Beat artifact in sub-Nyquist Signals

Pablo F. Panitta, AES Member

pfp2017@gmail.com

ABSTRACT

This work aims to study the artifact on signals whose frequencies are in the lower vicinity of the Nyquist frequency (called sub-Nyquist signals). We will begin by analysing the interaction between two analog signals and the effects that arise. Then, we will examine what happens in the digital domain regarding a signal's proximity to the Nyquist frequency. The implications for the graphical representation of signals and their impact on D/A conversion for playback will be addressed.

1. Addition of two analog signals

Suppose we have two analog sine wave signals and we sum them. What will we observe on the oscilloscope screen? If the frequency difference between the two is large (say, greater than 10x), the higher-frequency sine wave will appear superimposed on the lower-frequency one.
However, as this difference decreases, we begin to notice more and more that the higher-frequency wave appears "beaten" or modulated by the lower-frequency one. Figures 1, 2, 3, and 4 illustrate the effects just mentioned

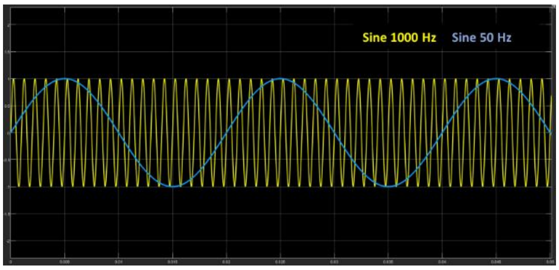


Fig 1. Two sinusoidal signals at 1000Hz and 50Hz

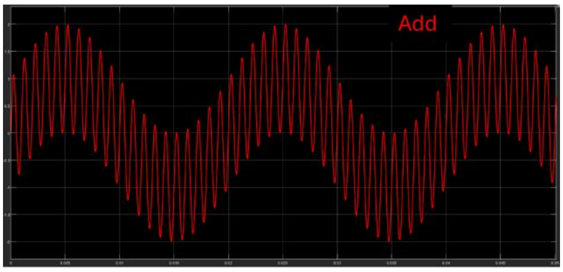


Fig 2. Result of summing two sinusoidal signals at 1000Hz and 50Hz

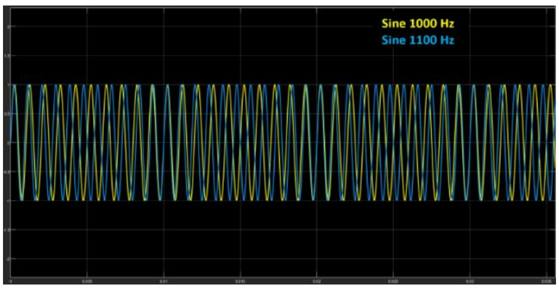


Fig 3. Two sinusoidal signals at 1000Hz and 1100Hz

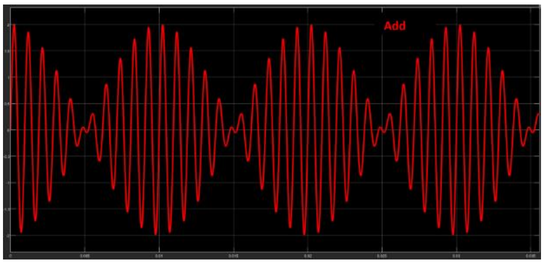


Fig 4: Result of summing two sinusoidal signals at 1000Hz and 1100Hz

It is interesting to note the effect that phase has on the sum of signals in both combinations. While phase does not play a significant role when the frequency difference is large, it does when the frequencies are close, producing cancellations or reinforcements in the magnitude of the sum.

Note that the beat frequency is equal to the difference between the frequencies of both signals (in this case, 100Hz).[1]

In both cases, a spectral analysis will show only the two pure sinusoidal frequencies and nothing else.

