Rapid Energy Decay (RED) loudspeaker drivers

"In space, no one can hear you scream" is an iconic quote from the first Alien movie, but why is it true? In space there is no air and without air there cant be any changes in air pressure, as all sounds are created by changes in air pressure, there cant be any sound in space.

Here are some discussion points:

(1) From a snapping twig to a 100 piece orchestra in full swing all sounds are "APE" ie Air Pressure Events, an increase in air pressure (compression) or decrease in air pressure (rarefaction).

(2) The only variables are: The duration of each APE which is measured in time, the level of each APE which is measured in Decibels (dB) and the frequency of the APE's which is measured in Hertz (Hz).

(3) APE's are defined by TIME not frequency, they have a measurable duration ie A start time when the increase in air pressure starts followed by a decay time to allow the air pressure to decrease and stabilise back to ambient. The step response and Cumulative Spectral Decay (CSD) plots indicate how accurate a loudspeaker driver or complete loudspeaker design is in the time domain.

(4) The frequency of a sound is the direct result of the number of APE's per second.

(5) Real sounds / APEs have incredibly short rise and decay times lasting only a few millionth of a second (microseconds), unfortunately, current "cost no object" loudspeaker drivers can only manage decay times measured in tens or even hundreds of milliseconds (thousands of a second) which is several orders of magnitude too slow.

(6) Any delayed resonance or "ghost echoes" caused by loudspeaker drivers "bouncing around" or resonating on their mechanical suspensions, after the initial electrical impulse, will distort the sound in a fundamental way. This time domain distortion is the single worst type of distortion in the audio chain and is instantly detected by our Human Auditory System (HAS).

(7) This fundamental lack of loudspeaker cone / diaphragm control is the reason conventional loudspeakers fail to recreate accurate life-like sounds in 3 dimensional space and this is the most important factor in all gaming audio.

The solution?

We need a mechanical loudspeaker driver which can react much faster, in an ideal world we need a mechanical driver which can react at the speed of electrical components... Now thats a tall order!

Obviously there is much more to APE's and the time domain in sound than these discussion points but its a fascinating subject which is only just beginning to get some wider acceptance. Audio industry guru John Watkinson has written several best sellers in the audio sector and the following pages are his latest thinking regarding the vital importance of the time domain in audio.

Of particular interest in the gaming market is the way in which our HAS locates 3 D sounds all around us and how our "Fight or Flight" response has evolved.

Over millions of years we (mammals) have evolved to detect **where** sounds are coming from and prioritise this information over identifying **what** a sound is. For example the instant a primate hears a sound the number one priority is to instantly flee away from that sound, only after it has reached a safe distance, does it have the luxury of identifying what the sound is ie a snapping twig from a predator or just a harmless falling coconut… Better safe than sorry! Today we still detect and process sounds in exactly the same way.

https://www.theregister.com/2014/07/02/feature_the_future_loudspeaker_design/

Design, DSPs and the debunking of traditional hi-fi

Wed 2 Jul 2014 // 13:01 UTC234 GOT TIPS?

John Watkinson BIO EMAIL

SHARE

Feature Today’s loudspeakers are nowhere near as good as they could be, due in no small measure to the presence of "traditional" audiophile products.

In the future, loudspeakers will increasingly communicate via digital wireless links and will contain digital processing. Indeed, the link between IT and loudspeakers is destined to grow.

But no progress can be made when science is replaced by bizarre belief structures and marketing fluff, leading to a decades-long stagnation of the audiophile domain.

It’s a scenario ripe for "disruption", as they say, and there's an opportunity for a profitable IT company to move into loudspeakers and deliver products having undreamed-of quality. Digital guru John Watkinson writes for us today with some, er, sound thinking on how IT should rule the waves.

Speaker design hasn't really moved with the times – Pink Floyd *Ummagumma* image by Hipgnosis (1969)

The criterion for loudspeaker performance is purely what the human ear will tolerate in different applications from tiny handheld IT devices upwards. Ultimately performance is limited by the laws of physics and communications theory, but thanks to psychological factors, aided and abetted by accounting and marketing, actual hardware often falls far short of what is technically possible.

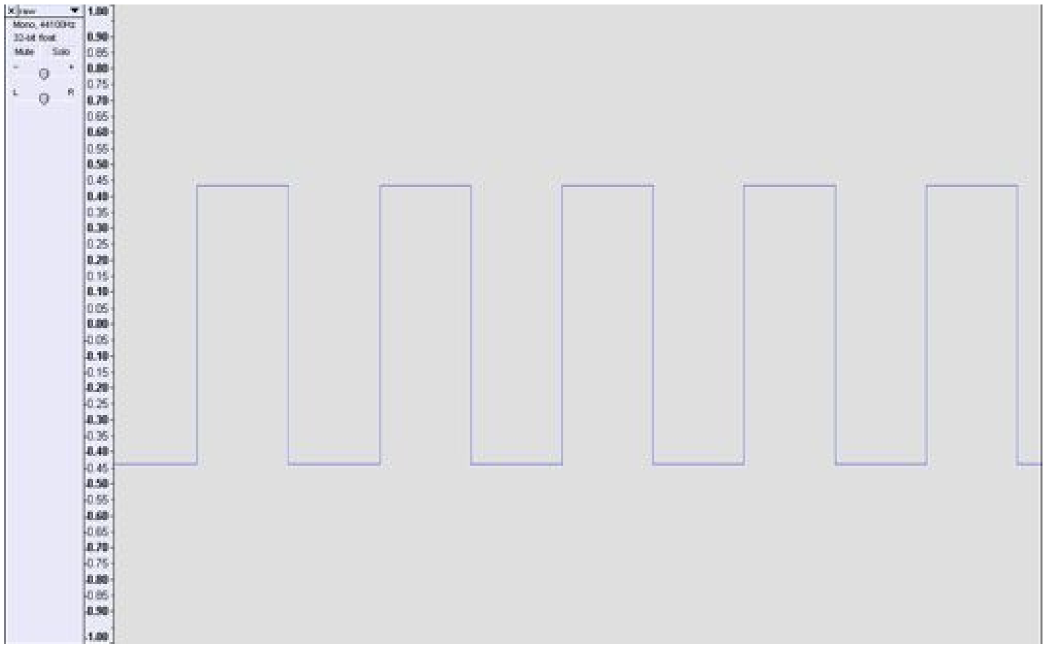
The inner ear is a peculiar transducer that is filled with liquid and may reflect our origins as sea creatures. Sound in air suffers an impedance mismatch at the surface of a liquid, yet the ear has evolved to have remarkable sensitivity by using an impedance-matching mechanism consisting of a series of bones acting as levers between the ear drum and the transducer proper. Such an unlikely arrangement would appear to result in a score of Darwin 1: Intelligent Design Nil.

It would seem logical that if the shortcomings in real loudspeakers could be made a little less than the shortcomings of our hearing, we would believe them to be perfect. So there is, in principle, no technical reason why a perfect-sounding loudspeaker shouldn't be made, even if it won't be hand-held. The rarity of such devices suggests that the reasons are *not* technical and that the application of logic is absent: almost the definition of audiophile behaviour.

Just as cartoons can elicit responses that trigger memories to make them more lifelike, low resolution audio can pull the same tricks

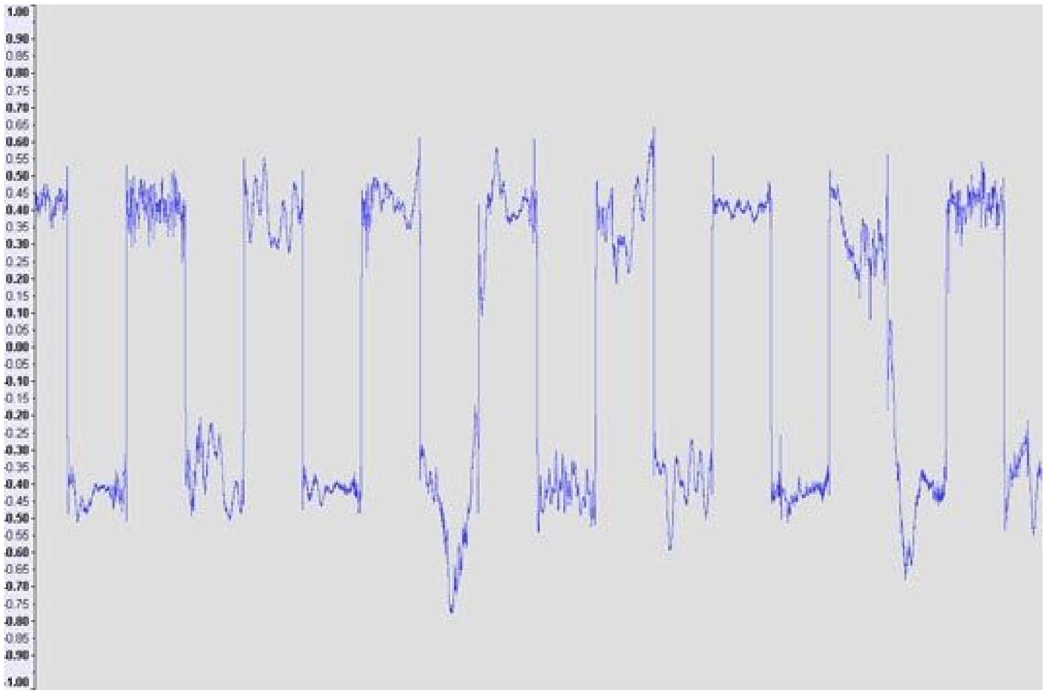
Many of the reasons are psychological. Although humans are equipped with a remarkable range of senses, they appear to be under-utilised most of the time. Cartoons, caricatures and souvenirs are all severely information-limited versions of the original sensation, yet they appear to elicit much the same satisfaction as a more faithful rendering, possibly because the true rendering is in the imagination and the reproduction simply acts as a memory jogger.

