

John Watkinson July 2014

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.

Article first published : www.theregister.co.uk/2014/07/02/feature_the_future_loudspeaker_design/?page=1

"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.



Loudspeaker perfection? Manger's MSMc1 is a step in the right direction

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.

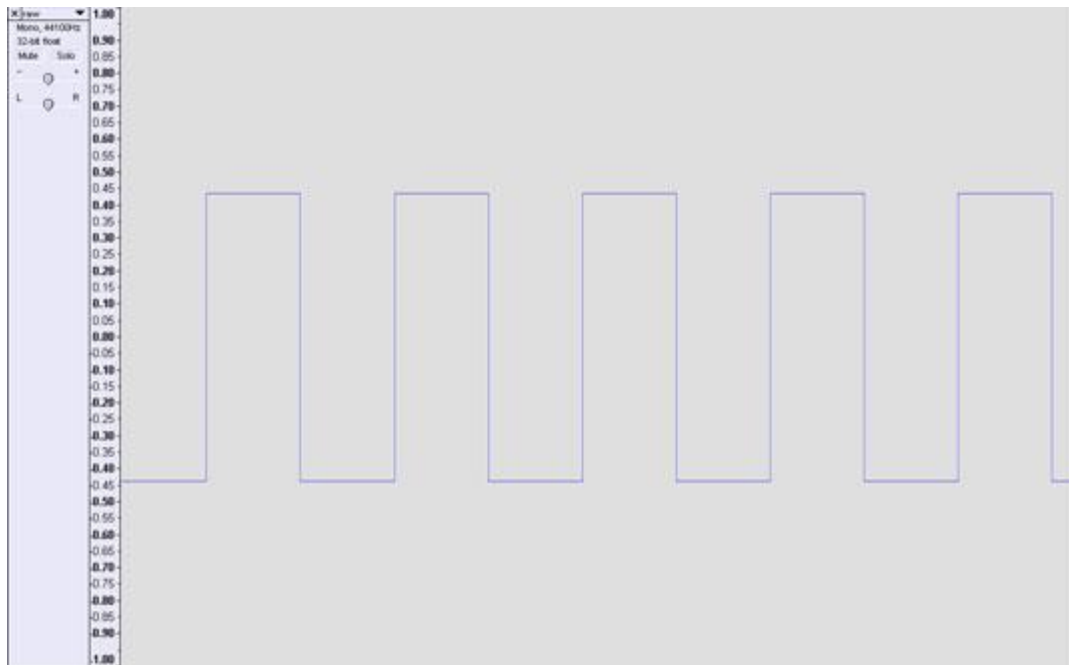
It would seem logical that if the shortcomings in real loudspeakers could be made a little less than the shortcomings of our hearing, we would believe them to be perfect. So there is, in principle, no technical reason why a perfect-sounding loudspeaker shouldn't be made, even if it won't be hand-held. 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

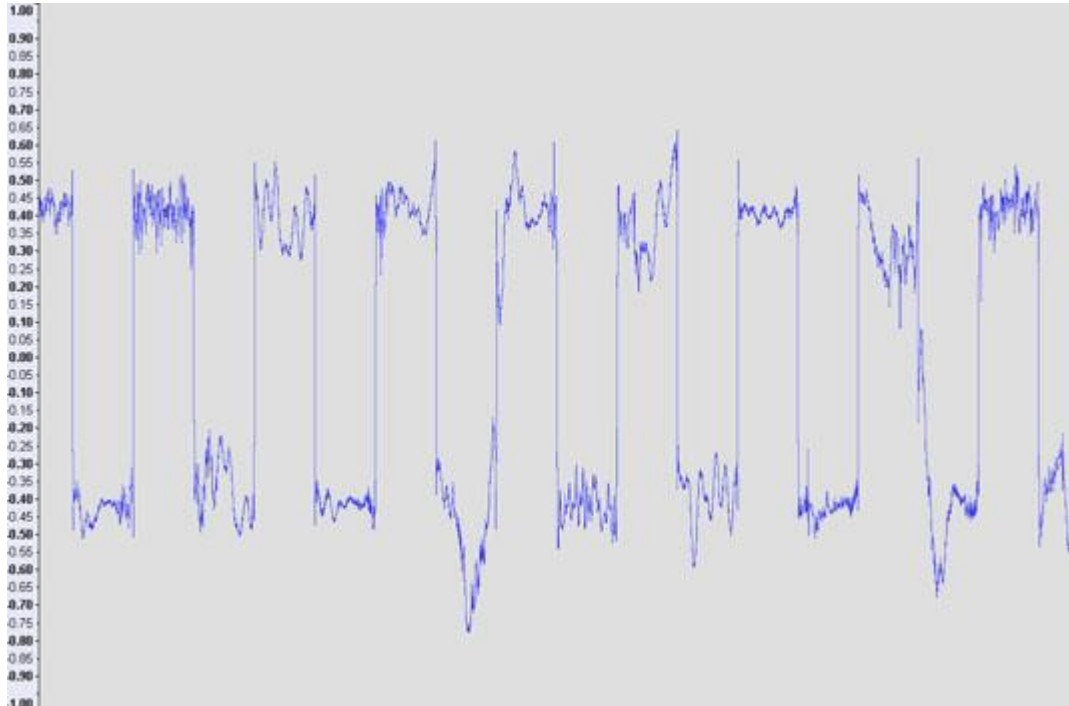
Many of the reasons are psychological. Although humans are equipped with a remarkable range of senses, they appear to be under-utilised most of the time. Cartoons, caricatures and souvenirs are all severely information-limited versions of the original sensation, yet they appear to elicit much the same 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 beancounters 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.



