Power Limitations in High Quality Sound Reproduction

- Neville Thiele[1]

Summary. Parameters that affect the accurate reproduction of sound at high levels are examined and suggestions made for their amelioration. Topics covered include the true peak amplitude of an audio signal, recovery of amplifiers after overload, the loudspeaker, prevalence and toleration of clipping signal peaks, maximum driver excursion x_{max}, high pass filtering, equalization and group delay.[2]

Introduction.

Loud sounds are one of the joys of life. Not continuously loud roaring sound, but sudden bursts of volume that leap suddenly and unexpectedly out of gentleness and quiet, the sudden surprise when a band strikes up. Variety of dynamics has always been one of the devices available to musicians and actors to engage the interest and enthusiasm of an audience, along with variety of pitch and timbre and rhythm and pace.

Yet too often, when sound is reproduced, that quality is diminished or lost. This happens for a variety of reasons. The dynamics of much popular music was crushed, beginning around 1957, by record manufacturers competing for loudness and thus commercial exposure through fixed-gain juke boxes in public places, and has remained so ever since. In radio and television the primary function of sound most often is to maximize the audibility of the sponsor's message. So, in deference to the all-powerful sponsor, the dynamics of programme before and afterwards must be graduated regardless of the original artist's intentions so that the changes are not apparent to annoy the listener.

Even when it is not inhibited by such limitations, sound reproduction with unfettered dynamics goes hand-in-hand all too often with distortion. The discussion below will deal with some of the reasons and suggest some ways of ameliorating such damage.

True Peak Amplitudes.

The maximum reproducible sinusoidal Sound Pressure Level (SPL) in the atmosphere at sea level is approximately 194dB. This is limited not by the maximum pressure part but by the limit of a complete vacuum on the rarification parts of the wave. The wave can cycle from this absolute vacuum level to double atmospheric pressure on peaks, thus maintaining a mean pressure of 1 atmosphere. Overpressures are possible as with explosions but not for repetitive waveforms. Fortunately audio reproduced for pleasurable human experience lies well below these levels.

When levels of live sound are quoted, e.g that normal conversation level at 1 metre is +74 dB SPL or that a symphony orchestra produces a peak sound level of +94 dB SPL, it is traditionally in terms of readings by meters that do not pretend to read true peak levels. The ratio of true peak level to meter reading varies somewhat, of course, with the nature of the programme material and the ballistics of the meter. However, the standards of broadcasting and recording studios provide a useful guide.

For many years, the VU meter was the indicator of programme level in many counties, e.g. China, Australia, France and the United States. All meters were calibrated to produce the same meter deflection with the level of a sine wave signal, called Alignment Level, throughout each installation of studios and output destinations, recorders, programme distribution channels or transmitters. Different installations used different Alignment Levels, e.g. 0 dBu, 4 dBu or 8 dBu, but the one level applied throughout each installation. Other countries, e.g. most of Europe, used Peak Programme Meters, but the same principle of a standardized Alignment Level applied. Then all amplifiers, recorders etc. throughout the installation had to be capable of handling the peak level of all analogue signals presented to them. This was obviously above alignment level, but by how much?

The Australian Broadcasting Corporation specified that its audio distribution amplifiers must produce less than ¼% THD at +12 dB above Alignment Level and not clip any signals below +16 dB (where clipping was determined as the rapid rise of distortion products exceeding audible limits and often taken as 2%). Answering an ABC contract, a highly respected manufacturer of mixing desks specified that clipping should not occur below + 20 dB above Alignment Level and reported a commercial advantage in doing so.

With the advent of digital transmission, an inflexible upper limit of the maximum sized word was enforced. For 16 bit audio this was 7FFF. This level was the maximum catered for could not be exceeded. The question was just how to relate this to the program levels in use. After much study, eventually maximum word Signal levels of 7FFF were specified worldwide as being +18 dB above Alignment Level for broadcasting and +20 dB for film sound.

Thus a figure of +18 dB or +20 dB is not an unreasonable, though admittedly "worst case", estimate of the "headroom" needed to reproduce sound truthfully, without distortion, and while not necessary for all kinds of audio programmes, it is a prudent figure to avoid distress, to the equipment and the listener, for every kind of material.

Then, even if we take the margin for "headroom" as +16 dB rather than +20 dB, the Sound Pressure Level (SPL) required at the listeners' ears become +90 dB SPL for speech and +110 dB SPL for orchestral music.

[1] This paper is one of a number of Neville's "works in progress" at the time of his passing. It has been completed in the spirit of Neville and with minimal change to his input. (Graeme Huon, ed).

[2] Based on the paper presented initially at International Symposium on Electro-Acoustical Technologies (ISEAT2011), Shenzhen, China, 2011 November 12-13

Now when a source radiates 1 acoustic watt into free space it produces +112 dB SPL at 1 metre distance. Thus assuming, for example, that for home listening the room provides for a listener an acoustic signal equivalent to 1 metre distance in free space (most likely pessimistic) and that the loudspeaker produces +92 dB SPL at 1 metre with 1watt in free space (quite likely optimistic) the amplifier would need to deliver peaks of 600 milliwatts for speech and 60 watts for orchestral music. A loud rock band would need an even bigger margin, and to cover greater listening distances such as public venues, even more again because of a drop of 6dB for every doubling of distance according to the free space inverse square law.

These levels allow for reproduction of the loudest expected sound. They are not the continuous sound level. For 5.1 cinema sound, when the loudest expected sound level is taken as 112dB this corresponds to a per-channel contribution of 103 - 105dB SPL delivered to the listener for each of the five channels, noting that channel signal contributions will not normally be phase-consistent and so will not directly add. Going by the Programme Level, the average reading during loud passages will be 18 – 20 dB below this, or approximately 83dB SPL.

Recovery of Amplifiers after Overload.

When some elements of a sound reproducing chain are overloaded and distort for a short time, it is usually assumed that, once the overload ceases, each element will return immediately to its previous undistorted state. This certainly applies, for example, to loudspeakers and magnetic analogue recording tape but it does not necessarily apply to amplifiers, which are present in great numbers in any reproducing chain. In our discussion below we will be concerned primarily with the power output amplifier, the final link in the chain that feeds the loudspeaker, and the mechanism that can easily convert the effect of a momentary overload, just a few milliseconds long, into a distorting mode that can persist for up to a second.

Most amplifiers use negative feedback as a powerful means of lowering distortion and ensuring the low output impedance needed for driving electromagnetic loudspeakers. Even designs claiming minimal or even no feedback often include internal current sensed voltage feedback through such mechanisms as series cathode, emitter or source resistors.

The feedback amplifier schematic Fig 1 presents an amplifier as comprising three parts, an input adding, or subtracting, stage which takes two inputs $e_{in,}$ from the input, and $e_{fb,}$ the feedback, and feeds their difference e_{in} - e_{fb} to an amplifier proper, of gain μ. The output e_{out} of this stage feeds the loudspeaker and also an attenuator that feeds a proportion of it, βe_{out} ,back to the input as e_{fb}. The result is to reduce the open loop gain μ, and with it the distortion and the output impedance, by a factor $1 + \mu\beta$.

The resulting voltages are typified by Fig 2, where the feedback factor $1+ \mu\beta$ is 10 and the amplifier is being driven to the brink of distortion at its maximum output signal voltage e_{out} of 10 units peak amplitude by a difference signal $e_{in} – e_{fb}$ of 1 unit peak. Fig 3 shows the input and the feedback voltages when the input voltage has increased by 1 dB to 11.22 units peak and the output voltage, driven to distortion, clips, hard for the sake of illustration. The difference voltage in Fig 4, after rising undistorted to 1 unit, then rises rapidly to an amplitude of 2.2 units peak, most likely beyond the capabilities of the early drive stages.

With the gain reduction factor increased to 100 and an overload of 1 dB as before, the input and feedback voltages in Figs 5 and 6 are so high as to rise rapidly "out of sight", even with a small display of difference signal. Then, with an increase of 1 dB in the input signal, the difference voltage, after rising undistorted to 1 unit height, goes during output clipping to a peak of 13.2 units, a level certain to drive one and possibly all the earlier stages of the amplifier into non-linearity. Modern power amplifiers often have gain reduction factors greater than 100 times, resulting in the possibility of significant internal overdrive with internal errors arising from overdrive or out of band signals.

Now when driving stages go non-linear, their mean d.c. voltages change and thus also the charge in any internal coupling capacitor(s). Additionally, when the inputs of both transistor and valve stages are overdriven, they act as diodes, with quite low forward resistance. This can then charge the coupling capacitor rapidly, easily driven by the great increase of voltage swing during output clipping, and when the series resistance is low during diode conduction and high when the conduction ceases, a large charge can accumulate that then will take a much longer time to discharge.

In this way a small momentary overload may cause an amplifier to mis-operate afterwards and continue to distort over a considerable time, sometimes for the a large part of an elapsed second or more. Thus an amplifier that performs well, and measures well in other respects with signals below overload, may easily be unsatisfactory for practical use.

Overload Restoring Time.

Such problems can be avoided, or at least mitigated, by careful design. On the other hand they may be comparatively hard to predict, or find, in a complex device, so it is comforting to know that a simple test is available for detecting and quantifying them *[1]* .

The amplifier under test is fed with a sine wave signal, e.g. at 1 kHz., through a 20 dB switchable attenuator, e.g. as in Fig 7, so that its output signal is 10 dB below full output, and is read on an oscilloscope with its time base set to a slow scan, e.g. 5 seconds total. The attenuation is then removed for 1 second, allowing it to rise 20 dB, well into overload. After 1 second, it is reapplied, and then ideally the output voltage would resemble the input in Fig 8. The output observed on the oscilloscope, however, may well resemble the output sections of the diagram, and the overload restoring time, during which the amplifier may continue to distort, is recorded as the time elapsed from the re-application of attenuation until the output has reached 1 dB below its final level.

