The Theory of Transient Intermodulation Distortion

MATTI OTALA AND EERO LEINONEN

Abstract—The existing theory of transient intermodulation distortion (TIM) is extended to cover the calculation of the duration of intermodulation bursts. It is shown that feedback values in excess of some 40 dB will cause large internal overshoots within the amplifier. The clipping of these overshoots due to the limited dynamic margins of the amplifier driver stages is shown to give rise to long periods during which the amplifier is in cut-off condition. The duration of these periods is calculated and the mathematical results are verified with digital and analogue simulation. Finally, the relationships of TIM, slew rate, and power bandwidth are discussed.

INTRODUCTION

THE application of strong negative feedback in audio amplifiers has become standard practice over the last decade because transformerless transistor amplifier circuits have enabled its easy use.

In the tube amplifier era it was the output transformer which, due to its complex transfer function, precluded the use of much more than some 20-30 dB feedback. Even then, amplifier designers were discussing a particular "veil" which seemed to appear in the sound when the feedback was very strong. At the same time, amplifiers having small-signal frequency ranges greater than 1 MHz were advocated as being the only ones giving adequate sound quality. The theory presented here will justify these conclusions to some extent.

Later, the inherent nonlinearity of early transistor amplifiers forced designers to use strong negative feedback to cope with the requirements of low harmonic and intermodulation distortion. Even when the importance of this basic reason decreased in later years, the apparent ease of the use of feedback as a cure-all for almost all amplifier sicknesses continued and commercial power amplifiers having some 60–100 dB feedback are not rare today.

At the same time, the debate of "transistor sound" versus "tube sound" continued and has become an object of intensive research. It should be strange in principle that a tube amplifier and a transistor amplifier having equal performances and specifications which are much better than necessary for the ear, should sound remarkably different. The only solutions of this apparent dilemma seem to be that

1) the present amplifier measurements are partly irrelevant in respect to the audible amplifier characteristics; and/or
2) the present amplifier measurement methods do not reveal all of the major sound degradation effects [10].

Recently, at least one basic distortion mechanism, which

does not appear with present amplifier measurement methods, the transient intermodulation distortion (TIM), has been discovered. It is a side effect of the use of a too strong negative feedback, which makes the modern transistor amplifiers, in particular, very susceptible to it.

The basic theory of TIM is straightforward and clear [1]. TIM has been shown to be relatively common in commercial audio amplifiers [2] and the ear seems to be very sensitive to it [3]. Methods for its elimination have been outlined [4] and several "TIM-free" amplifiers have been constructed [5]–[8]. An intensive public discussion of the effect has recently begun in several countries.

The purpose of this paper is to extend the theory of TIM to the calculation of the amplifier cut-off time, to explain some of the basic behavioral models for this phenomenon, and to try to establish practical limits to the application of feedback in audio amplifiers. The relationships of TIM and some other amplifier properties will also be discussed shortly.

BASIC CONFIGURATION

We will examine the basic feedback amplifier circuit of Fig. 1. Here β is the purely resistive feedback path around the amplifier A which has an open-loop gain A_0. C is the preamplifier which is imagined to incorporate the transfer function of the signal source, so that for the purpose of the analysis the input signal V_1 may be assumed to have infinite bandwidth. It is essential to note that though V_1 itself may have signal components of very high frequency, the preamplifier C acts as a low-pass filter, moderating the signal in such a way that the amplifier A input signal V_2 includes only components in the usual audio passband.

The preamplifier frequency response is assumed to be that shown in Fig. 2. Two alternative responses are shown, one corresponding to linear response and the second corresponding to treble boost.

The power amplifier frequency response is assumed to have the form of Fig. 3. Because of stability considerations, idealized -6 dB/octave slopes are assumed in the open-loop response. This may not always be true in practical amplifiers where the phase margin is not always $\pi/2$, and where different frequency compensation techniques may be combined to shape the response of Fig. 3. The results of the analysis may, however, be used as general guidelines, and the effects of the departures from the idealized model should be carefully analyzed separately.

