Automatica, Vol. 16, pp. 519-526
Pergamon Press Ltd. 1980. Printed in Great Britain

0005-1098/80/0901-0519 $02.00/0

Ms. 195

Step Response and Frequency Response Methods

H. RAKE†

Key words—Identification; Fourier analysis; frequency response; step response.

Abstract—In this tutorial very simple methods will be presented by which models for dynamic processes can be obtained. These methods employ step responses of the process or responses to other nonperiodic testsignals or process responses to periodic signals.

INTRODUCTION

ONLY methods giving linear models will be treated in spite of the fact that most processes of interest will not be linear. Thus a linear model is fitted somehow to the process.

The discussion will be restricted further to single input–single output (SISO) models. Many processes with more than one input and/or more than one output may be represented by a number of SISO-models connecting each input to each output. These models can often be determined by investigating all signal paths separately from each other.

Figure 1 shows the general setup of process and model and will be referred to when discussing different methods. The process is exited by the input (test) signal u provided by a signal generator. The measurable process response y will in general consist of the uncorrupted process output plus additional process and/or measurement noise n. Thus all noise is thought to be condensed into this noise added to the process output. It will be shown that different identification methods are more or less sensitive to noise. It will also become apparent that in almost any identification a large signal to noise ratio is desirable. Thus in the design of experiments and measurements all means should be employed to diminish noise and to enlarge usable signals.

Test-signal amplitudes in general are subject to restrictions. With regard to safety, product quality etc. large deviations from setpoints cannot be tolerated in most real processes. In addition to this, nonlinearities in many processes prevent the use of large signals for determining linear models.

The identification methods to be presented will be classified according to the test signals employed. This leads to two groups of methods one of which uses nonperiodic test signals whereas the other works with periodic signals. Both groups will be subdivided into methods giving nonparametric models and those giving parametric models respectively.

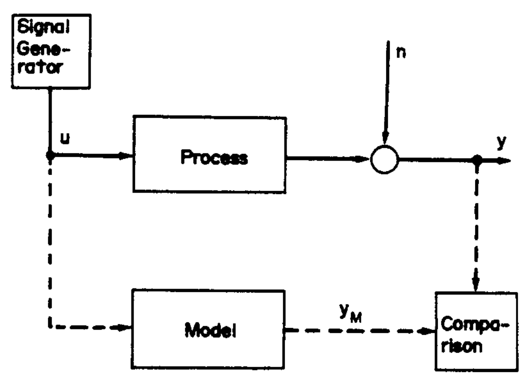


FIG. 1. General setup of process and model.

Description of processes and models

As implied by Fig. 1 process models will be chosen and rated on the basis of input and output signals. Thus only the transfer characteristics of process and model have to be described. Models are graded according to their ability to generate an output signal y_M which is very similar to the signal y generated by the process when model and process are exited by the same input signal u.

Models very often are described by their weighting function $h(t)$ or their frequency response function $H(j\omega)$. When no analytic expression for these functions is known a nonpara-

*Received 19 March, 1980. The original version of this paper was presented at the 5th IFAC Symposium on Identification and System Parameter Estimation which was held in Darmstadt, Federal Republic of Germany during September 1979.
The published Proceedings of this IFAC Meeting may be ordered from: Pergamon Press Limited, Headington Hill Hall, Oxford OX3 0BW, England.
This paper was recommended for publication in revised form by the Automatic Editorial Board.
†Inst. f. Regelungstechnik, RWTH Aachen, Aachen, Germany.

metric model is obtained, which is represented by a plot or a number of sample values. Best known member of this set of models probably is the frequency response represented by a Bode plot. Parametric models are described by the number and the values of parameters of differential equations, frequency response functions or transfer functions. From these analytic representations plots or values of interest of frequency response or weighting functions can be computed in general without much difficulty whereas the inverse process namely deriving parameters from nonparametric model descriptions is much more difficult.

