The Engineer's Guide to the Loudspeakers' Modelling

1 INTRODUCTION

It is not your fault that you don’t like nor understand mathematics nor how to use it in everyday endeavours. The teaching of math in schools and universities is a sadistic torture on painstaking memorizing of meaningless texts full of otherworldly symbols. There is absolutely no sense in learning endless definitions, lemmas and theorems coming from a few “first principles” which commonly include pathetic “obviously that”. Professional mathematicians can easily provide a few examples of functions like $\exp(-1/x^2)$ that turn most if not all of these proofs into a laughing stock.

The math is not a science but an instrument of problem solving. Math was not developed by abstract theologists. Geometry, Calculus, Fourier transform, etc were invented by problem solvers for solving very real practical problems. It was only later that it was discovered that the same instruments can be used for solving other, seemingly unrelated, problems. It was even later when some other intellectuals figured out a way to generalize these instruments, and combine most of it in a structure, following a common human desire to classify everything around us. Playing with this structure is an intellectual game like chess, go, or modern theoretical physics. Advanced brain gymnastics - yes. Science – no[1].

- Deriving a theory with math-rich formulations does not make the theory scientific.
- Science does not have a concept of authority, including “first principles”.
- A scientific theory must be derived from repeatable observations and reproducible measurements.
- A scientific theory may not contradict any reliably established fact of nature. It does not matter how many other facts it explains. A single contradiction invalidates any theory, however “accepted and established” it was.
- A scientific theory must not pretend to be extrapolatable beyond the set of observations, nor beyond some finite precision. No exact predictions of future, please.

The history of developments of audio recording, transmission, and reproduction skims is a shining display of both human ingenuity and human stupidity. Some of the devices, developed with a logarithmic scale ruler, are pure marvels. Some of the theories pretending to explain how and why these marvels work so well are, at the best, unrelated to the subject. Generally, the theory is hopelessly lagging behind the engineering practice… as usual.

It must be noted that we, humans, are pathetically bad with non-linearities. For linear systems, we have many elegant instruments. Whenever and wherever the systems become non-linear, we are at loss. The lack of an adequate theory of loudspeakers’ LTI distortions is not an awkward fault to be ashamed of but one of the many examples of this general rule, one to be acknowledged and admitted without feeling a loser.

A decent guide shall be a trilogy in five parts. This is a part I, which covers only few aspects of loudspeaker modelling, something simple to start with.

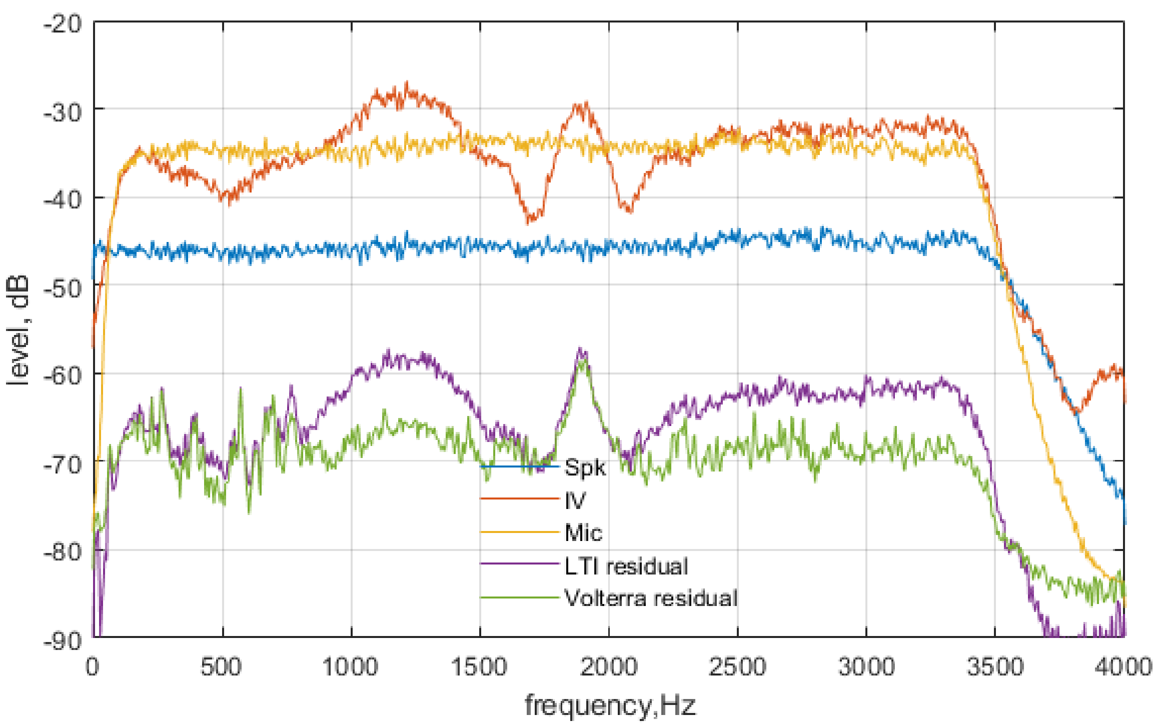
[1] The purely mathematical apparatus of Computer Tomography etc was developed by astronomers who studied remote planet systems by observing the periodically flickering dots on the sky. The pure mathematicians came later … to provide needless proofs that the Radon-2D-FFT/window/IFFT based back-projection algorithm is sub-optimal, and that algebraic methods suck.

2 OBSERVATIONS

Some anonymous pretenders boldly claim that "*It is well understood where the distortion is coming from, even at relatively low SPL: Nonlinear force factor Bl(x) and inductance Le(x), Le(i) of motor assembly (voice coil, iron path, magnet), nonlinear stiffness Kms(x) of mechanical suspension (surround and Spider), nonlinear losses Rms(v) of mechanical and acoustical system, Partial vibration of the radiator's surface (surround, cone, diaphragm, dust cap), ...*" If it is so WELL understood, why nobody so far could predict these distortions and pre-distort the output, same as in RF amplifiers in smartphones? Solely for the lack of trying? Not at all.

In reality, any smooth polynomial model can be expressed via a series of Volterra kernels. The same kernels can be found with System Identification tools by direct observation of the system's inputs and outputs. If Volterra kernels work, you can proceed to building an elegant theory why they do, and how to express them with less parameters based on some simplified physical theories (or pseudo-theories, which does not matter as far as they work). If Volterra kernels don't work – forget about it.

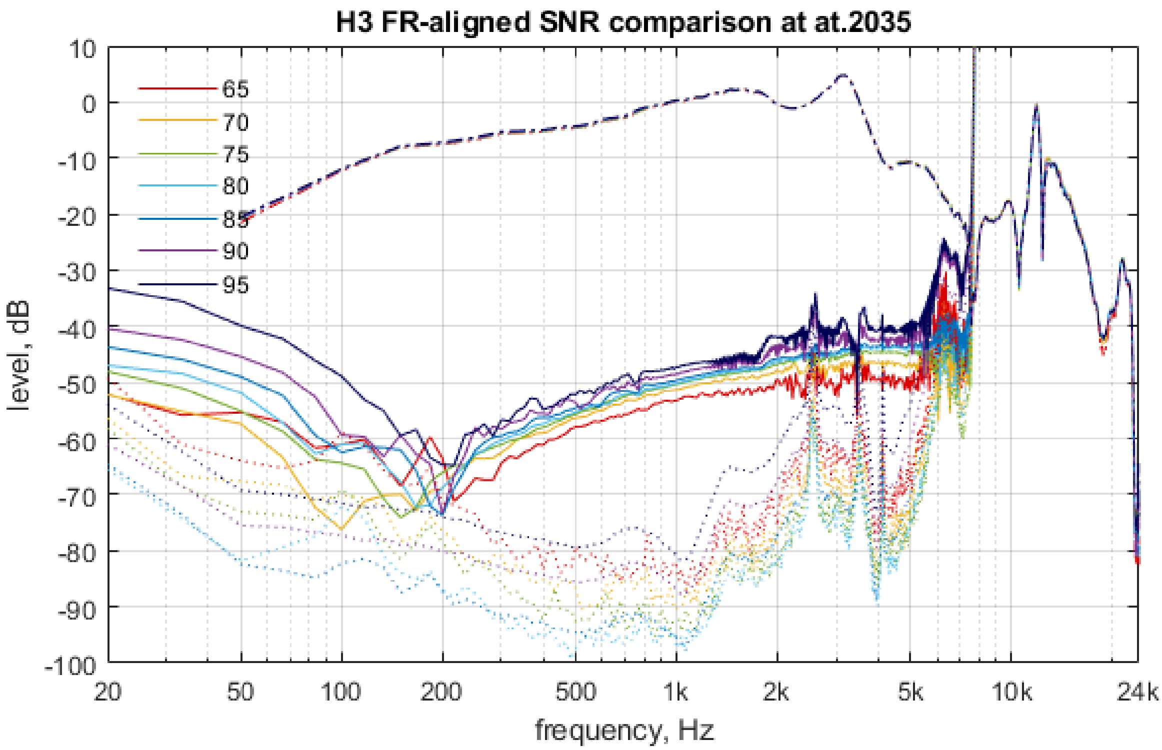
There have been many unsuccessful attempts to adapt Wiener-Hammerstein or 2nd / 3rd order Volterra kernels to remove LTI distortions, made by professional folks (see ICASSP proceedings) since the 90s, when computers became powerful enough to make required calculations. Every few years, a new entrant[2] to the field makes another attempt … with the same result as reported beforehand.



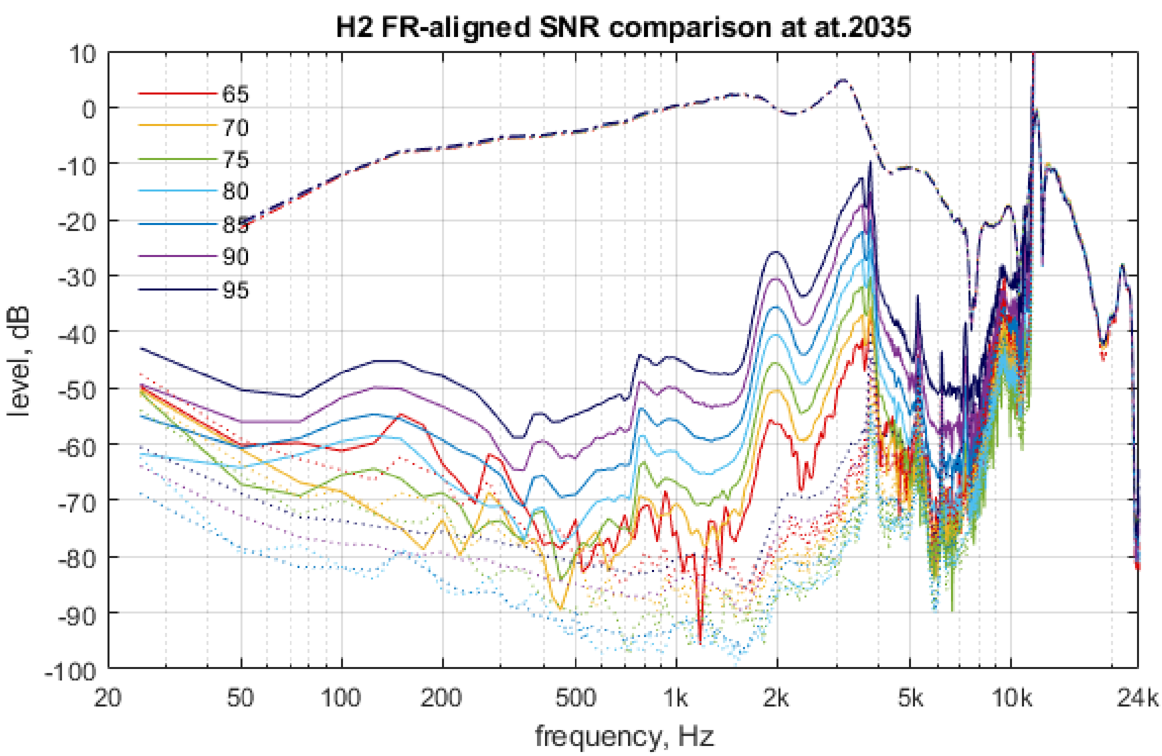
It may work but only for a certain driver in certain range of amplitudes, frequencies and excitations, to a small degree like 8dB… and miserably fails at the tiniest attempt to generalize it.

The most puzzling in the "practice-vs-theory" observations is H3 aka the 3rd harmonic. Below, solid lines are relative harmonics, dashed lines are FR, dotted lies are measurement confidence interval levels:

[2] including myself



At the same time, H2 aka 2nd harmonic, behaves "perfectly":



- In theory: As the excitation level grows from diminishingly low levels (surely a linear system), in full agreement with generalized Taylor expansion, if the excitation level rises by X dB, relative H2 must rise by the same X dB. The H3 should have risen by 2X dB.

- In practice: H2 ok, but H3 stubbornly refuses to obey our so elegant time-proven academy-accepted theories. Obviously, H3 mischiefs solely to annoy us… or, alternatively, H3 hints that we need to finally start using our brains?

3 MEASUREMENTS

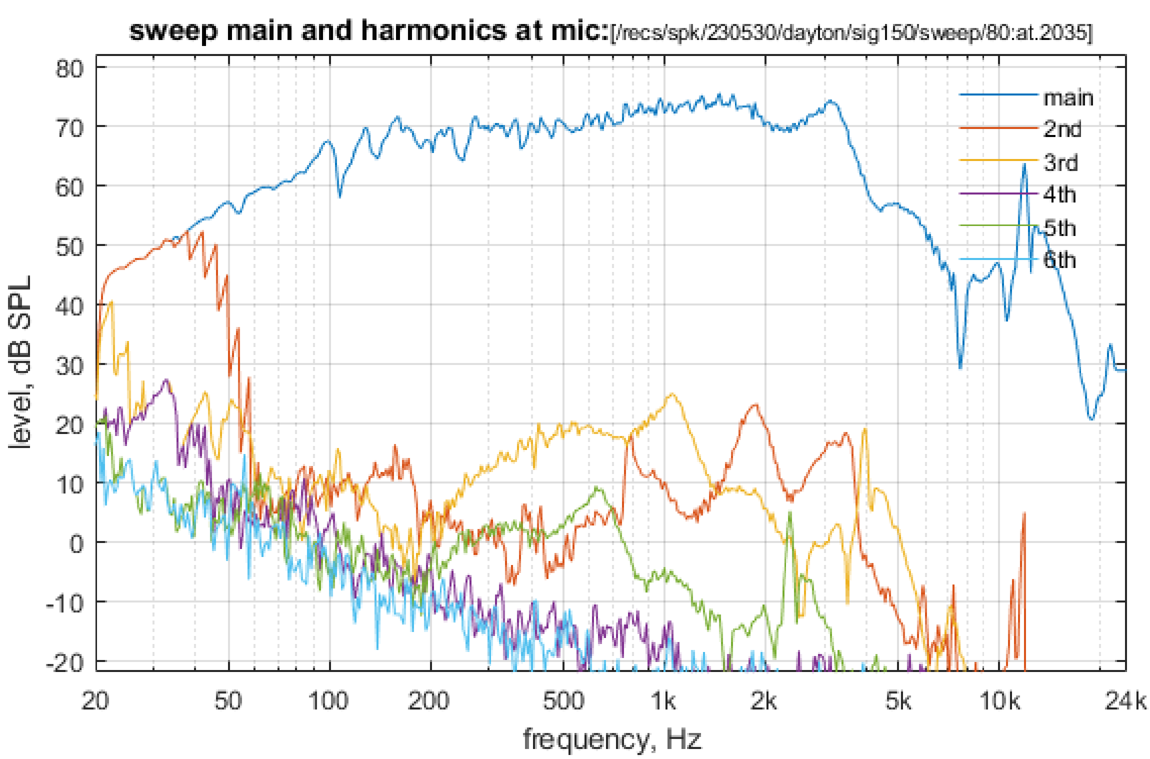
Without good relevant measurements, your task is hopeless. Let's go step by step.

3.1 PRECISION ESTIMATE

First of all, proper measurement shall always include an estimate of its precision. If you don't include it within your toolset, quite a few users will start making multiply repeated measurements to figure out the precision "manually". These users are not properly trained experimental physicists but professionals with deep understanding of the order of importance. Users are not as stupid as it may look at the first sight, some of them will inevitably turn out to be clever than me.

3.2 APPLES TO APPLES

If your loudspeakers are ruler-flat and you have an anechoic room, you can measure non-linear distortions using Angelo Farina's method[3][4]. It is a bit harder when you measure a non-flat driver in an echoey room. Usually, you get something like that:

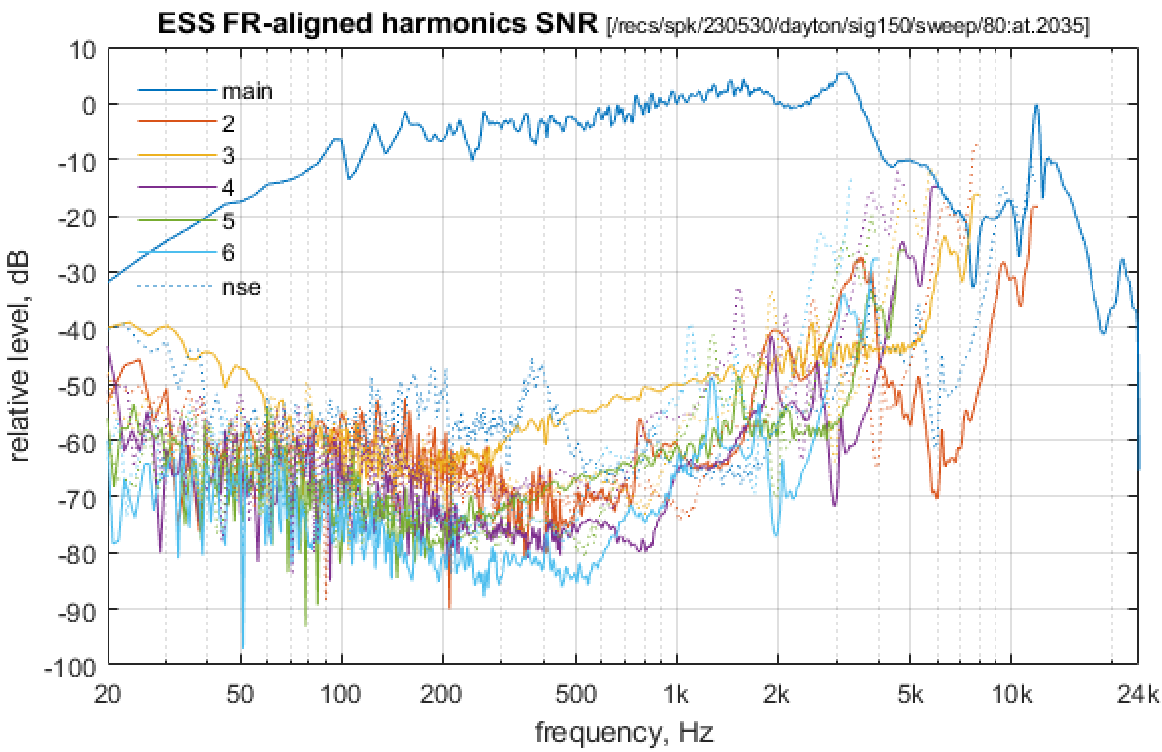


[3] For proper regularization of the Sine Sweep method, see [Kernel based approaches to linear system identification, by Gianluigi Pillonetto](#).

[4] MLS / Chirp / Sine Sweep were invented by Sidney Darlington in 1947 while he was working on radars in AT&T. As any genius invention, it was borrowed /stolen / reinvented countless times. Yet, he shall be distinguished as the original inventor.

You look at it and ask yourself – how do these wildly fluctuating curves fit into such small voice coil assembly? You'll need a very long IR to approximate them, won't you? Then, each tap on 48kHz is equal to 7mm in air. 100 taps, and you get 700mm. 1000 taps, and you get 7m. How do you fit 7m into a 25mm diameter voice coil?

The problem is that at a frequency f, you see H1(f), H2(f/2), H3(f/3), etc. Thus, you need to shift the Hk curves forth, subtract them from the H1 at the right place, align back (or not), and get something like that:



- ✓ This finally makes (more) sense because a sum of a low-pass and a high-pass in H3/H5/etc can be realisable by a physically meaningful short IIR.
- ✓ Add a proper denoising, and you'll get the picture of H3 two pages back.
- ✓ As H3 changes smoothly and slowly across frequency range, you don't have to sample the detailed measurements each 10Hz. Relatively few points shall be sufficiently representable for analysis.
- ✓ You do not need an anechoic chamber.

3.3 Resonances

- It is relatively easy to exclude room resonances – move your measurement mic sufficiently close to the driver & window IR – but not as close as to exceed mic non-linearity allowances.
- A bit harder to exclude cone and box resonances. Box resonances are high-Q, usually around 1…2kHz, and related to the standing waves inside as per box depth.
- Cone resonances are (hopefully) beyond the range of your interest.
- It is a bit harder to deal with Helmholtz resonances because some speaker designers love them. They add tunnels inside the core to cool the voice coil and use non-breathing cone caps. You'll have to modify the driver.

3.4 Noise

There are no microphones that are ruler-flat and low-noise, these are mutually exclusive.

You'll need a ¼" "measurement" microphone (~30dBA) and 1" low-noise condenser (Rode NT1 5th generation is 4dBA, low cost compared to B&K, highly recommended) which you shall equalize to fit the measurement mic FR at each specific measurement distance.

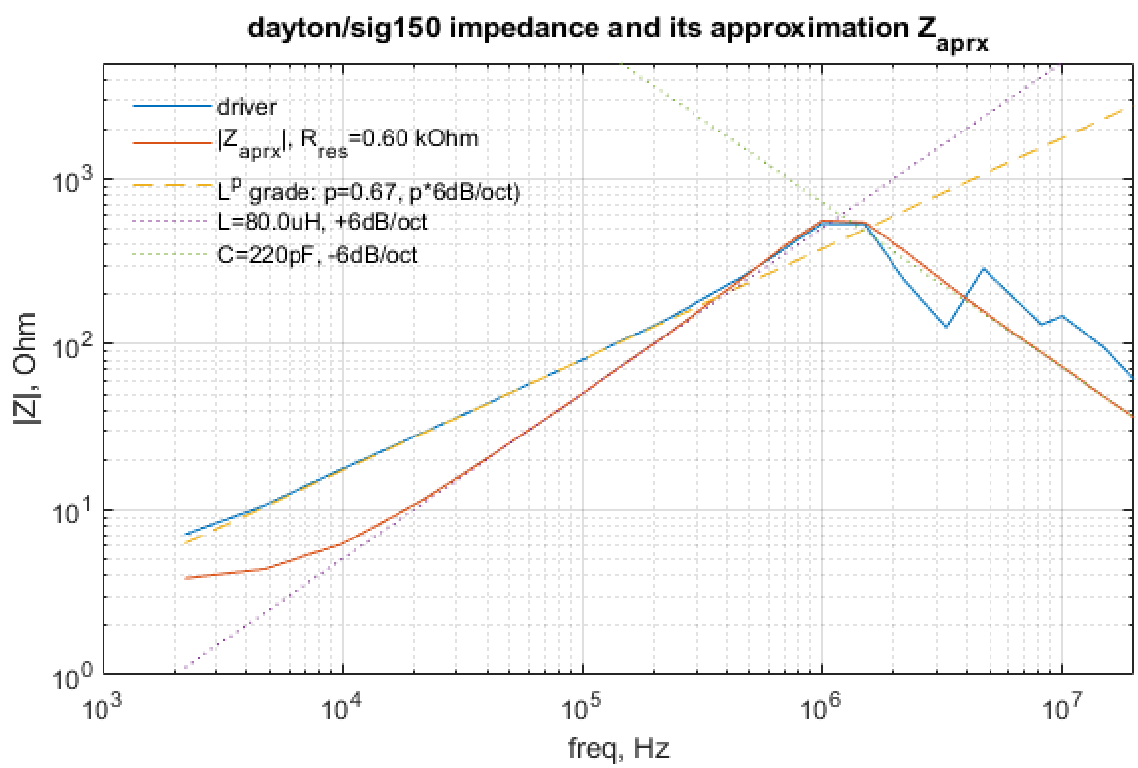
If the 1" condenser microphone lists 145 dB SPL as 1% threshold, you can count on having <=0.01% microphone distortions below 105 dB SPL.

- ✓ The low-cost ¼" electret "measurement" microphones are often ruler-flat but generally unusable for Hk measurements because their 1% distortions ceiling is about 105…110dB SPL. You can never be sure where the distortions are coming from.

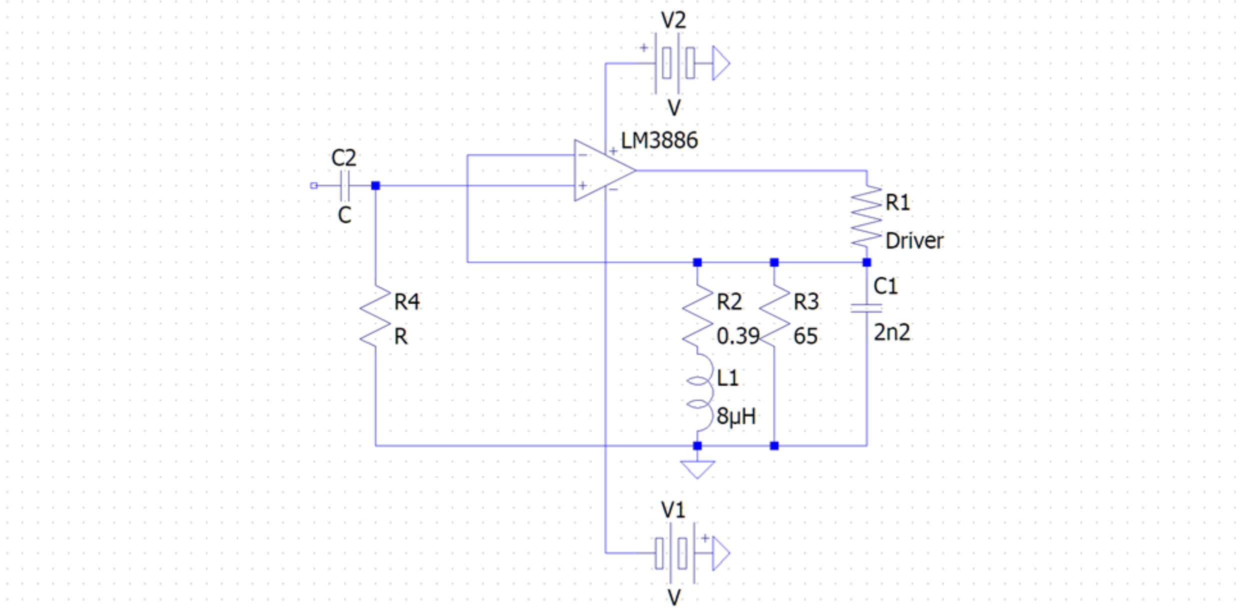
3.5 Current Drive

Modelling voltage drive loudspeakers is a hopeless endeavor. A voltage-driven driver is not a time-invariant device because it the voice coil DC resistance is not temperature-invariant. The voice coil temperature varies wildly and is out of control / observation. Making repeatable measurements becomes such a problem due to the dependence on previously experienced power, especially for tweeters, that you need to solve this problem first before working on anything else.

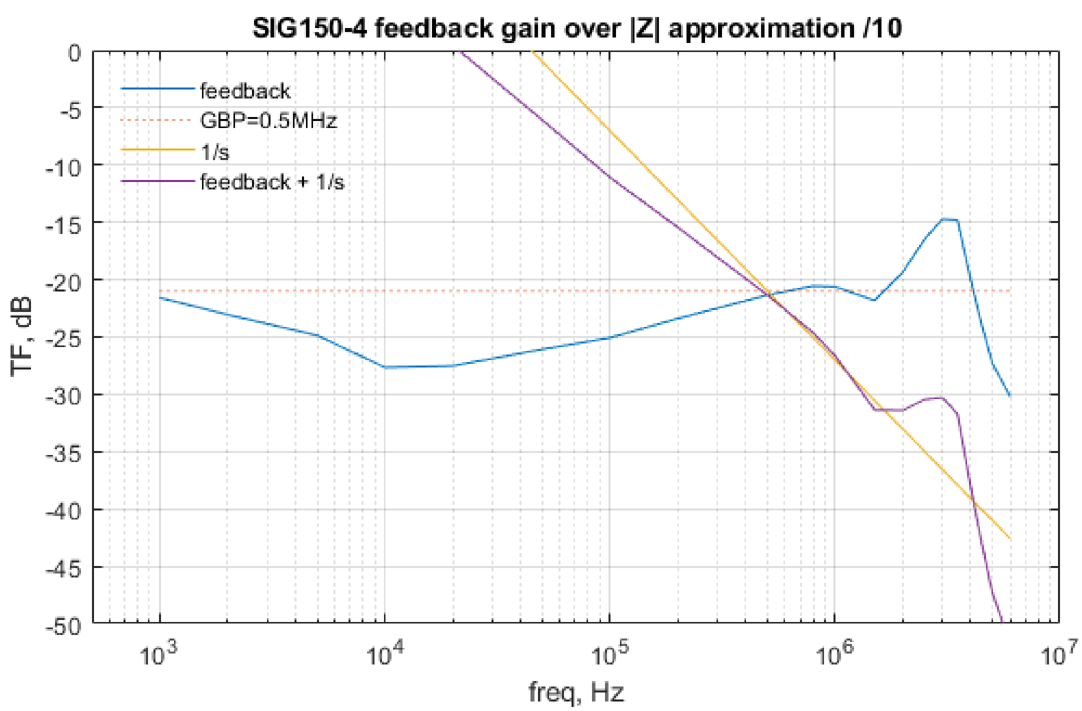
Current drive is a voltage drive amplifier with feedback on current sensor, which becomes "transconductance amplifier". The main obstacle is getting sufficiently deep feedback to consider it current drive. You'll need to measure the driver impedance from, say, 10kHz to 10MHz, something like that:



Then you construct the current sensor so that it mimics the driver's impedance in, say, 1:20 scale, and change the power amplifier feedback circuitry like that:



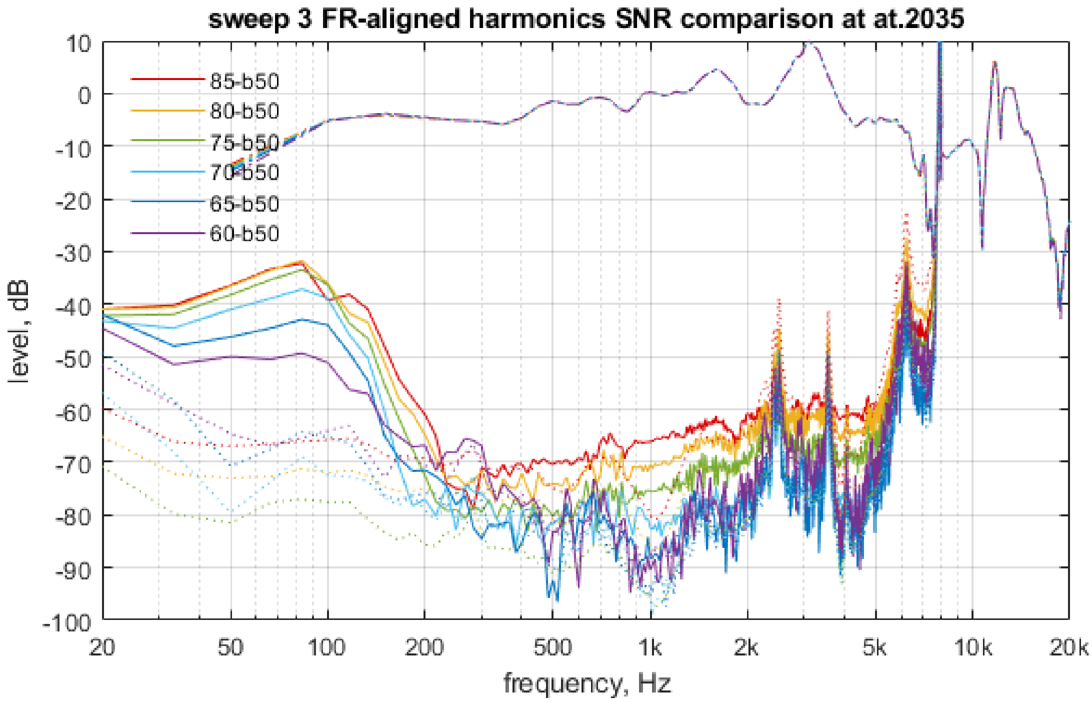
The measured feedback circuit gain shall look like that:



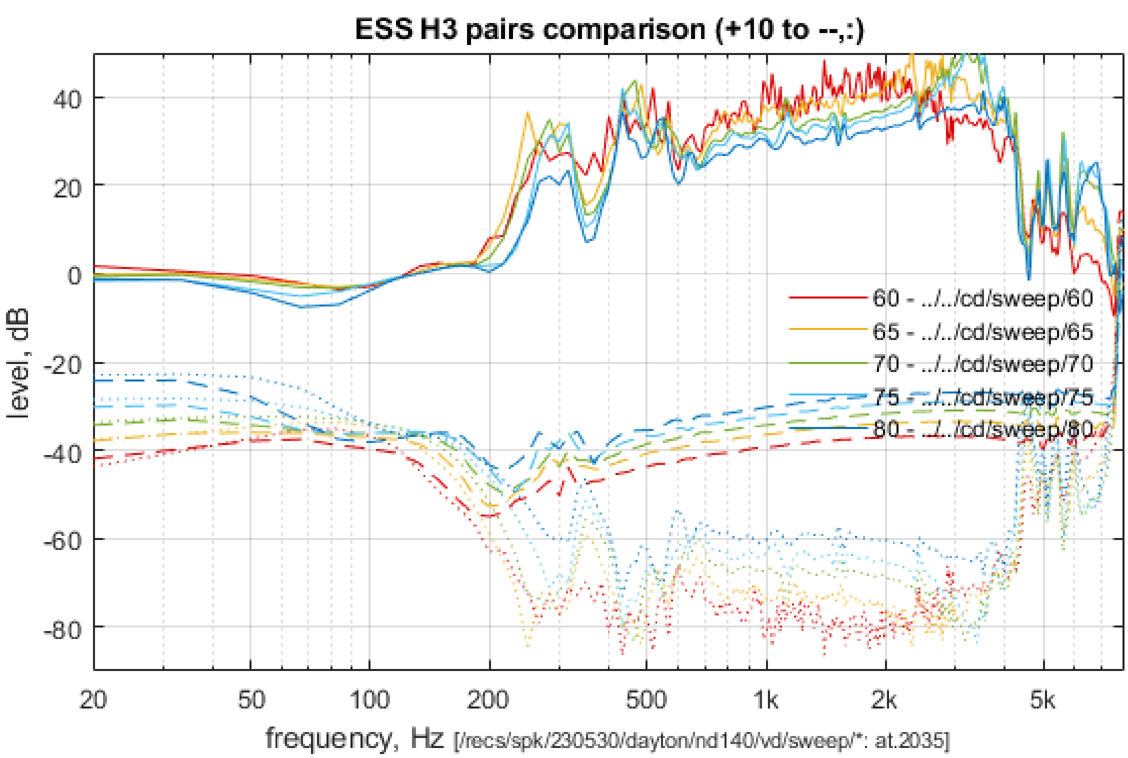
The amplifier shall be reasonably close to the driver to avoid ringing. Active loudspeakers are the way to go. The amplifiers with higher f_T provide lower distortions but you need to have a good understanding of control theory to dive into it. Adding nested feedback loops is obviously the next step in making the current drive's feedback deeper.

If you know of a better way to implement the current drive of loudspeakers, please let me know.

Drop of H3 by 20 dB or so is quite normal:

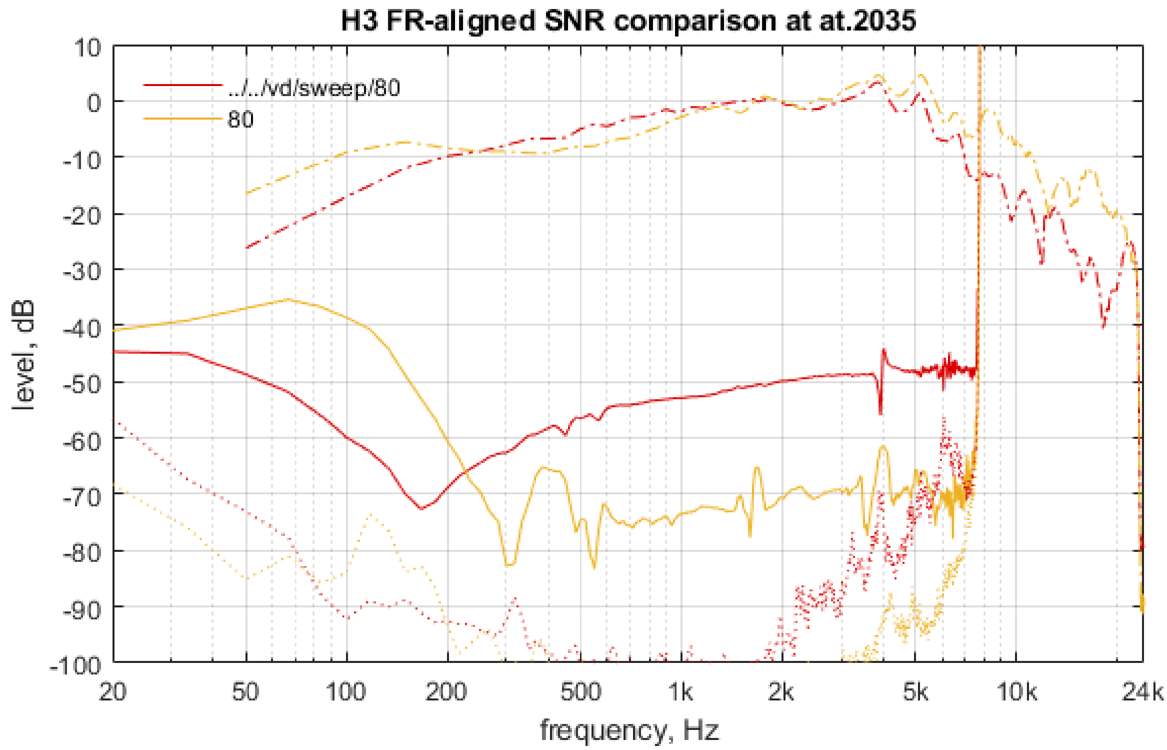


In some cases, the change of H3 is very dramatic, 40dB+:

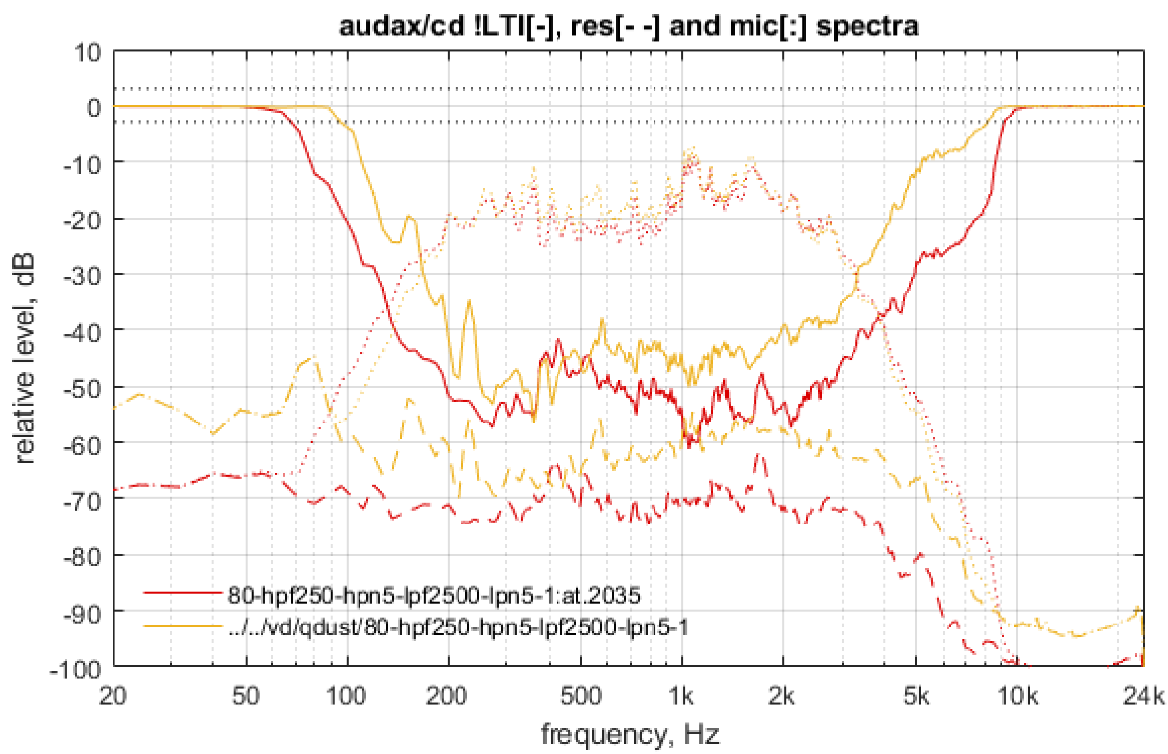


Here the dashed lines show H3 in voltage drive (for SPL varying from 60 to 80dB), dotted lines show H3 in current drive, and the solid lines show the H3 drop from voltage to current drive.

Here is the comparison of Audax HM130Z10 in voltage and current drive:



The harmonics by themselves are nearly harmless for single pitch instruments. However, for a complex sound they spread IMDs all over, the higher the order, the wider (see yellow dashed line below). "Wall of sound" recording technique was invented specifically to deal with high distortions in consumer loudspeakers. Musicians had to "fill the spectrum" to mask inevitable distortions at the user's end. Till now we experience the aftermath of the poor quality of loudspeakers in the 60s and 70s for many recordings of excellent musicianship were irreversibly polluted.

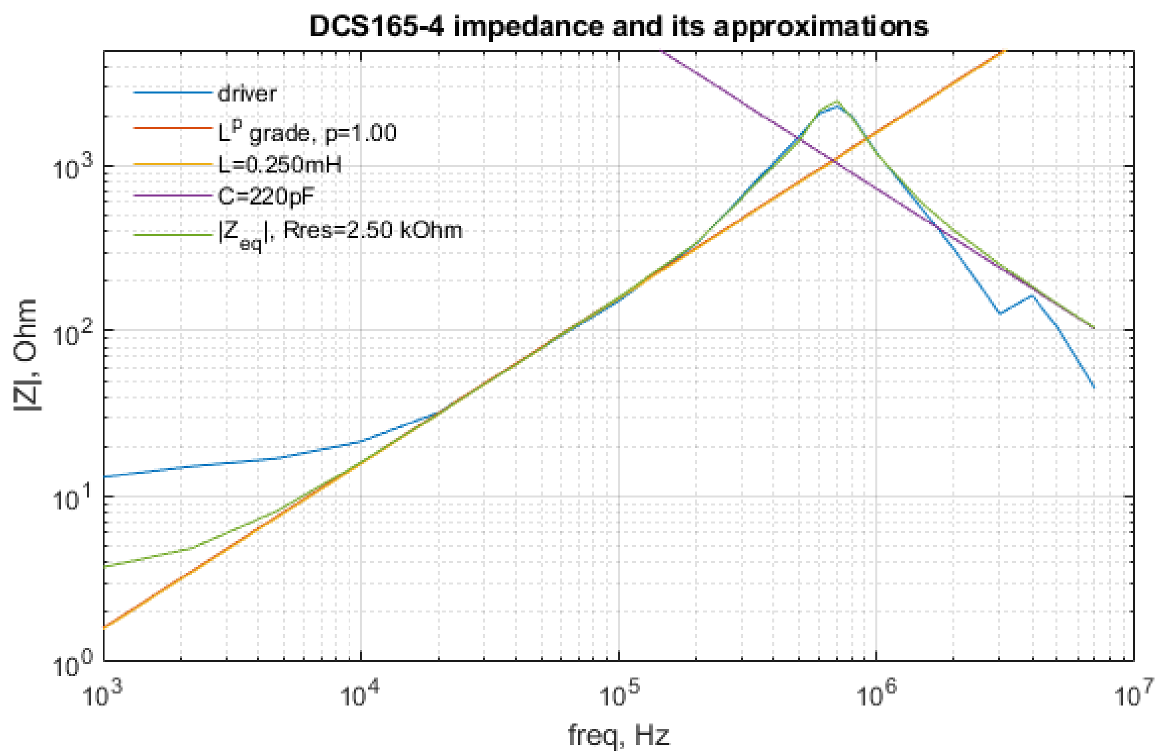


- ✓ In current drive, most drivers exhibit lower distortions - but some do not change much if any.
- ✓ In current drive, for most drivers, H3 drops with the SPL level[5] while it was nearly constant in voltage drive.
- ✓ In current drive, the distortions drop more for midranges than for woofers and tweeters but devil is in the details. You need to know where and how to measure the drivers implicitly designed for voltage drive.
- ✓ In current drive, the measurements become at least an order of amplitude more repeatable and reproducible. Not perfect yet☹

3.6 SHORTING COIL

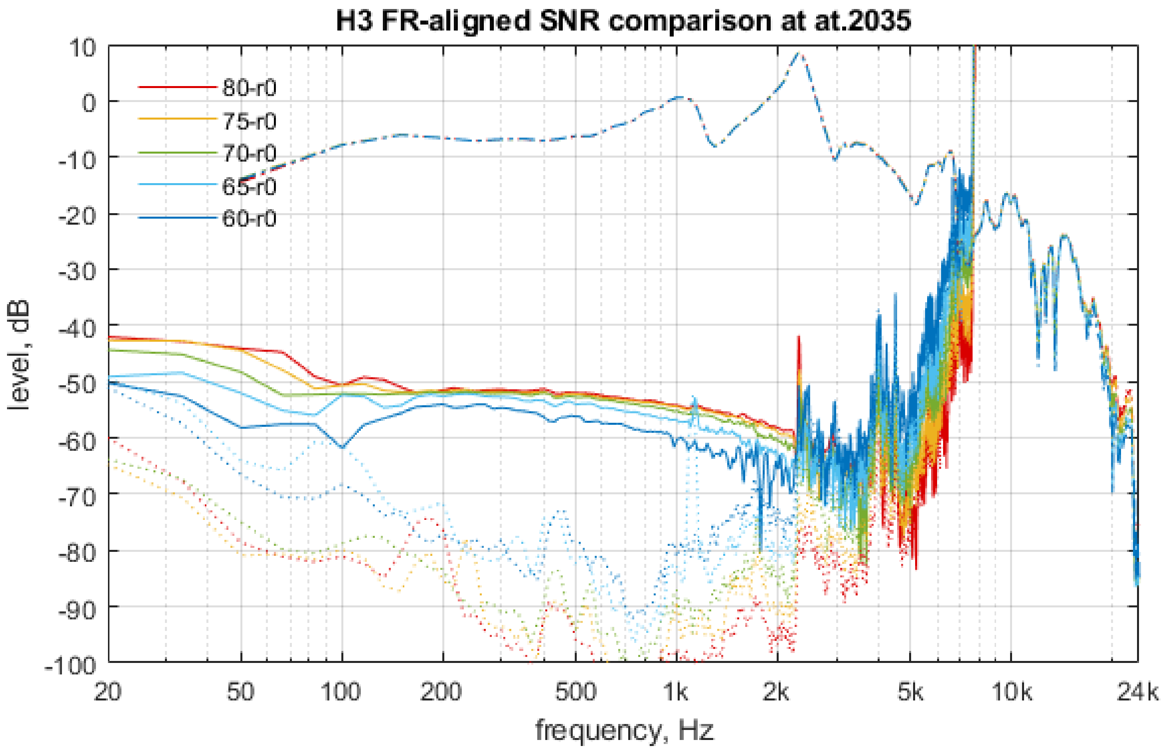
The changes in the voltage vs current drive behaviour depend on the shorting coil, more precisely, on the amount of it. Shorting coil is effectively an internal-to-driver feedback loop exploiting the linearity of electric conductance, with some energy losses in the shorting coil (ratio unclear).

There is a connection between the shorting coil "amount", slope of driver's impedance and the H3 slope, either in voltage or current drive. Let's consider Dayton Audio DCS165:

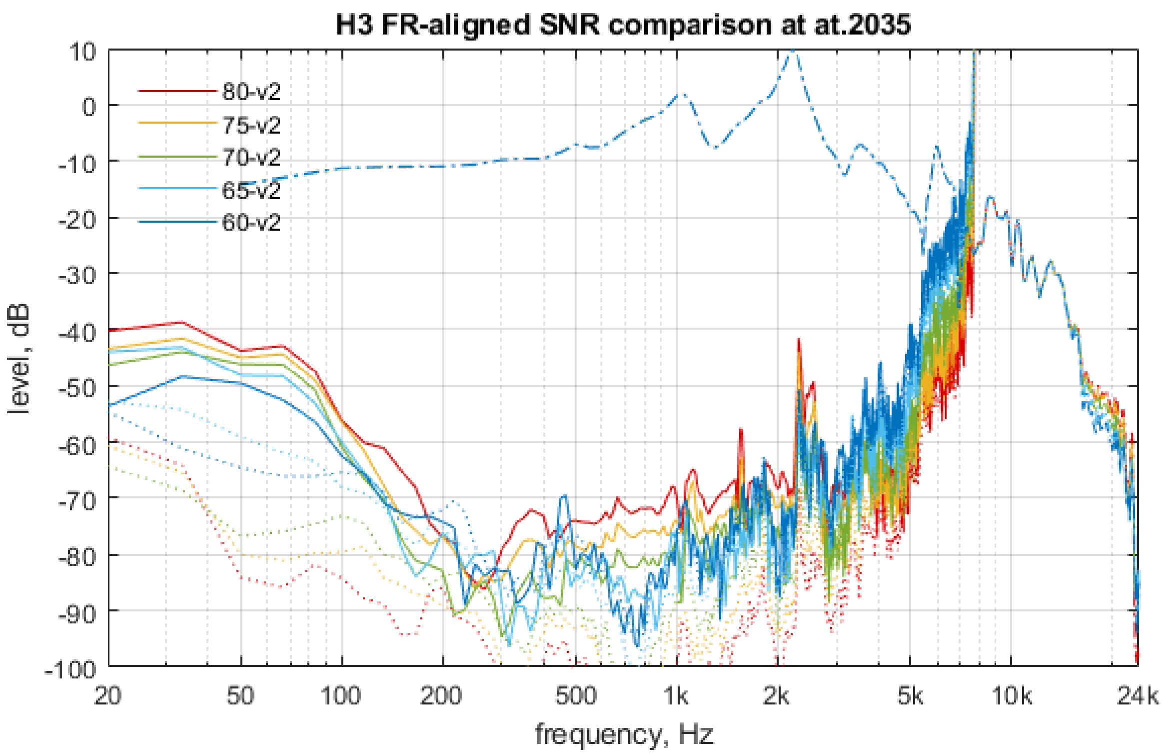


[5] … (my 2¢) also making sound smooth and sweet, Focal-like. I have no idea how Audax/Focal engineers achieved such phenomenal quality in voltage drive and I raise my hat in appreciation for their ingenuity.

In voltage drive, H3 is flat and goes down with increasing frequency:

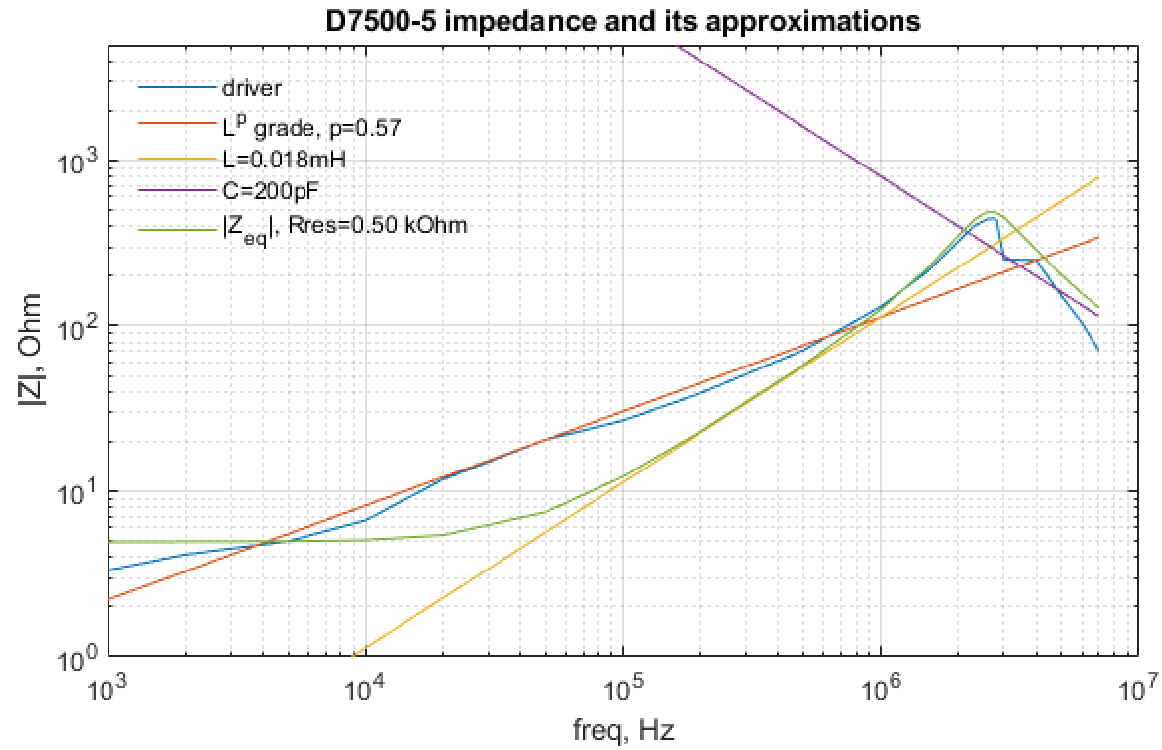


In current drive, H3 is going up with increasing frequency:

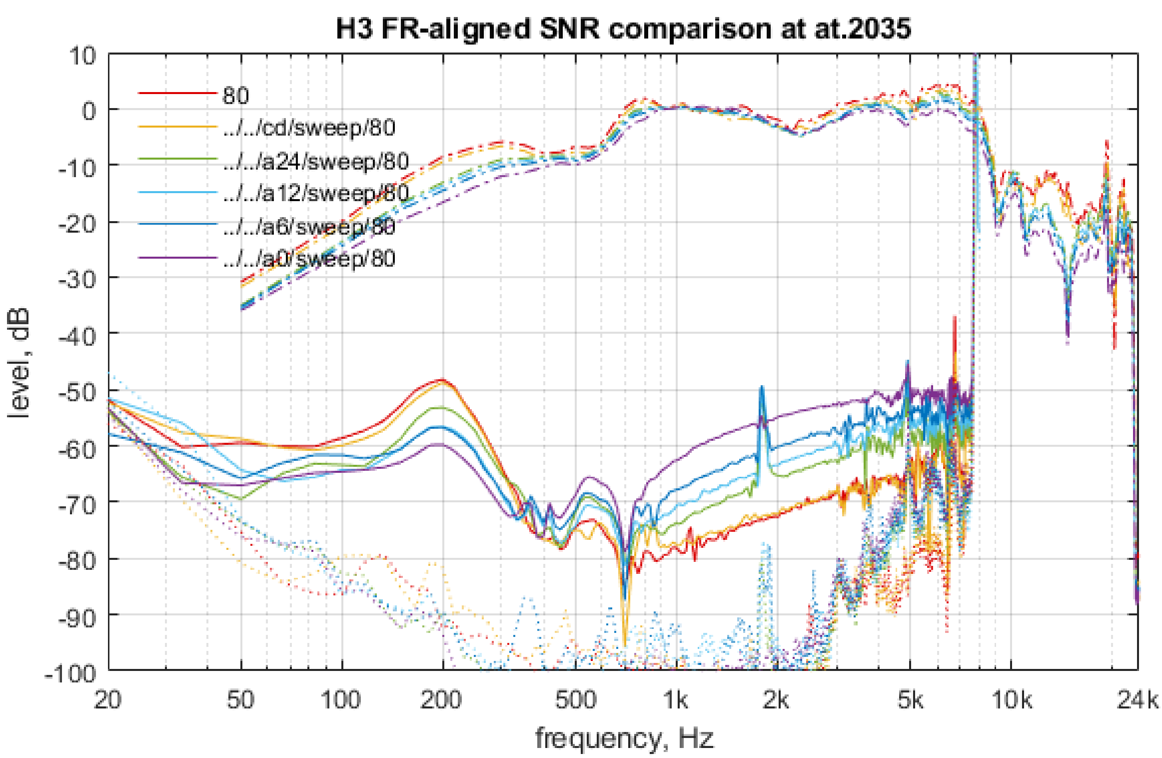


H3 is lowered by 30dB or so, basically, down to the noise level.

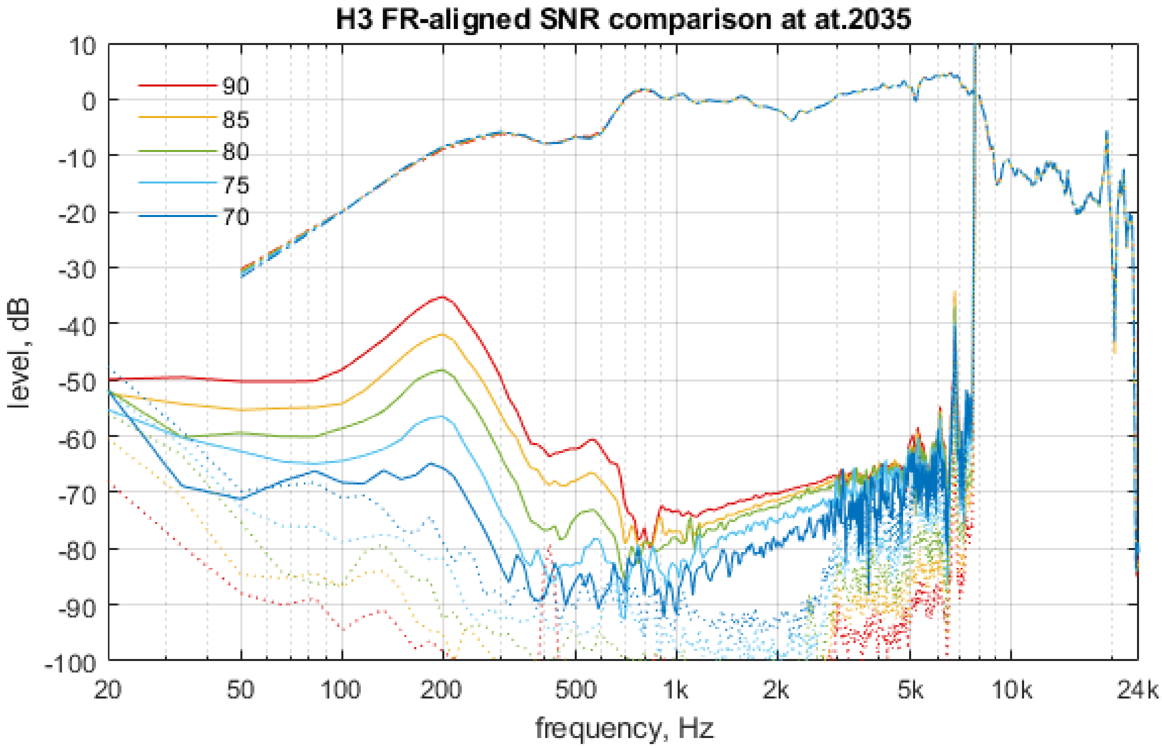
The behaviour of drivers with more massive shorting coils (lower p) is quite different:



Here we can see the transition of H3 from a curve to the straight line + 6dB/ octave, with adding a series resistor 6/12/24 Ohm (simulating deepness of current feedback) and then going into pure current drive, with two types of amplifiers (red line is for deeper feedback):



In current drive mode, H3 does tend to become a straight line, whatever SPL a driver produces (90 dB SPL is bloody loud):



By now, I have quite a bit of material on the "black-box" research re shorting coils but without being able to look "inside" the driver, it takes clearly disproportionate amount of efforts.

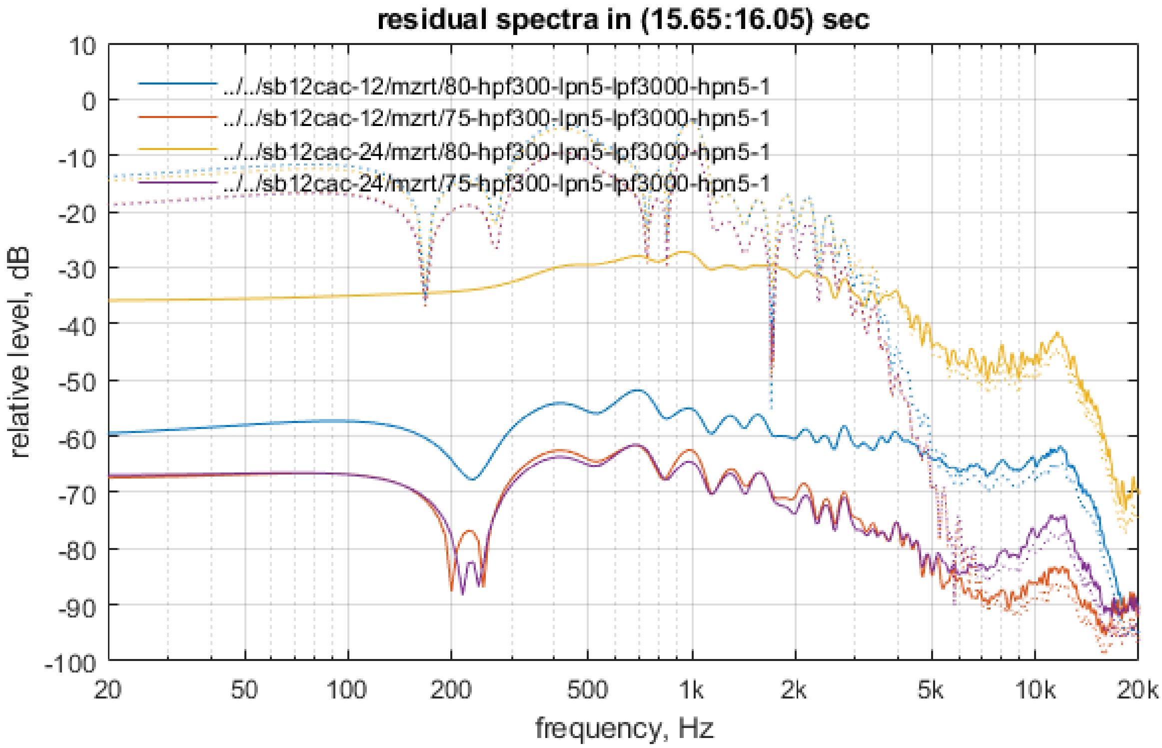
- ✓ Further "black-box" research into shorting coils appears unproductive
- ✓ Shorting coils drastically affect the complexity of the adequate loudspeaker's model
- ✓ To start with, you need something simple, and go step by step.

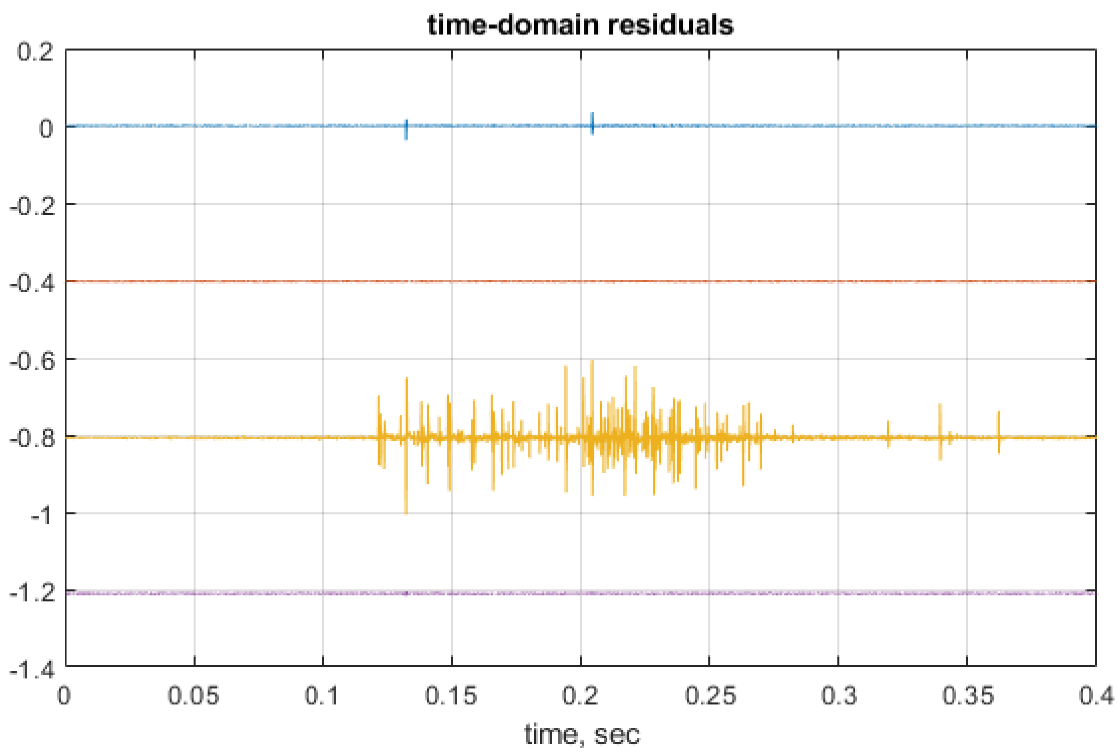
3.7 BARKHAUSEN NOISE

As far as I know, Bernd Häusler was the first who noticed and properly identified the effect about 20 years ago. Lars Risbo succeeded to identify Barkhausen noise while running mere sine sweeps. These people are truly remarkable. Unfortunately, both did not have proper instrumentation to separate the Barkhausen noise from the main and other distortions (nobody had).

Now it is easy to observe, separate, and measure it with [Loudspeakers for AEC: Measurement and Linearization - File Exchange - MATLAB Central]

The Barkhausen noise looks like a delta function shaped by FR. Here, SB12CACS25-4 playing tutti in a Mozart piano concerto:





From the theory point of view, hysteresis is the cumulative effect of Barkhausen noise. You have to be fully aware of the effect due to its stochastic nature, for it changes unpredictably from a run to a run. You need a lot of averaging to ensure repeatability and reproducibility.

- ✓ The effect strongly depends on the technology of the magnet and the magnetic circuitry.

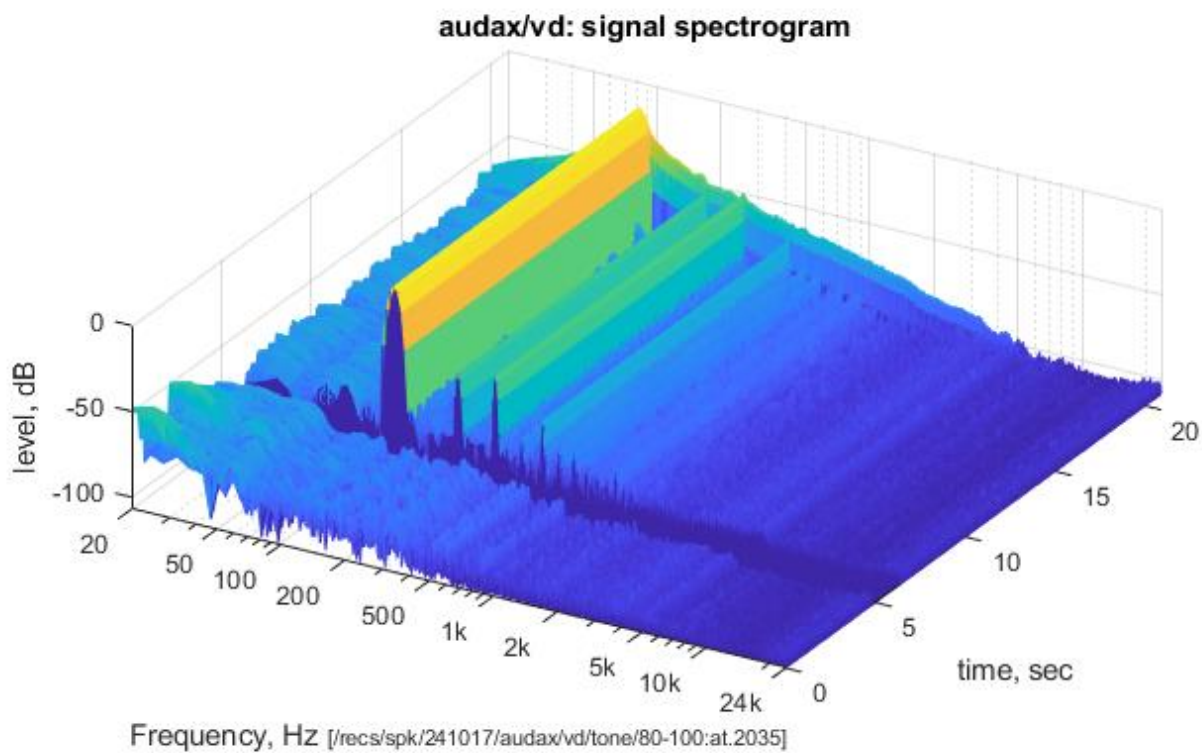
- ✓ So far, the Barkhausen noise has been observed to be lower in Nd magnets.
- ✓ So far, Barkhausen noise has been observed to be much lower in current drive mode.
- ✓ Again, further "black box" research is unproductive for same reasons.

3.8 Long Term Averaging

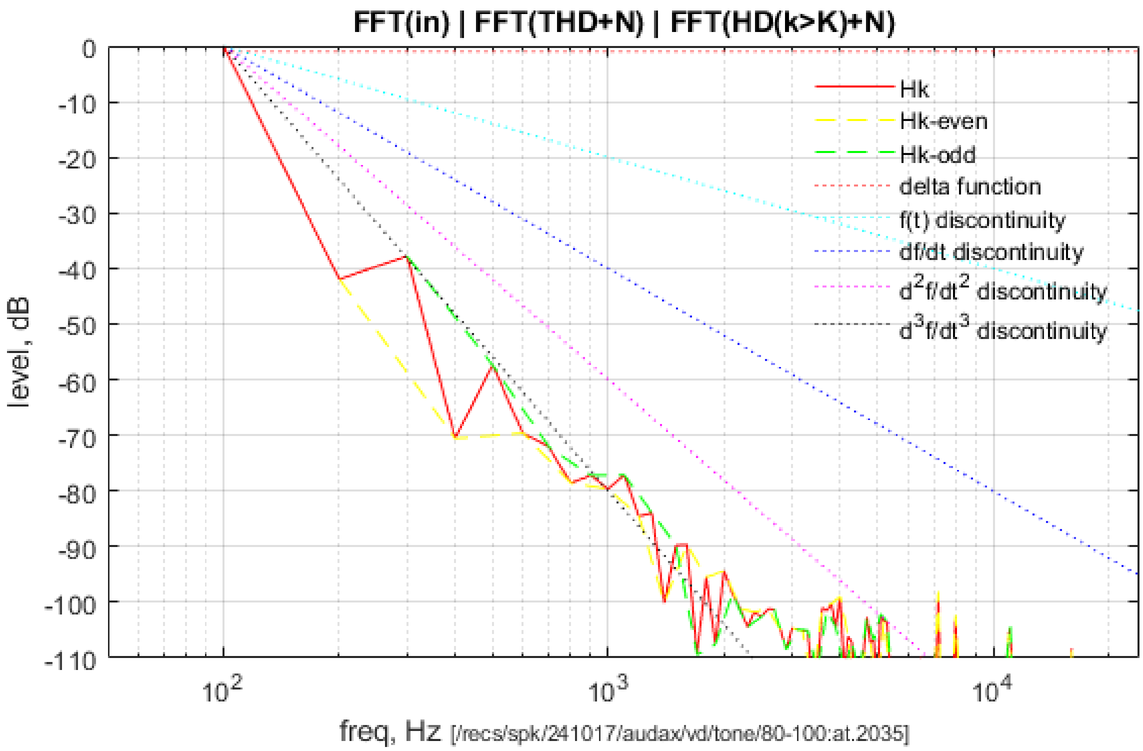
You surely learned in university that the noise's effects can be decreased with increasing the *N* number of observation as *sqrt(N)* and forgot it as a nightmare as soon as you exited the exam room. If the professors were teaching math as an instrument, you could have remembered and used it. Let's put all together:

- Low noise 1" microphone
- Smooth curve for H3, therefore you need to check only at a few points
- Reproducible measurements
- Long term averaging

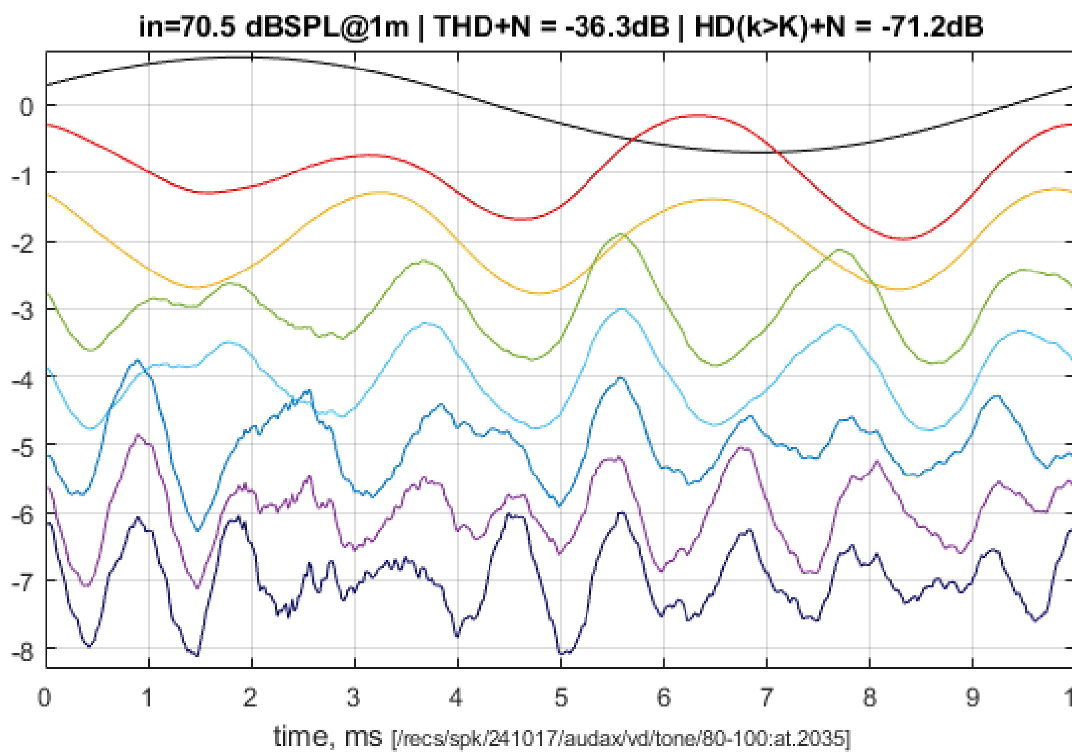
… and we get a novel experimental setup nobody tried before. What could we observe and measure? Let's look at the Audax HM130Z10, voltage drive, driving 100Hz on approximately 70dB SPL for 15 seconds. Using traditional toolset, we can have a look at the spectrogram (few rows removed to make it more transparent):



Using the proposed novel toolset, we sum all the 10ms periods and average the noise. As you could have remembered, the speed of harmonics' decay indicates where and how the time-domain function becomes discontinuous. If the function itself, it's *1/f*. If the first derivate, it is $1/f^2$. If there is a spike (like Barkhausen noise or Dirac delta function), the spectrum is flat. Same for the white noise. Now you can see why this driver has been thought of so highly since mid-90s:

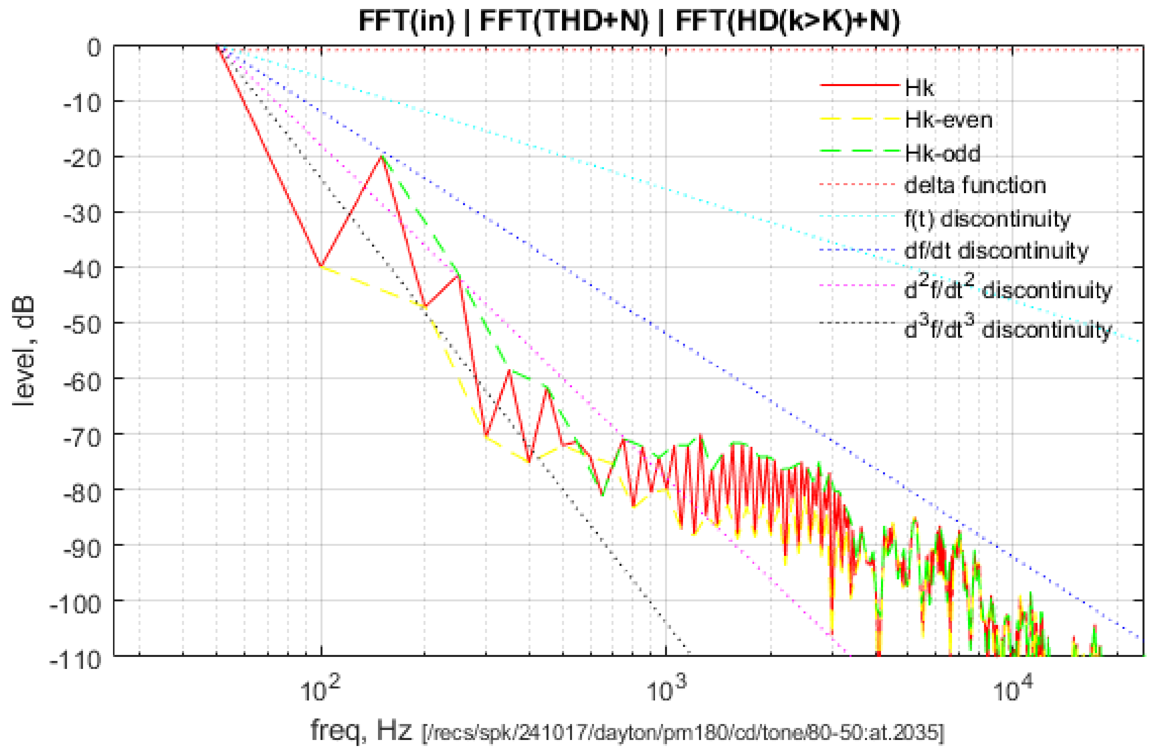


This way we get about 100 dB SNR relative to 70 dB SPL, i.e., -30 dB SPL noise. The time domain distortions are not so interesting:

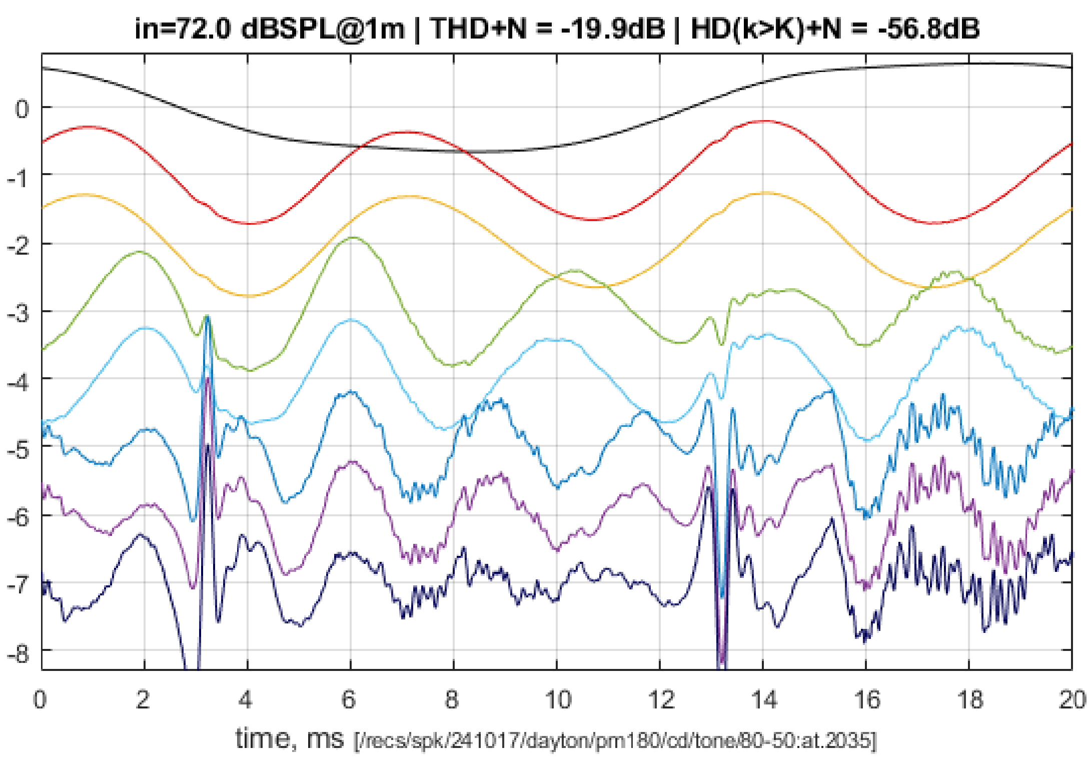


Here we see the main in black, and the residuals after subtracting the 2nd, 3rd, etc distortions, with color changing as in rainbow, red for the lowest, purple and dark-something for the highest.

For a change, let's consider the Dayton Audio PM180 driver, in current driving mode, 50 Hz on 70 dBSPL:

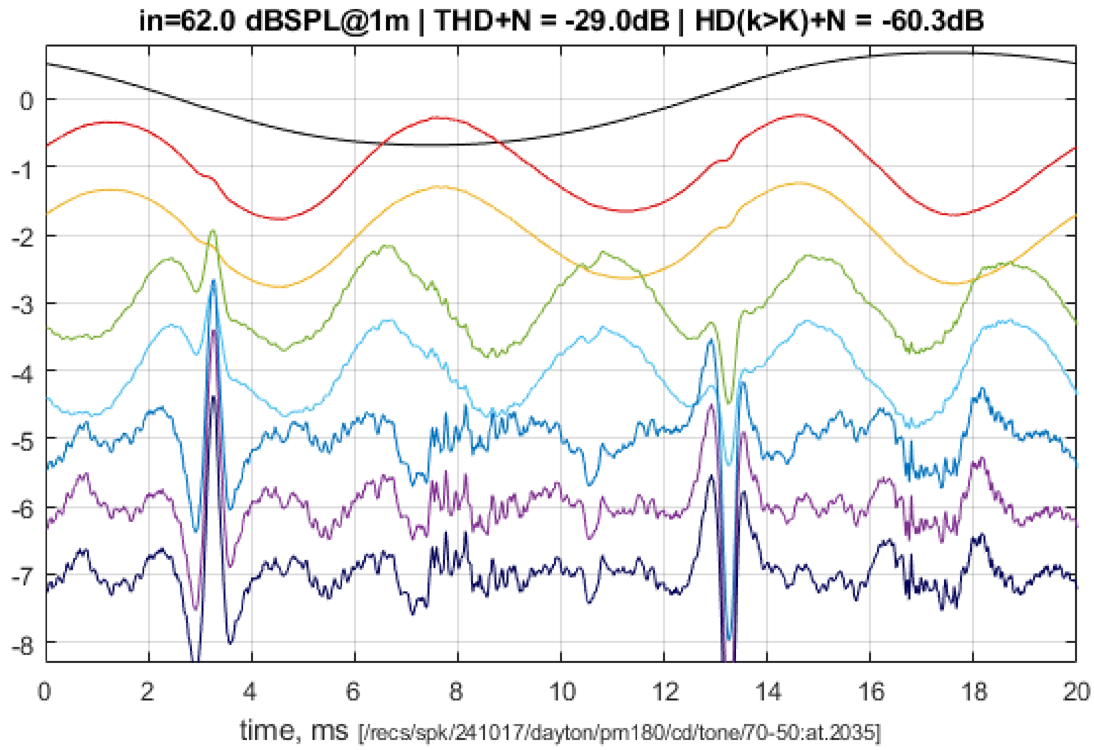


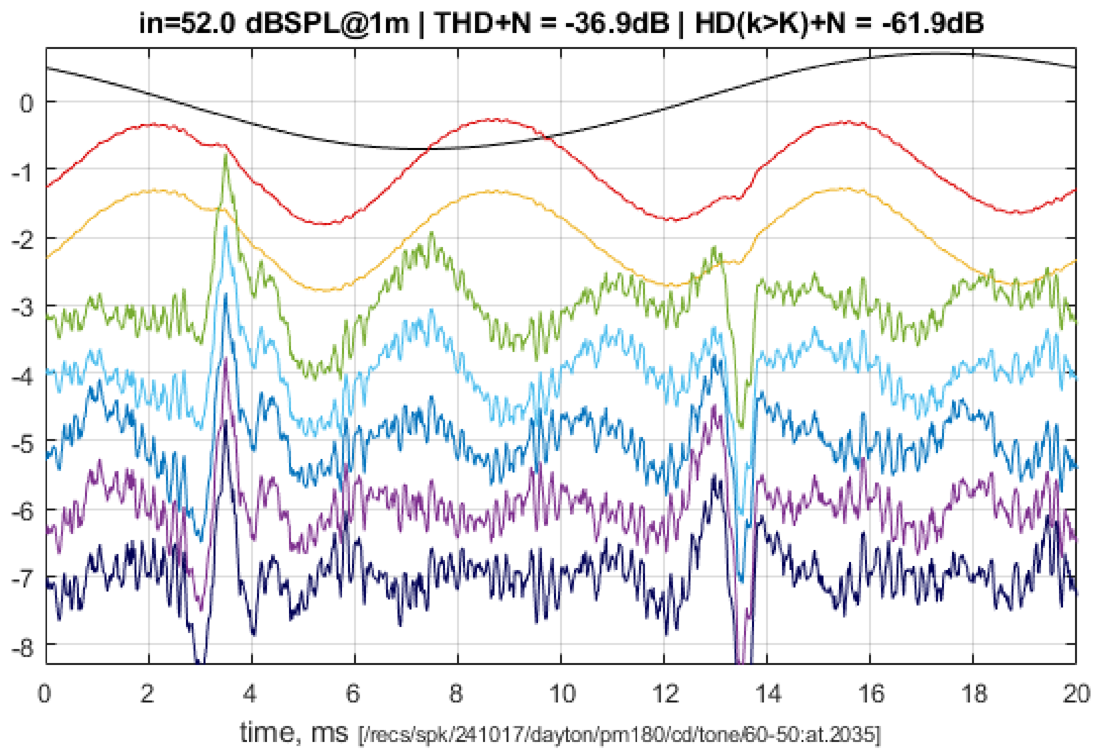
Here we see something different: flat slope at -70 dB (2 dB SPL). Why?

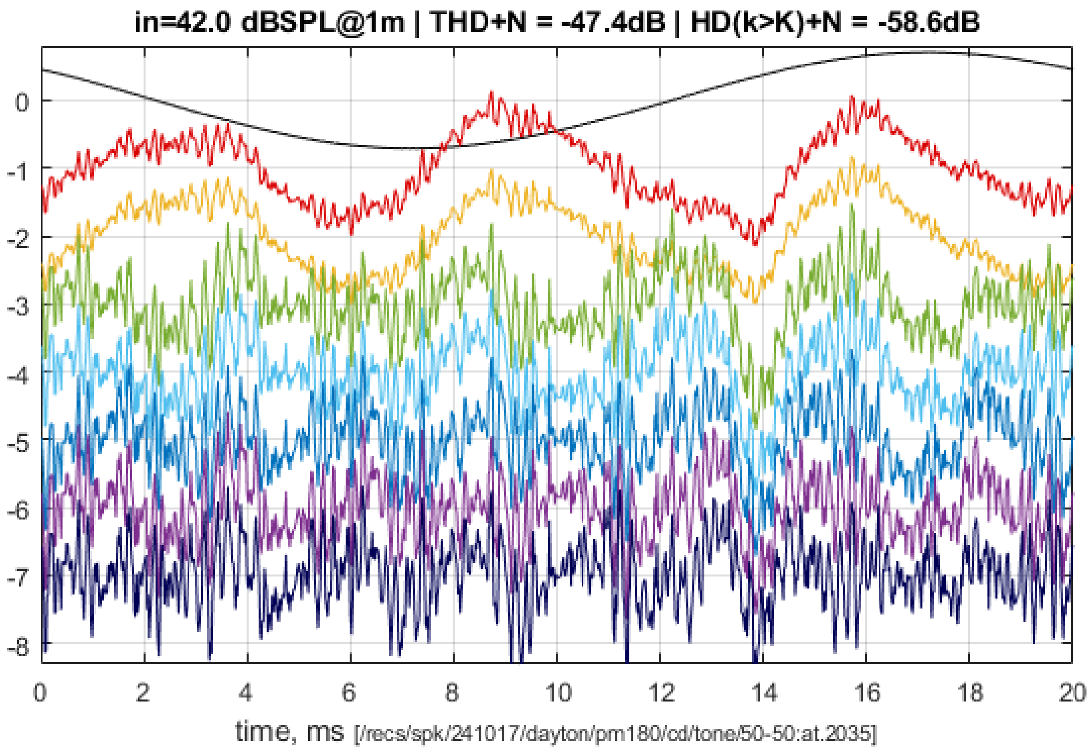


...because of the spikes at zero-crossings, which become noticeable after removal of the H3 and dominant soon thereafter. Are they just a random defect of this specific driver?

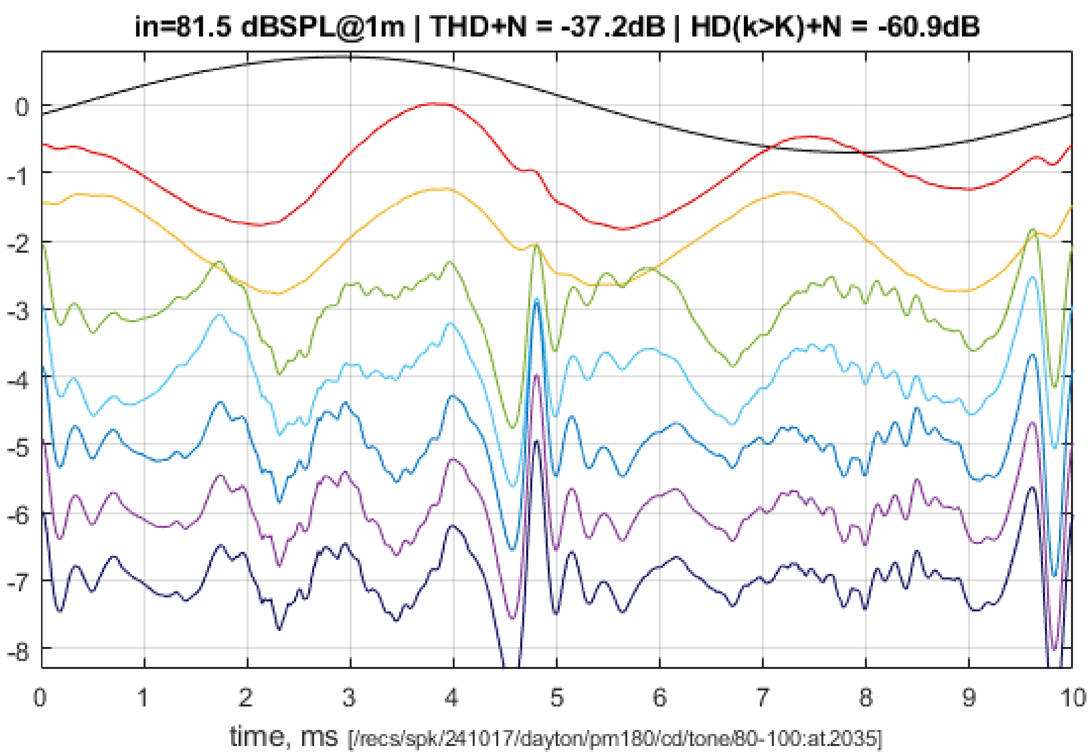
Let's dig deeper, varying into lower SPL:

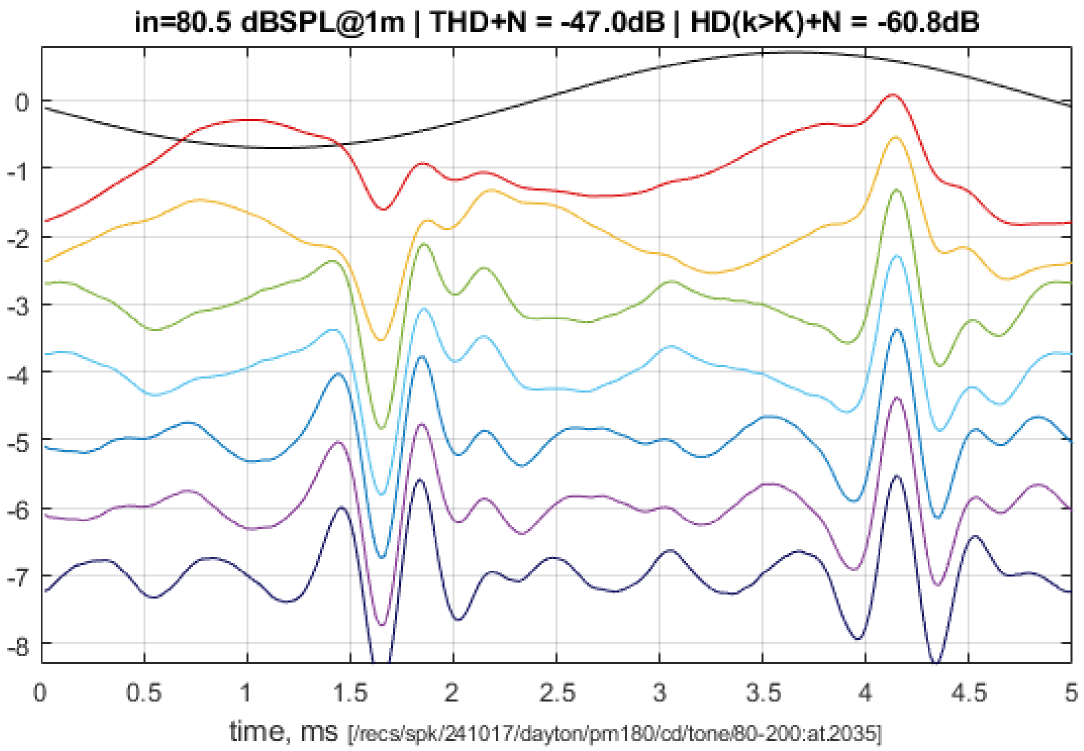


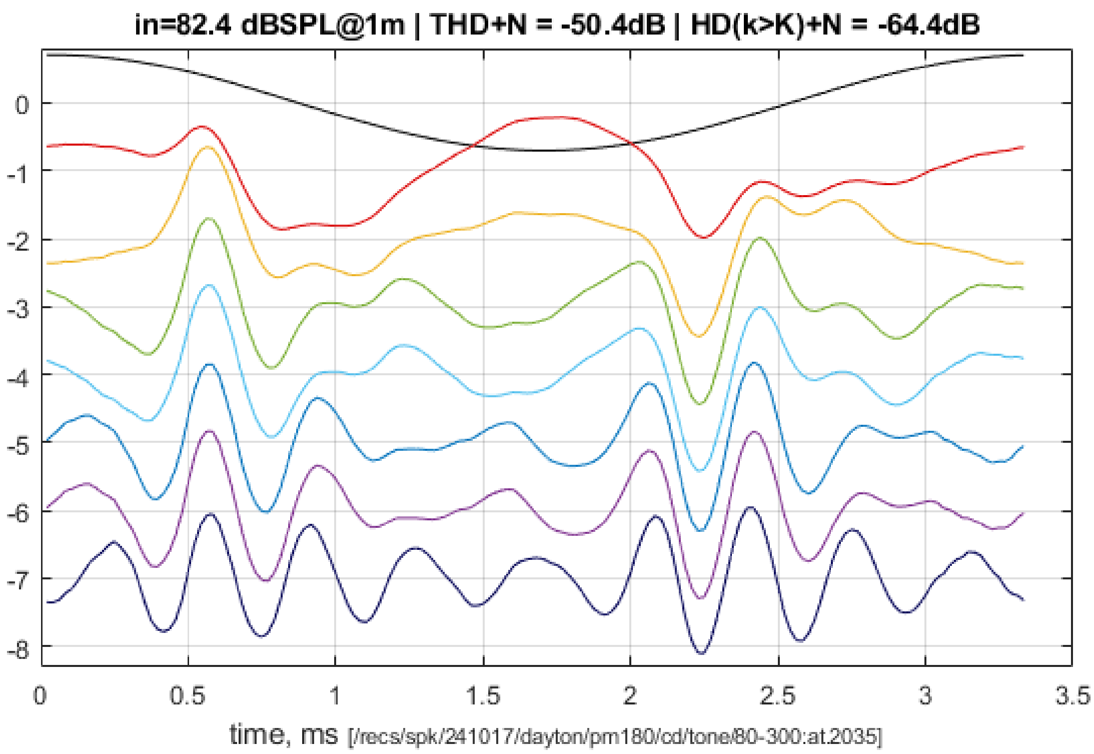




The spikes do persist and become pretty bad at 50 dB SPL but nearly disappear at 40 dB SPL. Are they just below the noise? (No.) Let's dig deeper varying frequency (->100->200->300Hz):







The spikes become wider, proportionally to the frequency. BTW, the same spikes do happen in the Audax driver too, but on much higher excitation.

What makes you the first people observing and measuring them? The same problem: teaching the math as theology and a blind belief that Fourier basis[6] is the eigen-vector basis whenever the excitation is harmonic. How come people go directly to the frequency domain without looking into the time domain first?

The reason of this belief is that the professors today were students yesterday, and that’s what they have been taught themselves, in quite a religious manner. Nobody asked questions, nobody was allowed to ask questions, and nobody listened to the questions when they have been asked.

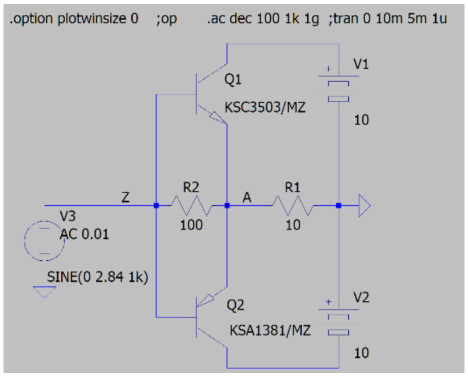
These spikes vary in amplitude from a driver to a driver. For some, it is much worse than for PM180.

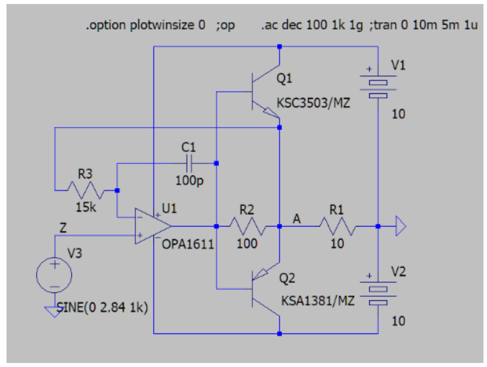
- ✓ These spikes are much worse than the smooth polynomial expansion products. They are about as bad as the Barkhausen noise. They produce a lot of high-order distortions that spill over and become very annoying, judging by the easily measured FSAF residual.
- ✓ As we remember, it is quite typical in the proper mathematics that the same instrument can be used for solving other, seemingly unrelated, problems.
- ✓ Where else could we have seen similar distortions?

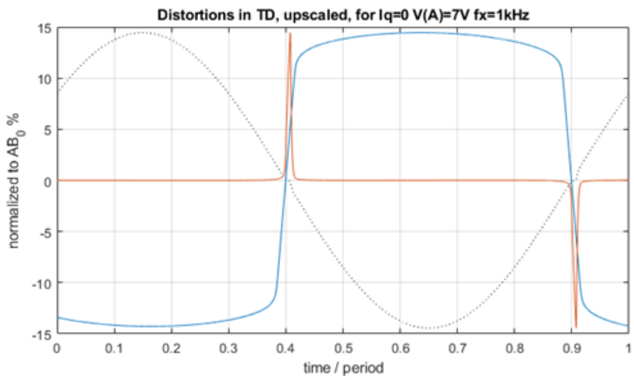
[6] In reality, the Fourier basis is not always reasonably close to the eigen vectors distribution even for harmonic excitation. We had to do the proper Karhunen-Loeve expansion analysis, to start with, instead of staring into chaotic Fourier spectra and wondering why it is so screwed up.

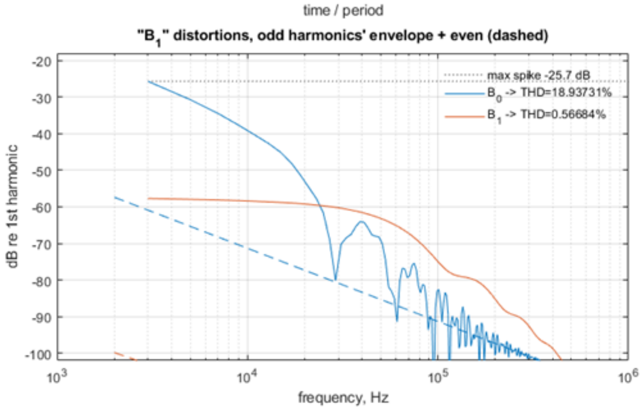
4 Crossover Distortions in Class AB Amplifiers

Actually, very close to the loudspeakers – crossover distortions in Class AB amplifiers:









These exhibit the same properties:

- Sharp delta-function-like spike at or close to zero-crossing points
- The spike exhibits a flat maximum around some point (~2V p2p) and becomes lower at smaller amplitudes.
- The spike does not change much, at absolute scale, when amplitude increases
- The spike widens for higher frequency excitation

Unfortunately, most university professors who teach feedback control do not understand it themselves, so they have nothing to transfer to the next generations. They repeat, a generation after a generation, that feedback decreases all distortions. Nope. The usual 1st order feedback loop (an integrator) takes a derivate of distortions. More precisely, it performs a 1st order high-pass filter at f_T. These spikes appear as the product of a derivative of a sigmoid-like static transfer function of the output dual emitter repeater.

If you use a 2nd order astatizm loop or add nested loops, the derivate order grows accordingly. The further details can be found at Simulation of Crossover Distortions in Class AB 0.00002% THD - File Exchange - MATLAB Central.

- ✓ I needed about 100MHz sampling rate to simulate an amplifier with f_T=1MHz.
- ✓ You do not have to run LTspice to model crossover distortions in loudspeakers. It's an excellent tool but a bit too heavy for our purpose.

