Is feedback the miracle cure for high-end audio?

Dr. Hans R.E. van Maanen (Temporal Coherence, info@temporalcoherence.nl)

Date of issue: 14 May 2017, extended 31 January 2019

Summary and Conclusions

Feedback seems like a miracle cure for all shortcomings of audio equipment. Yet, in the "high-end" audio community, many critics on feedback can be found. It is beyond discussion that the *specifications* of many semiconductor amplifiers are far superior to those of loudspeakers and valve (electron tube) amplifiers, yet this correlates not well with the *perceived* quality of the equipment. How come? Are certain phenomena overlooked and, if so, what can we learn to improve the *perceived* quality of equipment? This paper analyses some pitfalls and parasitic effects of feedback and gives directions for improvement of the perceived quality. This was confirmed by designing amplifiers, derived from this analysis, using unconventional lines in which listening by music experts was regarded more important than measurement results.

It showed that parasitic effects occur in amplifiers with global feedback, which are often disregarded, because these do not show up with the usual derivation of the equations for systems with feedback. These parasitic lead to the *introduction* of artefacts, which are *specific* for systems with feedback. This is surprising, as the common idea is that feedback only suppresses undesired phenomena, but is an unambiguous result from the analysis, presented in this paper, which also shows that the commonly used equations for feedback are incorrect.

Suppression of these parasitic effects requires linearization of the individual amplification stages as much as possible, in combination with a constant, but moderate, feedback factor over the entire audio range.

The testing of equipment using continuous sinewaves does often not reveal these parasitic effects as these only show up in the *dynamic* response of the amplifier to music-like signals. The simplistic approach that the sinewave response enables the prediction of the behaviour under all conditions ignores the conditions under which the Fourier theory may be applied and leads therefore to incorrect results and conclusions. Which is why there is a great need for well-defined dynamic test signals, but as long as these are not available, human hearing remains the best piece of measurement "equipment" which can be used.

---

1. Introduction

It remains an intriguing question: why do sound amplifiers different, even when their specifications are similar? This also happens with other electronic equipment: two SACD players from the same brand, but with a significant different price, were compared. Their virtually identical specifications were three orders of magnitude better than those of good loudspeakers, but the difference in sonic quality was obvious. In the (popular) literature on "high-end" audio, strongly opposing views on global feedback can be found. Some say "you cannot have enough feedback" (ref. 1), whereas others have serious criticism on feedback. Due to the lack of technical foundation of the criticisms and requirements from the high-end community, more technical / scientific oriented people easily tend to put these aside as remarks from people who smoked too much weed as their meters tell a different story. Yet, because the volume of criticism on amplifiers with feedback is so extensive, denying that there is at least some truth in it, is unrealistic. So where do the problems with feedback hide and could we get more clarity on this issue? In "Audio Power Amplifiers" (ref. 2), two interesting statements have been found:

"Many transistor amplifiers produce so little harmonic distortion that it is unlikely that it would be the reason for subjective differences between units. This leads to the conclusion that there must be other error sources than those that are usually analysed which have a major effect on audible quality."

"Most amplifiers today rely on high global negative feedback. When loop gain is high enough, input circuitry operates at low signal levels, allowing very simple circuits to be used without compromising linearity. This assumption is valid when operation is analysed using a single tone and steady state input signals. However, it is possible that this design philosophy results in amplifiers with poor dynamic, transient and overload recovery performance."

From another paper (ref. 3), the following interesting abstract was taken:

The subjective and objective evaluation of 5 high-quality vacuum-tube audio amplifiers is presented in this paper. As the reference the professional transistor amplifier has been used. The subjective evaluation has been done by the team of judges. It was found that the best sound quality is obtained by vacuum-tube amplifiers, the worst by the reference amplifier. The results of subjective evaluation are inconsistent with quality assessed by measurement of objective parameters: all amplifiers have comparable quality, but the best is the transistor amplifier because of lowest level of THD+N.

Both sources show that the usually measured properties do, diplomatically put, not correlate well with the perceived quality of amplifiers. So we will have to identify "*the other error sources*". The *temporal* response of systems to *dynamic* signals, like these occur in music, are rarely taken into account with these measurements, but when feedback is used, this could be crucial to understand their *perceived* quality. However, at first sight it is unclear why the common testing does not reveal the dynamic and temporal properties of the electronics. Revealing the cause of this paradox might also shed light on the above mentioned requirements by the high-end community.

After defining the different parameters of the analysis, we will have a better look at the criticisms of the high-end users, followed by an extensive analysis of local and global feedback in an amplifier. Subsequently, we will look at other error sources in amplifiers and their loading and the consequences for amplifier design.

2. Nomenclature / definitions used in this paper

In this paper, we will use a number of variables, which are sometimes differently defined than is common. To avoid confusion, we will define these here:

Open-loop gain μ:

$$\mu = \frac{V_{out}}{V_+ - V_-}$$

Open-loop gain at low frequencies (frequencies far below the open-loop cut-off frequency) μ_0:

$$\mu_0 = \frac{V_{out}}{V_+ - V_-}$$

The feedback factor β:

$$\beta = \frac{V_-}{V_{out}}$$

N.B. Note that both μ and β can be complex numbers and can be a function of frequency.

The closed-loop gain A:

$$A = \frac{V_{out}}{V_{in}} = \frac{\mu}{1+\mu\beta}$$

The feedback margin M:

$$M = \frac{\mu}{A} = 1 + \mu\beta$$

The open-loop bandwidth:
This parameter is determined by the -3 dB (open-loop cut-off) frequency (relative to the maximum value of μ) of the dominant pole of the open-loop gain.

The closed-loop bandwidth f_{max}:
This parameter is defined by the -3 dB frequency of the closed-loop gain A (relative to the maximum value of A).

3. Remarks from the high-end audio fora

On fora, which discuss high-end audio and its related equipment, often statements like

- Amplifiers with strong feedback do not sound 'musical'
- When an amplifier is processing rather error signal than the music, it does not sound natural
- Amplifiers with a constant feedback margin in the audio band sound better
- Feedback needs time to react, so it is always too late

- Amplifiers should be fast (up to 100 V/μs)
- Amplifiers should have ‘headroom’

can be found. At first sight, these statements are rather puzzling. For the first statement, all kinds of reasons are brought forward, some seem more (omen est nomen) sound than others, but it is never quite clear *why*. Some people even go as far that they do not want amplifiers with feedback anymore. The paradox is that often the same people prefer vinyl over digital audio, but there is no cutting disc in this world that does not use *motional feedback*.

The error signal can, of course, differ from the input signal, this is basically the “raison d’etre” of feedback. But when the differences are larger than the input signal, it means that there is a lot (or too much?) to correct. One could wonder what is actually happening and where the error in the signal comes from. This will be a topic which will be studied in more detail later on.

Why amplifiers which use a constant feedback factor in the audio band should sound better is not obvious. Yet, there are at least two good reasons for it, which will be discussed below.

Feedback can have an effect on the temporal properties of the amplifier, leading to audible time-smear. So the application of feedback needs to be aligned with the temporal requirements of sound reproduction.

The -absurd- requirement of speeds up to 100 V/μs probably finds its motivation in other undesirable properties. This can be understood as 5 V/μs already corresponds to 100 W in 8 Ω at 20 kHz, a 20 fold increase would mean 40 kW at this frequency. For a system at home, this is way beyond what is feasible and why should this be necessary?

The requirement for headroom is also rather confusing as an amplifier should work properly up to its clipping power. So when no clipping occurs, there should be no difference in the perceived quality of the amplifier, so there would be no need for ‘headroom’, which is nothing but underutilisation of the amplifier. When people find this necessary, something in the “linear” range of the amplifier must go wrong which is not easily detected.

In due course, these issues will be traced.

4. Local and global feedback inside the amplifier

The starting point of the discussion is that there is nothing *fundamentally* wrong with either local or global feedback (ref. 4), but that in practice there might be parasitic phenomena which are often overlooked, so a more detailed look into (amplifier) feedback is required to find these.

