Federal Agencies Digital Guidelines Initiative

ADC Performance Testing:
Low Cost Performance Test Setup
Test Method Round 2

Submitted by AVPreserve
September 1, 2016

Note from FADGI about the Scope of this Report

This report was submitted in September 2016 before the software development phase of the project began. As research and development progressed, it became clear that some of the concepts and ideas expressed within could not be executed or supported in the proposed manner and adjustments have been made along the way. Although some aspects have changed on the path from planning to execution of the final low cost testing system, FADGI sees value in making this report available as background material for the overall ADC Performance Testing project which continues to evolve.

Introduction

The purpose of this report is to perform the following activities excerpted from the original task order:

This activity shall build upon the (prior) 2015 proof of concept assembly of an ADC Test System (Partial, Low Quality). That system employed the following elements:

> *(1) NTi MR-PRO Minirator Audio Signal Generator device (furnished by the Library)*

> *(2) SpectraFOO and ARTA software packages (ODC funding provided by Library)*

> *(3) other elements (e.g., connecting cables)*

This requirement calls for the assembly of a working system, at a "beta" level of readiness, using the same elements used in 2015 or their equivalents, capable of testing ADCs against the 2015 draft FADGI low-cost performance guideline.

Vendor should propose a system that they feel will accomplish the goal, seeking to find a system that is also easy to obtain and affordable by archives of all types. That is, when adopted or used by future end-users who provide needed hardware and software themselves, the system elements ought to be low in cost (net outlay of less than $1,500) and, to the degree possible, including open source (or equivalent) elements and/or elements that will be practical for FADGI to provide at little or no cost.

At this beta level, the assembled system may be set up to require end-users to perform a series of manual operations, with minimal support from script or software modules that support only a segment of the process. Some support

shall be provided to extract the results of analytic testing of the output files, i.e., to analyze the files produced in the test and reporting on the performance of the system as compared to the metrics and pass-fail points specified in the 2015 Low Level Performance recommendation. Note that a separate requirement (C.3.2.1.1) concerns the development of software support for this system.

An additional element in this requirement is an analytic report that describes the system as developed with an explanation and/or rationale, together with discussion of problems encountered and how they were overcome. In addition, after carrying out the requirement in this section, and the field testing required in section C.3.1.1.4, the contractor shall draft a section for this report that assesses the success or shortcomings in the 2015 version of the FADGI low-cost performance guideline. If any refinements are indicated, these should be highlighted as recommendations in the report.

Reflections on Phase 1 Findings and Analysis

This phase of work (phase 2) began with a reflection on phase 1 findings, analysis and conclusions from 2015. This process led to new thinking on the framing and strategy for the low-cost test setup that was fundamentally important to the approach taken in this phase of work. Namely the primary shift consisted of our consideration of the purpose of the low-cost test system. The original vision for the relation between the high-level and low-cost test methods and systems was:

    a. the low-cost system would perform a subset of tests from the high performance test method;

    b. for the tests that the low-cost system performed, the relative quality of the test setups would dictate that the low-cost test system would only be capable of providing proof that an ADC was performing at a quality equivalent to the low-cost test system itself, and

    c. points a and b would be reconciled through a construct of quality levels using a star system (1 – 4 star quality where 4 is the highest quality) and the use of terminology speaking to the ability of a device to offer partial testing or complete testing in reference to its ability to perform all of the tests detailed in the high-level guideline. In this case a high-level test setup would be able to offer proof up to a quality level of 4 stars for the complete suite of tests, whereas a low-cost test system may be able to offer proof up to a 2 star quality level for a partial subset of tests.

In fact this proved to be a poor framing once discoveries were made in the last phase of work. Most importantly, the discoveries that the test methods employed by the low-cost test setup options were drastically different than those employed by the high-level test setup options, and that the low-cost

test setup results did not correlate with the high-level test setup results. We concluded in the last phase of work that this construct would not suffice, but it left us with a question of how to think about the purpose and utility of the low-cost test setup.

