ELECTRIC DISTORTION MEASUREMENT

In the context of the electrical domain, the term *distortion* may be broadly defined as any deviation of a signal in any parameter (time, amplitude, or wave shape) from that of an ideal signal. The term is usually associated with lack of fidelity. When referring to analog signals, distortion means shape alteration. For binary data transmission the term has a particular meaning commonly defined as a displacement in time of a signal from the time that the receiver expects to be correct. When applied to a system, distortion is a manifestation of its nonideal behavior in which the manipulation of at least some type of signal is involved.

DISTORTIONLESS SYSTEM

A periodic signal $s(t)$ may be represented by the following Fourier series:

$$s(t) = S_0 + \sum_{n=1}^{\infty} S_n \cos(n\omega_1 t + \varphi_n) \tag{1}$$

where S_0 represents the dc component of the signal, S_n is the amplitude and φ_n the phase of the nth harmonic, and ω_1 is the angular frequency of its fundamental component. If the signal is not periodic and it satisfies some conditions, the representation does not involve a series but an integral, the Fourier integral given by

$$s(t) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \overline{S}(\omega) e^{j\omega t}\, d\omega \tag{2}$$

where

$$\overline{S}(\omega) = \int_{-\infty}^{+\infty} s(t) e^{-j\omega t}\, dt \tag{3}$$

is the Fourier transform of $s(t)$.

According to the Fourier representation, each signal $s(t)$ may be viewed as the sum of an infinite number of sinusoidal signals having a well-defined amplitude, frequency, and phase. The representation is unique. No two signals have the same Fourier representation and vice versa.

An electric analog signal $s(t)$ is said to suffer from distortion when, for some reason, it is transformed into a signal $s'(t)$ that does not satisfy the condition

$$s'(t) = ks(t - t_0) \tag{4}$$

where k is a constant that accounts for a change in amplitude and t_0 is a time delay. Any signal that does not satisfy Eq. (4) is not an exact replica of $s(t)$.

There are several causes for signal distortion: conducted or radiated interference (harmonics in the power network voltage and current, crosstalk between conductors, for instance); signal manipulation (modulation, mixing, etc.); and nonideal behavior of the media or of the systems used to transmit and manipulate the signal.

Even though the first two causes are important, the last deserves further consideration because real transmission media and systems are always nonideal and thus may contribute significantly to signal distortion. To understand the mechanism of distortion produced in a signal by a medium or system, let us first consider that the media or systems are linear and time-invariant so that they are characterized by a transfer function in the frequency domain:

$$\overline{T}(\omega) = T(\omega) e^{j\theta(\omega)} \tag{5}$$

In such cases, both media and systems may be analyzed in terms of systems theory, and thus we will refer to both as a "system." Because the transfer function relates the output of a system to its input, it accurately describes the system's frequency response and thus the effect of such a system on its signals.

Considering that the input of the system is a periodic signal $x(t)$ represented by its Fourier series

$$x(t) = X_0 + \sum_{n=1}^{\infty} X_n \cos(n\omega_1 t + \varphi_n) \tag{6}$$

then the steady state output signal $y(t)$ is given by

$$y(t) = Y_0 + \sum_{n=1}^{\infty} Y_n \cos(n\omega_1 t + \Psi_n) \tag{7}$$

where

$$\begin{cases} Y_n = T(n\omega_1) X_n \\ \Psi_n = \varphi_n + \theta(n\omega_1) \end{cases} \qquad n = 0, 1, 2, \cdots \tag{8}$$

If $\overline{T}(\omega)$ is such that

$$T(\omega) = k \tag{9a}$$

and

$$\theta(\omega) = -t_0 \omega \tag{9b}$$

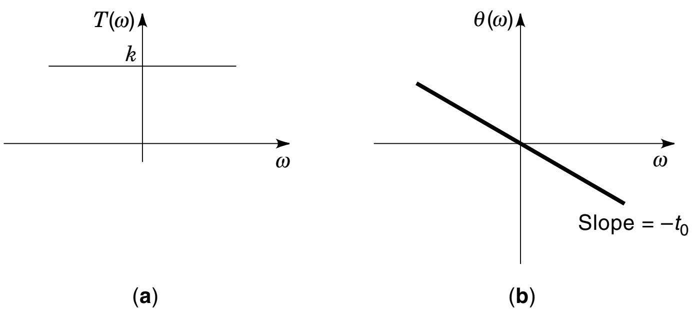

Figure 1. Transfer function of an ideal system. (a) Amplitude response; (b) phase response. A distortionless system is a linear and time-invariant system that has a constant amplitude frequency response k and a phase response that changes in frequency according to a straight line passing through the origin and whose slope is a time delay t_0.

then replacing Eqs. (9a) and (9b) in Eq. (7) yields

$$y(t) = k\left\{X_0 + \sum_{n=1}^{\infty} X_n \cos[n\omega_1(t-t_0) + \varphi_n]\right\} = kx(t-t_0) \quad (10)$$

The same result is obtained for a nonperiodic signal provided that it may be represented by a Fourier transform $\overline{X}(\omega)$ and that the transfer function of the system satisfies Eqs. (9a) and (9b). In fact, extending the concept of the transfer function to the domain of negative frequencies and taking into consideration that because $x(t)$ and $y(t)$ are real, $T(\omega)$ must be an even function of ω, and $\theta(\omega)$ an odd function of ω, one may write:

$$\overline{Y}(\omega) = \overline{T}(\omega)\overline{X}(\omega) = ke^{-j\omega t_0}\overline{X}(\omega) \quad (11)$$

$y(t)$ is obtained by replacing Eq. (11) in Eq. (2):

$$y(t) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} k\overline{X}(\omega)e^{j\omega(t-t_0)}\,d\omega = kx(t-t_0) \quad (12)$$

The analysis just presented leads to the important conclusion that a distortionless system must be linear and time-invariant and have an amplitude response constant for all frequencies and a phase response linear with frequency. The transfer function of an ideal system is represented in Fig. 1.