Most of the time, most people are remarkably uncritical; some are practically begging to be gulled and their needs are avidly met. The technological revolution that gave us radio and sound recording happened so long ago now that all of the true innovators have retired or passed away to be replaced by bean counters whose only skill is to make things cheaper and worse.



Square wave signal input applied to MP3 encoder – see below

The sound reaching a listener has passed through a communication channel that includes a number of stages that can restrict information capacity. With the advent of the Compact Disc, the bottleneck became the loudspeaker. The subsequent development of compression algorithms complicated matters.



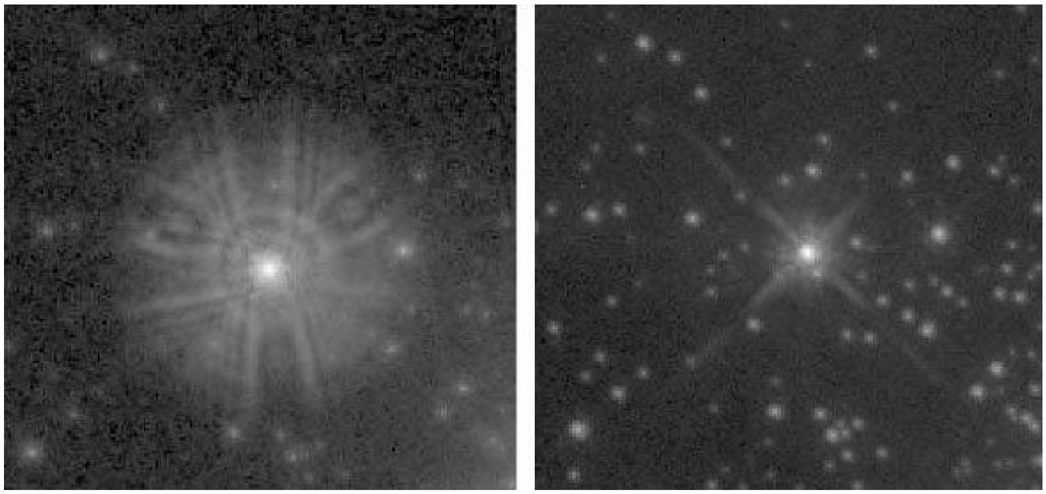
Square wave signal output follows LAME MP3 encoding. With the prevalence of poor speaker design, most listeners don't appear to notice this level of signal degradation. *Pics courtesy of* *Chipmusic.org forums*

We then had the absurd situation where codec designers claimed their compression algorithm was inaudible. However, what they had actually done was to reduce the information in the original signal down to the information capacity of the speaker. The information capacity of legacy loudspeakers is miserable, typically 10 per cent that of a CD. If the speaker is improved, the *inaudible* codec becomes *audible*.

Axis of evil

The specifications of loudspeakers are incomplete: any number of speakers having the same specification will all sound different. There is obsession with on-axis frequency response, but neglect of the equally important, possibly more important, parameters of time response and imaging.

I will explain below how human hearing requires accurate time information in sounds, yet in misguided attempts to extend the frequency range, the accuracy in the time domains may actually be damaged. Never mind the quality; feel the bandwidth.



Hubble Space Telescope shows point spread function (left) before servicing (right)

Stereophonic loudspeakers are intended to deliver a sonic image. In photography, SONAR and so on, there are agreed methods of testing image accuracy using concepts such as the point-spread function. Stated simply, an image with an infinite number of pixels would be perfect and each pixel would be a point. If each point were to be spread or smeared out by some defect, it’s the equivalent of making the pixels bigger and the sharpness of the image is lost.

Objective comparisons can be made which result in improvements. Unfortunately there is no standard for stereophonic sound imaging accuracy, no objective comparisons are possible and progress is impeded. Most legacy speakers have massive point spread functions due to diffraction from inept enclosure design and their stereophonic images are badly smeared.

This is just as well, because when the dominant sound sources are massively smeared, they will mask the fact that a compression codec has thrown away the ambience and reverberation. The mediocrity of legacy loudspeakers may be retained so that the poor quality of many compression algorithms and microphone techniques is not revealed. This also applies to earphones supplied with many portable IT based music players. Never mind the quality; look at the iconic styling.

Apple iPod Classic – a design icon but the range has never been lauded for sonic excellence

Conversely, audio codecs can be used to test and improve loudspeakers. Using a state-of-the-art speaker designed according to psychoacoustic criteria, it becomes immediately obvious how bad DAB, MiniDisc and MP3 are and that the only lossy codec that has any merit is AAC (at an adequate bit rate). It is not uncommon when demonstrating such speakers for people to assume that the signal source is some exotic high-bit-rate recording when it is simply a competently engineered CD.

To make such loudspeakers, the starting point has to be good knowledge of how the human auditory system (HAS) works, since that defines the problem. Once the problem is understood, the solution lies in the application of good engineering.

It is important to realise that the HAS evolved as a survival tool to help find food and a mate, whilst avoiding becoming a meal for something else. Given the dubious biological nature of the transducer itself, sophisticated mental processes have evolved to make the best of it.

The most important contribution hearing can provide to survival is the location of a potential threat and an estimate of its size. The HAS is very good at it, even in the presence of reflections. It does this a lot better than any modern microphone can, because microphones don't have brains.

Evolutionary Bond: our survival has depended on locating sound direction and identifying the size of the threat

Source: *Quantum of Solace*, EON Productions

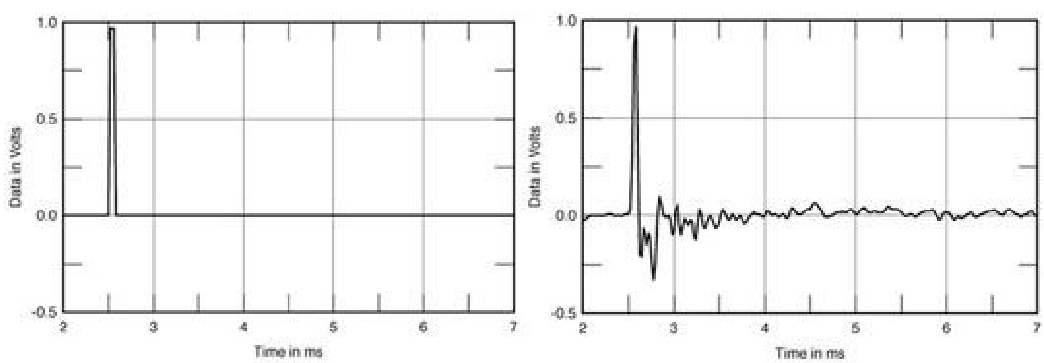
With two horizontally displaced ears, the most reliable directional information comes from the difference in time of arrival of wave fronts at the ears. The true source must be the one that results in the first version of a given sound. The HAS is working in the time domain, constantly attempting to correlate sounds from each ear to identify the first version and sounds from both ears to determine the direction. It can do this most effectively with transients, or events, since these can carry timing information.

The corollary is that a sine wave has no bandwidth and according to Shannon carries no information. This is easy to grasp. Once you have seen a few waves of a sine wave, you are not going to find anything new if the waveform continues indefinitely.

Back to square one

At the same time, the HAS is attempting to estimate the size of the sound source from the time constants. Small objects create shorter sounds than large objects. Record a cannon shot, speed the recording up by a factor of 10, and it sounds like a hand gun. Clearly if a loudspeaker has time constants of its own, it will interfere with any time analysis the HAS is attempting to perform.

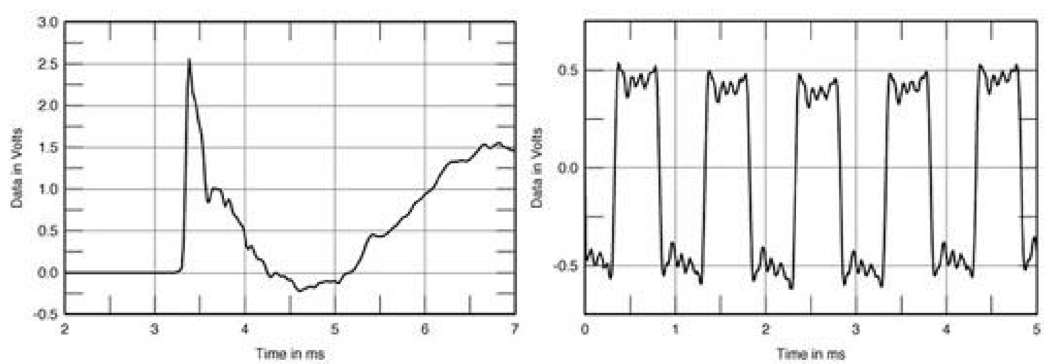
For example, in the majority of legacy loudspeakers, the acoustic source, which is the place where the sound *appears* to be generated, actually moves backwards several metres behind the speaker at low frequencies. This does not happen with real sound sources such as tympani.



Quad ESL-63 electrostatic speakers were costly but delivered impressive timing accuracy: the input pulse signal (left) is used to generate the impulse response of the speaker (right)

Only after the direction and size of a source has been determined does the HAS revert to the frequency domain to give us pitch and harmonic information. When the ear is working in the frequency domain on a sound having stationary statistics, the phase relationship between different harmonics can be changed and those changes will not be detected.

