Prolog to

Sampling—50 Years After Shannon

An introduction to the paper by Unser

It has been 50 years since Claude Shannon laid the foundation for information theory with the publication of "Communication in the Presence of Noise." In that paper, Shannon articulated the theorem that information could be quantified and coded by a mathematical process of sampling. The intervening years have proved Shannon's paper to be a major theoretical work, one that has had the greatest impact on modern electrical engineering. In this tutorial, the author revisits Shannon's original sampling paradigm to see how well it stands up to modern requirements and how it may be extended to accommodate today's larger selection of sampling functions. This includes an examination of standard sampling methodology as it relates to current technology, the application of Shannon's theorem to wavelet theory, methods for controlling approximation error, and variations and extensions of sampling theory.

An optimum information system is one in which the flow of information equals the capacity of the transmitting channel. The objective is to compact the information while minimizing the distortion, and thus utilize channel capacity to maximum efficiency. This capability assumes special importance in today's information-driven society. Shannon's theory provides that capability. The key to his theory lies in the orthogonality of the underlying basis functions. An analog signal (or lightwave) is imaged on a two-dimensional Cartesian plane, then sampled at key ordinates to develop an identifiable profile, or signature. These ordinates are then converted into a sequence of numbers that can be processed digitally, transmitted, and reconstructed with absolute fidelity at the receiving end. The approach works perfectly, provided that the input wave (or function) is bandlimited; i.e., it has no frequencies above the Nyquist limit. When the input is not bandlimited, it needs to be ideally low-pass filtered prior to sampling in order to avoid aliasing. In effect, this process is equivalent to performing the orthogonal projection of the input signal onto the subspace of bandlimited functions. In other words, it provides the minimum error representation of the input signal within the given bandwidth.

Paradoxically, although Shannon's theorem is crucial to modern signal processing and communications, it is rarely applied literally in practical applications. The primary reason is that the theorem is based on a nonrealistic model in which word signals and images are exactly bandlimited. It further assumes the existence of an ideal low-pass filter to suppress aliasing. Nevertheless, Shannon's idealized theory still supplies the underpinning for modern and extended versions of sampling. The explanation lies in the recent and intense activity occurring in the field of wavelet theory.

Ten to fifteen years ago, it appeared that sampling had reached maturity; research in this area had become very mathematically oriented and had less immediate relevance to signal processing and communications. However, subsequent investigations into wavelet theory demonstrated that the mathematics of wavelets could also apply to sampling. This prompted researchers to reexamine Shannon's theory with a view to adapting its principles to more generalized, practical formulations that would accommodate the newer technologies.

The paper next considers extending Shannon's theorem to other classes of functions. Doing so requires a sampling scheme that is practical yet retains the qualities of classic sampling theory. This is achieved by replacing Shannon's sinc-function with a more general template, the generating function. In this approach, functions that are being defined continuously are characterized by sequences of coefficients. These coefficients are not necessarily samples. Rather, they constitute a discrete signal representation that will be used to calculate signal processing or perform coding. To minimize error, the coefficients must ensure that the signal model faithfully approximates the input function. The optimal solution for obtaining such coefficients again relies on orthogonal projection. The algorithm employed uses straightforward signal processing, and the procedure is exactly the same as the one dictated in Shannon's theorem, with the exception that the filters are not necessarily ideal.

A paramount concern in any practical sampling procedure is that the results be consistent. Specifically, it should be possible to reconstruct a signal that yields exactly the same measurements as that of the originating system. Assuming that the measurements of a function within a nonlimited bandwidth are obtained by sampling the original, prefiltered version, it is then possible to approximate the original function by applying a suitable digital correction filter. The most basic form of sampling occurs when a signal is defined in terms of its sample values. The challenge lies in finding coefficients that will interpolate those values not only consistently but also without error. Since it is not always possible to reconstruct a signal perfectly, control of approximation error is therefore crucial to efficient signal processing. The basic device for accomplishing this lies in selecting a sampling step that will keep approximation error within an acceptable threshold. The premise is that approximation error will steadily decay, and eventually vanish, as the sampling step gets smaller. The ability to predict the actual rate of decay, called order of approximation, is a significant factor in wavelet and approximation theory.

Publisher Item Identifier S 0018-9219(00)03154-6.

The paper compares and evaluates three standard algorithms used in sampling: sampling without prefiltering, sampling with suboptimal prefiltering, and least squares sampling. The first approach corresponds to standard interpolation (the typical example is piecewise linear interpolation); the second uses the simplest possible analog filter (a box function); and the third uses the optimal prefilter. Of these options, the first is the least favorable, as the absence of filtering creates aliasing. The second, which uses oblique projection rather than orthogonal, lends flexibility to the sampling process and is only slightly suboptimal. The best, in terms of performance, is the third, since it is closest to the Shannon paradigm in its use of ideal filters. This comparison clearly emphasizes the importance of prefiltering for the suppression of aliasing.

In the final section of the paper, the author considers related topics that can be considered as variations and extensions of sampling theory. These include wavelets, generalized sampling, finite elements and multiwavelets, frames, and irregular sampling. He points out that the analysis tools and mathematics used in wavelet sampling are essentially the same as those used in modern formulations of sampling theory. In this sense, research into wavelet sampling has supplied positive feedback on sampling and generated renewed interest in this field. Generalized sampling includes such recent examples as motion-compensated analysis of television images and a process called super resolution, which attempts to construct high-resolution images from a series of samples derived from low resolution images.

Another interesting generalization is multisampling, which uses several generating functions instead of one. This corresponds to the finite-element, or multiwavelet, framework. With finite elements, the functions typically chosen are as short as possible and involve minimal overlap. Because of the importance of finite elements in engineering, the quality of this type of approximation has been studied thoroughly by approximation theorists. Last, a frame is basically a set of functions that span the entire signaling space but that are not necessarily linearly independent.

Irregular, or nonuniform, sampling is another area now undergoing intense research. A problem addressed by irregular sampling, and one that has been studied most extensively, is that of recovering a bandlimited function from nonuniform samples. In these circumstances, stable reconstruction of a signal is important, particularly because there are sets of samples that uniquely determine a bandlimited function, but for which the reconstruction is unstable. It is possible, however, to reconstruct bandlimited functions perfectly from nonuniform samples, and efficient methods for performing such reconstructions do exist.

Fifty years later, Shannon's sampling theory still appears alive and well, part of the basic knowledge of every engineer involved with digital signals or images. Far from being a closed subject, sampling is likely to grow in use and importance as analog systems increasingly give way to digital ones in the evolving information age. Researchers representing different disciplines—engineers involved in signal and image processing and mathematicians steeped in harmonic analysis, mathematical physics, and approximation theory—are joining efforts to advance research into sampling theory with substantial results. The author concludes that the general view of sampling that has emerged over the past decade will provide a unifying framework for understanding and improving many techniques that have traditionally been studied separately.

—Richard O'Donnell