Note however that after the overload ceases, not only is the output signal initially attenuated below its normal level; its centre line also is often displaced as in the diagram, denoting a d.c. shift that may return to normal with a damped L.F. oscillation. Such artifacts produce distortion and demand the designer's urgent attention.

A Simple Design Example.

The need for care and the reward for attention to recovery from overload are illustrated by the simple example below, which was published so long ago that it used valves, but illustrates the general principle clearly. It was published in the Proceedings of the Institution of Radio Engineers Australia, which had little circulation outside Australia.

It has received little interest elsewhere so it may be worth re-visiting. During the development of a television receiver, it was noticed that when the receiver reproduced a loud sound, the display on its picture tube also contracted vertically. But when the loud sound ceased, the display not only expanded again, it overshot to an even greater height before slowly contracting back to normal. Easy to understand; the audio power amplifier and the vertical output deflection stage were both supplied from a common h.t. rail, decoupled from the rest of the receiver by a 220 ohm resistor. Clearly the audio output valve drew higher current during loud passages, and afterwards drew abnormally less current until it returned slowly to normal. But why?

The circuit diagram of the amplifier is shown in Fig 9[3]. In the diagram, the initial values for the components C1, R2 and R3 are bracketed, 0.047 µF, 1MΩ and 10 kΩ. The "paralysis" or at least "semi-paralysis" that seemed to be the most appropriate description of the recovery process was produced when overload produced a high signal from the first valve that drove the control grid of the second, output, valve positive with respect to its cathode so that it acted as a diode. The diode conduction clipped the output and thus the feedback voltage, producing a much larger difference signal and output from the first valve, very soon exceeding its maximum capability in the manner that we have seen earlier. Then this large signal rapidly increased the d.c. charge on C1, the 0.047µF capacitor, through though R3, the small 10KΩ resistor.

When the loud sound ceased, the grid no longer conducted like a diode and C1 discharged its comparatively large extra charge at a more leisurely pace through the much larger R2, 1MΩ resistor, and the amplifier's operation returned comparatively slowly to normal.

The solution was to change the component values to the un-bracketed values shown. R3 was increased as much as possible, to 470 kΩ, R2 reduced to 470 kΩ so as not to exceed the recommended total grid resistance of 1MΩ, and capacitance C3 reduced as far as possible to 0.01 µF, which reduced its time constant with R3 to 4700 µs and thus moved its open-loop -3 dB frequency to the higher, but in the circumstances acceptable, figure of 34 Hz. As a result the overshoot of picture height virtually disappeared

But solving that problem also produced a second serendipitous effect. To quote from the initial publication *[2]*:

> "In listening tests conducted with the (modified) amplifier….. , ….and also with a push-pull amplifier using two 6BM8's designed on a similar basis, programme peaks could be seen to flatten in a cathode-ray oscilloscope about 3 dB before they were detected by ear, and even 6 dB overload on speech peaks did not produce audible distortion. This compares with tests on conventional amplifiers (e.g. the unmodified version of Fig. 1) where the onset of distortion was detected by oscilloscope and by ear at the same level.

While such subjective results must be assessed with caution, they do indicate the substantial improvement that can be obtained." In other words, the modified amplifier behaved, as far as perceived distorted distortion was concerned, as if its power output had been at least doubled.

Such "paralysis", as we called it, occurs to a greater or lesser extent in all amplifiers. The valve amplifier of Fig 9 that produced these results had a feedback gain reduction factor of 14 dB. High quality valve amplifiers can have gain reduction factors exceeding 26 dB. Modern transistor amplifiers have gain reduction factors of more than 40 dB, and have no output transformer to limit open loop gain at the lowest frequencies and minimize low frequency oscillation during recovery.

This increase in gain reduction by feedback increases the likely severity of paralysis, so modern power amplifiers need even more care in their design and in measurement of their performance. It is a useful strategy for rapid recovery to restrict overdrive limiting and recovery issues to the lower signal level sections of an amplifier where possible, rather than to have overload limiting occur in downstream high power stages where time constants can be significantly longer.

A further indication of the ease with which short high amplitude peaks can be formed that can paralyse unprepared amplifiers while remaining, of themselves, undetected by ear, may be gained from an investigation into the statistics of a signal that was considered (and rejected) for the measurement of intermodulation distortion.

---

[3] The behaviour of valves/tubes, bipolar and FET active devices and their similarities and differences can be universally modeled by the use of Charge Control Theory (Cherry). The theory underlying the example remains applicable to modern circuit designs

Four sine waves of equal amplitude that execute complete sequences of cycles in the same sampling period, with different numbers of cycles per sequence, 31, 65, 203 and 303, all unrelated by a common factor, were added

Level Number	**16**	**15**	**13**	**9**
Signal Level below Peak (dB)	-0.6	-1.2	-2.5	-6.0
Single Sine Wave	232	329	471	683
Sum of 4 Sine Waves	0	6	35	155

together. Each sequence of 2048 samples was then quantized into 32 levels. Because the resulting waveform is symmetrical about the centre line we need only examine half the samples that were of the one polarity, i.e. examine 1024 positive-going samples quantized into 16 positive-going levels (bins). The table presents the populations of samples that reach to different levels below peak.

Table 1. Numbers of samples (out of 1024) that reach various quantized levels in a signal summed from 4 equal amplitude sine waves.

This example reinforces our previous observations of the great difference between mean and peak levels of simple and complex audio programmes.

Clipping in High-Quality Magnetic Recordings.

The prospect of tolerating clipping, soft or hard, so long as it is non-paralytic may be offensive to an engineer aiming for the highest possible standards of sound reproduction, but it is important to realize that some clipping - and the important question of how much? – has often been tolerated in the past when it allowed better performance in other respects.

A case in point is magnetic analogue recording tape, whose development as a high quality medium always depended on minimizing background noise. It was therefore even more important than usual important to record as high a level as possible without audible distortion. Also high frequency components were pre-emphasized so that, even though it further increased the risk of overload, de-emphasis in playback minimized high frequency noise.

The compromise struck between recorded level and noise led to standards that allowed tests at 1kHz to produce 3% THD at 6 dB above Alignment Level in some US specifications and 2% THD at 8 dB above Alignment Level in specifications of the Australian Broadcasting Corporation, as well as pre-emphasis to lift high frequency content at a 6 dB per octave rate for frequencies above 2 kHz. Yet, as we have seen, those levels, at middle and low frequencies were 6dB and 4 dB respectively below generally accepted figures for true peak levels in programmes. The author is unaware of distortion figures measured at levels higher than these, but the consequences described below give some indication.

It was instructive to compare live music immediately with the same piece replayed from analogue tape *[3]*. A piano, in particular, loses the sharp edge of its percussive initial transients. Further, in the 1970’s, an Australian radio producer set out to restore jazz records of the 1930’s and 1940’s, using the Packburn equipment, an early device for reducing from monophonic 78 r.p.m. discs. The disc was played with a stereophonic cartridge, producing Left and Right “stereo” signals from the two walls of the groove. On a clean disc, the two signals were identical, but any noise or wear on one wall added a pulse to its signal that made its voltage output instantaneously higher than the other. The Packburn device compared the two signals continuously, and selected instantaneously the one with the lower voltage. Its results were remarkably successful.

But when the producer, not wishing to play a precious mint record a number of times while adjusting the Packburn’s clipping levels, tried to play each disc just once and record the signal on stereophonic magnetic tape for later manipulation, he found that no noise pulses had survived. Analogue tape, driven hard at low frequencies and even harder at the pre-emphasized high frequencies, could not reproduce such short pulses. In the end, he achieved his object with an early encoder that recorded stereo digitally on videotape.

Yet analogue magnetic tape, from the late 1940’s until well into the 1970’s, was accepted unquestionably as a recording medium of the highest quality, used in all broadcasting studios and for mastering nearly all LP discs and later, as ADD, many CD’s. It clipped, i.e. distorted, the signals, but only instantaneously on the highest peaks, free of paralysis, and that distortion remained unnoticed, or sometimes even preferred, by generations of listeners.

Loudspeaker Driver Distortion.

The Study of True Peak Amplitudes earlier in this paper asserted an SPL delivered to the listener, and quoted distortion levels in the programme chain that were below, or at, the threshold of audibility. What was temporarily ignored was any distortion introduced by the loudspeaker transducer itself. Unfortunately, loudspeakers do distort, and quite significantly when maximum SPL is required.

Power ratings for loudspeakers are commonly based on survivability rather than performance. It is therefore not uncommon for distortion levels well in excess of 10% when devices are driven to quoted power ratings.

For direct radiating loudspeakers in particular, this limiting distortion is most often cone displacement-dependent and so occurs at the lowest driven frequencies for each driver. Harmonics and intermodulation products generated by extremes of excursion cannot be limited by upstream electronic filtering and will be further emphasized by peaks occurring in the loudspeaker driver response itself, virtually independently of any upstream equalization.

Fortunately, when a loudspeaker is driven to distortion, it generally recovers instantaneously, somewhat in the manner of analogue magnetic tape when saturated. The perceptual impact of distortion is thus somewhat minimized but individual driver frequency response may extend far enough to have harmonics fall outside source spectrum masking and may be emphasized by increasing directivity of the driver at these higher frequencies. Further, some electromagnetic drivers exhibit a non-linear "pump-down" behaviour where the mean cone position shifts temporarily and recovers at a driver-dependent rate when the drive signal is reduced. In extreme cases, the voice-coil can be pumped completely out of the gap, dramatically reducing all output and running the risk of permanent mechanical or thermal damage.

When designing a system it is important to consider that the distortion figure for maximum expected Programme Level delivered to each listener must include and will often be limited by harmonic contributions of the loudspeaker drivers themselves.

