3.0 Features extraction procedures.

3.1 Systematic of time-dependent procedures

Irregular (stochastic) temporal variable procedures one can divide

zeitlich veränderlich			
deterministisch		stochastisch	
periodisch	nicht periodisch	stationär	nicht stationär
sinusförmig	komplex periodisch	ergodisch	nicht ergodisch

Figure 3.01: Systematic time-dependent procedures

The analytic treatment of such oscillation procedures requires statistic methods of treatment:

deterministically: A procedure is deterministically called, if an explicitly mathematical connection exists.
Example: sinusoidal course of motion, which is described mathematically by the time function $x(t) = A\sin(\omega_0 t)$.

stochastically: A stochastic procedure does not possess explicit mathematical description. There is no such procedure to determine an exact function value, since any long time-section is unique. Irregular procedures must be described therefore by statistic parameters and statistic knowing functions. With deterministic procedures the mathematical connection permits to forecast a function value with certainty. The use of statistic characteristics leads however to the statement, by which probability a function value arises.
With irregular procedures you must differentiate between stationary and non stationary stochastic processes. Their characteristics are determined by a quantity of time-functions.

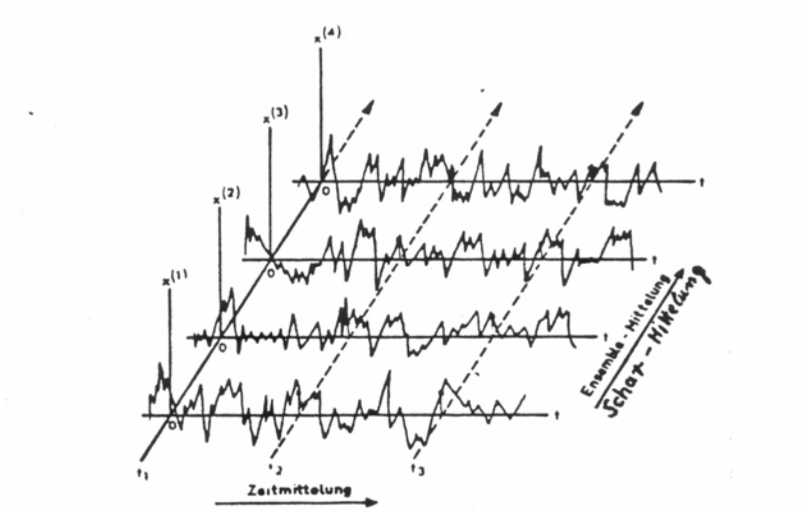


Figure 3.02: Explanation to quantity time-average of stochastical functions

A quantity is the series of time-functions, which are seized on same measuring points of several identical systems during one observation period (e. g. resistor random noise). In case the statistic characteristic values, which result transverse to the quantity at the time t_1, do not change with the time and are equal to those at the time t_2 so we have a stationary process. If the statistic characteristics of the quantity change during the movement of time $x(t) = x(t + \tau)$, so the procedure is not stationary.

The process is called ergodic, if the statistic characteristics, which are determined transverse to the quantity, are equal to those, which apply along a time average function. This unique function represents the statistic behaviour of all time functions. The quantity average values may be replaced by time average values. Stationary and ergodisch behaviour of stochastic processes is not always given with existent tasks. Nevertheless the application of the theory supplies useful results.

3.2 Statistic descriptions of stochastic processes

The use of the theory for stochastic processes leads to four statistic definitions, which describe the basic qualities of random processes.

a.) mean values,
b.) probability density function,
c.) autocorrelation function,
d.) spectral power density function.

The characteristics of a) to c) characterize the behaviour in the time section, while the characteristic of d.) supplies information in the frequency domain. The relations derived in 3.3 to 3.6 are valid for time-functions x(t) of an ergodic process.

3.3 Mean values

Physical procedures are to be divided into a static, time-independent component and into a dynamic component. The linear average value describes the static component.

$$\overline{X} = \frac{1}{N}\sum_{n=0}^{N-1} X_n$$

The variance describes the dynamic part.

$$\sigma_X^2 = \frac{1}{N-1}\sum_{n=0}^{N-1}(X_n - \overline{X})^2$$

Simplified calculation of variance, if being constantly measured and computed

$$\sum_n (X_n - \overline{X})^2 = \sum_n (X_n^2 - 2X_n\overline{X} + \overline{X}^2)$$

$$= \sum X_n^2 - 2\overline{X}\sum X_n + \sum \overline{X}^2 \qquad \sum \overline{X}^2 = N\overline{X}^2$$

$$= \sum X_n^2 - \frac{2}{N}\left(\sum X_n\right)^2 + N\left(\frac{\sum X_n}{N}\right)^2$$

$$= \sum X_n^2 - \frac{1}{N}\left(\sum X_n\right)^2$$

therefore

$$\sigma^2 = \frac{1}{N-1}\left(\sum_{n=0}^{N-1} X_n^2 - \frac{1}{N}\left(\sum_{n=0}^{N-1} X_n^2\right)^2\right)$$

Advantage:
The differences do not arise and the influence of an error in rounding-off will become smaller with $\overline{X}$. The mean square $\overline{X}^2$ value is measure of the intensity of a procedure.

$$\overline{X}^2 = \frac{1}{N}\sum_{n=0}^{N-1} X_n^2$$

The definition of σ^2 gives evidence that the variance is equal to the square mean value, if the linear mean value is set by a suitable transformation to zero.

$$\sigma^2 = \overline{X}^2$$

The root out of σ^2 supplies the standard deviation σ and the rms X_{eff}, which is also an intensity measure. Since the mean process improves the statistic security of the results and reduces the variance it is often used with EKG- and EEG analysis and serves the weakening of disturbing signals.