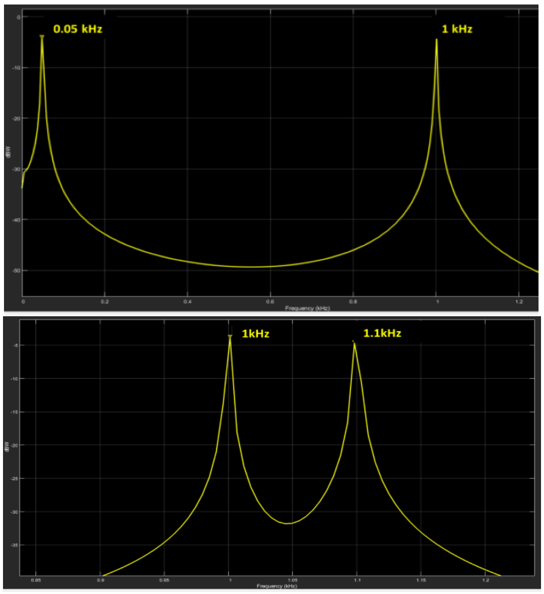


Fig 5. Spectrum of signals a) Scenario1, b) Scenario2

[1] The beat frequency is the difference between frequency signals. However, the frequency of each of the interlaced waves is half, i.e., 50Hz in our case. This relationship will be different in the sub-Nyquist effect we will see in the following sections.

2. Beat effect in digital domain

Now suppose we have an analog signal and proceed to sample it for digitization. For demonstration purposes, let's take the following values:

- *Signal frequency: 500 Hz*
- *Sampling frequency: 44.1kHz*

In Fig 6, we can see that the sampled signal is perfectly overlaid on the analog signal, showing no amplitude deviations.

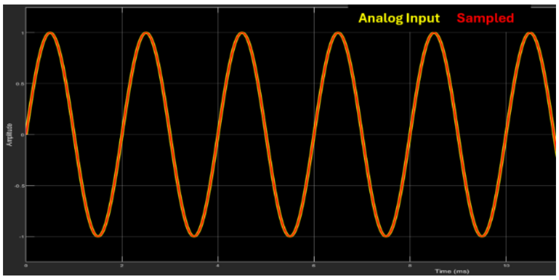


Fig 6. Sampling of an analog signal with a frequency far from Nyquist

Let's repeat the process, but now with a frequency close to Nyquist, for example, a signal frequency of 20 kHz.

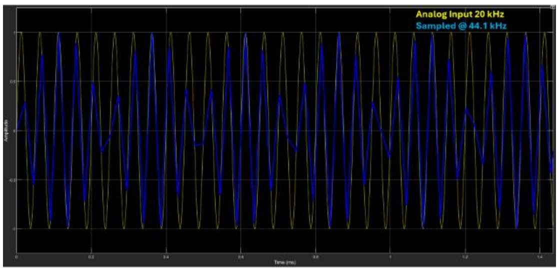


Fig 7. Sampling of an analog signal with a frequency close to Nyquist

In Fig 7, we can clearly observe a beat effect on the 20kHz signal. This looks very similar to the effect described in the previous section with two analog frequencies. But if we are working with only one frequency here… how can this be happening?

Let's try to understand what's going on in this case. I have a 20kHz analog signal at the input, which, when sampled at a rate of 44.1kHz and then converted back to analog[2], appears as a 20kHz signal but with a beat at 4.1kHz. Everything suggests that, based on what we saw earlier, there should be another signal "modulating" the input to produce the modulated signal. Let's confirm this with a spectrum analyser.

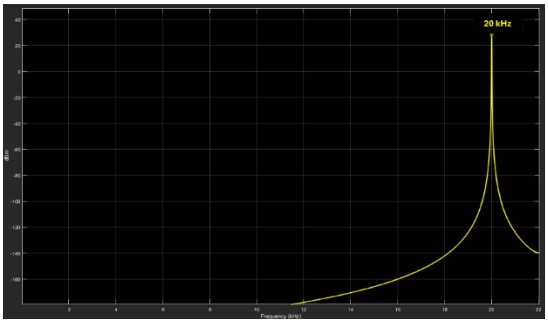


Fig 8. Spectrum of a 20kHz signal sampled at 44kHz

Undoubtedly, Fig 8 shows evidence of only one signal! This doesn't seem to make much sense. There is only one signal present in the spectrum, yet in the time domain, it shows clear signs of amplitude modulation (beat) that could only occur if another signal were interacting.