4.1 The input differential amplifier

The first stage is usually a differential amplifier, combining the functions of subtraction and amplification. Note that the subtraction sits *outside* of the feedback loop. There are three pitfalls with this circuitry. Firstly, the signal level *increases* with *increasing* frequency due to the *decreasing* open-loop gain μ, which can simply be derived from the equation of global feedback:

$$V_+ - V_- = \frac{V_{out}}{\mu} = \frac{V_{in}}{1+\mu\beta} = \frac{V_{in}}{M}$$

Assuming that the open-loop gain decreases with 6 dB/oct. (first order filtering) above the open-loop cut-off frequency, determined by the time constant τ, the voltage over the differential input stage is equal to:

$$V_i - \beta V_o = \left[\frac{V_i}{1+\mu_0\beta}\right] \cdot \left[\frac{1+j\omega\tau}{1+j\omega\,{}^{\tau}/_{(1+\mu_0\beta)}}\right]$$

In fig. 1, two examples are shown for different conditions to illustrate the phenomenon.

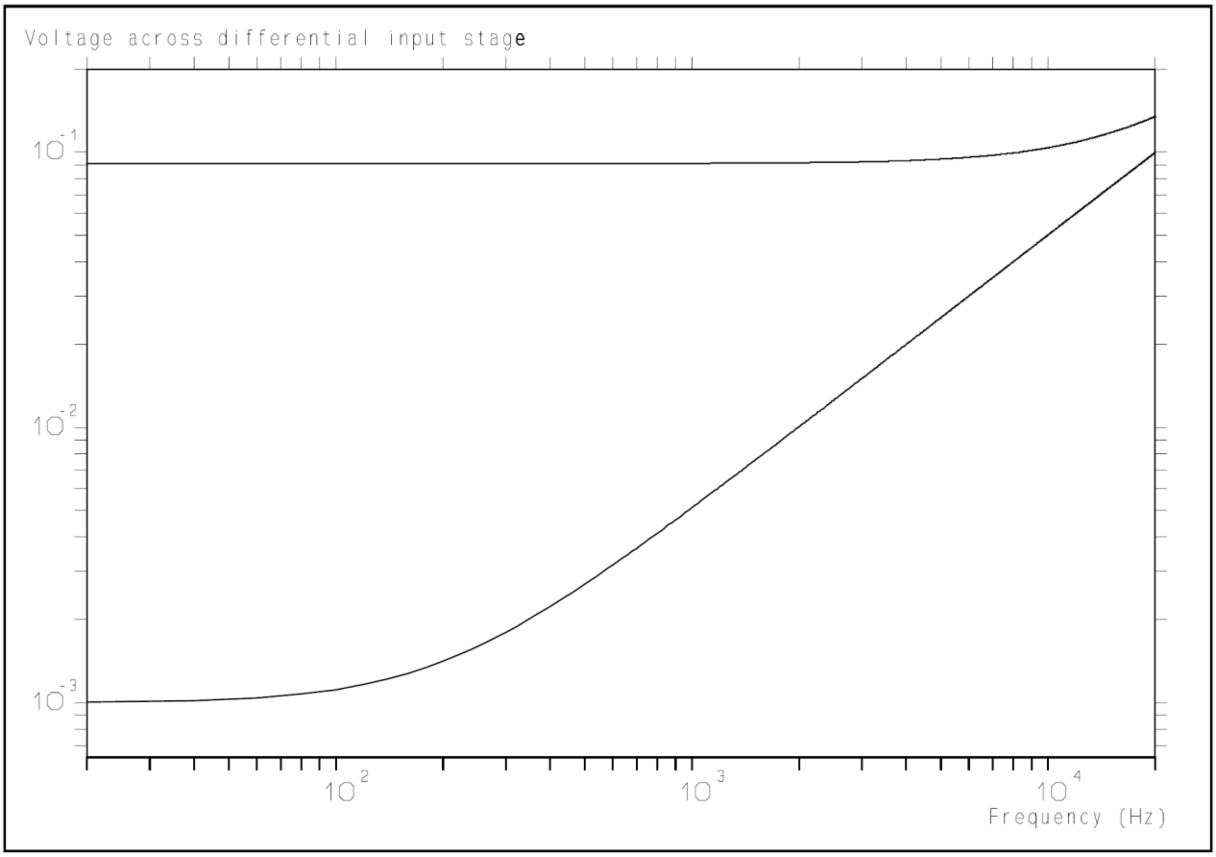

Figure 1: *The differential voltage at the input of the amplifier depends on the open-loop gain at low frequencies μ_0, the feedback factor β, the closed-loop bandwidth of the amplifier f_{max} and -of course- the input voltage V_{in}. Lower trace:* μ_0 = 10 000; β = 0.1; f_{max} = 200 kHz; V_{in} = 1 V, *upper trace*: μ_0 = 100; β = 0.1; f_{max} = 200 kHz: V_{in} = 1 V.

Note that the differential voltage across the input of the amplifier can increase significantly with frequency, so the second assumption of Kolinummi (ref. 2, see above) might be correct for low frequencies, but could be in error at higher frequencies. It will certainly lead to increasing distortion with frequency unless the differential input stage is perfectly linear or that the feedback margin M is constant in the audio band. This, however, puts an upper limit on the feedback margin M, as will be discussed later. Selecting the properties of the differential input stage on the requirements, set by the signal level at low frequencies, is likely to head for disaster.

Secondly, the collector-emitter voltages of the transistors of the differential amplifier vary when the same (in phase) AC voltage is applied to the two bases. The transistor properties depend on V_{ce}, so a "common mode" voltage will generate an output voltage when the transistors are not identical and/or the collector loads are different. **N.B.** Note that with high feedback margins, the common mode voltage becomes very large compared to the differential voltage:

$\frac{V_{in}}{V_+ - V_-} \approx \mu\beta$ and the common mode signal ≈ V_{in}. So the common mode voltage is approximately $\mu\beta$ times larger than the difference voltage.

Both effects are the largest with differential amplifiers with directly coupled emitters and the non-linearity of the base-emitter diode manifests itself, even with small (mV) excitations, in the perceived sound quality (ref. 5, see also fig. 1). Due to the exponential characteristic of the base-emitter diode, many higher harmonics (higher than the fifth) are generated, which are known to be irritating to the ear, even at low levels (ref. 6). Subsequently, the output signal of the differential amplifier is amplified by the following stages. A widespread misunderstanding is that the distortion of *all* amplifier stages is suppressed by the feedback margin *M*. This is incorrect (ref. 2) and this can easily be proven by splitting the amplifier in its separate stages. Assume three separate amplification stages, that each amplification stage is perfect and that after each stage the "misery" it generates is added. Thus the approximation of fig. 2 is obtained:

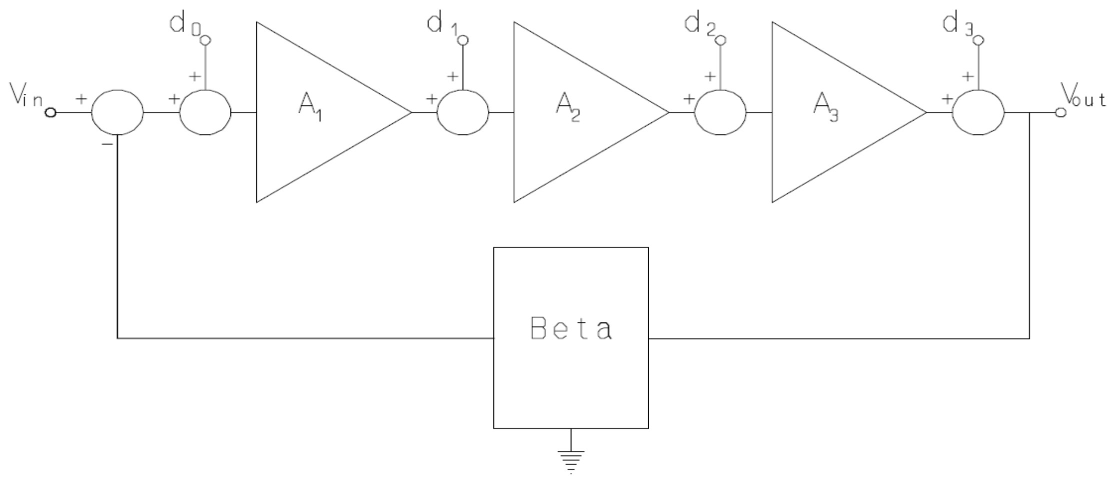

Figure 2: *The approximation of an amplifier with several stages to calculate its properties when global feedback is applied.*

