Parabola Volume 62 Issue 2 (2026), Article 9

Nyquist-Shannon Sampling Theorem for a wider audience: An intuitive proof sketch without Fourier

Wai-Yiu Keung[1]

1 Introduction

The Nyquist-Shannon Sampling Theorem [1, 2], often quoted in laymen terms as

"A band-limited signal can be perfectly reconstructed from samples taken at twice its highest frequency",

is commonly regarded as one of the most important mathematical results in the twentieth century. It bridges between a class of continuous domain function and discrete domain functions. In essence, it explains why physical quantities that are analogue in nature can be reliably represented by discrete samples, which may then be converted into digital data through subsequent quantization, encoding and binarisation processes. From the legacy of telephony and television broadcasting to the frontiers of 5G and beyond mobile networks and medical imaging diagnostics, the Sampling Theorem continues to play a central role in our digital world.

In a typical undergraduate curriculum, the Sampling Theorem is covered in the first course of Signals and Systems. The traditional path to understanding runs through Fourier analysis: decomposing signals into their frequency components, comparing continuous and discrete representations, and proving that sampling at twice the highest frequency preserves all information. While such trajectory preserves mathematical rigour, and illustrates the interplay between a signal's continuous and discrete counterparts, it also raises the barrier to entry. Students are expected to be comfortable with calculus and complex analysis methods in order to appreciate the theorem's beauty. As a result, while being foundational to electrical and computer engineering, this material often remains under-exposed to students from other STEM disciplines–including computer science, despite its vital role in virtually all digital technology.

Such absence, from the author's personal viewpoint, is likely due to the underlying pedagogical challenge. The conventional presentation of the theorem requires advanced mathematical machinery, which can obscure its simple intuition behind the result. This note presents an alternative pathway to engage with the Nyquist-Shannon Sampling Theorem–one that allows educators to bypass the prescribed barriers, while preserving the theorem's conceptual beauty. As to be described, the main building

[1]Wai-Yiu Keung is an Assistant Lecturer at the Department of Computer Science and Engineering, The Chinese University of Hong Kong.

blocks are pre-tertiary level mathematics together with a small number of trigonometric identities. In particular, no knowledge of calculus, including the concepts of derivatives and integrals, is required. The only additional concepts required are the time-frequency duality viewpoint on a pre-defined signal, and the idea of filtering in the frequency domain. As a result, the presented material is suitable for teaching at various levels of post-secondary education, and is broadly accessible to readers from diverse academic backgrounds.

2 Background knowledge

A *signal* is a physical quantity that varies over time. We use the functional notion $x(t)$ to denote a signal. Depending on context, $x(t)$ may carry various interpretations: vibration patterns of air particles, price of a stock product over different time, or the rainfall measurements across different seasons. This note will consider the specific use case of audio signals, but readers should keep in mind that the ideas apply much more generally.

2.1 Harmonic decomposition

A fundamental block of signal analysis is the Harmonic decomposition viewpoint. Consider periodic signals that repeat at $f_0\text{Hz}$. The most representative examples of such a class of signal are the sinusoid

$$g(t) = A\cos(2\pi f_0 t + \phi)$$

where $A \geq 0$ is an amplitude coefficient that controls the oscillation strength, and where $\phi \in [0, 2\pi)$ is the phase. Define

Definition 1. *The* amplitude spectrum*, or just the* spectrum*, of a signal is a function that describes how strongly the signal oscillates at different frequencies.*

Under such definition, we see that the spectrum of $g(t)$ may be described schematically as

$$G(f) = \begin{cases} A\,, & f = f_0\,; \\ 0\,, & f \neq f_0\,. \end{cases}$$

Notably, the value of ϕ does not change the oscillation *strength* at frequency f_0. In a more complete treatment, one also considers negative frequencies. Since real-valued sinusoids can be viewed as having symmetric frequency components at $\pm f_0$, we focus on non-negative frequencies $f \geq 0$ in the remainder of this note. Indeed, this corroborates with the intent of a spectrum to capture only the strength of oscillation at different frequencies. Furthermore, observe