So, what's going on?

3. Samples Distribution.

Suppose we sample a signal with frequency F at a sampling rate $F_s = 3F$. What we would get is the presence of 3 samples per signal cycle, and furthermore, these occur at the same phase of the signal. In other words, a periodic pattern is created, as we can observe in Fig 9.

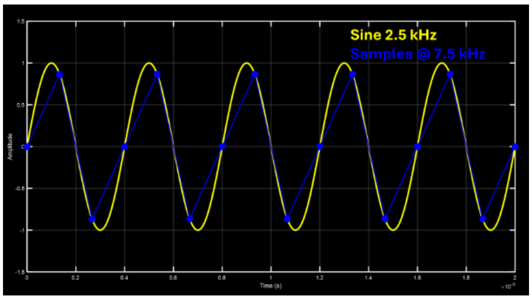


Fig 9. An analog signal sampled at 3 times its frequency. Sample distribution and linear interpolation for reconstruction

And if we did the same, but now with F_s=5F, there would be 5 samples per signal cycle, with the same characteristics as the previous signal, as can be seen in the following figure.

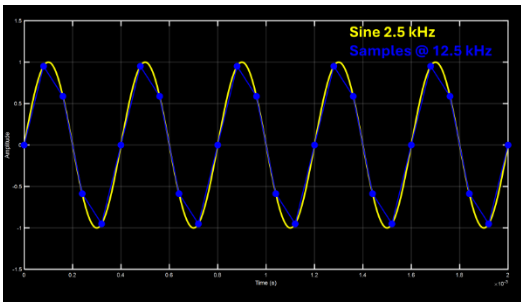


Fig 10. An analog signal sampled at 5 times its frequency. Sample distribution and linear interpolation for reconstruction

[2] Matlab's interpolation method is linear, which is why we see straight lines connecting the samples.

A particularity of the examples just mentioned, is that the number of samples per signal cycle is invariant. This means that, for example, in F_s=3F, there will be 3, and only 3, samples in any signal cycle. I won't find cycles with 2, 4, or any other number of samples. In other words, there is a relationship between F and Fs that we can represent with the following equation:

$$F = \frac{1}{n} * F_s \qquad (1)$$

Again, this means that **every 1 cycle**, I will have "n" (2, 3, 4, etc.) and **only "n"** samples. No more, no less. We can generalize the formula and rewrite it as:

$$F = \frac{m}{n} * F_s \qquad (2)$$

Here, _m indicates the number of cycles_ in which I will find a number of _n samples_. For example, if the ratio were $\frac{2}{5}$, it would mean that in 2 signal cycles, I will find exactly 5 samples.[3]

But there is another characteristic that has a particularly important consequence: **n samples occur at the same phases of the signal every m cycles**. Following the previous example, the samples will fall at the same phase of the signal every 2 cycles.
But where are we going with this? What does it have to do with what I see on the screen? A little more patience, we're getting closer.

4. Cyclic Patterns

Now let's add one last parameter to the formula, called ε, so that it results as shown in the following equation:

$$F = \frac{m}{n} * \ F_s - \epsilon \qquad (3)$$

What does this modification achieve? It cancels the fixed positioning characteristic of the samples within the signal cycle. Although there will still be n samples every m cycles, their location within the signal will vary across cycles.

And will the samples ever fall at the same phase of the wave again? In other words, will there be a frequency at which the cycle repeats? The answer is yes, and that is precisely the value of ε.

