Rating Nonlinear Behavior in Acoustic Echo Control Scenarios

Ingo Schalk-Schupp[1,2], Patrick Hannon[1], Friedrich Faubel[1], Markus Buck[1], Andreas Wendemuth[2]

[1] *Acoustic Speech Enhancement Research, Nuance Communications, Germany*
Email: {Ingo.Schalk-Schupp, Patrick.Hannon, Friedrich.Faubel, Markus.Buck}@nuance.com

[2] *Lehrstuhl Kognitive Systeme, Otto-von-Guericke-Universität Magdeburg*
Email: Andreas.Wendemuth@ovgu.de

Nonlinear behavior in the amplifier of embedded audio systems remains a challenge in acoustic echo control. Echo cancellation or suppression methods are usually evaluated using the well-known echo return loss enhancement as a performance measure and the equally well-known total harmonic distortion to characterize the severity of the system's nonlinear distortion. However, the latter fails to capture nonlinear behavior under excitation signals with amplitude and frequency distributions different from that of a simple sine wave.

We propose a novel nonlinearity measure and examine its usefulness in different scenarios in relation to existing measures.

Also, we present an appropriate procedure that comprises the measurements and calculations necessary to determine the proposed measure. Finally, data obtained from an example measurement series are presented and discussed.

1 Introduction

While several metrics for the severity of nonlinear behavior exist [1], their usefulness depends on the intended application.

The well-known total harmonic distortion (THD), and its more practically-motivated counterpart total harmonic distortion including noise (THD+N), can reflect a system's harmonic distortion given a fixed frequency. Usually, a sine signal at a base frequency of $f_0 = 1000\,\mathrm{Hz}$ is used as the excitation signal, and its amplitude is either constant or swept over a range of interest. The result is a distortion factor obtained from multiples of the base frequency.

Broadband measures like the signal-to-distortion power ratio (SDR) mentioned in [1] exist, which allow arbitrary excitation signals and take into account any type of distortion. They are usually designed to explain the measured signal linearly to the greatest extent possible, while the remnants are defined as the nonlinear part. Their respective powers are compared to obtain a characteristic for the amount of distortion introduced by a system given a specific excitation signal.

It is obvious that systems with a frequency-independent nonlinear behavior can be adequately described with simple THD class measurement, while systems with more complicated types of nonlinear behavior require a broadband measurement.

A special case, where higher harmonics are reflected at the Nyquist frequency, is covered by the THD+N approach. In the case that these harmonics are attenuated by an antialiasing filter for strong distortions, this slips the notice of both THD measures. This can lead to an apparent, but erroneous, decrease in THD for very high amplitudes.

In acoustic echo control (AEC) scenarios, the acoustic channel, as a part of the measurement path, generally exhibits an uneven frequency response. Sound reflections can lead to destructive interference at certain frequencies. It is therefore advisable to utilize broadband measures in these cases.

In [2], a measure was introduced that turns out to be appropriate for the problem at hand.

1.1 Model Definition

It is common to assume a parallel model for the separation of linear and nonlinear system components [1]. One degree of freedom remains that determines what is considered the linear system's output and hence, the nonlinear system's output as well.

Schüßler [3] separates both parts by defining their respective output signals to be uncorrelated. While this approach is favorable for the tractability of theoretical analyses, it has no underlying physical model: the linear subsystem explains the whole system's output as good as possible, which is a consequence of the well-known minimum mean square error (MMSE) approach.

We describe a different separation approach that is based on the observation that typical weakly nonlinear systems exhibit adequate linear behavior up to a certain excitation signal amplitude.

It is our basic idea that if a system is linear then its

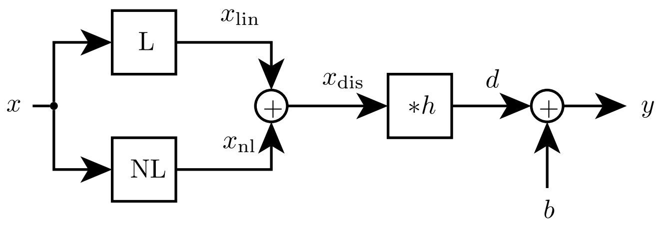

Figure 1: Echo path model. A parallel stage with linear and nonlinear subsystems in cascade with a linear room impulse response. The reference signal x is processed by both subsystems, resulting in a linear portion x_{lin}, and a nonlinear portion x_{nl}. Their sum is the distorted reference signal x_{dis}. Convolution with a common room impulse response h results in the echo signal d. Addition of local noise b yields the measurement signal y.

energy return

$$\frac{E_y}{E_x} = \frac{\sum_n y^{(\nu)}(n)^2}{\sum_n x^{(\nu)}(n)^2}, \quad \text{where } x^{(\nu)} := a^{(\nu)} x^{\mathrm{ref}} \qquad (1)$$

must necessarily be invariant to an arbitrary amplitude scaling factor a. Here, we disregard measurement noise.

