

**ATSC Implementation Subcommittee Finding:
Relative Timing of Sound and Vision for Broadcast
Operations**

The Advanced Television Systems Committee, Inc., is an international, non-profit organization developing voluntary standards for digital television. The ATSC member organizations represent the broadcast, broadcast equipment, motion picture, consumer electronics, computer, cable, satellite, and semiconductor industries.

Specifically, ATSC is working to coordinate television standards among different communications media focusing on digital television, interactive systems, and broadband multimedia communications. ATSC is also developing digital television implementation strategies and presenting educational seminars on the ATSC standards.

ATSC was formed in 1982 by the member organizations of the Joint Committee on InterSociety Coordination (JCIC): the Electronic Industries Association (EIA), the Institute of Electrical and Electronic Engineers (IEEE), the National Association of Broadcasters (NAB), the National Cable Television Association (NCTA), and the Society of Motion Picture and Television Engineers (SMPTE). Currently, there are approximately 160 members representing the broadcast, broadcast equipment, motion picture, consumer electronics, computer, cable, satellite, and semiconductor industries.

ATSC Digital TV Standards include digital high definition television (HDTV), standard definition television (SDTV), data broadcasting, multichannel surround-sound audio, and satellite direct-to-home broadcasting.

ATSC Implementation Subcommittee Finding: Relative Timing of Sound and Vision for Broadcast Operations

The end-to-end DTV audio-video production, distribution, and broadcast system is a complex array of digital processing, compression, decompression, and storage devices. Each component in the system imposes a latency on the audio and/or video signals flowing through it. System design goals often call for the relative audio-video latency through each component to be in the sub-millisecond range. Operationally, unequal delays can be imposed on the audio and video signals respectively, and these delays compromise audio-video synchronization.

One of the overarching goals of the DTV broadcasting system is to deliver audio and video in proper synchronization to the viewer. Because each digital audio and video component in the chain from production to reception imposes some degree of latency on the signals passing through it, and the delays imposed on the audio and video signals are typically unequal, each component harbors the potential to cause an audio-video synchronization error at its output. The overall audio-video synchronization error is the algebraic sum of the individual synchronization errors encountered in the chain. While a given synchronization error may cause either a positive or negative differential shift in audio-video timing, the video signal is typically subjected to greater delay than the audio signal, and the tendency is therefore toward video lagging behind audio. Thus, there is a requirement to monitor audio-video synchronization at various points within in the system and to make corrections, where required, in order to deliver to the viewer audio-video synchronization within the required tolerance.

In addition, there are points within the end-to-end chain that require A/V synchronization to be maintained, such as switching points, monitoring points, and transmission/encoding points.

For the purposes of this discussion, the end-to-end DTV system may be divided into four segments: acquisition and production/post production (contribution system), release facility and distribution system, local broadcast station, and home receiver. IS finds that steps must be taken to ensure that the audio and video signals delivered at the output stage of each of the four segments are synchronized within a tight tolerance (see below).

At the production-post production stage, audio-video synchronization errors can occur in the capture stage, in film-to-video transfer, and in editing. The product may be delivered on video tape or by various electronic means, but whatever the delivery medium, IS finds that steps must be taken to ensure that audio-video synchronization in the delivered product falls within the required tolerance.

The release facility segment contains a number of devices through which the DTV audio and video signals are passed, which variously impose compression and de-compression, processing, and storage and their attendant differential delays on the signals. The process of distributing the signals to affiliate stations typically requires compression and decompression steps. IS finds that it is incumbent on the release facility to correct the differential audio-video delays that the signals experience within the plant so that the initial timing relationship is restored to a tight tolerance before the signals reach the distribution encoder. IS also finds that synchronization to a tight tolerance should be maintained in any encode/decode process that is involved in delivering the signals to the affiliate station, so that the tight A/V synchronization can be monitored at switching and other points.

The affiliate station segment contains a number of devices that are similar to those encountered in the release facility segment and that generate the same types of differential audio-video delays, including switching and monitoring points. IS finds that audio-video synchronization should be restored to a tight tolerance before the signals are input to the broadcast station's ATSC audio and video encoding devices to assure that the presentation

time stamps placed on the audio and video access units by the encoder faithfully represent correct synchronization.

IS finds that under all operational situations, at the inputs to the DTV encoding devices, the sound program should be tightly synchronized to the video program. The sound program should never lead the video program by more than 15 milliseconds, and should never lag the video program by more than 45 milliseconds.

MPEG-2 models the end-to-end delay from an encoder's signal input to a decoder's signal output as constant. This end-to-end delay is the sum of the delays from encoding, encoder buffering, multiplexing, transmission, demultiplexing, decoder buffering, decoding, and presentation. Presentation time stamps are required in the MPEG bit stream at intervals not exceeding 700 milliseconds. The MPEG System Target Decoder model allows a maximum decoder buffer delay of one second. Audio and video presentation units that represent sound and pictures that are to be presented simultaneously may be separated in time within the transport stream by as much as one second. In order to produce synchronized output, IS finds that the receiver must recover the encoder's System Time Clock (STC) and use the Presentation Time Stamps (PTS) to present the audio-video content to the viewer with a tolerance of +/-15 milliseconds of the time indicated by PTS.

Although real aural and visual presentation devices typically have finite and different inherent delays, and may have additional delays imposed by post-processing or output functions, the System Target Decoder models these delays as zero. IS finds that such delays must be corrected before the audio and video signals are presented to the viewer.

The IS has undertaken additional work to develop tolerances for system design. Pending that finding, designers should strive for zero differential offset throughout the system.

Note: ITU R BT.1359-1 (1998) was carefully considered and found inadequate for purposes of audio and video synchronization for DTV broadcasting.