

For DTV audio quality, there's good news and . . . hope

PBS has convened the Quality Group to improve the quality of the public TV digital signals reaching viewers' homes. This article is part of the group's series on real-life issues in DTV production and broadcasting. The author is Bruce Jacobs, chief technologist of Twin Cities Public Television, former chair and long-term member of the PBS Engineering Technology Advisory Committee (ETAC).

The group is also sponsoring a series of low-registration-fee regional workshops on these topics (story at right).



Part 4: Audio demands attention

By Bruce Jacobs

This Quality Group feature is a little different. Instead of dealing with characteristics of DTV pictures, we'll cover three aspects of the ATSC digital standard on the audio side: lip sync, loudness, and multichannel sound.

Digital television offers the potential for dramatic improvements in audio quality. With careful management of these problem areas, we can make the most of that potential.

Audio synchronization

Let's start with a Quality Group Puzzler in the great tradition of *Car Talk*:

Julie's company makes jars. She thinks her factory is making jars of the right size, because all the lids that she tries on her jars fit just fine. Larry's company makes lids, and he thinks they are all the right size, because they fit on every jar that he lays his hands on. Priscilla makes preserves and has a problem. The lids she ordered from Larry don't fit the jars she ordered from Julie.

Whose fault is it? How should it be fixed?

While you are thinking about the answer, let's consider a clue from television.

We've had problems with lip sync ever since the early 1980s, when we started to add digital devices that delayed the pictures but not the sound. The Frame-sync and Digital Video Effects (DVE) were the first offenders. As long as the error was limited to a single frame or two, the human brain ignored the discrepancy and nobody noticed. Freddie could think his frame-sync was fine and Delores could be unconcerned about her DVE. But when we started using more and more of these devices in the chain at producers, networks, and stations, the tiny delays cascaded into an annoying gap. Whose fault was it?

A television 'frame' is about 1/30 of a second. People have slightly differing views on how much video delay is imperceptible, how much is perceptible but acceptable, and how much is simply unacceptable.

One reason is that individuals have different perceptual capabilities. A commonly held view, which PBS adopted, is that lip sync is acceptable by most people when the audio leads the video by fewer than two frames (1/15 of a second) or the audio trails the video by fewer than four frames (1/8 of a second). People more readily accept the audio trailing the video than the other way around because this is what we experience in real life, such as when we listen to someone speaking across a large room.

Now, for the Puzzler answer. Some callers thought Larry should adjust his lids to fit Julie's jars and Delores should adjust her station to compensate for Freddie's frame-sync. While this would solve one problem, it is the wrong answer, because it would create another problem that can be even worse: We might fix the immediate problem with the result that Larry's lids don't fit any other jars and Delores' station is now incompatible with any other frame-syncs.

A much better solution is to reach agreement on the proper dimensions for lids and jars, the proper amount of lip sync, a good way to measure it, and an acceptable toler-

ance. The problem is then solved universally.

In television, the correct solution is zero lip-sync error in every part of the system. If we can use equipment that *always* delays the audio by the same amount as the video, we don't even need a way to measure the sync. This kind of automatic correction became a lot easier when we converted our studios to digital. (When audio is embedded with the digital video, it is delayed by the same amount.)

Just as we were getting a handle on fixing the lip-sync problem, DTV came along and made it a whole lot worse. The reasons are complicated and the implications are sobering.

DTV broadcasting (and many of the other systems in our industry) uses MPEG compression to provide high-quality digital video within a practical number of bits. In MPEG, video signals encounter huge delays that vary greatly depending on picture complexity. This is due to the enormous amount of processing needed over multiple video frames and the need to change the number of bits used as the encoding difficulty varies. To keep audio and video in sync, the MPEG standard provides time-stamp references in both the audio and video packets, which *should* be used in the MPEG decoder to line them back up again.

Did you notice that I emphasize "*should*?" It turns out that this feature *rarely* has been used properly in the design of *millions* of MPEG decoders in the marketplace and there is no one with the authority to fix this situation.

