Sound reproduction, Human hearing & Drivers of design

An overview of audio matters, their perceptions and technical associations

Peter Hawkins
Hawkshead Designs Ltd

Unit 3 Mount Pleasant Eco Park
Porthtowan
Redruth
Cornwall
TR4 8HL
T: +44 (0)1209 890550

1. Purpose

This document aims to outline the fundamentals of sound reproduction equipment and human hearing. This is used to underpin the requirements of any audio equipment design.

As far as practical the document is intended to be "non-technical" although the subject at hand requires very deep technical understanding to be fully appreciated. In this sense it can only be considered an "overview".

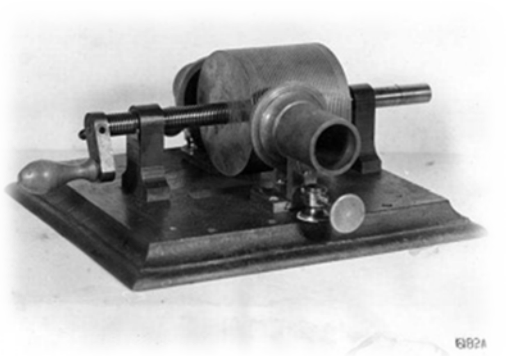
It could be considered flawed to attempt the design and build of any audio equipment without knowledge of the underlying parameters first.

2. Sound reproduction

2.1. Origins

Sound reproduction has origins that predate Edison's phonograph which was invented in 1877. The original goal was merely to be able to replay recorded sound. In the following decades, work and inventions combined to produce improvements in the quality of recording and the ability to reproduce that original sound.

The early stages of the inexorable quest for quality had little need to consider the individual listener. There were sufficient and obvious technical shortcomings that required solutions before the priorities and failings of human perception created parameters of their own.



Audio evolution has progressed in an uneven manner. The majority of progressions are small and randomly distributed. In the history of sound reproduction there have been steps which create dramatic changes in a relatively short timeframe. Some of these steps have been:

- cylinder phonograph to gramophone
- valve-based electronics to solid-state
- all analogue to partial digital systems

The understanding of human sound perception has also progressed in an uneven fashion. This has not necessarily been co-incident with the step changes in technology. In the early 1900's auditory research was dependent on "new" technology to make investigations. In turn this meant that some conclusions were drawn based on the limitations of the equipment rather than the true fundamentals of the experiment.

Two of the earliest foundation findings were the limits of human hearing in terms of frequency[1] response and time. These findings are enduring in many ways but overly simplistic and not fully accurate as will be seen. The findings are summarised as:

- Human frequency response 20 – 20,000Hz (or 20Hz – 20kHz)
- Human time distinction 1/10th second (or 100ms[2])

[1] Frequency is simply the number of times per second that something occurs. It is expressed as Hertz (symbol: Hz). In audio terms it is the pitch of a sound.
[2] A millisecond (ms) is 1/1000th of a second

2.2.Digital era

One of the big steps in sound storage and reproduction has been the move from all analogue systems to those including digital. It may seem odd to consider that there is not an all-digital system but the end point of sound reproduction in this context is a human (analogue domain) often preceded by an analogue transducer[3].

The CD format was introduced commercially in 1982, although penetration to the critical 10% of households was not met for another decade. CD allows for 74 minutes of recorded audio according to the original red book[4] standard. The provision of 80 minute recording time is achieved by squeezing the concentric storage areas of the disc together. This exceeds the original specification but is handled by modern replay devices.

There were research topics known at the time that suggested this new format may not be "perfect sound forever". However, there was considerable commercial value to its promotion and limits to the technology available to use more data at the replay end. The CD was defined as 16 bits with a sample rate of 44,100Hz.

The digital world works on the basis of taking the infinite variation of analogue and compartmentalising it as discrete number steps – there are no fractions of a step. These measurements of "size" are taken repeatedly at very regular intervals. In the case of CD the interval is every 1/44,100th of a second (otherwise expressed as 0.0227ms).

The 16 bit term is a description of how many discrete values exist between maximum and minimum. The values are recorded on a CD as pits and lands like a rough terrain that is read by the laser light and its associated (photodiode) "receiver" through the smooth protective polycarbonate layer. The actual storage and recovery adds another layer of complexity. It is sufficient to say that the numerical values for CD are recorded as 1's and 0's which is binary. 16 bits of binary is (16 lots of 2 x):

2 x 2 x 2 x 2 x 2 x 2 x 2 x 2 x 2 x 2 x 2 x 2 x 2 x 2 x 2 x 2 = 65,536 steps

[3] Transducer is a generic term for a device that turns one from of energy into another. In this context the primary transducers are microphones, (loud)speakers and bone conductors. These are sound to/from electrical energy transducers.

