DG Interpretation between July 2005 and December 2007 and on a study by the Fraunhofer Institute to define an objective evaluation method for assessing the minimum quality of digital video and audio sources (as used inter alia in videoconferences) required for simultaneous interpretation.

The requirements have been defined for a standard conference situation. In this scenario,

- Interpreters are sitting in the main location (in ISO2603 interpretation booths) together with the chairman of the meeting.
- One or more delegates is participating in a multilingual interactive communication process from a distant location linked to the main location via a videoconferencing system. Images from the remote location are visible to both participants and interpreters sitting in the main location. Remote original audio is mixed with local original audio. Remote participants receive images from the main venue.
- Remote and local participants receive interpretation from the same source (main venue).

The indicative values must be seen in the context of the entire audiovisual chain. They are not the minimum requirements for individual components or for any subset of the audiovisual transmission chain.

The values refer to the audiovisual quality as perceived in interpreters' booths.

Some values are related specifically to the equipment and brands used for the assessments. They may differ slightly with newer or different equipment.

J. ESTEBAN CAUSO,
Head of Unit

Items	Requirements
Image display (projection in the meeting room for both participants and interpreters)	1024x768 for 4:3, 1280x720 for 16:9. with a high luminosity/contrast projector, a display of 210x290 m at a maximum distance of booths from display of about 10 m. if a talking head or bust view is displayed. For a wider view of the room taking in several participants, a higher resolution chain (HD) is required. EN 12464-1 defines a minimum illuminance of 500 Lx for conference sites. If the conference room is lit by daylight, the actual illuminance may be much higher.
Image display (in the interpretation booth if appropriate) .	1024x768 for 4:3, 1280x720 for 16:9. with a high luminosity/contrast LCD display. This is applicable to talking head or bust views. For a wider view of the room, taking in several participants, a higher resolution chain (HD) is required. The number of screens and display arrangements (split screen, picture in picture, etc.) shall be adapted to suit the specific situation)
Image frame rate	≥ 25 fps
Echo canceller operational interval	Audio management system with "soft" set up (freq. response +0/-0.4dB 20Hz-20000Hz, sampling rate 48Hz, Dynamic range 20Hz-20000Hz, 0db gain: 107dB, A/D converter 24-bit)
A/V synchronization	Audio should not be in advance of video by more than 25 ms; video should not be in advance of audio by more than 95 ms (ITU-R BS.1359), adjusted on the display/earphones combination used by interpreters Projectors may introduce a considerable video delay; this should be compensated by adding audio delay (in interpreters' earphones).
Codec with MPEG-2 compression Video speed compression Framing Video GOP size Audio codec Encoding speed Sampling	SD (PAL) resolution 2500 kbps IBP about 50 Layer 2 256 kbps 48Hz
Codec with MPEG-4 compression Video speed compression Framing Video GOP size Audio codec Encoding speed Sampling	SD (PAL) resolution 1500 kbps (or less, depending on codec implementation) IBP about 30 AAC 128 kbps 48Hz
Microphone	Compliant with latest version of IEC 60914 standard
Communications	Virtually error free and jitter free network communication (QoS necessary if shared traffic with limited bandwidth).
Camera	3.5-inch CCDs, 0.75 lux low light, 800 lines resolution) Optics (indicative): focal length 7.3 to over 100mm (14x), Maximum Aperture f/1.9 (to 69.7mm), f/2.1 (at 100mm), horizontal Field of View: 7.3mm(47° 20), 100mm (3° 36) and Filter 82mm P = 0.7
Lighting	Should be appropriate to avoid shadow effect and allow a clear perception of facial expressions and body language.