Headphone Essentials 5:

The quest for a reference (over-ear) headphone

Note: this document is part of an instructional series. If you would like to start with more foundational background information, see *Basics of Musical Sound* and Basics of Headphone Sound at Headphone Essentials. This document makes extensive use of headphone frequency response graphs. To learn your way around them you'll find an in-depth exploration in The Skinny on Headphone Frequency Response, but you'll find a quick intro to the topic below as Appendix A.

Lately as I write this, the phrase ***reference headphones*** has been cropping up on forums. This typically seems to refer to a particular frequency response, presumably one that accurately reproduces the frequency response of a recording — also called *flat* or *neutral*. But logically, a reference headphone should also correctly reproduce all other aspects of the audio signal, such as detail resolution, dynamic range, dynamic impact, spatial dimensions and spatial placement as we'll explore below. But first —

1. The reference frequency response enigma

Background

If recordings were engineered to a known frequency response standard, the concept of a reference frequency response would make some sort of sense. In fact, the current practice for mass market recordings is to create a frequency response that is the least problematic possible on as many of the playback devices as possible that the target audience will likely employ, from sound bars to ear buds. The typical consumer playback device is lacking in some amount of both bass and treble. The built-in speakers of a smartphone or laptop typically have no bass at all below 100 hertz, while consumer enhanced car audio often seems to have nothing *but* bass response. How you engineer for extremes like that is an arcane mystery.

Only in the recordings of all-acoustic performances, such as classical, orchestral, jazz and similar are we likely to find a more traditional approach to accurately re-creating the live experience. Yet, this is typically achieved by mixing the raw recordings generated by microphones placed both near to and at a distance from the sound sources. The mix is then judged for accuracy by listening to playback on loud speakers (not headphones) in a room acoustically treated to remove unwanted frequency emphases. So far as I know this is not done by any consideration of recorded signal strengths and distances, but simply by memory and by ear.

Crucially, recording engineers working on all-acoustic content still have to assume that the same extremes of playback equipment will need to be accounted for. All of which brings us to falling back on the old trope of good loudspeakers in a good room. In other words, a mythical ideal 2-channel stereo system. The best we can hope for is that most recordings have been engineered to not disappoint the 2-channel enthusiast.

Even from this minimal analysis of recording studio issues two points fall out that will prove very useful going forward:

The Principle of Sufficient Bass: When the goal is accurate/reference reproduction of recordings extra bass and sub-bass do not need to be added by the playback system. The recording engineer will have already added a sufficient amount to compensate for the bass roll-off of the large majority of playback systems.

The Principle of No-Trouble Treble: anatomical differences in human treble response at the eardrum mean that there will always be a segment of the population that is sensitive to any given frequency past about 2 kilohertz — especially at over-loud playback levels. Recording engineers have to be very conservative in how much treble loudness they introduce. Better the listener hears too little treble than too much.

Loudspeakers and ear gain

The loudspeakers used in a typical recording studio are called ***studio monitors*** and produce roughly the same loudness for all frequencies from bass to treble — but only when they are placed in an echo-free environment or are equalized to replicate one:

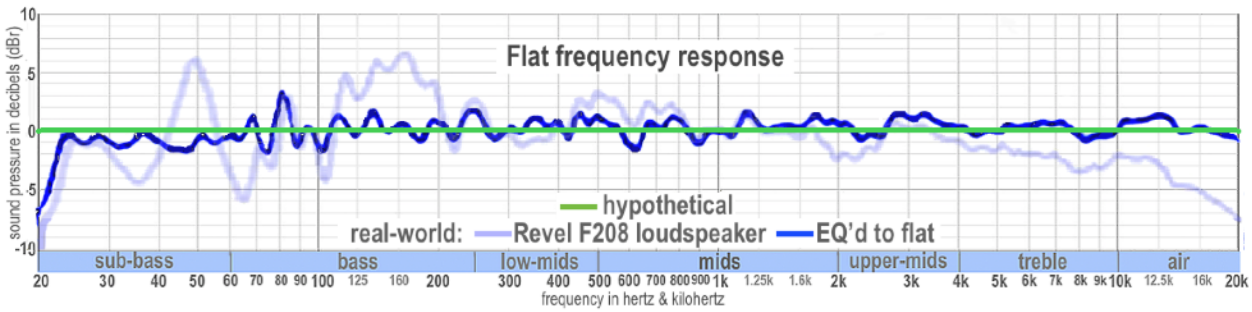


Fig. 1: hypothetical flat versus a very good approximation (source: TenBrook)

This is likely the origin of the common vocabulary that flat = accurate. We might therefore assume that a headphone should also be designed to generate the same loudness from bass to treble. We might expect that when a recording that sounds accurate on studio monitors is played back on such a headphone, the result will also sound accurate. But, in fact, any early attempts to tune headphones to equal loudness for all frequencies would have demonstrated that the result is very obviously not even close to being either flat or accurate.

The problem proves to be two-fold. First, what measures flat to a microphone placed in the vicinity of a loudspeaker in an echo-free environment does not also measure flat when measured at the human eardrum:

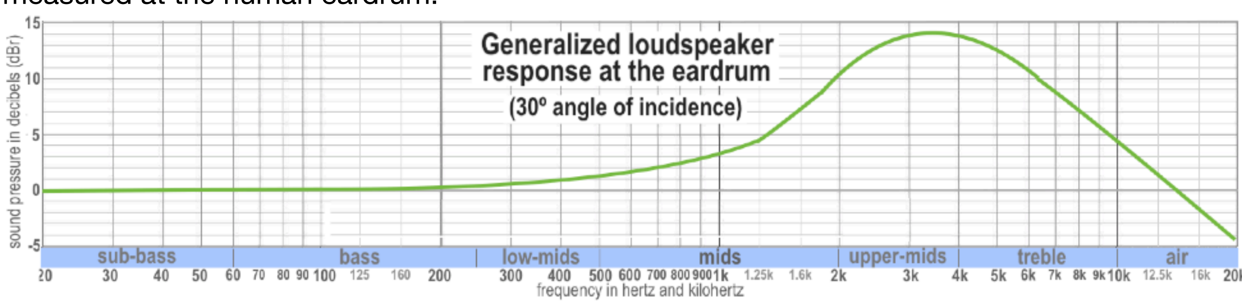


Fig. 2: approximation of Fig. 1's EQ'd loudspeaker but measured at the eardrum

The green line in Fig. 2 reports the same source sound as the green line in Fig. 1. The large hump on the right in Fig. 2 is the result of the selective amplification of the incoming sound waves as they interact with our heads, torsos, outer ears and ear canals. The longer and therefore deeper-pitched sound waves are reflected off our bodies and ears to some degree. But the shorter, higher-pitched sound waves proceed unimpeded. This is typically, if not quite correctly, called ***ear gain***. (Think of the phrase *ear gain* as shorthand for total sound pressure measured at the eardrum.) Nature really wants us to hear high frequency sounds extra clearly.

Ear gain is what the inner ear transmits to the brain, not what we finally perceive. Once the brain has extracted all that high frequency detail, it massages the signal that it sends to our conscious experience back to being perceptually flat in loudness — but with the extra detail now baked in. Again, our brains receive and process the input signal of Fig. 2, but what we consciously hear is right back to Fig. 1. The massive high frequency ear gain that the brain is expecting as input is also missing from the sound produced by a headphone naively engineered to match a loudspeaker's (anechoically) flat response.

But notice the connection with the Principle of Sufficient Bass: there is no elevation on the left side of Fig. 2. Our brains are designed to make do with flat line bass.

Note: we need to bother with graphs of the response at the eardrum, not only to understand why headphones can't be designed to a simple flat line response, but also because there is no way to measure how headphones sound in an open room, as is done with loudspeakers. They are designed and only function properly when coupled to a human head.

So now we might think the solution is to engineer headphones to have the ear gain shown in Fig. 2. After all, this is what a flat loudspeaker sounds like to the human ear. But headphones engineered to sound like Fig. 2 proved to have the opposite problem: they were now too loud at the peak of the ear gain hump.

Room gain to the rescue

Sennheiser solved this problem in the 1990's with the release of first the Sennheiser HD 580, soon followed by the similar-sounding HD 600:

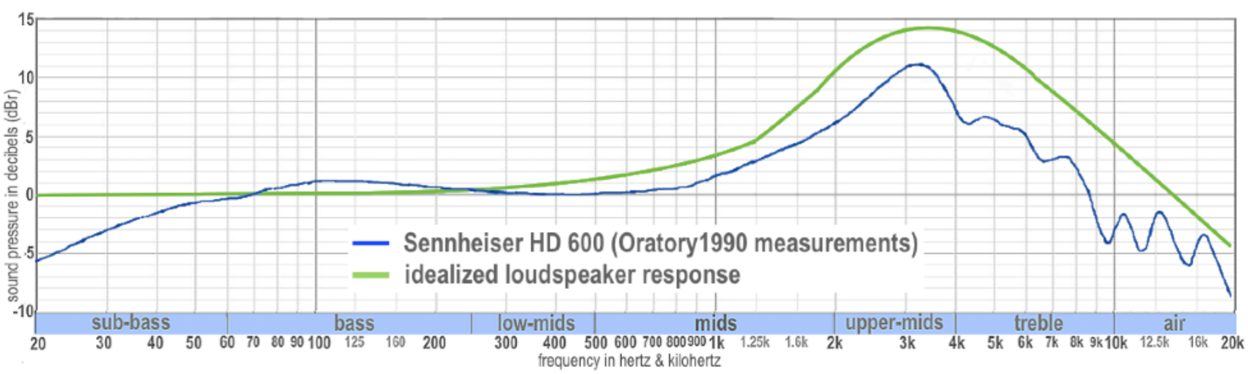


Fig. 3: (source: all Oratory1990 headphone measurements here)

While the sub-bass falls off on the left and the high frequencies on the right look jagged (inevitably, due to the unique shape of each human ear), this was so close to re-creating the sound of most recordings in what the majority of headphone enthusiasts found to be an accurate way that the HD 600's are still in wide use today, including in recording studios around the world.

Sennheiser's secret sauce (presumably arrived at by thinking along these lines) was to honour the good loudspeakers in a good room mantra by introducing a further variable called ***room gain***. Fig. 2 shows us what ear gain does to the flat line response of an ideal loudspeaker measured in an ***anechoic***, or echo-free, environment — essentially equivalent to the outdoor world. In fact, music is rarely created in an anechoic environment, but rather in enclosed spaces ranging from concert halls to night clubs to recording studios, each with its own natural reverberation.

The key concept to room gain is that bass frequencies literally bounce off the walls, while mids and high frequencies do so much less. This is just the opposite of what we encountered with ear gain. In fact, an over-the-top excess of bass at one seating location and a nearly complete loss of bass near-by (where bouncing bass waves collide rather than reinforce) is the bugaboo of many a home stereo system. A whole section of the audio industry exists to provide room sound modification devices, such as absorbers and diffusers, to minimize such problems. But what is the frequency response of a hypothetically well-treated, problem-free room? Again, the system owner is trying to re-create the original sound of the live performances from the audio signal contained in a variety of recordings at some hypothetically ideal audience seating location.

The Danish acoustics firm Brüel & Kjær (pronounced much like broil and care) researched this topic back in 1974, resulting in the Brüel & Kjær Optimum HiFi room gain curve:

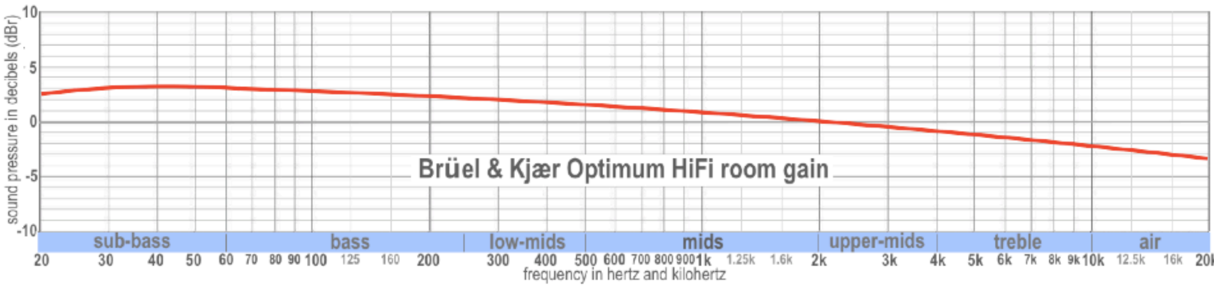


Fig. 4: (source: Brüel & Kjær)

Many alternate room gain curves have been proposed, including a straight-line one decibel per octave slope and the one used in the very different (and very controversial) Harman over-ear target. Both these alternates have more bass than the B&K curve. But, again: recording engineers can't assume the end listener will be using a bass-heavy playback system. Recording engineers certainly can't assume home stereo listeners will have the sub-woofers needed to address sub-bass roll-off.

