Audio/Video Synchronization Standards and Solutions A Status Report

Patrick Waddell/Graham Jones/Adam Goldberg

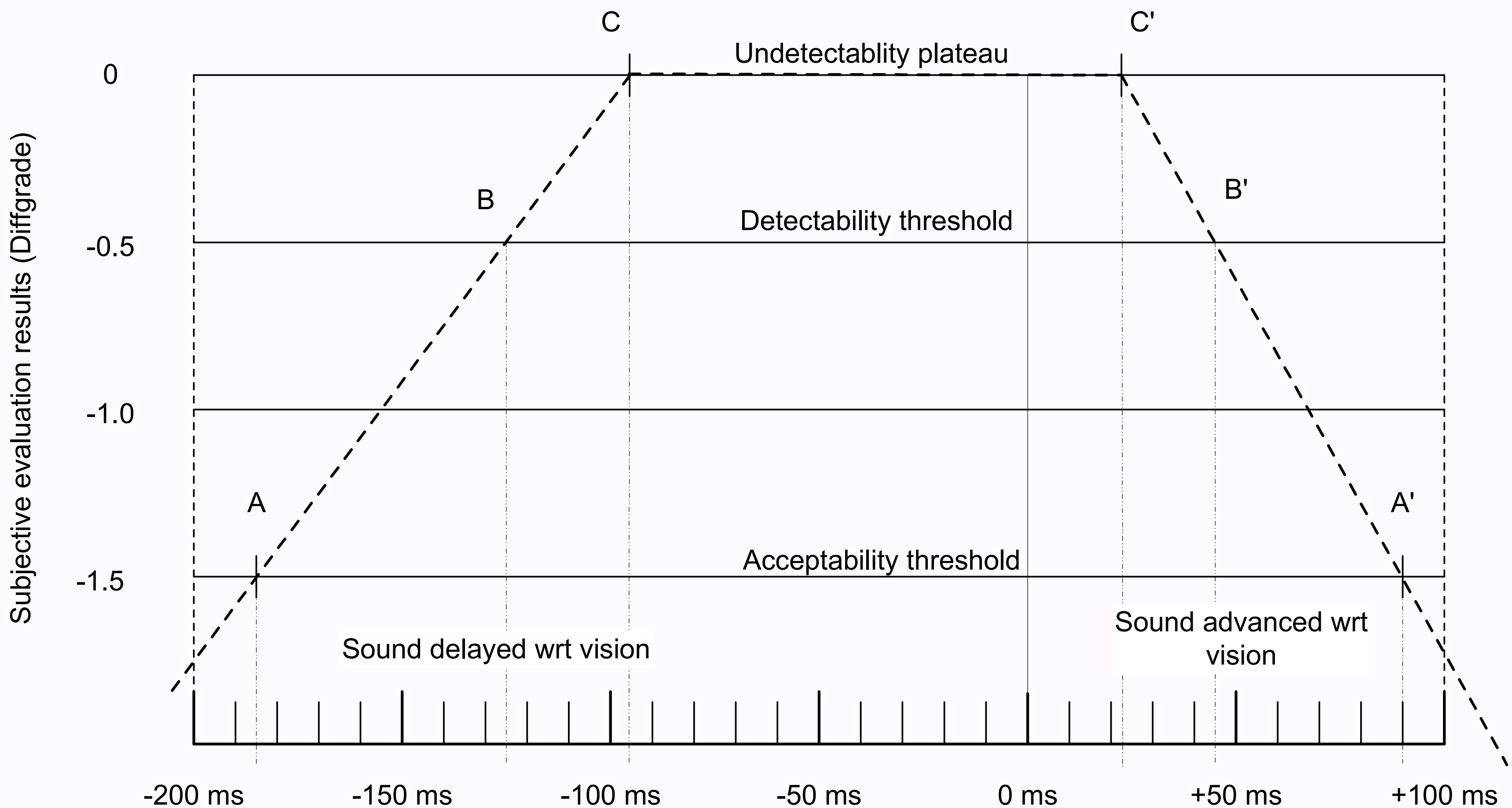
Advanced Television Systems Committee

ITU-R BT.1359-1 (1998)

Only International Standard on A/V Sync

- Subjective study with *EXPERT* viewers
  - SDTV not HDTV images
  - CRT displays, of course
- At first glance it seems loose: +90 ms to -185 ms as a "Window of Acceptability"
  - In their terms, positive values are audio advanced relative to video, negative is delayed relative to video
  - We will examine these results more closely…
  - The numbers were statistically significant for each point
- Remember, the measurements were *very* carefully made
  - Expert viewers
  - 20" CRT monitors
  - fixed viewing distances

ITU-R BT.1359 Figure 2



ITU-R BT.1359 Figure 2

ITU-R BT.1359 Figure 2

- Let’s quickly look at Figure 2 versus Fixed Pixel Display rates
  - 30/1.001 Hz (or 33.3 ms per image)
  - 25 Hz (or 40 ms per image)
- This may be informative…

