

The Origins of the Sampling Theorem

Hans Dieter Lüke, Aachen University of Technology

ABSTRACT

Fifty years ago the publications of Claude E. Shannon brought the sampling theorem to the broad attention of communication engineers. This article demonstrates how practitioners, theoreticians, and mathematicians discovered the implications of the sampling theorem almost independent of one another.

In 1948 and 1949, Claude E. Shannon published the two revolutionary papers in which he founded the information theory [1, 2]. In [1] the sampling theorem is formulated as "Theorem 13":

Let f(t) contain no frequencies over W. Then

$$f(t) = \sum_{-\infty}^{\infty} X_n \frac{\sin \pi(2Wt - n)}{\pi(2Wt - n)}$$

where

$$X_n = f\left(\frac{n}{2W}\right).$$

It was not until these papers were published that the theorem known as "Shannon's sampling theorem" became common property among communication engineers, although Shannon himself writes in [2] that

This is a fact which is common knowledge in the communication art.

A few lines further on, however, he adds:

... but in spite of its evident importance [it] seems not to have appeared explicitly in the literature of communication theory.

The following analysis takes the above statement as its starting point. It will become apparent that mathematicians, practitioners, and theoreticians in communication engineering came across the implications of the sampling theorem almost independent of one another, and that the links between them did not emerge until later stages of this development.

THE PRACTITIONERS

In communication engineering, the first experiments with time-division multiplexing (TDM) in telephony led to the questions of how and how often it is necessary to sample a continuous-time signal.

The attempt to transmit more than one signal simultaneously over a single wire began shortly after the early commercial successes with telegraphy in the 1840s. The first proposals for TDM using synchronously rotating commutators derive from F. C. Bakewell (1848), A. V. Newton (1851), and M. B. Farmer (1853). Technically more accomplished methods were then developed by B. Meyer (1870), J. M. E. Baudot (to 1874), as well as P. Lacour and P. B. Delany (1878) [3, 4]. It is significant not only that methods were used in which complete telegraphic signals from different transmitters were placed in chronological order (e.g., Baudot), but that certain systems were also equipped with fast rotating commutators which were able to transmit at least two samples of each elementary signal (e.g., Delany). This technique makes additional synchronization between transmitter

and sampler unnecessary. One of these fast rotating commutators, the "distributor" of the telegraphy system by F. J. Patten (around 1891), was used for the

first demonstration of TDM of telephone signals. The inventor's name was Willard M. Miner. He had his method patented in 1903 following many years of preliminary experiments [5]. Figures 1 and 2 from [6] show the circuit diagram and the "Patten Distributor" which was used. Miner determined the required sampling rate experimentally [6]:

It will be understood, then, that the apparatus devised by Mr. Miner, while in its general form the same as that heretofore used for multiplex telegraphy — or telephony for that matter — such apparatus is run at a much greater speed so as to bring the frequency of the closures of connection upon the several branches or sub-circuits up to a rate approximating in greater or less degree the rate of the vibrations of the overtones characterizing speech. A rate of closure of 1,000 or 2,000 per second will not answer the purpose, but as the rate increases and passes beyond 3,000, improved results become apparent, and are markedly better when a rate of 3,500 or 3,600 per second is reached; the best results being obtained with a rate of about 4,300 per second.

Miner thus assumed that the sampling rate would coincide approximately with the upper frequency components of speech. In actuality, his telephone apparatus will have had a cutoff frequency of barely more than 2 kHz, which fulfills the requirements of the sampling theorem.

Since a theoretical clarification of the sampling process was not forthcoming, pronouncements concerning the sampling rate in publications as well as patent applications for TDM of speech signals remained similarly vague right up until the 1930s. For example, L. von Kramolin in 1923 writes in a patent on TDM:

... therefore it is possible to work with a switching speed which lies beyond the limits of audibility, whereby the switching noise in the individual telephones is avoided and noiseless communication is possible.

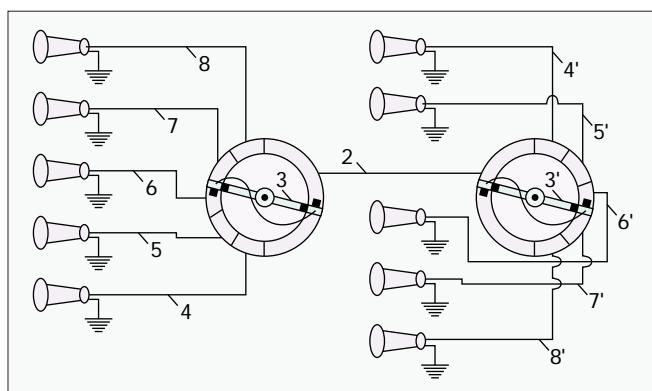


Figure 1. Circuit diagram.