Notice the italics again on "*rarely*." Many decoders ignore these time stamps completely. Many more use the time stamps only when a user changes the channel and then ignore the time stamps until the next channel change, assuming that the audio-to-video relationship will not drift. The problem is, it drifts!

These MPEG compression chips, with the time-stamp synchronization function missing, were used widely in both cheap consumer gear and pricey professional equipment. We learned this the hard way when the first video server implementation at PBS used professional Bitlink IRDs, with this varying lip-sync problem, to decode the server files. To keep the audio in sync, the PBS staff had to reset the decoder a few times a week to prompt it to resync the video and audio packets!

The good news is that PBS converted to new servers, eliminated the faulty IRDs and solved that problem. And the other good news is that the consumer industry saw the error in their ways a year or two ago and started making MPEG chips correctly — so consumers using newer sets will suffer one



More than 80 technical and production personnel from stations and production organizations gathered in San Francisco for the first of the Quality Group's regional workshops in San Francisco early this month. Planners considered it a rousing success. Speakers, including John Watkinson (pictured) covered DTV technology from field production to over-the-air broadcasting.

The group's second workshop will be held at Twin Cities Public Television in St. Paul, July 8-9, with expert speakers including tech director Mark Shubin and Canon representative Larry Thorpe. Additional workshops are being planned, including one in the Northeast. For more information or to register, check: <http://pbsconnect.org/qualitygroup>.

less reason for lip-sync problems. But think about it: There are still *millions* of faulty decoders in use by consumers and professionals, with lip sync that drifts. Talk about a moving target!

To add to the lip-sync problem, digital television brought forth a host of popular new digital displays that *universally* have significant video delay, sometimes as much as *four frames!* While newer consumer systems can add an audio delay to match the video delay, we cannot trust that all consumers have them and will properly adjust them!

The best we can hope for is that we do our job right, that the older MPEG decoders eventually will be replaced, and consumer display and audio equipment will become more integrated, resulting in better sync all the way to the viewers.

If only it were even that easy. To top it all off, even some professional equipment has obscure, poorly documented settings that can result in erratic variations of lip sync, amounting to 2 or 3 frames. The only remedy seems to be regularly measuring your systems' behavior, end to end, and tracking down the cause of any erratic behavior that you detect.

In response to sync problems, manufacturers have focused on the detection and measurement of lip-sync errors. There are various ways to measure lip sync — here are three:

- For encoding systems, Miranda and Evertz have introduced products that use *regular programming* to run sync tests. These

products can compare the signals at two ends of a distribution system and report the total sync error.

- If you lack this equipment, you can embed a simple test ID with a frame with a white dot and low level tone. Play it in a break and record it after passing through a system. Load it into a video editor and scrub the timeline, observing the audio waveform and white flash to measure the sync error.

- For displays, Syncheck is a very useful low-cost device.

Summary: Why lip-sync problems are confounding

- Errors show up only when there are tight shots of real people speaking.

- Our brains are accustomed to late audio but not early audio.

- You can have a serious sync error in one system but you can't see it until multiple errors cascade into something painfully noticeable.

- New professional and consumer digital displays introduce sync errors.

- MPEG sync errors can drift.

Tips for maintaining proper lip sync:

- Be diligent in designing facilities: Provide an audio delay to accompany every video delay, even if it *looks* okay without.

- Don't fix one delay upstream by adding an opposite delay downstream. Fix it at the source. Two wrongs don't make it right.

- Make sure you can accurately monitor sync problems by maintaining video and

audio monitoring calibrated to have perfect sync. Because digital video displays themselves have video delays of up to four frames, you may need to add an audio delay unit to every monitor that you use for monitoring quality.

- Beware of obscure MPEG equipment settings and measure your systems over time. Wrong settings can result in erratic lip sync.

- Beware the single complaining consumer. Their new flat-screen display might be delaying the video by multiple frames, while their set-top-box and surround-sound system aren't delaying the audio all. The single complaint may mean there's a problem in a single home.

