Measuring loudspeakers

Meditation on impulse response gating

M. Batík, September 2026

There are still a lot of misconceptions circulating in the audio community about the effects of impulse response gating on the measured frequency response. Let me offer a simple explanation of what's actually going on.

What is gating?

Gating, as a process, means applying a time window to the measured impulse response: only the initial segment inside the window is retained; the remainder is set to zero[1].

The whole purpose of gating is to suppress reflections that contaminate the measured data. As a result, anything beyond the arrival of the first reflection is discarded and replaced by zeros. The quality of the data obtained in this way depends entirely on how much (or if at all) the true impulse response has been altered by the gating. All of this will be explained in the following paragraphs.

Obtaining the frequency response

Imagine a loudspeaker, our device under test, placed in a perfectly anechoic environment, far from any reflective surfaces, so that we can record 500 milliseconds of a reflection-free impulse response. However unrealistic this assumption may be, it will serve us well in understanding the essential concepts.

By this time (500 ms), the recorded signal is essentially just silence — the impulse response of the loudspeaker had died out long before. Let's say the impulse response takes 20 ms to decay below the measurable levels. For simplicity, we'll call this the duration of the impulse response — in the sense that extending the time record longer makes no difference.

Now, to convert a time signal into the frequency domain — i.e., to obtain the frequency response from an impulse response — we normally use the Fast Fourier Transform (FFT[2]). We can think of it as a black box that takes time samples (the sampled values of a signal) and outputs frequency-response samples (magnitudes and phases). The frequency step of the output samples is determined by the total duration of the input: a sampled signal of duration T seconds transforms to a frequency step of 1/T Hz. The output samples always start at 0 Hz (the DC component) and continue in steps of that size, which is also called the bin width.

---

1 For now, let's set aside any tapering of the window — i.e., how gradual or abrupt the transition to zero is — we'll come back to it later, it's only a secondary effect.

2 FFT is an efficient implementation of the Discrete Fourier Transform (DFT). For all our purposes here, the two terms can be used interchangeably.

For instance, 5 milliseconds of data will give us frequency points every 200 Hz (1/0.005 = 200), 50 ms of data gives the frequency step 20 Hz, etc.

Thus, with 500 ms of input data, the frequency resolution will always be 2 Hz.

For all subsequent analyses we will always use 500 ms of sampled data as the input to the FFT. If we gate the impulse response with a shorter window, we will pad the remaining data with zeros (i.e. silence), so that the total duration remains 500 ms.

Having the complete data

Let's go back to our idealized scenario - as a first step, we start with a full 500 ms “gate.” Obviously, it has no effect on the impulse response (remember, it has already long died out by that time). From the FFT we obtain the true frequency response with 2 Hz resolution (starting at 0 Hz, as always). There’s nothing more to wish for.

And if we wanted 1 Hz resolution, we would simply feed the FFT with data twice as long — that is, with another 500 ms of zeros added to what we already have. Nothing would be made up: all this information is already present in the data (we haven’t erased anything). There would simply be more frequency points describing the same data, the same stored information. It’s like drawing a circle with an ever-increasing number of points, until they form an essentially continuous line — there’s no limit to the number of points, as the “new” points add no new information about the circle; they only provide a denser sampling of it, as we would say.

Obviously, we can safely shorten the gate down to the point where it still doesn’t “cut” into the tail of our impulse response — it merely “touches” it but doesn’t erase any part of it. Since we always pad the data out to 500 ms with zeros, our frequency step won’t change; it will stay at 2 Hz. And neither will the frequency response change, no matter how short the actual impulse response or the gate is.

We said it takes 20 ms for the impulse response to die out, but it could just as well be 5 ms and nothing would change. We would still have our 2 Hz resolution (given by the 500 ms of data) and still the full, perfectly accurate picture of the true frequency response down to 0 Hz. There could be a reflection arriving at 6 ms — and that would already be a perfectly realistic situation. But again, this holds only as long as we don’t touch the (non-zero) impulse response data.

Now let’s try making the gate shorter and shorter and see what happens…

Accuracy loss

Everything starts to change the moment we begin erasing the tail of the impulse response with our gate, as we make the gate shorter and shorter. The process is gradual — the more of the tail we truncate, the greater the effect — but it always affects the entire frequency response, which will no longer be the true one. How different does it become? Let's see.

We deliberately avoid any mathematics here (that would involve the Convolution theorem[3]), so here's a hopefully intuitive description of what happens when we erase part of the IR's tail:

Every time we truncate an impulse response by applying a gate, we inevitably also affect the calculated frequency response. It works like this: Imagine we take the frequency response of the gate itself[4], which resembles a low-pass filter of a certain bandwidth, and within this bandwidth we "smear" the entire original frequency response — almost as if we took a dry sponge and smudged a chalk writing on a blackboard. The wider the strokes, the more blurred the picture becomes. It works essentially the same way with the gate: the shorter the gate in time, the wider the "sponge strokes" in frequency. The exact width of this smoothing depends on both the length and the shape of the gate, but typical gates have a width of around 1/T Hz, where T is the gate length in seconds.

It is important to note that this smearing always occurs across the entire frequency range. Unlike ordinary smoothing, which works relative to octaves, the gate smoothing works across a fixed interval in hertz, sliding over the original frequency response and "blurring" whatever lies underneath. That is one reason why the effect appears strongest at low frequencies. The other reason is that, unlike low frequencies, higher frequencies only rarely have time-domain tails long enough to be effectively truncated (although it can happen and some high-Q resonances somewhere in the midrange may get virtually lost in the "blurred" data if we truncate the impulse response too much).