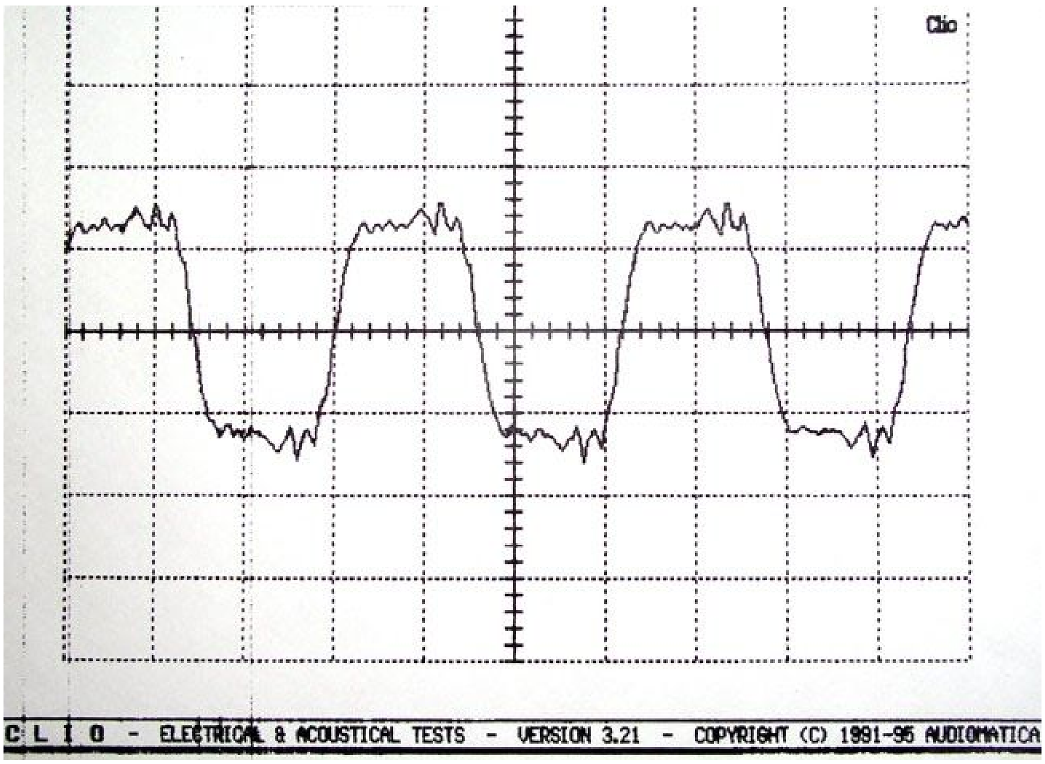
With this in mind, most speaker designers incorrectly argue that time accuracy is never necessary in a loudspeaker. They are simply not aware that time accuracy is vital when the ear is working in the time domain. Their expertise lies in making coffins for monkeys. Think what would happen to a RADAR set if the signals were not time accurate.



Quad ESL-63 step response is calculated from the impulse response above and performs well – albeit with some bass emphasis. The square wave output is a different class over conventional loudspeakers

One simple way of checking a signal path for time accuracy, or phase linearity, is to see how it responds to a square wave. A square wave only remains square if the Fourier components maintain the same time relationship. Amplifier designers routinely test with square waves to prove the quality of their designs. Loudspeaker designers never test with square waves because they maintain it's not necessary. Self-evidently one group is in denial.

The great majority of legacy loudspeakers will fail a square wave test spectacularly. Creating a time accurate speaker that will reproduce a square wave is only a matter of finding engineering solutions to the problem. The image above shows the acoustic output from a square wave input of an experimental time-accurate speaker I designed about 15 years ago. The difficulty is not in doing it but in realising it is necessary.



John Watkinson's experimental speaker design square wave test output

Since air cannot sustain a pressure change, the step response of a time accurate loudspeaker should consist of a sharply rising leading edge followed by decay back to ambient pressure. Again, most legacy loudspeakers fail this test spectacularly, displaying a step response like an empty furniture truck hitting a pothole and performing a comprehensive demolition job on the input waveform.

One of the few transducers that exhibits a good step response – and consequent realistic reproduction of percussion – is the electrostatic loudspeaker. Unfortunately, for good performance, these must be large and sited well away from walls and this is not appropriate for many domestic circumstances. Another is the moving coil device developed by Josef Manger that was specifically designed with accurate time response to meet the imaging requirements of the human auditory system.

Manger MSW transducer: designed with a very fast rise time and low linear and non-linear distortion

Since disturbed air pressure leaks away back to ambient, it should be clear that at low frequencies there is more time for this to happen. To generate low frequency sounds, a significant displacement of air is necessary, obtained by a surface having a large area

Blurred lines

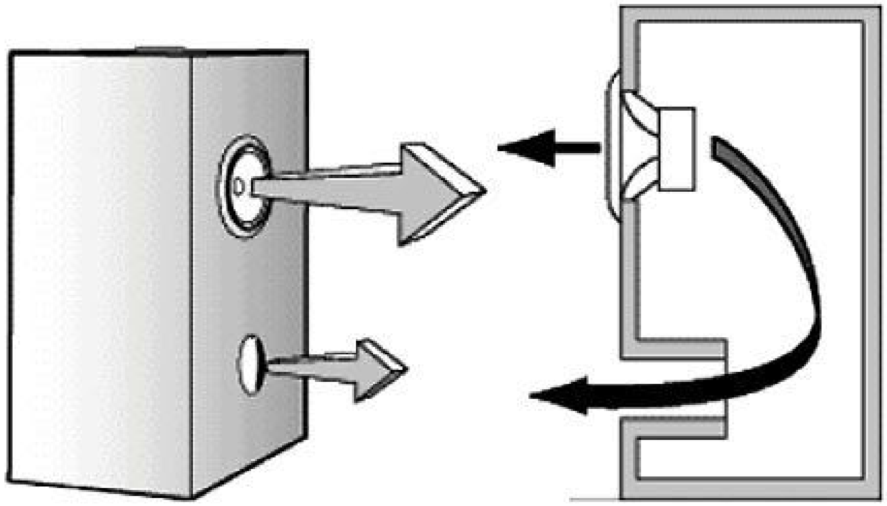
If such a surface moves in isolation, the air will simply flow around the edges from one side to the other and there will be little radiation. It is necessary to have some sort of enclosure to prevent that happening. The enclosure needs appreciable volume, otherwise the air inside will act like a stiff spring and restrict the movement.

Clearly low frequency reproduction requires physically large devices. It is simply not possible to radiate low frequencies from iPhones and tablets when the volume of the product is less than the displacement of a large woofer, which is why such devices need to be used with earphones for music.

Small speakers are never going to deliver a significant bass response

Don’t expect any leaps in the sound from the speakers in small IT devices: Moore’s Laws doesn’t apply to acoustics. On the other hand if the main purpose of the hand held device is speech communication, there is no requirement for low frequencies.

The advent of the flat screen TV has fuelled demand for equally flat loudspeakers. Whilst impossible with a legacy approach, there is no fundamental obstacle to more modern techniques and materials achieving good results. The problem is that the legacy loudspeaker industry cannot disrupt itself and the disruption has to come from outside.



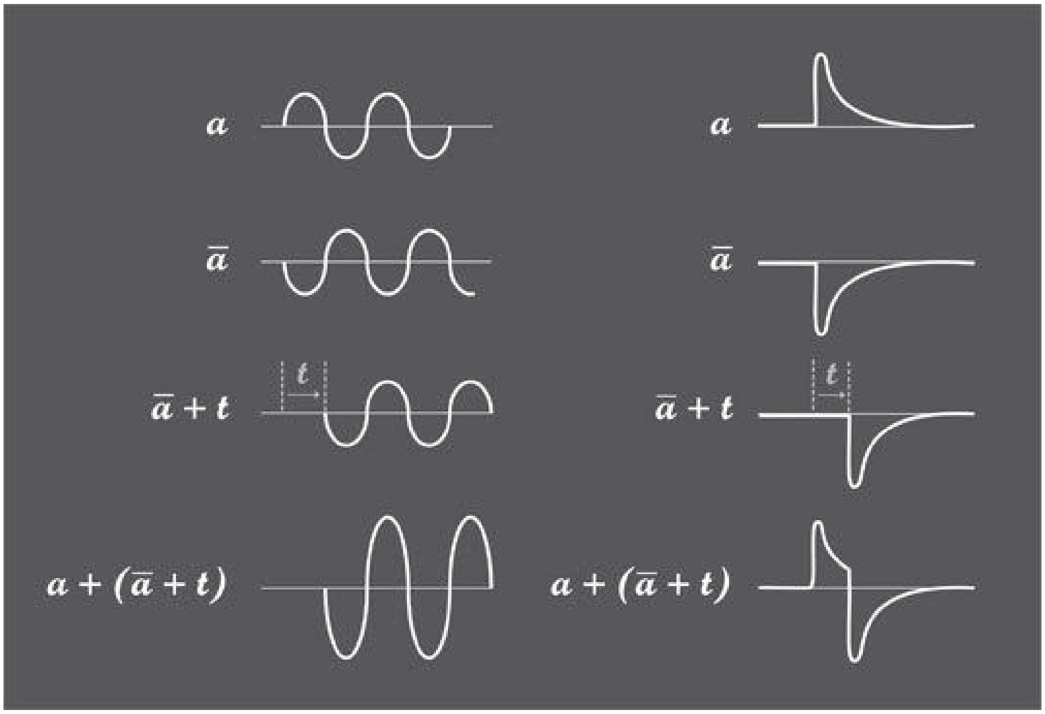
Phase-inverting, bass reflex speaker design

Back in the dark ages, magnets were made of alnico (aluminium, nickel and cobalt alloys) whose magnetic characteristics dictated long thin magnets. Small voice coils sat in the centre of cones made of flimsy paper. It made sense, then, to try to reduce the probability of the cone flexing by using resonant techniques such as bass reflex enclosures.

