Practical Test & Measurement

Smarter Measurements for Smart Speakers

In this article, we describe techniques to characterize the frequency response, output level, and distortion of the device under test to enable direct comparisons between Internet of Things (IoT) smart speakers and conventional speakers.

By
Daniel Knighten and Glenn Hess
(Listen, Inc.) and (Indy Acoustic Research)

Since the launch of the Amazon Echo in 2015, the adoption of smart speakers has been perhaps the fastest in the history of consumer electronics, with shipments predicted to exceed 100 million units by 2024. While some manufacturers, such as Amazon and Apple, produce devices to work exclusively with their infrastructure, others such as Sonos are building smart speakers that will work with any voice control system. Although smart speakers have only been around for a relatively short period of time, players such as Apple, Sonos, and Harman are now active in this marketplace. Audio quality is a strong decision-making factor for customers and a key selling point for manufacturers, with each manufacturer incorporating its own unique speaker design and positioning for optimum sound quality. Design and improvement of far-field microphone response, which affects the speaker's ability to understand in a noisy environment, is another area where manufacturers are seeking to differentiate themselves with audio design advances.

With both sound and listening quality at the forefront of designers' minds, robust audio testing of both the speaker and the microphone array is essential. This presents new challenges in performing objective measurements as there is typically no direct path to inject or extract response signals from the embedded speakers and microphones. Rather than playing back and recording signals to local storage media, they store and playback signals from Internet-based cloud services. Furthermore, the use of voice recognition for control makes activating these devices challenging. These characteristics necessitate various additional steps to be taken in the testing process to ensure accurate and meaningful measurements.

Here, we discuss how to provide an objective evaluation of a smart speaker's audio performance by describing techniques to characterize the frequency response, output level, and distortion of the device under test (DUT) to enable direct comparison between Internet of Things (IoT) smart speakers and conventional speakers. We also discuss the selection of test signals, and the challenges around injecting and extracting signals from the device. As an example, we specifically demonstrate how to measure basic speaker and microphone audio parameters of an Amazon Echo Dot. The techniques described are applicable to any smart speaker or IoT device with an audio component.

Challenges

When measuring conventional speakers and microphones, the DUT is usually directly stimulated with a sinusoidal test signal swept across a frequency band. Measurements, such as frequency

response and distortion, are then derived from the response.

The fundamental challenges in testing the speakers and microphones in a smart speaker are:

- There is no direct output path for the microphone(s) or input path for the speaker. The devices record and playback from the "cloud," an Internet-based service of some kind (e.g., Amazon Music).
- Smart speakers are typically voice activated and have minimal or no manual controls to facilitate testing.
- Smart speakers are non-synchronous audio devices, therefore subject to sampling rate error or mismatch between the device and the measurement system.

Retrieving the Response Signal

In the case of Amazon Alexa-based devices, while we are unable to directly extract the microphone signal, a recording of everything captured by the device is available from a website. Trial and error was used to discover that Alexa devices record for about 8 seconds after the activation word is detected. A test signal was constructed that swept over the frequency range of interest within this window. For other brands of speakers, a method of accessing the recorded signal must be discovered, and the time for which it records ascertained to ensure that the test signal is fully included.

During the test, a composite signal consisting of a voice recording of the activation word followed by a stepped sine sweep is used. After capture of the recording by the device, the audio recording is downloaded from the associated website. This captured recording can then be analyzed to determine the microphone's characteristics.

Storing and Playing Back the Stimulus

To test the speaker function, a test signal can be uploaded to the account with the associated Internet music service as if it was a musical recording. The DUT can then be commanded to playback the test signal. From there, the playback of the stored test signal is captured by the measurement system. Specifically in the case of Amazon Music, the stored audio signal had to be encoded in the MP3 format. We verified that the MP3 encoding itself was not degrading the test signal by encoding and decoding the test signal, and using the transfer function analysis between the original signal and the encoded and then the decoded version.

