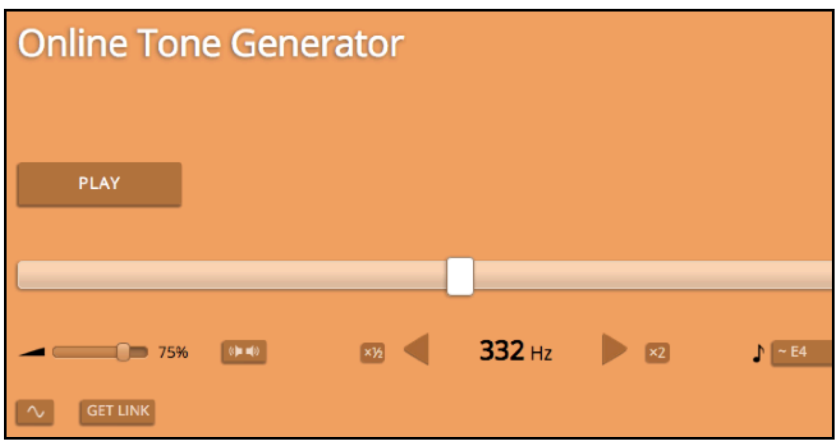
Headphone Essentials 4:

The Skinny on Headphone Frequency Response Graphs

If you are reading this document, you are likely interested in finding a headphone that matches your listening preferences. The single most important factor is the headphone's frequency response (FR). This is often called its tuning or tonality. Other factors, such as its dynamic responsiveness, sound staging, imaging and detail reproduction are also important considerations. But if the tuning is not agreeable those other factors are not likely to rescue it.

If you would like to start with more foundational background material on the nature of sound, see my tutorials Basics of Musical Sound and Basics of Headphone Sound at http://daystarvisions.com/Music/index.html.

Getting started



To get started go to https://www.szynalski.com/tone-generator/. Make sure the volume on your computer is initially set to a fairly low level, then press the space bar once to turn sound on, then move the slider fairly slowly and evenly left and right. If the sound does not change in loudness you're hearing what's called a flat response. If it grows louder and softer as you move the slider, that's a non-flat response. If possible, you may want to try this with more than one headphone or speaker to see how they vary (remembering to start at a low loudness level with each one). By the way, as is typical of tiny speakers, the one built into my laptop doesn't produce any sound at all to the left of about 100 Hertz.

Now let's say we want to record what we're hearing when sound is being produced by a particular pair of headphones. Let's say we start at the far left and hear nothing until maybe 22 Hertz. Then we hear it gradually getting louder until 67 Hertz, then quickly getting quieter until 83 Hz, then starting to climbing again. We might draw a line like so to represent this:

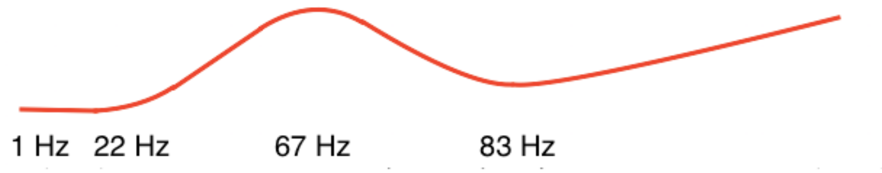


Fig. 1: freehand loudness line

What we're intuitively doing is recording the Hertz numbers from left to right, just as they appear when we move the slider from left to right. Then we're showing the change in loudness as an up and down movement. And that's all there is to a frequency response graph:

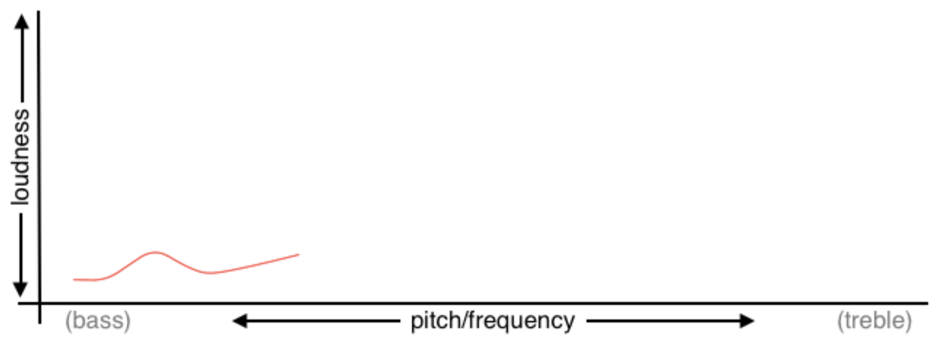


Fig. 2: frequency response graph axes

We can see that if we want to record the actual Hertz numbers we would need a very long graph from left to right, since the Hertz numbers go from 1 to 20,154. The math trick to deal with that is to use a logarithmic scale. This simply means the distance keeps getting shorter as the numbers keep getting larger:

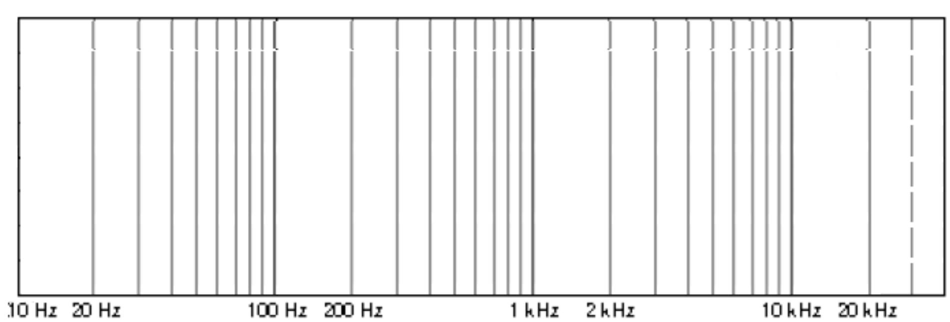


Fig. 3: frequency response graph with logarithmic Hertz scale

Fig. 3 has a logarithmic scale from left to right. From 10 to 100 there are guide lines on the tens. From 100 to 1000 they're on the hundreds. From 1000 to 20,000 they're on the thousands. Notice that for convenience we use the kHz notation to get rid of the increasingly large number of zeros, so 1 kHz = 1000 Hz. The scale concentrates on the range of 20 to 20,000 Hz since this is the range of human hearing. (Keep in mind that the highest frequency you can hear decreases gradually with age. So 20,000 Hz is highly optimistic.) Also notice that the distance between 10 Hz and 100 Hz is exactly the same as the distance between 100 Hz and 1000 Hz. And both are exactly the same as the distance between 1000 Hz and 10,000 Hz.

That takes care of the left/right scale. For up/down we can use logarithms again. But the units we need for loudness are notated as dB, which stands for decibels:

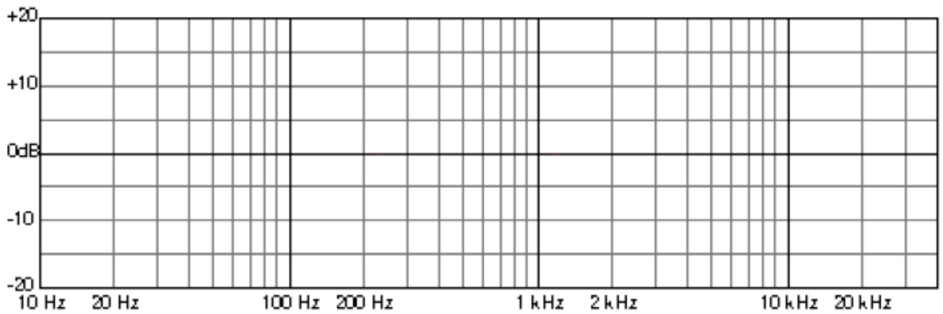


Fig. 4: FR graph with decibel loudness scale

In this case, any 10 decibel change indicates a doubling of loudness. And 1 decibel is about as fine a discrimination as we humans can reliably make. This might seem like an improvement until we actually try to use it. We can get a precise frequency number from the Online Tone Generator. But how do we judge how high or low to place each loudness measurement?

Measurements done right

For this we need some fancy, technical, and extremely expensive measuring equipment ($20,000+) to achieve reliable results. Fortunately, there are several good samaritans doing such measurements and posting them on the web. An excellent example is crinacle.com (https://crinacle.com/graphs/headphones/graphtool/). Here's the Crinacle measurement for the still-popular Audio-Technica ATH-M50x headphone:

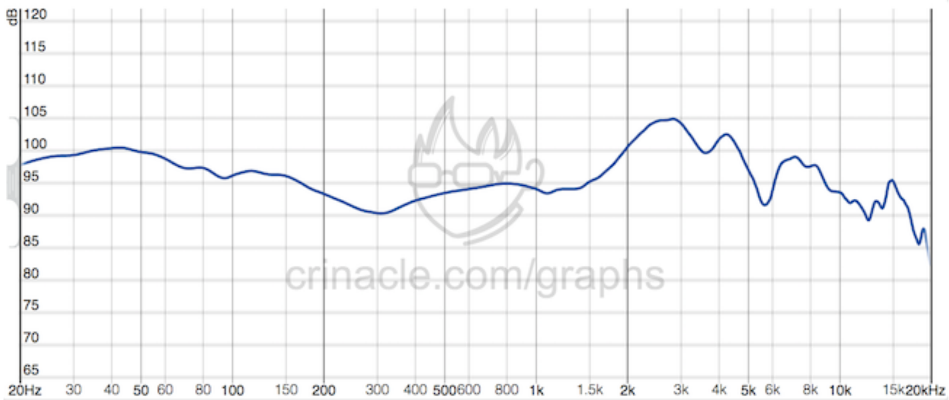


Fig. 5: Crinacle FR graph for Audio-Technica ATH-M50x

Unfortunately, there's a slight gotcha. The blue line shows changes in loudness going across the human hearing range of frequencies. But, since it shows actual measurements, the line is much more of a jagged rollercoaster ride than your ears would lead you to believe. Also, our ears and brains are designed to hear the faintest possible sounds that are critical for survival. The peculiar shape of our ears are in fact natural amplifiers for high frequency sounds. Here are the results of a fairly recent study:

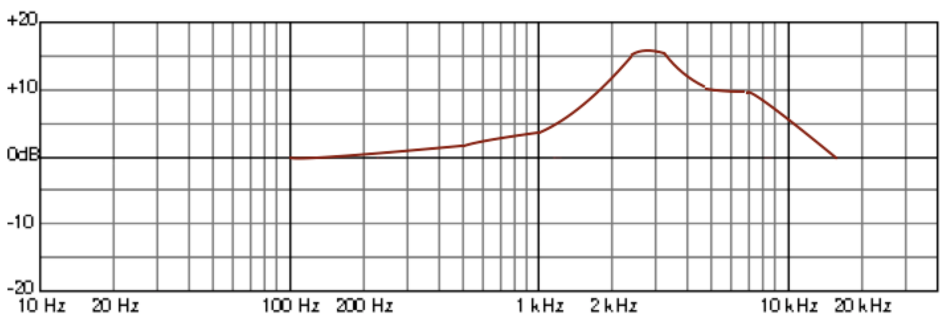


Fig. 6: Hammershøi and Møller, ear amplification in a diffuse sound field, 2008

In theory a headphone that exactly re-produced this measurement curve would produce a perfectly even, or flat-line, sound if we used it to listen to the Online Tone Generator. In other words, our brains translate the curve shown in Fig. 6 into the subjective "curve" shown in Fig. 7:

