Testing Smart Devices

This article addresses the challenges and solutions for testing the wider class of "smart" connected, voice-activated and sensor-equipped devices, proposing solutions to implement basic acoustic tests and real-world tests of greater complexity.

By
Steve Temme
(Listen, Inc.)

Smart devices such as smart speakers, hearables, wearables, robots and voice-controlled car entertainment systems are notoriously complex to test. They have numerous interfaces from Bluetooth to cloud-based/Wi-Fi, and contain much signal processing, both on the record side (e.g., beamforming and background noise filtering) and on the playback side (e.g., loudness and active noise cancellation). This means that their characteristics change according to "real-world" conditions such as the physical environment and background noise. Furthermore, their multifunctional nature means there are many aspects of the device that may need to be tested, ranging from voice recognition to music playback or even operation as a hands-free telephone. Due to their complex nonlinear use cases, these devices often need to be tested at different levels and different environmental conditions (e.g., different physical setups and background noise, different signals etc.).

Although as yet there are no standards for testing smart devices, we can borrow principles and test configurations from many other audio devices and use existing standards such as the European Telecommunications Standards Institute (ETSI) and The Institute of Electrical and Electronics Engineers (IEEE) background noise and telephone test standards. Flexibility of the test system, and experience with testing a wide range of acoustic devices is critical to enable a device to be completely characterized. In this article, we will discuss how to implement basic acoustic tests and then discuss some of the more complex real-world tests that may be carried out on smart speakers and other smart devices, along with the techniques and standards that may be used. Finally, we present a check list of the test-system functionality you should look for when choosing a system to fully characterize a smart speaker or other smart device.

Basic Acoustic Tests

Basic acoustic tests essentially measure the device under static conditions to measure microphone array and speaker characteristics (e.g., frequency response and distortion). While the measurements themselves are similar to standard loudspeaker and microphone measurements, injecting the stimulus and extracting the recorded signal present challenges. Unless you happen to be the device's manufacturer and have access via a debugger or other remote access tool, or are testing speaker Bluetooth performance only, you have a true black box test solution. This is generally the case when comparing the performance of different speakers, for example, for reviewing or competitive benchmarking. Without knowing what is inside the device under test (DUT), you need to get signals in and out of it, and the only way to do this is like an actual user—via the cloud using voice commands.

To further complicate matters, the test signal needs to be in the cloud to enable playback. Each manufacturer's ecosystem is different in how it plays back from the cloud. Some enable you to upload your own recordings (although you should bear in mind that these will probably be compressed). Others require them to be on a media streaming platform (e.g., Spotify). For microphone testing, some systems such as Alexa enable you to access recordings you have made; others do not for security and privacy reasons, which makes microphone

