PSYCHOACOUSTICS EXPLAINED

In this section we will lay down the psychoacoustic laws and how we can use them to improve our productions, as well as how we can use this knowledge to treat our room, to provide a better listening environment.

WHAT ARE PSYCHOACOUSTICS

Psychoacoustics can be thought of as the way we hear and perceive sound within the brain. This is a very in depth subject; however, in this section we have collated all of the main points and expanded on them with the electronic music producer in mind, We will cover these main points:

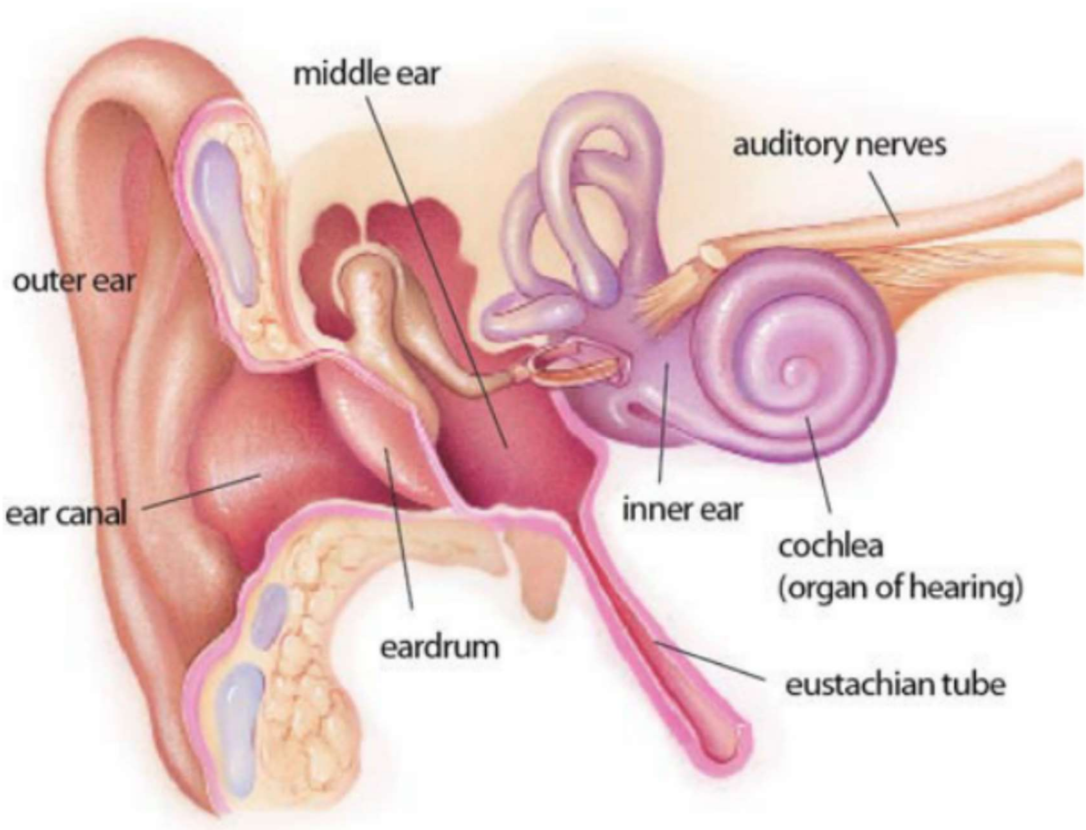
PERCEPTION IS NOT LINEAR
THE HAAS EFFECT (PRECEDENCE EFFECT)
MASKING
CRITICAL BANDS
NATURAL RESPONSE TO LOUD NOISES
EQUAL LOUDNESS CURVES
FLETCHER-MUNSON TO ALTER DEPTH
HEARING IN RMS
LOCALISATION
SOUND LAYERING
DISTANCE PERCEPTION