A fairly general parametric model can be described by the differential equation

$$a_n y_M^{(n)}(t)+\ldots+a_1\dot{y}_M(t)+a_0 y_M(t)=b_0 u(t-\tau_d) + b_1\dot{u}(t-\tau_d)+\ldots+b_m u^{(m)}(t-\tau_d) \qquad (1)$$

or by the corresponding frequency response

$$H(j\omega)=\frac{b_0+b_1 j\omega+\ldots+b_m(j\omega)^m}{a_0+a_1 j\omega+\ldots+a_n(j\omega)^n}\exp(-j\omega\tau_d). \qquad (2)$$

Solving the differential equation for the input signal

$$u(t)=\delta(t)\begin{cases}=0 & \text{for } t\neq 0\\ \neq 0 & \text{for } t=0\end{cases}$$
$$\int_{-\infty}^{\infty}\delta(t)\,\mathrm{d}t=1 \qquad (3)$$

with $\delta(t)$ being the unit impulse function gives the so-called weighting function

$$y_M(t)=h(t) \qquad (4)$$

as corresponding output function. This weighting function if represented by a plot or a number of sample values can be regarded as a nonparametric model description and be used in the convolution integral

$$y_M(t)=\int_0^{\infty}h(\tau)\cdot u(t-\tau)\,\mathrm{d}\tau \qquad (5)$$

to compute the output $y_M(t)$ for arbitrary inputs $u(t)$.

Of course the differential equation can be solved by Laplace- or Fourier-Transform techniques employing transfer functions or frequency response functions.

Signals

It has been mentioned before that two different types of input signals u (Fig. 1) will be taken into account namely periodic and nonperiodic ones. In case of a nonperiodic input (test) signal the process being in equilibrium operation, corresponding to zero initial conditions, is disturbed. The process may then settle into a new equilibrium which can be equal to the former one or start drifting out of the region of normal operation. Which way the process will behave depends on the properties of the process and the input signal.

Periodic signals in general are employed in such a way that the process is running in a steady state with the output oscillating with the same frequency (or frequencies) as the input and all transients having subsided.

In contrast to the deterministic input signal well known in advance the noise n (Fig. 1) has to be treated as a stochastic signal about which only its statistical properties if any at all might be available. In many cases of practical interest the noise can be approximated by stochastic signals with Gaussian amplitude distribution and a spectral density which is either independent of frequency (white noise) or a not too complicated function of frequency (coloured noise, e.g. low pass-filtered noise). In most cases the noise will be considered as having zero mean

$$E[n]=0 \qquad (6)$$

and its intensity will be described by its variance

$$\sigma_n^2=E[n^2] \qquad (7)$$

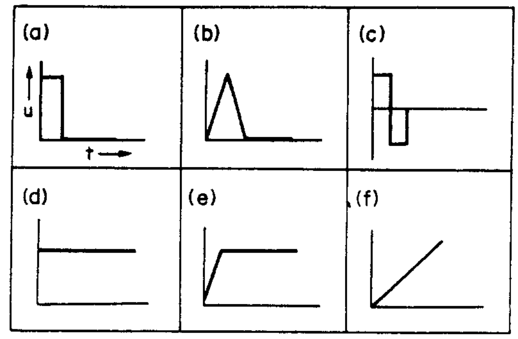
with σ as so called standard deviation or root mean square (r.m.s.) value.

METHODS FOR NONPERIODIC TEST SIGNALS

Nonperiod test signals

Nonperiodic test signals in common use are: impulses of different shape and relatively short duration; step functions; and ramp functions. Table 1 shows a collection of nonperiodic test signals.

TABLE 1. NONPERIODIC TEST SIGNALS



When choosing a test signal for a specific application the properties of the signal, the possibilities for generating and applying the signal and the kind of information wanted from the test have among others to be taken into account. Thus universal rules are hard to formulate.

Nonparametric models from nonperiodic signals

As stated before nonparametric process models of common interest are: the weighting function $h(t)$; and the frequency response function $H(j\omega)$.