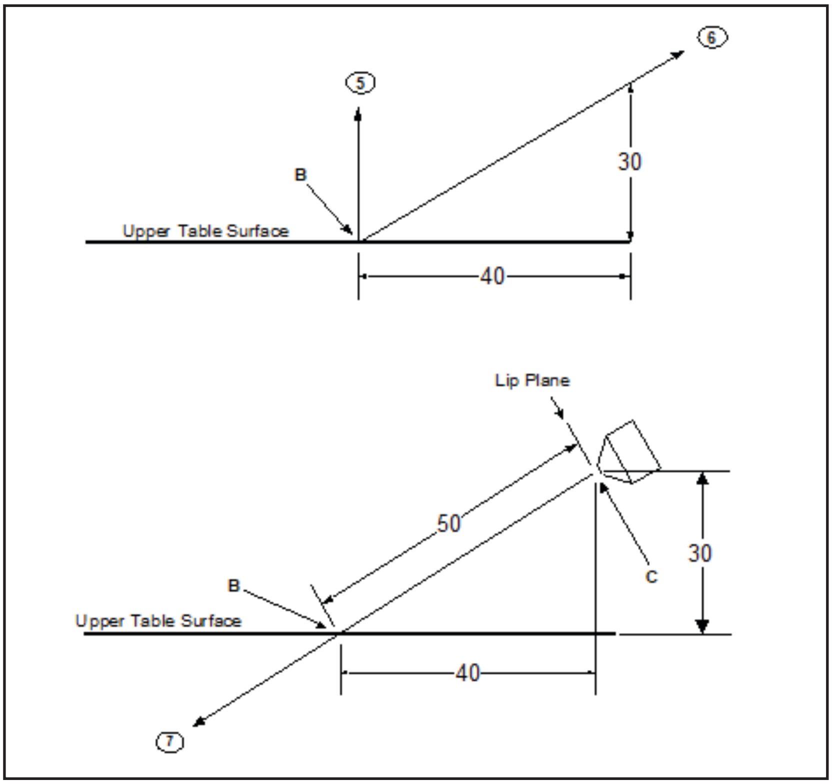

Figure 1. This is the IEEE 1329-2010 standard test setup with a table (dimensions in centimeters).

testing challenging. Although the actual physical testing setup is very similar from device to device, for each one it is necessary to understand out how to wirelessly route the signal. Furthermore, each device needs activating with a different wake word, has a different delay compensation, and records for a different amount of time after it hears the wake word. This, therefore, needs figuring out (largely by trial and error) for each speaker that you need to test.

Let us first discuss the physical setup. At this point there are no test standards specifically written for smart speakers. However, their practical use case closely resembles speakerphones, so the physical geometry recommended by IEEE 1329-2010 is a good starting point. Following this standard, tests would be conducted inside an anechoic chamber with the DUT placed on a table (see **Figure 1**).

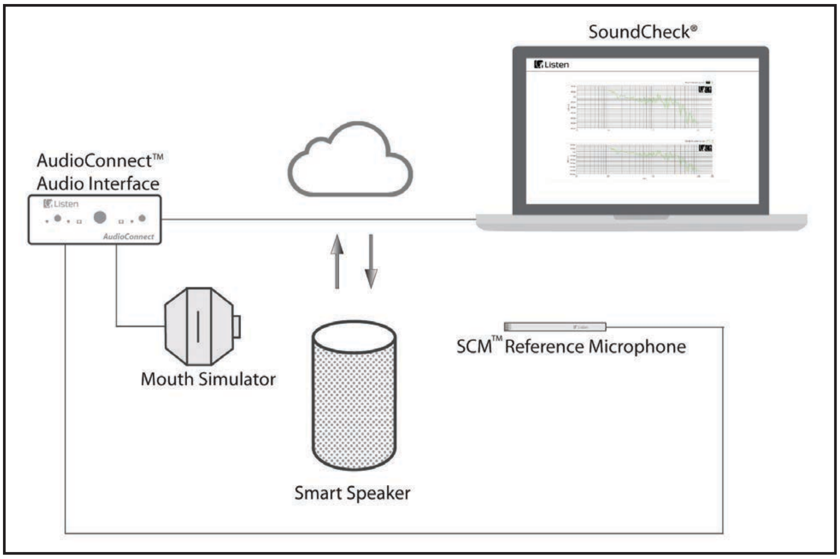

Figure 2: This is a viable test configuration for the basic loudspeaker measurement of a smart speaker.

The setup includes a speaker or artificial mouth to play the voice activation commands, a reference microphone to record the signal, and of course, the DUT (see **Figure 2**).

Either a mouth simulator or loudspeaker may be used to play the voice commands, but generally loudspeakers are preferred—measurements are made at a distance to represent a voice command and loudspeakers typically can play louder and are easier to equalize as they have a flatter frequency response.

Let's now discuss the steps necessary to create a test. It is ideal to automate the test as much as possible, so wake words and test stimuli should be pre-recorded and uploaded so that they can be triggered by the test system.

First, we need to record the activation signal. This might be something like "Alexa, play test signal 1." Naturally the system's appropriate wake word should be used. If a wide range of speakers from different manufacturers is being tested, it makes sense to record the wake words separately from the command so that any combination can be selected.

The test stimulus also needs to be created and uploaded to the cloud. Bear in mind that it might be compressed when it is uploaded, which could introduce some distortion. You can verify that the MP3 encoding itself does not degrade the test signal by encoding and decoding the test signal, and using transfer function analysis between the original signal and the encoded version. We'll go into test signals in more depth later, but for basic speaker performance testing we tend to use a standard Stweep (stepped swept sine sweep).

Once the stimuli and voice commands are created and uploaded, the test sequence simply needs to play the wake word followed by the activation command via the speaker and capture the response with the reference microphone. Stimulus and analysis are then compared as in a traditional measurement to objectively qualify characteristics such as frequency response and distortion.