Definition 2. *For integers $k = 0, 1, \ldots$, the signals*

$$g_k(t) = A_k \cos(2\pi(kf_0)t + \phi_k)\,,$$

called the k-th harmonics *of f_0, are also repeating at the fundamental period $T_0 = 1/f_0$, although they oscillate faster as k increases.*

This leads us to the following thought: *Can we represent a general periodic signal by adding up a few harmonics?* This is indeed possible shall one look into the notion of Fourier series. That would nonetheless go beyond the scope and objective of this note. Instead, we pose the following simplified conjecture.

Conjecture 3. *A generic periodic signal $x(t)$ repeating at $f_0\mathrm{Hz}$ can be decomposed into a linear combination of its harmonics, i.e.,*

$$x(t) = \sum_{k=0}^{K} A_k \cos(2\pi k f_0 t + \phi_k)\,, \tag{1}$$

where K is a (finite) integer.

In other words, the spectrum of a periodic signal may be visually interpreted as

$$X(f) = \begin{cases} A_k\,, & f = kf_0\,; \\ 0\,, & \text{otherwise}. \end{cases} \tag{2}$$

Geometrically, the spectrum $X(f)$ is a series of equally-spaced stems of height A_k at the frequencies kf_0. See Figure 1.

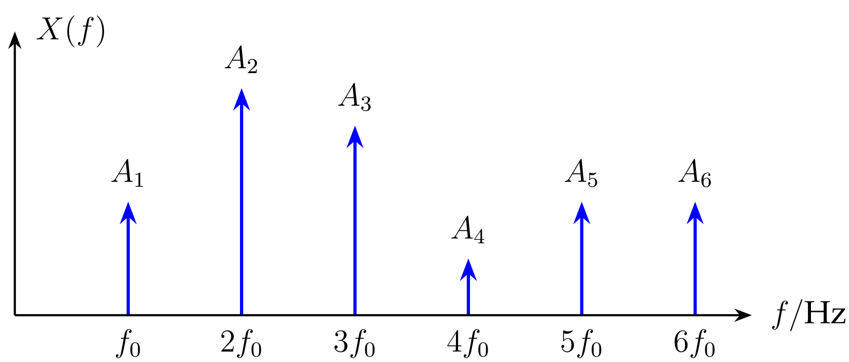


Figure 1: Spectrum illustration of a periodic signal at $f_0\mathrm{Hz}$.

Remark 4. *The value of K can, in principle, be very large. Larger values of K correspond to signals with more rapidly oscillating components at $Kf_0\mathrm{Hz}$. In audio applications, this is unlikely the case because the human audible band sits between $20\mathrm{Hz}$ and $20000\mathrm{Hz}$. Therefore, such K is often finite in practice. When a signal has a finite K, we say that it is* band-limited *to $Kf_0\mathrm{Hz}$.*

Conjecture 3 might require a level of mathematical literacy from the reader. For educators having audiences coming from a non-technical background, the illustration method used in Figure 1 is a more effective way to communicate the prescribed ideas.

2.2 Filtering as a frequency domain operation

A typical treatment to the topic of filtering in Signals and Systems is to introduce it as a subclass of systems with "good properties". Specifically, filters are often introduced as a special case of a linear and time-invariant (LTI) system. Here, we bypass such a barrier by defining filtering directly in the frequency domain.