Figure 2 with Fixed Pixel Display Timings Shown

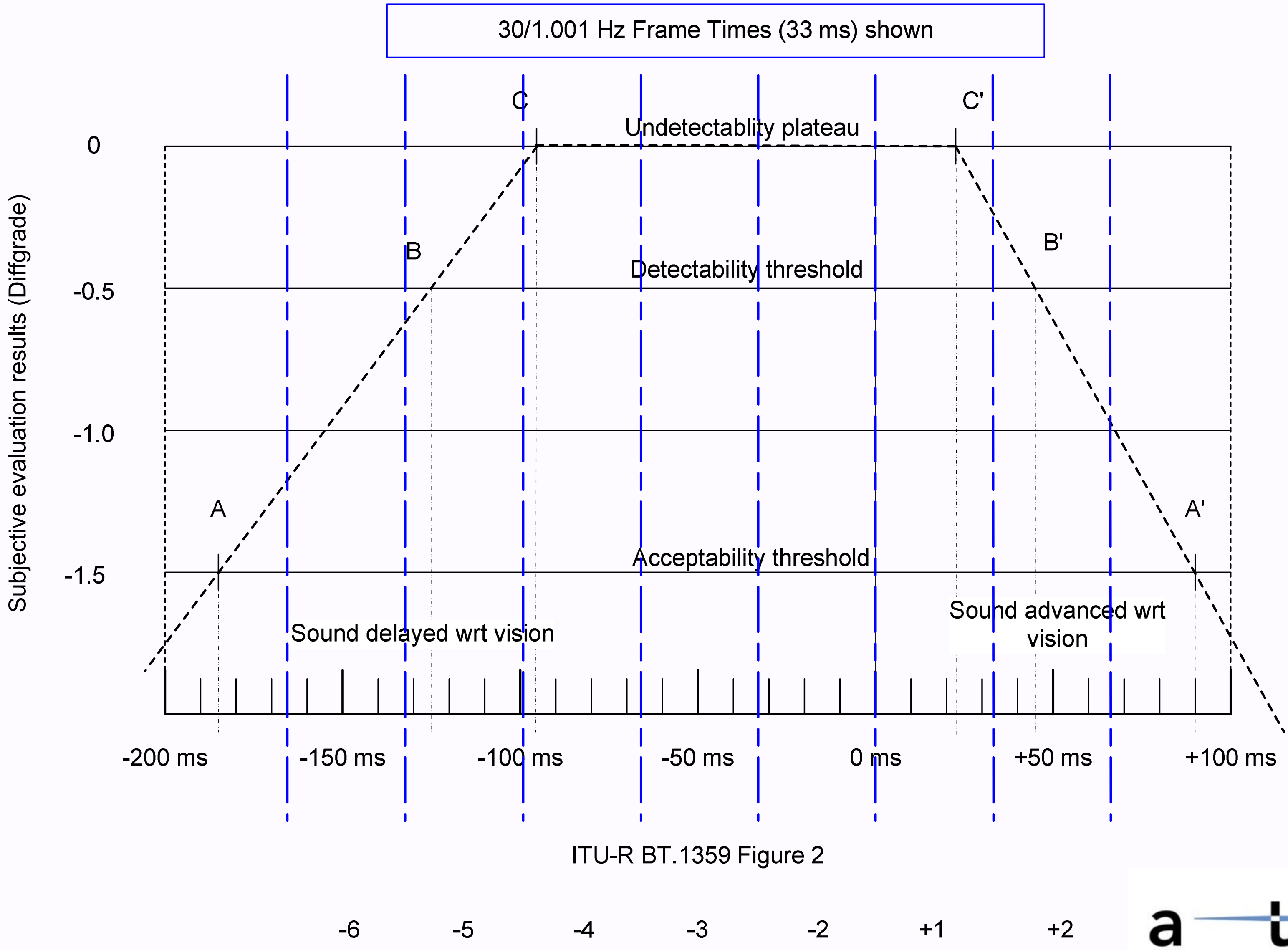
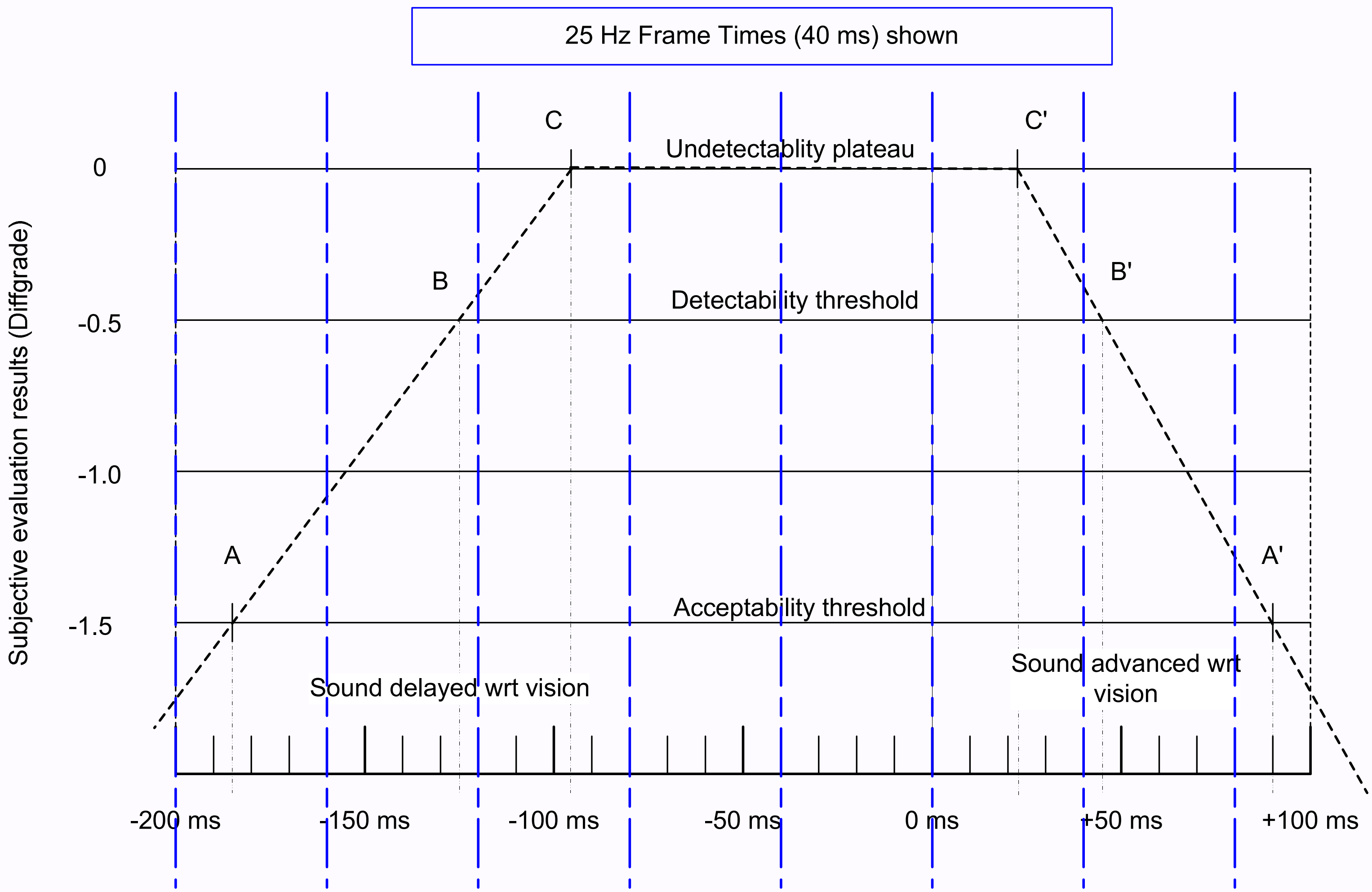