When feedback β is applied in the amplifier, the gain of the amplifier decreases from A_0 to

$$A = A_0/(1 + \beta A_0) \tag{1a}$$

and the open-loop upper cut-off frequency ω_1 appears to be
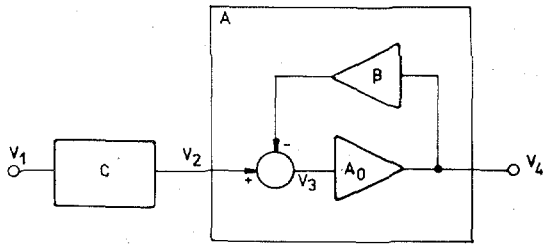
Fig. 1. The basic amplifier configuration. A is the power amplifier having an open-loop gain of A_0, β is the purely resistive feedback path around it, and C is the preamplifier.
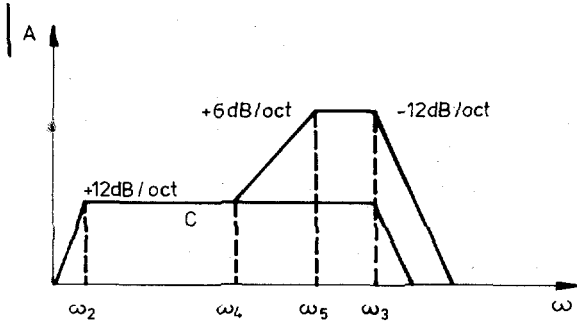

Fig. 2. The preamplifier frequency responses used in the analysis. The straight curve corresponds to linear frequency response and the peaked one corresponds to treble boost. The amount of treble boost used in the analysis is $\epsilon = \omega_4/\omega_5 = 0.05$.

increased to

$$\omega_7 = \omega_1(1 + \beta A_0). \tag{1b}$$

For the sake of simplicity, it is assumed that the lower cut-off frequencies equal zero, i.e.

$$\omega_2 = \omega_0 = \omega_6 \approx 0. \tag{2}$$

The effect of this assumption on the results is negligible if the upper and lower cut-off frequencies are sufficiently far apart, as is always the case in high-fidelity audio reproduction.

Provided that $\omega_1 < \omega_3$, i.e., the amplifier A open-loop frequency response is smaller than the preamplifier C frequency response, the voltages $V_1 \cdots V_4$ in Fig. 1 assume the form shown in Fig. 4 for a step function preamplifier input voltage

$$V_1 = v_1/s. \tag{3}$$

The main amplifier input voltage V_2 then has a rise-time governed by C. However, there is a marked overshoot in V_3, caused by the tendency of the feedback to compensate for the intrinsic slowness ω_1 of the amplifier. It is this overshoot that in fact decreases the apparent small-signal rise-time of the feedback amplifier.

DEFINITION OF TIM

The overshoot in V_3 may be up to a thousand times greater in amplitude than the nominal input signal. It may be substantially greater than the internal dynamic margin of the amplifier and is therefore prone to be suppressed or clipped
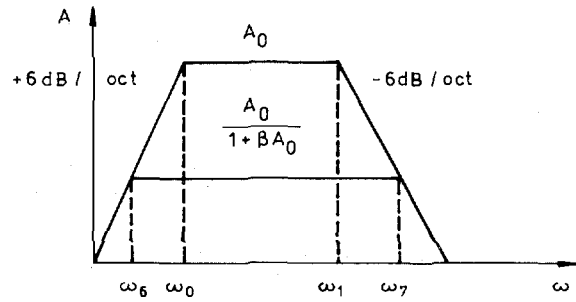

Fig. 3. The power amplifier open-loop frequency response (upper curve) is modified by feedback (lower curve). -6 dB/octave slopes are assumed because of stability considerations.

within the amplifier, usually in the stage preceding the compensation. During this suppression, the amplifier gain for other simultaneous signals will decrease, causing momentary intermodulation distortion bursts. These momentary intermodulation bursts are called TIM because they have the appearance of one signal influencing the amplitude of another ($=$ intermodulation), and are caused by the time and amplitude characteristics of the input signal rather than by the amplitude characteristics alone, as is the case with ordinary intermodulation distortion.

BOUNDARY CONDITIONS

We will denote the following:

$$\gamma = \omega_1/\omega_3 \tag{4a}$$

and

$$\alpha = 1 + \beta A_0. \tag{4b}$$

In addition, we will take two auxiliary normalized parameters

$$\epsilon = \omega_4/\omega_5 \tag{4c}$$

denoting the treble boost, and

$$T = \omega_3 t \tag{4d}$$

where t is the real time and T the normalized time.

To find the relevant α-γ combinations for practical amplifiers, we may stipulate that a minimum requirement for any high-quality audio amplifier will be a closed-loop cut-off frequency ω_7 of, say, 30 kHz. The maximum value of the closed-loop cut-off frequency ω_7 depends on the quality of the circuit and its components, but 1 MHz seems to be already relatively difficult, albeit possible to obtain in practice. These boundary values govern the frequency compensation applied in the amplifier in order to ensure stability.