Nevertheless the **Harman** research provides some valuable cross-checks to what we're doing. To explore further, see Appendix B.

So we'll go with the time-honoured B&K room gain — but subject it to cross examination. Just bear in mind that most generic playback systems have heavy roll-off in the sub-bass and often in the mid-to-upper treble, as well.

Combining ear gain with room gain

Now we have curves for both ear gain and room gain, two essential ingredients for sound reproduction accuracy. When we combine the two by adding their sound pressure levels at each frequency, we get the following:

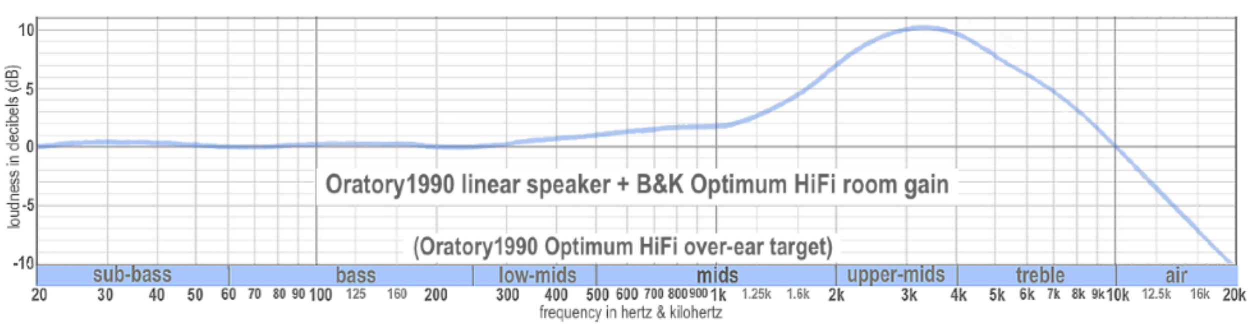


Fig. 5: proposed reference over-ear headphone frequency response target (source: Oratory1990)

The overall shape strongly resembles that of Fig. 2. But crucially, notice the 3 decibel reduction at the peak of the ear gain hump (from 13 dB to 10 dB). Also notice that the B&K room gain curve tapers off just enough on the bass end to keep the bass line essentially as flat in Fig. 5 as it is in Fig. 2. I suspect that, far from being mere coincidence, this is a significant bit of evidence in favour of the B&K room gain research.

Listen in: it would be a shame to read all this content as a purely theoretical discussion. I have getting-started tutorials for the two most common forms of EQ, 10-band graphic and parametric. At the moment, by tuning your over-ear headphones to approximately match the candidate frequency response in Fig. 5, then listening especially to acoustic instruments at moderate loudness, you can hear for yourself what that sounds like. Just keep in mind that the question is not personal preference but accuracy/realism.

Comparing this to the Sennheiser HD 600's frequency response …

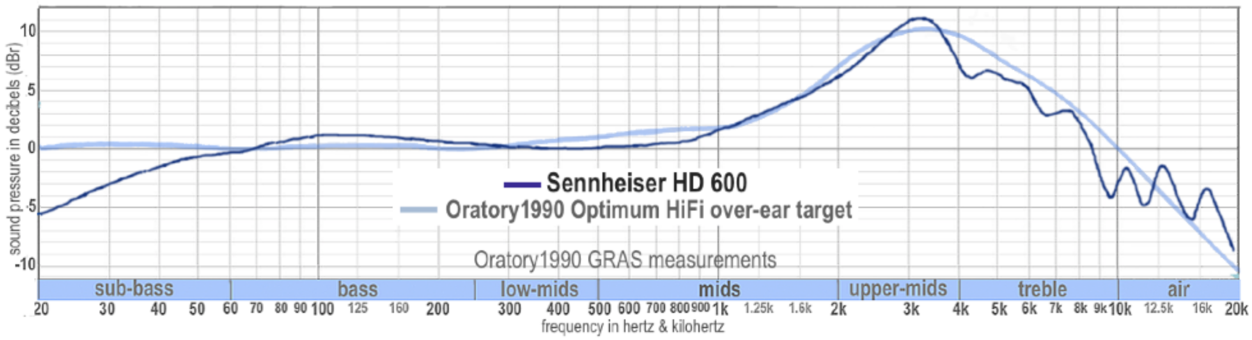


Fig. 6

… now we see quite a close fit, at least by headphone frequency response standards. We see:

- increasing roll-off in the sub-bass (due to technical challenges inherent in designing an open-back headphone, but in a region seldom used for acoustic music)
- a slight overshoot between 70 and 300 hertz, followed by a slight recession from 300 hertz to 1 kilohertz (I suspect both derive from an engineering compromise needed to keep the sub-bass roll-off from being even worse than it is)
- a slight overshoot around the 3 kilohertz peak
- and a moderate shortfall through most of the treble (as we'll see below, this treble shortfall varies from person to person, will be just right for many and will avoid the sensitivities of others).

These are all departures from sheer perfection that discerning headphone enthusiasts have commented on over time.

Some other, more recent, headphones that are also well-regarded for accurate frequency response are the Hifiman Sundara, the Focal Clear and the ZMF Auteur:

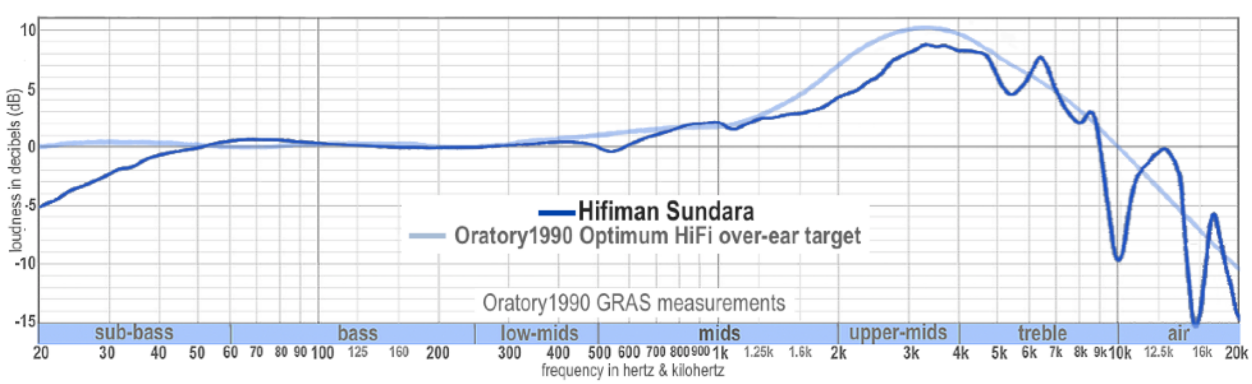


Fig. 7

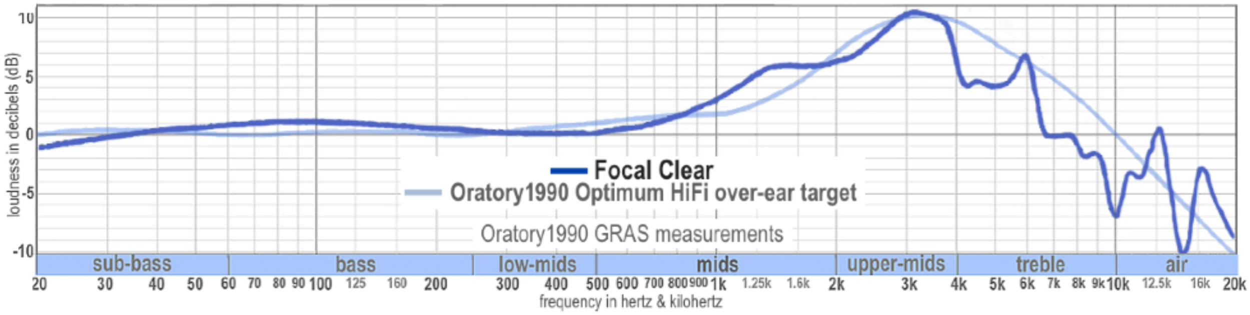


Fig. 8

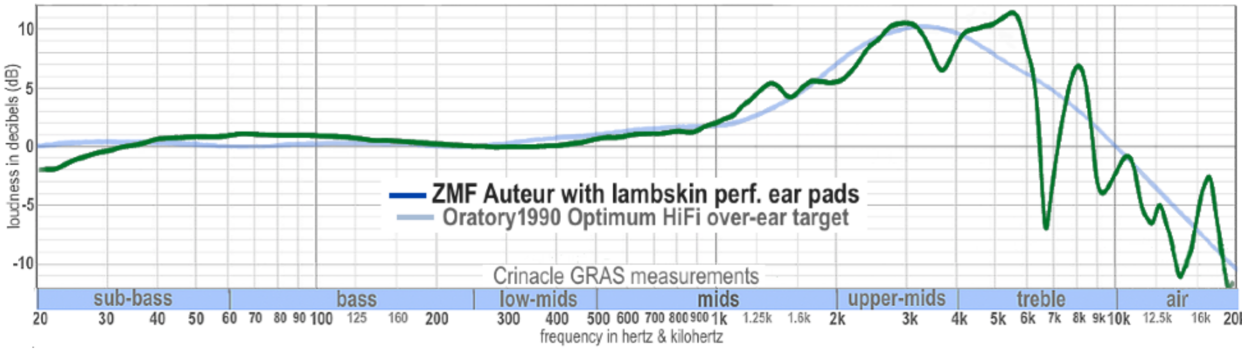


Fig. 9 (source: all Crinacle measurements here)

Again, we see that what the headphone enthusiast community endorses as neutral comes very close to our theoretical prediction. Comparing them to the Fig. 5 Optimum HiFi reference curve simply highlights the most frequent critiques of those headphones for where they are not quite perfect, such as the Sundara's recession from 1 to 3 kilohertz or the Focal Clear's overall treble recession combined with a few spikes.

The HD 600, the Sundara, the Clear and the Auteur are all headphones that the broad consensus of the headphone community refers to as neutral, or relatively un-coloured. They are certainly not regarded as being overly bassy, but instead as being neutral in the bass, except for any sub-bass roll-off. Similarly, the overall ear gain amount of these headphones is not broadly criticized as being too loud or too quiet. There are many headphones with more bass loudness — and the headphone community duly singles them out for having this particular feature. Similarly, headphones with extra ear gain loudness are quickly critiqued for that fact. So also, headphones with lower overall ear gain loudness, such as the HD 650 — while they may be preferred for loud listening to electric guitar riffs — are singled out for having this quality.

As a matter of vocabulary, many long-time headphone enthusiasts refer to anything in the ballpark of the Oratory1990 Optimum HiFi reference curve as ***diffuse field tuning***. This arises from what I suspect was a bit of cagey marketing by Sennheiser. Diffuse field is an alternate approach to quantifying ear gain and comes up with similar, but not identical, results to linear loudspeaker. Diffuse field measurements tend to have an even larger (15 decibels) ear gain peak than the 13 decibels we see in Fig. 2. The design team at Sennheiser led by Axel Grell had surely figured out the need to factor in room gain. But presumably Sennheiser was not about to give away that secret to their competition. So their literature specifically, and misleadingly, specifies their headphones as simply being diffuse field tuned.