- **Energy and power of a signal**

The referred momentary power of any signal x(t) is defined:

$$P_x(t) = |x(t)|^2$$

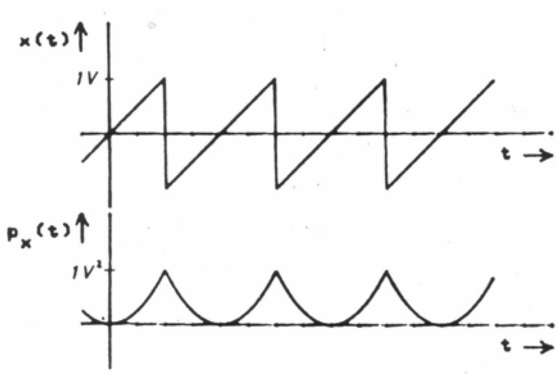


Figure 3.03: Signal and momentary power

The momentary power is used for calculating the mean power:

$$S_x = \lim_{T\to 0} \frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} p_x(t)\,dt = \lim_{T\to 0} \frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} |x(t)|^2\,dt = \overline{|x(t)|^2}$$

In case of pulse-type procedures this formula leads on $S_x = 0$. The rms of a signal x(t) determined by the root from its power.

$$x_{rms} = +\sqrt{S_x}$$

If the mean power S_x of a signal is not infinitely large and larger than zero, the signal is called power-limited:

$$0 \langle S_x \langle \infty$$

To calculate the mean power of periodic signals you only have to compute over one period T_0

$$S_x = \overline{|x(t)|^2} = \frac{1}{T_0} \int_{-\frac{T_0}{2}}^{\frac{T_0}{2}} |x(t)|^2\,dt$$

We compute the momentary and mean power of a saw-tooth signal and a cosine signal.

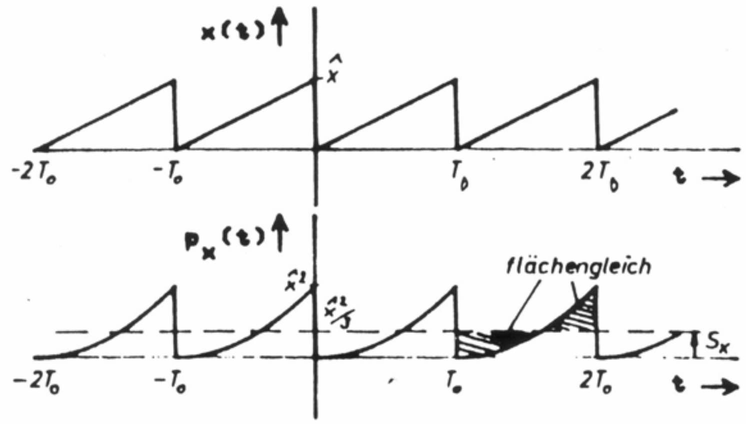


Figure 3.04: Momentary and mean Power of a saw-tooth signal

$$S_x = \overline{|x(t)|^2} = \frac{1}{T_0}\int_0^{T_0} |\hat{x}\frac{t}{T_0})|^2\,dt = \frac{\hat{x}^2}{T_0}\frac{t^3}{3T_0^2}\Big|_0^{T_0} = \frac{\hat{x}^2}{3}$$

