Time domain accuracy in human hearing and loudspeaker design - By John Watkinson.

https://www.thebroadcastbridge.com/content/entry/7066/loudspeaker-technology-part-1

Broadcasting began with sound radio, and sound was soon added to early silent movies. Television began with both sound and picture. And sound continues to be an important part of television broadcasts, whether truly broadcast or carried over some IT structure.

Today there is scarcely an information device, be it a cell phone, tablet, laptop, Satnav, or game that isn't capable of emitting some kind of sound. Whilst the broadcaster or disseminator of sound has no control over how it will finally be reproduced at its destination(s), it is at least necessary to have the ability to monitor the quality of what is leaving, as well as checking during prior production steps.

There is an argument that any originator of sound, such as a broadcaster or recording studio, should have monitoring equipment that is as good as or better at revealing faults than anything the consumer is likely to listen with. I suggest that is only an argument, and it has not universally been won.

Defining sound

With few exceptions, the loudspeaker exists to provide some kind of stimulus to the human auditory system (HAS). In the same way that making television and cinema work requires some knowledge of the human visual system, then making loudspeakers work most effectively requires knowledge of the HAS and that will be considered in this series.

There is an old saying that if a tree falls in the forest and there is no-one around to hear it, does it make a sound? The answer depends on perspective and how sound is defined. To the physicist, sound is the propagation of pressure disturbances through some kind of gas, from blast to ultrasonics. By that definition, the tree made a sound. If sound is defined as an auditory system detecting an acoustic stimulus, then there was no sound.

The HAS can only operate over a certain range of frequencies and can only tolerate a certain range of levels. Perhaps the old saying could be replaced by asking what happens if someone blows a dog whistle and there is no dog?

Whilst physics may be mathematical, it is also extremely reassuring. The laws of physics are unmoved by national boundaries, the passage of time and the utterances of politicians or marketing departments. Once people get involved, with their variable education, hopes, emotions, traditions and so on, it all gets rather messy and success comes to those who can see through the mess.

Sound versus perception of sound

One of my many hats is that of an acoustician and loudspeaker designer, and it is often assumed that I will be interested in hi-fi. Well, in a sense I am interested in hi-fi. Not in the products, but in the psychology. At one time the technology needed for accurate reproduction of sound didn't exist, and it became the norm to strive for an improvement through endless listening and tinkering. Then tradition took over and the subjectivist tinkering became more important than the goal.