So we see that a headphone with a flat line bass not only agrees with the anecdotal evidence that stands behind the continued popularity of the Sennheiser HD 600, not to mention the Hifiman Sundara, the Focal Clear and the ZMF Auteur. It not only agrees with the B&K research. But importantly, it represents a reference target that is already hard enough to achieve in a (traditional) open-back design without introducing the pie in the sky requirement of yet more sub-bass and/or mid-bass, as the Harman and some other options suggest. Let's delve into this last point a bit more deeply …

Theoretical vs achievable

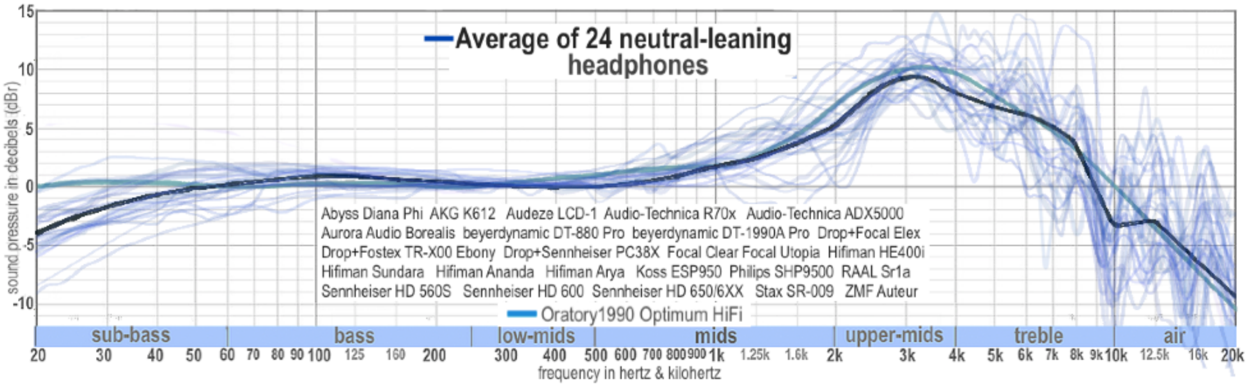


Fig. 10 (source: Cotton)

In Fig. 10 I've averaged the measurements of the 24 closest matches I could find to the Fig. 5 Optimum HiFi reference curve. All of these are open-back construction. Even four of the five planar magnetic driver models have some sub-bass roll-off. If a reference headphone is going to be reference quality in all aspects, not just frequency response, the engineering compromises involved mean that it will most likely be open-back, wired and passive (no built-in electronics). Apparently, the more tweaking that is done to address irregularities in the frequency response the more compromise there will be to other areas such as detail and spatial reproduction.

To illustrate this, here are the frequency response graphs for the price-no-object flagship headphones from three well-regarded manufacturers:

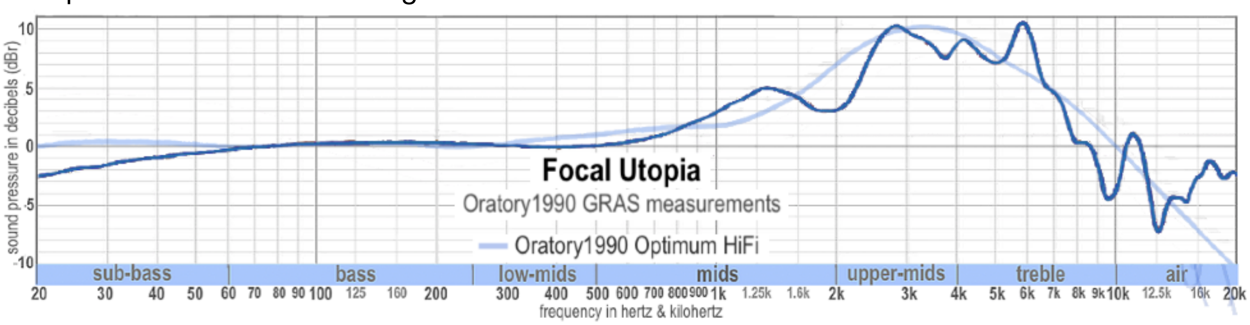


Fig. 11: $4000 dynamic driver over-ear

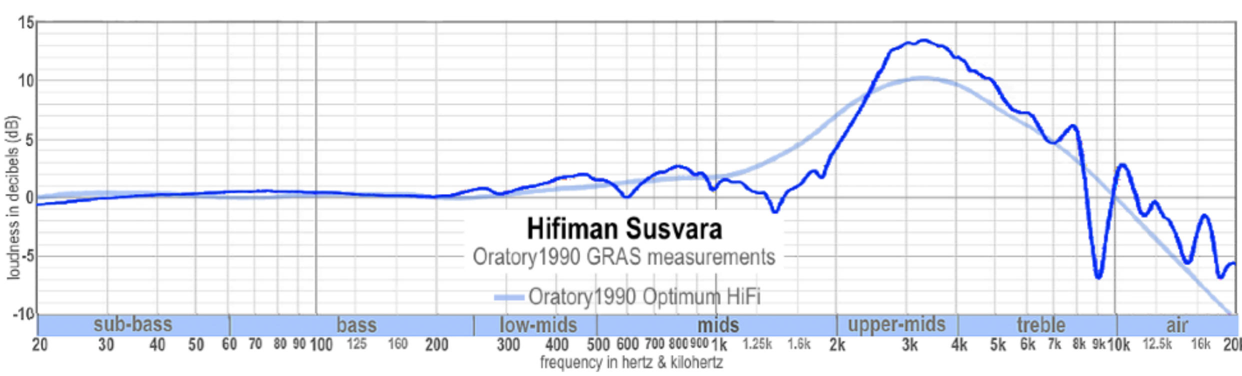


Fig. 12: $6000 planar magnetic driver over-ear (Hifiman's $60,000 Shangri-La electrostat has an even wonkier frequency response)

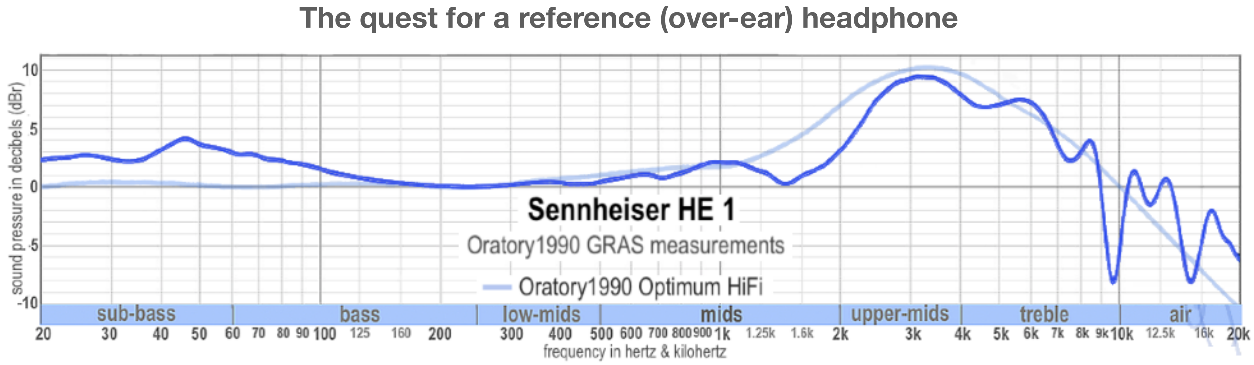


Fig. 13: $65,000 electrostatic driver over-ear

We might expect companies to put their best engineering efforts into their flagship models. But, even ignoring target compliance, there remains the sheer wonkiness of the various paths these four headphones take to get from the left end of the graph to the right end. Only the Utopia has the smoothness of curve that we see in the Sennheiser HD 600, before the inevitable jaggedness of the treble. As mentioned and as we'll discuss extensively below, frequency response is just one sonic attribute needed to reproduce an audio signal accurately. (Some others are dynamic range, dynamic speed, detail, spatial presentation size and spatial placement accuracy.) Headphone engineering is a matter of juggling the often mutually contradictory factors that are needed to optimize all these variables. Frequency response irregularities are the consequence of the trade offs the design engineers deemed necessary to create an optimum overall mix of sonic attributes.

Do you hear what I hear?

Fig. 14: some random ear photographs (source: WikiCommons)

A quick glance at a random sampling of human outer ears tells us that, unlike with the human eyeball, evolution made little attempt to standardize size and shape — yet ear size and shape have a dramatic effect on the shape of the ear gain curves we see in Figs. 2 and 5. Both those ear gain curves derive from a carefully studied average of thousands of measured human outer ears and ear canals. Few of us, if any, happen to have ears that match this average shape. Yet to some extent it doesn't matter. Neuroscience tells us that the human brain simply corrects the raw signals we receive from our ears to match some built-in template that maximizes survival and communications potential. What this amounts to is that all the graphs and data laboriously analyzed above take us as far as the human eardrum. Past the eardrum lies the cochlea, which becomes increasingly degraded with age. Past the cochlea is the brain which, as we're seeing, treats the raw signal from the cochlea almost with contempt.

Clearly, given ear gain, acute high frequency perception is a key survival factor. It's cliché that the distant snapping of a twig could be the only warning our ancestors might have got of an approaching predator. Yet evolution hasn't gone so far as to completely standardize our high frequency perception. As we witness on any audio gear forum, the variation in individual sensitivity to various treble frequencies is huge. Apparently, we can all hear enough highs for survival. But hearing even fainter high pitched sounds at various frequencies doesn't seem to have been a survival disadvantage, and so was not selected against by evolution.

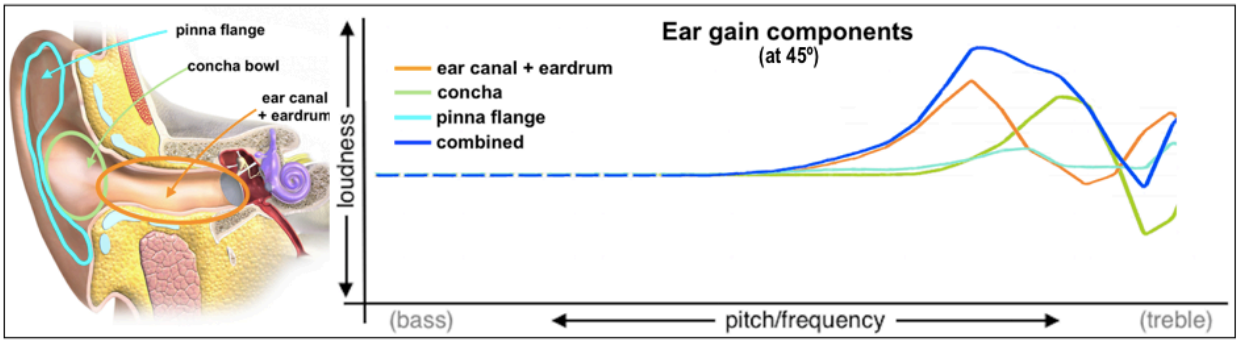


Fig. 15: contribution of the various inner and outer ear regions to total ear gain (excludes head and torso) (source: Wikipedia & InnerFidelity)

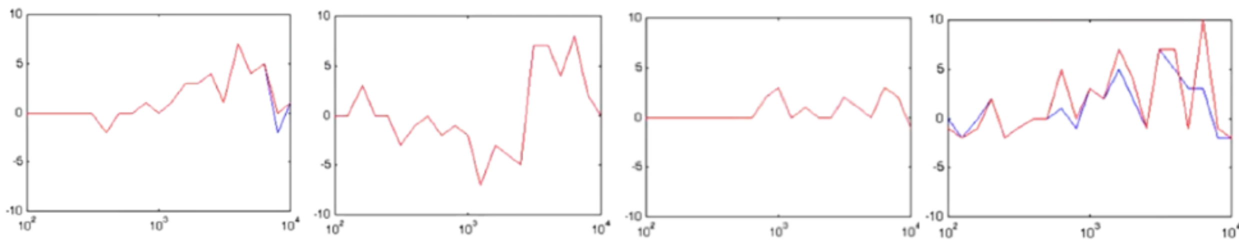


Fig. 16: measurements of four different people's hearing at the ear drum done with insert microphones (source: David Griesinger, includes audio)

Fig. 15 shows how the ear canal's and the outer ear's contributions add up to total ear gain. But because the sound sources illustrated are at a single, 45° angle, the aggregate curve shown in blue is not fully comparable to the loudspeaker curve in Fig. 2, which is for combined left and right 30° angles. Fig. 16 shows measurements at the ear drum for four different individuals and is again not directly comparable to Fig. 2 or Fig. 15, this time due to differences in the type of measurement equipment used. What is relevant is how wildly differently the combined inner and outer ear shapes and sizes affect the same acoustic input as we see in Fig. 16.

Our brains automatically correct for many aspects of our individual ear gains. But it doesn't take much time on a headphone forum to discover that people have wildly differing treble sensitivities. I have yet to have anyone of any age notice the bizarre nine decibel loudness spike of my beyerdynamic DT 1990 around 7½ kilohertz. Yet, read any forum thread or watch a few YouTube reviews of this same model. Half the reviewers froth at the mouth at how piercing it is, while the other half strain to even hear it. The DT 1990 bothers me not at all, but the earlier DT 990 has a 6 kilohertz peak that drives myself and many others up a wall. Fig. 16 shows us that one person may have a concha peak in one frequency band, while the next might have a strong dip there. For headphone design the Principle of No-Trouble Treble remains in full force.

There is a tedious and complex process for measuring an individual's eardrum response, called the ***head-related transfer function***, or ***HRTF***. Correcting for individual variance is particularly beneficial for zeroing in on the precise cues our brains use to locate sound sources in the surrounding three dimensional space. What that doesn't tell us is how HRTF relates to what we actually hear after our brains get through with their post-processing.