In the 1930s several TDM systems for telephony were developed. However, as Cattermole remarks [7]:

The situation about 1936, then, was that sampling and TDM telephony were known empirically though the theory was rudimentary

Some authors, such as M. Marro in 1938, appear to give too-low sampling rates for speech transmission. Marro uses broad sampling pulses for a duplex TDM system. Here, the effect must be taken into account that if the sampling pulses are broadened, it is possible to reduce the sampling rate and still achieve the same word intelligibility. This dependency was examined quantitatively by G. A. Miller and J. C. R. Licklider; their results are shown in Fig. 3 [8]. According to these results, word intelligibility with a dropping sampling rate is only reduced monotonely for very short sampling impulses (up to a relative width of approximately 6 percent of the sampling period). For sampling impulses of greater relative width, by contrast, the intelligibility again increases in the range of a sampling rate of 10 to 100 Hz.

THEORETICIANS

Theoretical communication engineers did not begin working on the problem of sampling until surprisingly late. H. Nyquist and K. Küpfmüller in 1924 proved that the number of telegraph signals which can be transmitted over a line is proportional to the product of transmission time and bandwidth. R. V. L. Hartley in 1928 generalized this result with respect to multilevel transmission. Also in that year, Nyquist derived his famous theorem on distortionless transmission of telegraphic (digital) signals. But the distortionless transmission at Nyquist rate and the error-free interpolation of sampling pulses of an analog signal are different problems, even though there are some mathematical similarities. Therefore, these works cannot be regarded as sources for the sampling theorem, especially during the 1920s and 1930s.

The first scientist to formulate the sampling theorem precisely and apply it to problems of communication engineering is probably V. A. Kotelnikov. In his work "On the transmission capacity of 'ether' and wire in electrical communications," published in 1933, he proves the sampling theorems for lowpass signals as well as for bandpass signals [9]. He uses these theorems in the course of his work to show that the bandwidth of an analog signal cannot be reduced by modulation methods. The lowpass sampling theorem is formulated as follows:

THEOREM I

Any function F(t) which consists of frequencies from 0 to f₁ periods/s may be represented by the following series:

$$F(t) = \sum_{-\infty}^{+\infty} D_k \frac{\sin \omega_1 \left(t - \frac{k}{2f_1} \right)}{t - \frac{k}{2f_1}}, \quad (1)$$

where k – integer

$$\omega_1 = 2\pi f_1$$

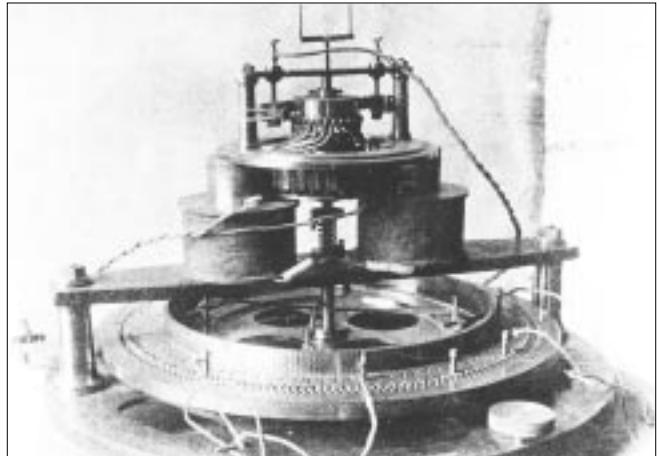
D_k – constant which depends on F(t).

Conversely, any function F(t) which is represented by the series in Eq. (1) only consists of frequencies from 0 to f₁ periods/s.

⋮

THEOREM II

Any function F(t) which consists of the frequencies from 0 to f₁ may be transmitted continuously to any desired degree of precision using numbers which follow on with 1/2f₁ s. From the mea-



■ Figure 2. The "Patten Distributor."

