Testing Voice-Controlled & Smartphone Integrated Infotainment Systems

Tutorial

Presented at the AES International Conference on
Automotive Audio
2019 September 11–13, Neuburg an der Donau, Germany

This paper accompanies the tutorial presented at the above conference, and is intended to provide the reader with additional technical details beyond what is presented in the slides / verbally.

Testing Voice-Controlled & Smartphone Integrated Infotainment Systems

Steve F. Temme[1]

[1] *Listen, Inc., Boston, MA, 02118, USA*

Correspondence should be addressed to Steve Temme (stemme@listeninc.com)

ABSTRACT

Voice-controlled and smartphone integrated vehicle infotainment systems are notoriously complex to test. They have numerous connections from wired to wireless and contain much signal processing, both on the record and on the playback side. This means that their characteristics change according to ‘real world’ conditions of the vehicle’s environment, including cabin acoustics and background noises from road, wind and motors. Furthermore, their multifunctional nature means that there are many aspects of the device that may need to be tested, ranging from voice recognition to music playback and operation as a hands-free telephone. Due to their complex non-linear use cases, these devices often need to be tested at different levels and different environmental conditions.

1 Introduction

Infotainment systems have become increasingly challenging to test. They have many possible interfaces; hard-wired or auxiliary input, radio, CD, memory card, hard drive, USB, Bluetooth, smartphone (including Apple CarPlay and Android Auto) and even voice. They contain much signal processing, both on the record side (e.g. beamforming, background noise filtering, voice activity detection, and on the playback side (e.g. loudness, compression, equalization, and active noise cancellation). Some even have wake word detection, e.g. “Hey Siri”, “OK Google”, and “Alexa”. Due to their complex non-linear use cases, these devices often need to be tested at different levels and in different environmental conditions, for example with different background noises and different test signals.

To further complicate matters, the test signal may need to be in the cloud to enable playback for testing voice recognition systems. Each manufacturer’s ecosystem is different in how it plays and records. Smartphone integrated infotainment systems usually require an internet connection with voice services in order to process commands. On the playback side, some enable you to upload your own recordings such as iTunes (although bear in mind that these will probably be compressed). Others require them to be on a media streaming platform such as Spotify. For microphone testing, some systems such as Alexa allow access to recordings made; others do not for security and privacy reasons, which makes microphone testing challenging. Although the actual physical testing setup is very similar from vehicle to vehicle, for each it is necessary to understand how to wirelessly route the signal. Furthermore, each device needs activating with a different wake word, needs different delay compensation, and records for a different amount of time after it hears the wake word. This needs figuring out (largely by trial and error) for each infotainment system that you need to test.

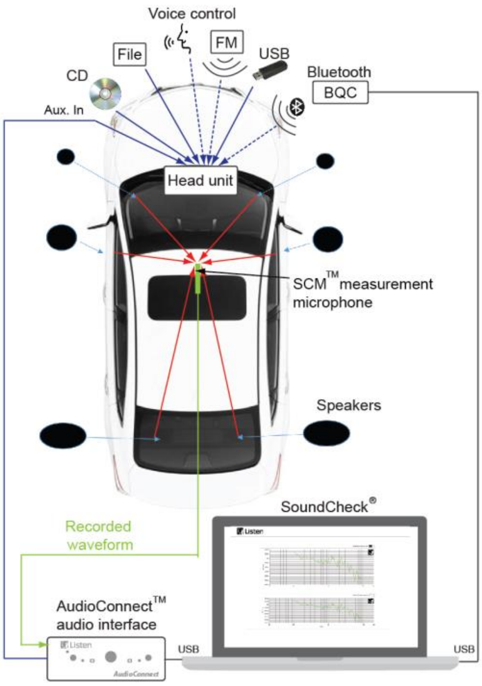

Figure 1. Numerous interfaces required for testing infotainment systems

Although, there are currently no standards for testing infotainment systems with smartphone integration, principles and test configurations can be borrowed from many other audio devices and use existing standards such as IEC for loudspeakers, IEEE/TIA/ITU for speakerphones, and ETSI for background noise. 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. This paper focuses on how to implement basic acoustic tests and some of the more complex real-world tests along with the techniques and standards that may be used.

