Audio Engineering Society

Conference Paper

Presented at the International Conference on
Headphone Technology
2019 August 27–29, San Francisco, CA, USA

Testing Audio Performance of Hearables

Steve F. Temme[1]

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

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

ABSTRACT

Smart headphones or "hearables" are designed not only to playback music but to enhance communications in the presence of background noise and in some cases, even compensate for hearing loss. They may also provide voice recognition, medical monitoring, fitness tracking, real-time translation and even augmented reality (AR). They contain complex signal processing and their characteristics change according to their smartphone application and 'real world' conditions of their actual environment, including background noises and playback levels. This paper focuses on how to measure their audio performance under the many various real-world conditions they are used in.

1 Introduction

Hearables are notoriously challenging to test. They have various interfaces ranging from hardwired to wireless (e.g. Bluetooth) and may 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). This means that their characteristics change according to 'real world' conditions such as their physical environment and background noises. Some even have wake word detection, e.g. 'Hey Siri'. 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 or even operation as a telephone headset or hearing aid. 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 background noise, different signals etc.

Although, there are currently no standards for testing smart devices such as hearables, we can borrow principles and test configurations from many other audio devices and use existing standards such as; IEC for headphones [1], IEEE for headsets [2], ETSI for background noise [3], TIA/ITU for telephone test [4] and ANSI for hearing aids standards [5]. 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 discusses 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. Test system requirements for measuring voice enabled hearables will also be discussed.

2 Basic Acoustic Tests

Basic acoustic tests essentially measure the device under static conditions to measure microphone or microphone array and earphone characteristics such as frequency response, distortion and left/right tracking for stereo headphones. If possible, it is a good idea to measure both wired and wireless performance for comparison. This requires a calibrated headphone amplifier for wired headphone responses and a calibrated Bluetooth interface for wireless headphone responses (Fig. 1). In addition, a calibrated mouth simulator, equalized for a constant sound pressure output level versus frequency, and ear simulators are required for both send and receive acoustical responses.

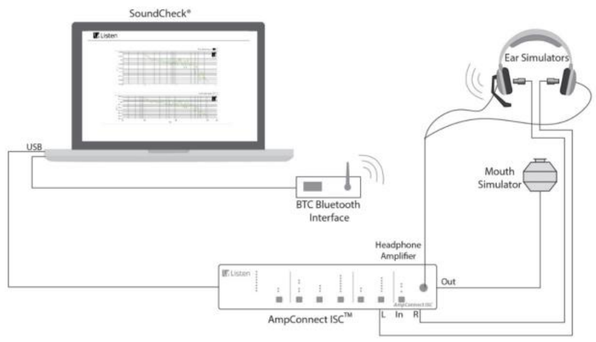

Figure 1. Test configuration for basic wired and wireless (Bluetooth) headset measurement

In some cases, it is not possible to use a standard wired or wireless connection. Some headphones require USB-C or Lightning connections while some hearables, such as Alexa, Siri, Google and Cortana enabled devices, will only respond to speech activation through wake words. While the measurements themselves are similar to standard headphone and microphone measurements, injecting the stimulus and extracting the recorded signal present can be challenging. Unless you happen to be the manufacturer of the device and have access via a debugger or other remote access tool, or are testing headphone Bluetooth performance only, you have a true black box test solution. This is generally the case when comparing performance of different headphones, for example, for reviewing or competitive benchmarking; the only way to get signals in and out of the devices is like an actual user – using real music and speech.

To further complicate matters, the test signal may need 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 such as Spotify. For microphone testing, some systems such as Alexa allow you to access recordings you have made via their website, whereas others do not for security and privacy reasons. This makes microphone testing challenging. Although the actual physical testing setup is very similar from one headphone brand to another, 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. These factors need figuring out (largely by trial and error) for each headphone that you need to test.