In a nonlinear system, the output must be related to the input by means of a n-degree polynomial commonly called a transfer characteristic. In those systems distortion always occurs, because a sinusoidal input produces an output with an enriched spectral content.

DISTORTION TAXONOMY AND TERMINOLOGY

As already discussed, electric distortion produces a change in a signal parameter and, in the context of the Fourier representation of a signal, it corresponds to one or several of the following alterations: appearance of energy in new frequencies (nonlinear amplitude distortion); nonproportional change in the amplitudes of the fundamental and harmonic components of the signal (frequency distortion); or change in the phase of the frequencies of the signal (phase or delay distortion). These are the three principal types of distortion. The first is characteristic of nonlinear systems and thus is called nonlinear distortion. The other two are characteristic of dispersive systems and are called linear distortions.

The distortion produced in electrical signals becomes a problem when the information they convey is altered or even lost or when those distorted signals interfere with other signals. Because of the different origins of distortion and the need to characterize and evaluate the performance of systems, the terminology related to distortion includes several expressions that are worth examining because some of them define parameters used in specifying systems distortion performance.

Telecommunications is probably the electrical subdomain where distortion causes the most problems. Two main factors contribute to that: (a) the use of components and devices that are intrinsically nonlinear (mixers, loudspeakers, etc.) or potentially nonlinear (e.g., amplifiers) affect the spectra of the signals they process; (b) the congestion of the frequency spectrum. Thus it is only natural to expect the terminology of distortion to be directly related to sound and video signals.

The following terms are strongly supported by Ref. 1. The list is not exhaustive but aims to include those terms most relevant in electrical distortion measurements.

Amplitude nonlinearity (amplitude distortion) is the phenomenon through which frequency components appear at the output of a system. They are dependent on the characteristics of the input signal but not present in it. Amplitude nonlinearity produces two types of amplitude distortion: harmonic distortion and intermodulation distortion. Crossover distortion caused by the nonlinear characteristics of a device that changes operating modes, such as a push-pull amplifier, is an example of amplitude distortion.

Harmonic distortion is the amplitude nonlinearity expressed in terms of the ratio of the harmonics in the output signal to the total output signal when a sinusoidal input signal is applied. It is caused by nonlinearity in the system's transfer characteristic. In harmonic distortion measurements, a single sinusoidal signal is applied to the system, and wave analysis at harmonic frequencies determines the percentage distortion.

Harmonic distortion of the nth *order* is the harmonic distortion expressed in terms of the ratio of the rms output signal due to the component of harmonic order n to the total rms output signal. Usually, the difference between the total rms and the rms of the fundamental is very small, and thus the latter is also used in the ratio. *Total harmonic distortion* is the harmonic distortion expressed in terms of the ratio of the rms output signal due to distortion to the total rms output signal. *Weighted total harmonic distortion,* used in sound system equipment, is the total harmonic distortion measured with a frequency weighting. *Noise harmonic distortion* (or noise distortion) is the harmonic distortion when one-third octave-band filtered noise is used as the input signal.

Intermodulation distortion is the amplitude nonlinearity expressed in terms of the ratio of the input signal of frequencies $pf_1 + qf_2 + \cdot\cdot\cdot$ (where $p, q, \ldots$, are positive or negative integers) to the total output signal, when (at least two) sinusoidal input signals having the fundamental frequencies f_1, $f_2, \ldots$, are applied at the input. In radio frequency (RF) power amplifiers, for instance, the major causes of intermodulation distortion are crossover effects, gain reduction at high current, and device saturation.

In intermodulation distortion measurements, two sinusoidal signals of different frequencies f_1 and f_2 are applied to the system. *Modulation distortion* is the intermodulation distortion where the input signal is composed of a large-amplitude, low-frequency signal f_1 and a small-amplitude, high-frequency signal f_2. In some systems two kinds of modulation distortion are present, both having the same spectral components and differing only in phase: (a) amplitude modulation distortion caused by the amplitude modulation due to nonlinearity; (b) frequency modulation caused by frequency modulation having no relationship to nonlinearity. In such cases, it is necessary to distinguish between these two types of distortion. The reference output at which the distortion occurs is taken as the arithmetic sum of the output signals at frequencies f_1 and f_2.

Modulation distortion of the nth order is the modulation distortion in terms of the ratio of the arithmetic sum of the rms output signal components at frequencies $f_2 \pm(n-1)f_1$ to the rms output signal component at frequency f_2. *Total modulation distortion* is the modulation distortion in terms of the ratio of the arithmetic sum of the rms output signal components to the rms output signal at frequency f_2. *Difference-frequency distortion* is the intermodulation distortion where the input signal is composed of two sinusoidal signals f_1 and f_2 of similar or equal amplitude. The difference in the frequency of the two signals is less than the lower of the frequencies. The reference output at which the distortion occurs is taken as the arithmetic sum of the output signals at frequencies f_1 and f_2. *Noise intermodulation distortion* is the intermodulation distortion where one-third octave-band filtered noise is used as the input signal. *Transient intermodulation distortion,* important in characterizing loudspeakers, results from nonlinear response to steep wavefronts. It is measured by adding square wave (3.18 kHz) and sine wave (15 kHz) inputs with a 4 : 1 amplitude ratio and observing the multiple sum and difference-frequency components added to the output spectrum.