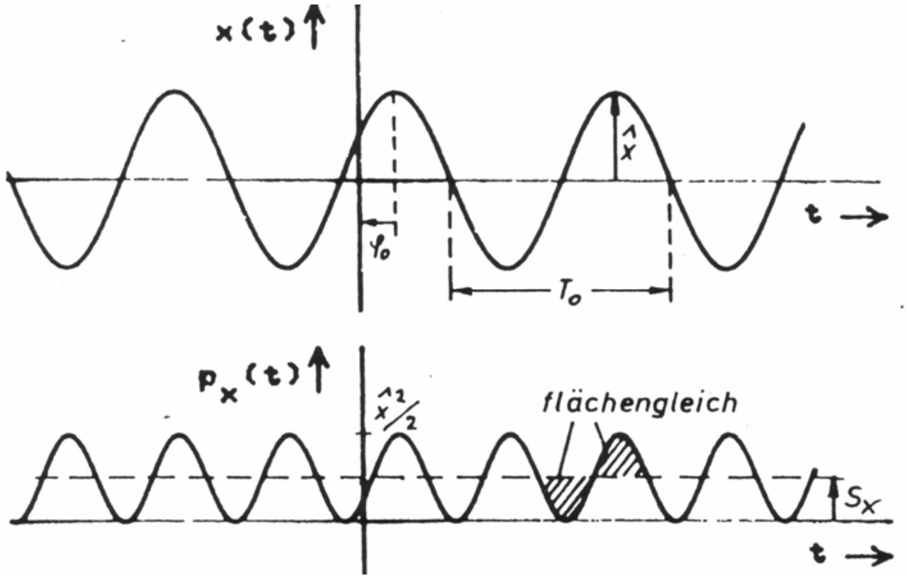


Figure 3.05: Momentary and mean Power of a cosine signal

$$S_x = \overline{|x(t)|^2} = \frac{1}{T_0}\int_0^{T_0} \hat{x}^2 \cos^2 \omega_0\, t)\,dt = \frac{\hat{x}^2}{T_0}\int_0^{T_0} \frac{1}{2}(1+\cos(2\omega_0\, t))\,dt$$

$$= \frac{\hat{x}^2}{T_0}\left(\frac{1}{2}t + \frac{1}{4\omega_0}\sin(2\omega_0 t\right)\Big|_0^{T_0} = \frac{\hat{x}^2}{2T_0}T_0 = \frac{\hat{x}^2}{2}$$

Energy:

If the energy E of a signal x(t) possesses a finite value, the signal is energy-limited:

$$0 < E_x < \infty$$

e.g. impulses are energy-limited signals.

3.4 Probability-density function

(distribution density function)

The most perfect description of an irregular procedure represents the probability – density- function. It gives information about the distribution of the individual values of the time-function in the time-interval by indicating the probability, in which an irregular procedure x(t) at any time can take a certain value between x and $x + \Delta x$.

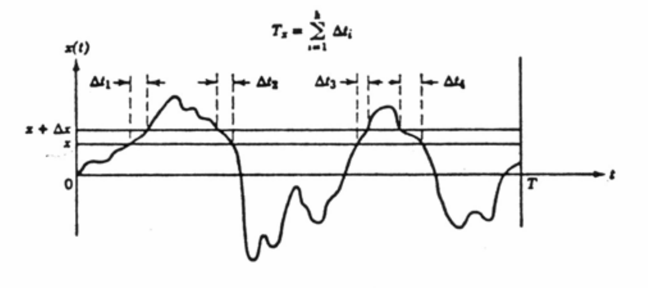


Figure 3.06: Explanation of the probability

The density of probability is defined:

$$p(x) = \lim_{\Delta x \to 0} \frac{\text{probabilty}(x < x(t) < x + \Delta x)}{\Delta x} = \lim_{\Delta x \to 0} \frac{1}{\Delta x} \left(\lim_{x \to \infty} \frac{\sum_{i=1}^{k} \Delta t_i}{T} \right)$$

The function p(x) describes basically stochastic signals. But usually in digital signal processing you begin with known determined signals to compare them later with stochastic signals.

- **Probability-density of a sine-signal**

For a sine-signal $x(t) = x_0 \sin(\frac{2\pi}{T_0} t)$ you can calculate p(x) as it is shown in the following diagram

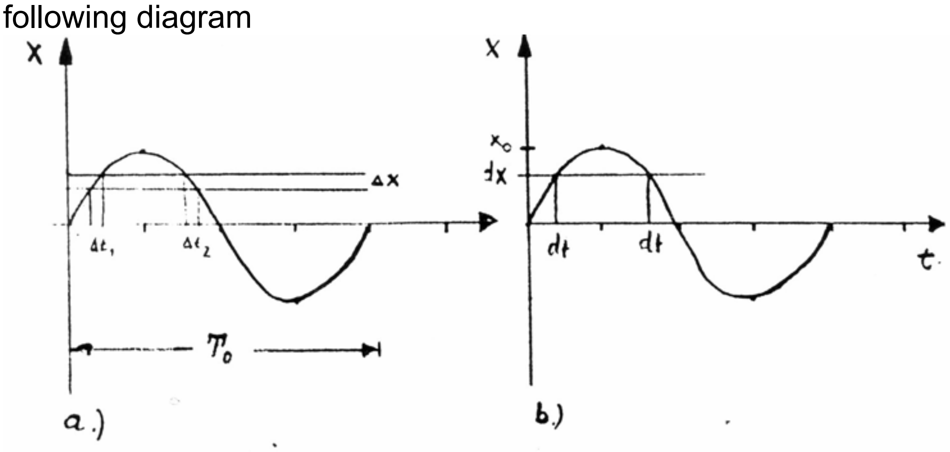


Figure 3.07: Derivation of probability-density of a sine signal

Figure 3.07 shows two time intervals Δt_1 and Δt_2, which describe the function in interval Δx. After the transition to infinitesimal size (dx and dt - see fig. 3.07b), in the range of values $-x_0 \le x \le x_0$ of the signal x(t) 2 portions dt can be indicated, which supply a contribution within the period duration T_0 to p(x). By using the definition of the probability-density we see:

$$p(x) = \frac{1}{T_0} 2 \left| \frac{dt}{dx} \right| \quad \text{it is valid} \quad \frac{dx}{dt} = \frac{2\pi}{T_0} x_0 \cos(\omega_0 t).$$

we calculate the reverse value $\left| \frac{dt}{dx} \right| = \frac{T_0}{2\pi x_0 \, |\cos(\omega_0 t)|}$ and replace it like this

$$|\cos(\omega_0 t)| = \sqrt{1 - \sin^2(\omega_0 t)} = +\sqrt{1 - \left(\frac{x(t)}{x_0} \right)^2}$$