We noted above that low frequency sounds literally bounce off the walls (plus, of course, floor and ceiling) of a room causing the bass amplification known as room gain. We also noted that this effect is not consistent. At one point in a room the amplification will be in full force. At other points it will be diminished or non-existent. And, of course, there are rarely any wall reverberations outdoors. Yet, the next time you're out walking with a friend, notice that the sound of his or her voice remains the same whether you're indoors or outdoors and whether you're in one room or another room, let alone one place in a room versus another place. This simple, unnoticed facet of daily auditory experience demonstrates just how powerfully your brain is manipulating the raw signals from our two eardrums to create the sounds we consciously experience. This is called ***psychoacoustic adaptation***.

We would love for our brains to also edit out the drone of a plane or train engine, the hum of fluorescent lighting, the screech of fingernails on chalkboards and the sibilance of over-amplified S's and T's. Evolution goes the extra mile where necessary to aid our survival. It sadly cares not a whit for our peace of mind. However much psychoacoustic adaptation benefits our switching from one audio playback device to another, it seems to fall far short of compensating for all their various deficiencies. It seems that only those playback deficiencies that fall within the auditory ranges that have the potential to negatively impact survival are lucky enough to get addressed by brain processing.

Frequency response provisional results

Part of the problem in identifying a reference over-ear headphone frequency response is the question of reference in what regard? Are we looking to hear through a given recording back to what we'd have heard if experiencing the original performance? If so, at what seating position? according to a population average HRTF? Plus, our brains tend to cut a decibel or two of slack due to psychoacoustic adaptation/accommodation, as mentioned above. Given these considerations, the Fig. 5 reference curve has to be taken as only a starting point. For true validity it would need to be adjusted for each person based on having an expensive HRTF created or by listening to multiple recordings or test tones. So a more realistic version of Fig. 5 might look something like this:

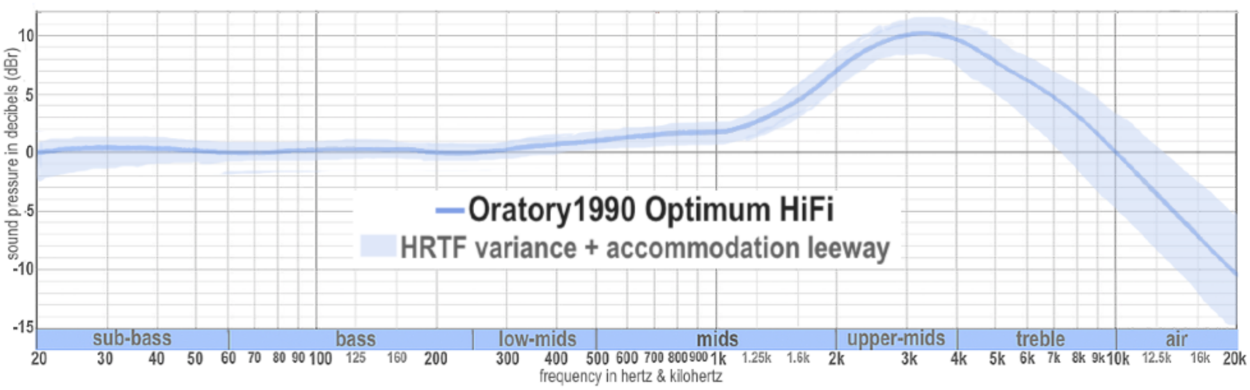


Fig. 17

(In particular, the whole downslope of the ear gain on the right from 3½ kilohertz on will have a unique jaggedness for each person. The broad swath there in Fig. 17 just blurs all those individual jaggies together.) The upshot is that if we desire to hear as closely as possible to what the original performance — or whatever the production studio wants us to accept as the final audio signal — sounded like, we can either live with the frequency response problems of a given headphone situated on our own heads, or we can apply EQ.

— And all the above has been just to home in on one auditory aspect of a hypothetical reference headphone playback. Time to move on to the others, which I'll lump together as …

Reference technicals

While frequency response is arguably the most immediately obvious variable that contributes to our perception that we're hearing an accurate reproduction, the other sonic variables can hardly be ignored. Unfortunately, frequency response is only one of two of the sonic variables we'll be discussing that anyone currently seems to know how to measure. So all discussion regarding the other variables reverts to subjective impressions, which, as we all know, rarely elicit agreement from person to person.

Distortion

By far the largest type of distortion present in audio is ***frequency response distortion***, which is precisely what we were considering in the first section with all those frequency response graphs. While the Sennheiser HD 600 may well weigh in at less than five or ten percent frequency response distortion, few headphones come anywhere near that close.

Beyond frequency response distortion, the usual distortion metric is ***THD+N***, which stands for ***total harmonic distortion plus noise***. ***Harmonic distortion*** just means that the audio system is adding integer multiples of the frequency of the original sine wave. It's adding them to a hypothetical sine wave input signal and they were not part of the original signal. Harmonic distortion is inaccurate but not typically unpleasant. ***Noise*** is any further addition to the input signal beyond harmonic distortion. Noise typically sounds like a hiss. When harmonic distortion and/or noise is present in a music track, research, including large scale ABX testing on-line, consistently shows that it needs to be on the order of five percent to be at all perceptible over a typical recording. Yet headphones typically contribute less than one percent THD+N, except in the deep bass where five percent or even more may occasionally be reached …

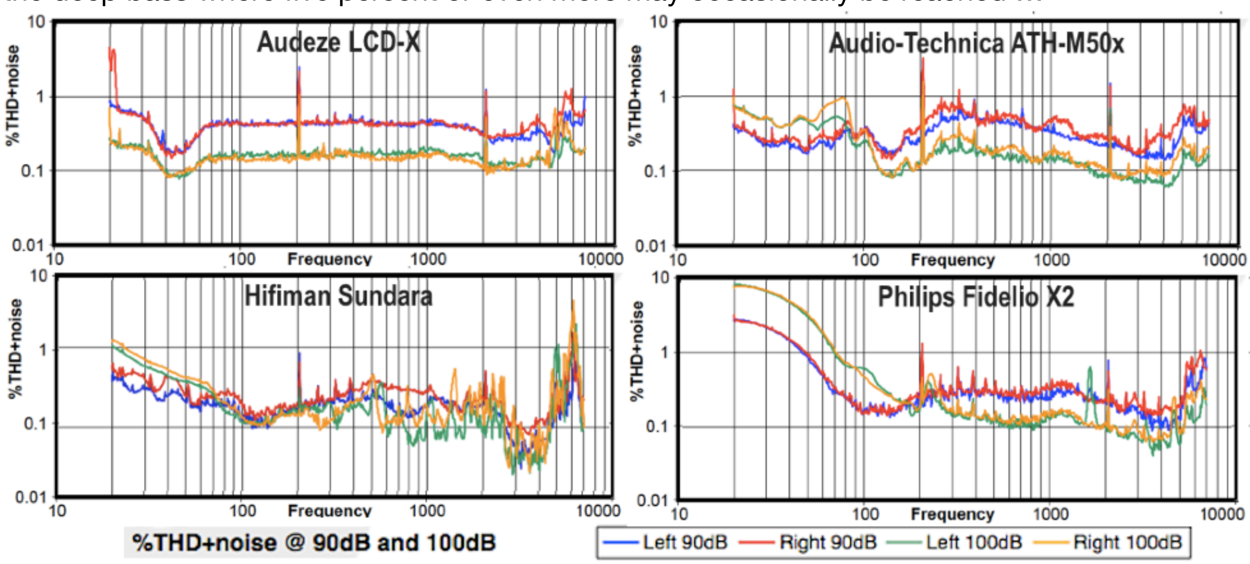


Fig. 18: THD+N measurements for four headphones (source: InnerFidelity)

… but our hearing is even less sensitive to distortion in these frequencies. Note that THD+N measurements like the ones in Fig. 18 are typically done at loudnesses higher than safe listening levels (except as crescendos).

The upshot here is that THD+N has long since ceased to be a factor in any even moderately competently designed headphone, especially when compared to frequency response distortion. THD+N rarely rises above one percent. The one exception to this is that boosting the bass loudness, whether by EQ or selective amplification, of those headphones that are pushing the five percent distortion threshold in the bass will push the distortion into audibility.

Detail resolution

Music performances are typically mic'd relatively closely to the instruments/vocalists as well as at a distance. So ultra-fine details are potentially in the final release of the recording. These are details that a typical concert hall audience member would never hear. Headphone enthusiasts, waxing effusive about their ultra-resolving headphones, talk about hearing the spittle in a singer's voice, the individual hairs on a violinist's or cellist's bow, the faint rustle as the page is turned in a classical musician's score. While these sounds may add to a sense of you-were-there realism, to the composer and performer they're simply unwelcome distractions from the intended soundscape. So, at the risk of incurring the wrath of a whole army of audiophiles, I'd like to suggest that there may be an optimum level of detail recovery in a playback system when the goal is simple music enjoyment. And that optimum might actually be less than some theoretical maximum.

Once again, accurate frequency response plays a significant role in detail retrieval. Louder portions of the frequency range drown out quieter portions. A dip below the reference target curve at any point in the frequency response graph means the faintest signal in that range will go unnoticed. Rises above an accurate level for a given frequency means that the loudness level the listener chooses will put all other frequencies, that are being accurately reproduced, lower in loudness than they otherwise would be.

Further, bass frequency sounds tend to be inherently lacking in focus compared to higher frequencies. The sense that bass notes are being heard clearly typically comes not from the fundamental tone but from the strength of their supporting overtones, which will typically fall in the mid-range and even higher. Cranking up bass loudness often acts more to obscure bass note detail and clarity than to enhance it.

Similarly, listening at unhealthily high loudnesses simply triggers a protective mechanism, called the ***stapedius reflex***, that clamps down the tiny bones that connect the eardrum to the cochlea, muting the sound pressure level and undoubtedly wreaking havoc with any detail that might be present. While 85 decibels (A-weighted) is often quoted by audio enthusiasts as the maximum safe loudness level, the stapedius reflex may well kick in before that. So the loudness level for maximum detail recovery will vary from person to person, but in general will likely be well below 85 decibels.

Getting back to reference headphone design, although frequency response is a significant contributor to detail retention, it is clearly not the only factor. While this seems as though it could be measured, I'm not aware that such testing is being done. Logically, all one would need to do is devise a sound track with precise but faint tones mixed in at seemingly random intervals along with much louder sounds. One would then have the test subject report when s/he hears the start and stop of each faint sound event. Systematic testing could isolate which design factors maximize detail perception. The same methodology could equally be used to evaluate production headphones. Lacking such results, all we have is the subjective reporting of enthusiasts and reviewers. Some reviewers even distinguish themselves by explaining that they are sophisticated enough to not mistake a mere excess of high frequencies for the overall detail retrieval of a headphone model.

Dynamic range

Dynamic range simply refers to the span from the quietest portion to the loudest in a given recording. It's rare, but some headphones compress or limit this range. This ***dynamic compression*** isn't necessarily a bad thing. Especially in western classical music, the range composers employ from pianissimo to fortissimo can span a whopping 75 decibels (Dr. Sean Olive measurements). This means that the listener has to choose whether to listen to the recording at an annoyingly quiet level so the loudest passages aren't ear-splitting, or listen at a more comfortable level and suffer the consequences when the loud parts arrive. But to have the headphone impose its own loudness limit means that it's not faithful to the recording and, hence by definition, not reference.

My understanding is that one of the chief culprits for dynamic range inaccuracy is when damping material is used to correct the frequency response, plus, the damping material is located between the listener's ear and the sound producing mechanism (called ***drivers*** or ***transducers***) inside the ear cup of the headphone. This is called front damping. The problem is that front damping tends to blunt/remove faint detail in certain frequency ranges as a side effect.

Dynamic range also seems measurable, but we don't see this measurement being performed. Whether this is due to the rarity of the problem or due to some non-obvious difficulty in making the measurement I don't know. It seems as though all one would need to do is compare the dynamic range of the digital recording to that of the recording played back through the headphone.

Dynamic impact AKA slam

This is colloquially referred to as ***punch*** or ***slam*** and is often seen as a particularly bass range feature. The prevailing idea seems to be that punchiness or its lack results from the ***speed*** of the headphone's drivers — how responsive the drivers are to rapid changes in loudness. But all modern headphones, beyond airline freebies, are quite capable of producing sound right out to the 20,000 hertz limit of human hearing. This is one vibration every 0.00005 seconds, which is far shorter than our ability to separate one sound from another. So actual speed is not an issue.

Another, related, theory is that dynamic impact is related to ***impulse response***, or ringing. Does the headphone continue sounding one note after the next note is beginning? Headphones indeed vary in this regard. But the physics is well known. Impulse response is simply a mathematical permutation of frequency response, derivable via the fast Fourier transform. The more accurate the frequency response, the more accurate the impulse response. So if that's indeed the basis of punch and slam, then we already have a good measurement tool for this quality. And if so, simply EQ'ing out frequency loudness spikes, which corrects for impulse response ringing, would increase punch/slam. This seems dubious.

