SONY®

Introduction to DIGITAL AUDIO

by Luc Theunissen

TRAINING TEXT

SONY®

Introduction to DIGITAL AUDIO

by Luc Theunissen

TRAINING TEXT

CONTENTS

INTRODUCTION

The possible application of digital signal processing in the audio field has been known and studied for many years: the first theoretical description of pulse-code modulation in telephony actually was written more than 40 years ago!

For application of PCM in High-Fi and professional audio however, even today we are not so far from the limits of what is technologically possible. Therefore, only relatively recently, it has become possible to develop commercially viable digital audio products.
As a consequence, only a small group of high-level specialists are nowadays familiar with the theory and applications of digital audio technology.

This booklet is intended as a way to familiarize a newcomer in the field with some of the theoretical and technological aspects of PCM-audio, in the simplest possible way. We hope it will encourage towards further studies in this interesting field.

SONY SERVICE CENTRE (Europe) N.V.

PART I

Analog - to - Digital Conversion

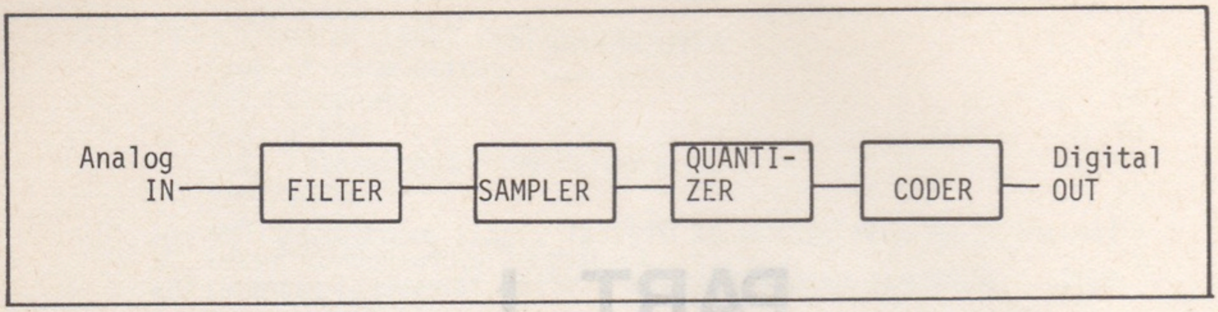

Fig. 1-1. Basic steps of analog-to-digital conversion

PART I

ANALOG-TO-DIGITAL CONVERSION

I. OUTLINE

An analog signal, be it audio or any other, has to be converted to a digital signal before it can be digitally processed. Although the principles of A/D- and D/A-conversion may seem relatively simple, in fact, this conversion between the analog and the digital domain is, technologically speaking very difficult and will unavoidably cause degradation of the original signal. Consequently, most of the time this stage will be the limiting factor that determines system performance.

Conversion from the analog to the digital domain is done in several steps, which however are not always clearly defined :

- filtering
- sampling
- quantization
- coding.

(see Fig. 1-1.)

(a)

(b)

