# A SCIENTIFIC BASIS FOR CHOOSING LOUDSPEAKERS AND HEADPHONES FOR RECORDING AND BROADCAST

By Dr. Sean E. Olive
Acoustic Research Fellow
Harman International

Recording and broadcast engineers require accurate loudspeaker and headphone monitors to make precise and reliable assessments of sound quality. Accurate monitors must not emphasize, add or remove any audio frequencies in the sound mix that might be inadvertently "compensated" for during the mixing and mastering stages. Once these "compensations" are permanently encoded into the recording they will be perceived as distortions and colorations when reproduced through neutral or accurate monitors. Unfortunately, this scenario is all too common and represents a significant quality-control problem that plagues our audio industry, which Toole [1] refers to as audio's "circle of confusion".

The solution is to adopt a meaningful standard that defines what an accurate monitor is, and is shared by both the recording/broadcast and consumer audio industries. Such a standard would improve the consistency and quality of recordings and allow consumers to finally hear artist intended. As we will learn in this article, recent scientific evidence shows that listeners recognize and prefer accurate monitors, and their preferences can be predicted based on the monitors' acoustic performance. In other words, there is a scientific basis for choosing loudspeaker and headphones for recording and broadcast.

This brings us to the three main questions addressed in this article:

1) Do listeners agree on what an accurate loudspeaker and headphones and prefer them?
2) Can we measure and predict the perceived sound quality of loudspeakers and headphones to quantify their accuracy?
3) Does the recording and broadcast industry have standards in place to ensure that the loudspeakers and headphones used provide a consistent high standard of accurate monitoring?

## 1. Listeners Prefer Accurate Loudspeakers and Headphones

Significant research has been focused on the perception and measurement of loudspeaker sound quality over the past three decades. In the 1980's, Dr. Floyd Toole [2,3,4] at the National Research Council of Canada conducted controlled, double blind listening tests on loudspeakers. Listeners gave the highest fidelity ratings to loudspeakers having the flattest, smoothest frequency response measured over a wide range of angles. This was perhaps the first documented evidence that listeners recognize accurate sound and prefer it.

Toole and the author joined Harman International in the early 1990's where the research continued. State-of-the-art subjective [5,6,7] (see Fig. 1) and objective loudspeaker test facilities (see Fig. 2) along with a listener training program [8,9,10] were established to understand listeners' perception of loudspeaker sound quality and how to measure it. Over a 20-year period, hundreds of loudspeakers covering a wide range of prices, brands, and design-types were evaluated by large numbers of trained and untrained listeners from different cultures, geographical regions, and demographic groups [11,12] In 2012, the same research began in earnest for headphones. Some of the key findings from these collective studies are summarized below:

1) Both trained and untrained listeners preferred loudspeakers and headphones that they described as sounding the most neutral or accurate. This was true when tested across different cultures, genders and age groups regardless of whether listeners were trained or untrained [11, 12,13]

2) When listeners are asked to adjust the bass and treble control of an accurate monitor according to taste, younger and less trained listeners prefer slightly more bass and treble. Older listeners (> 55 years) prefer less treble and bass which is likely a compensation for age-related hearing loss [14,15]
3) The quality and spectrum of the program or recordings can influence the perception of monitor sound quality. Well-recorded programs with broad spectra produce the most discriminating and ratings [8, 16]
4) Listeners' loudspeaker and headphone preferences can be accurately predicted based on their measured frequency response as discussed in the next section.

Given the evidence that most listeners can identify accurate loudspeakers and headphones and prefer them, can we predict their perceived sound quality and accuracy based on their acoustic performance?

## 2. Predicting Loudspeaker Sound Quality From Objective Measurements

Three independent loudspeaker studies by Toole [2,3,4], Klippel [17] and Olive [18,19] found that the measured frequency response of the loudspeaker is the best indicator how good it sounds. Both phase and non-linear distortion measurements were found to be less useful, predictors of sound quality[1]. Loudspeakers radiate sound in all directions that eventually arrive at the listeners as direct and reflected sounds that influence the listeners' perception of its sound quality. Therefore, it's necessary to measure the loudspeaker at many angles around its primary axis. At Harman, we measure loudspeakers every 10 degrees in both horizontal and vertical orbits, and calculate spatial averages of the curves are used to predict the quality of direct, early and late reflected sounds can in a listening room (see Fig. 3).