Reconsideration of these findings and realities in the beginning of phase 2 led us to think of a new purpose for the low-cost test. While it is not sufficient for the types of nuanced testing necessary for activities such as selection of the best ADC amongst a group of ADCs, or for examining precise characteristics and behaviors of a given ADC, it is sufficient for identifying failures and defects in an ADC. In other words, it won't tell you which is the best ADC in an ADC performance "shootout" but it will tell you when your ADC is failing. One might ask whether or not it is still worth it to move forward with the low-cost test method and system given that it has more limited utility. We believe that it is certainly worth it because we have found that the biggest obstacle to performing test and measurement in preservation facilities is: the lack of standard test methods and performance metrics; the high cost of test and measurement systems, and the complexity of available test and measurement systems. The net result is that most organizations performing audio digitization for preservation have historically done no test and measurement within their operations, internally or externally. We have also found that even very simple tests turn up issues in ADC malfunction, system configuration, and driver and firmware issues. Therefore, we feel that a simple to use, low-cost test system that is capable of performing test methods and reporting results in conformance with a standard is of great utility.

Once we reached this conclusion, it informed our thinking on the test setup itself and a reprioritization on the attributes of a low-cost test setup. Simplicity and cost became greater priorities than precision and accuracy. In phase 1 of this effort we ended up with a test system consisting of an MR-Pro test signal generator and the software application ARTA used both as a signal generator and signal analyzer. Although the ARTA was capable of generating all of the test signals and more, using our previous framework and thinking we preferred the MR-Pro because of the consistent reference point. ARTA is less consistent as a signal generator because it requires a DAC to produce the analog signal, and users may use any number of DACs, each with their own unique performance characteristics. With our new line of thinking we realized that the downsides of using the MR-Pro in this context are:
- there are certain signals it cannot generate which means that users would have to upload supplementary audio files in order to use adding another component and dependency on the system;

- it requires a non-real-time approach in which files are captured after running through an ADC and then played back in real time on the computer and utilizing a convoluted signal routing scheme to get the signals into ARTA. This added a great deal of complexity and opportunity for issues, and
- it is simply more components in the system, more complex overall, and potentially more expensive.

With our new framework and thinking we believe that removing the MR-Pro from the equation and using only ARTA as a signal generator and analyzer along with a user-provided-DAC yields the highest chance of achieving the goals of the low-cost test guideline.

Basic Setup

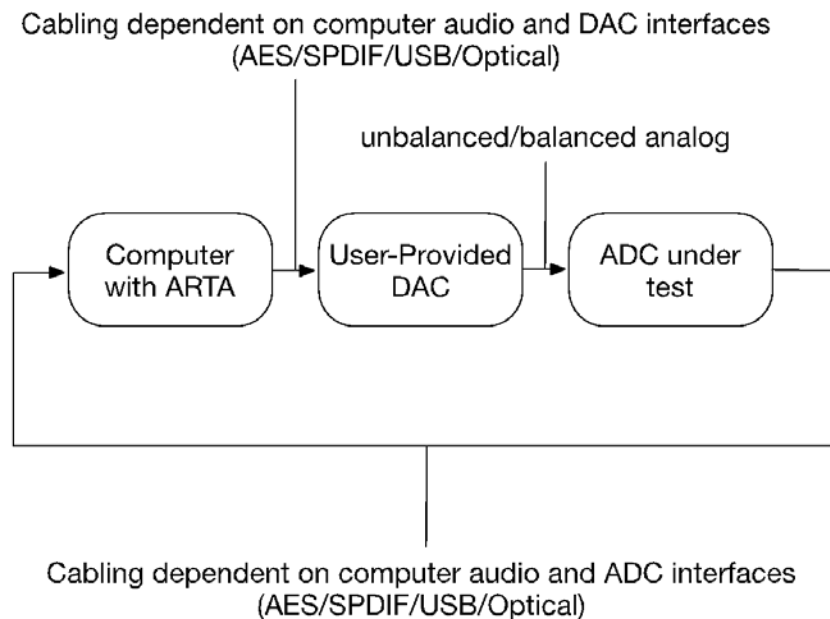The new low-cost test setup is detailed in figure 1 below.

Figure 1: Basic Diagram of Test Setup

This setup removes the MR-Pro as the signal generator and replaces it with ARTA which is run on a host computer in conjunction with a user-provided DAC. For the purpose of our testing we utilized a Sound Devices USBPre2, although our use of this device is not intended to suggest a recommendation or

requirement to use this device. ARTA was used to generate the signals as well as analyze them in real-time.

As the notes in the figure suggest, the cabling used for the setup will be dependent on the interfaces (e.g. PCI soundcard with digital audio IO, USB, Firewire) found on the computer, the DAC and the ADC.

