multiplying circuit[1] for Q_1 as shown in Fig. 10(b). The feedback voltage must be taken from the emitter of Q_2, however, because Q_1 will not tolerate any additional loading. This system will provide about 6 db more gain at the expense of 50 per cent more distortion and some variation of power gain with transistor parameters. If the signal source has a high impedance, this approach may have merit.

Conclusion

The supporting data in this report show that a single-ended two-watt amplifier, which is very noncritical of transistor parameters and which has performance acceptable for high-fidelity applications, is practical. High-temperature I_{CO} in the power transistor must be controlled, however, or performance is seriously degraded.

Nonlinear Distortion Reduction by Complementary Distortion*

J. ROSS MACDONALD†

Summary—**Nonlinear distortion produced in a given circuit can be reduced by pre- or postdistorting the signal applied to or from the circuit. Such complementary distortion cannot reduce the original distortion to zero in practice because of distortion of distortion, but it can result in greatly reduced output distortion over a limited amplitude range. General results for the design of pre- or postdistortion circuits are given, and the mathematical results are illustrated by comparing the total harmonic distortions obtained with pre- and postdistortion corrections of increasing complexity applied to a simple nonlinear circuit.**

Introduction

The correction of an undesired frequency distortion, such as a droop in loudspeaker output at low frequencies, by means of a complementary response in the applied signal is well known and often used. Somewhat less well known and understood is the corresponding technique for reducing nonlinear distortion. It is usually stated or implied[1,2] that nonlinear distortion such as that arising from the response characteristic of Fig. 1(a) can be cancelled by passing the distorted signal through a circuit having the complementary response characteristic of Fig. 1(b). For example, it is often expected that if the distortion arises only from a square-law term, it may be completely cancelled by subsequent transmission through a network yielding square-law distortion of equal magnitude but opposite sign. The present work shows that complete cancellation is impossible because of distortion of the original distortion and that over-all distortion reduction is only possible over a limited range of input signal amplitude.

Negative feedback is commonly regarded as the great panacea for distortion. Nevertheless, there are instances where its application for nonlinear distortion reduction is inconvenient or impossible. Such instances often occur at the beginning or end of a signal transmission system. In the audio field, it is difficult to generate an error signal to correct any nonlinear distortion arising in a record pickup. Loudspeaker nonlinear distortion, at the opposite end of the system, is usually more important because of its greater magnitude. Because of reverberation and phase shifts, it is not generally practical to apply negative feedback between the sound output of a loudspeaker and its driver. On the other hand, feedback derived from a separate winding on the voice coil will be imperfectly related to the actual sound output. In this instance, where negative feedback is impractical or inefficient, complementary nonlinear distortion can greatly reduce the distortion present in the speaker output.

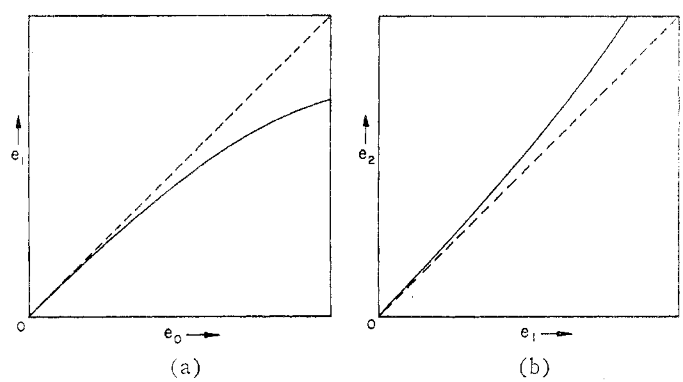

Fig. 1—(a) Typical input-output characteristic for a nonlinear circuit. (b) Input-output characteristic complementary to that of (a).

There are two ways by which complementary distortion correction may be applied. The usual way, which will be designated postdistortion, is that illustrated in Fig. 1. Here the complementary distortion acts on the originally distorted signal. Comparable but not identical results can be obtained, however, if the signal is first in-

* Manuscript received by the PGA, July 6, 1959.
† Texas Instruments Incorporated, Dallas, Tex.
[1] P. A. Reiling, U. S. Patent No. 2,293,628, issued August 18, 1942.
[2] G. Guanella, U. S. Patent No. 2,776,410, issued January 1, 1957. A means alternative to complementary distortion correction is described in this patent.

tentionally predistorted and then passed through the original distorting circuit. Correction of loudspeaker nonlinear distortion is an instance where pre- but not postdistortion is applicable.