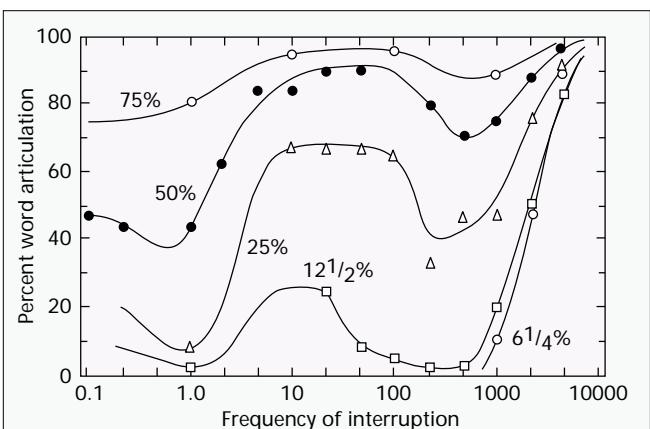
surement of the value of F(t) for t = n/2f₁ (n is an integer), we indeed obtain

$$F\left(\frac{n}{2f_1}\right) = D_n \omega_1 \quad (8)$$

because all terms of the series in Eq. (1) for this value t tend to zero, with the exception of the term with k = n that equals D_nω₁, which may easily be established after calculation of the indefinite point. Thus, we may recognize the next D_k according to the progress after each 1/2f₁ s. If we transmit these D_k one after another, following each 1/2f₁ s, we can reconstruct F(t) termwise according to Eq. (1) to any degree of precision.

Since this noteworthy work has never been published in internationally accessible form, the publications on the theoretically exact formulation of the sampling theorem in the literature of communication engineering came about independent of one another. Thus, H. Raabe deduced the sampling theorem in his Ph.D. thesis and published it in 1939 [10]. This publication is especially relevant to the practical application of the sampling theorem, since it is here that the influence of sampling impulses of finite duration is taken into account in the form of "natural sampling." Raabe summarizes his findings:

For the demonstrated conditions of transmission, the sampling frequency is determined by the range of signal frequencies. If these are kept below half of the sampling frequency, all of the noise frequencies remain above this limit and can be kept away from the receiver by a lowpass filter. The transmission of a signal may thus be completely distortionless, if the sampling frequency is twice the highest signal frequency. The upper limitation of the signal fre-



■ Figure 3. Results of quantitative examination.

quencies is therefore a vital condition of distortionless transmissibility by time division multiplex transmission.

This work also contains a special sampling theorem for bandpass signals. Raabe's work is cited in a related publication by W. R. Bennett from the year 1941 [11], and Bennett's work is in turn cited by Shannon in [2] as one of the sources of the sampling theorem.

Lastly, it should be mentioned that the sampling theorem is also treated in 1949 in the Japanese book *Hakei Denso (Signal Transmission)* by I. Someya. Hence, the term "Someya's Theorem" may be found in some Japanese literature.

THE MATHEMATICIANS

For mathematicians, the sampling theorem is a special theorem from the field of approximation theory. Approximation theory asks, for example, what functions can be represented by a linear sum of given basic functions such as algebraic or trigonometric polynomials and with what approximation error. One possible approach is to determine these linear sums such that they take on at definite points the same values as the function which is to be approximated.

In this sense, sampling theorems make it possible to establish how this task of interpolation can be solved, especially for functions bounded in the frequency domain with a vanishing approximation error.

A first approach in this direction was described as early as 1765 by J. L. Lagrange. Lagrange determines a linear sum in harmonical sinus functions in such a way that it coincides with the function which is to be approximated at n equidistant points. Generalizing this approach, it may be said that knowledge of $2n + 1$ equidistant functional values of one period is sufficient in order to represent a periodic function which may be described by a trigonometric series, each with n sine and cosine terms as well as one constant. This well known theorem may be viewed as a sampling theorem for bandlimited periodic functions.

The first proposal for the interpolation of equidistant functional values using the $\sin(x)/x$ function was published in 1908 by C.-J. de la Vallée Poussin in the *Bulletin Académie Royale de Belgique*. However, the special significance of this interpolation for bandlimited functions is not yet explored in this work.

E. T. Whittaker's paper "On the functions which are represented by the expansions of the interpolation theory" in 1915 must therefore be regarded as the first work ever to address the sampling theorem for all bandlimited functions [12]. Whittaker addresses the problem of achieving the smoothest possible interpolation without singularities and without "rapid oscillations" for given tabular values of a function $f(x)$. In order to fulfill the latter condition, he shows that under certain conditions it is possible to interpolate given sampling values at intervals of w such that the Fourier transform of this interpolation function does not contain any terms with periods less than $2w$. This interpolation named "cardinal function" $C(x)$ by Whittaker has the form

$$C(x) = \sum_{r=-\infty}^{\infty} \frac{f(a + rw) \sin \frac{\pi}{w} (x - a - rw)}{\pi (x - a - rw)}$$

(a is an arbitrary shifting term) with the following characteristics:

We defined it originally as that unique function of the cotabular set, which has no singularities in the finite part of the plane and no constituents whose period is less than twice the tabular interval w .