Activating Smart Speakers

In general, smart speakers are always listening with their internal microphone(s) active. They respond to an activation word or wake-up phrase followed by a command string. To test the audio performance of such a speaker, the device must first be activated, and then a suitable test signal applied to keep the smart speaker active and operating in a normal state (see **Figure 1**).

Each smart speaker platform has its own activation phrase, and trial and error is needed to establish how long the device records, and then to get it to playback a specific signal. Once these parameters are established, the device can be made to either play back the desired stimulus for speaker testing or record a signal for microphone testing. When the response is analyzed, it is parsed so that the activation phrase is ignored.

Overcoming Sampling Rate Error

Traditional measurement systems rely on the

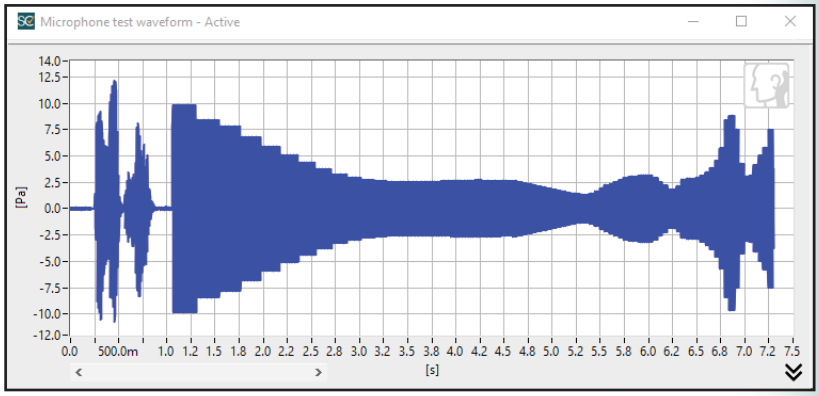

Figure 1: Activation word and test signal waveform for microphone measurement

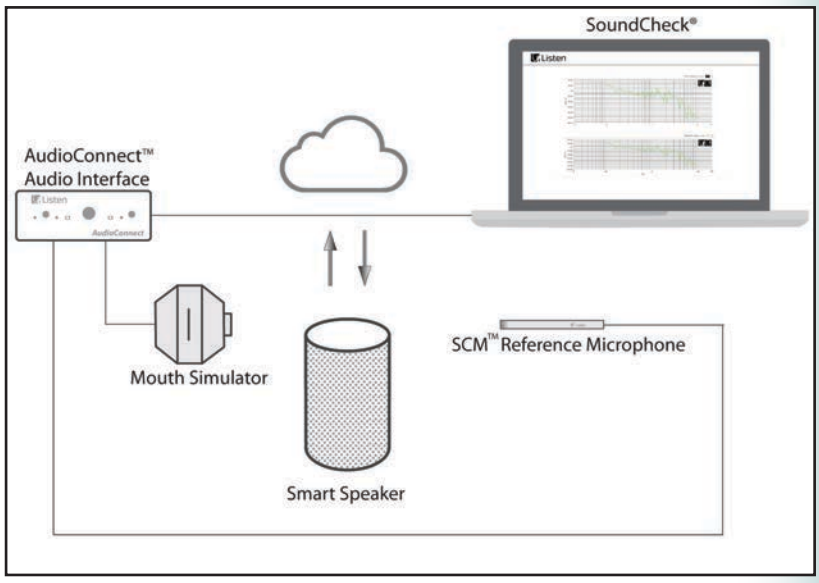

Figure 2: Measurement setup for testing the loudspeaker

DUT having a synchronous input and output. As a category, smart speakers are intrinsically open loop, which means that there is no synchronous input and output. Instead as described above, the device can record a signal or playback a signal but it does not have a synchronous signal path.

This introduces the possibility for sampling rate error. That is, the device may record a signal to a file with a sample rate of 44.1 kHz, but it may have been recorded at 44.09 kHz or another similar rate due to skew in the actual rate of the clock crystal 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.

