Summary of comments received (second installment)

Regarding the guideline: *Audio Analog-to-Digital Converter Performance Specification and Test Method: Guideline (High Level Performance), Draft, June 19, 2012*

Regarding the guideline's introductory text: *Audio Analog-to-Digital Converter Performance Specification and Test Method: Introduction, Draft, June 19, 2012*

Summary compiled July 12, 2012

> *Editor's note: The following comments are from George Blood, an expert in audio technology and the president of George Blood Audio and Video, a company that provides audio and video preservation services from their base of operations in Philadelphia. These comments were submitted in a series of emails directly to Chris Lacinak, the principal author of the two documents, and his replies have also been summarized below. The dialog has been modestly edited for this presentation.*

General comments

Comment: Who is the intended audience for the introduction? I feel it sits between the more technical among us - for example, members of the Association of Recorded Sound Collections and the International Association of Sound and Audiovisual Archives Technical Committee -- and the less technical -- most FADGI members, small institutions trying to use this as a guide to purchases. The more technical folks will challenge the document as needing more rigor . . . and the less technical will just try to match up the numbers.

Response: We did try to hit a point between the more technical and less technical. I'm not sure that the more technical will challenge it as needing more rigor. The IASA specs, for instance were admittedly best guesses and this guideline builds on that work, adding significantly more rigor. We think it is a major step forward from prior work on this topic. However, we will welcome challenges for more rigor, as it provides an opportunity to point out the how little exists today and the need to do some work. If people are passionate it would be great to direct that energy toward continued progress.

Comment: Many people will pick up this document, start reading and their eyes will glaze over the technical stuff; then they'll jump to the charts and just read the numbers. As such I'm concerned there isn't more extensive and rigorous testing. You acknowledge as much, but given how this document is likely to be applied in decision making, I think this is a critical weakness that should be addressed.

Response: I think this guideline and the introduction, as they stand today, add to and advance the state of audio test and measurement and are a valuable contribution. On the other hand, I would never argue that they are perfect or that additional testing isn't a worthwhile endeavor, and I embrace your thoughts on the value of additional testing. I am all for additional research and development that continues to add to this body of work.

Comment: I recommend two additional sections, one explaining how theory and real world differ, and another with more about the basis of how the new specs were derived.

For the section on real world vs. theoretical performance limits here are a few issues to consider:

- if a 24 bit converter has a 144dB dynamic range, why is Dynamic Range of 120dB adequate (because you can't get 24 bits of resolution at room temperature because the thermal noise of atoms is louder than 130dB, etc.)
- why don't we test below 20Hz (issues of performance as we approach DC)
- general discussion of perception (hardly anyone can hear frequencies to 20kHz, Fletcher-Munson and perception of frequency response a different levels and hence A-weighting, etc.)

This might be a good place to make the case for not getting hung up over a dB here or there. As a practical matter there will be a greater deviation from perfection in any given transfer. It's simply not possible to know what the highest peak is in a recording until you've digitized it. You're more likely to leave extra dB on the table with poor level setting than comparing spec sheets over a few dB.

Response: These are good and interesting proposed additions. This is a guideline and not a standard, but we have followed the example provided by other Audio Engineering Society (AES) standards. The intro document is what would normally be an appendix on a standard. And in my experience there is always an ongoing discussion in standards creation about the where to draw the line between providing the necessary information with which to interpret and use the standard vs. creating an educational text that teaches people who don't know otherwise. How this discussion typically gets resolved and addressed in standards is by pointing to other standards, information documents, and texts that provide the proper context that people feel are missing. So, I agree that the topics you raise and probably several others could be very helpful to readers, and my instinct at this point would be to identify and reference other texts that address these fundamental test and measurement topics rather than to address them within this document.

Comment regarding the two preceding responses: I recently served on the Minimum Capture Taskforce for the American Library Association Preservation and Preservation and Reformatting Section (PARS). The premise was there's not enough time to digitize everything, we shouldn't waste time re-digitizing items done by others. Rather than setting a *best* target, the goal was to define a minimum at which another institution had digitized an object that would be good enough that it wouldn't need to be redone. We took just the approach you advocate, citing (usually via URL) to relevant standards, targets, discussions and such. By the way, this report is currently under review at the PARS Executive level and FADGI members should watch for it later this year.

Comment: Regarding the new specifications: I'd like to see more explanation for how specs were derived. I'm left with the impression that they were empirically derived, or to put it another way: the specs are based on what's being delivered in the marketplace already, rather than upon some scientific basis. This may be perfectly adequate, but I don't think the case is made.

Response: We do explain how the specs were derived in the *Introduction*. The reality is that in many cases the specs were based on an informed estimate of what was simultaneously stringent

but realistic. At the heart of your comment is an important distinction - theory vs. practice, or top down vs. bottom up. What I realized early on is that approaching this question from a theoretical perspective gets you nowhere fast. We can easily come up with performance metrics that represent an unquestionably amazing converter on the basis of authenticity and integrity. We could even factor in the limitations of the source media in order to anchor things to reality a bit more (although "limitations of the source media" is a Pandora's box of its own). At the end of this exercise there's a risk that we end up with performance specifications that no converter meets and we haven't really done anyone any good.