An obvious and fairly simple way for obtaining the weighting function $h(t)$ would be to use

$$u(t)=K\cdot\delta(t) \tag{8}$$

as an input signal, giving an output

$$y(t)=K\cdot h(t)+n(t) \tag{9}$$

which in case of negligible noise $n(t)$ will give $h(t)$ without further effort. In practice the Dirac impulse of equation (8) has to be approximated by impulse functions with short but finite duration and limited amplitude (Table 1a, b). If these impulse functions are of a duration that is short in comparison to the duration of the weighting function, systematic errors resulting from use of these signals instead of a Dirac impulse might be tolerable. The (approximated) weighting function then can be computed as

$$\hat{h}(t)=\frac{y(t)}{\int_0^\infty u(t)\,\mathrm{d}t}. \tag{10}$$

In cases where impulse functions of non-negligible duration have to be used, inversion of the convolution integral equation (5) leads to a (numerical) deconvolution method for the determination of the weighting function. Though feasible in principle this method is not used very often due to the facts (among others) that: weighting function models are not very common; deconvolution may suffer significantly from ill conditioned inputs and/or non-negligible noise.

As model representations via frequency response functions are used very often, methods for deriving frequency responses from input–output signal pairs are of common interest. Frequency response functions are calculated from Fourier transforms of input and output signals.

$$\hat{H}(j\omega)=\frac{Y(\omega)}{U(\omega)}=\frac{\mathscr{F}[y(t)]}{\mathscr{F}[u(t)]}. \tag{11}$$

Equation (11) will give satisfying results only if the noise (Fig. 1) is negligible as noise components in the output signal $y(t)$ will be treated in the same way as genuine process outputs are.

Fourier transforms of signals measured are usually computed by digital computers via the Discrete Fourier Transform (DFT), which can be realized in a computationally efficient manner by algorithms using the so called Fast Fourier transform (FFT). Without going into too much detail, it seems worthwhile to bear in mind that the Fourier transform $F(\omega)$ of a function of time $f(t)$ is

$$F(\omega)=\int_{-\infty}^{\infty} f(t)\cdot\exp(-j\omega t)\,\mathrm{d}t \tag{12}$$

and that it exists only if

$$\int_{-\infty}^{\infty} |f(t)|\,\mathrm{d}t<\infty. \tag{13}$$

Here signals $u(t)$ and $y(t)$ are to be transformed. By suitable definitions one can ensure

$$y(t)=u(t)=0 \quad \text{for } t<0 \tag{14}$$

and thus

$$Y(\omega)=\int_0^\infty y(t)\cdot\exp(-j\omega t)\,\mathrm{d}t. \tag{15}$$

For numerical evaluation of equation (15) the signal $y(t)$ is sampled at equidistant points $n\cdot T$ in time and sample values $y(n)$ are stored and processed. For practical reasons functions of time, in order to fulfill equation (13) and thus be Fourier-transformable, have to tend to zero for large $t>T_N=N\cdot T$. Such signals can be represented by a finite number of samples $y(n)$, $n=0,\ldots,N-1$ and their DFT can be obtained by evaluating the finite sum

$$Y(k\cdot\Delta\omega)=\sum_{n=0}^{N-1} y(n)\cdot\exp\left(-j\frac{2\pi kn}{N}\right) \qquad k=0,\ldots,N-1 \tag{16}$$

with

$$\Delta\omega=\frac{2\pi}{N\cdot\Delta t} \tag{17}$$

instead of the infinite integral equation (15).

Equation (16) indicates, that the DFT $Y(k\cdot\Delta\omega)$ is defined for N discrete values of frequency and thus the associated frequency response is defined for the same N discrete frequencies. Evidently the DFT of the input signal can be obtained in the same manner as described above for the output signal. Often standard input signals will be used whose Fourier transform is known beforehand or can be calculated analytically or looked up in a suitable table. In such cases the output signal only has to be processed.

From equation (13) is to be seen that step functions and step response functions tending to a finite nonzero value for large time values are not Fourier-transformable. This difficulty can be circumvented by not transforming the signals themselves but their first derivatives. One can prove that using simple differences

$$y'(n)=y(n+1)-y(n) \qquad (18)$$
$$u'(n)=u(n+1)-u(n)$$

in order to obtain

$$\hat{H}(k\cdot\Delta\omega)=\frac{\mathscr{F}[y'(n)]}{\mathscr{F}[u'(n)]} \qquad (19)$$

does not introduce systematic errors as input and output signal are preprocessed in the same way.