These employ a mass of air in a tube or port that resonates with the air in the enclosure to amplify the sound from the back of the cone. Whilst the output is increased and the low frequency response is extended, this is achieved at the expense of wrecking the speaker's time response.

PMC's OB1i – one of many takes on Transmission Line loudspeaker design

In the transmission line loudspeaker, the back wave from the woofer is delayed by guiding it through a folded pipe that causes a delay. At some frequency, the delay will be equal to half a cycle and the delayed back wave emerging from the pipe will be in phase with the radiation from the front. It is only in the case of a sine wave that a delay is indistinguishable from an inversion and we know a sine wave carries no information. In the case of a transient, the transmission line speaker destroys the waveform. The baby is thrown out and the bathwater is retained.



Transmission line speaker design debunked: On the left, a sine wave a leaves the front of the speaker. An inverted sine wave $\bar{a}$ leaves the rear. Rear wave is delayed by transmission line to become $\bar{a} + t$. When this emerges from the transmission line it is in phase with *a* and adds up. On the right, a transient is applied instead. What comes from the speaker is unrecognisable because a time delay only looks like an inversion to a symmetrical and continuous signal. Unfortunately, most of the information in audio is in the transients.

The only woofer design that is capable of being made time-accurate is the sealed enclosure. Modern drive units with stiff carbon fibre cones and large voice coils overcome break up due to internal pressures.

Woofers are always omni-directional because they are so small compared to the wavelength at which they work. But as frequency rises, a large diaphragm becomes too directional and it is necessary to switch to a smaller drive unit called a tweeter. The two drive units are supplied with the appropriate parts of the input spectrum by a set of filters known as a crossover network.

Passive crossover designs abound but will always have inherent delays

It should be an obvious requirement that if the two outputs of the crossover are added back together the result should be the input waveform. Unfortunately the majority of crossovers simply fail to meet that criterion. Passive crossovers will never be able to meet it. Active crossovers, in which the filtering is performed in analogue or digital electronic circuits at signal level, can meet the criterion but often don’t because they have simply copied the filtering of a passive crossover.

moving an appreciable distance.

Making waves with IT

One of the tenets of audiophile systems is that they are assembled from components, allegedly so that the user can "choose" the best combination. This is a complete myth, because when the amplifier designer has never met the loudspeaker designer, the use of active crossovers optimised for the speaker is precluded.

Nordost Valhalla 2 Reference Speaker Cable will set you back £10k – WTF?

The main advantage of component systems is that the dealer can sell ridiculously expensive cables, hand-knitted by Peruvian virgins and soaked in snake oil, to connect it all up. That some of these are supplied with arrows denoting the direction of signal flow defies description. Fortunately, the electrons can't see the markings and behave normally.

I think it is interesting to contrast the small IT device with considerably larger audiophile speaker systems. IT devices generally make a clean job of the bandwidth that can be realised by filtering out the frequencies that cannot be reproduced to avoid distortion.

Clearly in iPhones and tablets, the designer has complete control and so can use some of the processing power of the device to improve the sound. The foibles of the impossibly small transducers can be equalised in time and frequency. An impression of a bass response can be obtained by frequency doubling so that missing bass frequencies are reproduced as a second harmonic.

Yamaha's YSP-1400 soundbar DSP can be remotely configured for room size and listener position from an app

When the sound from a tablet has rapidly become so good *considering* the serious constraints of size, weight, power and cost it is a sad reflection on the squalid state of audiophilia that the sound of a legacy loudspeaker has made little progress for years despite those constraints being absent.

Science makes progress, pseudo-science doesn't. That leaves the door open for IT companies to take over hi-fi markets. One obvious tool IT can bring to the party is DSP-based room correction, so that the variations in response due to inevitable standing waves in the room can be compensated.

A full-range flat speaker, Earo's Wally show's what's possible in loudspeaker design, but it'll cost ya

Legacy loudspeakers are omni-directional at low frequencies, but as frequency rises, the radiation becomes more directional until at the highest frequencies the sound only emerges directly forwards. Thus to enjoy the full frequency range, the listener has to sit in the so-called sweet spot. If one moves off axis, the sound becomes increasingly deficient in treble. But it is this off-axis sound that excites the reflections in the room.

If the reflections are too different from the direct sound because of the treble deficiency, the HAS will not be able to correlate them to determine the true source of the sound and they will damage the image. As a result legacy loudspeakers with sweet spots need extensive room treatment to soak up the deficient off-axis sound. Such dead rooms are oppressive and not consistent with domestic living arrangements.

Legend prototype omni-directional speakers

In contrast, omni-directional speakers radiate accurate sound in all directions, so the HAS can easily tell the direct sound from the reflections. They do not need extensive room treatment and work well in locations from cement block store rooms to luxury yachts. They only need room correction at low frequencies.

Despite their clear advantages, they remain uncommon because when time accuracy is needed and high frequencies are to be radiated all around, internal computing and equalisation is necessary and carpenters don't know how to do that. Disruptive technology like this is not especially hard to make in quantity or at different sizes and price points, but it won't come from traditional manufacturers, just as the iPod did not. ®

John Watkinson is an international consultant on digital audio and a Fellow of the AES (Audio Engineering Society). He is the author of numerous books on audiovisual and avionics systems, regarded as industry bibles.

https://www.thebroadcastbridge.com/content/entry/7125/loudspeaker-technology-part-2-the-time-domain-and-human-hearing

Loudspeaker Technology Part 2: The Time Domain and Human Hearing

November 1st 2016 - 12:05 PM

By John Watkinson

Noted audio engineer for The Doors, Bruce Botnick, relies on three JBL M2 Master Reference Monitors in his studio.

In this second part of his loudspeaker series, John Watkinson considers the importance of the time domain to human hearing.

In Part 1, I mentioned that the recording and distribution of audio had been transformed by the application of IT. This means that audio must contain information and it follows immediately that anything audio passes through can be considered as an information channel – including hard drives, networks, loudspeakers and human hearing, all of which will be seen to have actual or effective information capacity or bit rates.

But what form does audio information take? To answer that, we have to go back to what hearing is for. In evolutionary terms, electronic entertainment and IT has happened in the last few milliseconds. Long ago, hearing was a means to survive and evolution rewarded species that evolved better means to avoid being eaten, to find food and a mate. A sense of hearing would benefit its owner in that respect.

What would be the most important information that a hearing mechanism could tell an early living being? Pretty obviously the location of a source of sound must be at the top of the list, closely followed by the size of the sound source. Is this sound a threat or does it reveal our next meal? In the absence of speech or music, the concept of establishing pitch was of limited importance; indeed the frequency domain was of little importance and means to deal with it evolved later.

Those prehistoric means to establish direction and size are still with us, hard wired into the Human Audio System (HAS) and functional at birth. The failure to consider the aspects of audio reproduction that these mechanisms interpret, results in loss of realism and listening fatigue.

Let us consider how direction is established, first in principle and then in the presence of reflections. Figure 1 shows that in the case of an off-centre sound source, the distance from the source to each ear is different. The finite speed of sound means that whatever waveform arrives at the nearer ear will arrive at the more distant ear with a predictable delay. This introduces the first hard-wired mechanism in the HAS, which is a variable delay and a correlator. Any new waveform, the onset of a sound, recognised by either ear will result in the other ear trying to find that waveform at a later time. The time difference tells us the direction.

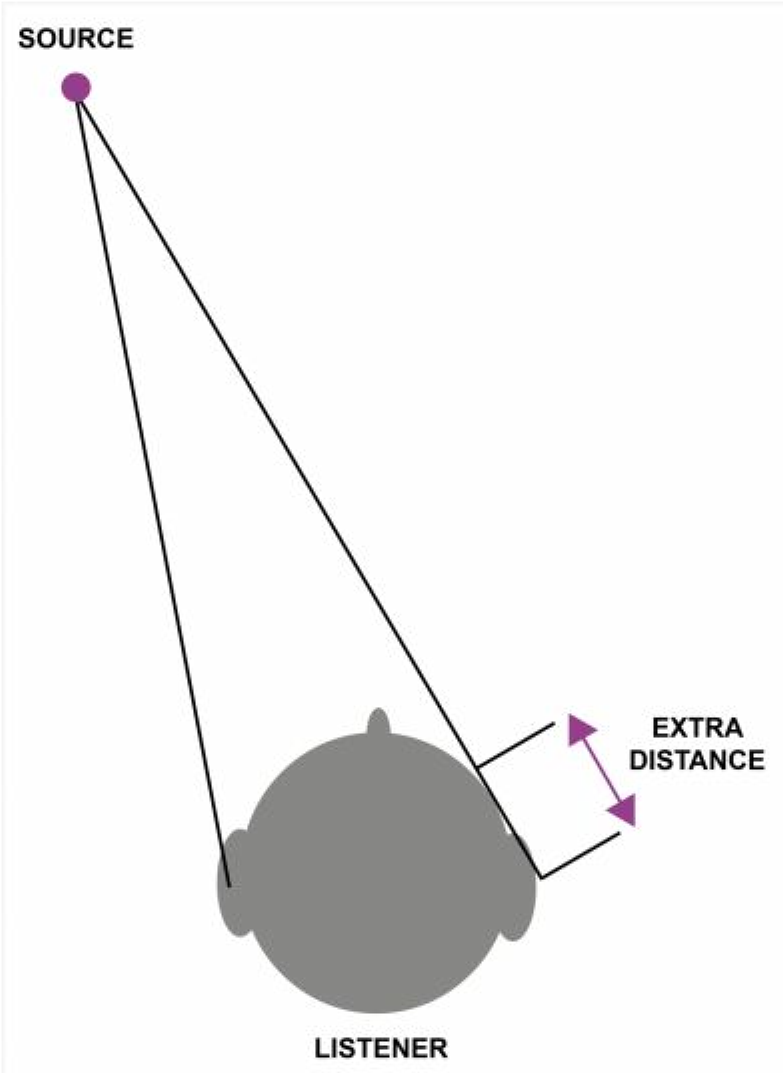


Figure 1. When a sound source is off to one side of the listener, the sound arrives at the two ears at different times. The HAS can identify the same transient at both ears and measure the time difference.