Let us first discuss the physical setup. At this point there are no test standards specifically written for hearables. Their practical use case, however, closely resembles headsets, so the physical setup recommended by IEEE 269 [2] is a good starting point. Following this, tests should be conducted with the hearable worn on a head and torso simulator (HATS) shown in Fig. 2.

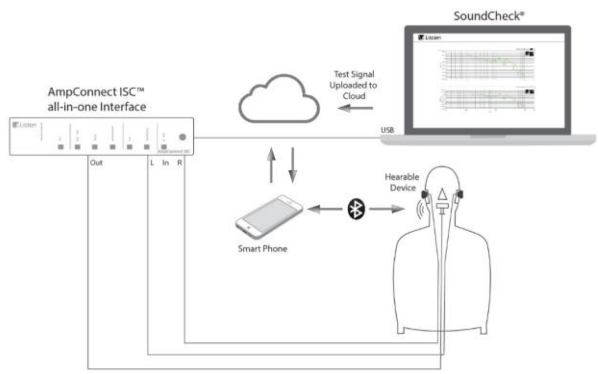

Figure 2. Test configuration for stereo headphone measurement using voice activation

The setup includes an artificial mouth (contained within the HATS) to play the voice activation commands to the hearable'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 dictates which uploaded test signal to stream/play back through the smart phone via Bluetooth to the hearable's earphones. The test measurement system uses a triggered record to capture both the left and right earphones' playback signals that can then be analyzed by the test system.

2.1 Playback

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 different voice services is being used, 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 may be compressed when it is uploaded which might 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 performing a transfer function analysis between the original signal and the encoded version to check for any degradation. We'll discuss test signals in more depth later but for basic earphone performance testing, we tend to use a standard stepped 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 mouth simulator and capture the response with the ear simulators. Stimulus and analysis are then compared as in a traditional measurement to objectively qualify characteristics such as frequency response and distortion (Fig. 3 & Fig. 4).

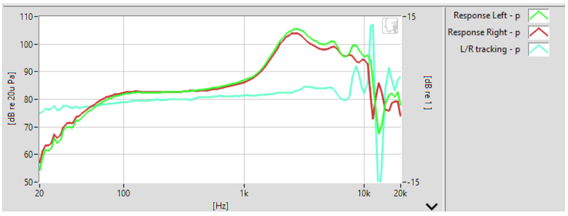

Figure 3. Smart headphone frequency and left/right tracking responses

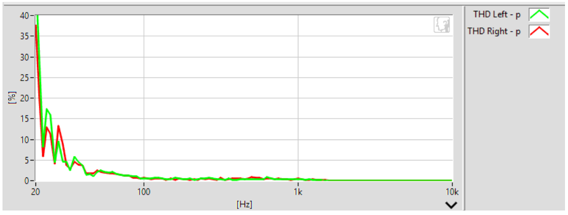

Figure 4. Smart headphone distortion response

2.2 Open Loop Measurements

One of the main challenges that test system manufacturers have faced in developing accurate smart headphone tests, is overcoming sampling rate error. Traditional measurement systems rely on the device under test having a synchronous input and output. As a category, hearables connected to a smart phone 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 crystal clock used in the A/D converter. 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 (Fig. 5).

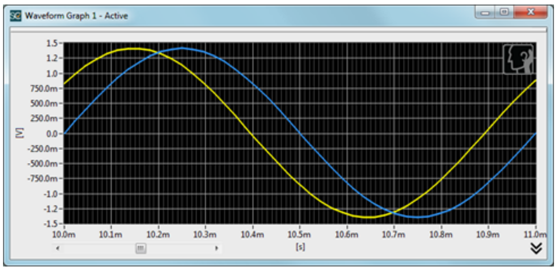

Figure 5. Open loop measurements with differing sampling rates

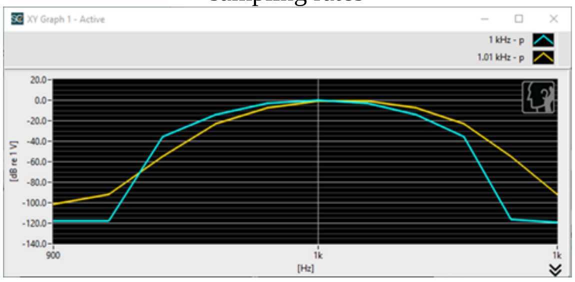

Figure 6. Recorded waveform needs to be resampled to match sample rate of test signal