[4] Red book refers to the specification for recording/replay of compact discs using red laser light. Subsequent technologies use blue light and are accordingly "blue book".

Sound is fundamentally based on amplitude and frequency. Amplitude is the measure of height of the wave (similar to its volume or intensity) and frequency is the pitch. Amplitude exists about a zero or mid-point. This means that the step range must be shared between positive excursion and negative excursion such that there are notionally 32,768 steps for each direction.

The sample rate is the number of times per second that the amplitude measurement is taken. The first issue of sample rate is that frequencies above half the rate cannot be reproduced. In CD terms this means that the limit is 22,050Hz (usually referred as 22kHz).

There are strange behaviours to digital sounds that attempt to break the immutable limits of amplitude and frequency, whereas analogue systems tend to fade away gracefully beyond their intended limits.

It may appear obvious that the samples gathered in the recording process must be reproduced in the replay process. It may not be obvious that time is an analogue (at least in this context) and that the precision of that sample timing is all important. The sample only has a valid value at a very specific time. If the recording process were perfect and the replay system wandered in terms of time, by as much as 1/1,000,000th (1µs)[5] then the resultant sound would be “damaged” even though it was numerically perfect.

The consequences of analogue misbehaviour tend to be harmonious. This means that the analogue errors or artefacts add sound of their own which is inherently linked to the sound that is intended. Digital errors tend to be anything but harmonious based on the incongruity of trying to define a varying signal in discrete steps and based on a time interval that also changes when it should not.

2.3. Digital audio goal - CD

The digital audio goal based on CD was initially quality of sound. It did reach the “limits” of the time in the sense that Philips only had a 14 bit DAC[6], whereas Sony drove the specification for the 16 bit DAC that they had at the time.

[5] A microsecond is one millionth of a second
[6] DAC – Digital to Analogue Converter

The CD red book specification uses a robust mechanism for ensuring that data is not lost. There is an inherent system for self-clocking the data to ensure that the timing errors are minimised.

In spite of all this, or because of all this, the manufacturers of turntables saw their market under considerable threat. In the main they were small scale operators, especially when compared with the mega-corporations of Sony/Philips et al. The turntable market and its products had matured over many decades, with several incremental steps to aid performance. Yet it is considered that there was as much development effort in turntables in the entry years of CD than all previous development summed together.

There is something of a moral to this intrusion of the turntable in the digital audio goal. Yes, there was a commercial battle between the proponents of the new technology versus the established purveyors of old technology. Turntable development was, and still is, able to make use of the advanced measurement systems and expertise that had also evolved. If CD had been the all-encompassing quality success at the outset, turntables would have died out more completely and more quickly.

Improvements in CD quality (and for this read digital audio technology as a whole) came about through an understanding that audiophiles[7] have had for years. The resurgence of turntable development was driven by audiophiles who already knew that the pops and clicks of vinyl are often due to poor systems. When the underlying problems are resolved, the performance of vinyl is freed from the limits of standard CD fare.

The reality became clear that the quality output from vinyl in many ways exceeded that of CD. The precise manner of these ways took some while to unravel and some are still debated. For turntables the improvements were down to the enhanced extraction of information from the record. This is achieved by materials choice and design for everything from the platter to the suspension and motor drive and very much includes the tone-arm and cartridge. The analogue derived vinyl replay could be made to perform exceedingly well with none of the artefacts of CD.

[7] Audiophile – Latin: Audio (I hear) + Greek: Philos (Loving) an individual who seeks high quality audio reproduction through the use of specialised equipment

The wonder of CD was mostly perceived in the initial mass market as being free from the pops and clicks of vinyl records and the vagaries of fluff/dust on the stylus ("needle"). The ease with which the disc could be handled and tracks skipped or replayed added to the convenience angle. So the major battle was pitched at mass market acceptability criteria, not solidly in sound quality.

The human approach to sound is nowhere near as trivial as is first understood. Human perception and recognition is a wide scope of quality acceptability. At the low end of the spectrum humans can take very much reduced data to create intelligibility. However, this is not to be confused with the gulf of difference between recognition and appreciation. Appreciation can be purely recreational but can also be educational or experiential. This requires orders of magnitude [8] difference in performance.

Many things in audio are based on a logarithmic scale. An example may be an amplifier whose power output is 50W (Watts) for each of 2 channels. The promotion of a 100W power for each of 2 channels may look like a factor of 2 improvement. The reality is that in power terms, it is only 3dB [9] louder (like 3 clicks of a volume control) than the former model. To gain something in the order of twice as loud would require an amplifier of around 500W.

