CHAPTER 4

Systems, automata, and grammars

Hendrik G. STASSEN and Willem J.M. LEVELT
Laboratory for Measurement and Control, University of Delft
Max-Planck-Projektgruppe für Psycholinguistik, and University of Nijmegen

1. Systems

1.1. Introduction

The study of today's complex problems – whether they originate from physics or engineering, from physiology or biology or medicine, from economics or industrial management, from psychology or sociology – leads to a growing tendency of specialization toward different disciplines. As a result we are able to build up an enormous amount of basic knowledge in particular areas or disciplines. However, this specialization makes the communication between the disciplines more and more difficult, or often even impossible. Yet the need to solve real world problems, which generally may be characterized by their strongly multi-disciplinary character, demands a high degree of communication between these disciplines. We are anxious, therefore, to develop one common language. For this, the systems approach may serve, since in every scientific research project we can recognize three essential, and very common phases:

(1) The study and formulation of real world problems, resulting in one or another qualitative and/or quantitative model;

(2) The study of the model behavior; that is, the performance of a sensitivity analysis of those factors which may influence the model results, so that finally, on the basis of the model, predictions can be made in newly designed situations;

(3) The interpretation and translation of the model results to the original real world problem.

For such communication, system theory can be very helpful because

it can be considered as a universal tool in formulating and solving a great variety of problems. Before we are able to explain the real contribution of system theory to experimental psychology we first must come to a generally acceptable definition of what a system is. This is not easy, because most definitions in literature have been restricted to a particular field of interest. In general we may say that a system is a part of the real world, separated from its environment, and that it may or may not have a relation to this environment. This means that the environment may act on the system, and vice versa. The chosen boundaries of the system are arbitrary, and are dependent on the investigator's interests and goals. System theory can contribute to the formulation of models in order to describe the system behavior by supporting:

- Methods to formulate system models in completely different fields, largely by recognizing analogies;
- Methods for the analysis and identification of systems, and for the quantification of the interactions between system and environment;
- Methods to classify different systems.

We re-emphasize that system theory can contribute significantly to model formulation. In particular by the analysis of input-output relations we will be able to understand the structure, the parameters, and thus the dynamics of the system under study.

1.2. General system definition

A more precise definition of a system is the following: *A system is a bounded part of the environment in which a certain structure is specified, and which may have an interaction with its environment.* We now define the interaction between system and environment by *inputs* and *outputs.* That means that the environment acts on the system by the inputs, whereas the system acts on the environment by outputs. In the case where the inputs and outputs are defined as a function of time, we call them *signals.* Input signals can be divided into *non-controllable inputs* or *disturbances* or *noises,* and *controllable inputs* or *control signals.* The most commonly used notations for the control signals, disturbances and outputs are $u(t)$, $v(t)$, and $y(t)$, respectively (fig. 1).

We can represent a system by means of a block diagram. Here the block represents the system itself, the control inputs are entering the block from the left, the disturbances from above, and the outputs are leaving the block to the right. In addition the *initial conditions,* that is the

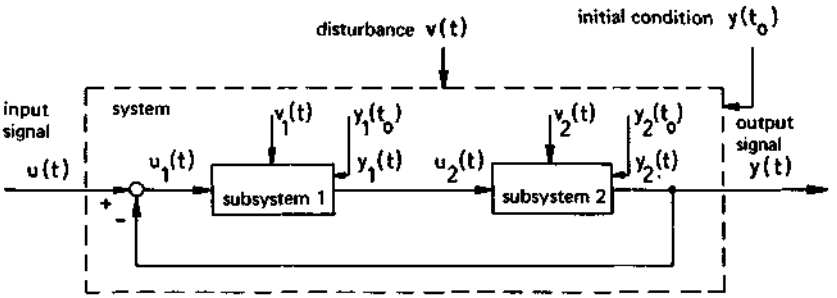

Fig. 1. An example of a system with control input $u(t)$, disturbance $v(t)$, output $y(t)$, and initial condition $y(t_0)$.

condition of the system before the inputs acted upon it, enters the block from the right.

In the context of this chapter we will only deal with *causal* systems, that is, we will deal with systems where the outputs are the result of the inputs. We will call those outputs *responses*.

1.3. Signal description

The description of systems can easily be reduced to the description of signals and their mutual relations, without any loss of generality, since the system simply transfers the input into an output. The system only performs an operation on a signal. Therefore we will first describe the different types of signals.

1.3.1. Signal characterization

The way we characterize signals is dependent on the properties of interest. A possible breakdown is the following:

– *Deterministic* versus *stochastic:* A deterministic signal $x(t)$ is a function for which the amplitude is uniquely defined for each value of t; a stochastic signal $\bar{x}(t)$ is a function of time t which cannot be defined in such a way; it is defined in terms of statistical properties such as probability density functions or the moments derived herefrom (the notation $\bar{x}(t)$ refers to a stochastic signal, whereas $x(t)$ stands for a deterministic one, fig. 2).

– *Continuous* versus *sampled:* A continuous signal is defined for all values of t, whereas a sampled one is just defined at particular instants of time (fig. 3).

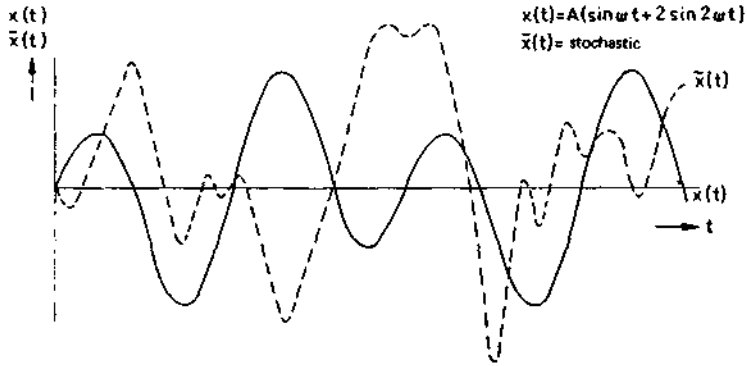

Fig. 2. An example of a deterministic and a stochastic signal.

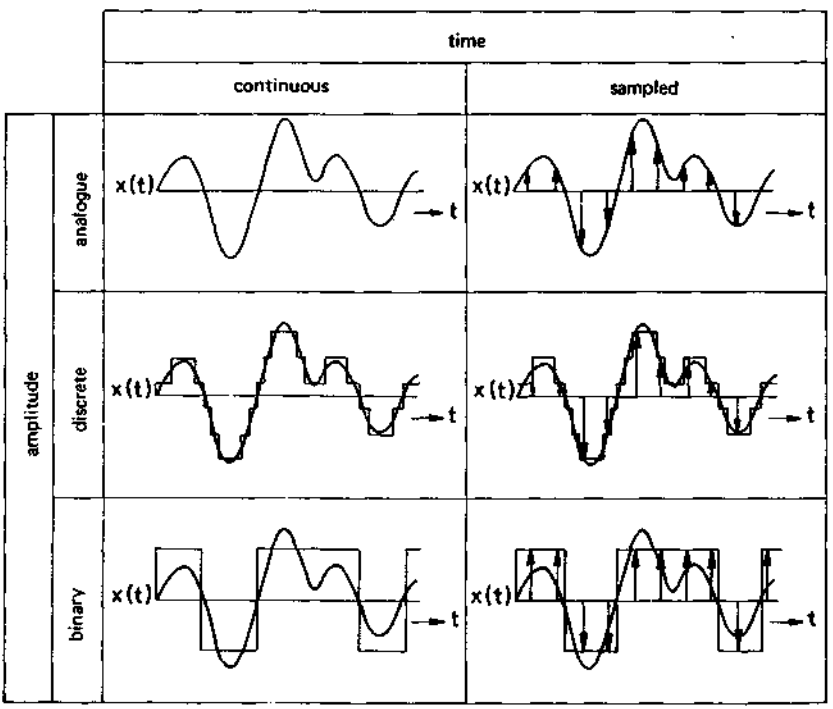

Fig. 3. Characterization of signals. In this example the mean values of the signals are assumed to be zero.

– *Analogue* versus *discrete* or *binary:* A signal is called analogue when the signal amplitude can have an infinite number of values within a certain limited interval, whereas a discrete signal will have a finite number of values. A particular case of a discrete signal is the one that has just two values; this is called a *binary signal.*

– *Periodic* versus *non-periodic:* A signal $x(t)$ is said to be periodic with a finite time period T when $x(t) = x(t + T)$ for every value of t. A non-periodic signal will not possess this property; we can consider it as a periodic signal with a periodic time T equal to infinity (fig. 4).

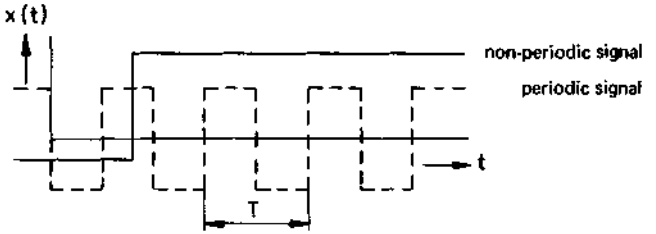

Fig. 4. Example of a periodic signal and a non-periodic signal.

1.3.2. Decomposition of signals: The Fourier series

In order to describe the process of the *decomposition* of signals, we can use the afore-mentioned distinction between deterministic and stochastic signals. Since deterministic signals, such as the sinusoidal function, the step function, the ramp function and the impulse function (fig. 5), easily can be described as functions of time, these functions can be approximated by the summation of a finite or infinite number of *subsignals.* The special significance of such subsignals is that linear systems all have the property that the response on the summation of a set of subsignals equals the summation of the individual responses to each of the subsignals. This extremely important property implies that

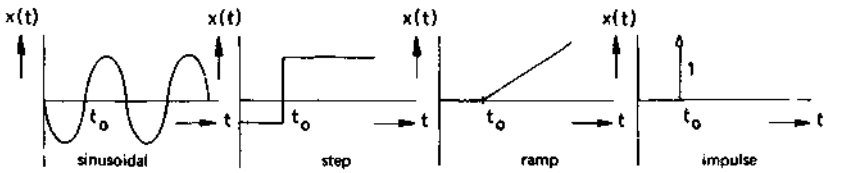

Fig. 5. Some examples of often used deterministic signals.

if the response of a system on a set of subsignals is known, the response on any arbitrarily chosen input can be derived directly.

We also can consider the decomposition process as equivalent to building a *model* of the signal, where by choosing the *structure of the model*, that is by choosing the subsignals, a set of *unknown parameters* must then be determined (fig. 6b). The procedure of the decomposition process is as follows: Assume that we will approximate the signal $x(t)$ by the summation $\tilde{x}(t)$ of a number of a priori chosen subsignals $u(t)$, each provided with an unknown coefficient c_k. Then it follows that:

$$\tilde{x}(t) = \sum_{k=0}^{N} c_k u_k(t). \tag{1}$$

Furthermore, let us assume that we would like to fit the approximation over the interval $[t_1, t_2]$ according to the *criterion*:

$$J = \int_{t_1}^{t_2} |x(t) - \tilde{x}(t)|^p \, w(t) \, dt. \tag{2}$$

Now, the *optimal approximation* $\tilde{x}(t)$ of $x(t)$ can be found by *minimizing* the criterion function, in which the function $w(t)$ is called the *weighting function*, the function $x(t) - \tilde{x}(t)$ the *error function*, and the interval $[t_1, t_2]$ the *approximation interval*. The *exponent* p determines to what extent the error contributes to the value of the criterion function. Most often *quadratic criteria* are used, that is $p = 2$, since in this form the mathematical derivation is very simple. Then the optimal solution is

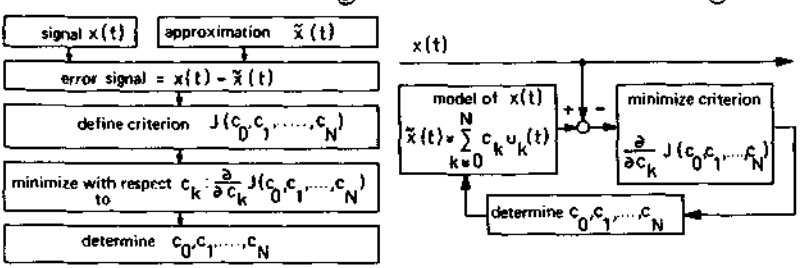

Fig. 6. Decomposition of a signal into subsignals.

obtained by setting to zero the partial derivatives of the cost function J with respect to the unknown parameters c_k, hence,

$$\int_{t_1}^{t_2}\left[x(t) - \sum_{i=0}^{N} c_i u_i(t)\right]u_k(t)w(t)\,dt = 0;$$

$$k = 0, 1, \ldots, N. \tag{3}$$

The equation obtained in this way is called the *normal equation*; from this the coefficients c_k can be derived. Much dull and needless arithmetic can be avoided in determining the coefficients c_k by an intelligent choice of the set of subsignals $u_k(t)$, as well as by choosing a reasonable weighting function $w(t)$. If we choose those according to:

$$\int_{t_1}^{t_2} u_k(t)u_l(t)w(t)\,dt = \begin{cases} 0 & \text{for } k \neq l; \\ d_k & \text{for } k = l, \end{cases} \tag{4}$$

where d_k is a constant, it follows directly from the normal Eqs (3) that:

$$c_k = \frac{1}{d_k}\int_{t_1}^{t_2} x(t)u_k(t)w(t)\,dt. \tag{5}$$

Functions $u_k(t)$ as given by the Eqs (4), where the weighting function $w(t) = 1$, are called *orthogonal* over the interval $[t_1, t_2]$. A great variety of functions will satisfy the Eqs (4), but it is indisputable that the most commonly used one is the sinusoidal function, finally resulting in a Fourier series. If we approximate a given signal $x(t)$ over the finite approximation interval $[t_0, t_0 + T]$ by $\tilde{x}(t)$, then it follows that:

$$\tilde{x}(t) = \sum_{k=0}^{\infty}\left[a_k \cos k\omega t + b_k \sin k\omega t\right] \tag{6}$$

with $\omega = 2\pi/T$. The approximation (6) is called the *Fourier series*. Following the procedure just mentioned with $p = 2$, $w(t) = 1$, and $c_k u_k(t) = a_k \cos k\omega t + b_k \sin k\omega t$, we obtain:

$$a_0 = \frac{1}{T}\int_{t_0}^{t_0+T} x(t)\,\mathrm{d}t,$$

$$a_k = \frac{2}{T}\int_{t_0}^{t_0+T} x(t)\cos k\omega t\,\mathrm{d}t \quad \text{for } k = 1, 2, \ldots,$$

$$b_k = \frac{2}{T}\int_{t_0}^{t_0+T} x(t)\sin k\omega t\,\mathrm{d}t \quad \text{for } k = 1, 2, \ldots \tag{7}$$

By eliminating the coefficients a_k and b_k, that is by substituting the Eqs (7) into Eq. (6), we obtain the *Fourier series*. Without proof, the following important properties of the Fourier series are to be mentioned:

- The coefficients a_k and b_k are dependent solely upon k.
- In choosing the approximation interval $[t_0, t_0 + T]$, the initial time t_0 is arbitrary.
- Extension of the approximation $\tilde{x}_N(t)$, based on the summation of N subsignals, to $\tilde{x}_{N+1}(t)$ will result in a lower value of the cost criterion J, and thus in a better approximation; hence:

$$\lim_{N\to\infty} J_N = 0, \quad \text{and} \lim_{N\to\infty} \tilde{x}_N(t) = x(t).$$

