Headphone Essentials 2:

Basics of Headphone Sound

Note: this document is part of an instructional series. If you would like to start with more foundational background information on the nature of sound, see *Basics of Musical Sound* at Headphone Essentials.

Audio reproduction involves several sound variables. Here are the four most commonly discussed as they apply to headphones:

1. Frequency response

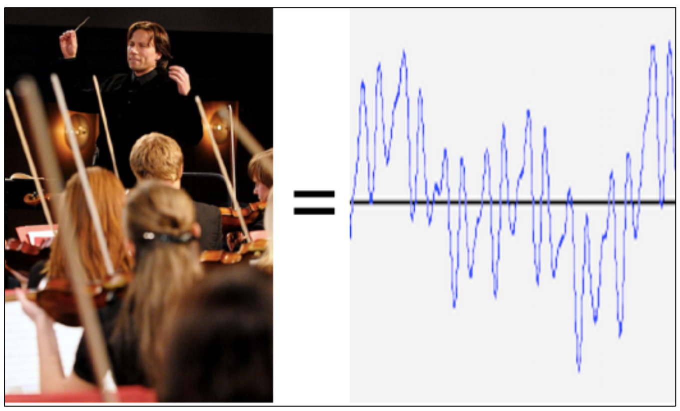

Fig. 1: millisecond view of a complex orchestral passage
Photo credit: Johannes Jansson, License

Frequency response refers to the bass-tenor-alto-soprano sort of thing. Tuba, tenor sax, piccolo for another set of examples. The common vocabulary for this in audio reproduction is bass-mids-treble, corresponding to low frequencies, middle frequencies and high frequencies respectively. *Upper-mids* is the headphone enthusiast's term for a narrow but important range of high frequencies between the mids and treble. Vocally, this is above even coloratura soprano and reaches into the whistle range. Technically, the frequency ranges are defined as follows:

Frequency ranges

Bass: 16-256 Hz* (approximate notes: C0 to middle C4)
Sub bass: 20-60 Hz, mid bass: 60-125, high bass: 125-256
Mids: 256-2048 Hz (C4-C7)
Low mids: 256-512, Mid mids: 512-1024, high mids: 1024-2048
Upper-mids: 2048-4096 kHz
Treble: 2048 kHz-16,384 kHz (C7-C10)
Low treble: 2048-5120, mid treble: 5120-10,240, high treble: 10240-20,480

You can hear what the various frequency ranges sound like at:
Online Tone Generator

*Frequency is measured in Hz = hertz = air pressure changes per second

Photo credit: Vincent Escudero, License

No headphone exists which matches any known standard of frequency response accuracy. Few even come close. Some reproduce all or part of the bass too loudly or not enough, some reproduce all or part of the middle frequencies too loudly or not enough, etc. Primitive EQ (equalization) applets are available in both Window's Media Player and in MacOS/iTunes. If you have EQ available to you, experiment with raising and lowering the various adjustment points while listening to music to get a feel what this means. But be aware that your existing headphone has its own imbalance that is contributing to what you are hearing.

Commonly used **frequency response vocabulary**

- **Neutral/accurate/flat**: all frequencies from sub bass to high treble are subjectively even in loudness.
- **Bright** or **cool/cold**: the high frequencies are louder than the low and middle frequencies.
- **Warm**: the lower frequencies are louder than the middle or high frequencies.
- **Dark**: the high frequencies are rolled off (diminish in loudness).
- **Mid-focused** or **mid-forward**: the middle frequencies are louder than the low or high frequencies.
- **V-shaped** or **U-shaped** or **fun tuning**: bass and treble both louder than the mids. The mids in turn are then described as being **recessed** (relatively quiet) in comparison.
- A **relaxed** tuning (frequency balance): something like warm/dark in that the higher frequencies are toned down (but often the bass and mids are fairly neutral).

Audio professionals have developed a colourful vocabulary over the years to describe too much loudness in specific frequency response areas (courtesy M0N 6019):

Boom: 80 - 160 Hz
Mud: 100 - 400 Hz
Honk: 500 - 1000 Hz = 1 kHz
Fatigue: 2.5 - 5 kHz
Sibilance: 5 - 8 kHz
Hiss: 10 - 20 kHz
Grit: 1 - 2.5 kHz

As you might expect, for music and video production work the closer to frequency response accuracy/neutrality the better.