Here the "cotabular set" refers to the set of all possible functions with the same tabular or sampling values, whereby the "tabular interval" then corresponds to the sampling interval.

Furthermore, Whittaker demonstrates that the cardinal function is the *only* interpolation function with these characteristics. This also implicitly postulates that every function whose Fourier transform is limited to frequencies $< 1/2w$ may be described by sampling values at intervals of w , and can uniquely be interpolated again in the form of the cardinal function.

W. L. Ferrar pointed out in 1925 that sampling and interpolation of the cardinal function itself again lead to the same function, irrespective of how the sampling values are shifted in time. He refers this important invariance property of the sampling theorem as "consistency." Subsequent publications on the cardinal function are cited by J. M. Whittaker (not to be confused with E. T. Whittaker) [13]. Shannon then refers to this book in [2].

By way of conclusion, it should be mentioned that an extensive tutorial review of the development of the sampling theorem after Shannon was published by A. J. Jerri in 1977 [14].

CONCLUSION

The sampling theorem for lowpass functions plays an important role in communication engineering as a connecting link between continuous-time and discrete-time signals. The numerous different names to which the sampling theorem is attributed in the literature — Shannon, Nyquist, Kotelnikov, Whittaker, to Someya — gave rise to the above discussion of its origins. However, this history also reveals a process which is often apparent in theoretical problems in technology or physics: first the practitioners put forward a rule of thumb, then the theoreticians develop the general solution, and finally someone discovers that the mathematicians have long since solved the mathematical problem which it contains, but in "splendid isolation."

REFERENCES

- [1] C. E. Shannon, "A mathematical theory of communication," *Bell Sys. Tech. J.*, vol. 27, 1948, pp. 379–423, 623–56.
- [2] C. E. Shannon, "Communication in the presence of noise," *Proc. IRE*, vol. 37, 1949, pp. 10–21.
- [3] K. E. Zetsche, *Geschichte der elektrischen Telegraphie*, Berlin: Springer Verlag, 1877.
- [4] A. E. Granfeld, *Die Mehrfach-Telegraphie auf einem Drahte*, Vienna: A. Hartleben's Verlag, 1885.
- [5] W. M. Miner, "Multiplex Telephony," U.S. Pat. 745 734. Filed, Feb. 26, 1903.
- [6] W. M. Miner., "Recent developments in multiplex-telephony," *Elec. World and Eng.*, vol. 42, 1903, p. 920.
- [7] K. W. Cattermole, *Principles of Pulse Code Modulation*, London: Iliffe Books, 1969.
- [8] G. A. Miller and J. C. R. Licklider, "The intelligibility of interrupted speech," *JASA*, vol. 22, 1950, pp. 167–73.
- [9] V. A. Kotelnikov, "O propusknoj sposobnosti 'efira' i provoloki v elektrosvjazi," ("On the transmission capacity of 'ether' and wire in electro-communications"), *First All-Union Conf. Questions of Commun.*, Jan. 14, 1933.
- [10] H. Raabe, "Untersuchungen an der wechselzeitigen Mehrfachübertragung (Multiplexübertragung)," *Elektrische Nachrichtentechnik*, vol. 16, 1939, pp. 213–28.
- [11] W. R. Bennett, "Time division multiplex systems," *Bell Sys. Tech. J.*, vol. 20, 1941, pp. 199–221.
- [12] E. T. Whittaker, "On the functions which are represented by the expansions of the interpolation-theory," *Proc. Roy. Soc., Edinburgh*, vol. 35, 1915, pp. 181–94.
- [13] J. M. Whittaker, *Interpolatory Function Theory*, Cambridge Univ. Press, 1935.
- [14] A. J. Jerri, "The Shannon sampling theorem — its various extensions and applications: a tutorial review," *Proc. IEEE*, vol. 65, 1977, pp. 1565–96.

BIOGRAPHY

HANS DIETER LUKE [SM] (lueke@ent.rwth-aachen.de) received a Dipl.-Ing. in 1960 and a Dr.-Ing. in 1964 from the Aachen University of Technology (RWTH), Germany. From 1966 to 1972 he was with the AEG-Telefunken Comp., Ulm. Since 1972 he has been a full professor of electrical communications at RWTH, Aachen. His research activities are in signal theory, image coding, and acoustic signal processing. He is author of the books *Signalübertragung* (Springer-Verlag, 1975, 1998) and *Korrelationssignale* (Springer-Verlag, 1992). He served as a member of the Board of Directors of the Informationstechnische Gesellschaft (VDE) from 1984 to 1987, and as member of the Senate of German Research Council (DFG) from 1987 to 1993.