The amplification of the i^{th} stage is A_i and its "misery" is d_i. Immediately after the subtraction of the input and the feedback signal, d_0 is added. This represents the "misery" the subtraction circuit contributes, as it is not perfect either. (**N.B.** "Misery" includes more than just distortion, but also noise, hum, signals coming from the power supply rail, etc. But for the sake of simplicity, we will mostly talk about distortion, but the reader should keep in mind that it includes other unwanted signals as well.) It is now easy to see that:

$$\mu = A_1 \cdot A_2 \cdot A_3$$

$$V_{out} = (((V_{in} - \beta V_{out} + d_0) \cdot A_1 + d_1) \cdot A_2 + d_2) \cdot A_3 + d_3$$

Reshuffling yields:

$$V_{out} = \frac{\mu \cdot V_{in}}{1+\mu \cdot \beta} + \frac{\mu \cdot d_0}{1+\mu \cdot \beta} + \frac{A_2 \cdot A_3 \cdot d_1}{1+A_1 \cdot A_2 \cdot A_3 \cdot \beta} + \frac{A_3 \cdot d_2}{1+A_1 \cdot A_2 \cdot A_3 \cdot \beta} + \frac{d_3}{1+A_1 \cdot A_2 \cdot A_3 \cdot \beta}$$

$$V_{out} = \frac{\mu \cdot V_{in}}{1+\mu\beta} + \frac{\mu \cdot d_0}{1+\mu\beta} + \frac{A_2 \cdot A_3 \cdot d_1}{1+\mu\beta} + \frac{A_3 \cdot d_2}{1+\mu\beta} + \frac{d_3}{1+\mu\beta}$$

$$V_{out} = \frac{\mu \cdot V_{in}}{M} + \frac{A_1 \cdot A_2 \cdot A_3 \cdot d_0}{M} + \frac{A_2 \cdot A_3 \cdot d_1}{M} + \frac{A_3 \cdot d_2}{M} + \frac{d_3}{M}$$

So the suppression of the distortion of the pre-stages is less than the feedback margin *M* and the distortion of the subtraction stage is amplified as much as the input signal (**N.B.** This is equivalent to saying that it sits outside the feedback loop). The consequence is that the distortion of the previous stages is amplified by the subsequent stages and that the distortion products are distorted as well, resulting in an increase of the harmonics of the harmonics. Note that the same holds for the sensitivity to supply voltage variations and the other contributors to the “misery” as all are amplified first before the feedback can act. The improvements, ascribed to global feedback, need to be analysed in more detail because the gratuitous application of the well-known equations is “cutting corners” and will lead to incorrect results. Which is why the internal workings of an amplifier will be discussed now. As a bonus it will be shown that the commonly used feedback equations are fundamentally incorrect. This has severe consequences for the design of amplifiers.

4.2 The modulation depth of the amplification stages

The third pitfall requires a more extensive description. Two contributions to the output signal of the subtraction stage can be distinguished: first of all the difference between the input signal and β times the ideal (= misery free) output signal and secondly the misery at the output of the amplifier, also multiplied with β. In equation, where $V_{sub} = V_+ - V_-$:

$$V_{sub} = \frac{1}{1+\mu\beta} V_{in} - \beta \cdot d_t$$

and in which d_t is the total misery at the output of the amplifier.

The contribution of the distortion alone can be significant, compared to that of the input signal. The larger the open-loop gain, the less favourable the ratio between the “signal” and the “misery” gets. So are all underlying assumptions of the feedback theory (like the quasi-linearity of the individual amplification stages), still valid? In the popular “high-end” literature, you can read that “when an amplifier is rather processing distortion instead of music, you end up with a non-musical system”. This does not seem to be just hot air:

when $\frac{d_t}{V_O} \geq \frac{1}{\mu\beta}$, the amplifier is processing misery rather than signal.

Often, it is remarked that this cannot happen as the misery is suppressed by feedback margin, but that is incorrect. As we have seen above, the suppression of the misery depends on where in the amplifier the misery is generated and the suppression decreases with increasing frequency, whereas μ decreases (see fig. 1). So the above mentioned condition can be fulfilled, depending on the design of the amplifier, with complex, multitone signals. This condition should be kept in mind when an amplifier is designed and it should be used when a design is evaluated.

The strength of the difference signal increases with increasing frequency (see fig. 1). All amplification stages, prior to the cut-off capacitor, will have to process a signal with an increasing strength with frequency. This can mean that these stages will be operated in a strong non-linear way and the lower the open-loop cut-off frequency, the sooner the increase with frequency starts. As the closed-loop bandwidth is determined by the open-loop gain, its cut-off time constant and β, these parameters cannot be chosen completely independent of each other, but the choices made do have an influence on the non-linearity of the individual stages and the reduction of their distortion. Below, we will discuss some options to optimise these choices.

4.3 Fundamental error in the commonly used feedback equations

The basic assumption, underlying the derivation of the commonly used feedback equations, is that u_0, the open-loop gain of the amplifier at low frequencies, is a constant. However, distortion is caused by a *non-linear relation* between the input and output voltages of the amplifier in open-loop (see fig. 3). So u_0 is ***not*** a constant and thus the basic assumption is incorrect and therefore the equations are incorrect too: the equations describe the reduction of the distortion of the amplifier, caused by the non-constant value of μ_0, by assuming that μ_0 is a constant. Mathematically this makes, of course, no sense at all. (**N.B.** A constant μ_0 and a distorting amplifier are mutually exclusive!). At best, one obtains an approximate result, but it is obvious that the larger the difference between the *assumption* and *reality* (or the stronger the non-linearity of the input-output relation), the less accurate the approximation and the stronger the interactions will be. Some examples of such undesirable interactions will be described below.

4.4 Interaction between the non-linear characteristic and the limited bandwidth

The non-linear (open-loop) input-output relation of an amplifier also means that the “instantaneous” μ (= dV_{out}/dV_{sub}) is not a constant, but (also) a function of the input voltage (ref. 6) and as illustrated in fig 3.