Low Frequency Reproduction and x_{max}

Every loudspeaker engineer is aware that the power that can be delivered at low frequencies is limited by the volume of air that its diaphragm can move. Thus its maximum excursion x_{max} depends on the square root of W_{AC}, the radiated acoustic power, inversely on f^2, the square of the frequency being radiated and inversely on d_c^2, the square of the diameter, and thus inversely on the area, of the cone.

$$x_{max} = \frac{2.73\times10^9\sqrt{W_{AC}}}{f^2 d_c^2} \qquad -(1)$$

where x_{max} and d_c are in mm, W_{AC} in watts and f in Hz *[4]*.

Eqn (1) shows that the reproduced acoustic power depends only on the mechanical dimensions of the cone, on the product $x_{max}d_c^2$ and thus on the volume of air that the driver can pump. It is quite unaffected by anything the designer can do in other parts of the driver or in the amplifier. And for a given driver, its power capability falls rapidly with decreasing frequency. This excursion limit can be ameliorated somewhat, decreased to about a half, by using a vented box, whose vent takes over from the driver the radiation of power at the lowest frequencies.

As a guide to the dimensions involved, we consider the excursions needed to radiate an acoustic power of 0.1 Watt, that would produce a sound pressure level (SPL) of 102 dB at 1 metre. A low frequency driver of 300 mm diameter reproducing that level at 30 Hz would need an x_{MAX} of 10.7 mm, while a tweeter of 25 mm diameter at 2000 Hz would need an x_{MAX} of 0.35 mm.

We should also point out that if the driver efficiencies were 1%, i.e. rated at 92 dB SPL at 1 metre with 1 watt input, the amplifier would need to produce 10 electrical watts for that level, but with efficiencies at the quite common figure of 0.2%, i.e. with drivers rated at 85 dB SPL at 1 metre with 1 Watt input, the amplifier would have to deliver 50 watts of peak electrical power.

X_{MAX} depends on a number of features of the driver design, among them the design of the centre pole and front plate, whether the voice coil is under- or over-hung, and variation with motion of the compliance of the cone surround and spider, so it is not surprising that it varies greatly between drivers of similar dimensions. Fig 10 summarises the values of x_{MAX} for a sample of 111 drivers available on the European market during 2011. The cone diameters are nominal vales. The true piston diameters would be a little smaller. The summary is not claimed to be exhaustive or even typical but it does indicate how great the variation can be in practice. The greatest variation is among the smallest of the low frequency drivers, the best of which make heroic but inevitably doomed attempts to reproduce really loud undistorted signals at frequencies down towards 50 Hz from "3 to 5 inch woofers".

There is another consideration related to x_{MAX}. At frequencies below cutoff, where the radiated power diminishes rapidly, at a 12 dB per octave rate with a totally closed box and at a 24 dB per octave rate with a vented box, the cone excursion remains high, and proceeds to a maximum. Thus any programme content at "sub-sonic" frequencies, not necessarily frequencies that a human cannot hear but rather that the loudspeaker cannot reproduce, can exercise the cone so that they intermodulate with audible sounds and produce distortion.

This is illustrated by Fig 11, which shows how in both the un-filtered closed box (dashed curve, labelled 2') and in the un-filtered vented box (solid curve, labelled 4) the excursion goes up to and remains at a high maximum value at the lowest frequencies. However, filtering the response to one order higher, to 3rd order from 2nd in the closed box and to 5th order from 4th in the vented box, reduces the out-of band excursion to a great extent. Filtering the response to two orders higher, to 4^{th} order from 2^{nd} in the closed box and to 6^{th} order from 4^{th} in the vented box reduces the out-of band excursion even further. Filtering to orders higher, in spite of its greater cost and group delay error, produces little further relief.

However such filtering, though simple in principle, cannot be applied in all circumstances. The performance of a closed box is easily improved, its response extended somewhat and its excursion reduced by the simple insertion of a capacitor in series *[5]*. The capacitance needs to be large, hundreds of microfarads, and a given driver ideally can be fitted into only one box volume to produce a given response. But it is a very satisfactory solution for a closed box loudspeaker whose performance is assessed on a stand-alone basis. Sadly though, no equivalent simple solution is available for a vented box.

It is important also, in selecting a driver for a loudspeaker system, to be sure that its x_{MAX} is adequate for the intended application, as drivers can vary greatly in this respect. Fig 10 results from an analysis of parameters specified for drivers available in Europe during 2011. It is of interest, not only because it allows a (very rough) estimate of mean values, but also for the very wide range of values that are specified.

Active Filtering of a Loudspeaker.

The overall response of a sound reproduction system combines the response of all system parts from the source signal through all electronics and ultimately through the air to the listener. The elements of this response path can be combined in different ways and still achieve the desired result subject to satisfactory design of the component parts and the system as a whole.

When loudspeaker and amplifier are designed together as a single “active loudspeaker”, great flexibility is available for design *[6]*, especially in handling low frequency power. Great changes in response become possible by cascading small, inexpensive additional circuits with the input of the main amplifier. The overall response then becomes the product (or sum, when they are expressed in dB) of the individual responses of filter(s) and loudspeaker(s).

Whilst an active system requires electronic filtering followed by multiple amplifiers and loudspeaker connection circuits and so is inherently more complex, there are several potential advantages of splitting the audio into filtered bands before presentation to amplifiers and so to the loudspeakers.

- The peak signal requirement for each band need be met for that band only, rather than catering for the peak requirement for the full bandwidth. As an example, in most systems, lower frequency system power (voltage swing and current) requirements usually dominate. Where a single amplifier is approaching maximum signal swing capability for low frequencies, there will be limited voltage swing available for all bands, and so the risk of clipping and intermodulation distortion products is increased when having to meet additional demands in the other bands.. With an active system, full headroom will remain available to other bands even when the lower band has reached the limit of voltage/current. On typical program material this can result in significantly increased dynamic range and increased usable undistorted output.
- High order filters can easily be constructed electronically. High order passive electrical filters are far more challenging.
- Very accurate and physically compact filters can be constructed because there is no need for inductors. Inductance footprint is a problem particularly for high order filters and at low frequencies where inductance values can be large. Almost any desired response can be achieved with passive designs (L/R ratio) but the amount of copper required quickly becomes prohibitive).
- Distortion introduced by non-linear magnetic materials such as iron and ferrite is avoided.
- Filter loss is minimized (and so overall efficiency is improved) as there are no inductors with attendant series resistance.
- Stray permeability and magnetic fields have minimum influence. Where inductances are used, care needs to be paid to cross-coupling between coils in particular.
- Drive levels can easily be adjusted between bands. With passive crossovers, changing drive levels can necessitate a redesign of the crossover to adjust termination impedances.
- Variations in driver impedance do not matter. With a passive design, the driver termination impedance directly affects the response.
- Each loudspeaker driver sees an optimum low source impedance from its dedicated amplifier.
- Amplifier size can be scaled to individual driver requirements. Tweeters often require far less power than lower frequency band drivers for a given SPL.
- Amplifier clipping will only apply in the band of the amplifier. Any amplifier clipping in passive designs will generate harmonics that can damage high frequency units.
- Dynamic band-limiting strategies can be more simply (electronically) implemented.

In some instances a hybrid solution having both active (or digital) electronic and passive filtering can be used to advantage.

Electronic filtering can be implemented either digitally, active-electronically or passive electronically. For analogue designs, the simplest (and quite effective) procedure is to electrically filter the loudspeaker response to one greater order. This is done simply by suitably proportioning a capacitor-resistor coupling network. It also requires, of course, appropriate design of the loudspeaker/box combination for a suitable complementary response. We shall see later that to produce an overall response that is substantially flat before going into steep attenuation, a gently attenuating response, e.g. of a CR network, needs to be mated with the complementary peaked response.

It is also prudent to have the input CR network, calculated on the assumption that it is fed from a low impedance, isolated by a buffer amplifier. Otherwise its carefully estimated response might be impaired by unexpected impedance of the device feeding it.

When two orders of filtering or more are to be added, the design will demand creation of complex poles in the filter. These can be realized very simply by a Sallen and Key high-pass filter *[7]*, again with the precaution of an input buffer amplifier. But now, greater flexibility is possible. If our loudspeaker is in a closed box, the overall 4th order transfer function may be split into two 2nd order functions, either of which may be allotted to the loudspeaker/box combination, while the other is realized in the electrical filter. The design of both the box transfer function and the electrical filter will need to accurately implement the overall desired response. The two possibilities of loudspeaker driver/box response are then realized with very different alignments of driver Q_T and box volume. This can be advantageous where there are constraints on box size or available driver parameters, for example.

If the loudspeaker is in a vented box, the overall 6th order transfer function can be broken into three 2nd order factors. Any one of these can now be allotted to the electrical filter and the other two factors multiplied together for the response of the loudspeaker driver/box combination, which now allow three sets (alignments) of parameters that are very different, especially in respect of box volume and driver Q_T.

In principle, of course, the filter need not be active. Procedures have been developed *[8,9]* for passive filtering of loudspeakers. Their components however, particularly the inductors, while eminently practical for tweeters filtered from 2000 Hz or higher more, become expensive, cumbersome and prone to distortion in applications to filtering below 100 Hz, as previously identified.

Equalizing a Loudspeaker.