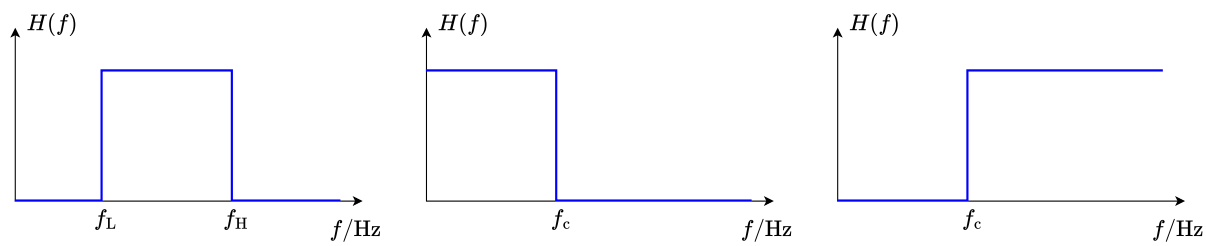
Definition 5. *A filter $H(f)$ is a frequency domain function designed to modify the input signal spectrum $X(f)$. The output spectrum is obtained by point-wise multiplication:*

$$Y(f) = H(f) \cdot X(f)\,. \tag{3}$$

Such definition utilises the convolution property of the Fourier transform, and alleviates the bar of explaining convolution, without explicitly stating so. As a standard family of idea filter examples, consider

$$H(f) = \begin{cases} 1\,, & f_{\mathrm{L}} \le f \le f_{\mathrm{H}}\,; \\ 0\,, & \text{otherwise.} \end{cases} \tag{4}$$

By choosing f_{L} and f_{H} appropriately, one obtains low-pass, band-pass and high-pass filters. The spectral shape of the main types of filters are illustrated in Figure 2.



(a) Band-pass filter (BPF) (b) Low-pass filter (LPF) (c) High-pass filter (HPF)

Figure 2: Main types of filters

Example 6. *Consider a low-pass filter (LPF) with $f_{\mathrm{L}} = 0\mathrm{Hz}$ and $f_{\mathrm{H}} = 4.5 f_0 \mathrm{Hz}$. If the signal $x(t)$ from Equation* (1) *passes through this filter $H(f)$, then only harmonics with $kf_0 \le 4.5 f_0$ remain at the output. That is,*

$$y(t) = \sum_{k=0}^{4} A_k \cos(2\pi k f_0 t + \phi_k)\,,$$

meaning that harmonic components above $4f_0\,\mathrm{Hz}$ are removed from the output $y(t)$.

As a side note, filtering is quietly shaping our daily auditory experiences in both classical and modern technology. In traditional telephony, a band-pass filter is applied on the voice-band $[250, 4000]$Hz so that the transmitted voice is free of background noises like sharp mechanical clatter or low engine rumbles. Similar principles guided the algorithmic designs of noise-cancelling algorithms running in earbuds, in conjunction with adaptive models that differentiate desired voices from background noise in real time. In sound engineering, multi-band equalizers process isolated sub-bands (e.g., Sub-bass, Midrange, Treble) for sound quality adjustment in audio headphones.

3 The Nyquist-Shannon Sampling Theorem

We are now ready to introduce the Sampling Theorem. The need of a theory for sampling is motivated by the need of digitisation. The digitisation process aims to represent an analogue (continuous) message signal by a discrete sequence of ones and zeroes — a so-called *bit stream* in the digital domain. By doing so, we convert a signal from the natural domain to a digital signal.

3.1 Advantages of digitization

The advantages of digitisation deserve a further explanation. Unlike analog signals, where a little static could ruin the quality, a digital bit is either a “1” or a “0”. As long as the system can distinguish between a “high” and a “low” voltage, the message remains perfect. This renders an advantage in using digital representation in fighting against noise, which is an inevitable phenomenon in signal transmission. Furthermore, digital signals can be mathematically “squeezed” (like MP3s or FLAC files) to take up less storage space without losing perceived quality; and it is significantly easier to secure a bit stream using encryption methods than it is to scramble an analogue radio wave. A typical digitization process is illustrated in Figure 3.