However, it is not as simple as it looks—although the good news is that with the right test system, the complicated analysis happens "under the hood" so there is very little that the user needs to do other than make sure their test system has been told to analyze it correctly.

One of the main challenges that test system manufacturers have faced in developing accurate smart speaker tests, is overcoming sampling rate error. Traditional measurement systems rely on the DUT having a synchronous input and output. As a category, smart speakers are intrinsically open loop, which means that although the device can record or playback a signal there is no synchronous signal path. This introduces the possibility for sampling rate error.

In other words, the device may record a signal to a file with a sample rate of 44.1 kHz, but it may have been recorded at a slightly different sample rate (e.g., 44.09 kHz) due to skew in the actual rate of the crystal clock used to drive the sampler. A similar error can occur when playing back test signals. That is, the test stimulus may be sampled at 44.1 kHz, but due to error in the playback sample rate the file is actually played back at a slightly faster or slower rate. This sampling rate error results in the component tones of the test stimulus being shifted to either a higher or lower frequency. This shift can then lead to measurement errors due to loss of coherence between the stimulus and response signals.

To overcome this sampling rate error, an algorithm is applied, which searches the beginning of the response waveform for a "trigger" that is used to provide a reference point for alignment and shifting of the stimulus and response signal. This trigger may be a steady-state sinusoid at a pre-set level and frequency, or for more robust performance that is less susceptible to false triggers, a log chirp (SoundCheck offers both). The signal is then shifted to DC using a heterodyne filter and all other frequencies are filtered out. The output of the heterodyne filter includes the phase information, which is ultimately used to estimate actual playback or recording sample rate of the response signal. With this information, the entire response waveform is resampled to the correct stimulus sample rate prior to analysis. This frequency shift step corrects for sampling rate error in the device and makes testing them straightforward.

With the test system taking care of all these, frequency output and distortion can be displayed in exactly the same way as for a wired speaker, or a speaker connected via a Bluetooth interface or directly via a debug port (see **Figure 3** and **Figure 4**).

Of course, it is also necessary to measure the microphone array in the speakers. The setup for this is very similar, except that in this case, we need to play the wake word and the test stimulus through the mouth simulator, then download the test stimulus from the cloud and route it back into the test system using a virtual audio cable (see **Figure 5**). Once the signal has been brought into the test system this way, as with the speaker test, the appropriate calculations are applied to overcome frequency shift and error, and it can be analyzed as normal (see **Figure 6 and Figure 7**).

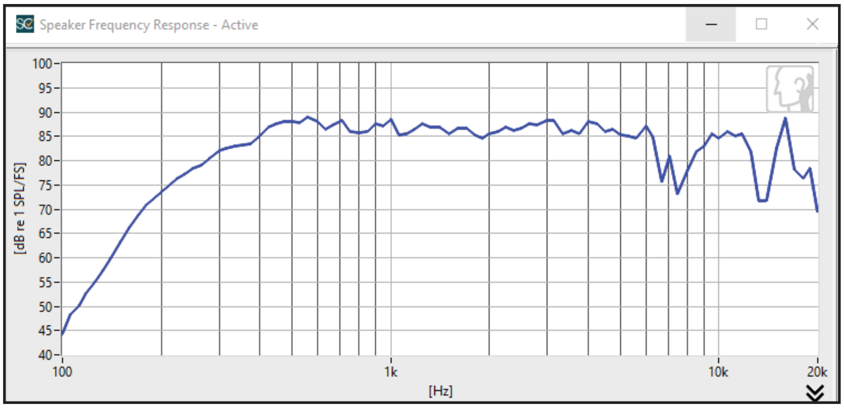

Figure 3: This graph shows the frequency response of a smart speaker.