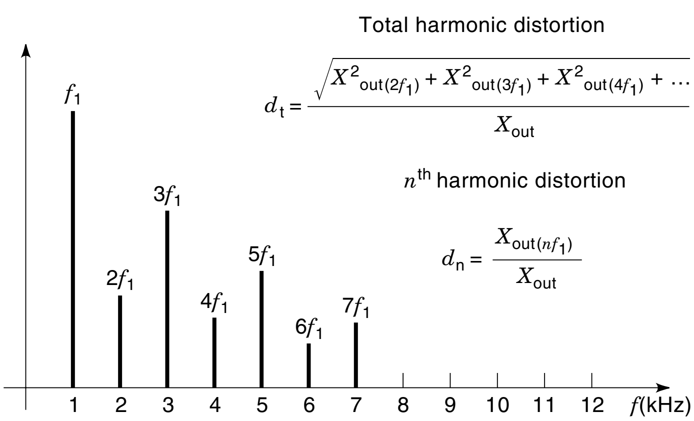

Figure 2. Harmonic distortion. Total harmonic distortion, d_t, and nth harmonic distortion, d_n, are two parameters for characterizing the harmonic distortion of a system. In the expressions, X_{out} represents the rms of the output signal (an electric voltage, in general), $X_{out}(f_1)$ represents the rms of the fundamental in the output signal, and $X_{out(nf_1)}$ represents the rms of the harmonic component of nf_1 in the output signal. The input of the system is a sine wave signal of frequency f_1.

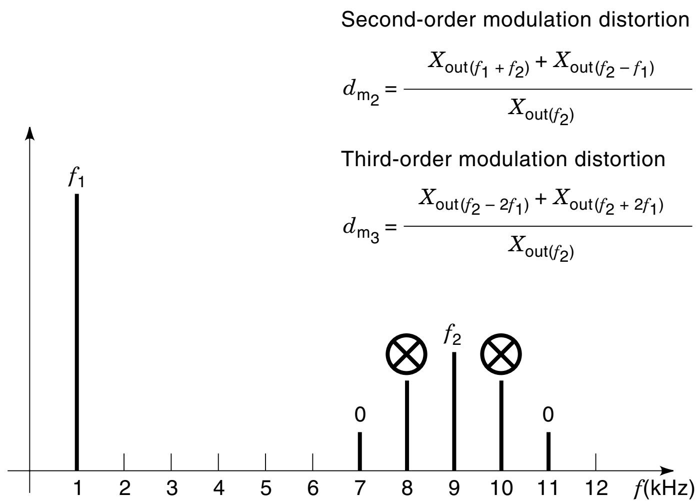

Figure 3. Modulation distortion. Modulation distortion of the nth order d_{m_n} is a parameter for characterizing the intermodulation distortion of a system. The input of the system is a signal with frequencies f_1 and f_2. Refer to the legend of Fig. 2 for the meaning of the variables in the expressions.

In Figs. 2–5, excerpted from Ref. 1, some of the concepts previously discussed are presented in graphical and mathematical form.

Frequency distortion is the effect on a signal that results from variation in the amplitude response of a system as a function of frequency. Some authors also use attenuation distortion or amplitude distortion to designate this effect. If the amplitude response assumes values between k_{max} and k_{min}, the parameter $(k_{max} - k_{min})/[(k_{max} + k_{min})/2]$ may be used to express frequency distortion.

Phase or delay distortion results from the deviation from a constant slope of the output phase versus frequency response of a system. This produces echo responses in the output that precede and follow the main response and a distortion of the output signal when an input signal having a large number of frequency components is applied. When the phase characteristic of a linear system assumes the value θ at frequency f_1, the system introduces at that frequency a time delay $t_{d_1} = \theta_1/2\pi f_1$ between the input and the output. If the system is not ideal, the time delay $t_{d_2} = \theta_2/2\pi f_2$ introduced at frequency f_2 differs from t_{d_1}. In that case, the derivative of the phase with respect to frequency is not constant. In our opinion, the maximum value of that derivative over any frequency interval expressed in time units better characterizes phase distortion. Some authors (2) designate this parameter as envelope delay distortion. The experimental evaluation of this parameter may be cumbersome or even impossible, which leads to the implementation of alternative methods. Details are presented in a forthcoming section.

SIGNAL DISTORTION MEASUREMENT

Electric distortion measurements are usually carried out by examining signals (electric voltage, as a rule) in the frequency domain (harmonic analysis). However, both frequency-domain and time-domain instrumentation are used for that purpose. Frequency-domain instrumentation analyzes of signals by using analog filtering techniques. Wave analyzers, such as the frequency selective voltmeter, the heterodyne tuned voltmeter, the heterodyne harmonic analyzer (wavemeter), and the heterodyne spectrum analyzer, are examples of this type of instrumentation designed to measure the relative amplitudes of single-frequency components in a complex signal. Time-domain instrumentation analyzes by time sampling the signals and subsequent numerical handling of the sampled data commonly using the fast Fourier transform (FFT) algorithm. The FFT spectrum analyzer is an example of time-domain instrumentation.

Special-purpose instruments, such as the one whose block diagram is presented in Fig. 6 (total harmonic distortion meter), directly display many distortion measurements. The spectrum analyzer is, however, the general-purpose instrument most often used to measure distortion. With it, the entire spectrum within its frequency band is analyzed even though second and third harmonic measurements are enough for many applications. The most common spectrum analyzers are the superheterodyne spectrum analyzer and the FFT spectrum analyzer.

In the FFT spectrum analyzer the input signal is converted, the samples converted from analog to digital, and a FFT is performed. As a result, magnitude and phase spectra of the input signal are obtained. The main advantages of FFT spectrum analyzers compared to superheterodyne spectrum analyzers are their ability to measure phase and the possibility of characterizing single-shot phenomena. Their limitations are related to the frequency range (limited by the ADC maximum conversion rate) and sensitivity (related to quantization noise). FFT spectrum analyzers are very easily implemented by using PC-based automatic measuring systems with plug-in data acquisition boards or by using digitizers and digital oscilloscopes having computer interfaces like RS232 or IEEE488. Now manufacturers are including FFT spectrum analyses as an additional feature of digital oscilloscopes.