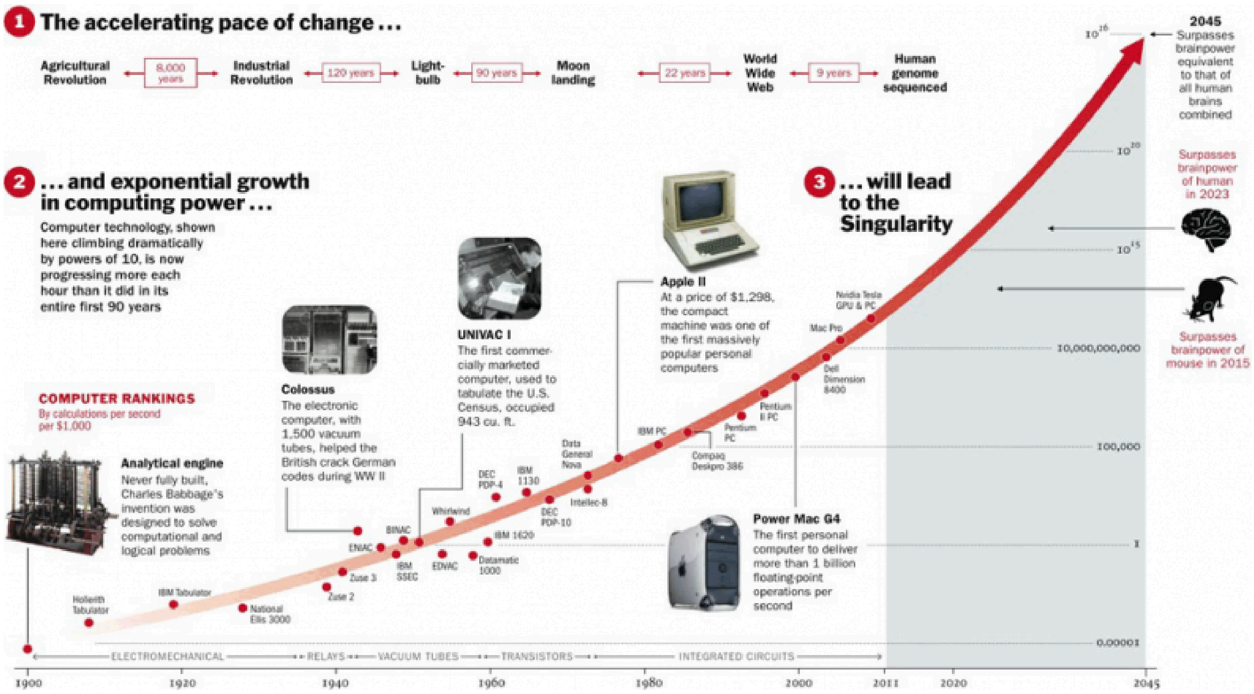
In a parallel to travelling, rather than arriving, it seems better to tinker with audio than actually achieve anything. As a result hi-fi has not achieved anything, except possibly a reputation, and indeed cannot achieve anything, because it is a process and not a goal.

Audio enthusiasts all have different ideas about what makes the perfect speaker. Shown here is a set of Ixoost-Xilo-51 Bluetooth speakers.

As a friend recently pointed out, the last thing a hi-fi enthusiast wants is a perfect loudspeaker, as then he would have nothing to do. In order to avoid arriving at the perfect loudspeaker, modern understanding of the HAS must be ignored or denied, and physics must be replaced with mythology so tradition can be kept alive and the snake oil can keep flowing.

Those of us who do not subscribe to such thinking should not be too smug. No one has ever looked at a TV or a cinema screen and confused it with the real thing. It may not even be necessary. With a few exceptions, such as reversing an RV or monitoring a prison wall, watching screens is a form of escapism, and realism isn't on the menu. That goes for sound too. As with so many human senses, we don't perceive what is actually there, but instead we perceive what social and other pressures suggest is there.

This means that many people will happily listen to sound quality that is significantly worse than what is possible because it is only serving as a reminder, in the same way that a nasty die-cast replica of the Eiffel Tower transports the owner back to Paris. Given the ubiquity of mediocre sound, I also suspect that many people simply don't believe that anything better is available. Even if the conscious accepts mediocre sound, the sub-conscious doesn't and that is the origin of listening fatigue, which mysteriously disappears once a quality threshold is crossed.



While much of electronics continues to benefit from Moore's Law, the physics of speakers limit its application. Image Reddit.

The incompatibility of Moore's Law

Looking on the bright side, tangible progress in loudspeakers has come from the requirements of a completely different market, that of IT. The relentless advance described by Moore's Law has seen anything to do with computing or communications shrink in size, cost and power consumption. Whatever can be digitised is fair game, including audio waveforms. Unfortunately, Moore's Law does not apply to acoustics. Even so, the enormous pressure to reproduce audio from ridiculously small devices such as iPads and tablets has led to some remarkable results.

In almost all other areas, loudspeaker design seems to be stuck firmly in the mud without enough traction to escape. One driver for progress in larger loudspeakers may be the cinema. Projection has just about reached the limit for the amount of light power that can be shoehorned through a gate. The future of very large bright screens may lie in direct radiation. This is great for the picture, but such screens do not let the sound pass through, and will require cinema sound to be re-thought.

The ubiquity of IT also means that the loudspeaker of the future will not have any terminals, but will instead obtain the sound to be reproduced via some wireless means. By definition this requires the loudspeaker to be active, which in itself has advantages. Those advantages form a subject in itself.