Figure 3: The digitization process

3.2 The Sampling Theorem

As seen, the first step of the digitization process involves the sampling block. Our interest lies in the following sub-system:

Figure 4: The sampler of interest

Note that the input $x(t)$ is a continuous function while the output $x[n]$ is a discrete sequence. Given a sampling interval $T_s > 0$, which corresponds to the reciprocal of the sampling frequency $f_s = 1/T_s\text{Hz}$, the discrete samples are given by taking values of $x(t)$ at

$$x[n] := x(nT_s)\,, \qquad n \in \mathbb{Z}\,. \tag{5}$$

Due to the fact that there are infinitely many real numbers between two integers, there are infinitely many values of $x(t)$ between two samples, e.g., $x[n]$ and $x[n-1]$, that are not directly observable from the sequence $x[n]$. This motivates our central question:

Question 1. *How should we choose f_s, or equivalently T_s, such that the pre-sampled signal $x(t)$ can be* ***perfectly recovered*** *from the samples $x[n]$?*

A heuristic, simple, yet non-definitive answer to Question 1 is to sample $x(t)$ with a *sufficiently small* sampling interval T_s or, equivalently, a sufficiently high sampling frequency f_s, so that the samples $x(nT_s)$ and $x((n-1)T_s)$ are closer to each other. As a natural comment, if a signal changes more rapidly, then a higher sampling frequency is needed to maintain the "closeness" between adjacent samples. This is indeed the major intuition formalised by the Nyquist-Shannon Sampling Theorem.

Recall from (1) that the fastest oscillation in $x(t)$ is determined by its highest frequency component, denoted by $W\mathrm{Hz}$ and known as the *bandwidth* of $x(t)$. To draw the connection between the bandwidth and the sampling requirement, we present

Theorem 7. *Consider a signal $x(t)$ with no frequency component higher than $W\mathrm{Hz}$. Then, the samples $x[n] = x(n/f_s)$ for $n \in \mathbb{Z}$ can perfectly reconstruct $x(t)$ provided that*

$$f_s > 2W\,. \tag{6}$$

The critical frequency threshold, $2W\mathrm{Hz}$, is known as the Nyquist rate*.*

Before presenting our proof sketch, let us discuss the implications of Theorem 7. The condition (6) is not merely a mathematical curiosity. It draws a line between proper and improper sampling. If the sampling frequency falls below the Nyquist rate, then the signal structure at the higher frequencies can be corrupted, inducing a phenomenon known as *aliasing*. Even at the boundary $f_s = 2W$, perfect reconstruction can fail, depending on sample alignment.

Example 8. *Let $x(t) = \cos(2\pi f_0 t - \pi/2) = \sin(2\pi f_0 t)$ be a pre-sampled signal. We choose the boundary Nyquist rate $f_s = 2f_0\mathrm{Hz}$ from* (6)*. The samples then read*

$$x[n] = x(nT_s) = \sin\left(\frac{2\pi f_0 n}{2f_0}\right) = \sin(\pi n) = 0 \qquad \textit{for all} \qquad n \in \mathbb{Z}\,,$$

yielding a complete loss of information in $x(t)$.

We see that perfect reconstruction can fail when $f_s \leq 2W$, i.e., the under-sampled sense. Violation of (6) causes high-frequency details to merge with lower-frequency patterns, yielding an irreversible loss of information. In essence, the theorem guarantees that the infinite fineness of the pre-sampled signal is captured exactly by a finite sequence of samples, provided that the sampling frequency is properly chosen.

3.3 An oversimplified proof of Theorem 7

The conventional proof of the Sampling Theorem is widely available in the literature, e.g., in the classical textbook [3]. Typically, it requires the notions of the Fourier decompositions for the pre-sampled signal and the impulse train signal, and it then applies the multiplication rule in Fourier transforms to evaluate the frequency domain behaviour of the sampled signal. We aim to avoid these in our proof, as they are relatively advanced for 1st-year university students and general readers.