It ought to be clear that unambiguous measurement of the time shift between two waveforms is only possible if the waveform is transient. Trying to do it with a pure tone or sine wave suffers two difficulties. Firstly all cycles look the same so correlation can be found at a number of time shifts. Secondly in the real world pure tones jump to the nearest standing wave or eigentone in the room so the location of the source is concealed. This is hardly an issue in the real world where the majority of sounds such as footfalls, doors closing, objects falling, are transient.

This explains why wailing sirens on emergency vehicles are not very smart and why the vehicle often can't be located until the flashing lights are visible. The use of blue lights is equally dumb as human vision is least sensitive to blue. Given that sub-optimal applications are the rule rather than the exception in acoustics, which has always been a Cinderella subject, we should not be surprised to find mediocrity in a lot of legacy loudspeakers. Equally it's not appropriate for me to complain if I can't advance solutions.

It is interesting to consider the problem from a communications theory standpoint. Sine waves are pure tones and so have no bandwidth. Thus their information capacity is zero. The bandwidth of a transient is large. It follows that most all of the information in audio is carried by transients and that failing to consider the time domain accuracy of a loudspeaker may seriously compromise its information capacity. This is one of the reasons we hear speakers that all have the same frequency response yet all sound different. We find, for example, speakers that are good on violins but lousy on percussion.

Getting back to the plot, not only can the HAS measure the delay between the versions of a sound at the two ears, but it can also insert that delay prior to adding the sound from the two ears, so that sound from that direction is emphasised and sounds from other directions are diminished. This is known as attentional selectivity, or in familiar terms, the cocktail party effect, which allows the listener to pay attention to one source in preference to others. The two ears have been made into a simple phased array.

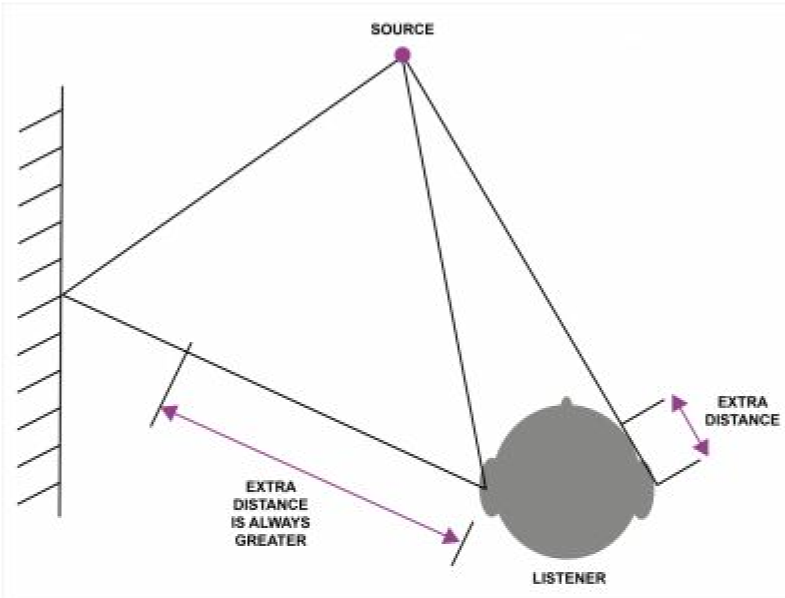


Figure 2. Reflections don't reduce the ability to determine direction, because they arrive too late. Instead the reflections and the direct sound are time shifted to make the source more audible.

The HAS deals with reflections using an extension of the same mechanism. Figure 2 shows that reflections must have travelled by a longer path and must have suffered a delay that is greater than that due to ear separation. Thus after the onset of a sound and after the inter-aural delay has been detected, the correlator continues to run and looks for further versions of the sound, which will indicate the presence of reflections.

The reflections are recognised, and the delay is used to give us a sense of the distance from the reflecting surface, or in an enclosed space, an idea of the size of that space. The fact that reflections are recognised for what they are means that they do not diminish the accuracy of the initial location of the source via the direct sound. Now here comes the clever part. Provided the reflections are not too late, the HAS time-aligns all the reflections with the original sound and adds them, so that the original sound can be heard better in a reverberant environment. This is known as the Haas effect. Compare that with a microphone which has no such mechanism and where reflections make things worse.

This is why amateur sound recordings are invariably terrible because it is not understood that the microphone doesn’t hear as living things do. The poor microphone doesn’t have a brain and, instead of thinking for it, the amateur recordist seeks to emulate it.

Here we find one of the great contradictions of audio. Ask an acoustician whether it is the reverberant sound or the direct sound that conveys the most power to the listener in an auditorium and he will correctly say it’s the reverberant sound. Take the reverberation out of a concert hall and the audience will condemn it.

For critical listening, why are legacy loudspeakers traditionally played in acoustically treated, practically dead spaces? The brief answer is that legacy speakers cannot reproduce the time domain correctly and cannot excite reverberation correctly, so the Haas effect cannot work and the reflected sounds become a distraction that has to be absorbed. It doesn’t have to be like that. A fuller answer will emerge as this series progresses.

Real sources of sound are frequently physical objects that have been set into motion. If this is someone stepping on a twig which then breaks, the vibration and the sound will be transient. Imagine some surface suddenly moving forward in a step-like manner. A sound transient having increased pressure will be radiated. However, the atmosphere cannot sustain local pressure differences, so the over-pressure leaks away. The speed with which it leaks away is the time constant of the transient sound. The waveform of a hand gun firing has the same shape as the waveform of a Howitzer firing, except that the latter has a much longer time constant. So if you need a big gun sound effect, just record a pistol shot and slow it down.

The time constant is a function of the size of the radiating object. Larger objects block the path by which pressure equalises to a greater extent and cause longer time constants. In real life, the HAS can measure the time constant of a transient and estimate the size of the source from it. Most legacy loudspeaker designs do not allow this mechanism to operate because they superimpose fixed time constants of their own which come down on the sound like a Pythonesque foot.

What we are concluding is that for realism, audio waveforms need to have their phase linearity preserved. Readers familiar with television technology know that this is paramount for video waveforms and won’t be surprised at all.

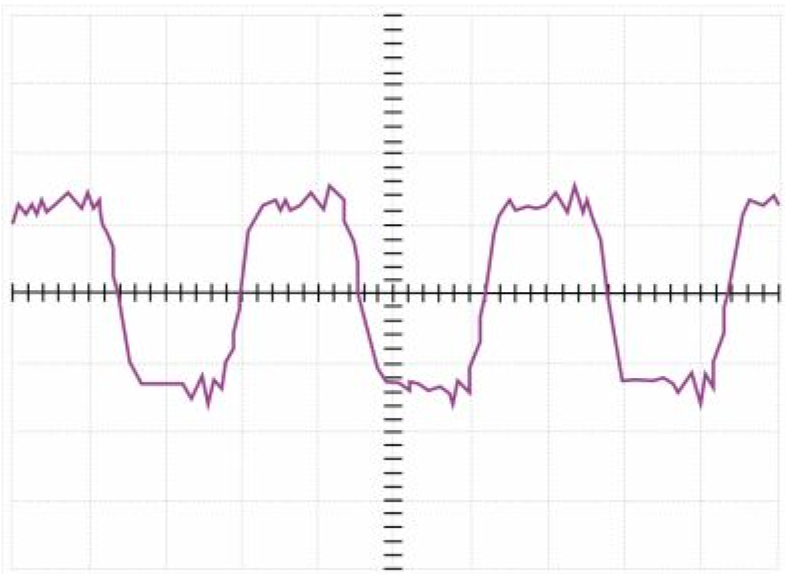


Figure 3. The square-wave response of a loudspeaker designed to meet the time-accuracy criteria of human hearing.

Audio amplifier designers test their products using square waves, or strictly speaking, band limited square waves and often publish the results. According to Fourier, a square wave consists of a series of harmonics which are closely defined in amplitude and phase. A system that can reproduce a square wave at its output not only has a flat frequency response, but is also phase linear and capable of carrying the time domain information the HAS requires.

Very few loudspeaker manufacturers publish the square wave response of their products, usually because the output waveform is unrecognisable. However, just to illustrate that it is possible, Figure 3 shows the square wave response of a speaker I designed about 15 years ago.

In Part 3 of this series, we will begin by looking at how the frequency domain comes into play in human hearing.

https://www.thebroadcastbridge.com/content/entry/7401/loudspeaker-technology-part-3-the-frequency-domain-and-human-hearing

Loudspeaker Technology Part 3: The Frequency Domain and Human Hearing

December 13th 2016 - 12:05 PM

By John Watkinson

The Bang & Olufsen BeoLab 90 speaker celebrates the company's 90th anniversary. Retail price for these innovative speakers, about $90,000.