The practical application of complementary distortion requires that the original distortion dependence on amplitude be known over the amplitude range of interest. If the input-output response characteristic is very irregular or has strong discontinuities in slope or value, it cannot be well represented by a rapidly convergent power series, and complementary distortion correction will not be practical. It will therefore be assumed that the input-output characteristic is a smooth function and can be represented by a power series of the form

$$e_1 = \sum_{m=1}^{M} a_m e_0^m, \tag{1}$$

where M may be finite or infinite, e_0 is the input signal and e_1 the output. The zero-order, or dc term, has been omitted for simplicity.

In general, the a_m coefficients will be functions of frequency, and the complementary distortion circuit will then also have to be frequency dependent to yield maximum distortion reduction over the amplitude and frequency ranges of interest. In practice, however, the coefficients may be frequency independent over much of the range. While it is possible to incorporate frequency-sensitive nonlinear correcting elements in the complementary distortion circuit to compensate for the frequency dependence of the original distortion, such complications will not be pursued further herein.

The coefficients of the input-output-characteristic power series must be known to allow design of the complementary distortion circuit. Haber and Epstein[3] have given equations which allow these coefficients to be calculated from the results of harmonic distortion measurements together with measurements of the polarity of the harmonics. As shown below, these coefficients may then be used to determine the corresponding coefficients in the power series describing the pre- or post-complementary distortion circuit. Finally, the resulting nonlinear characteristic can be realized in a practical circuit using diodes and resistors and other components and techniques well known in the analog computer art. It should be emphasized that both pre and post-distortion techniques are also applicable when the aim is not as linear amplification as possible but instead a close approximation to some more complicated functional relationship between input and output, such as, for example, square-law output with minimum linear and higher-than-second-order output terms.

Mathematical Analysis

For postdistortion, (1) may be used to represent the characteristic of the device or circuit whose distortion is to be reduced by a subsequent complementary distortion circuit. The characteristic of the latter may then be written

$$e_2 = \sum_{n=1}^{N} b_n e_1^n. \tag{2}$$

Here the b_n's must be determined in terms of the a_m's to minimize resultant distortion. For convenience, in the predistortion case (1) will be used to represent the initial complementary distortion while (2) will then describe the original distorting device or circuit. Thus, in this case, the b_n's are assumed known and the a_m's are to be determined as functions of them.

Substituting (2) in (1) yields

$$e_2 = \sum_{n=1}^{N} b_n \left[\sum_{m=1}^{M} a_m e_0^m \right]^n \equiv \sum_{s=1}^{NM} c_s e_0^s, \tag{3}$$

where the c_s's are new coefficients whose values, determined from (3), appear in the second column of Table I. Now for zero output distortion in e_2, we desire $e_2 = c_1 e_0$. To obtain the values of a_m or b_n which make the higher order c_s's zero, we can set these c_s's to zero and solve them individually to obtain the desired a_m's for predistortion or the b_n's for postdistortion. This procedure becomes very arduous as the order increases, and a preferable method is to use reversion of series.[4]

On setting $e_2 = c_1 e_0$ equal to (2) and solving for e_0, we obtain

$$e_0 = \frac{1}{c_1} \sum_{n=1}^{N} b_n e_1^n. \tag{4}$$

If now (1) is reverted to yield e_0 in terms of e_1, one obtains (see Appendix) an infinite series like (4) and comparison of terms yields the postdistortion b_n's directly in terms of the a_m's. The results up to fifth order are given in the third column of Table I. Note that when less than an infinite number of correction terms are used, there will remain residual distortion which, however, may be much lower than that originally present.

A similar procedure can be carried out for predistortion. Eq. (4) may be written as

$$c_1 e_0 = \sum_{n=1}^{N} b_n e_1^n.$$

Next, this series may be reverted to yield e_1 in terms of $(c_1 e_0)$ and the result compared with (1). Using $c_1 = a_1 b_1$, the equations of column 4 of Table I are obtained. Columns 3 and 4 of Table I are the basic results of the present work. Although expressions connecting the coefficients have only been stated to the fifth order, higher-order terms may be readily obtained from the corresponding known expressions tabulated in the reversion of series method.[4]

[3] F. Haber and B. Epstein, "The parameters of nonlinear devices from harmonic measurements," IRE Trans. on Electron Devices, vol. ED-5, pp. 26–28; January, 1958.