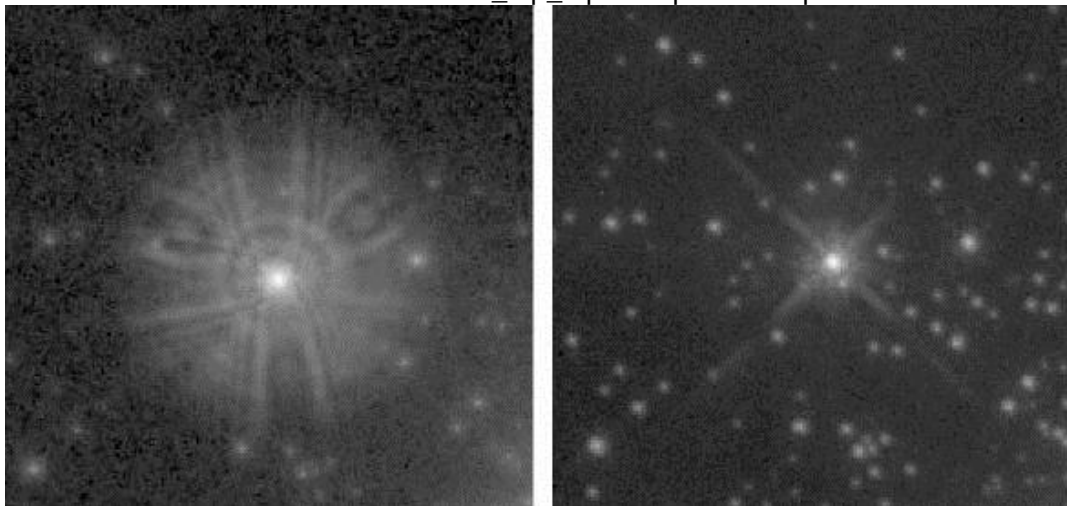
Square wave signal output follows LAME MP3 encoding. With the prevalence of poor speaker design, most listeners don't appear to notice this level of signal degradation. *Pics courtesy of Chipmusic.org forums*

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

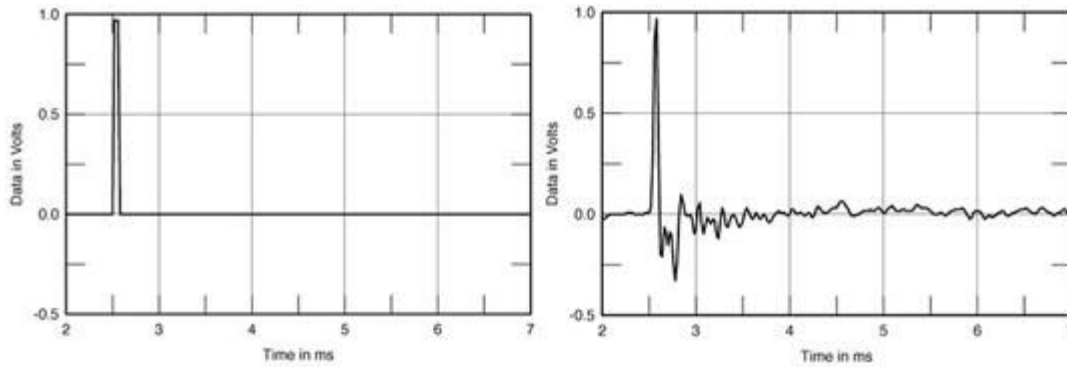
With two horizontally displaced ears, the most reliable directional information comes from the difference in time of arrival of wave fronts at the ears. The true source must be the one that results in the first version of a given sound. The HAS is working in the time domain, constantly attempting to 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.