In Part 2, of John Watkinson's series of articles on loudspeakers, the critical time-domain operation of human hearing was considered. In Part 3, he explains how the frequency domain interacts with the time domain and why they are a crucial concern in any accurate loudspeaker design.

The bit rate of a CD is about 1.5Mbs. The human nervous system simply isn't capable of that sort of data rate, or anything like it, so one of the jobs of the physical ear is to perform some prior analysis of incoming sounds before nerve impulses are created. The Human Visual System must do something similar with images, of course. That topic was considered in my article, "How we see."

The basic transduction method of the human ear is that tiny hairs are deflected by the flow of fluid and the deflection is sensed by nerves. To increase sensitivity, some of the hairs are active: they amplify the fluid movement by moving in sympathy.

A transducer filled with fluid is not an obvious solution for a land-dwelling being, and may indicate that life began in water. A technical problem for a fluid-filled transducer in air is the mechanical impedance mismatch between sound travelling in air and sound travelling in fluid. Most of the sound energy in the air would simply reflect from that mismatch.

Instead, sound arrives at, and vibrates, the eardrum, whose motion is geared down by a system of tiny bones, or ossicles, that act as an impedance convertor. Small forces and high velocities at the eardrum are converted to higher forces and lower velocities at the output, which is a piston-like bone that excites the fluid-filled mechanism of the inner ear.

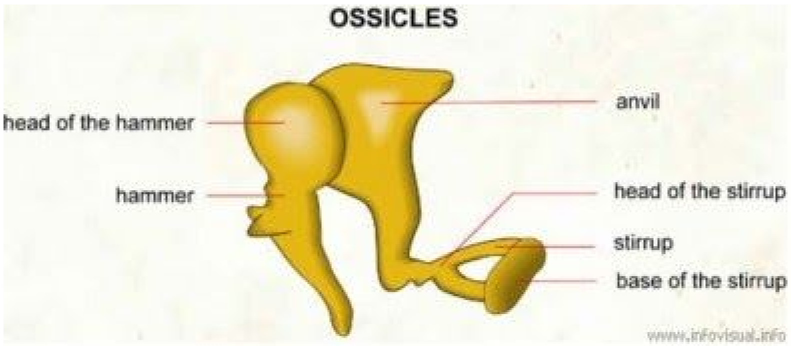


Figure 1. The ossicles are three bones in either middle ear that are among the smallest bones in the human body. They serve to transmit sounds from the air to the fluid-filled labyrinth.

If someone were to propose a microphone design working on that basis, everyone would die laughing because of the obvious shortcomings. The truth of the matter is that in some respects the HAS is not very good. The story put about by hi-fi enthusiasts that the ear is some miraculous device that can hear problems that no instrumentation can detect is a huge joke. It does, however, justify the sale of products (generically known as snake oil) that claim to produce an improvement that no instrumentation can detect. The improvement to the vendor’s bank balance is beyond dispute.

The inner ear is a small tube hollowed out of the skull having a flexible diaphragm dividing it lengthways. This is known as the basilar membrane. The membrane and the surrounding fluid together create a mechanism that can respond to transient and stationary sounds. (Here, stationary is used in the statistical sense that the spectrum is not time variant).

The membrane and attached fluid has mass and associated compliance and damping. It is capable of both transmission line behavior and resonant behavior, but at different times.

A transient sound will be supplied as a time-domain fluid pressure waveform to the outer end of the Basilar membrane. As the disturbance travels along the membrane at finite speed, nerve cells trigger in different places at different times. Thus a very sharp transient, having maximal bandwidth, can be handled by nerves having a low firing rate because the transmission line spreads the event out in time. When the HAS seeks to correlate two transient waveforms for location purposes or to identify a reflection, what it is actually doing is looking for a pair of similar patterns of nerve firings, which is a lot easier for a low-speed biological process.

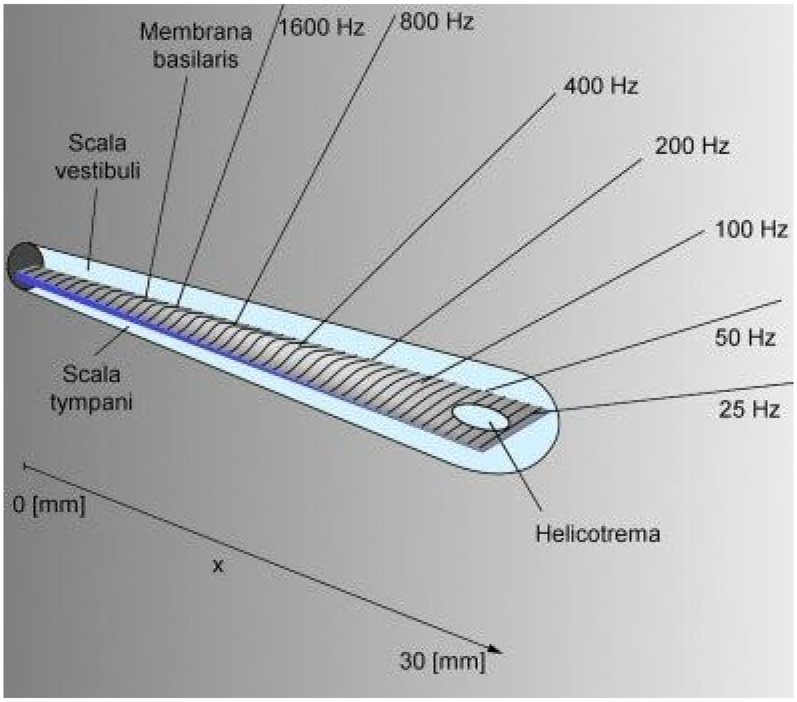


Figure 2. Human ear basilar membrane. Image: Kern A, Heid C, Steeb W-H, Stoop N, Stoop R, Biophysical Parameters Modification Could Overcome Essential Hearing Gaps.

Only after the sound source has been located, and its size estimated, will the HAS transfer over to operate the more evolutionarily recent frequency domain analysis mechanism.

The basilar membrane is far from uniform. Near the middle ear it is light and stiff, further away it becomes gradually heavier and looser, so that it has a range of resonant frequencies along its length, from 20kHz near the middle ear (in the young) to 20Hz at the pointed end.

There is simply no evidence of any adult HAS response to sounds above 20kHz, and clearly engineering audio systems with a response much above that makes no sense. On the other hand there is no law against people buying hopelessly over-specified products on the basis of unsubstantiated beliefs.

When acting as a frequency analyzer, the basilar membrane only provides amplitude information for each frequency it detects. There is a well known demonstration in which some stationary waveform is synthesised and whilst listening the phase relationship between the frequency components is varied (this is linear distortion: it changes the waveform but not the harmonic content) and no-one listening is any the wiser.

From this test many people conclude that the ear is phase-deaf at all times and that the time response of loudspeakers doesn't matter. That conclusion is *totally erroneous*. Whilst the ear may be phase-deaf on stationary sounds like tones, as we have seen these convey little information. More importantly, when the ear is working in the time domain it is highly sensitive to linear distortion and if this is too great it will impair the ability of the HAS to process time-domain information.

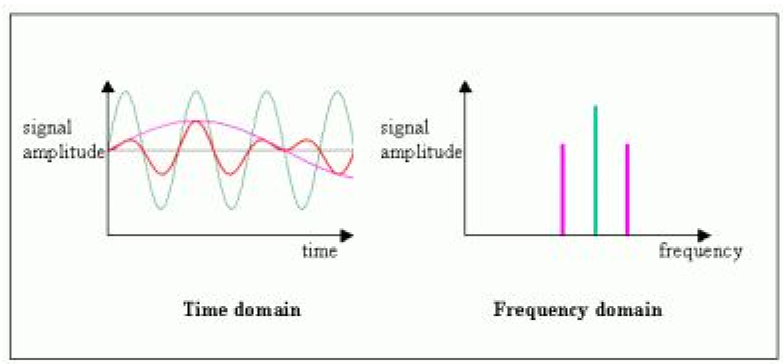


Figure 3. Relationship between time and frequency domains.

It is easy to see that if a loudspeaker has time constants of its own, that will impair the ability of the HAS to estimate the size of sound sources using time constants in the audio signal. This may also explain why certain designs of loudspeaker appear to work better (or at least show fewer deficiencies) on certain types of music. If one considers musical genres in which all of the instruments are electric or electronic, the signals concerned will contain no information about the size of an acoustic sound source because there is no such source. It follows that a loudspeaker that superimposes time constants of its own will do no damage to such recordings.

There is no shortage of speakers that sound great on rock music yet are incapable of reproducing female speech with any realism. The unfortunate lady sounds like she is inside a tea chest. Smaller speakers are considered better for speech.

The corollary, of course, is that an accurate loudspeaker that does not superimpose its own views on what the sound waveform should look like can be used for all types of sound. Equally, all accurate loudspeakers sound surprisingly similar.

Required speaker performance

It may be that we have come far enough through the working of the human ear to attempt some sort of a specification for a realistic or accurate loudspeaker. An adequate frequency response is obvious, as is freedom from harmonic distortion on stationary signals, so I won't dwell on that. However, if we believe all that stuff about how the HAS works in the time domain, and we should, it immediately follows that linear distortion is not acceptable in a loudspeaker. In other words all frequencies should take the same time to pass through the speaker, such that the input waveform is preserved.