In the literature, the Fourier series as given by Eq. (6) is often given in a different way. Based on Euler's formula,

$$\cos \omega t = \tfrac{1}{2}(\mathrm{e}^{j\omega t} + \mathrm{e}^{-j\omega t}); \sin \omega t = \frac{1}{2j}(\mathrm{e}^{j\omega t} - \mathrm{e}^{-j\omega t}), \tag{8}$$

a complex version of the Fourier series can be formulated:

$$x(t) = \sum_{k=-\infty}^{\infty} r_k\,\mathrm{e}^{jk\omega t} \quad \text{with } \omega = 2\pi/T, \tag{9}$$

$$r_k = \frac{1}{T}\int_{t_0}^{t_0+T} x(t)\,\mathrm{e}^{-jk\omega t}\,\mathrm{d}t, \quad \mathrm{k} = 0, \pm 1, \pm 2, \ldots, \tag{10}$$

where the coefficients r_k are defined as follows:

$$\left.\begin{aligned} r_0 &= a_0, \\ r_k &= \tfrac{1}{2}(a_k - jb_k), \quad k = 1, 2, \ldots, \\ r_{-k} &= \tfrac{1}{2}(a_k + jb_k), \quad k = 1, 2, \ldots \end{aligned}\right\} \tag{11}$$

It follows that:

$$x(t) = A_0 + \sum_{k=1}^{\infty} A_k \cos(k\omega t - \varphi_k), \tag{12}$$

$$\left.\begin{aligned} A_0 &= a_0, \quad \varphi_0 = 0, \\ A_k &= \sqrt{a_k^2 + b_k^2}, \quad \varphi_k = \operatorname{arctg} b_k/a_k. \end{aligned}\right\} \tag{13}$$

We can consider the components of the Fourier series as sinusoidal subsignals with a radial frequency $k\omega$, and amplitude A_k, and a phase shift φ_k. Usually the quantities A_k and φ_k are plotted as functions of the radial frequency $k\omega$; the diagram obtained in this way is called the *spectrum* of a signal. For periodic signals with a finite time period T, the spectrum only exists for radial frequencies $k\omega = k2\pi/T$; such a spectrum is called a *line spectrum*. The representation of the quantity r_k, in the form of $|r_k|$ and $\arg\{r_k\}$ results in a *line amplitude spectrum* and a *line phase shift spectrum*.

Finally, one other important property of the Fourier series should be mentioned, namely the *Theorem of Parceval*. This theorem can be seen as a direct conclusion of the combination of Eqs (6) through (13):

$$\frac{1}{T}\int_{t_0}^{t_0+T} x^2(t)\,\mathrm{d}t = a_0^2 + \sum_{k=1}^{\infty} \tfrac{1}{2}(a_k^2 + b_k^2) =$$

$$= \sum_{k=-\infty}^{\infty} |r_k|^2 = A_0^2 + \sum_{k=1}^{\infty} \tfrac{1}{2}A_k^2. \tag{14}$$

The theorem shows that the mean squared value of $x(t)$ at the interval observed is equal to the summation of the squared Fourier coefficients, hence the theorem symbolizes a power balance.

1.3.3. The Fourier and Laplace transform

Based on the complex series, Eqs (9) and (10), valid for a finite approximation interval $[t_0, t_0 + T]$, we now can extend the interval to infinity by choosing $t_0 = -\frac{1}{2}T$, and consequently by taking $T \to \infty$. It then follows that:

$$x(t) = \int_{-\infty}^{\infty} X(v)\, e^{j2\pi vt}\, dv, \tag{15}$$

$$X(v) = \int_{-\infty}^{\infty} x(t)\, e^{-j2\pi vt}\, dt, \tag{16}$$

where the quantity $X(v)$ is called the *Fourier transform* of $x(t)$. As explained for the complex Fourier series, we can represent the Fourier transform by the quantities $|X(v)|$ and $\arg\{X(v)\}$ as a function of the frequency v. The figure obtained in this way is called a *continuous amplitude density spectrum* since:

- With the transition of $T \to \infty$ the spectral lines of magnitude c_k spaced at distances of $2\pi/T$ will become infinitesimally close to each other.
- The dimension of $X(v)$ will no longer be that of an amplitude, but that of an amplitude density, since $X(v) \triangleq c_k T = c_k/v$.

Analogous to the Fourier series, the Fourier transform can be seen as the decomposition of a signal into subsignals; however, we now deal with the decomposition of non-periodic signals into sinusoidal subsignals.

Many methods in signal theory and system theory make use of these transforms. By means of a *one-to-one relation* we transform the *original function* into the *transform image;* we say that we transform from one *domain* to another. In this case we call the domain of $x(t)$ the *time domain*, and that of the transform $X(v)$ the *frequency domain.* Due to the one-to-one relation between original function and transform image we also can derive from the transform $X(v)$ the original function $x(t)$. This operation is called the *inverse transform.* Thus we obtain for the Fourier transform:

$$F\{x(t)\} = X(v) = \int_{-\infty}^{+\infty} x(t)\, e^{-j2\pi vt}\, dt, \tag{16}$$

and for the inverse Fourier transform:

$$F^{-1}\{X(\nu)\} = x(t) = \int_{-\infty}^{+\infty} X(\nu)\, e^{j2\pi\nu t}\, d\nu \tag{15}$$

The great importance of these transformation techniques can be illustrated best by noting the following points:

- The transform or decomposition of the original function into subsignals often provides better insight into the original function.
- Particularly complicated and difficult calculations such as the solution of integral equations can be performed much more easily in the newly obtained domain. In spite of the fact that we now have to add a transformation and an inverse transformation to our calculations, we often gain significantly in time in following this procedure (fig. 7).

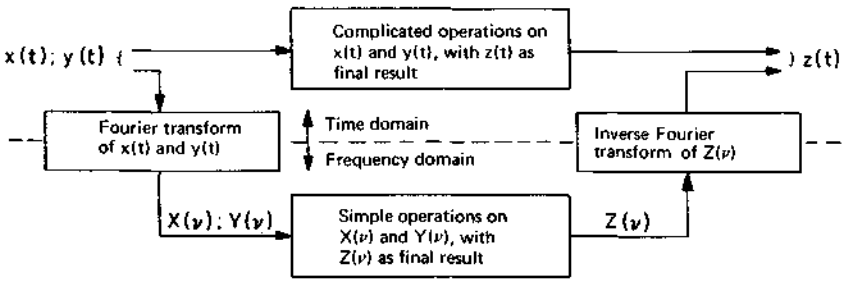

Fig. 7. Application of the Fourier transform in order to simplify particular mathematical operations.

It is important to specify under what conditions we are allowed to use the Fourier transform, or, to say it in another way: For what kind of original functions does the Fourier transform exist, so that the Eqs (15) and (16) will converge at all times? It can be proven that a sufficient, but not a necessary, condition for the existence of a Fourier transform of the original function $x(t)$ is, that this function is absolutely integrable, thus,

$$\int_{-\infty}^{+\infty} |x(t)|\, dt < \infty. \tag{17}$$

The class of functions fulfilling this property is thus restricted to the *normal functions*; this implies that a large number of very important functions such as $x(t) = e^{at}$ with $a > 0$ will not have a Fourier transform. This problem can be solved by calculating the Fourier transform of the function $e^{-\lambda t}x(t)$, with the constant λ being a real constant. In the case that the function $x(t)$ does not converge for $t \to \infty$, but very strongly converges for $t \to -\infty$, we now can make an intelligent choice for the factor $e^{-\lambda t}$, so that the integral is forced to converge for $t \to \infty$ whereas it still converges for $t \to -\infty$. By introducing the complex quantity s, called the *complex frequency*, $s = \lambda + j2\pi\nu = \lambda + j\omega$, we obtain from Eqs (16) and (15):

$$L_{\Pi}\{x(t)\} = X_{\Pi}(s) = \int_{-\infty}^{\infty} x(t)\, e^{-\lambda t} e^{-j2\pi\nu t}\, dt =$$

$$= \int_{-\infty}^{\infty} x(t)\, e^{-st}\, dt, \tag{18}$$

$$L_{\Pi}^{-1}\{X_{\Pi}(s)\} = x(t) = e^{\lambda t} \int_{-\infty}^{\infty} X_{\Pi}(s)\, e^{j2\pi\nu t}\, d\nu =$$

$$= \frac{1}{2\pi j} \int_{\lambda - j\infty}^{\lambda + j\infty} X_{\Pi}(s)\, e^{st}\, ds. \tag{19}$$

This newly defined transform is called the *two-sided Laplace transform*. The factor $e^{-\lambda t}$ forcing the integral to converge is only effective at one of the intervals $t > 0$ or $t < 0$, whereas at the other interval the effect is just the opposite; it will make the integral diverge. Thus, based upon the original functions which converge after multiplication by the factor $e^{-\lambda t}$ with $\lambda > 0$, the two-sided Laplace transform will exist. Unfortunately, only very few time-functions will meet this property. However, a very important class of functions, namely the functions which are equal to zero for $t < 0$ will do so. For those functions the factor $e^{-\lambda t}$ will force the integral at the interval $t > 0$ to convergence, whereas the factor $e^{-\lambda t}$ does not affect the integral in the interval $t < 0$. In this way we obtain the *one-sided Laplace transform:*

$$L\{x(t)\} = X(s) = \int_0^\infty x(t)\,\mathrm{e}^{-st}\,\mathrm{d}t, \tag{20}$$

$$L^{-1}\{X(s)\} = x(t) = \frac{1}{2\pi j}\int_{\lambda-j\infty}^{\lambda+j\infty} X(s)\,\mathrm{e}^{st}\,\mathrm{d}s. \tag{21}$$

A sufficient condition for the existence of the one-sided Laplace transform is that the original functions are *normal* and *exponential of the order* q for $t > t_m$. That means $|x(t) > t_m| \leqslant M\,\mathrm{e}^{qt}$, where the quantities M and t_m are arbitrary but finite constants. Analogous to our discussion of the Fourier series and Fourier transform, we can consider the Laplace transform as the decomposition of the original function into subsignals, where the subsignals in this case are sinusoidal signals with increasing amplitudes. This behavior is caused by the factor $\mathrm{e}^{\lambda t}$.

Until now we have only discussed the description of continuous signals; the application of digital computers, however, forces us to consider also the sampled signals. The treatment of the sampled signals does not differ from those of the continuous ones. The direct extension of continous signal theory to the discrete case is embodied in theory of automata. Because of the fact that sampled signals have only been defined at particular instants of time, the Laplace transform does not exist in the way we discussed. A modified version of the Laplace transform called the *Z-transform* was developed specifically to handle the sampled signals.

Within the scope of this handbook we believe it is better not to devote too much space to all properties and calculation techniques for determining Fourier and Laplace transforms. We therefore only summarize the most important properties in table 1. Because of its great importance in system and signal theory two properties will be elucidated. We will start with the *convolution integral:*

$$L\left\{\int_0^t x_1(\tau)x_2(t-\tau)\,\mathrm{d}\tau\right\} = X_1(s)X_2(s),$$

$$L\{x_1(t)x_2(t)\} = \frac{1}{2\pi j}\int_{c-j\infty}^{c+j\infty} X_1(\sigma)X_2(s-\sigma)\,\mathrm{d}\sigma. \tag{22}$$

Table 1
Transform properties.

Property	*Fourier transform*	*One-sided Laplace transform*
transform	$F\{x(t)\}=X(\nu)=\int_{-\infty}^{\infty} x(t)e^{-j2\pi\nu t}dt$	$L\{x(t)\}=X(s)=\int_{0}^{\infty} x(t)e^{-st}dt$
inverse of the transform	$F^{-1}\{X(\nu)\}=x(t)=\int_{-\infty}^{\infty} X(\nu)e^{j2\pi\nu t}d\nu$	$L^{-1}\{X(s)\}=x(t)=\frac{1}{2\pi j}\int_{\lambda-j\infty}^{\lambda+j\infty} X(s)e^{st}ds$
linearity property	$F\{c_1x_1(t)+c_2x_2(t)\}=c_1X_1(\nu)+c_2X_2(\nu)$	$L\{c_1x_1(t)+c_2x_2(t)\}=c_1X_1(s)+c_2X_2(s)$
delay property	$F\{x(t-a)\}=e^{-j2\pi a\nu}X(\nu)$ $F^{-1}\{X(\nu+a)\}=e^{-j2\pi at}x(t)$	$L\{x(t-a)\,U(t-a)\}=e^{-as}X(s)$ $L^{-1}\{X(s+a)\}=e^{-at}x(t)$
scaling property	$F\{x(at)\}=\frac{1}{\lvert a\rvert}X(\frac{\nu}{a})$ $F^{-1}\{X(a\nu)\}=\frac{1}{\lvert a\rvert}x(\frac{t}{a})$	$L\{x(at)\}=\frac{1}{a}X(\frac{s}{a})$, with $a>0$ $L^{-1}\{X(as)\}=\frac{1}{a}x(\frac{t}{a})$, with $a>0$
transform of a derivative	$F\{\dot{x}(t)\}=j2\pi\nu X(\nu)$ $F\{x^{(n)}(t)\}=(j2\pi\nu)^nX(\nu)$	$L\{\dot{x}(t)\}=sX(s)-x(0)$ $L\{x^{(n)}(t)\}=s^nX(s)-s^{n-1}x(0)-\ldots-x^{(n-1)}(0)$
transform of an integral	$F\{\int_{-\infty}^{t} x(\tau)d\tau\}=\frac{1}{j2\pi\nu}X(\nu)+\frac{1}{2}X(0)\delta(\nu)$	$L\{\int_{0}^{t} x(\tau)d\tau\}=\frac{1}{s}X(s)$

multiplication with t^n	$F\{t^n x(t)\} = \left(\frac{j}{2\pi}\right)^n \frac{d^n X(\nu)}{d\nu^n}$	$L\{t^n x(t)\} = (-1)^n \frac{d^n X(s)}{ds^n}$
division by t	$F\{\frac{x(t)}{t}\} = j2\pi \int_{\nu}^{\infty} X(\xi)\,d\xi$	$L\{\frac{x(t)}{t}\} = \int_{s}^{\infty} X(\sigma)\,d\sigma$
transform of a periodic function with time period T: $x(t)=x(t+T)$	$F\{x(t)\} = \sum_{k=-\infty}^{\infty} \{\frac{1}{T}\int_0^T x(t)e^{-j2\pi kt/T}dt\}\delta(\nu - \frac{k}{T})$	$L\{x(t)\} = \frac{\int_0^T x(t)\,e^{-st}\,dt}{1-e^{-sT}}$
convolution property	$F\{\int_{-\infty}^{\infty} x_1(\tau)x_2(t-\tau)d\tau\} = X_1(\nu)X_2(\nu)$	$L\{\int_0^t x_1(\tau)x_2(t-\tau)d\tau\} = X_1(s)X_2(s)$
	$F\{x_1(t)x_2(t)\} = \int_{-\infty}^{\infty} X_1(\xi)X_2(\nu-\xi)d\xi$	$L\{x_1(t)x_2(t)\} = \frac{1}{2\pi j}\int_{c-j\infty}^{c+j\infty} X_1(\sigma)X_2(s-\sigma)d\sigma;$ $Re(s)>\max(\lambda_1,\lambda_2,\lambda_1+\lambda_2);\ \lambda_1<c<Re(s)-\lambda_2$
boundary properties	$\lim_{\nu\to\infty} X(\nu) = 0$	$\lim_{s\to\infty} X(s) = 0$
		$\lim_{s\to\infty} sX(s) = \lim_{t\downarrow 0} x(t)$
		$\lim_{s\to 0} sX(s) = \lim_{t\to\infty} x(t)$

Eq. (22) illustrates the importance of the transform techniques: A convolution integral in one domain is, after transformation, a simple multiplication in the other. The second important example is the transform of the derivative of a function:

$$L\{x^{(n)}(t)\} = s^n X(s) - s^{n-1}x(0) - \ldots - x^{(n-1)}(0). \tag{23}$$

This equation is the basis for the solution of differential equations, and thus for the description of systems, since a differential equation in time domain becomes an algebraic equation in the frequency domain. The solution of the differential equation with the help of the Laplace transform will lead to the complete solution, including both the homogeneous and the forced solution (integration interval $[0, \infty]$). The solution obtained by the Fourier transform only will provide us the forced solution (integration interval $[-\infty, \infty]$). In general we can conclude that the Laplace transform is quite applicable to solving differential equations, as well as to solving ordinary linear differential equations either with constant or time-varying coefficients, and to linear partial differential equations. The Fourier transform is very suitable for the decomposition of functions into subfunctions. By using the properties given in table 1 and/or any book on transform techniques which gives pairs of time functions and their transforms, we are able to transform almost any ordinary function. Table 2, finally, summarizes those original time functions and their transforms most commonly used in system and signal theory.