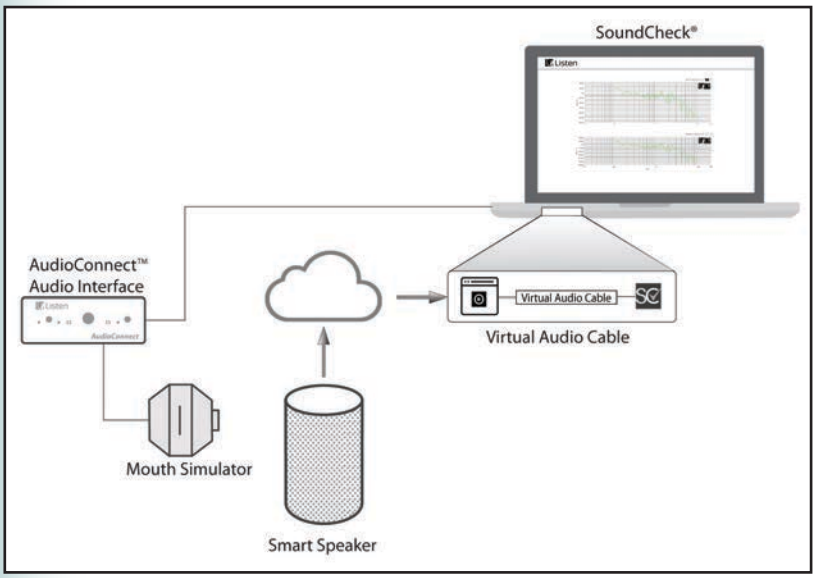

Figure 3: Measurement setup for testing the microphone array

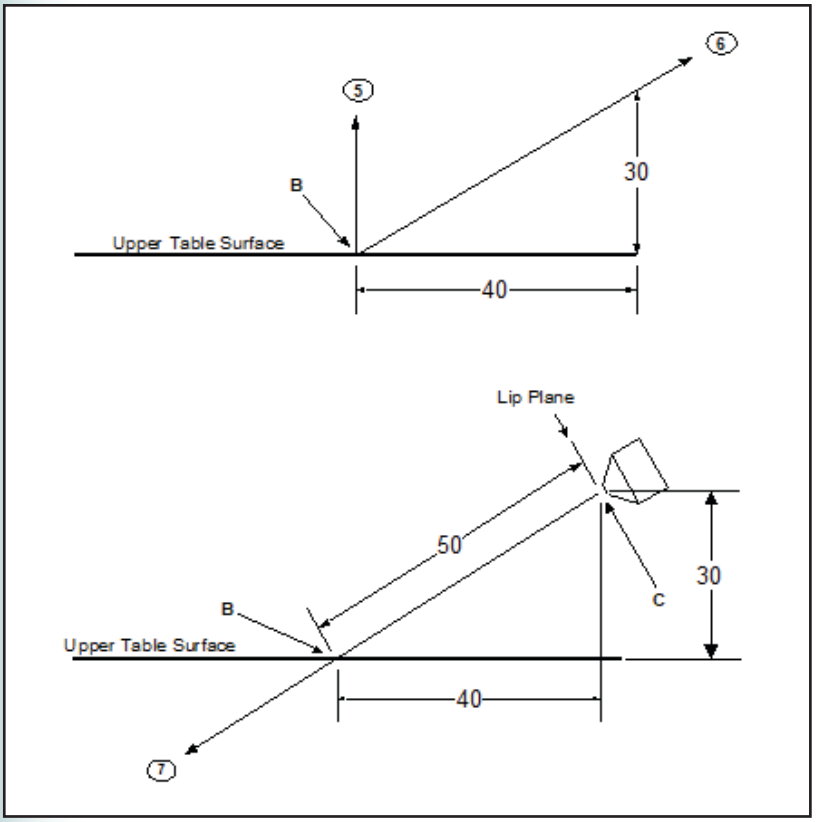

Figure 4: The IEEE 1329-2010 test setup with table (dimensions shown in cm)

This sampling rate error will result 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 the response signals.

To overcome this sampling rate error, an algorithm is applied, which searches the beginning of the response waveform for a steady-state sinusoid at a preset frequency. 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.