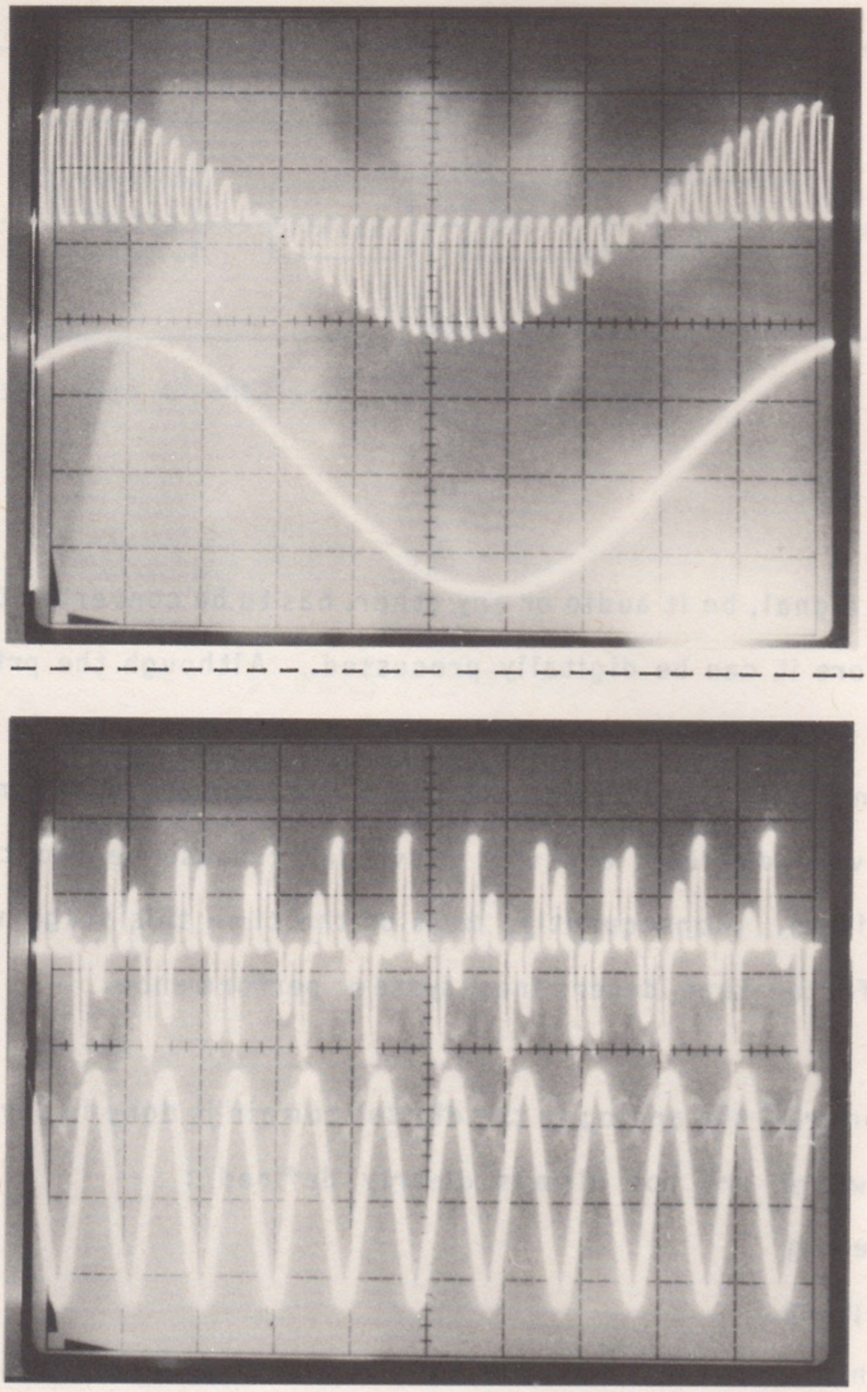

Fig. 1-2. Two examples of sine waves together with their sampled versions (f_S = 44.056kHz)

(a)1kHz sine wave

(b)10kHz sine wave

Although sampling in (b) seems much coarser than in (a), in both cases restitution of the original signal is perfectly possible.

Below, we will explain and discuss these basic steps of analog-to-digital conversion, along with the signal degradations that each conversion block is likely to introduce.

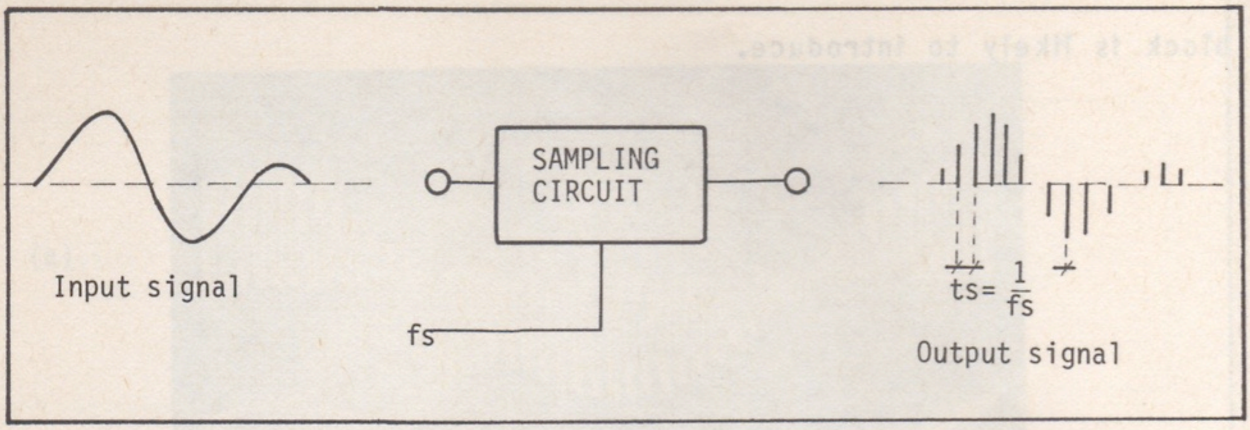

Fig. 1-3. Principle of sampling

II. SAMPLING

1. Principle - The Nyquist Theorem

By definition, an audio signal varies continuously in time. To enable it to be converted into a digital signal, it is necessary that the signal is first sampled, i.e. at certain points in time a sample must be taken of the input value, which subsequently can be converted. The fixed time intervals between each sample are called sampling intervals (t_s).