1.3.4. Description of stochastic signals

In section 1.3.1 we indicated that stochastic signals are not explicitly defined as functions of time; the description of such signals can only be achieved in terms of probabilities. We will use for this the probability density function, well-known in probability theory. Given a random variable $\bar{x}(\zeta)$ based on an ensemble ζ, we define the probability density function as:

$$f_{\bar{x}}(x) = \lim_{\Delta x \to 0} \Pr\{x < \bar{x}(\zeta) \leqq x + \Delta x\}/\Delta x. \tag{24}$$

The quantity $f_{\bar{x}}(x)\Delta x$ indicates the probability that the random variable $\bar{x}(\zeta)$ has a value between x and $x + \Delta x$. If the random variable $\bar{x}(\zeta)$

Table 2
Original functions with corresponding transforms.

Original function	*Fourier transform* $\chi(\nu)$	*One-sided Laplace transform* $\chi(s)$
* $\delta(t-t_0)$	$e^{-j2\pi\nu t_0}$	e^{-st_0}
$\delta(t)$	1	1
** $U(t-t_0)$	$\frac{1}{2}\{\frac{1}{j\pi\nu}e^{-j2\pi\nu t_0}+\delta(\nu)\}$	e^{-st_0}/s
$U(t)$	$\frac{1}{2}\{\frac{1}{j\pi\nu}+\delta(\nu)\}$	$1/s$
1	$\delta(\nu)$	
*** $t\,U(t)$	$\frac{1}{2}j\delta'(\nu)-\frac{1}{(2\pi\nu)^2}$	$1/s^2$
t	$j\delta'(\nu)$	
$t^a\,U(t)$	$j^n\delta^{(a)}(\nu)+\frac{\Gamma(a+1)}{(j2\pi\nu)^{a+1}}$ a integer >0	$\Gamma(a+1)/s^{a+1}$
t^a	$j^n\delta^{(a)}(\nu)$ a integer >0	
$e^{-bt}U(t)$	$1/(j2\pi\nu+b)$	$1/(s+b)$
$te^{-bt}U(t)$	$1/(j2\pi\nu+b)^2$	$1/(s+b)^2$
$t^a e^{-bt}U(t)$	$\Gamma(a+1)/(j2\pi\nu+b)^{a+1}$	$\Gamma(a+1)/(s+b)^{a+1}$
$e^{j2\pi\nu_0 t}$	$\delta(\nu-\nu_0)$	
$\sin at$	$\frac{1}{2}j\{\delta(\nu+\frac{a}{2\pi})-\delta(\nu-\frac{a}{2\pi})\}$	$a/(s^2+a^2)$
$\cos at$	$\frac{1}{2}\{\delta(\nu+\frac{a}{2\pi})+\delta(\nu-\frac{a}{2\pi})\}$	$s/(s^2+a^2)$
sign t	$\frac{1}{j\pi\nu}$	

* $\delta(t)$ = impulse function.
** $U(t)$ = unit step function.
*** $tU(t)$ = ramp signal.

also is a function of time, hence $\bar{x}(t;\zeta)$ is a function of time t and ensemble ζ, a set of time functions arises which is called a *stochastic process*. The realization $\bar{x}(t;\zeta_0)$ of $\bar{x}(t;\zeta)$ belonging to the element ζ_0 of the ensemble ζ is thus a time function; the stochastic process at a particular moment is a random variable $\bar{x}(t_0,\zeta)$. Probability theory is applied to random variables, and thus can be applied also to a stochastic process at time t_0. In this case we obtain a *probability density function* as a function of time t, hence:

$$f_{\bar{x}}(x,t) = \lim_{\Delta x \to 0} \Pr\{x < \bar{x}(t;\zeta) \leqq x + \Delta x\}/\Delta x. \tag{25}$$

The stochastic process is said to be *stationary* if the probability density function obtained for the ensemble does not depend on time, so that

$$f_{\bar{x}}(x,t) = f_{\bar{x}}(x,t+\Delta t) = f_{\bar{x}}(x) \text{ for all } \Delta t. \tag{26}$$

From these probability density functions we can derive a number of important statistical properties called *moments*, such as the *first order moment* or *mean value* or *mathematical expectation* $\eta_{\bar{x}}$:

$$\eta_{\bar{x}} = E\{\bar{x}(t;\zeta)\} = \int_{-\infty}^{\infty} x f_{\bar{x}}(x)\,\mathrm{d}x \tag{27}$$

and the *second order moment*, such as the *mean square error* or *variance* $\sigma_{\bar{x}}^2$:

$$\sigma_{\bar{x}}^2 = E\{[\bar{x}(t;\zeta) - \eta_{\bar{x}}]^2\} = \int_{-\infty}^{\infty} [x - \eta_{\bar{x}}]^2 f_{\bar{x}}(x)\,\mathrm{d}x. \tag{28}$$

The Eqs (27) and (28) give only static information of the stochastic process; dynamic information such as frequency distribution and time dependence can be derived from the *joint probability density function* $f_{\bar{x}\bar{x}}(x_1, x_2; \tau)$ of the stationary stochastic process at two different instants of time t_0 and $t_0 + \tau$; thus from $\bar{x}(t_0;\zeta)$ and $\bar{x}(t_0+\tau;\zeta)$:

$$f_{\bar{x}\bar{x}}(x_1,x_2,\tau) = \lim_{\substack{\Delta x_1 \to 0 \\ \Delta x_2 \to 0}} \Pr\{x_1 < \bar{x}(t;\zeta) \leqq x_1 + \Delta x_1,$$
$$x_2 < \bar{x}(t+\tau;\zeta) \leqq x_2 + \Delta x_2\}/\Delta x_1 \Delta x_2. \tag{29}$$

In the same way as in probability theory the *correlation coefficient* between two random variables has been defined, we can define such a coefficient between two stationary stochastic processes at a time difference τ. Such a statistical quantity will provide us with information about the interdependence of the two processes as a function of time difference τ. We can define three different second-order moments, being: the *average product function* $R_{\bar{x}\bar{x}}(\tau)$:

$$R_{\bar{x}\bar{x}}(\tau) = E\{\bar{x}(t;\zeta)\bar{x}(t+\tau;\zeta)\} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_1 x_2 f_{\bar{x}\bar{x}}(x_1, x_2, \tau)\,dx_1\,dx_2, \tag{30}$$

the *covariance function* $C_{\bar{x}\bar{x}}(\tau)$:

$$C_{\bar{x}\bar{x}}(\tau) = E\{[\bar{x}(t;\zeta) - \eta_{\bar{x}}][\bar{x}(t+\tau;\zeta) - \eta_{\bar{x}}]\} =$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} [x_1 - \eta_{\bar{x}}][x_2 - \eta_{\bar{x}}] f_{\bar{x}\bar{x}}(x_1, x_2; \tau)\,dx_1\,dx_2, \tag{31}$$

and finally the *correlation function* $K_{\bar{x}\bar{x}}(\tau)$:

$$K_{\bar{x}\bar{x}}(\tau) = C_{\bar{x}\bar{x}}(\tau)/\sigma_{\bar{x}}^2. \tag{32}$$

The characterization of the stochastic process by Eqs (27) through (32) is based on an averaging across the ensemble. In system and signal theory, however, we deal with time functions or signals. We therefore have to build a bridge between the statistical quantities obtained for the ensemble and a description in the time domain. This is achieved by estimation theory, by defining the following *estimators* of the stationary stochastic process. These estimators are time-averages of realizations of the stochastic process.

$$\hat{\eta}_{\bar{x}}(\zeta) = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} \bar{x}(t;\zeta)\,dt, \tag{33}$$

$$\hat{\sigma}_{\bar{x}}^2(\zeta) = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} [\bar{x}(t;\zeta) - \hat{\eta}_{\bar{x}}(\zeta)]^2\,dt, \tag{34}$$

$$\hat{R}_{\bar{x}\bar{x}}(\tau;\zeta) = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} \bar{x}(t;\zeta)\bar{x}(t+\tau;\zeta)\,\mathrm{d}t, \tag{35}$$

$$\hat{C}_{\bar{x}\bar{x}}(\tau;\zeta) = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} [\bar{x}(t;\zeta) - \hat{\eta}_{\bar{x}}(\zeta)][\bar{x}(t+\tau;\zeta) - \hat{\eta}_{\bar{x}}(\zeta)]\,\mathrm{d}t, \tag{36}$$

$$\hat{K}_{\bar{x}\bar{x}}(\tau;\zeta) = \hat{C}_{\bar{x}\bar{x}}(\tau;\zeta)/\hat{\sigma}_{\bar{x}}^2(\zeta). \tag{37}$$

The estimators Eq. (33) through Eq. (37) can be associated with the statistical properties Eq. (27) through Eq. (32), so that we finally obtain a relation between ensemble average and time average (*ergodicity*).

When considering a stationary stochastic process with a mean value $\eta_{\bar{x}}$:

$$E\{\hat{\eta}_{\bar{x}}(\zeta)\} = \eta_{\bar{x}}, \tag{38}$$

and with a variance of the estimator $\hat{\eta}_{\bar{x}}$, Eq. (33), according to Eq. (39):

$$E\{[\hat{\eta}_{\bar{x}}(\zeta) - \eta_{\bar{x}}]^2\} = 0, \tag{39}$$

we call that stochastic process *ergodic with reference to the mean value*. If the estimator Eq. (33) satisfies the Eq. (38) the estimator is said to be *unbiased;* if it also satisfies Eq. (39) the estimator is called *consistent*. The same is true for the estimators Eq. (34) through Eq. (37). It can be proven that all of the estimators considered here are consistent. In the case that all estimators which can be defined for the process under consideration satisfy similar equations as given by Eqs (38) and (39), we speak of a *strictly ergodic stochastic process*. It is important to realize that an ergodic process implies that the process is stationary; the inverse is certainly not true. The ergodic property of a stochastic process is the basis for the description of time functions, since this property associates the description of a signal in the time domain with the calculus of statistical quantities in the ensemble domain.

The decomposition of stochastic signals in subsignals is not meaningful in the present context. On the one hand a transformation of a stochastic process $\bar{x}(t;\zeta)$ is not possible because there does not exist a direct relation between the process and the time t. On the other hand, the

transformation of just one realization $\bar{x}(t;\zeta_i)$ of the stochastic process $\bar{x}(t;\zeta)$ at the interval $[0, T]$ provides only information about that single realization, and not about the stochastic process as a whole. A quantity which certainly provides information about the complete stochastic process is the so-called *power density spectrum* $S_{\bar{x}\bar{x}}(\nu)$ which is defined as the Fourier transform of the average product function $R_{\bar{x}\bar{x}}(\tau)$:

$$S_{\bar{x}\bar{x}}(\nu) = F\{R_{\bar{x}\bar{x}}(\tau)\} = \int_{-\infty}^{+\infty} R_{\bar{x}\bar{x}}(\tau)\, e^{-j2\pi\nu\tau}\, d\tau, \tag{40}$$

$$R_{\bar{x}\bar{x}}(\tau) = F^{-1}\{S_{\bar{x}\bar{x}}(\nu)\} = \int_{-\infty}^{+\infty} S_{\bar{x}\bar{x}}(\nu)\, e^{j2\pi\nu\tau}\, d\nu. \tag{41}$$

The definition of the quantity $S_{\bar{x}\bar{x}}(\nu)$ can be explained as follows. According to Eq. (42) the average product function $R_{\bar{x}\bar{x}}(\tau)$ of the stochastic process $\bar{x}(t;\zeta)$ can be rewritten as:

$$R_{\bar{x}\bar{x}}(\tau) = E\{\hat{R}_{\bar{x}\bar{x}}(\tau)\} = E\left\{\lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} \bar{x}(t;\zeta)\bar{x}(t+\tau;\zeta)\, d\tau\right\}. \tag{42}$$

From the definition of the Fourier transform and Eq. (40) we can derive that:

$$S_{\bar{x}\bar{x}}(\nu) = F\{R_{\bar{x}\bar{x}}(\tau)\} = E\left\{\lim_{T\to\infty} \frac{1}{2T} |\bar{X}(\nu;\zeta)|^2\right\}, \tag{43}$$

in which the quantity $X(\nu;\zeta)$, the Fourier transform of a particular realization of the stochastic process $\bar{x}(t;\zeta)$, means that:

$$\bar{X}(\nu;\zeta) = \int_{-\infty}^{\infty} \bar{x}(t;\zeta)\, e^{-j2\pi\nu t}\, dt =$$

$$= \lim_{T\to\infty} \int_{-T}^{T} \bar{x}(t;\zeta)\, e^{-j2\pi\nu t}\, dt. \tag{44}$$

In other words, the quantity $\bar{X}(\nu;\zeta)$ is the amplitude density spectrum of the realization $\bar{x}(t;\zeta)$, whereas $|\bar{X}(\nu;\zeta)|^2$ represents the power density as a function of the frequency ν. The quantity $S_{\bar{x}\bar{x}}(\nu)$ then is an ensemble average of the time average power density $\lim_{T\to\infty}(1/2T)|\bar{X}(\nu;\zeta)|^2$. Thus the power density spectrum $S_{\bar{x}\bar{x}}(\nu)$ is a statistical property in the frequency domain of the stochastic process $\bar{x}(t;\zeta)$; it indicates in what way the power of the process is distributed over the frequency span.