In our example test, the actual signal used to measure the frequency response and the distortion

Resources

K. Brandenburg, "MP3 and AAC Explained," presented at Audio Engineering Society (AES) 17th International Conference on High-Quality Audio Coding, August 1999.

G. Hess and D. Knighten, "The Challenges of IoT Smart Speaker Testing," presented at the Audio Engineering Society (AES) 143rd Convention, October 2017.

IEC 61094-4, "Measurement Microphones Part 4: Specifications for Working Standard Microphones," November 1995.

IEEE 1329-2010, "Standard Method for Measuring Transmission Performance of Speakerphones," October 2010.

ITU-T Recommendation P.51, "Artificial Mouth," August 1996.

S. Temme, et al., "The Challenges of MP3 Player Testing," presented at the Audio Engineering Society (AES) 122nd Convention, May 2007.

of the device was a stepped sine sweep. For the microphone measurements, the amplitude at each frequency step of the sweep was equalized to produce an equal pressure response of 89 dB SPL (-5 dBPa) at the mouth reference point of a mouth simulator, per IEEE standard 1329-2010. The sweep provided 1/12 octave resolution from 100 Hz to 10 kHz. For the speaker test, the stimulus signal had an amplitude of –13 dBFS. A 1/3 octave resolution sweep was used from 20 to 100 Hz and 1/12 octave resolution from 100 Hz to 20 kHz.

Analysis was performed using a heterodyne filter, which extracted the fundamental and harmonic distortion products from the response signal.

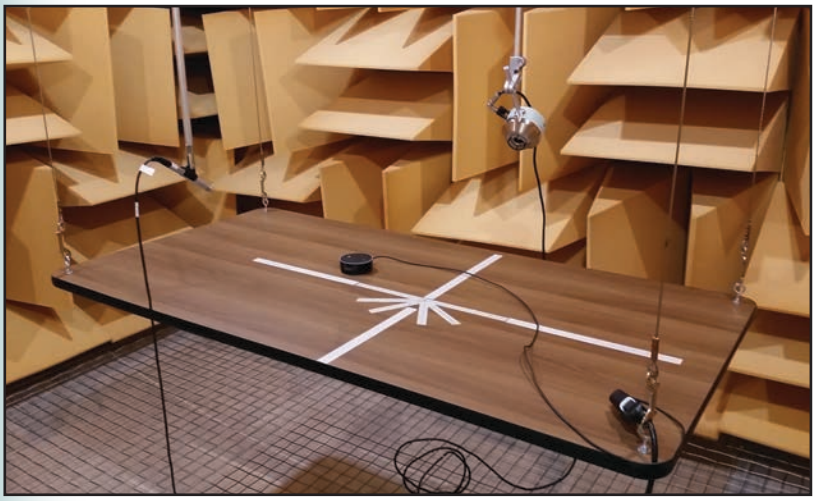

Photo 1: We followed the IEEE 1329-2010 test setup and conducted the test in a full anechoic chamber).