As mentioned before noise is a big problem in system identification with nonperiodic signals. All methods using and processing such signals are feasible only in cases of high signal to noise ratio. This can easily be seen for the determination of frequency response functions. If the process has a frequency response $H(\omega)$ and the noise $n(t)$ gives a Fourier transform

$$N(\omega)=\int_0^{T_N} n(t)\cdot\exp(-j\omega t)\,\mathrm{d}t \qquad (20)$$

the measurable process output signal will have a Fourier transform

$$Y(\omega)=U(\omega)\cdot H(\omega)+N(\omega) \qquad (21)$$

and the model frequency response will be

$$\hat{H}(\omega)=\frac{Y(\omega)}{U(\omega)}=H(\omega)+\frac{N(\omega)}{U(\omega)}. \qquad (22)$$

Thus the additional noise leads to additional errors in the frequency response. By averaging several response functions one might attenuate the effects of noise to a certain degree.

Parametric models from aperiodic signals

Parametric models usually are described by their frequency response function

$$H(\omega)=\frac{b_0+b_1 j\omega+\ldots b_m(j\omega)^m}{a_0+a_1 j\omega+\ldots a_n(j\omega)^n}\cdot\exp(-j\omega\tau) \qquad (23)$$

with a_j, b_j, m, n, τ as parameters.

Many authors have proposed methods for determining parameter values from time functions of the process output provided the process input is a well determined signal. Of primary importance are step functions as input and step responses as output signals. Here the discussion will be limited to this signal pair though there are other input signals (Table 1) with interesting properties.

Almost all methods for the evaluation of step response functions use a small number of characteristic values of the response function. One of the oldest and best known is that introduced by Küpfmüller giving a first order lag model with time delay and a frequency response

$$\hat{H}(\omega)=\frac{K}{1+j\omega\tau_1}\cdot\exp(-j\omega\tau). \qquad (24)$$

The parameters K, τ_1, τ can be derived from the step response as indicated in Fig. 2 with u_0 being

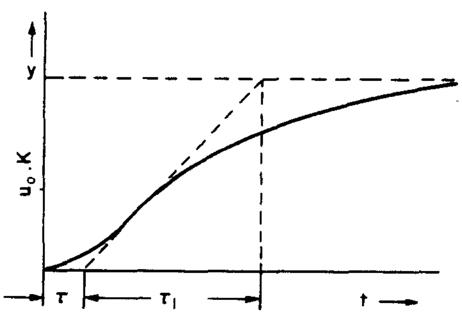


FIG. 2. Evaluation of step responses according to Küpfmüller.

the amplitude of the input step signal. Another method has been developed by Strejc giving parameter values for

$$\hat{H}(\omega)=\frac{K}{(1+j\omega\tau_1)^n}\cdot\exp(-j\omega\tau). \qquad (25)$$

Equation (25) describes a model of order n with equal time constants and a time delay. The gain factor K can be obtained in the usual way by dividing the final value y_∞ of the step response Fig. 3 by the input step amplitude u_0. The remaining parameters τ, τ_1 and n are found by construction of the tangent with maximum slope and measurement of several intermediate values as indicated in Fig. 3. Then with the aid of Table 2 the model order is estimated from T_u/T_g and y_i/y_∞. Usually an integer value just below or equal to that indicated by these quotients will be chosen and T_u, τ and T_i adjusted if necessary. From T_i, T_u, T_g and T_m as measured according to Fig. 3 and quotients T_i/T, T_u/T, T_g/T and T_m/T in Table 2 corresponding with the order n chosen before four values of T are obtained. If these do agree fairly well an average is to be computed

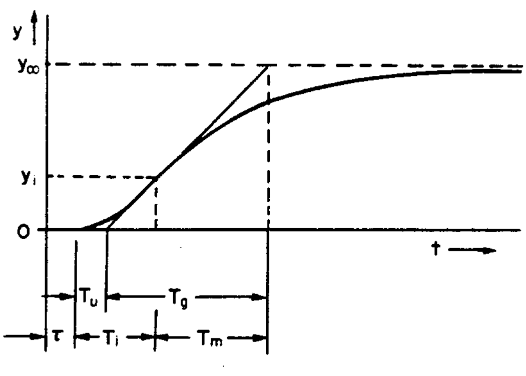


FIG. 3. Evaluation of step responses according to Strejc.