This is an extensive topic and it is only possible in this first part to give some idea about where the series is going. If it is accepted that the goal is to please the Human Auditory System, we need to start there, to look at how it works, what's critical and what doesn't matter. One important fact that will emerge is that few people ever hear a sound source of any kind, including a loudspeaker, the way it really sounds.

In the real world, the original sound cannot be separated from the response of the environment: reflections, reverberation and so on. It will therefore be important to see how the HAS deals with that, because it will also tell us how to design loudspeakers that work well in reverberant environments, like musical instruments do, rather than only in specially treated rooms.

The HAS is directional, can readily determine where direct sounds and reflections are coming from and can separate them. If we wish to make any progress with stereophonic or surround sound, that mechanism needs to be understood and respected. Stereo and surround sound create sonic images, whereas TV and cinema create visual images. The difference is that in TV and cinema there are standards for measuring the accuracy of the image, whereas in audio there are none.

No one believes that the pixel count is the only parameter that determines the quality of a display; there are plenty of others I don't need to mention. However, when it comes to loudspeakers the only parameter of interest seems to be the frequency response. Often this means it must be extended at all costs even if it means damaging other parameters that are not measured. Clearly there is more to loudspeaker design than frequency response.

By looking at the parameters that affect the degree of respect shown to the HAS, it is possible to create a specification or goal for an ideal speaker. Once that specification is understood, meeting it is a straightforward engineering challenge.

Can the challenge be met? Of course. The human ear is imperfect and if the imperfections in the speaker are less than the imperfections in the ear, then they will be indetectible.

Part 2

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

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.