Figure 7 shows a simplified block diagram of a superheterodyne spectrum analyzer. After the input attenuator, the signal is applied to a low-pass filter, whose function is analyzed later. The output of the filter is applied to a mixer. Here the signal is mixed with the output of a voltage-controlled

$$X_{\text{out}(f_1)} = X_{\text{out}(f_2)}$$

$$X_{\text{out}_{\text{ref}}} = X_{\text{out}(f_1)} + X_{\text{out}(f_2)} = 2X_{\text{out}(f_2)}$$

$f_2 - f_1 = 80$ Hz, for instance

$f_{\text{m}} = \dfrac{f_2 + f_1}{2}$ is a preferred one-third octave band center frequency (for instance, 10kHz)

Second-order difference-frequency distortion

$$d_{\text{d}_2} = \frac{X_{\text{out}(f_2 - f_1)}}{2X_{\text{out}(f_2)}} = \frac{X_{\text{out}(f_2 - f_1)}}{X_{\text{out}_{\text{ref}}}}$$

Third-order difference-frequency distortion

$$d_{\text{d}_3} = \frac{X_{\text{out}(2f_2 - f_1)} + X_{\text{out}(2f_1 - f_2)}}{X_{\text{out}_{\text{ref}}}}$$

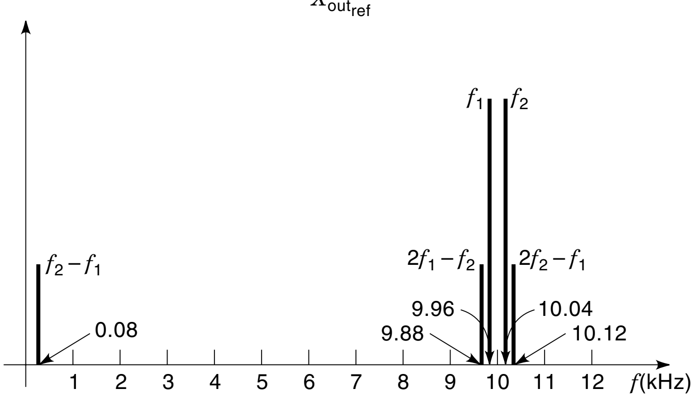

Figure 4. Difference-frequency distortion. Difference-frequency distortion of the nth order d_{d_n} is a parameter for characterizing the intermodulation distortion of a system when a signal having two closely spaced frequency components of similar amplitudes is supplied to the input. In general, this parameter indicates the in-band distortion introduced by the system. Refer to the legend of Fig. 2 for the meaning of the variables in the expressions.

$X_{\text{out}(f_1)} = X_{\text{out}(f_2)}$ with, for instance,

$f_1 = 8$ kHz $f_2 = 11.95$ kHz

$$X_{\text{out}_{\text{ref}}} = X_{\text{out}(f_1)} + X_{\text{out}(f_2)} = 2X_{\text{out}(f_2)}$$

$f' = f_2 - f_1 = 3.95$ kHz $f'' = 2f_1 - f_2 = 4.05$ kHz

Total difference-frequency distortion:

$$d_{\text{d}_{\text{tot}}} = \frac{\sqrt{(X'_{\text{out}})^2 + (X''_{\text{out}})^2}}{X_{\text{out}}}$$

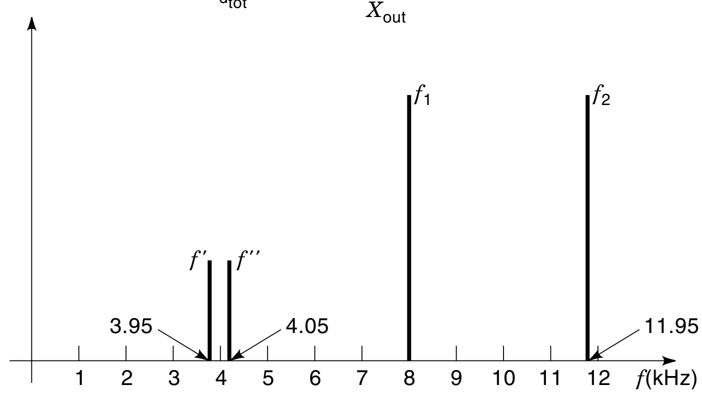

Figure 5. Total difference-frequency distortion. Total difference-frequency distortion $d_{\text{d}_{\text{tot}}}$ is a parameter particularly relevant in assessing the out-of-band distortion introduced by a system. Note that the frequencies of the two-tone signal supplied to the input are not closely spaced as in Fig. 4. Refer to the legend of Fig. 2 for the meaning of the variables in the expressions.

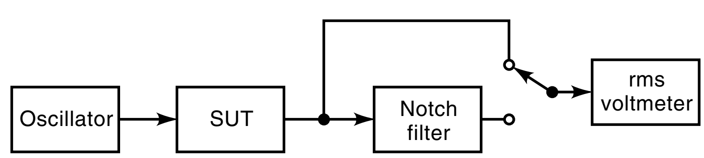

Figure 6. Block diagram of a typical fundamental-suppression total harmonic distortion meter. Total harmonic distortion by the system under test (SUT) is evaluated by internally computing the ratio between the rms values of the output voltage and its value upon suppression of its fundamental frequency. The instrument includes the oscillator, and the rms values are measured by an rms responding voltmeter.

oscillator (VCO). The ramp generator sweeps the VCO linearly from f_{min} to f_{max}. Because the mixer is a nonlinear device, its output contains the two original signals and also their harmonics, the sums and differences of the original frequencies, and their harmonics. When any of the frequency components of the mixer output falls within the passband of the filter, a nonzero voltage is applied to the envelope detector and after amplification, to the vertical plates of a cathode-ray tube (CRT), producing a vertical deflection of the electron beam. As the ramp that commands the VCO is also applied to the horizontal plates of the CRT, the horizontal axis can be calibrated in frequency.