Once we start using an active filter to shape the cut-off characteristic of a loudspeaker, it is only a small step further to consider equalization, using active filters to correct deficiencies in response at low frequencies generally. The transfer function F(s) of a high pass filter may be generalized as

$$F(s) = \frac{s^2 T_d^2}{1+sx_d T_d + s^2 T_d^2} \qquad - \quad (2)$$

where x_d is a "shape" parameter, sometimes written as $1/Q_d$, that determines the shape of its response around its characteristic frequency f_o , and $T_d = 1/\omega_o = 1/2\pi f_o$. If now we have a device with a 2nd order transfer function

$$F(s) = \frac{s^2 T_n^2}{1+sx_n T_n + s^2 T_n^2} \qquad - \quad (3)$$

such as a closed-box loudspeaker, or perhaps one 2nd order factor of a vented loudspeaker's 4th order transfer function, and would like to change it to that of eqn (2), all we have to do is cascade with it the equalizing biquadratic function whose 2nd order denominator is the wanted transfer function and its 2nd order numerator is the denominator of the unwanted transfer function

$$F(s) = \frac{1+sx_n T_n + s^2 T_n^2}{1+sx_d T_d + s^2 T_d^2} * \frac{T_d^2}{T_n^2} \qquad - \quad (4)$$

where T_d^2/T_n^2 is simply a numerical gain figure.

Such a response is readily realized using a state variable biquadratic filter, which can implement any 2nd order biquadratic transfer function *[10]*. In Fig 13, the first two op. amps. are connected as integrators, with capacitors between their output and negative input, the first leaky, with resistance shunting its capacitance. The third stage is a unity gain phase inverter. Together these three stages determine the denominator of the transfer function. Signals from the input to the filter and the outputs of the first three stages are combined in a 4th adder stage.

The coefficients of the numerator of its transfer function are determined by the proportions of the components feeding the adder, or by their absence. Then the output function is high-pass, low-pass, all-pass or the more general quadratic function that we need for equalization. depending on which of those coefficients is positive, negative or zero, The apparent complexity of the filter is due to its need for about ten resistors, depending on its use, but otherwise it needs only two capacitors and four op. amps. that can be had in a quad package.

This device can realize any kind of biquadratic function. Nevertheless, equalizers suitable for a number of, but by no means all, equalizers for loudspeakers can also be realized by a Sallen and Key filter incorporating a bridged-T network as in Fig 14. It can only be used when, in eqn (4)

$$\frac{x_n}{x_d} > \frac{T_n}{T_d} \qquad (5)$$

but this limits its application less than might at first appear. It still needs two capacitors but only one op. amp. and four resistors (*or a digital implementation – ed*).

The procedure can be illustrated by the example of Fig. 15, which was realized using the biquad equalizer of Fig 14. A closed-back loudspeaker with a system Q of 1.33 (and thus a 2nd order transfer function in which

$x_N = 1/Q = 0.75$) and a cut-off frequency f_O of 100 Hz (so that $T_O = 1/2\pi f_O = 1592$ µs) is to be equalized to a 3rd order Butterworth response with a cut-off frequency of 50 Hz ($T_O = 3183$ µs). The overall equalizer response is produced by cascading a 2nd order response with an x_d of 1, realized by the biquad of Fig. 14, with the 1st order response of a simple CR network whose CR product is 3183 µs. The two response components, shown separately in Fig. 16, illustrate the considerable increase in level at low frequencies in the active stage and hence the need to place the circuit element realizing the peaky response *later* in the chain than the drooping CR response to avoid overload in the equalizing stage.

Figs. 17 and 18 illustrate the general principle that sharp cut-off responses are always implemented by cascading peaky component responses with gently sloping, e.g. CR, component responses.

In Fig 17, which shows the components of a 3rd order high-pass Butterworth response that is maximally flat in its pass-band, the peak is only 1.3 dB high, but to produce a 3rd order Chebyshev response that ripples between 0.0 dB and 0.5 dB, as in Fig 18, the response peaks 5 dB high, and for greater magnitudes of ripple and/or higher orders the peaks of component responses go even higher.

Loudspeaker equalizers can also be implemented digitally of course, an especially convenient option when the input signal is already presented digitally.

In planning the equalization of a loudspeaker, particularly for extending its low frequency response, it is important not to try to achieve too much just “because it can be done”. The limitations that can arise from excessive increases of driver excursion, amplifier power and group delay must always be kept in mind, also the extent of drift with temperature and time between the responses of the system being equalized and its equalizer.

Group Delay Error.

Discussions about the use of higher order transfer functions to handle, or control, the low frequency response of a loudspeaker inevitably lead to the topic of group delay error. We should first make it clear that when sound propagates from a transmitter, there will always be a delay before it reaches the receiver because of the finite velocity of sound propagation in air.

The sound quality is only impaired when what we will call a “group delay error” occurs, when some frequency components are delayed differently from others and arrive at different times. It has been said that while a small group delay error, when some components of a signal arrive a little before others, is inaudible, whereas if some components arrive today and others tomorrow, it is clearly distorted. (If all the signals are equally delayed for example by starting the piece later, it is not perceived as distorted)

All filters produce some group delay error, however small, and any increase in the order of a filter or the steepness of its attenuation will surely increase its group delay error, see Fig 19. Some listeners aver that they can certainly hear the group delay error produced by a vented box and insist on using a closed box. Whether that is true when the frequency response is substantially flat within the pass-band before it falls away in the stop-band remains a matter of debate.

In their classic reference, Blauert and Laws *[11]* measured thresholds of perception of group delay as going from a minimum of 1 ms around 2000 Hz to 2.0 ms at 8000 Hz and 3.2 ms at 500 Hz. We may surmise that the threshold at frequencies of 50 Hz and lower will be rather more than 3.2 ms, but by how much?

It is very difficult, if not impossible, to measure group delay at very low frequencies under practical acoustical conditions. Equipment is, or used to be, available for measuring group delay at frequencies down to 300 Hz, which was the lower frequency limit for analogue telephony circuits. That equipment modulated sine waves at frequencies across the range under investigation with a much smaller fixed frequency ω_M, measured the phase difference $\Delta\beta$ between the two sidebands of the modulated signal $2\omega_M$ apart, and calculated the group delay $d\beta/d\omega$ as the ratio of two small quantities $\Delta\beta/\Delta\omega = \Delta\beta/2\omega_M$.

A similar facility is in fact available from equipments that test using a pulse input with the system output pulse analysed by the Fast Fourier Transform. Such a scheme can read amplitude and phase at fixed frequency intervals Δf determined by the sampling rate divided by the number of points sampled. $\Delta\beta$ is found as the difference in the phase measured at two adjacent frequencies that are Δf apart.

However, measurement of group delay T_G down to the lowest frequencies, 10 Hz or 20Hz, to a precision δT_G of 1ms with a modulation frequency Δf of 2Hz, i.e. $\Delta\omega$ of 12.6 radians/sec, would need to read a phase angle $\Delta\beta$ to the high precision of $2\Delta\omega\cdot\delta T_G$ i.e. 0.025 radians or 1.4°. The experience of a trusted colleague is that while such a scheme had worked well in simulations, his tests of group delay in practical systems at the lowest audio frequencies had proven unsatisfactory due to high levels of background noise.

Plots of group delay in Fig 19 calculated for Butterworth responses of 2nd, 4th and 6th order, comparing cutoff frequencies of 50 Hz and 25 Hz, demonstrate that a 6th order 25 Hz filter produces a maximum delay of more than 40 ms around 25 Hz. By comparison a 2nd order 25 Hz filter produces little more than 10 ms maximum, which might give some comfort to closed box enthusiasts. However, Fig 19 illustrates altogether that-

(i). group delay increases as the order of the response, and thus the slope of out-of-band attenuation, increases

(ii). group delay decreases rapidly at higher frequencies with the inverse square of frequency, and

(iii). while lowering the cutoff frequency produces a *greater* maximum delay at the lowest frequencies, it produces *less* group delay at higher frequencies

But we still await a definitive figure for the threshold of perception of group delay at the lowest audio frequencies.

Conclusion.

The paper has tried to elucidate some of the problems of handling audio power generally and particularly at low frequencies; some methods of solving these problems and the new problems that may arise from previous solutions. Some aspects will surely be known to the expert reader but it is hoped that a sufficient quantity of the material is novel enough to have made the effort worthwhile.

Acknowledgement.

The author gratefully acknowledges the helpful comments of Graeme Huon in editing the paper.

References.

1. Publication 268-3, Sound System Equipment – Amplifers, International Electrotechnical Commission, Geneva, 1988 – Section 18.5, Overload Restoring Time

2. A.N.Thiele - Recovery of Amplifiers after Overload - Proc. IRE Australia, 20, 5, 1959 May, pp. 257-261

3. I.H. Shearman – Assessment of Loudspeaker Quality – Proc, IREE Australia, Vol. 31 No.6, 1970 June, pp.165-176

4.. A.N.Thiele - Loudspeakers in Vented Boxes - Proc. IRE Aust, Vol. 22, No. 8, 1961August, pp. 487 – 508

& J Audio Eng Soc, Vol. 19, Nos. 5 & 6, 1971 May & June, pp. 382 - 392 & 471 - 483.
[eqn (1) of the present paper is adapted to m.k.s.units from eqn (81) of ref. 4]

5.. Neville Thiele - Closed Box Loudspeaker with a Series Capacitor – J Audio Eng Soc, 58, 7/8, 2010 July/August, pp 577-582, also presented at ISEAT2011, Shenzhen, China, 2011 November 12-13

6.. A.N.Thiele – Loudspeakers Enclosures and Equalisers – Proc IREE Australia, 34, 11, 1973 November, pp. 425-448

7.. R.P. Sallen & E.L.Key – A Practical Method of Designing RC Active Filters – IRE Trans. Circuit Theory, CT 2(1), 1955, pp. 74-85

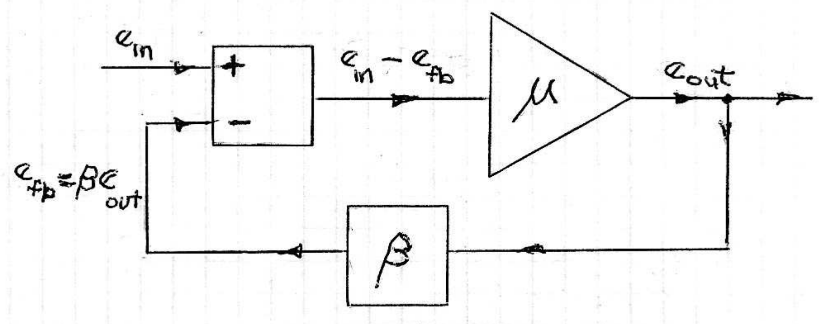
8.. Neville Thiele - An Inexpensive Precise Passive Crossover System – Proc. AES 21st International Conference, St. Petersburg, Russia, 2002 June 1-3, pp 257-267.