2 Basic Acoustic Tests

Basic acoustic tests essentially measure the infotainment system under static conditions to measure speaker and microphone characteristics such as frequency & impulse response, and distortion. If possible, it is a good idea to measure both wired and wireless performance for comparison. This requires a calibrated auxiliary input for wired measurements and a calibrated Bluetooth interface for wireless measurements. In addition, a calibrated mouth simulator 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. A Head & Torso simulator is the ideal choice since it has both a calibrated mouth and ear simulators to truly represent a typical passenger in a vehicle. However, many automotive test engineers still use an array of 6 measurement microphones (Figure 2) to represent an average driver and passenger in a vehicle.

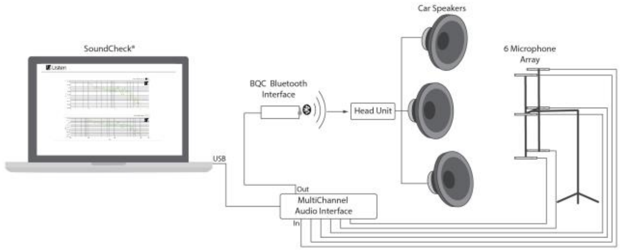

Figure 2. Test configuration for basic wireless (Bluetooth) in-vehicle loudspeaker measurements

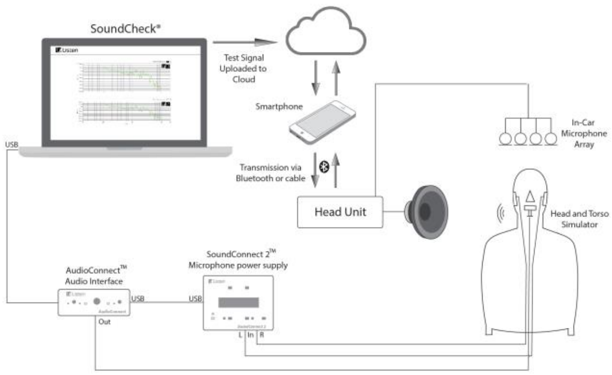

Figure 3. Test configuration for testing infotainment speaker system using voice activation

The setup in Figure 3 includes an artificial mouth (contained within the HATS) to play the voice activation commands to the infotainment's microphone(s) which is typically connected via Bluetooth to a smartphone. The smartphone communicates to the proprietary voice services in the cloud e.g. Alexa Voice Services (AVS) that tells the voice services which uploaded test signal to stream/playback through the smartphone via Bluetooth to the 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.

2.1 Playback

Let's now discuss the steps necessary to create a playback test using voice activation. It is ideal to automate the test as much as possible, so 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. First, 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 voice recognition systems from different manufacturers is being tested, it makes sense to record the different voice 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 may be compressed when it is uploaded which might 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, and using transfer function analysis between the original signal and the encoded version. More in-depth discussion about test signals to follow later but for basic speaker performance testing, it is preferred to use a standard stweep (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 artificial mouth, and capture the response with the reference microphone(s) or artificial ears. Stimulus and analysis are then compared as in a traditional measurement to objectively qualify characteristics such as frequency response, impulse response and distortion.

With the test system taking care of all these interfaces, the 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.

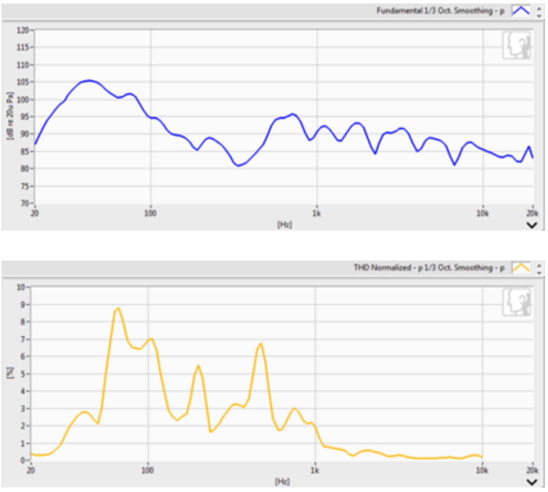

Figure 4. Frequency and distortion responses of an infotainment multichannel playback system

2.2 Open Loop Measurements

One of the main challenges that test system manufacturers have faced in developing accurate smart system tests, is overcoming sampling rate error. Traditional measurement systems rely on the device under test having a synchronous input and output. As a category, infotainment systems are often 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 (Figure 5). 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.