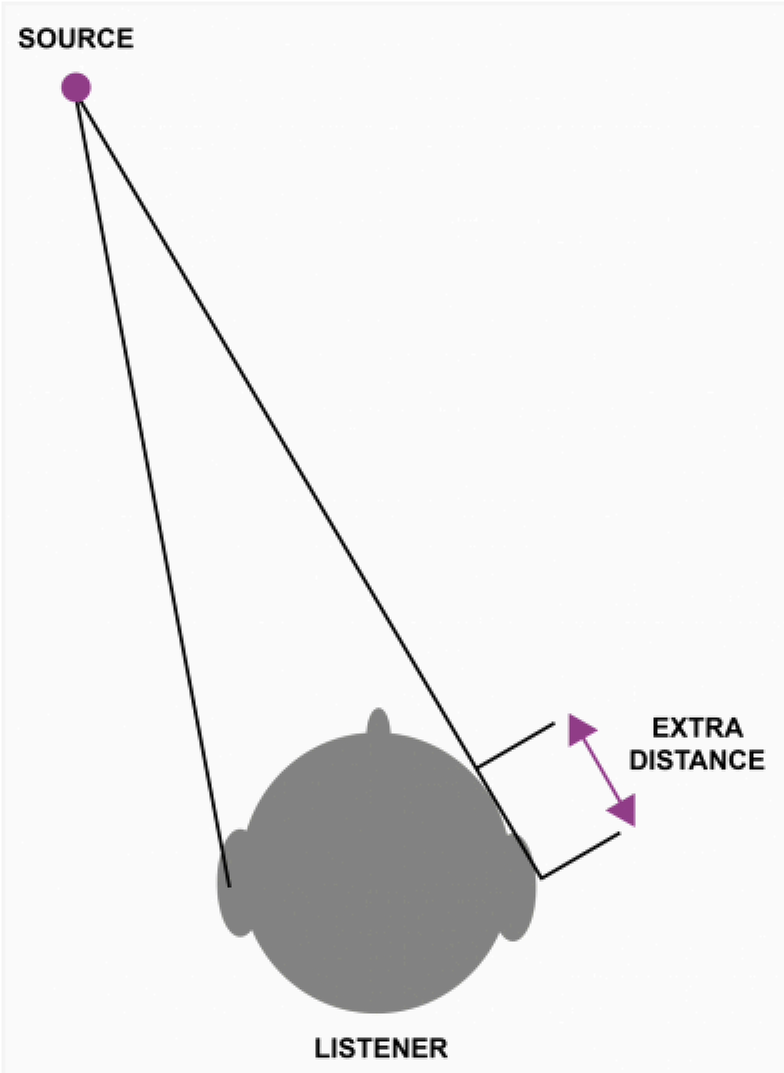


Figure 1. When a sound source is off to one side of the listener, the sound arrives at the two ears at different times. The HAS can identify the same transient at both ears and measure the time difference.

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.

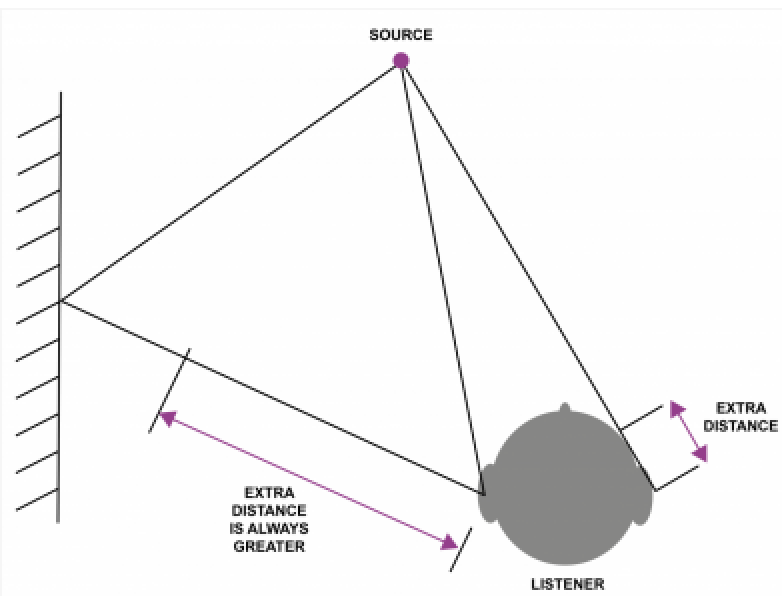


Figure 2. Reflections don't reduce the ability to determine direction, because they arrive too late. Instead the reflections and the direct sound are time shifted to make the source more audible.

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.

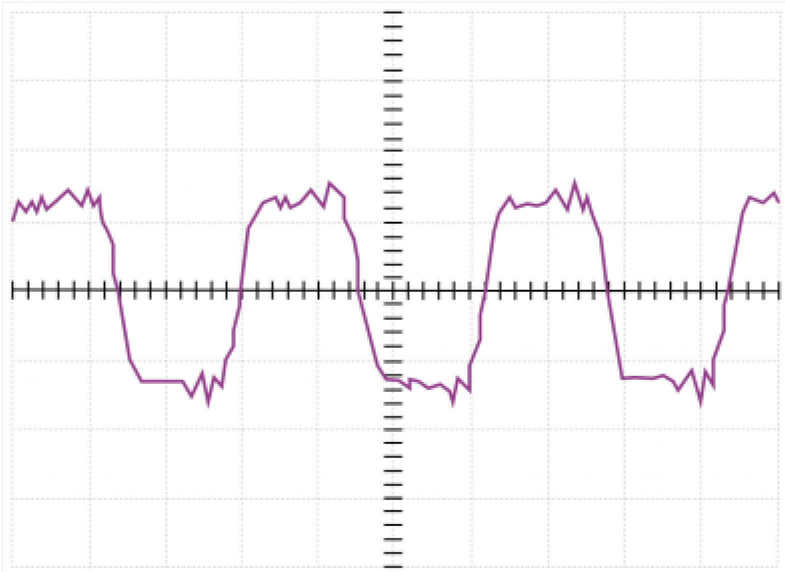


Figure 3. The square-wave response of a loudspeaker designed to meet the time-accuracy criteria of human hearing.