9.. Neville Thiele - Passive All-Pass Crossover System of Order 3 (Low-Pass) + 5 (High-Pass), Incorporating Driver Parameters - J Audio Eng Soc, 50, 12, 2002 December, pp. 1030-1038

10.. Gobind Daryanani – Principles of Active Network Synthesis and Design - John Wiley and Sons, New York, 1976, pp. 343-350, 363

11.. J.Bauert & P Laws – Group Delay Distortion in Electroacoustical Systems – JASA, 63, 5, 1978 May, pp. 1478-1483

-o0o-



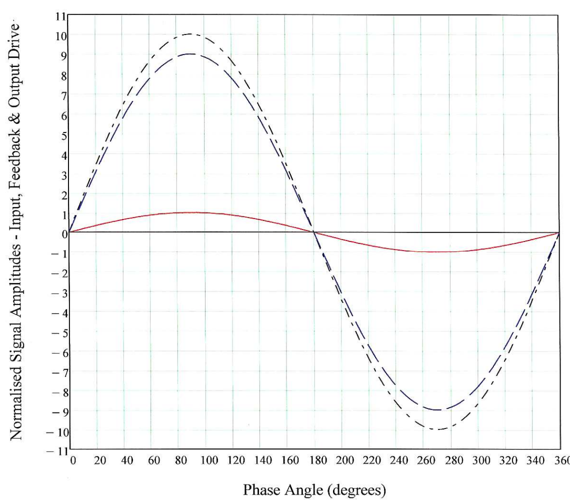
$$Feedback\ Amplifier\ Overall\ Gain = \frac{\mu}{1+\mu\beta}$$

$\mu = amplifier\ gain$

$\beta = feedback\ ratio\ (<1)$

$1+\mu\beta = feedback\ gain\ reduction\ factor$

Fig 1. Schematic Diagram of a Generalized Amplifier with Negative Feedback

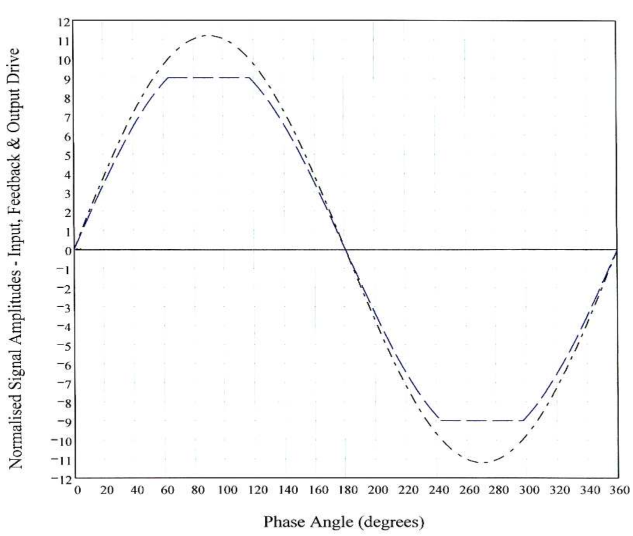


Maximum Output - Linear Operation

Input signal (dash-dot curve) - 10 units peak : Feedback signal (dashed curve) - 9 units peak

Input signal, after subtraction of feedback (solid curve) - 1 unit peak

Fig 2. Signals within a Feedback Amplifier with 20dB (10 times) Gain Reduction by feedback

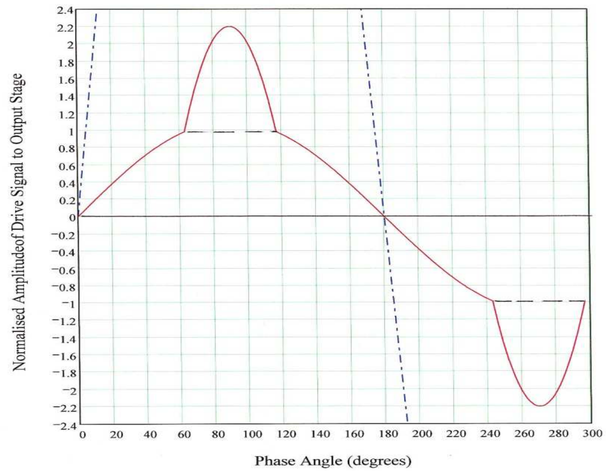


Input Signal 1.0 dB above maximum output (clipping)

Input signal (dash-dot curve) -11.22 units peak : Feedback signal (dashed curve) – clipped at 9 units peak

Hence Input signal, after subtraction of feedback (solid curve) rises during clipping to 2.2 units peak

Fig 3. Signals within a Feedback Amplifier with 20dB (10 times) Feedback Gain Reduction



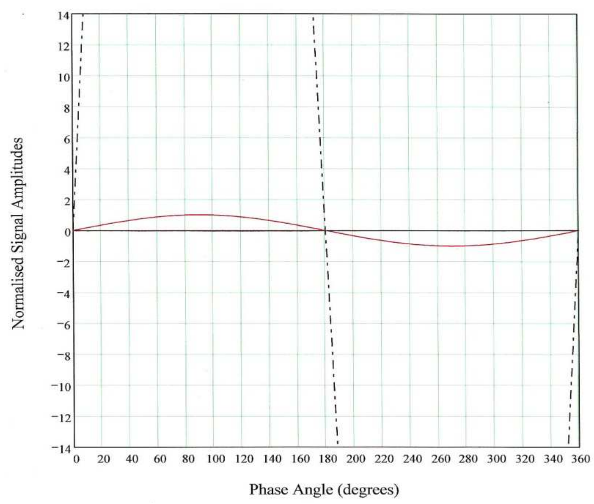
20dB (10 times) Feedback Gain Reduction

Input signal (dash-dot curve) -11.22 units peak : Feedback signal is clipped at 9 units peak

Hence Input signal, after subtraction of feedback (solid curve) rises during clipping to 2.2 units peak

Output Clipping Level (dashed curve) - 1 unit : Note: Vertical scale magnified from Fig 3

Fig 4. Signals within a Feedback Amplifier with input signal 1.0 dB above maximum

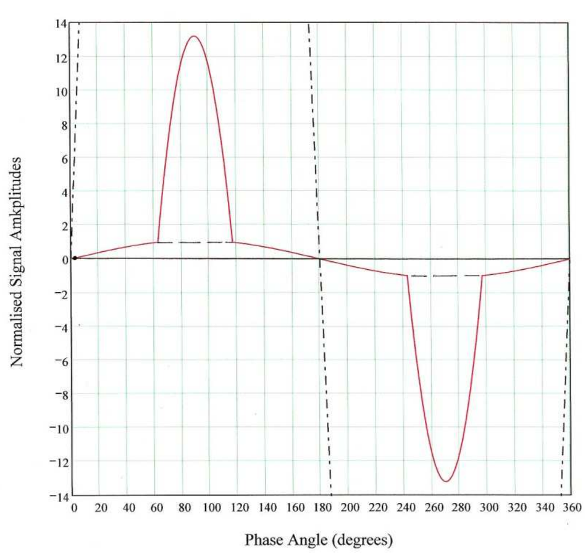


Maximum output – Linear Operation

Input signal (dash-dot curve) -100 units peak : Feedback signal (dashed curve) - 99 units peak

Input signal, after subtraction of feedback (solid curve) - 1 unit peak

Fig 5. Signals within a linear operating feedback Amplifier - 40dB (100 times) feedback gain reduction



Input signal (dash-dot curve) - 112.2 units peak : Feedback signal is clipped at 9 units peak

Hence Input signal, after subtraction of feedback (solid curve) rises during clipping to 13.2 units peak

Output Clipping Level (dashed curve) - 1 unit

Fig 6. Signals possible within a Feedback Amplifier with 40dB (100 times) Feedback Gain Reduction and input signal 1.0 dB above maximum

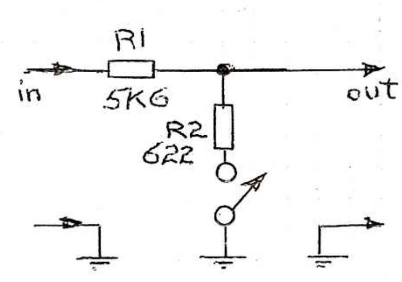
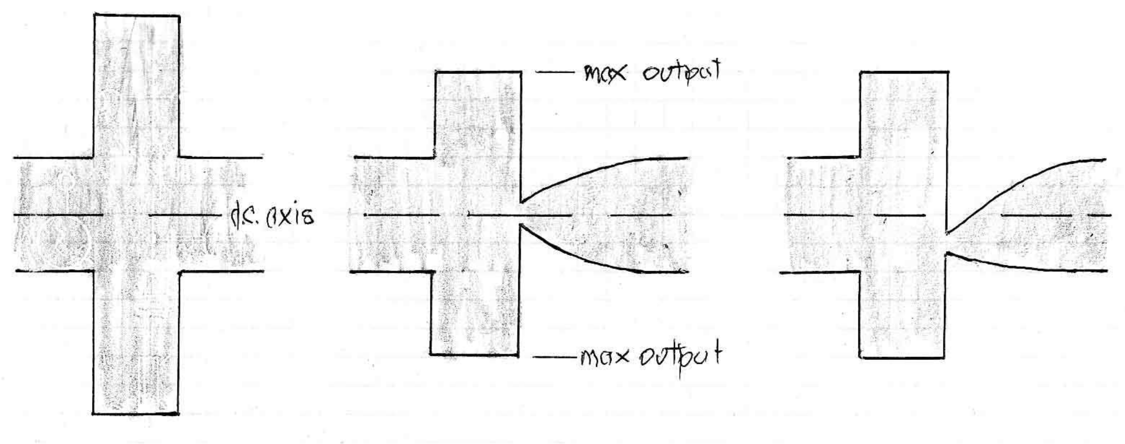