The central frequency f_{if} and the bandwidth of the intermediate-frequency filter are chosen so that at any time the only frequency component at the output of the mixer that is within the band of the filter is the difference between the frequency f_{vco} of the VCO and that of the input signal f_{signal}. This implies that f_{if} must be out of the input band. Otherwise, apart from the difference of frequencies we could have a component of the input signal within the passband of the filter. In this case, because the output of the mixer includes the original input signal, this would produce a constant vertical deflection of the CRT during all the frequency scanning of the VCO.

To display frequencies near 0 Hz, the lower frequency f_{min} of the VCO must be equal to f_{if}, because $f_{if} = f_{vco} - f_{signal}$. To display all the frequencies within the band of the spectrum analyzer, the maximum output frequency f_c of the VCO must be $f_{max} = f_{if} + f_c$.

Now let us suppose that we have a spectrum analyzer with an input range of frequencies from 0 to 3 GHz. In this case, f_{if} could be, for instance, 3.5 GHz, and then the output frequency of the VCO should vary from 3.5 GHz to 6.5 GHz. Suppose that we have an input signal with two frequency components, one at 1 GHz (f_{s1}) and the other at 2 GHz (f_{s2}). When the ramp begins, the beam is deflected to the left of the CRT screen, and the VCO oscillates at 3.5 GHz. As the ramp amplitude grows, the beam is moving to the right of the screen, and f_{vco} increases. Suppose that at a given moment f_{vco} = 3.6 GHz at the output of the mixer. Then we have the following components: 1 GHz (f_{s1}), 2 GHz (f_{s2}), 3.6 GHz (f_{vco}), 2.6 GHz ($f_{vco} - f_{s1}$), 4.6 GHz ($f_{vco} + f_{s1}$), 1.6 GHz ($f_{vco} - f_{s2}$), and 5.6 GHz ($f_{vco} + f_{s2}$). Because the bandwidth of the bandpass filter is much less than 0.1 GHz, none of the components appear after the filter, and so no vertical deflection occurs at the screen. This is the case until f_{vco} reaches 4.5 GHz. Then the components at the output of the mixer are 1 GHz (f_{s1}), 2 GHz (f_{s2}), 4.5 GHz (f_{vco}), 3.5 GHz ($f_{vco} - f_{s1}$), 5.5 GHz ($f_{vco} + f_{s1}$), 2.5 GHz ($f_{vco} - f_{s2}$) and 6.5 GHz ($f_{vco} + f_{s2}$). The output of the bandpass filter is no longer zero. There is a component at 3.5 GHz ($f_{vco} - f_{s1}$) whose amplitude is proportional to the amplitude of the input signal component of 1 GHz. This produces a vertical deflection when the horizontal deflection produced by the ramp corresponds to 1 GHz. During the rest of the sweeping the vertical deflection would be zero except when the frequency of the VCO reaches 5.5 GHz. At that time the difference between f_{vco} and f_{s2} is within the band of the filter, and a vertical deflection proportional to the amplitude of f_{s2} appears on the screen.

Now we are ready to understand the function of the low-pass filter at the input. Suppose that a component at 8.5 GHz (f_{s3}) is present in the input signal and that the spectrum analyzer does not have the low-pass filter. When f_{vco} = 5 GHz the difference between f_{s3} and f_{vco} is 3.5 GHz and falls within the passband of the filter. So it produces a vertical deflection on the screen proportional to the amplitude of f_{s3}. The problem is that this component would be displayed when the ramp reaches the voltage level corresponding to 2 GHz. This would give the user the erroneous indication that a 2 GHz component is present in the input signal, instead of the 8.5 GHz component really present. Then the function of the low-pass filter is to prevent these high frequencies from getting to the mixer. The bandwidth of this filter should be equal to the range of the spectrum analyzer.

Superheterodyne spectrum analyzers are not real-time instruments. They need the input signal to remain unchangeable during the sweep time, and storage CRTs are necessary to display the spectrum of the input signal.

To resolve signals with closely spaced frequency components, spectrum analyzers have bandpass filters with bandwidths as narrow as 10 Hz. Such narrow filters are difficult (or impossible) to achieve, especially at high center frequencies as in our example at 3.5 GHz. Adding mixing stages solves this problem. Figure 8 shows a simplified block diagram of a spectrum analyzer with two mixing steps. The output of the bandpass filter centered at 1 MHz differs from zero only when $f_{vco} - f_{signal}$ is equal to the central frequency of the first bandpass filter (3.5 GHz in the example), plus or minus the bandwidth of the last bandpass filter (the bandwidth of the 1 MHz filter in the example).

In some spectrum analyzers the signal is converted after the last bandpass filter from analog to digital (at a much lower rate then if it were converted at the instrument input), and then digital filtering is performed, allowing implementation of very narrow filters.

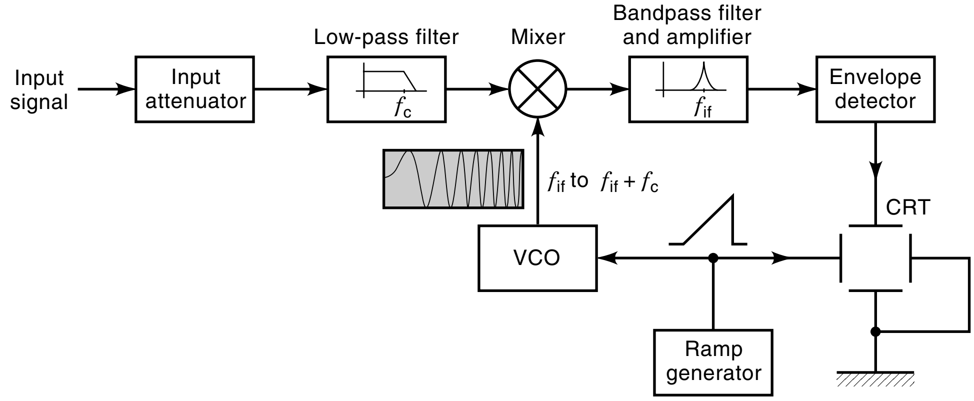

Figure 7. Simplified block diagram of the superheterodyne spectrum analyzer. The superheterodyne spectrum analyzer allows measuring the amplitude or rms values of the frequency components of an electric voltage in a defined frequency band. Those values are obtained by measuring the output voltage of a narrow-passband, fixed-frequency filter when, upon being heterodyned, the spectrum of the input voltage passes the frequency window of the filter.