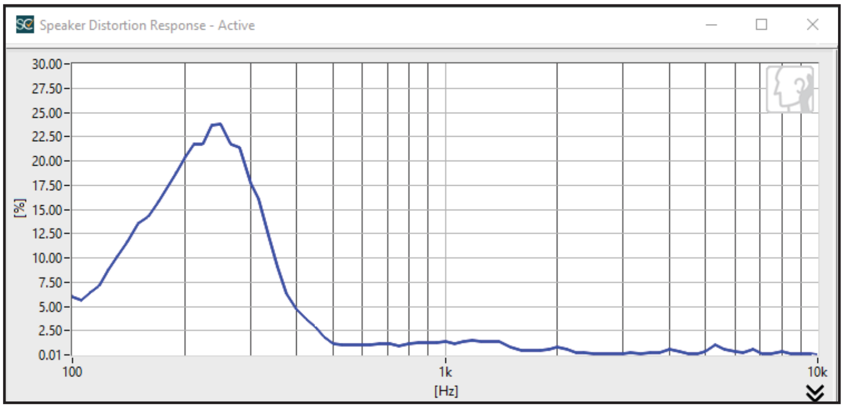

Figure 4: This graph shows the distortion response of a smart speaker.

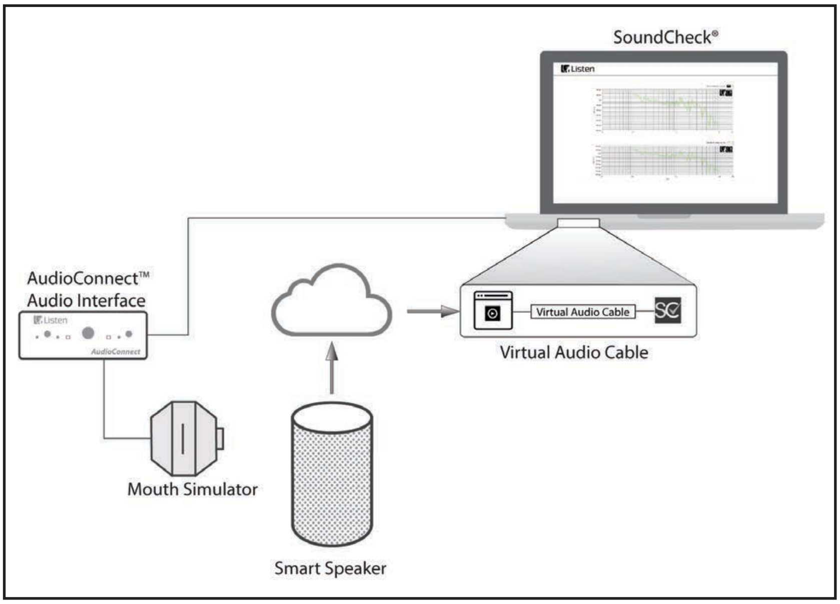

Figure 5: Here is a viable test configuration for a basic microphone measurement of a smart speaker.

Advanced, or Real-World Testing of Smart Devices

While the basic speaker test previously described is ideal for measuring frequency response and distortion, it barely scratches the surface in terms of the breadth and depth of tests that one might want to carry out on a smart device in the R&D lab. Consider that smart devices are generally nonlinear and their response probably varies according to where it is in the room, what the background noise is, how loud the background noise is, whether it is being used to play music or to communicate voice, and more. These considerations alone bring up a multitude of tests that could yield valuable information about the device's performance. While the specific tests that are carried out will vary according to the manufacturer's design objectives, we can discuss some of the test methodologies that we might use and some of the industry standards that provide a useful starting point.

A significant difference from basic tests is that advanced tests generally require the use of speech or music as test signals. This means a test system that can support such test signals is essential, and it is also important that the system can measure the

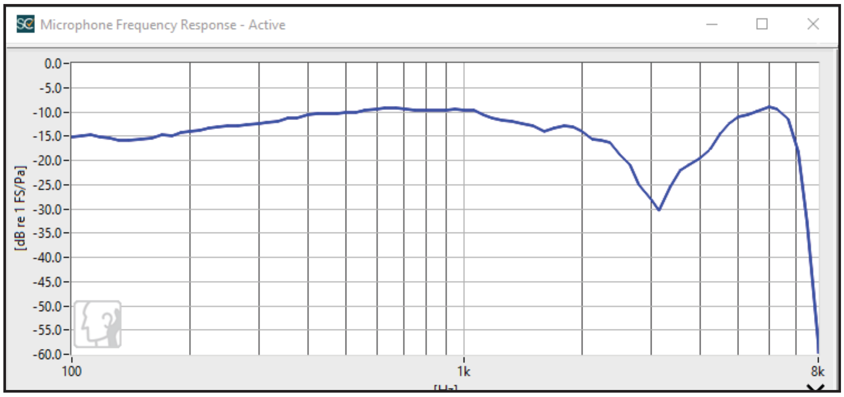

Figure 6: This graph shows the frequency response of a smart speaker microphone array.