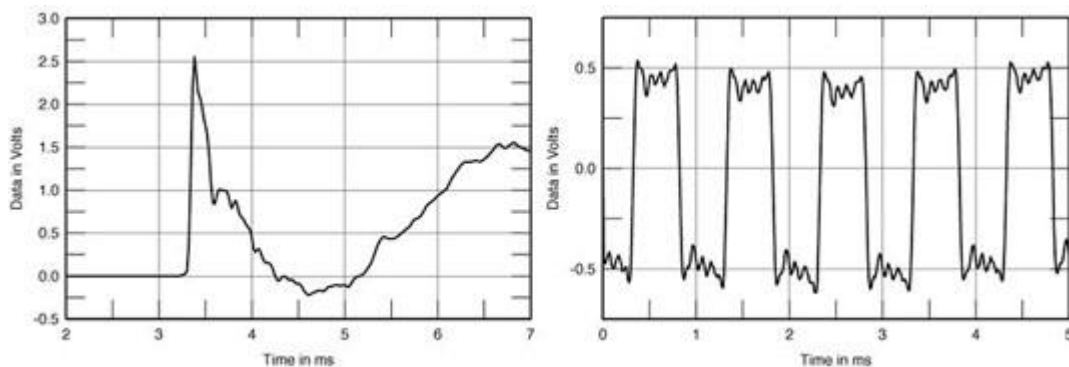
For example, in the majority of legacy loudspeakers, the acoustic source, which is the place where the sound *appears* to be generated, actually moves backwards several metres behind the speaker at low frequencies. This does not happen with real sound sources such as tympani.



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.

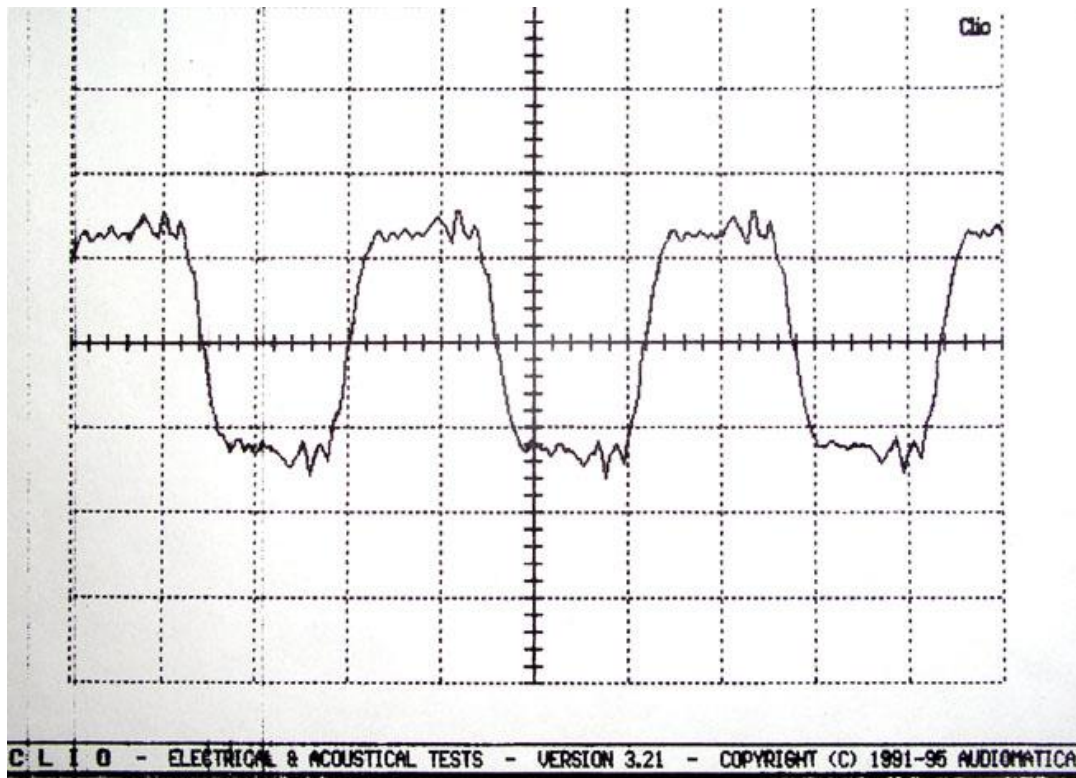
With this in mind, most speaker designers incorrectly argue that time accuracy is never necessary in a loudspeaker. They are simply not aware that time accuracy is vital when the ear is working in the time domain. Their expertise lies in making coffins for monkeys. Think what would happen to a RADAR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