The result is:

$$p(x) = \frac{1}{\pi x_0 \sqrt{1 - \left(\frac{x}{x_0} \right)^2}}$$

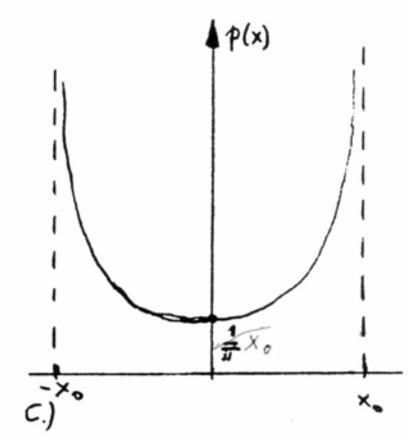


Figure 3.08: Probability-density of a sine-function x_0= 1

- **Probability-density of a triangular signal**

A Further example shows us the probability-density of a triangular signal with the amplitude x_0 and periodic duration T_0.

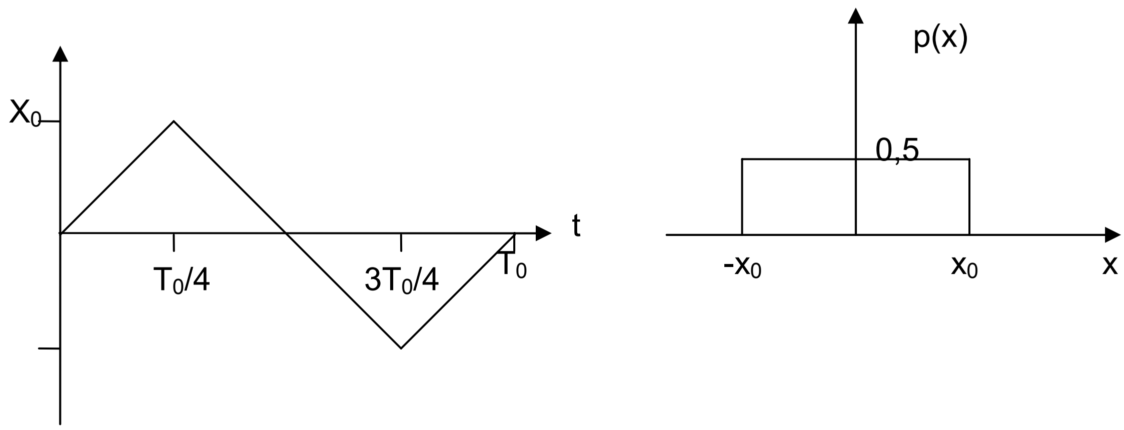


Figure 3.09: Derivation of probability-density of a triangular signal

Just like we did with the sinus we look for $\frac{dx}{dt}$ and find out that

$$|\frac{\Delta x}{\Delta t}| = |\frac{dx}{dt}| = \frac{4x_0}{T_0} \qquad \text{is constant, except for} \quad t = \frac{T_0}{4} \text{ and } t = \frac{3}{4}T_0.$$

As in the range of value like in the sinus function during the periodic duration T_0 two parts dt are to be found, the same connection is valid:

$$p(x) = \frac{1}{T_0} 2 |\frac{dt}{dx}| = \frac{1}{2x_0}$$

Figure 3.09 shows on the right the probability-density for x_0=1 as a rectangular uniform distribution with the constant value 0.5.

In order to lead up to the stochastic signals, four typical time functions are represented in figure 3.10a. Figure 3.10b shows the probability-density functions

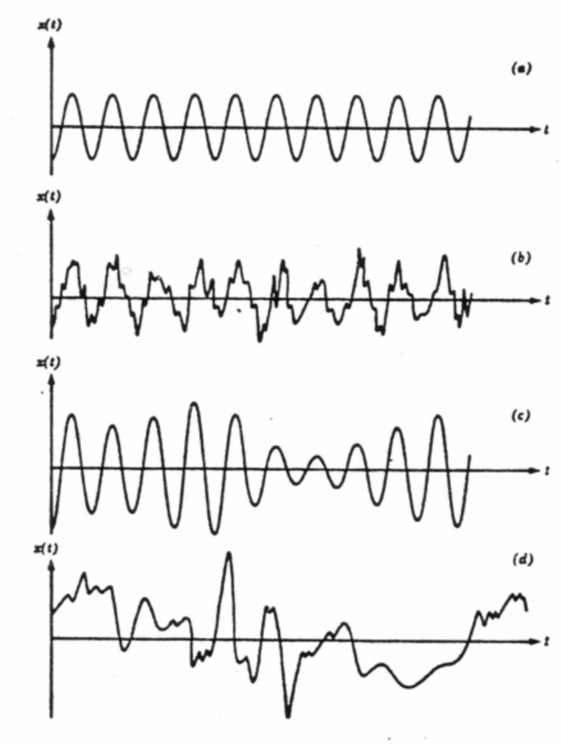


Figure 3.10:
Four typical time functions
a.) Sine
b.) Sine with noise
c.) Narrow band noise
d.) Wide band noise