This sampling rate error results in the component tones of the test stimulus being shifted to either a higher or lower frequency (Fig. 6). 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’ 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 (Fig. 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 (Fig. 6). This frequency shift step corrects for sampling rate error in the device and makes testing them straightforward.

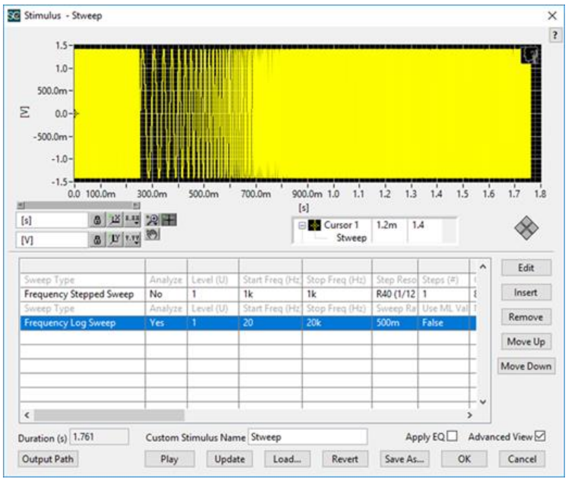

Figure 7: 1 kHz frequency tone prepended to stepped sine stimulus to calculate sampling rate and simplify record triggering

With the test system taking care of these issues, frequency output and distortion can be displayed in exactly the same way as for a headphone connected via a Bluetooth interface, directly via a debug port, or conventional wired headphones.

2.3 Recording

It is also necessary to measure the microphone or microphone array in the hearable. 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 hearable’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 (Fig. 8). Once the signal has been brought into the test system this way, as with the headphone test, the appropriate calculations are applied to overcome frequency shift and error and it is analyzed as normal (Fig. 9).

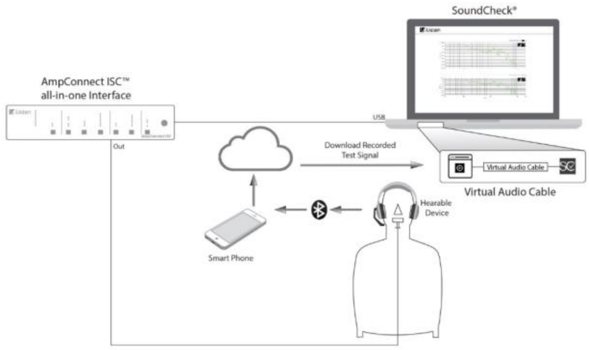

Figure 8. Test configuration for headset microphone measurement using voice activation

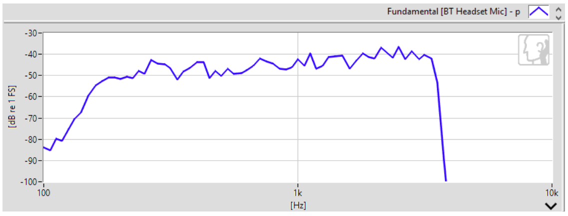

Figure 9. Frequency responses of hearable microphone array

3 Advanced, or 'Real World' Testing of Hearables

While the basic hearable test described above 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 hearable in the R&D lab. Consider that smart devices are generally non-linear 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 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, it is a good idea to discuss some of the test methodologies that might be used 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 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 Noise Cancellation

The test configuration for measuring noise cancelling headphones is shown in Figure 10. 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.