[4] H. B. Dwight, "Tables of Integrals and Other Mathematical Data," 3rd Ed., The Macmillan Co., New York, N. Y., p. 11; 1947.

TABLE I

EXPRESSIONS FOR c_i, b_i, AND a_i

i	c_i	Postdistortion, $c_i = 0$ b_i	Predistortion, $c_i = 0$ a_i
1	a_1b_1	b_1	a_1
2	$a_2b_1 + a_1^2b_2$	$-\left(\frac{b_1}{a_1^2}\right) a_2$	$-\left(\frac{a_1^2}{b_1}\right) b_2$
3	$a_3b_1 + 2a_1a_2b_2 + a_1^3b_3$	$\left(\frac{b_1}{a_1^4}\right)(2a_2^2 - a_1a_3)$	$\left(\frac{a_1^3}{b_1^2}\right)(2b_2^2 - b_1b_3)$
4	$a_4b_1 + (a_2^2 + 2a_1a_3)b_2 + 3a_1^2a_2b_3 + a_1^4b_4$	$\left(\frac{b_1}{a_1^6}\right)[5a_2(a_1a_3 - a_2^2) - a_1^2a_4]$	$\left(\frac{a_1^4}{b_1^3}\right)[5b_2(b_1b_3 - b_2^2) - b_1^2b_4]$
5	$a_5b_1 + [2(a_1a_4 + a_2a_3)]b_2 + [3(a_1^2a_3 + a_1a_2^2)]b_3 + [4a_1^3a_2]b_4 + a_1^5b_5$	$\left(\frac{b_1}{a_1^8}\right)[6a_1^2a_2a_4 + 3a_1^2a_3^2 + 14a_2^4 - a_1^3a_5 - 21a_1a_2^2a_3]$	$\left(\frac{a_1^5}{b_1^4}\right)[6b_1^2b_2b_4 + 3b_1^2b_3^2 + 14b_2^4 - b_1^3b_5 - 21b_1b_2^2b_3]$

The above reversion procedure shows that for either pre- or postdistortion correction to yield $e_2 = c_1e_0$ exactly, an infinite number of complementary correction terms will be required. For postdistortion, for example, each correction term acts on the original distortion to create higher-order distortion terms which, in turn, require the presence of higher-order correction terms to eliminate them and so on. Further, the larger the number of correction terms, the smaller in general the amplitude range over which reduced distortion is obtained. Nevertheless, a finite number of pre- or postcorrection terms can effect a very significant improvement in nonlinear distortion over a finite and important amplitude range.

EXAMPLE

The square-law distortion case, being simplest, will be used to show how complementary distortion may be applied for over-all distortion reduction. This case is useful also as an example since it can be solved directly, as shown in the Appendix. For predistortion, we shall take $N=2$ so that $e_2 = b_1e_1 + b_2e_1^2$, while $M=2$ for postdistortion yields $e_1 = a_1e_0 + a_2e_0^2$. As an illustration, we shall investigate in both cases how the residual total harmonic distortion (THD) varies with normalized amplitude when $c_2 = 0$ only (one correction term), when $c_2 = c_3 = 0$, and when $c_2 = c_3 = c_4 = 0$.

For both pre- and postdistortion, the procedure is to substitute the values of a_m or b_n which make the desired c_s's zero into the higher-order c's and thus evaluate the residual, nonzero distortion. The results of such a calculation with no distortion terms omitted are summarized in Table II. For convenience, we have set $a_1 = b_1 = 1$ in the results of Table II. The quantities a_1 and b_1 are merely scale factors such as amplification factors, and no significant generality is lost by taking them unity. It should be emphasized that these results are pertinent only to the square-law distortion case, as shown by the appearance of only a_2 and b_2. More complicated results would be obtained for cubic distortion or for a combination of quadratic and cubic distortion.