PERCEPTION IS NOT LINEAR

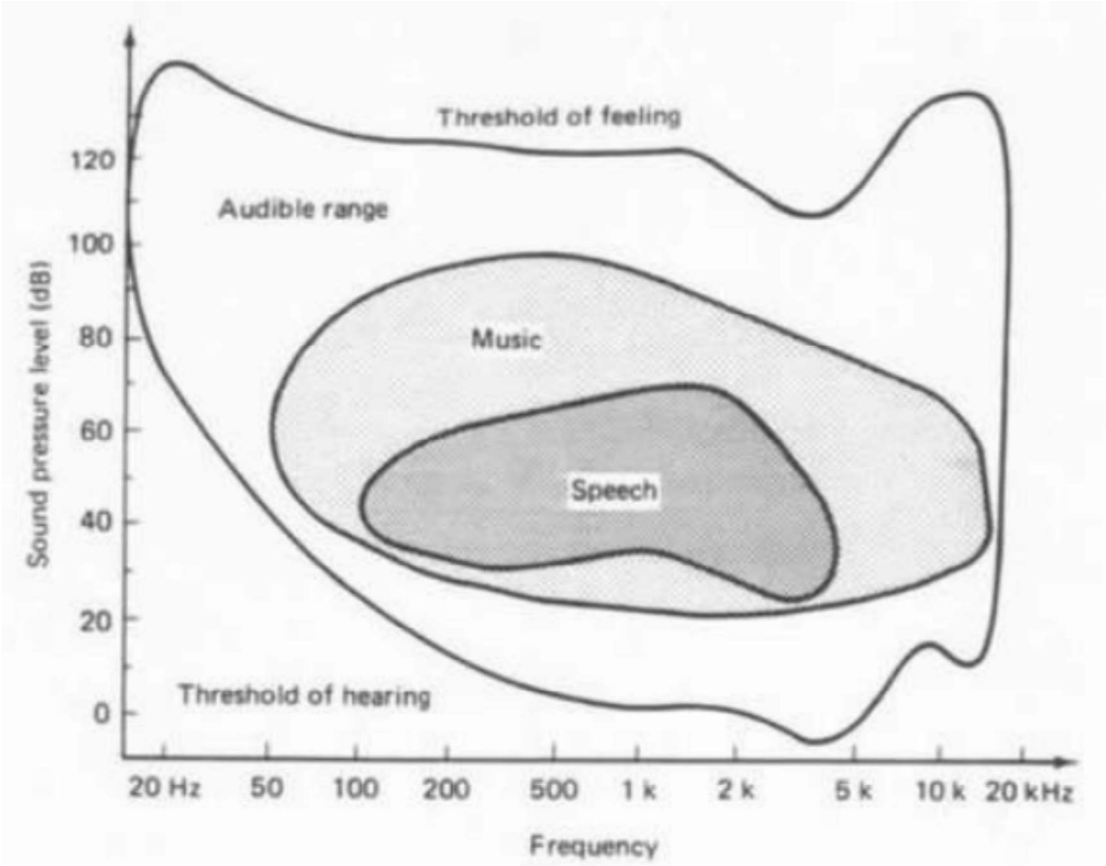
Our first point is that perception is logarithmic. This was discovered and researched by Gustav Fechner, who introduced the concept of psychoacoustics.

The way we perceive weight, light and most importantly, in our case, sound. Is non-linear. This can make it tricky to measure, which will be explained shortly.

Firstly we need to understand frequency ranges. The human ears can detect frequencies in the range of 20Hz - 20KHz. This is due to their inner and outer structure, as well as how our brain processes the neural signals sent from the cochlea.



Anatomy of the human ear



The frequency range of human hearing: Threshold of hearing at 0dB SPL to the threshold of pain at 120dB SPL.

The perceived loudness of sound at different frequencies is also non-linear. Our ears will need a very low frequency, of say a 100Hz sine wave, to be much louder for us to perceive it as the same volume as a sine wave at 1KHz.

DYNAMIC RANGE

The dynamic range of human hearing is defined by the threshold of human hearing, (The lower limit.) and the threshold of pain. (The upper limit.) The threshold of pain is relatively unbiased to frequency, and is capped at around 120dB SPL, which is roughly equivalent to being stood next to a jet as it takes off. The threshold of hearing however, is much more frequency dependant, which can be proved with equal loudness, or fletcher-Munson curves, which will be explained shortly.

The threshold of human hearing can be defined by a 1KHz sine wave measured at 20 micro-Pascal's, which is the equivalent of 0dB SPL.

Just to put this into perspective, the blood rushing through your veins is louder than this, and can be heard if you are in an anechoic chamber. (A chamber that has almost no reverberations.) Also known as a silent room.

Armed with this basic knowledge, we can now tackle some of the phenomenon of the psychoacoustic domain.