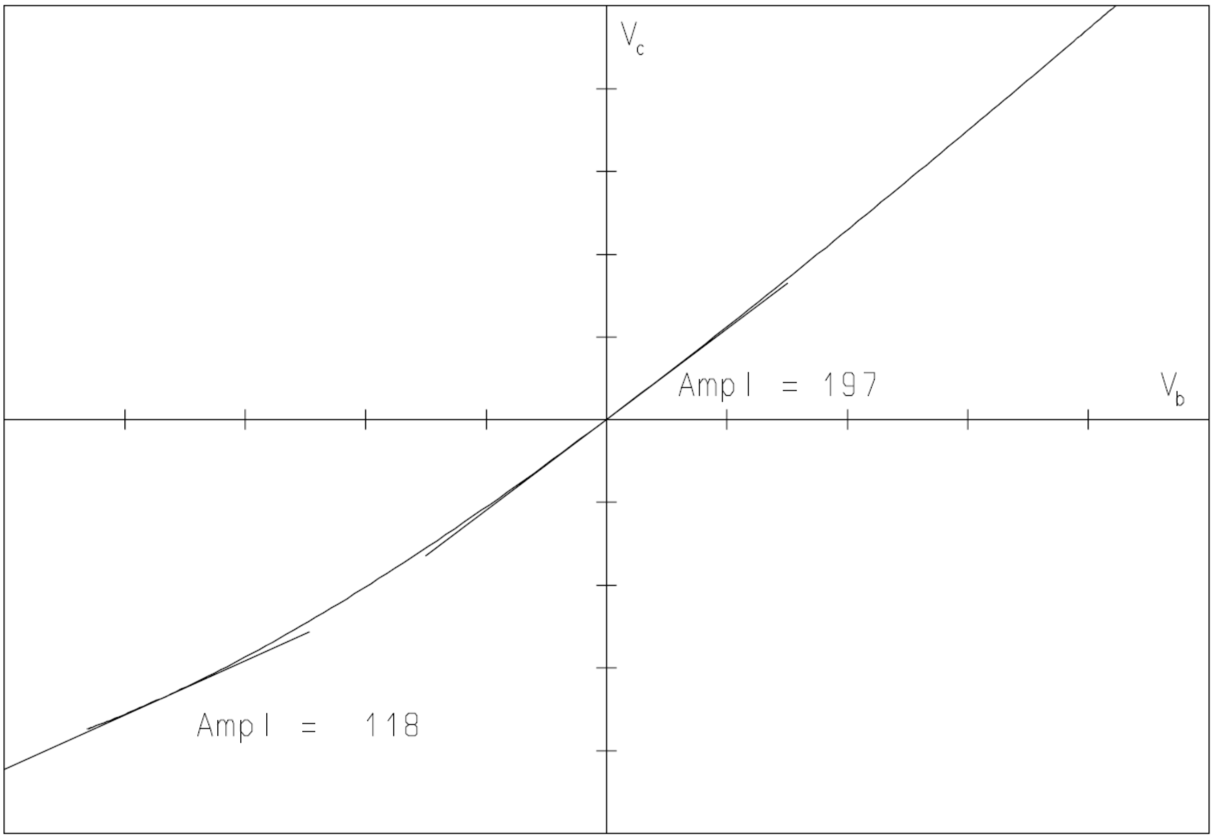

Figure 3: *With a non-linear input-output relation, the "instantaneous" amplification depends on the input signal value. Around 0 Volts, it is about 197, at lower voltages it reduces to below 120.*

Commonly, the feedback equations are derived by separation of the different effects. But this is incorrect as than the *interaction* between the two is swept under the carpet. An example of an interaction is illustrated as follows: take a low frequency signal with a large amplitude and a high frequency signal with a small amplitude. Intermodulation between the "low" and the "high" frequency is generated, but as the "instantaneous" bandwidth of the *closed-loop* amplifier varies (modulated by the low frequency, see fig. 3), also a *phase* modulation of the "high" frequency results (**N.B.** Phase modulation shows a close resemblance to *frequency* modulation). This has been noted before (ref. 7), but the assumptions in that paper concerning the closed loop bandwidth are rather optimistic, as will be discussed below. Similar interactions are to be expected between e.g. supply rail variations and the input signal. Such interactions are hard to put in equations, yet these cannot, and should not, be neglected. Numerical simulations can reveal these and provide a semi-quantitative result, useful for optimisation. (**N.B.** This effect is a direct consequence of the use of (global) feedback and thus an example of an artefact, which is *introduced* by feedback.) These interactions also plead for a linear (open loop) input-output relation and a high open-loop cut-off frequency. (**N.B.** These effects will *not* be detected with a *continuous sinewave sweep*, simply because it requires two frequencies simultaneously, which does not happen with a continuous sinewave sweep.)

4.5 Memory effects caused by the non-linear input-output relation

The non-linear open-loop input-output relation leads to a parasitic effect: the distortion generates a (small) DC component in the output signal. The cause can best be understood by an example, for which the basic circuit of fig. 4 will be used. **N.B.** The examples shown will be exaggerated to a certain extent to illustrate the phenomena more clearly.

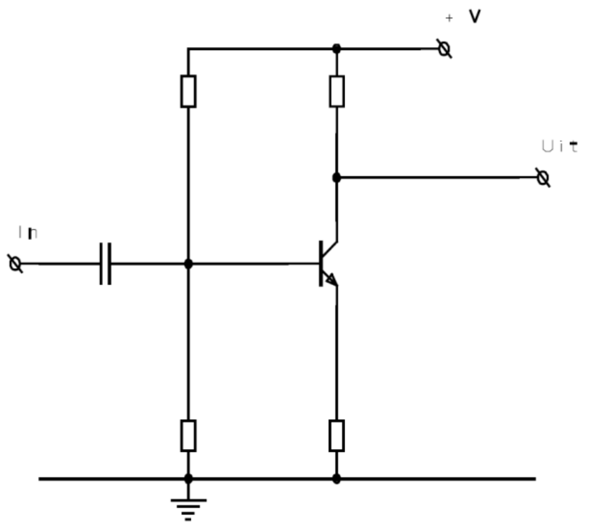

Figure 4: *The basic circuit used to illustrate the generation of a DC voltage due to distortion.*

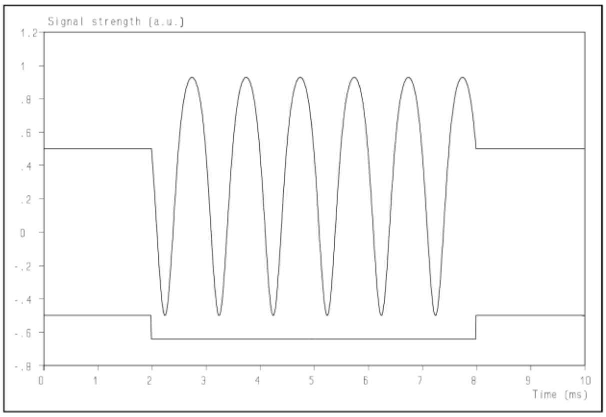

Figure 5: *Upper trace: the distorted output signal of a tone-burst. Lower trace: the average value of the upper trace signal, averaged of a single cycle.*

The distortion is caused by the non-linear properties of the base-emitter diode. In the positive part of the input cycle, it conducts *more* above the setpoint value than it conducts *less* in the negative part of it. As a result, the collector voltage drops further below the setpoint value during the positive part of the input cycle than it rises during the negative part. This is illustrated in fig. 5 (upper trace) and as a result, the collector voltage, averaged over a cycle, is lower than the setpoint value (lower trace). So a non-steady signal (as is common in music) is accompanied by a “DC” component, which is related to the input signal, but this is rather a low frequency signal, related to the *envelope* of the input signal and thus it does not consist of harmonics of the input signal. The value of the “DC” component will, of course, increase with the amplitude of the input signal and it does so more than proportional. This phenomenon is illustrated in fig. 6.