TABLE II

EXPRESSIONS FOR NONZERO c_i'S IN VARIOUS PRE- AND POSTDISTORTION CASES

	$c_2=0$		$c_2=c_3=0$		$c_2=c_3=c_4=0$	
	Pre	Post	Pre	Post	Pre	Post
c_3	$-2b_2^2$	$-2a_2^2$	0	0	0	0
c_4	b_2^3	$-a_2^3$	$5b_2^3$	$5a_2^3$	0	0
c_5	0	0	$-4b_2^4$	$6a_2^4$	$-14b_2^4$	$-14a_2^4$
c_6	0	0	$4b_2^5$	$2a_2^5$	$14b_2^5$	$-28a_2^5$
c_7	0	0	0	0	$-20b_2^6$	$-20a_2^6$
c_8	0	0	0	0	$25b_2^7$	$-5a_2^7$
c_9	0	0	0	0	0	0

Next, the results of Table II together with

$$e_2 = \sum_1 c_s e_0^s$$

may be used to obtain the harmonic distortion terms in each case. If one takes $e_0 = A \cos \omega t$ and expresses all powers of $\cos \omega t$ as harmonic terms, one finally obtains the finite series of harmonics pertinent to each case. For example, for $c_2 = 0$ and predistortion, the result is

$$e_2 = e_0 - 2b_2^2e_0^3 + b_2^3k_0^4, \tag{5}$$

or

$$e_2 = [A - (3b_2^2A^3/2)] \cos \omega t + (b_2^3A^4/2) \cos 2\omega t - (b_2^2A^3/4) \cos 3\omega t + (b_2^3A^4/8) \cos 4\omega t, \tag{5}$$

where dc terms have again been omitted.

In general, we may write

$$e_2 = \sum_{s=1} c_s e_0^s \equiv \sum_{r=1} h_r \cos r\omega t \equiv A \sum_{r=1} g_r \cos r\omega t, \tag{6}$$

where the h_r's are the harmonic amplitudes and $g_r = h_r/A$. Next, let $x = Aa_2 \equiv Aa_2/a_1$ and $x = Ab_2 \equiv Ab_2/b_1$ for post- or predistortion, respectively. Here x is a normalized distortion amplitude variable, as shown by the result

$$\begin{aligned} e_1 &= a_1 e_0 + a_2 e_0^2 = a_1 e_0[1 + (a_2/a_1)e_0] \\ &= a_1 A \cos \omega t[1 + x \cos \omega t]. \end{aligned} \tag{7}$$

In the present case, x clearly measures the relative importance of the original square-law or second-harmonic distortion term. For small x, in fact, x equals twice the original, uncorrected THD. The normalized harmonics, g_r, can now be expressed entirely in terms of x and the results are summarized in Table III.

Rather than compare the predictions of these different cases by comparing individual harmonic behavior graphically, the THD's of the various cases will be compared as functions of x. We may write

$$\text{THD} = \left[\frac{\sum_{r=2} h_r^2}{\sum_{r=1} h_r^2}\right]^{1/2} = \left[\frac{\sum_{r=2} g_r^2}{\sum_{r=1} g_r^2}\right]^{1/2}. \tag{8}$$

Note that g_1 may reach zero in most of the cases of Table III. In such cases, the THD will be unity and there will be no fundamental component present. The THD is sometimes written as

$$\text{THD}' = \left[\sum_{r=2} h_r^2/h_1^2\right]^{1/2}, \tag{9}$$

and would be infinite at the points where h_1 or g_1 were zero. The distinction between the two forms is only important for high values of THD anyway, and this is generally not the region of most physical interest.

