Delay alignment of top- and sub loudspeaker systems; a survey of new and old methods.

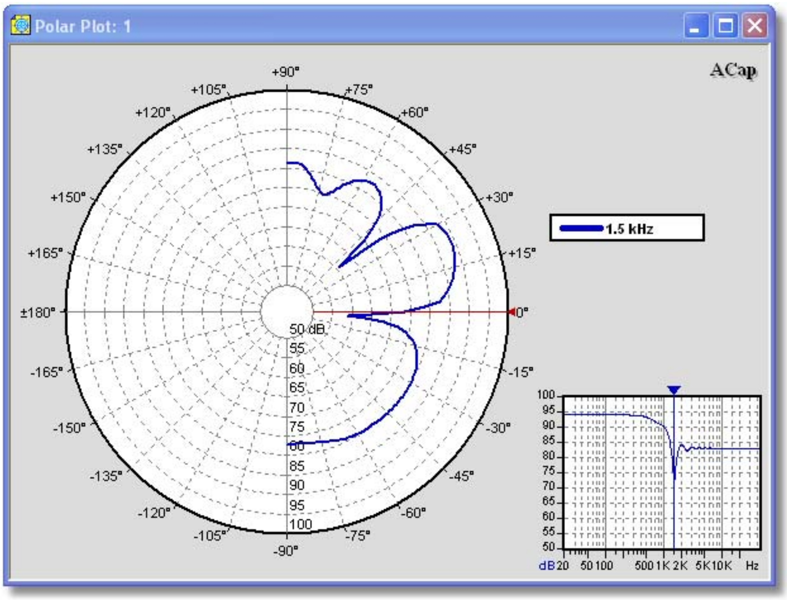
Dan Bävholm*
Johny Grenander**
*WaveCapture, db@wavecapture.com
**WaveCapture, Johny@wavecapture.com

A couple of popular methods are presented with their benefits and drawbacks. Commonly used methods are using wrapped phase and impulse response. With real time FFT analysis, magnitude and time domain can be analyzed simultaneously. Filtered impulse response and Cepstrum analysis are helpful tools when the spectral content differs and make it hard to analyse the impulse response.
To make a successful time alignment the measurements must be anechoic. Methods such as multiple time windowing and averaging in frequency domain are presented. Group-delay and wavelets analysis are used to evaluate the measurements.

Introduction

In the first section of this presentation, the problem will be defined. Why it is so important to align the delay between different drivers in a loudspeaker system?
Misalignment will cause so-called time smearing; where the sound from different bands doesn't arrive at the same time. This might be inaudible for continuous waves, but it will cause transient distortion. One can imagine if the top system is placed 340m behind the sub-bass system, what a snare drum will sound like. The transient will arrive one second after the boom from the drum.
Do we really hear time and phase distortion?
This controversy dates back to 160 years ago, when Ohm [1] formulated his "phase law". According to that law, only the power spectrum will determine the sound's characteristic. This implies that our human auditory system should work similarly to a real time analyzer (RTA). Both Helmholtz [2] and Rayleigh [3] had later doubts about the validity when it came to transients.

Another issue to bring up concerning time misalignment is uncontrolled behaviour in the directivity of a loudspeaker system. If the drivers are placed non-coaxial and one band is misaligned in time there is a risk that the lobe will divert in an unwanted direction in the crossover region [5, D'Appolito]. In the crossover region the magnitude will be close to equal between the drivers. They will interact if their directivity lobs are subsets of each other. This is actually used for beam steering. Note that highly directive drivers, such as high-frequency drivers with waveguides for line-arrays, don't interact [6, Meyer].



A problem occurs if the directivity lobe directs towards the ceiling in the crossover region. This can cause reflections with worse speech intelligibility as a result. The on-axis frequency response will then show a dip in the crossover region.

Fig 1. Directivity lobe can divert towards ceiling in the crossover region

The last issue to bring up concerning time misalignment is that only a small discontinuity in the phase curve (or a shift in group-delay) will cause the waveform of for example a square-wave to alter its crest factor. Transients can be enhanced and cause overloading of the amplitude range.
In these two figures 2 & 3, a square wave is built up by four odd harmonics according to Fourier [4].

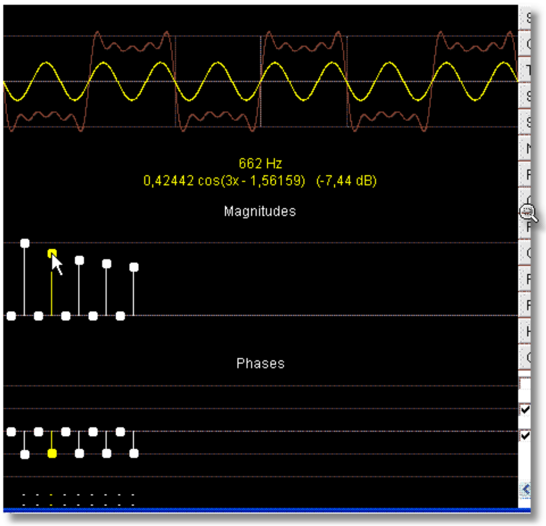


Fig2. All harmonics are in phase

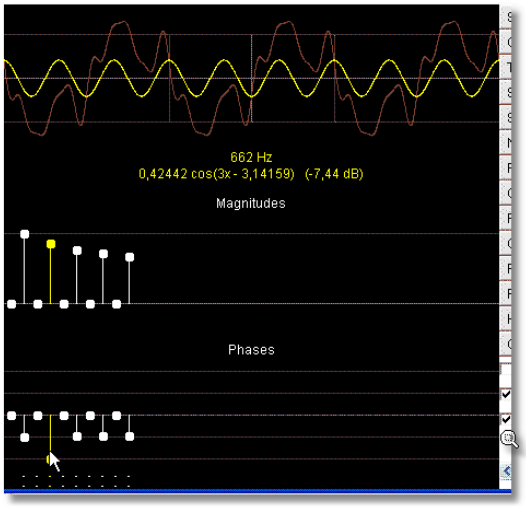


Fig 32. 3rd harmonic with misaligned phase

Note that a small phase shift at the 3rd harmonic will distort the amplitude so the square wave is hardly recognisable.

Perceptual tolerances of group-delay distortion

Many psychoacoustic studies have been done [7] – [10] of the audibility of group-delay distortion. First out are a row of German studies of Zwicker & Feldtkeller (1967) and Fleischer (1975 & 1977). Most well known is however Blauert & Laws 1978. That dependency of test signal, the use of loudspeaker or headphones and the training of the test persons can explain the differences in the results.
A good tutorial of phase distortion is found in [11, Preis].

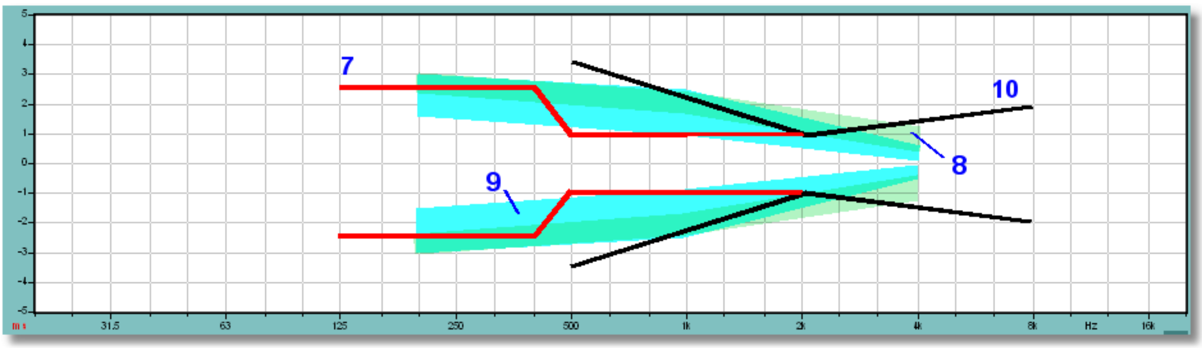


Fig 4. Perceptual thresholds for group-delay distortion from a number of investigations.
[7] Zwicker & Feldtkeller, 1967
[8] Fleischer 1977
[9] Fleischer 1975
[10] Blauert & Laws, 1978

Most results indicates a couple of periods misalignment tolerance in the vocal band between 300 – 2000Hz

It must be pointed out that it isn't the actual phase shift as a function of frequency or phase delay or group delay itself that matters [12, Liphitz et al]. It is rather how much the phase shift or group delay differs from a pure time delay.
An example of misuse can occur when comparing a digital signal processor to a similar analogue one. The digital unit has a latency of 1ms, corresponding to -7200 degrees at 20kHz and the analogue has –45 degrees. Both have a bandwidth to 45kHz. Which is the best?

We can't judge this because the phase shift is a pure delay, which didn't cause any phase distortion. Maybe the digital unit is better, because the lack of high-pass filter in the circuits, which can cause extra group-delay at the lowest frequencies in the analogue unit. We can't judge this from the phase quantity itself, but from the group-delay curve or the phase deviation from pure delay.

One of the measures is phase delay:

$$t_{\varphi} = \frac{-\varphi}{2\pi f} \text{ , s}$$

The phase is here in radians. The phase delay indicates the number of covered turns per frequency.

And the other is group delay:

$$\tau g = -\frac{\partial \phi}{\partial \omega} \text{ , s}$$

Group delay is the local deviation of the phase. This means that irregularities in the phase curve will be visible in the group-delay versus frequency curve.

The human auditory system

Recent research about the human hearing actually confirms that the auditory system is time sensitive. It isn't only binaural hearing (two ears) that makes it possible to determine direction of sound, but the cochlea's ability to detect travelling waves. Gone is the assumption that the cochlea acts a power spectrum detector.

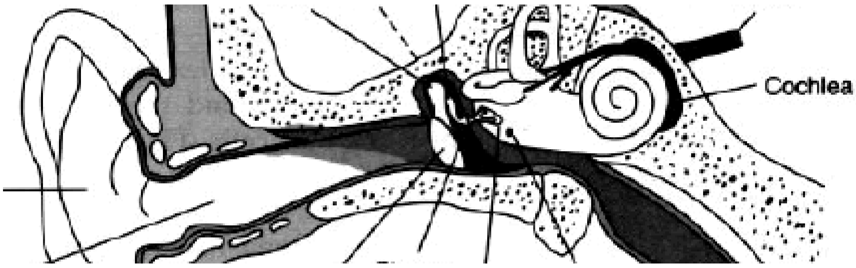


Fig 5. The human Auditory system

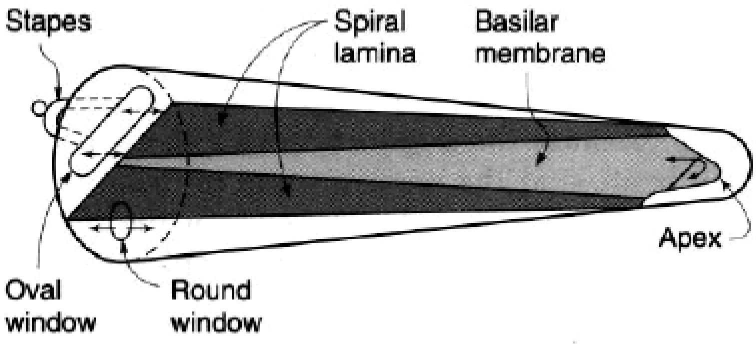


Fig 6. The snail shaped structure can be unrolled to be more apparent

If Helmholtz's resonance theory is applied to the Basilar membrane, a travelling wave can be detected [13].

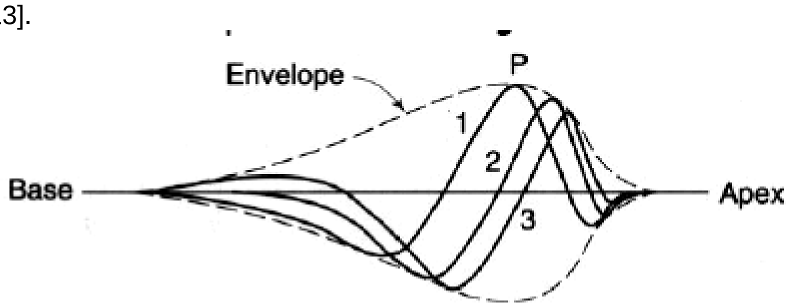


Fig 7. Bekeksy's travelling wave

The tactile organs in the cochlea will detect magnitude and time position of the travelling wave. From this the frequency will also be detected, as low frequencies will have its peak near the apex. Note that harmonics will be masked by the travelling wave. (A reason why we can accept compressed audio, such as mp3 etc.)

How we can visualize time smearing

Delay misalignment will cause so-called time smearing; which means that the sound from different bands doesn't arrive at the same time. We have seen that the group-delay versus frequency (group-delay curve) is a powerful tool to judge delay misalignment. An even more powerful graph is the Wavelet analysis plotted in a two- or three-dimensional view, called Scalogram. Scalogram is an interpretation of the signal energy distribution. It is displayed as a two-dimensional intensity plot, where the x-axis shows the time, the y-axis shows the frequency and the colour scale shows the magnitude in decibel.
Wavelet was investigated by a group of French mathematicians and others at the end of 80's [14] – [18].