The gradual steps to improving CD audio have incrementally addressed the issues of timing, filtering, resolution and resolved or reduced strange artefacts and other generators of non-harmonious spurious sound. These improvements are required in dB steps, which often mean considerable work for limited perceived improvement. A combination of technology and improved understanding of human auditory processes and perception have allowed the digital technology to make a worthy adversary for the turntable in the quality arena.

There are two key matters to be understood. Firstly, the red book approach was based on "best performance" (for the time) and intended to address the flaws and inconveniences of turntables for vinyl records. Secondly there was a mass market drive to capitalise on the "digital" future. The rewards of that marketing are in some ways still being felt today. Quality and auditory perception related performance are attributes of follow-on products for niche markets and applications.

[8] An order of magnitude is 10 times, 2 orders of magnitude is 100 times… as such it is effectively a logarithmic scale

[9] dB – Decibel, a logarithmic scale often used in audio to match the perception steps of human hearing

2.4.Formats and numbers

The "Digital Era" section may give the impression that the future of audio is misplaced in the digital domain. There is no doubt about the versatility and usability of the digital domain and these features have no real equivalent in the analogue domain. However, the over-simplification of the digital world has led to many misconceptions. It is the purpose of this document to lead an informed path to the requirements of a digitally based system.

In the world of digital audio there are generally no pops and clicks but there are a myriad of other problems. These are compounded by the style of promotion used for mass market acceptance or desirability. It has long been the case that numbers sell. A 100W amplifier is "better" by market assessment than a 50W amplifier. This approach becomes more troublesome as formats and their associated numbers fall into the window of user consideration.

The 16 bit (resolution), 44.1kHz (sample rate) of CD red book could be considered "old hat". There is even a commonplace parlance which would term this 16/44.1, giving glib or perhaps nil awareness of its implications. Studios and professional applications were primarily based on 48kHz sample rate since they were freed of the history of CD. Instead they became based on the easy multiples (or more truthfully divisions) of accurate local crystal clocks for precision interval timing of the samples.

The numbers then increased from 16/48 to 18/48 and 20/48, whilst there was also an evolution to faster sample rates. 24 bit data is hard to truly achieve in application since the outcome must exist in an analogue world. There is a fundamental trade-off between speed and accuracy in many systems, which very much applies to audio.

Thus 24 bits at 48kHz poses many problems if accurately measured output performance is the arbiter. Moving from 24/96 to 24/192 creates a level of theoretical performance which has yet to be fully achieved in application practice, anywhere. On the whole, the full scope of all aspects of 24/96 promise is not delivered. With the right technical understanding, 24/96 structure can help with certain aspects of audio management.

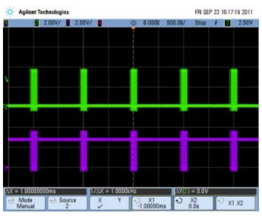
It comes as no surprise in a market driven environment that many proud boasts are made of 24/96 in budget consumer products. To pass the "market" test legally, all that has to be demonstrated is the 24 bits of data are transferred at a time interval of 10.4μs (0.0104ms). There is no requirement to state that the output falls way short of the performance implied.

The implied performance of 24 bits is 144dB of dynamic range, whilst studio products are probably achieving 20 bits at best and 21 bits is the level of achievement in test lab conditions. To handle this, the 10.4µs sample rate requires data in excess of 6,144,000 bits per second and a clock rate in excess of this, with a preferred timing accuracy of better than 0.0406µs (0.0000406ms).

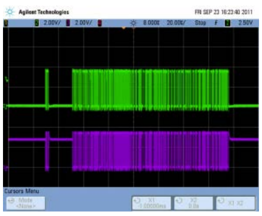
Once all the digital goals are achieved, then the analogue audio electronics must maintain the dynamic range. The analogue domain must not add to distortion and is at the mercy of its power supply to assist in this regard. All analogue operations emanate from a DAC which is in the presence of very high speed digital contamination by radio frequency signals that are fundamentally part of the system.

2.5.Other systems

Broadcast TV is still being handled in the analogue domain as the switch over occurs in the UK to “digital” terrestrial (as opposed to satellite). When this moved from the original Mono (monaural) to Stereo the process was called NICAM[10] and is broadly comparable to a 14 bit system.