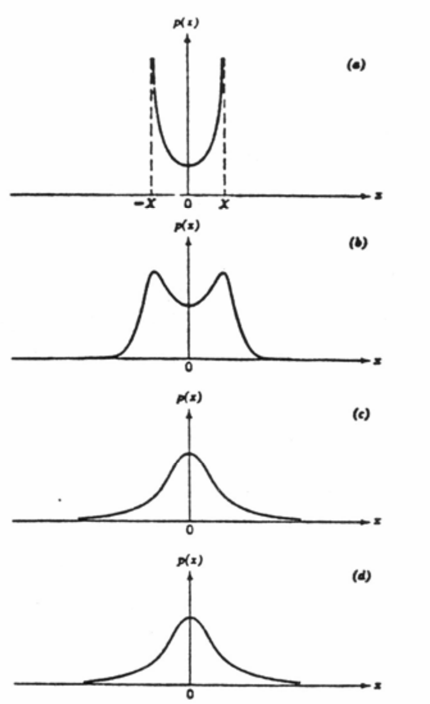


Figure 3.11:
Probability-density function
a.) Sine
b.) Sine with noise
c.) Narrow band noise
d.) Wide band noise

In figure 3.11b you can see the influence of the sinus in the probability density.
In figure 3.11c and 3.11d the probability-density functions are very similar.

The Gauss' stochastic process in figure 3.11d embodies the ideal case of an irregular process. The probability-density function belonging to this process is clearly determined by the indication of the linear average value and the standard deviation. Only the Gauss process possesses this characteristic. The probability-density function makes you recognize if the signals are too weak, also in case the signals are over-regulated.

3.5 Correlation and autocorrelation function ACF

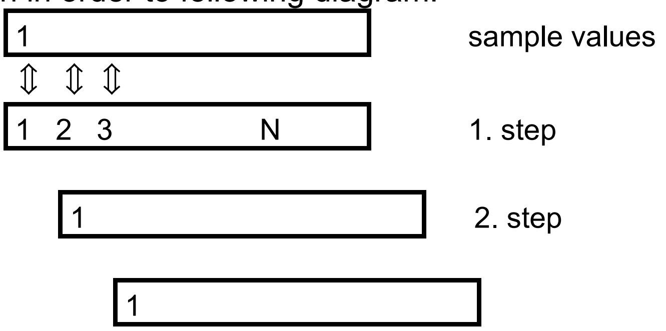
The correlation function results from the multiplication of **two** different functions.

Definition: $$CF(k) = \frac{1}{N-k} \sum_{n=0}^{N-k-1} x_1(n) \cdot x_2(n+k)$$

The autocorrelation function calculates the connection from **one** function. It results from the multiplication of function values x(n) and x(n+k). x(n+k) is a function shifted with $k\Delta t$

Definition: $$ACF(k) = \frac{1}{N-k} \sum_{n=0}^{N-k-1} x(n) \cdot x(n+k)$$

You get the squared mean value for k= 0. Compare the autocorrelation-, correlation- and convolution function, which are described in chapter 2, page 2.
Numerical calculation in order to following diagram:



This function examines the degree of the internal relationship of different sections of a time function $x(n\ \Delta t)$. If $x(n\Delta t)$ possesses irregular behaviour, dependence is not present, the function ACF(k) is monotonous falling, $ACF(\infty)$ strives against $\overline{x}^2$. If periodic portions are contained in $x(n\Delta t)$, ACF(k) shows also periodic behaviour. Therefore the autocorrelation function serves among other things the recognition of periodic signals, which are embedded into a stochastic environment.

Characteristics of the ACF function

- the ACF is an even function.
- the initial value of the ACF is exceeded by no other value of the function.
  $ACF(0) \leq ACF(k)$, $AKF(0) = \overline{x}^2$ equal the variance of the signal.
- a standing dc component a_0 increases the values of the ACF by the constant value a_0^2.
- the ACF of a periodic signal has the same periodicity.
- the ACF does not contain a phase information.
- the larger the range of a stochastic signal, the faster the ACF falls to zero.

Figure 3.12 shows the autocorrelation functions of the time-signals presented in figure 3.10.