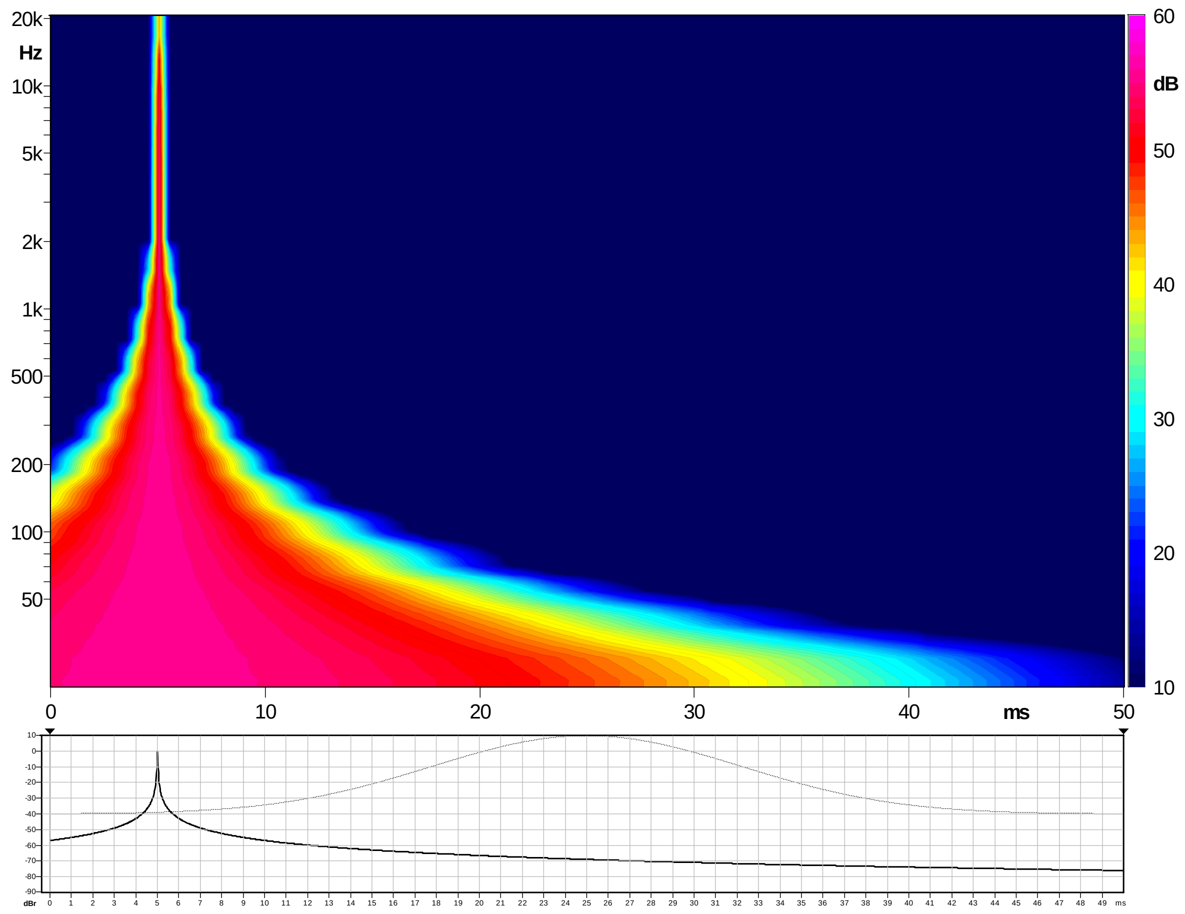


Fig 8 . A Wavelet analyse of a perfect impulse (Dirac) in a Scalogram. Note, this is the desired result if a linear phase system is the target.

Wavelets offer the same frequency resolution, i.e. constant relative bandwidth, but increased time resolution at higher frequencies.
Wavelets are considered as more hearing-like frequency-time analysis than traditional Spectrogram. (Compare to the perceptual threshold of group-delay distortion above).

Wavelets have constant relative bandwidth and therefore the time resolution will be higher at higher frequencies. This is the reason why wavelet Scalograms show a “standing trumpet”, see picture of a perfect impulse at 10ms.

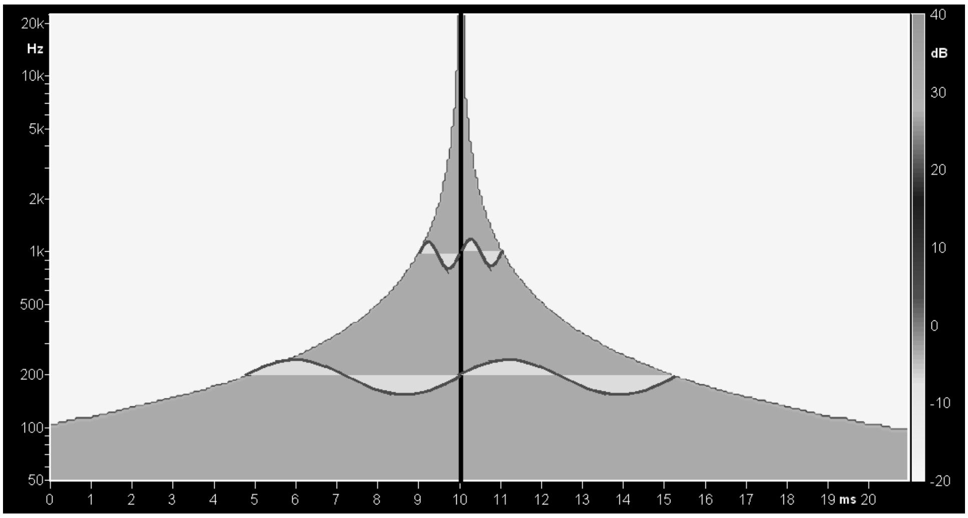


Fig 9. Gaussian window requires two periods

It is the type and size of the time window that determines the shape of the trumpet. A minimum of one period is needed to determine the magnitude of a wave, if the window is rectangular. Therefore the maximum time resolution for a rectangular window can’t be less than one period at each frequency of interest. However rectangular windows gives some artefacts such as spectral leakage. A Gaussian filter on the other hand is free of spectral leakage and will provide an undistorted time display. Another reason to use it is that a Gaussian time window will also have a Gaussian shape in the frequency domain.
Gaussian windows require two periods to determine the magnitude due to the flat bell. This is shown in the picture as the sine waves at 200Hz and 1kHz. (This is due to the uncertainty principle saying that $\Delta t \cdot \Delta\omega = 1/2$, which also can be written $2 \cdot \Delta t = 1/\Delta\omega$, i.e. two periods are needed for each frequency period).

Wavelet analysis and Scalogram is also very useful for analyzing acoustics. The energy distribution over time due to reflections from the room boundaries is clearly displayed.

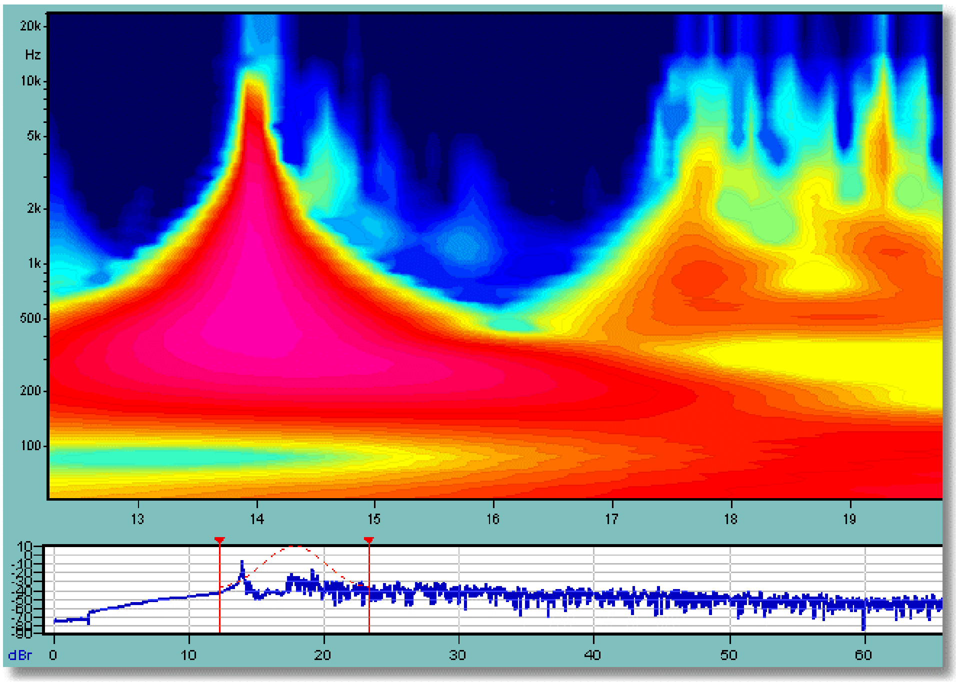
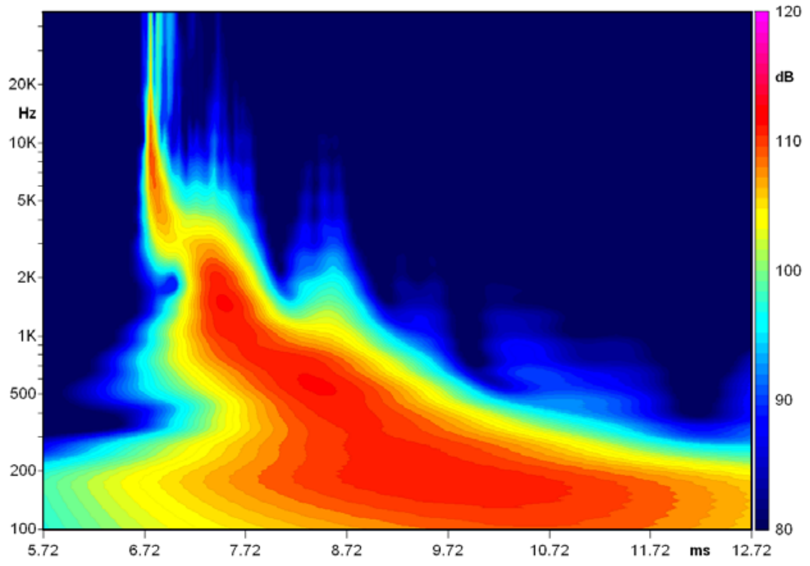


Fig 10 . A loudspeaker close (1m or 3.4ms) to a diffuser in a control room. The scatter between 2-5kHz close to the loudspeaker is box resonances.



The next picture is a monitor prototype suffering from severe time-smearing. The scatter closer than 2 ms (70cm) to the direct sound is due to a diffraction or cone break-up from the loudspeaker itself.

Fig 11. Monitor prototype with severe time smearing.

If we study the multi-windowed group-delay of the same measurement we can notice the same time-smearing. However it's harder to notice the box resonances and cone break-ups.

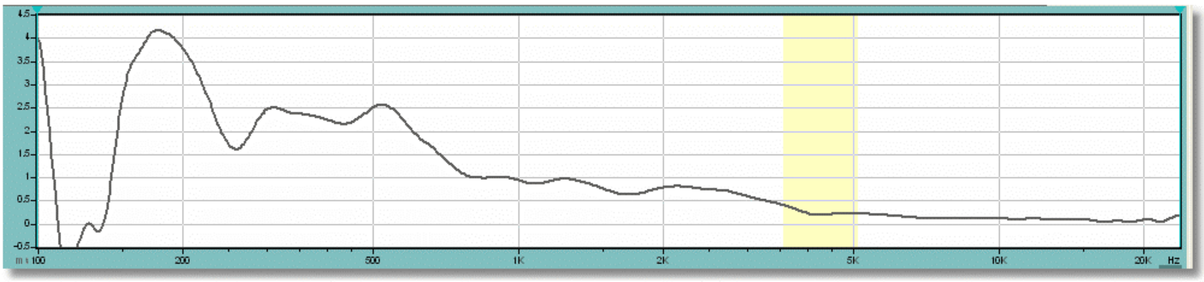
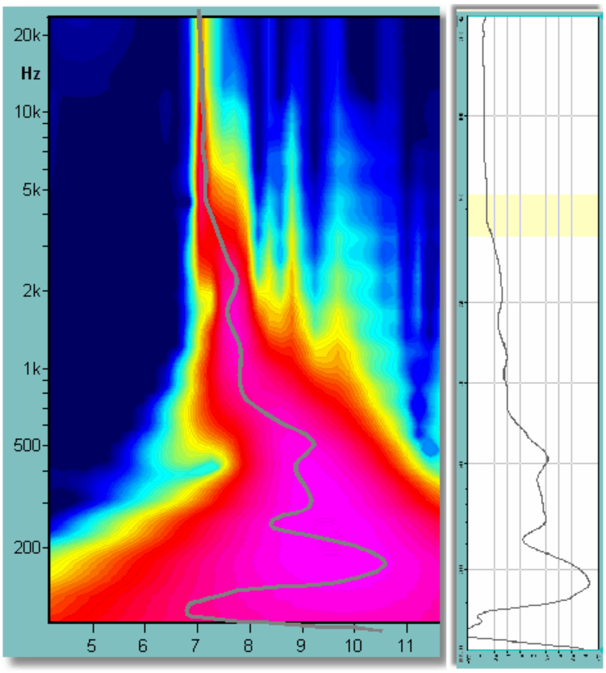


Fig 12. Group-delay plot of a loudspeaker with severe time smearing. The group-delay graph shows the time smearing of 0.7ms between 1 – 3 kHz and a 2.3ms offset between 150 to 500Hz. Referred to the time arrival above 3k.



One can actually take a MultiWindowed measurement and turn the group-delay curve about 90deg counter-clock wise and flip it around. It corresponds to the ridge of the wavelet analysis.

The ***vertical curve*** represents the **ridge** of the scalogram (highest amplitude). Here is the energy focused in time and represents the same value in time as the group-delay. However there are more details in the group-delay curve below 300Hz.

Fig 13 .Group-delay visualized in the Scalogram

To visualize the time smearing one can see the three vertical lines, representing the time offset in the 3-way loudspeaker. The loudspeaker symbols represent how the listener will be physically exposed to the three driver's misalignment.

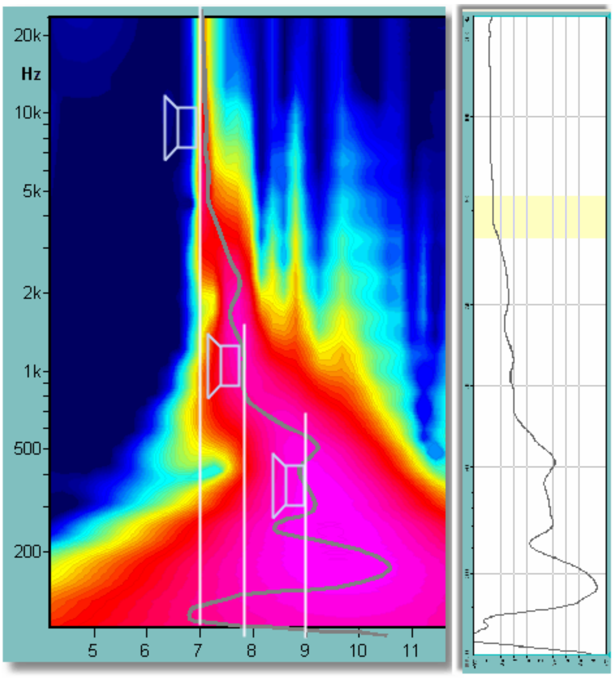


Fig 14. 3-way loudspeaker with misaligned delays

Minimum phase system

However it isn't always possible to achieve a perfect impulse response. This requires that the system is has a linear phase target.
A good compromise is a minimum phase system. This is the most common alignment, as it requires a minimum of latency. **When the magnitude is flat and minimum-phase, all sine waves start at the same time.** This gives a group-delay that has more delay at lower frequencies. At low frequencies the perceptual threshold is quite high, so this is hardly audible.