An anechoic chamber

THE HAAS EFFECT (THE PRECEDENCE EFFECT)

This is where two near identical sounds, (or an early reflection of a sound) that are no further than 35ms apart, and are within 10 dB of each other, will be interpreted as one sound.

A good way to test this theory is to stand 10m from a concrete wall and clap. You should be able to hear the reverberated signal separately as it bounces off the concrete wall and back to your ears.

This is because the reverberated sound is roughly 60ms or 20 metres (10m there & 10m back) apart from the original sound. (Sound travels at 343m per second (or 1125ft per second.) This approximates to around (1ms = 1ft) or (3ms = 1metre) If you keep clapping, whilst moving closer to the wall, try and stop when you can no longer tell the difference between the two sounds and they just become one unified sound.

You will be roughly 5m from the wall. This is due to being half the distance, the clap now returns to your ears in half the time, which is 30ms.

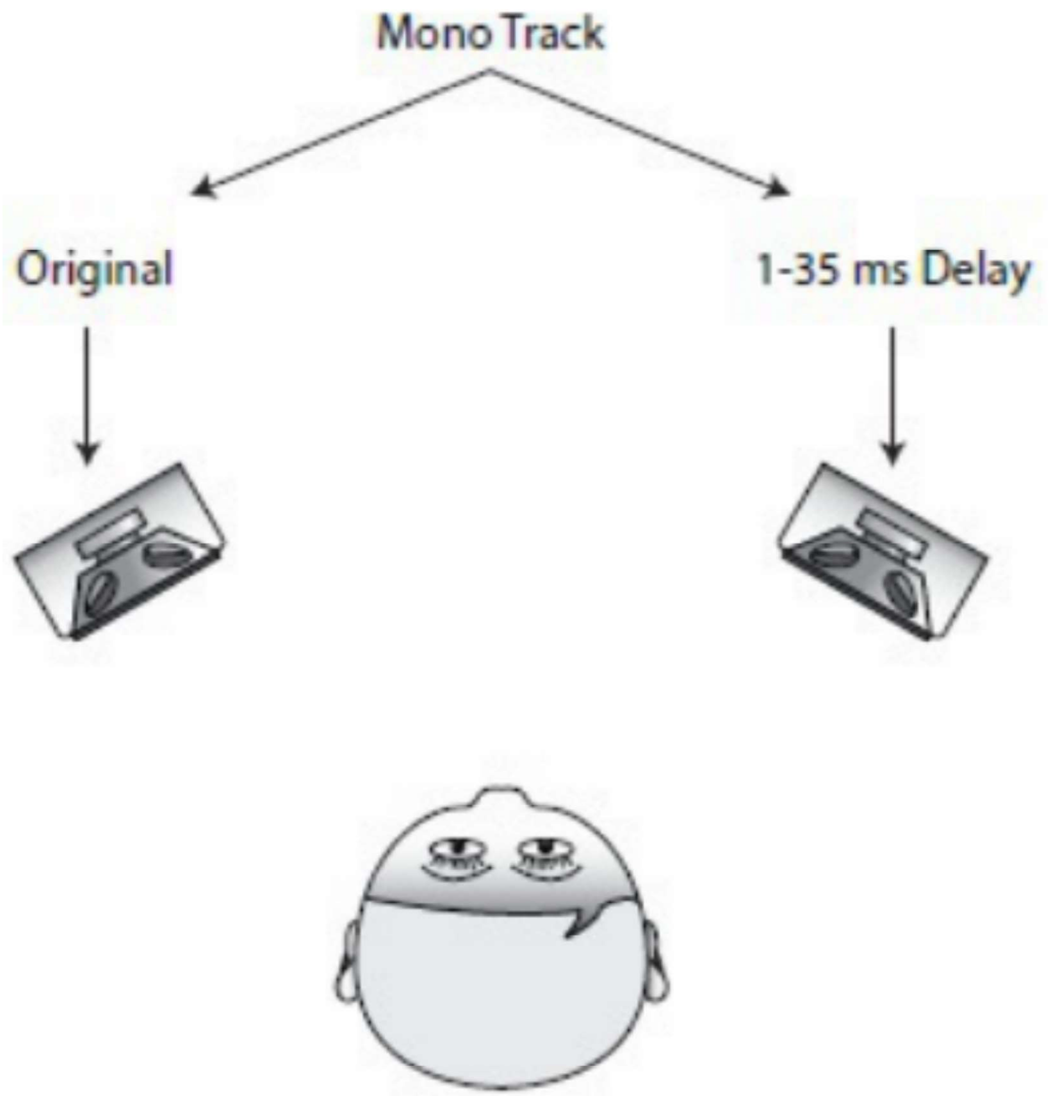
This is the time window where humans can no longer distinguish between the two separate sounds. (The outer time frame of the Haas effect.)