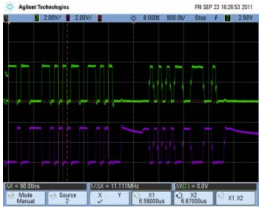


USB audio does not in itself define a performance but there are implications in the discussion above. Observation of the data transaction in USB may give some idea of what is happening and how potentially damaging this can be. The top image shows bursts of data. There are 1,000 crunched up sets of data per second. It appears busy until it is compared with the permanent flow of data from CD 44,100 times per second.



Once the time image is expanded it becomes apparent that the amount of data in this supposedly small cluster is quite intense. The last image takes this a step further to look at the “size” of one bit[11]. It is 90ns (0.00000009 seconds) long. The technology that lies within USB and its associated components is without question a remarkable and effective technology.

[10] NICAM – acronym for Near Instantaneous Compound Audio Multiplex
[11] Bit – basic unit of information in computing/software



These images are all taken from a PC playing MP3 file through an external USB based audio device. In spite of all the technology, it should be remembered that the amount of data passed across from which to create the final sound is now much less than a 16/48 or 16/44.1 system. Less data cannot mean anything other than reduced quality.

On the subject of quality it is also worth looking briefly at the enormous impact of the various Apple™ products (i-pod, i-pad etc.). Iconic styling and clever marketing, including how the specifications are written, have ensured that the uptake is phenomenal. The specification for these products amounts to barely 20Hz-20kHz, with frequent capability being less than 18kHz. The output performance (analogue) is on a par with 14 bit technology.

This is not a slur on the brand value and market suitability of the products. It is an engineering observation and is only to be expected when considering the essentially limited resources available. The devices are all hand held, battery powered and feature rich including a power hungry display. All of this has to be achieved at an attractive market price, which is targeted for affordability not affluence.

In the context of a commercial proposition, the Apple products are exceptionally well placed. In audio performance terms they meet an acceptability of the masses but are not even equivalent to a budget household stereo system.

3. Human hearing

This section is not intended to be complete dissemination of material relating to:

- ear structure
- fundamentals of human hearing
- basic spatial awareness and signal interpretation

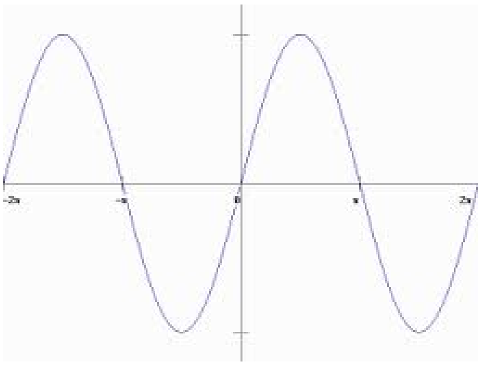
It is assumed that the above is either known or adequate covered in other publications.

Human hearing in this context is looking at the performance capabilities and needs. The focus is on a Tomatis therapeutic goal. The mention of psychoacoustics is necessary as it is the interpretation of sound that is being tested and trained by the Tomatis method.

3.1.Basic range - frequency

Early experiments on hearing sought to "test the human" as a means of defining when the reproduction equipment was good enough. These experiments were limited to the available technology at the time and could not explore the greater implications of underlying topics.

These experiments firstly sought to look at frequency response. This is the range of frequencies or fundamental tones that the ear can receive. The outcome of 20Hz–20kHz is still primarily valid today. It is recognised the higher frequencies are limited to 16–18kHz with advancing age.



These tests were all done with sine waves, which is a single pure tone. Humans and indeed animals, almost never use single tones. Many common sounds including voices, car engines and musical instruments are rich with harmonics. A harmonic sound contains a single pitch overlaid with differing proportions of 2X pitch, 3X pitch, 4X pitch and so on. The sound picture with harmonics is remarkably different from that based on pure tones.

3.2.Basic range – sound level

Amplitude is the size (or could be considered volume) of a sound wave. Humans rarely hear “nothing” and any degree of Tinnitus will leave a residual “ringing” that detracts from that so called perception of silence. Silence in an open natural environment is hard to achieve and yet perceived differently to that of being in a man-made anechoic chamber. Silence is a somewhat vague notion of nothing.

In engineering terms silence can be defined as 0dB (SPL)[12] and at this level the ear drum is moving 1/100 of the length of an air molecule. If this base point is used then the other sounds are scaled as follows:

10dB	Engineered quiet room
15dB	Pin drop from 10mm at 1m distance
30dB	Very still quiet night/not near any city
40dB	Whisper
60dB	Conversation
90dB	Stereo on loud but not maximum
120dB	Front row in loud rock concert
130dB	Very loudly played brass instrument at about 1m
140dB	Pain threshold/hearing damage even at short duration

There are many curves that have been produced for equal loudness perception in humans. This means that the perceived amplitude (volume) of some frequencies (tones) is very different from other frequencies. The most prolific of these curves is probably the Fletcher-Munson curves. Using these curves, the “silence” above is only discerned in a range of about 800Hz – 8kHz. An increase in amplitude will create an increase in sound pressure level such that all frequencies are audible with a degree of equality at around 90dB (SPL).