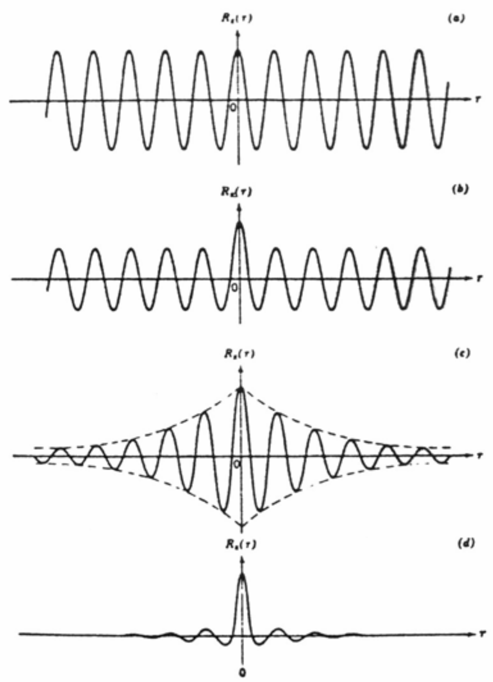


Figure 3.12: Autocorrelation functions
a.) Sine
b.) Sine with noise
c.) Narrow band noise
d.) Wide band noise

Signals such as pulse series, bearing noises and transmission noises often can be identified more clearly by ACF than by time-synchronous averaging or by frequency-analysis. You can use the presented procedures with linear and two-dimensional signals. E. g. the two-dimensional Fourier transformation offers some insights, to help you judge the process procedure in a better way. Look at this example from [Meyer-Broetz70] to understand, what a two-dimensional signal means.

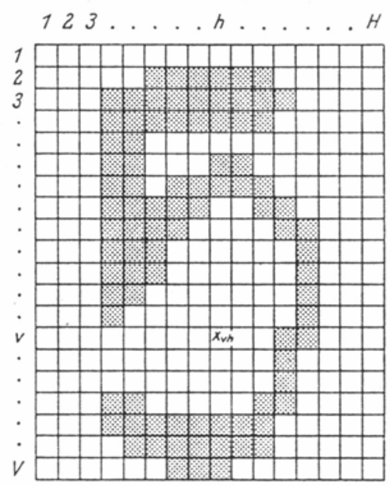
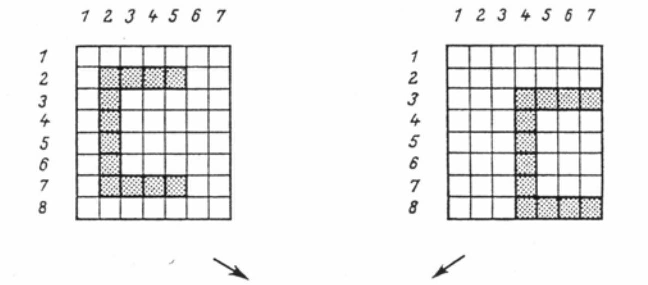


Figure 3.13: Gray-values of a two-dimensional signal range domain 8 bit.

Application of the ACF to a two-dimensional signal:
Obviously the ACF has its maximum, if the shifted picture lies accurately on the first picture. You can use this characteristic for the recognition and to find a transformation independent of position.



Calculate the autocorrelation function

	-6	-5	-4	-3	-2	-1	0	1	2	3	4	5	6
-7	·	·	·	·	·	·	·	·	·	·	·	·	·
-6	·	·	·	·	·	·	·	·	·	·	·	·	·
-5	·	·	·	1	2	3	4	3	2	1	·	·	·
-4	·	·	·	1	1	1	2	1	1	1	·	·	·
-3	·	·	·	1	1	1	3	1	1	1	·	·	·
-2	·	·	·	1	1	1	4	1	1	1	·	·	·
-1	·	·	·	1	1	1	5	1	1	1	·	·	·
0	·	·	·	2	4	6	12	6	4	2	·	·	·
1	·	·	·	1	1	1	5	1	1	1	·	·	·
2	·	·	·	1	1	1	4	1	1	1	·	·	·
3	·	·	·	1	1	1	3	1	1	1	·	·	·
4	·	·	·	1	1	1	2	1	1	1	·	·	·
5	·	·	·	1	2	3	4	3	2	1	·	·	·
6	·	·	·	·	·	·	·	·	·	·	·	·	·
7	·	·	·	·	·	·	·	·	·	·	·	·	·

Figure 3.14: Two grey value distributions, which differ in translation and their common autocorrelation function.

We come to a description of the sample invariant of position.