5 Sigmoid Feedback Model for Loudspeakers[7]

All you need is to implement the following loop in the language of your choice:

```
ryk=o.yk; % into a register R0
racc=o.acc; % into a register R1
for k=1:lenin
    racc=racc+x(k)-ryk; % dynamic
    ryk=racc-o.cfgTB*tanh(o.cfgTA*racc);
    y(k)=ryk+o.cfgP2*ryk^2+o.cfgP3*ryk^3;
end
o.yk=ryk; % out of a register
o.acc=racc; % out of a register
```

- "*x(t)*" is the input
- "*lenin*" is the LENgth of the INput
- "*acc*" is the integrator state
- "*yk*" is the output state
- "*cfgTB*" and "*cfgTA*" are the parameters of sigmoid (here implemented as hyperbolic tangent)
- "*cfgP2*" and "*cfgP3*" are the parameters of polynomial expansion
- "*y(t)*" is the output

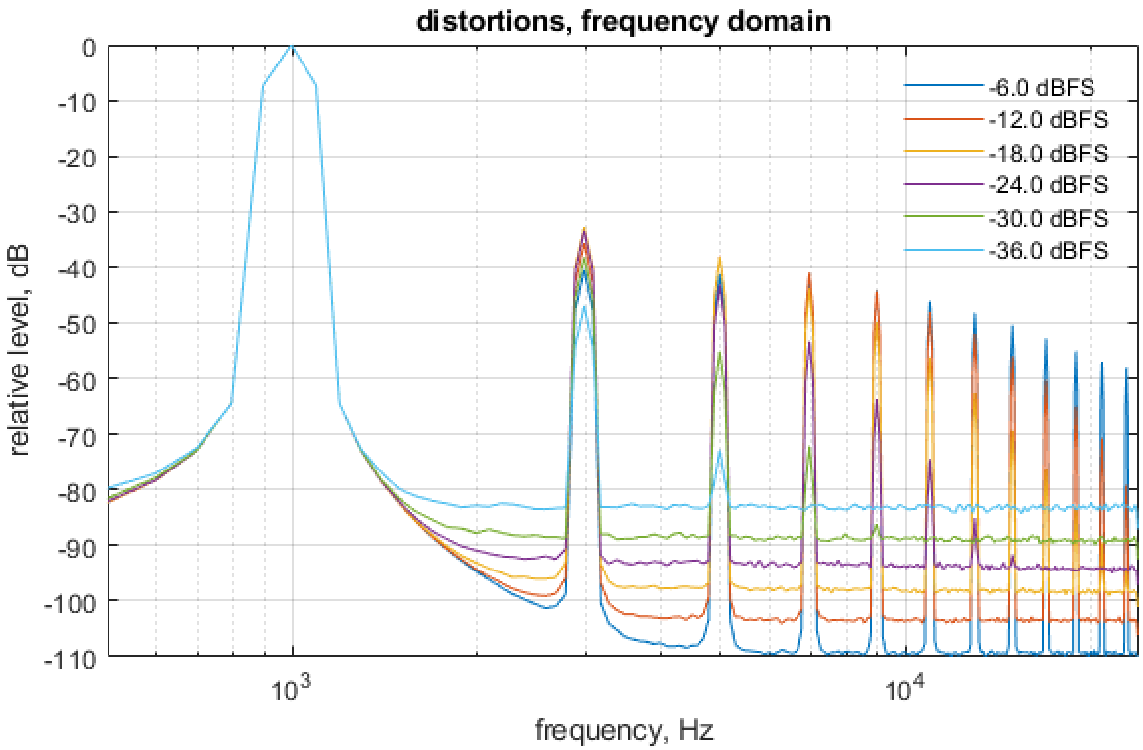
These 3 (three) lines inside the loop are all you need to get a good grasp on the issue. To model the real loudspeaker driver, you may need to expand them a tiny bit and/or fit the sigmoid to the driver-specific hysteresis curve.

- ✓ You will need to learn about Legendre polynomials and other intricacies to combat convergence issues.
- ✓ You will need higher sampling rates (than 48kHz) to simulate the crossover distortions properly unless you want to run into numerous sampling artifacts and precision issues, much like as in Class AB amplifiers.
- ✓ Lennard Ljung, with his MATLAB's System Identification Toolbox, including Gray-Box Modelling, is your best friend.
- ✓ It may not be as trivial as I described, and you will learn a lot of real math – not the nonsense you were tortured with in university.

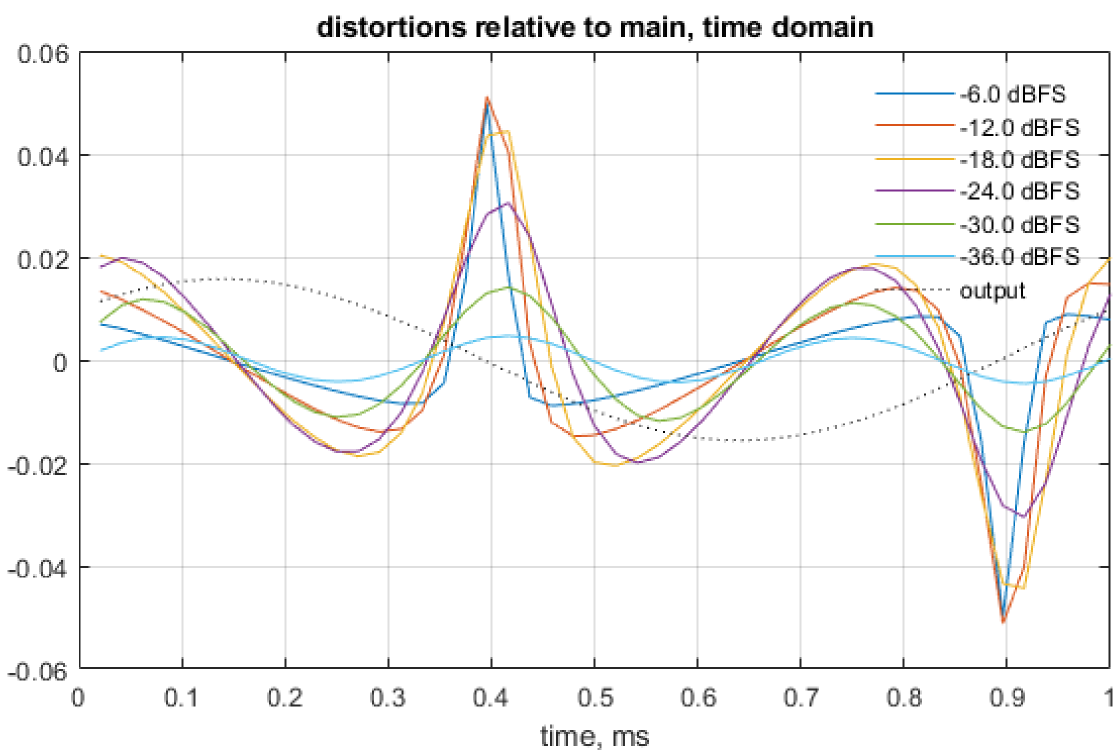
---

[7] Do not ask me where exactly the op-amp and the emitter repeaters are hidden in the loudspeaker drivers. I am only an adaptive control scientist. Ask a theoretical physicist – a good one can provide you with at least three mutually exclusive theories any time, day or night, pretty much immediately, and each theory would fully explain any set of experimental data – or the direct opposite of it.

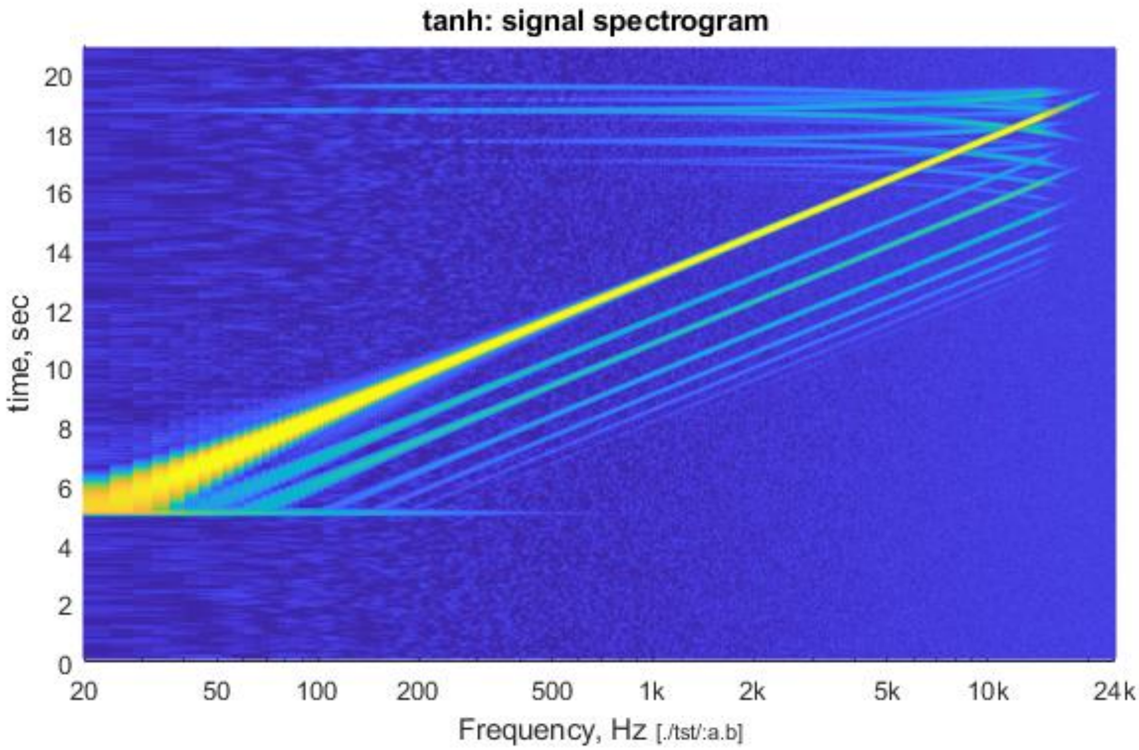
So, how it looks? Let's feed a sinewave of "1kHz" with varying amplitude (test_nlm.m):



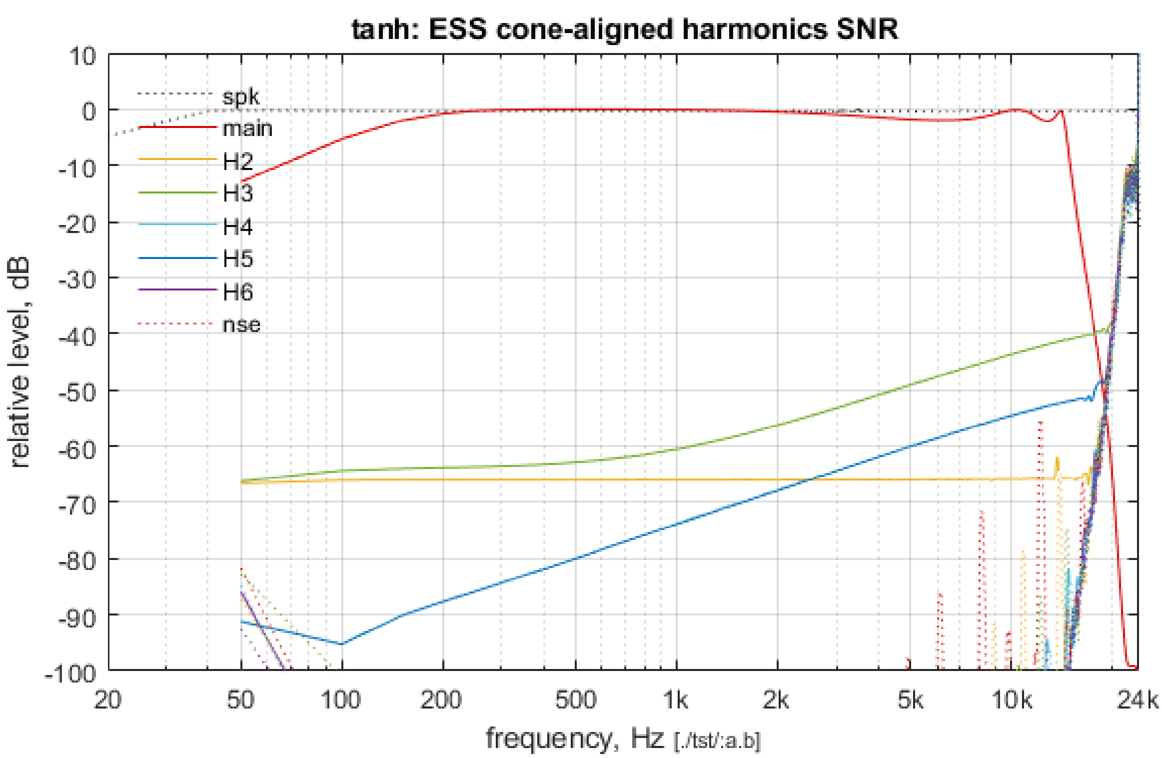
As we can see, H3 does not change much.



Let's feed a sine sweep into this model (+some noise and LTI filtering for fun, test_nlm2.m):



The low sampling frequency, missing 100x factor of LTspice, shows up.



This should look somewhat familiar, shouldn't it?

6 CONCLUSION

Most people don't have a problem connecting two dots. When you need to connect a dozen of dots and the result won't happen unless you connect all of them in a proper constellation, most people have problems… but this is exactly what mathematics is supposed to help with. I am not a genius like Newton or Einstein and I have no other explanation for so late "discovery" of crossover distortions in loudspeakers than the paranoidly stubborn theological approach to the teaching of mathematics in schools and universities.

So far, the perfect sound reproduction chain has been thought of to include perfect source, perfect power amplifier, and perfect voltage driven passive loudspeaker. If all recording studios and listeners have perfect sound reproduction chains, each listener's experience shall be perfect. That's the ultimate target, or so it was thought of.

Most people understand that any attempt of WWII-era air defense 88mm artillery to shoot high altitude flying F-35 is laughable stupid. F-35 flies there faster than shells, it can track shells and outmanoeuvre them automatically, and within a few second precision bomb the enemy's battery into pieces. Yet, the same people have troubles comprehending why audio industry can not produce perfect sound reproduction chains and sell them for pennies.

- This perfect sound reproduction chain is an excellent example of a pure program control paradigm, totally unrealistic and hopelessly archaic.
- The feedback control paradigm works much better. There are few vendors making active current drive monitors thought after by connoisseurs.
- The adaptive control paradigm is even better, like in F-35. You feed a digital source into a DSP which calculates predistortions and feeds the output into a current driving amplifier on the driver itself (compact assembly a must-have) with a sensing microphone to adapt higher-frequency driver's model for pre-distortions and directly performing digital adaptive loopback control of lower frequencies. Drivers will lose uniqueness and characters, become generics to be designed as a Lego block and software will soon take over.

Proper modelling of the drivers designed for passive loudspeakers and voltage drive is an art for the art's sake. When there will be drivers designed for current drive and ease of modelling, there will be some practical sense in spending the time and efforts.

Unfortunately, most commercial organizations believe that math shall be free and in no circumstances would they invest time and money into a team of PhDs who would work hard for years to get a decent practical result. These CEOs and CTOs do not understand that it is much easier for a mathematician to earn the living by writing databases (in a fraction of time, relative to a "normal" developer) and use the rest of time on their individual theoretical pursuits for pure pleasure.

There has never been a lack of theories why loudspeakers' distortions exist, including that they don't.

7 ACKNOWLEDGEMENTS

This work is an evolutionary step and by no means pretends to be error-free or contain any "Truth". This work would not happen without the people who have exhausted all other possibilities (a few of which blissfully failed); Esa Merilainen and Dr. Pascal Brunet, with whom I strongly disagree but value their efforts infinitely; Dr. Wolfgang Klippel who advanced the audio field a long way forward. This work is an addition to his research, not the replacement.