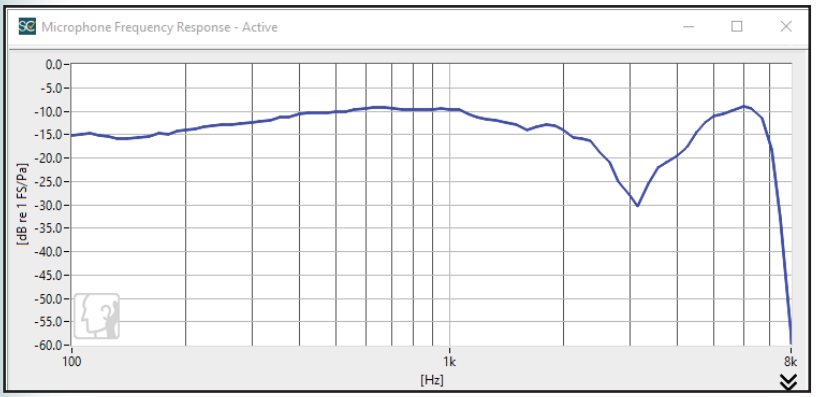

Figure 5: Microphone frequency response

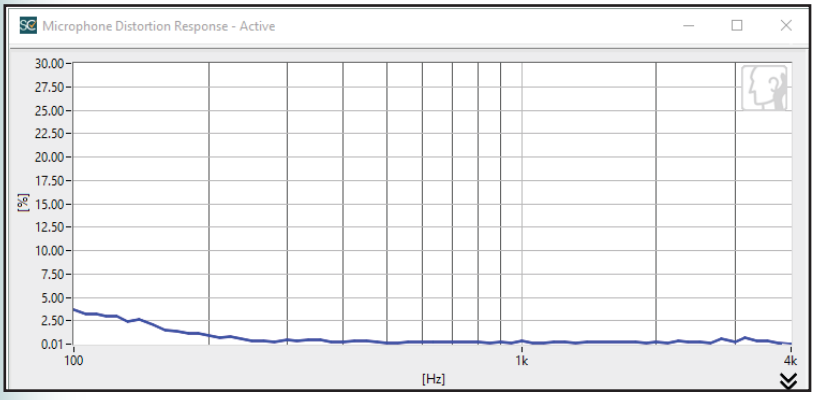

Figure 6: The microphone distortion response

Measurement Equipment

The measurement setup is similar to a regular loudspeaker measurement. At the heart of the setup is Listen, Inc.'s SoundCheck test system. This generates and plays all test signals, as well as acquiring and analyzing the response. An AudioConnect audio interface plays the signal from the computer to the mouth simulator, and similarly from the microphone to the computer (see **Figure 2**). The artificial mouth is required to generate the activation command and to play the test signal for a microphone measurement (see **Figure 3**).

Physical Setup

At this point there are no test standards specifically written for smart speakers. However, their practical use case closely resembles speakerphones, so for these measurements we chose to test using the physical geometry recommended by IEEE 1329-2010 (see **Figure 4**). All our tests were conducted inside an anechoic chamber with the DUT placed on a table (see **Photo 1**).

Device Microphone Measurements

Microphone system performance includes measurements of frequency response, sensitivity, and distortion. The microphone measurements were made by:

- Applying the test signal using a calibrated mouth simulator
- Retrieving of the voice interaction recording from the associated website and transferring it to the software system
- Analyzing the recording with the SoundCheck software

The microphone frequency response is relatively flat except for the dip at 3 kHz (**see Figure 5**). This is due to the device's microphone proximity to the table surface causing a "table bounce" effect. The sensitivity is measured in decibels F_S/Pa, where F_S is relative to digital full scale in the waveform recorded by the device,

and relative to the sound pressure in Pascals at the mouth reference point.

The total harmonic distortion is relatively low above 200 Hz, but increases at lower frequencies due to the residual distortion of the mouth simulator (see **Figure 6**).

Device Speaker Measurements

Speaker system performance includes measurements of frequency response, sensitivity, and distortion. The speaker measurements were made by:

- Uploading test signal to the associated online music library in MP3 format
- Dictating the activation phrase to play test signal (test operator or artificial mouth)
- Capturing the response signal with a measurement microphone
- Analyzing the captured signal with the SoundCheck software

The frequency response is relatively flat in the passband from 400 Hz to 6 kHz (see **Figure 7**). The 400 Hz roll-off is due to physical limitations of the device's loudspeaker. The sensitivity is measured in decibels re 20 µPa/FS, where output is sound pressure level (SPL) at the reference mic and input is relative to digital full scale in the signal played back by the device (see **Figure 8**).

Total harmonic distortion (THD) is relatively low above 500 Hz, but increases at lower frequencies due to limitations of device's loudspeaker.

Conclusions

Smart speakers present three main testing challenges. The practical challenge of playing back test signals and recording response signals, controlling the DUT, and sampling rate mismatch between the DUT and the test system. However, as we have demonstrated, all these challenges can be overcome and conventional results can be extracted.