3.6 Clipped autocorrelation function ACF

Since there were made good experiences on the classification using the characteristics of the autocorrelation function, it is practical to use a simplified or clipped autocorrelation function. This method was successfully used in recognition of isolated words [Calavrytinos78]. In the following one the definition of the clipped autocorrelation function and the characteristics are described. If one wants to use the coefficients of the autocorrelation function just alike this for recognition purposes, it is the disadvantage, that they depend sensitively on fluctuations of the sampled values $x(n\Delta t)$, since they enter the equation as a product [Jesorsky76]. Two unwanted effects result out of this fact. On one side the fluctuations in a coefficient ACF(1) are strengthened. The second effect derives from the connection between autocorrelation function and power density spectrum. Since the autocorrelation function is the inverse Fourier transformation of the power density spectrum, the amplitudes of the individual Fourier components enter the calculation squarely. If therefore a Fourier component with a large amplitude is present, it will cover all the other Fourier components due to the square formation. These difficulties are decreased, introducing a clipped autocorrelation function.

$$ACF_{cli}(k) = \sum_{n=0}^{N-k-1} x(n)\,\mathrm{sign}(n+k)$$

An additional advantage is the considerable reduction of computation-time (no multiplication). The dynamic range is approximately given by the square root from the dynamic range of the standard autocorrelation. Clipping "sign" is a high-grade non linear operation. Approximately clipping in the time domain corresponds to a dynamic reduction in the spectral domain. In the following text those characteristics of the clipped autocorrelation are described, which are substantial for its use as distinguishing features:

- it reaches the maximum with k = 0, which coefficient ACF(0) is called magnitude average value and represents an estimated value for the intensity of the signal segment.
- it depends linear on the absolute value of the scaling factor, if the signal segment is strengthened as a whole.
- it keeps the periodic structures, which are present in the signal element.

The signal must be band-pass filtered, e. g. within a language-signal 300-3000Hz. The lower cut off frequency is extremely important, because the clipping operation is very susceptible against a DC voltage part and low frequency components.

3.7 Spectral power density function

If the autocorrelation function supplies the statistic characteristics of linear and square average values in the time domain, the power density function informs about the spectral distribution of the square average value in the frequency domain. Correlation function and power density function are connected over the Fourier transformation. This connection is well-known as a Wiener Khintchin theorem. Further data are to be found in [Winkler77]. Figure 3.15 the shows you the power density functions of the time-signals presented in figure 3.10.

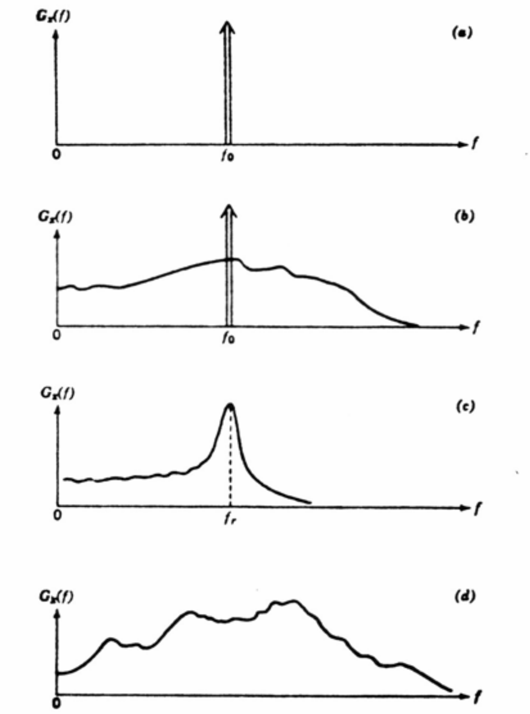


Figure 3.15: Spectral power density function
a.) Sine
b.) Sine with noise
c.) Narrow band noise
d.) Wide band noise

3.8 Quasi moments

Another global description is the computation of statistic moments (especially the quasi-moments) from the amplitude distribution, the autocorrelation function or the spectral function, which are used directly as features. The linear power density spectrum can be regarded as such due to its similarity with a distribution density (only positive values; expressed maxima) with an appropriate standardisation. The standardisation consists of a division of the spectral values by the total output, whereby the total area below the spectral function takes the value 1. A distribution density exactly demands this condition. The used quasi-moments are calculated according to [Winkler77] out of the central moments nth order (further examples see in MATLAB Statisic Toolbox).

$$Q_n = \mu_n = E\left\{(x-\mu)^k\right\}$$

Presupposed the mean value is equal 0 and the variance σ is equal 1, it is valid:

$$\begin{aligned} Q_3 &= \mu_3 && \text{skewness} \\ Q_4 &= \mu_4 - 3\mu_2^2 && \text{excess} \\ Q_5 &= \mu_5 - 10\mu_2\mu_3 \\ Q_6 &= \mu_6 - 15\mu_2\mu_4 + 30\mu_2^3 \end{aligned}$$

The straight quasi-moments particularly mark the flatness and the odd symmetry of a distribution density. The skewness is zero, if the distribution density is symmetrical, the excess describes the deviation from the normal distribution concerning an extension or a compression of the Gauß function, see figure 3.16.