Of course we should also define an estimator for the power density function. It seems logical to base such an estimator on the average product function, hence:

$$\hat{S}_{\bar{x}\bar{x}}(\nu;\zeta) = \lim_{T\to\infty} \int_{-T}^{T} \hat{R}_{\bar{x}\bar{x}}(\tau;\zeta)\, e^{-j2\pi\nu\tau}\, d\tau =$$

$$= \lim_{T\to\infty} \frac{1}{2T} \left| \int_{-T}^{T} \bar{x}(t;\zeta)\, e^{-j2\pi\nu t}\, dt \right|^2. \tag{45}$$

It can be shown that this estimator is unbiased, so that $E\{\hat{S}_{\bar{x}\bar{x}}(\nu;\zeta)\} = S_{\bar{x}\bar{x}}(\nu)$, but, the estimator is not consistent. Assuming that the stochastic process is normally distributed, the variance of the estimator is:

$$E\{[\hat{S}_{\bar{x}\bar{x}}(\nu;\zeta) - E\{\hat{S}_{\bar{x}\bar{x}}(\nu;\zeta)\}]^2\} = S^2_{\bar{x}\bar{x}}(\nu) \tag{46}$$

and thus differs from zero. For practical application therefore, we always use the unbiased and consistent estimator $\hat{S}_{\bar{x}\bar{x}}(\nu;\zeta)$ according to:

$$\hat{S}_{\bar{x}\bar{x}}(\nu;\zeta) = \frac{\int_{\nu-\Delta\nu}^{\nu+\Delta\nu} \hat{S}_{\bar{x}\bar{x}}(\alpha;\zeta)\, d\alpha}{2\Delta\nu}, \quad \text{with } \Delta\nu \neq 0 \tag{47}$$

The estimator Eq. (47) is unbiased and consistent due to the averaging of an infinite number of sinusoidal subsignals over a finite frequency interval $2\Delta\nu$.

An example of the description of some signals either in the time domain or the frequency domain is given in fig. 8.

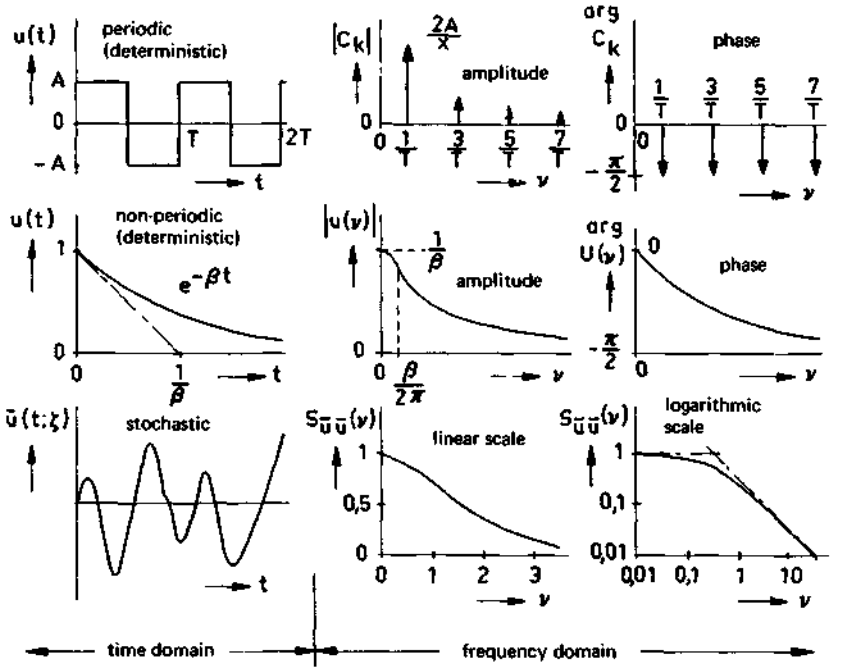

Fig. 8. Examples of deterministic, periodic and non-periodic, and stochastic signals in the time and frequency domains.

1.4. System description

Having discussed the basis of signal theory, we now can proceed with system theory. Therefore we will classify the different systems according to their specific properties.

1.4.1. Classification of systems

Table 3 represents a classification of systems: In this table the properties listed in a row are mutually exclusive, those in a column, however, are not mutually exclusive.

– *Static* versus *dynamic*. A *static system* is a system of which the output $y(t)$ at instant of time t_e is only dependent on the value of the input $u(t)$ at that particular instant, and the instant of time t_e itself. For a *dynamic system* the output $y(t)$ at the moment t_e is a function of the history of $u(t)$ for $t \leqslant t_e$, and also of each t smaller than t_e. Thus:

$$\text{static:} \quad y(t_e) = f\{u(t_e), t_e\} \tag{48}$$

$$\text{dynamic:} \quad y(t_e) = f\{u(t: t_0, t_e), t; y(t_0)\}. \tag{49}$$

The notation $u(t: t_0, t_e)$ means that the signal $u(t)$ is considered over the interval t_0 through t_e. The quantity $y(t_0)$ is called the *initial condition.*

– *Concentrated* versus *distributed:* A *concentrated system* is characterized by variables such as inputs and outputs which are dependent on time as well as spatial coordinates. The input-output relations are described by partial differential or difference equations. A *concentrated system* has only inputs and outputs which are functions of time; it is described by an ordinary differential of difference equation.

– *Constant* versus *time-varying:* A system is described by input-output relations, of which the *structure* as well as the *system parameters* determine the system behavior. In the case that the system parameters are time-dependent it is said that the system is *constant*; if the parameters are time-dependent the system is said to be time-varying. The system response $y(t_e)$ of a time varying system thus becomes:

$$y(t_e) = f[u(t: t_0, t_e), t; y(t_0)]; \tag{50}$$

while that of a constant system is:

$$y(t_e) = f[u(t: t_0, t_e); y(t_0)]. \tag{51}$$

Here it should be noted that an equal distinction can be made with regard to the system structure.

Table 3
Classification of systems.

Classification	Mutually exclusive properties	
Non-mutually exclusive properties	static	dynamic
	concentrated	distributed
	constant	time-varying
	deterministic	stochastic
	linear	non-linear
	continuous	discrete
	scalar	multi-variable

– *Deterministic* versus *stochastic:* A *deterministic system* is a system for which structure and parameters are explicitly defined as a function of time, whereas for a *stochastic system* the structure and parameters can be explained only in terms of probabilities.

– *Linear* versus *non-linear:* Let us assume that the response of a system with initial condition $y_I(t_0)$ and input $u_I(t : t_0, t_e)$ is equal to $y_I(t_e)$, and that the response to initial condition $y_{II}(t_0)$ and input $u_{II}(t : t_0, t_e)$ is equal to $y_{II}(t_e)$. Then the system is said to be linear if:

– The system is *linear with respect to the input:*

$$f[c_I u_I(t : t_0, t_e), t ; y(t_0)] - f[c_{II} u_{II}(t : t_0, t_e), t ; y(t_0)] = \\ = f[c_I u_I(t : t_0, t_e) - c_{II} u_{II}(t : t_0, t_e), t ; 0], \tag{52}$$

that is, the difference between the system response to the inputs $c_I u_I(t : t_0, t_e)$ and $c_{II} u_{II}(t : t_0, t_e)$ with equal initial conditions $y(t_0)$ should be the same as the system response to input $c_I u_I(t : t_0, t_e) - c_{II} u_{II}(t : t_0, t_e)$ with initial condition $y(t_0) = 0$.

– The system is *linear with respect to the initial condition:*

$$f[u(t : t_0, t_e), t ; c_I y_I(t_0)] - f[u(t : t_0, t_e), t ; c_{II} y_{II}(t_0)] = \\ = f[0, t ; c_I y_I(t_0) - c_{II} y_{II}(t_0)], \tag{53}$$

or, the difference between the system response to equal inputs $u(t : t_0, t_e)$, but different initial conditions $c_I y_I(t_0)$ and $c_{II} y_{II}(t_0)$ should be the same as the response on an input $u(t : t_0, t_e) = 0$ and an initial condition $c_I y_I(t_0) - c_{II} y_{II}(t_0)$.

A linear system must satisfy both properties for every value of $y(t_0)$, $u(t : t_0, t_e)$, t_0, t_e and c. If a system does not satisfy either or both of these properties it is called a *non-linear system.*

– *Continuous* versus *discrete: Continuous systems* are systems whose inputs and outputs are continuous, while *discrete systems* have sampled inputs and outputs. Continuous systems are described by differential equations, discrete systems by difference equations.

– *Scalar* versus *multi-variable.* A *scalar system* is a single-input, single-output system. A system with more than one input and/or output is said to be multivariable (fig. 9). For convenience vectorial variables are often introduced. The description of a multi-variable system with r inputs and m outputs,

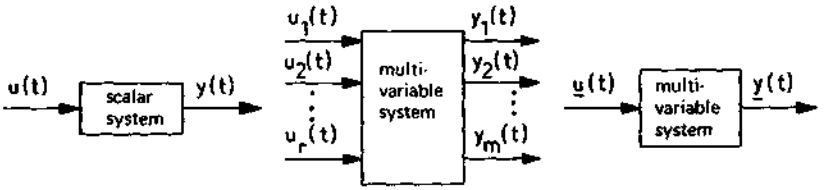

Fig. 9. Block diagram of a scalar and multi-variable system.

$$\begin{aligned} y_1(t_e) &= f_1[u_1(t : t_0, t_e), \ldots, u_r(t : t_0, t_e), t; y_1(t_0) \ldots y_m(t_0)], \\ &\vdots \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \vdots \\ y_m(t_e) &= f_m[u_1(t : t_0, t_e), \ldots, u_r(t : t_0, t_e), t; y_1(t_0) \ldots y_m(t_0)], \end{aligned} \tag{54}$$

can now be easily formulated as follows:

$$\underline{y}(t_e) = \underline{f}[\underline{u}(t : t_0, t_e), t; \underline{y}(t_0)]. \tag{55}$$

This formulation is the basis for the *state vector description.*

A particular class of dynamic systems is that of constant discrete systems. These systems are commonly called *automata*, and they are related to the *grammars* discussed in the second part of this chapter.

The classification into linear and non-linear systems is by far the most important one. The description of linear systems can be achieved in a closed analytical way. In general this is not possible for non-linear systems and it will differ from system to system. We will therefore discuss linear systems in more detail.

1.4.2. Linear systems

The two linearity properties, Eqs (52) and (53), together form the *superposition principle*, a most important principle. The superposition principle teaches us that the response of a linear combination of subsignals is equal to the corresponding combination of the responses to these subsignals (fig. 7). This property is true for the inputs as well as for the initial conditions. As a consequence the principle shows us that a linear system can be characterized by just one function: the impulse response $h(t; t_0)$ to the impulse function $\delta(t - t_0)$.

1.4.2.1. Description by means of impulse response. The system response can principally be determined by forcing a system with an impulse. In practice this is actually not possible, since a pure impulse does actually

not exist and can only be approximated. However, in theory this is a very satisfactory way of describing a linear system in the time domain (fig. 10), and therefore it is often used. From the definition of causal systems it follows directly that the impulse response $h(t; t_0) = 0$ for $t < t_0$. If in addition the system is constant we obtain $h(t; t_0) = h(t - t_0)$.

Fig. 10. The impulse response.

An important property, derived from the definition of the impulse function, is

$$u(t) = \int_{-\infty}^{\infty} u(\tau)\delta(t - \tau)\,\mathrm{d}\tau, \tag{56}$$

which says that any input to a system can be considered to be composed of the summation of impulse functions each having an area $u(\tau)\,\mathrm{d}\tau$. This means that according to the superposition principle the output of a linear system can be written as:

$$y(t) = \int_{-\infty}^{\infty} u(\tau)h(t - \tau)\,\mathrm{d}\tau. \tag{57}$$

Thus for causal systems we obtain:

$$y(t) = \int_{-\infty}^{t} u(\tau)h(t - \tau)\,\mathrm{d}\tau = \int_{0}^{\infty} u(t - \tau)h(\tau)\,\mathrm{d}\tau = u(t)*h(t). \tag{58}$$

Equations such as Eq. (58) are called *convolution integrals*. According to Eq. (22) we can transform this equation into a product of the Laplace

transforms $U(s)$ and $H(s)$. Based on the Eqs (57) and (58), for any linear system the system response to an arbitrary but deterministic input can be achieved, since by measuring the output $y(t)$ of an input $u(t)$ we are able to calculate with Eqs (57) and (58) the impulse response. In the case that the inputs are stochastic we cannot simply apply the Eqs (57) and (58) since the input $\bar{u}(t;\zeta)$ is not explicitly known as a function of time. However, of course, Eq. (57) is also valid for a constant system with a stochastic input $\bar{u}(t;\zeta)$; hence:

$$\bar{y}(t;\zeta) = \int_{-\infty}^{\infty} \bar{u}(t-\theta;\zeta)h(\theta)\,\mathrm{d}\theta. \tag{59}$$

By multiplying both the members of Eq. (59) with the term $\bar{u}(t-\tau;\zeta)$, and by taking the mathematical expectation, using the definition of the average product function, we obtain:

$$R_{\bar{u}\bar{y}}(\tau) = \int_{-\infty}^{\infty} R_{\bar{u}\bar{u}}(\tau-\theta)h(\theta)\,\mathrm{d}\theta = R_{\bar{u}\bar{u}}(\tau)*h(\tau), \tag{60}$$

Here, we call the function $R_{\bar{u}\bar{y}}(\tau)$ the *average cross product function:*

$$R_{\bar{u}\bar{y}}(\tau) = E\{\bar{u}(t;\zeta)\bar{y}(t+\tau;\zeta)\} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} uy f_{\bar{u}\bar{y}}(u,y;\tau)\,\mathrm{d}u\,\mathrm{d}y. \tag{61}$$

The function $R_{\bar{u}\bar{y}}(\tau)$ represents the relation between the stochastic processes $\bar{u}(t;\zeta)$ and $\bar{y}(t+\tau;\zeta)$ as a function of the time difference τ. In the same way it can be shown that $R_{\bar{y}\bar{y}}(\tau) = R_{\bar{y}\bar{u}}(\tau)*h(\tau)$, so that:

$$R_{\bar{y}\bar{y}}(\tau) = R_{\bar{u}\bar{u}}(\tau)*h(-\tau)*h(\tau). \tag{62}$$

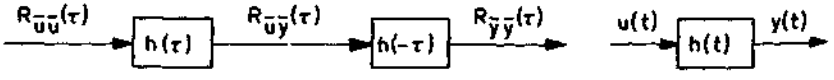

Fig. 11. System description by means of the impulse response for deterministic and stochastic imputs.

1.4.2.2. Description by means of differential equations. The determination of the system response $y(t)$ can also be achieved by solving the differential equation which gives the relation between input and output of the system. In general a linear concentrated system is given by the following ordinary differential equation:

$$a_n \frac{d^n y(t)}{dt^n} + a_{n-1} \frac{d^{n-1} y(t)}{dt^{n-1}} + \ldots + a_1 \frac{dy(t)}{dt} + a_0 y(t) = \\ = f(t) = b_m \frac{d^m u(t)}{dt_m} + \ldots + b_1 \frac{du(t)}{dt} + b_0 u(t). \tag{63}$$

For time-varying systems the coefficients $a_i (i = 0, 1, \ldots, n)$ and $b_j (j = 0, 1, \ldots, m)$ will be functions of time; for constant systems those coefficients will be time-independent. In general physical systems will satisfy the relation $m \leqslant n$. The integer n is said to be the *order of the system.* The solution of the differential equation will provide us the *total response.* If the function $f(t) = 0$, we obtain the *natural response.* If the initial conditions are equal to zero we obtain the *forced solution.* The most general method and, incidentally, the easiest, makes use of the Laplace transform. The calculation can be summarized as follows:

- Assume that $f(t) = 0$, and determine the Laplace transform of Eq. (63). As a result we obtain the Laplace transform $Y(s)$, so that after calculating the inverse Laplace transform we obtain the natural response.
- In the case that $f(t) \neq 0$, and with zero initial conditions, we obtain after transformation of Eq. (63)

$$[a_n s^n + a_{n-1} s^{n-1} + \ldots + a_1 s + a_0] Y(s) = \\ = [b_m s^m + b_{m-1} s^{m-1} + \ldots + b_1 s + b_0] U(s). \tag{64}$$

  From Eq. (64) the response $y(t)$ can be calculated by taking the inverse transformation of Laplace transform $Y(s)$.
- The total response is now achieved by adding the forced response to the natural response.