ARTA could be installed on, and run from, either the user's digital audio workstation or an independent computer. Which selection makes most sense, functionally and budget wise, will be up to each organization. One factor which may drive this decision is the fact that ARTA will only run on a computer running the Windows operating system.

Similarly, an organization may use a dedicated DAC for testing or may use the same DAC that it uses in the day-to-day operations, assuming the functionality and interfacing works correctly, for the DAC in this setup.

With these options in mind the cost of this setup may be as low as the cost of ARTA alone, approximating $100 to $200 USD depending on the version. Some organizations may want or need to buy additional components to complete their test setup as well, but in most cases we believe that test setups can be procured for less than the stated goal $1500, and in many cases we believe that they can be procured for far less.

Considerations and Challenges

There are several aspects of this test setup in its current form that are in need of special consideration and which presented some challenges to us in our work.

- **Single channel functionality**
  ARTA currently only allows testing of one channel at a time. This does not impact the quality of the testing at all, but it does essentially double the time that it takes to perform the testing.

- **Real-time operation**
  ARTA, like most test and measurement systems, is intended to be operated in real-time. The original vision for this work was that the test signals could be generated, routed through the ADC, and then captured as audio files. Following this the files could be loaded into corresponding test and measurement software and analyzed with the push of a button. This would allow a client of a digitization service provider to receive the captured test signal files from the service provider and test the files within the client organization as a form of due diligence and quality control. In the

last phase of work we did establish a method to make this happen, and it is feasible to maintain this method as an option to fulfill this scenario should FADGI find this desirable. However, the one test that we were unable to ever establish a working non-real-time method for was the frequency response test, and this remains problematic. Perhaps in an alternative non-real-time approach the service provider provides a graph and numeric result for the frequency response test and the captured audio files for everything else.

- **Dynamic Range Test**
  ARTA does not have a specific dynamic range test. The manual provides one thought on achieving a measurement for this, stating:

  "If there is no signal at the card input, then RMS shows the input channel S/N ratio."

  In our last round of work we felt that this was not an ideal measure due to the fact that the noise floor of ADCs can behave differently in the presence of signal than they do with no signal input. Therefore we devised an algorithm that subtracts THD+N from THD in order to derive a meaningful result under the presence of signal.

- **Gain Architecture**
  There are two primary places to adjust gain prior to the ADC input. These are within the ARTA user-interface and the DAC output gain. Through testing and analysis we found that the best performance was achieved when ARTA's gain was set to -0.5 dBFS and the DAC output gain was used to achieve the specified level in the test method. This is represented in the user instructions below.

- **Low Level Tests**
  There are two tests that require a signal at -60 dB. We found that this was difficult to achieve given the accuracy of the lower range of the gain potentiometer's throw on multiple DAC devices. We realized that this will often be the case, making it difficult to perform tests requiring the -60 dB level. Therefore we adjusted the -60 dB levels to -40 dB within the test method. While it is not equivalent it is a meaningful reference point, and the consistency and accuracy achieved by the higher level will yield better results over time.

- **Calibration**
  The ARTA manual offers instructions for calibrating the input and output devices that it interfaces with in order to perform signal generation and

analysis. There were a couple of different issues encountered in the calibration of devices.

- o ***Driverless USB devices***

  Many USB ADC and DAC devices are what is referred to as driverless. This means that they are plug-and-play and do not require any additional drivers or software to interface with the computer. The upside to this is the lack of dependency on third party software to have applications interface with the devices. The downside, as we found in this project, is that there is limited ability to configure and control the devices beyond what is made available in the standard Windows interface. This made level setting throughout the signal path very difficult because it is unclear in the standard Windows audio interface control panel, where in the range of 0 – 100 the null value is. In other words, when inputting or outputting audio to and from ADCs and DACs, if you want the internal levels within the Windows operating system to pass signal without adjusting the level in any way it is unclear what the level within the control panel should be set to. By comparison, ADCs and DACs with their own drivers and software provide much more control over signal path within the operating system. Having spent much time investigating how to make this work, we found ourselves with an approach that worked in these instances but with no definitive good answers on how to address this issue.

- o ***Volt meter requirement***

  ARTA's calibration protocol requires use of a voltmeter in order to set references for input and output levels. While not the most difficult piece of equipment to use, the requirement for an additional piece of equipment, the cost of the equipment and the added complexity seems like a potential barrier to adoption of this process.