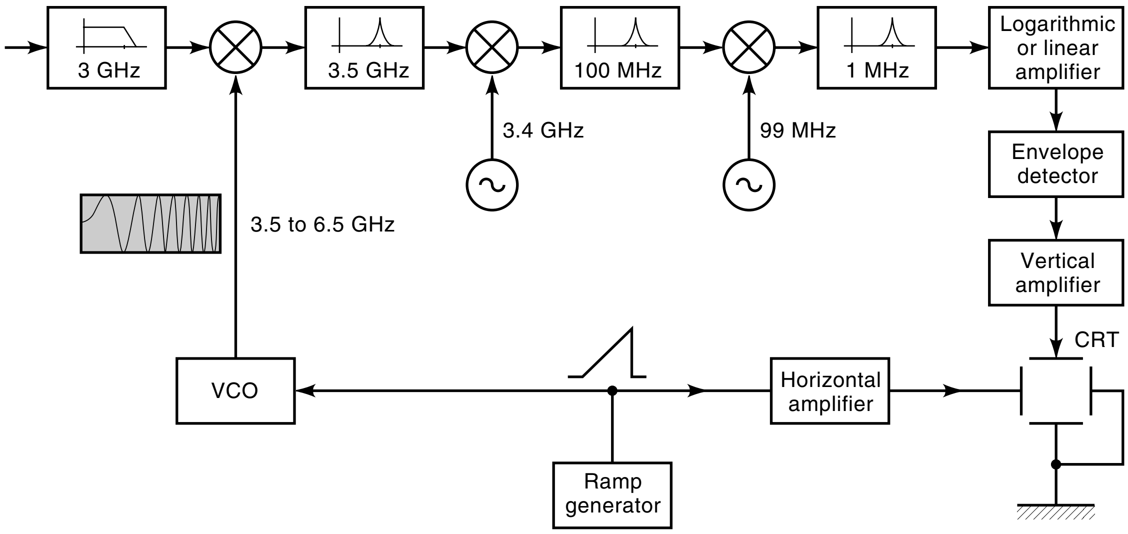

Figure 8. Block diagram of the superheterodyne spectrum analyzer with two mixing steps. The inclusion of multiple mixing stages in a superheterodyne spectrum analyzer allows high-resolution spectral analysis of high frequency voltages and also analog-to-digital conversion of the voltage representing the components of the input voltages.

Sensitivity is the measure of the smallest amplitude that the spectrum analyzer can display. The ultimate limitation in measuring a low-level signal is the random noise generated by the instrument itself. The thermal noise generated in the circuit elements is amplified by the different gain stages, added to the noise they generate, and displayed on the screen as a noise signal below which one cannot make measurements. The instrument sensitivity is determined by measuring the noise level on the display without any applied input signal. Signals at lower levels cannot be measured because they are masked by the noise. Even though the input attenuator and mixers have little effect on the actual system noise before the first gain stage, they do have a marked effect on the ability of the instrument to measure low-level signals because they attenuate the input and so they reduce the signal-to-noise ratio. Choosing the minimum input attenuation maximizes the instrument's sensitivity.

To minimize the internal spectrum analyzer's harmonic distortion, signal levels should be kept as low as possible at the input of the spectrum analyzer mixer. This means that to increase accuracy, the input attenuator of the spectrum analyzer must be used to decrease the level of the signal applied to the mixer when high-level signals are applied. However, this reduces the signal-to-noise ratio and so the instrument's sensitivity.

The bandwidth of the bandpass filter (resolution bandwidth) affects sensitivity. The spectrum analyzer generates random noise of constant amplitude over a wide range of frequencies. Because part of the internally generated noise is present at the input of the bandpass filter, the noise present at the output also decreases and sensitivity increases when the filter bandwidth decreases.

The dynamic range of a spectrum analyzer is defined as the difference between its maximum input voltage and its noise level.

Frequency resolution is the ability of the spectrum analyzer to separate closely spaced input signal frequency components. It depends on the bandwidth of the narrowest filter in the chain (see Fig. 8). As the VCO is swept in frequency, the input of the bandpass filter is also swept. Unless two distinct input signal frequency components are far enough apart when compared with the filter bandwidth, the traces they produce on the screen fall on top of each other and look like only one response.

Band-pass band filters, as with band-limited circuits, require finite time to respond to an input stimulus. Because the rise time of a filter is inversely proportional to its bandwidth, the narrower the resolution of the filter, the greater the time it needs to respond to the input. If the VCO is swept too quickly, there is a loss of displayed amplitude in the screen of the spectrum analyzer. This means that when a narrow filter is selected, sweep time must increase, otherwise the instrument's accuracy is affected. It can be shown (3) that the sweep time must decrease with the square of the bandwidth to assure that the time when the mixer output is within the passband is of the order of magnitude of the rise time of the filter. This means that each time the resolution bandwidth is reduced by a factor of 10, the sweep time goes up by a factor of 100. If we select a very narrow filter, the sweep time becomes prohibitive. For instance, a bandwidth of 30 Hz in a 10 division display with 50 MHz/div selected, leads to a sweep time of 34 days!!! Some spectrum analyzers automatically set sweep time to the span and bandwidth resolutions selected to maintain the instrument's calibration. Others allow the user to select sweep time also, but when this is too small it indicates that the display is uncalibrated.