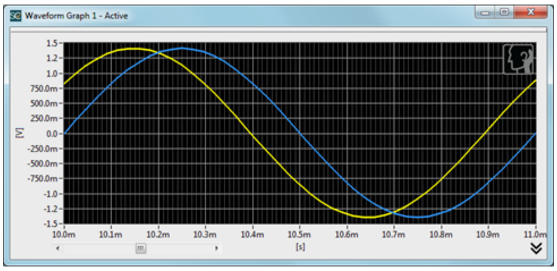

Figure 5. Open loop measurements with differing sampling rates

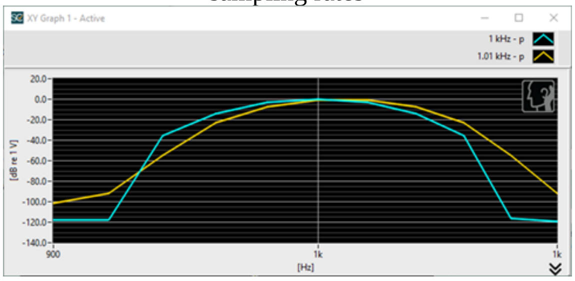

Figure 6. Resampling Recorded Waveform to Match Sample Rate of Test Signal

This sampling rate error results in the component tones of the test stimulus being shifted to either a higher or lower frequency (Figure 6). This shift can then lead to measurement errors due to loss of coherence between the stimulus and response signals. It also changes the phase response.

To overcome this sampling rate error, an algorithm is applied which searches the beginning of the response waveform for a 'trigger' which 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 (Figure 7), 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 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.

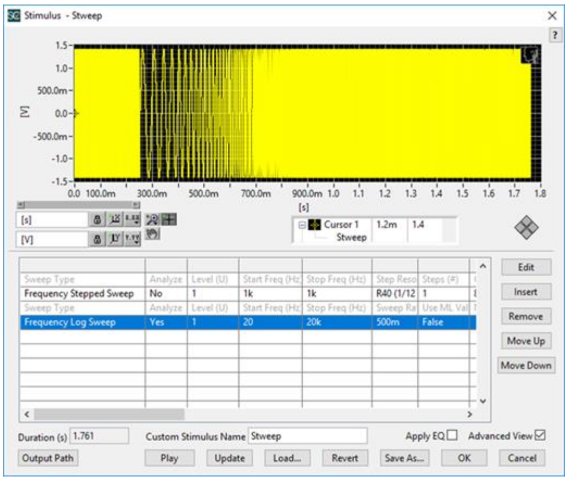

Figure 7: 1kHz frequency tone prepended to stepped sine stimulus to calculate sampling rate and simplify record triggering

2.3 Recording

Of course, it is also necessary to measure the microphone(s) or microphone array in the vehicle. The set up for this is very similar, except that in this case, the wake word and the test stimulus is played through the mouth simulator, captured by the infotainment's microphone(s), then the recorded test stimulus is downloaded from the cloud and routed back into the test system using a virtual audio cable (Figure 8). 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.

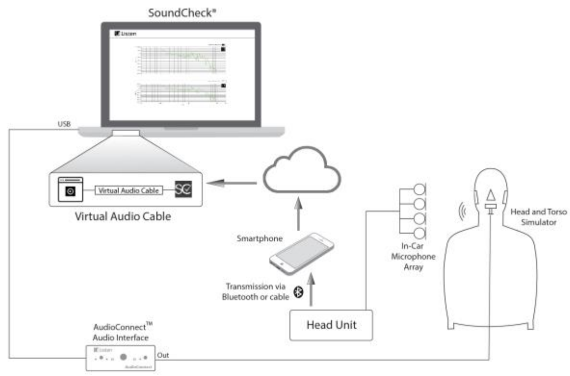

Figure 8. Microphone(s) measurement using voice activation and virtual audio cable

3 Advanced or 'Real World' Testing of Infotainment Systems

While the basic tests described above 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 an infotainment system in the R&D lab. Consider that smart systems are generally non-linear and their response varies according to many factors such as type of background noise, volume of background noise, whether the system is playing music or communicating voice, and more. These considerations alone bring up a multitude of tests that could yield valuable information about the systems' performance. While the specific tests that are carried out will vary according to the manufacturer's design objectives, some of the test methodologies that might be used and some of the industry standards that provide a useful starting point, should be discussed.

A significant difference from basic tests is that 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.

4 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.

One test method is to create a series of specially titled musical tracks and upload them to the particular 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'[1] and widely used in telecoms voice recognition. Pre-recorded versions of these can be downloaded online, or your own can be recorded (for example, 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 (Figure 9).