Although the sampling operation may seem to introduce a rather drastic modification of the input signal, (since it ignores all the signal changes that occur between the sampling times), it can be shown that the sampling process itself in principle removes no information whatsoever, as long as the sampling frequency is at least twice the highest frequency that is present in the input signal.

This is the famous Nyquist criterion on sampling (also called Shannon criterion or Shannon theorem).

The correctness of the Nyquist criterion can be understood when we consider the spectra of the input and output signals (Fig. 1-4.).

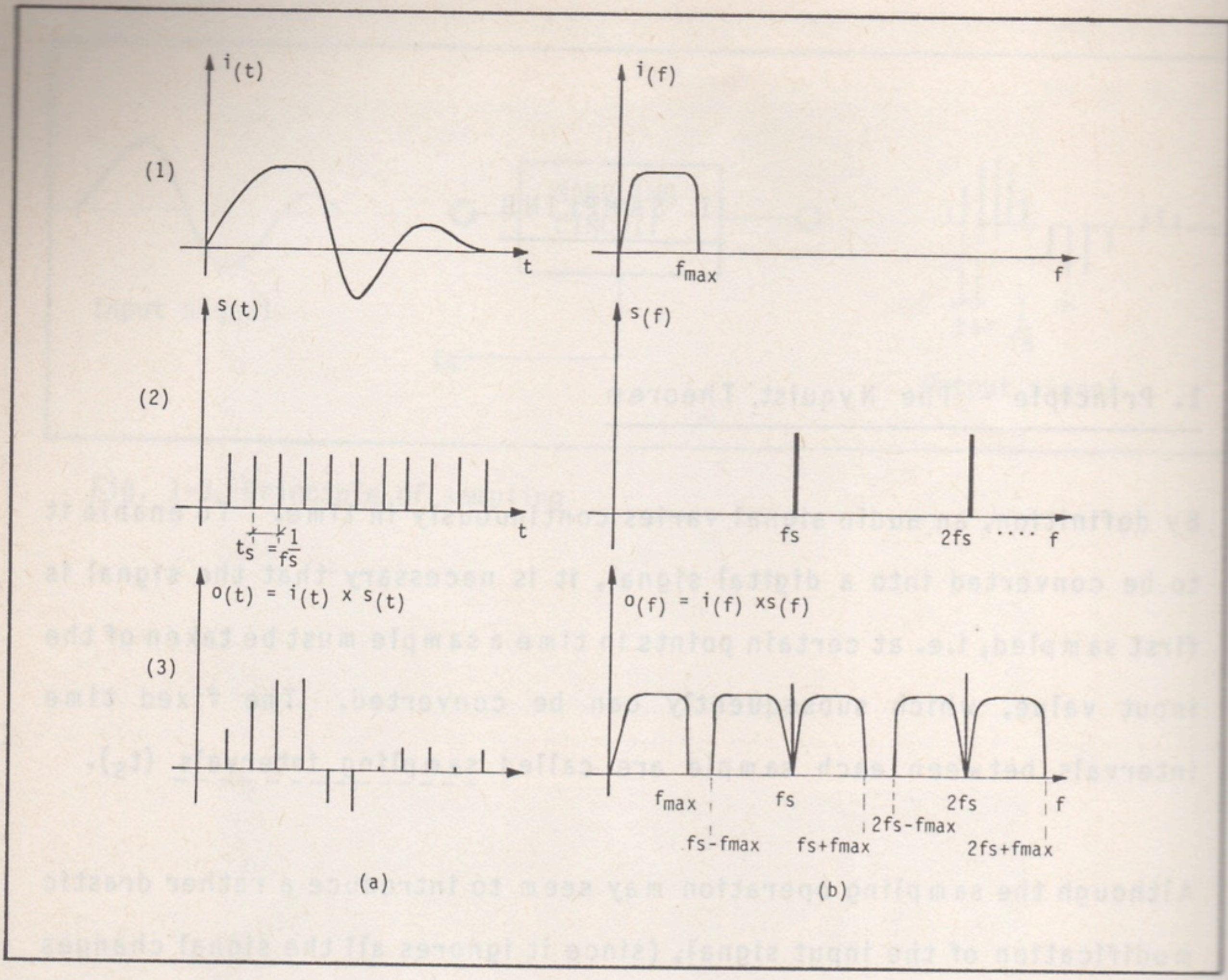

Fig. 1-4. Time domain - (a), and frequency domain in representation (b) of sampling process