Anecdotally, however, at least one aspect of punchiness seems to be the raw physical vibration often felt in the chest when listening to a loudspeaker system with the bass loudness cranked up. By its very nature the traditional headphone design can't deliver this sensation. Experiments are being performed with the addition of bone conduction and similar extensions to the basic headphone concept. But apparently this is in the early stages of development.

Beyond chest vibration, for a reference headphone, the question is which headphones have an accurate amount of punchiness. There is one brand of headphone that has the reputation for dynamic compression and another brand (Focal) which has a reputation for taking punchiness to the next level. What's needed for reference is the level of punchiness that matches the original performance, or at least the recording thereof. At present all we have is reviewer opinion, but fortunately, at least in this one area, reviewers tend to agree.

Spatial extent AKA sound stage

Since the advent of commercially available stereo playback in the 1960's, re-creating the location of the instruments and voices that produce the sounds in a recording has been a big focus among enthusiasts. Acoustic music recordings are typically engineered to do a good job at this when played back on a loudspeaker system in a typical house room. But the "room" headphones make use of is the tiny cavity inside the ear cup surrounding each ear. This results in the sounds being more-or-less correctly distributed but in a small space inside of and just surrounding the listener's head. What's going on is that the brain teases out different audio cues, such as the difference in left/right loudness, to create the total stereo effect. These cues are diminished, distorted and often just plain missing from headphone listening. This is because the recording microphones are placed for ease of re-creating accurate sound cues in a room loudspeaker set-up, not a headphone set-up. A headphone specific recording approach called ***binaural*** exists but isn't generally used in commercial recordings.

A few headphones, such as the Sennheiser HD 800S, are almost universally reputed to have a particularly wide separation of sounds from left to right. But a sound field exists in three dimensions, not just one. Closer/farther distancing is typically described as layering. Up/down in only infrequently discussed. Not only are the audio cues needed to generate these effects largely obliterated by headphone reproduction, the exact frequencies the brain is using to create these effects are different for each person (HRTF dependent). So one person'e reportage on the sound staging of a given headphone is going to differ from another's. Likewise, measuring and engineering for this property are similarly hobbled.

There seems to be a general consensus as to which headphones have a close-to-head sound field and which reach a bit further out. But beyond that, reviewer descriptions are all over the map. And, in any case, headphone sound fields are no match for real world distances.

Spatial placement AKA imaging

Here we're talking about not the size of the space that sounds are distributed over, but the accuracy of their placement within that space. However, exactly the same considerations as for spatial extent apply. There's little more to add.

Timbre

Another sonic attribute that's increasingly in vogue as a descriptor for headphone sound is called ***timbre*** (and is pronounced tam ber in English). Timbre simply refers to the characteristic sound of different instruments or voices when playing the same note. However, headphone reviewers have taken to using the word to refer to the sound of the headphone itself. The theory is that a headphone with a metal or metal-coated diaphragm in the driver may have a metallic edge to its reproduction, while a headphone with a plastic diaphragm (often mylar) in the driver may have a plastic edge to its reproduction, and so on. This may be true, but if so this has to be largely buried in the frequency response. So any headphone that is accurate in frequency response will necessarily have no self-timbre.

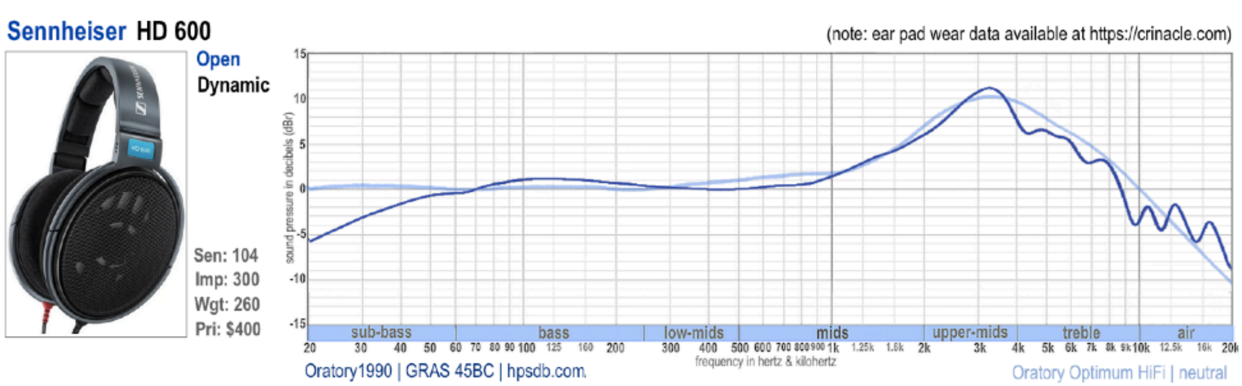
Putting it all together

So what we can say is that a good reference headphone will have:

- a frequency response close to the Oratory1990 Optimum Hifi curve shown back in Fig. 5
- harmonic and noise distortion levels lower than five percent (as is typical)
- at least as much detail as a concert goer would hear during an all-acoustic performance in a good concert hall
- uncompressed dynamic range, plus neither too little nor too much dynamic impact
- a reputation for having a continuous stereo sound field outside the listener's head with precise placement of sounds within that field
- the lack of any sense of adding a metallic or plasticky colouration to the sound it produces.

Additionally, EQ will help correct its inevitable deficiencies in frequency response.

As I write this there are several candidates to choose from. Here are some that stand out for me (but note that I only have hands-on and ears-on with the first two):

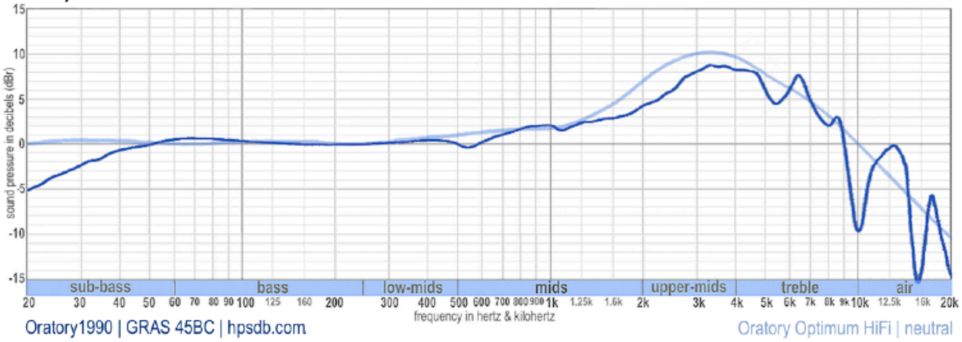


This meets all criteria excellently, except for having an unabashedly close-to-head sound stage. Not only is the roll-off in the far sub-bass where very little un-amplified music lives, it's also representative of the typical playback device a recording engineer is targeting.

Hifiman **Sundara (2020 revision)**

Open
Planar

Sen: 108
Imp: 37
Wgt: 372
Pri: $350

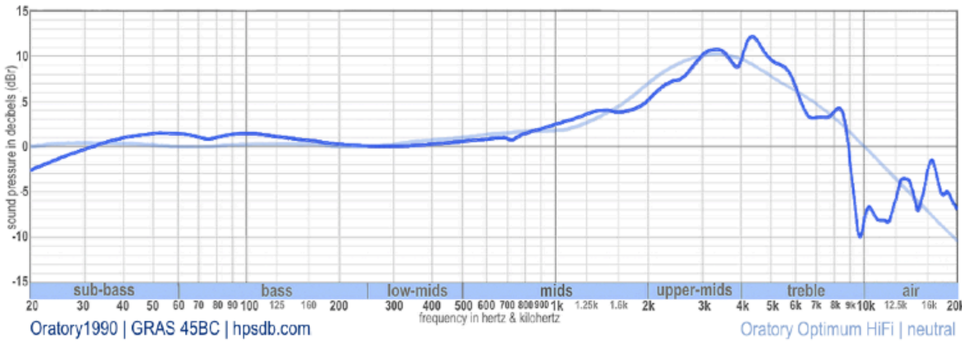


This meets all criteria excellently, except for the obvious frequency response recession from 1 to 4 kilohertz that subtly reduces clarity. EQ recommended.

Sennheiser **HD 560S**

Open
Dynamic

Sen: 110
Imp: 120
Wgt: 240
Pri: $200

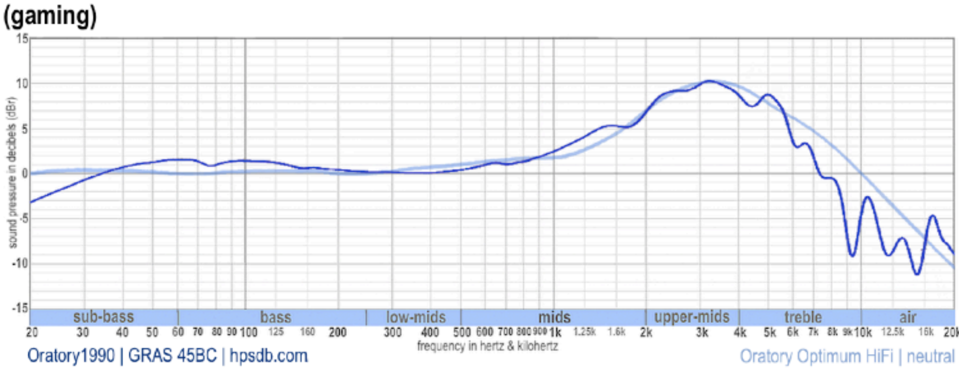


This meets most criteria excellently, except for having a frequency response peak at 4 kilohertz, which may need EQ depending on your sensitivity, plus a bit less than really top-notch detail retrieval. Going by reviews: sound stage is at least reasonably wide, and placement is at least average in accuracy.

Drop+Sennheiser **PC38X (gaming)**

Open
Dynamic

Sen: 109
Imp: 28
Wgt: 253
Pri: $169



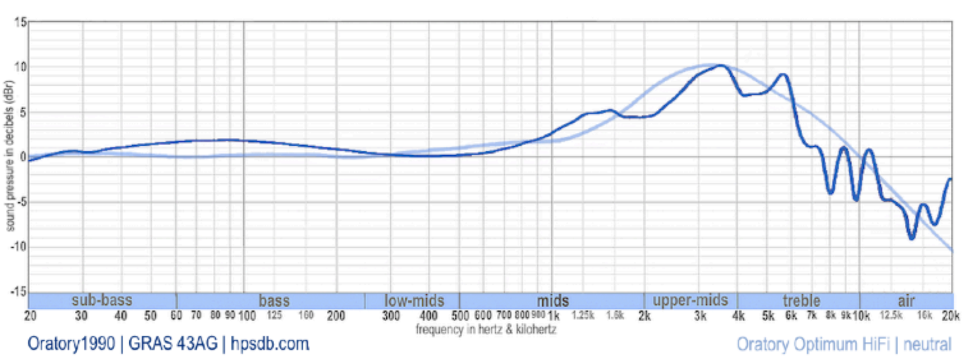
This has an excellent frequency response, and is at least decent in most technicals but is reputed to fairly average at detail retrieval and spatials.