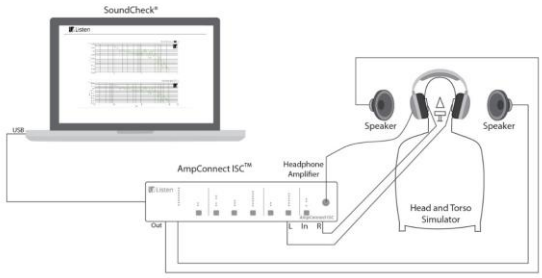

Figure 10. Test configuration for measuring ANC headphones

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, but noise cancellation is dis-engaged. 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 supraaural 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 (Fig. 11).

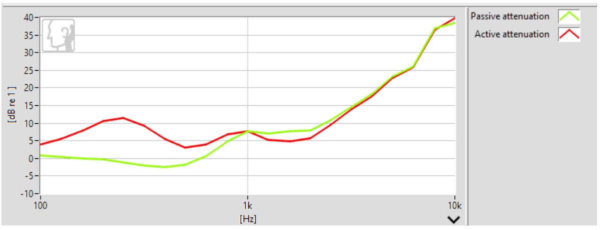

Figure 11. Typical Noise Cancelling Measurement Results

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 supposed to account for different listening environments e.g. on a plane with full noise cancelling on versus walking on a sidewalk next to a busy street where one should be more aware of their environment to avoid getting hit by a car, bike or even another person. In this case, a measurement showing the different levels of noise cancellation versus frequency for the different environments is a good idea. Again, this is a case for using a calibrated, simulated background noise system such as specified by ETSI to recreate the different test environments in a very repeatable way.

Also, some noise cancelling headphones will actually alter the frequency response of the music being played when the noise cancelling circuit is engaged. To test for this effect, one can measure the frequency response using the methods outlined above, both with and without the noise cancellation turned on, and compare the resulting response curves. Differences, if they are present, will typically be in the 1 – 3 kHz range

5 Speech 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 in Fig. 2.

There are various methods of measuring speech recognition, but one of the most straightforward 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' 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 (Fig. 12).

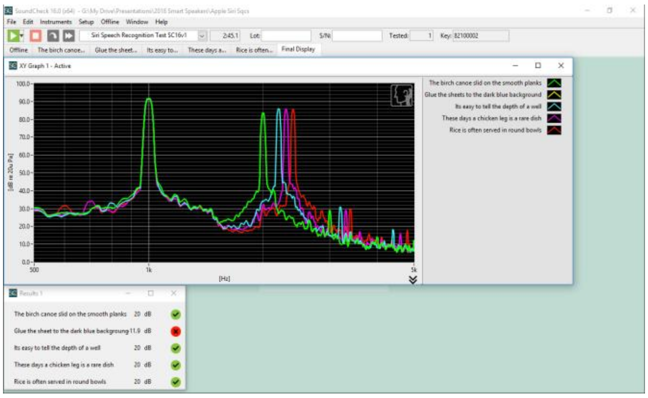

Figure 12. Graphs showing results of end-to-end speech recognition

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

6 Background Noise

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, for example through purchasing a dedicated system or by programming your own, it is advantageous to use a format that allows the ETSI standard library to be completely integrated with, and controlled by, your test system (Fig. 13). This means that the background noise can be played 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. Background noise is necessary for a wide range of smart device tests including voice recognition, SNR optimization of microphone arrays, beamforming directionality studies and more (Fig. 14).