The amplitude accuracy of the spectrum analyzer depends on several factors. The input attenuator and the first mixer must present a flat frequency response over the entire band of the instrument. In a low-frequency instrument, ±0.5 dB of deviation from a flat response is a typical value, but for a spectrum analyzer with a frequency range of tens of GHz, ±4 dB is an acceptable value. The fidelity of the logarithmic characteristic of the log amplifiers and the linearity of the envelope detector characteristic also affect amplitude accuracy. Impedance mismatch is also a source of error at high frequencies. Spectrum analyzers do not have perfect input impedances. In most cases, however, uncertainty is relatively small. When this is not the case, the use of a well-matched attenuator at the instrument input solves the problem.

The measurement of signal distortion must consider the characteristics of the signal. The change with time of the harmonic components of a signal may create four types of signals: (a) signals with quasi-stationary components; (b) signals with fluctuating components; (c) signals with rapidly changing components; and (d) signals with interharmonics and other spurious components. Continuous real-time analysis is required for (b). For (c) continuous real-time measurement is absolutely necessary because the value of each component is meaningful only when obtained through statistical analysis of a set of values measured over time. Very precise requirements are necessary to get reproducible results.

The use of a superheterodyne spectrum analyzer is not incompatible with real-time analysis, but it requires (1) the possibility of external control of the analyzer's local oscillator; (2) that the analyzer have an analog-to-digital converter at the output of the bandpass filter, a digital memory, and a CPU so that successive values of each frequency component may be stored and processed. These features are common in many spectrum analyzers now commercially available.

SYSTEMS DISTORTION MEASUREMENT

The evaluation of distortion introduced into electrical signals by systems, such as electronic devices and equipment, consists of one or several of the following basic measurements involving electric voltages:

1. Linear distortion (frequency and phase distortion): measurement of the amplitude and phase of a system's output voltage as a function of the frequency. The system is driven by a sine wave voltage whose amplitude is kept constant and whose frequency is swept in the range of interest. The output voltage amplitude and the phase shift between the input and output voltages are measured. Frequency distortion is evaluated by dividing the difference between the maximum and minimum indications of the voltmeter by half their sum. The result is expressed either as a percentage or in logarithmic units (dB). Delay distortion, expressed in time units, commonly μs or ms, is determined by dividing each phase shift (in rad) by the corresponding angular frequency (in rad/s) and selecting the maximum difference of the obtained ratios. The test setup for linear distortion measurements is shown in Fig. 9.
2. Harmonic distortion (nonlinear): measurement of the absolute values or values relative to the fundamental of the harmonics in the output voltage of the system when a sine wave voltage is applied to the system. In some cases, when the nonlinearity of the system depends heavily on frequency, the system is subjected to a band-limited noise signal. Harmonic distortion is usually expressed either as a percentage or in logarithmic units (dB or, sometimes, dBc, that is, dB relative to the fundamental, when the harmonics have small amplitudes compared to the fundamental). Figure 10 represents a possible test setup for harmonic distortion measurement.
3. Intermodulation distortion (nonlinear): the system is supplied with at least two sine waves of different frequencies, and the frequency components in the output voltage are measured. As in the case of harmonic distortion measurement, when the nonlinearity of the system depends heavily on frequency, the system is subjected to a band-limited noise signal. Intermodulation distortion is usually expressed either as a percentage or in logarithmic units (dB). The test equipment for intermodulation distortion measurement is shown in Fig. 11.

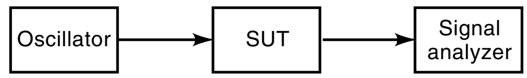

Figure 10. Setup for electric harmonic distortion measurement. The signal analyzer is generally a spectrum analyzer with an input voltage attenuator.

Testing to assess the performance of a system depends on the application. In sound systems, where harmonics and intermodulation products that fall in the audible spectrum produce distortion, harmonic and intermodulation distortion measurements are mandatory. Because the human ear is relatively insensitive to delay distortion, however, this type of electrical distortion needs no attention. In video and data signals, delay distortion constitutes one of the most limiting impairments, which means that it must be measured.

Following the ideas just presented, the measurement of electric distortion has been a subject of standardization. Thus the test of a loudspeaker, an audio amplifier, or a TV receiver involves many different distortion measurements according to specific methods and procedures. Standards IEC 268 Parts 3,

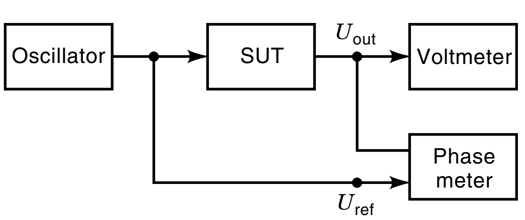

Figure 9. Setup for linear electric distortion measurement. The system under test (SUT) is subjected to a sine wave voltage. The voltmeter may be rms or peak responding. The function of the voltmeter and of the phase meter may be performed by a vector network analyzer.

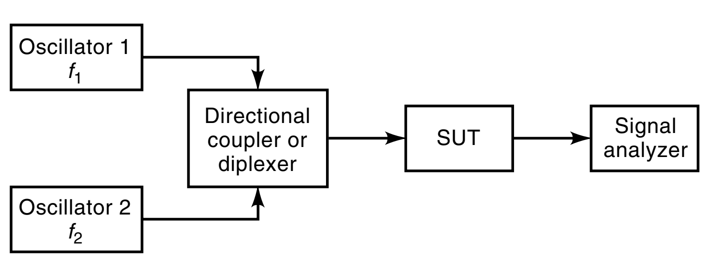

Figure 11. Setup for electric intermodulation distortion measurement. The two oscillators are connected to the system under test (SUT) through a directional coupler or a diplexer. The oscillators may be replaced by a multitone generator.