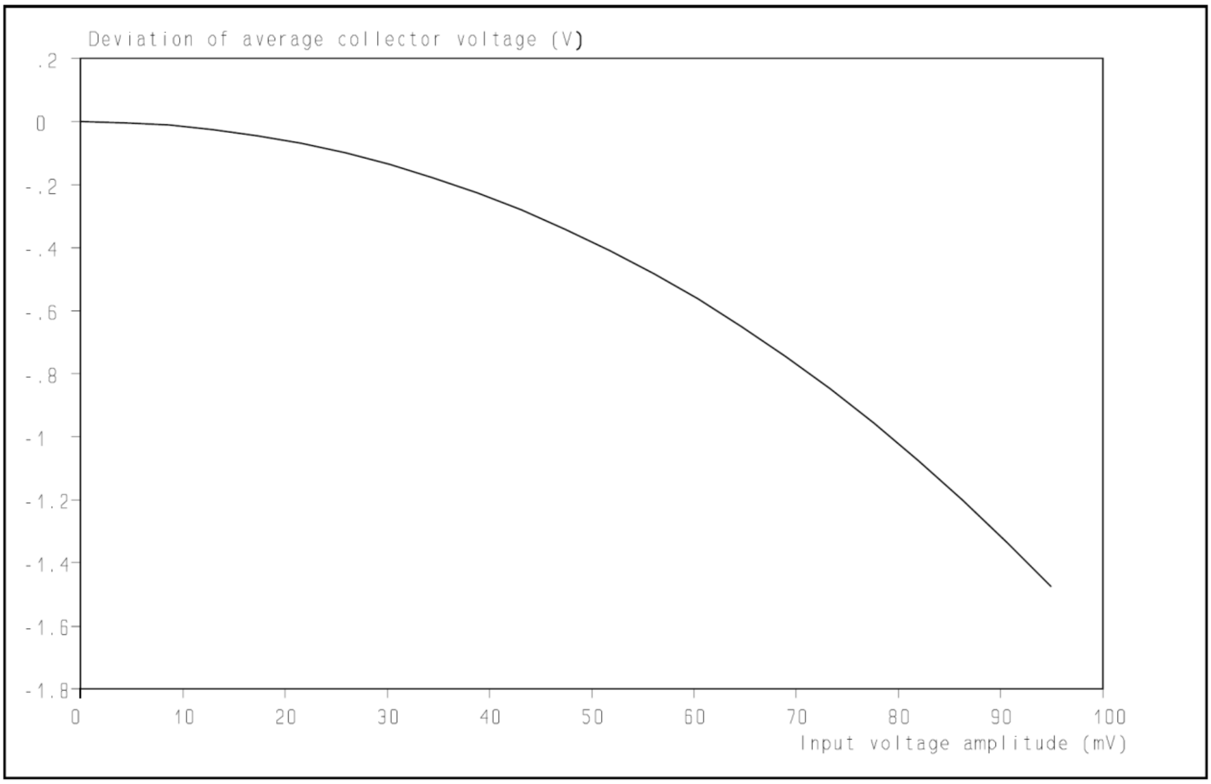

Figure 6: *Deviation of average collector voltage as a function of the input voltage. Amplification of stage ≈ 200 times, supply voltage 36 V.*

N.B. Only if the input-output relation is perfectly symmetric, this effect will not occur. This is not going to happen in reality. **N.B.** In virtually all modern amplifiers, the amplification stages are DC-coupled. So shifts in the operating points of one stage will proceed in the operating points of the following stages and will also show up at the output of the amplifier. This is an undesirable artefact as it introduces low-frequency signals, which are not present in the original signal (music) and, by experience, this leads to an unrest in the reproduced sound. It might also be the background of the desire of high-end listeners for “headroom”: as the phenomenon is rapidly going stronger with the input signal, by using the amplifier only far below its maximum rating, the artefact will be rather small and therefore not as irritating. However, it is an artefact, which can be avoided by a proper design of the amplifier.

The consequences of this mechanism are enhanced by the memory effect of the capacitor, used to limit the open-loop bandwidth of the amplifier as shown in fig. 7. Fig. 8 shows that the collector voltage is turned into a sort of modulated DC signal, which is not really surprising as the generated DC component is not attenuated by the low-pass filtering, whereas the AC of the input is. In this case, the frequency of the input signal is 1 kHz and the cut-off frequency is set at 200 Hz. It is obvious that the lower the cut-off frequency is chosen, the worse this phenomenon becomes. So this is another argument to keep the open-loop cut-off frequency as high as possible, which also tends to a condition in which the feedback margin in the audio band is constant. As can be seen in fig. 9, where the cut-off frequency is set at 20 kHz, the AC component is still dominant in such a case.

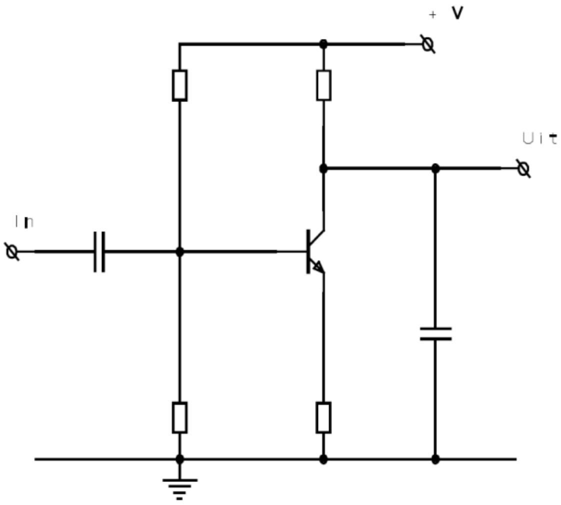

Figure 7: *The addition of a capacitor to limit the open-loop bandwidth of the amplifier creates a memory for the "DC" signal, generated by the distortion. Compare with fig. 4.*

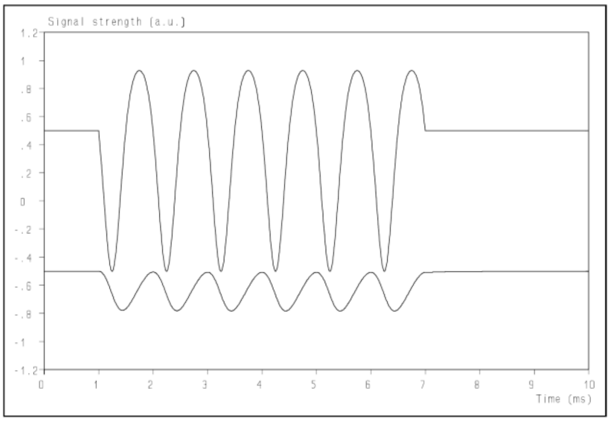

Figure 8: *The collector voltage without (upper trace) and with the capacitor of fig. 7 added (lower trace). Note that the collector voltage of the lower trace is rather a modulated DC signal than an AC signal. Input signal 1 kHz, cut-off frequency 200 Hz. Compare with fig. 5.*

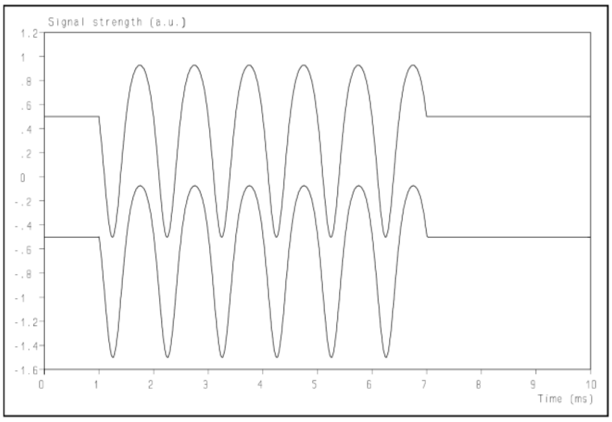

Figure 9: *The same situation as of fig. 8 is shown, except that the open-loop cut-off frequency is set at 20 kHz instead of 200 Hz.*

N.B. This effect will also *not* be detected with *continuous sinewaves*, simply because it is caused by a short term varying input signal amplitude, as happens with music, which does not occur with a continuous sine wave. A tone-burst signal might reveal it in some, more extreme, cases. **N.B.** Note that this mechanism generates an artefact which is specific for amplifiers with global feedback as the required reduction of the amplification with frequency demands the introduction of the cut-off capacitor. So we can conclude that feedback does not always suppresses "misery", but can also *introduce* new "misery"! Another advantage of a high open-loop cut-off frequency is that the "recovery time" after an overload condition is short.

Figure 6 shows that the deviation of the operating point increases rapidly with the input signal amplitude, a quadratic fit approximates the curve quite well. As it also gives rise to the varying charge of the cut-off capacitor and thus results in relatively slow envelope signals at the amplifier output, it could be experienced as a "slow" response, unrelated to its rise time. However, when this phenomenon would be suppressed, the amplifier will behave a lot better up to its clipping level and a "normal" rise time would suffice for reproduction in the living room.