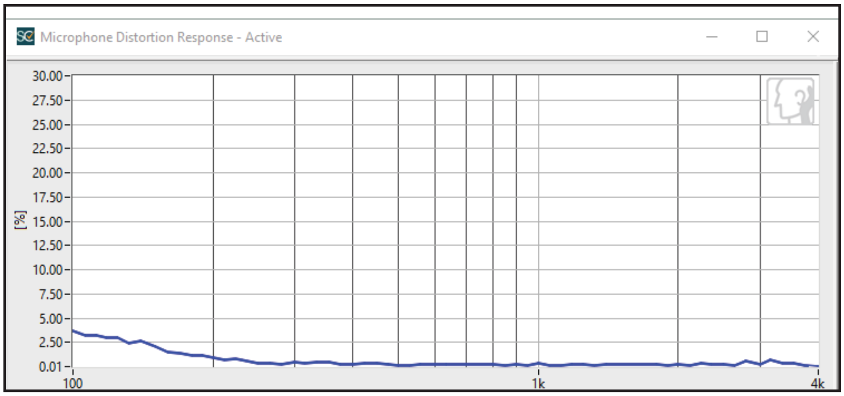

Figure 7: This graph shows the distortion response of a smart speaker microphone array.

active speech level of speech-based signals to ensure calibration and standardization as commonly used in telephony measurements.

Voice Recognition

Speech recognition tests are important for evaluating the device's ability to understand voice commands. The physical setup is the same as in the basic test outlined above.

One option is to create a series of specially titled musical tracks and upload them to the cloud. The track titles are the words or phrases that you want to test, and the actual content of the tracks is a single, dual, or multitone that enables you to identify the track by its audio content. Generally, "Harvard Sentences"—a collection of phonetically balanced sentences that use specific phonemes at the same frequency they appear in English—are used for the track titles. These are recommended by "IEEE Recommended Practices for Speech Quality Measurements for Voip, Cellular, and other telephone systems" and widely used in telecoms voice recognition. Pre-recorded versions of these can be downloaded online, or your own can be recorded (e.g., if you are testing the device with different accents). These track titles are used to request playbacks, and a limit is used to detect whether the correct, or indeed any, signal was played back.

There are many variations on this test, including testing different voices and accents, testing different speaker locations, testing an active talker versus an interfering talker or background noise levels, testing with different types of background noise, and more. Where background noise is needed, ETSI standard ES 202 396-1 contains binaural recordings made in different noise environments (e.g., cafe noise, traffic noise, subway etc.). While there are many ways of playing these (e.g., through purchasing an expensive dedicated system or by programming your own), SoundCheck offers an optional ETSI standard ES 202 396-1 module. This integrates the ETSI standard library with the test system, playing the background noise at calibrated levels that can be controlled and adjusted as part of a pre-programmed test sequence. This test sequence may include loops to incrementally increase the volume or change the noise and repeat the test, making it simple to create a test that increases the background noise by fixed levels until the voice is no longer accurately recognized. In addition to being a very cost-effective approach, it also significantly simplifies the physical test setup and reduces test development time. Background noise is necessary for a wide range of smart device tests including voice recognition, signal-to-noise ratio (SNR) optimization of mic arrays, beamforming directionality studies, and more.

Indy Acoustic Research

Independent acoustic design and test lab Indy Acoustic Research, LLC (IAR) has developed extensive expertise in testing smart devices such as voice-activated speakers. With a solid background in speakerphone, loudspeaker, and microphone tests, IAR's team has a thorough understanding of how to test the various elements of smart devices. In addition to contracted testing of smart devices, IAR has also invested in developing new test methods for characterizing smart speakers.

Smart speakers are evaluated in an anechoic chamber using SoundCheck and a test setup very similar to that described in the article. The device under test (DUT) is placed on a suspended table, a calibrated Brüel and Kjäer mouth simulator is used to generate the stimulus for the microphone tests, and a speaker monitor is used to generate white, pink, and ETSI noise to enable measurement of performance with background noise and interfering talkers.