Audio loudness

After all that bad news about lip sync, we need some good news. With the topic of loudness, it's all good news!

In analog television, the rules for audio loudness were pretty simple. The FCC said, "Don't over-modulate or we'll fine you." So everybody made sure they were loud, without going over the limit. This practice provided reasonably consistent loudness and a horribly restricted dynamic range.

The DTV standard came equipped with a much better plan: to give viewers consistent loudness *and* a wide dynamic range. Simply put, every operator of an encoding system is required to set it to match the *average* loudness of the incoming audio signal.

Every consumer decoder is required to use this information to insure that the *average* perceived loudness is consistent from show to show and from channel to channel.

The beauty of the system is that average loudness is consistent, and yet programs can have dramatic musical crescendos or sound effects much louder than the average.

When the system was first implemented, however, many of us in the industry didn't understand how it was supposed to work. Some broadcasters blatantly disregarded the loudness configuration requirement. One thing that was certain: Consumer equipment did behave as the broadcaster instructed it — right or wrong!

Today the news is all good. Last year's approval of the Advanced Television Systems Committee A/85 recommended practice (RP), has resolved the conflict. The RP has been endorsed by all U.S. broadcasters and cable MSOs and is available for free: www.atsc.org/cms/standards/a_85-2009.pdf.

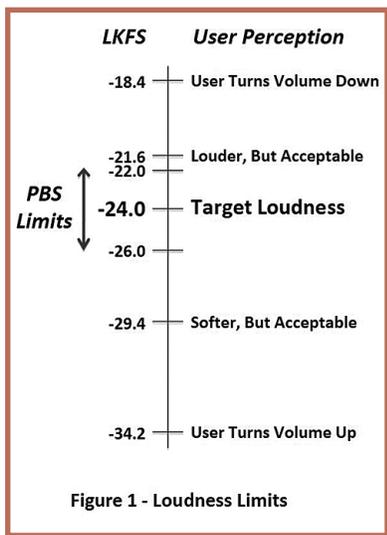
Under this consensus, broadcasters and cable systems will set their encoder "Dialnorm" parameter to represent the average loudness at the encoder input. This insures that consumers hear consistent loudness.

The RP has an important additional requirement that the FCC did not put into their specs: a standard loudness level for program submissions. The level chosen, -24 LKFS (a measure of perceived loudness in decibels below full scale), is the same level PBS established for our program submissions three years ago! Even better, there's an international consensus behind the same measuring method and target level.

While it's good practice to adjust loudness for a consistent target level, consumer perception allows some tolerance before the users head for the remote. Figure 1 shows the range within which PBS producers need to operate and the tolerances for consumer comfort.

Tips for proper loudness:

- Run your plant in unity gain for PBS network programming.
- Set your emission encoder to a dialnorm value of -24.
- Master local content so the loudness matches national content (that is, with peaks close to -10 dBFS)
- Install a new audio processor that applies mild correction for content that is out of spec.
- Purchase at least one Bs.1770(LKFS) meter.



- Try to achieve -24 LKFS ± 2 dB.
- Take daylong average readings of every station in the market and share your findings with your peers.

Multichannel sound

The topic of multichannel sound brings good news, bad news, and finally, more good news.