Assuming that the preamplifier upper cut-off frequency is

$$\omega_3/2\pi = 30 \text{ kHz}$$

we obtain explicitly

$$1 \leqslant \alpha\gamma \leqslant 33.3. \tag{5}$$

This leads to Fig. 5 which depicts the possible values of the γ as function of the feedback α. The following analysis is carried out using the circled values of the α-γ combinations.
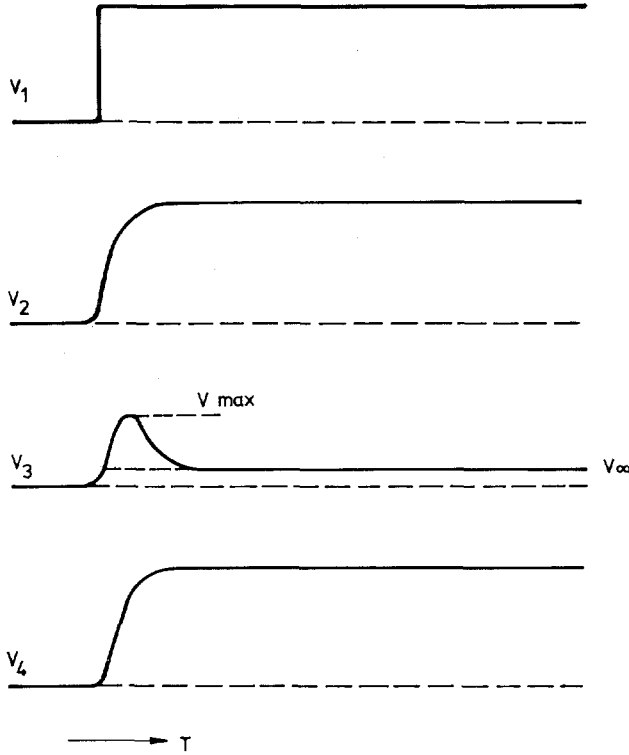
Fig. 4. The response of the circuit of Fig. 1 for a unit step function input v_1/s under the assumptions made in the text.

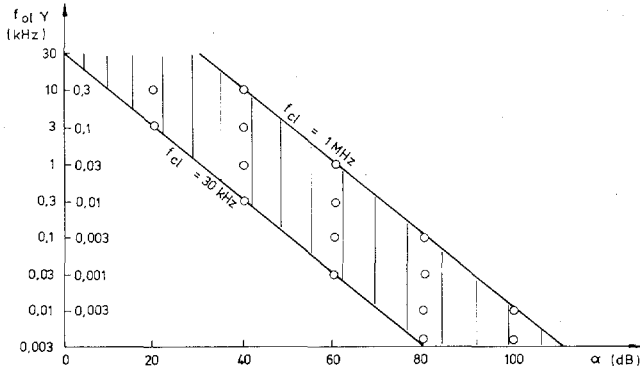Fig. 5. The possible values of the open-loop upper cut-off frequency f_{01} as function of feedback α. The upper line corresponds to a closed-loop frequency response f_{01} of 1 MHz and the lower line to that of 30 kHz. In converting these responses to γ-values, an upper cut-off frequency of 30 kHz has been assumed for the preamplifier. The circles denote the α-γ-pairs used in the analysis.

THE OVERSHOOT PHENOMENON

Based on the previous assumptions, assuming a voltage gain of unity for the preamplifier, and taking $\omega_5 = \omega_3$, the internal drive voltage V_3 of amplifier A becomes

$$V_3(s) = \frac{(s + \omega_1)\omega_3^2 \, V_1(s)}{[s + (1 + \beta A_0)\omega_1] \, (s + \omega_3)^2} \quad (6a)$$

for the linear frequency response, and

$$V_3(s) = \frac{(s + \omega_1)\omega_3^3 (s + \omega_4)V_1(s)}{[s + (1 + \beta A_0)\omega_1] \, \omega_4(s + \omega_3)^3} \quad (6b)$$

for the treble boost case.