Co-incidentally (or not!) 90dB is the approximate volume of a TV or sound system when the user is “listening”/paying attention. This presupposes that there are no social constraints on the use of this volume, such as other household members. Conversation would require raised voices at this level, causing stress and a detraction from the attention paid to the sound.

[12] SPL – Sound Pressure Level

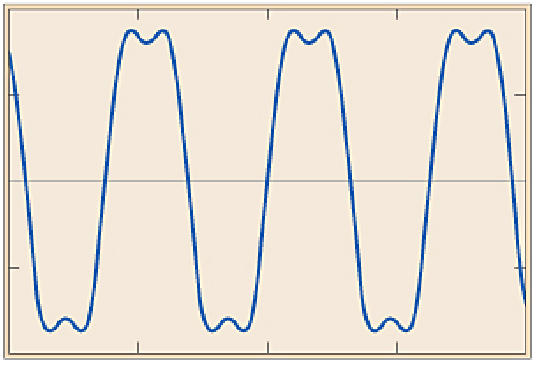
In other words the hearing response is tailored and not surprisingly centred on the human vocal range. This is not the only adaptation to a special set of priorities. Hearing and its perception is a survival mechanism in origin. This means that there are other attributes of sound importance to humans that extend beyond simple frequency and amplitude.

3.3.Basic range –time: phase

A man standing close by and talking quietly might generate 50dB SPL near the listener's ear(s). The same man standing much further away may be shouting to achieve the 50dB SPL. The "speaking" man remained the same, as did the amount of sound reaching the listener but something changed, since the perception of the two scenarios is very different.

The man talking quietly is taken as a base line for the scenario. His voice will be unique in its tonal character and importantly its minute degrees of variation in emphasis on certain sound types. The man standing further away is stressing his vocal chords and moving more air (locally to him) so as to achieve the target of 50dB at the listener. The listener is hearing the same level of sound but the tonal character and emphasis will have changed.

The distinction between musical instrument types and the strength of the notes played are all based on the same discernment premise as above. To this end there are some ore definitions of basic range with which to contend.



The image depicts a sound wave that is rich in harmonics. Without a mathematical treatise it is not easy to portray the considerable quantities of 3rd, 5th, 7th & 9th harmonics present. It is more visually evident to simply observe the "dimple" that occurs at the peak and trough of every wave.

Provided that the fundamental frequency and its harmonic frequencies maintain their time relationship, they will look like the waveform above. If the harmonics are delayed in proportion to their own frequency (say 30μs for 3rd harmonic, 50μs for 5th harmonic etc.) then the waveform will not look like the above. The newly "time shifted" waveform will contain the same energy, the same frequency components at the same amplitudes… it will neither look nor sound the same.

The relationship just described is called phase. It is a time relationship based on the coherence of the fundamental frequency with its harmonics. Sometimes phase is expressed in degrees because in terms of mathematics the proportional nature of the frequencies and time allow analogy with circular motion and angles. With 0° phase shift there is no time difference and the harmonics will be utterly consistent (faithful in reproduction) to the fundamental to which they belong.

There are no immediate definitions of the acceptable phase angle or time determination for humans. Early experiments of around 1932 used a pair pure sine waves of the same frequency and amplitude to be fed one each to the ears of a number of human listeners. It was shown that varying the phase of one wave in relation to the other caused no discernible change in any listener. Thus it was concluded that phase played no part. This has since been shown to be very simplistic and false for anything other than a pure sine wave.

Observation and general technical direction shows that, in amplification, the response for 20kHz that has a phase shift of 45° at 50kHz is "acceptable" but detectable. This corresponds to 2.5μs of time shift. It should then be observed that:

- 50kHz is inaudible in direct terms
- The time period of a 20kHz waveform is 50μs
- A time difference of 2.5μs represents 1/20th of 50μs

Preserving phase is a subject matter to be covered later. However, the simple implication is that a system whose parameters are defined purely on a 20Hz–20kHz frequency response cannot re-create the level of adequacy needed for attentive human audition.

3.4.Basic range –time: location

Continuing the theme of hearing as a survival mechanism, it is not sufficient to be aware that a man is talking or shouting at a distance. It is a benefit to know that this is a man and not a predatory animal but it is of some importance to know the location of the sound source.