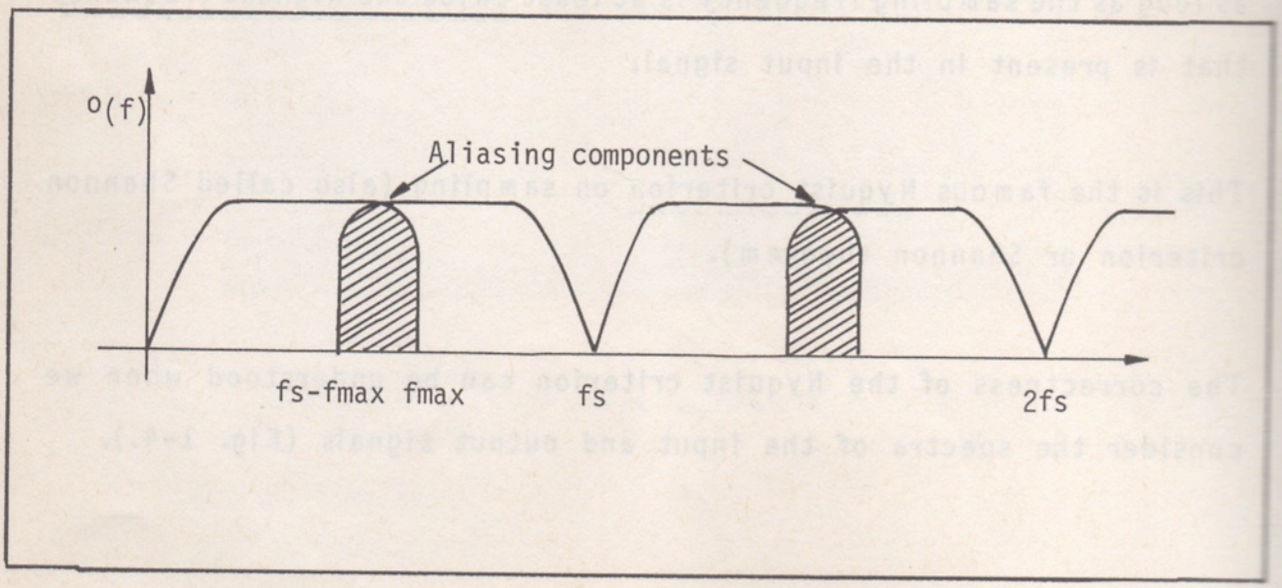

Fig. 1-5. Principle of aliasing

If we consider an input signal $i_{(t)}$ which has a maximum frequency f_{max}, its spectrum may have any form between 0 Hz and f_{max} (1); the sampling signal $s_{(t)}$, having a fixed frequency f_s, can be represented by one single line at f_s (2). The sampling process is equivalent to a multiplication of $i_{(t)}$ and $s_{(t)}$, and the spectrum of the output signal, as shown in (3), can be seen to contain the same spectrum as the input signal, together with repetition of the same spectrum, modulated around multiples of the sampling frequency.

As a consequence, proper low-pass filtering can completely isolate and thus completely recover the input signal.

From Fig. 1-4., it can also easily be understood that f_s must be greater than $2x\,f_{max}$. If this would not be the case, the original spectrum would overlap with the modulated part of the spectrum, and consequently be inseparable from it (Fig. 1-5.)

For example, a 20kHz-signal sampled at 35kHz would produce a 5kHz difference frequency. This phenomenon is known as <u>aliasing</u> and must be avoided by all means.

For this reason, a very sharp cut-off filter (anti-aliasing filter) is absolutely required in the signal path to remove all the unwanted harmonics out of the input signal <u>before</u> the sampling takes place. If this is not done, aliasing will cause unremovable distortion components.