With a step function input voltage (3), these voltages are in the normalized time domain

$$V_3(T) = \frac{v_1}{\alpha} = \left\{ 1 + \frac{(\alpha - 1) \exp(-\alpha\gamma T)}{(1 - \alpha\gamma)^2} \right.$$
$$\left. + \left[\frac{(\gamma - 1)T}{1 - \alpha\gamma} - \frac{1 + \gamma(\alpha\gamma - 2)}{(1 - \alpha\gamma)^2} \right] \alpha \exp(-T) \right\} \quad (7a)$$

for the linear frequency response, and

$$V_3(T) = \frac{v_1}{\alpha} \left\{ 1 - \exp(-T) + \frac{(\alpha - 1)(\epsilon - \alpha\gamma)}{\epsilon(1 - \alpha\gamma)^3} [\exp(-\alpha\gamma T) \right.$$
$$- \exp(-T)] + \left[\frac{(\gamma - 1)(\epsilon - 1)T^2}{2\epsilon(1 - \alpha\gamma)} \right.$$
$$\left. - \frac{(\alpha\gamma - 1)(\gamma\epsilon - 1) - (\gamma - 1)(\epsilon - 1)}{\epsilon(1 - \alpha\gamma)^2} T \right] \alpha$$
$$\left. \cdot \exp(-T) \right\} \quad (7b)$$

for the treble boost case. Note that these equations differ from those given in the previous paper [1] due to a more realistic assumption of the preamplifier frequency response (Fig. 2).

The calculation of the possible overshoot values is straightforward and yields the values in Fig. 6. Here the ratio of the maximum value of the overshoot

$$V_{\max} = \max V_3(T) \quad (8a)$$

and the steady-state signal value

$$V_\infty = V_3(\infty) \quad (8b)$$

are given as function of the relative amplifier slowness with feedback α as a parameter. The solid lines between the arrowheads depict the possible combinations of α and γ, as shown in Fig. 5. The lower group of curves applies for the linear frequency response and the upper group for treble boost with $\epsilon = 0.05$, corresponding to some 12.5 dB of boost.

Fig. 6 shows that, for this idealized model, feedback values greater than 40 dB inevitably lead to rather large internal overshoots within the amplifier. Feedback values beyond 60–80 dB lead to overshoots, which may be impossible to handle with the usual dynamic margins of the amplifiers, unless very special circuit designs and extremely careful dimensioning are employed.

THE CLIPPING PHENOMENON

To analyze the clipping, amplifier A may be divided as in Fig. 7. Here amplifiers A_1 and A_2 represent the driver stages and the output stages, respectively, and for the purpose of the analysis they are assumed to have no effective poles or
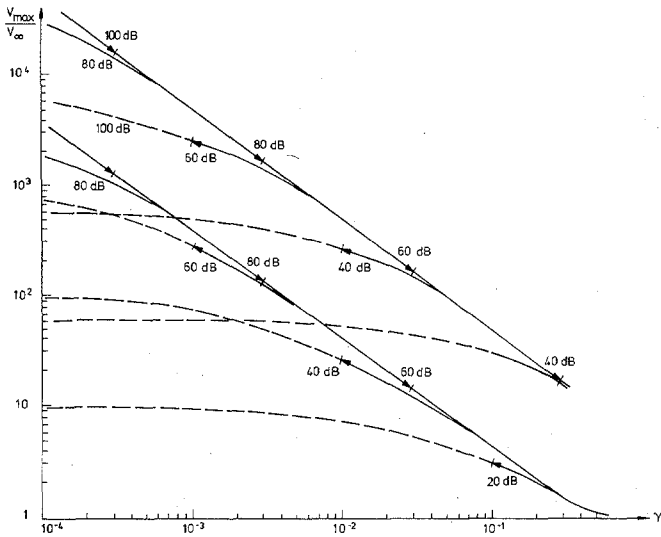
Fig. 6. The maximum value V_{max} of the overshoot of $V_3(T)$ as a function of the frequency ratio γ with feedback as a parameter. The solid lines between the arrowheads depict the possible combinations of α and γ, shown in Fig. 5. The lower group of curves applies for the linear frequency response and the upper group for the treble boost with $\epsilon = 0.05$, corresponding to approximately 12.5 dB of boost.
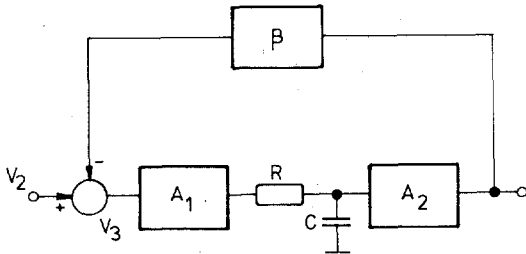

Fig. 7. The division of the power amplifier. The "slowness," usually the frequency compensation circuit, is simulated by the R and C. For the purpose of the analysis, the gain of amplifier A_2 is taken to be unity.

zeros in their transfer functions. The open-loop cut-off frequency ω_1 is assumed to be produced by a passive R-C network connected between the two amplifiers. These assumptions are self-consistent with the frequency response shown in Fig. 3, and in practical amplifiers the physical counterpart of the R-C network is almost invariably the lag compensation network. Voltage V_3 (T) is amplified in amplifier A_1 before being integrated in the R-C network. The amplified overshoot in V_3 (T) then represents the extra energy needed to speed up the charging of capacitor C.