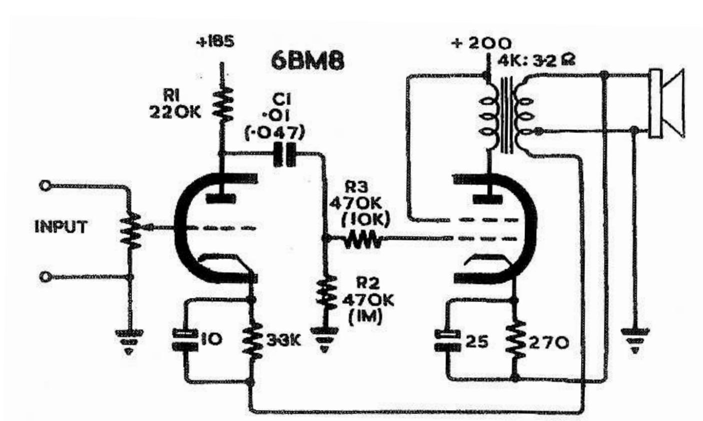
Fig 7. Switched 20 dB Attenuator



Input *Output*

Without d.c. axis shift With d.c.axis shift

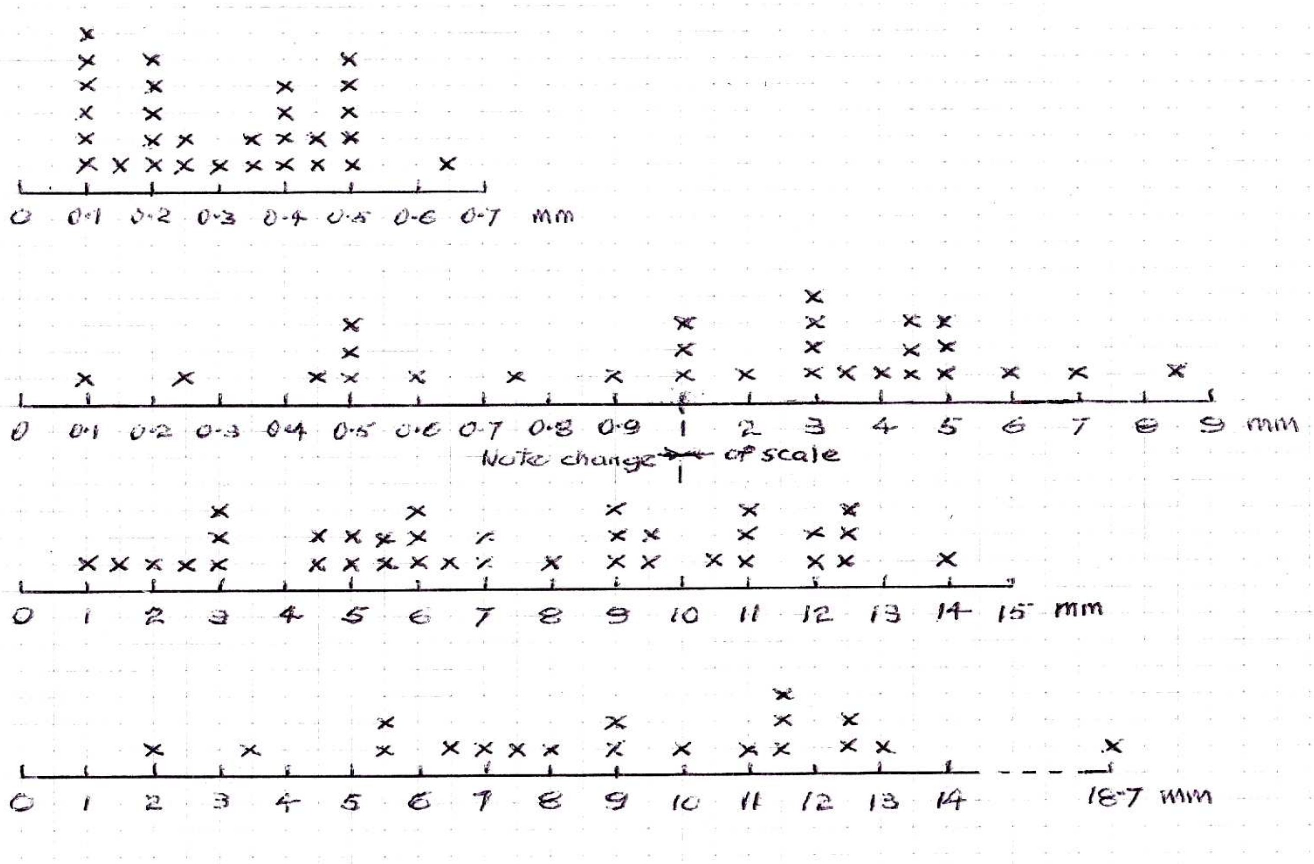
Fig 8. Envelopes of. Signals at Slow Scan Rates (e.g.1 cm/sec) in Tests of Amplifier Recovery after Overload



Original component values of C1, R2 & R3 that produced paralysis shown in brackets.

Component values that minimized paralysis un-bracketed.

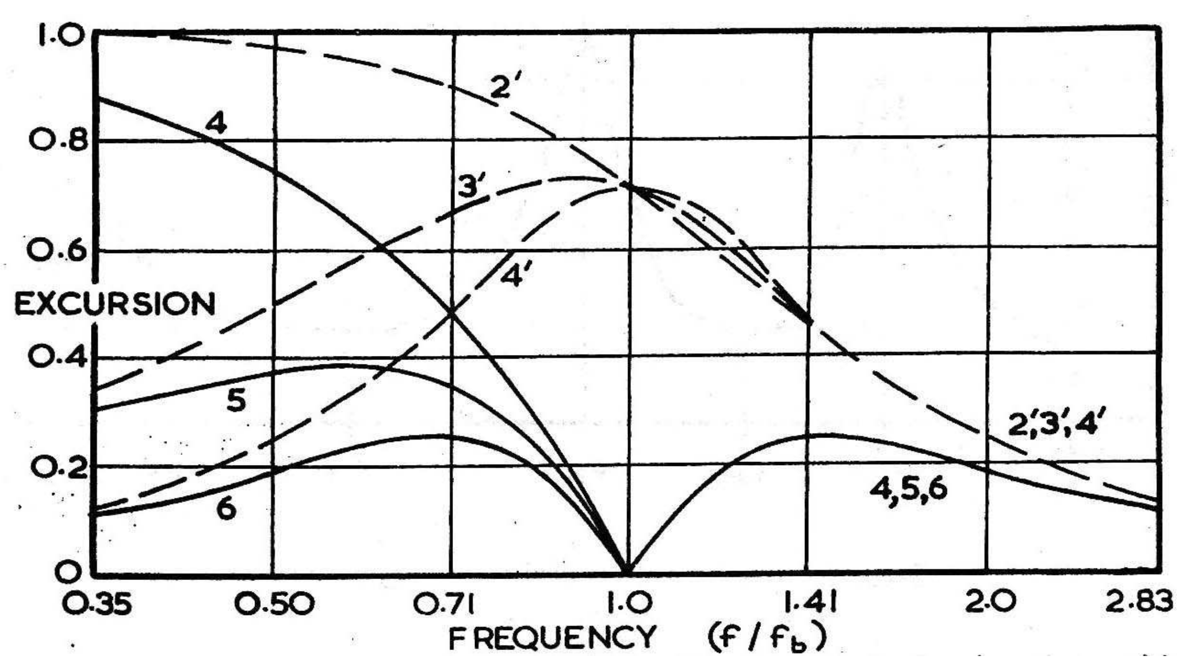
Fig 9. Circuit Schematic of Valve Amplifier (from the 1959 publication).



