Audio Engineering Society

Convention Paper

Presented at the 115th Convention
2003 October 10–13 New York, New York

Auditory Perception of Nonlinear Distortion - Theory

Earl R. Geddes[1] and Lidia W. Lee[2]

[2]GedLee Associates, Northville, MI 48167, USA

[1]Eastern Michigan University, Ypsilanti, MI 48197, USA

ABSTRACT

Historically distortion has been measured using specific signals sent through a system and quantified by the degree to which the signal is modified by the system. The human hearing system has not been taken into account in these metrics. Combining nonlinear systems theory with the theory of hearing, a new paradigm for quantifying distortion is proposed.

1. Background

A system that does not pass a signal through to its output that is indistinguishable from the input signal is said to distort it. Some signal modifications are desirable, like equalization and frequency response changes to improve the sound quality, but some distortion is undesirable, like most loudspeaker resonances or the nonlinear distortion of a signal. Linear distortion refers to distortions of the frequency response of the system that are linear in their action – they are level and signal independent. Nonlinear distortion of a system refers to distortion that is created as a result of a nonlinear transfer of the signal from the input to the output. Its action is both level and signal dependent. For the purposes of this paper we are only concerned with nonlinear distortion.

A system is said to be nonlinear if its input and output are not linearly related in a mathematical sense. Such systems do not obey the principle of superposition and can have frequency responses that are signal dependent. In fact, even the concept of frequency response is a linear one and its application to a nonlinear system must be done with care. A single tone input to such a system does not produce a single tone at its output, but a multiplicity of tones. Nonlinear systems cannot be analyzed with classical linear systems theory and as such they pose a significant impediment to systems analysis.

This paper will investigate the theory of nonlinear systems with the intent of determining a way to quantify the auditory perception of the sound quality of these systems. We will start with a brief review of the general theory of nonlinear systems and describe the classical measurement metrics that are used to define the sound quality of these systems. We will then introduce some concepts from psychoacoustics that will shed some skepticism on the use of these metrics for the evaluation of sound quality. Finally a new approach to evaluating a nonlinear system will be formulated that is consistent with the psychoacoustic criteria that the classical measures do not account for. A second paper [1] will then show the results of clinical tests which use each of these metrics on simulated nonlinear systems.

2. Nonlinear Systems

The general theory of nonlinear systems is well developed and we would refer the interested reader to one of the numerous texts on the subject for a more in depth discussion [2,3]. In a good audio system linearity is usually one of the design goals, although, there are very real situations where one might want the system to be nonlinear. Our interest in these papers is restricted to those systems where linearity and low distortion is a goal. However, the results that we will show do shed some light on some of the aspects of intentional nonlinear systems that could be useful.

When one considers all of the components in an audio system, it is clear that there are a multitude of mechanisms that can create distortion. Although there are many mechanisms, there is fortunately a fairly simple method for describing the majority of them. We will review that method here.

We will define a function that relates the instantaneous output level of some quantity versus the instantaneous input level of this same quantity and call this function the nonlinearity transfer function $T(x)$. When this relationship is not a straight line, then the system is said to be nonlinear. Several examples of this function are shown in Fig.1. These curves denote what is known in the literature as a memory-less nonlinear transfer function, memory-less because it has no frequency dependence or memory in time. This curve is also sometimes called a static nonlinearity. The importance of this distinction will become apparent later.

The input-output relationship can be between any two variables of the same type, displacement, velocity, current, voltage, any variable of the system. The only requirement is that this "block" must be placed in the domain in which it is defined and it should relate identical quantities (its possible to deviate from this requirement but we will not do so here). These are not the most general restrictions, but they make our analysis simpler without compromising its validity.

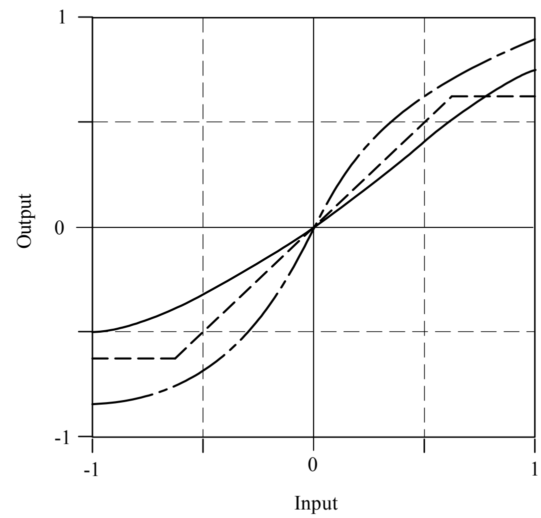

Figure 1 - Nonlinear transfer functions

Fig.1 shows three examples of nonlinear transfer functions. Note that the hard clipping transfer characteristic (dashed line) is completely linear as long as the magnitude of the input remains below .6. However, if we allow the input to go to .7 or 1.0, then the distortion becomes highly dependent on the actual input level. This makes clear an important point that we must always consider. What values are we going to allow as inputs? Or outputs?