We can use this to define any deviation to be a result of the system's nonlinear portion. Additionally, we arbitrarily choose the linear parallel path to be neutral, that is:

$$x_{\mathrm{lin}} := x \quad \Rightarrow \quad x_{\mathrm{nl}} = x_{\mathrm{dis}} - x, \qquad (2)$$

thus necessarily modifying the definition of the cascaded room impulse response h, the modification of which we now call $\tilde{h}$ [2].

2 Determination of the Linear Operating Range

Typically, weakly nonlinear systems exhibit linear behavior at sufficiently low amplitudes up to measurement precision. Beyond a certain amplitude, nonlinear behavior sets in gradually or abruptly. However, at very low amplitudes, the signal-to-noise power ratio (SNR) deteriorates due to a constant background and measurement noise. Hence, completely clean measurements cannot be obtained.

2.1 Principle

By measuring the system's energy response over the full amplitude range, a coarse demarcation of these areas is obtained. Figure 2 shows a typical energy return curve with a pole around zero amplitude documenting low SNR, a constant range marking linear operation, and a downward bend, which indicates clipping characteristics.

2.2 Measurement Signal

It is advisable to use a broadband signal as the excitation or reference signal, since a frequency dependent nonlinear behavior might be present. Moreover, because saturation effects are a commonly encountered type of nonlinearity, a signal with a well-defined amplitude interval is desirable. Inter-sample peaks are inherent to many analog signals generated from digital signals like random or pseudo-random sequences. Amplifier saturation is usually triggered by the presence of high analog amplitudes. It is therefore inconvenient to characterize this nonlinear behavior with such signals. Conversely, an interpolated digital sine signal never exceeds its nominal amplitude, even for frequencies near the Nyquist frequency.

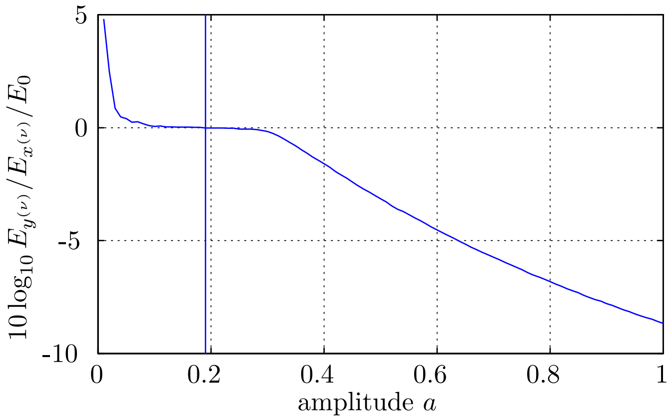

Figure 2: Relative energy return for low noise power with outliers. Energies are normalized by E_0 at a position chosen in the flat curve range. An example amplitude in the obvious linear range between low SNR and nonlinear clipping is marked with a vertical line.

Hence we choose an exponential sine frequency sweep for a number of equidistant amplitudes in the range from zero to digital full scale. As Weinzierl et al. [4] state, the sweep's shape has a possible effect on SNR depending on the type of background noise. Its power spectral density (PSD) also constitutes a frequency weighting of the assessed nonlinear effects. An exponential shape has a pink spectral magnitude distribution and was chosen to emphasize these effects according to naturally occurring signals, such as speech.

2.3 Procedure

The reference signal $x(n)$ is generated at a sample frequency f_{s} with at least 16 bits amplitude resolution. Each sweep ranges from f_0 to the Nyquist frequency $f_{\mathrm{Ny}} = f_{\mathrm{s}}/2$ and takes $\Delta t_{\mathrm{sweep}}$ to complete. A pause of duration $\Delta t_{\mathrm{pause}}$ is allowed between individual sweeps. We divide the available symmetric digital amplitude scale (for 16 bits signed integer resolution, this would be $a_{\min} = -2^{15} + 1$ to $a_{\max} = +2^{15} - 1$) into a number n_a of equidistant amplitudes:

$$a^{(\nu)} := \frac{\nu}{n_a} a_{\max} \quad \text{for all } \nu \in [1, n_a] \cap \mathbb{N}. \qquad (3)$$

All parameters should be chosen according to the situation at hand. For example, the pause duration $\Delta t_{\mathrm{pause}}$ should allow for the longest expected room response, and the sweep duration $\Delta t_{\mathrm{sweep}}$ should be extended if higher background noise is expected.