Line 1: 25 mm tweeters
Line 2: 65–150 mm.diameter
Line 3: 150–250 mm.diameter:
Line 4: 250–400 mm.diameter

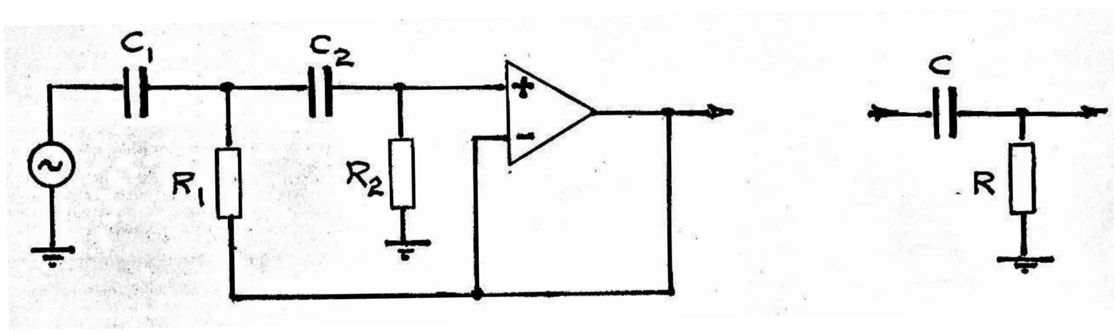
*Compiled from HOBBY HiFi, Klebe, Germany, Ausgabe 1-6/2011

Fig 10. Summary of Specifications for Maximum Excursion x_{max} of 111 Drivers grouped according to nominal cone diameters*



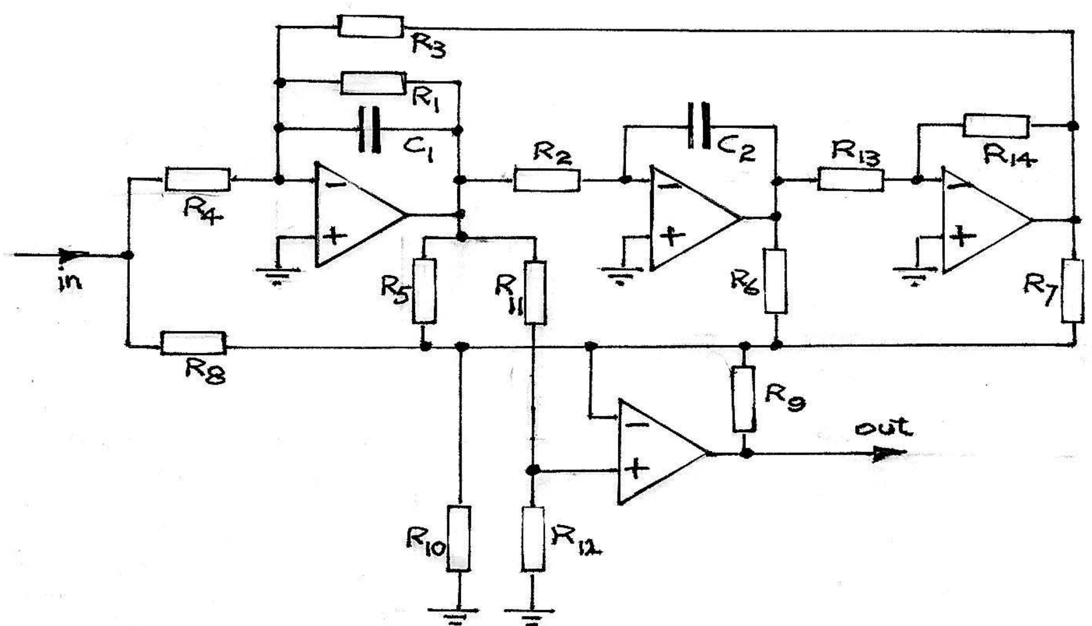
With closed box, orders 2', 3' & 4' (dashed curves): With vented box, orders 4, 5 & 6 (solid curves)
Curves 2' & 4 – without electrical filter: Curves 3', 4', 5 & 6 – with electrical filter in amplifier
Excursion of 1.0 is arbitrarily taken as low frequency compliance limit

Fig 11. Cone Excursions vs. Frequency, Butterworth Responses into room



Sallen & Key Active 2nd order high-pass filter 1st order high-pass filter

Fig 12. Active High-Pass Filters Cascaded with Amplifier Input



All possible resistors are shown. Some resistors, from amongst R5, R6, R7, R10, R11 & R12, will always be omitted, according to wether the required response is 2nd order high-pass, low-pass, notch, all-pass or bi-quadratic.

Fig 13. State Variable Bi-Quadratic Filter (or Equalizer)

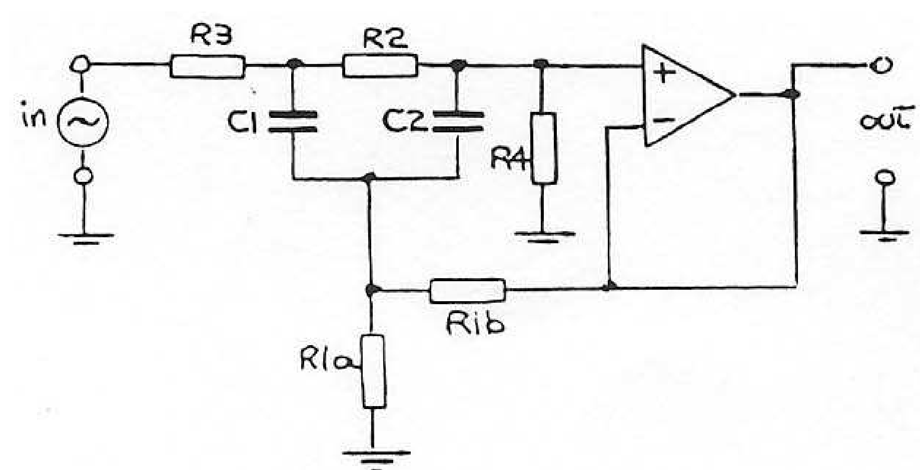
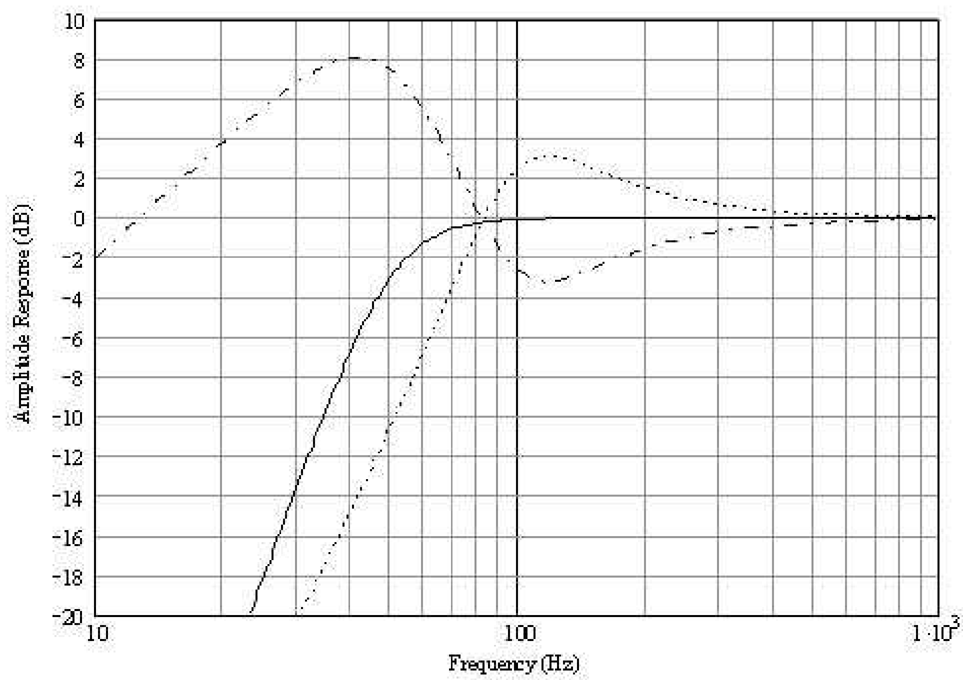


Fig 14 Active Biquadratic Filter Using a Bridged-T Network, Sallen & Key (unity gain) Configuration



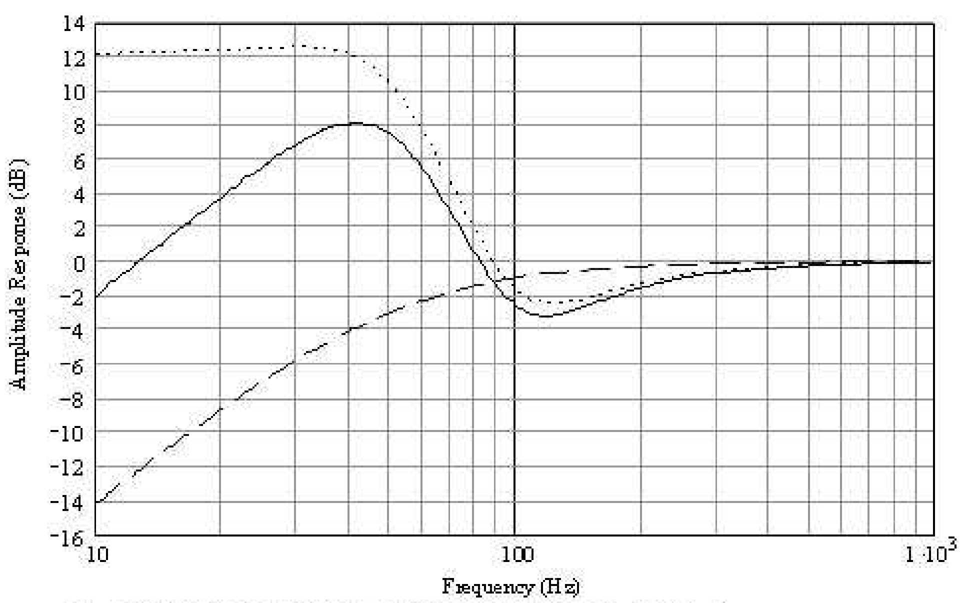
Qt = 1.33 (x = 0.75) and Fo = 100 Hz, equalized to a 3rd order Butterworth (maximum bandwith) Response into air, with a 2nd order factor, x = 1 and f_O = 50 Hz, and a 1st order factor at 50 Hz.

