TECHNICAL GUIDELINES FOR REMOTE INTERPRETATION
(update March 2011)

APPLICABLE STANDARDS

The booths used, whether fixed or mobile, shall be compliant with ISO standards 2603-1998 and 4043-1998 respectively. When mobile booths are used, they shall be installed in such a way as to replicate the interpreters' normal working environment. The audio circuit must be consistent with the standards ISO 2603, ISO 4043 and CEI 60914.
Details of the video circuit and the related aspects are set out below.

TECHNICAL SET UP

The technical set up shall be designed according to the type of meeting and the characteristics of the meeting rooms. There are two situations to consider: the place where sound and images are captured (source room) and the place where interpretation is provided (target room)

SOURCE ROOM

Lighting

The room must be adequately lit to allow the shooting of clear, high-quality images, well-defined colours and sufficient depth of field for each type of shot. If the level of lighting is not sufficient, additional lighting shall be provided to avoid dark areas and excessive contrasts, ensuring that images are not shot against the light .

Cameras

The cameras must provide images of all the participants. They should be displayed for interpreters in such a way as to be relevant and spatially coherent with the sequence of interventions in the source room (for example: if A is talking to B, and A and B are not displayed on the same screen, images should be presented in accordance to their positioning in the room; B on the left side if A , when talking to B, is turning to the left..

The number and position of the cameras shall be determined by the type of situation, the number of participants and the size and shape of the source room so that the following shots can be provided:

- Total and/or recomposed partial views of the meeting room as a whole so that the position of speakers can be determined
- View of the speaker (close-up)
- View of the chairman/Judge
- Other relevant views, such as:
- a face-on foreground shot of each participant who is likely to speak or to be referred to by the speakers, or
- the next registered speaker or
- the previous speaker or
- a part of the room where a relevant event is occurring, or
- the participants listening to the interpretation

All images should have at least HD 720 quality and preferably HD 1080 when a larger audience than a couple of people is displayed.

TARGET ROOM

Lighting

Lighting in the target room must offer a good degree of visual comfort and be adjustable.
The booths shall be equipped with continuously adjustable individual lights, in accordance with the ISO standards for office spaces.

Screens for the interpreters

At least
- one 22-24" screen per interpreter, placed directly in front of the booth showing the active speaker. However, if the booth is equipped with a display, this may be used to display the active speaker
- one 22-24" screen placed directly in front of the booth , displaying the chairman's dais
- one over-50" screen providing an overall view of the room or more (smaller) screens providing partial views of the room (taking into account the spatial coherence).

The above set up is just indicative. For each specific situation a careful analysis should be made to define which images are relevant for the interpreters, and how the screens are laid out in the most appropriate way.
The distance between the screen and the interpreters eye will be determined by the size of the screens : between +/- 1 m and 3.5 m. If mobile booths are used, the position of the screens shall ensure that the view is not obstructed by the window frames. All screens should have a size ratio of 16:9 with a resolution $\geq$ 1920 x 1080 pixels.
Any visual pollution caused by screens intended for an adjacent booth shall be avoided.
The individual screens per interpreter may receive either images from the meeting room or presentations (PowerPoint, texts, overhead projection or videoconference) and show them at the maximum resolution.

Image quality according to SSIM (Structural Similarity) index should result in values between 100 and 0.9 (perfect or nearly perfect image) when comparing

the original broadcast-quality image and the image displayed on the interpreters display.

Sound-Image Transmission and Distribution

No perceivable quality loss is acceptable when transmitting and distributing images and sound Processing of the video signal shall be kept to the strict minimum required to minimize delays and any deterioration in the image quality, such as a loss of clarity or the appearance of artefacts.

The signal transmission delay between the source room and the target room must not exceed 300 milliseconds.

Audio quality, according to ITU-R BS-1116-1 [BS.1116] should have a worst-case perceived quality not less than 3.5 (which translates into an impairment being in between 'perceptible' and 'slightly annoying').

Lip Synchronization

With reference to Recommendation ITU-R BT.1359-1 "RELATIVE TIMING OF SOUND AND VISION FOR BROADCASTING" the relative timing must be kept within the "Undetectability plateau" (figure 2) : audio should not advance video for more than 22 ms and video should not advance audio for more than 95 ms (-95 ÷ + 22 ms range).

As projectors, displays, the videoswitch matrix, videoprocessors and, in general, the image delivery chain may introduce a considerable video delay, the incoming interpreters' audio has to be adjusted accordingly.
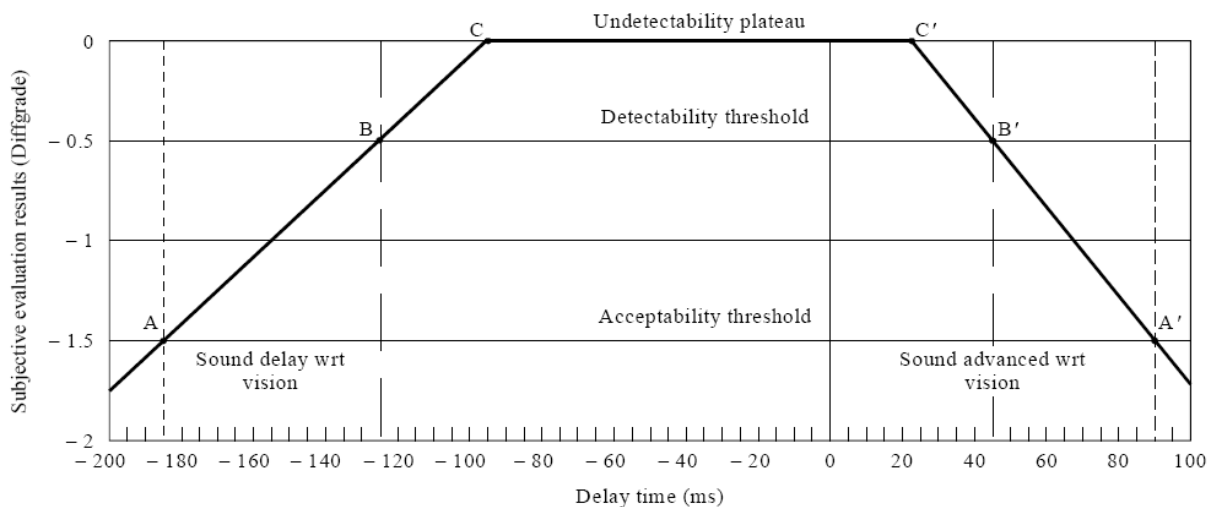
1359-02

Undetectability plateau [Rec. ITU-R BT.1359]

José Esteban Causo
Head of unit