After recording the signals, an energy return curve plot as in figure 2 provides information about the linear and nonlinear ranges of the device under test (DUT): Linear operation at low amplitudes exhibits a flat return curve, while the energy return might decay for higher amplitudes due to clipping. A reference amplitude a^{ref} with corresponding index ν^{ref} is chosen from the higher linear range to provide linear operation at a good SNR.

3 Linear System Identification

The linear system part can be identified by performing measurements at the chosen reference amplitude a^{ref}.

This measurement can either be performed as a separate step after the linear operating range has been determined, or the previous measurement can be performed with the signal described below simultaneously.

As described in [3], a robust way to identify a linear system is to use a full-band periodic excitation signal

$$\underline{x}^{\mathrm{per}} := \left[\underline{x}^{\mathrm{T}}, \ldots, \underline{x}^{\mathrm{T}}\right]^{\mathrm{T}}, \tag{4}$$

where

$$\underline{x} := [x_0, \ldots, x_{N-1}]^{\mathrm{T}} \tag{5}$$

is an exponential sine sweep signal of length $N = 2^m$, $m \in \mathbb{N}$ and amplitude a^{ref}. The excitation signal $\underline{x}^{\mathrm{per}}$ is then a P-periodic signal of length NP. A synchronous measurement provides the microphone signal $\underline{y}$, which we partition accordingly:

$$\underline{y}_p := \left[y_{pN}, \ldots, y_{(p+1)N-1}\right]^{\mathrm{T}}. \tag{6}$$

Barring the first and last partitions, and not taking measurement noise into account, the partitions are periodic. We reduce random uncorrelated noise by averaging the partitions:

$$\underline{\hat{d}} := \frac{1}{P-2} \sum_{p=1}^{P-2} \underline{y}_p. \tag{7}$$

The system's transfer function h can then be estimated using this estimate and a single excitation period in the discrete Fourier domain:

$$X_k := \mathrm{FFT}(\underline{x})_k, \quad \hat{D}_k := \mathrm{FFT}(\underline{\hat{d}})_k, \tag{8}$$

$$\underline{\hat{H}}: \ \hat{H}_k := \frac{\hat{D}_k}{X_k}, \tag{9}$$

where k is the sub-band index.

Due to the signal's periodic nature, the sample-domain room impulse response (RIR) can be directly obtained by an inverse discrete Fourier transform:

$$\underline{\hat{h}}: \ \hat{h}_{n'} := \mathrm{IFFT}(\underline{\hat{H}})_{n'}, \tag{10}$$

where n' is the filter tap.

This RIR estimate is later used in its discrete Fourier domain only.

4 Nonlinear System Assessment

Having identified the purely linear system, we can now construct would-be linear system responses $\underline{\hat{d}}$, which would be observed if the whole system were linear. Any deviation from these predictions is classified as contribution of the nonlinear sub-system as described in figure 1.

We use or re-use excitation signals $\underline{x}$ scaled by a set of equidistantly distributed amplitudes $a^{(\nu)}$ as defined in (3):

$$\underline{x}^{(\nu)} := a^{(\nu)} \underline{x}, \tag{11}$$

as well as the corresponding measurements $\underline{y}^{(\nu)}$. We then define the linearly predicted signal $\underline{\hat{d}}^{(\nu)}$, which should be proportional to the excitation amplitude, for each amplitude index ν:

$$\underline{\hat{d}}^{(\nu)} := \frac{a^{(\nu)}}{a^{\mathrm{ref}}} \underline{y}^{(\nu^{\mathrm{ref}})}. \tag{12}$$

Here, we rely on the reference amplitude a^{ref} and corresponding index ν^{ref} obtained in section 2.

We can then calculate the nonlinear sub-system's response after convolution with the RIR in the discrete Fourier domain:

$$\underline{\hat{Y}}_{\mathrm{nl}}^{(\nu)} := \underline{Y}^{(\nu)} - \underline{\hat{D}}^{(\nu)}. \tag{13}$$

Also, by approximately inverting the RIR's amplitude response, we get an estimate for the nonlinear path's direct output PSD:

$$\hat{\Phi}_{X_{\mathrm{nl}}^{(\nu)},k} := \frac{\left|\hat{Y}_{\mathrm{nl},k}^{(\nu)}\right|^2}{\max\left(\epsilon, \left|\hat{H}_k\right|^2\right)}, \tag{14}$$

where ϵ is a regularization term.

Finally, we estimate the reference-to-nonlinear power ratio (RNLR) per amplitude as defined in [2]:

$$\widehat{\mathrm{RNLR}}^{(\nu)} := \frac{\sum_{k=0}^{N-1} \Phi_{X^{(\nu)},k}}{\sum_{k=0}^{N-1} \hat{\Phi}_{X_{\mathrm{nl}}^{(\nu)},k}}, \tag{15}$$

where the reference PSD is:

$$\Phi_{X^{(\nu)},k} := \left|X_k^{(\nu)}\right|^2. \tag{16}$$

The RNLR characterizes the amount of distortion in the distorted excitation signal x_{dis}. High values indicate near-linear behavior, while low values result from strong nonlinear distortions.