With this in mind, we can use this knowledge to enhance the stereo image of a sound. For example, we can take a clap sound and duplicate it. Pan one clap hard-left, and the other one hard to the right.

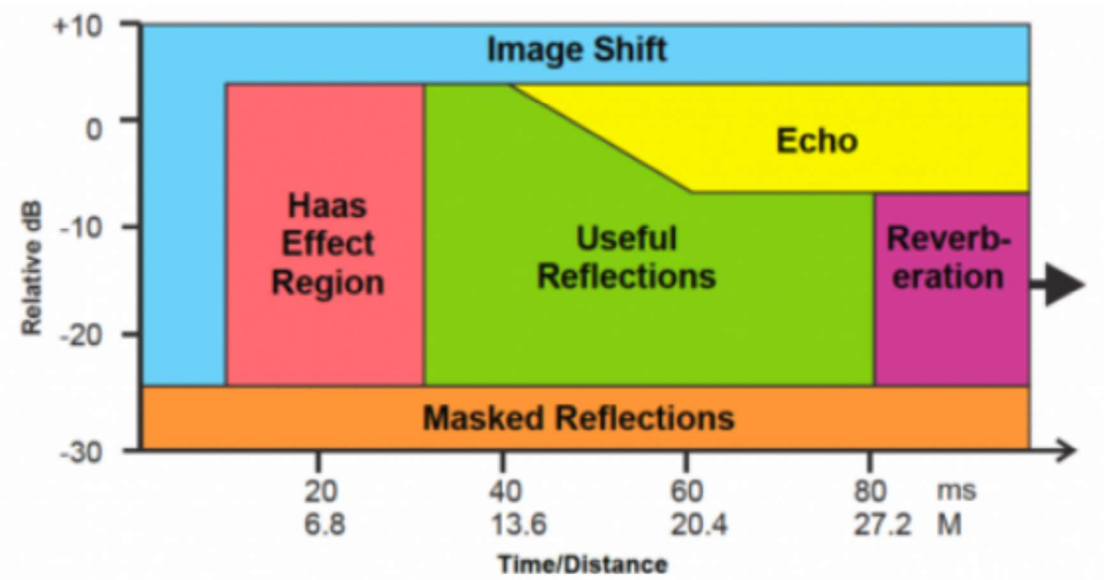
At the moment, all this will do is increase the loudness of the sound. However, if we now delay one of these signals by between 5-35ms, we will be changing the frequency information in each separate channel, making our sound very wide.

As long as we keep this below 35ms, our ears will still think that this is one sound. A caveat to this is that if we go above 5ms, we will start to hear the sounds separate. Going below 5ms will result in metallic and 'phasey' or 'whooshy' sounds.

The Haas window can change slightly, depending on if the type of sound is a transient, or a sustained note. But generally it will be anywhere between 0-40ms.



The Haas effect. A mono track is sent to both left ad right, but one channel is delayed up to 35ms.



A graph showing the Haas effect window in milliseconds, and meters. Within this window, two sounds will be perceived as being one.

MASKING

If two sounds with very similar frequency content play at the same time, and in the same stereo field within a mix, the louder sound will drown out, or 'mask' the quieter sound.

For example: If two people are having a conversation, we can hear them perfectly. However, if they are on a busy street, and a bus was to drive past, there's a high chance that we would struggle to make out their conversation.

This is due to their voices being masked by similar frequencies from the bus, and surroundings. So basically masking can be thought of as the 'drowning out' of sounds, due to louder sounds, (and their harmonics) in the same frequency ranges.

CRITICAL BANDS

This states that masking only occurs over a certain bandwidth, for example: if a narrow band of white noise at 2KHz is played over a 2KHz sine wave, we won't hear the sine wave.
As we alter the bandwidth, frequency, or amplitude of the sine wave, the masking effect will change.