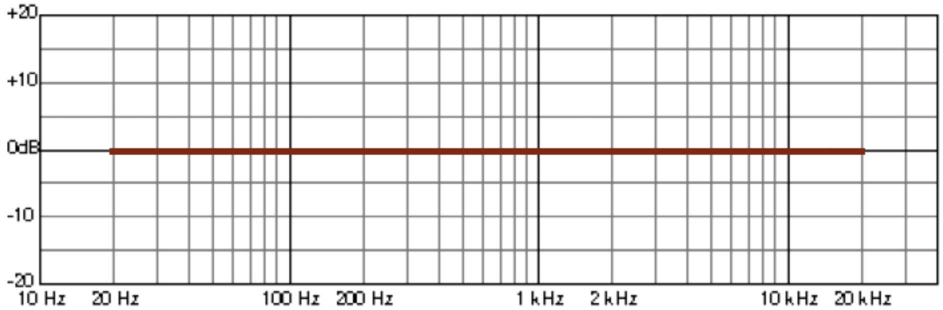


Fig. 7: subjective experience of a perfectly even loudness across the frequency range

If we overlay the H & M ear amplification curve from Fig. 6 onto the M50x measurements of Fig. 5 we get:

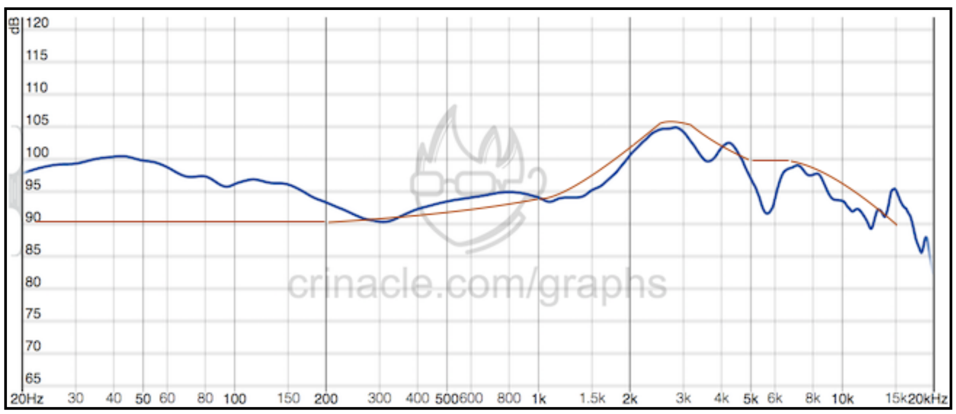


Fig. 8: ear amplification (red) vs M50x frequency response (blue)

The sheer number of ups and downs in Fig. 8 may look like a problem. But it's actually about average for today's headphones (especially if they're a closed-back design like this one is). Our ears can't hear the small up/down changes, and manufacturers simply don't have the technology to do better.

What we should hear if we use a pair of M50x to listen to the Online Tone Generator is an increase in loudness where the blue line climbs above the red line. Plus we should hear a decrease in loudness wherever the blue line drops below the red line. And that's pretty much what I do hear when I do that. Except for whatever reason I don't hear the huge rise above the red line on the left side of the graph from 20 to 250 Hz. From there on I do hear the rises and dips shown.

Another excellent source, especially for lower-priced headphone measurements, is Rtings.com, https://www.rtings.com/headphones/1-4/graph#295/4011. Here's their measurement presentation for the M50x:

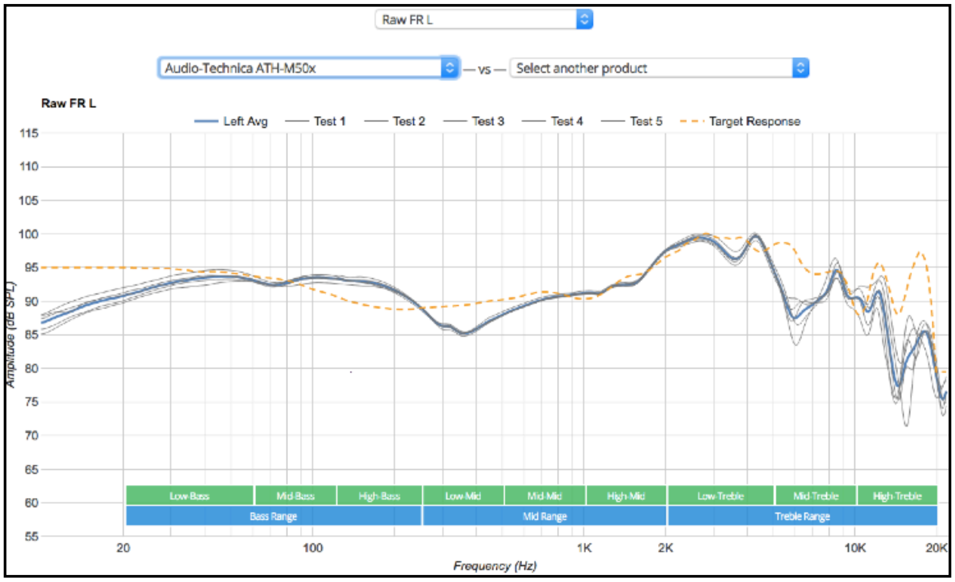


Fig. 9: Rtings.com measurements for M50x

One difference is that Rtings show measurements between 10 and 20 Hz on the far left. Mostly, we would perceive these as vibrations rather than sound. Another difference is the dotted orange line. This is their (non-standard) reference for an ideal headphone sound response.