What you see when you look at the test data is that even high end converters tend to excel in certain areas based on design decisions which have implications to performance in other areas. This was a reality I came to terms with in performing these tests and also in speaking with ADC manufacturers. There is no perfect converter, and multiple highly educated people will argue about which parameters are most critical for transparency, authenticity, and integrity.

We wanted to create a guideline immersed in practicality, that can be picked up and used today, and so we took more of a bottom up approach. We decided to test actual converters expected to provide a variety of performance, and to use real data as the basis for navigating through what really matters and to what extent it matters. We still waded through the theoretical factors mentioned above, but they don't come along with hard data to help define the finer points of distinction, so they served as guides rather than serving as the basis of any decision point. The current recommendations represent all the input I could garner from people that were knowledgeable and willing enough to provide input on where the test method and performance specs needed adjusting. Some of these points represent ongoing complex disagreements between experts in the field.

We continue to welcome input on where things may be off, but at this point I can't find anyone that disagrees with these recommendations based on theory, practice or otherwise. After this document is approved, I am going to push it through AES SC-02-01 as a standards project. This document has already been informed by the key players in this group, but my hope is that pushing it through as an AES standard will draw others out to provide additional input and will further expose and flesh out some of the disagreements between experts. Of course, we'll update the guideline appropriately but as slow as the standards process is we didn't want to tie this guideline to the AES standards process.

Comment regarding preceding response: I strongly agree. But I'm a bit cynical that people will fail to understand the distinction between "it has to meet these numbers" and "this is a work in progress." The latter, which I believe is your position, is that this guideline represents the current state of the art, building on what was done in TC-04, and on a path to the next best understanding in the future. Of course, I also feel that we all need to see more baseline data than five converters before this moves to true standards work in AES.

Comment: Here are three tests that I think could add significantly to the weight of the paper without "real" broad based representative testing:

- Test the same ADCs under different environmental conditions. How much does it matter that the ADC is in a rack and running much hotter than 70-74 degF ambient?

- Can you provide data on unit variation? The Library of Congress has thirteen same-model converters from one manufacturer, bought in two lots (as I recall). That would be a sample set of very like units that are readily available. If you can't take them out of production and put them on a bench to run the test suite, does the manufacturer keep unit-level performance tests at the factory?
- I have five of the same make-and-model as the Library, bought over a number of years. They're kept in near identical racked conditions, and are moved occasionally (such as for location recordings). Could putting my units through this test suite provide data for how converters age? Especially if the manufacturer has factory production baseline data on each unit.

Response: This is an interesting idea and, in my opinion, strengthens the appeal for more bottom-up real-world practical testing rather than more top-down theoretical lab-based testing. I think that this question is a great one, and actually supports our thoughts on the next round of work, which aims to apply the recommended test method and performance specs. Of course, the FADGI circumstances mean that only moderate levels of support can be offered for more testing, with luck sufficient to catch the more glaring issues.

The real test will be the test of time as this is implemented on a wider scale over a period of months and years. In thinking about this further, maybe it's worth adding some language recognizing the prospective limitations of performance specs based on limited testing and theory, and that continued efforts will aim to test the performance specifications in practice to further vet the recommendations. I hadn't been thinking along the lines of the types of information you are interested in, but I think your suggestions are good when positioned as "other value-added outcomes of this guideline and routine testing." Those are certainly good questions to pose and to keep an eye out for in continued testing.

Let me backtrack a bit again in this new context to your statement about the lack of rigor being a critical weakness of the guideline and posit that this is true of most standards. They almost always precede adoption and are further informed and refined based on real-world testing over time. With this in mind, I would argue that we've done more testing than most published standards already, and while we expect for the guideline to be refined by actual adoption and use (some of which will be embarked upon in a small way by FADGI), I feel good about publishing the guideline more or less in its current state

Comments on narrower technical matters

Comment: I'd like to see some testing of phase linearity. Anti-aliasing filter design decisions, for instance, impact many things and may explain many of the design trade offs you acknowledge. Could this kind of test expose and explain those trade offs?

Response: The way we approached this was to start with the IASA TC-04 specs. Then I matched them up to test methods from AES/IEC test methods and noted the disconnects. Then I asked a number of Test and Measurement experts what was critical and missing from TC-04 that should be added, inside or outside of AES/IEC test methods. I parsed through the responses and came to a decision about the final test methods and specs. Phase linearity simply never came up.

Upon your suggestion though, I'll revisit this parameter. Of course, we now have the data we have for the purpose of this project, but it may inform future work.

Comment regarding preceding response: Phase linearity is always a messy topic. I think in large part because its audibility is poorly understood, and because it's a topic that's more difficult to understand. The phase plots I've seen, notably record/play on analog machines and loudspeakers, make me cringe. How could we possibly accept such un-flatness? Because it's a necessary evil, an artifact of making other things much better, such as anti-aliasing filters and low pass filters to keep radio frequencies out of 1/4-inch tape.

Comment: I'd like to see more dynamic measures other than Amplitude Linearity. As you say it's hard to reduce such a test to single number, but it does represent more real world operating conditions.

Response: Thank you for the comment.

Comment: On page 10 in the *Introduction*, Dynamic Range/S/N Ratio. I'm wondering about the choice of word in the first sentence: "Dynamic range is an approximation to the perceived noise in the converter." How is this different from measured noise? What is weighting and why do we do it?

Response: I think that any person knowledgeable in matters of audio will make the connection between use of the word *perceived* in the first sentence and the text in the following paragraph that states: *This is passed through a weighting filter (A-weighting, band limited to 20 kHz, for these measurements) to simulate the ears sensitivity at low levels.* I'd prefer not to get into explaining too much, as I think that would open a Pandora's box. My preference would be to simply point to other texts for readers who need and wish to know more.

Comment: On page 20 in the *Introduction*, there is a comment in first sentence at top of the page about a human observer hearing sounds below the noise floor. Curiously this is the result of the increase in resolution provided by the dithering effect of the noise itself.

Response: This seems to be a musing more than a comment in need of response, but I'll respond anyway and say that it is true with or without dither and has to do with a characteristic of the brain which focuses on coherent patterns and effectively pulls them out of the noise floor. Tying this back to the comment on dynamic range, dynamic range (and therefore the noise floor) is a wide-band performance measure. Whereas the noise floor is wide-band, particular frequencies exhibiting coherent patterns can be detected beyond it. This is true in the analog or digital domain, and with or without dither.

Comment: On page 20 in the *Introduction*, Common Mode Rejection Ratio (CMRR). I think it's worth arguing the importance of this parameter based on the nature of the *input* to the ADC, namely the lower quality, less sophisticated, less "modern", output stages from many of the analog playback devices. It's their lesser design this parameter addresses.

Response: Valid point, but we're necessarily staying away from matters related to pre-ADC. Certainly not because we don't care or their not important but because we need to draw the line somewhere and once we go down that road there are lots of other things we need to address as well.

Comment: On page 25 in the *Introduction*, or more generally about the empirically derived test limits. What do your expert advisors say about the values stated? A little narrative could help legitimize the values given. I need something between a sentence and a short paragraph that bridges the perception of (a) "we picked these numbers because that's what the test results said was possible [good enough]" and (b) we received input from high level experts on the validity of drawing a line between the range of test values. I worry that you may be so close to the whole process that you do not see that the document doesn't broadcast your understanding as well as it should.

Response: Thank you for the comment.

Comment: On page 26 in the *Introduction*, Sync Input Lock Range. In dismissing this parameter perhaps discuss its legitimacy later in a digital signal chain. Since the ADC, even externally clocked, is going to be early in a chain, the jitter in its clock source should be low. After successive devices that's not likely to be true. This doesn't argue for keeping this parameter, only supporting it's value as a global issue. This is a real nagging issue with early videotape-based formats. We have found doing D-to-D transfers of those early digital F1 format recordings is exceeding challenging in terms of clock factors.

Response: A couple of responses here. To your first point, you are speaking to jitter accumulation. Sync Input Lock Range speaks to the ability of a device to lock to an external clock that differs from the devices own internal clock. It answers the questions "will or won't it lock" or "how long will it take to lock." But it doesn't speak to the performance of the clock once locked or to managing jitter accumulation. The more relevant parameters for jitter accumulation are Sync Input Jitter Susceptibility and Jitter Transfer Gain performance specs. For these reasons Sync Input Lock range is not seen as a critical factor. To your second point, we're focused on A-to-D here, and not on D-to-D.

Thinking about your comments is helpful; we will revise our text in order to add clarity.

Comment regarding preceding response: I agree. It is true that D-to-D is another lively discussion topic where the most common solution is to cheat.

Comment: On page 28 in the *Introduction*, THD+N. I wonder how much distortion is the result of the notch filter. That's beyond the scope of the paper but I'm still curious about it.

Response: Reasonable question, but I'll respectfully say that it doesn't much matter. The notch filter is necessary and the limits are created in reference to the test method. As long as the test method is used correctly the results are meaningful for an individual device and as a means of comparison between multiple devices.

Comment regarding preceding response: I will argue that the results are meaningful as a means of comparison between multiple units. The higher the Q of a filter the more damage it inflicts on the pass band. That damage is, by definition, distortion. Different test equipment will have different notch filter designs and produce different results. Of course it's a necessary evil. But how much it matters is probably a short discussion with one of your experts. It could be as simple as considering that the test is a broadband measure; as the percentage of the entire passband, the distortion imposed by the notch filter is trivial. Perhaps this is a topic for further testing.