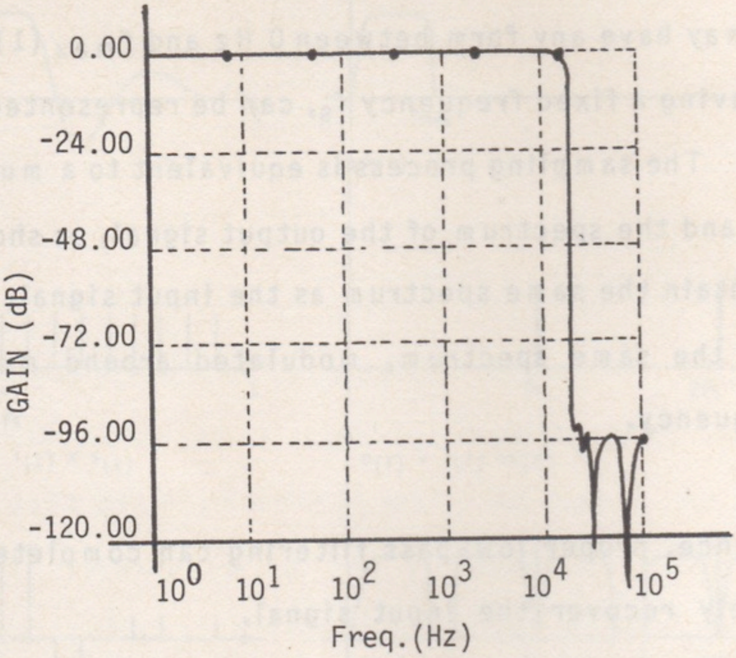

Fig. 1-6. Characteristic of anti-aliasing filter

2. The problem of the sampling frequency

2-1. Minimum value

In the digital audio field, proper selection of a convenient sampling frequency is extremely important.

It should indeed be avoided to select the sampling rate unnecessarily high, since this tends to increase the hardware costs dramatically.

On the other hand, since ideal low-pass filters do not exist, a certain safety margin must be incorporated in order to avoid any frequency higher than $0.5f_s$ from passing through the filter with insufficient attenuation.

Therefore, if we wish to reproduce a flat audio bandwidth of 20 Hz - 20.000 Hz, a sampling frequency of 44 kHz can be considered as a minimum, giving 22 kHz as extreme frequency where the "anti-aliasing filter" already gives sufficient attenuation (60 dB) to make eventual aliasing components inaudible (Fig. 1-6).

2-2. Use of 44.056/44.1 kHz

Another important criterion for selection of a sampling frequency lies in the fact that, to arrange the digital information in a video-like signal, (as it is done in all the PCM-adapters which use a standard helical-scan video recorder as a storage medium) there must be a fixed relationship between sampling frequency (f_s), the horizontal video frequency (f_h) and the vertical frequency (f_v). For this reason, these frequencies must be derived from the same master clock by frequency division, or in other words, f_s and f_h should have an as-low-as-possible Least Common Multiple.

In the NTSC system, f_h = 15.734,2657... Hz (a non-integer value, due to the necessary relationship with the NTSC chroma and audio subcarriers), whereas in the European PAL or SECAM systems, f_h = 15.625,0 Hz.

Calculations have shown that, for the NTSC system, a frequency of 44.055944... Hz would come closest to this ideal, whereas for the PAL system, a frequency of 44.100 Hz was also quite feasible.

The difference between these two frequencies is only 0,1%, which is negligible for normal use (the difference translates as a pitch difference at playback, and 0,1% is entirely imperceptible).

As a consequence, 44.056 has been adapted as sampling frequency in the EIAJ-standard for PCM-adapters for EIAJ, while 44.1 will be used by adapters for the CCIR-system, as well as for the future Compact Digital Audio Disc.

2-3. Use of 50.4 kHz

For studio recording, mostly using the multiple-track recording technique, the helical-scan recorders are not so ideal, therefore stationary-head recorders are being designed, giving up to 48-channel recording capability.

One of the aspects of such studio recorders is that the tape speed must be adjustable, in order to allow easy synchronization between several machines, correct tuning, etc. If we consider a speed tuning range of 10%, a sampling frequency of 44 kHz could become less than 40 kHz, which is insufficient to comply with the Nyquist criterion. Therefore, such machines should use a higher sampling rate, for example 50 kHz, which at the lowest speed would still give a satisfactory 45 kHz.