After grappling with these issues for some time we ultimately decided that the calibration protocol itself is not critical. There are three things that are critical: 1) that the output level of the ADC specified for each test is achieved, 2) that there is no imposed overloading of the signal path, and 3) that the DAC does not clip before the ADC clips. If these three criteria are met then the need to perform the calibration protocol are mitigated and can be bypassed. This doesn't mean that a user does not have to give any consideration to this topic. They will need to make sure that levels can be set accordingly without causing overloading. One area where we ran into issues with this is highlighted in the following challenge.

- **ADC and DAC factory settings and operating levels**
  The most popular market for the purchase of DACs, and to a lesser extent ADCs, is the consumer high fidelity market for use in home environments. The use cases and signal paths for this user are different than those associated with studio and digitization operations. This reality came to light as an issue when using an ADC and DAC from a manufacturer which shipped their devices configured for the home market. The operating levels of the devices were not suitable for professional digitization operations. In order to achieve the appropriate operating level we needed to open up the devices and configure the dip switch settings internally. Without doing this we were unable to achieve the appropriate levels on the ADC without clipping the DAC output. This type of issue is potentially tricky for some users who are not in the practice of performing calibration of their systems. The best possible way to mitigate the risk of this happening is through raising awareness of this as a possible issue in the supporting information accompanying the test method and metrics.

- **Using DACs as part of signal generator signal path**
  One of the main challenges that we grappled with in the previous phase of work when considering the sole use of ARTA was the fact that it requires the use of a DAC to generate signals. Coupled with the fact that it seems unreasonable to recommend a specific DAC for this purpose, the realities presented lead to variable DACs being used, each with their own performance limitations. This means that the results will encompass both the performance limitations of the DAC and the ADC. With the shift in thinking, away from nuanced assessment of performance and toward identification of outright failure, this becomes less of an issue. However we still felt the need to provide performance metrics to use as a baseline and these became difficult to set with the known variability in mind. In an instance where a user tests an ADC with their system and receives a failure, following a pattern of routine testing the same ADC with the same system and passing, the results are clear and there are no issues. If a user sets up their system for the first time and receives failures from test number 1 this does place a burden on them to investigate further. They will need to determine if: 1) there is a malfunction in the ADC and/or DAC individually or cumulatively, or 2) the ADC and/or DAC are legitimately of such low quality that they fail to meet the performance specifications identified individually or cumulatively.

  ARTA does provide a mechanism for mitigating this issue in some ways and to some extent, referred to as frequency compensation. This is employed by first measuring the frequency response of the DAC and then applying an inverse curve to the DAC output, essentially achieving a "flat"

output with regard to frequency response. This is a useful feature of ARTA but it requires a test and measurement device capable of testing the DAC in order to plot the inverse frequency compensation curve. While it is feasible that this could be done through procured or borrowed equipment, or through a third party with such equipment we anticipate that use of the frequency compensation feature will be limited due to the added complexity and effort.

While the above points are worthy of consideration we believe that this test method and system is sufficient for fulfilling the stated purpose.

- **Difference between Peak and RMS levels on the IMD tests**
  We used the additional peak meter for the IMD tests. However there is a pattern that when using the previous test at -1dBFS output for a less complex signal, the IMD tests are also really at -1dBFS even though non-true-peak meters report otherwise. So the conclusion was to avoid adjusting levels further after setting the level for the test preceding the IMD tests.

- **Metering**
  Within the circle of manufacturers of audio hardware and software, the inconsistency of metering across products is a well-known problem. We encountered this in the low-cost test with the intermodulation distortion (IMD) tests. Each of these tests require a twin tone signal that should equal -1dBFS at the output of the ADC. What we found through investigation and a variety of metering tools is that with this complex signal, ARTA's meters and other software and hardware meters inaccurately reported the output level of the ADC. Whereas ARTA reported the output at -2.7dBFS, a true peak digital meter constructed by Phillip Sztenderowicz indicated that the output was in fact -1dBFS.

  However, what we found was that the level setting in the test that precedes the IMD tests (Crosstalk) requires a -1dBFS output at the ADC using a series of simple tones. When leaving the levels set according to this test the IMD twin tone signal output is also at -1dBFS. Therefore the levels can be left untouched following the Crosstalk test and the IMD tests will be performed with the correct levels.

User Instructions

The guideline and performance metrics are addressed in other documents within this deliverable. The text below serves as an aid to users in putting the test method to use and documenting the results.