Often the response is split up in a different way, namely the *transient* and the *steady state solution.* The *transient phenomenon* is that part of the total response that converges to zero as $t \to \infty$; the part that does not equal zero for $t \to \infty$ is said to be the *steady state solution.*

Finally we show the relation between the impulse response and the

solution of the differential equation. According to Eq. (58) a causal linear constant system can be described by:

$$y(t) = \int_0^\infty u(t-\tau)h(\tau)\,\mathrm{d}\tau, \tag{65}$$

where $u(t)$ is the input. After Laplace transformation and using Eq. (22) it follows that:

$$Y(s) = H(s)U(s), \tag{66}$$

where the function $H(s)$ is said to be the Laplace transform of the impulse response $h(t)$. This function is called the *transfer function*. From Eq. (64) we learn with regard to the forced response that if Eq. (64) equals Eq. (66) we obtain:

$$H(s) = \frac{Y(s)}{U(s)} = \frac{b_m s^m + b_{m-1}s^{m-1} + \ldots + b_1 s + b_0}{a_n s^n + a_{n-1}s^{n-1} + \ldots + a_1 s + a_0}, \tag{67}$$

In this equation the polynomial $(a_n s^n + a_{n-1}s^{n-1} + \ldots + a_1 s + a_0)$ is said to be the *characteristic equation*. It can be shown that, indeed, Eq. (67) holds, and thus that the transfer function is:

$$H(s) = L\{h(t)\} = \int_0^\infty h(t)\,\mathrm{e}^{-st}\,\mathrm{d}t. \tag{68}$$

The same equation can be derived for the Fourier transform, be it that the initial conditions are left out:

$$Y(\nu) = H(\nu)U(\nu), \tag{69}$$

with:

$$H(\nu) = F\{h(t)\} = \int_{-\infty}^\infty h(t)\,\mathrm{e}^{-j2\pi\nu t}\,\mathrm{d}t. \tag{70}$$

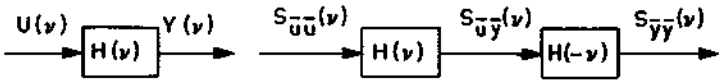

Fig. 12. Description of linear constant causal systems by means of transfer functions.

1.4.2.3. Description by means of the transfer function. Eqs (66) and (69) describe a linear constant causal system with deterministic inputs (fig. 12); this is a description in the frequency domain. The description of a system with stochastic inputs follows directly by applying the Fourier transformation to Eqs (60) and (62):

$$S_{\bar{u}\bar{y}}(\nu) = H(\nu)S_{\bar{u}\bar{u}}(\nu), \tag{71}$$

$$S_{\bar{y}\bar{y}}(\nu) = S_{\bar{u}\bar{y}}(\nu)H(-\nu) = S_{\bar{u}\bar{u}}(\nu)H(\nu)H(-\nu) = |H(\nu)|^2 S_{\bar{u}\bar{u}}(\nu). \tag{72}$$

Here it should be mentioned that Eq. (71) gives phase as well as amplitude information of the transfer function $H(\nu)$, whereas Eq. (72) contains just amplitude information. The transfer function $H(\nu)$ may also be considered as a *frequency response*, obtained by the response to a sinusoidal input $u(t) = a\cos 2\pi\nu t$. It can be shown that the ratio between the amplitudes of input and output of the linear constant system is equal to $|H(\nu)|$, whereas the phase difference between input and output equals $\arg H(\nu)$.

1.4.2.4. Stability. The stability of a system is determined by its dynamics. On the basis of the impulse response we can define the stability of a system as follows: A system is *stable* if, and only if, the impulse response tends to zero for $t \to \infty$. Thus from

$$h(t) = L^{-1}\{H(s)\} = L^{-1}\left\{\frac{b_m s^m + b_{m-1}s^{m-1} + \ldots + b_1 s + s_0}{a_n s^n + a_{n-1}s^{n-1} + \ldots + a_1 s + a_0}\right\}, \tag{73}$$

it follows that the roots of the characteristic equation

$$a_n s^n + a_{n-1}s^{n-1} + \ldots + a_1 s + a_0 = a_n(s - s_1)(s - s_2)\ldots(s - s_n) = 0 \tag{74}$$

must have a negative real part: $R_e(s_i) = \lambda_i < 0$ for $i = 1, 2, \ldots, n$.

1.4.2.5. Complex systems. The boundary of a system is arbitrarily chosen; often a combination of many subsystems forms a single new

system. By means of a block diagram the structure of such a new system can clearly be explained. A number of different structures can be recognized.

– *Cascade series of subsystems:* In a cascade series the output of the first subsystem is the input of the second subsystem (fig. 13). The impulse function $h(t)$ of the complex system obtained in this way then becomes:

$$h(t) = h_1(t) * h_2(t), \tag{75}$$

so that for the transfer function $H(s)$ results:

$$H(s) = H_1(s)H_2(s). \tag{76}$$

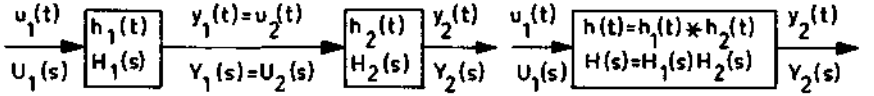

Fig. 13. Cascade series of subsystems.

– *Parallel series of subsystems:* In a parallel series of subsystems the input drives both the subsystems, after which the responses are added (fig. 14). It can be shown that the overall response of the system will be:

$$h(t) = h_1(t) + h_2(t), \tag{77}$$

$$H(s) = H_1(s) + H_2(s). \tag{78}$$

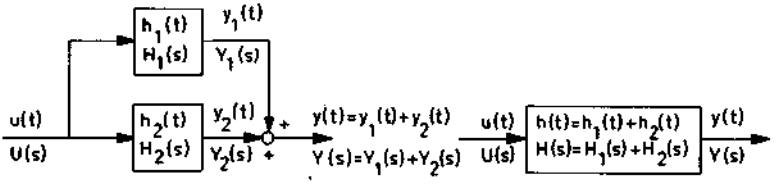

Fig. 14. Parallel series of subsystems.

– *Closed loop systems:* For a closed loop system we feed back the output $y_1(t)$ of the subsystem $H_1(s)$ in the *forward loop* via the *feedback loop* with subsystem $H_2(s)$. We now can derive that (fig. 15):

$$H(s) = \frac{H_1(s)}{1 + \hat{H}(s)}, \quad \text{with } \hat{H}(s) = H_1(s)H_2(s). \tag{79}$$

The transfer function $\hat{H}(s)$ is called the *open loop gain*. After applying the inverse Laplace transformation we obtain:

$$h(t) = L^{-1}\left\{\frac{H(s)}{1 + \hat{H}(s)}\right\}. \tag{80}$$

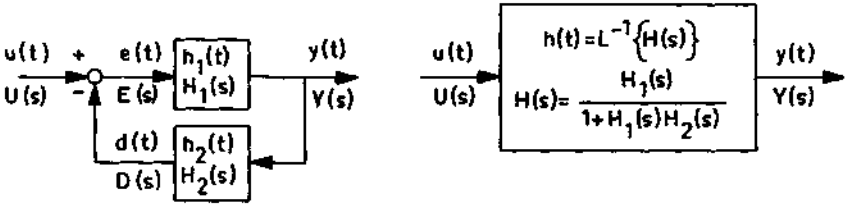

Fig. 15. A closed loop system.

1.4.3. Non-linear systems

As mentioned before, a *non-linear system* is defined as a system that is not linear, that is, it does not satisfy the superposition principle. This definition is a very poor one, however, since it only excludes the possibility that a particular system possesses a certain property, be it an important one; it specifies no other property. The behavior of non-linear systems is described by non-linear algebraic, differential or difference equations. Rarely are we able to find a closed analytical solution; a generally applicable method to solve these non-linear equations may be expected never to be found. In system theory, however, we mainly are interested in the behavior of a system around a certain *operating point*, so that by *linearization around the operating point* a useful description can be derived. In this way we obtain a *linear description of the non-linear system*, valuable around a particular operating point. This approach makes it possible to apply the theory of linear systems to non-linear systems.

A more profound study of non-linear systems will teach us that:

- The behavior, static as well as dynamic, will depend strongly on the input.
- In general the system will generate *higher harmonics* of the forcing input; sometimes even the fundamental frequency will disappear in the output.

1.4.3.1. Linearization of constant static systems. Linearization of a non-linear system around an operating point leads to a linear description. Observe now the constant static non-linear system:

$$y(t) = g\{u(t)\}, \tag{81}$$

where the function $g\{u(t)\}$ represents a non-linear, continuous, and differentiable function. Then, according to the Taylor series around the operating point $g\{\eta_u\}$, we obtain:

$$y(t) = g\{\eta_u\} + \frac{dg}{du}\bigg|_{\eta_u} [u(t) - \eta_u] + \tfrac{1}{2}\frac{d^2g}{du^2}\bigg|_{\eta_u} [u(t) - \eta_u]^2 + \ldots, \tag{82}$$

By neglecting the higher harmonics, it follows that:

$$y(t) = g\{\eta_u\} + \frac{dg}{du}\bigg|_{\eta_u} [u(t) - \eta_u] \tag{83}$$

thus for small variations around $g\{\eta_u\}$ we obtain the relation:

$$k(\eta_u) = \frac{dg}{du}\bigg|_{\eta_u}. \tag{84}$$

The *gain* $k(\eta_u)$ is a function of the operating point $g(\eta_u)$ of the non-linearity, or, with reference to the input it is a function of the mean value η_u of input $u(t)$. Here we should note that this linearization technique can be applied in a significant way only if the variations in $u(t)$ around the mean value η_u are relatively small, such as in closed-loop systems. For open-loop systems these variations may often be too large, so that such a simple linearization technique leads to too great a simplification. In those cases it is worthwhile to apply the *method of the statistical linearization* or the *method of Booton*. This method is based on the following philosophy. Assume that the input $\bar{u}(t;\zeta)$ has a mean value equal to zero, and that we can approximate the linear system by an equivalent gain $k_e(0, \sigma_{\bar{u}})$, that is, a linear constant system. Then, by minimization of the difference between the output $\bar{y}(t;\zeta) = g\{\bar{u}(t;\zeta)\}$ of the non-linear static system and the output $\bar{y}^*(t;\zeta) = k_e\bar{u}(t;\zeta)$ of the equivalent linear system according to a quadratic criterion, we obtain the least-square estimate of the equivalent gain (fig. 16). It can be shown that with:

$$\frac{\partial}{\partial k_e} E\{[g\{\bar{u}(t;\zeta) - k_e\bar{u}(t;\zeta)]^2\} = 0 \tag{85}$$

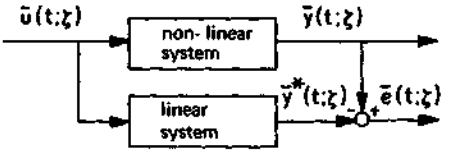

Fig. 16. Statistical linearization of a constant static non-linear system.

we obtain the following relation for $k_e(0, \sigma_{\bar{u}})$:

$$k_e(0, \sigma_{\bar{u}}) = \frac{\int_{-\infty}^{\infty} ug(u) f_{\bar{u}}(u)\, du}{\int_{-\infty}^{\infty} u^2 f_{\bar{u}}(u)\, du} = \frac{1}{\sigma_{\bar{u}}^2} \int_{-\infty}^{\infty} ug(u) f_{\bar{u}}(u)\, du. \tag{86}$$

A number of remarks should be made here.

- Eq. (86) shows that the equivalent gain $k_e(0, \sigma_{\bar{u}})$ is a function of the variance $\sigma_{\bar{u}}^2$; in fact one should know the probability density function $f_{\bar{u}}(u)$ in order to be able to determine $k_e(0, \sigma_{\bar{u}})$. The method discussed here is not restricted to inputs with a mean value equal to zero, although for $\eta_{\bar{u}} \neq 0$ the calculation becomes much more complicated.
- Because of the fact that the equivalent gain is only a constant system, the method is restricted to non-linear constant static systems.
- The methodology followed in obtaining the optimal equivalent gain is similar to that applied to the decomposition of signals into subsignals; it is of interest to compare figs 6 and 16.

1.4.3.2. Describing function method. In order to describe non-linear constant dynamic systems, one often uses the *describing function method.* This method is based on the idea that the non-linear system can be replaced by a linear dynamic system with a transfer function $G(\nu)$. If now the non-linear system is driven by an input $u(t) = A \cos 2\pi\nu t$, we can describe the output by a series of sinusoidal subsignals with the help of the Fourier series, Eq. (6). By comparing the input with the output's fundamental harmonic, the first term of the Fourier series, we can define a transfer function, the *describing function.* Also, this method is a linearization around a certain operating point; a linear relation between the input and the output's higher harmonics, the other terms of the Fourier series, does not exist. The method can be applied to systems with

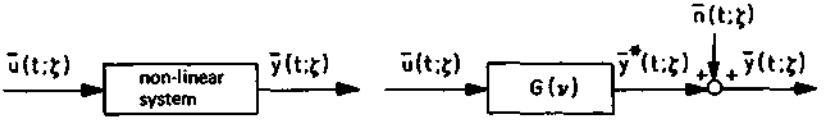

Fig. 17. The describing function: A linear description of a non-linear constant dynamic system.

deterministic as well as stochastic inputs. In the latter case the non-linear system is thought to be approximated by a linear dynamic system to which output a signal $\bar{n}(t;\zeta)$, called the *remnant*, is added (fig. 17). This remnant is thus the difference between the outputs $\bar{y}(t;\zeta)$ and $\bar{y}^*(t;\zeta)$ of the non-linear and equivalent system, respectively. The optimal dynamic gain $G(\nu)$ again is achieved by minimization of the variance of the remnant with respect to the parameters of the transfer function $G(\nu)$. It can be shown that minimization of this variance $\sigma_{\bar{n}}^2$ yields a result for which the remnant $\bar{n}(t;\zeta)$ is uncorrelated with the input $\bar{u}(t;\zeta)$, and thus $R_{\bar{u}\bar{n}}(\tau) = 0$ for any value of τ. It can be derived that:

$$S_{\bar{u}\bar{y}}(\nu) = S_{\bar{u}\bar{y}^*}(\nu) = G(\nu)S_{\bar{u}\bar{u}}(\nu), \tag{87}$$

$$S_{\bar{y}\bar{y}}(\nu) = |G(\nu)|^2 S_{\bar{u}\bar{u}}(\nu) + S_{\bar{n}\bar{n}}(\nu). \tag{88}$$

Eq. (87) provides us the describing function $G(\nu)$, so that by substituting $|G(\nu)|$ into Eq. (88) the spectral density $S_{\bar{n}\bar{n}}(\nu)$ of the remnant can be obtained.