Our ability to survive in the evolutionary sense and to operate on a daily basis comes down to the same basic process. The sound of interest is:

- Observed with greater "focus" (brain processed preferentially)
- Based on a complex pattern of frequencies and their harmonics
- Heard at some distance

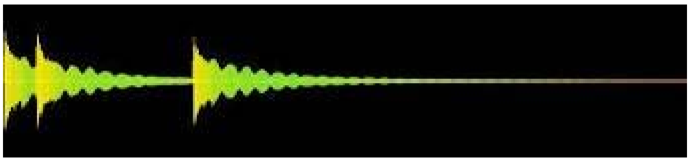
The brain function related to hearing is a vital part of listening. The sound pressure waves reaching the ear are not all treated equally. It is the translation and interpretation of these waves that become the listening process. Some of this process is based on life experience and some of it is more ingrained and developmental.

The developmental patterns of sound awareness and recognition occur from the first month after birth, although sounds will have been heard from within the womb. The temporal lobe of the brain deals with learning, language and smell. The temporal lobe becomes more receptive and active at around 3 months old. It can take another 3–4 months before location awareness develops.

Part of the processing for spatial location (distance, angle from you and height above/below you) is based on the subtle differences between the sounds received in the right ear compared with those received in the left ear. Time differences are the most commonly known dissimilarities between the sound receptions at each ear for location. This may be referred to as Interaural Time Difference (ITD). There are also some variances in the sound levels of different frequencies, known as Interaural Level Differences (ILD).

The time it takes for sound to travel across the width of a head is around 660μs. The perception of time difference to achieve location at a distance requires acuity in the order of 3μs.

3.5. Envelope of sound



The considerations of hearing up to this point have been based on frequency, intensity or amplitude and time. The image is a representation of a bell being rung three times. The frequencies and harmonics are not visually clear but are evident when heard. The time interval between the first two strikes is not great but readily discerned. The interval to the last ring is greater and also easily interpreted.

One of the many facets of hearing is that sound is important for its shape or envelope. The very abrupt change of the bell from its “static” condition to its resonance after the strike is a critical to the comprehension of the sound. This is very much the case in language where initial consonants often shape an identifiable envelope in an expressed word.

The sound continues in its range of frequencies and harmonics but these vary in intensity. There is a process of constructive and destructive interference occurring in the bell causing peaks and troughs of emitted sound energy. Gradually the total energy is dissipated and the sound fades in intensity as a result.

It is understood that the envelope of a sound has as much to do with language comprehension as time and frequencies.

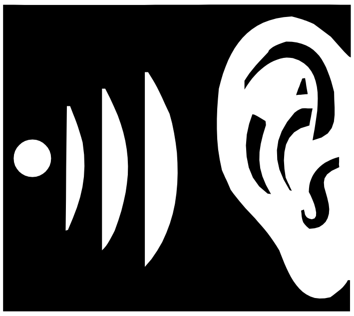
3.6.Summary

Through the previous headings there is now a series of basic criteria that offer some insight to the requirements of full stimulation of the human hearing system:

Dynamic range (at output):	Better than 100dB
Equates to:	20μV noise for 2V signal and
	Better than 17bit resolution (at the output)
Timing resolution:	In the order of 5μs
	(requires high speed analogue)

4. Human Perception

Sound perception is scientifically called psychoacoustics. Unfortunately this term is often perceived as a “mysticism” behind which all manner of ill-contrived ideas can be masqueraded. Despite the erosion of the term,



psychoacoustics is a necessary part of understanding the human approach to sound. It becomes particularly relevant when considering an engaging audio experience.

Psycho-acoustics covers the brain’s interpretation of received sound. It has been seen that this process involves everything from time differences to frequencies and the envelope shape of waveforms. These are primarily "learned" in healthy babies in the first 6–7 months of life.

4.1.Harmonic acceptability

Many pieces of audio electronics quote distortion as a figure of merit. All too often the figure is not defined any further, which is scientifically and factually unsatisfying. Other literature (marketing) takes an approach that gives less credence to some artefacts making the overall distortion figure appear to be lower.

It is common and widely accepted that anything other than the desired signal should be measured. This gives rise to Total Harmonic Distortion and Noise (THD+N). This is an effective measurement in purely electronic terms but still does not fully achieve the somewhat different goals here.

Two different devices could come out with the same THD+N figure, which can be expressed in % or dB. One device could be noisy, producing a gentle hiss on audition. The other could be quieter but with more distortion artefacts. This will sound different to the noisy device.

