Broadcast Project Research Ltd

"Smartlips" Products

Audio/Video (A/V) Sync Test Sources and Equipment

A/V TEST SOURCES:-

1) "TSG" (Television Test Generator) DVD

625-line Video test material on a two layer DVD, with multichannel audio in Dolby Digital and DTS formats.

(Includes a comprehensive explanatory booklet).

2) "SMARTLIPS GENERATOR"

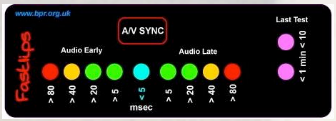
Physical Blip and Flash generator for Studio and Location testing.

A/V MEASUREMENT TOOLS:-

2) "SMARTLIPS"

Hand held video display Photometer, Sound Level Meter, and intelligent off-screen A/V sync meter. Works on all systems, HD or SD.

3) "FASTLIPS"



An A/V sync meter using electrical inputs. (Including the optional plug-in Serial Digital adapter with audio de-embedder). Works on all systems, HD or SD.

More details of all these items may be found on:- www.bpr.org.uk

A/V sync error resources:-

- EBU R-37 and Tech 3305 recommendations are freely available from:- www.ebu.ch
- The SMPTE currently have the S22 group dealing with Lipsync issues:- www.smpte.org

About ChEFF:-

The Chief Engineers of Facilities Forum represents the views of the UK TV facilities industry on engineering and operational matters. It provides a focal point for dialogue with the broadcasters, standards bodies, manufacturers, equipment suppliers and between its own member companies.

ChEFF has close ties with FEITIS, the European facilities organisation, and with the EBU:-

. http://cheff.sohonet.co.uk

About Broadcast Project Research Ltd (BPR):-

Broadcast Project Research Ltd is a Studio-based Research and Design Group based at Teddington Studios in the UK. They specialise in Research into areas such as the Perceptual Effects of Audio and Video, along with a strong emphasis on Quality Control (QC) and Code-of Practice Standards. They produce a small range of QC and test equipment, including "Gordon" the photoepilepsy monitor used by all UK broadcasters, whilst the News Jem is the latest in their range of digital video and multichannel audio interface products. Further details may be found at: - www.bpr.org.uk

About the ITFC

The itfc provides a unique range of post-production facilities and access services (Subtitling, Audio Description & Signing) to some of the biggest names in the broadcast, cinema and DVD industries, and has been established for more than 25 years:- www.itfc.com

A Pocket Guide to Lipsync

Ever since Cinema audiences first laughed at out-of-sync "Talkies" in the 1920's, lipsync has been just a problem waiting to happen in Television.

In film practice, lipsync tolerances of the sound being no more than a frame early or up to two late were soon established. But Television production process have now became so complex that a tighter tolerance is necessary at each stage in order to maintain the same basic figure at the end of the chain in the viewers' home. The EBU and SMPTE have recently tightened their recommended tolerances at each stage down to 5-msec sound early and up to 15-msec sound late.

The question then arises as to how this value can be ascertained and maintained in day-to-day Quality Control operations.

This Leaflet has been prepared by an editing group from the UK companies BPR and ITFC, in conjunction with the ChEFF facilities forum.

In Analogue

Except for situations such as News contribution, where audio and video could be sent by different paths, the only place where pure analogue SD television could experience lipsync problems was during post-production where the audio and video streams diverged. The use of a blip-and-flash on the tape heading “VT Clock” provided a “Rosetta Stone” for downstream reference purposes, and in sound dubbing the sync could be re-aligned by eye and ear using the low latency CRT displays.

When Digital video synchronisers were introduced, the resultant variable video latency placed the audio early by up to a television frame or two. This was a noticeably unnatural situation for all viewers. The solution here was to use an audio delay in parallel with the synchroniser and coupled to it. Of course the audio delay cannot be tightly coupled to the video delay, as you cannot drop or repeat a frame length of audio, as you could with the video. Most systems therefore used a slow stretch or compaction of the audio stream, but nonetheless the error was effectively corrected at source. Correcting these errors at source is still a desirable condition in the increasingly complex workflow of HD.

High Definition

First experiences with High definition Television are definitely showing a heightened susceptibility to lipsync errors. Why this is so is not hard to see. Our ears strain to understand not just the Syntax of the dialogue, but the whole body language of the human subject that is exposed by the higher amount of visual (mainly facial) information.

Pity the poor Actor though, they have not been so up close and confidential since the Theatre of the 18th Century.