The overshoot in V_3 (T) will be clipped at the output of amplifier A_1 if A_1 does not have a sufficient dynamic overload margin. A_1 usually consists of the input stages of the amplifier, which are commonly designed for best possible signal-to-noise ratio and thus often do not have much headroom above the normal signal levels. The measured overload margins (above V_∞) of several commercial amplifiers [2] ranged between 2 and 15.
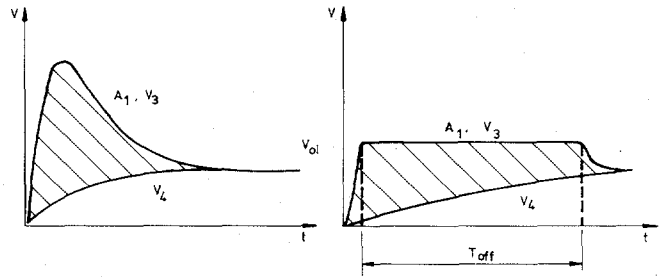

Fig. 8. The mechanism of clipping. In order to charge the same energy into the capacitor C of Fig. 7, the shaded areas must be equal in the case of no clipping and in the case of clipping at a specified overload level V_{01} of the amplifier A_1.

The clipping phenomenon is shown in Fig. 8 where, for the sake of clarity, the gain of amplifier A_2 is taken to be unity. The left figure is the unclipped case. If clipping occurs at an overload level V_{01}, as in the right figure, the cut-off time T_{off} becomes much longer than might be expected. This is because in the absence of the overshoot it now takes a longer time to charge capacitor C to the final value of the output voltage V_4. The shaded areas in both figures must then be equal because of energetic boundary conditions, they represent the charge in the virtual capacitor C.

The calculation of the cut-off time T_{off} may be performed with a piecewise numerical integration computer program based on the flow chart shown in Fig. 9. The voltages (Fig. 1) are governed by the following equations:

$$V_3(T) = V_2(T) - \beta V_4(T - \tau) \tag{9}$$

$$V_4(T) = V_4(T - \tau) + [A_0 V_3(T)$$
$$- V_4(T - \tau)] [1 - \exp (- \tau/RC)], \tag{10}$$

τ being the time increment in the program. These equations assume $A_1 = A_0$, $A_2 = 1$ for convenience as in Fig. 8. However, the results are independent of the location of the stage where clipping occurs, as well as of the gain distribution between A_1 and A_2, the important parameter being the normalized overload level V_{01}/V_∞ for the clipping stage.