It has been shown that white noise can be readily ignored by human perception. Indeed the effect is so marked that noise is added and shaped in some digital systems to create a "better" result. It starts to account for why some things can measure badly and sound "good".

Similarly the harmonic structure of sounds is important. Naturally (real life/voices etc.) there is an array of harmonics to accompany almost every sound. These harmonics, in their different proportions, make the sound identifiable in its characteristic. However, there are often mechanical or physical constraints to the range of harmonics presented to the ear. Consequently, natural sound may have harmonics but often limited in number.

Transformation of sound to and from the digital domain is fraught with processes that can cause artefacts that are not at all harmonious. Digital systems and their associated analogue can also produce harmonic distortions that reach out to 24th harmonic or more.

The artefacts that are not harmonious are "alien" and have no parallel in nature. The interpretation of these sounds is, at best, fatiguing to the listener but often causes a desire from the listener to have the sound removed or simply “turned off”.

Harmonic distortion that includes multiples that are “too high” is more complex. The precise balance and nature of the harmonic array affects the listening outcome. A natural array of say 2nd, 3rd, 4th & 5th harmonics that is then augmented with less natural 18th, 19th, 20th & 21st harmonics will initially cause interest in the reproduced sound. The unnatural presentation will often be aggravated by associated amplification/headphones which are dealing with unusually high frequencies. Fatigue and frustration are common listener experiences, with the sound being described as “etched”, “gritty” or even “tinsel-like”.

Systems that produce these “alien” artefacts and harmonics that are “too high” may not be very noisy and may not have a great deal of the undesirable distortions. As a result the THD+N can be relatively low. This is the converse to the above situation, where things can measure well but sound bad.

4.2.Masking

Auditory masking is a far more common experience for many listeners with the proliferation of MP3 files used in portable playback devices.

The concept of masking can be understood by understanding a hand clap in two different environments. The hand clap could take place in a library. Here there may be the muted sounds of people moving quietly, pages gently turning and small noises of chair movement. The hand clap is a significant and intrusive sound in the environment. It would cause some listeners to undergo a rapid shift of awareness away from their current function towards the newly injected “disturbance”.

The same hand clap at the same intensity could take place on a busy street. Here the sounds are loud and powerful emanating from traffic, multiple conversations and other sirens/safe to walk signals. In this context the hand clap is neither significant in size nor a danger competing for attention. Quite simply, it can be ignored and make little difference to the overall “sound picture”.

Human perception does not weight all sounds, envelopes or frequencies equally. Similarly the context of a given sound within the whole array makes a difference.

MP3 files are significantly reduced in data size compared with their classic full range (PCM[13]) counterparts. The data compression is lossy meaning that data has to be thrown away. The data rate and bit size is not directly comparable to PCM because of this. Part of the data that is discarded relates to sounds that are considered (by mathematics) to be masked by the more prominent sounds.

From an intelligibility or simple recognition perspective the MP3 masking is satisfactory in outcome. Unfortunately human perception has another dynamic.

Instead of the hand clap on the busy street, it is possible that it is a conversation that needs to be had or a mobile phone call taken. The intensity of noise could suggest that the conversation is "masked out" or greatly removed from awareness. Since the conversation is directly relevant to the listener in that environment, great effort will be made to "focus" on the desired sound or voice.

It is the ability of human perception to change dynamic that makes masking by fixed rules problematic. Even, or perhaps more importantly, in reproduced music/sound the presentation of a complete sound picture allows the listener to focus on different instruments or aspects of the sound.

The listener's ability to dynamically change perception is a small part of the creation of a believable sound image. In the context of Tomatis where a new exercise or training is required, this attention-pleasing plausibility and dynamic perception is fundamental.

4.3.Air

The concept of "air" in audio terms might be a challenging one. In everyday life air is noiseless and only becomes an issue when it is moving. In movement outside, wind may be directly perceived on its passage through leaves on a tree. Indoors draft might whistle through a crack in a door or window. These are movements of air…

[13] (L)PCM – (Linear) Pulse Coded Modulation is the standard form of coding for CD, DVD & Blu-ray + computers. It is a direct digital representation of the analogue waveform.

Listening to recorded sound through speakers in a living room has many constraints and problems. Not least of these are the flat ceiling and walls of the room. These create sound reflections that are not part of the original sound and disturb the context of the auditory spatial clues for listening. Auditioning in this environment is like listening to a box (speaker) within a box (room) where the difference in proportion between the two boxes is too small to be ignored.

The "air" to be discussed is that perception of space, even when constrained in a "boxed" listening room. The achievement of this auditory misdirection[14] is based on several effects at one time.