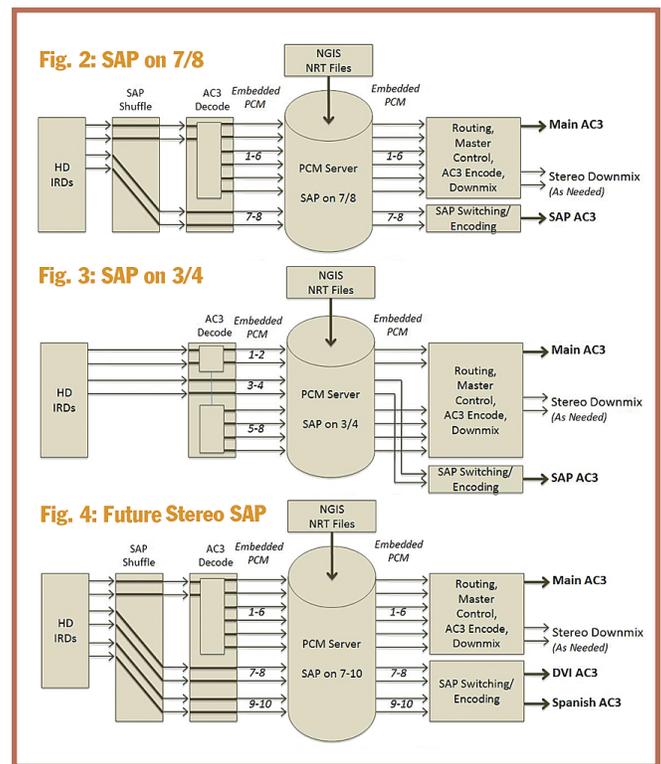
The good news is that DTV gives us the ability to broadcast audio in 5.1 surround. The bad news is that it's not very easy to implement in master control. The good news is that the non-real-time delivery of public TV's forthcoming Next Generation Interconnection System (NGIS) will make it easier.

The easy part of 5.1 broadcasting is the encoding. Simply buy a single 5.1 AC3 encoder, hook it up in place of the internal stereo encoder in your MPEG system, and be sure to configure the MPEG encoder to allow for the external encoding latency. (There's that lip-sync thing again.)

The expensive part is getting 5.1 content from the satellite receivers and through your master control. The most commonly used method is to place embedded AC3 decoders on the output of every HD IRD and configure the broadcast server to record six channels of PCM audio. Modern servers and master control switchers generally have this capability. But the cost of multiple IRD decoders can add up. You also need to provide a downmixer for each stereo output of master control (Fig. 2, above).

It's a bit trickier for those stations that have legacy server recordings with SAP on pair 3/4. Some stations record the 5.1 as AC3 on pair 1/2, with automatic AC3 decoders on every server output. Others wanting the simplicity of PCM reshuffle the 5.1 to appear on pairs 1/2, 5/6 and 7/8 (Fig. 3).

Stations may want to plan for the possibility of future delivery



and broadcast of stereo Descriptive Video and Spanish services (Fig. 4). By a fortunate coincidence, the programming most likely to be delivered with stunning 5.1 audio is also most likely to be delivered as files by public TV's NGIS. This will allow stations to implement 5.1 broadcasting without having to buy AC3 decoders for the IRDs used for ordinary real-time satellite reception. All that is needed is a compatible server, switcher, 5.1 encoder and downmix (Figs. 2-4).

Tips for 5.1 surround broadcasting:

- Start planning for 5.1. Compelling content and compatible home systems are becoming more commonplace.
- Get your master control and server ready.
- Get geared up to start using the NRT system when it goes online.
- Budget for a 5.1 broadcast encoder. ■

The Quality Group is funded by CPB through PBS's yearlong Video Technical Quality Improvement Program. Quality Group members include: Wendy Allen, PBS; David Felland, WMVS-WMVT, Milwaukee; Gerry Field, American Public Television; Frank Graybill, WNET, New York; Terry Harvey, WSIU, Carbondale, Ill.; Bruce Jacobs, Twin Cities PTV; Chris Lane, WETA, Washington; Dave MacCarn, WGBH, Boston; Tim Mangini, WGBH/Frontline; Ernie Neumann, Northern California Public Broadcasting; Mark Schubin, Metropolitan Opera (and other clients); Steve Scheel, PBS; Greg Tillou, National Educational Telecommunications Association (NETA); Steve Welch, NCPB; Eric Wolf, PBS; Ann Tucker (project manager), PBS; and Jim Kutzner (group manager) PBS.

Questions or suggestions for the Quality Group: qq@pbs.org.