Let's look at the graphs of some examples:

[3] $\frac{m}{n} \leq 0.5$ to satisfy Nyquist criterion

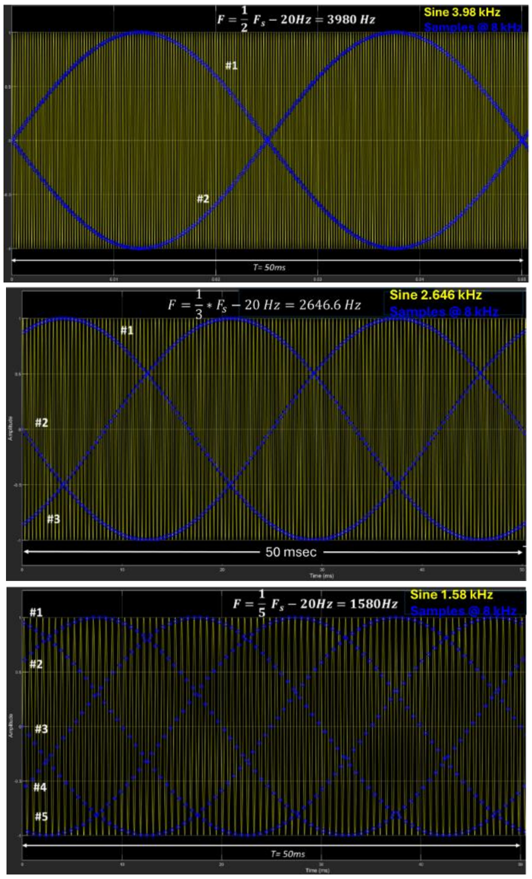


Fig 11. An analog signal (yellow) and its samples a) F = Fs/2 + 20, b) F = Fs/3 + 20, c) F = Fs/5 + 20. For all cases, Fs=8 kHz

In Fig 11, we can observe the distribution of samples *"forming" sinusoidal waves whose number corresponds to the value of the factor "n" and with a frequency of value ε.*
In other words, in the formula $F = \frac{F_s}{2} - 20$, 2 waves will be formed; in $= F = \frac{F_s}{3} - 20$, 3 waves; and in $F = \frac{F_s}{5} - 20$, 5 interlaced waves. All of them have a frequency of 20 Hz (as is the value of ε).

Note that these waves are the product of a graphical distribution of the signal's samples, and all have a frequency of ε. This means that ε is the frequency at which the samples occur again at the same phase of the signal. It is easy to see that if ε=0, the variation frequency is 0, and therefore, the samples always occur at the same phase of the signal.

5. The Reconstruction Process

It is striking that if I wanted to reconstruct the signal by connecting the samples, I would not be obtaining the same original signal. As observed in Fig 12, if I apply the sampling process to a pure sinusoidal signal (A/D) and then, from those samples, reconstruct the signal (D/A) by connecting the samples, I will arrive at a sinusoidal signal but with a beat, which obviously is not the same as the original signal. But this would be a clear contradiction to the Nyquist theorem, which states that if I meet the condition of sampling at a rate higher than twice the maximum frequency present, I could reconstruct the original signal exactly.

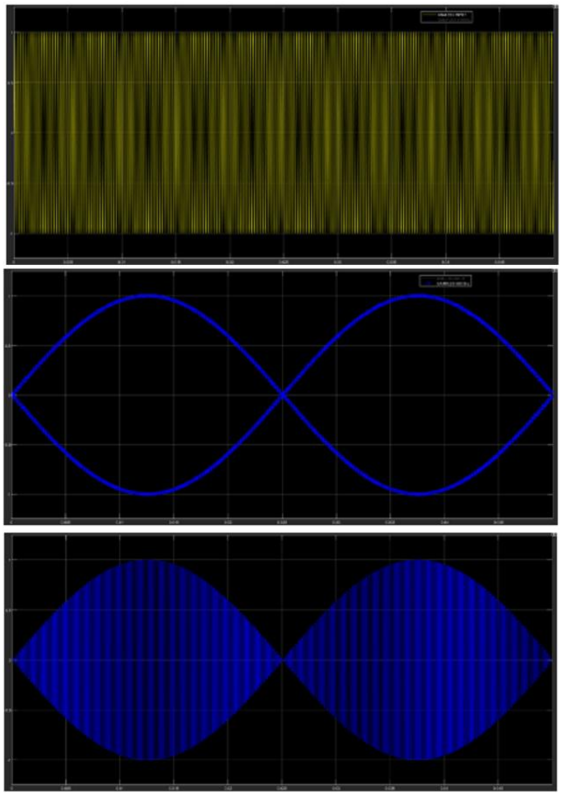


Fig 12. a) Analog signal b) Samples of the analog signal c) Reconstruction process. It can be observed that a) and c) are different