A common way to linearize an amplification stage is the addition of an emitter resistor (see fig. 4). Surprisingly, this does not reduce this parasitic effect proportionally. When all the obtained amplification would be used for feedback to reduce the phenomenon, it actually is worsening the situation. This is shown by table 1, below:

Nominal amplification	Shift in average collector voltage	Minimal value using amplification for feedback
240	1.6 V	6.67 mV
43	0.46 V	10.70 mV
9.83	0.13 V	13.22 mV

Table 1: *Shift of the average collector voltage of a single amplification stage as a function of the amplification when the amplification is controlled by local feedback, using an emitter resistor.* $V_{cc} = 36$ *V,* $I_c = 1$ *mA.*

So the linearization of the individual amplification stages requires other means than using a relatively large emitter resistor. The design team of "Temporal Coherence" has succeeded in finding a solution for this problem.

Another limitation, introduced by feedback, is more well-known: the slew rate of an amplifier (the maximum of dV_{out}/dt). This is caused by the current, required to charge the cut-off capacitor by the amplification stage. The larger the capacitor, the lower the slew rate is. This also pleads for a small capacitor (and thus a wide open-loop bandwidth). Note that the slew rate limitation requires that the signal source is limited in spectral content and / or that an input filter is required to ensure this limitation is realised under normal operating conditions.

4.6 Limitations to the feedback margin

By now, we have encountered several effects which plead for a linear open-loop gain with a high cut-off frequency. In the popular high-end literature, this is often mentioned, based on listening experiences, but without technical or scientific underpinning worth mentioning. In the above, at least some indications can be found which support this and which make it plausible that the exchange of local and global feedback does *not* lead to amplifiers with the same *perceived* quality. It rather indicates that it is better to linearize the open-loop gain as much as possible and to apply a moderate *global* feedback. The additional advantages are i) that the cut-off frequency of the open-loop gain can be chosen high so that a constant feedback margin in the audio range (< 20 kHz) becomes feasible, ii) that the amplifier is mainly processing music and not misery and iii) the recovery time after overload short is.

The feedback margin and the closed-loop bandwidth of an amplifier are linked to maintain stability with global feedback. Optimum stability is reached when the open-loop gain decreases with 6 dB/oct. above a certain frequency and that the phase shift is -90º. The consequence is that the maximum feedback margin at 20 kHz equals the closed-loop bandwidth divided by 20 kHz. This follows directly from the gain-bandwidth product equation. When a high feedback margin at 20 kHz is required, the amplifier needs to have a wide closed-loop bandwidth. Although this is -theoretically- no problem, there are a number of practical problems like i) the cut-off frequency of the power transistors and ii) the processing of radio station signals (and whatever other high frequency litter is around nowadays, contributing to the misery), which can lead to annoying and noticeable artefacts. As the amplifier gets more and more non-linear with increasing frequency, it will generate more harmonic and intermodulation products of the high frequencies which, by itself, may not be audible, but could have a detrimental effect on the total sound image (**N.B.** Note that with linearization of the individual amplification stages, this effect is smaller!). Therefore, the feedback margin cannot be increased "up to infinity", not even to "stratospheric heights". To a certain extent, the strict relation between feedback margin and closed-loop bandwidth can be circumvented: if (in the Nyquist-plot) the point (-1, 0) does not lie within the curve of the $\mu\beta$ product, the amplifier will be stable. With clever tricks, this can be achieved, even if the slope of μ is not a nice -6 dB/oct. and/or the phase shift deviates from -90º. But the price is that the stability of the amplifier will depend on its load (and a loudspeaker is rarely a pure resistor, see below) and that its impulse response is degraded, so additional time smear is introduced (**N.B.** Papers on time smear can be found on www.temporalcoherence.nl). See, for an extreme case, figs. 10 and 11. **N.B.** A similar time-smear as shown in fig. 10, introduced by a tweeter, as illustrated in fig 12, proved to be clearly audible and to degrade its perceived quality (ref. 8)). Reduced temporal resolution might also be the basis for the requirement of "fast" amplifiers as it gives rise to the loss of detail, which can easily be interpreted as "slow" response.

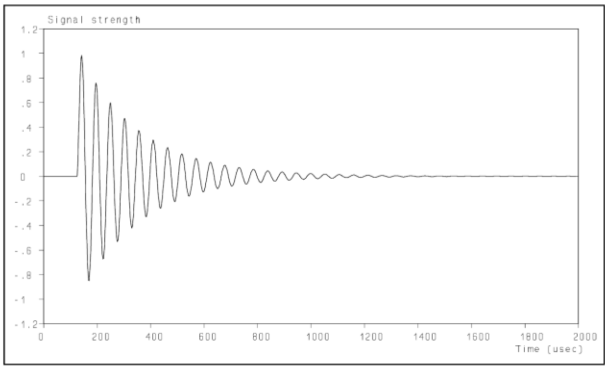

Figure 10: *Impulse response of an amplifier with feedback on the edge of stability.*

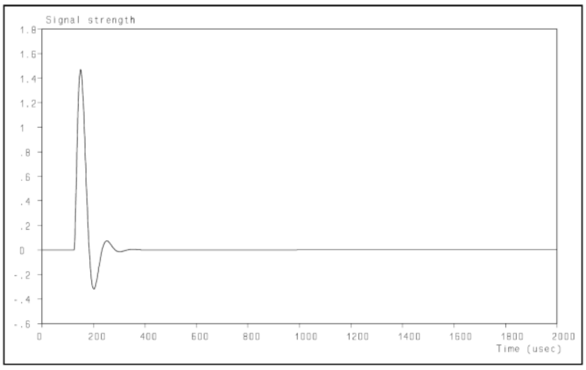

Figure 11: *Impulse response of an amplifier with a lower feedback margin, and thus a higher distortion level, but with a higher temporal resolution.*

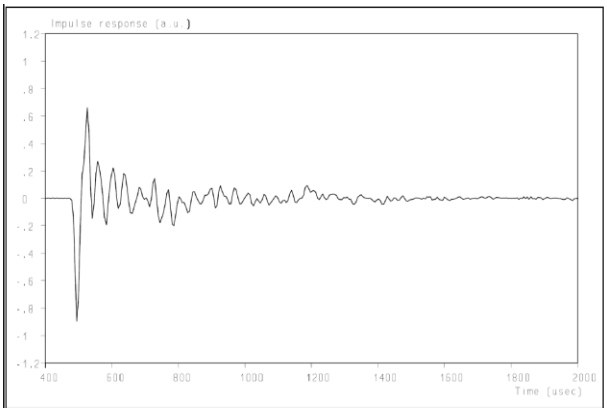

Figure 12: *A similar response as of fig. 10, this time from a tweeter, proved to be responsible for a clearly audible degradation of the perceived quality.*

So it is possible to realise some additional feedback margin, but there are limits on what can be achieved at 20 kHz. It is possible to increase it inversely proportional to the frequency at lower frequencies, but with the disadvantages discussed above. Kolinummi (ref. 2) found that above the open-loop cut-off frequency, the distortion increases far more rapidly than is to be expected, based on the -6 dB/oct. slope, possibly caused by the distortion of the distortion products which are less and less suppressed by the feedback and by the increase of the differential voltage at the input stage, thus enhancing the distortion by the differential amplifier (see fig. 1). So, again, it is probably the best to strive for a constant feedback margin over the entire audio range.