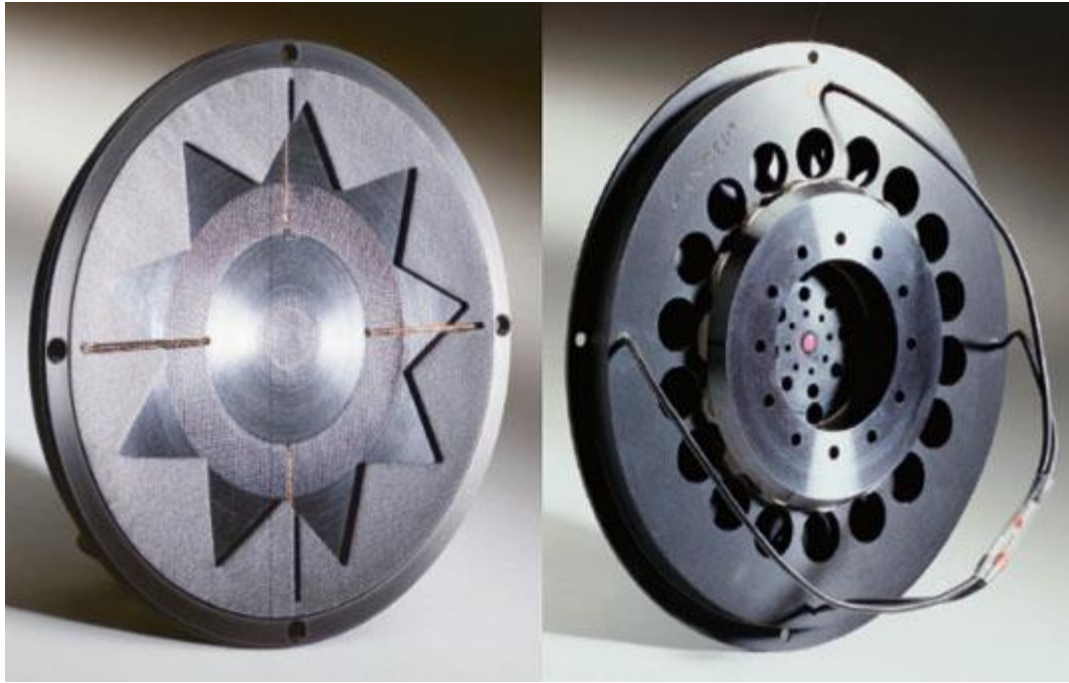
The great majority of legacy loudspeakers will fail a square wave test spectacularly. Creating a time accurate speaker that will reproduce a square wave is only a matter of finding engineering solutions to the problem. The image above shows the acoustic output from a square wave input of an experimental 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 moving an appreciable distance.

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.