The explanation lies in how the reconstruction of the analog signal from its samples is performed. The interpolation process consists of reconstructing the signal value between samples, and this is *not achieved by simply "connecting" the samples graphically*. There are various ways to "connect" (interpolate) the samples graphically. For example, one could be to do it linearly (straight line), through an N-order polynomial (curve), or some other type.[4][5] But these methods do not reconstruct the original signal exactly.

The theoretical method to fully recover an analog signal (i.e., without loss) is through a convolution process between the samples and the SINC function. It is entirely beyond the scope of this work to explain the reconstruction (interpolation) process in detail, but we can think of it as follows: if I wanted to recover a frequency-limited signal (that meets Nyquist requirements), I could filter it with a low-pass filter to keep the frequency band below $\frac{F_S}{2}$ and reject all spectral components above it. In Fig 13, we can see the process.

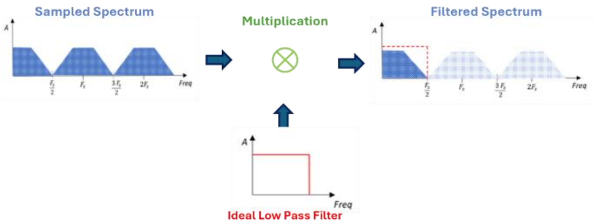


Fig 13. Signal reconstruction process in the frequency domain

The mechanism of this filtering process but viewed from the time domain (always remember the frequency-time duality), would mean a convolution between each of the samples and the SINC function, which is the time-domain representation of an ideal low-pass response.

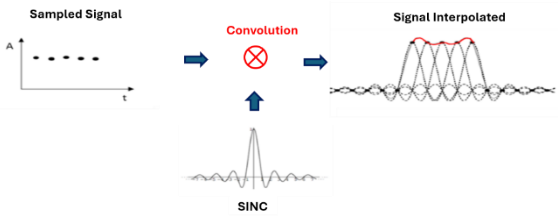


Fig 14. Signal reconstruction process in the time domain

In theory, this process provides perfect reconstruction. But the problem lies in the impossibility of its implementation, as the SINC function extends from $-\infty$ to $+\infty$. Since this is not feasible for practical purposes, we must truncate it in time. This temporal truncation of the function will have an impact in the frequency domain, meaning that the low-pass filter response will no longer be ideal, and this will introduce "distortions" in the reconstruction of the analog signal. And it is precisely these errors that make the reconstructed signal like the original.

In Appendix III, you will find a practical demonstration of the impact of the SINC length on the reconstruction process.

6. Is this artifact real?

Having understood the reason for the "weird" shapes we see in the reconstruction of certain frequencies, it is worth asking whether this is a real effect or simply a visual effect. The answer is that this effect is as real as digital audio itself, so it is not only a visual perception but also an auditory one.

And verifying this is not difficult at all. You can simply create a session at some audible Nyquist frequency, generate a sinusoidal signal (or any periodic signal), and listen. For example, we can test

[4] There are several interpolation methods such as Lagrange (N-order), cubic splines (3rd order), Bezier splines, etc.

[5] Some software allows choosing the graphical interpolation mode (e.g., Ozone RX).

with an Fs = 8000Hz and use a generator to create sinusoids at F = 3998 Hz and F=3999Hz. What do you expect to hear?

7. Impact on digital audio

But if this is the case…how is it not a complete disaster for everything we hear through digital audio? Well, there are two important aspects to consider in the process explained:
1) While not exclusive, this is *noticeable in periodic signals*. The beat effect is evident when there are periodic repetitions of samples in the signal cycles (as explained in section 4).