TABLE 2. CHARACTERISTIC VALUES OF STEP RESPONSES

n	$\frac{T_g}{T}$	$\frac{T_u}{T}$	$\frac{T_u}{T_g}$	$\frac{T_i}{T}$	$\frac{y_i}{y_\infty}$	$\frac{T_m}{T}$	$\frac{T_m}{T_g}$
1	1	0	0	0	0	1	1
2	2.718	0.282	0.104	1	0.204	2.000	0.736
3	3.695	0.805	0.218	2	0.323	2.500	0.677
4	4.463	1.425	0.319	3	0.353	2.888	0.674
5	5.199	2.100	0.410	4	0.371	3.219	0.629
6	5.699	2.811	0.493	5	0.384	3.510	0.616
7	6.226	3.549	0.570	6	0.394	3.775	0.606
8	6.711	4.307	0.642	7	0.401	4.018	0.599
9	7.146	5.081	0.709	8	0.407	4.245	0.593
10	7.590	5.869	0.773	9	0.413	4.458	0.587

and the time constant is obtained by

$$\tau_1 = \frac{T}{n}. \qquad (26)$$

In case these values for T disagree a different (usually higher) order n and choice of another (usually smaller) delay time τ might alleviate this disagreement.

A still more detailed method has been developed by Schwarze. With the aid of tables and graphs the user is put into a position where he can evaluate step response, and ramp response functions and determine parameters for several different model structures. For reasons of brevity only one model structure and one set of intermediate values to be extracted from a step response will be discussed.

The parameters K, T, b and n of the frequency response

$$\hat{H}(\omega) = \frac{K}{(1+j\omega T)\cdot(1+j\omega bT)^{n-1}} \qquad (27)$$

describing a model of order n can be obtained via the time values T_{10}, T_{50}, T_{90} measured as indicated in Fig. 4. From these e.g. the quotient T_{90}/T_{50} is computed and used to determine a suitable model order with the aid of Fig. 5. From this figure one might find that $T_{90}/T_{50}=2$ in-

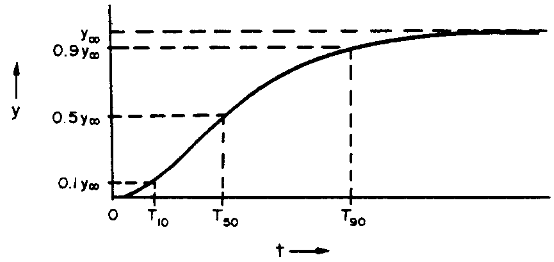


FIG. 4. Evaluation of step responses according to Schwarze.

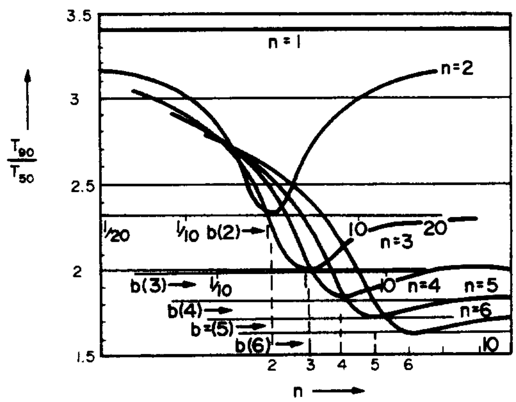


FIG. 5. Determination of model order.

dicates a model order of $n=3$ with equal time constants ($b=1$) or one of $n>3$ with $b<1$. Having chosen n (usually as small as possible) graphs similar to Fig. 6 and Fig. 7 can be used to determine the factor b and a corresponding set of quotients T_{90}/T, T_{50}/T and T_{10}/T from which three values of the time constant T are to be obtained. In case of good agreement all para-