One of our criteria has to be that the loudspeaker must be able to reproduce a (band-limited) square wave, because that is the simplest test we have for linear distortion. However painful it may be to break with tradition, that is a fundamental requirement and anything that prevents it has to be abandoned and an alternative found.

Because all loudspeakers radiate into a more or less reverberant environment, it is vital that they should radiate more like real sound sources do. This means that it is no longer acceptable that a loudspeaker only meets some performance criterion on axis whilst ignoring what happens off axis. It is perhaps pertinent to ask why loudspeakers are deemed to have an axis when people, instruments and natural sound sources don't.

Perhaps the concept of an axis is undesirable in loudspeakers. This implies that the sound quality radiated in any direction should be as good as in any other direction.

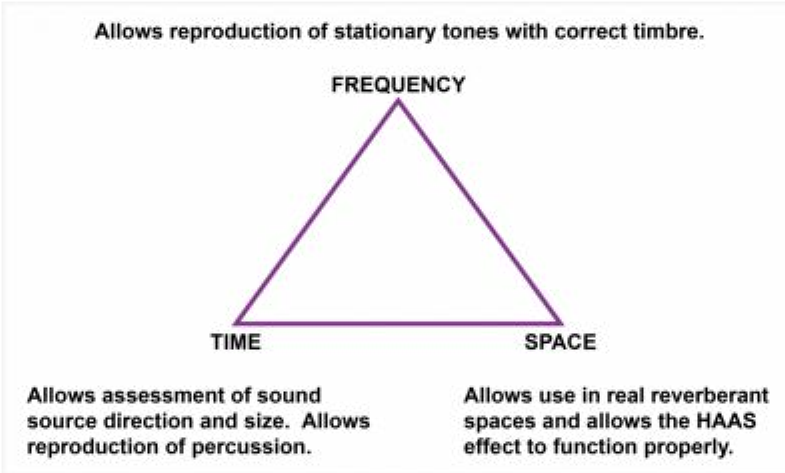


Figure 4. The three important domains in which a realistic loudspeaker must meet performance criteria; Time, Space and Frequency. Neglect of any one will nullify excellence in the other two. Legacy speakers concentrate on the frequency domain and so they always sound like loudspeakers.

Figure 4 shows the three key domains in which a loudspeaker must meet performance criteria. These are time, space and frequency. The time domain sets criteria for linear distortion, the space domain sets criteria for directivity and the frequency domain needs no comment. Typically, legacy speakers address the frequency domain only and the fact that such speakers all sound different is because they fail to address the other two domains in different degrees.

In summary, a flat frequency response is needed so that the timbre or tonality of the original sound is unimpaired. Lack of linear distortion allows transient sounds such as percussion properly to be reproduced and allows the ear to determine the size of sources. Good directivity means that the quality of reverberant sound is sufficiently close to the direct sound that the ear can recognise reflections for what they are and the Haas effect can operate.

https://www.thebroadcastbridge.com/content/entry/7477/loudspeaker-technology-part-4-the-frequency-domain-and-human-hearing

Loudspeaker Technology Part 4: The Frequency Domain and Human Hearing

December 27th 2016 - 09:00 AM

By John Watkinson

EV speaker replacement diaphragm.

In Part 3 of this series on speaker technology, we saw that accurate loudspeakers need to consider the time, space and frequency domains. Now it is useful to consider what that means in terms of arriving at some kind of specification for a real loudspeaker.

The development of digital audio was a tremendous boon for audio quality because by recording data using error correction, the sound quality became independent of the medium. Audio could be stored on tapes, hard drives, RAM and optical discs or transmitted down wires, radio links or optical fibres without any loss of quality beyond that due to the initial conversion. Basically digital audio is time accurate because the sound waveform can be preserved. Microphones can, and often are, made time accurate and audio power amplifiers traditionally have been tested with square waves to prove it.

This led to the bizarre situation in which sound waveforms could be captured, stored and amplified very accurately and delivered to a legacy loudspeaker that would destroy them. There seemed to be a schism between manufacturers of microphones, recorders, desks and amplifiers who regarded phase linearity as important and loudspeaker manufacturers who said it didn’t matter or it couldn’t be done. Clearly both points of view could not be correct.

The physics do matter

One of the ways in which one knows one has become a cynic is when the realisation dawns that someone who claims something is unnecessary or impossible is doing so simply to avoid having to admit that they don't now how to do it. Since that realisation, I only accept impossibility when the laws of physics need to be violated.

There is thus a marvellous symmetry whereby manufacturers, who hold to be impossible things that physics does permit, sell products to hi-fi enthusiasts, who hold views that physics does not permit.

All of the evidence suggests it is those who hold *time accuracy* to be important that are correct. A modern understanding of human hearing suggests that it is theoretically important and the dramatic increase in realism that is obvious to any unbiased listener when the original sound waveform accurately traverses the entire reproduction chain confirms that it is practically important.

Photo 1. There appears to be no shortage of creativity in the field, hence the wide range of cabinetry designs offered in today's loudspeakers. The above system is called Pnoe and is produced by Arcadian Audio. Cost $25,000.

Those who are familiar with digital imaging know that the smaller the pixels are the sharper the image becomes. In audio the equivalent is that the smaller the acoustic source is, the sharper the image. Acoustic source has an idiomatic meaning in the context of loudspeakers: it is the place from which the sound produced by a loudspeaker appears to come, in three dimensions. Ideally the acoustic source should be a fixed and vanishingly small point. In most legacy designs it is neither, for reasons which we shall explore.

There is another aspect of the spatial domain that is important. This is that the frequency response should be the same in all the directions in which sound is radiated. That is the same as saying the *directivity pattern* is independent of frequency. If this is not the case, the speaker may fail to excite reverberation that can be identified as such by the ear because it will be coloured.

In some respects meeting an advanced specification such as that outlined here is difficult. But in other respects it is easier, because if all of the technologies, architectures and components that cannot meet the specification are discounted, the final choice must be made from a smaller list.

Speaker diaphragm

Out of many ideas that have been tried, the most successful way of reproducing sound is the moving diaphragm. As was explained in an earlier article, air cannot sustain a local pressure difference. It leaks away and the more time there is available and the smaller the source, the more powerful the leakage. Thus a naked diaphragm, what we call a dipole, of moderate size oscillating at low frequency radiates next to no sound because the pressure increase on one side and the reduction on the other side are cancelled by air moving around the edge.

To reproduce the lowest audible frequencies, the dipole has to be tens of feet across and this is not feasible. Thus all practical reproduction of the lowest audible frequencies requires some baffle or enclosure that prevents the radiation from the two sides of the diaphragm from cancelling. That forms a subject in itself.

For a given diaphragm area, the tendency for air pressure to leak away means that in order to obtain a flat frequency response, the amplitude of motion of the diaphragm must rise as frequency falls, at 12dB per octave to be precise.

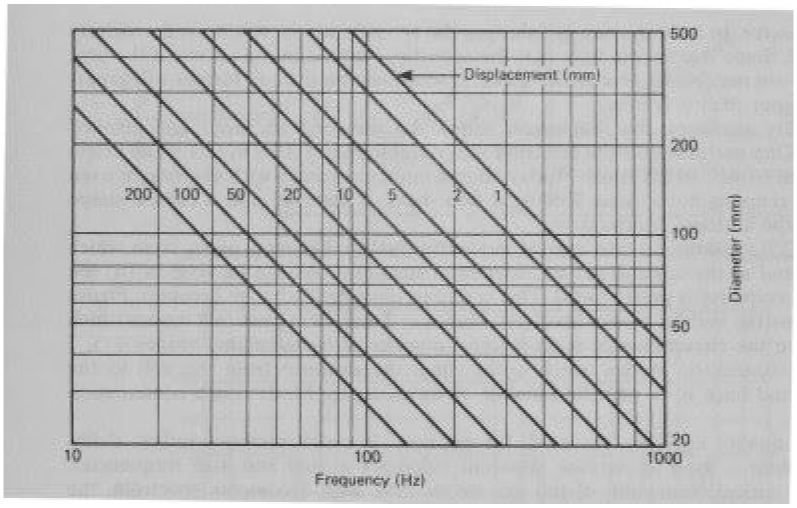


Figure 1 shows the peak displacement needed by a diaphragm of a given diameter to radiate 1 Watt at a given frequency. Note how the amplitude rises as frequency falls.

A good speaker needs a large diaphragm as this chart illustrates. As the frequency falls, the speaker needs either a large diaphragm or large radiator displacement.

Figure 1 illustrates the link between frequency, displacement and amplitude for constant power. It will be evident that as frequency falls, the designer is pushed towards large diaphragm area or large displacement. It should immediately be obvious why we must not expect much low frequency sound from iPhones or tablets. It is also clear why we can see woofers moving, but not tweeters.

Doubling the diameter of the diaphragm quadruples the area, and so the displacement can be reduced by a factor of four. This trade-off gives rise to the concept of *volume velocity*, which is the product of the diaphragm area and the velocity. All combinations that have the same volume velocity radiate the same power. Volume velocity is a misnomer, because it is not a vector quantity.

Another difficulty is that the sensitivity of the HAS (Human Auditory System) to low frequencies is not very good, so there is no point in having an extended low frequency response if sufficient level cannot be created.

Self-evidently, for good ability to radiate power, a large diaphragm is a good thing.

As noted, the tendency of air to leak away is frequency dependent, so any diaphragm moving with constant velocity would display a rising frequency response which would be no good for sound reproduction. More precisely, the level radiated would be proportional to frequency, or rise at 6dB per octave, and thus the power radiated would be proportional to the square of the frequency, or rise at 12db per octave.