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.

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

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.

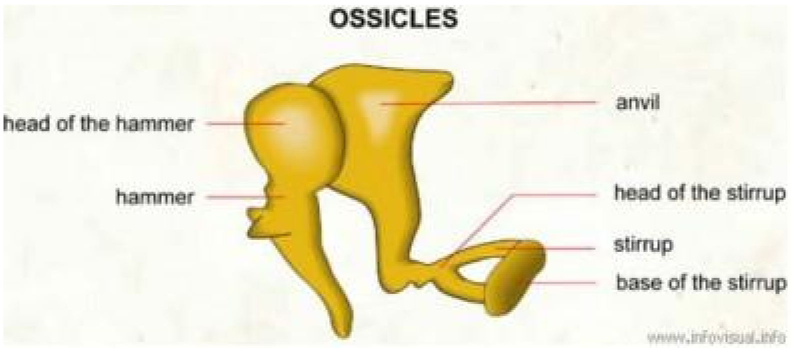


Figure 1. The ossicles are three bones in either middle ear that are among the smallest bones in the human body. They serve to transmit sounds from the air to the fluid-filled labyrinth.

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.

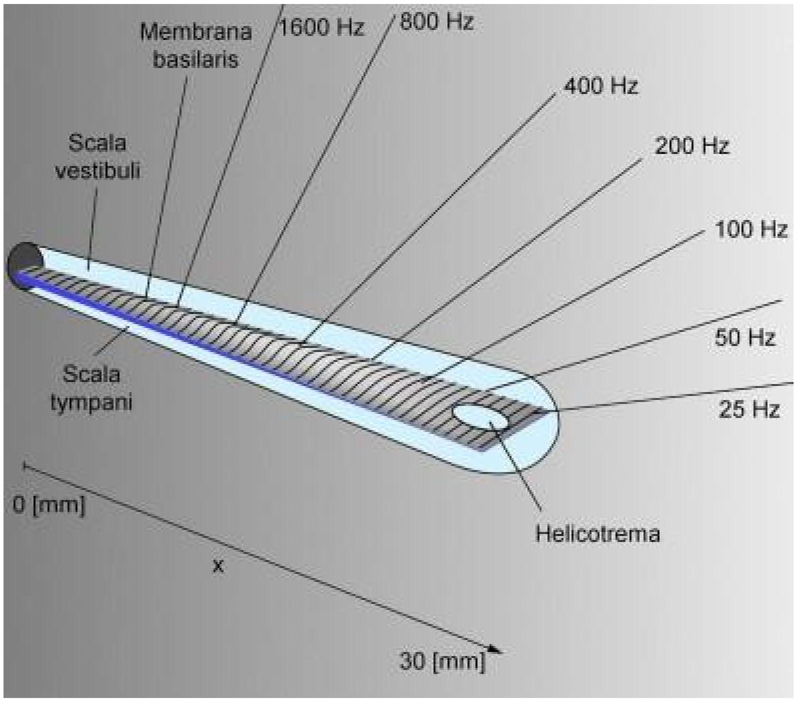


Figure 2. Human ear basilar membrane. Image: Kern A, Heid C, Steeb W-H, Stoop N, Stoop R, Biophysical Parameters Modification Could Overcome Essential Hearing Gaps.

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.

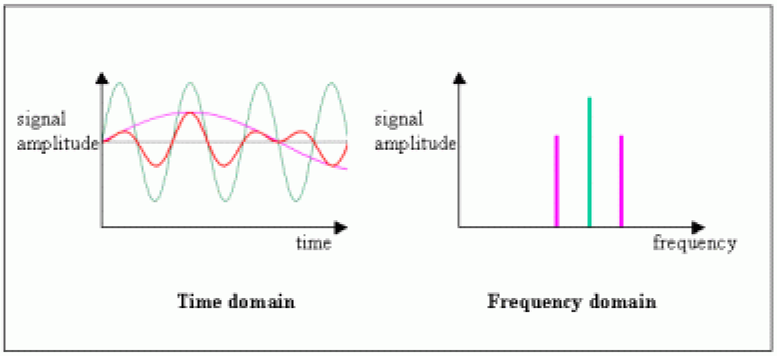


Figure 3. Relationship between time and frequency domains.