A continuing challenge is that every smart speaker is different. Unlike conventional loudspeakers where an identical test can be used on a selection of speakers, a test will likely need to be tailored for each speaker due to differences in activation words, recording time, method of extracting the recording, and so forth.

There is also much additional work to be done exploring tests involving non-stationary signals, including noise, modulated noise, and actual speech and music. In addition, it would be of value to devise a test to explore how well the devices can detect their activation phrases while playing back music, in the presence of noise, and with multiple interfering talkers. Finally, a test that could independently evaluate the actual accuracy of the speech recognition algorithms would be valuable.

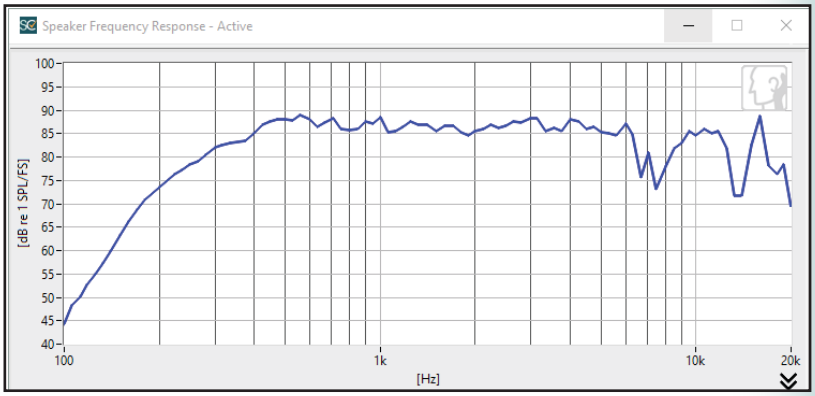

Figure 7: The speaker frequency response

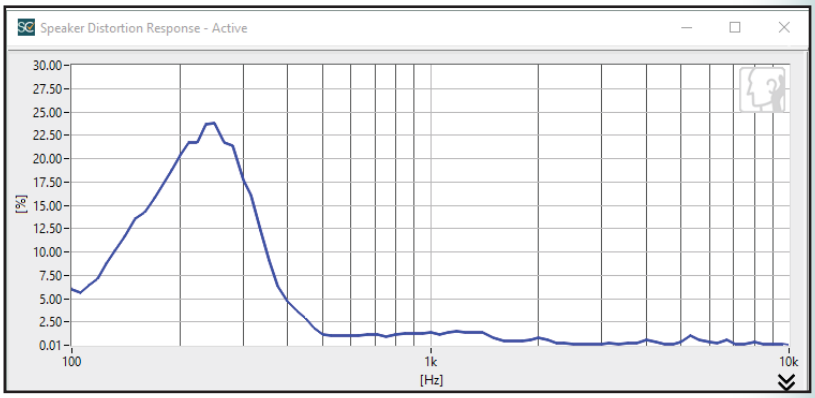

Figure 8: The speaker distortion response

About the Authors

Daniel Knighten has a wealth of experience testing all types of audio and audio electronic devices, and offers applications support to Listen Inc.'s global customer base. Prior to Listen, he held various sales, engineering and support positions at Audio Precision for more than 10 years, where he acquired a wealth of practical experience testing a wide range of electroacoustic and electronic devices. In addition to his role at Listen, he is also director and founder of Portland Tool & Die, a company that designs and manufactures audio test interfaces, which are exclusively distributed by Listen.

Glenn Hess has been involved in electro-acoustic testing of communication systems and more recently consumer audio devices for more than 30 years. His career has included engineering positions at Bell Laboratories, MWM Acoustics, and Harman International. He has served as technical editor and contributor on various Institute of Electrical and Electronics Engineers (IEEE) and Telecommunications Industry Association (TIA) industry standards for more than 25 years. And, he currently serves as vice chairman of the IEEE Working Group for Communication Electroacoustics. Glenn is a co-founder and general manager of Indy Acoustic Research (IAR) formed in April 2015. IAR provides acoustic engineering consulting for communication, consumer electronics, and specialty products.