TABLE III

NORMALIZED HARMONIC AMPLITUDES AS FUNCTIONS OF x FOR VARIOUS CASES

No Corrections		$c_2 = 0$		$c_2 = c_3 = 0$		$c_2 = c_3 = c_4 = 0$	
Case	A	B Pre	C Post	D Pre	E Post	F Pre	G Post
g_1	1	$1 - \frac{3}{2}x^2$	$1 - \frac{3}{2}x^2$	$1 - \frac{5}{2}x^4$	$1 + \frac{15}{4}x^4$	$1 - \frac{5x^4}{16}(28 + 35x^2)$	$1 - \frac{5x^4}{16}(28 + 35x^2)$
g_2	$\frac{x}{2}$	$\frac{x^3}{2}$	$-\frac{x^3}{2}$	$\frac{5}{8}x^3(4 + 3x^2)$	$\frac{5x^3}{16}(8 + 3x^2)$	$\frac{35}{16}x^5(3 + 5x^2)$	$-\frac{35}{16}x^5(6 + x^2)$
g_3	0	$-\frac{x^2}{2}$	$-\frac{x^2}{2}$	$-\frac{5}{4}x^4$	$\frac{15}{8}x^4$	$-\frac{35}{16}x^4(2 + 3x^2)$	$-\frac{35}{16}x^4(2 + 3x^2)$
g_4	0	$\frac{x^3}{8}$	$-\frac{x^3}{8}$	$\frac{x^3}{8}(5 + 6x^2)$	$\frac{x^3}{8}(5 + 3x^2)$	$\frac{x^5}{32}(84 + 175x^2)$	$\frac{x^5}{32}(168 + 35x^2)$
g_5	0	0	0	$-\frac{x^4}{4}$	$\frac{3}{8}x^4$	$-\frac{x^4}{16}(14 + 35x^2)$	$-\frac{x^4}{16}(14 + 35x^2)$
g_6	0	0	0	$\frac{x^5}{8}$	$\frac{x^5}{16}$	$\frac{x^5}{16}(7 + 25x^2)$	$-\frac{x^5}{16}(14 + 5x^2)$
g_7	0	0	0	0	0	$-\frac{5}{16}x^6$	$-\frac{5}{16}x^6$
g_8	0	0	0	0	0	$\frac{25}{128}x^7$	$-\frac{5}{128}x^7$
g_9	0	0	0	0	0	0	0

The dependence on x of per cent THD calculated from (8) with the help of Table III is shown in Fig. 2. The cases shown pertain to predistortion only. There is no difference between the pre- and postdistortion THD's in cases B and C. The limiting slopes for small x are, reading upwards, 4, 3, 2, and 1. Note that F yields minimum THD for $x<0.31$, while D is better than BC for $x<0.21$, and A (no correction) is better than BC for $x>0.53$. In general, the more correction terms present, the smaller the value of x below which the corrected THD is less than the uncorrected value. Nevertheless, when x is small, several correction terms can greatly reduce the over-all distortion. For an uncorrected THD of 2 per cent, three-term predistortion correction (case F) reduces the THD to about 0.0011 per cent.

Finally, Fig. 3 shows a comparison between the pre- and postdistortion cases for $c_2=c_3=0$ and for $c_2=c_3=c_4=0$. The quantity R is the ratio of the postdistortion THD to that for predistortion. These curves show that postdistortion results in somewhat lower THD than predistortion when two correction terms are used, but that the reverse is the case when three correction terms are used. In the region of considerable distortion improvement however, say for $x\leq 0.2$, there is no significant difference between the pre- and postdistortion corrections. Thus, which of the two methods to use in practical cases can be determined solely on the basis of applicability or simplicity.

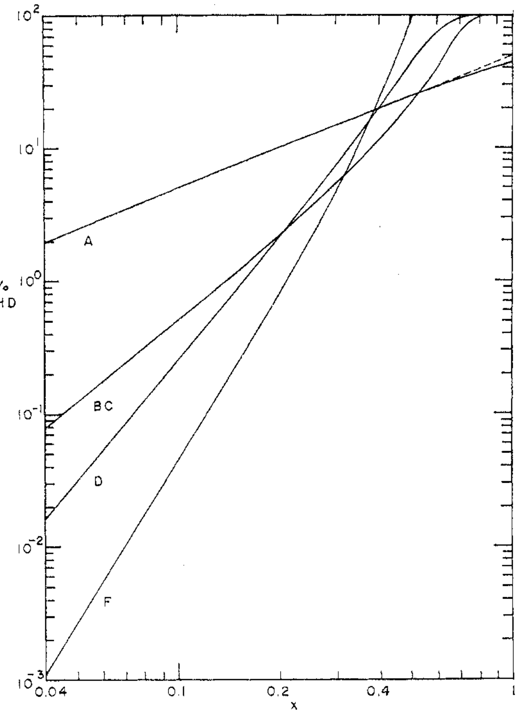

Fig. 2—Log-log plots of per cent total harmonic distortion in various cases (defined in Table III) vs the normalized distortion variable x for original square-law distortion only.