Venturing into pricier territory:

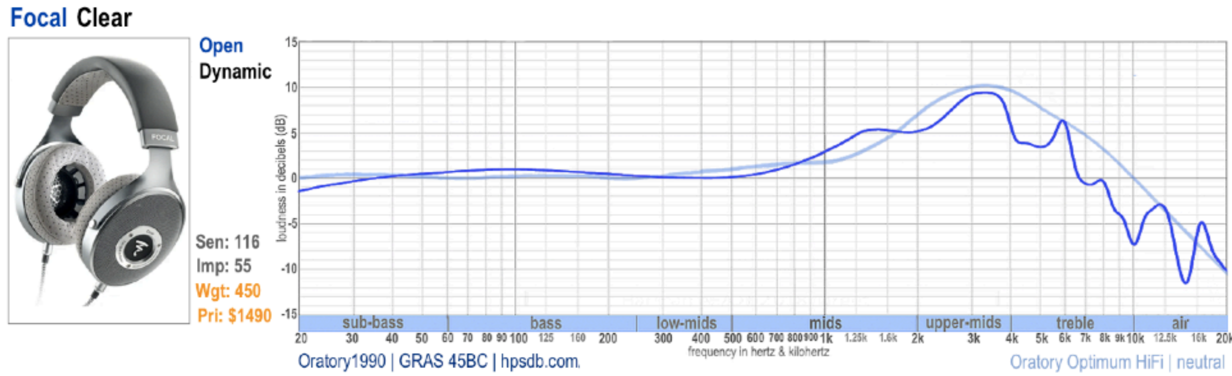
Drop+Focal **Elex**

Open
Planar

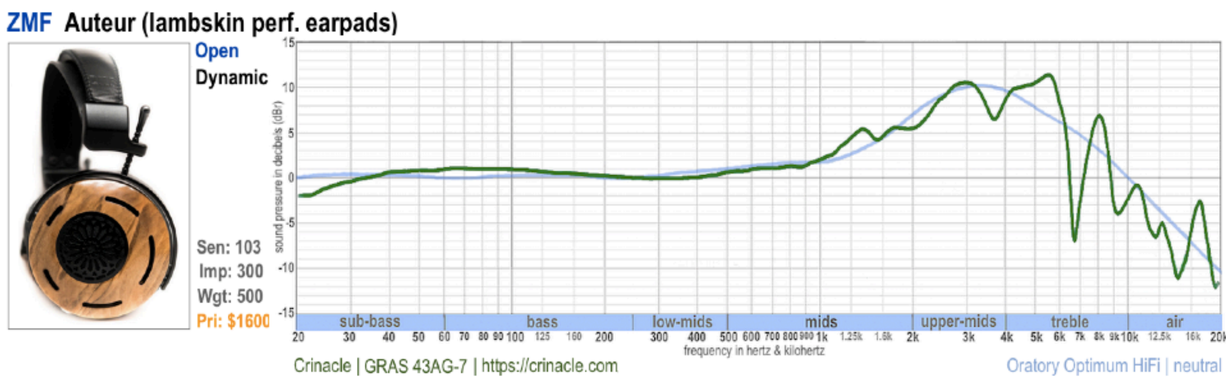
Sen: 115
Imp: 80
Wgt: 450
Pri: $700



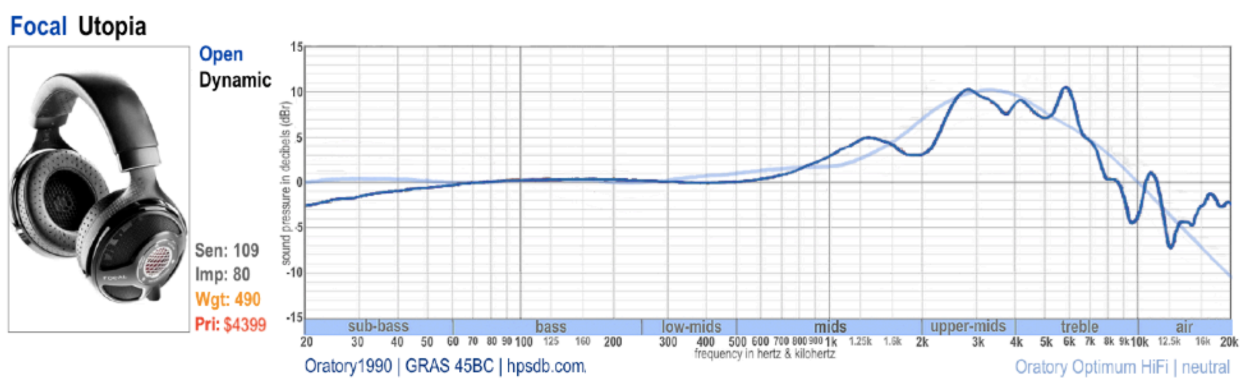
The Elex meets most criteria excellently, except for a bit of overall frequency response irregularity, a somewhat small sound stage and what may be a slight excess of punchiness / dynamic impact.



If this is in your budget, as with the Elex, this has some overall frequency response irregularity, a somewhat small sound stage and what may be a not-so-slight excess of punchiness. But detail retrieval is reputed to be excellent. (Note: the Focal Clear MG is very different and not tuned as reference, so would need EQ).



If this is in your budget, about the only criticism seems to be the few peaks in the ear gain part of the frequency response.



If price is truly no object, this has some overall frequency response irregularity, a somewhat small sound stage and what may be a not-so-slight excess of punchiness. But detail retrieval is even more excellent than that of the others. (Also, I have to note the repeated reports of driver failure as I write this.)

Beyond the above, any headphone with good technicals but with less than optimal frequency response can be used when EQ is an option. One example, from my own collection, is the beyerdynamic DT 1990. The traditional Audeze LCD line-up is another example. A typical Audeze LCD headphone has a frequency response like this:

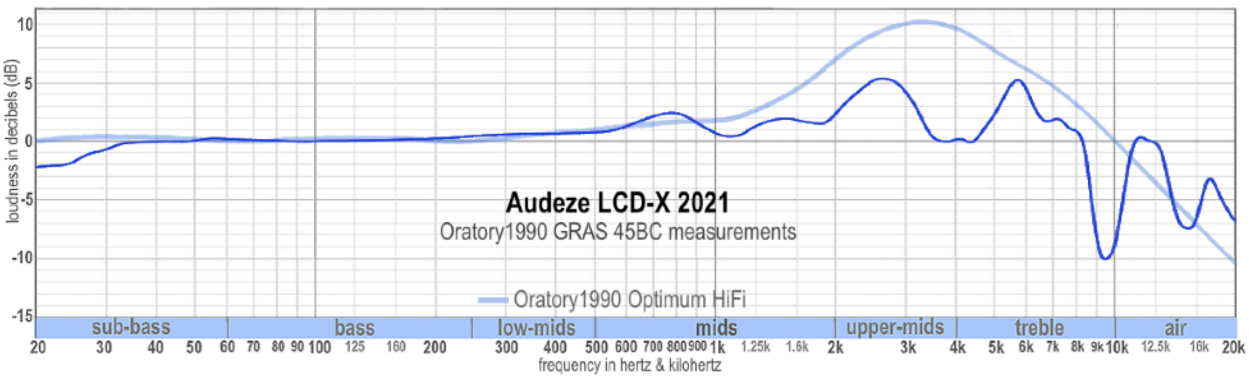


Fig. 19

They're renown for three things — heavy weight, excellent sound technicals and a huge recess throughout the ear gain. My theory is that Audeze has chosen to maximize all other sonic performance parameters, leaving frequency response to the user to EQ in. In fact, they market a custom EQ solution as an option to their headphones. Their latest releases are their new top-of-line LCD-5 and the more affordable MM-500, both with much lower weight and much greater ear gain. Hopefully, this trickles down to their lower-priced models over time.

Appendices

Appendix A: getting the hang of frequency response graphs

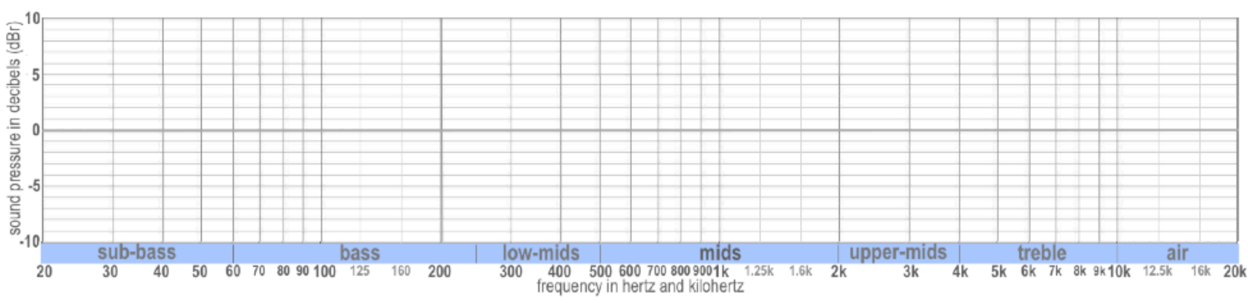


Fig. A1

This explainer makes extensive use of the kind of two dimensional, X axis plus Y axis graph you first made acquaintance with in high school algebra, but with a few differences. First, the X axis plots sound vibration frequency, otherwise known as pitch, as in bass to treble. The unit for sound vibrations is cycles per second, or hertz. The Y axis plots loudness as measured in the form of sound pressure, called SPL for sound pressure level. This is expressed in units called decibels.

The other difference is that the scales are done in ***logarithms***. This means that instead of showing an equal distance between numbers, the distance gets systematically smaller as the numbers get larger. Human hearing covers the range from 20 vibrations per second to 20,000 vibrations per second, so a non-logarithmic graph with each number separated from the next by a single millimetre would have to be 20 metres long. If instead, the graph only showed markings for every 50th or 100th unit, important one-digit differences at the bass end of the scale would be missed. Notice that past the 1000 hertz line the notation switches from hertz to kilohertz. So 2k on the horizontal scale is called 2 kilohertz is the same as 2,000 hertz. On the Y axis decibels are intrinsically logarithmic, so the vertical scale shows an equal distance between each decibel number.

To put this all together, the following graph shows the frequency response of a Beats Solo 3 headphone compared to a well-known curve for appropriately boosting bass frequencies:

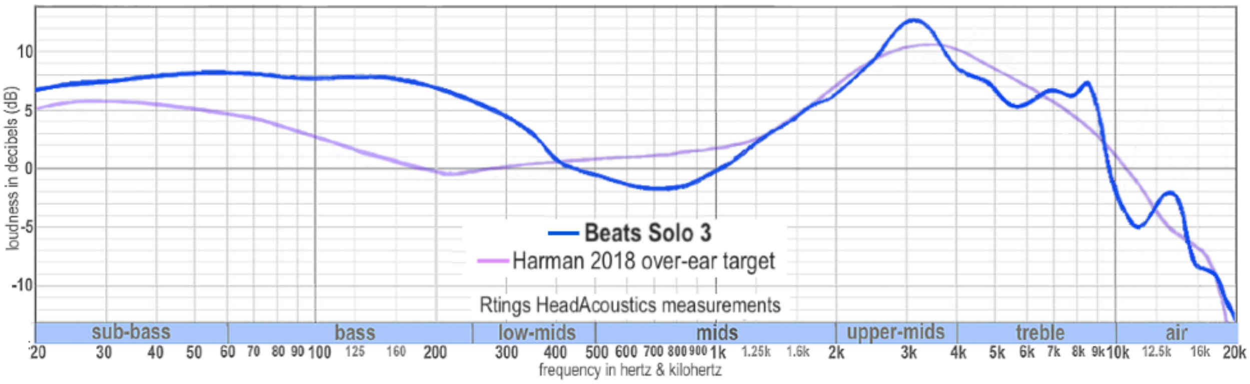


Fig. A2 (source: Rtings.com)

From the blue line of the measurements we see that entire bass range and into the lower mid-range plays back sound much more loudly than frequencies in the middle of the mid-range, such as 700 hertz. Similarly, some very high frequencies around 3 kilohertz are even louder than the loudest bass sounds. 3 kilohertz is higher than the soprano vocal range and well into the whistle range, or very nearly the highest notes on the piano or piccolo.

The frequency response, or tuning, of the Solo 3 is called a ***fun*** or ***V-shaped*** tuning and maximally appeals to the younger male demographic. The central dip in loudness diminishes a large swath of the vocal range and of most musical instruments. The Beats Solo 3 is very typical of mass market, non-reference headphones.