Using the curves in Fig. 3, the author developed a mathematical model that calculates errors in the frequency response of the loudspeaker and predicts its sound quality preference rating in a controlled listening test [18,19]. The further the frequency response of the loudspeaker deviates from ideal linear behavior (flat on-axis, smooth off-axis) the lower its predict score. The accuracy of the predictions range from 86% (based on 70 different loudspeakers) to 99% with bookshelf loudspeakers with restricted low frequency output. It is noteworthy that the loudspeaker curves related to the direct and reflected sounds are equally weighted in the model, indicating that a meaningful loudspeaker standard must consider both components – not just the sound produced 0-30 degrees off-axis. Bass quality is also weighted 30% in the model, a warning that low frequency interactions between the room and the loudspeaker must be dealt with in order to maximize the listeners' pleasure.

Alternative predictive sound quality models based on loudspeaker measurements commonly practice in the industry (e.g. in-room loudspeaker measurements, low-resolution 1/3-octave measurements, and sound power) were also considered. All of them were found to produce less accurate predictions.

### 2.1 Predicting Headphone Sound Quality From Objective Measurements

A recent study by Jeroen Breebaart measured 283 headphones and found there was no correlation between their retail price and measured frequency response: the best objective indicator of how good the sound [20]. Clearly headphones are still in their wild-west stages of their evolution. Without sound guidance from science and standards, their performance and sound are highly variable. Fortunately, this is rapidly changing.

---

[1] Most well designed loudspeakers do not produce significant audible distortion until they are driven beyond their physical limits. Most listening tests are conducted at comfortable playback levels that do not reach those limits.

Three recent independent studies [21,22,23] have concluded that the current recommended around-the-ear (AE) headphone target response – the so-called the diffuse-field (DF) calibration) - is less preferred to alternative choices. Listeners perceive it as sounding too thin and bright. A preferred target curve one that simulates at the eardrum the in-room response curve of an accurate loudspeaker (like the one in Fig. 3) calibrated in a reference room. The scientific rationale is that most stereo recordings are optimized under these conditions, and such recordings will sound best through headphones tuned to the same target curve as the loudspeaker. This has been confirmed in several studies so far, and more are underway.

Predicting the perceived sound quality of a headphone from objective measurements is the current focus of Harman headphone research [24,25]. Much like loudspeakers, the measured frequency response of the headphone is an excellent indictor of how good its sounds (see Fig. 4). In a recent study we asked 71 listeners rate their sound quality preferences for 30 different models of in-ear (IE) headphones with a hidden reference (a modified version of the Harman target curve based on an accurate loudspeaker in a room. Using only the measured frequency response we were able to predict the listeners preference ratings with 91% accuracy.

Together the summarized research studies provide a scientific basis for choosing accurate loudspeakers and headphones. The listening tests show evidence that listeners recognize and prefer the more accurate models, and their preferences can be largely predicted based their measured frequency response.

## 3. International Standards For Choosing Loudspeakers and Headphones

Two popular current standards used for choosing and calibrating professional monitors in recording, broadcast, and audio research are the International Telecommunications Union's ITU-R BS. 1116.3 [26], and the European Broadcast Union's EBU Tech. 3276, 2nd edition [27]. Both are essentially identical and when examined in light of the scientific research presented in the previous sections are outdated and flawed. There are many issues throughout the standard but most egregious errors are the following:

1) It is recommend loudspeakers be measured using pink noise with 1/3-octave resolution. This is insufficient resolution to accurately measure the level and Q of resonances in the loudspeakers, which is needed to predict its sound quality [18,19].
2) The recommendation only qualifies the loudspeaker's frequency response over a ± 30° (horizontal) and a ± 10°, (vertical) window around its primary axis, essentially of ignoring what listeners sitting off-axis hear, and the quality of the perceptually important early and late reflected sounds.
3) No details are given on how or where the loudspeakers should be measured (anechoic chamber, control room, outdoors??) making it impossible to practice the standard in any consistent or meaningful way.
4) An operational room response window with wide tolerances is specified to account for acoustic interactions between the loudspeaker-room. However, no target curve is given meaning the loudspeaker can equalized to any arbitrary target response within the window potentially transfoming an accurate loudspeaker into an inaccurate one.
5) The recommended headphone must have a diffuse-field target curve as defined by ITU-R BS.708 [28] despite the current evidence that its doesn't sound accurate (too bright and thin in the bass) for stereo reproduction, and is less preferred to a headphone target curve based on an accurate loudspeaker calibrated in a reference listening room.

In summary, the ITU/EBU standard only guarantees the loudspeaker and headphone will produce sound – but not necessarily good or accurate sound.

Ironically the consumer loudspeaker industry currently has a better standard for measuring loudspeakers in ANSI/CIA-2034-A [29], "A standard "Standard Method of Measurement for In-Home Loudspeakers" recommended anechoic measurements based on the loudspeaker research at the

National Research Council and Harman International. The standard produces measurements like those shown in Fig. 3 that can used to interpret and predict how good and accurate the loudspeaker sounds.

## 4. Where Do We Go From Here?

We've learned that most listeners regardless of training, age or culture recognize and prefer accurate loudspeakers and headphones that can be predicted by measuring the acoustic performance of the devices. Unfortunately, these types of measurements are not usually available from product manufacturers and reviewers, and the current international standards for recording and broadcast do not specify them. Consequently, there is little spectral uniformity between the loudspeakers and headphones used to create and reproduce the art, meaning the consumer rarely hears the music as the artist intended. We are trapped in a "circle of confusion". So, what can be done to eliminate the confusion and bring order, quality control and accountability to our industry?

One solution would be to for the ITU/EBU to adopt the current ANSCI/CIA-2034 standard since the loudspeaker measurements are proven to characterize how good the loudspeaker sounds in small to medium sized rooms. Some tight targets with tolerances (e.g. ± 1.5 dB) would be added for on-axis, listening window sound power curves (see Fig. 3), and perhaps a minimum fidelity score (85%) based on the predictive model [18,19] could be specified. To control loudspeaker-room bass variations the industry needs to adopt a standardized in-room loudspeaker target response and calibration procedure to ensure a more uniform quality of bass in produced and reproduced sound [1].

The ITU/EBU diffuse-field headphone standard neither sounds accurate nor is it compatible with stereo recordings produced through accurate loudspeakers in a listening room. Replacing it with a headphone target response based on an accurate loudspeaker-room target curve like the Harman headphone target response [24,25] would improve the sound quality and spectral uniformity of recordings when switching between loudspeakers and headphones. With these new standards in place, recording and broadcast professionals, as well as audio enthusiasts would finally have a scientific basis for choosing their loudspeakers and headphones, and improving the reliability and quality of their audio experiences.

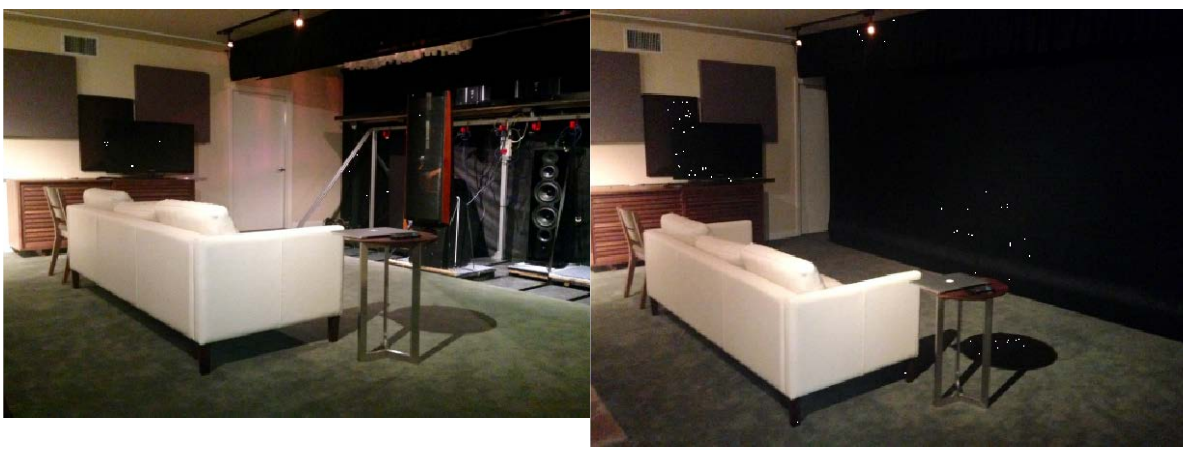

Figure 1 Hundreds of loudspeakers have been subjectively evaluated in Harman's multichannel listening lab (MLL) where listeners make double blind comparisons of different loudspeakers in the same position via an automated speaker shuffler. Important listening test nuisance variables like sighted biases, relative loudness differences, and loudspeaker/listener position are controlled or removed from the test.

Figure 2 Acoustic measurements of the loudspeakers like the JBL M2 monitor are measured in the large anechoic (reflection-free) chamber at the Harman Acoustic Laboratory in Northridge, California. The acoustic measurements allow accurate predictions of listeners' sound quality preferences in listening rooms.

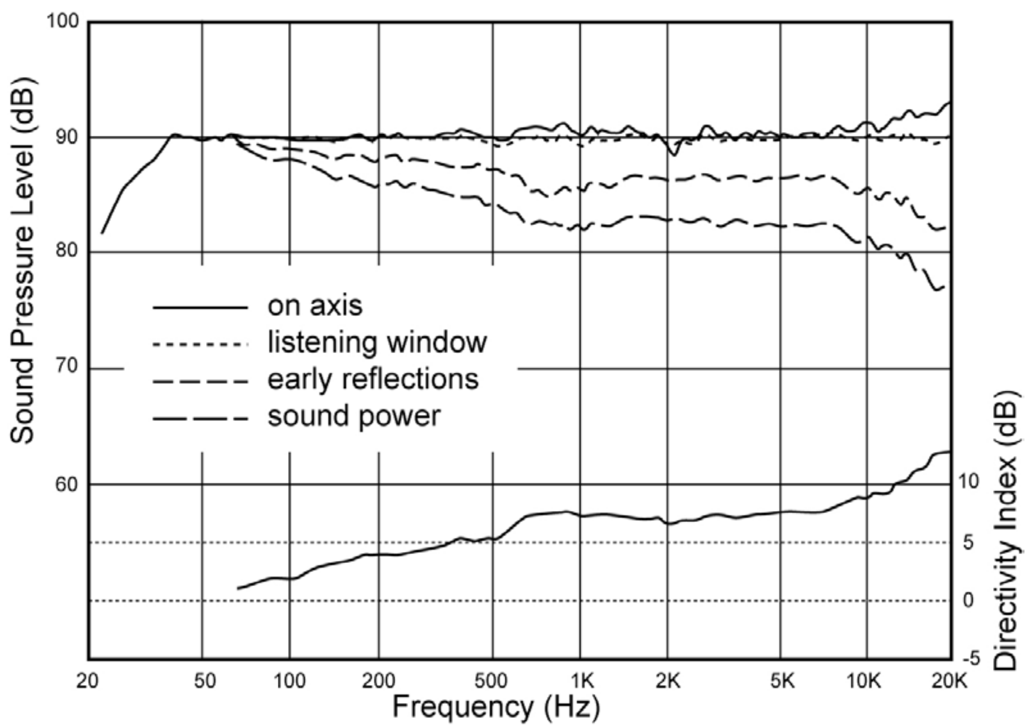


Figure 3 The so-called spinorama of the JBL M2 after it's been measured in the anechoic chamber. The set of frequency response curves represent the quality of direct sound produced by the loudspeaker for

someone sitting on-axis, within the listening window, and the early and late reflected sounds. The bottom curve is the directivity index (from Toole [4])

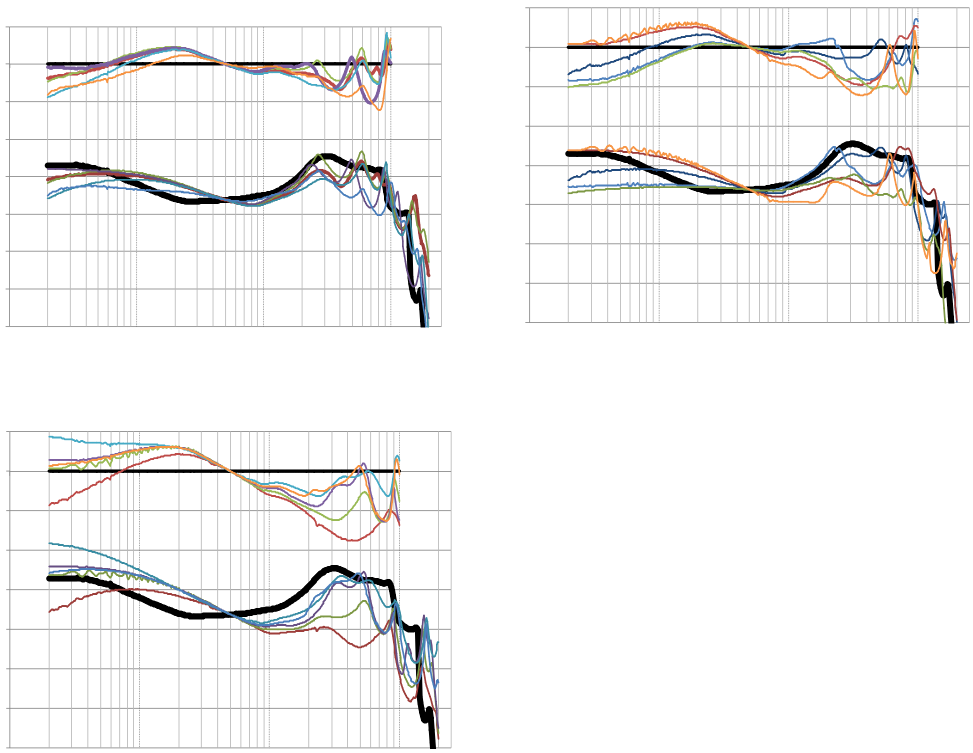

Figure 4 The frequency responses of 15 in-ear headphones plotted in three categories of sound quality (good, mediocre and poor) based on preference ratings of listeners **[24].** At the top of each graph is the error response curve: the difference in frequency response between the headphone and the preferred Target curve.