5 Results

We performed the described measurements in a reverberant workshop with both stationary and occasional transient noise. Repeated measurements allowed for robust results. The DUT was a 12 W consumer grade audio amplifier driven at its specified line level on one channel. The amplifier was known to produce distorted signals at high levels and was consistently set to maximum volume. A DI box was utilized to attenuate the amplifier's output power down

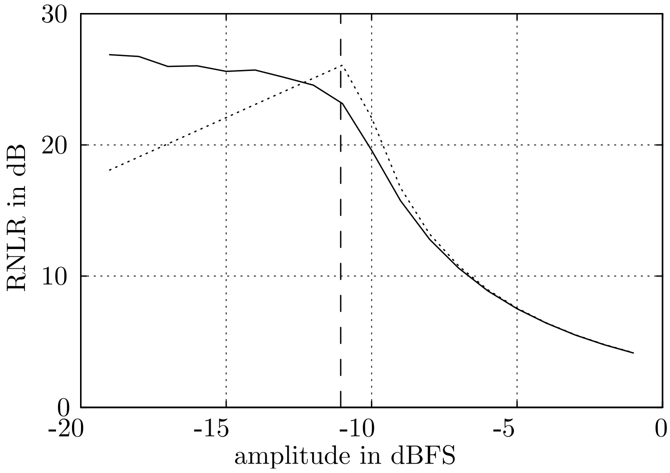

Figure 3: RNLR as a function of the excitation amplitude. The solid line is calculated from actual measurements, while the dashed line is generated from a digitally clipped sine signal, to which artificial random noise was added. The clipping threshold $a_{\text{clip}} = 0.28$ is indicated by a vertical line.

to line level and feed it to an active loudspeaker. The audio signal was picked up by a commercial-grade condenser microphone at a distance of approximately 2 meters from the loudspeaker. Synchronous recordings were achieved by digitally looping back the excitation signal in the soundcard and recording both microphone and reference channels synchronously.

The RNLR curve plotted in figure 3 describes a flat bank in the lower amplitude range and a strong bend downwards at a medium amplitude. It then continues to approach zero RNLR.

This behavior can be explained with amplifier clipping as is demonstrated by the analysis of an artificially clipped signal also shown in figure 3. An RNLR of zero corresponds to “full clipping”, that is, complete deletion of the signal. In this case, the deviation power is equal to the reference power, hence, the power ratio is zero decibels. Inaccuracies can result from any type of deviation from perfectly linear behavior. This includes background noise, the effects of which can be controlled if its PSD is known beforehand, as described above. Limited precision of the estimated RIR is detrimental, but this is mitigated by robust periodic excitation and adequate filter length. Zeros in the inverted RIR PSD can limit the precision of the nonlinear excitation signal power estimate in (14).

As shown in [2], the RNLR is a useful parameter, alongside SNR and echo-to-noise power ratio (ENR), for comparing operating conditions in echo control scenarios. It is an advantage that the excitation signal can be tailored to reflect a desired PSD. Moreover, the RNLR is also invariant to the RIR to some extent. It can therefore help to make different algorithms’ echo return loss enhancement (ERLE) performance comparable under different scenarios without the need for different researchers to use the exact same excitation signals. This cannot generally be done with a THD measure alone.

6 Conclusion

We introduced a novel measurement method useful for characterizing nonlinear behavior of the amplifier or loudspeaker in acoustic echo control scenarios. Through proper design, it is capable of identifying the room impulse response as well as nonlinear effects in a single measurement session. The RNLR characteristic benefits from robust room identification and can be used to numerically compare, characterize, or specify, for example, hands-free telephony scenarios with echo cancellation.

References

[1] G. Enzner, “From Acoustic Nonlinearity to Adaptive Nonlinear System Identification,” in *Proc. of Speech Communication; 10. ITG Symposium*, 2012.

[2] I. Schalk-Schupp, F. Faubel, M. Buck, and A. Wendemuth, “Approximation of a Nonlinear Distortion Function for Combined Linear and Nonlinear Residual Echo Suppression,” in *Proc. of IWAENC*, 2016.

[3] H. W. Schüßler and Y. Dong, “A new method for measuring the performance of weakly nonlinear systems,” in *Proc. of ICASSP*, 1989, pp. 2089–2092.

[4] S. Weinzierl, A. Giese, and A. Lindau, “Generalized Multiple Sweep Measurement,” in *126th AES Convention*. Audio Engineering Society, 2009.