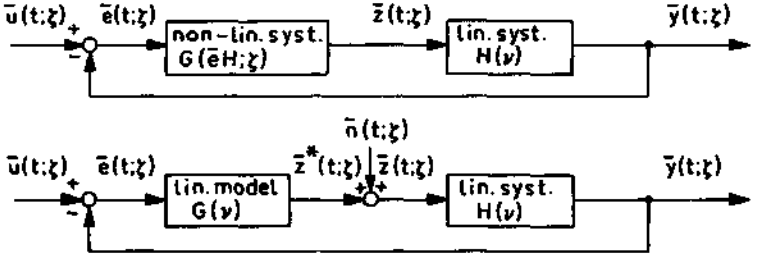

Fig. 18. The describing function method applied in a closed system.

The method as formulated by Eqs (87) and (88) can only be applied if the non-linear element is not linked up in a closed loop system, since for closed loop systems (fig. 18) the remnant $\bar{n}(t;\zeta)$, due to the feedback, is always linearly correlated with input $\bar{e}(t;\zeta)$ to the non-linear element;

thus $R_{\overline{zn}}(\tau) \neq 0$ for all values of τ. The describing function $G(\nu)$ can be determined in two different ways:

– *Indirect method:*

$$S_{\overline{uy}}(\nu) = G_{tot}(\nu) S_{\overline{uu}}(\nu), \quad \text{with } G_{tot}(\nu) = \frac{G(\nu)H(\nu)}{1 + G(\nu)H(\nu)}. \tag{89}$$

– *Direct method:*

$$G(\nu) = S_{\overline{uz}}(\nu)/S_{\overline{uz}}(\nu). \tag{90}$$

The indirect method only can be applied if the linear system $H(\nu)$ is known; this method is certainly the easiest one.

The describing function method is of great importance in the mathematical description of the human operator's behavior in manual control (see vol. II, chapter 10).

1.4.4. Description by means of state variables

The description of systems by means of state variables has been mentioned already in the discussion of scalar and multi-variable systems. This approach has become very popular in the last decade, on the one hand because the description method is conceptually simple for describing very complex systems, and on the other hand because this method can be easily applied when digital computers are used. The basic philosophy is that any dynamic linear system can be described by a set of first-order linear differential equations.

If for a given system the input $u(t)$ is known over the interval $[-\infty, t]$, then we can determine the output $y(t)$. If the input $u(t)$, however, is known over only the finite interval $[t_0, t]$, the output $y(t)$ can only be determined if we have knowledge of the initial condition $y(t_0)$ (except for constant static systems). Hence, we can define the state of a system as follows: *The state of a system at a particular time instant t_0 is the set of numbers which, together with the input signal over the interval $[t_0, t]$, determines the output of a system at the time instant t.* From this definition we obtain the following relation, called the *output equation of the system:*

$$\underline{y}(t) = \underline{f}\{\underline{x}(t_0), \underline{u}(t : t_0, t)\}; t > t_0. \tag{91}$$

In this equation the quantity $\underline{y}(t)$ represents the *output vector*, the quantity

$\underline{u}(t:t_0, t)$ the *input vector* over the interval $[t_0, t]$, and $\underline{x}(t_0)$ the *state vector*, defined at the instant of time t_0. From the definition it is clear that the state vector also could be defined at the time instant t_1, in which case it easily can be shown that for $t_1 > t_0$ the state $\underline{x}(t_1)$ is completely defined by $\underline{x}(t_0)$ and $\underline{u}(t:t_0, t_1)$. We formulate this by means of the *state equation* (92):

$$\underline{x}(t) = \underline{g}\{\underline{x}(t_0), \underline{u}(t:t_0, t)\};\ t > t_0. \tag{92}$$

Eqs (91) and (92) can be used to describe all of the different classes of systems mentioned before, except for the distributed systems.

In the context of this Handbook it is impossible to discuss the description by means of state variables in a general way; we will restrict ourselves to the *differential systems*, which are those systems where the Eqs (91) and (92) can be rewritten as:

$$\dot{\underline{x}}(t) = \underline{g}\{\underline{x}(t), \underline{u}(t), t\} \tag{93}$$

$$\underline{y}(t) = \underline{f}\{\underline{x}(t), \underline{u}(t), t\}. \tag{94}$$

The *system equations* (93) and (94) are much simpler if we derive them for linear systems, obtaining:

$$\dot{\underline{x}}(t) = A(t)\underline{x}(t) + B(t)\underline{u}(t) \tag{95}$$

$$\underline{y}(t) = C(t)\underline{x}(t) + D(t)\underline{u}(t) \tag{96}$$

where $A(t)$ is said to be the *system matrix*, $B(t)$ the *input matrix*, $C(t)$ the *output matrix*, and $D(t)$ the *direct transfer matrix*. For linear constant systems the matrices of Eqs (95) and (96) become independent of time, so that it follows:

$$\dot{\underline{x}}(t) = A\underline{x}(t) + B\underline{u}(t) \tag{97}$$

$$\underline{y}(t) = C\underline{x}(t) + D\underline{u}(t). \tag{98}$$

The meaning of the matrices A, B, C, and D can be elucidated by means of the block diagram of fig. 19:

– Matrix A is placed in the feedback loop, and thus acts on the state variable $\underline{x}(t)$ which is fed back to the integrators. The matrix A entirely determines the dynamic behavior of the system.

– Matrix B determines to what extent and in what way the input

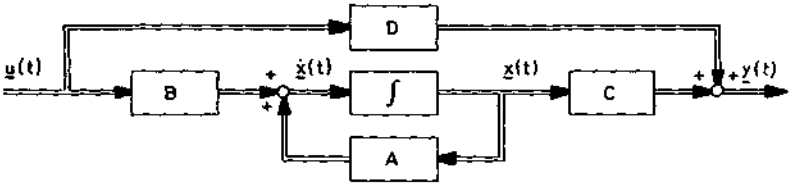

Fig. 19. The state description of a linear constant system.

$\underline{u}(t)$ controls the system. In conjunction with matrix A, matrix B determines whether the system is controllable or not. A *fully controllable system* is a system where from an arbitrary initial state $\underline{x}(t_0)$ any other final state $\underline{x}(t_e)$ can be reached by forcing the system with a certain input $\underline{u}(t : t_0, t_e)$ where the interval $[t_0, t_e]$ must be finite with $t_e > t_0$.

– Contrary to the idea of controllability is the idea of observability. Observability is determined by the matrices A and C. It is said that a system is *fully observable* if from the state vector $\underline{x}(t_0)$ and the output $\underline{y}(t : t_0, t_e)$ over the finite interval $[t_0, t_e]$ the state vector $\underline{x}(t_e)$ can be determined uniquely ($t_0 < t_e$).

– The direct transfer matrix D determines to what extent and in what way the input affects the output without passing through the closed loop.

The ideas of observability and controllability will be used in the subsequent sections on automata and grammars.

Without proof we state that the solution of the set of differential equations (97) and (98) with initial state vector $\underline{x}(t_0)$ will be:

$$\underline{x}(t) = \mathrm{e}^{A(t-t_0)}\underline{x}(t_0) + \int_{t_0}^{t} \mathrm{e}^{A(t-\tau)}B\underline{u}(\tau)\,\mathrm{d}\tau \tag{99}$$

$$\underline{y}(t) = C\,\mathrm{e}^{A(t-t_0)}\underline{x}(t_0) + C\int_{t_0}^{t} \mathrm{e}^{A(t-\tau)}B\underline{u}(\tau)\,\mathrm{d}\tau + D\underline{u}(t), \tag{100}$$

where $\mathrm{e}^{A(t)}$ is called the *transient matrix*. The Eqs (99) and (100) give the response $\underline{y}(t)$ of the system on an input $\underline{u}(t)$ in the time domain.

As was the case for scalar systems, a description in the frequency domain can here be derived again. By applying the Laplace transformation to Eq. (97) we obtain:

$$s\underline{X}(s) - \underline{x}(t_0) = A\underline{X}(s) + B\underline{U}(s), \tag{101}$$

and thus:

$$X(s) = (sI - A)^{-1}\underline{x}(t_0) + (sI - A)^{-1}B\underline{U}(s). \tag{102}$$

The matrix $(sI - A)^{-1}B$ is said to be the *transfer matrix from* $U(s)$ *to* $X(s)$. Transforming Eq. (98) and substituting the result into Eq. (102) it follows:

$$\underline{Y}(s) = C(sI - A)^{-1}\underline{x}(t_0) + [C(sI - A)^{-1}B + D]\underline{U}(s). \tag{103}$$

The matrix $C(sI - A)^{-1}B + D$ is called the *transfer matrix from* $U(s)$ *to* $Y(s)$. This matrix can be compared with the well-known transfer function $H(s)$ as discussed in section 1.4.2.3.

Recent developments in the field of the description of the human operator's behavior in manual as well as supervisory control often are based on description in terms of state variables (see vol. II, chapter 10).

1.5. Models and parameter estimation techniques

A major goal of using system theory is to model real world problems; that is, we try to formulate mathematical models on the basis of which predictions can be made under a variety of circumstances. In the foregoing we almost implicitly assumed that just a little or no information at all was available. In such cases systems can be described by impulse responses, transfer functions and averaged product functions. Often, however, we have some knowledge of the system under study, which makes it possible that we define the *structure* of the system at hand in advance. In that case the problem of identifying a system is reduced to the determination of *a set of parameters* given a certain model structure.

The choice of the structure of the model is of great importance; this choice, among other factors, depends on:

- The objectives to be realized in determining the model.
- The insights and ideas of the investigator.
- The information available about the system under study.
- The required accuracy of the model.
- The observability of the system under study.

Determination of the unknown parameters is mostly achieved according to a standardized pattern (fig. 20). A model of the system under study is proposed; the structure is chosen and the parameters α_i must then be

determined. Given a certain input $\bar{u}(t;\zeta)$, the output $\bar{y}^*(t;\zeta;\alpha_i)$ of the model is compared with the output $\bar{y}(t;\zeta)$ of the system under study. According to a certain criterion we now minimize the error between $\bar{y}(t;\zeta)$ and $\bar{y}^*(t;\zeta;\alpha_i)$ by varying the parameters α_i. It follows that:

$$\frac{\partial}{\partial \alpha_i} E \int_{-\infty}^{\infty} w(t) \left|\bar{y}(t;\zeta) - \bar{y}^*(t;\zeta;\alpha_i)\right|^p \mathrm{d}t = 0, \tag{104}$$

where the quantity $w(t)$ is called a *weighting function*, and where the exponent p mostly is chosen to be two. The minimization of the cost function, Eq. (104), can be achieved analytically or by an iterative procedure. In general, the solution of Eq. (104) will generate a set of equations which are non-linear in the parameters. Only with $w(t) = 1$, and $p = 2$ the set of equations will be linear, and thus solvable in an analytical way. It should be noted that the method, as illustrated in fig. 20, can only be applied to open loop systems. Variants of this method have been developed in order to be able to handle closed loop systems. Finally one very important remark should be made. In general the definition of a model is based on certain assumptions; this means that the application of the model is limited to situations for which the assumptions are valid. We should therefore always check whether the circumstances in which we plan to apply the model satisfy these assumptions.

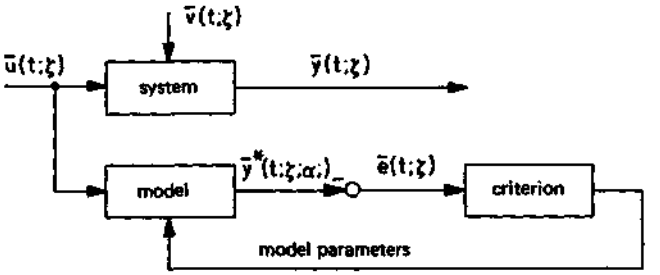

Fig. 20. Parameter estimation of the model parameters in an open loop system.

2. Automata and grammars

Automata are dynamic systems which are discrete and constant. They are dynamic, because their behavior is not only dependent on the last value of the input signal, but also on the series of preceding input signals. They are discrete systems, since they have a discrete time axis $t = 0, 1, 2, \ldots$; they are constant since both system parameters and structure are independent of t. Furthermore, automata have quantized signals: signal values are elements of a finite set.

We shall first discuss some automata of increasing complexity (2.1). The choice is determined by the relations these automata are entertaining to the grammars that are treated in 2.2. All automata to be discussed are observable systems (cf. section 1.4.4), for which it is inconsequential whether the discription proceeds from the state, or from the output signal. As we will proceed from the state, the notion 'output signal' will not be used anymore.

2.1. Some automata

2.1.1. Finite automata

A finite automaton is a system characterized by the following five entities. X is a finite non-empty set of *states*. At any moment the automaton must be in just one of these states. One of these states, x_0, is called the *initial state*, and F, a non-empty subset of X, constitutes the set of *final states*. The possible values of the input signal form a non-empty finite set V; this set is also called the automaton's *vocabulary*. Finally, there is a (*state*) *transition function*, δ, which indicates how the automaton changes state under influence of particular input signals: $\delta(x_i, v_j) = x_k$ means that the automaton in state x_i changes to state x_k at input of vocabulary element v_j. Figure 21a shows the transition-diagram of a finite automaton with two states x_0 and x_1, where V is binary (0 or 1), and with transition function $\delta(x_0, 1) = x_1$, $\delta(x_0, 0) = x_0$, and $\delta(x_1, 0) = x_0$.

In order to explain the workings of this as well as the other automata in this section, we shall use the notion 'controllability', which was introduced in section 1.4.4. There, a system was called 'fully controllable' if there is always an input signal such that the system can make a transition from any initial state $x(t_0)$ to any final state $x(t_e)$. In the theory of automata full controllability (i.e. from *any* initial state) is irrelevant. Important

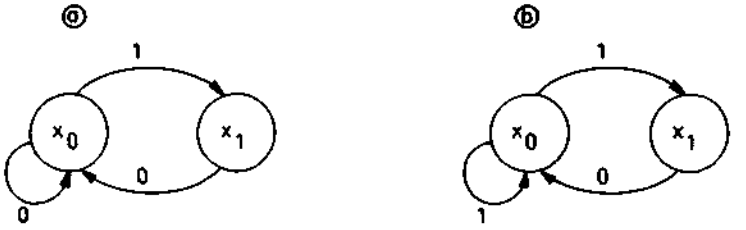

Fig. 21. Transition diagram for deterministic (a) and non-deterministic (b) finite automaton.

is controllability from one defined initial state x_0. This state is called controllable, if there exists a string of input elements s (a signal) which can bring the automaton from there to a final state $x_f \in F$. If such is the case, the automaton is said to *accept* input string s. This can be written as follows $\delta(x_0, s) = x_f$, where $s \in V^*$ (the set of strings of vocabulary elements), and $x_f \in F$. For example, fig. 21a shows that one can proceed from x_0 to x_1 by presenting vocabulary element 1, but also by presenting strings 01, 001, 101, etc. The *language* accepted by automaton A is the set of accepted strings: $T(A) = \{s | \delta(x_0, s) \in F\}$; these are the strings by which the initial state is controllable. The automaton of fig. 21a is controllable with any string consisting of an arbitrary number of 0-elements, followed by an arbitrary number of sequences, 10, followed by 1; in short: $T = \{0^*(10)^*1\}$. Two automata, A_1 and A_2, are *equivalent* if $T(A_1) = T(A_2)$. The languages accepted by finite automata are called *regular languages.*

Apart from deterministic finite automata, as in fig. 21a, there are also non-deterministic ones. The transition function of these automata gives a *set* of possible transitions for any pair of state and vocabulary elements: $\delta(x_i, v_j) = \{x_h, \ldots, x_k\}$. Fig. 21b gives an example. There $\delta(x_0, 1) = \{x_0, x_1\}$. From x_0, the automaton can either go to x_0 or x_1 at the input of 1. Each input string therefore corresponds to a set of paths in the transition diagram. The automaton *accepts* a string if there is at least one path from x_0 to a final state for that string. One can prove that for each non-deterministic finite automaton there is an equivalent deterministic automaton (the inverse is trivially true). It follows that the non-deterministic finite automata generate the same class of regular languages as deterministic finite automata. A *probabilistic finite automaton* is a generalization of the non-deterministic finite automaton, where a probability is assigned to every possible transition. It is a

stochastic system as defined in section 1.4.1. Markov sources are a subclass of probabilistic finite automata.