2. Dynamics

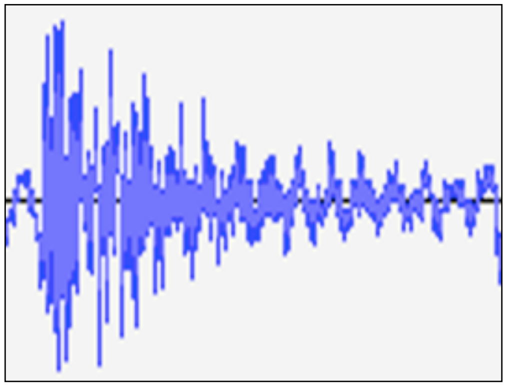

Fig. 2: milliseconds view of a drum strike

Another audio reproduction variable is *dynamics*. This is how responsive the headphone is to changes in the loudness of a given frequency. Most headphones do a very good job with dynamics. A headphone that cannot achieve the full range from quietest to loudest has *dynamic compression*. Given how often mass market music is deliberately dynamically compressed in the recording studio, we might be surprised that anyone is concerned when a playback component such as a headphone does this as well. One limiting factor is whether the actual sound producing element in the driver can move up and down at each frequency sufficiently far to create the required amount of sound pressure. This is called *excursion*. The other factor is whether the full amount of sound pressure from the driver reaches our ears. If something in the headphone absorbs some of the sound pressure this is called over-*damping*. A headphone that cannot respond quickly enough to sudden changes in loudness would be slow in *speed (see below)* and so would lack a certain sense of precision, but this is rare.

Related to dynamics is a phenomenon often called *slam*. Very loud bass sounds can sometimes be felt as literal bodily vibrations, often especially in the chest. Headphones obviously cannot reproduce this sensation since the sound they produce is confined entirely to the ears plus, to some degree, the surrounding head bones.

Photo credit: Vincent Escudero, License

3. Detail

Another variable is *detail reproduction*. Expensive headphones often excel at allowing the listener to hear the faintest details in a recording. Some listeners may mistake extra loudness of the high frequencies for overall detail.

Another common mistake is to assume detail reproduction is a result of *transducer speed*, meaning how quickly an abrupt sound stops sounding. A transducer is the sound producing component inside any headphone and is often called the *driver*. If transducer speed were an issue headphone transducers wouldn't be able to vibrate fast enough to reproduce the highest frequencies they're almost always rated for — at least 20 kHz or 20,000 vibrations per second. Expressed another way, this is one vibration every 0.00005 seconds, which should surely be enough resolution to satisfy anybody.

Surprisingly, simple accuracy of frequency response plays an especially important role in detail reproduction. Perhaps you're used to listening to a particular track on a pair of headphones that have a weak (recessed) frequency response over a certain range of sounds. When listening to the same track on a different pair of headphones you may hear a quiet instrument or voice that you hadn't been aware of as being in the track. Yet all that's happening is likely that the second pair of headphones plays that frequency range at a louder level. A related issue that is sometimes discussed is *acoustic masking*, a louder frequency band drowning out a neighbouring quieter band.

Digging a bit more deeply, fine detail reproduction is a by-product of a precisely tuned transducer. In Fig. 3 we see two more millisecond traces, this time of an electronically generated minimum duration impulse. The trace on the left shows the performance of a by-gone Sennheiser headphone, the HD 222 in trying to reproduce this response. The trace on the right shows the same headphone reproducing the same impulse, but this time the headphone has been equalized to produce an accurate frequency response:

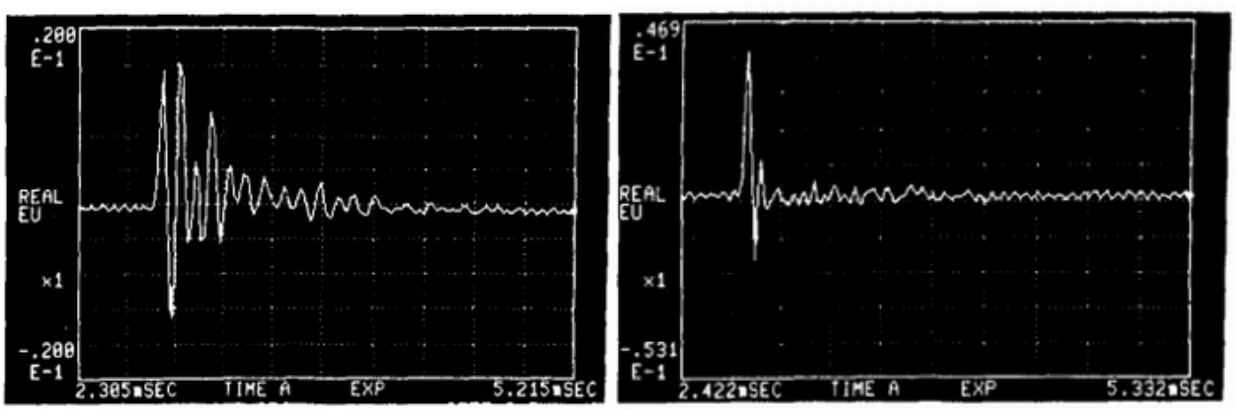

Fig. 3: milliseconds views of an extremely brief sound signal (courtesy B&K via Oratory1990)

Clearly, the un-equalized headphone produces a reverberant sound that lasts visibly longer than the same headphone equalized — an acoustic smearing or blurring of the sound. The equalization is compensating for the lack of sufficiently precise physical damping around the driver at certain frequencies.

Tech note: "For mininimum-phase systems [such as headphones], an improvement in the frequency response will lead to an improvement in the time domain since the two domains are mathematically related via the Fourier/Hilbert transforms. A high-Q resonance in the frequency response will exhibit itself as a ringing in the time domain. Equalize the resonance with an inverse filter so that the frequency response is now flat, and the ringing will disappear in the time domain." - Sean Olive

4. Sound Stage and Imaging

Photo credit: domdomegg, License

Sound stage is an illusion created by the proper recording and playback of at least 2-channel (stereo) sound (although monaural playback can produce a surprising sense of space). Sound stage ideally gives you the sense of sounds coming from all three dimensions of space — left-right, back-front, up-down. *Imaging* is how precise or pinpoint the location of each sound source is within the sound stage.

Commercial recordings are engineered to provide this spatial illusion via playback on loudspeakers. Headphones are handicapped at reproducing sound stage by their inability to do *crosstalk* (the fact that the left ear can hear sound from the right speaker and vice versa). Other sound localization cues are also diminished due to headphone reproduction. There is a special recording technique called *binaural* that targets headphone listening, but this is little used in the recording industry. Despite these obstacles most headphones do create a stereo spatial illusion in which instruments and voices seem to be spread out around us. However, this sound space mostly seems to be too close to, and often right inside, our heads.

Headphone models that produce even a slightly larger sound stage (even if only in the left/right, or width, dimension) are prized. Unfortunately, this is often achieved by introducing frequency response inaccuracies. Another stratagem of placing the transducers at an angle to the ears unfortunately has very limited effect. Using over-sized ear pads to introduce a distance between the listener's ears and the transducer seems to have a relatively small but more reliable effect.

Other considerations

The phrase *sound signature* is used for the combination of frequency response, dynamics, detail, sound stage, etc. for a given headphone. In other words — how it handles sound reproduction as opposed to its build quality, comfort, etc. The *tuning,* or more rarely *tonality*, of a headphone is just its frequency response.

Finally, here's an important but potentially confusing point. Many headphone enthusiasts are not looking for accuracy, especially in the area of frequency response. Enthusiasts tend to treat frequency response like a menu of flavours at a restaurant. Buy one headphone that does bass-emphasis really well, buy another headphone that does V-shape really well, etc. With the other sound variables the attitude is generally the more the better — the more dynamics, detail retrieval, sound stage size and imaging accuracy the better.

✣ ✧ ✣ ✧ ✣ ✧ ✣

And that's a wrap. Be sure to proceed to the next wild and crazy instalment in the epic Headphone Essentials series: *HE-3 The Basics of Headphone Types and Tech* (http://daystarvisions.com/Music/index.html).