Basic performance tests on smart speakers include frequency response, sensitivity (output versus input), total harmonic distortion (THD), and total harmonic distortion plus noise (THD+N). The calculations used to create these graphs are carried out within SoundCheck, so once the test signal is captured and routed into SoundCheck, there is no additional operator intervention. The effect of background noise interference is also quantified by adding interfering noise played via a speaker monitor in the room. Background noise is chosen from a range of options—ETSI noise, pub noise, office environment, or noisy children—and the noise level is selected. Measurements are repeated, and using a post-processing step in SoundCheck, a difference response is created based on the baseline and noise-degraded response.

All data is exported to an Excel file containing the raw numerical data, graphical results and testing comments. Retaining all test data in one place helps with any future re-analysis or comparison. For example, if it is necessary to compare a new, redesigned model to an older one it is easy to run the tests for the new model and compare it to identical saved data from the older model. Storing all data including wav files even makes it possible to listen to the comparative recording quality as well as compare measured parameters.

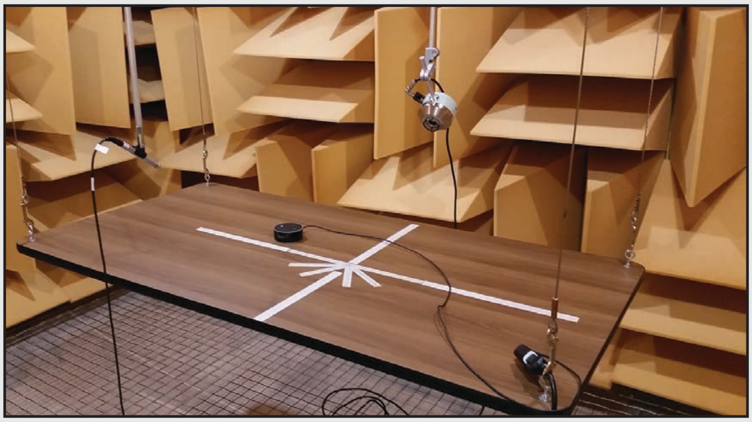

Photo 1: A smart speaker test is being conducted in an anechoic chamber at Indy Acoustic Research.

Voice Quality

Smart speakers often function as handsfree communication devices. For such use cases, we can lean heavily on telephone test sequences such as the Telecommunications Industry Association (TIA) standards for handsfree devices (TIA920-B). In fact, a smart speaker can be tested as a speakerphone exactly as defined in the TIA standard. Although the test setup is identical to the basic smart speaker test previously outlined, the test sequences for measurement to the standard are highly complex and either require considerable expertise in this area to create them, or they need to be purchased as an off-the-shelf package.

Choosing a Test System

Smart devices represent the convergence of tests from speakers, surround systems, microphones and recording systems, telephones, hearing aids, MP3 players and more. It is, therefore, advantageous to select a system from a manufacturer that truly understands these varying applications and has the various test stimuli, algorithms and post-processing functionality for these in their toolkit, along with the expertise. A checklist of the test system functionality that is important for characterizing smart devices should include:

- Tests both speakers and microphones
- Can create a compound stimulus with frequency and/or log chirp trigger to enable accurate open-loop testing
- Can be set to analyze only parts of the response signal, so that the trigger tone can be eliminated from the analysis
- Ability to use speech and music as test signals, calibrate levels, and equalize a mouth simulator (requires the inclusion of active speech algorithms)
- Background noise generation—although this can be generated externally, testing is much simpler and faster if it can be created by the test system and integrated with the test sequence
- Can accept signal via virtual audio cable in order to route in signals from the cloud through the computer's multimedia interface (e.g., WDM or Core audio)
- Availability of pre-written TIA test sequences if device needs to be tested as speakerphone (may be manually programmed with sufficient expertise)
- Sophisticated sequence writer that allows duplication of sub-sequences and loops - essential when repeating a test at various levels
- Bluetooth interface available, if testing over Bluetooth is desired
- Mentor A2B interface compatibility and control from within test system if testing automotive smart devices (e.g., in-car voice control)

It is a bonus if the test system is also available in a lower-cost, lighter version so that basic tests can be simply transferred to the production line rather than being recreated in a new system.