The input strings in V^* which are not accepted by the automaton form the *complement* CT of the language T. The complement of a regular language is itself regular, i.e. for every finite automaton A, there is another finite automaton A', such that $CT(A) = T(A')$.

2.1.2. Push-down automata

A push-down automaton (*PDA*) is a system which, other than the finite automaton, has an infinite set of states X. In order to describe this set, one may factorize X into two parts: a finite set $T = \{t_0, t_1, \ldots, t_m\}$ of 'states' in a stricter sense, plus a memory store of infinite size. This store can contain strings $(\chi, \psi, \omega, \ldots)$ of so-called memory symbols taken from a finite *push-down vocabulary* $\Gamma = \{\gamma_0, \gamma_1, \ldots, \gamma_n\}$. This push-down store can be conceived of as operating in the following way: The automaton always starts with just γ_0 in the store. New memory symbols are put on top of old ones, 'pushing' them 'down', so that only the top-element is removable: it is not allowed to add or remove elements at other places in the stack.

In factorized form, the initial state x_0 of the *PDA* is the pair (t_0, γ_0), where t_0 is now called the initial state (in the strict sense). $F \in T$ is the set of final states. The automaton is a system $(V, T, \Gamma, t_0, \gamma_0, F, \delta)$, where δ denotes the set of transition rules. These transition rules determine what happens in a given state, and with given top element, when a new input element appears, i.e. the rules determine the next state, and the memory change. They are written as $(t_i, v_j, \gamma_k) = (t_l, \chi)$, which means that at the input of v_j, with γ_k as top element in memory, the state changes from t_i to t_l, and γ_k is replaced by the string (or better: stack) of memory elements χ (this may be the null-string, which amounts to simply removing γ_k). A string s of input elements is accepted by the automaton if a final state $t_f \in F$ is reached from (t_0, γ_0). The language $T(PDA)$, accepted by the push-down automaton is the set of accepted strings: $T(PDA) = \{s | \delta(t_0, s, \gamma_0) = (t_f, \chi), t_f \in F, \chi \in \Gamma^*\}$. The languages accepted by *PDA*'s are called *deterministic languages*. The complement of a deterministic language is also deterministic. Regular languages form a strict subset of deterministic languages.

Analogous to the non-deterministic finite automata, one can design non-deterministic push-down automata (*NPDA*). Such automata can, for each state, top element, and input element, 'choose' from a set of

transitions. This type of automaton is more powerful than the deterministic automaton. The languages accepted by *NPDA's* are called *context-free languages.* They include deterministic languages as a strict subset. The question whether the complement of a context-free language is also context-free has been proven to be unsolvable, but in any case the complement is context-sensitive (see next section).

2.1.3. Linear-bounded automata

The linear-bounded automaton (*LBA*) can be conceived of as a tape with a movable reading and writing head, which can be in any of a number of states (cf. fig. 22). The tape is used to write down the input string, and also serves as memory space. It is characteristic for the *LBA* that this 'working space' on the tape is exactly the same size as the input string: for small inputs there will be little memory space, for large inputs there will be much memory space, i.e. there are no intrinsic bounds on memory size.

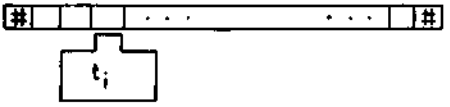

Fig. 22. A linear-bounded automaton.

The *LBA*, which is characterized by the entities $V, T, \Gamma, t_0, F, \delta$, and $\#$, starts reading the tape at the left, i.e. at the first input symbol; the *LBA* is then in state t_0. Depending on what is read, it will change state, and it can replace the element that has been read by another one. This may be an element of V, or an additional memory element (from the finite memory vocabulary Γ). Also, it will select a new position k, i.e. it may go one position to the right ($k = +1$), one position to the left ($k = -1$), or remain unchanged ($k = 0$). Thus, each pair of state and tape symbol will cause three changes: a change of state, a change of tape symbol, and a change of position. For each pair of state and tape symbol the transition rules δ describe what these three changes will be. The *LBA* is said to accept an input string if it reaches the right boundary symbol ($\#$), and then enters a final state ($\in F$). *LBA's* are always non-deterministic: for every combination of state and tape symbol a *set* of transitions is specified. The languages accepted by *LBA's* are called *context-sensitive languages.* It is still unknown whether their complements are also con-

text-sensitive. Context-free languages form a strict subset of context-sensitive languages.

2.1.4. Turing machines

A Turing machine *TM* differs in only one respect from the *LBA*: the tape for reading and writing is of infinite length to the left and to the right. The transition rules, however, are the same as for the *LBA*: for each pair of tape symbol and state they determine the new tape symbol (to be written), the new state, and the new reading position ($k = -1, 0$, or $+1$). Apart from non-deterministic *TM*s there are also deterministic ones. It is also true that each non-deterministic *TM* is equivalent to a deterministic one.

In view of its very simple structure, it is surprising that *TM*'s can do any operation a modern digital computer can do. The inverse is even not the case, except if one assumes that the computer firm can make unlimited amounts of additional memory space available. A Turing machine can perform any explicit symbol operation, it seems. In fact, this is even possible on a *TM* with no more than two states, t_0 and t_1. Therefore, nowadays, the notion of ('effective' or 'mechanical') *procedure* is defined as 'capable of being carried out by a Turing machine'. Lack of space forbids further discussion of this important notion of procedure, see however Minsky (1967).

A *TM* is said to *accept* an input string s if this string brings the automaton from initial state t_0 to a final state $t_f \in F$. The languages accepted by Turing machines are called *recursively enumerable languages*. This name indicates that the strings or 'sentences' of such a language can be 'enumerated', i.e. there is a procedure by which the sentences of language T (and no other strings) are successively generated in such a way that each sentence of T will be enumerated after a finite number of elementary operations. (The complete enumeration of an infinite language will nevertheless take an infinite number of operations!). Assume one has a *TM*, with language $T(TM)$, and an arbitrary string $s(\in V^*)$. If $s \in T(TM)$ this fact can be determined by means of a finite number of operations, due to the recursive enumerability of T. One says that s can be *recognized*.

All languages mentioned earlier are recursively enumerable as well, but there are recursively enumerable languages which are not context-sensitive (or context-free, or regular). The complement $CT(TM)$ of a $T(TM)$, moreover, is not necessarily recursively enumerable. This means that if $s \in CT(TM)$, there is no guarantee that this fact can be recognized

by a Turing machine. Or stated otherwise: it is not true that for all recursively enumerable languages there exists a procedure to decide for an arbitrary string s whether that string belongs to T *or not*. Languages for which such a procedure *does* exist are called *decidable* or *recursive*. These are recursively enumerable languages that have a complement which is also recursively enumerable.

2.2. Grammars and automata

2.2.1. Grammars and Turing machines

A recursively enumerable language can not only be described by a Turing machine, but also by means of a grammar. A grammar G can be characterized by the entities V, H, P, and S. V is a finite *terminal vocabulary* (with terminal elements $a, b, \ldots$); H is a finite *nonterminal vocabulary* (with auxiliary symbols, or variables $A, B, \ldots$) with a special *start symbol* S. And P, finally, is a finite set of *production rules*. V and H are disjoint: $V \cap H = \varphi$, whereas their union, $V \cup H = \Gamma$, is sometimes called the grammar's (unspecified) *vocabulary*. The rules of P are ordered pairs of strings (α, β), mostly written as $\alpha \rightarrow \beta$, where the first string (α) consists of one or more elements of Γ, and the second (β) of zero or more elements of Γ. To put it differently: $\alpha \in \Gamma^+$ (the strings of positive length over Γ), and $\beta \in \Gamma^*$ (the strings over Γ, including the null string λ). Thus $P \subset \Gamma^+ \times \Gamma^*$.

The rule $\alpha \rightarrow \beta$ means that string α can be replaced by string β in any context. Such a replacement is indicated by $\Rightarrow$. So, for instance, given rule $\alpha \rightarrow \beta$, one can replace string $\gamma\alpha\delta$ by $\gamma\beta\delta$; this is written as $\gamma\alpha\delta \Rightarrow \gamma\beta\delta$. More generally one writes $\xi \overset{*}{\Rightarrow} \psi$ (ψ is a derivation of ξ) if there is a sequence of zero or more replacements by which ξ can be transformed in ψ (zero replacements if $\xi = \psi$). A *sentence* generated by grammar G is any string of terminal elements, which can be derived from S by production rules of G. Thus, string σ is a sentence generated by G if there is a derivation $S \overset{*}{\Rightarrow} \sigma$, with $\sigma \in V^*$. The language $L(G)$, generated by G is the set of generated sentences, or $L(G) = \{\sigma | S \overset{*}{\Rightarrow} \sigma\}$. Example: let $G = (V, H, P, S)$, with $V = \{a(\text{pes}), b(\text{ake}), c(\text{akes})\}$, $H = \{N(\text{oun}), VP$ (verb phrase), $MV(\text{main verb}), S(\text{entence})\}$, and with production rules $P = \{S \rightarrow N\,VP,\ VP \rightarrow MV\,N,\ VP \rightarrow MV,\ N \rightarrow a,\ N \rightarrow c,\ MV \rightarrow b\}$, then one can make the following derivation from S: $S \Rightarrow N\,VP$, $N\,VP \Rightarrow a\,VP$, $a\,VP \Rightarrow aMV\,N$, $aMV\,N \Rightarrow abN$, $abN \Rightarrow abc$, or in short: $S \overset{*}{\Rightarrow} abc$. Since a, b, and c are terminal elements, the string a, b, c, or

apes bake cakes, is a sentence in $L(G)$. The reader can verify that the other sentences in $L(G)$ are: *apes bake*, *cakes bake*, and *cakes bake apes*.

It can be proven that the grammars defined in this way generate the class of recursively enumerable languages: For each recursively enumerable language $T(TM)$, there is a grammar G such that $L(G) = T(TM)$, while each language $L(G)$ is recursively enumerable. Thus, each Turing machine has an equivalent grammar, and inversely.

Chomsky has proposed to distinguish some progressively restrictive classes of grammars. We follow his classification.

2.2.2. Context-sensitive grammars and linear-bounded automata

The first restriction on the production rules is that 'shortening' rules are excluded. If the length of string α is written as $|\alpha|$ (thus, for instance, $|abc| = 3$), this restriction means that for all production rules $\alpha \to \beta$ in P it should be the case that $|\alpha| \leqq |\beta|$. Grammars which satisfy this restriction are called *type*-1 or *context-sensitive* grammars. Any language which can be generated by a context-sensitive grammar is a 'context-sensitive' or 'type-1' language.

We saw earlier that linear-bounded automata accept just the context-sensitive languages. It has been proven that context-sensitive grammars are equivalent to *LBA*'s. This can be grasped intuitively if one realizes that, during the generation of a sentence by means of a context-sensitive grammar, one can never obtain a string which is longer than the final sentence (otherwise one would need shortening rules to arrive at that sentence). In the same way, the *LBA* can never produce a string on its tape which is longer than the input sentence. Thus essentially the same restriction holds for both systems.

2.2.3. Context-free grammars and push-down automata

The second restriction is somewhat stronger. Apart from $|\alpha| \leqq |\beta|$ moreover $\alpha \in H$; consequently, α is a single auxiliary symbol. In this way we have a *type*-2, or *context-free*, grammar. Its rules appear as $A \to \beta$, where A is a variable, and β a string in Γ^+. Any language which can be generated by a context-free grammar is called a context-free language. A derivation by means of a context-free grammar can be easily made visible in the form of a *derivation-tree*, or *phrase marker*.

Fig. 23 presents the production rules of a context-free grammar G, and a derivation of the sentence *abcd*. Next to it the corresponding derivation tree is shown. One should notice that there is also another

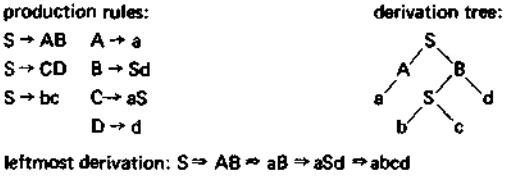

Fig. 23. Derivation of *abcd* with corresponding derivation tree.

derivation for the same sentence which corresponds to the same tree-diagram, namely $S \rightarrow AB \rightarrow ASd \rightarrow Abcd \rightarrow abcd$. Each variable is rewritten in the same way in both derivations (i.e. by the same rule), the only difference is that the rewriting is somewhat earlier or later. The derivation given in fig. 23 is called a *left-most derivation* of *abcd*, since at each step the left-most variable is rewritten. Each derivation tree corresponds to one and only one left-most derivation (if there is one in the grammar). Sentence *abcd*, however, has yet another left-most derivation by the rules in fig. 23. This must correspond, therefore, to a different derivation tree. Derivation and tree are presented in fig. 24.

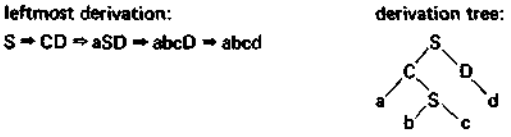

Fig. 24. Alternative derivation of *abcd* with derivation tree.

If a context-free grammar allows for two or more left-most derivations (phrase structures) for a sentence, as is the case in the grammar G under concern, that grammar is called ambiguous. If a context-free grammar is non-ambiguous, it will generate a deterministic language. Such a non-ambiguous context-free grammar is also called a '$LR(k)$-grammar'. A language is ambiguous if all of its grammars are ambiguous.

It has been shown that context-free grammars are equivalent to non-deterministic push-down automata. For every $T(NPDA)$ there is a context-free grammar CFG for which $L(CFG) = T(NPDA)$, and conversely. A similar equivalence exists between PDA's and $LR(k)$-grammars.

There are many other formal systems for the description of context-free languages. Examples are categorical grammars and dependency grammars, which will not be treated here.

2.2.4. Regular grammars and finite automata

The third and last restriction concerns β in the context-free rule $A \rightarrow \beta$. In this case β can have one of two forms only: $\beta \in V$, i.e. β is a single terminal element, or $\beta = aB$, with $a \in V$ and $B \in H$, in other words a terminal element followed by a variable.

The grammars resulting from this restriction are called *type*-3 or *regular* grammars (also: *finite state grammars*). Its rules thus have the forms $A \rightarrow a$ or $A \rightarrow aB$. It has been proven that regular grammars generate regular languages. Regular grammars are equivalent to finite automata: They define the same class of languages. Every finite language (i.e. with a finite number of sentences) can be generated by a regular grammar, and is, therefore, regular.

2.3. Chomsky's hierarchy of languages

Fig. 25 gives the relations of strict inclusion between the languages defined by Chomsky.

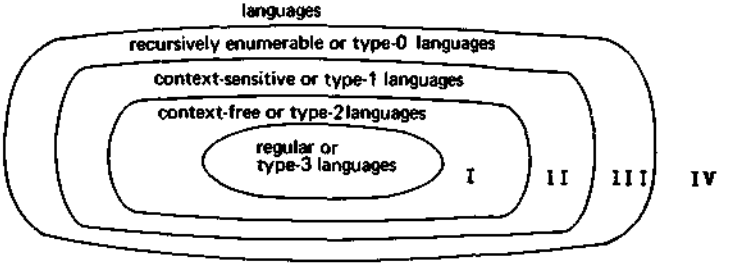

Fig. 25.

In this section a brief discussion will be given of the areas indicated by I, II, III, and IV, i.e. languages which are context-free but not regular (I), context-sensitive but not context-free (II), etc.

Area I. What is characteristic of a language which is not regular, i.e. which cannot be generated by a regular grammar? In contrast with regular languages, these languages are *self-embedding*. This requires some explanation. A grammar is called self-embedding if there is a variable B in H for which $B \overset{*}{\Rightarrow} \alpha B \gamma$, where $\alpha \neq \lambda$, and $\gamma \neq \lambda$. In words: The rules of the grammar are such that there is a variable B from which a string can be derived which contains B, but not at the left or right extreme. A language is called self-embedding if every grammar generating

the language is self-embedding. An example of a self-embedding language is $\{ww^R | w \in V^+, V = \{a, b\}\}$, i.e. a language consisting of symmetrical sentences, such as *aa*, *abba*, *baab*, *abbbba*, *abaaba*, etc. This is called a 'mirror-image' language. Another example is the language $\{a^n b^n | n \geqq 1\}$, consisting of sentences beginning with n *a*'s, followed by n *b*'s. Chomsky's proof that natural languages are not regular, and therefore not acceptable by finite automata (or Markov sources for that matter), was based on the demonstration that natural languages are self-embedding.

Area II. Context-sensitive, non-context-free languages are not so uniformly characterizable. It has been shown, however, that various languages belong to this category. An example is $\{a^n b^n c^n | n \geqq 1\}$, strings of *a*'s, followed by an equal number of *b*'s, followed by equally many *c*'s. Another example is the language consisting of string repetitions: $\{ww\}$, where w is any string of terminal elements, and sentences therefore consist of repetitions of such strings. These examples have been used to prove that natural languages are non-context-free (Levelt, 1974).

Area III. This category contains especially the non-decidable (or non-recursive) type-0 languages. These are recursively enumerable languages with complements that are not recursively enumerable. There are, however, decidable or recursive type-0 languages which still are not context-sensitive. Transformational grammars (see vol. II, chapter 7) are type-0 grammars: they can contain string-shortening production rules. It has been proven that Chomsky's transformational grammars generate at least the class of recursively enumerable languages, whereas natural languages are hopefully of a more restrictive sort. Other transformational grammars are more restrictive. Joshi's adjunction grammars, for instance, generate decidable languages (see Levelt, 1974).

Area IV. Not all sets of strings over a finite vocabulary (i.e. languages) can be described by means of a grammar. Saying that a natural language is type-0, or generable by a Chomskyan transformational grammar, is saying no more than that the language has a grammar.

2.4. Probabilistic grammars

The notion of grammar can be generalized by assigning probabilities to the production rules. In this way one obtains the so-called probabilistic grammars. This principle has been worked out especially for regular and for context-free grammars. A probabilistic grammar defines a probability distribution over the sentences of a language. One can

derive conditions for which the language is 'normalized', i.e. has a total probability of 1. The probabilistic grammar is a powerful instrument for the analysis of the so-called corpus, i.e. a set of observed strings (sentences, behavior sequences, etc.). Inferring a grammar from a corpus and estimating the probabilistic parameters for a given grammar constitute the subject of the theory of grammatical inference.

2.5. Grammaticality and controllability

A string s is called 'grammatical', given type-i grammar G, if $s \in L(G)$; it is 'ungrammatical' if $s \in CL(G)$. Also, and in view of the equivalence relations between grammars and automata discussed in the previous sections, one can say that s is grammatical if it is accepted by its equivalent automaton. Therefore, just those strings are grammatical by which the automaton can be controlled from the initial state. The linguistic notion of grammaticality, therefore, is closely related to the systems notion of controllability. In the same way, the system theoretical notion of observability is closely connected to grammatical inferability (cf. Levelt, 1975).

2.6. Psychological applications and schematic overview

The theory of automata and grammars has not only been applied in psycholinguistics (cf. vol. II, chapter 7), but also in the (ethological) analysis of behavioral sequences (see for instance Bodnar and Van Baren-Kets, 1974), in pattern recognition research (cf. two issues of *Pattern Recognition*, vol. 3(4) (1971), and vol. 4(1) (1972)), in the analysis of learning theories (Suppes, 1969), in the psychology of thinking (Suppes, 1973), and in memory research (Anderson and Bower, 1973).

Table 4, finally, summarizes the most important grammars, automata, languages and their complements. For the column 'language' there is strict inclusion from bottom to top: Every finite language is regular, every regular language is deterministic, etc., whereas the converse does not hold. For the column 'grammar' the same strict inclusions hold, with one exception: There exist 'ambiguous' regular grammars which are, therefore, not $LR(k)$; these grammars, however, do not generate ambiguous languages.

Table 4.
Scheme of grammars, language and automata.

Grammar	Equivalent automaton	Language	Complement of language
type-0	Turing machine	recursively enumerable or type-0	language
recursive (decidable)		recursive (decidable)	type-0
type-1, or context-sensitive	linear-bounded automaton	type-1, or context-sensitive	?, but at least type-0
type-2, or context-free (resp. dependency, categorical)	non-deterministic push-down automaton	type-2, or context-free	context-freeness undecidable, at least context-sensitive
LR(k)	deterministic push-down automaton	deterministic or non-ambiguous	deterministic
type-3, or regular (finite state)	finite automaton	regular, or finite state	regular
		finite language	regular

Literature

General area of system theory

1. Kok, J.J., Stassen, H.G., Signaaltheorie. T.H. Delft, Afdeling der Werktuigbouwkunde; Delft; vol. I, 1971; vol. II, 1972; vol. III, 1975; vol. IV, in preparation.

Probability

1. Papoulis, A., Probability, random variables and stochastic processes. New York: McGraw-Hill, 1965.
2. Cramer, H., Mathematical methods of statistics. Princeton: Princeton University Press, 1945.

Estimation theory

1. Deutsch, R., Estimation theory. Englewood Cliffs: Prentice-Hall, 1965.

Stochastic processes

1. Papoulis, A., Probability, random variables and stochastic processes. New York: McGraw-Hill, 1965.
2. Laning, J.H., Bating, R.H., Random processes in automatic control. New York: McGraw-Hill, 1956.

Decomposition of deterministic signals

1. Papoulis, A., The Fourier integral and its applications. New York: McGraw-Hill, 1962.
2. Courant, R., Hilbert, D., Methoden der mathematischen Physik I, II. Heidelberg: Springer-Verlag Berlin, 1968.

Transforms

1. Papoulis, A., The Fourier integral and its applications. New York: McGraw-Hill, 1962.
2. Lighthill, M.J., Introduction to Fourier analysis and generalized functions. New York: Cambridge University Press, 1958.
3. Hsu, H.P., Outline of Fourier analysis. New York: Unitech Division, 1967.
4. LePage, W.R., Complex variables and the Laplace transform for engineers. New York: McGraw-Hill, 1961.
5. Doetsch, G., Handbuch der Laplace Transformation I, II, and III. Stuttgart: Birkhauser, 1950–1956.
6. Spiegel, M.R., Theory and problems of Laplace transforms. Schaum's Outline Series. New York: McGraw-Hill, 1965.
7. Jury, E.J., Theory and application of the z-transform method. New York: Wiley, 1964.
8. DeRusso, P.M., Roy, R.J., Close, C.M., State variables for engineers. New York: Wiley, 1967.

Extensive tables of Laplace and z-transforms may be found in the following books:

9. Abramowitz, M., Stegun, J.A., Handbook of mathematical functions. New York: Dover Publications, 1965, pp. 1019–1030.
10. Tou, J.T., Digital and sampled-data control systems. New York: McGraw-Hill, 1959.

Decomposition of signals of stochastic processes

1. Papoulis, A., Probability, random variables and stochastic processes. New York: McGraw-Hill, 1965.
2. Davenport, W.B., Root, W.L., An introduction to the theory of random signals and noise. New York: McGraw-Hill, 1958.
3. Laning, J.H., Battin, R.H., Random processes in automatic control. New York: McGraw-Hill, 1956.
4. Blackman, R.B., Tukey, J.W., The measurement of power spectra from the point of view of communications engineering. New York: Dover Publ., 1958.
5. Jenkins, G.M., Watts, D.G., Spectral analysis and its applications. San Francisco: Holden-Day, 1969.

Linear systems

1. DeRusso, P.M., Roy, R.J., Close, C.M., State variables for engineers. New York: Wiley, 1965
2. Kuo, B.C., Analysis and synthesis of sampled-data control systems. Englewood Cliffs: Prentice-Hall, 1963.

State vector description

1. Zadeh, L.A., Desoer, C.A., Linear system theory: the state space approach. McGraw-Hill, 1963.
2. Kwakernaak, H., Sivan, R., Linear optimal control systems. New York: Wiley, 1972.

Non-linear systems

1. Gibson, J.E., Non-linear automatic control. New York: McGraw-Hill, 1963.
2. Csáki, F., Modern control theories: Nonlinear optimal and adaptive systems. Budapest: Akadémiai Kiadó, 1972.

Parameter and state estimation

1. Eykhoff, P., System identification, parameter and state estimation. New York: Wiley, 1974.
2. Nahi, N.E., Estimation theory and applications. New York: Wiley, 1969.

Automata and grammars

General

1. Hopcroft, J.E., Ullman, J.D., Formal languages and their relation to automata. Reading, Mass.: Addison-Wesley, 1969.
2. Levelt, W.J.M., Formal grammars in linguistics and psycholinguistics. Vol. I.: An introduction to the theory of formal language and automata. The Hague: Mouton, 1974.
3. Minsky, M.L., Computation: Finite and infinite machines. Englewood Cliffs: Prentice-Hall, 1967.

Applications

1. Anderson, J., Bower, G., Human associative memory. Washington: Winston, 1973.
2. Bodnár, F.A., Van Baren-Kets, E.J., Sequentiële analyses van gedragsobservaties bij jonge kinderen. Nederlands Tijdschrift voor de Psychologie, 1974, 29, 27–66.
3. Levelt, W.J.M., Formal grammars. Vol. II.: Applications in linguistic theory. Vol. III: Psycholinguistic applications. The Hague: Mouton, 1974.
4. Levelt, W.J.M., Systems, skills and language learning. In: Van Essen, A.J., Menting, J.P. (Eds.), The context of foreign language learning. Assen: van Gorcum, 1975.
5. Suppes, P., Stimulus-response theory of finite automata. Journal of Mathematical Psychology, 1969, 327–355.
6. Suppes, P., Theory of automata and its application to psychology. In: Dalenoort, G.J. (Ed.), Process models for psychology. The Hague: NUFFIC, 1973.

Lyst of symbols

a_k	– k^e Fourier coefficient
a_n	– coefficients of differential equation
A_k	– amplitude
$A(t)$	– system matrix
b_k	– k^e Fourier coefficient
$B(t)$	– input matrix
CT	– complement of language T
$C(t)$	– output matrix
$C(sI-A)^{-1}B+D$	– transfer matrix from $\underline{U}(s)$ to $\underline{Y}(s)$
$C_{xy}(\tau)$	– covariance function

$\hat{C}_{\bar{x}\bar{y}}(\tau)$ – estimator of covariance function $C_{\bar{x}\bar{y}}(\tau)$
$D(t)$ – direct transfer matrix
e – base of natural logarithm
$_eA(t)$ – transient matrix
$E\{\cdot\}$ – mathematical expectation
$f_{\bar{x}}(x)$ – probability density function
$f_{\bar{x}\bar{x}}(x_1, x_2; \tau)$ – joint probability density function
F – Fourier transform
F^{-1} – inverse Fourier transform
$g\{u(t)\}$ – non-linear system
$G(v)$ – transfer function of the linear model of the non-linear system $g\{u(t)\}$
G – grammar
$h(t, t_0)$ – impulse response of linear time-dependent system
$h(t)$ – impulse response of linear constant system
$H(s)$ – transfer function
H – finite non-terminal vocabulary
I – matrix
j – index for the imaginary part
J – criterion function
$k(\eta_{\bar{u}})$ – gain
$k_e(0, \sigma_{\bar{u}})$ – gain according to Booton
$K_{\bar{x}\bar{y}}(\tau)$ – correlation function
L – one-sided Laplace transform
L^{-1} – inverse one-sided Laplace transform
LBA – linear-bounded automaton
$m(t)$ – remnant
$NPDA$ – non-deterministic push-down automaton
Pr – probability
PDA – push-down automaton
p – finite set of production rules
$R_{\bar{x}\bar{y}}(\tau)$ – average product function
S – start symbol of grammar
s – Laplace operator of complex frequency
$(sI-A)^{-1}$ – transfer matrix from $U(s)$ to $x(s)$
$S_{\bar{x}\bar{x}}(v)$ – power density spectrum
t – time
$tU(t)$ – ramp function
T – time period
$T(A)$ – language accepted by automaton A
TM – Turing machine
$u(t)$ – control or input signal
$u_k(t)$ – orthogonal signal
$u(t)$ – input vector
$U(t)$ – unit step function
$v(t)$ – noise

V – terminal vocabulary of grammar, vocabulary of automaton
$w(t)$ – criterion weighting function
$x(t)$ – deterministic signal
$\bar{x}(t)$ – stochastic signal
$\bar{x}(t;\zeta)$ – stochastic process
$\bar{x}(\zeta)$ – random variable
$\underline{x}(t)$ – state vector
x_0 – initial state
$X(\nu)$ – Fourier transform of $x(t)$
$X(s)$ – (one-sided) Laplace transform of $x(t)$
$y(t)$ – output signal
$y(t_0)$ – initial condition
α_i – model parameters
Γ – push-down vocabulary, memory vocabulary of *LBA*, vocabulary of grammar
$\delta(t)$ – impulse function
δ – state function
ζ – domain of ensemble
$\eta_{\bar{u}}$ – average value of $\bar{u}(t;\zeta)$
λ – convergence abscis, null string
ν – frequency
$\sigma_{\bar{u}}^2$ – variance of $\bar{u}(t;\zeta)$
τ – time constant or time difference
Φ_k – phase of the k^{th} component
ω – radial frequency