Early speaker design

The solution to this problem was one of the most seminal discoveries in the history of audio which set out the principle on which the great majority of loudspeakers work to this day. In the 1920s Edward Kellogg and Chester Rice were working at General Electric on the problem of getting enough sound level from radio receivers. They understood the physics of moving masses.

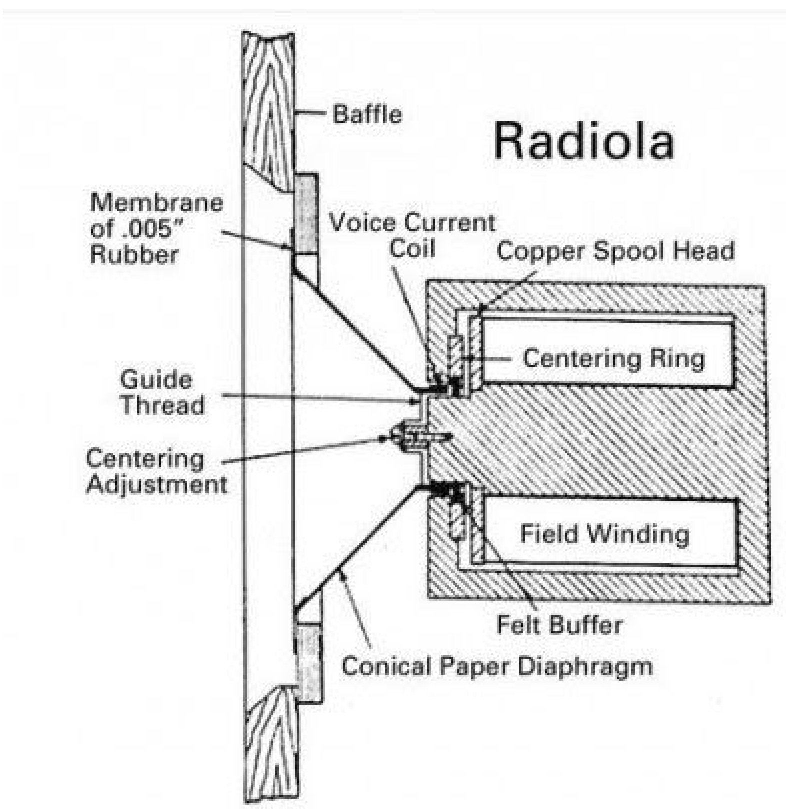


Figure 2. Basic Rice-Kellogg moving coil speaker.

It is sufficient to describe the position of a diaphragm with respect to time. The rate of change of position (the first derivative) is the velocity and the rate of change of velocity (the second derivative) is the acceleration. Position, velocity and acceleration with respect to frequency are linked by 6db per octave functions. What Rice and Kellogg did was to realise that if the *acceleration* of their diaphragm was held constant with frequency, the velocity would fall at 6db per octave which would cancel out the rising radiation efficiency and result in a flat overall frequency response.

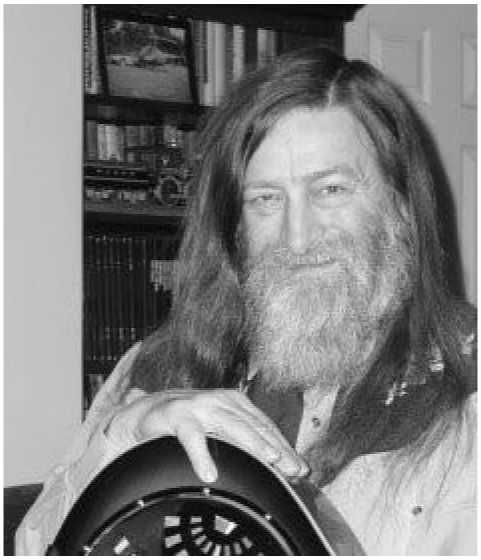
It follows from Newton's Laws that the acceleration of masses, diaphragms included, is proportional to the applied force. What Rice and Kellogg needed was a motor that applied a force proportional to the audio input waveform. They found that solution in the *moving coil motor*, shown in Figure 2, which is a subject into itself.

Interestingly the Rice-Kellogg loudspeaker launched in 1926 as the Radiola model 104 was also an active loudspeaker as it necessarily contained an audio amplifier powerful enough to drive the moving coil.

The significance of the Rice-Kellogg speaker is more than academic. By making it possible to reproduce sound at realistic levels, they essentially created the audio industry.

With the exception of inhospitable locations, the speed of sound where anyone would want to live is about 340 metres per second: the equivalent of about one foot per millisecond, which is easier to remember, or a mile in five seconds. In contrast the speed of light *in vacuo* is about a million times faster, or one foot per nanosecond.

The lowest frequency anyone can hear is about 20Hz, and oddly, hi-fi enthusiasts don't challenge it. 20Hz corresponds to a wavelength of about 17 metres or about 56 feet. The highest frequency a young person can hear is about 20kHz, although some hi-fi enthusiasts believe, and it is a belief, that higher frequencies can be heard. 20kHz corresponds to a wavelength of about 17mm or about three quarters of an inch.

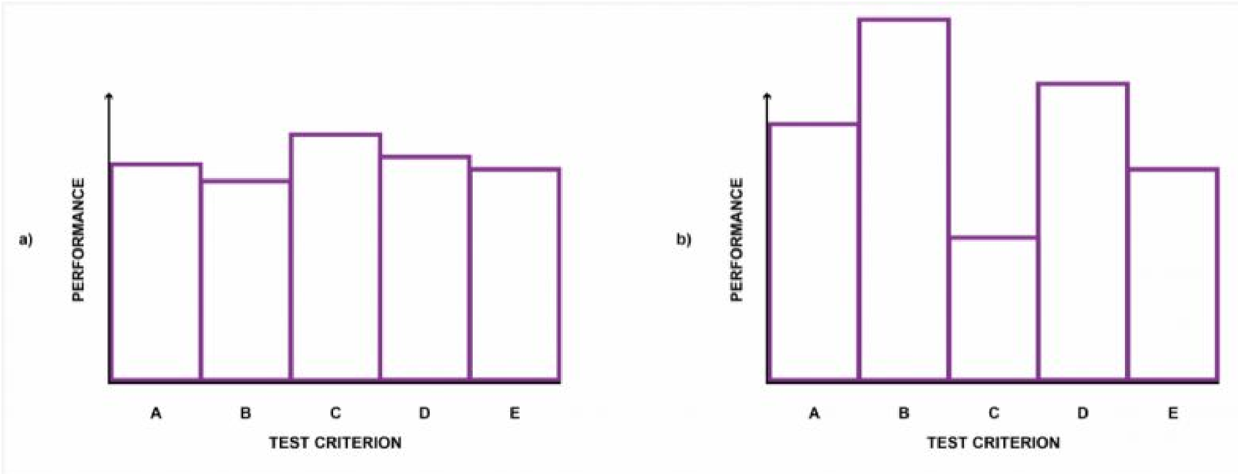


John Watkinson, Consultant, publisher, London (UK)

This extraordinary range of wavelength, spanning some ten octaves, ranges from wavelengths that are considerably larger than most everyday objects to wavelengths that are considerably smaller. We must expect some change in behaviour because of that.

In contrast, visible light exists over a range of less than an octave and the wavelengths are always significantly smaller than everyday objects, making visible light behave much more consistently than sound. For example it is possible to obtain deep shadows when light encounters an obstacle. That simply doesn’t happen with sound.

The ten-octave range of wavelength fundamentally affects loudspeaker design and it will be necessary to consider wavelength related effects to see why. We will do that in the next instalment.

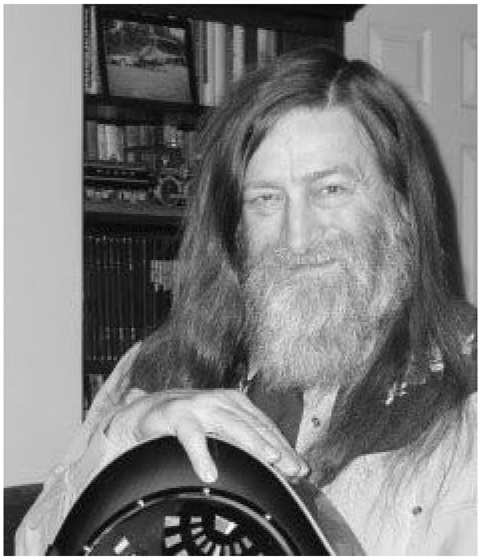


In Figure 5, the a) graph shows measured speaker performance across a set of tests. The results are uniform. In graph b), test C results are lower than other results. This means that any manufacturing costs expended in obtaining the high values in tests B and D are squandered.

In any product design, it is the weaknesses that irritate the user and cause resentment. Thus for any product cost, the best performance/price ratio will be where the product performs equally well across the range of tests. In other words do nothing badly. This can be seen in Figure 5, graph a). In Figure 5, graph b) the product performance is dragged down by results C, which means the money spent on obtaining the high B and D performance is wasted.

No realism will be obtained until all three domains are addressed. The first serious attempt at addressing the time and directivity domains was the Quad ELS-63 designed over half a century ago. It still gives a good account of itself.

The best value will be obtained when failings in the three domains are balanced. Figure 5, The graph a) portion illustrates the hallmark of good industrial design, where all relevant factors perform about the same so the product is not let down by a weakness. This is almost anathema to the world of hi-fi, in which vast sums are spent on minutiae and glaring deficiencies are totally neglected.



John Watkinson Consultant, publisher, London (UK)

Editor’s note;