2) The effect is extremely noticeable in *frequencies very close to Nyquist* and loses impact as we move away from it. If my Nyquist frequency is 22050 Hz, the effect begins to fade from 20 kHz downward. In other words, these are high frequencies for the average human hearing.[6]

In Fig 15, we can see the impact in relation to the analog frequency being sampled. As the frequency F moves away from $\frac{F_S}{2}$, the effect diminishes (the beat loses intensity).

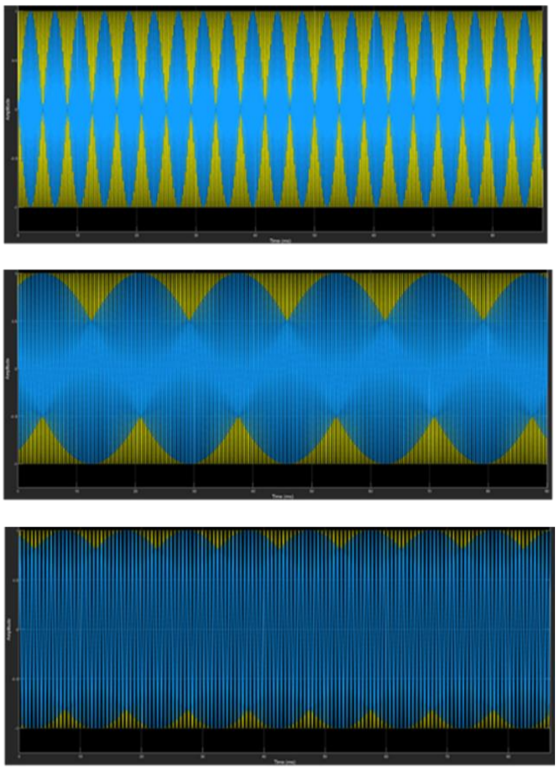


Fig 15. Beat at different frequencies. In all cases $F_s = 8\ kHz$. 1mv peak

For these reasons, although it is curious to see it on DAW screens, this effect is not a problem for the purpose of digital audio.

8. Conclusions

This paper has addressed the sub-Nyquist artefact present on digital audio, explained the reasons and evaluated its impact on digital audio.

9. Bibliography

[1] I.Amidror, "Sub-Nyquist artefacts and sampling moiré effects", *Ecole Polytechnique Fédérale de Lausanne (EPFL)* (2015).

[2] Kerry Schutz, "Analyzing the Sampling-Induced Beat Frequency Effect", *Mathwork* (2021).

[3] Julius. O. Smith III, "Bandlimited Interpolation", *CCRMA – Standford Univ.* (2016).

[4] Jamie.A.S.Angus, "Bandlimited Interpolation" Modern Sampling: A Tutorial", *AES Vol 67*, No 5 (2019).

10. A1. Measurements

The following displays the measurements made at the output of a Universal Audio Apollo Twin sound card. The measurement instrument used was a Tektronix TBS 1052B, at different sampling and signal frequencies.

The measurement process consisted of generating sinusoids at different frequencies and measuring the analog audio output of the card. This is what the speakers will actually reproduce.

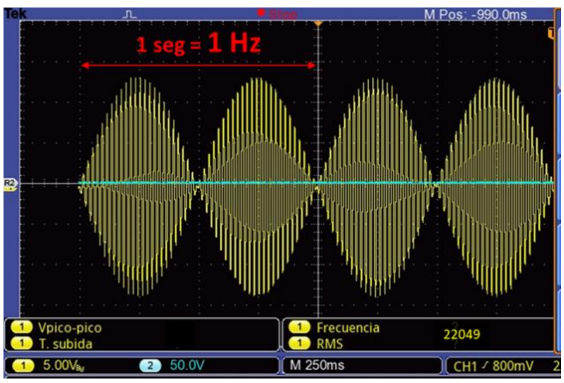


Fig 16 Fs=44100 Hz , F=22049 Hz. The beating can be seen at a frequency of 0.5 Hz: Note that the frequency of each interlaced sine wave is 1 Hz.

[6] Additionally, the beat can only be heard if it is just a few Hertz.