4, 5, and 6 for sound system equipment, IEC 244 Parts 4 and 4A for radio transmitters and IEC 244 Part 5 for television transmitters are texts where the reader may find useful information concerning the measurement of electric distortion in those systems.

The following are some considerations on equipment and measuring methods for distortion measurement:

1. The level of total harmonic distortion of the source of signals shall be at least 10 dB below the lowest level of distortion to be measured.
2. To correctly measure the distortion of a system, one must consider the distortion introduced by the test setup. For that purpose, it is good practice to calibrate the setup before testing the system under test (SUT). A measurement on the setup alone provides values that are used as correction factors.
3. The signal analyzer is often a spectrum analyzer. Network analyzers are also very much in use, particularly for testing RF and microwave devices (4). Available spectrum and network analyzers have three useful features: (a) an internal oscillator that may be used to excite the SUT; (b) digital interfaces for remote control that allow automated measurement procedures; (c) an internal CPU useful for reducing data and presenting distortion parameters.
4. Delay distortion is expressed by a parameter that is a function of the derivative of the phase with respect to frequency $d\theta/d\omega$. Several methods leading to different parameters are in use (4,5). One method very commonly implemented in network analyzers consists of measuring $d\theta/d\omega$ at two frequencies, one of which is a reference. Then delay distortion is expressed as the difference of the two derivatives.

   Measuring the phase shift introduced by a system requires a reference signal. This may be a problem when testing tuners, for instance. The AM/FM-delay method for group delay measurement (4) is one solution to overcoming that difficulty.

   Another delay distortion measurement problem arises when the input and output of the system are far apart as is the case with some communications systems. Solutions for this problem are discussed in (5).

Apart from the distortion due to the nonlinearity or the nonideal frequency response of a system considered in this article, a signal may be distorted because of the interference of signals of the same system, for example, cross talk and cross-modulation. All of these types of distortions are included in what we designate an intrasystem distortion. A signal in a system, however, may be distorted by a signal from another system. This type of distortion that is designated as intersystem distortion is produced when a coupling between the two systems exists either by conduction or radiation. Intersystem distortion measurement and evaluation is an electromagnetic interference (EMI) or electromagnetic compatibility (EMC) problem and is thus beyond the scope of the present article. EMI/EMC is presently an extremely important domain of electrical engineering and it will be even more important in the future. The proper operation of electrical and electronic equipment requires increased attention to interference and susceptibility issues. Interested readers may refer to Refs. 6 and 7.

Legislation and standards for measurements on EMI/EMC have been produced all over the world. References 8–14, are examples of standards that may assist the reader in evaluating intersystem distortion. For a more detailed list, readers should consult Ref. 7.

BIBLIOGRAPHY

1. Standard IEC 268-2, Sound system equipment—explanation of general terms, International Electrotechnical Commission, Geneva, Switzerland, 1987.
2. R. L. Freeman, *Telecommunications Transmission Handbook,* New York: Wiley, 1991.
3. Spectrum Analysis Basics, Application Note 150, Hewlett-Packard Company, Palo Alto, CA, 1989.
4. RF and microwave device test for the 90s, *Seminar Papers,* Hewlett-Packard, Palo Alto, CA, 1995.
5. B. M. Oliver and J. M. Cage, *Electronic Measurements and Instrumentation,* New York: McGraw-Hill, 1975.
6. B. E. Keyser, *Principles of Electromagnetic Compatibility,* Norwood, MA: Artech House, 1985.
7. V. P. Kodali, *Engineering Electromagnetic Compatiblity,* Principles, Measurements and Technology, Piscataway, NJ: IEEE Press, 1996.
8. Standard EN 50082-1, *Electromagnetic compatibility general immunity standard.* Part 1: Residential, commercial and light industry, Electrotechnical Standardization European Committee, Brussels, Belgium.
9. Standard EN 50082-2, *Electromagnetic compatibility general immunity standard.* Part 2: Industrial environment, Electrotechnical Standardization European Committee, Brussels, Belgium.
10. Standard CISPR 20, *Limits and methods of measurement of the immunity characteristics of radio broadcast and television receivers and associated equipment,* International Electrotechnical Commission, Geneva, Switzerland, 1990.
11. Standard CISPR 22, *Limits and methods of measurement of radio disturbance characteristics of information technology equipment,* International Electrotechnical Commission, Geneva, Switzerland, 1993.
12. Standard IEC 555-2, *Disturbances in supply systems caused by household appliances and similar electrical equipment.* Part 2: Harmonics, International Electrotechnical Commission, Geneva, Switzerland, 1982.
13. Standard IEC 555-3, *Disturbances in supply systems caused by household appliances and similar electrical equipment.* Part 3: Voltage fluctuations, International Electrotechnical Commission, Geneva, Switzerland, 1982.
14. Standards IEC 1000-4-x, *EMC: Test and measurement techniques,* International Electrotechnical Commission, Geneva, Switzerland.
15. S. Haykin, *Communication Systems,* New York: Wiley, 1994.
16. Joseph J. Carr, *Elements of Electronic Instrumentation and Measurement,* 3rd ed., Englewood Cliffs, NJ: Prentice-Hall, 1996.
17. A. D. Helfrick and W. D. Cooper, *Modern Electronic Instrumentation and Measurement Techniques,* Englewood Cliffs, NJ: Prentice-Hall, 1990.

PEDRO M. B. SILVA GIRÃO
ANTÓNIO M. CRUZ SERRA
HELENA M. GEIRINHAS RAMOS
Instituto Superior Técnico

ELECTRIC FIELD IONIZATION. See FIELD IONIZATION.

ELECTRIC FILTERS. See ELLIPTIC FILTERS; FILTERING THEORY; NONLINEAR FILTERS.