Figs. 10-12 show the calculated cut-off time as a function of the normalized amplifier overload level V_{01}/V_∞. The curves for the linear frequency response (solid lines in Figs. 10-12) follow the simple rule of clipping shown in Fig. 8. The case of treble boost is, however, different, as is shown in Fig. 13. As can be seen, this case includes an undershoot in $A_1 V_3$ and results in double-sided clipping for some values of V_{01}. This is responsible for the odd shape of the dotted curves in Figs. 10-12. If T_{off} is small, there is a large difference between the linear response and treble boost cases. If, however, T_{off} becomes long enough, the effect of the treble boost has already faded off before the amplifier recovers from the cut-off. This is the reason why the curves for the linear response and the treble boost case join together in the upper portion of the curves.
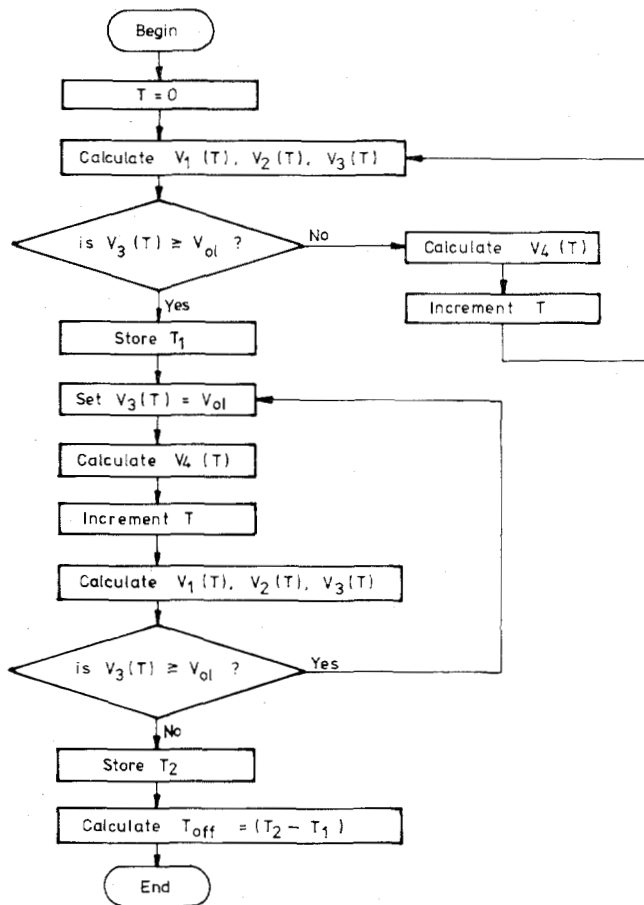
Fig. 9. Flow diagram of the program used for the calculation of the cut-off times from (9) and (10).
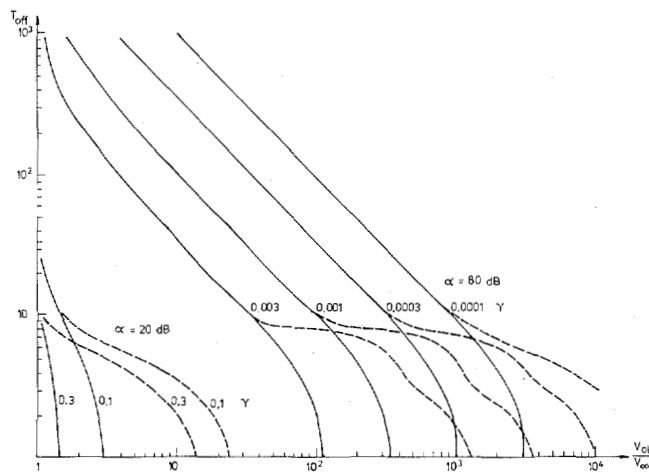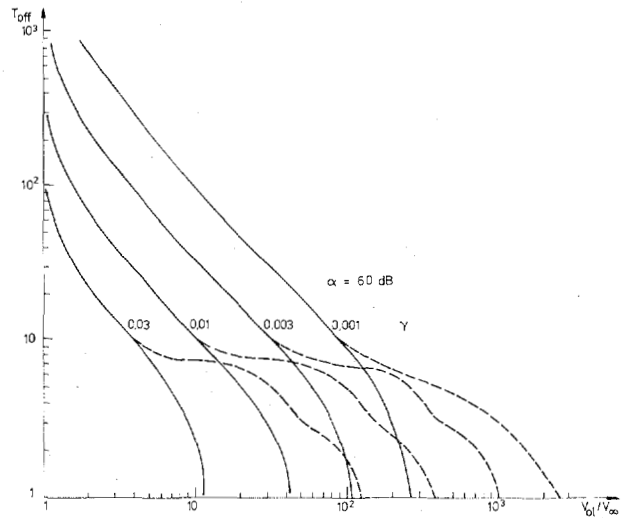

Fig. 11. The equivalent curves as in Fig. 10; the feedback value is 60 dB.
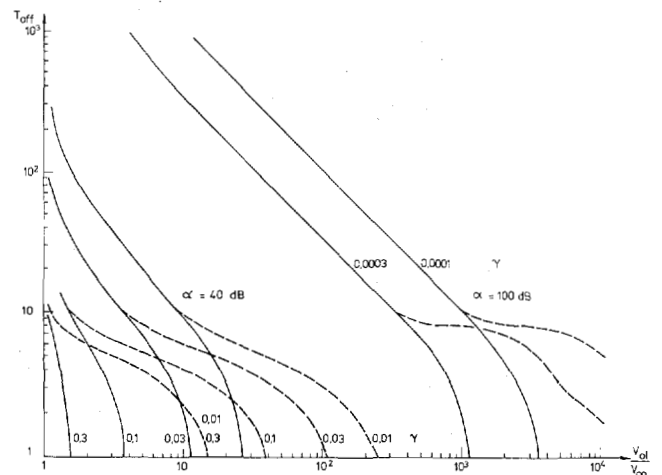
Fig. 12. The equivalent curves as in Fig. 10; the feedback values are 40 and 100 dB.

Fig. 10. The length of the clipping T_{off} as a function of the amplifier driver stage overload margin V_{01}/V for feedback values of 20 and 80 dB. The solid lines represent the linear frequency response and the dotted lines the treble boost with $\epsilon = 0.05$.