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.

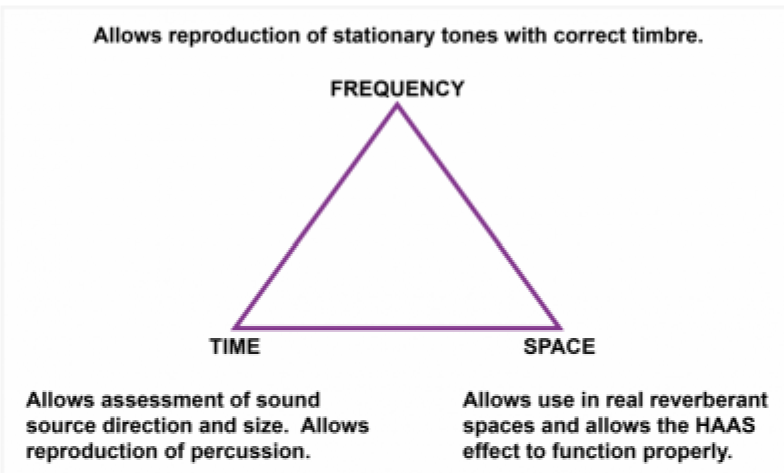
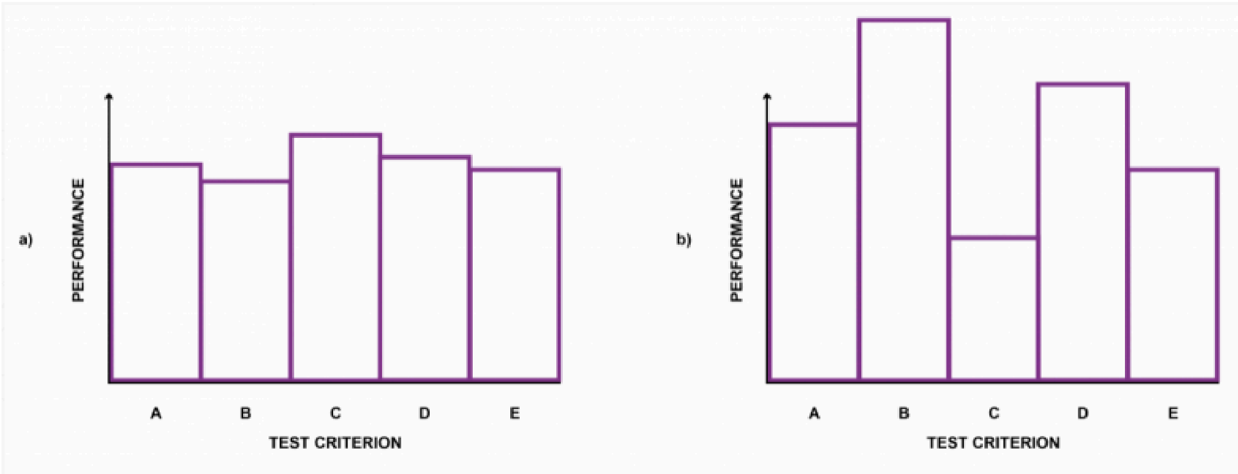


Figure 4. The three important domains in which a realistic loudspeaker must meet performance criteria; Time, Space and Frequency. Neglect of any one will nullify excellence in the other two. Legacy speakers concentrate on the frequency domain and so they always sound like loudspeakers.

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.



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.

Part 4

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

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.