Notice that we still haven't even mentioned any FFT — it's because the loss of resolution and accuracy is a completely independent thing (although sharing some very deep principles), determined solely by the truncation of the impulse response, whether we do any subsequent transform or not.

To sum up, with our 500 ms of (zero-padded) input data, we still have 2 Hz resolution data starting at DC, but of a smoothed/smeared version of the frequency response. If we didn't do any zero-padding, we would still get frequency-response points lying on the same smoothed underlying curve — there would only be far fewer of them. The longer the portion of the IR we erase with the gate, the more smoothed this underlying frequency response becomes; however, we can still take an arbitrarily dense sampling of it (as we did with the circle).

3 For those mathematically inclined: a multiplication of signals in the time domain (i.e. gating) is equal to a convolution of their spectra in the frequency domain. The spectrum of the gate is then also called a convolution (or smoothing) kernel, which has a sinc-like shape. Side lobes of the kernel can also introduce ripple in the "smoothed" response.

4 The gate itself is a time signal like any other, with its counterpart in the frequency domain (via FFT).

Here also comes one, maybe not so obvious, question, and in fact quite an important point: how is it possible that this process of losing the accuracy is gradual, as we "hit" the IR's tail and progressively erase it? How can it be that if we shorten the gate from the length where it still does nothing, by just a few more samples, it is suddenly supposed to smooth the entire frequency response with a window of, say, 200 Hz (for a ~5 ms gate), even when it erases only a tiny part of the response? It is because any time-limited signal contains only a limited amount of frequency detail to begin with. This is the fundamental relationship between time and frequency. If the true impulse response is only T seconds long in itself, it inevitably contains only frequency detail limited by a 1/T Hz wide “smoothing window.” It is exactly the same as if the response had already been gated by a (rectangular) window of the same length. Therefore, truncating that 5 ms response with a gate only slightly shorter than that, say 4.9 ms, is equivalent to applying a 204 Hz smoothing window to a response that has already been smoothed with a 200 Hz window. It adds a little further smoothing, but not much.

As it should be clear by now, it’s not the gate length *per se* that’s limiting us to get accurate data, neither it's the FFT. It’s always the truncation of the original impulse response, leading to smearing of the frequency response. If we don’t truncate the IR, no matter how short gate we use, we always have accurate data without any low-frequency or resolution limits. An extreme example would be a perfect impulse, i.e., a single non-zero sample followed by all zeros: there’s no distinction of the frequencies at all (the frequency response is just a flat line — the same value for all the frequency samples), but we can still extract information about any frequency.

In other words, the FFT and its length has no effect on the accuracy, resolution or low-frequency limit of the data, which is either always accurate, or always smeared, for any FFT length, zero-padded or not. A longer, zero-padded FFT only allows us to densely sample the underlying curve, accurate or smeared — the FFT algorithm works always the same.

Knowing what's available

So, how can we know whether we have a long enough gate when reflections are present and obscure our view of the true tail of the impulse response? It’s the amount of decay the IR has managed to achieve before the arrival of the first reflection that can help us. The more it has died out, the smaller the chance that we are erasing something significant.

If the impulse response dies out so that the frequency response stops changing while we are still within the reflection-free time (as we shift the gate’s ending mark), we can be essentially sure that we have all the data we need. Once again, this gives us data of arbitrary resolution, down to arbitrary frequencies, no matter how long or short the actual gate is.

If the impulse response hasn’t had enough time to die out sufficiently — i.e., if the frequency response still changes significantly throughout the entire reflection-free time (as we slide the gate’s ending mark) — we can be sure that the frequency response is heavily smeared, especially in those frequency regions where the response changes with the gate length.

Summary

After going through this paper the reader should be able to answer the following question: If a 5 ms window has been used to gate an impulse response, what can we say about the obtained frequency response?

We can say with certainty only one thing: that the frequency response is fundamentally limited in its features by a ~200 Hz smoothing window. But without examining whether the impulse response managed to die out within the gate time, we don’t know if it's the true and accurate response, or a smeared one.

It would be wrong to say that below 200 Hz there are no valid data (which is one of the widely spread misconceptions). The gate length alone just doesn’t imply anything like that. And even if the IR was dramatically truncated, the frequency response becomes "blurred" everywhere, not only below 200 Hz.

Remarks and further observations

a You may have noticed that many programs for acoustic measurements try to mark a frequency point below which the gated data is supposed to be unreliable, based solely on the gate length. It often sort of works, but it should hopefully be clear by now that this is far from representative of what is really going on. Once we truncate the impulse response, the entire frequency response gets smeared — not only below some frequency threshold. Conversely, if we manage to capture enough of the IR, no matter how short it is, there may still be perfectly accurate data even below the indicated frequency, and at arbitrary resolution. This really depends on the actual data and how much of it has been erased, not on the gate length alone.

b We haven’t discussed, or even mentioned in any detail, any particular shape of the gating function (or window). That’s because it isn’t particularly crucial here. There are many different types, each with its own name and its own properties in how it affects the (truncated) frequency response. In our sponge analogy this would correspond to something like the exact trajectories of the strokes — details that don’t change the basic fact that they simply smudge the picture.