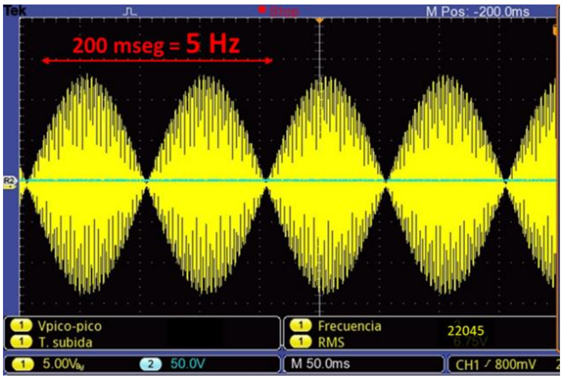


Fig 17 F_s=44100 Hz , F=22045 Hz. The beating can be seen at a frequency of 2.5 Hz: Note that the frequency of each interlaced sine wave is 5 Hz.

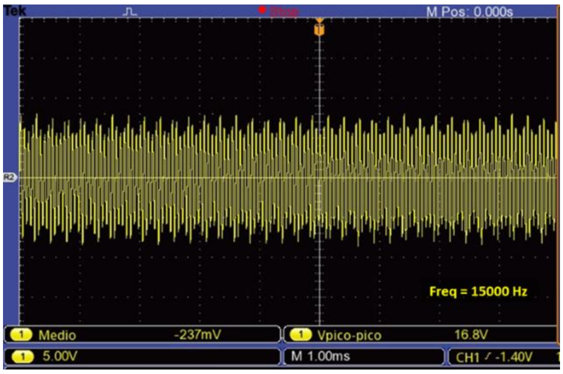
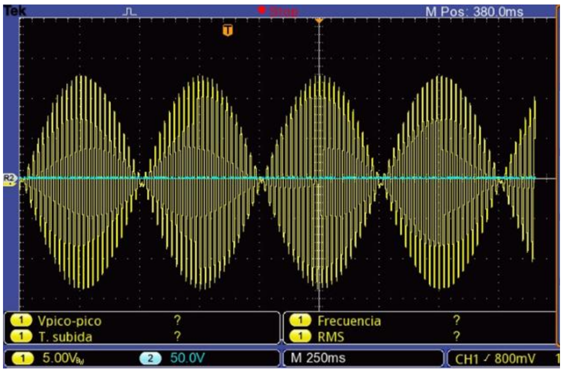


Fig 18 $F_s = 44100\ Hz$ *,* $F = 15000Hz$ *No significant beating can be observed. It is confirmed that the beating effect decreases as the Nyquist frequency is moved away.*



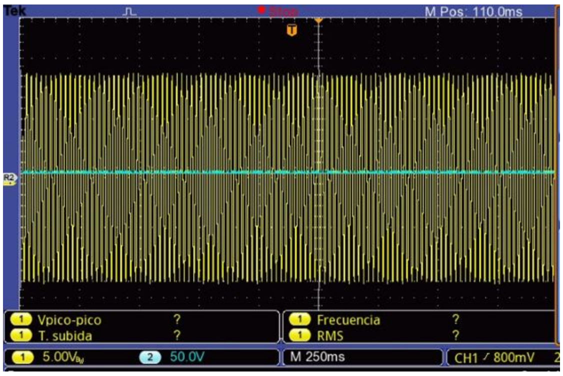


Fig 19.a) F_s=44100 Hz, F=22049 Hz. b) F_s=88200,F=22049 Hz The impact can be seen for the same frequency at different sampling rates. For 88,100 Hz, the impact is negligible, while for 44,100 Hz, it is noticeable.