Initial Loudspeaker 2nd order response, x = 0.75 and f_O = 100 Hz (dotted curve)

Equalizer response – biquadratic stage plus 1st order network with CR product of 3183 µs (dash-dot curve)

Combined Equalized response (solid curve)

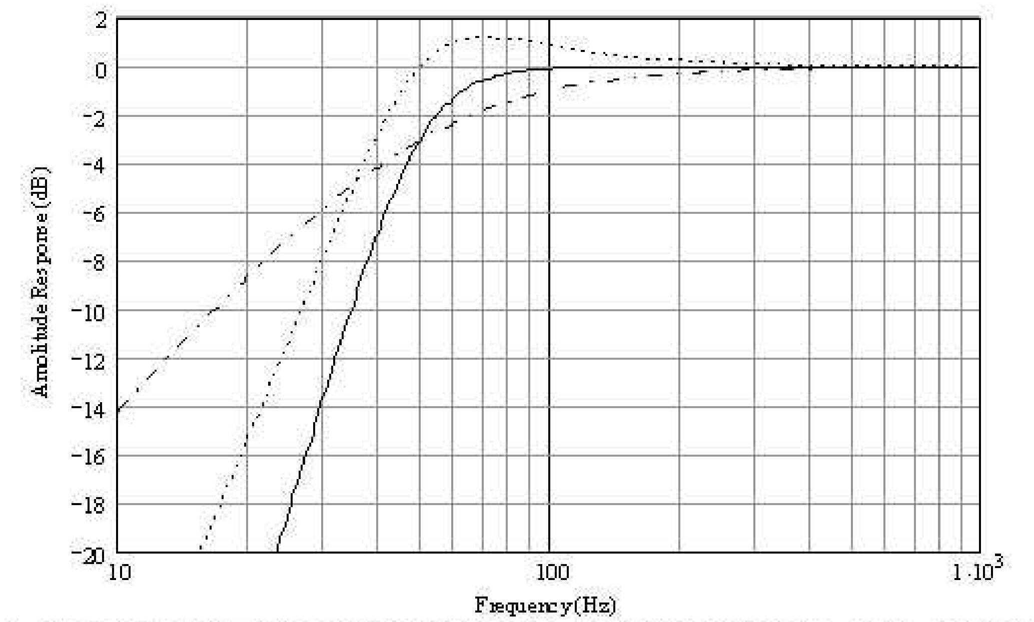
Fig 15 Equalization of a Closed-Box Loudspeaker



Overall equalizing response (dash-dot curve in Fig 15) – solid curve

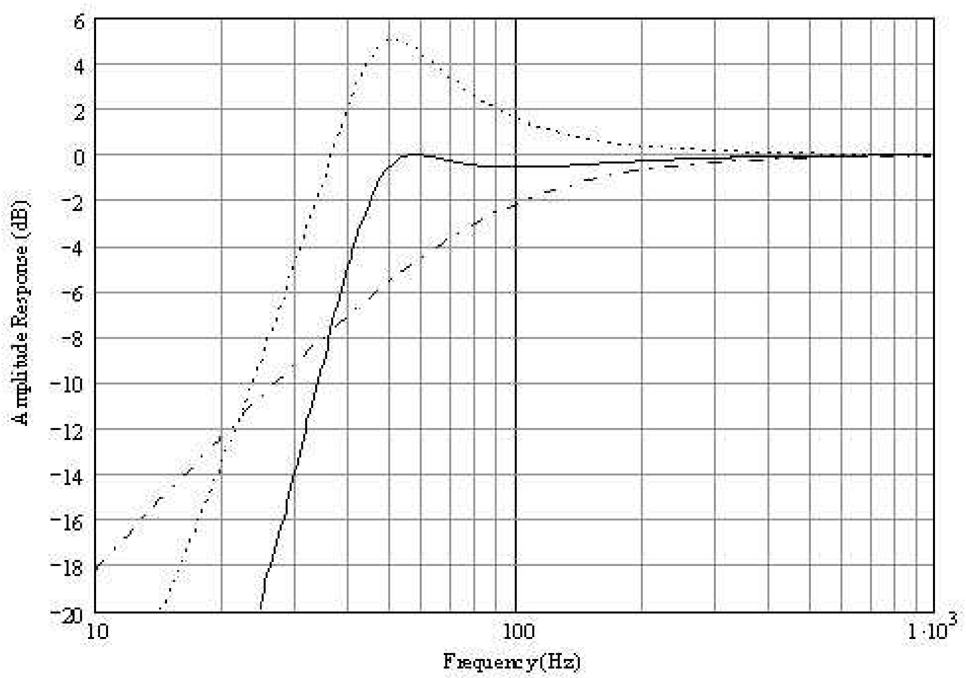
1st order CR Network – dashed curve: biquadratic active equalizer – dotted curve

Fig 16. Factors of Equalizing Response



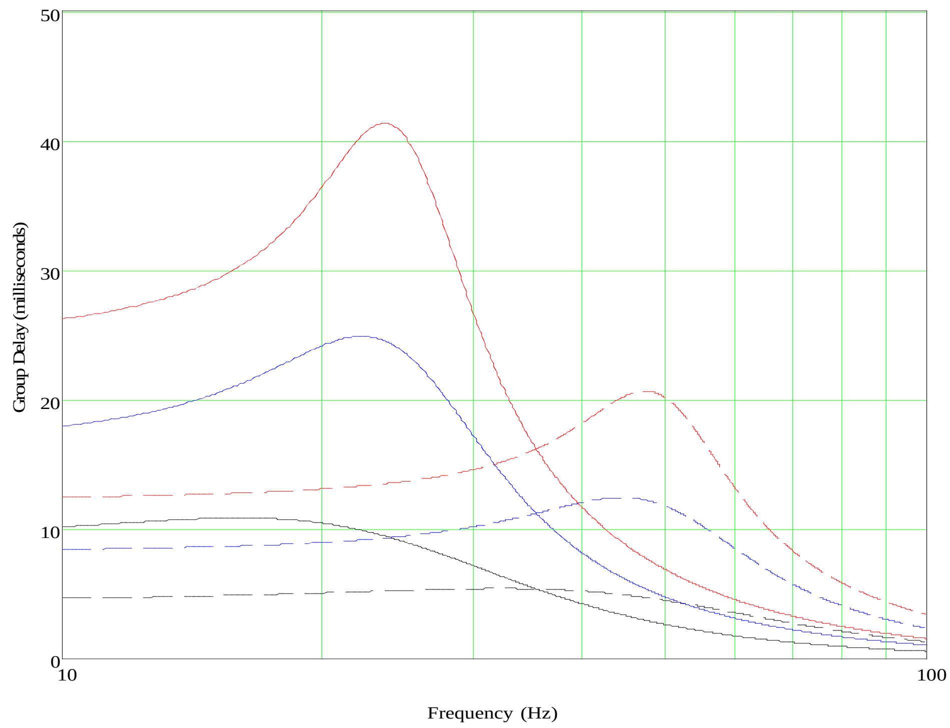
f_O = 50 Hz, 1st order factor, T_O = 3183 µs – dash-dot curve:
2nd order factor, x = 1.000, T_O = 3183 µs – dotted curve
Overall 3rd order response – solid curve

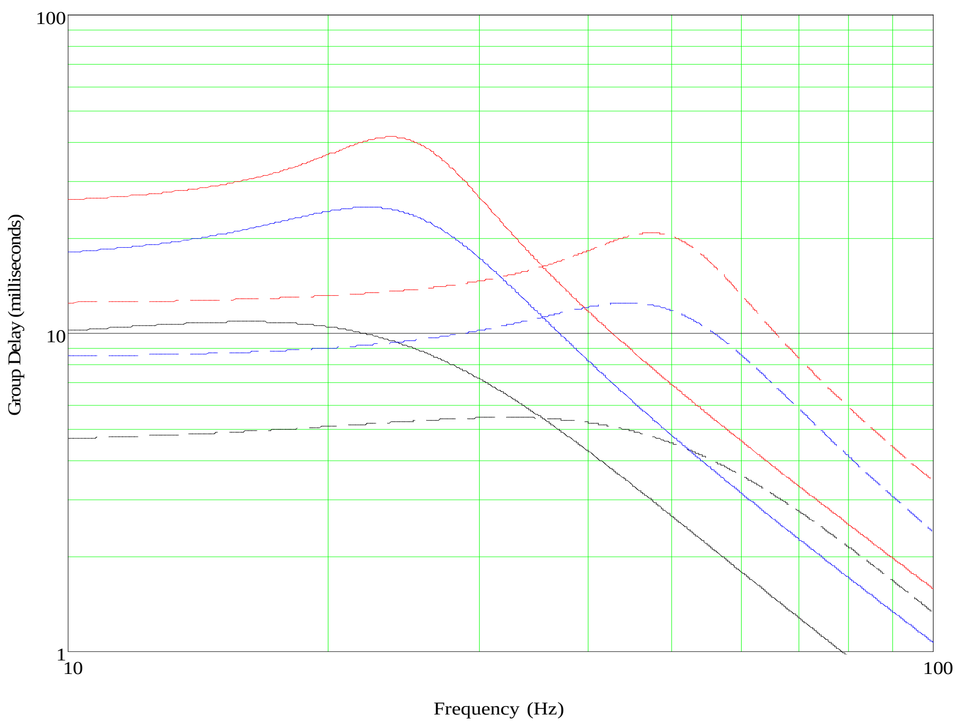
Fig 17 Factors of 3rd Order Butterworth (Maximally Flat) High-Pass Frequency Response



1st order factor CR = T_O = 1993 µs – dash-dot curve
2nd order factor, x = 0.5869, T_O = 3402 µs – dotted curve
Overall 3rd order response – solid curve

Fig 18. 3rd order Chebyshev (equal-ripple) Response (within 0.5 dB to 50 Hz) and Its Factors





Cutoff Frequencies, 25 Hz and 50 Hz

Solid curves: f_O = 50 Hz dashed curves: f_O = 25 Hz

6th order – uppermost curves: 4th order – middle curves: 2nd order – lowest curves

Upper diagram 19a – Linear plot of group delay: lower diagram 19b - Logarithmic plot of group delay

Figs. 19a, 19b. Group Delay of Butterworth Filters of 2nd, 4th and 6th Order