A third site with excellent data is Oracle1990's reddit page: https://www.reddit.com/r/oratory1990/wiki/index/list_of_presets:

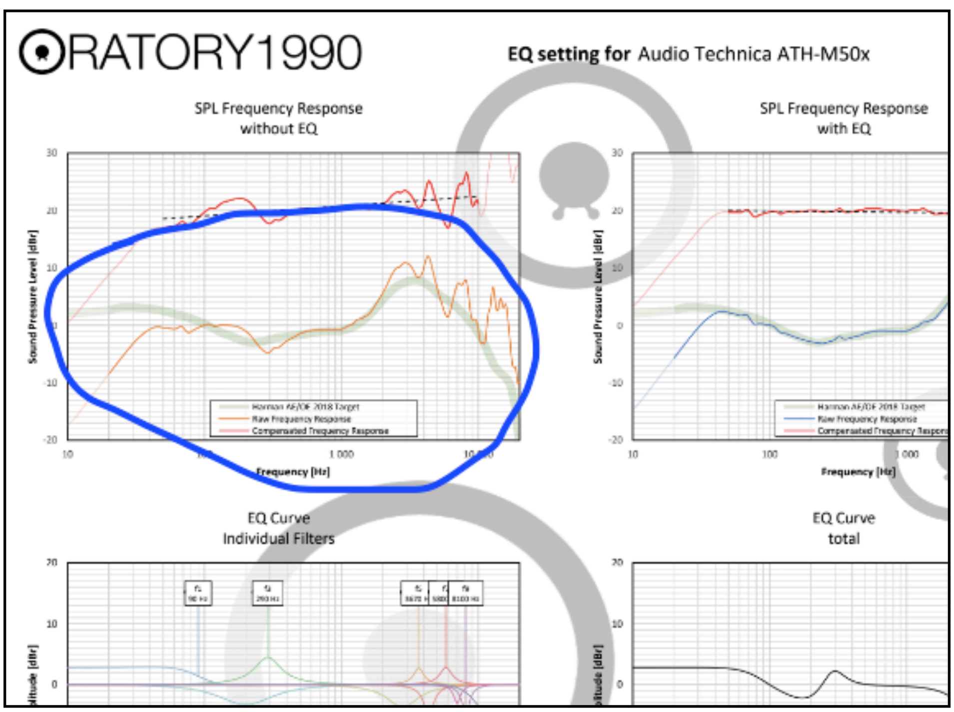


Fig. 10: Oratory1990 page for M50x with FR graph circled in blue

This site focuses on headphone EQ (equalization), which is beyond the scope of this tutorial. But the measurements we want are on the top left of each headphone page.

Here the frequency response is the orange line. The broad green line is yet another headphone reference curve called the Harman Target.

(Note: As I write this, Oratory1990 is using the most well-regarded and most accurate equipment. So when you find the same model headphone being measured by more than one of these sites, you might want to take the Oratory1990 measurement as being the most definitive.

Finally, there is an older site, innerfidelity.com (https://www.innerfidelity.com/headphone-measurements), with an extensive collection of headphone measurements that look like this:

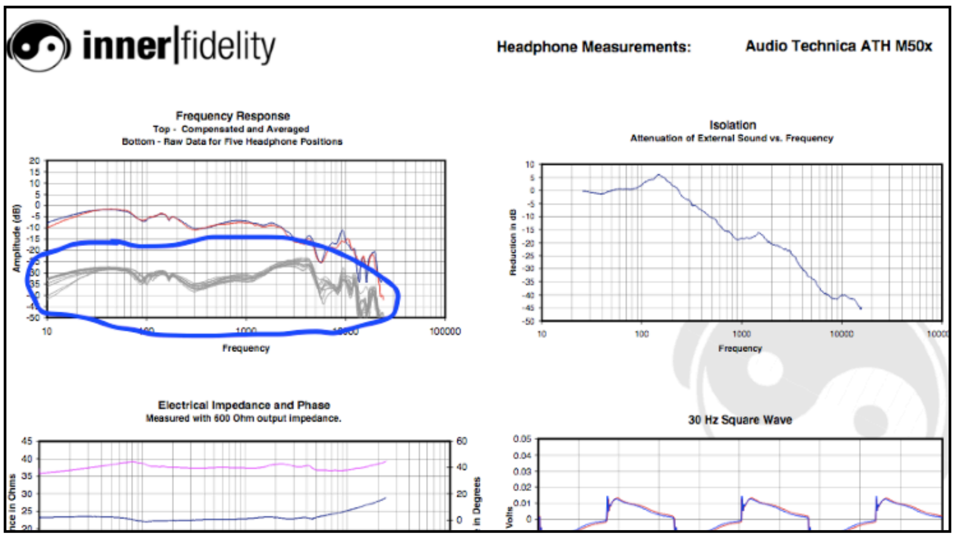


Fig. 11: innerfidelity.com measurement page for M50x, FR graph circled in blue

Again, the FR graph is upper left. But confusingly it contains sets of two lines. The grey lines are multiple trials of measurements for the M50x with the headphone moved slightly for each one. (Rtings, Crinacle and Oratory1990 use the same multiple-trial procedure.) But the red line above the grey lines is a modification of the grey lines. It adjusts for the properties of the measurement equipment by comparing them to a now-obsolete reference. Simply ignore the red line.

I present innerfidelity last because the site changed hands a few years ago and no new headphones have been measured since. Plus, the equipment used was an older design and somewhat inaccurate for higher frequencies. So I recommend preferring one of the three previous sites, if the headphone you're researching can be found there.

Measurements done less right

The above are competent and useful sources of information. Far more often, however, you'll encounter measurement graphs done by amateurs using inexpensive equipment.

Here's Oratory1990's measurements for the recent AKG K371:

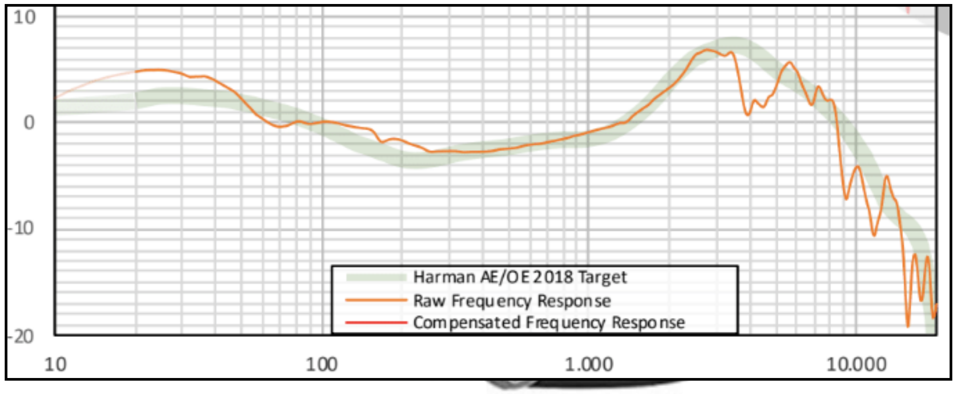


Fig. 12: Oratory1990 FR graph for AKG K371

This headphone was designed to reproduce something called the Harman Target frequency response as closely as possible. As we can see it does so admirably. Here is an amateur measurement of the K371 I found by using Google:

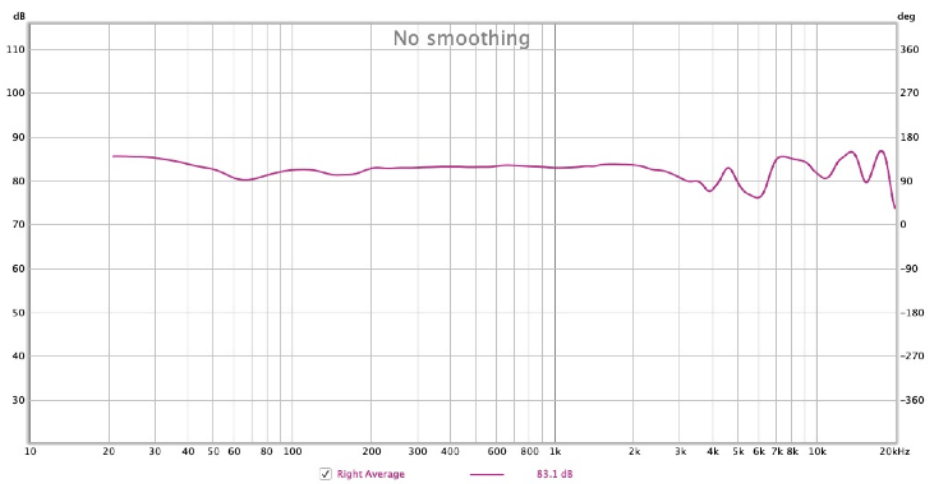


Fig. 13: MiniDSP EARs FR graph for AKG-K371

This was done using an inexpensive and popular tool called the MiniDSP EARs. We may scratch our heads as to how to relate the purple line in Fig. 13 to the orange line in Fig. 12. They both claim to show the frequency response of the same model headphone. The first thing we need to know is that we're not seeing the actual data in Fig. 13. What we're seeing is a plot of the *difference* between the actual measurements and some unspecified reference curve. And, as we've already seen, reference curves vary all over the proverbial map. The graphs we've been looking at previously were what's called raw or uncompensated. Fig. 13 is called a *compensated* FR graph.

The MiniDSP EARs tool comes with two different reference curves. One called HPN (headphone neutral) is somewhat similar to the diffuse field curve. The other, called HEQ (headphone equalization), is somewhat similar to the Harman Target. Given the relative flatness of Fig. 13 we can make an educated guess that the HEQ curve was used. But that information is rarely provided. In Fig. 12, looking from left to right, we see three obvious discrepancies between the Harman Target and Oratory1990's measurements of the K371. There is a rise from

20 to 50 Hz, a smaller rise from 100 to 250 Hz and the dramatic dip from 3.5 to 5.5 kHz. Looking at Fig. 13, only the first of these shows up. From 100 to 250 Hz we see nothing but a slight dip. From 3.5 to 5.5 kHz we see a gentler dip, rise, dip.

Here's another MiniDSP EARs graph of the K371:

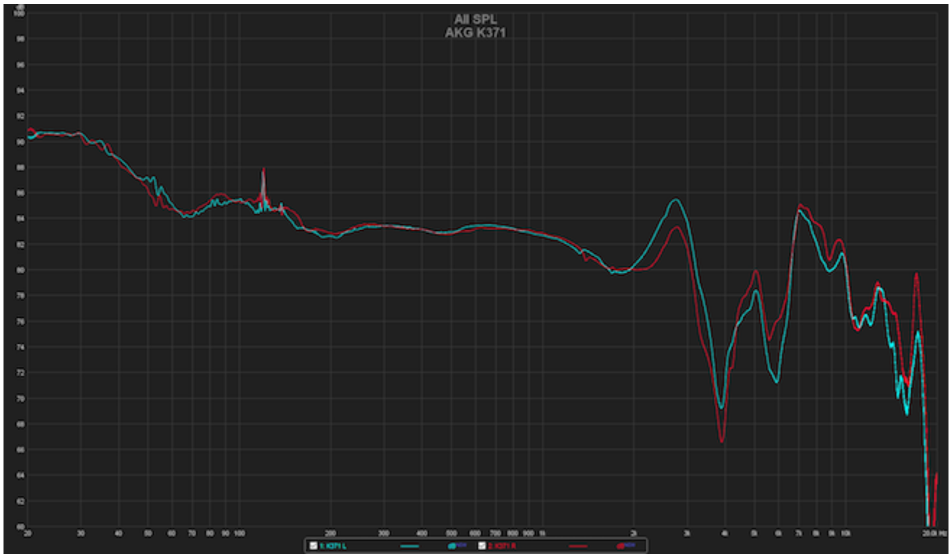


Fig. 14: Metal571 MiniDSP EARs FR graph for AKG-K371

This one was done by YouTube headphone reviewer Metal571. Metal is a careful and thoughtful guy. He makes sure to tell us that the MiniDSP EARs graphs are seriously limited (especially in the region around 4.5 kHz). He also makes sure to tell us that he's using the HPN compensation. Don't be confused by the dramatic black background. That's purely cosmetic. The overall downward slope of the measurement curve is due to the increasing difference between the diffuse field curve and the Harman Target curve in the ear amplification area above 1 kHz.

But another major difference between Fig. 13 and Fig. 14 is the vertical scale. In Fig. 13 each horizontal reference line is 10 dB above or below the next. In Fig. 13 they're only 2 dB apart. In consequence, Fig. 13 tends to downplay the rises and falls of the measurement curve. Fig. 14 highly exaggerates them.

This scaling effect can be used to great effect to put a headphone in a good light. The headphone forum head-fi.com occasionally produces FR graphs of headphones released by sponsoring companies. The head-fi measurement equipment is the best that money can by and they know how to use it. Accuracy is not at issue.

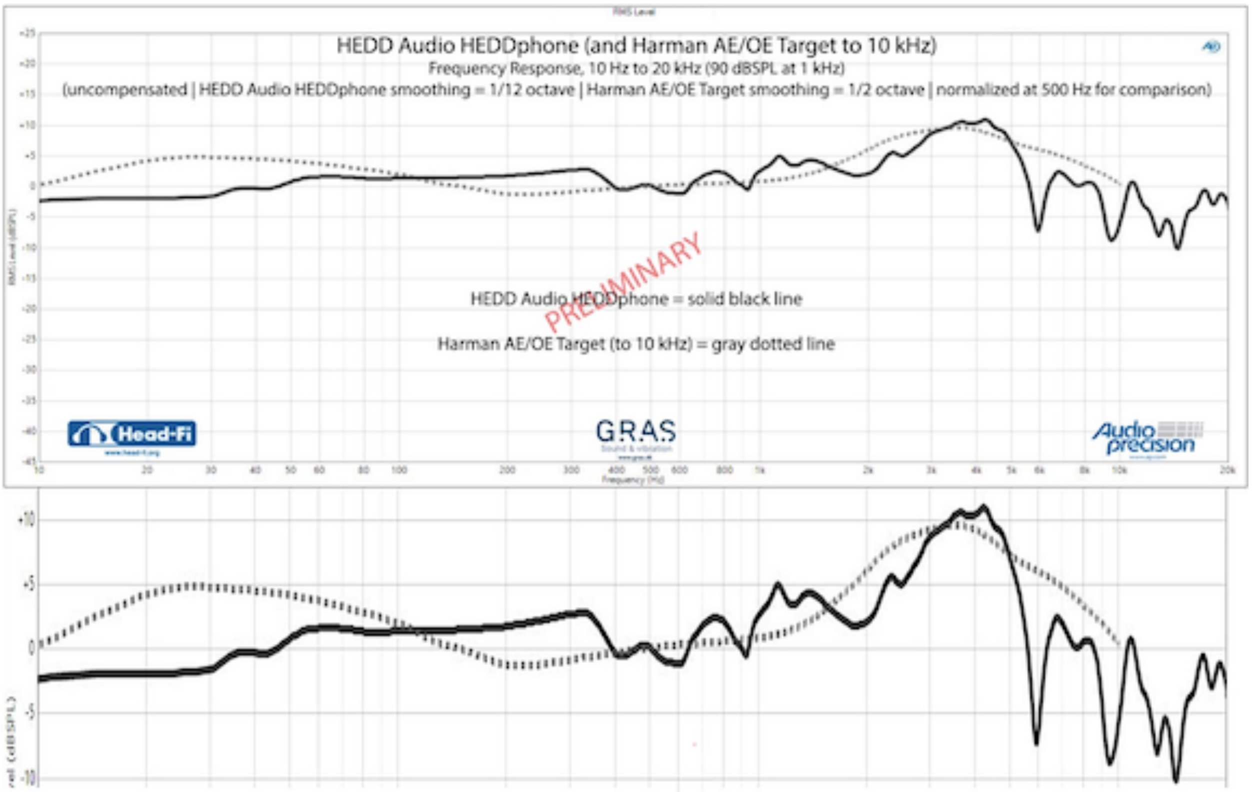


Fig. 15: Head-fi measurements for the HEDDphone

The upper graph in Fig. 15 is directly from the head-fi website. This particular headphone currently sells for $2500US. The solid black line are the raw measurement data. The dotted black line is the Harman Target. Notice how shallow the Harman curve is compared to what we're used to from the green line version in Oratory1990 graphs. For the lower graph in Fig. 15 I've simply taken the upper graph and stretched it vertically so the horizontal reference lines are now a more typical distance apart. The Harman Target curve now looks normal. Same data, dramatically different presentation. Another example of vertically compressed presentation is the MiniDSP EARs graph in Fig. 13 (while Fig. 14 is just the opposite).

The small print in the Head-Fi graph legend says the data are presented with 1/12th octave smoothing. This is frequently done. It simply acknowledges the limitations of even the best measuring equipment and measuring techniques. However, from time to time you'll see FR graphs with much smoother measurement curves, even as extreme as 1/3rd octave. When the graph line looks smooth and simple, be suspicious that more aggressive smoothing has been done.

In short, it's the wild west out there in the land of headphone FR measurements. We have to be wary of:

- Amateur-grade test equipment
- Unspecified compensation curves applied to the data
- Loudness scale compression applied to the data
- Excessive smoothing applied to the data.

Do even good measurements agree?

Even if we choose to focus our attention on the three or four quality measurement sites for frequency response graphs that doesn't mean their results are completely interchangeable.

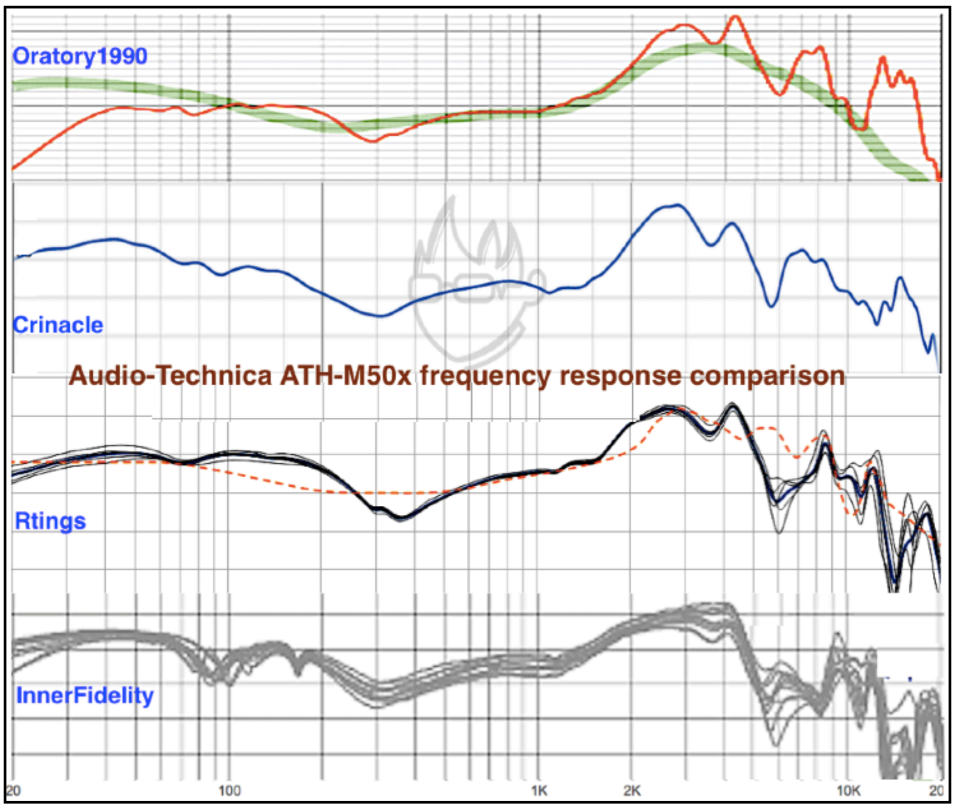


Fig. 16: four-way comparison of measurements of the same headphone model

In Fig. 16 we see a lot of agreement but also some disagreement. There will often be disagreement in the very lowest frequencies. This is due to the difficulties getting the ear pads to rest completely on the side of the measurement head surrounding the artificial ear. Equally, there are disagreements in the high frequencies above 2 kHz. This is at least in part due to differences in the artificial ear canal's properties from one measurement head to another.

But another factor is unit-to-unit product variation. Suppose you bought more than one unit of exactly the same headphone that you find right beside each other on a store's shelves. Suppose you then sent them to Oratory1990, Crinacle or Rtings asking for measurements. Experience shows there would be noticeable differences between the three units. The amount of difference tends to vary from manufacturer to manufacturer. Sennheiser has a reputation for minimal variation, for example. So when we look at Fig. 16 we have no idea how much of the discrepancies between the four measurement curves is due to the test equipment being used and how much is due to unit-to-unit differences.

Yet another factor is what we call silent revisions. Manufacturers occasionally change internal components or assembly procedures over time in the production of a particular headphone model. A common scenario is for a manufacturer to experience a significant number of returns after releasing a new model headphone. The manufacturer will investigate what the failure was then fix the problem, if possible, in the production of that model going forward. (It's always wise to hold off for several months on purchasing an exciting but brand new product.)