To begin with, we define an idealised continuous-time representation of sampling to reason about spectral replication. Let $\tau > 0$ be a small constant that is close to zero and is strictly smaller than T_s. Define

$$x_s(t) = d(t) \cdot x(t)\,, \qquad d(t) = \begin{cases} 1\,, & t \in [nT_s, nT_s + \tau)\,; \\ 0\,, & \text{otherwise.} \end{cases}$$

as the sampled signal model in the continuous time domain. Remark that the actual sampled data are the discrete sequence $x[n]$ in Equation (5). Here, we employ $d(t)$ as a symbolic periodic mask capturing the sampling operation, and we assume that its harmonic decomposition is interpretable and consistent with (1). With a period T_s, we have

$$d(t) = \sum_{j=0}^{J} B_j \cos(2\pi j f_s t + \phi_j)\,,$$

following (1). To keep our study neat, we consider the case in which $\phi_j = 0$ for all j. Interested readers may attempt to follow the same track provided here, and to develop the theorem for the general case of non-zero phases. One will see the extension is trivial. We further assume that B_j's are known and $B_0 \neq 0$. [2] We are now able to express, in an explicit fashion, the sampled signal model as

$$x_s(t) = B_0 x(t) + \sum_{j=1}^{J} \underbrace{B_j \cos(2\pi j f_s t) \cdot x(t)}_{z_j(t)}\,. \tag{7}$$

Consider the frequency domain contents of $x_s(t)$. We see from this expression that the term $B_0 x(t)$ would share the spectral shape $X(f)$ with the pre-sampled signal, subject to a scaling factor B_0. Our next interest is to characterise the frequency components of the signals $z_j(t)$. For $j \geq 1$, consider

$$z_j(t) = B_j \cos(2\pi j f_s t) \cdot x(t) = \sum_{k=0}^{K} A_k B_j \cos\big(2\pi(j f_s)t\big) \cos\big(2\pi(k f_0)t + \phi_k\big)\,.$$

Upon using the product-to-sum identity[3], we arrive

$$z_j(t) = \sum_{k=0}^{K} \frac{A_k B_j}{2} \cos\big(2\pi(j f_s + k f_0)t + \phi_k\big) + \frac{A_k B_j}{2} \cos\big(2\pi(j f_s - k f_0)t + \phi_k\big)\,.$$

We make the following claim.

[2] Readers with a technical background would notice B_0 represents the direct current (DC) value of $d(t)$, which is indeed non-zero as $B_0 = \tau / T_s$.

[3] Namely, $\cos(A)\cos(B) = \frac{1}{2}\cos(A+B) + \frac{1}{2}\cos(A-B)$.

Claim 9. *The spectrum of $z(t) = \sum_{j=1}^{J} z_j(t)$ equals zero for $f \leq W$, provided that $f_s > 2W$.*

Proof. The signal $z_j(t)$ is a linear combination of the harmonic components at the frequencies

$$jf_s, \qquad jf_s \pm f_0, \qquad jf_s \pm 2f_0\,, \quad \ldots\,, \qquad jf_s \pm Kf_0\,.$$

Recall that the bandwidth of $x(t)$ is $W = Kf_0\text{Hz}$. Among these frequencies, the smallest value is $jf_s - Kf_0 = jf_s - W$. Since $j \geq 1$, the smallest value among all $\{jf_s - W\}_{j=1}^{J}$ is $f_s - W$, i.e., when $j = 1$. Hence, the smallest possible frequency component of $z(t)$ locates at $(f_s - W)\text{Hz}$. Since $f_s > 2W$, the lowest frequency range occupied by $z(t)$ shall satisfy $f > 2W - W = W$. This completes the proof. □