A narrow band of white noise at say 8KHz will have little to no masking effect on a 2KHz sine wave, regardless of amplitude because these frequency ranges are too far apart. On the other hand, if we pushed the bandwidth of the white noise to maximum, meaning that it is the same amplitude throughout the entire frequency spectrum then we are not going to hear the sine wave over the white noise.

NATURAL EAR PROTECTION TO LOUD NOISES

The human ear naturally shuts down to protect the cochlea when it is subject to loud noises. We can think of this as a self-defence mechanism to protect our hearing.

We can replicate this by playing a loud noise, then quickly reducing the sound, before bringing it back up again. This is done throughout music production to make sounds seem louder than they are. From big EDM drops right down to on a transient level.

Some producers get up real close in waveform view and create small fades, or attenuations in volume in between the transient and tone of a sound to further accentuate and empower its impact. This is also done very well in film sound design, for sounds such as explosions and gunshots.

EQUAL LOUDNESS CURVES

As we mentioned in the introduction of this article, the ear perceives the loudness of a sound dependant on the frequency. This was discovered by Fletcher and Munson, who did a lot of research to produce the fletcher-Munson curves, (also commonly known as equal loudness curves.) their findings were that at high volumes, a sound will appear to have more high and low end.
Monitoring at low volumes is better for balancing, and gives a more true representation of the sound. (As well as being less damaging to the ears.)

Drawing from this knowledge that at loud levels the low and high frequency content sounds louder, we can also figure out that by Slightly Scooping out mid-frequency content, & boosting highs & lows when listening at low volumes, we can make a mix be perceived as louder and closer. When in reality all we have done is tricked the brain into thinking its listening at a louder level by reducing the mid frequencies slightly.

FLETCHER-MUNSON TO ALTER DEPTH

To make sounds appear further away, roll off the high frequencies. This simulates sound dissipation through air.

This is because our ears can sense depth perception by the ratio between mid frequency content and high & low frequency content.

As sound travels through air, the high frequencies are dampened or attenuated faster than the mid and low frequencies due to the size of their wavelength. (Which is inversely proportional to frequency).

EARS HEAR SOUND IN AN AVERAGE WAY (RMS)

A sustained sound at any given volume will be perceived as louder than a transient at any given volume. (Ears perceive loudness by RMS values.) This is why, when we use compression to raise the body of a sound we think it's louder or punchier than beforehand, even though the peak level is the same due to raising its RMS value.

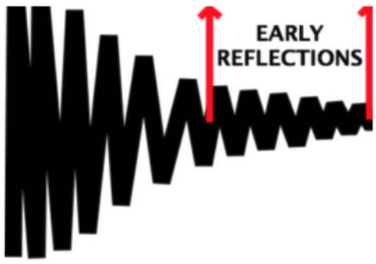
Fundamentally, this rule governs the whole concept of why we use compression to reduce dynamic range, as well as why we use as much head room within a mix as possible to achieve the loudest perceived signal possible.
By combining the precedence effect and the way our ears detect loudness in an RMS type manner, we can beef up sounds using a reverbs early reflections. By adding these early reflections we will be boosting the RMS value of the body of the sound. And as we know, a higher RMS value equates to a louder (and fuller) perceived

sound, due to making the sounds waveform more similar to a sustained note rather than a transient.

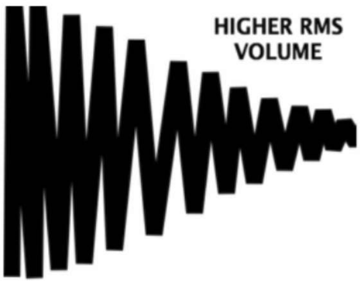
Provided our early reflections are <35ms from the original signal, we will hear this as one unified sound rather than an initial sound and its reflection.

Making sure we don't include the late reflections is important, as this will prevent us from altering the depth perception of the sound.

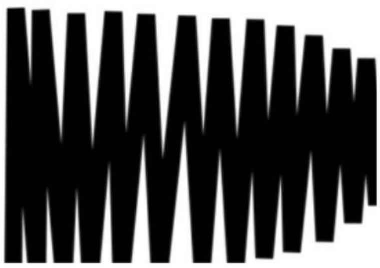
A transient with a high transient peak, whilst having a low RMS value.



Increasing the RMS value by increasing the gain of the early reflections only



The new waveform that is perceived as louder, due to its higher RMS value.



A sustained waveform, which will be perceived as being very loud due to having an RMS value, which will be very similar to its peak value.