The normalized overload level V_{01}/V_{∞} may be measured for a practical amplifier by identifying the stage in which the clipping occurs and checking the type of clipping [2]. Using very small square-wave input signal, the overshoot-to-signal ratio V_{max}/V_{∞} is measured at the output of the clipping stage. The input signal level is then slowly increased until this ratio begins to decrease, meaning that the overshoot is being supressed. This amplitude of the overshoot is then V_{01}. The input signal is further increased, until the amplifier reaches its full output voltage swing. The voltage of the tail of the square wave, measured at the output of the clipping stage, is then V_{∞} at the full output. It must be noted, however, that for most amplifiers V_{01} and V_{∞} are different for positive and negative signals. From these measurements, the ratio V_{01}/V_{∞} at full output may be calculated. Measuring the amplifier α and γ, and using Figs. 10–12, the normalized T_{off} at full output may then be read. For lower output levels, V_{∞} is proportionally decreased and T_{off} follows the same curve to the right.

In order to relate the curves to practical amplifier design, the denormalization of T is necessary. In most cases, it is sufficient to consider $\omega_3/2\pi = 30$ kHz as a realistic upper cut-off frequency for the signal source and the preamplifier. In this case, the normalized T-values must be multiplied by roughly 5 μs in order to get the real cut-off times.
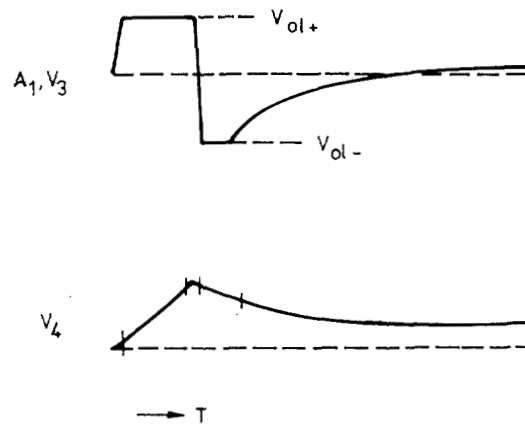
Fig. 13. The clipping phenomenon in the case of treble boost. At certain levels of the signal and/or the overload margin, a double-sided clipping occurs. This explains the odd shape of the dotted lines in Figs. 10–12 for the feedback values of 60 to 100 dB. The regions of cut-off are marked with lines in the $V_4(t)$ curve.

TABLE I
CALCULATED CUT-OFF TIMES FOR THE POSTULATED 100 W AMPLIFIER

Power	t_{off} (µs) linear	t_{off} (µs) treble boost
100 (W)	160	160
10	48	48
1	0,1	36
100 (mW)	0	20
10	0	0,1

As an example, consider a 100 W amplifier having a closed-loop small-signal frequency response of 100 kHz, preamplifier frequency response of 30 kHz, 80 dB of feedback, and driver stage overload margin of 100 at full output. γ is then 0.0003 and $V_{01}/V_{\infty} = 100$. The cut-off times for different power levels will then be as listed in Table I. As can be seen, the cut-off is still present at surprisingly low output power levels, though it becomes rather short.

Recently a more direct method of the measurement of TIM has been suggested [10]. It is based on the measurement of intermodulation spectra between a low-frequency square wave and a high-frequency sine wave.

ACCURACY OF THE ANALYSIS

The results of the analysis have been subjected to extensive checking [9].

Digital checking of the overshoot values was performed by calculating the various circuit potentials for step function excitation using the ANP 3 computer program. The results agree with those calculated directly from (7) to an accuracy of 1 percent. This small discrepancy is caused by the piecewise nature of the programs.

Digital checking of the cut-off times was performed with a numerical simulation program, specially developed for this purpose [9]. The discrepancies between the simulation and the results presented in this paper were of the order of 1 percent, again depending on the piecewise program structures.

Analog simulation was performed with a specially designed simulator [9]. The overshoot values could be reliably checked up to 60 dB of feedback and the cut-off time results from 0.6 to 400 T. The estimated accuracy of the simulator was of the order of 3 percent and the discrepancies between calculated values and the simulated ones were of the order of 5 percent. A major source of error in the analogue simulation was the poor interloop signal-to-noise ratio, caused by the necessity to reserve a large "headroom" for the overshoots.

These results indicate that the calculations should in principle be correct and that the physical nature of the phenomenon has been correctly interpreted.

TIM, SLEW RATE, AND POWER BANDWIDTH

The slew rate, the power bandwidth and the TIM all characterize in one way or another the high-frequency capabilities of an amplifier, and it would be desirable to predict the one from the other. Though this might in some cases be possible, the definitions of these characteristics are so different that no generalizations can be drawn.