The Cinema experience is slightly different, as the audience is subjected to a fairly fixed audio-visual genre, involving a visual update rate of just 24 segmented frames per second, and they are often a long distance from loudspeakers and screen. Remember that sound travels only one foot in a millisecond, but the light from the screen takes barely a nanosecond to travel the same distance.

Lipsync in Production

In SD Production there was the possibility of both a video delay in the vision mixer (especially with effect insertion) and to some extent processing in the digital sound mixer.

In the case of HD, there are two additional factors, firstly HD cameras have significant video latency at the output of the CCU, and secondly the vision mixer and even the displays used for viewing in the Gallery also have a significant and possibly variable latency. On the tape or output file however, the significant factor is the camera plus mixer latency, and if a matching sound delay is introduced at this stage, the output file or tape can be considered to be in sync with the header blip and flash. Although unlikely at this stage in the production process, any compressed audio contribution or distribution streams will need careful attention.

Production Advice-

Liberal use of VT clocks to EBU R37 will help sorting out problems later in the chain. Introduce audio delay to match the overall video channel latency. Only introduce viewing compensation audio delay for local areas. Deal with each output stream individually, as cross or down-conversion can introduce different problems in each stream.

Lipsync in Post-Production

VT clock timing should be checked at ingest on every clip, and also on completion of any copying or format conversion. Indeed it is often in the simplest mechanical transfer where the sync problems are introduced. For instance Aspect Ratio Conversion (ARC) can add two frames of video delay. Any compressed format contribution or distribution for audio or video should be treated carefully as well.

Post Production Advice-

Maintain your VT clock on all clips and check the A/V timing in the electrical domain. Compensate for viewing latency only in the local domain. Watch out for problems introduced by any format transfers, especially where audio or video compression is introduced.

Lipsync in Reception

The EBU Project Group B/MCAT found many problems with early set top boxes, and issued Tech Doc 3311, “Guidelines for Multichannel Audio in DVB”, and this freely available document is worth study where receiver problems are encountered. In current installations we can assume that the predominant problem with A/V sync will occur once the audio and video signals diverge, and although many current display devices are capable of delaying audio carried on the HDMI interface, some cannot accept multichannel audio streams via this route.

Reception Advice -

Ensure the STB meets EBU requirements for audio timing. Ensure display latency is accounted for. Ideally this should be done within the display itself, as latency could vary markedly if the emission coding changes from programme to programme or channel to channel.

Measuring Lipsync

With Timecode

Separate audio and video timecodes are an ideal way of showing offsets, and indeed of correcting sync errors. However, although timecode translation has been used for years in systems such as the DTRS tape, reliable and consistent separate A/V timecode systems are rarely found along the overall signal chain.

Buried data based

Hidden metadata markers, which can be used throughout the signal chain for both measurement and correction, have been proposed for many years. However with the wider production base that exists today, it is unlikely that any one system could become sufficiently widely available to be used reliably. There is also the danger of wrong marking causing more trouble downstream than the initial error, or of legacy markers being left in place and causing errors.

Eye and Ear

The human eye and ear method only works up to a point around one or two frames from the correct timing, at which point even experts will have a problem with the direction of error. As with all sound monitoring, one operator can view many screens but listen to only one, so this is undoubtedly a labour intensive method.

Electrical layer

Embedded or separate audio signals can be easily checked with respect to video streams, thus removing any influence of a viewing monitor. An oscilloscope and signal decoders of known latency for both audio and video are a minimum requirement, along with a blip and flash “Rosetta Stone” carried on the programme material.

Visual layer

Checking errors off the screen is ultimately the most satisfying method as it covers the whole signal chain. With streaming media It requires a photometer and sound level meter based approach, and typically an oscilloscope to display the results. A blip and flash “Rosetta Stone” is essential.

Non real time file based checks using jog wheel searches do not require the blip and flash, but they do require a degree of confidence in the relative reading offset and consistency of the workstation and program reading the video and audio files, especially where compressed media are involved. Remember as well, that some bass musical instruments such as Tubas only start to sound maybe 5 frames after the keys have been pressed.

Fixing the errors, Static or Dynamic?

Static errors are typically introduced by a start up slippage, or decoder or transcoder latency, whereas any vision synchronisation function or wild sync tracks will introduce dynamic errors. Making repeated and extended measurements is the only way of exposing the full range of possibilities. If automatic correction cannot be applied, err on the side of audio late, as that is a more natural effect.

Above all, dealing with all sync errors where they occur, is a far better solution than trying to mop up large and probably variable errors at the end of a process.