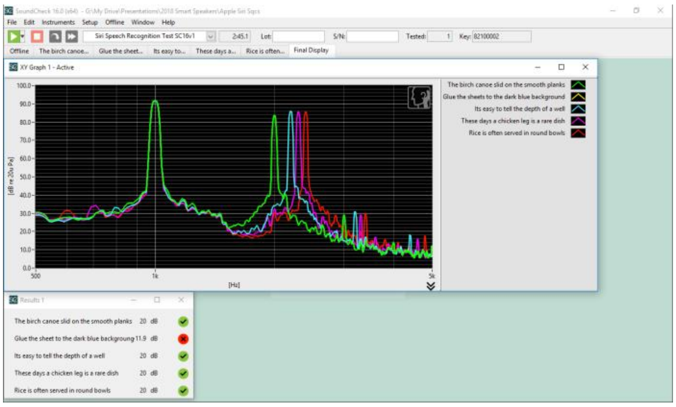

Figure 9. Graphs showing results of end-to-end speech recognition

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, etc.

5 Background Noise

Background noise is necessary for a wide range of smart system tests including voice recognition, SNR optimization of microphone arrays, beamforming directionality studies and more.

Where background noise is needed, ETSI standard ES 202 396-1 [2] contains binaural recordings made in different noise environments e.g. 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. 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.

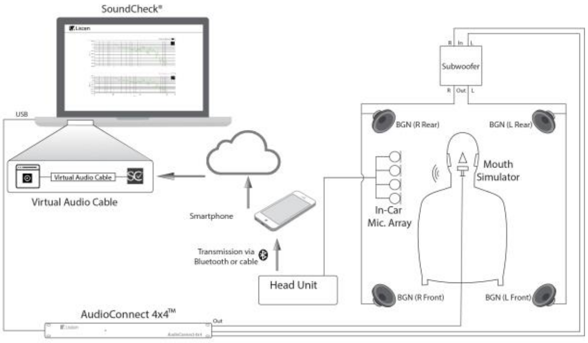

Figure 10. Test configuration for hands-free measurements with ETSI standard background noises

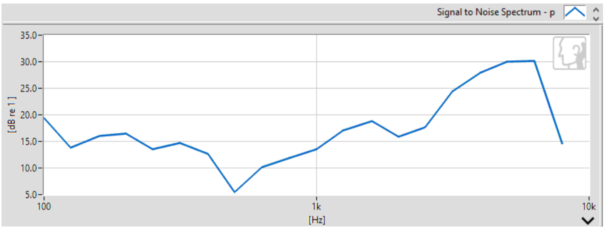

Figure 11. SNR for a microphone array measured with ETSI standard background noises

6 Voice Quality

Infotainment systems often function as hands-free communication devices. For such use cases, we can lean heavily on telephone test standards such as the TIA standard for hands-free telephones (TIA920-B). In fact, an infotainment system can be tested as a speakerphone exactly as defined in the TIA standard. Although the test setup is identical to the basic smart system test 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 need to be purchased as an off-the-shelf package.

7 Measurement System Requirements

Voice-controlled and smartphone integrated infotainment systems represent the convergence of tests from loudspeakers, microphones, telephones, and more. Between these, 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. Below is a checklist of the functionalities that should be evaluated before selecting a test system:

- 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 analyse only parts of the response signal, so that 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 (may be manually programmed with sufficient expertise)
- 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

8 Conclusions

This paper has described a variety of ways to measure the audio performance of modern day infotainment systems. Since most of these systems have voice control and smartphone integration with algorithms built-in, it is necessary to perform measurements that include these functionalities. These complex functionalities require open-loop measurements that include the entire signal path and use speech and music as test signals under real-world conditions.

References

[1] IEEE 1329-2010, "Standard Method for Measuring Transmission Performance of Speakerphones", October 2010

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

[3] Karlheinz Brandenburg, "MP3 and AAC Explained", presented at AES17th International Conference on High-Quality Audio Coding, August 1999

[4] Steve Temme et al., "The Challenges of MP3 Player Testing", presented at the AES 122nd Convention, May 2007

[5] Glenn Hess et al., "Challenges of IoT Smart Speaker Testing", presented at the AES 143rd Convention, October 2017

[6] Steve Temme, "Testing Audio Performance of Hearables", presented at the International Conference on Headphone Technology, August 2019

[7] 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