Figure 2 with Fixed Pixel Display Timings Shown

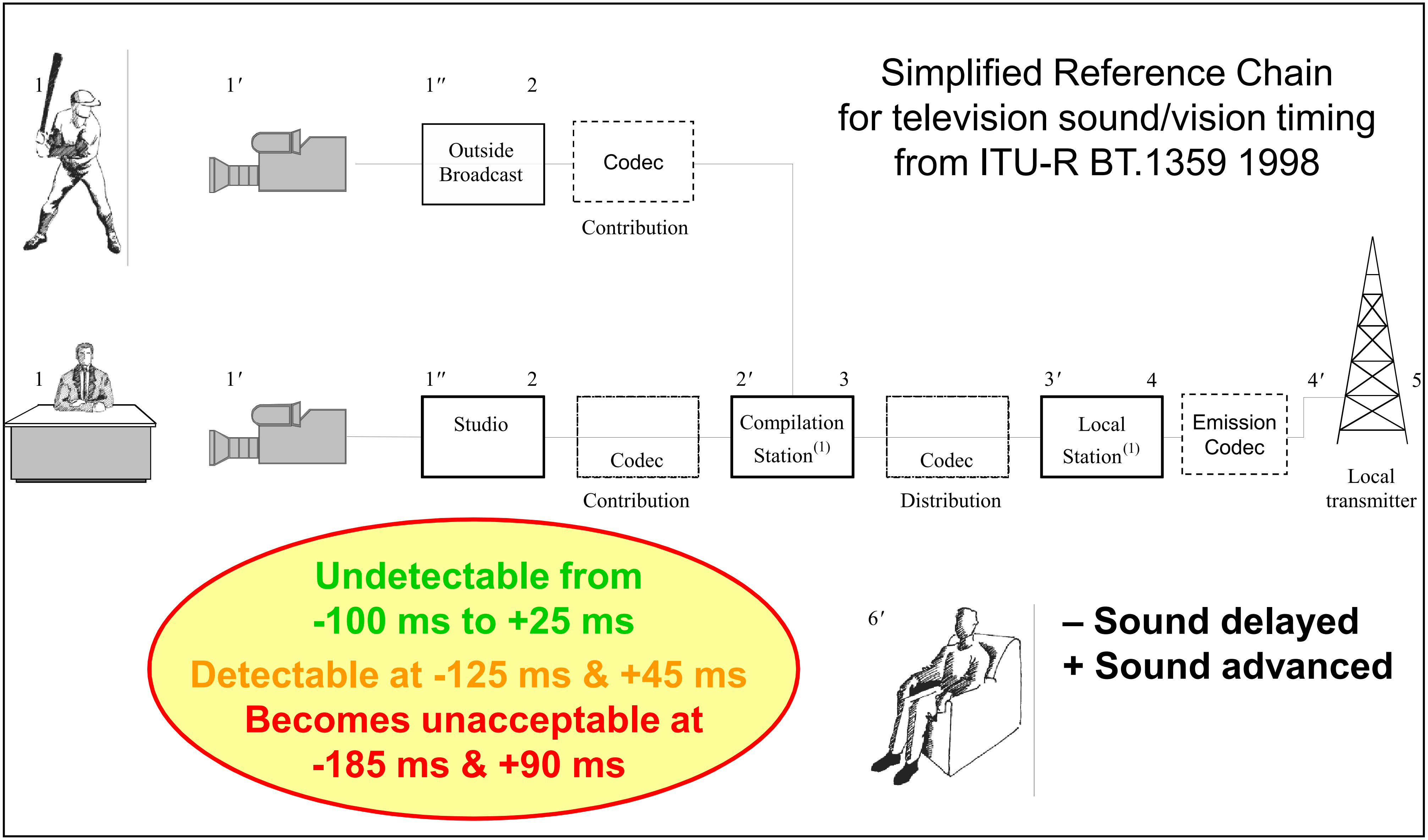


ITU-R BT.1359 Figure 2

Fixed Pixel Display Timings

- Interesting results
- Note that both charts assumed interlaced video
  - So 1080P/60 or 1080P/50 display times are half that shown
- The measured values with CRTs line up fairly well with FPM times for detectability
  - Most of the ITU study measurements were with 25 Hz video (except the Japanese, who used 30 Hz)
- Note that the Acceptance threshold is merely 2 frames advanced for either frame rate!
  - Our brains are used to sound being delayed in nature (by distance)
  - Our brains are confused when sound precedes the vision!

Lip Sync is an End-to-End Issue

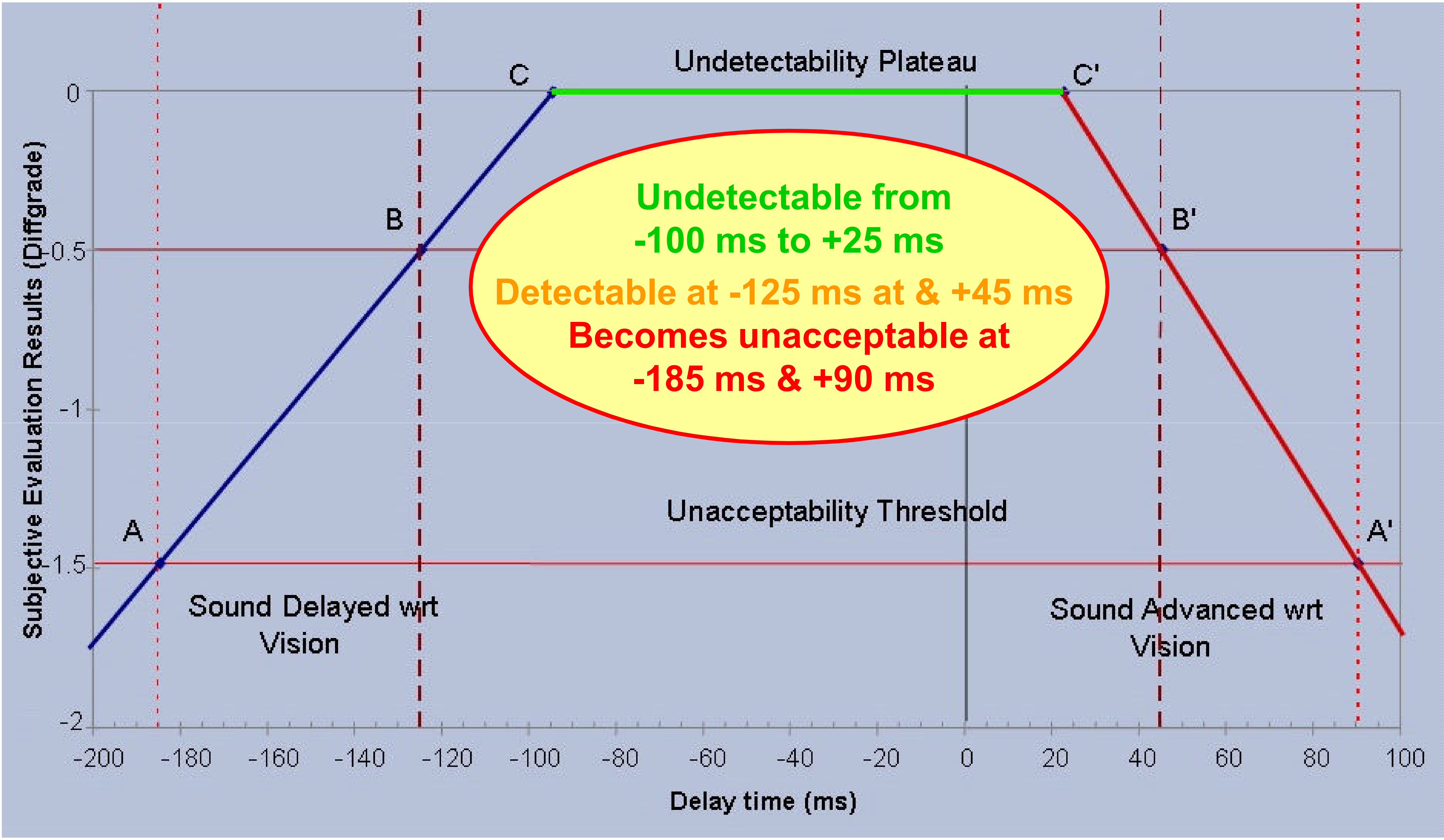


(1)

Subjective Tests

- Subjective tests for the ITU-R BT.1359 standard were carried out in Australia, Japan and Switzerland in 1995 and 1996
  - Used PAL and NTSC video
  - Tube cameras, 22” CRT displays
  - 6x picture height
- New tests carried out this year by JEITA in Japan
  - HD, CCD cameras, large flat panel displays, 3x picture height
  - Results to be published later this year
  - Will possibly show lower threshold levels
  - ITU standard may need to be revised ??

ITU-R BT.1359 Thresholds



Recommended Tolerances

At the input to the transmitter/emission encoder

ITU BT.1359	1998	-30 ms	+22.5 ms
ATSC IS/191	2003	-45 ms	+15 ms
EBU R37	2007	-60 ms	+40 ms
		– Sound delayed	+ Sound advanced

Undetectable from
-100 ms to +25 ms

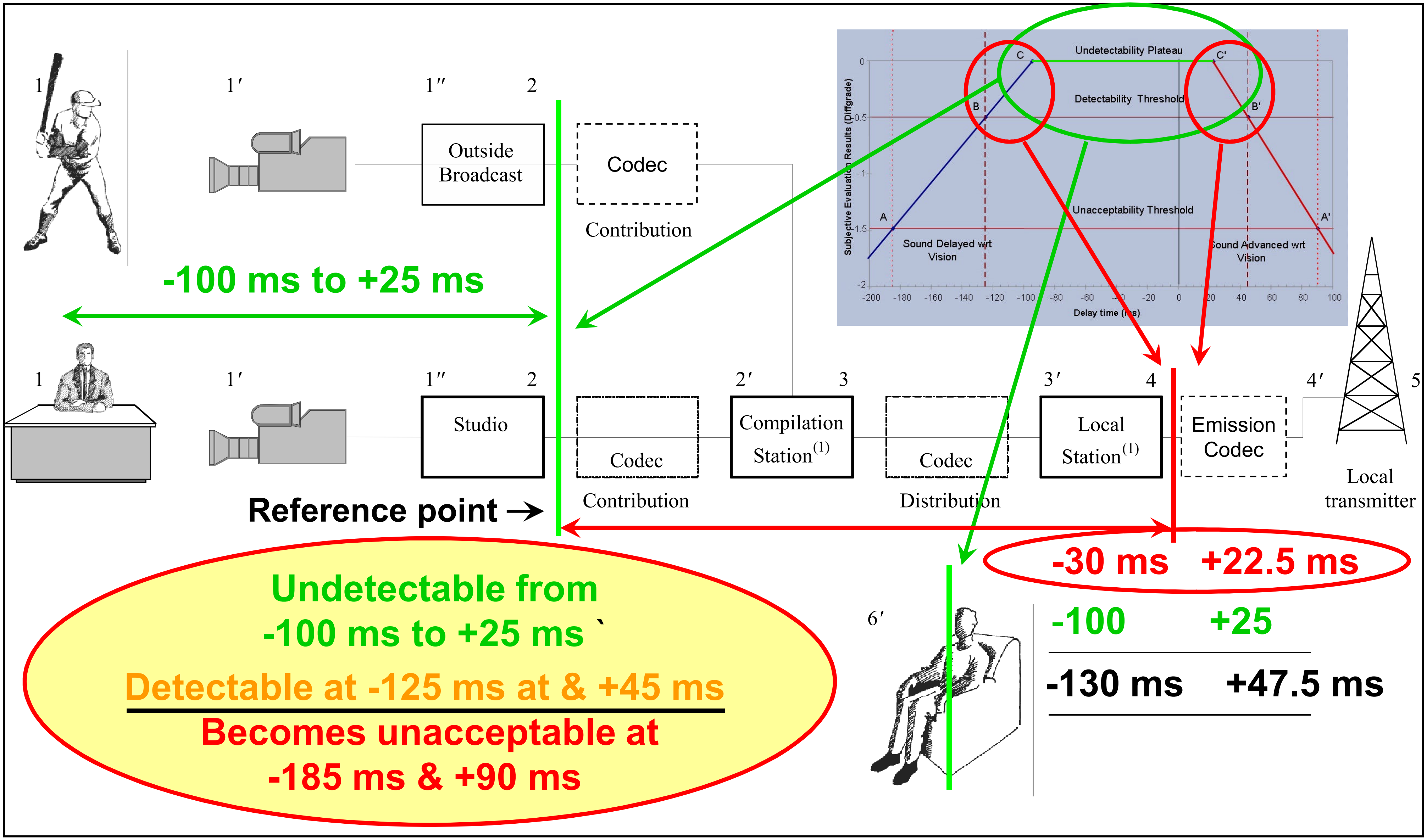
Detectable at -125 ms at & +45 ms

Becomes unacceptable at
-185 ms & +90 ms

ATSC and EBU tolerances are for absolute A/V timing errors

ITU tolerance is for the A/V timing difference in the path from the output of the final program source selection element to the input to the transmitter for emission

Link Budget



(1)

Advanced Television Systems Committee

Broadcaster Tolerance

- Given the level of uncertainty of A/V sync coming out of production and the:
  - Variability of consumer devices
  - Variability in viewing conditions
- In order to have reasonable expectation that viewers will see acceptable lip sync:
  - The broadcaster has no choice but to target a very low or zero error through the chain from reference point to emission encoder
  - There is little or no spare budget to allocate!

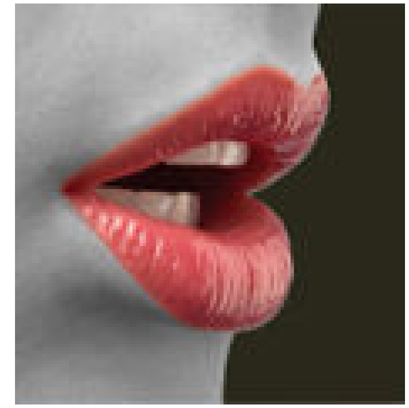
Correct Sync Errors Where they Occur

- Good system design can correct for known and predictable differential delays
  - Solid state cameras
  - Frame synchronizers
  - Vision switchers, format converters, etc.
  - Flat panel monitors with associated audio monitoring
- Fixed and variable delay compensation
  - Available from various manufacturers
  - Control signals from some video devices allow automatic delay switching
  - Care needed to avoid audio artifacts
- Some errors in the chain cannot be predicted or corrected automatically where they occur

Out of Service Measurement

- Clapper board
- Electronic clapper boards
- Beep-flash systems
- Sarnoff Visualizer™

In Service Measurement



- Pixel Instruments LipTracker ™
- Asaca TuLips ™
  - Both use sophisticated analysis of lip movements and associated audio sounds to establish an absolute measurement of sync error at any point in the chain
  - Applicable when moving lips are clearly visible
  - May not be very practical for real world broadcast systems

a t s c
Advanced Television Systems Committee

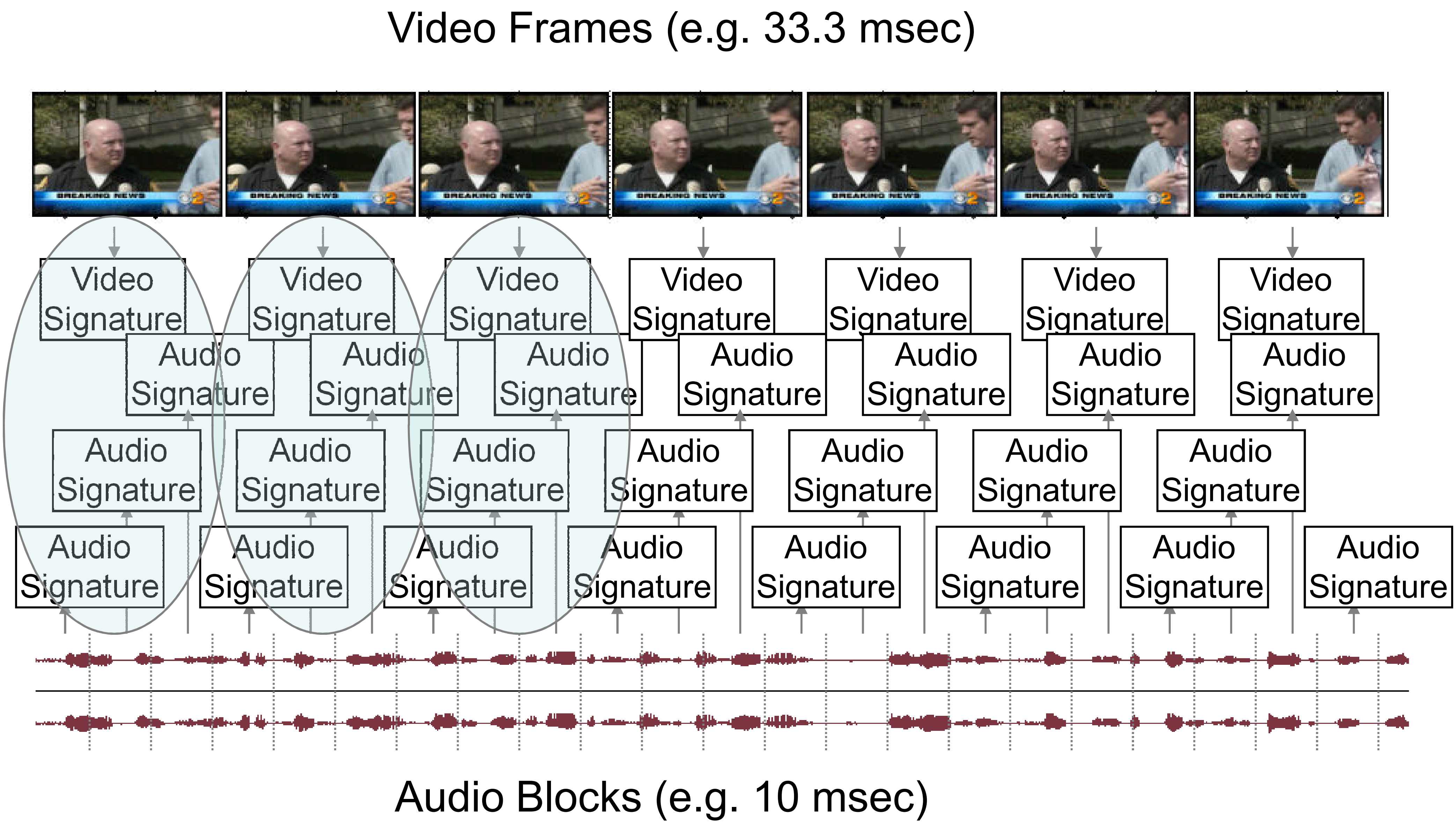
What Is Needed?

- A dynamic in-service method that can respond in near real time
  - Works *while content is playing* - not a calibration method
- Not reliant on any specific signal format or interface so it can be carried through all the different parts of the entire signal chain
  - Particularly needed for the professional parts of the delivery chain
  - Possible application for consumer devices

A/V Signature / Fingerprint / DNA

- Extract features from <u>both</u> audio and video and combine together in an *independent data stream*
- Use <u>fingerprinting</u> methods that are resilient to processing of the audio and video signals
  - Designed to allow typical types of processing (data rate compression, format changes, etc.)
- This data stream may be called an *A/V Sync Signature, Fingerprint, or "DNA"*
  - Relies on generating the signature at a point where A/V sync is *known to be correct*
  - From that point on the system is designed to *measure and maintain* the relative audio/video timing that was present when the signature was generated

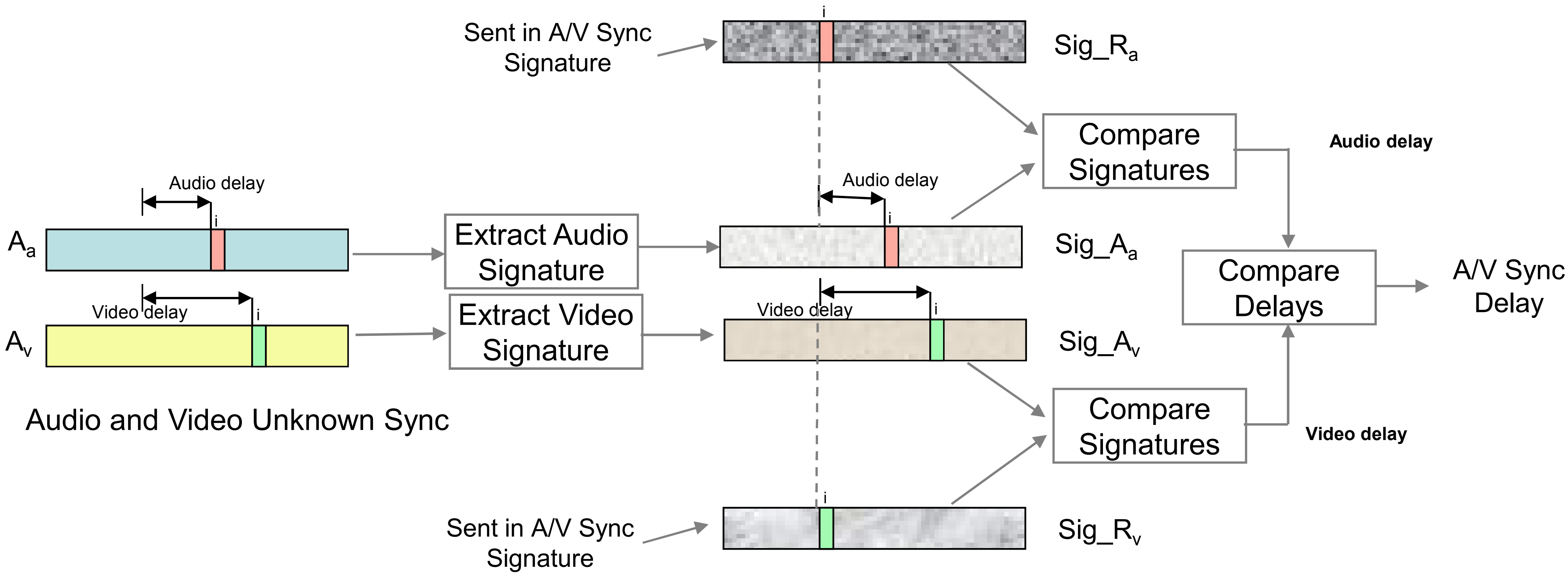
A/V Synchronization Signature



Slide courtesy of Dolby

atsc
Advanced Television Systems Committee

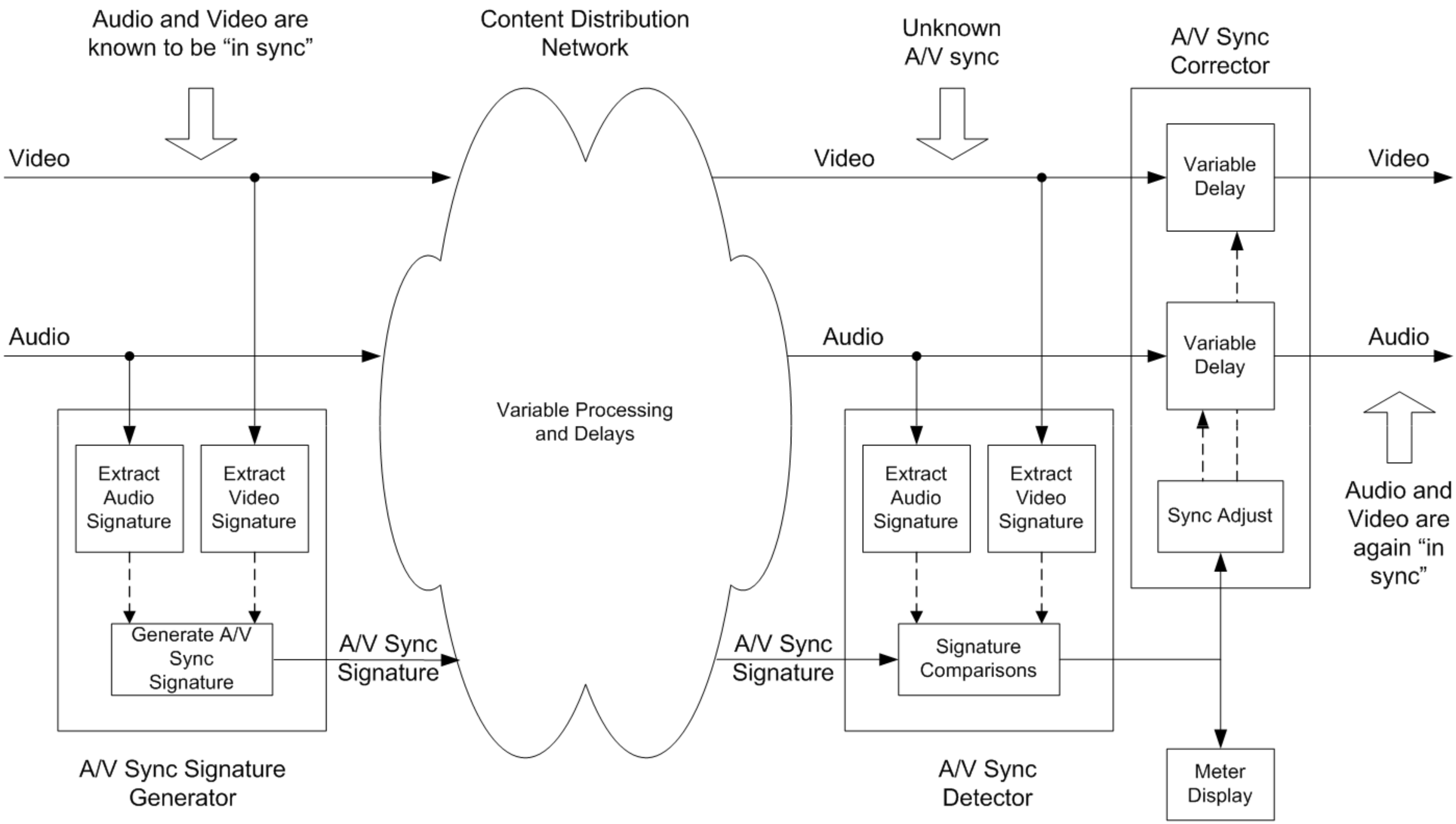
A/V Sync Signature Comparison



- Difference between audio delay and video delay is the A/V sync error

Slide courtesy of Dolby

A/V Sync Correction

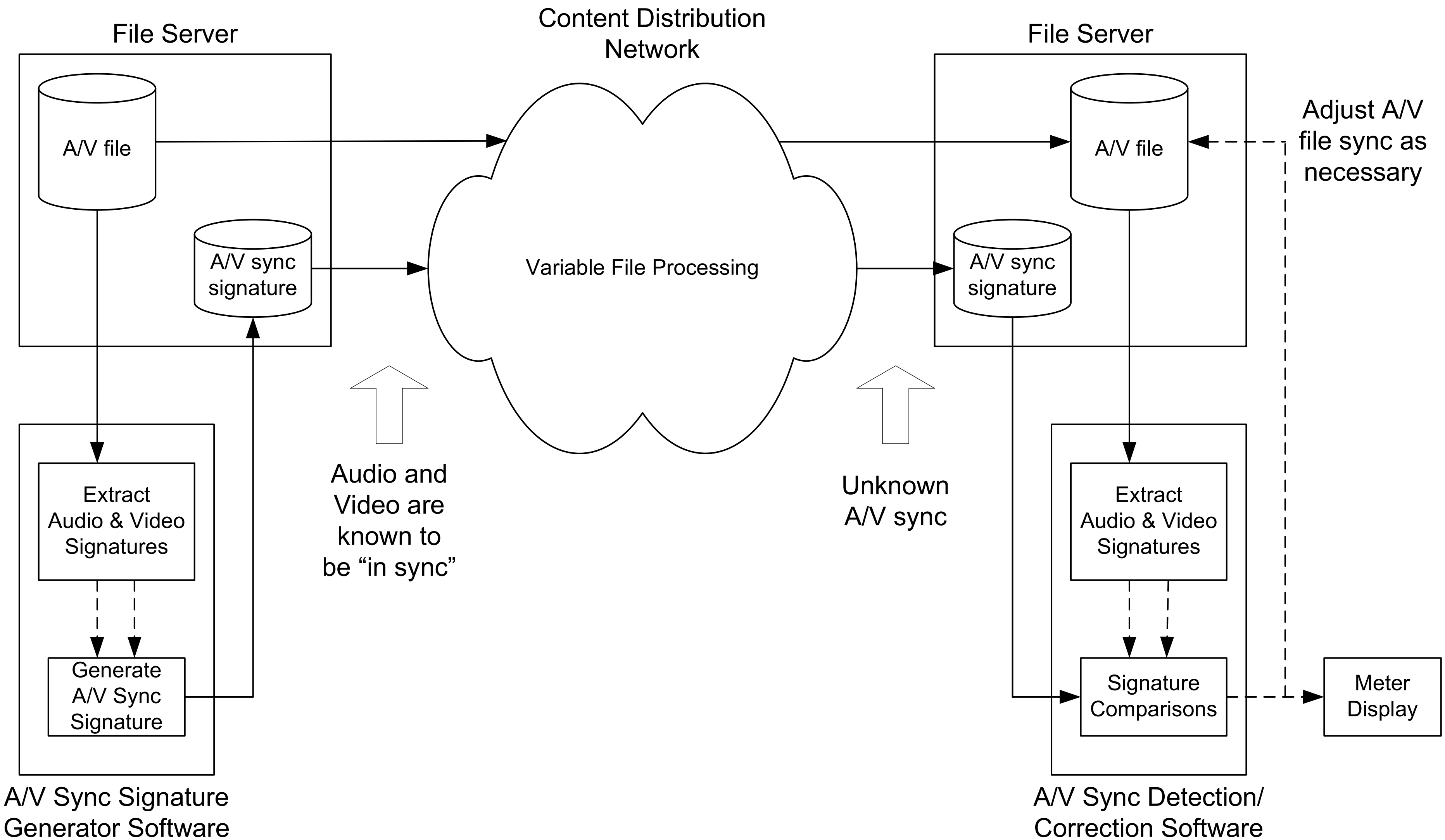


Dolby A/V Signature Real-Time System

Slide courtesy of Dolby

Advanced Television Systems Committee

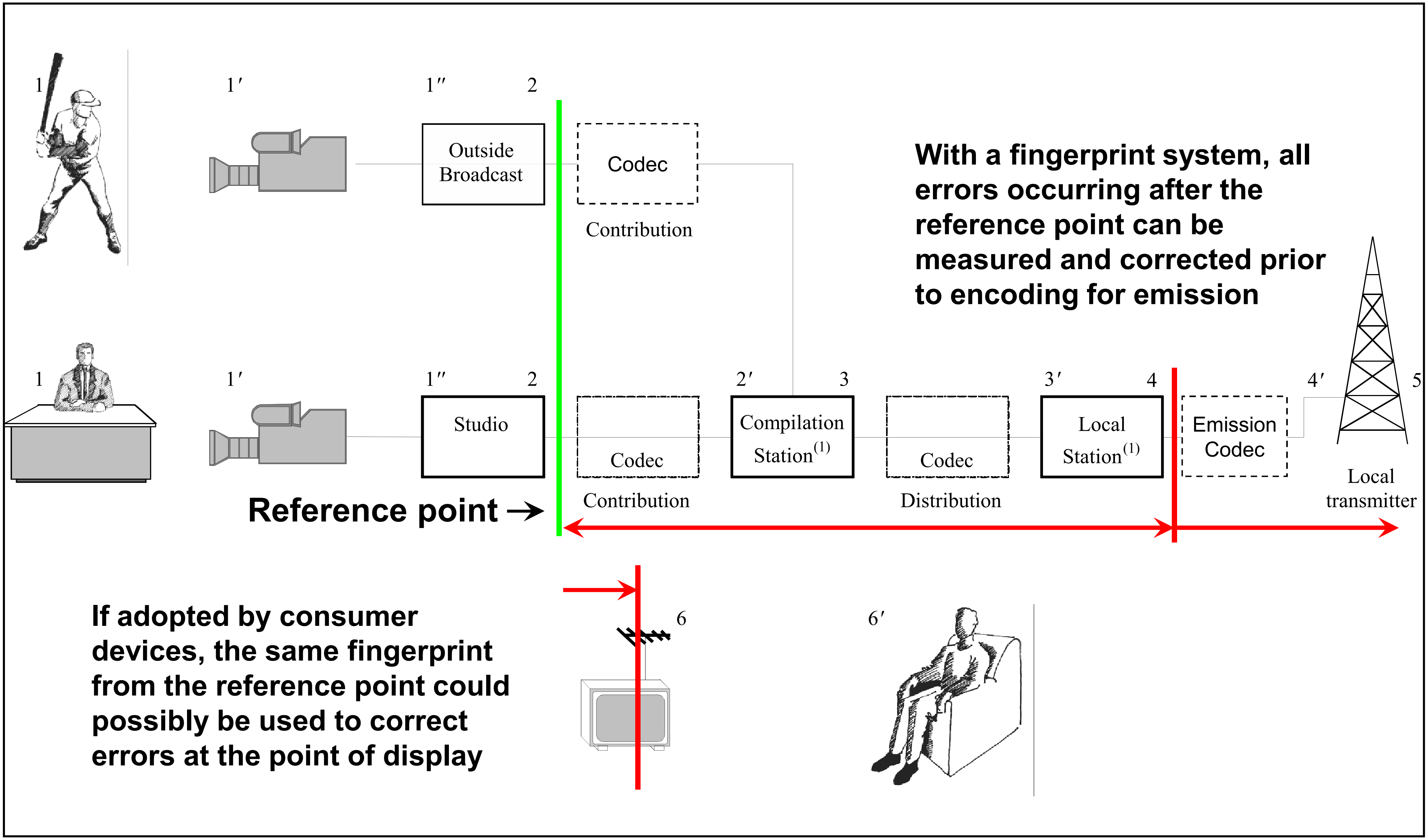
A/V Sync Correction



Dolby A/V Signature File-based System

Slide courtesy of Dolby

Broadcast Chain



(1)

Advanced Television Systems Committee

Products/ Technologies

- Evertz IntelliTrak™
- Miranda Densite HLP-1801
- Sigma Electronics Arbalest™
- K-Will QuMax 2000™
- Dolby A-V Signature
  - All use A-V signature / DNA / fingerprint metadata
  - All assume correct sync at the input reference point
  - All measure errors at downstream point, enabling errors to be corrected automatically

A Standardized Fingerprint?

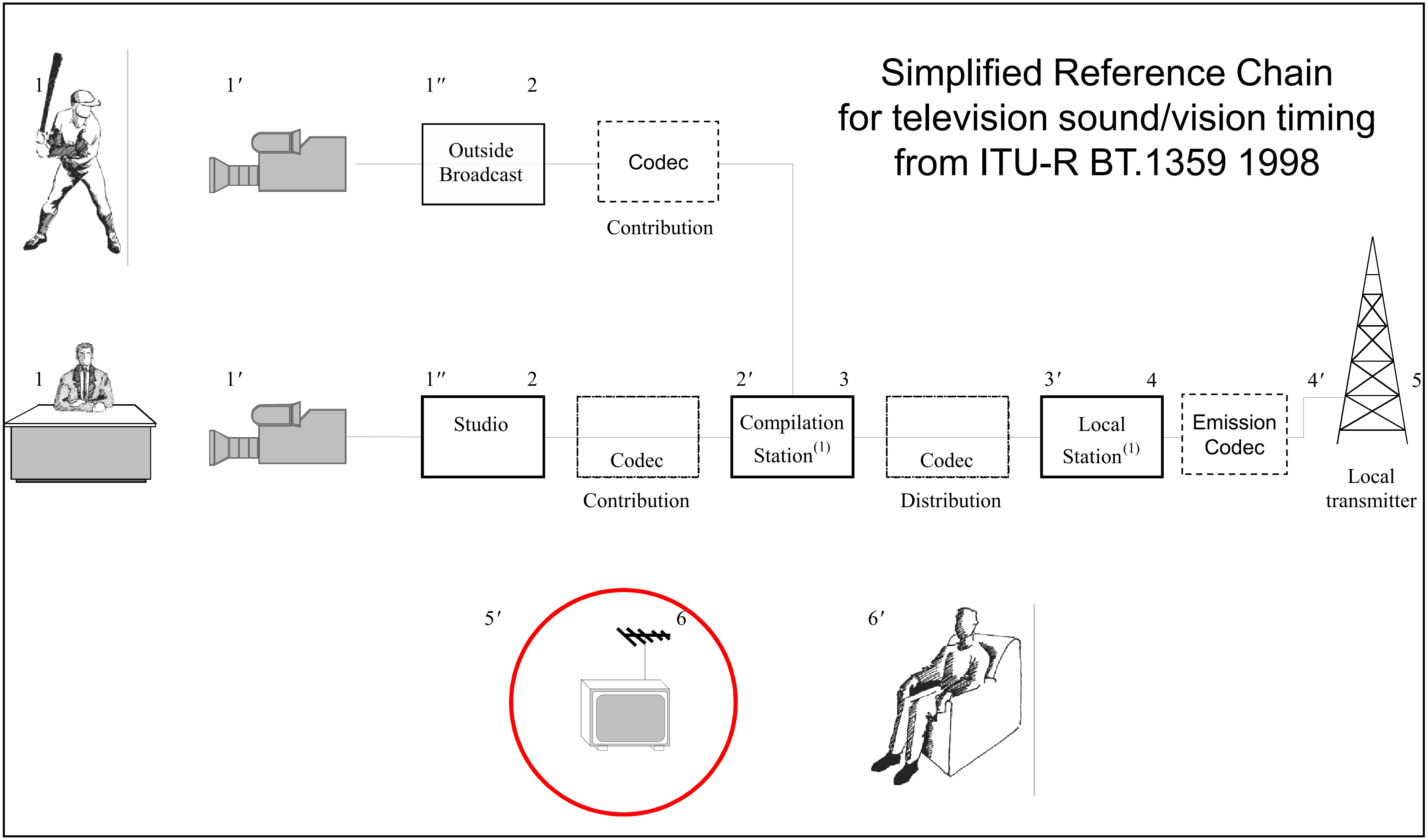
- Entire program chain usually not under control of broadcaster
- From user's perspective, it is highly desirable for equipment from different manufacturers in different parts of the chain to interoperate
- Is standardized fingerprint metadata for A-V sync the solution ?
- Standardized transport methods ?
- Seeking input from broadcasters and users on what they want from manufacturers

SMPTE 22TV Standards Work

A-V Sync Measurement and Assessment

- Project scope: Define recommended techniques for audio-video synchronization error measurement, and techniques and environment for synchronization assessment
- Specific tasks: Determine requirements for consistent out-of-service measurements and in-service assessments and measurements of audio-visual synchronization errors, as may be necessary and practical.

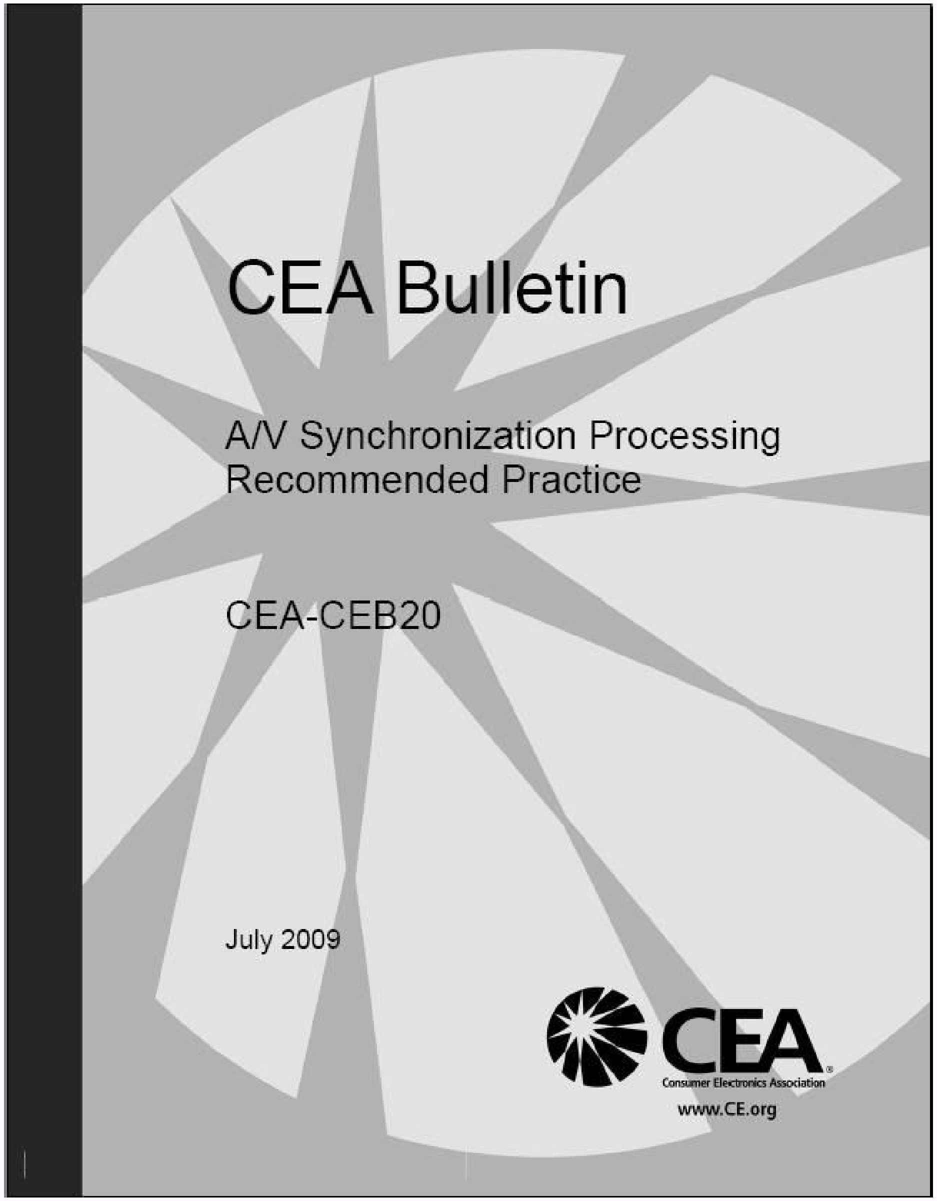
DTV Receivers



(1)

Advanced Television Systems Committee

CEA-CEB20


CEA Bulletin

A/V Synchronization Processing
Recommended Practice

CEA-CEB20

July 2009

CEA
Consumer Electronics Association
www.CE.org

a t s c
Advanced Television Systems Committee

CEA-CEB20

- "A/V Synchronization Processing"
  - "… outlines the steps that an MPEG decoder should take to ensure and maintain audio/video synchronization. Such synchronization is necessary for end-viewer satisfaction."
- Written assuming the reader has a fundamental understanding of MPEG-2 Systems, but not of "real world" conditions

Real-world Conditions

- Why is this important?
  - Designers often are not aware of the types of input disruptions that are common and the consequences of those to decoding
  - Designers forget seemingly obvious things, such as PCR wrap-around
  - Designers may not understand the importance of frequent cross-checking of clock samples between separate audio and video decoder ICs

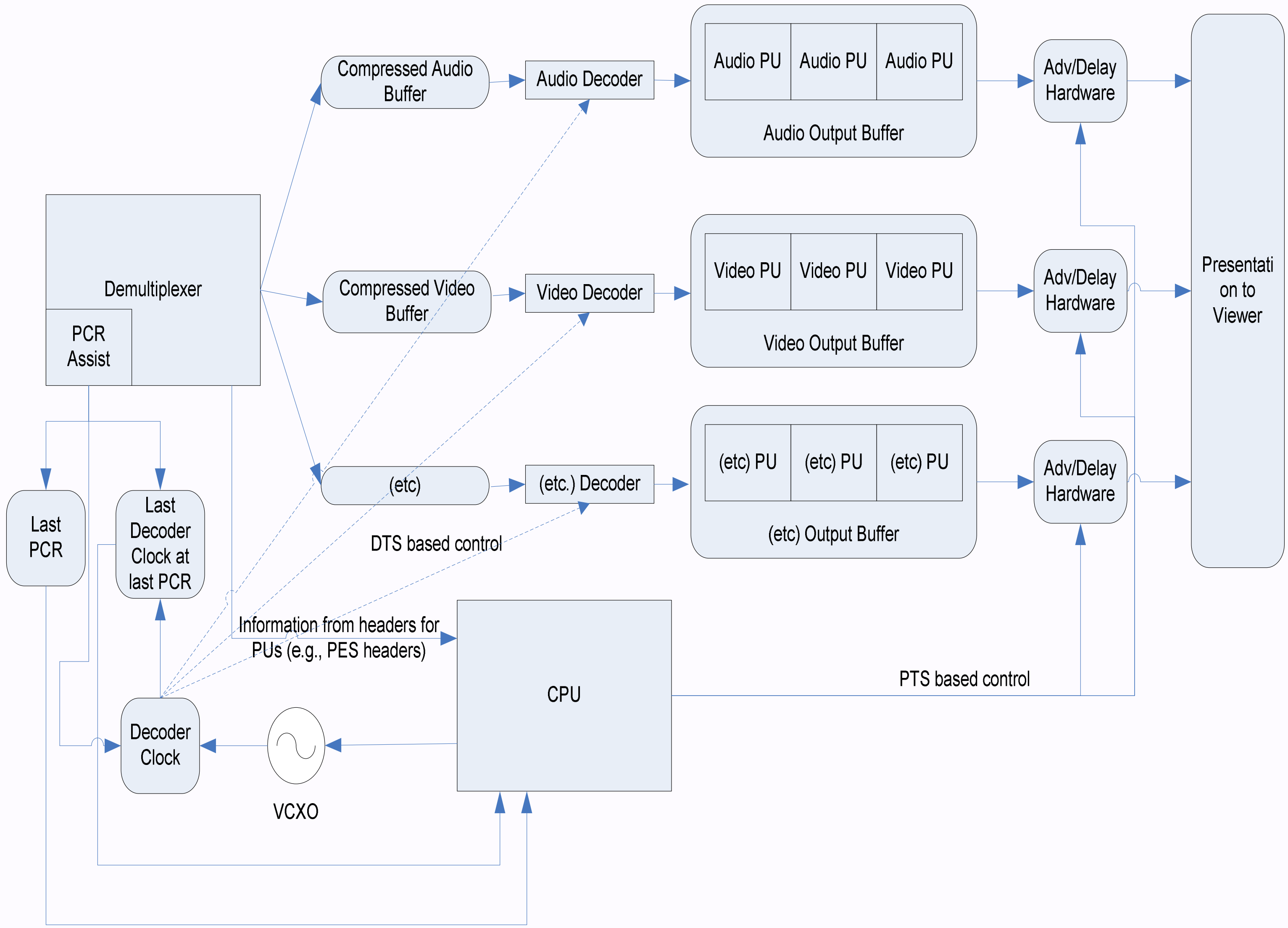
Real-world Conditions

- The industry continues to see new entrants into the decoder market
  - Both for professional as well as home use
  - Even experienced engineers (with traditional video/audio backgrounds) make horrible assumptions about MPEG
- While CEB20 will assist, it cannot be regarded as a “panacea”

CEB20 Major Topics

- Receiver Architecture Model
- Decoder Clock Startup and Maintenance
- Presentation Time Processing
- Advanced Transport Stream Processing for Recording or Remote Playback
- Carriage of MPEG-2 TS over IP networks

Receiver Hardware Reference Model



Receiver Architecture Model

- Demultiplexer PCR Assist
  - How the demux hardware can assist keeping clocks accurate
- Decoder Clock
- Hardware for buffer management
  - Identifies issues with variance in buffer sizes between SDOs (DVB vs. ATSC/SCTE)
  - Discusses maintenance of A/V sync at a high level
- Audio and Video Output Clocks

Decoder Clock Startup and Maintenance

- Startup
- Disturbances to the MPEG Transport Stream
- Major Adjustments
  - System Time-Base Discontinuity
  - Recommended Decoder Clock Error Event Recovery Method
- Minor Adjustments

Presentation Time Processing

- Startup
- Practical Considerations
  - This is a key area… and needs attention paid to it
- Adjustments
- Major Adjustments

Advanced Transport Stream Processing for Recording or Remote Playback

- Partial Transport Stream Recording
  - Recovery of SPTS from MPTS
  - Clock maintenance in such a situation
- Maintaining Inter-packet Timing Relationships During Playback of Recorded Content
  - Critical for recovered SPTS
  - Pointers to two documented methods of doing this

THANK YOU