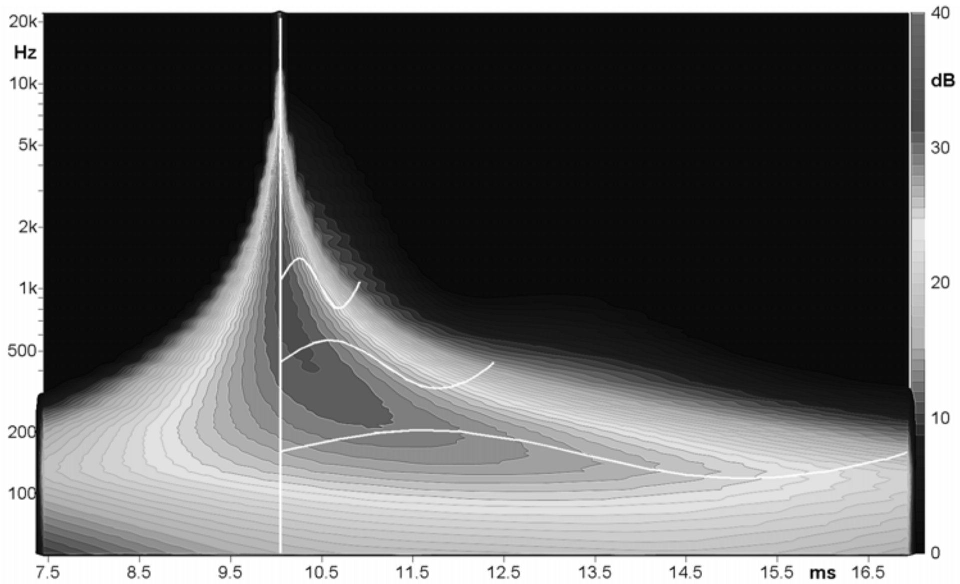


Fig 15. Scalogram from wavelets analysis of a Minimum Phase system.

Linear phase

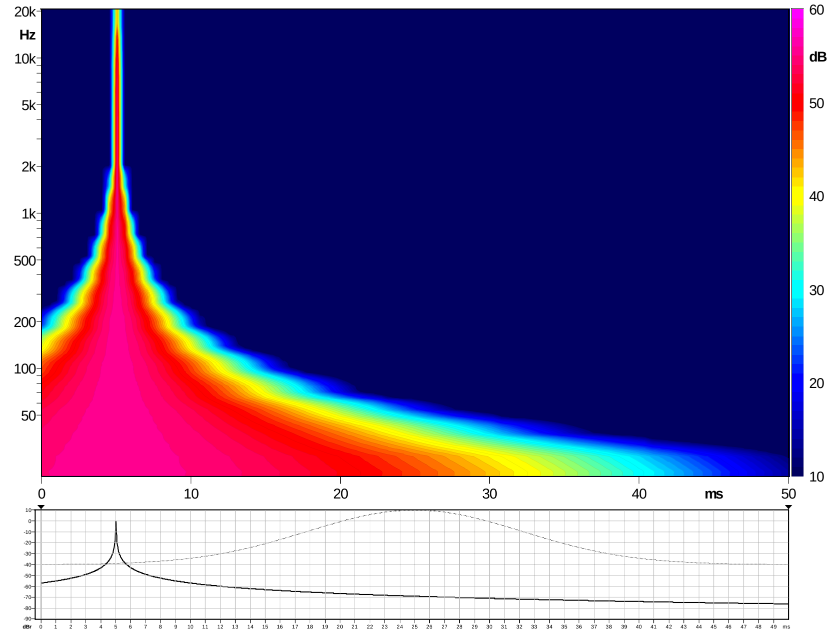


Fig 16. Scalogram from wavelets analysis of a Linear Phase system.

However, a minimum phase system doesn’t reproduce transients perfectly. (With one exception; if the has Bessel low-pass characteristic, the group delay will be flat even for a min-phase system). The common opinion is that the spatial resolution in a stereo system will be improved in a linear phase system.
When the magnitude is flat and linear-phase, all sine waves have their zero crossing at the same position in time. The resulting group-delay is flat. Please note that the group-delay can never be 0s when linear-phase (or minimum-phase) this is only true for a dirac which is a signal that have infinite bandwidth or 1 sample duration.
Therefore there is always latency “cost” to make a linear phase system.

Maximum phase

A third variant of group-delay target is the maximum-phase. To our knowledge it has never been described in the literature before.
When the magnitude is flat and maximum-phase, all sine waves stops at the same time. This gives a group-delay that has less delay at lower frequencies. The question is the practical use of such a target. One could be to compensate for a chain of high-pass filters in a distribution chain. But the latency cost is high as seen in the graph.

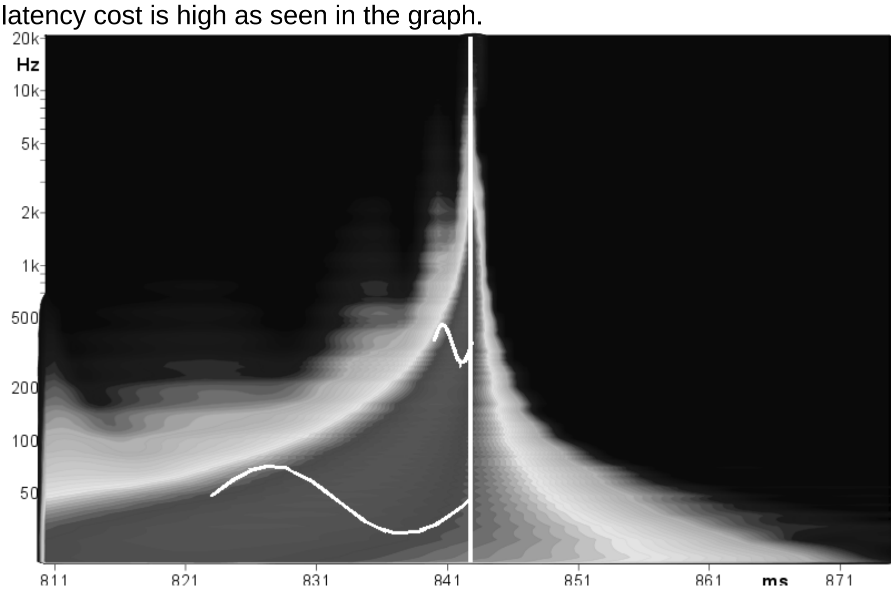


Fig 17. Scalogram from wavelets analysis of a Maximum Phase system. Note the 845ms latency!

Understanding the frequency and time graph

Even under nearly perfect conditions, such as a medium size studio room, the response is blurred with "noise"

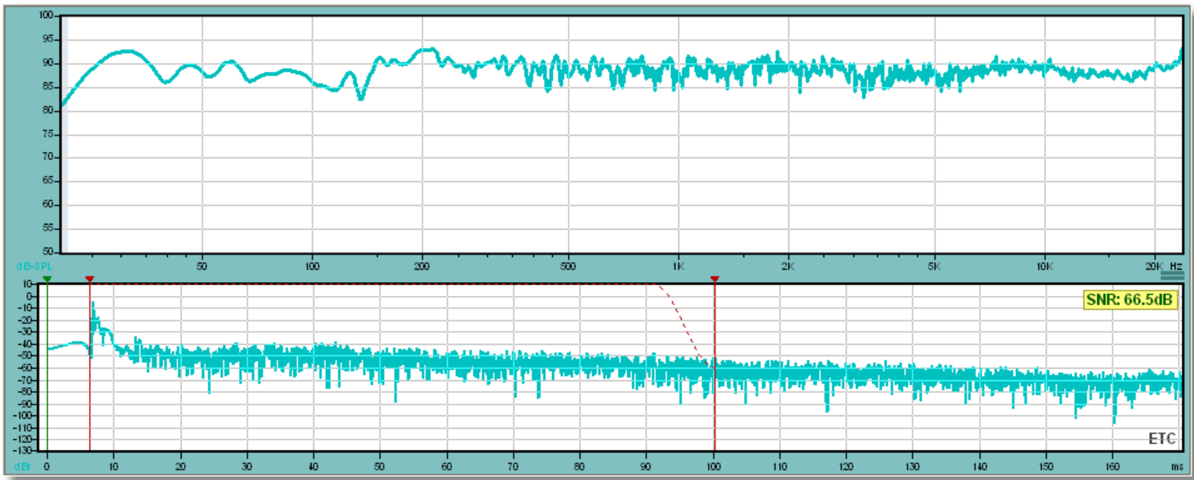


Fig 18. A high performance studio monitor in a medium size studio, the lower graph is the Energy Time Curve, ETC.

The standard procedure in many measurement systems is to apply "Smoothing" in the frequency domain to remove the "noise" and make the response more visible.
However, smoothing is useful for removing random noise, such as thermal noise, but not time related comb filter results.

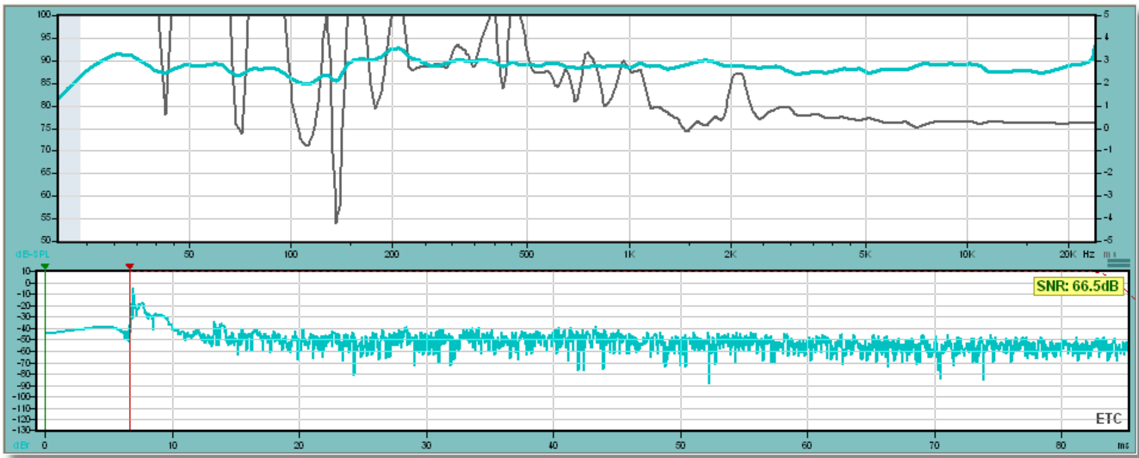


Fig 19. PPO smoothing applied on the magnitude and group-delay curves

The drawback, as can be seen in the graph above the group-delay (black curve) will be distorted with smoothing.
This isn't "noise" but multiple reflections, which cancel out the direct sound due to the comb-filter effect.

Lets see where this "noise" is coming from. A source (loudspeaker) is placed in a room with walls, floor and ceiling. A listener is placed with a distance nearly equal to the source and walls.
If the source excites a pulse it will take a time (the propagation delay or Initial delay) until it reach the listener. This first sound is called the Direct sound, because it isn't influenced by the room. If the source isn't totally directional, but usually omnidirectional, the sound will also hit the walls etc.
If the sound diverts omnidirectional from the source and it bounces against a hard surface (early reflections) it can actually be at higher magnitude, due to the fact that the surface is a half-sphere and not omnidirectional. When the sound has been reflected several times against the surfaces, the energy will decay in a reverberation field.

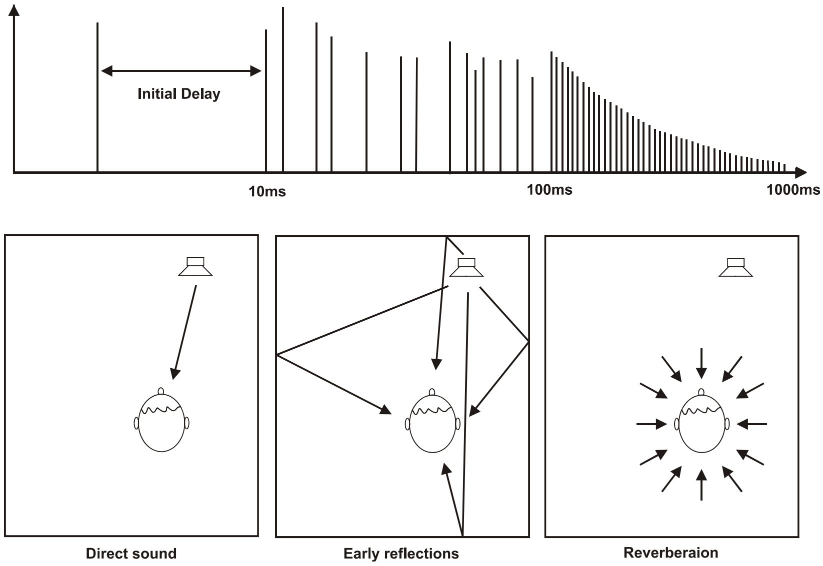


Fig 20 . Direct sound, Early reflections and Reverberation field

The early reflections will interact with the direct sound in the listener position; when the waves has travelled distances so they are in phase they will enhance the magnitude, but on the frequencies there they are out of phase with the direct sound, the wave will be cancelled. The pattern of enhanced and cancelled waves is therefore highly frequency and position dependent. This pattern describes of a sc. Comb-filter response. See fig 22. below.

Summation examples

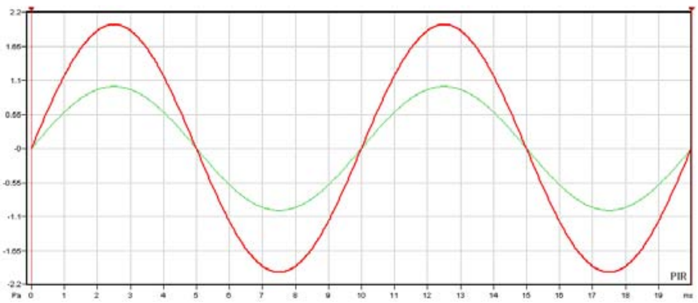


Fig 20. – 0 deg (blue behind the green trace) + 0 deg (green) = sum (red) 2.0

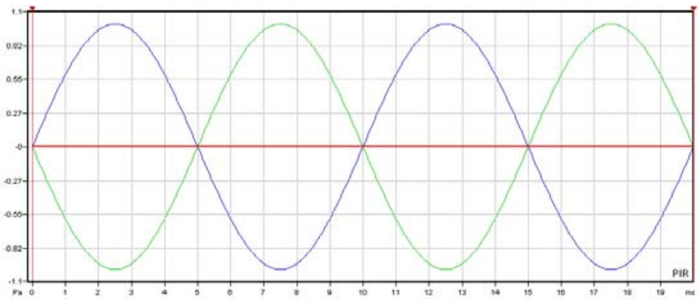


Fig 21. – 0 deg (blue) + 180 deg (green) = sum (red) 0.0

14.87500000	0.00000000
14.89583333	0.00000000
14.91666667	0.00000000
14.93750000	0.00000000
14.95833333	0.00000000
14.97916667	0.00000000
15.00000000	1.00000000
15.02083333	0.00000000
15.04166667	0.00000000
15.06250000	0.00000000
15.08333333	0.00000000
15.10416667	0.00000000
15.12500000	0.00000000
15.14583333	0.00000000
15.16666667	0.00000000
15.18750000	0.00000000
15.20833333	0.00000000
15.22916667	0.00000000
15.25000000	0.00000000
15.27083333	0.00000000
15.29166667	0.00000000

To understand what we see in the graphs, we can create a synthetic impulse response by putting “1” in a text file with the time-axis defined. Measurement noise and reverberation noise is then removed from the graphs. The complex sum of two impulse responses, separated by 10ms is shown in the graph below.

The resulting curve is however highly dependent on the time between the impulses. The time is dependent on the distance from the reflection’s surfaces in a room. Therefore the resulting response is very position dependent if reflections are included in the measurement. If the measurement is position dependent there is no meaning in doing an overall equalization of a loudspeaker.

There is a risk to misalign a delay setting if the reflection is taken as a reference instead of the direct sound.

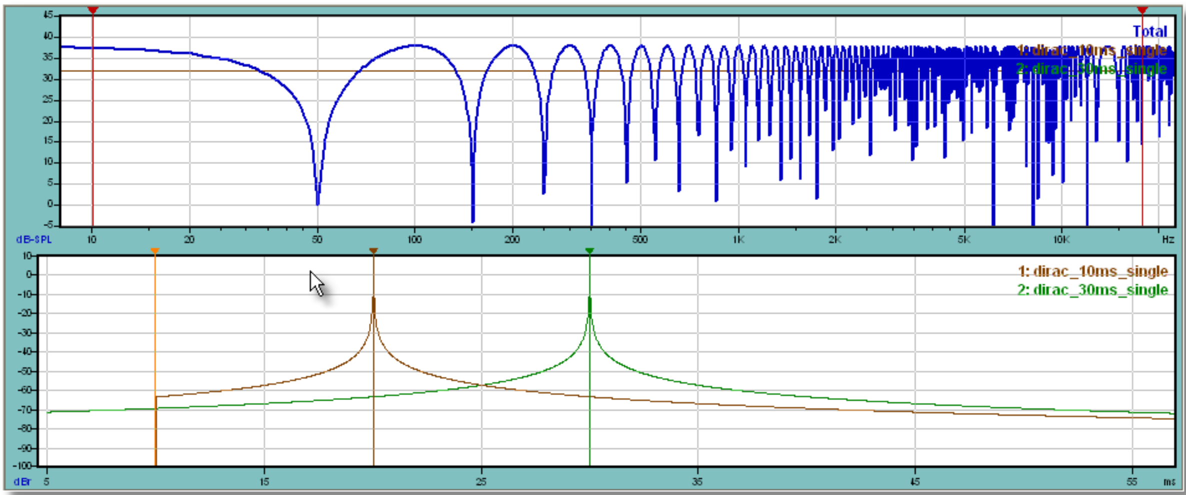


Fig 22. Two impulses separated by 10ms sum up in a comb filter.

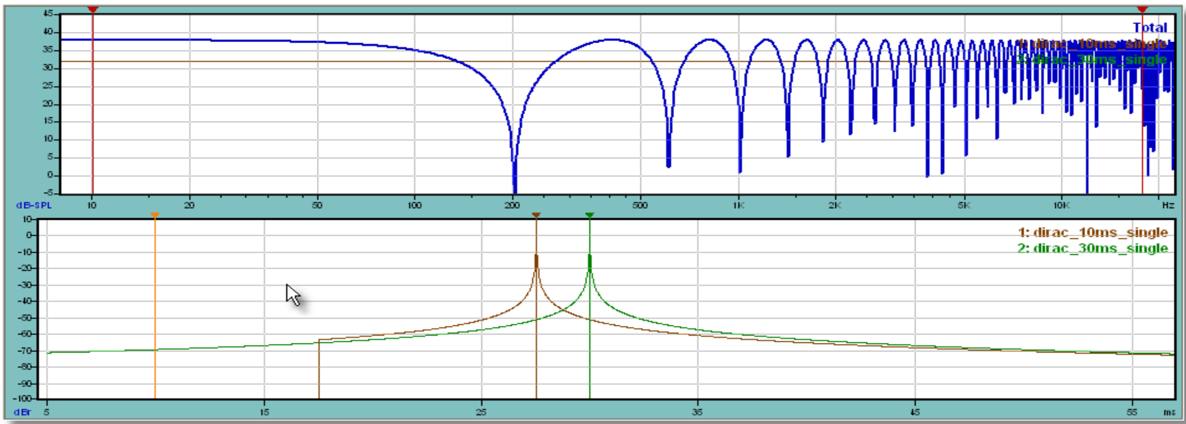


Fig 23. Two impulses separated by 2.5ms sum up in a comb filter.

We can also make a synthetic reverberation by adding multiple impulses at 8, 9 and 10ms after the main impulse.
Here we can see something close to the real measurement and we can draw the conclusion that there isn't so much "noise", but the response is a result of multiple impulse arrivals.

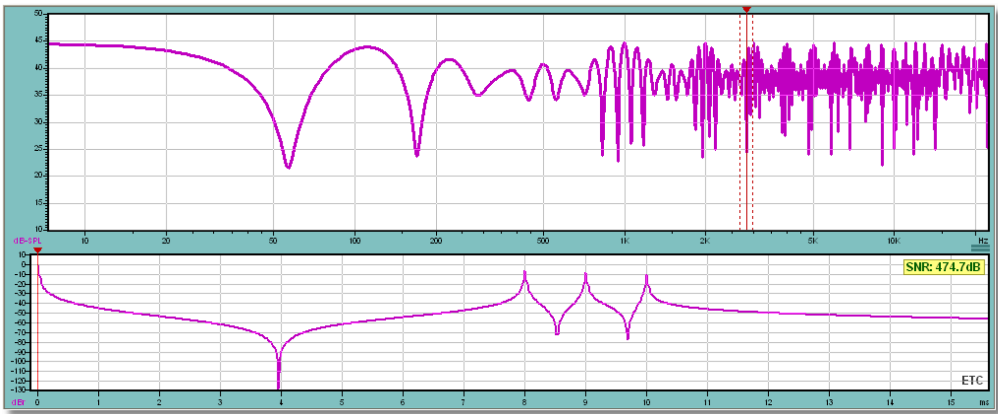
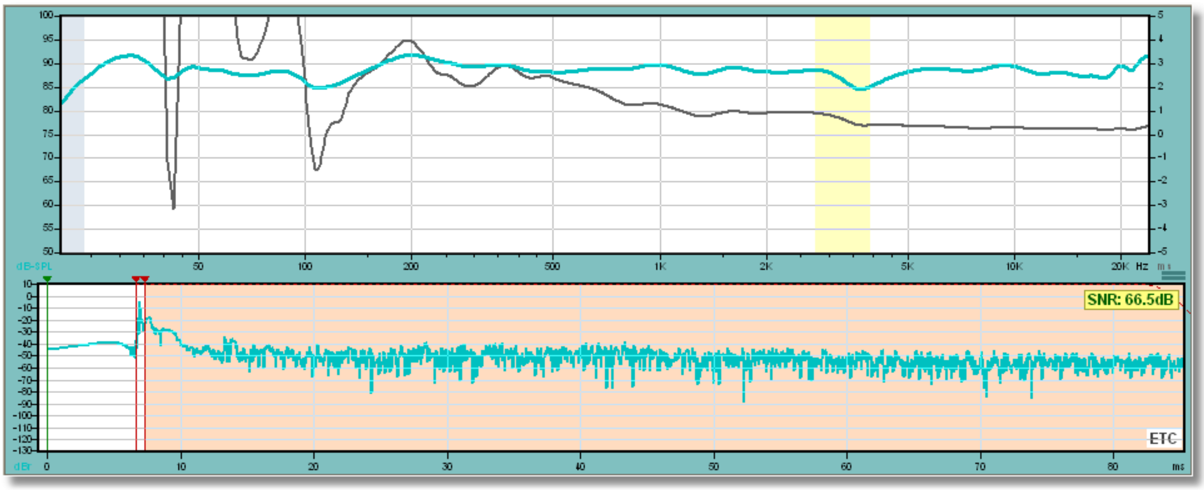


Fig 24. A synthetic IR with direct sound and early reflections

Windowing

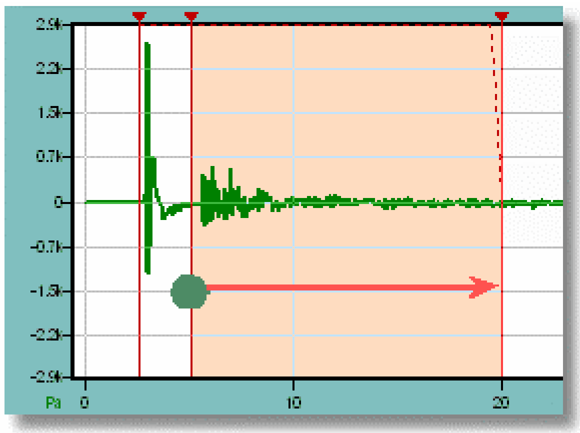
If the room reflections are attenuated by powerful windowing, the anechoic response can be seen and the loudspeaker's group-delay will be more useful.



The transients in the group-delay is the jump between the phase of the direct sound and a reflection of a specific frequency at which the reflection is dominated over the direct sound. (This can happen if the sound wave reflects against a hard surface. If the sound diverts omnidirectional from the source and it bounces against a hard surface it can actually be at higher magnitude, due to the fact that the surface is a half-sphere and not omnidirectional).

By using Multiple windowing most of the group-delay transients can be attenuated and this is a good sign that most reflections are removed from the measurement.

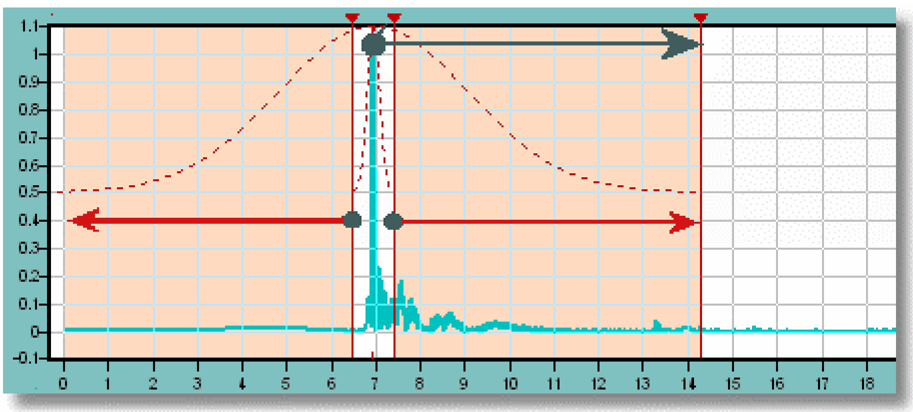
Multi-windowing starts with a small window (here a half window), washing out the reflections, corresponding to the highest frequencies. In the next step the window slides with twice the size, corresponding to half the highest frequency. These two first time slices are moved to the frequency domain by FFTs and merged together. The finer the time slices are, the more points in the frequency domain and more *"Merge" per octave*. As the wider windows correspond to the lower frequencies, the higher frequency contents in them are from reflections etc. and are washed away.



Wavelets con't

Wavelets analysis is very similar to Multiple time windowing.

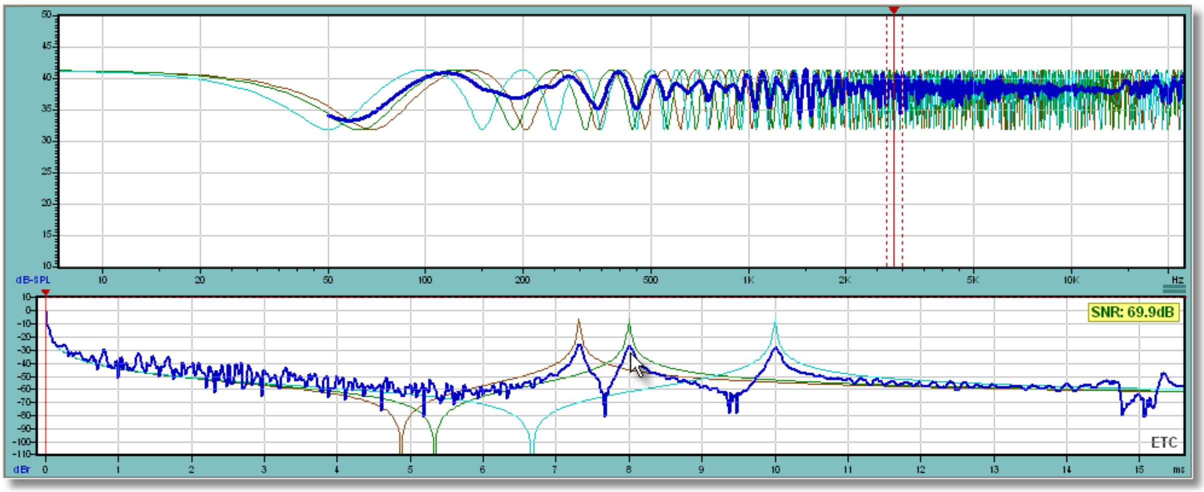
The wavelets analysis usually starts with a Gaussian window (here a symmetric window). Each time slice in the scalogram is made up of merged FFTs from corresponding time windowing of the PIR. The number of Merge Per Octave, will determine the frequency resolution.



The wavelet scalogram is made of a number of time slices. These time slices are made in the same way as a Sliding CSD, by sliding the whole multiple window packages stepwise.

Note: Wavelets are very computing intensive. Even with the lowest frequency resolution with 20 merging bands, it's 2000 of 32k points FFTs to calculate. With Ultra high quality and 320 frequency bands there will be 256 000 FFTs to make!

Averaging of multiple positions

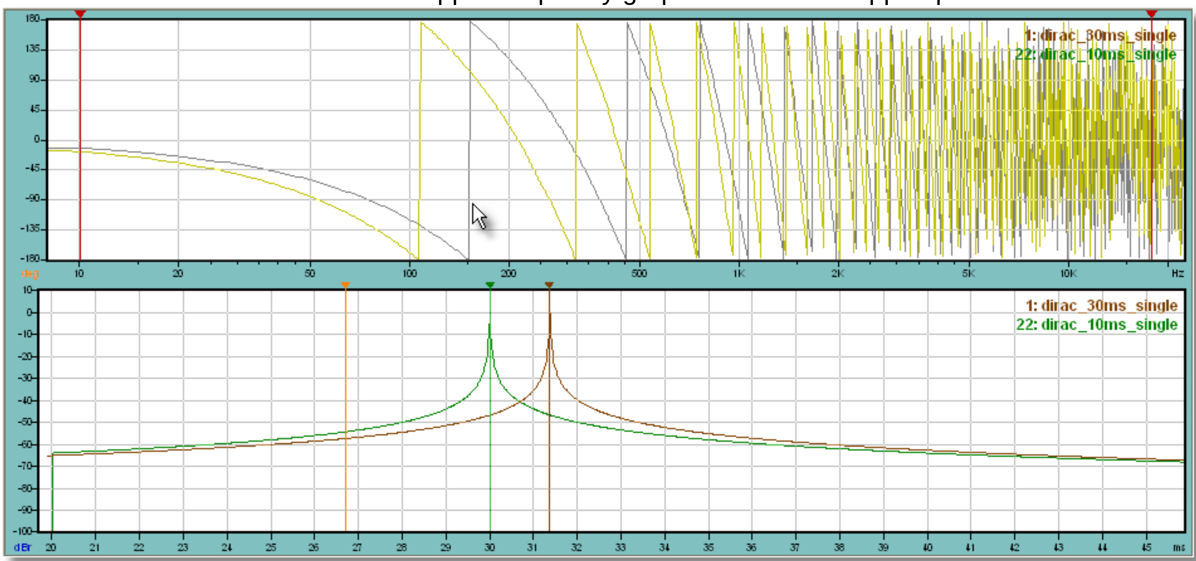


By making multiple measurements in different positions, the reflections will spread out in time. As seen in this example they are only spread by one or a few milliseconds, which corresponds to a distance less than a meter. The averaged magnitude response is the power average of the three measurements. To be able to recreate the time graph, the group-delays are separately arithmetically averaged and then combined in the time graph with the magnitude.
As seen in the time graph, the three reflections are suppressed by approx. 25dB in the result, compared to the original level in the ETC.

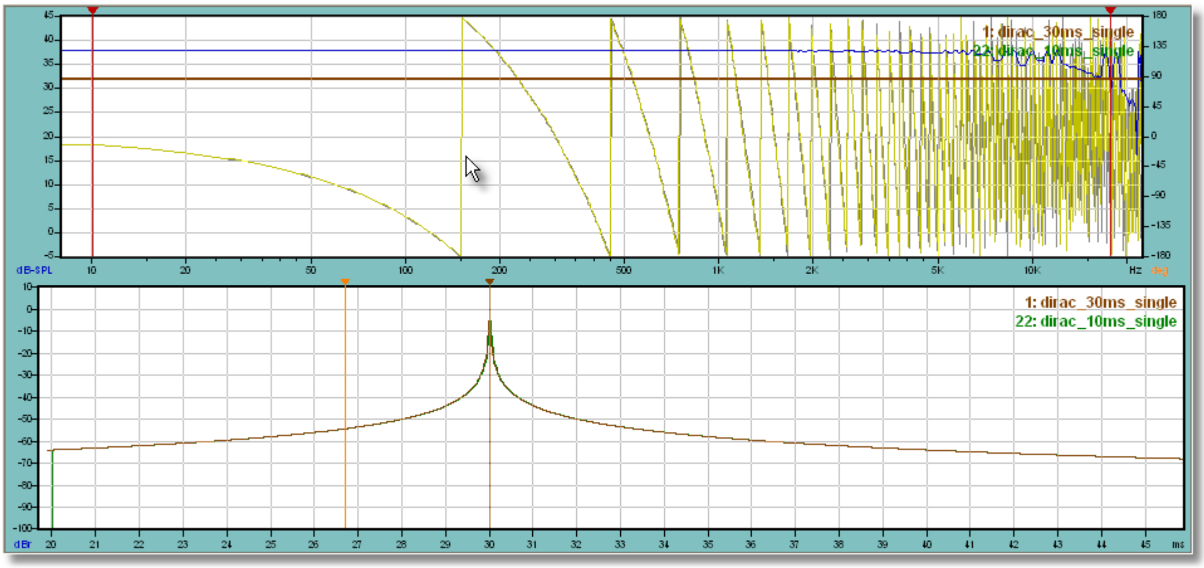
Tools for delay alignment

Wrapped phase

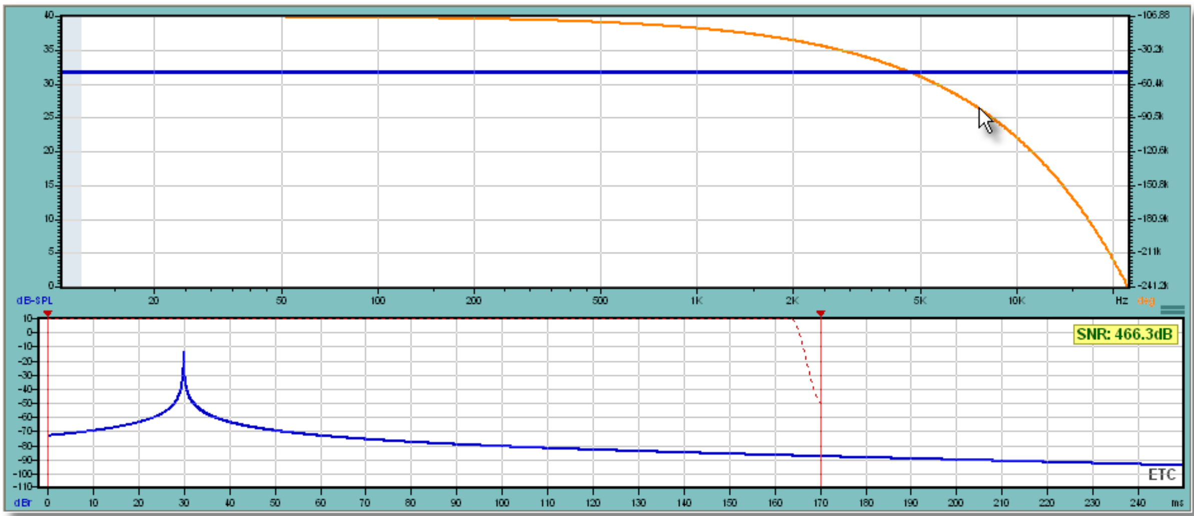
Using wrapped phase for matching two signals is a well-established method. However it has it pitfalls and it is misused in many cases. The time graph (lower) shows two single impulses in ETC form, which have an offset of 1.4ms. The upper frequency graph shows the wrapped phase of the two.



By aligning the delay of one of them, the two impulses will match both in time and phase. The blue line is the summed complex frequency response.



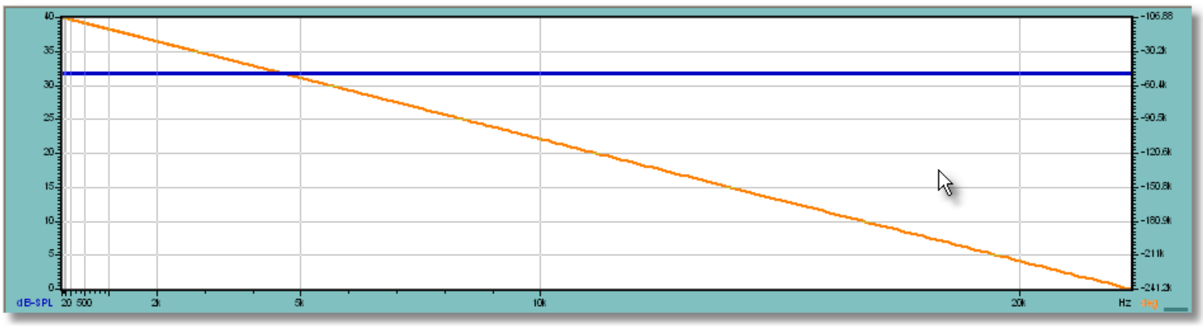
To understand the wrapped phase graph, one must study the unwrapped phase. See the orange curve.



The problem in this graph is the scale. Already at 1kHz the phase is 10 800 degrees. (the period time is 1ms at 1kHz (t=1/f) and the propagation delay is 30ms, so the phase is 30 x 360 degrees. At 20kHz the phase is 216 000 degrees. This means that it's very hard to see any details in such a graph.

Phase-Delay

The reason why the graph is slightly curved is due to the logarithmic frequency scale. A delay is proportional to the frequency, so with linear frequency scale the phase curve will be a straight slope.



(The steepness of the slope represents the delay. In this case both the phase and group-delay are the same.

$$t_{\varphi} = \frac{-\varphi}{2\pi f}\ ,\ \mathrm{s}$$

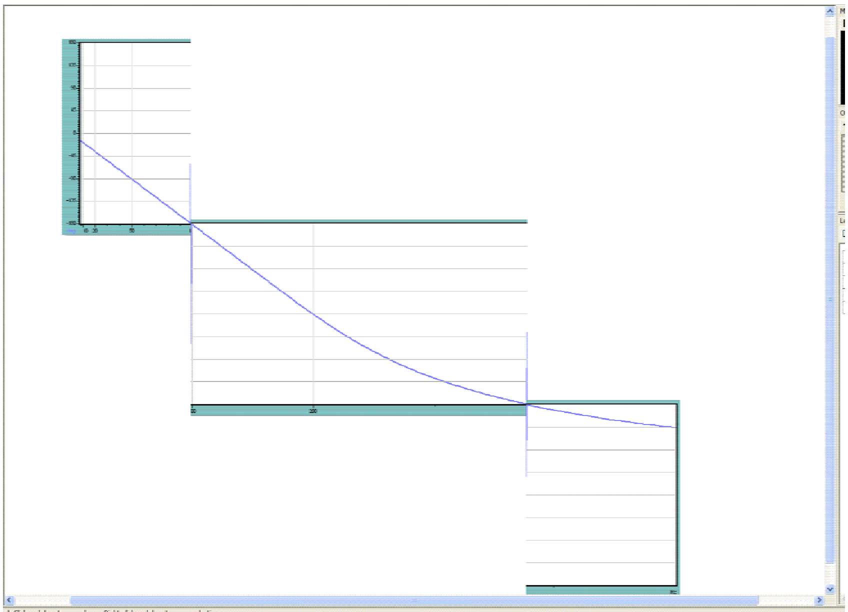
In this case we use degrees instead of radians, so the delay is;

$$t_{\varphi} = \frac{-\theta}{360\,f} \text{ , s}$$

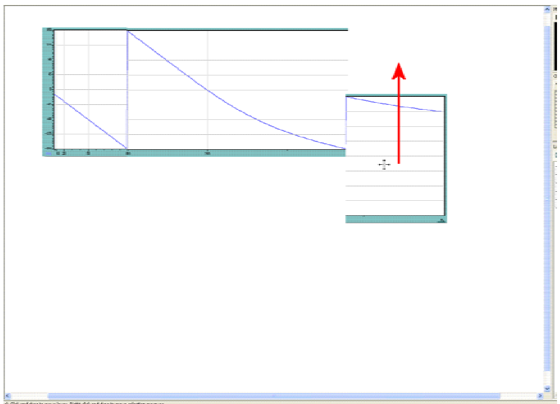
(In this case Tp= 216000/ (360 x 20kHz) = 0.03s= 30ms.)

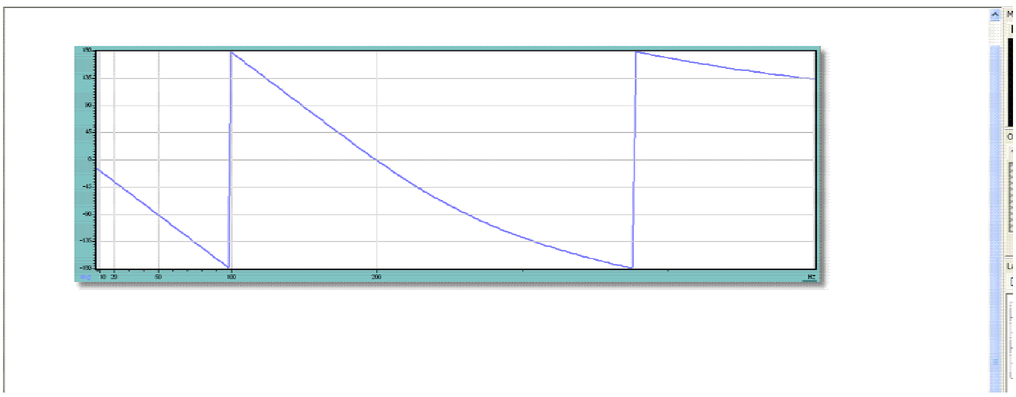
Wrapped phase con't.

In the following row of pictures is shown how a wrapped phase curve is created from the same unwrapped phase curve.



The unwrapped phase is zoomed vertically to be able to see the details

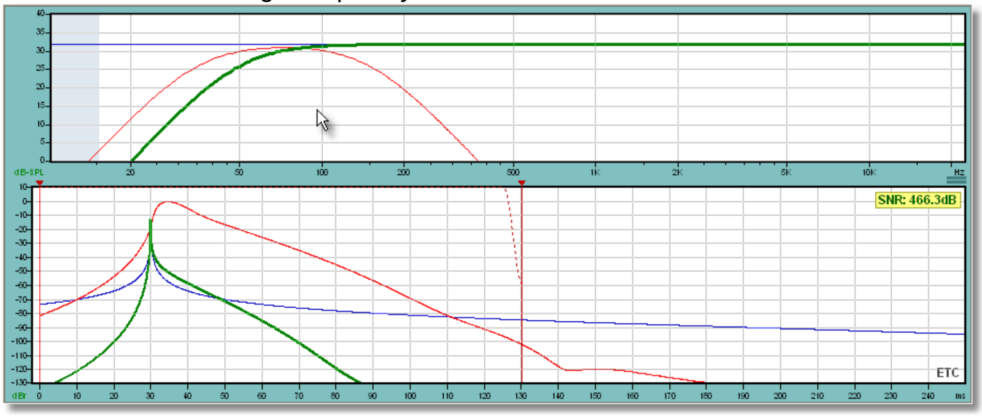




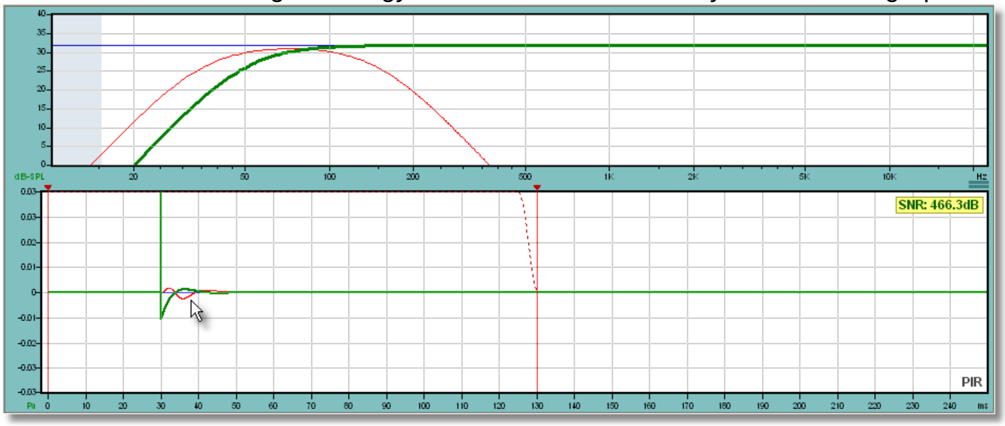
Each +/-180 deg segments are cut vertically and interlaced to a common +/-180 deg scale. The vertical lines are just the jump between the half turns.

Impulse response calculations; ETC & Cepstrum

Besides the IR curve itself, the time graph [22] can be processed to show other curves with different features. The one we use here are Energy Time Curve [21] and Cepstrum [19]. Other useful curves are Step response [23], Log-squared, Schröder etc.
The graph below shows unfiltered perfect IR (Dirac) (Blue), a high-pass filtered dirac (green) and a band-pass filtered dirac (red).
Note how the high-pass filter delays the bottom energy (low freqs) and how the sub-bassETC is rounded off from its high frequency content.



The same graphs with the Impulse response itself. Note that the pure dirac and HP-filtered looks the same. The sub-bass signal energy is smeared out and hardly visible in the graph.



A not so common calculation, but very useful is the **Cepstrum** analysis. The benefits over PIR view is that the noise is suppressed, so reflections are clearly visible [19, 20]. This is very useful for setting the time windows. Cepstrum is the inverse FFT of the logarithmic magnitude with the phase preserved. As the logarithmic scale doesn't contain negative values, the time domain will be "rectified". The peak is normalized to one. (**Ceps**trum is a spectrum of a spectrum and the word is coined by paraphrasing **Spec**trum. Other words in the Cepstrum domain are Quefrency (Frequency) and Rahmonics (Harmonics)).

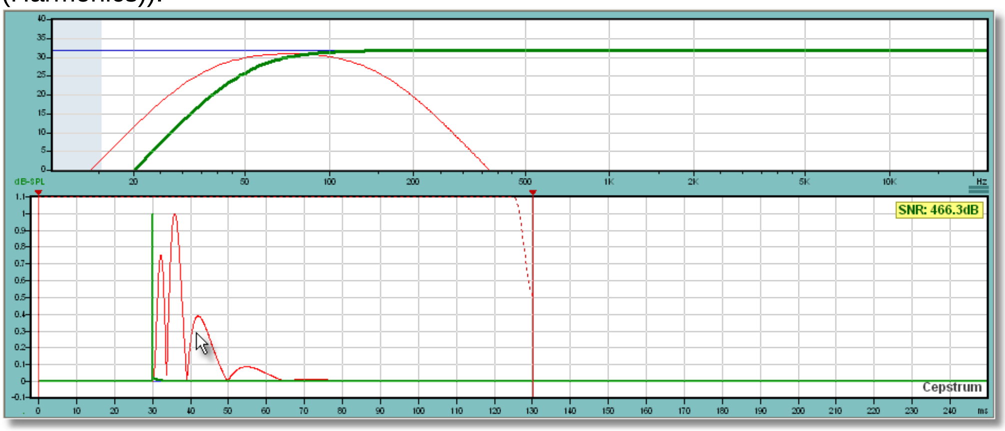
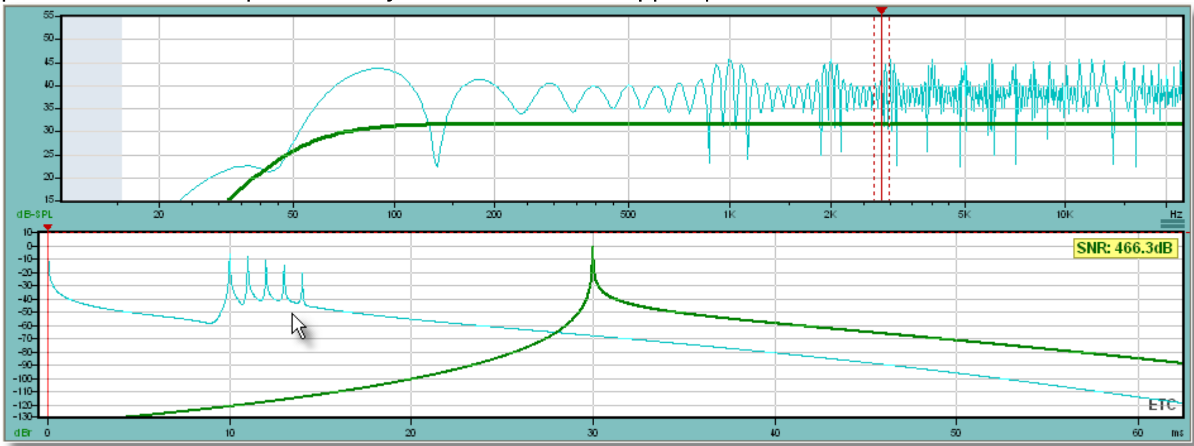


Fig 25. The feature here is that the sub-bass is visible and it can be seen how the energy is smeared out in time.

Making delay alignment

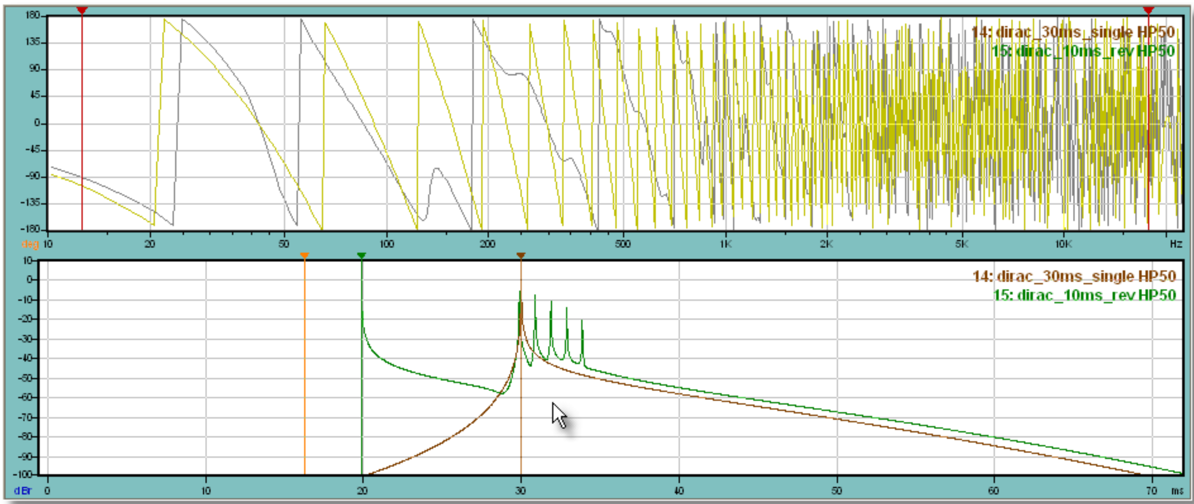
Case 1; two full-range loudspeakers in a delay system

The alignment of two full-range systems in time isn’t so easy, as the real world loudspeakers aren’t so perfect as the one represented by the Dirac in the wrapped phase section above.

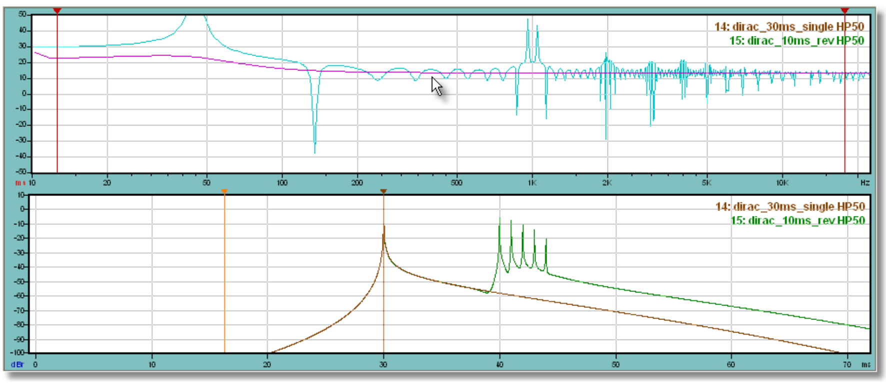


Most sound reinforcement loudspeakers have a low frequency cut-off around 35Hz or above. The curves above are both filtered by a high-pass filter around 50Hz. The measurement is often influenced by reflections from the room as seen in the blue curve above.

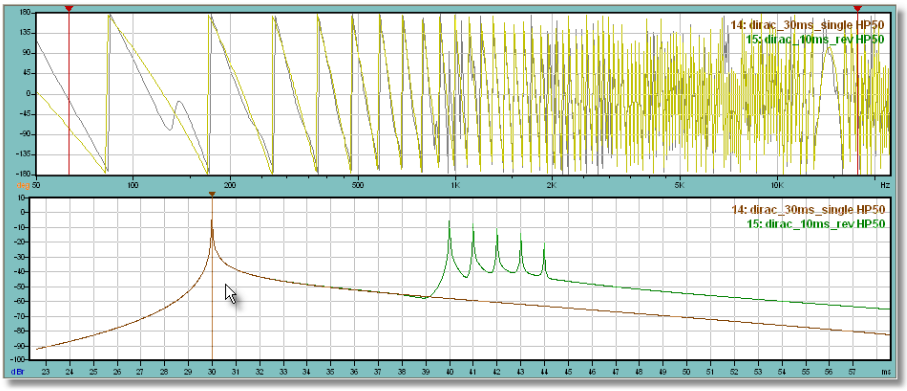
The next graph shows the wrapped phase curves from the two systems in the top graph and the ETC in the bottom graph. Very often mistakes are made when trying to match the wrapped phase. The phase can became several turns off, with time smearing and side-lobe errors as results.



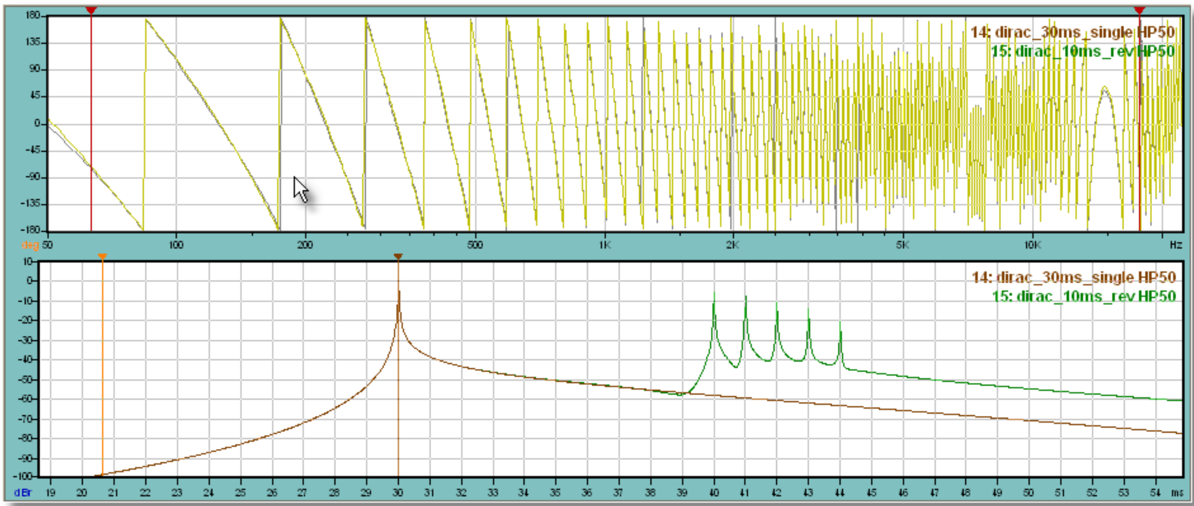
The reason is often the cut-off frequency, which makes an increasing group-delay at low frequency



and will bend the low frequencies of the phase. This is very hard to see in the phase graph, but if we look at the group delay instead we can see how the curves divert below 150Hz.

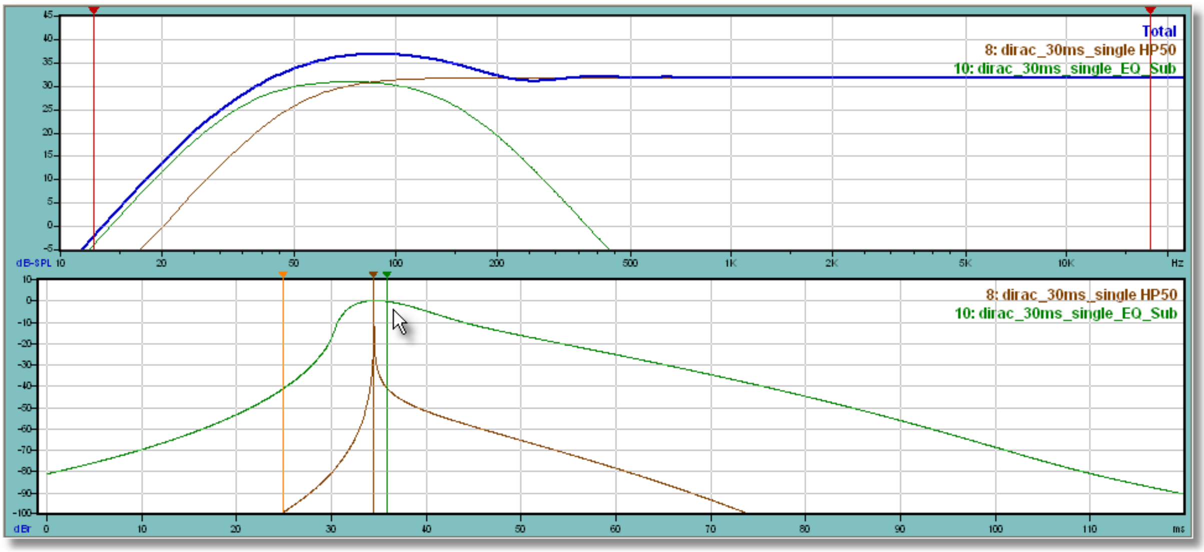


If the high-pass slope or the LF cut-off is very steep its very important to raise the start frequency of the FFT. Otherwise the phase calculation will start down in the noise floor with invalid results.
In this graph the FFT start frequency is raised to 50Hz. It's obvious that it helps when trying to find the match to the wraps.
As seen in the wrapped phase above, there are some extra oscillations due to the reflections in one of the IR. ***If a strong reflection is present close to the desired IR, there is a risk that one aligns to the reflection instead.***
It is therefore very important to window the time graph so the measurement will be anechoic. In the time graph below Multiple windows are applied to clean the IR.

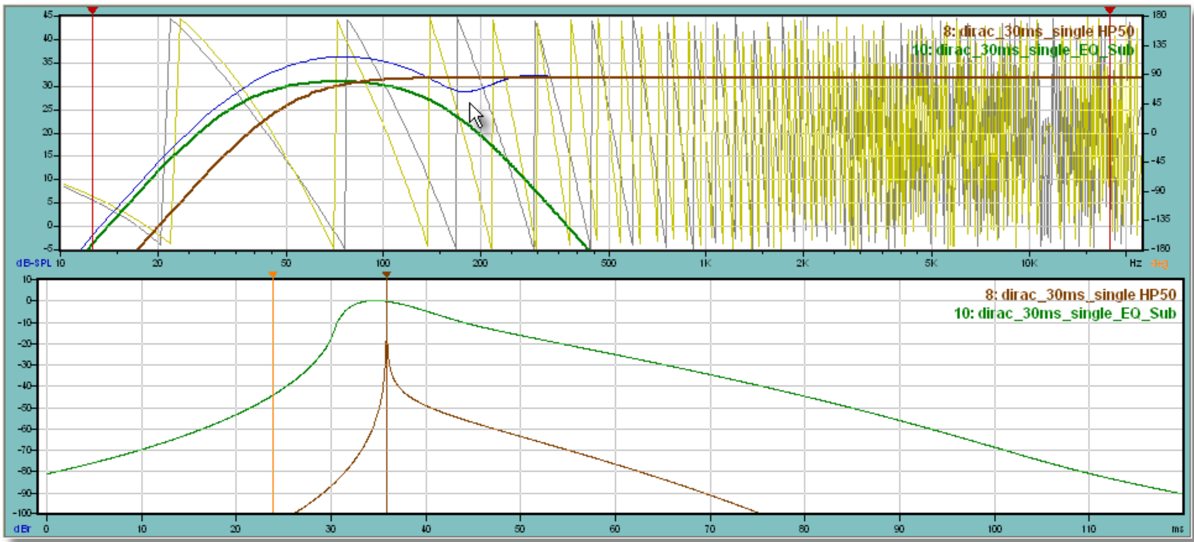


Case 2; aligning sub-bass to a top-system

There are a number of issues to keep in mind when delay aligning a sub-bass system to a top-system. Due to the lack of high frequency energy in the sub-bass, the impulse response is smeared out in time and at low level. It's therefore hard to get a well-defined impulse to adjust in the time domain. (refer to Impulse calc, ETC & Cepstrum section fig. n)
The difference in spectral content makes it hard to visually align the delay with assistance of ETC or Cepstrum.
In the graph below the delay is aligned for flattest amplitude. However, the peaks of the two ETC, as indicated with the cursors, are separated around 2ms. This is due to that we want to achieve a minimum-phase system where both band's impulse starts is matching. (If we match the ETC peaks we will match the zero-crossing and have a linear-phase system, but in this case only in the crossover region, which isn't desirable.

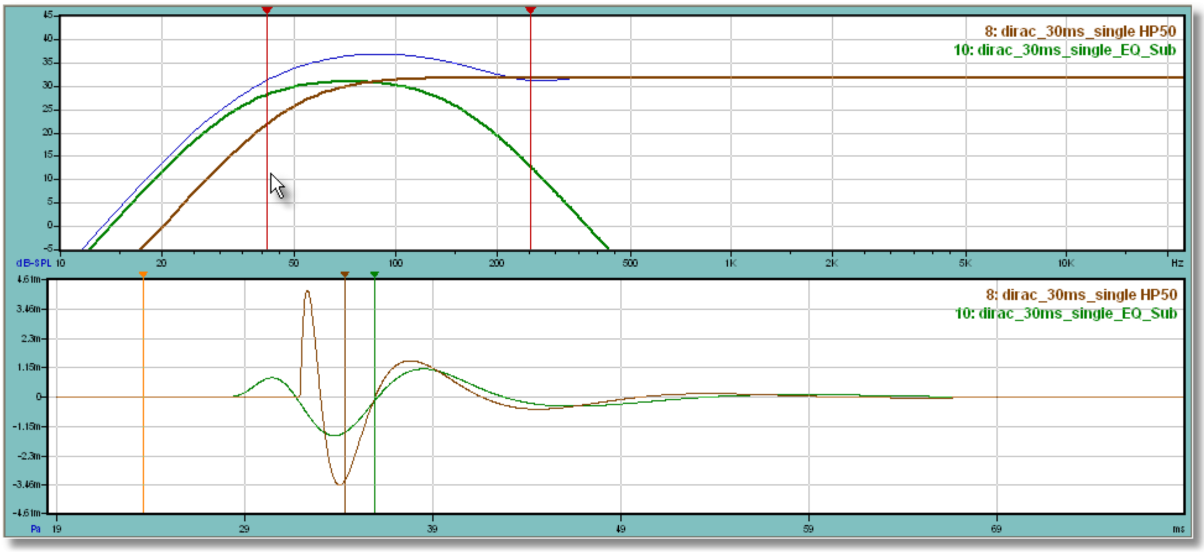


Note that it's desired that the HF-energy come slightly before or on time compared to the sub as seen in the ETC graph to achieve a min-phase system.
When using the wrapped phase alignment method attention must be paid to that the phase for the sub is only valid in the 30 –300Hz range. Outside this range there is a risk that the signal is down in the noise floor. One can try to match the wraps around 100Hz, but with a risk of picking the wrong turn.



Matching group-delay in the crossover point can be successful, if the measurement is anechoic. Otherwise the method has the same limitation as wrapped phase, due to the band-limited signal in the sub-bass.

A more successful method is to apply a filter on both impulse responses, so the shape of the IR (and Ceptrum) will be the same. This will enhance the time domain display for the crossover region; the PIRs can be filtered so the high frequency energy and the raised peak will be attenuated. The red cursors in the frequency graph are set to define the crossover region. Here the cursors are set one octave up and one octave down from the crossover point.



As seen in the graph, the sub-IR has the same magnitude range as the HF-energy is filtered away from the top-system. The delay can be found by studying the total complex magnitude sum (blue curve) for maximum flatness and at the same time have a match of the filtered impulse responses.

Note that with a crossover frequency of 100Hz, the period time is 10ms. The summation will be within 3dB tolerance if the phase difference is within 90 degrees. 90 degrees is a quarter of a period, ie. 2.5ms. So the delay has to be set within +/-1.3ms. The mechanically tolerance is approx +/- 40cm. This also implies that the measurement must be done on an enough distance from the sub/ top combination, otherwise the difference in height will add to the tolerance (if we assume that the subs are on the floor and the tops are flying).

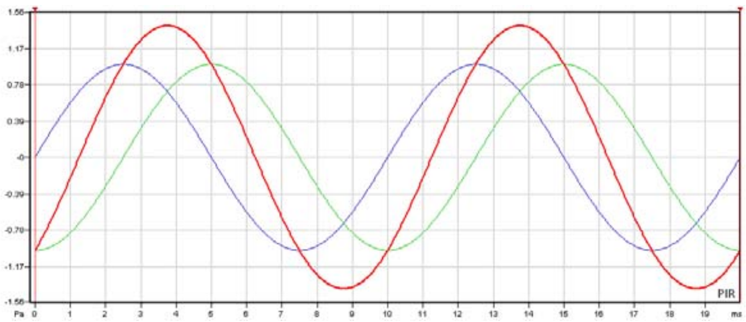


Fig 26. – 0 deg (blue) + 90 deg (green) = sum (red) 1.4142

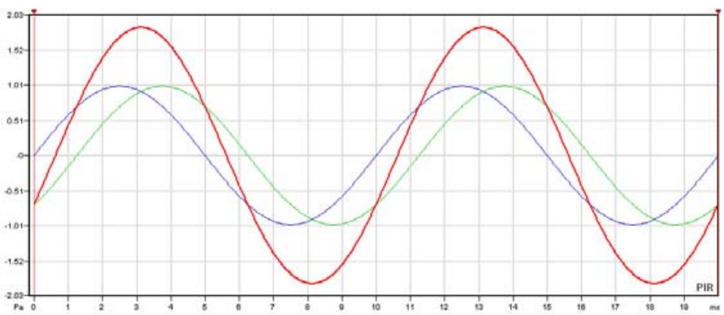
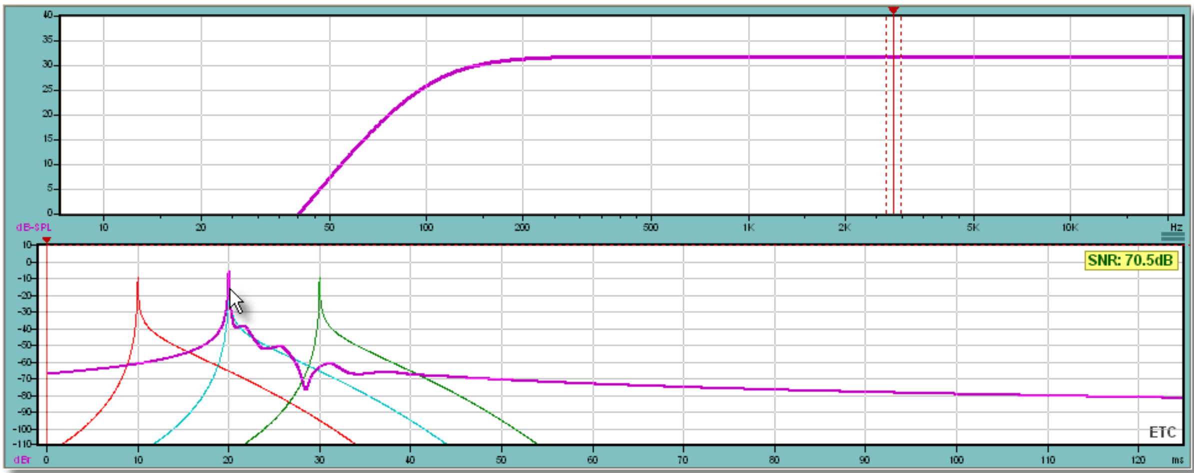
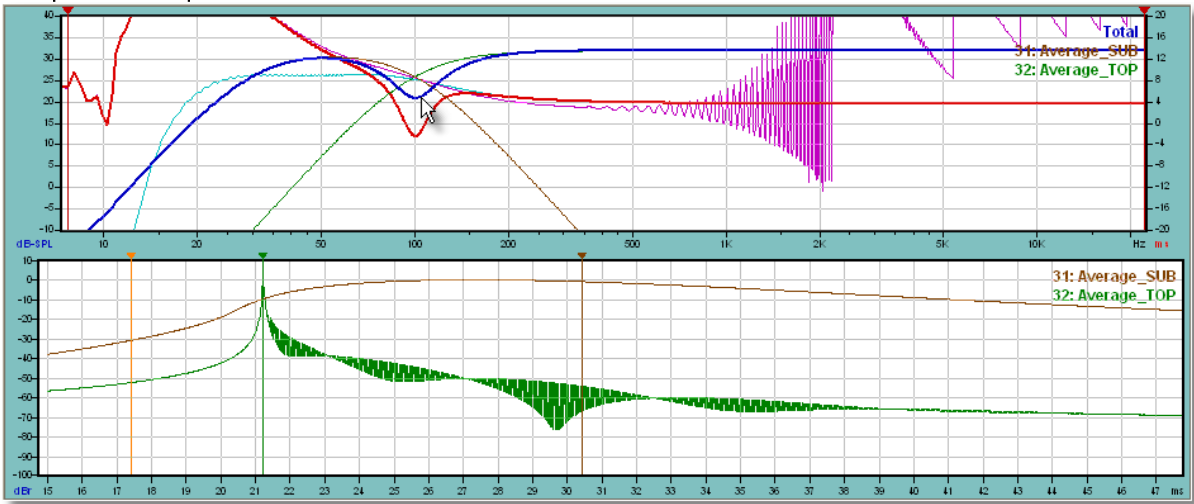


Fig 27. – 0 deg (blue) + 45 deg (green) = sum (red) 1.8478

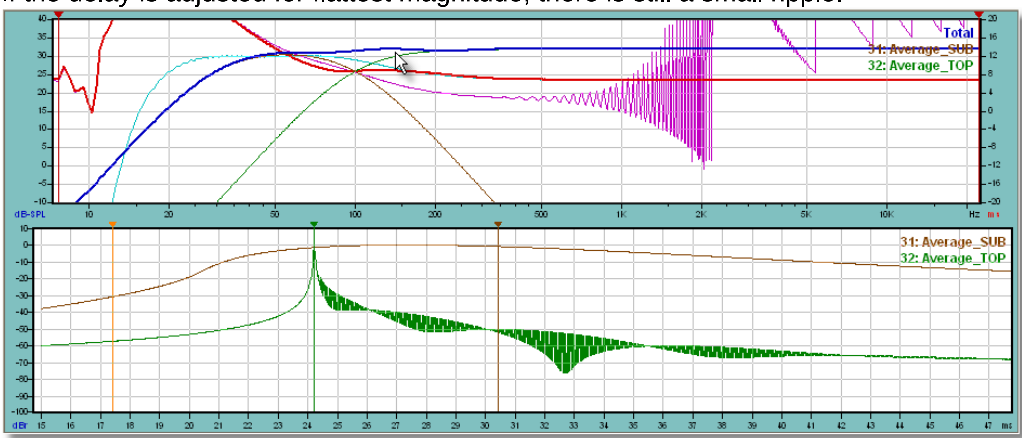
Case 3; aligning delay with averaging positions



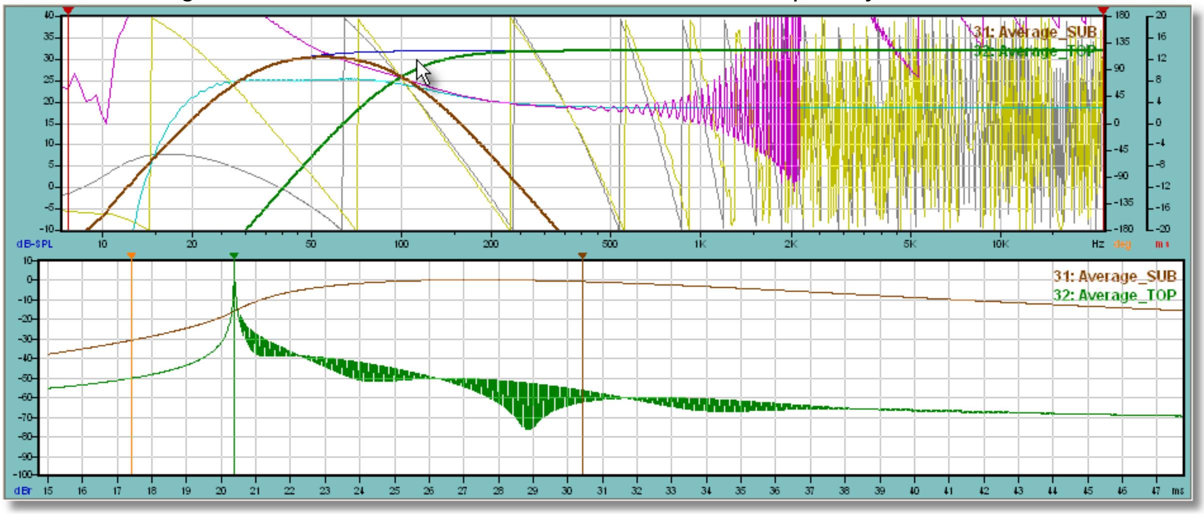
Due to room-reflections the measurements are highly position dependent. The risk is therefore that delay alignment and equalization will only be optimized in one position. To avoid this it's desired to make an average over many positions in the entire audience area. As the different position has different delay and therefore different phase, the frequency domain vectors can't be summed complex, as comb-filter cancellation will occur. Power averaging, which is the root-mean square of the frequency vectors, is the standard solution to take out the phase of the equation. However, usually the time domain data is lost as no phase is available. In the graph above the impulse is restored as the phase is arithmetically averaged.
This implies that a number of positions can be averaged and used for delay alignment and "best compromise" equalization.



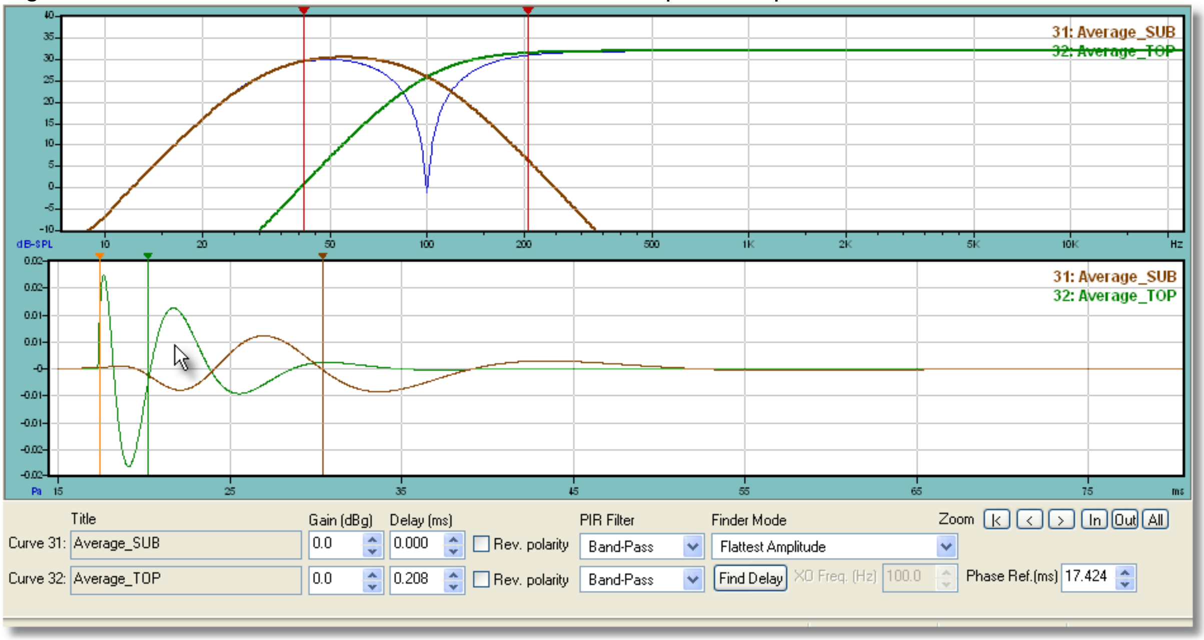
In this graph the measurements are made in three positions approx 3.4m apart and averaged. A sub-bass (brown curves) between 30 –100Hz and a Top-system (green) from 100Hz and up are displayed. In the frequency graph the complex sum (blue) of the two bands and the group-delay are shown. The group-delay is aligned to match at the crossover-point. However the sum doesn't sum up flat.
If the delay is adjusted for flattest magnitude, there is still a small ripple.



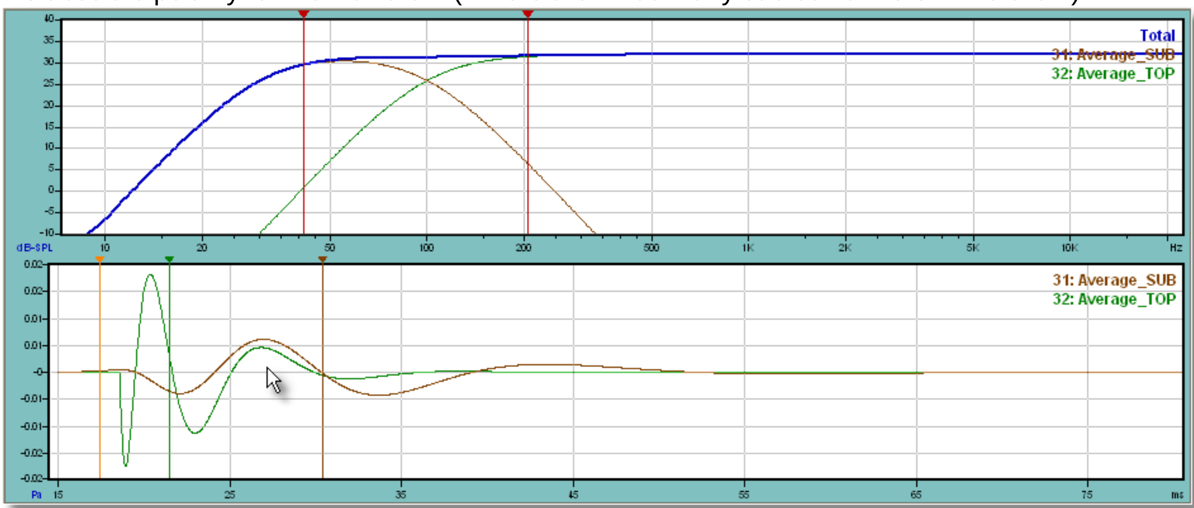
In this case no good match can be found unless one of the band is polarity reversed.



Both group delay and wrapped phase now aligns at the crossover point.
A faster way to come to this conclusion is to use the impulse response filtering and define a crossover region, in this case 50 –200Hz and look at the filtered impulse responses.



By trying to match the top-system’s second oscillation it’s clear that the match can only occur if one of the system is reversed. We prefer to reverse the top system therefore we hopefully have preserved the absolute polarity for the kick-drum (if there aren’t too many sub-sonic filters in the chain).



With the top’s polarity reversed we can achieve a perfect summation.

Conclusions

- Equalize and set the gain structure of top- and sub-system before any time alignment. Do new measurement after the equalization. Equalization will change the group-delay (and phase) and make the delay alignment invalid.

- Group-delay will be significant for sharp filters at low frequencies. Room-mode equalization requires high Q filters. Be aware where to place them in the signal chain. Preferable first in the chain, before the crossover and delay units.

- Use multiple windowing of the time domain to achieve an anechoic measurement. Otherwise there is a risk of aligning the delay to a reflection.

- To achieve an even quality over the audience area, use multiple positions averaging of the measurements. This also helps to attenuate reflections and to achieve an anechoic measurement.

- Adjust the start frequency for the FFT calculation so the phase calculation start well over the noise floor.

- Use ETC or Cepstrum to find a coarse value of the delay. ETC and Cepstrum are normalized, which helps to localize the energy.

- A minimum-phase system requires that the top-system be slightly before the sub-system in the ETC-graph.

- Use the complex magnitude summation to judge if the systems will sum up perfect. Any ripple will indicate that the delay is misaligned several turns.

- Group-delay curve matching can be misleading as high-pass filter and crossover filter add group-delays. The same problem applies to wrapped phase matching. On top of that it's hard to judge "extra turns" in the wrapped phase due to echoic IR.

- Wrapped phase can be used for the fine-tuning within a phase turn, to get the values of incremental samples.

- If filtered impulses are available, the oscillations can be matched as a fine-tuning of the delay. A rule of thumb is that the second oscillation in the top-system shall match the highest amplitude oscillation in the sub-system IR.

- Overall equalization can be done after delay the alignment, but only if it is made for all involved systems, before the crossover and delay units in the signal chain.

References

[1] Ohm G.S., "Uber die Definition des Tones..." *Ann. Phys. Chem.*, vol 59 & 62, (1843 & 1844)

[2] Helmholtz H.L.F. von, "On the sensations of Tone" *transl. AJ Ellis (Longmans, London 1885)*

[3] Rayleigh B. "The theory of sound" *vol. 2. 2nd ed. (Macmillan, London, 1896)*

[4] http://www.falstad.com/mathphysics.html

[5] D'Appolito J. "A geometric approach to eliminating lobbing error in multiway loudspeakers" *AES preprint 2000, (1983 Oct.)*

[6] Meyer J. "DSP Beam steering with modern line arrays", *Technical Report 2002 Meyer Sound Labs.*

Group-delay
[7] Zwicker and Feldtkeller “Das Ohr als Nachrichtenempänger” , *S. Hirzel, Stuttgart, 1967*

[8] Fleisher H. “Hörbarkeit von Phasenunterschieden bei verschiedenen Arten der Schalldarbietung” *Acoustica vol. 36 1976-1977*

[9] Fleischer H. “Gerade wahrnehmbare Phasenänderungen bei Drei-Ton-Komplexen” *Acoustica vol.32, 1975*

[10] Blauert J. and Laws P. “ Group Delay distortion in electroacoustic systems” *J Acoust. Soc. Am. Vol.63 1978 May*

[11] Preis D. “Phase distortion and phase equalization in audio signal processing – A tutorial review” *JAES, vol. 30. 1982 Nov.*

[12] Liphitz S., Pocock M. Vanderkooy J. ”On the audibility of midrange distortion in audio systems” *JAES vol 30. 1982 Sept.*

Hearing
[13] Gurnsey, “Hearing: Review Part 1” *AuditionReview.pdf (2003*)

Wavelets
[14] J. M. Combes et al, "Wavelets, Time-Frequency Methods and Phase Space", *Proc. Int. Conf. Marseille, France, Dec (1987)*

[15] A. Grossman, R. Kronland-Martinet and J. Morlet, "Analysis of Sound Patterns Through Wavelet Transforms", *Int. Journ. of Pattern Recognition and Artificial Intelligence.*, vol. 1, *No. 2, World Scientific Publishing Company, (1987).*

[16] F. Hlawatsch and G. F. Boudreaux-Bartels, "Linear and Quadratic Time-Frequency Signal Representations", In *IEEE Signal Processing Magazine*, *vol. 9, No. 2, (April 1992).*

[17] M. Vetterli and C. Herley, "Wavelets and Filter Banks: Theory and Design", *IEEE Transaction on Signal Processing,* vol. 40, No. 9, (Sep. 1992).

[18] P. D. Hatziantoniou and J. N. Mourjopoulos, “Generalized Fractional-Octave Smoothing of Audio and Acoustic Responses, “ *JAES vol. 48, no. 4 (April 2000).*

Cepstrum
[19] Randall R.B. & Hee J. “Cepstrum analysis” *Bruel & Kjaer Technical Review no.3, (1981)*

[20] Bauman P.D. Liphitz, Scott & Vanderkooy, “Cepstral techniques for transducer measurements” *AES preprint 2172 (1984 Oct.)*

ETC
[21] Heyser R. C. “ Determination of loudspeaker signal arrival times” *JAES vol.19 (1971 Oct)*

Impulse response
[22] J.M. Berman & L.R. Fincham, “The Application of Digital Techniques to the Measurement of Loudspeakers”, *JAES vol. 25, no. 6 (June 1977)*

[23] D. Preis, “A Catalog of Frequency and Transient Responses”, *JAES vol. 25, no. 12 (December 1977)*