General Setup

> ***Initial Setup***
> ● Open the test method document (*ADC_Low_Cost_performGuide_2016-09-01)*. Follow the instructions within the document and use for each test a point of reference.
>
> ● For documenting and reporting, utilize the Excel spreadsheet named *Low_Cost_Test_Template* and begin by filling out the information under the sections labeled "General Information", "Digital to Analog Converter" and "Equipment Under Test". Note that there are multiple tabs. When performing a test, navigate to the tab for the test being performed and populate the cells highlighted in green. These results will automatically populate the Summary tab.
>
> ● Ensure that your hardware and software are configured so that:
>   1) there is no overloading of the signal path imposed by improper configuration or settings
>   2) the DAC does not clip before the ADC clips
>
> ● Ensure that all of the OS settings for your audio devices are operating at the correct sample rate and bit depth.
>
> ● Set the ADC clock to internal and the correct sample rate
>
>
> ***Open and Configure Steps***
> a. Open the Steps program that comes with ARTA
>
> b. Within the *Setup -> Measurements* configure the following parameters:
>
>   Single Channel - Level
>   Response Channel: select the channel under test
>   Sampling Frequency: Set accordingly
>   Min Integration time: 8
>   Transient time: 4
>   I/O Delay: 0
>   Intra burst Pause: 4
>   Start Frequency: 20
>   Stop Frequency: 20000
>   Frequency: 1/12 octave
>   Generator Level: -0.5

Test Frequency: 1000
Mute Generator Switch-off Transients checked

Select "Generate" button and adjust the DAC gain until the
meter reads -20

c. If Frequency Compensation is used within ARTA for the DAC being
used be sure to document this in the notes section of the report.

d. Within *Setup -> Graph* configure the following parameters:
Magnitude Range
Graph Top: -18
Graph Range: 4

Frequency Range
High: 22000
Low: 10

Frequency Response
a. Within the main Steps window start the measurement by selecting the
red triangle button.

b. Once complete, on the graph find and note the value at 1 kHz (note
that the Steps scale is in dBV and not dBFS. This will not impact the
test).

c. On the graph find and note the largest variance from 1 kHz between 20
Hz and 20 kHz.

d. Calculate and document the difference between the values in dB.

e. Within the *Overlay* menu select "Set as Overlay"

f. Run the test again for other channels being sure to select the
appropriate channel under *Setup -> Measurement.*

g. Within the *Edit* menu select "Copy" and within the dialog window that
appears select "Save to File" and save the graph to input into the
results report.

Open and configure ARTA
a. Open ARTA and select SPA

b. Within *Setup -> Measurement* configure the following parameters

        Averaging: Linear
        Max Averages: 4096
        FFT Size: 32768
        Window: Kaiser7
        Sampling Rate: set accordingly

c. Within *Setup -> Spectrum Scaling* configure the following parameters
        Scaling: dBFS
        Voltage units: dBu
        Pressure: dB re 20 uPa
        Distortion
                check the following:
                        THD
                        THD+N
                        IMD
                        Multitone TD+N
        Normalize with full power: unchecked
        Low cut-off: 20
        2nd and 3rd Order IMD: checked
        Frequency weighting: unchecked

THD+N

a. Within *Generator-> Configure*, under the Sine/square generator section

        Set the frequency to the specified frequency

b. Adjust the DAC gain until ARTA shows the specified level at the ADC output

c. In the main ARTA window run the test for each channel at each frequency and level combination

d. Document the THD+N results displayed in ARTA

e. Repeat the test and document the results for each channel being sure to select the appropriate channel under *Setup -> Measurement*

Dynamic Range

a. Generating 997 Hz at -40 dBFS (the last tone and level generated in the THD+N test), document both the THD and THD+N values.

b. Repeat for each channel being sure to select the appropriate channel under *Setup -> Measurement*

Crosstalk

a. Within *Generator-> Configure* , under the Sine/square generator section

      Set the frequency to the specified frequency

b. Adjust the DAC gain until ARTA shows -1 dBFS at the ADC output

c. Place a shorting plug into the channel under test (the not the channel being driven with the signal generator)

d. Measure the value at the specified frequency on the channel under test (the shorted channel) by placing the cursor at the frequency, adding 1dB and documenting the value.

e. Repeat and report for each frequency on each channel being sure to select the appropriate channel under *Setup -> Measurement*