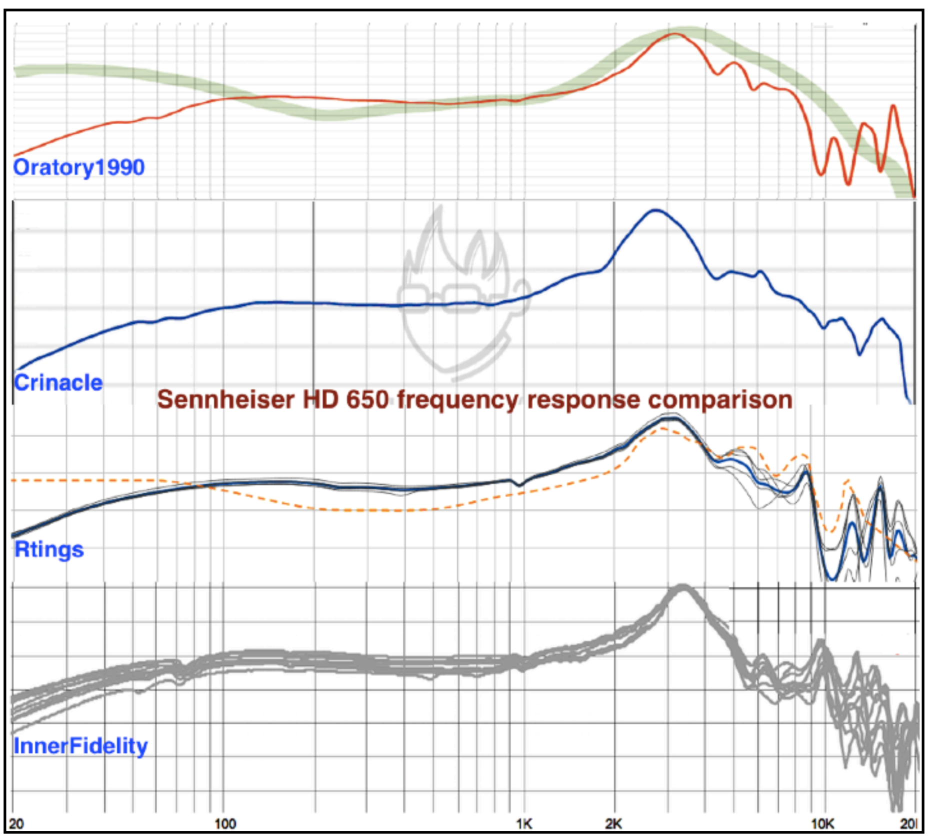
Silent revisions, however, can occur at any point in a product's life cycle. Just one cause is a component supplier going out of business. So any difference between an Oratory1990 vs Crinacle vs Rtings, etc. measurements may be due to actual differences in the sample of the headphone model being tested.

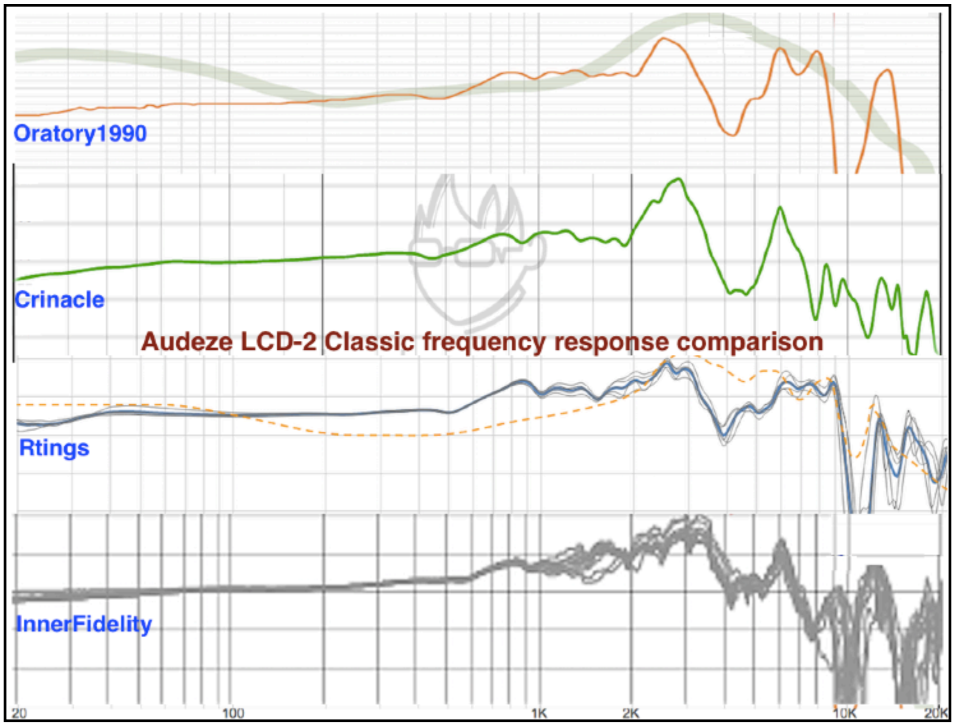
✣ ✧ ✣ ✧ ✣ ✧ ✣

And that's it. You're now far more knowledgeable about headphone frequency response graphs than 90% of the Internet's self-appointed expert commentators. Be sure to proceed on to the next exciting instalment of my must-read Headphone Essentials series: HE-5 The whole flat/neutral/Harman headphone thing (http://daystarvisions.com/Music/index.html).

Bonus

Here are two more four-way comparisons to illustrate both their similarities and their differences. Again, we don't know which combination of measurement equipment variation, unit-to-unit variation and silent revisions to attribute any discrepancies to:




Oratory1990
Crinacle
Audeze LCD-2 Classic frequency response comparison
Rtings
InnerFidelity
20
100
1K
2K
10K
20K