11. A2. Product of Two Analog Signals.

According to the trigonometric identity:

$$Sin\ (A) + Sin(B) = 2 * Sin\left[\frac{(A+B)}{2}\right] * Cos\left[\frac{(A-B)}{2}\right]$$

we can create the same beat effect <u>*multiplying sinusoidal signals*</u> (instead of summing them), modifying the frequencies to satisfy this identity.[7] In our example of A=1000Hz and B=1100Hz, the new frequencies to achieve the same effect would be:

$$\frac{1000Hz+1100Hz}{2} = 1050Hz$$

$$\frac{1100Hz-1000Hz}{2} = 50Hz$$

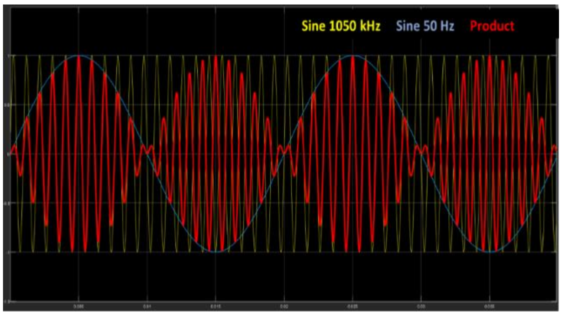


Fig 20. Beating effect through the product of signals

Note the beat frequency remains 100Hz (as in the summing example), which is precisely twice the frequency of the modulator.

12. A3. Interpolation Process of a Sampled Signal.

It has been mentioned that, even when meeting the Nyquist criterion, we will have an imperfect reconstruction of the signal due to the limitation of having to truncate the SINC signal in the convolution process. We have also noted that this phenomenon is noticeable at frequencies close to Nyquist and that as we move away from it, the effect diminishes. Below, three graphs are shown in which we can see how increasing the SINC span, significantly improves the reconstruction.[8]

---

[7] This is the principle used in the amplitude modulation (AM) process. A low-frequency modulating signal (e.g., voice) modulates a much higher-frequency carrier for transmission.

[8] This is the result of a simulation, but it accurately reflects the stated theoretical concept. Obviously, this cannot be applied in the real world, as in DA converters, this duration is not adjustable.

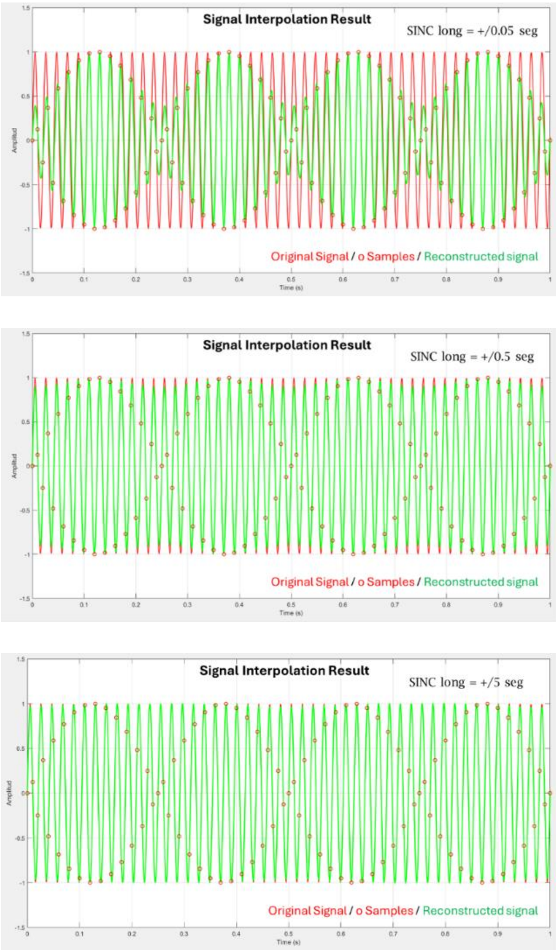


Fig 21. Effect of SINC length on the reconstruction result of the original signal a) SINC_span=+/- 50 msec; b) SINC_span=+/- 500 msec; c) SINC_span=+/- 5 sec

The example presented corresponds to a F=48Hz and Fs=100Hz. While these are not usable values in audio, they are perfectly valid for demonstrating the effect in question.

13. A4. Hardware & Software

Simulations and measurements performed in this paper were done by:

- MATLAB™.
- Universal Audio, Apollo Twin™.
- Tektronix TBS 1052B.