IMD LF

a. No gain adjustments will be made for this test following the Crosstalk test

b. Within *Generator -> Configure*

      Under the Two sine generator section

            Select "User" and populate Freq1 with 41, Freq2 with 7993 and Magn with 4

c. Select "Two sine" in the *Gen* pull-down menu in the Spectrum Analyzer window

d. Run the test by selecting the red triangle button

e. Document the SMPTE IMD measurement reported by ARTA

f. Repeat the test for each channel being sure to select the appropriate channel under *Setup -> Measurement*

IMD HF

a. No gain adjustments will be made for this test following the IMD LF test

b. Maintain "Two sine" in the *Gen* pull-down menu

c. Within *Generator -> Configure*, under the Two sine generator, User section

   populate Freq1 with 18000, Freq2 with 20000 and Magn with 1

d. Run the test by selecting the red triangle button

e. Document the DFD2 and DFD3 measurements reported by ARTA

f. Repeat the test for each channel being sure to select the appropriate channel under *Setup -> Measurement*

Spurious Inharmonic

a. No gain adjustment is needed following IMD HF

b. Select "Sine" in the *Gen* dropdown menu of the main ARTA window

c. Within *Generator -> Configure*

   Set Frequency to 997 Hz

d. Within *Setup -> Measurement* set the following parameters
   Type: Linear
   Max averages: 32

e. Within the main ARTA window select the red triangle button to run the test. After 32 averages the test will stop.

f. Cursor along the graph finding the most significant inharmonic (not a multiple of the stimulus) spurious component outside of the stimulus skirt (the immediate area surrounding the stimulus).

g. Document the RMS level reported by ARTA

h. Within the *Overlay* menu select "Set as Overlay"

i. Repeat for each channel being sure to select the appropriate channel under *Setup -> Measurement*

j. Within the *Edit* menu select "Copy" and within the dialog window that appears select "Save to File" and save the graph to input into the results report.

Conclusions and Next Steps

Performing repeated tests using the test method and user instructions followed above we devised a revised set of performance specifications. It bears repeating that these performance specifications are more about identifying failure in ADCs than they are about nuanced evaluation. Because of this the performance metrics are looser in order to provide some leeway for lesser performing devices. Having said this, these performance metrics will identify devices that perform poorly enough to be deemed unusable for preservation purposes. The revised test method and performance specifications devised as a result of testing and evaluation in this phase of work can be seen in *ADC_Low_Cost_performGuide_2016-09-01.* The accompanying instructions on how to use this test setup to implement the test method, yielding appropriate results can be seen in *ADC_Low_Cost_User_Instructions.* A spreadsheet for documenting raw results that contains embedded calculations for deriving results conforming to the test method and performance specifications and a summary report can be seen in *Low-Budget_Test_Template.*

Performing this work provided some insights into welcomed improvements and potential activities for next steps. Specifically these include:

1. Automatically perform the test method and sequence documented in *ADC_Low_Cost_performGuide_2016-09-01*
Performing these tests manually is a time consuming process that will be a barrier to them being adopted or at least performed routinely. Now that the test method and calculations are established, turning this into an automated process

would be an excellent and worthwhile next step. Of course, this leads to the question of how the results would be documented, leading into the next point.

2. Automatically perform the calculations and generate the report represented by *Low-Budget_Test_Template*.
Related to the automatic performance of tests, the automatic generation of a report resembling a populated version of the Low Budget Test Template would be extremely advantageous.

3. Measure 2 channels at the same time
Currently ARTA only tests one channel at a time. This more than doubles the time it would take if two channels were able to be tested at the same time. Whether using a manual or automated test protocol, this feature would add a great deal of efficiency.

4. Make calibration easier
The calibration process is currently complex enough as to discourage users from using it. As discussed earlier, while it's not critical to perform calibration in order to perform these tests, it is best practice and could avoid some missteps. Making the calibration process easier could range from altering the methodology used to simply providing more guidance, information and instruction.

5. Provide accurate metering for the IMD tests.
As discussed in the section on challenges and considerations, the twin tone signals in the IMD tests are reported incorrectly using the ARTA meters. A true -1 dBFS peak signal reads as -2.7 dBFS in ARTA. While the current user instructions navigate around this issue, it would be best to have the meters accurately report the level for these tests.

6. Steps scale and frequency response measurement
Within ARTA's Steps program there is currently no option to have the scale use dBFS. While not critical, it would be ideal to be able to have an option for the scale in Steps to report dBFS.