If we scale the output values to be unity when they reach some predefined level, x_{peak} (we would have preferred the term x_{max}, but its historical usage, which is inconsistent with our usage here, prohibits us from doing that), then we can see that the output scale would go from -1 to 1. What we use as an input scale is not arbitrary and it should be arranged so that the output level never exceeds ±1 for any valid input level. If this is not done, then there can be an ambiguity (and a failure of the applicability of the theory) in the series expansions that we will use. If the system has a scalable gain, then we can always scale the gain to accomplish this task. If, on the other hand, the system gain is set by the systems characteristics then we must be careful in selecting the allowed input range so as to insure a valid mapping curve.

The curves shown in Fig.1 can be expanded in many different ways. For example we could expand the curves into Legendre Polynomials and study theses expansions, (there are some very good reasons to do that [4]). We could also expand them as Chebycheff Polynomials, or Laguerre Polynomials (as Weiner does [2]) etc. Finally, a Fourier series is even applicable and we will actually use that approach, among others. For our purposes right now, a simple polynomial expansion is attractive because of its simplicity.

We will let

$$T(x) = \sum_n a_n x^n \quad 1)$$

The solid curve in Fig. 1 then has the equation

$$T(x) = .8x + .1x^2 - .2x^3 \quad 2)$$

Here, the gain of the system is given as .8, but since the output does not come close to either ±1, we should actually adjust the gain to be 1.1 to better normalize the curves, or we should readjust the allowed input scaling. The results of any nonlinear analysis depend on the choice of x_{peak} and the gain values. The choice of too large a value for x_{peak} and/or too small a gain will result in a significant change in the order coefficients. It should be apparent that the value of x_{peak}=1.0 in the transfer characteristic of Eq.(2) can never be reached with the input levels as defined. It is always desirable to define x_{peak} in such a way that one of the limits is reached when the limit of the input is reached, i.e. the maximum output should always be normalized to the maximum input. When this is done we will get a more consistent and useful specification for the system.

The first term in equation Eq.(1) is called the *offset* term. None of the curves in the figure have an offset. The second term is the *gain*. The third term is know as the second order or *quadratic* nonlinearity. The fourth term is know as the third order or *cubic* nonlinearity. Higher order terms are simply defined by their powers, i.e. fifth order, x^5 and so on. There is no limit to the number of orders that can be required to represent a given transfer characteristic. For example, the two curves with sharp slope discontinuities would require very high orders to fit them over the range of -1 to 1, which is an extremely important point, as we shall see.

As an example of the effect of a nonlinearity in a system, consider a nonlinear transfer function with only gain and a *quadratic* nonlinearity. We know that for a sinusoidal excitation, the output will contain harmonics of the input. Given an input $x(t)$, this can be shown as follows

$$x(t) = A\cos(\omega t) = \frac{A}{2i}\left(e^{-i\omega t} + e^{i\omega t}\right) \quad 3)$$

then the output $y(t)$ will be

$$\begin{aligned} y(t) &= T(x(t)) = a_{.1}\,x(t) + a_2 x(t)^2 \\ &= a_1 A\left(\frac{e^{-i\omega t} + e^{i\omega t}}{2i}\right) + a_2 A^2\left(\frac{e^{-i\omega t} + e^{i\omega t}}{2i}\right)^2 \\ &= a_1 A\left(\frac{e^{-i\omega t} + e^{i\omega t}}{2i}\right) + a_2 A^2\,\frac{1}{2}\left(\frac{e^{-i2\omega t} + e^{i2\omega t}}{2i} + 1\right) \end{aligned} \quad 4)$$

The output contains the original input scaled by the gain a_1 and a second harmonic, at 2ω scaled by $a_2A^2/2$. If the output is normalized to the input then A can be taken as unity. There is also an offset term in the output that results from all even order nonlinearities. The use of complex exponentials is desirable due to the simplicity of taking powers, but we need to remember that we must always use two complex exponentials (both signs) or we will get an incorrect result. (Why?)

