This article discusses tools and techniques that are available to accurately measure the audio performance of voice-controlled and connected devices under the many various real-world conditions they may be used. It covers basic acoustic measurements such as frequency and distortion response, which have always been carried out on conventional wired systems, and the more complex real-world tests that apply specifically to voice-activated devices, along with the techniques and standards that may be used.

Although smart speakers, hearables and automotive infotainment systems are often categorized as different types of devices, they all have one thing in common—voice control. Furthermore, they all present similar challenges and complexities of test. These include many possible interfaces ranging from hardwired to wireless (Bluetooth, cloud-based), smartphone, voice (“Hey Siri,” “OK Google,” “Alexa”), and in the case of automotive, even USB, memory stick, and CarPlay/Android Auto.

Testing these devices also involves considerable signal processing both on the record side (e.g., beamforming, background noise filtering, and voice activity detection) and on the playback side (e.g., loudness, compression, equalization, and active noise cancellation). This results in changes to their characteristics according to real-world conditions such as the physical environment, playback levels, and background noise, so it is often necessary to test such devices at different levels and different environmental conditions.

Their multifunctional nature means that there are also 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, telephone headset, or hearing aid. Voice recognition usually needs to also be tested with different languages, levels, and environmental conditions (e.g., with and without background noise, different signals etc.)

Most of the tests discussed in this article are relevant to all smart devices including smart speakers, hearables, and automotive infotainment, and some hearable-specific additional tests are also detailed. Finally, the article presents a check list of the test-system functionality required to fully characterize voice controlled audio devices.

Basic Acoustic Tests: Wired and Wireless Measurements

Basic acoustic tests measure the device under static conditions to measure one or multiple microphones and speakers or earphones characteristics such as frequency and impulse response, distortion, and for stereo headphones, left/right tracking. Usually, both wired and wireless performance are measured for comparison. This requires a calibrated input (e.g., a headphone amplifier or auxiliary input) for wired responses and a calibrated Bluetooth interface for wireless measurements.

In addition, a calibrated mouth simulator or source speaker, equalized for a constant sound pressure output level versus frequency, and a calibrated measurement microphone or ear simulators are required for both send and receive acoustical measurements. Although a single measurement microphone usually suffices for smart speaker tests, a head-and-torso simulator (HATS) is recommended for hearables and infotainment systems measurements, since it has both a calibrated mouth and ear simulators to truly represent a typical listener. However, for
hearables, ear couplers offer a lower-cost option, and for automotive testing, an array of six measurement microphones is often substituted (see Figure 1).

There are currently no standards for testing most smart devices. That said, standards and guidelines exist for many devices that are functionally very similar (speakers, headphones, hearing aids, telephones, etc.), so it makes sense to borrow from these. Such standards include International Electrotechnical Commission (IEC) and European Norm (EN) for loudspeakers and headphones; Institute of Electrical and Electronics Engineers (IEEE) for headsets; IEEE/Telecommunication Industry Association (TIA)/International Telecommunication Union (ITU) for telephone tests; American National Standards Institute (ANSI) and IEC for hearing aid standards; and European Telecommunications Standards Institute (ETSI) for background noise.

Figure 1: The test configuration for basic wired and wireless infotainment measurement

Figure 2: Test configuration for basic speaker measurements via the cloud for a smart speaker (a); a hearable (b); and automotive infotainment (c)
Open-Loop Measurements

Although, on the face of it, playback and record measurements via the cloud look simple, there is much going on “under the hood” of the measurement system. One of the main challenges that must be overcome is correcting for sampling rate error. Traditional measurement systems rely on the device under test (DUT) having a synchronous input and output. Voice-activated hearables, infotainment systems, and 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 in-fact been recorded at a slightly different sample rate (e.g., 44.09 kHz) due to skew in the actual rate of the clock crystal used to drive the sampler (see Figure 1).

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, and also changes the phase response.

To overcome this sampling rate error, an algorithm is applied that searches the beginning of the response waveform for a “trigger,” which is used to provide a reference point for alignment and shifting of the response to the stimulus signal. This trigger may be a steady state sinusoid at a pre-set level and frequency (see Figure 2), or for more robust performance that is less susceptible to false triggers, a log chirp. 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 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 straightforward, enabling frequency and distortion to be displayed in exactly the same way as for a conventional wired device.

Figure 1: Open-loop measurements with differing sampling rates are shown in the time and frequency domain.

Figure 2: A 1 kHz frequency tone is prepended to stepped sine stimulus to calculate sampling rate and simplify record triggering.
Measurements Via the Cloud

Measurements on smart devices are very similar to standard loudspeaker and microphone measurements, but injecting the stimulus and extracting the recorded signal presents challenges. The only way to do this is like an actual user—via the cloud using voice commands.