5. The consequences of the loudspeaker impedance

A loudspeaker's impedance is complex, meaning that the voltage and the current are not in phase at most frequencies, which is rather different from the properties of a resistor. So it can happen that the amplifier has e.g. to deliver a *positive* output voltage and a *negative* output current. This creates problems with the control voltages for the power transistors, which may lead to audible artefacts because this needs to be arranged via the feedback. Related to this is the given that a loudspeaker *under dynamic use* is not time-independent because it can store energy (contrary to a resistor!), which can be returned to the amplifier at any time. How the amplifier will react to this phenomenon will strongly depend on its design. But impedance compensation is always attractive. **N.B.** More information on the impedance of loudspeakers and how to compensate its variations can be found in ref. 9 and on the website of "Temporal Coherence" www.temporalcoherence.nl

6. Power supplies

All amplifiers require a power supply, which should be able to deliver the required power and peak currents. But how *stable* should the *voltage* on the power supply rail be? It can be noticed that a varying voltage on the power supply rail will lead to shifting operating points of the individual amplification stages. We have already seen that global feedback is less effective for the first amplification stages. Unknown is how the voltage supply rail variations interact with the non-linearities of the different amplification stages. Also, a "memory" effect is created: the history of the input signal is also determining the voltage at the power supply rail because of the amount of current that has been delivered to the amplifier. This can lead to an “unrest” in the sound stage and to a kind of cross-talk when both channels of a stereo amplifier use the same power supply (**N.B.** Again, these effects are *not* found by using *continuous sinewaves*, simply because it requires a varying amplitude of the input signal, which does not happen with a continuous sine wave analysis). The use of regulated power supplies will solve the majority of these problems, cq. prevent these. The question is just how good this stabilisation needs to be.

N.B. An important working hypothesis, resulting from this work, is that shifting operation points of the individual amplification stages is detrimental to the perceived quality of the reproduced sound. It leads to “unrest” in the reproduced sound, probably because it adds things, like the envelope of the input signal, to the output signal. So the operation points of the individual amplification stages should be kept “rock steady” under *dynamic* conditions. (**N.B.** in this context, *dynamic* refers to the properties of music-like signals, not to AC with a constant amplitude!) Two important sources of such shifts have been identified: i) the non-linear properties of the individual amplification stages and ii) a varying power supply voltage. Elimination of both leads, in general, to a better sounding amplifier.

N.B. One should realise that there is also a frequency dependency, which is often overlooked. The unregulated power rail is re-charged by the (double) power line frequency (50 Hz in Europe, 60 Hz in North America) and this is slow, compared to the usual test frequency of 1 kHz. But at low frequencies, e.g. 20 Hz, the opposite occurs: the current the amplifier requires, varies slowly, compared to the re-charging repetition frequency. So the load of the power supply has rather a DC character when the sonic frequency is low as it, at the peaks, almost doubles the average load of a 1 kHz signal of the same amplitude.

7. Phenomena which cannot be determined by using continuous sine waves

Although some have already be mentioned above, there are several phenomena, which cannot be determined using continuous sine waves of a constant amplitude, against the common knowledge that this should be possible, based on Fourier theory. The following list is not complete, but several important properties are listed:

- Shifting operating points of the individual amplification stages
- Changes in the charge of the cut-off capacitor due to shifting operating points
- Phase modulation caused by a varying closed-loop bandwidth
- Time smear, introduced by the amplifier
- Variations of the power supply voltages due to the changing signal strength / power delivery

From the above reasoning, it should be clear why these cannot be determined by using continuous sine waves of a constant amplitude. The reason why it is in conflict with the common knowledge is that the Fourier theory requires the fulfilment of a number of conditions, which are often disregarded. It also explains the points that Kolinummi (ref. 2) has mentioned. This will be further elucidated in the Appendix.

8. Unconventional development procedure

The development team of “Temporal Coherence” is supported by a listening team, its members do not have a technical background and they are unable to pinpoint the things they hear in a technical sense. But they have a background in music (as musicians) and are thus very familiar with “natural” sounds. So when they develop an opinion on our products, it is always based on the *perceived* quality. Through the years, they have pointed at audible phenomena, generated by our equipment. In most cases, the cause was a mystery at first and discovering it was often a challenging quest, the results of some are reported above.

When the development team thought they had identified the culprit, action was taken to correct the problem, the learnings were in due course used to design amplifiers along the lines, outlined in this paper. This resulted in equipment with a clearly higher rating of the *perceived* quality. The improvements manifested itself in e.g. a better control of the low frequencies, more dynamics, more detail, better sibilance and, yes, a more “musical” sound. Or, if you wish, it sounded less like reproduced music and more like the actual thing. This was confirmed by both our listening team and others (not connected to “Temporal Coherence” in any way, like reviewers of HiFi magazines). It is tempting to assume that the explanations, given here, describe correctly the causes of the improvements, but this is not crystal clear. We do know that the improvements are there, but we have no proof that the presented explanations are realistic and complete. This will require additional experiments, which not necessarily have to be done by ourselves. And if anybody can provide other or additional reasons why designing amplifiers along these lines improves the perceived quality, we will welcome these.

9. Conclusions

It can be concluded that there is nothing fundamentally wrong with (negative) feedback, but systems with feedback are sensitive to a number of parasitic effects which usually get little, if any, attention. So feedback is certainly *not* a miracle cure for all the imperfections of an amplifier, *it can actually introduce new imperfections*. Control and limitation of these parasitic effects requires a careful design of the amplifier with implicit linearization of the individual amplification stages, else audible artefacts will degrade the sound reproduction. The designer should realize that global feedback can only be applied to a limited extent and that the dynamic behaviour of the amplifier to music-like signals is (far) more important for the perceived quality than distortion figures, in line with the findings of refs. 2 and 3. These probably explain a part of the audible differences between amplifiers or other electronic audio equipment, which cannot be understood from the distortion figures and has given feedback a bad name in certain high-end circles. Such artefacts are therefore hard, in many cases not at all, measurable using *continuous* sinewaves. As music is a textbook example of such a *dynamic* process, this is likely to be crucial for the determination of the *perceived* quality of an amplifier. So more complex test signals, which simulate non-steady conditions, as occur in music, are needed. As long such test signals are not agreed on, human hearing is still the best “measurement” instrument available.

Most of the parasitic effects, presented in this paper, are caused by the non-linearity of the individual amplification stages. So these should be suppressed by the linearization of the individual amplification stages by means of local feedback (or other means). By making the open-loop gain of the amplifier as wide-band as possible, yet leaving the closed loop bandwidth not too wide, a constant feedback factor in the whole audio band can be achieved. A moderate global feedback will lead to an optimum design of electronics for high-end audio applications by improving its dynamic response to music-like signals.

The operation points of the individual amplification stages should be kept “rock steady” and this also requires regulated voltages on the supply rails. This is an essential requirement for high-end amplifiers.

Although the development paths of “Temporal Coherence” have resulted in a significant improvement of the perceived quality by introducing modifications along the lines described in this paper, there is no proof that the explanations are correct or complete. Yet, it has been shown that these phenomena do occur and the it seems logical to conclude that these have a negative effect on the perceived quality of amplifiers. Else it would be hard to understand the improvement when these are suppressed.

It can be remarked that critical comments from high-end audio enthusiasts are often scornfully put aside by technology experts as “non-scientific” small talk from freaks who do not understand the theory. The author does not share this view as too often critical remarks from people with “golden ears” did make sense, albeit that it was initially absolutely unclear what the technical or scientific background was. Such remarks did help the development team to further improve the equipment, even though it would have been very hard to show the effect of the individual steps in a scientific way. But the progress over the years is beyond discussion.

Acknowledgements

The author wants to thank especially Menno van der Veen, Jan Didden and Guido Tent for their contributions to the fruitful discussions which have led to this paper. He also wants to thank Prof. Dr. Ir. Ronald van Zolingen for his critical review of the concept and his recommendations for improvements, which certainly helped a lot to raise the quality of this paper. Another special thanks to the members of the listening team of “Temporal Coherence” who have carefully listened to the products. Their critical comments have been *crucial* in the further development and the continuing improvement of the *perceived* quality of the products. Their ears are still the best measurement equipment until we know what we should measure.