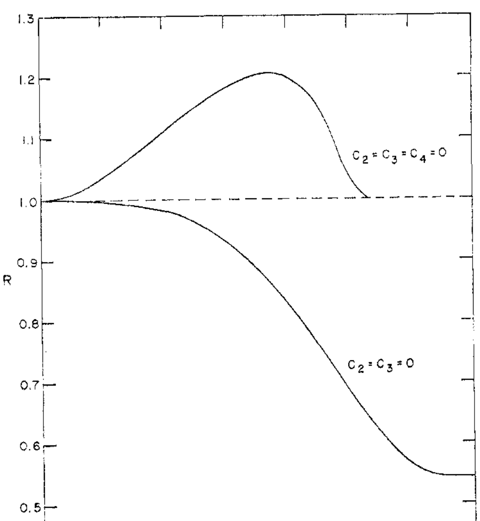

Fig. 3—Total harmonic distortion with postdistortion divided by that with predistortion vs x.

Appendix

Given simple square-law distortion of the form $e_1=a_1e_0+a_2e_0^2$, we can revert this finite series directly to obtain

$$e_0 = -\frac{a_1}{2a_2} + \left[\left(\frac{a_1}{2a_2}\right)^2 + \frac{e_1}{a_2}\right]^{1/2}. \tag{10}$$

If $|4a_2e_1/a_1^2|<1$, expansion yields

$$e_0 \cong (1/a_1)e_1 - (a_2/a_1^3)e_1^2 + (2a_2^2/a_1^5)e_1^3 \cdots. \tag{11}$$

Comparison with (4) yields $c_1=a_1b_1$ and the values of b_2 and b_3 given in the postdistortion column of Table I (with $a_3\equiv 0$).

The series expansion in (11) on which the reversion solution is based is only convergent for $|4a_2e_1/a_1^2|<1$. This condition may be written

$$4\left| a_2(e_0/a_1) + a_2^2(e_0/a_1)^2 \right| < 1, \tag{12}$$

or

$$\left| (a_2A/a_1)\cos\omega t + (a_2A/a_1)^2\cos^2\omega t \right| < \tfrac{1}{4}. \tag{12a}$$

Now the only maxima of the left-hand side, considered as a function of (ωt), which will satisfy the inequality is that obtained for $\cos\omega t=0$. Thus, the most restrictive condition following from (12a) is

$$\left| x + x^2 \right| < \tfrac{1}{4}. \tag{12b}$$

Replacing the inequality by an equality and solving for x yields $x=(\sqrt{\tfrac{1}{2}}-\tfrac{1}{2})=0.207$. This result shows that in the present case the reversion solution with an infinite number of power-law correction terms is convergent for $x<0.207$. For smaller x, an infinite number of complementary distortion terms may be used to correct completely (in principle) for the original distortion. When $x\geq 0.207$, complementary distortion correction with an infinite number of correction terms is impossible, but Fig. 2 shows that some improvement with a finite number of terms is still possible for x somewhat greater than 0.207.

In the present simple case, there is an alternative to the reversion solution which will allow complete distortion correction for a different range of x. We wish to obtain $e_2=c_1e_0$, with all higher-order terms zero. From (10), this equation may be written

$$e_2 = c_1\left[\frac{-a}{2a_2} + \left\{\left(\frac{a_1}{2a_2}\right)^2 + \frac{e_1}{a_2}\right\}^{1/2}\right]. \qquad (13)$$

If c_1, a_1, and a_2 are known, the output e_1 from the nonlinear circuit whose nonlinearity is to be corrected may be passed through an analog computer which operates on e_1 in accordance with (13) to produce the undistorted output $e_2=c_1e_0$. Eq. (13) can only be applied when the radicand is not less than zero. This condition leads to

$$(4x \cos \omega t)[1 + x \cos \omega t] \geq -1. \qquad (14)$$

The most restrictive condition is $\cos \omega t = -1$, leading to $4x-4x^2=1$. The solution of this equation is $x=\frac{1}{2}$, the maximum value permitted.

It should be emphasized that a closed-form reversion of the type illustrated by (10) is only possible in the simplest cases. In general, the series-reversion method must be used with a finite number of correction terms, leaving finite residual distortion. Further, as the complexity of the distortion to be corrected increases, it is likely that the range over which effective distortion reduction can be produced will diminish.

Acknowledgment

The author is most grateful to D. R. Powell for checking the complicated algebra of this work.