Any signal could be substituted in Eq.(3) and substituted into Eq.(1) with an arbitrarily complex transfer function and the output of this nonlinear system would be specified. However, unlike a linear system, we cannot find the output from a complex signal as the sum of the outputs of simple signals. This simplification, which forms the basis of all linear systems theory, is based on the validity of superposition, which, as we said before, does not hold.

The above procedure also does not allow for a frequency dependence in the nonlinearities such as would occur in a loudspeaker. To handle frequency dependencies we must extend the procedure shown above. We can consider the individual nonlinear terms in Eq.(1) to act as individual transfer function blocks all acting in parallel as shown in the figure below.

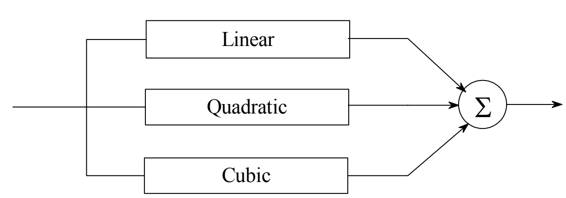

Figure 2 - Nonlinear system with parallel paths

In this figure we have only shown the first three orders for simplicity. The general case would have n parallel blocks where n is the highest orders under consideration. In order to handle a frequency dependence we put transfer functions into each leg of this single input single output system as shown below.

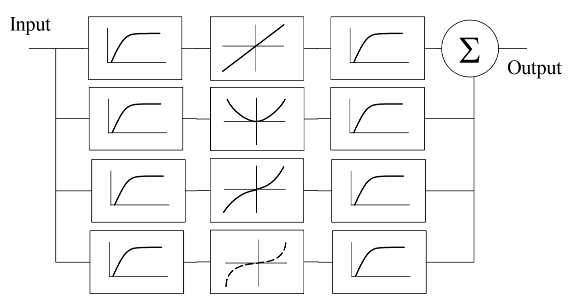

Figure 3 - Schematic Diagram of Nonlinear system

Fig.3 is almost the completely general case of a nonlinear system, although it cannot account for some of the more complex nonlinear effects. Fortunately nearly all nonlinear systems in audio will allow modeling according to Fig.3. The top leg in the above drawings shows two blocks one of which is superfluous since this is the linear leg and the pre and post frequency responses can be made one and the nonlinear transfer function block simply dropped. They have been included to emphasize that the nonlinear legs actually act very much like the linear one.

Individual frequency response blocks before and after the nonlinear transfer characteristic are not always required. These frequency response blocks are used in most cases to take the variables into the domain of the nonlinearity. For example, in a loudspeaker the nonlinearities are predominately in the cone displacement, but the input is a voltage. In this case the first (pre) frequency responses would take the voltage signal into a displacement domain signal where the nonlinear transfer functions operate. This is a simple low pass filter that principly depends on the linear systems parameters (the Theile-Small parameters). For all practical purposes, in a loudspeaker, all of the "pre" – blocks are the same[4].

In an amplifier with crossover distortion and clipping there would not be any frequency response blocks at all – unless the amplifier was of an unusual design.

The point to be made here is that with a specification of the frequency responses of each of the legs shown above one has a complete description of the nonlinear system. In the general case the frequency response functions both before and after the nonlinearity block are required, but, as stated, these are usually not necessary. In the case of a loudspeaker we can adequately specify the distortion by characterizing the frequency response blocks that follow each of the nonlinearity blocks. These frequency responses are those of the nonlinear orders not of the harmonics, the two are quite different things. That is because the orders generate a multiplicity of harmonics – every other one below its order. For example a sixth order nonlinearity will generate a sixth harmonic, a fourth harmonic and a second harmonic.

It is possible to find the order frequency responses from the harmonic frequency responses by using characteristics of the Legendre functions. By expanding the nonlinear transfer function in Legendre Polynomials we can find the relationships between the orders, which are the same as the Legendre Polynomials, and the harmonics. We will not actually show this relationship here since it is not really required for the point of this paper. It can be found in [4]. This relationship does form the cornerstone of measurement techniques of the actual orders.

A review of our approach to nonlinear systems thus far is easily stated. By showing the frequency responses of the nonlinear transfer functions orders, which are related to the harmonics of a sine wave, virtually all of the information that is required to understand how the underlying nonlinear system behaves is available. This is true only for signals which lie within the region of applicability of the expansions involved, but this is not a real limitation since we are free to define that region at will. We only have to be careful not to attempt to apply this analysis to signal outputs which exceed the limits that we have defined.

3. Distortion Metrics

Now that we have seen how we can analyze a nonlinear system in terms of multi-path nonlinear transfer functions, it is desirable to have some way to relate this analytical mathematical approach to the less quantitative subjective perception aspects of the of distorted system.

What we are seeking is a metric of distortion perception. A metric is a value (or it could be a function or multi-valued but a single value is usually desirable) which is given as an attribute of a relationship to indicate its scaling in some predefined context. For instance, temperature can be a metric in the context of human perception heat content. We can describe the perception of temperature in words like hot, warm, cool, or cold and since temperature also has an exact scientific scaling (conveniently), it is a simple matter to map from the subjective metric to the physical one. We must always remember however, that the subjective terms are relative and a precise mapping is often difficult to obtain. Whenever human perception is involved, metrics can only ever be statistically relevant. We can talk about the subjective perception of something, like temperature, which will be stable for people as a group but we can almost never have a subjective metric that will be precisely true for an arbitrarily selected individual.

Historically, the audio community has viewed distortion metrics in the context of the nonlinear systems response to a sinusoid or sometimes, two or more sinusoids - basically a signal based metric. The current metrics that are used for distortion are, Total Harmonic Distortion (THD); Inter-Modulation Distortion (IM), multi-tone inter-modulation, etc., which are all usually expressed as a percentage – the ratio of the distortion by-products to the total system output. In an absolute sense this view of distortion is satisfactory. If our goal were to eliminate all distortion then clearly any measure of its level is adequate. But it is neither reasonable nor desirable to set as our goal the complete elimination of all distortion. From a cost effective standpoint, reducing distortion below perceivable levels is a complete waste of time and money. It may also be that we might want a scale by which to compare two levels of distortion in order to make tradeoff decisions. In this context, we will show that the signal-based metrics fall far short of the mark, for they fail to correlate with, or even consider, subjective impression.

It is well know that one can manipulate the actual waveform of a sound signal to a rather large extent without there being an appreciable effect on the sound quality. This knowledge is a direct result of the massive amount of work recently done in the area of perceptual coders (i.e. MP3). These coders make substantial alterations of the waveform in an attempt to conserve data while having only a very small perceptual effect. Therefore, it seems intuitively obvious to question the belief that a distortion measurement which is based purely on the mathematical difference between specific input and output waveforms, without any regard for the human hearing system, would yield a reliable metric? We think not.

The need for a reliable distortion metric is obvious. With it we could do psychoacoustic studies and determine the same kind of mapping for distortion perception that we described for temperature. Without it we can only guess at what effect a change in a distortion number will have on the perception of that change. To be useful the metric must be consistent and reliable – the same number must mean the same thing in every context and there must be a close correlation between the metric and the response that it is intended to scale.

This is precisely where the signal-based distortion metrics fail. In our next paper we will show that .01% THD of one type of nonlinear system can be perceived as unacceptable while 10% THD in another example is perceived as inaudible. Even one of these simple examples is sufficient to invalidate THD as a viable metric for discussion of the perception of distortion. Furthermore, 1% THD is not at all the same as 1% IM, but we will show that neither correlates with subjective perception. While some of the signal-based metrics may be "better" than others, it is our opinion they all fall short of what we are seeking.

How then does one establish a metric for the quantification of distortion that is consistent, reliable and obtains a high degree of correlation with subjective impression? Based on what we known about the human hearing system and what we have learned about the nonlinear systems analysis, we will propose such a metric.

4. The Psychoacoustics of Distortion Perception

We would like to briefly review a few concepts in the theory of the human hearing system in order to support our usage of these concepts here. The reader may wish to consult other texts if the psychoacoustic terminology or the concepts being used are unfamiliar [5].

One reason that the perception of nonlinear distortion is so complex is that the hearing mechanism itself is not linear and taken as a "system" it is also quite complex. It should thus be expected that it will be a difficult task to ascertain what levels and types of nonlinearity the ear can perceive and even more difficult will be the scaling of the subjective impression of these nonlinear functions.

The attribute of hearing that will overwhelmingly dominate the perception of distortion is masking. Masking is also the principal effect used in the creation of all modern techniques of perceptual coders (MP3, AAC, etc.). When masking effects allows us to reduce signal data by 90% or more, in a way that is subjectively benign, then one has to suspect that masking would have a profound effect on the perception of nonlinear distortion. Masking has no analog in linear systems theory, and it is not very intuitive since it does not occur in systems other than the ear.

From our knowledge of masking there are two fundamental characteristics that are of importance.

- Masking is predominately upward toward higher frequencies although masking does occur in both directions.
- The masking effect increases – masking occurs further away from the masker – at a substantial rate with excitation level.

Given these characteristics we will propose the following three *Distortion Perception Principles*.

1. Distortion by-products that are created upward in frequency are likely to be less perceptible (masked to a greater extent) than those that fall lower in frequency.
2. Distortion by-products that lie closer to the excitation are less likely to be perceived (they are masked) than those that lie farther away (masking is a localized effect – it only occurs in the vicinity of the masker).
3. Distortion by-products of any kind are likely to be more perceptible at lower signal levels than at higher signal levels.

The following discussion relies on these "principles," as its foundation. If one accepts these principles as valid, then what we say in the following sections should have substantial validity.

We should also note the following facts.

- Odd and even orders do not interact, odd orders generate only odd harmonics, even orders generate only even harmonics.
- An *n*th order nonlinearity generates *n*th order harmonics and every other harmonic below it.
- Harmonics of pure tones are generated only above the input signal (this is true only for nonlinear transfer functions which can be represented by Eq.(1), which, fortunately for us, is true for most everything that we will talk about.)
- For multi-tones an nth order nonlinearity causes sidebands at $\pm n$ times the modulation frequency and every other value of n below it, as well as harmonics (as above).

These may all seem obvious since we have already provided the mathematical foundation for these points but the important points comes next.

Consider Fig.4 where typical modulation distortion products are shown for tones in four situations: a low order nonlinearity and a high order nonlinearity, at a low signal level and a high signal level. Approximate masking curves of the principle masker tone are also shown. We can see that the higher order distortion products are not masked as well as the lower order ones and that the masking effect is greater at the higher signal level. The low order distortion at a high signal level is completely masked in this figure. The high order distortion is never masked, but it would clearly be more audible at low levels.

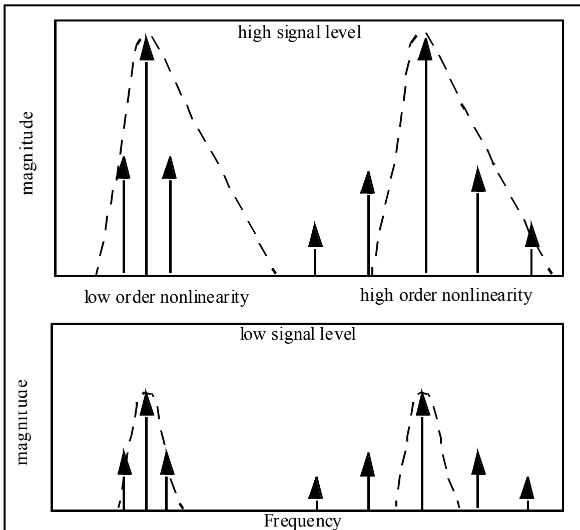

Figure 4 - Masked Distorted Signals

If we take these facts and join them up with our Perception Principles then we can make the following statements, which are, perhaps, not exact, but they are, none the less, more valid than not.

- The masking effect of the human ear will tend to make higher order nonlinearities more audible than lower order ones.
- Nonlinear by-products that increase with level can be completely masked if the order of the nonlinearity is low.
- Nonlinearities that occur at low signal levels will be more audible than those that occur at higher signal levels.

Again these may seem intuitively obvious.

These statements give rise to our hypothesis for a new approach to quantifying nonlinearity (distortion):

> We propose a metric that is based on the shape of the nonlinearity curve that has the following features:

- It should be more sensitive to higher order nonlinearities than lower order ones.
- It should be weighted towards greater values for nonlinearities at lower signal levels.
- It must be immune to changes in offset and gain (first order slope) since these are inaudible effects.

To meet the first objective we propose using the second derivative of the nonlinear transfer function since this function increases in value according to the square of the order. To alleviate a sign problem with this value we propose squaring this term. This function also addresses the third requirement above. To meet the second requirement we propose using a cosine-squared function which is unity for small values of the signal and zero for the largest ones. Finally we would propose integrating this function from -1 to 1 (the range of the output signal) to yield a single number which we will call G_m – the GedLee metric. The exact equation is as shown below.

$$G_m = \sqrt{\int_{-1}^{1} \left(\cos\left(\frac{x\pi}{2}\right)\right)^2 \left(\frac{d^2}{dx^2} T(x)\right)^2 dx} \qquad (5$$

Eq. 5 represents the central hypothesis of this paper – a proposed metric for distortion which is based in the general theory of nonlinear systems and takes into account the characteristics of human hearing.

It is important to note that Eq.(5) is actually a property of the *system*, not of a *signal* sent through the system. It is completely independent of the actual input signal sent through the subject system and is thus valid for any signal. This is a very attractive feature since, like linear systems theory, it is desirable to be able to describe the performance of the system in a way that is independent of the actual signal sent through the system.

We should say a word about Eq.(5) regarding how it is applied. In the form shown it requires a knowledge of the shape of the nonlinear transfer function $T(x)$, as defined in Eq.(1). As shown it is basically a frequency independent measure. There is no ambiguity in performing the calculations at a particular frequency, but in a real system $T(x)$ can be a frequency dependent, i.e. $T(x, f)$, in which case $G_m(f)$ will also be frequency dependent. This is completely analogous to THD being frequency dependent, even though we usually specify its value at only a single frequency. The analysis of frequency dependent G_m's will not be taken up in this set of papers. This must be left for future work. In all that follows we will assume that the systems under discussion have a G_m which is either independent of frequency or we are considering a single frequency or worst case value.

Consider now an example of the failure of THD to differentiate between loudspeaker distortion and amplifier distortion as we alluded to above. If the amplifier has crossover distortion then this type of nonlinearity violates both of our principles – it is both very high order and it increases (as a proportion of the linear terms) with decreasing signal level. Based on our hypothesis, one would expect that this type of distortion would be highly objectionable and it is. Now consider the loudspeaker example. Unless it has some severe design or manufacturing problems, it will mostly have lower orders of nonlinearity and the distortion will typically rise with level. Based on our principles, we should expect this type of distortion to be fairly benign, almost inaudible, and this is in fact what we find to be true (for comparable levels of THD for the loudspeaker and the amplifier). Generally speaking, electronics and mechanics have different nonlinear characteristics. It is not at all uncommon to see very high orders of nonlinearity in electronics, but it is rare to see these very high orders in mechanical systems. Our new view of distortion explains a lot of the THD based metric paradoxes.

As we have said, this new metric G_m is based on the actual shape of the nonlinear transfer function itself, but it should be noted that there is a strong relationship between the frequency response curves of the orders as defined in Fig.(2) and the shape of the nonlinear transfer function. It should be clear that to get a good estimate of G_m one needs to know the nonlinear orders to a fairly high order. This is because of the second derivative in the equation for G_m, which makes it more sensitive to the higher orders than the lower ones.

It is not uncommon to see discussions of 2nd and 3rd order nonlinearity, but it is rare to see much discussion of the higher order ones. In fact we have even seen this problem simplified to a discussion of symmetric versus non-symmetric nonlinearities. Since we are hypothesizing that increasing orders are more audible than lower orders then limiting the discussions to only the lower two orders is seriously flawed and discussing nonlinearity in the simplified terms of symmetric or non-symmetric is even worse.

The *root cause* of distortion is the nonlinearity of the system and the correct way to discuss nonlinearity is with the orders of its nonlinear transfer function. When one views the distortion problem in this way, signal based distortion metrics (THD, IMD, etc.) become irrelevant. As we will show in the next paper, the relationship of the signal-based metrics to subjective impression is virtually nonexistent. It is our hope that the audio community will give the outdated notion of THD, IM, signal types, etc. (signal-based concepts) as these are all just symptoms of the real problem – nonlinearity.

5. ACKNOWLEDGMENT

This work was supported by a grant by Eastern Michigan University and GedLee LLC.

6. REFERENCES

[1] L.W. Lee and E.R. Geddes, "*Auditory Perception of Nonlinear Distortion*," presented at the 115th Convention of the Audio Engineering Society, (October, 2003)

[2] Schetzen, M., *The Nonlinear Theories of Volterra and Weiner*, Krieger Publishing Co., Malabar, Florida, 1989.

[3] Bendat, J.S., *Nonlinear Systems Techniques and Applications*, John Wiley & Sons, New York, NY, 1998.

[4] Geddes, E.R, *Audio Transducers*, GedLee Publishing, Novi, MI, 2002.

[5] Moore, B.C.J., *Psychology of Hearing*, Academic Press, San Diego, CA, 1997.