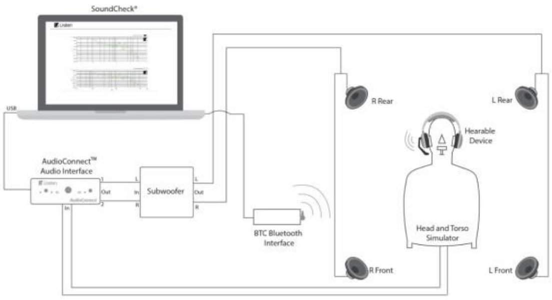

Figure 13. Test configuration for headset measurements with ETSI standard background noises

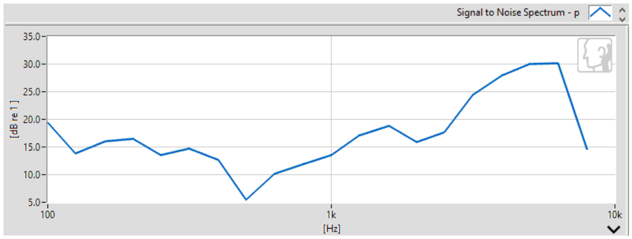

Figure 14. SNR for a smart headset measured with ETSI standard background noises

7 Voice Quality

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

8 Hearing Assistance

Since some hearables can also act as “personal sound amplification products (PSAPs)” or essentially hearing aids, it makes sense to measure their amplification or compression algorithms along with their corresponding attack and release times.

This is usually measured according to the ANSI/IEC standard in an anechoic test box (Fig. 15) where the loudspeaker is equalized to have a flat frequency response and the hearable is inserted into a 2cc coupler or “artificial ear”.

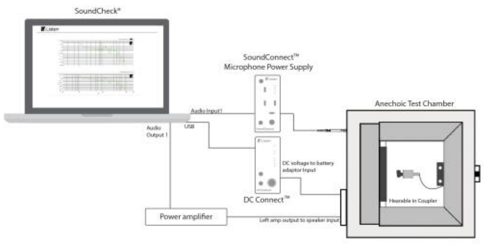

Figure 15. Test configuration for measuring hearing aids

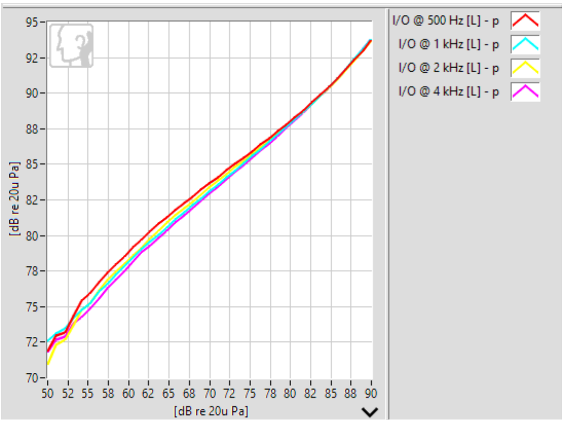

Figure 16. Hearable measurement of Input vs. Output

The ANSI S3.22 standard [5] describes the method for testing hearing aids with AGC (automatic gain control) and could be applied to hearables that offer this functionality. The device is placed in an anechoic box, and an amplitude sweep from 50-90 dBSPL played through the test box speaker. This sweep is repeated at four frequencies (500 Hz – 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 (Fig. 16).

9 Measurement System Requirements

Hearables represent the convergence of tests from headphones, microphones, telephones, telephony, hearing aids, 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 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 analyse 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 – 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)
- Sophisticated sequence writer that allows duplication of sub-sequences and loops - essential when repeating a test at various levels
- Has a calibrated Bluetooth interface available

10 Conclusions

This paper has described a variety of ways to measure the audio performance of modern day smart headphones/hearables. 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] IEC 60268-7: Sound system equipment – Part 7: Headphones and earphones

[2] IEEE 269: Standard Methods for Measuring Transmission Performance of Analog and Digital Telephone Sets, Handsets, and Headsets

[3] 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

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

[5] ANSI S3.22 Specification of hearing aid characteristics

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

[7] BS EN 50332-1 Sound system equipment – Headphones and earphones associated with portable audio equipment – Maximum SPL measurement methodology and limit considerations