Consider the sampled signal model (7). With Claim 1 holding, we see that the information of $x(t)$ suffers no interference from $z(t)$ in the spectral sense. Mathematically, we have

$$x_s(t) = B_0 x(t) + z(t)$$

with $x(t)$ being band-limited to $W\text{Hz}$, and $z(t)$ contains only harmonics with $f > W$. This means that they are well-separated in the spectrum. Sampling creates extra harmonics, but none of them lie below $W\text{Hz}$ when $f_s > 2W\text{Hz}$. Therefore, simply discarding all harmonics above W allows a perfect recovery from $x_s(t)$ to $x(t)$. This completes our proof sketch. As an ending remark, the reconstruction process is conducted by passing the sampled signal through a low-pass filter with a cutoff frequency at $W\text{Hz}$ using a pass-band amplitude of $1/B_0$. The end-to-end sampling and reconstruction process is illustrated in Figure 5.

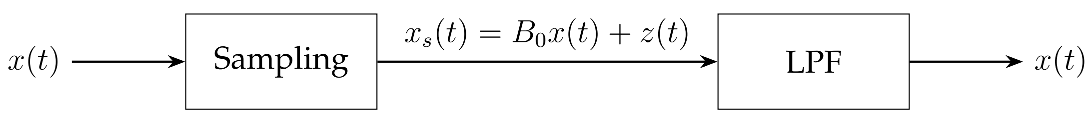


Figure 5: The Sampling and Reconstruction Process

3.4 Limitations and scope

This note is intentionally pedagogical: the goal is to make the Nyquist-Shannon Sampling Theorem accessible to general readers. The scope and limitation of the presented materials should be remarked. We assume band-limited periodic signals that admit a finite harmonic representation, which allows us to reason about sampling through frequency arithmetic and trigonometric identities. The sampling mask $d(t)$ is used as a symbolic abstraction consistent with the sampling process, rather than a standard theoretic object. This allows us to escape from a formal definition of the Dirac-delta function and the impulse train function. Filtering is treated as ideal spectral selection, i.e., Equation (3), without using the notions of convolution. The condition $f_s > 2W$ is presented as a sufficient criterion for perfect reconstruction; issues of boundary cases, noise robustness, quantisation and non-ideal filters are beyond the scope of this note.

4 Closing remarks

The Nyquist-Shannon Sampling Theorem is often introduced in an mathematically inclined fashion, with its core insight hidden behind layers of Fourier analysis and convolution. Such a rigorous way of delivery is necessary for advanced study in signal and information processing. At the same time, it may however discourage beginners without a strong technical background from following the track far enough to realise the theorem's underlying simplicity.

This note has presented an alternative pathway toward introducing the Sampling Theorem using simpler concepts in harmonic decomposition, elementary trigonometry and frequency domain reasoning. The core intuition underlying the theorem is conveyed without requiring Fourier transforms and convolution as prerequisites. The sampling process is framed as a higher-order harmonic generation process, and the induced harmonics do not overlap with the spectral contents of the pre-sampled signal. In this view, the Sampling Theorem naturally becomes a consequence of spectrum separation, and reconstruction is seen as a process to remove the higher frequencies, which is technically done by employing a low-pass filter.

For students at an early stage of their academic journey, and for readers encountering signal processing through modern applications such as data science, artificial intelligence, or multimedia systems, this perspective serves as an entry point of contact with one of the most influential results in applied mathematics and engineering. Once the intuition is established, the reader is better prepared to appreciate the more formal treatments found in classical texts. Ultimately, the value of the Nyquist-Shannon Sampling Theorem lies in its conceptual simplicity, which is equally important with its mathematical rigour. The author hopes that this note helps bring that simplicity to a wider audience.

References

[1] H. Nyquist, Certain topics in telegraph transmission theory, *Transactions of the American Institute of Electrical Engineers* **47** (1928), 617–644.

[2] C.E. Shannon, Communication in the presence of noise, *Proceedings of the IRE* **37** (1949), 10–21.

[3] A.V. Oppenheim and R.W. Schafer, *Discrete-Time Signal Processing*, 3rd ed., Pearson, 2010.