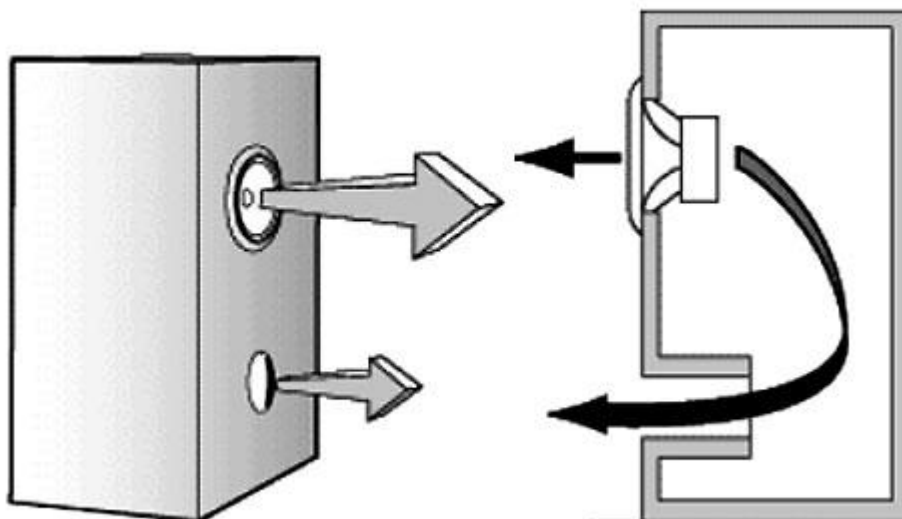
Clearly low frequency reproduction requires physically large devices. It is simply not possible to radiate low frequencies from iPhones and tablets when the volume of the product is less than the displacement of a large woofer, which is why such devices need to be used with earphones for music.



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

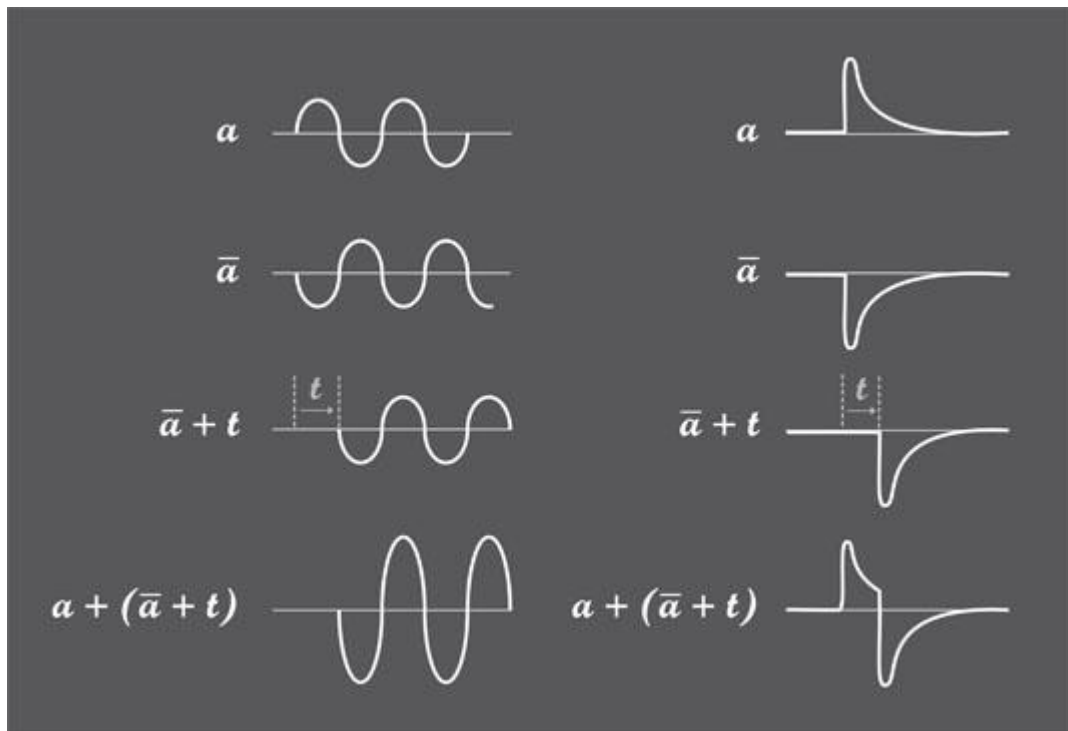
Back in the dark ages, magnets were made of alnico (aluminium, nickel and cobalt alloys) whose magnetic characteristics dictated long thin magnets. Small voice coils sat in the centre of cones made of flimsy paper. It made sense, then, to try to reduce the probability of the cone flexing by using resonant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

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.