Reality check

Equally important, however, is to understand the fudge factors that are at work:

- ❖ No two units of the same headphone measure identically, and likely not within a few percent of each other. Here are the measured frequency responses of ten units of the same model produced by a company that has an excellent reputation for consistency:

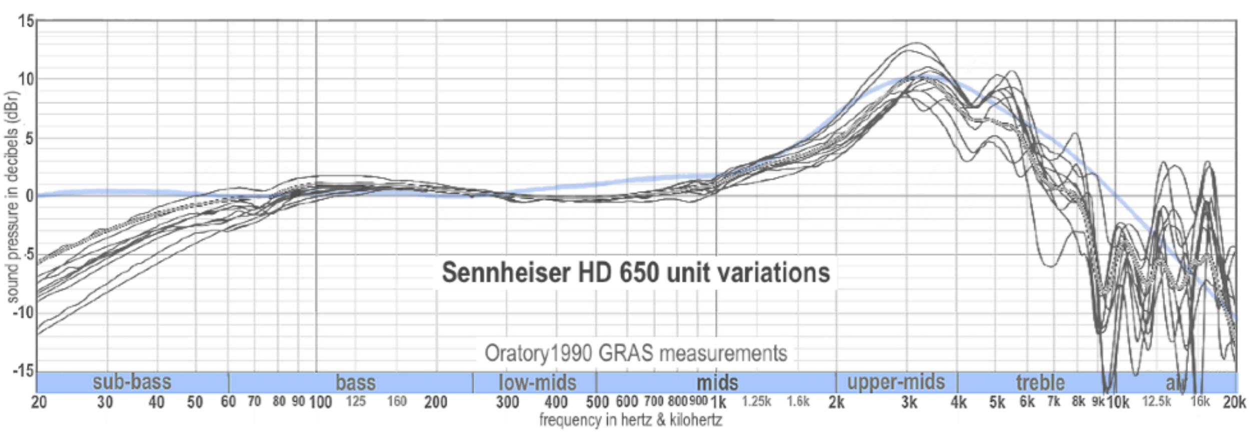


Fig. A3 (source: Oratory1990)

- ❖ Another problem is that simply shifting the same headphone to slightly different positions on your head produces similarly variable results:

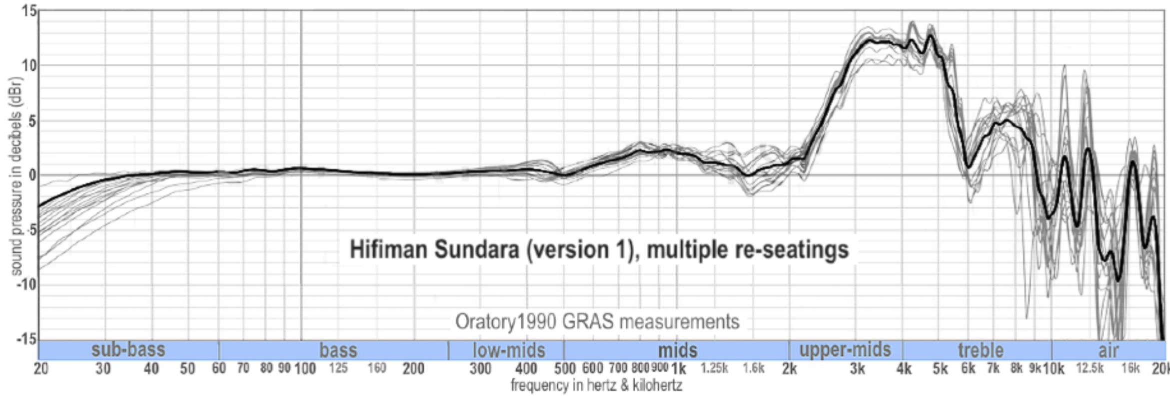


Fig. A4 (source: Oratory1990)

❖ Yet a third problem is the difference in frequency response for the same headphone when worn by different individuals. This is discussed in **Do you hear what I hear?** above and illustrated by Fig. 16 in that section. Due to the very nature of the GRAS measuring equipment used throughout this document there is no graph comparable to Figs. A3 and A4 and done on GRAS equipment to illustrate this issue. But so far as I can tell, such a graph would be very similar to A4, except perhaps with less spread in the sub-bass.

The upshot of all three of these factors is that any one headphone measurement graph such as we see throughout the frequency response section of this document is quite accurate, it represents a snapshot of a particular unit of a particular model headphone on a statistically average human head. If you were to purchase a new unit of the same model headphone measured in a particular graph, not only would the graph differ by several percent compared to what you are hearing, it would vary by several percent just by shifting its position on your head however slightly up, down, left and right.

Appendix B: Harman target development and depth

One simple thread through all the above is that recordings targeted at accurate reproduction of the original music event are engineered on loud speakers with the intent of re-creating that accuracy on home loud speaker systems. Dr Floyd Toole was instrumental in the research that identified the approach of using flat frequency response speakers in a well-treated room. Once Dr Toole was hired by the audio conglomerate Harman International to oversee their loud speaker frequency response research, Harman hired his protégé, Dr Sean Olive, to research what the equivalently accurate frequency response for headphones might be. While combining diffuse field ear gain measurements with room gain led to the well received Sennheiser HD 600, a more direct approach is simply to turn to the sound produced by anechoically flat loud speakers placed in a well-treated room.

So the first step, handled primarily by Todd Welti, was to measure the eardrum frequency response of flat-measuring loud speakers without room gain. Todd Welti did this by using an industry standard ear, head and torso simulator (GRAS 45CA) recording Harman's well-regarded Revel 208 loudspeakers in a room carefully equalized to a flat line response:

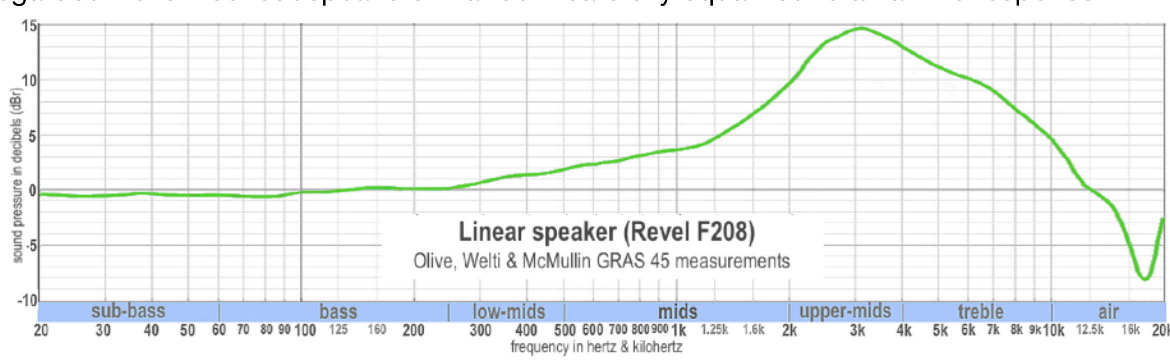


Fig. B1: Todd Welti's linear speaker measurements, circa 2013 (Herstens)

As mentioned, this is similar to, but not identical to, the ***diffuse field*** (equal loudness from every direction) data Sennheiser likely used. The Harman team then added neither the common one decibel per octave room gain correction (pale green line in Fig. B2), nor the B&K Optimum HiFi correction (pale orange line), but a correction with even more bass, based on Harman room correction research, namely, the magenta RR1 line in Fig. B2:

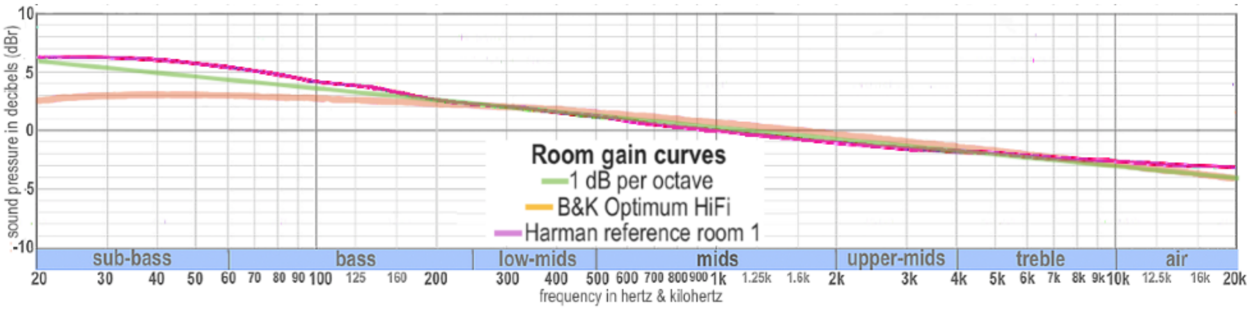


Fig. B2 (source: TenBrook)

The result, when tweaked a bit by hand, was the basis for the original Harman over-ear target:

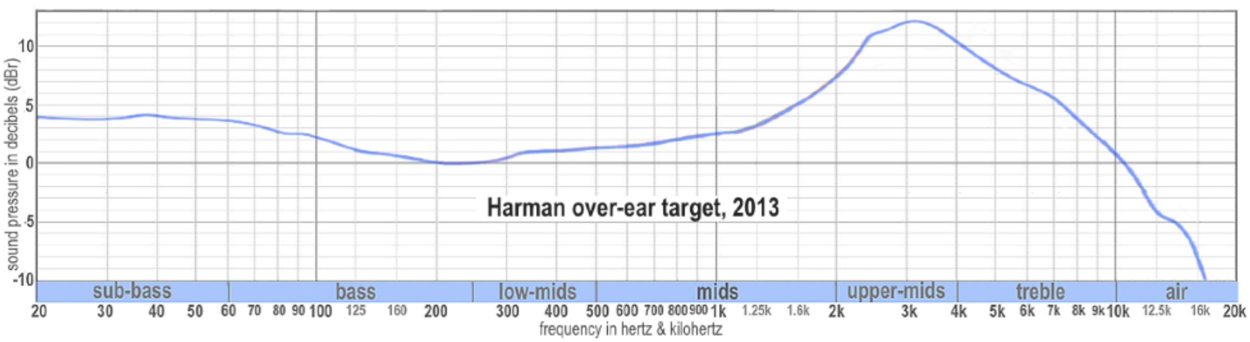


Fig. B3: First published Harman over-ear headphone target (source: Herstens)

FIg. B3 is what astonished headphone enthusiasts encountered one day in 2013 on the extremely popular and well-regarded (and sadly, now defunct) headphone site, InnerFidelity.com. The controversy over its boosted bass has hardly taken a breather ever since. While the Harman team's initial tests validated that listeners did indeed prefer this target over other relevant options, extensive subsequent research produced the final Harman over-ear headphone target …

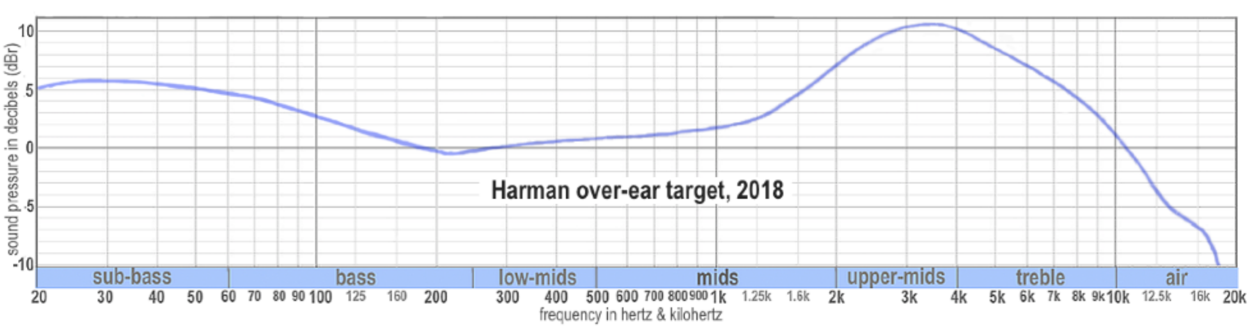


Fig. B4: 2018 Harman over-ear target

… which has even more bass, but somewhat less treble, than the 2013 version.

From 250 hertz on up the Harman curve in Fig. B4 is a very close match to the Oratory+B&K reference curve in Fig. 5. However, the Harman curve was developed by studying the preferences of almost 250 people from a variety of demographics. The agreement from 250 hertz on up provides another bit of validation for the Oratory+B&K reference, since a lot of acoustic research tells us that there is a wide divergence between individual eardrum response in this region. But what about that Harman bass shelf? Here is the room gain needed to produce the Harman target from a flat loudspeaker response back in Fig. 2:

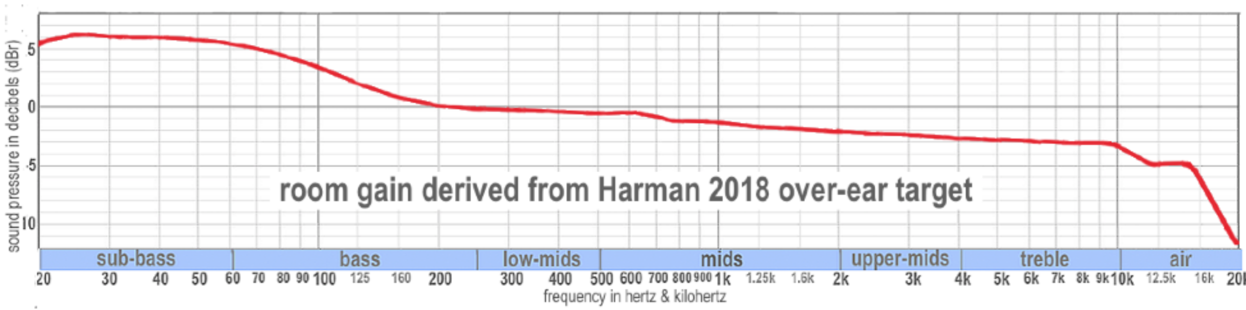


Fig. B5

I can well imagine this might be typical of the sound field from a wall of amps at a rock concert. But typical of Carnegie Hall or La Scala? In fact, the large hump on the left is the average of a wide range of individual bass response preferences:

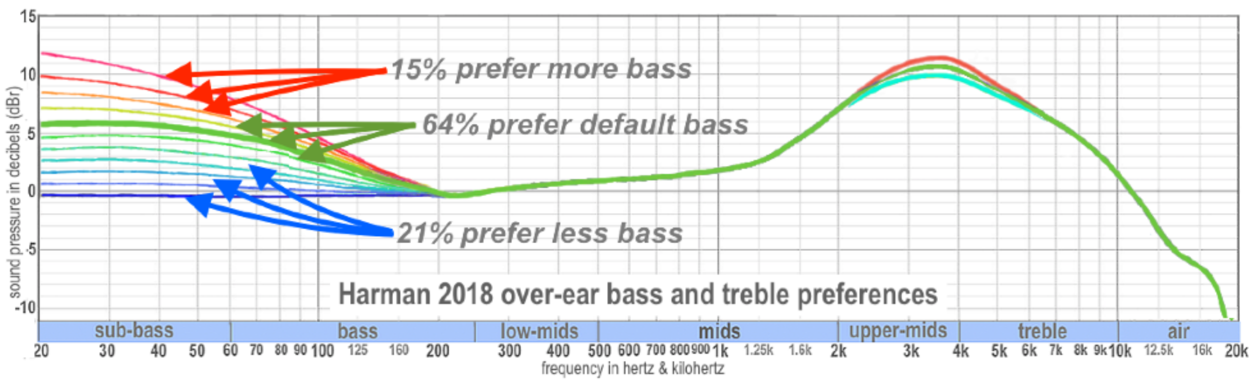


Fig. B6 (source: Oratory1990)

Olive and company further parsed this range of bass preferences by age and sex:

Distribution of Listener Categories Within Each Class (in %) Based on Preferred Headphone Sound Profile			
Category	**Class 1: Harman Target Lovers**	**Class 2: More Bass Is Better**	**Class 3: Less Bass Is Better**
Males	0.69	0.18	0.13
Females	0.56	0.04	0.40
Trained	0.70	0.30	0.00
Untrained	0.65	0.10	0.25
Age (years)			
20s	0.69	0.17	0.15
30s	0.74	0.13	0.13
40s	0.67	0.10	0.24
50+	0.30	0.20	0.50

Fig. B7 (source: Olve)

They also parsed by nationality (U.S., Canada, Europe and Japan) but found little to report.

So we see that the bass loudness preference shown in Fig. B4 is simply the average of a wide range of preferences that were recorded. These preferences vary by demographic, with young males tending to prefer more bass than the target. While females plus older people of both sexes tended to prefer less. Put another way, if you barred testosterone-enriched males from participating in your study, your results would probably look much more like this combination of a linear loudspeaker response together with the 1 decibel per octave room gain beloved of 2.1 channel stereo:

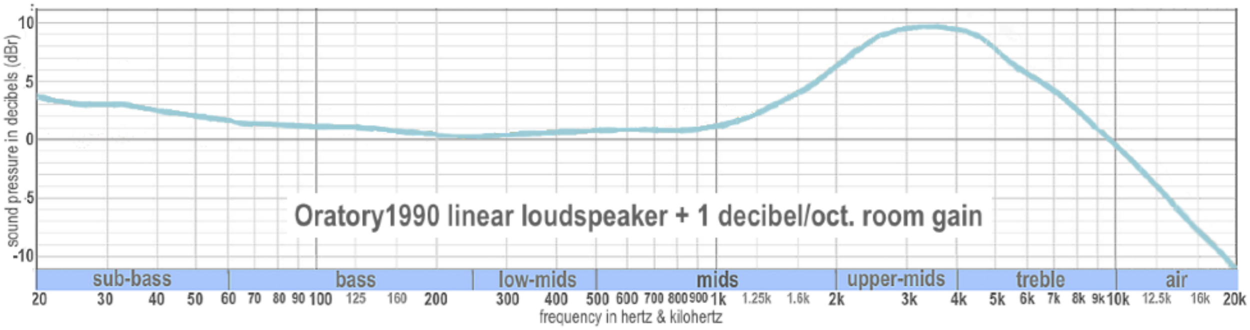


Fig. B8

Yet another, if minor, consideration re the Harman bass boost is that some people find the frequency range just past the bass (say 200 to 400 hertz) to sound a bit thin (quiet). In other words, the contrast of the loudness in bass to the relative quiet of the lower mid-range causes a problem for them that wouldn't otherwise be there.

In any case, there need be no argument with the Harman target. For closed-backs and DSP-based on-ear and over-ear headphones targeting the general consumer listening to a wide mix of genres I'm sure it serves admirably, and would do so even better if it came in three demographic bass level versions. But we're looking for what's reference rather than what's preference. Our brains are still evolved to process a flat-line bass signal. And recording engineers still have to mix music tracks for playback on everything from sound bars to car audio. Hopefully, our considerations above have demonstrated that the Oratory Optimum HiFi target remains a better approximation of the middle ground among them all.

Appendix C: alternative neutrals

For comparison, here are three headphones that have been touted as appropriate for studio use:

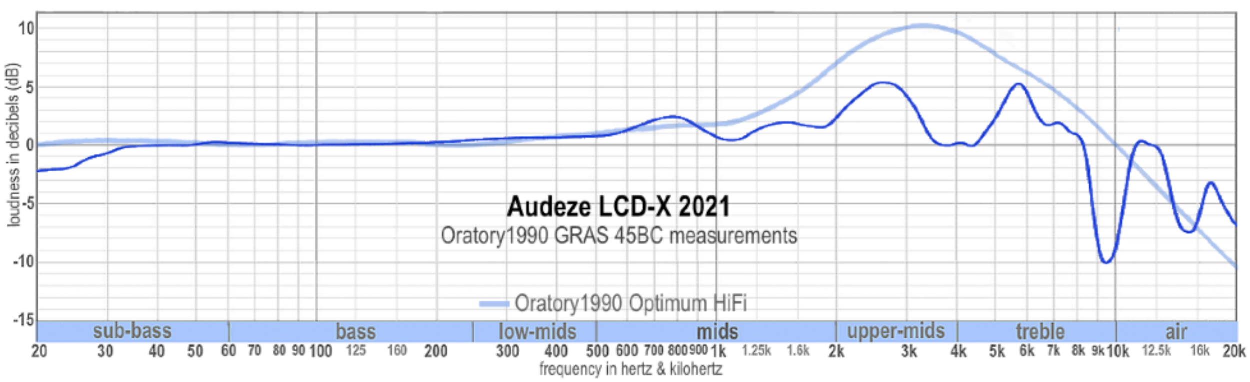


Fig. C1

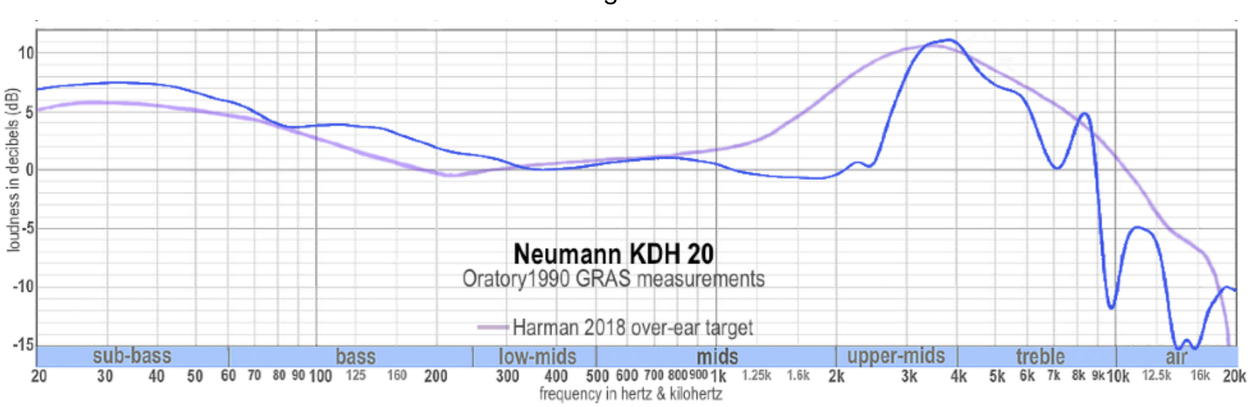


Fig. C2

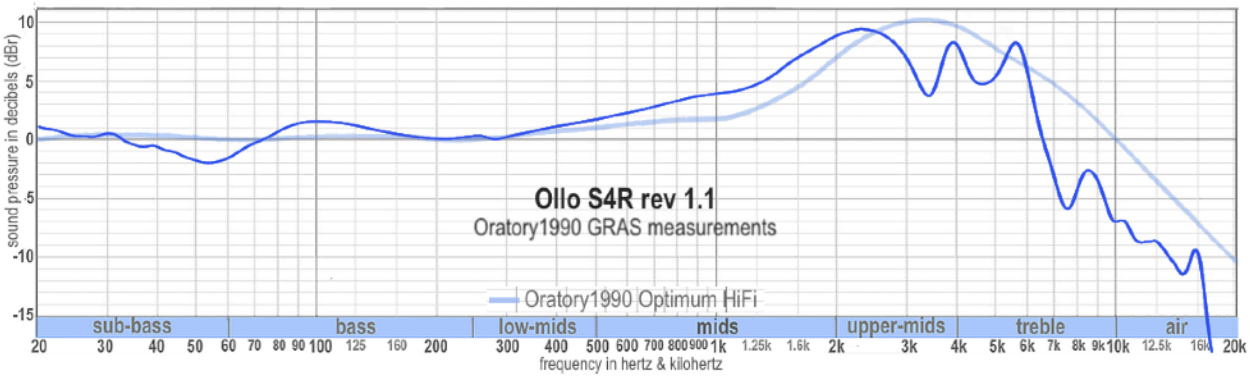


Fig. C3

It's instructive to EQ a pair of headphones to even roughly match any of these measurements then listen to some all-acoustic music recordings.

Appendix D: Further reading and viewing

Some source material:

1. **Brüel & Kjær**, "Relevant loudspeaker tests in studios in Hi-Fi dealers' demo rooms in the home etc." (1974), https://www.bksv.com/media/doc/17-197.pdf
2. **Tyll Herstens**, "Headphone Measurements Explained - Frequency Response Part One", https://www.stereophile.com/content/innerfidelity-headphone-measurements-explained
3. **Warren TenBrook**, "Warren TenBrook's Summary of Head Measurements at Harman", https://web.archive.org/web/20190711222725/https://www.innerfidelity.com/content/warren-tenbrooks-summary-head-measurements-harman
4. **Warren TenBrook**, "An Acoustic Basis for the Harman Listener Target Curve", https://web.archive.org/web/20160924084851/http://www.innerfidelity.com/content/acoustic-basis-harman-listener-target-curve
5. **Chris Plack**, "The Musical Ear", https://nmbx.newmusicusa.org/The-Musical-Ear/
6. **David Griesinger**, "Binaural hearing, Ear canals, and Headphones" (2016), https://www.youtube.com/watch?v=a-JGAobDwGs
7. **Oratory1990**, "Differences between the Harman Curve and Diffuse Field", https://www.reddit.com/r/headphones/comments/78x77b/initial_impressions_of_2016_audeze_lcd2f_with/doyj84e/
8. **Oratory1990**, "Full list of EQ settings" [and measurements], https://www.reddit.com/r/oratory1990/wiki/index/list_of_presets and https://headphonedatabase.com/oratory
9. **Sean E. Olive**, Do Listeners Agree on What Makes a Headphone Sound Good?, https://www.youtube.com/watch?v=f1EVZVDaeLw
10. **Olive and Welti**, AES 139 Presentation 2015 Olive and Welti Preferred Bass and Treble Levels, https://www.youtube.com/watch?time_continue=2&v=ySQV5OR71e4&feature=emb_logo
11. **Sean E. Olive**, "Modeling And Predicting Listeners' Headphone Preference Ratings (2021)", https://www.youtube.com/watch?v=62fdLy5OC9A (relevant before 28:50)
12. **Sean E. Olive**, "The Perception and Measurement of Headphone Sound Quality- What Do Listeners Prefer?", https://acousticstoday.org/he-perception-and-measurement-of-headphone-sound-quality-what-do-listeners-prefer-sean-e-olive/
13. **Wikipedia**, "Head-related transfer function", https://en.wikipedia.org/wiki/Head-related_transfer_function
14. **Wikipedia**, "Acoustic Reflex" (Stapedius reflex), https://en.wikipedia.org/wiki/Acoustic_reflex
15. **D. Hammershøi and H. Møller**, "Determination of Noise Emission from Sound Sources Close to the Ears", *Acta Acustica*, Vol.94 No.1, January 2008