The slew rate is defined as a maximum dV/dt at the output of an amplifier for a very large input signal. Under these conditions the amplifier is in a complete slope overload condition, and is extremely nonlinear. It is a common practice to boost the slew rate specification by designing the amplifier to have a "soft" gradual slope overload characteristic. The amplifier then becomes excessively nonlinear long before the slew rate is reached. As TIM is produced already when the overshoot is *suppressed*, certain amplifiers may show a marked tendency to produce TIM at signal risetimes far below the slew rate [10]. Therefore, TIM can be predicted from the slew rate specification only in those amplifiers, which have a "hard" clipping in the driver stages. In the authors' experience, marked TIM may begin already at less than one tenth of the slew rate in some amplifiers.

The power bandwidth is defined at a specified total harmonic distortion level, usually 1 percent. The TIM may be predicted from the power bandwidth provided that the feedback is low, say, 20–30 dB. If, however, the feedback is higher, the whole comparison becomes absurd as high-frequency inter-loop clipping occurs already at much lower output THD levels than 1 percent. From the authors' experience, an amplifier having 60 dB feedback and 50 kHz power bandwidth showed already marked TIM for program material having an upper cut-off frequency of 20 kHz.

As can be seen, the interrelation of these three characteristics depends on the design principles of the amplifier. A low slew rate or poor power bandwidth may indicate the presence of TIM, but the converse is not generally true. Equally, a high slew rate or good power bandwidth does not automatically make TIM less likely to occur. The presence of TIM in an amplifier may also be suspected from an unrealistically low THD specification, which may indicate the possibility of large feedback and, consequently, heavy compensation [10].

CONCLUSIONS

The basic mechanism of TIM has been discussed. The following points have been shown.

1) If feedback value in excess of, say, 60 dB is used, it may lead to large internal overshoots within the amplifier, caused by the necessary heavy compensation. Depending on the construction of the amplifier, these overshoots may be up to several thousand times larger than the nominal values of the signal, although the amplifier input signal is in the usual low-frequency audio range.

2) As the amplifier driver stages usually have a limited dynamic range with regard to these overshoots, they will most probably be clipped in the driver stages preceding the amplifier "slowness." The cut-off time is lengthened by energetic conditions.

3) If treble boost is used, the cut-off time may increase by roughly an order of magnitude. This may also result in two-sided clipping.

4) The TIM is related to the slew rate and the power bandwidth, but prediction of one from another is possible only in cases where a low value of feedback is used and the amplifier driver stages have hard clipping characteristics.

REFERENCES

[1] M. Otala, "Transient distortion in transistorized audio power amplifiers," IEEE Trans. Audio Electroacoust., vol. AU-18, pp. 234–239, Sept. 1970.
[2] M. Otala and R. Ensomaa, "Transient intermodulation distortion in commercial audio amplifiers," J. Audio Eng. Soc., vol. 22, no. 4, pp. 244–246, 1974.
[3] H. Lewitt et al., "Perception of slope-overload distortion in delta-modulated speech signals," IEEE Trans. Audio Electroacoust., vol. AU-18, pp. 240–247, Sept. 1970.
[4] M. Otala, "Circuit design modifications for minimizing transient intermodulation distortion in audio amplifiers," J. Audio Eng. Soc., vol. 20, no. 6, pp. 396–399, 1972.
[5] J. Lohstroh and M. Otala, "An audio power amplifier for ultimate quality requirements," IEEE Trans. Audio Electroacoust., vol. AU-21, no. 6, pp. 545–551, Dec. 1973.
[6] J. Linsley Hood, "The straight wire with gain?" Studio Sound, pp. 22–27, Apr. 1975.
[7] W. M. Leach, "Transient IM distortion in power amplifiers," Audio, pp. 34–42, Feb. 1975.
[8] J. L. Stuart, "An approach to audio amplifier design," Wireless World, Part I, pp. 387–391, Aug. 1973; Part II, pp. 439–446, Sept. 1973; Part III, pp. 491–494, Oct. 1973.
[9] E. Leinonen, "Digital and analogue simulation of the transient intermodulation distortion," Electron. Lab., Tech. Res. Centre Finland, Oulu, Finland, publication S-16/1976, p. 14.
[10] E. Leinonen, M. Otala, and J. Cure, "Method for measuring transient intermodulation distortion," preprint 1185, presented at the 55th Conv. AES, NY, 1976; J. Audio Eng. Soc., to be published.