Masking is one of the primary effects. The sound that comes directly from the speakers is more dominant than the reflections along the ceiling or walls. The temporal lobe processes the dominant sound and "ignores" the remainder. The masking process is useful in this situation, especially since the "reflected" sounds are "right information" but delivered at the wrong time.

A sound system that is designed with the facets of human hearing correctly interpreted has the potential to offer some significant auditory spatial cues. These cues are the signature elements of sound that can be used by the brain to make interpretations of space, volume and localisation of instruments or voices.

To enhance the acoustic experience, audiophiles often spend a considerable time placing their loudspeakers within the room. They may also use curtains or other less acoustically reflective surfaces to reduce the impact of the reflected secondary sounds. All these adjustments result in a cleaner presentation at the audition point, with less acoustic clutter for the temporal lobe to process.

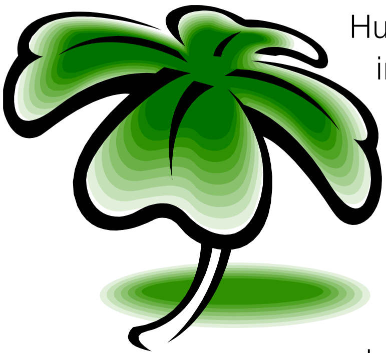
The outcome of accurately timed and reproduced audio is a presentation of auditory spatial cues. These cues allow the temporal lobe to recreate a concept of size, scale and context of all the different sounds being reproduced. The audiophile approach removes some derelictions of the sound reproduced in the room, which enhances the human perception process.

[14] Misdirection in audio perception is allied to the kind of misdirection applied in close up magic. It takes advantage of the limits of human mind to give falsified understanding of what is observed by sight and sound

The difference between a soloist in a church and an individual in a hushed library is largely a matter of space and acoustic environment. A recording of these two venues requires some skill in the making. However, the captured sounds would contain a huge array of small nuances affecting the sound.

If these nuances are reproduced faithfully, then the auditory cues are recovered and presented. If the presentation of cues is executed in a considered environment, then the cues are more easily discerned and the experience is enhanced. The perception that is achieved through this process is natural and has "air" about it.

4.4. Suspension of belief



Humans perceive reality through the senses. An irony exists in that reality is a perception and that reality is an individual view. There is neither time here, nor immediate need for an excursion down a philosophical and sociological path. Substantial evidence supports the notion that whilst the same physical interactions and historic events exist, separate individual experiences may not be identical.

If all the facets of human hearing are combined with the elements of human perception (and sociological factors), the stage is set for a higher realm of listening experience.

The process of going to a large concert hall to hear a full orchestra embodies a number of sociological and logistic requirements. On arrival at the hall the size of the building and the social tension created by expectation create an emotionally perceived environment. Once the music starts, the scale, range, volume and localisation of the whole performance is clearly manifest.

The perception of the concert hall experience is then based on the individual's reaction to it and memories of it.

A recording of the same performance could be made and taken to a domestic setting. In this setting, there needs to be social acceptance, comfort (seating, adequate warmth) and a feeling of well-being and potential to relax. These factors coincide with the most important human needs expressed in Maslow's Hierarchy of Needs.

Given the right environment in the domestic and social setting, a reasonable opportunity for the reproduction of the recorded performance exists. Sound reproduction equipment (Hi-Fi) that adequately caters for the perception and hearing criteria discussed earlier is employed.

The listener is immediately visually aware that the orchestra is not in the room. There is also a cerebral reconciliation that the physical evidence of room size does not match with the occupancy at the concert hall. Depending on the individual focus, the sound may appear to be coming from the speakers not from a large array of musicians in a concert hall.

The light in the domestic room is reduced (not dark) and the music continues. The listener recalls memory experience of the real event and understands the auditory and perceptive experiences of the current event. At some point, before many minutes, the listener will engage with the current reproduction. The level of this engagement is dependent on a many factors many of which are highly specific to the individual.

The important factor in this engagement process is that the reality of the situation has been "adjusted". If questioned literally about the setting and the music, the listener would immediately return from the adjusted, relaxed state to a highly aware state. Any answers would be based on the tangible reality of the size of the room and the sound from the speakers.

In the "adjusted" reality and without any interruption, the listener will continue to absorb the sound and its auditory cues. The emotional and psychological connection with the reproduction remains intact. At this point, the interpretation of room size and sound origin is put on hold.

The listener is no longer bothered that the sound may emanate from a pair of speakers. Similarly the orchestra cannot be in the living room but the experience is sufficient that it does not need to be. The "realities" are intellectual considerations that can be suspended for the duration of the experience.