An artificial mouth, often contained within the HATS, plays the voice-activation commands to the device’s microphones, which are typically connected via Bluetooth to a smartphone and/or a head unit. In the case of a smart speaker, an artificial mouth or small speaker can be used, and a smartphone is not required as the device itself has the Internet connection. The smartphone communicates to the proprietary voice services in the cloud, for example, Alexa Voice Services (AVS), and tells the voice services which uploaded test signal to stream/playback through the smartphone via Bluetooth to the hearable or infotainment system. The test measurement system uses a triggered record to capture the loudspeakers playback signals that can then be analyzed by the test system.

Although the actual physical testing setup is very similar from device to device, the method of wirelessly routing the signal may differ. The test signal must generally be in the cloud to enable playback, but each manufacturer’s ecosystem is different in how it plays back from the cloud. Some allow your own recordings to be uploaded, whereas others require them to be on a media streaming platform (e.g., Spotify).

For microphone testing, some systems (e.g., Alexa) allow you to access recordings you have made; others prevent this for security and privacy reasons, which makes making testing those microphones challenging. Each is activated 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 must be established for each device under test (DUT), and is usually automated by using record triggering and autodelay.

The physical setups can be taken from the test standards for that particular device. For hearables, the physical configuration for headsets recommended by IEEE 269, in which tests are conducted with the hearable worn on a HATS, is recommended. For smart speakers, since their practical use case resembles speakerphones, the physical geometry recommended by IEEE 1329-2010, which involves testing on a table inside an anechoic chamber, is a good starting point.

In the infotainment world, TIA920 presents a useful test configuration. In all three cases, the artificial mouth, which may be contained within HATS, plays the voice activation commands to the device’s microphones that are typically connected via Bluetooth to a smartphone, or directly to the cloud in the case of smart speakers. The smartphone communicates to the proprietary voice services in the cloud (e.g., AVS) that tells the voice services which uploaded test signal to stream/playback through the smartphone via Bluetooth to the device. The measurement system uses a triggered record to capture the loudspeakers playback signals that can then be analyzed. Example test configurations are shown in Figure 2.
Playback

When creating a playback test using voice activation, it is ideal to automate the test as much as possible. Wake words and test stimuli should be pre-recorded and uploaded to the music service in the cloud so that they can be triggered by the test system.

The activation signal consists of the system’s wake word, followed by a command such “play test signal 1.” If a wide range of voice recognition systems from different manufacturers is being tested, it makes sense to record the different voice wake words separately from the playback command so that any combination can be selected.

The test stimulus must also be created and uploaded to the cloud. There is a chance that it might be compressed when it is uploaded, which may introduce some distortion. It can be verified that the MP3 encoding itself does not degrade the test signal by encoding and decoding the test signal, then using transfer function analysis between the original signal and the encoded version. Test signals will be discussed in more detail later, but for basic playback performance testing, a standard stepped sine sweep is preferred.

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 mouth simulator and capture the response with the reference microphones or artificial ears. Stimulus and analysis are then compared as in a traditional measurement to objectively quantify characteristics such as frequency response, impulse response and distortion.

With the test system taking care of all the mechanics of getting the signals in and out of the devices and processing the waveforms, the frequency output and distortion for an earphone, car infotainment system or smart speaker can be displayed in exactly the same way as for a basic wired speaker, or a speaker connected via a Bluetooth interface (see Figure 3).

Recording

It is also necessary to measure the microphones or microphone array in the smart device. The setup for this is very similar, except that the wake word and the test stimulus is played through the mouth simulator or source speaker, captured by the device’s

Figure 4: Test configurations for measurement of the microphone within a smart speaker (a); a hearable (b); and an automotive infotainment system (c)
microphones, then the recorded test stimulus is downloaded from the cloud and routed back into the test system, either by recalling the recorded wave file or using a virtual audio cable (see Figure 4). Once the signal has been brought into the test system this way, as with the playback test, the appropriate calculations are applied to overcome frequency shift and error and it is analyzed as normal.

**Nonlinear Behavior**

While the basic tests described earlier are ideal for measuring frequency response and distortion, they barely scratch the surface in terms of the breadth and depth of tests that one might want to carry out on a smart device such as a smart speaker, hearable, or infotainment system in the R&D lab.

Smart devices and systems are generally nonlinear and their response varies according to many factors such as type of background noise, volume of background noise, whether the device is playing music or communicating voice, and more. These considerations bring up a multitude of tests that could yield valuable information about the systems’ performance. While the specific tests that are carried out vary according to the manufacturer’s design objectives, some of the test methodologies that might be used, and the industry standards that provide a useful starting point are relevant for the majority of voice-controlled devices.

**Figure 5:** Graphs showing results of end-to-end speech recognition

---

Ready for any Challenge!