However, to allow direct (digital) dubbing between such stationary head recorders and helical-scan recorders (which with their two-track capability are ideal and reliable master recorders), conversion between the higher and the lower (44.1 kHz) sampling rate must be possible. To do this in an economical way, a simple mathematical relationship between both frequencies is important. A frequency that has an integer relationship (8/7) with 44.1 kHz is 50.4 kHz. Also, 50.4 kHz has a good relationship with various TV and film standards (25 frames/sec), so it is also possible to make digital data blocks coincide with frames. For these reasons, Sony as well as some other Japanese manufacturers selected 50.4 kHz for their stationary-head machines.

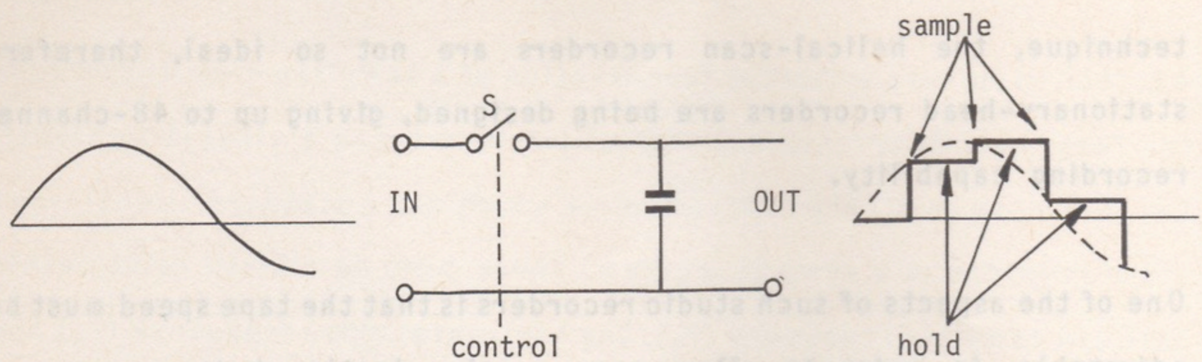

Fig. 1-7. Basic sample-hold circuit

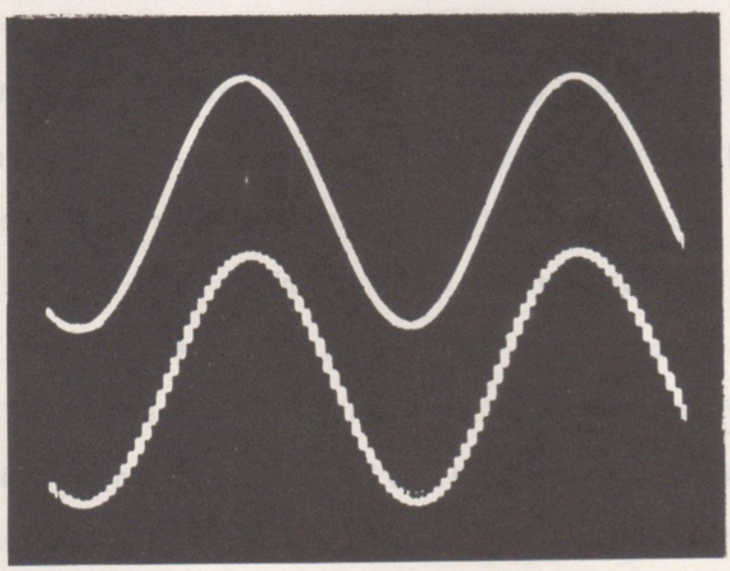

Fig. 1-8. Sine Wave before (upper) and after (lower) Sample-Hold

2-4. Use of 32 kHz

Quite a few European broadcasters are already utilizing digital transmission links between studios and transmitters. However, since the FM-broadcast frequency range is limited to 15 kHz anyhow, the EBU (European Broadcasting Union) has chosen 32 kHz as an (economical) standard sampling frequency. The same frequency will also be adopted by the Japanese P.T.T. However, this complicates the conversion problems again.

3. Sample-Hold Circuits

In the practice of analog-to-digital conversion, the sampling operation is performed by so-called Sample-Hold Circuits, that, once the sample has been taken, store the sampled analog voltage a certain time, during which the voltage can be converted by the A/D convertor into a digital code (see further).

The principle of a Sample-Hold Circuit is relatively simple (Fig. 1-7.).

A basic Sample-Hold Circuit is in fact a 'voltage memory' device that stores a given voltage in a high-quality capacitor. To sample the input voltage, switch S closes momentarily; when S opens, capacitor C holds the voltage until C closes again and passes the next sample. Fig. 1-8. shows a sine wave at the input and the output of a S/H circuit.