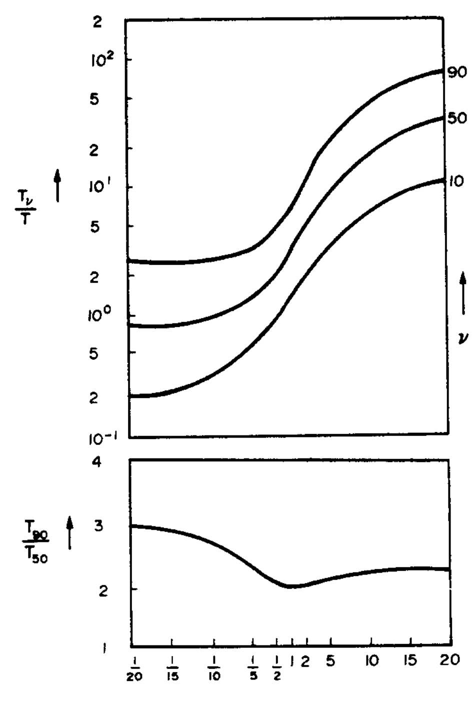


FIG. 6. Determination of time constant and factor b for $n=3$.

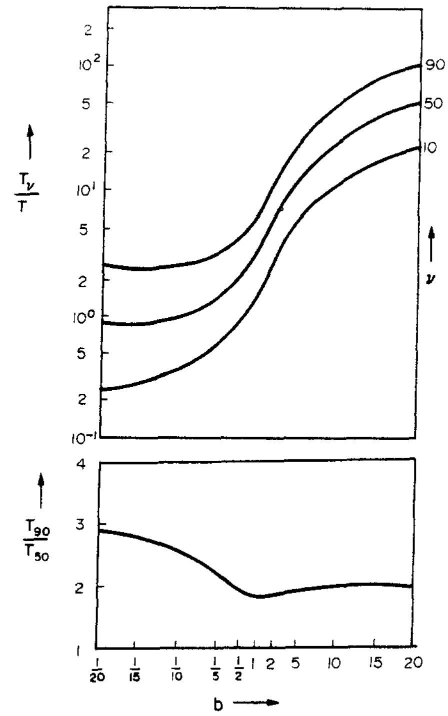


FIG. 7. Determination of time constant and factor b for $n=4$.

meters have been determined. In case of significant disagreement a different order n should be tried.

A quite different method uses limit theorems of the Laplace transform to determine the coefficients of the model transfer function

$$\hat{H}(s)=\frac{b_0+b_1s+\ldots+b_ms^m}{1+a_1s+\ldots+a_ns^n} \tag{28}$$

from the unit step response $\eta(t)=y(t)/u_0$. As

$$\lim_{t\to\infty}\eta(t)=K_0=\lim_{s\to 0}\hat{H}(s)=b_0 \tag{29}$$

one has

$$b_0=K_0 \tag{30}$$

with K_0 from Fig. 8 as indicated there. With the integral

$$\eta_1(t)=\int_0^t[K_0-\eta(\tau)]\,\mathrm{d}\tau \tag{31}$$

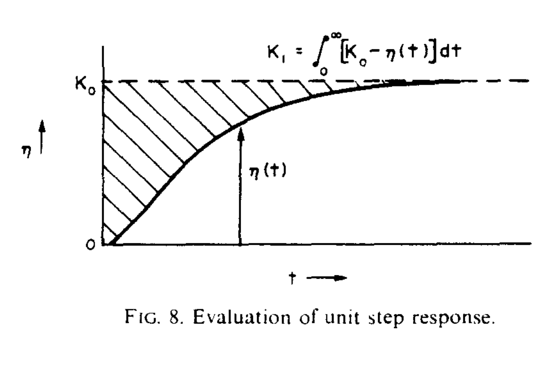


FIG. 8. Evaluation of unit step response.

and correspondingly

$$\hat{H}_1(s)=\frac{1}{s}[K_0-\hat{H}(s)] \tag{32}$$

the limit theorem gives

$$\lim_{t\to\infty}\eta_1(t)=K_1=\lim_{s\to 0}\hat{H}_1(s)=K_0\cdot a_1-b_1 \tag{33}$$

Another step of integrating etc. leads to

$$b_2+K_1\cdot a_1-K_0\cdot a_2=K_2 \tag{34}$$

and the general formula

$$(-1)^rb_r+K_{r-1}a_1-K_{r-2}a_2+\ldots+(-1)^{r-1}K_0a_r=K_r \tag{35}$$

giving a system of linear equations.