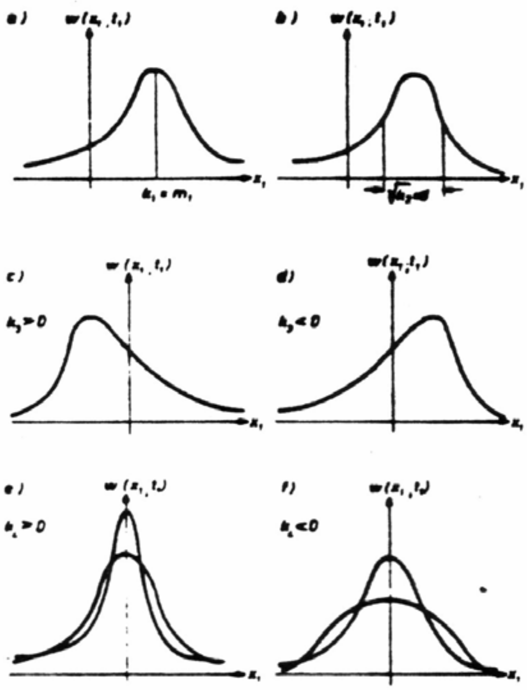


Figure 3.16: a) mean value. b.) variance c.) and d.) skewness e) and f) excess

3.8 Cepstrum analysis

This procedure has attained practical meaning with the discovery of echoes in signals, with the separation of multiplicatively linked consequences as well as with the development of signals convoluted with one another [Paul82].

- **The powercepstrum is defined as the power density spectrum of the logarithmic power density spectrum of a signal.**

If you omit the logarithm, you receive the squared autocorrelation. Now this spectrum is made whiter by the logarithm of the power density spectrum and so weak spectral lines are emphasized opposite the strong ones. In the acoustic pattern recognition the Cepstrum is often used for the classification of vehicles and for the automatic determination of the firing order frequency of combustion engines [Thomas72].

Example

The incitating time-signal consists of periodically sequential explosions in the distance T, while the transfer function of the system can be regarded as wide-band against the clock frequency $\frac{1}{T}$. The system causes a low-frequency wave shape in the logarithm spectrum. The periodicity of the incitation is expressed by the high frequency wave shape with the frequency period $\frac{1}{T}$. The richly harmonic structure in the original spectrum of $F(\omega)$ causes a periodicity in the new time-function C(q). Therefore the spectrum of the logarithmic spectrum possesses a distinctive line, which corresponds to the high frequency wave shape.

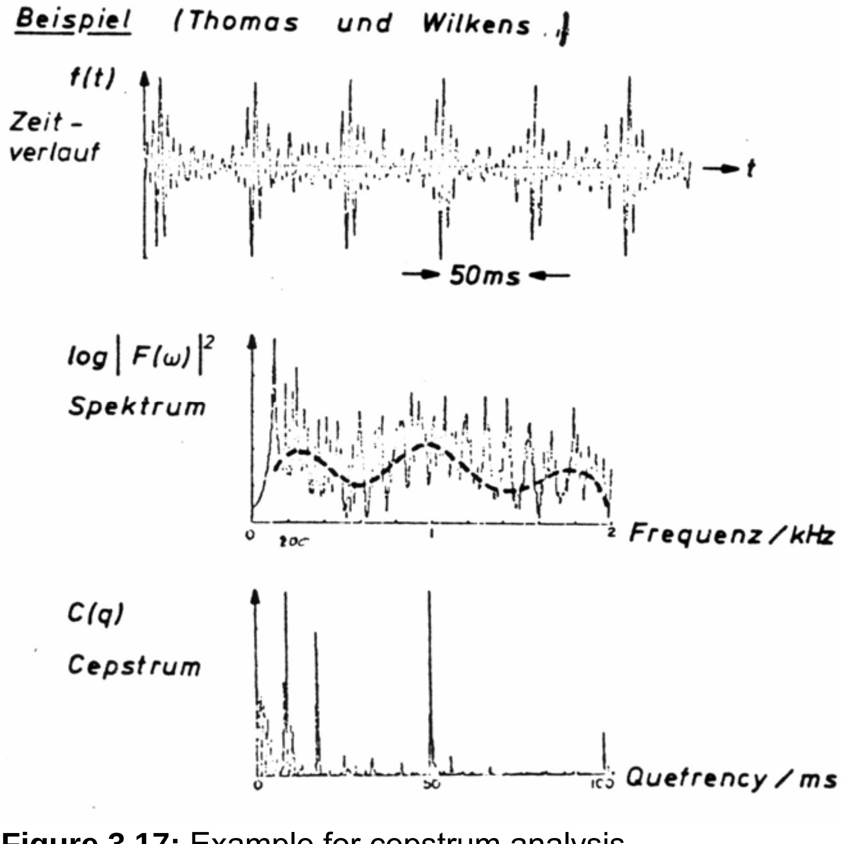


Figure 3.17: Example for cepstrum analysis

Finally another publication [Weck83] is left to be mentioned. In this paper the Cepstrum-analysis shows characteristics of the process and the machines independently of the task of monitoring. This paper examines the noises of a drilling process. The Fourier analysis did not supply a clear description of condition, whereas the characteristic values for the machines and process conditions could be derived by the Cepstrum-analysis.

Keep in mind:

Compute the spectrum from a time-signal. Compute the logarithm of the spectrum magnitude. This signal is regarded once again as a time-signal and the spectrum of this is to be calculated again. If you discover separation-effective features as an observer, the procedure is suitable very well for the pattern recognition task.