APPENDIX

Often disregarded Conditions for the application of the theory of the Fourier Analysis

---

The Fourier Theory is an elegant piece of mathematics, which describes how a signal in time domain can be transformed into a signal in the frequency domain. In a bit more mathematical terms it says that the Fourier Transform (FT) converts a function from the time domain in a *one to one projection* to a function in the frequency domain. The “one to one” projection means that for each function in the time domain there is only one (and only *exactly* one!) corresponding function in the frequency domain. Which is why there is also the way back: the *Inverse Fourier Transform* (IFT) converts a function from the frequency domain in a one to one projection to the time domain. As a side remark it can be noted that the general Fourier theory discusses back and forth transforms of functions from the independent variable x to functions of the independent variable $1/x$. For more details see textbooks like refs. 10, 11 and 12.

The Fourier theory has a number of interesting applications, which are (and will be) used to analyse all kinds of phenomena in audio. However, a number of requirements are connected to the *correct* application of the Fourier theory. Often disregarded conditions are that the system, to which it is applied, is *linear* and *time-invariant* (also internally!). The latter means that the *properties* of the system under study should, a.o., *not depend on its input signal*. When these conditions are not fulfilled, the theory cannot, or only partially, be applied correctly and the results need to be interpreted with great care. Some examples of this will be presented.

Let us first look at the frequency response. A signal in time can be written as an infinite series of sine and cosine waves with increasing frequency. When the response of a system in *frequency domain* is known, its response to any signal in *time domain* can be calculated: the signal in time domain is, by means of an FT, converted to the frequency domain, where it is multiplied with the frequency response curve of the system (resulting in the output signal in frequency domain) which is, by means of an IFT, converted to the time domain. This works well as long as the system is linear and time-invariant. But when the system is not linear, the outcome is incorrect because some phenomena do not show up. Well known examples are envelope detection and intermodulation. These are specifically generated by the non-linearity and when this is not added explicitly, these effects are not found in the output signal. In order to show up, the signal in time domain needs to be multiplied with the non-linearity first before the FT and the IFT are applied. But the calculation of e.g. a non-linear filter is very hard and it can only be done approximately with a limited accuracy. We have encountered that in this paper e.g. when we looked at the phase modulation due to the open-loop gain dependency on the input signal. Another example is given in fig. A-1.

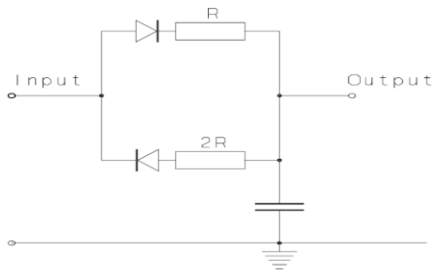

Figure A-1: *Example of a circuit where the Fourier theory is not capable to predict the response, even if the continuous sine wave response would be known by measurement. The reasons are the non-linear behaviour (due to the diodes in the circuit) and the memory, created by the capacitor as its charge will depend on the input signal in the past.*

Another example how one can be fooled by this phenomenon is human hearing. It has been determined that humans cannot hear continuous sinewaves above 20 kHz and from that given it was concluded that, based on the linear Fourier Theory, it is “thus” not necessary to reproduce any frequency above 20 kHz as these would not have any influence on the perceived sound. But for a non-linear system this conclusion is incorrect as, e.g. because of intermodulation, frequencies above 20 kHz can be of influence on the perceived sound (refs. 13 and 14). Human hearing is strongly non-linear and an important consequence is that its *temporal resolution* is much higher than can be expected, based on the measurements with continuous sinewaves. Some experiments indicate that it is an order of magnitude better (refs. 15 and 16). Therefore it should not come as a surprise that especially metal percussion instruments show this clearly, but, alas, most microphones and tweeters are insufficiently at level with the temporal resolution of human hearing. With the result that many people have the opinion that the “high resolution” digital formats are unnecessary because they cannot hear any difference, without asking *where* the limiting factors are to be found ☹. More information on this subject can be found at www.temporalcoherence.nl

When a system is not time-invariant, this usually means that the properties of the system depend in one way or another on what happened in the past. Because the Fourier theory is an *integral* method, one of the implicit conditions is that the system behaves, over the time span of the calculations, the same. An amplifier does in many cases- not behave like that, so before applying Fourier theory, one should ask how well it is approximating a time-invariant system. Memory effects can sneak into the design, which make the behaviour of the amplifier a function of the input signal in the past, even if overload is excluded. Components which are involved in memory effects often are capacitors and coils because these are able to store energy. It should not come as a surprise that these play a role in the discussed memory effects: the cut-off capacitor, the power supply and the loudspeaker loading. In combination with non-linear effects, the memory-effect can become more pronounced (see e.g. fig. A-1). During the design of high-end audio electronics, the potential memory-effects should be seriously considered, especially the parasitic ones, which easily escape attention.

It should be clear that when the condition of time-invariant properties is not fulfilled, results, based on the Fourier theory, can be thrown straight into the waste paper basket. Regretfully, this condition is rarely respected and without hesitation, the frequency response, determined with continuous sinewaves, is interpreted as if it were from a time-invariant system. But too often this is not the case, which is why the behaviour of the amplifier with *dynamic* signals differs from the (desired) behaviour, expected on results obtained with *steady, continuous* signals. To reproduce complex and dynamic signals like music well, the amplifier needs to be -next to a large number of other conditions- also as much as possible time-invariant. If not, artefacts will show up which manifest themselves mostly in the *time domain* and lead to a degradation of the sound stage and thus of the *perceived* quality. Which brings us back to the three citations in the Introduction of this paper.

References

1. Bruno Putzeys, “The F-word or, why there is no such thing as too much feedback”, Linear Audio Volume 1 – Circuit design, pp. 112 – 133, April 2011, Linear Audio (Belgium)

2. Dr. Arto Kolinummi, “Audio Power Amplifiers”, “subtitle “Towards inherently linear amplifiers”. Linear Audio (Belgium, 2011), ISBN 9789490929091

3. Andrzej Dobrucki et. al, “Subjective and Objective Evaluation of the Audio Vacuum-Tube Amplifiers”, AES Convention, Munich, Germany, May 2009

4. H.S. Black, “Feedback Amplifiers”, Bell Laboratories Record, June 1934, Vol. XII, No. 10

5 .M. Leach, “The Cascode Differential-Amplifier Input Stage”, http://users.ece.gatech.edu/mleach/lowtim/instage.html

6. M. Otala, “Non-linear distortion in audio amplifiers: Why do some amplifiers pass static distortion tests but fail listening tests?”, Wireless World, January 1977, pp. 41 – 43

7. Robert R. Cordell, “Phase Intermodulation Distortion Instrumentation and Measurements”, Journal of the Audio Engineering Society, Vol. 31, No. 3, 1983 March, pp. 114 - 123

8. Hans R.E. van Maanen, “Presentation Hans van Maanen AES Berlin 2017”, www.temporalcoherence.nl Workshop3.pdf

9. Hans R.E. van Maanen and E.T. Zonneveld, “An extended Model for Impedance and Compensation of Electro-dynamic Loudspeaker Units and an Algorithm for their Determination, paper 3823, 96th AES convention, Amsterdam (Netherlands), 1994

10. A. Papoulis, “The Fourier Integral and its Applications”, McGraw-Hill Book Company, New York (1962)

11. A. Papoulis, “Signal Analysis”, McGraw-Hill Book Company, New York (1984)

12. V. Čižek, "Discrete Fourier Transforms and their Application", Adam Hilger Ltd., Bristol (1986)

13. V.M. Eguíluz, M. Ospeck, Y. Choe, A.J. Huspeth and M.O. Magnasco, "Essential Nonlinearities in Hearing", Physical Review Letters 84, 22 (May 2000)

14. David Griesinger, "Perception of mid frequency and high frequency intermodulation distortion on loudspeakers, and its relation to high-definition audio", www.davidgriesinger.com/intermod.ppt

15. Milind N. Kunchur, "Temporal resolution of hearing probed by bandwidth restriction", Acta Acustica, Vol. 94 (2008), pp. 594 – 603

16. Milind N. Kunchur, "Audibility of temporal smearing and time misalignment of acoustic signals", Technical Acoustics, Vol. 17 (2007)