Discover an Audio Analyzer you can trust for your application:

- Sound Level Meter class 1
- Level RMS / THD+N
- 1/1 or 1/3 octave RTA (up to 1/12 optional)

Distortion Analyzer  |  RealTime Analyzer  |  1/12 octave Analyzer

amatix@nti-audio.com  |  (503) 684 7050  |  www.nti-audio.com/XL2
A significant difference from basic tests is that such advanced tests generally require the use of speech or music as test signals. This means that a test system that can support such test signals is essential and it is also important that the system can measure the active speech level of speech-based signals to ensure calibration and standardization as commonly used in telephony measurements.

**Speech Recognition**

Speech recognition tests are important for evaluating the system’s ability to understand voice commands. The physical setup is the same as in the basic test outlined above. The test method usually involved creating a series of specially titled musical tracks and uploading them to the appropriate cloud service (e.g., iTunes). 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 are 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 (see Figure 5).

There are many variations on this test, for example, testing different voices and accents, testing an active talker versus an interfering talker or background noise levels, testing with different types of background noise, and so forth.

**Background Noise**

Background noise is necessary for a wide range of smart system and device tests including voice recognition, signal-to-noise ratio (SNR) optimization of microphone arrays, beamforming directionality studies and more. Where background noise is needed, ETSI standard ES 202 396-1 contains binaural recordings made in different noise environments (e.g., café noise, subway, road noise, traffic noise, sirens etc.). These can be played using a costly dedicated system, or with sufficient expertise, by programming your own playback method.

However, some systems offer integrated ETSI background noise modules, which fully integrate the ETSI standard library with the test system, playing the background noise at calibrated levels, which can be controlled and adjusted as part of a pre-programmed test sequence (see Figure 6). Such test sequences may include loops to incrementally increase the volume or change the noise and repeat the test, making it simple to implement a test that increases the background noise by fixed levels until the voice is no longer accurately recognized (see Figure 7). In addition to being a very cost-effective approach, it also significantly simplifies the physical test set-up and reduces test development time.

**Voice Quality**

Smart speakers, infotainment systems, and hearables often function as hands-free communication devices. For such use cases, we can lean heavily on telephone test sequences such as the TIA standard TIA920, which covers both hands-free devices (as in automotive applications and smart speakers) and headset devices (e.g., hearables). In fact, smart speakers and infotainment systems can be tested as a speakerphone exactly as defined in...
the TIA standard, and a hearable can be tested as a headset exactly as defined in the TIA standard.

Although the test setup in each case is identical to the basic acoustic test setups outlined above, the test sequences for measurement to the standard are highly complex and either require considerable expertise in this area to create them, or they may be purchased as an off-the-shelf package.

**Noise Cancellation**

Although most commonly used for hearables, noise-cancellation tests also have applications in automotive infotainment systems for reduction of cabin background noise. The test configuration for measuring noise cancelling headphones is shown in Figure 8. The test procedure can be broken down into three measurements and three calculations. First, the headphones are removed from the head and torso, the noise signal is played, and the spectrum measured through the open ear. This un-occluded ear spectrum is used as the baseline for the noise. Next, passive isolation is measured by placing the headphones onto the head and torso, playing the noise signal again and measuring the noise through the artificial ear. Finally, the measurement is repeated with the active noise cancellation (ANC) circuit turned on.

These measurements result in three spectra, which are used to calculate the three attenuation parameters. Passive Isolation, which quantifies how much noise is attenuated simply by the headphones being worn, is calculated by subtracting the un-occluded baseline curve from the second measurement where the headphones are in place.

![Figure 8: A test configuration for measuring ANC hearables](image-url)
but noise cancellation is disengaged. Passive isolation can be significant across the frequency band but is most prominent at higher frequencies. Circum-aural headphones and in-ear earphones provide a reasonable level of passive isolation, whereas supra-aural headphones and earbuds typically offer much less.

Active Attenuation, which quantifies how much noise is reduced (or sometimes increased) when the active cancellation circuit is engaged, is calculated by subtracting the second measurement (noise cancellation off) from the third measurement (noise cancellation on). Typically, this will be most prominent in the lower frequencies.

Finally, total attenuation is calculated by subtracting the curve with noise cancellation turned on from the baseline measurement without headphones. This represents the end-user’s experience of the device, combining both passive and active components in attenuating background noise (see Figure 9).

Often this test is optimized by performing each measurement multiple times and averaging the resulting spectra. This accounts for variations in the fit of the headphones, which for some devices can impact the passive and even the active portions of the noise cancellation performance.

Some ANC headphones allow control of the degree of noise cancelling typically via a smartphone app. This is designed to account for different listening environments (e.g., on a plane with maximum noise cancellation versus walking on a sidewalk next to a busy street where one should be more aware of the environment). In this case, measurements can be configured to show the level of noise cancellation versus frequency for the different environments. Again, this is a case for using a calibrated, simulated background noise system such as that specified by ETSI to recreate the different test environments in a very repeatable way.