In case all b_j in equation (28) but b_0 are zero one has

$$\begin{aligned}K_0&=b_0\\K_1&=K_0\cdot a_1\\K_2&=K_1\cdot a_1-K_0\cdot a_2\end{aligned} \tag{36}$$

which can be solved recursively without much effort.

Of course determination of K_i for $i\geqq 2$ can be difficult and prone to errors limiting the area of useful application to low order models.

Equation (33) provides an important information for models with frequency response

$$\hat{H}(\omega)=\frac{K}{\prod_{i=1}^{n}(1+j\omega\tau_i)}. \tag{37}$$

As in this case

$$K=K_0 \tag{38}$$

and

$$a_1=\sum_{i=1}^{n}\tau_i \tag{39}$$

the area in Fig. 8 from which K_1 is to be obtained is equivalent to the sum of all time constants multiplied by the gain K. This property of the step response may be used for first approximations as well as for checking purposes.

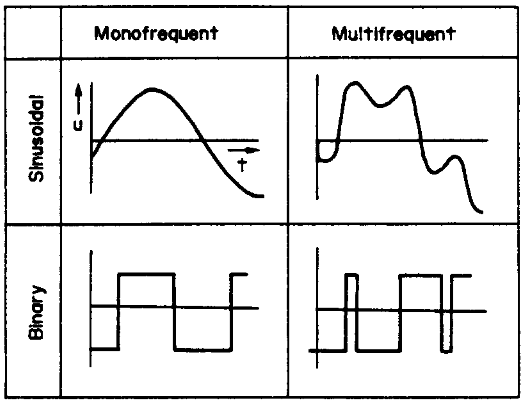
METHODS FOR PERIODIC TEST SIGNAL

Periodic test signals

All measurements with periodic test signals are conducted with the process in steady state, i.e. all transients produced by initial conditions have vanished.

Test signals in common use are simple sinusoidal or binary monofrequent signals or multifrequency signals containing significant signal amplitudes at more than one frequency. The latter signals can be of sinusoidal or binary type (Table 3).

TABLE 3. PERIODIC TEST SIGNALS



Nonparametric models from periodic signals

Models determined from periodic input and output signals are usually of the frequency response type. This type of model is most easily determined. In the simplest case of a monofrequent sinusoidal input and an uncorrupted output of a linear process, amplitude and phase of the frequency response can readily be computed from a recording of both signals by a two channel recorder (Fig. 9) giving

$$|\hat{H}(\omega)| = \frac{Y}{U}, \quad \phi(\omega) = -\frac{t\phi}{T} \cdot 360^0. \tag{40}$$

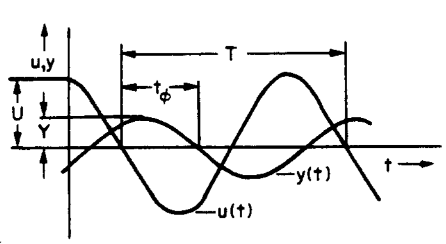


FIG. 9. Evaluation of sinusoidal response.

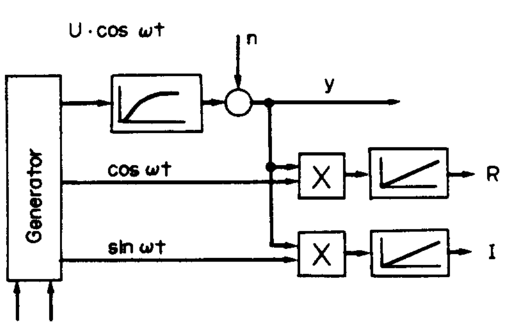


FIG. 10. Determination of frequency response by correlation.