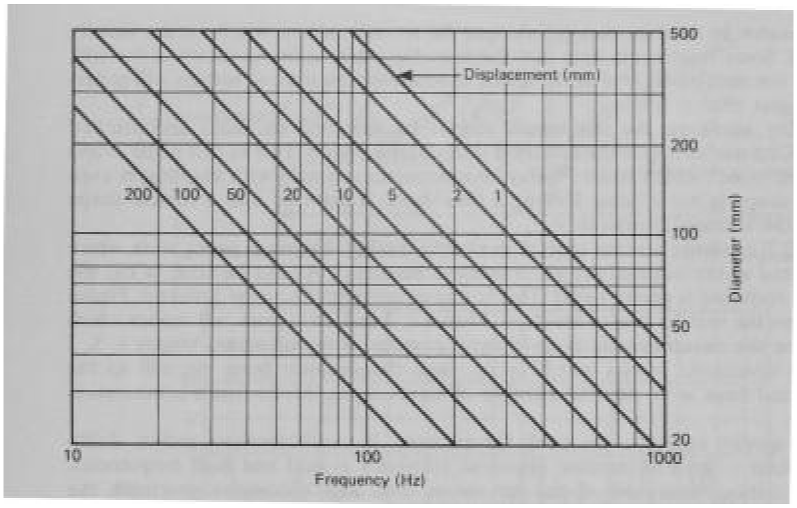


Figure 1 shows the peak displacement needed by a diaphragm of a given diameter to radiate 1 Watt at a given frequency. Note how the amplitude rises as frequency falls.

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.

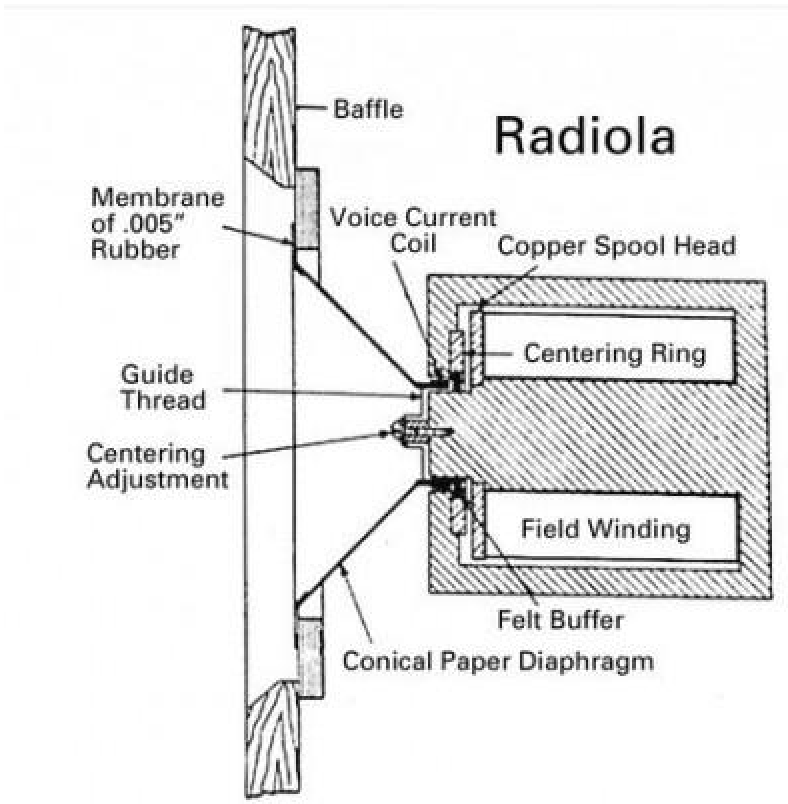


Figure 2. Basic Rice-Kellogg moving coil speaker.

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.

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.

Parts 1, 2 and 3 of this loudspeaker series can be found at the links

Editor Note: Readers may wish to read other Watkinson articles on The Broadcast Bridge. Two are listed below: