Audio Engineering Society

Convention Paper 6910

Presented at the 121st Convention
2006 October 5–8 San Francisco, CA, USA

This convention paper has been reproduced from the author's advance manuscript, without editing, corrections, or consideration by the Review Board. The AES takes no responsibility for the contents. Additional papers may be obtained by sending request and remittance to Audio Engineering Society, 60 East 42nd Street, New York, New York 10165-2520, USA; also see www.aes.org. All rights reserved. Reproduction of this paper, or any portion thereof, is not permitted without direct permission from the Journal of the Audio Engineering Society.

Assessment of Nonlinearity in Transducers and Sound Systems – from THD to Perceptual Models

Alex Voishvillo[1]

[1] JBL Professional, 8400 Balboa Blvd., Northridge, CA 91329, USA
avoishvi@harman.com

ABSTRACT

Research of the audibility of loudspeaker nonlinear distortion has not shown good correlation between traditionally used metrics (harmonics and intermodulation) and subjective performance. The problem of sound fidelity-related methods to assess nonlinearity in transducers has not been solved. Wide application of low-bit rate compression systems (MP3, etc.) demanded the development of objective measurement methods based on perceptual models. These methods however have not been used for measurement of loudspeakers and they may not be optimal for that due to the different nature of the nonlinearity in transducers. Recently perceptual models created specifically for the assessment of nonlinearity in transducers have emerged. In this work analysis of the old and new methods, their comparison, and the prospects for future developments are discussed.

If any manufacturer or group of manufacturers can carry out the necessary research required to correlate listening tests with various methods of measuring nonlinear distortion, it will be a great and valuable service for the industry.

Herman Hosmer Scott, "Intermodulation Measurements", JAES, 1953

1. INTRODUCTION

The history of the search for objective methods of measurement of nonlinear distortion in sound equipment is as old as the audio itself. Research into the audibility of sound system nonlinear distortion has not revealed good correlation between the traditionally measured nonlinear distortion of a loudspeaker and its subjective performance. Conventional methods of nonlinear distortion measurement using sweeping tones provide information that only partially correlates with perceived sound quality. The results of various studies often suggest different thresholds of distortion audibility. The problem of credible, sound fidelity-related methods to assess nonlinearity in sound systems (and in loudspeakers in particular) has not yet been solved. The problem includes many aspects and it is very complex. The problem is also interdisciplinary and it encompasses such areas as theory of dynamic nonlinear systems, electroacoustics and psychoacoustics. There is an abundant literature on theory of nonlinear systems and psychoacoustics, but most of it is written by specialists in their fields and is addressed to the specialists in these fields. The theory of nonlinear dynamic systems is based on rather complex mathematical apparatus not used on a regular basis by loudspeaker engineers, and psychoacoustics is based on specific knowledge, terminology and principles that may not be well familiar to loudspeaker engineers either.

This work attempts to bridge the needs of loudspeaker engineer and the pertinent knowledge in the aforementioned unrelated disciplines and give a concise review of the existing approaches to assessment of nonlinearity in audio. Also a history of the subject is briefly reviewed and covers last 90 years of the evolution of the technology beginning from measurement of the total harmonic distortion (THD) at the dawn of audio and ending with modern objective methods based on perceptual models and psychoacoustics of human auditory system. The glossary and mathematical definitions of different measurement metrics are given in the Appendices.

2. BACKGROUND

The objective of any method of measuring nonlinear distortion in sound equipment is to provide a numerical and graphic representation of the nonlinear properties of a device under test (DUT), such that certain judgments about its performance can be made and different devices can be compared objectively. Two impediments complicate this task. The first is the enormous amount of data needed to precisely describe the behavior of a dynamic nonlinear device such as a loudspeaker or horn driver. This differentiates any nonlinear system from a linear one, whose behavior is completely described by the impulse response or complex transfer function. (Both characteristics depend on space coordinates in the case of a sound-radiating device). The second impediment is the very complex reaction of the human auditory system to speech or musical sounds contaminated by distortion products.

Underestimating the complex and "multidimensional" nature of nonlinearity, and the intricacy of the human hearing response to a distorted musical or speech signal can lead to wrong conclusions. A typical misconception in the analysis and assessment of nonlinearity in audio is the confusion between the symptoms of nonlinearity obtained through the application of certain testing signals (typically sinusoidal ones) and the reaction of the audio system to non-stationary speech and musical signals. Audibility thresholds obtained for distorted tonal signals are sometimes erroneously considered meaningful data from the standpoint of the expected sound quality of the musical signals reproduced. The reaction of a nonlinear system to a particular testing signal such as pure tone, is sometimes mistakenly generalized as the nonlinear system's and hearing system's reactions to a musical signal.

Very often the reaction to sinusoidal signal (i.e. harmonics) is thought to have affinity with nonlinear products accompanying reproduction of speech and musical signals. In case of a single sinusoidal input signal, the "contamination" signals are the harmonics, whose level, phase and order depend on the properties of particular nonlinear system and the level of the input signal. In the case of a non-stationary non-periodic input signal such as speech or music the output "contamination" signal is essentially the non-stationary, non-periodic, and sometimes noise-like signal as well. Obviously, it is pointless to approach to such "contamination" or "distortion" signals from the standpoint of harmonic and intermodulation products.

Another typical example of a popular misconception is that a pronounced second harmonic distortion produced by a loudspeaker is essentially benign because this distortion generates signals in the

octave harmonic consonance with the fundamental tones. Further, that this distortion may even ameliorate subjectively perceived sound quality. Proponents of this misconception ignore the fact that once the sinusoidal signal is replaced by a real musical signal, the same nonlinear loudspeaker exhibiting the dominating second-order harmonic distortion will generate an enormous amount of instantaneous intermodulation products that are not harmonically related to the input signal. This line of thought about "mellifluous" second order distortion is an example of a popular misunderstanding and simplistic interpretation of the complex phenomena of nonlinearity and the perception of nonlinear distortion by the human hearing system.

There is a dilemma in measuring, graphing, and interpreting nonlinear distortion as it relates to subjective detectability. On the one hand the assessment of nonlinear distortion needs the analysis of much more information than is required to assess a linear system. On the other hand, this information should be presented in a comprehensible graphical or numerical manner. These two requirements may contradict each other. Furthermore, the graphed or numerical data should be pertinent from the standpoint of distortion audibility. The final goal of a loudspeaker nonlinear distortion measurement is to obtain data that conveys information about the nonlinearity so that this information can be related unambiguously to the perceived sound quality of a loudspeaker under test, and thus the performance of different loudspeakers can be compared objectively. The measurement data must be "manageable". In spite of the seeming simplicity of these goals, and a nearly 90-year history of numerous efforts of many researchers these goals have never been achieved.

During the last two decades so a new approach to the objective assessment of low-bit compressed audio and speech signal has emerged. This approach, based on psychoacoustic models of the human hearing system, allows one to process a speech or musical signal in a way that is similar to the functioning of the human auditory system. Taking into account such auditory system's properties as temporal and frequency masking, these perceptual methods make it possible to determine which components of the objectively distorted signal are psychoacoustically audible and which are not. There is a number of variations of this approach. The brief history of the subject will be reviewed in the "History" section. In general, the first works in the late 1980s targeted assessment of quality of speech signals in communication systems. These methods received fast development in the 1990s to assist objective evaluation of sound quality of codecs for low-bit compression of musical and speech signals (MP3, AAC, WMA, ATRAC, etc.). These methods, however, have never been used to assess sound quality of electroacoustical transducers whose nature of nonlinearity is significantly different from that of the low-bit compression schemes. In addition, the "output" of such approach does not have a "familiar" graphical presentation in terms of some frequency or level-dependent responses, and it does not put a bridge between the nonlinear effects responsible for the nonlinear contamination of loudspeaker-reproduced sound and the audibility of nonlinear distortion.

In addition to these methods of objective evaluation of codecs, several new methods were developed and reported lately by Tan, Moore, Zakharov, and Matilla. These methods exploit perceptual models for evaluation of nonlinear and linear distortions (and their combinations) produced by artificial nonlinearities and real electroacoustical transducers. Due to their importance, these works will be considered in detail in the "History" section of this work.

During last few years new ideas based on analysis of musical signals that passed through the static nonlinearity of the known character led to new metric better correlated with perceived sound quality than traditional harmonic and two-tone intermodulation distortion characteristics. This approach, originated by Geddes and Lee is based on idea that higher order nonlinearity produced distortion that are likely to detected by hearing system and on the assumption that a static nonlinearity that generates distortion at low-level of reproduced signal is significantly more detrimental to subjective sound quality than a nonlinearity that becomes significant at high levels of signal and does not adversely affect low levels of the input signal. The crux of the approach is the fact that the low level musical signals are poor maskers, especially in the presence of high-order nonlinearity, and in addition, the probability of the occurrence of the low-level signals is significantly higher that that of the high level ones. Independently from Geddes and Lee similar results were obtained a year ago in informal experiments and reported in JBL Pro. The way it is, Gedd-Lee metric is applicable only for analysis of static polynomial nonlinearities whose mathematical description is known. The approach will be reviewed in detail in the section on the history of the subject.

3. PROPERTIES OF NONLINEAR SYSTEMS

Although the mathematical concepts to describe dynamic nonlinear systems such as Volterra series expansion were developed at the beginning of the 20th century, if not earlier [1] (let alone nonlinear differential equations that were introduced centuries ago), the nonlinear theory for engineering applications began with the works of Wiener in the mid 1940s. The theory of the weakly nonlinear dynamic systems based on works of Volterra and Wiener was developed around 1950s – 1960s. At about the same period of time, the psychoacoustics of human hearing system matured and the basic generally accepted properties such as excitation patterns, critical bands and masking effects were researched and well understood. However, the theory of the "nonlinear loudspeaker" lagged behind the mainstream nonlinear theories. For example, the works of Schetzen on Volterra and Wiener nonlinear theory were published in the 1960s – 1970s and were later compiled in his book in 1980s [2], whereas the first work on application of Volterra series to loudspeaker analysis was published by Kaiser in the mid 1980s [3]. Significant progress in the loudspeaker nonlinear theory began in the early 1990s with the works of Klippel [4, 5, 6, and 7].

When such terms as "nonlinearity", "nonlinear system", or "nonlinear distortion" are met in audio literature, they often produce an image of a graph where the X-axis shows the level of the input signal and the Y-axis corresponds to the level of the output signal, and the relationship between them being some polynomial or "broken" line that characterize such effects as for example soft or hard clipping. However, such graphs describe only the simplest form of the static memory-less nonlinearity characterized by the instantaneous dependence of the output signal on the input one. In other words the level of the output signal at a particular moment of time is a nonlinear function of the level of the input signal at exactly the same moment of time. Such system immediately "forgets its past" and "lives one moment at a time". The fact of the lack of memory implies that linear and nonlinear characteristics of such system are not frequency-dependent. – Fig. 1

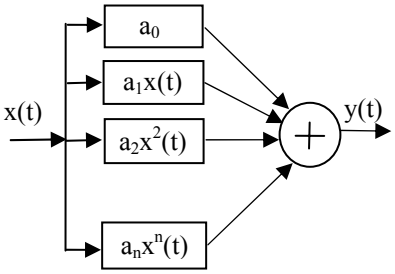


Figure 1: Static memory-less polynomial nonlinear system. $a_0 – a_n$ are coefficients characterizing weights of nonlinear products of different orders.

Such nonlinearities can rather accurately describe various nonlinear effects in electronic equipment (for example dependence of the anode current on the grid's voltage in a vacuum tube), but they cannot be applicable to analysis of nonlinear effects in electro-mechano-acoustical transducers that are not described merely by algebraic polynomials. Transducers such as woofers or horn drivers (and correspondingly loudspeaker systems encompassing them) exhibit much more complex nonlinear behavior that cannot be described by polynomials. In general they can be described to a certain degree of accuracy by nonlinear ordinary differential equations (woofers) or even more complex nonlinear partial differential equations (horns) [4, 5, 6, 7]. In this case simple one-dimensional "input-output" graphs are simply not applicable.

If the level of distortion products in the output sound pressure is not high (weak nonlinearity), such transducers can be described by Volterra series which can be considered as a Taylor series with memory, or as an extension of the concept of the impulse response and transfer function of a linear system to the multidimensional space [2, 3]. In a weakly nonlinear dynamic system not only the linear component of the signal has a memory (i.e. it "remembers" the levels of the past input signal) but the nonlinear components of various orders have memory as well, and mathematically this dependence is rather complex [2]. Output signal of a static polynomial nonlinear system is a result of raising the input signal to power (square, cube, etc.), multiplying these terms by constant coefficients characterizing the weights of nonlinear products of different orders, and summing them. In a weakly nonlinear Volterra system raising to power is substituted by convolution integrals and the different orders of nonlinearity (second, third, etc.) are

represented by the corresponding multidimensional convolution integrals of different orders – Fig. 2.

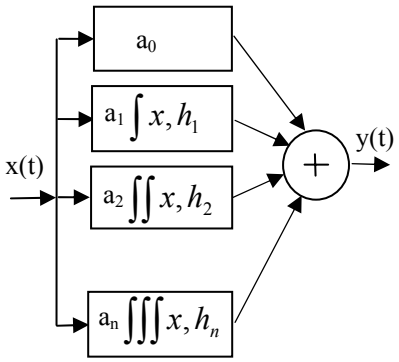


Figure 2: Dynamic weakly nonlinear system presented by Volterra series. h_1 is a linear impulse response, h_2,.. h_n are multidimensional impulse responses called Volterra kernels.

The linear part is represented by a familiar impulse response $h_1(t)$, whereas the nonlinear parts are represented by corresponding multidimensional impulse responses $h_2(t)$ $h_3(t)$… $h_n(t)$ called Volterra kernels. To transform these multidimensional impulse responses in frequency domain, corresponding multi-fold Fourier transform should be used. For example, second order frequency response corresponding to $h_2(t)$ looks like a three-dimensional graph with two horizontal axes both representing positive and negative frequency. A diagonal cut across this surface shows the level of second harmonic, all other cuts represent various sum and difference intermodulation products of second order [2]. Higher-order (3, 4, ...*n*) distortion frequency responses do not have graphical representation due to their multidimensionality. This example shows that the reaction of a dynamic weakly nonlinear system is significantly more complex than a static one. To make matters worse, the strongly nonlinear systems (such as a woofer at the very large level of voice coil excursion or a horn driver at very high levels of SPL) exhibit nonlinear behavior that cannot be described by the Volterra series, because the latter diverge in the presence of strong nonlinearity. Strongly nonlinear systems may exhibit even more complex behavior such as for example, bifurcation and chaos. Typical example of strongly nonlinear behavior in woofers is the voice coil jump-out effect also known as rectification.

Due to the multidimensional nature of dynamic nonlinear systems, a simple sinusoidal input signal cannot produce reaction that would represent the nonlinear system in its entirety. In other words, measuring output harmonics as a function of frequency and input level is not sufficient to predict reaction of a nonlinear system to an arbitrary input signal. In certain static nonlinear systems though, the information about harmonics measured at different levels of input signal may lead to restoration of the polynomial function governing input-output and reaction to an arbitrary input signal may be predicted. Much more complex signals are required to fully identify a dynamic nonlinear system. There are many methods of nonlinear identification that employ for example Gaussian noise or multitone signals.

4. HISTORY

The history of the subject is given in detail in the JAES article by Czerwinski, Voishvillo, Alexandrov, and Terekhov "Multitone Testing of Sound System Components – Some Results and Conclusion" [8]. It does not make sense to refer to every publication reviewed in the "History" section of this work. A total of 87 sources are reviewed there, covering the period from 1913 to 1998. Here, only the most significant works and the works published after 1998 will be mentioned.

In general, the history of the objective and subjective assessment of nonlinearity in audio equipment may be divided into several periods. The earliest "naïve" period that began at the dawn of audio was based mainly on measuring harmonic distortion, THD, THD+noise, and two-tone intermodulation distortion. The accuracy of the objective and subjective tests was significantly limited by the modest capabilities of early audio and testing equipment. Experiments with artificially induced distortion were based on static nonlinearities. Listening tests were often based on sinusoidal signals adversely affected by the static nonlinearity. This early period was characterized by a limited understanding of the nature of nonlinear dynamic systems and by the lack of proper information about psychoacoustics of the human auditory system. One of the earliest works is the 1929 publication of Janovsky [9], who attempted to find thresholds for the subjective detection of harmonic and intermodulation distortion. He used a generator of reference tones, a gramophone (a rudimentary record player), a distorting device based on single-end triodes, a measuring device, and headphones. Janovsky searched for a threshold of distortion audibility on a single pure tone, on two tones,

and on musical signals coming from 78 rpm shellac records, hardly a reference source of high fidelity musical material. Nevertheless, in a certain way Janovsky's work was ahead of its time. In particular, he came to conclusion that the difference intermodulation products are more critical than harmonics from the standpoint of audibility. His observation is correct; stronger audibility of the difference products having frequencies lower than the masking fundamental tones is explained by the non-symmetry of masking curves, i.e. by the stronger masking of the frequency components located higher on the frequency scale than the masking tones.

The next period of objective and subjective assessment of nonlinearity in audio equipment is characterized by significant attention paid to the two-tone intermodulation distortion measurements and the search for correlation between listening test results and the intermodulation distortion characteristics. Also in this period the critical role played by high-order nonlinearity was better understood. Intermodulation is a broad term referring to the generation of various spectral components at the output of a nonlinear system after more than one sinusoidal tone is applied to its input. In most of the research publications of that period intermodulation distortion was considered as a reaction to the two-tone input signal. There were two widely used methods of taking two-tone intermodulation distortion measurements. One was based on the application of two closely spaced tones and measuring the difference intermodulation product of the second order (CCIF method) which gained popularity first in Europe and was later popularized in United States by Scott [10]. The second method, initially conceived to measure distortion in cinema equipment, (SMPTE method) was developed by Hilliard [11]. The method uses a low-frequency tone and a lower level high-frequency tone to produce modulation of the high-frequency tone by the low-frequency one. Better correlation of intermodulation distortion measurements with subjectively perceived sound quality (versus harmonic distortion and THD) was reported by both Scott and Hilliard. Hilliard's method typically used 60 Hz and 3 kHz tones, the amplitude of the high-frequency tone being 12 dB lower than that of the low frequency tone. The CCIF method usually used 3 kHz and 3.05 kHz tones of equal amplitude or sweeping of the two closely spaced tones f_1 and f_2 and measuring second–order difference intermodulation product located at the frequency $f_2 - f_1$.

One of the important publications of that time was an article by Shorter [10], who paid attention to the role played by high-order nonlinear products on subjective perception of distortion. Shorter came to conclusion that nonlinear systems exhibiting harmonics of higher order sounded worse than the nonlinear systems generating low order harmonics but having similar THD. Shorter decided that higher order harmonics are more noticeable and suggested weighting them in such a way that each higher order harmonic received a higher weighting coefficient. It was dubbed in literature the Weighted Harmonic Distortion Coefficient. He measured harmonic distortion and performed listening tests (piano recording) of six systems (A, B, C, D, E, F) that had different level of RMS sum of harmonics and different distribution of the harmonics' levels. The RMS sum of harmonics' measurements of these six systems (on a single mid-band frequency) and the results of subjective listening test are shown in Table 1.

System	Subjective classification	RMS sum of harmonics
D	Bad	3.7%
B	Perceptible	3.3%
C	Just perceptible	2.6%
A	Bad	2.3%
E	Just perceptible	0.6%
F	Not perceptible	0.4%

Table 1 RMS sum of unweighted harmonics. (Shorter, [12]).

As it follows from the Table 1 the correlation between measured distortion and subjective assessment of distortion was poor. Shorter applied two methods of weighting the higher order harmonics. One was borrowed from the recommendations of the Radio Manufacturers' Association (RMA) [13]. This method was proposed as early as in 1937, an ancient time by the modern audio scale. The second weighting method was originally introduced by Shorter himself. According to the RMA method harmonics were multiplied by the weighting coefficient $n/2$, where n is the order of harmonic. Application of the $n/2$ weighting rearranged the hierarchy of the systems and made correlation between the subjective and objective assessment of distortion significantly better – see Table 2.

System	Subjective classification	Weighted RMS sum of harmonics (by n/2)
D	Bad	6.7%
A	Bad	5.1%
B	Perceptible	5.1%
C	Just perceptible	2.8%
E	Just perceptible	1.3%
F	Not perceptible	0.8%

Table 2 RMS sum o harmonics weighted by n/2. (Shorter [12])

However, the systems A (bad) and B (perceptible) had the same level of objective metric (5.1%) whereas the subjective quality was rather different. Shorter came up with a "stronger" weighting coefficient $n^2/4$ and its application obtained even more accurate scaling – Table 3:

System	Subjective classification	Weighted RMS sum of harmonics (by $n^2/4$)
A	Bad	19.4%
D	Bad	16.5%
B	Perceptible	8.6%
E	Just perceptible	4.5%
C	Just perceptible	3.3%
F	Not perceptible	2.2%

Table 3 RMS sum of harmonics weighted by $n^2/4$. (Shorter, [12]).

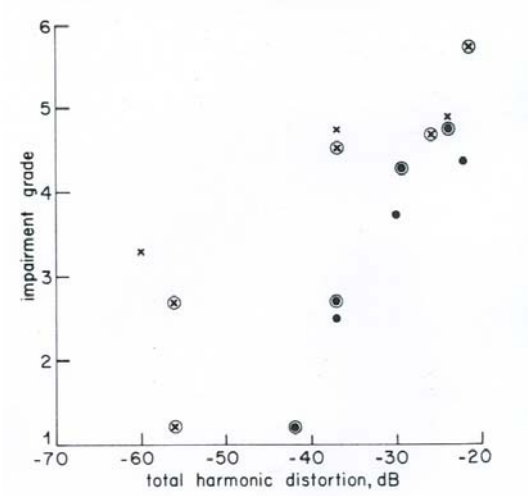
Shorter mentioned reasonably that measuring THD without detailing its contents does not have much value. Results obtained by Shorter seem to be a strong proof of significance of high-order distortion products. Although Shorter paid attention to the intermodulation products and mentioned that their number might be significant in the presence of high order nonlinearity, he seems to underestimate the dominating role played by intermodulation products as compared to harmonics. High-order nonlinearities produce significantly more intermodulation products than they do harmonics. For example, for the same level of an input multitone signal the number of intermodulation products in a system characterized by the high-order nonlinearity is dramatically higher than the number of intermodulation products in a system characterized by the lower orders of nonlinearity. Meanwhile the number of harmonic products in both systems is not that different [8]. Although Shorter's method does not seem to be a universal remedy, it may deserve a second look especially if weighting is applied to the intermodulation products of different orders generated for example by a multitone stimulus or two closely spaced sweeping tones.

A number of researchers believed that the role played by harmonics on sound quality is insignificant; see for example [13], [14], [15]. This opinion was based on the assumption that the harmonics coincide in frequency with natural overtones of musical instruments and therefore "blend" into their timbre. On the contrary, the intermodulation products have dissonant frequencies, and in particular, the difference tones are poorly masked because of the non-symmetry of the auditory systems' masking curve, and as such, they may be more annoying.

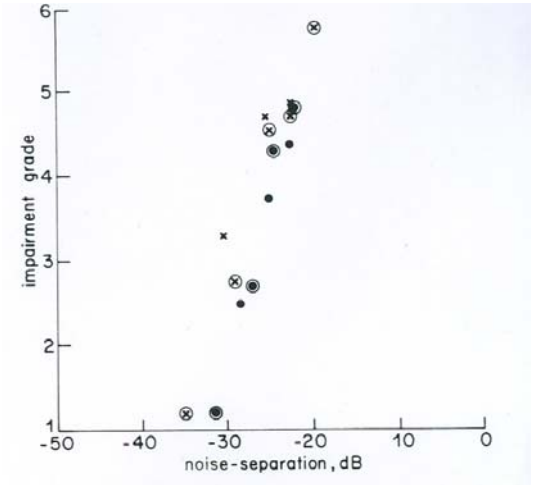
Some works paid attention to the stronger vulnerability of complex polyphonic music to nonlinear distortion compared with a sound of a single musical instrument – see for example Scott's works [16], [17]. This can be explained by the wider and more dense spectrum of the polyphonic music and therefore by the larger number of intermodulation products, their wider spectrum and thus, worse masking. Several works were based on a search for a quantitative relationship between the levels of two-tone intermodulation products and harmonic products [18], [19]. Hilliard in one of his publications on application of SMPTE intermodulation measurement method to test film-recording equipment [11] mentioned that the experimentally obtained ratio between the harmonic and intermodulation terms is approximately 4 to 1. The word "approximately" was eventually dropped and the ratio of 4 to 1 became essentially a law, though a wrong one. Strictly speaking, for a polynomial static nonlinearity of a finite order such relationship could be possibly analytically derived, but for a dynamic nonlinear system such as a woofer or a horn driver a universal relationship expressed in algebraic terms simply does not exist.

Work in the 70's was characterized by attention paid to more complex testing signals. In particular, Belcher from BBC advocated the use of multitone stimulus [20], [21]. Although the multitone had been known for decades before Belcher's work (first source

found is dated by 1913 [22]), it was Belcher who experimentally proved that multitone gives better correlation between objective test data and subjective perception of distortion. Using objective measure of multitone distortion which he called "noise separation" defined as a ratio between the peak level of all distortion products to the peak level of the signal+distortion products and using six point subjective impairment scale (from "imperceptible" to "unusable"), Belcher demonstrated that his metric gives a better correlation with subjective assessment of nonlinearity than THD. Fig. 3 shows comparison of "objective-subjective" assessments corresponding to Belcher's objective metric and to THD.



a)



b)

Figure 3: Comparison of correlation between subjective and objective assessment of sound quality impairment using THD (a) and multitone (b) (Belcher, [21, 22]).

Later period and modern "enlightenment" era of the objective assessment of nonlinearity in audio equipment has been characterized by the application of Volterra models [3, 28], NARMAX recursive models [29], [30], and the application of coherence and incoherence functions to assess nonlinearity in audio equipment. The latter functions have mainly been used in the assessment of sound quality in hearing aid systems [31], [32].

As an objective measure of nonlinearity multitone stimulus has a significant advantage over a single sweeping tone because it is faster; it generates a variety of harmonic and especially intermodulation products that can be classified by the order, level, and phase [8]. Moreover, using special mathematical manipulations with the response to multitone stimulus applied at different input levels the overlapping spectral components can be separated [27]. The drawback of multitone techniques is possibly generation of the excessive number of intermodulation products that makes it difficult to compare two or several characteristics belonging to different measured devices. To avoid this problem, a new measure, MTND (multitone total nonlinear distortion coefficient) was introduced in [33]. This metric is based on the averaging of all harmonic and intermodulation spectral components in a frequency domain "moving window" so that a continuous, frequency-dependent response is obtained. The level of this curve at a particular frequency depends on the level and "density" of the distortion spectral components in the vicinity of this frequency. However, no search for correlation between subjective audibility of distortion and MTND was attempted. Fig. 4 shows simulated multitone reaction and MTND of two 8-in woofers, one with a long coil and short top plate (a) and the other one with a short coil and long top plate. Other parameters of the woofers are identical. Mathematical definitions of MTND are given in Appendix 2.

As follows from the graphs, the woofer (a) produces more intermodulation between 100 and 500 Hz which is explained by a stronger dependence of the *Bl*-product and the voice coil inductance on the voice coil displacement. It would be practically impossible to overlay two multitone distortion graphs; however, using MTND curves it becomes quite possible. If the MTND curve is related to the level of loudspeaker frequency

response, values in percent can be obtained. Such curve may look like a THD, but there is a fundamental difference between them; THD response is blind (deaf?) to intermodulation, whereas the MTND characteristic takes them all into account.

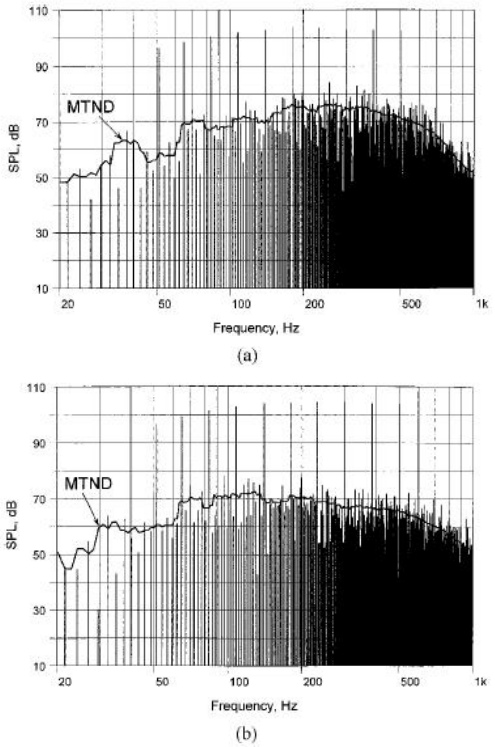


Figure 4: Multitone reaction and MTND (solid curve) of two 8-in woofers. (a) – long voice coil and short top plate, (b) – short voice coil and long top plate. (Voishvillo, Terekhov, Czerwinski, and Alexandrov [33]).

Schmitt researched audibility of nonlinear distortion in direct radiating loudspeakers and horn drivers. [34], [35]. Her work was principally different from her predecessors' work because she used nonlinear dynamic DSP models that simulated nonlinear physical effects in transducers. Analyzing thresholds of distortion audibility Schmitt came to the conclusion that perceptibility depends on the testing signal. She stated that the minimum threshold of audibility of nonlinear contamination of musical signals in direct radiating loudspeakers corresponds to about 5% in terms of the peak value ratio. Schmitt indicated that nonlinear contamination of musical signals (she used several musical excerpts with different degree of stationary and transient parts) causes "sound discoloration, intonation failure, modification of dynamic, temporal structure, and space perception, and noise". With regards to horn drivers the minimum threshold of audibility of nonlinear contamination corresponded to 2% of the peak value ratio. The threshold strongly depended of the type of testing signal – Fig. 5 and Fig. 6.

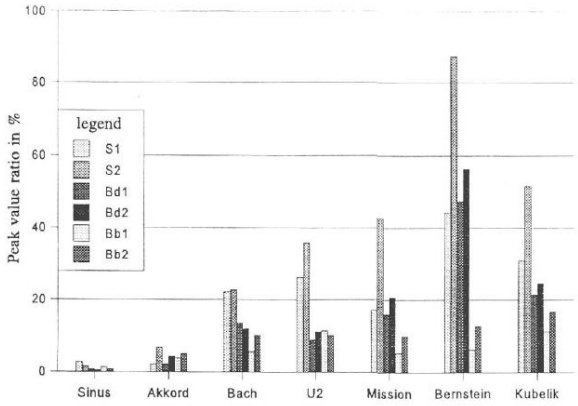


Figure 5: Threshold of audibility of nonlinear distortion in direct radiating loudspeaker in dependence on test signal. (Schmitt, [34]).

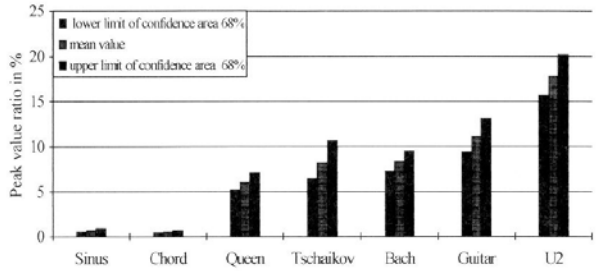


Figure 6: Threshold of audibility of nonlinear distortion in horn driver in dependence on test signal. (Schmitt, [35]).

Tan, Moore, Zakharov, and Mattila published a series of articles in JAES [36], [37], [38], and [39] where they attempted to search for the correlation between the audibility of the distortion in musical and speech signals, and objective measures of nonlinearity. In particular, in [36] they used multitone stimulus resembling the one used in the aforementioned work by Cerwinski et. al. [8] (ten tones logarithmically spaced across the audio frequency range). The authors used artificial static nonlinearity such as hard symmetrical clipping, hard asymmetrical clipping, soft symmetrical clipping, soft asymmetrical clipping, central clipping (also called zero-crossing or dead-zone nonlinearity), and full-range distortion produced by raising the instantaneous absolute magnitude of the signal to a

power from 0.5 to 2. They also investigated band-limited distortion applied to signal in four frequency bands covering audio frequency range. The authors developed a metric called Distortion Score (DS) – Fig. 7.

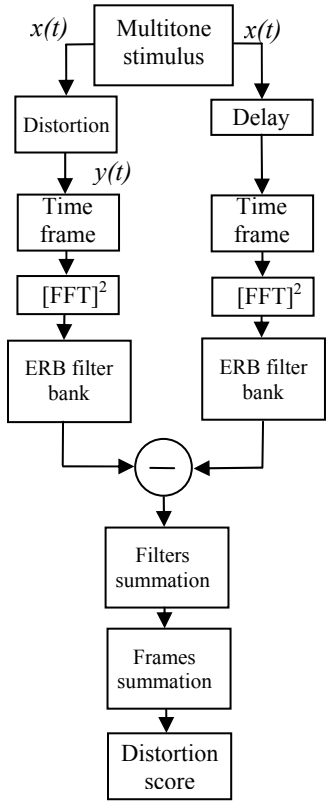


Figure 7: Simplified schematic diagram of obtaining nonlinear distortion perceptual metric Distortion Score (Tan, Moore, and Zakharov [36]).

●Both the input and the output of each nonlinear system were time-aligned and analyzed in a series of successive non-overlapping 30-ms frames.

●The spectra of input and output were split between 40 non-overlapping frequency bands, each of which was one ERB (mean equivalent rectangular bandwidth of the auditory filter) The ERB is conceptually similar to the traditional critical band of hearing system but differs slightly in numerical values [40].

●For each frame the difference in level in each ERB was calculated and the absolute value of the difference was summed across the audio frequency range. This gave a perceptually relevant measure of the difference in spectrum between the input and the output, which the authors dubbed Distortion Score (DS).

●The values of the DS across all 30-ms frames were summed to give an overall measure of distortion.

For the broad-band artificial distortion (hard limiting) the authors obtained very good correlation between DS and perceived sound quality. On the distortion produced by artificial dynamic nonlinearities and by real transducers the correlation was moderate. Fig. 8 shows the inverse correlation between the DS metric and the mean subjective ratings for static distortion. The trend's slope is negative because higher levels of DS correspond to lower subjective ratings. Correlation is -0.97. Fig.9 illustrates similar plot for the artificial dynamic nonlinearities and nonlinearities in real transducers. Correlation is 0.59.

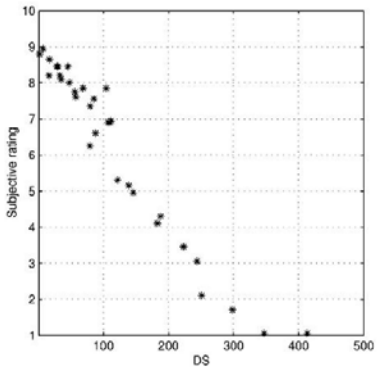


Figure 8: Mean subjective ratings of distortion produced by static nonlinearities as a function of Distortion Score metric. Musical signal. (Tan, Moore, Zakharov [36]).

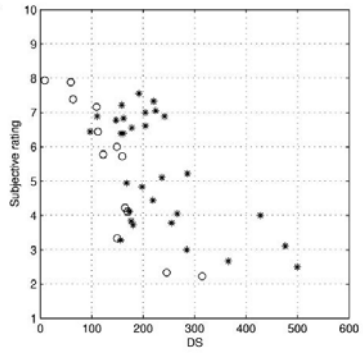


Figure 9: Mean subjective ratings of distortion produced by artificial dynamic nonlinearities and nonlinearities in real transducers as a function of Distortion Score metric. Musical signal. ¤ Artificial distortion, * Real distortion in transducers (Tan, Moore, Zakharov [36]).

The authors came to conclusion that the zero-crossing and the soft clipping had only small effect on the ratings, whereas the hard clipping and the full-range distortion had large effect. The conclusion about low audibility of zero crossing distortion contradicts the Gedd-Lee metric that is based on assumption of high sensitivity of auditory system to nonlinear distortion occurring at low levels of signal. The authors concluded that their DS metric based on the use of a multitone signal can predict the perceptual effects of nonlinear distortion reasonably well.

In their next work [37] the authors used more sophisticated perceptual model that did not use multitone signal but was based on analysis of input and output speech and musical signals. – Fig. 10.

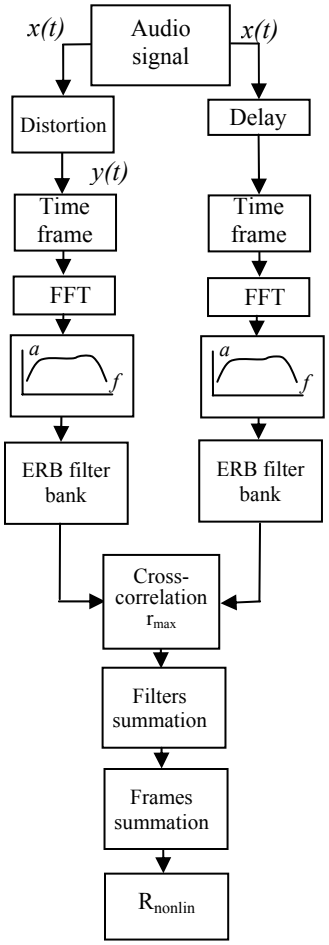


Fig. 10. Simplified schematic diagram of the improved perceptual model, *x(t)* – original signal, *y(t)* – distorted signal (Tan, Moore, and Zakharov [36]).

The improved model provided a measure of distortion that was better correlated with subjective quality than the Distortion Score described in [36]. The model operated in the following way:

- The input signal (speech or music) and the output distorted signal were time-aligned and analyzed in a series of successive non-overlapping frames.
- The original and distorted signals were equalized to simulate the effect of transmission through the outer and middle ear. The equalization resulted in a relative attenuation of frequency components below 500 Hz and above 5000Hz.
- The equalized signals were fed (separately) to an array of 40 band-pass filters each having a bandwidth of one ERB.
- For each filter the short-term normalized cross-correlation between the response to the undistorted signal and the response to the distorted signal was calculated. The filters' output was split into 30-ms non-overlapping frames. The cross-correlation was calculated between the waveform of a particular frame at the filters' output in response to the distorted signal.
- For each frame and each filter, the maximum value of the normalized cross-correlation samples r_{max} was calculated. Lower values of r_{max} corresponded to stronger influence of distortion components. The value r_{max} =1 was indicative of distortion-free reproduction.
- For each frame, the weighted values of r_{max} were summed across filters to give an overall measure of correlation for that frame. The average correlation determined in this way was then averaged across frames. This gave a measure R_{nonlin} which decreased with increasing distortion.

The authors tested their model on the same types of artificial static distortion they used in their previous work [36]. In addition, similarly to their previous work, the authors used the band-limited distortion, real distortion produced by transducers and the mixture of the artificial and real distortion.

Analyzing the results of their research the authors concluded that their model can accurately predict the subjective ratings of perceived distortion for both real and artificial nonlinear systems. The correlations between obtained and predicted ratings were high (0.93 for speech and 0.98 for music signals). Fig. 11 shows results obtained in experiment similar to the one carried out in the authors' previous work and shown on Fig. 8 (artificial static nonlinearity), but using new perceptual model. Correlation is 0.98. Fig. 12 shows results

obtained in experiment similar to the one carried out in authors' previous work and shown on Fig. 9 (artificial dynamic and real transducers' nonlinearities), but using new perceptual model. Correlation is 0.92.

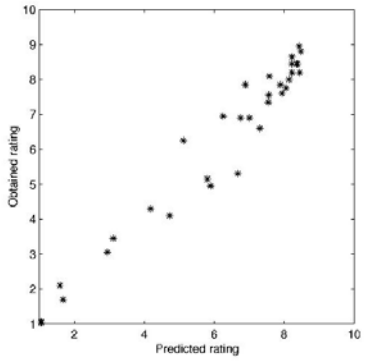


Figure 11: Mean subjective ratings of distortion produced by static nonlinearities as a function of predicted rating using metric R_{nonlin}. Musical signal. (Tan, Moore, Zakharov [37]).

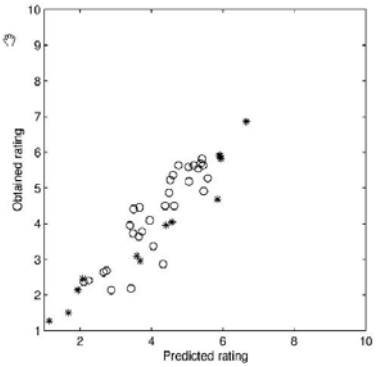


Figure 12: Mean subjective ratings of distortion produced by artificial dynamic nonlinearities and real transducers as a function of predicted rating using metric R_{nonlin}. Musical signal. (Tan, Moore, Zakharov [37]).

In [38] the same authors attempted to evaluate nonlinear distortion in the presence of the linear distortion. For evaluation of linear distortion the authors used a method described in their other work [39]. The perceptual model for the linear distortion detection was based on excitation patterns of a sound pressure signal. [40]. It can be calculated using the known characteristics of the auditory filters. Each auditory filter represents the tuning of the auditory system at one center frequency. The excitation pattern can be defined as the output level of the auditory filters plotted as a function of filter center frequency. It was assumed that the larger the change in excitation pattern produced by a given spectral linear distortion, the lower will be the perceived naturalness associated with that linear distortion. A related approach was used by Kates [41], [42] to assess coloration produced by room reflections. Detailed description of the linear perceptual model based on calculation of excitation patterns and the process of quantifying the excitation patterns differences is given in [39].

The authors used again 40 ERB filters covering audio frequency range. Linear distortions were introduced as ripples on the overall frequency response. The nonlinear model used by the authors was similar to the one described above [37]. Spectral ripples with depth (peak-to-valley ratio) 10 dB and rate 0.5 ripples per ERB number (approximately 0.4-octave wide) over several different frequency ranges were combined with hard symmetrical clipping (when the signal was clipped 10% of all time) in combination with pre and post filtering to apply distortion in several different frequency bands.

Another set of combinations of linear and nonlinear distortion included spectral ripples applied over the frequency ranges at different parts of the overall frequency band combined with waveform distortion produced by raising the absolute value of the instantaneous amplitude of the broad-band signal to power 0.7, while preserving the sign of the amplitude. This type of distortion was applied in several different frequency bands.

After numerous experiments the authors came to conclusion that the perceptual effects of nonlinear distortion were generally greater than those of linear distortion, except when the linear distortions were severe. The listening tests' results were compared with the predictions of the new model based on a weighted sum of predictions for nonlinear distortion alone [37] and linear distortion alone [39]. The resulting correlations were 0.89 for speech and 0.95 for music. Second set of experiments included both artificial distortion and the distortion produced by the transducers. The listening tests results were again compared with the predictions of the new model. There was a very good agreement between the obtained and predicted ratings; correlations were 0.85 for speech and 0.90 for music. The real transducers used in experiments were four different compact broad-band loudspeakers

(13 – 20 mm in diameter) mounted in various acoustical enclosures used in headphones and in "hands-free" telecommunication devices. The authors also noted that the obtained and predicted ratings were relatively low for the stimuli subjected to "real" distortion. This is partly explained by the fact that the real transducer all had a limited frequency range and irregular frequency responses.

Geddes and Lee performed a research study on the correlation between characteristics of static nonlinearity and subjective perception of sound quality. They developed original Gedd-Lee metric (see Appendix 2 for details) based on the known properties of a static nonlinearity [43], [44]. The crux of this approach is the assumption that a nonlinearity affecting the low-level signal is more detrimental to the sound quality than a nonlinearity that "kicks-in" at higher signal levels, such as hard or soft clipping. The statistical distribution of a typical musical signal resembles a Gaussian curve in that the low-level signals have a much higher probability of occurrence compared to the large level signals. In addition, the high-level signals are better maskers than the low-level signals. Geddes and Lee also assumed that the curvature of a nonlinear static polynomial transfer function is a significant factor in sound quality, because a high curvature of static nonlinear transfer function is responsible for generation of wide-spectrum high-order nonlinear products that are poorly masked by the human auditory system. The authors reported a good correlation of their metric with subjectively perceived sound quality (contrary to harmonic and two-tone intermodulation distortion). Fig. 13 shows the relationship between the subjective rating and Gedd-Lee metric obtained for 18 stimuli. The initial set consisted of 21 stimuli, but for the metric values larger than 10 significant variance was observed. For the 21 stimuli the correlation for three different metrics (THD, IMD, and Gedd-Lee metric) was as follows: THD: r = -0.423, IMD: r = -0.345, Gedd-Lee metric: r = 0.68. Removing three values of Gedd-Lee metric greater than 10 brought the correlation level of the latter to 0.95.

The metric is partly based on a consideration of masking effects in the auditory system, but those effects are not explicitly modeled by the metric. This approach, however, does not make it possible to apply it directly to measure static nonlinear systems whose parameters and structure are not known in detail ("black boxes"). In addition, the method is not applicable to analysis of dynamic nonlinearities whose properties are not described by a mere algebraic polynomial function.

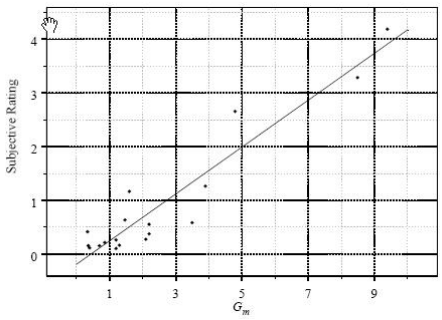
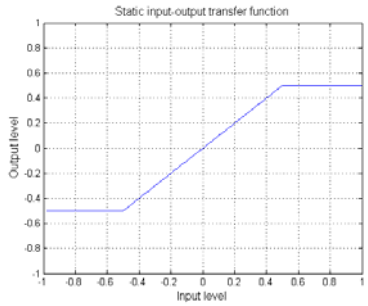


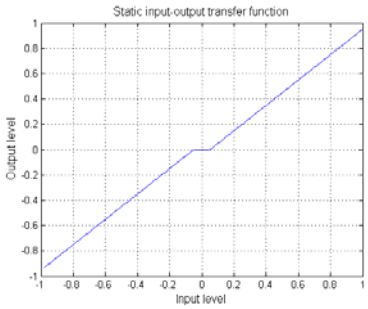
Figure 13: Values of Gedd-Lee metric versus subjective rating. (E. Geddes and L. Lee, [43, 44]).

Advantage of the Gedd-Lee metric is its simplicity and the fact that this metric is a strong proof of a critical role played by high-order nonlinearity and nonlinearity impairing low-level signal. Gedd-Lee metric could possibly be extended by splitting the signal into frequency bands and applying the metric independently in each band. In their work the authors indicate that the metric can be extended to the frequency-dependent nonlinearities by expressing the polynomial functions describing nonlinearities as a function of input level and frequency.

According to the Gedd-Lee metric the zero-crossing distortion should be more detrimental to sound quality than the distortion that occur only at large levels of signal such as hard clipping. Independently from the authors' work, the informal experiments with audibility of distortion produced on musical signal by the zero-crossing and hard limiting were conducted at JBL Pro by the author of this publication. The 17-second musical excerpt (Paul Anka, "It's My Life", CD "Rock Swings", 2005, Verve Records) was turned into a 16-bit, 44.1 kHz wav file. The Matlab program read the file and processed it by imposing hard clipping (50% of the +/- 1 maximum amplitude was "chopped off"), and zero-crossing where the original signal having amplitude less than 0.05 was equaled to zero. The THD corresponding to the hard clipping was 22.6% and the THD, corresponding to the zero crossing was only 2.9%. Fig. 14 shows both static nonlinear input-output functions.

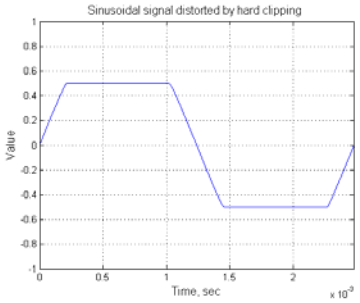


a)

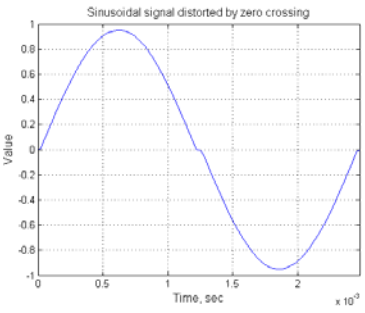


b)

Figure 14: Static input-output functions corresponding to hard clipping (a) and zero crossing (b).

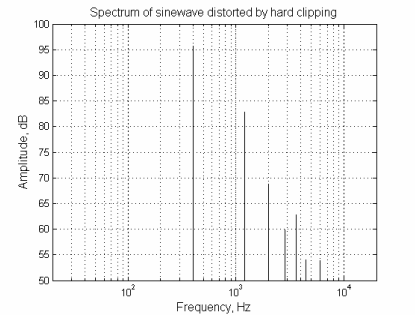


a)

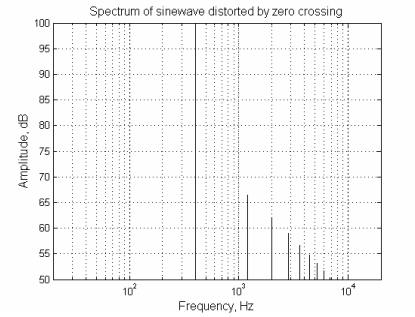


b)

Figure 15: Waveforms of 400 Hz sinusoidal signal distorted by hard clipping (a) and zero-crossing (b).



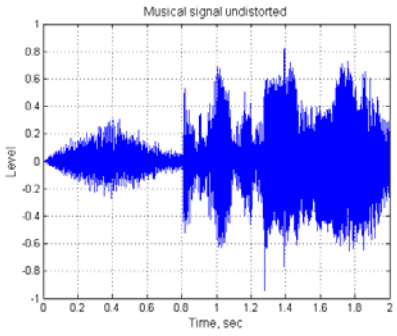
a)



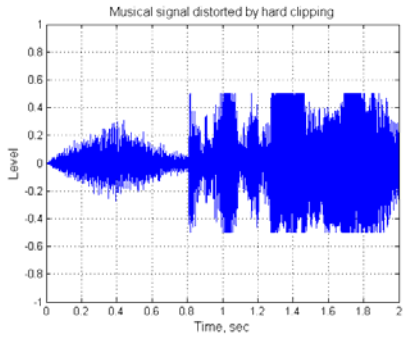
b)

Figure 16: Spectra of sinusoidal signal distorted by hard clipping (a) and zero-crossing (b).

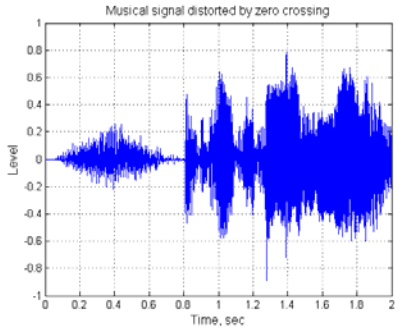
Fig. 17 shows the 2 seconds long waveforms of original music signal (a), original signal adversely affected by hard clipping (b) and zero crossing (c).



a)



b)



c)

Figure 17: Waveforms of musical signal (a) - undistorted; (b) - distorted by hard clipping and (c) – distorted by zero-crossing.

All presented objective data is indicative of much stronger deterioration of signals' waveforms and spectra by the hard clipping that is reflected in THD which is nearly 10 times higher for the case of hard clipping. The musical signal and sinusoidal signal before and after the nonlinear processing were stored, their levels were adjusted for equal loudness and they were reproduced via headphones and loudspeakers. The subjective difference on sinusoidal signal was not very strong, whereas on the musical signal it was dramatic; the zero crossing produced intolerable deterioration of sound quality, but the hard clipping produced only rare unpleasant, but tolerable effects. This observation is another proof that THD may be totally inaccurate as a measure of the nonlinear distortion, and subjective comparison of sinusoidal signals affected by various nonlinearities is pretty much pointless.

There is a vast area of methods developed for objective assessment of sound quality in low-bit compression systems (mp3, AAC, WMA, ATRAC, etc.). These methods, similarly to [36] – [38] use perceptual approach. The methods evolved into two standards, one for evaluation of the compressed musical audio signals [45], it is called PEAQ (Perceptual Evaluation of Audio Quality), and the other one, for narrow-band speech signals [46], called PESQ (Perceptual Evaluation of Speech Quality). These methods are intended for use only in low-bit compression systems that deal with electrical signals but not with the sound pressure signals radiated by electroacoustical transducers.

The history of the subject is given in detail in [45]. First works were focused on evaluation of sound quality in speech codecs. The first perceptual methods of assessment of sound quality in wide band audio signals appeared in late 1980s. Significant number of works employing different psychoacoustical models was published during last two decades. There are two basic approaches to perceptual models; one of them uses noise-to-mask ratio concept and the other one is based on so-called internal representations [45]. The former method explicitly exploits auditory system's masking effect.

Masking is one of the most important concepts of psychoacoustics. When a particular sound pressure signal reaches basilar membrane through the middle ear and the oval window in the cochlear [40], it excites mechanical movement of basilar membrane and the signal's frequency spectrum is mapped along the length of the basilar membrane in such a way that higher frequency components correspond to vibration near membrane's "entrance" at the oval window, whereas the lower frequency components excite mechanical vibration with maximum at the end of the membrane. This movement of the basilar membrane evokes neural activity corresponding to the vibrating parts of the membrane. Psychoacoustically, the excitation pattern can be defined as a reaction of auditory filters to the incoming sound pressure signal and this reaction is a function of filters' center frequencies [40]. In general excitation patterns look like "spreaded" replicas of signal's spectrum and waveform. For example, time domain excitation pattern of a short pulse reminds a triangle with faster ascending part and slower descending part, whereas the excitation pattern of a sinusoidal signal in frequency domain looks like an asymmetrical triangle with a steeper left-hand slope. Excitation patterns work as maskers for weaker signals located close in spectrum (or in time) to the masker. For example, if a second lower-level sinusoidal signal produces weaker excitation pattern that falls below the former one, the second signal becomes inaudible. This

is the crux of masking effect. Masking effect can be observed in frequency domain (simultaneous masking), and in time domain (nonsimultaneous masking). The letter can produce forward masking (weaker signal following a stronger one becomes inaudible), or backward masking, when the weaker signal preceding the stronger signal is masked by the latter one.

The noise-to-mask ratio (NMR) measurement method, also known as the masked threshold concept and noise signal evaluation, operates with error (distortion) signal extracted from the distorted signal as a difference between the original signal and the distorted one. This error signal is compared (across corresponding band-pass filters simulating auditory filters and across analysis time frames) with the masking patterns obtained from ear model. If the level of the error (distortion) signal is lower than the corresponding masking pattern produced by the original signal in the same frequency band and time frame, the error signal's sample is considered inaudible. The psychoacoustical distances obtained for particular auditory filters and time-frames are superimposed and a single number characterizing sound quality is obtained - Fig 18.

Fig. 18 shows simplified block diagram described in [47]. This method derives error signal in both time and frequency domains. Audio signal is applied to the codec under test and to the delay line to compensate for codec's latency. First, the difference between the distorted and undistorted signals is calculated in time domain to derive distortion signal *e(t)*. Then both the reference signal $x(t-\tau)$ and the distortion signal *e(t)* are transformed into frequency domain in each consecutive 11.6 ms-long analysis frame. Distortion signal's spectrum and reference signal's spectra are calculated. Then the spectrum of the reference signal and the distortion signal's spectrum are split between band pass filters each having a bandwidth of one critical band to simulate corresponding processing of human auditory system. Then a perceptual model, using basilar membrane's excitation patterns calculates the masked threshold of the reference signal. The energy of the error (distortion) signal is calculated and compared with the masked threshold in each critical band. The ratio of the error energy to the masked threshold is the NMR (noise-to- mask ratio) for a particular critical band. The numbers obtained for each band and time frame are summed to give an overall metric of sound quality.

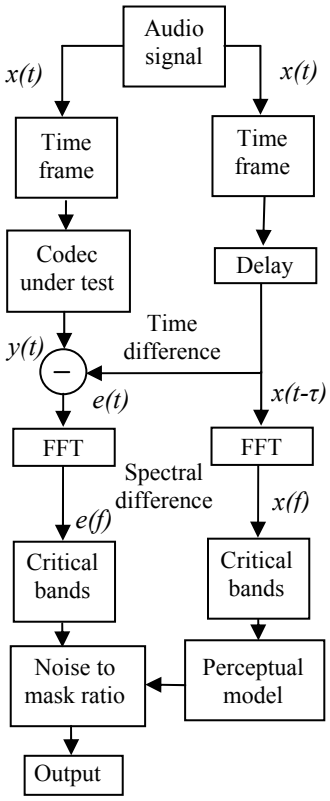


Figure 18: Simplified block-diagram of Noise-Mask Ratio method of measurement codecs' sound quality (Beaton, Beerends, Keyhl, and Treurniet [47]).

In the second method based on internal representations, the information indicative of the perceived quality is derived through comparison of simulations of basilar membrane's neural excitation patterns evoked by the original and distorted signals. The difference between excitation patterns is processed across the audio frequency range covered by auditory filters and across the time frames used in analysis. The ultimate metric is a number that is highly correlated with the subjective quality score. This approach is based on rather physiological functions of auditory system than on psychoacoustical ones. In general, this family of methods is based on so-called internal representations [45], [47]. The methods described in [36] and [37] also belong to this family. The internal representation maps initial audio signal on the physiological reaction of basilar membrane to this signal. For the stationary signals the internal representation is better described in frequency domain. For nonstationary signals internal

representation is carried out in time domain. Most important component of internal representation is the spreading of signal in time and frequency domain. The spreading is directly related to the masking in frequency and time domain. Internal representation also includes such effects as psychoacoustical compression explained by the nonlinear relationship between the intensity of sound and perceived loudness. It also includes transformation of signal's spectrum by outer and middle ear [40], [47]. This transformation corresponds to decrease auditory system's sensitivity to distortion that occurs in lower and higher parts of audio frequency range. Fig. 19 shows simplified block diagram of a measuring system based on the internal representations according to [47].

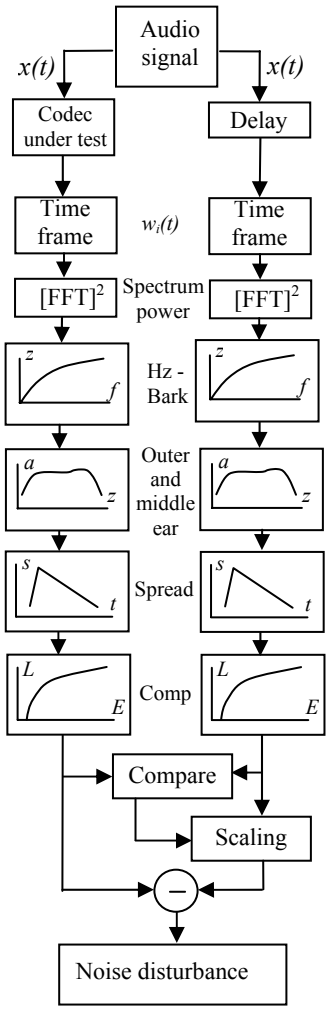


Figure 19: Simplified block diagram of the measurement method based on internal representation (Beaton, Beerends, Keyhl, and Treurniet [47]).

Original signal *x(t)* and distorted signal *y(t)* are windowed with a window $w_i(t)$ and transformed in frequency domain by FFT. Afterwards, the power spectrum calculated and transformed into psychological frequency scale expressed in Barks. Then the signal is equalized to simulate transformation of spectrum in outer and middle year. After that the signal is spreaded according to the basilar membrane's physiological reaction and compressed due to the nonlinear relationship between perceived loudness and the intensity of sound. Then the signals are compared and the difference across the auditory filters and time frames, called Noise Disturbance, is obtained.

In PEAQ the processing of signal and extraction of objective perceptual criteria is more complex than in the aforementioned examples. PEAQ evolved from seven different competing systems using different flavors of perceptual effects and absorbed various features of these methods. In addition, there are two versions of PEAQ, the basic one that has an advantage of higher computational speed, and the advanced one that uses more complex perceptual model. Basic model uses both methods of measurement – the internal representations and the noise-mask ratio. Basic method uses FFT-based approach having poorer temporal resolution. The advanced method uses both the FFT-based and the filter-bank methods, but only the internal representations approach. General approach inherent to both versions is illustrated on Fig. 20. The reference original *x(t)* and the distorted signal *y(t)* are compared after corresponding time-alignment. Concurrent time-frames of the original and nonlinearly processed signals are mapped to a basilar membrane representation, and differences are analyzed further as a function of frequency and time by a cognitive model. This model extracts perceptually relevant features which are used to compute a measure of sound quality. The model operates with several intermediary output variables. A selected set of output variables is mapped to an objective metric reflecting sound quality [45].

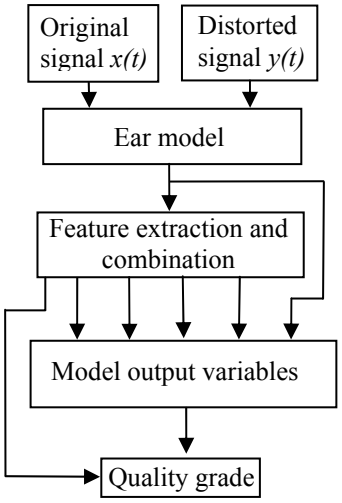


Figure 20: General structure of the PEAQ perceptual measurement model [45].

Fig. 21 illustrates measurement process in more detail.

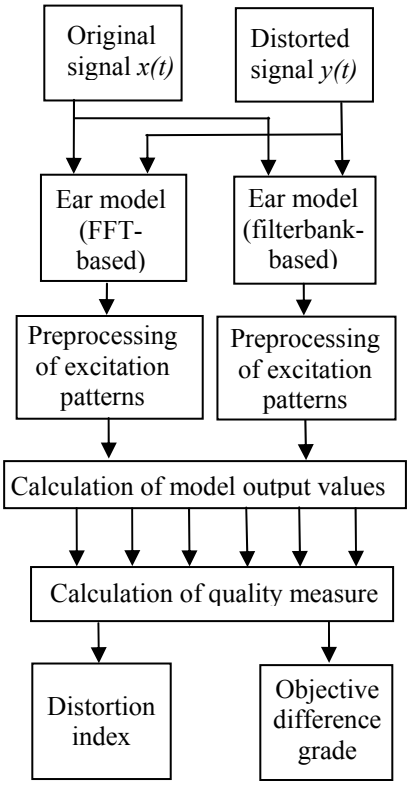


Figure 21: Simplified block diagram of the overall model of PEAQ measurement process [45].

Detailed description of both versions of PEAQ is given in detail in [45] and it is not a goal of the current work to analyze every aspect of it. PEAQ is a very complex system that includes, in addition to the aforementioned components, the cognitive part, adaptations, weightings, and variety of output variables that are processed by an artificial neural network to extract output objective metrics. Even without going into details and peculiarities of the measurement process, the reader can easily see that the distance dividing this system and the THD testing practiced in loudspeaker industry is measured by light years.

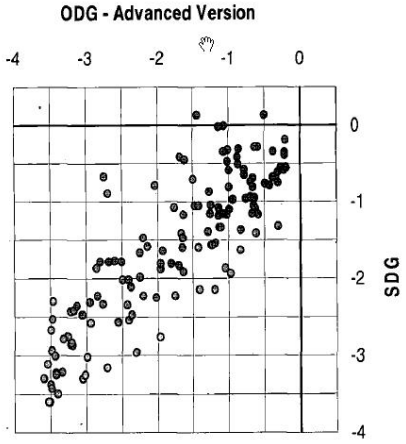


Figure 22: Result of using advanced version of PEAQ [45].

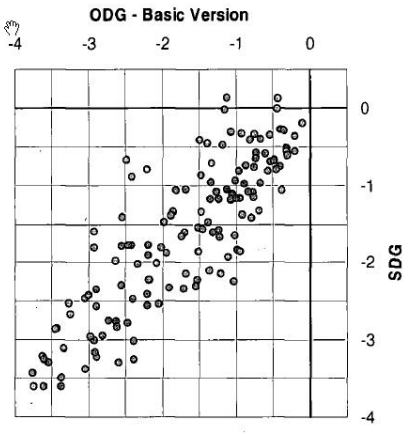


Figure 23: Result of using basic version of PEAQ [45].

Fig. 22 and Fig. 23 show the results of using both versions of PEAQ for assessment of sound quality of six codecs using 136 sets of different musical signals [45]. Fig. 22 corresponds to the advanced version of PEAQ,

and Fig. 23 corresponds to the basic version. Vertical scale corresponds to the SDG (subjective difference grade) obtained through listening tests, and the horizontal axis corresponds to the ODG (objective difference grade) provided by PEAQ. In general, correlation between subjective and objective evaluations is very high and PEAQ measurements can successfully substitute listening experts. The advanced version provides more accurate results.

5. POSSIBLE FUTURE DEVELOPMENTS, CONCLUSION

There are at least two possible directions in improving measurement of nonlinear distortion in transducers and loudspeaker systems. One of them is the adaptation of modern powerful perceptual methods specifically for measurement of degradation of loudspeaker sound quality caused by various nonlinear effects. The second approach could be based on revision of existing conventional methods that use objective metrics (harmonic distortion, two-tone intermodulation, multitone, etc.) through modifying them (by introducing weighting functions, selecting specific intermodulation components, manipulating with levels of testing signals, etc.) so that to obtain new responses having better correlation with subjective measures of nonlinearity. These responses could be presented along with the traditional objective ones. Whereas the objective methods may give information about level of certain symptoms of nonlinearity and indicate possible physical causes of nonlinear distortion in transducers and loudspeaker systems, the "psychoacoustically modified" objective methods would convey partial information about possible audibility of dynamic distortion signal produced by these nonlinear effects. For example, using fifty year old Shorter's idea about weighting higher-order harmonics could provide information more pertinent from the standpoint of perceived sound quality than traditional THD and harmonic measurements. Comparatively simple metrics could be built around following basic commonly accepted principles:

●Difference intermodulation products are poorly masked by human auditory system.

●High-order nonlinear products are indicative of the presence of high-order nonlinearity that generates distortion signal having wide and dense spectrum. This signal is likely to be poorly masked, and therefore, could be audible.

●Nonlinear effects that adversely affect the low-level input signal are more critical than those that impair input signal only at large levels.

●Sensitivity of distortion perception decreases towards both ends of the audio frequency range.

For example, a measurement method based on two closely spaced sweeping tones could select the difference intermodulation products of different orders, give the higher-order products corresponding weighting coefficients, and combine these sub-metrics into a single frequency-dependent response. Application of such metric at different levels of input signal with a high resolution at low levels could possibly convey information about presence of nonlinear effects that manifest themselves at low levels of signal and therefore are prone to produce audible distortion. Equalization of the output signal to simulate the change of distortion products' spectrum by the outer and middle ear would be responsible for frequency dependence of distortion perception. Intuitively, such a metric should be psychoacoustically more accurate than regular two-tone intermodulation and traditionally used harmonic distortion and THD. Similarly, it is possible to post-process the reaction to the multitone stimulus applied at different input levels in such a way so that the difference intermodulation products of different orders would be selected, weighted, and processed in the manner similar to the derivation of MTND.

With regards to the perceptual measurement models developed for assessment of codecs' sound quality, the methods using NMR (noise-to-mask ratio) measurement technique seems to better satisfy needs of loudspeaker engineer than the methods based on internal representations. The former provides explicit information about perceptual distance between the masker (undistorted signals) and the masked distortion signal. This may provide valuable information for engineer about acceptable levels of input signal, maximum voice coil excursion, maximum loudspeaker SPL, etc. The latter methods provide an accurate data that is indicative of the sound quality of the DUT, but they do not give information about perceptual distance between the levels of undistorted signal and the levels of the distortion signal.

Straightforward application of existing perceptual methods (created specifically for assessment of codecs' sound quality) for measuring loudspeakers may be problematic due to the radically different nature of linear and nonlinear distortion in codecs and in

transducers. In codecs the original signal is impaired by the quantization noise and these systems usually have flat frequency response with exception for the roll-offs at the boundaries of the spectrum of the processed musical or speech signals. Meanwhile, transducers may have significant irregularity of frequency response that may adversely affect accuracy of assessment of nonlinear distortion. In loudspeakers the signal is distorted by the nonlinear physical effects that make the radiated sound pressure signal dependent on the voice coil and diaphragm position, on the instantaneous values of the voice coil current, and on the instantaneous level of sound pressure in the compression chamber and horn (in horn drivers). Loudspeaker reproduction may suffer from limiting produced by saturation of suspension and decrease of *Bl*-products; there is no concept of level-dependent distortion in codecs.

Low-bit compression systems actually withdraw some information from the signal (this is a virtue of the low-bit compression), whereas the nonlinearities in loudspeakers add extra information in the form of dynamic distortion signal accompanying the initially undistorted input signal. It is known from psychoacoustical experiments that adding information to the audio signal (new distortion products) produce stronger cognitive effect than withdrawing information from the signal [45]. There are no publications that the author is aware of where an attempt is made to use the approaches standardized in PEAQ and PESQ to predict the degradation of sound quality by nonlinearities that occur in loudspeakers. Such experiments would present great importance. In addition, the comparison of the results obtained from the perceptual models developed in [36] – [38] and the PEAQ and PESQ systems for measuring transducers would be important as well.

One of the possible developments of the research works done by Tan, Moore, Zakharov, and Matilla is using nonlinear dynamic models of direct radiating loudspeakers and horn drivers rather than real transducers and artificial nonlinearities that do not occur in real loudspeakers. Using dynamic models has advantages over using real transducers; the models always give repeatable consistent results, they are not adversely affected by measurements' possible artifacts, the models are more flexible because they may or may not include various nonlinear effects. It is impossible to "turn on and off" different nonlinear mechanisms in real transducers.

The authors did not interpret their nonlinearities in terms of commonly used metrics such as THD, harmonic, and two-tone intermodulation distortion (let alone more complex objective metrics such as reaction to multitone stimulus and coherence (or incoherence) functions). It would give extra information and would help to "map" their results against regular measurements done by loudspeaker engineers to test their transducers. For example, it is not clear how the hard clipping distortion occurring 10% of all time of the signal existence translates into the regularly used metrics. The authors refer to the frequency scale expressed in ERB numbers which is a legitimate scale in psychoacoustics (as well as the other perception-related Bark frequency scale). Interpreting the results in regular frequency logarithmic scale would be a great help for audio and loudspeaker engineers who do not deal on a regular basis with perception-related frequency scales.

The authors used only one speech and one musical signal (piano, bass, and drums piece of jazz music). Since the perception of nonlinear distortion may depend on a particular signal used in experiments, by testing their models on different signals the authors would obtain valuable information about robustness of their measurement system. The authors came to conclusion that the zero crossing nonlinearity does not produce audible effects. This contradicts the basic idea behind the Gedd-Lee metric and observations made at JBL Pro that the nonlinearity that adversely affects low-levels signals is very detrimental to sound quality. The zero crossing is a typical example of such nonlinearity.

In all perceptual methods, and in the approaches described in [36] – [38] in particular, the objective metric is always a single number, be it DS, R_{nonlin}, or S_{nonlin}. These metrics result from "lumping" the perceptual differences between the reference signal and distorted signal in time and frequency domains. Such metrics are simple in interpretation; they make it easy to compare them with corresponding subjective metrics, and could be possibly convenient for customers of loudspeaker industry if some of the metrics were commonly accepted. However, they do not reveal much detail for engineers. It seems worth trying not to "lump" the sub-metrics in the frequency domain across all auditory filters into a single number, but present them as a function of auditory filters central frequency. This would give a "frequency response" of the perceptual metric. This "perceptual distortion frequency response" might give engineers clues what components of their transducers and loudspeaker systems are psychoacoustically troublesome, and what are

essentially innocuous. Such approach would be very valuable for loudspeaker engineers since it may lead to the new approaches in design of loudspeakers based on criteria that are directly related to the perception of nonlinearly distorted output sound pressure signal, similar to design of the low-bit compression systems. This may help to design more economical transducers where the resources are not wasted on minimization of traditional distortion metrics poorly related to sound quality. Plotting such perceptual frequency responses against the level of input signal might give information about "psychoacoustical" maximum SPL produced by a particular loudspeaker or its "psychoacoustical X_{max}" – maximum voice coil excursion. All the notes about modifications of methods described in [36] – [38] are applicable to other perceptual methods used for assessment of codecs' sound quality.

The Gedd-Lee metric could possibly be extended to a multi-frequency band version to obtain frequency-dependent characterization of subjectively perceived nonlinear distortion. In this case the metric's term responsible for generation of high-order nonlinear products (see Appendix 2) may possibly be modified using information obtained through the application of the low-level input signal at small incremental steps and analyzing output level in each frequency band independently.

The number of different approaches is virtually infinite. Comparatively simple "semi-perceptual" measurement methods based on commonly accepted principles should be at least more psychoacoustically relevant than traditionally measured harmonic distortion and could possibly provide pertinent information about audible distortion without using the "heavy artillery" of true perceptual models. It certainly does not prevent investigation of viability of using powerful and complex perceptual methods for assessment of nonlinearity in loudspeakers. Giving the loudspeaker industry better testing equipment based on new principles would be, as Herman Scott said 55 years ago, *a great and valuable service for the industry.*

6. REFERENCES

1. V. Volterra, "Theory of Functionals and of Integral and Integro-Differential Equations", Dover publications, Inc., New York, 1959 (Translated from French).
2. M. Schetzen, *Volterra and Wiener Theories of Nonlinear Systems*, (Krieger Publ., Malabar, FL, 1989)
3. A. Kaizer, "Modeling of the Nonlinear Response of an Electrodynamic Loudspeaker by a Volterra Series Expansion", *J. Audio Eng. Soc.*, vol. 35, pp. 421 – 433 (1987 June).
4. W. Klippel, "Dynamic Measurement and Interpretation of the Nonlinear Parameters of Electrodynamic Loudspeakers", *J. Audio Eng. Soc.*, vol. 38, pp. 949 – 955 (1990 Nov.).
5. W. Klippel, "Nonlinear Large-Signal Behavior of Electrodynamic Loudspeakers at Low Frequencies", *J. Audio Eng. Soc.*, vol. 40, pp. 483 – 496 (1992 May).
6. W. Klippel, "Modeling the Nonlinearities in Horn Loudspeakers", *J. Audio Eng. Soc.*, vol. 44, pp. 470 – 480 (1996 May).
7. W. Klippel, "Nonlinear System Identification for Horn Loudspeakers", *J. Audio Eng. Soc.*, vol. 44, pp. 811 – 820 (1996 September).
8. Multitone Testing of Sound System Components – Some Results and Conclusion, Part 1: History and Theory", JAES Vol. 49, No 11, 2001, November, page 1011 – 1027.
9 W. Janovski, "The Audibility of Distortion (in German), *Elek. Nachr.-Tech.*, vol. 6, pp. 421 – 439 (1929, Nov.).
10 10. Scott, "Measurement of Audio Distortion", Communications, pp. 25 – 32, 52 – 56 (1946, June).
11. J. Hilliard, "Distortion Test by the Intermodulation Method", *Proc. IRE*, vol. 29, pp. 614 – 620 (1941 Dec.).
12. D. Shorter, "The Influence of High Order Products on Nonlinear Distortion", *Electron. Eng.* vol. 22, pp. 152 – 153 (1950).
13. "Specifications for testing and Expressing Overall Performance of Radio Broadcast Receivers – Part 2 – Acoustic tests". Published by Radio manufacturers Association. Revised 1937, p. 5 – "Distortion Factor".
14. A. Peterson, "Intermodulation Distortion", *Gen. Radio Experim*., vol. 25 (1951 Mar.).
15. C. LeBel, "A New Method of Measuring and Analyzing Intermodulation", *Audio Eng.*, vol.

35, pp. 18 – 21, 30, 131 (1951 July); reprinted in *J. Audio Eng. Soc.*, vol. 16, pp. 386 – 389 (1968 Oct.).
16. H. Scott, "Audible Audio Distortion", *Electronics*, vol. 18, pp. 126 – 131 (1945 Jan.).
17. H. Scott, "Intermodulation Measurements", *J. Audio Eng. Soc.*, vol. 1, pp. 56 – 61 (1953 Jan.).
18. T. Roddam, "Intermodulation Distortion", *Wireless World,* vol. 46, pp. 122 – 125 (1950 Apr.).
19. M. Callendar and S. Mathews, "Relations between Amplitudes of Harmonics and Intermodulation Frequencies", *Electron Eng.*, pp. 230 – 232 (1951 June).
20. R. Belcher, "Audio Nonlinearity: An Initial Appraise of a Double Comb-Filter Method of Measurement", Rep. 1977/40, BBC Research Dept. (1977 Nov.).
21. R. Belcher, "A New Distortion Measurement (Better Subjective-Objective Correlation than Given by THD), *Wireless World*, pp. 36 – 41 (1978 May).
22. H. Reins, "Multitone Systems", Proc. IRE, vol. 1, pt. 4, pp. 5 – 41 (1913 Dec.).
23. D. Jensen and D. Sokolich, "Spectral Contamination Measurements", presented at the 85th Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts)*, vol. 36, p. 1034 (1988 Dec.), preprint 2725.
24. R. Cabot, "Fast Response and Distortion Testing", presented at the 90th Convention of the Audio Engineering Society, *J. Audio Eng. Soc*, *(Abstracts)*, vol. 39, p. 385 (1991 May), preprint 3045.
25. J. Vanderkooy, S. Norcross, "Multitone Testing of Audio Systems", presented at the 101st Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts),* vol. 44, p. 1174 (1996 Dec.), preprint 4378.
26. J. Rich, "A New Class of In-Band Multitone test Signals", presented at the 105th Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts)*, vol. 46, p. 1037 (1988 Nov.), preprint 4803.
27. S. Boyd, Y. Tang, and L. Chua, Measuring Volterra Kernels", *IEEE Trans. Circuits Sys*., vol. CAS – 30, pp. 571 – 577 (1983 Aug.).
28. M. Reed, M. Hawksford, "Identification of Discrete Volterra Series Using Maximum Length Sequences", *IEEE Proc. Circuits Dev. Sys.*, vol. 143, pp. 241 – 248 (1996 Oct.).
29. H. Jung, K. Kim, "Identification of Loudspeaker Nonlinearities Using the NARMAX Modeling Technique", *J. Audio Eng. Soc.* vol. 42, pp. 50 – 59 (1994 Jan./Feb.).
30. A. Dobrucki, P. Pruchinski, "Application of NARMAX Method for Modeling of the Nonlinearity of Dynamic Loudspeakers", 106th AES Convention, Preprint 4868.
31. O. Dyrlund, "Characterization of Nonlinear Distortion in Hearing Aids Using Coherence Function", *Scand. Audiol*., vol. 18, pp. 143 – 148 (1989).
32. J. Kates, "On Using Coherence to Measure Distortion in Hearing Aides", *J. Acoust. Soc. Am.*, vol. 19, pt.1, pp. 2236 – 2244 (1992 Apr.).
33. A. Voishvillo, A. Terekhov, E. Czerwinski, and S. Alexandrov, "Graphing, Interpretation, and Comparison of Results of Loudspeaker Nonlinear Distortion Measurements", *J. Audio Eng. Soc.* vol. 52, pp. 332 – 357 (2004, April).
34. R. Schmitt, "Audibility of Nonlinear Loudspeaker Distortions and their Audibility", presented at the 98th Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts),* vol. 43, p. 402, (1995 May), preprint 4016.
35. R. Schmitt, "Causes of Nonlinear Compression Driver Distortions and their Audibility" , presented at the 99th Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts),* vol. 43, p. 1095, (1995 Dec.), preprint 4131.
36. C. Tan, B. Moore, N. Zacharov, "The Effect of Nonlinear Distortion on the Perceived Quality of Music and Speech Signals", *J. Audio Eng. Soc.* vol. 51, pp. 1012 – 1031 (2003, November).
37. C. Tan, B. Moore, N. Zacharov, and V. Mattila, "Predicting the Perceived Quality of Nonlinearly Distorted Music and Speech Signals", *J. Audio Eng. Soc.* vol. 52, pp. 699 – 711 (2004, July/August).
38. C. Tan, B. Moore, N. Zacharov, and V. Mattila, "Measuring and Predicting the Perceived Quality of Music and Speech Subjected to Combined Linear and Nonlinear Distortion", *J. Audio Eng. Soc.* vol. 52, pp. 1228 – 1244 (2004, December).
39. B. Moore and C. Tan, "Development and Validation of a Method for Predicting the perceived Naturalness of Sounds Subjected to

Spectral Distortion", *J. Audio Eng. Soc.* vol. 52, pp. 900 – 914 (2004, September).
40. B. Moore, "Psychology of Hearing", Academic Press, 1997
41. J. Kates, "A Perceptual Criterion for Loudspeaker Evaluation", *J. Audio Eng. Soc.* vol. 32, pp. 938 – 945 (1984, Dec.).
42. J. Kates, "A Central Spectrum Model for the Perception of Coloration in Filtered Gaussian Noise", *J. Acoust. Soc. Am.*, pp. 1529 – 1534 (1985).
43. E. Geddes, L. Lee, "Auditory Perception of Nonlinear Distortion – Theory", presented at the 115th Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstract)*, vol. 52, p. 1226 (2003 Dec.), preprint 5890.
44. E. Geddes, L. Lee, "Auditory Perception of Nonlinear Distortion", presented at the 115th Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstract)*, vol. 52, p. 1226, 1227 (2003 Dec.), preprint 5891.
45. ITU BS.1387-1, "Method for Objective Measurement of Perceived Audio Quality", International Telecommunications Union", Geneva, Switzerland, 2001
46. ITU-T P.862, "Perceptual Evaluation of Speech Quality (PESQ): An Objective Method for End-to End Speech Quality Assessment of Narrow-Band Telephone Networks and Codecs", International Telecommunications Union, Geneva, Switzerland (2001).
47. R. Beaton, J. Beerends, Michael Keyhl, and W. Treurniet, "Objective perceptual measurement of Audio Quality", *Collected Papers on Digital Audio Bit-Rate Reduction,* N. Gilchrist and C. Grevin, Eds. (Audio Engineering Society, New Your, 1996).

APPENDIX 1. GLOSSARY

Nonlinearity – A broad definition that implies nonlinear properties of a physical or mathematical system that generates new spectral components not occurring in the original input stimulus.

Symptom of nonlinearity – A measured objective reaction of a nonlinear system to a certain input stimulus. For example, the second harmonic of the output signal corresponding to the input sinusoidal signal is a symptom of even-order nonlinearities. There may be a variety of different symptoms all produced by the same nonlinear system.

Nonlinear system – A physical or mathematical system that is characterized by nonlinear behavior, i.e. by the generation of new spectral components not presented in the original signal. Important property of the nonlinear system is the dependence of the output signal's spectrum on the level of the input signal. Superposition is not applicable to the nonlinear system. Contrary to the linear system, the sum of output signals (each of them being a reaction to individual input signal) in the nonlinear system does not equal to the output signal which is the reaction to the sum of the same input signals.

Nonlinear effects (nonlinear mechanisms) – Physical effects that cause generation of new spectral components not presented in the original input signal. For example, in a direct-radiating loudspeaker the dominating nonlinear effects (mechanisms) are the dependence of the suspension's stiffness, voice coil inductance, and the *Bl*-product on the instantaneous position of the voice coil in the magnetic gap.

Nonlinear distortion - A broad definition, often used improperly. Sometimes it is understood as an objectively measured symptom of nonlinearity (for example, the level of harmonic or intermodulation products), sometimes it is addressed to as a nonlinear effect, sometimes it is thought of as a measure of subjective psychoacoustic reaction to the signal impaired by nonlinearity.

Nonlinear contamination - An extraneous signal produced by a nonlinear system and occurring at the output of the nonlinear system along with the undistorted component of the output signal. For example, in case of the sinusoidal input signal, the nonlinear contamination is harmonics. In case of the noise or the non-stationary, non-periodic input signal such as speech or music, the contamination is also a non-stationary signal. A nonlinear system characterized by high order nonlinearity may produce enormous number of instantaneous intermodulation products turning contamination into a noise-like signal if the spectrum of the input signal is wide and dense.

Nonlinear distortion audibility - Psychoacoustical perception of nonlinear contamination of output signal not presented in the input signal. Psychoacoustical reaction to nonlinearly distorted part of the original

signal is associated with harshness, roughness and subjective unpleasantness. Audibility of nonlinear distortion depends on many factors including properties of a nonlinear system (i.e. nature of nonlinear effects generating nonlinear contamination) and the properties of the signal (spectral, probabilistic, and temporary structure of the original signal).

Static (memoryless) nonlinearity – A nonlinearity that has no memory. In other words, the instantaneous level of the output signal of the static nonlinear system is only a function of the level of the input signal at the exactly same moment of time. Broad class of the static nonlinear systems can be presented by algebraic polynomials.

Dynamic nonlinearity (nonlinearity with memory) - A nonlinear system that has memory. The instantaneous level of output signal at a particular moment of time is a multidimensional nonlinear function of the input signal (and in some systems, also the output signal) at this and the previous moments of time. In a broad class of the weakly dynamic nonlinear system, the instantaneous level of the output signal depends on the present and past levels of only the input signal and such systems are always stable. In a broad class of the strongly nonlinear dynamic systems the instantaneous level of the output signal depends also on the past levels of the output signal. Such systems are recursive and they may be unstable. Memory is associated with the frequency-dependent characteristics of a dynamic nonlinear system.

Weakly nonlinear system – A nonlinear system that is characterized by the low level of the generated spectral components, roughly no more that a few percent of harmonic distortion. It is assumed that there is no energy exchange between the fundamental and the distortion products in a weakly nonlinear dynamic signal. A broad class of the weakly nonlinear dynamic systems can be approximated by the Volterra series.

Strongly nonlinear system – A nonlinear system that may produce a high level of nonlinear distortion products including high-order products. Strongly nonlinear dynamic systems may also be characterized by the bifurcation, dynamic instability, and chaotic behavior. All strongly nonlinear systems cannot be approximated by Volterra series because latter do not converge in the presence of a strong nonlinearity. Broad class of strongly nonlinear dynamic systems can be presented by nonlinear differential and difference equations, by recursive NARMAX models, or by the neural networks.

Nonlinear distortion assessment metrics – Quantitative description of the nonlinear contamination presented in the output signal of a nonlinear system. Metrics depend on the nature of the input testing signal. For example, for a single sinusoidal signal the metrics might be the total harmonic distortion coefficient, or second harmonic, third harmonic, and subharmonics, etc. For a two-tone input signal the metrics can be the difference or sum intermodulation products, or their sum. Metrics can be the functions of frequency or level of input signal. Metrics can be related to objective assessment of nonlinearity (for example Volterra kernels) or they can be perception related (for example DS or R_{nonlin}).

NARMAX (Nonlinear AutoRegressive Moving Average with eXogenous input) – recursive model of the discrete dynamic nonlinear systems. This model stems from nonlinear difference equations. A sample of the output signal at a certain moment of time is a nonlinear function (usually polynomial one) of the present and past samples of the input signal, and the past samples of the output signal (meaning feedback or recursive properties). It is also assumed that the output signal may contain noise. A linear recursive system may be described by the ARMAX model. The difference is that the linear system's output is a just a sum of the present and past input signal's samples and past output signal samples.

Volterra Series – non-recursive model of weakly nonlinear dynamic systems. In a discrete Volterra model a sample of the output signal at a certain moment of time is a nonlinear function of the present and past samples of the input signal. Volterra series may be considered as Taylor series with memory or as an extension of the linear convolution integral to multidimensional space. Multidimensional impulse responses are called Volterra kernels.

Nonlinear identification – derivation of information about nonlinear system making it possible to predict the system's reaction to an arbitrary input signal. Nonlinear identification should not be confused with measurement of nonlinear distortion that provides only partial information about nonlinear system's behavior. For example the measurement of the first, second and third harmonics of a loudspeaker at certain level of input voltage does not provide enough information to predict loudspeaker reaction to an arbitrary real signal. Much

more information is required for that. For comparison, if a loudspeaker hypothetically does not have any nonlinear effects, the information about its amplitude and phase frequency responses (i.e. complex transfer function) is sufficient to predict loudspeaker's reaction to an arbitrary signal; transfer function can be turned into an impulse response, and the convolution of the latter with the input signal provides output signal (at a certain point in space where the transfer function was measured).

Basilar membrane – A membrane inside the cochlea that mechanically responds to the sound pressure signal coming through the outer and middle ear. Vibrations of basilar membrane are transformed into the corresponding physiological neural reaction that is processed by the brain. High frequency signal causes vibration of the membrane closer to its narrower and stiffer "entrance", whereas the low-frequency signals cause vibrations at the wider and more compliant apex of the membrane. Frequency scale of signals is mapped across the length of the membrane.

Masking – psychoacoustical suppression of weaker signal by the stronger one when the weaker signal becomes inaudible. Masking can be simultaneous (in frequency domain) and unsimultaneous (in time domain). In time domain masking the weaker signals that follow or precede the masker can be inaudible. Masking curves are nonsymmetrical; in frequency domain masking curve produced by the sinusoidal signal has a shape of triangle with steeper left-had side. Similarly, masking curve produced by a short impulse in time domain is a asymmetrical triangle with steeper ascending side.

Excitation pattern – pattern of neural excitation caused by a sound signal. Expressed as a function of auditory filters center frequency.

Auditory filters – array of band-pass filters that characterize the peripheral auditory system. The characteristics and parameters of auditory filters are usually found through masking experiments.

Critical bandwidth – an effective bandwidth of the auditory filter. Was first described by Fletcher in researching of masking of a single tone by a broadband white noise. He discovered that only a narrow frequency band of noise located in the vicinity of the tone's frequency contribute to the masking of the tone. He called this frequency band the critical bandwidth.

Bark scale – psychoacoustical scale. The scale range from 1 to 24 and corresponds to 24 critical bands that cover audio frequency range. The subsequent band edges are (in Hz): 0, 100, 200, 300, 400, 510, 630, 770, 920, 1080, 1270, 2000, 2320, 3150, 3700, 4400, 5300, 6400, 7700, 9500, 12000, 15500.

ERB – equivalent rectangular bandwidth of the auditory filter. Frequency scale expressed in terms of ERB is a perceptually-relevant scale resembling the Bark scale but numerically different. Audio frequency range corresponds approximately to 40 ERB numbers. Each ERB corresponds to about 0.9 mm distance on basilar membrane.

APPENDIX 2. NONLINEAR DISTORTION ASSESSMENT METRICS

THD (total harmonic distortion) **–** an objective metric, probably the least psychoacoustically accurate.

According to recommendations of IEC-268-5 standard the THD is expressed as:
in percentage:

$$d_t(f) = \frac{\sqrt{p_2^2(f) + p_3^2(f) + ... + p_n^2(f)}}{P_t(f)} x100\%$$

in decibel: $L_{dt}(f) = 20\lg\frac{d_t(f)}{100}$

where $p_2(f)$, $p_3(f)$,...$p_n(f)$ are the RMS levels of harmonics of corresponding order and

$p_t(f) = p_1(f) + p_2(f) + p_3(f) + ... + p_n(f)$

is the sum of all harmonics up the order *n*.

Even if as an objective measure of nonlinearity THD is far from perfection. It does not give information about particular harmonics. The same THD level may correspond to dominating second and third harmonics, or it may correspond to the presence of higher order harmonics that are indicative of the high-order nonlinearity associated with audible distortion. Another problem with using THD is that in high-frequency transducers harmonics cannot be measured over the entire frequency range. For example, the second harmonic can be measured only to about 10 kHz, the third to 7 kHz, the fourth to 5 kHz, etc.

Harmonic Distortion – an objective metric. Harmonics are measured separately and are not lumped into THD response. Individual harmonics may convey some important objective information. For example the second harmonic tells about the degree of non-symmetric nonlinear effects, whereas the third harmonic gives information about symmetric nonlinear effect (such as for example limiting of the spider). Although the current standards such as IEC 268-5 and AES2-1984 (r. 2003) recommend measuring only the second and the third harmonics, it seems reasonable to measure at least first five. While the second and third harmonics give information about non-symmetric and symmetric nonlinearities, the levels of the fourth and fifth are indirectly indicative of the "curvatures" of corresponding nonlinearities. If for example the level of the fifth harmonic is close or exceeds the level of the third one, the device under test may be prone to hard limiting. In addition, higher-order harmonics are indicative of the presence of higher order nonlinearity known for generating audible distortion products.

IEC Modulation Distortion – an objective metric. The IEC 268-5 standard defines the two-tone modulation distortion generated by the second and third-order nonlinearities as nonlinear products of frequencies: $f_2 \pm f_1$ and $f_2 \pm 2f_1$ $(f_2 > f_1)$. The standard mentions that measurement of higher order intermodulation components have generally not been found valuable. The modulation distortion of the second order is determined by the formula:

as a percentage: $d_2 = \frac{p_{f2-f1} + p_{f2+f1}}{p_{f2}} \text{x} 100\%$

in decibels: $L_{d2} = 20\lg\left(\frac{d_2}{100}\right)$

The modulation distortion of the second order is determined by the formula:

as a percentage: $d_3 = \frac{p_{f2-2f1} + p_{f2+2f1}}{p_{f2}} \text{x} 100\%$

in decibels: $L_{d3} = 20\lg\left(\frac{d_3}{100}\right)$

The standard recommends presenting the results of measurement as a function of a reference voltage, being the RMS value of a sinusoidal voltage having the same peak-to-peak value as the test signal applied to the loudspeaker terminals. The standard does not specify the frequencies and their amplitude ratio. The AES2-1984 (r. 2003) does not mention measurement intermodulation at all.

CCIF intermodulation distortion – an objective metric using psychoacoustical assumptions that the difference intermodulation products are poorer masked than harmonics and sum intermodulation products due to the inharmonic dissonant nature of difference products and because of the asymmetry of masking curves. The method recommends measurement only the second-order difference intermodulation product:

$$d_{CCIF} = \frac{p_{f2-f1}}{p_{f2}}$$

The method recommends close spacing between tones, for example 3 kHz and 3.05kHz.

IEC Difference Frequency Distortion – an objective metric conceptually similar to the CCIF intermodulation distortion. The method uses two closely spaced tones f_1 and f_2 (usually $f_2 - f_1 = 80$ Hz). The second order difference frequency distortion shall be determined by the formula:

In percentage: $d = \frac{p_{f2-f1}}{p_{f1} + p_{f2}} \text{x} 100$

In decibel: $L = 20\lg\left(\frac{d}{100}\right)$

SMPTE intermodulation distortion – an objective metric based on application of two-tone testing signal. Method recommends using one low-frequency tone (typically 60 Hz) and high frequency tone (typically 3 kHz). The low-frequency tone has four times higher level than the high-frequency one. Modulation of the high-frequency tone by the low-frequency one is measured.

Weighted Harmonics Coefficient method – objective method using some psychoacoustics-related assumptions. The method relates to measurement of harmonics. It is assumed that higher-order harmonics are poorly masked and are accompanied by wide spectrum of dissonant intermodulation products. The metric I s a weighted RMS sum of harmonics that are multiplied by the weighting coefficients either $n/2$ or $n^2/4$ where *n* is the order of harmonic.

Coherence function – objective metric expressed as a ratio of the square of the cross-spectrum (between input

and output) to the product of the autospectra of input and output

$$\gamma^2(f_i) = \frac{\left|G_{xy}(f_i)\right|^2}{G_{xx}(f_i)G_{yy}(f_i)}$$

the autospectra and crossspectrum are calculated as follows:

$$G_{xx}(f_i) = \lim \frac{1}{N} X_n(f_i) X_n^*(f_i)$$

$$G_{yy}(f_i) = \lim \frac{1}{N} Y_n(f_i) Y_n^*(f_i)$$

$$G_{xy}(f_i) = \lim \frac{1}{N} X_n(f_i) Y_n^*(f_i)$$

where *X(f)* and *Y(f)* are complex spectra of the input and output signals *x(t)* and *y(t)* respectively, and * denotes complex conjugation.

Gedd-Lee metric – a metric that indirectly uses psychoacoustical principles. The metric is based on the analytical or numerical representation of static nonlinearity (relationship between the levels of input and output signal) that has three features. The metric is sensitive to the higher order nonlinearity because the latter produces wider and denser spectrum distortion products that are likely to be audible (because of the poor masking). The metric is also sensitive to the nonlinearity impairing low-levels signals and generating distortion products that are likely to be audible. The metric is not sensitive to the offset and gain, because these effects are inaudible.

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

The first term under the integral is responsible for the higher sensitivity to the "low-level nonlinearity", whereas the second term provides information about degree of curvature of the static transfer function *T(x)*.

MTND (multitone total nonlinear distortion) – an objective metric that results from post-processing of the nonlinear reaction to multitone stimulus. The goal of MTND is to minimize excessive information provided by multitone testing and make it more "manageable". The distortion spectral components are averaged in a spectral "sliding window" such as rectangular or Hanning and the averaged value of distortion products is plotted at the frequency corresponding to the center of the window. One of the possible ways to calculate MTND where Hanning window is used, is presented in the expression

$$d_{MTND}(f_i) = 20\log\left(\sqrt{\sum_{k=i-K/2}^{K/2} \left\{D_k\left[\cos\left(\frac{\pi\left|f_i - f_k\right|}{\Delta f}\right)+1\right]\frac{1}{2}\right\}^2} / p_0\right)$$

dB, SPL

where Δf is the width of the frequency window, f_i is the window center frequency, D_k is the amplitude of a sound pressure distortion product at the frequency f_k and *k* is the number of spectral component; $p_0 = 2\text{x}10^{-5}$ Pa.

There are other possible ways to express MTND coefficient, for example:

$$d_{MTND}(f_i) = 20\log\left(\sqrt{\frac{1}{K}\sum_{k=i-K/2}^{K/2} D_k^2} / p_0\right) \text{ dB, SPL}$$