The not-so-simple problem of significant noise and/or process nonlinearity can be handled with a setup according to Fig. 10 which is the principle of a couple of commercially available frequency response analysers. By multiplying the process output by reference signals obtained from the test signal generator and integrating the products over an integer number K of test signal periods T one gets

$$R = \frac{K \cdot T \cdot U}{2} Re\,\hat{H}(\omega) + \int_0^{K \cdot T} n(t) \cdot \cos \omega t \, dt$$

$$I = \frac{K \cdot T \cdot U}{2} Im\,\hat{H}(\omega) + \int_0^{K \cdot T} n(t) \cdot \sin \omega t \, dt. \tag{41}$$

As can be shown the average of the integrated noise in equation (41) is zero and the variance of these integrals is bounded. Thus measurements taken over a sufficient number of periods can be made as accurate as desired even in cases of low signal to noise ratio.

If the power spectral density S_{nn} of the noise is known the standard deviation of the frequency response as determined according to Fig. 10 and equation (41) is

$$\sigma_{\hat{H}}(\omega) = \frac{2}{U \cdot |H(\omega)|} \sqrt{\left[\frac{S_{nn}(\omega)}{K \cdot T}\right]} \tag{42}$$

which can be decreased by increasing measurement time $K \cdot T$.

The method described above and illustrated by Fig. 10 can easily be interpreted as determination of the basic component of a Fourier series expansion of the process output. In fact all modern transfer function analysers are organized as Fourier analysers processing by digital hardware signal sample values taken at equidistant points in time. With such equipment or by processing signals recorded in a suitable manner by general purpose digital computers binary input signals can be employed. This can be advantageous as simple switching control elements only are needed to apply this test signal to the process. Another interesting possibility is the use of multifrequency sinusoidal or binary test signals (Table 3). The process output resulting from such input signals will of course contain signal components with different frequencies. These have to be separated by Fourier series expansion giving several e.g. up to ten values of the frequency response at predetermined frequencies of interest.

Though treated rather briefly here frequency response measurements with periodic signals if applicable have to be regarded as the most powerful method at hand for determining process models with high accuracy even from noisy measurements.

Parametric models from period signals

If parametric (frequency response) models are wanted from measurements with periodic signals the usual way is to determine in a first step a sufficient number of frequency response values i.e. a nonparametric model. This nonparametric model is used as a basis for further operations.

An impressive number of methods have been developed which aim at the determination of values of coefficients in analytic frequency response functions. One rather simple but fairly effective graphical method is based on the Bode plot of frequency responses. From the results of transfer function measurements a Bode plot is drawn and this in turn is approximated by elementary frequency response functions. This trial and error method can be made more effective by use of prefabricated drawing aids etc.

Computer oriented methods can be divided into interpolating methods and error minimizing methods. The former imply all measured values to be correct and determine analytic expressions which satisfy each and every measurement. In contrast the error minimizing methods determine parameters of an analytic expression. The structure of this expression usually has to be specified in advance. Parameters then are determined in such a way that a measure of total error e.g. the (weighted) sum of errors squared becomes minimal.

CONCLUSION

Identification and process parameter estimation will always be different things to different people. As there are many different reasons for experimental process identification the practising cost-conscious scientist and engineer needs to know at least in principle a variety of methods at hand from which he has to select a suitable one. For reasons of simplicity and economy, methods using deterministic periodic and nonperiodic test signals deserve close attention.

REFERENCES

Eykhoff, P. (1974). *System Identification.* John Wiley, London.

Isermann, R. (1971). *Experimentelle Analyse der Dynamik industrieller Prozesse.* Identifikation I. Bibliographisches Institut, Mannheim.

Schwarze, G. (1964). Algorithmische Bestimmung der Ordnung und Zeitkonstanten bei P-, I- und D-Gliedern mit zwei unterschiedlichen Zeitkonstanten und Verzögerung bis 6. Ordnung. *Messen, Steuern, Regeln* 7, 10–19.

Strejc, V. (1959). Näherungsverfahren für aperiodische Übergangscharakteristiken. *Regelungstechnik* 7, 124–128.

Strobel, H. (1975). *Experimentelle Systemanalyse.* Akademie-Verlag, Berlin.

Wilfert, H.-H. (1969). *Signal- und Frequenzganganalyse an stark gestörten Systemen.* VEB Verlag Technik, Berlin.