Some noise cancelling algorithms actually alter the frequency response of the music being played when the noise cancelling circuit is engaged. This effect can be evaluated by measuring the frequency response using the above methods, both with and without the noise cancellation turned on, and comparing the resulting response curves. Differences, if they are present, will typically be in the 1 kHz to 3 kHz range.

Hearing Assistance

Fairly specific to hearables is the integration of hearing enhancement for personal sound amplification product (PSAP) or hearing aid functionality, and it is desirable to measure
their amplification or compression algorithms along with their corresponding attack and release times. This is usually measured according to ANSI/IEC hearing aid standards in an anechoic test box (see Figure 10), where the loudspeaker is equalized to have a flat frequency response and the hearable is inserted into a 2cc coupler or “artificial ear.”

The ANSI S3.22 standard describes the method for testing hearing aids with automatic gain control (AGC) and can be applied to hearables that offer this functionality. The device is placed in an anechoic box, and an amplitude sweep from 50 to 90 dBSPL played through the test box speaker. This sweep is repeated at four frequencies (500 Hz to 4 kHz) in octave increments. The output of the hearing aid is plotted vs. the input on the graph to quantify the gain added for low level signals and compression at high levels for hearing loss compensation (see Figure 11).

Conclusions and Test System Recommendations

We have described a variety of ways to measure the audio performance of voice controlled audio devices including smart speakers, headphones/hearables, and automotive infotainment systems. These measurements represent the convergence of tests from loudspeakers, headphones, microphones, telephones, hearing aids, and more, and in addition must be made with a complex open-loop signal path, which may involve smartphones, head units and signals stored in the cloud.

When selecting a test system, based on these requirements, a huge number of test stimuli, algorithms, analysis options, and signal paths options are necessary, not to mention many “useful-to-have” features such as integrated background noise and a test configurator that makes it easy to run many iterations of a test continuously. We have compiled a checklist of the functionalities that will likely be needed and should be carefully evaluated before selecting a test system. The check list of functionalities includes:

- Tests both the input and output components of the device (i.e. loudspeaker or earphones 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

About the Author

Steve Temme is founder and President of Listen, Inc., manufacturer of the SoundCheck audio test system. Steve founded the company in 1995, and for the past 25 years the company has remained on the cutting edge of research in audio measurement, regularly introducing new measurement techniques, algorithms, and hardware to enable testing of all kinds of devices ranging from basic transducers to complex audio systems. In recent years, Listen’s product development focus has been on creating the algorithms, interfaces and test methods to measure today’s many and varied wireless and cloud-communicating devices. Prior to founding Listen, Steve worked for many years as an acoustic test and measurement applications engineer at Bruel & Kjaer, and also as a loudspeaker design engineer at Apogee Acoustics. He holds a BSME from Tufts University, has authored numerous papers on acoustic testing, and has lectured extensively throughout the world.
the trigger tone can be eliminated from the analysis

- Able 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—native Windows multimedia interface
- Availability of pre-written TIA test sequences if device needs to be tested to telephony standards
- Test programming flexibility – one that allows easy duplication of sub-sequences and loops is extremely helpful when repeating a test at various levels
- Offers a calibrated Bluetooth interface
- Mentor A2B interface compatibility and control from within test system if testing automotive smart devices (e.g. in-car voice control)

Relevant Standards

ANSI S3.22 Specification of Hearing Aid Characteristics

BS EN 50332-1 Sound System Equipment – Headphones and earphones associated with portable audio equipment—Maximum SPL measurement methodology and limit considerations

ETSI ES 202 396-1: Speech and Multimedia Transmission Quality (STQ); Speech quality performance in presence of background noise; Part 1: Background noise simulation technique and background noise database

IEC 60118-7: Electroacoustics—Hearing aids—Part 7: Measurement of the performance characteristics of hearing aids for production, supply and delivery quality assurance purposes

IEC 60268-7: Sound system equipment – Part 7: Headphones and Earphones


IEEE Recommended Practice for Speech Quality Measurements


TIA 920.130: Telecommunications Communications Products Transmission Requirements for Digital Interface Communications Devices with Headsets

The “Must Have” reference for loudspeaker engineering professionals.

Back and better than ever, this 7th edition provides everything you need to become a better speaker designer. If you still have a 3rd, 4th, 5th or even the 6th edition of the Loudspeaker Design Cookbook, you are missing out on a tremendous amount of new and important information! Now including: Klippel analysis of drivers, a chapter on loudspeaker voicing, advice on testing and crossover changes, and so much more!

A $99 value!
Yours today for just $39.95.
www.cc-webshop.com