Audio Engineering Society

Convention Paper 7174

Presented at the 123rd Convention
2007 October 5–8 New York, NY, USA

Measurements and Perception of Nonlinear Distortion – Comparing Numbers and Sound Quality

Alex Voishvillo[1]

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

ABSTRACT

The discrepancy between traditional measures of nonlinear distortion and its perception is commonly recognized. THD, two-tone and multitone intermodulation and coherence function provide certain objective information about nonlinear properties of a DUT but they do not use any psychoacoustical principles responsible for distortion perception. Two approaches to building psychoacoustically-relevant measurement methods are discussed; one is based on simulation of the hearing system's response similar to the methods used for assessment of codecs sound quality. The other approach is based on several ideas such as distinguishing low-level versus high-level nonlinearities, low-order versus high-order nonlinearities, and spectral content of distortion signals that occur below the spectrum of an undistorted signal versus one that overlaps the signal's spectrum or occurs above it. Several auralization examples substantiating this approach are demonstrated.

1 INTRODUCTION

The discrepancy between traditional measures of nonlinear distortion and its perception is commonly recognized. THD, two-tone, multitone intermodulation distortion, and coherence function provide certain objective information about nonlinear properties of a device under test (DUT) but they do not use any psychoacoustical principles responsible for distortion perception and audibility. This work attempts to illustrate why the traditional measures of nonlinearity have poor correlation with sound quality. In the previous works of the author the history of the subject and a review of the new methods were carried out [1, 2]. The current work has an emphasis on demonstration of various effects related to the audibility of distortion.

If sound reproduction could be considered as a communication system, then three major components of this system would be the "transmitter" (loudspeaker), the "signal" – reproduced music or speech, and the "receiver" – human auditory system. Sometimes the properties of these components are thought to be simpler than they are in reality. A loudspeaker is a complex dynamic nonlinear system and its description and characterization cannot be carried out by using simple testing signals such as a sweeping tone and measuring its harmonic distortion [3]. Musical or speech signals differ significantly from simple tonal testing signals statistically, in frequency and in time domains. And finally, the human auditory system "receives" and "processes" musical and speech signals in a much more complex manner that a mere frequency analyzer – Fig. 1.

We typically think that the "transmitter" – loudspeaker produces intermodulation (IM) and harmonic distortion, then the musical or speech signal is accompanied by these distortion products and we hear these irritating harmonics and IM components – Fig. 2. Sometimes we forget that the harmonics and intermodulation products are specific frequency components generated by a nonlinear system when a sinusoidal signal or a group of sinusoidal signals is applied to this system. In reality, during reproduction of a musical or a speech signals, the distortion signal does not exist in a form of only harmonics or intermodulation products, it rather "accompanies" the reproduced signal

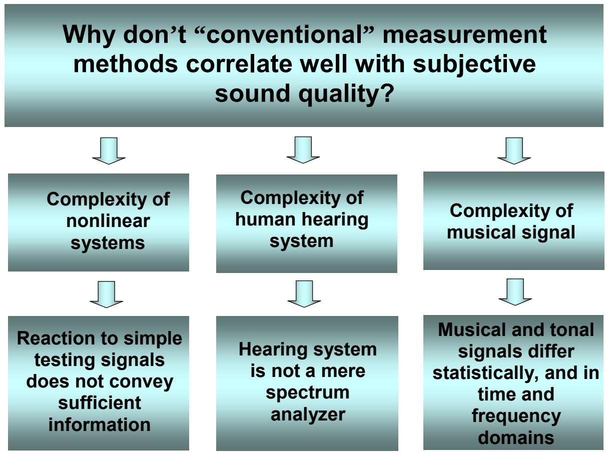

Figure 1: Three reasons why conventional distortion measurement methods do not correlate well with subjective sound quality.

and constantly changes its instantaneous spectrum and waveform. The audibility of this non-stationary distortion signal depends on many factors such as spectral content, temporal and probabilistic properties of the signal, and properties of the human auditory system. A significant role in the distortion perception is played by masking. Testing loudspeaker by a sweeping tone provides only partial objective information about its nonlinear properties, tonal testing signals are a far cry from the real musical or speech signals and the hearing system is an enormously complex nonlinear time-variant system characterized by numerous physiological, psychoacoustical, and cognitive effects – Fig. 3.

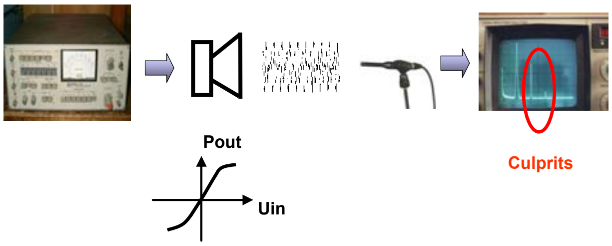

Figure 2: Stereotypical picture of loudspeaker distortion measurement.

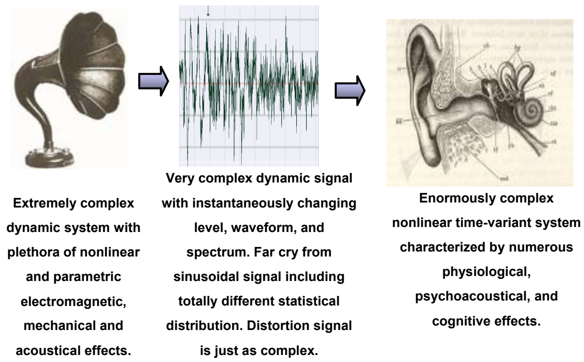

Figure 3: Three major components of sound reproduction and perception.

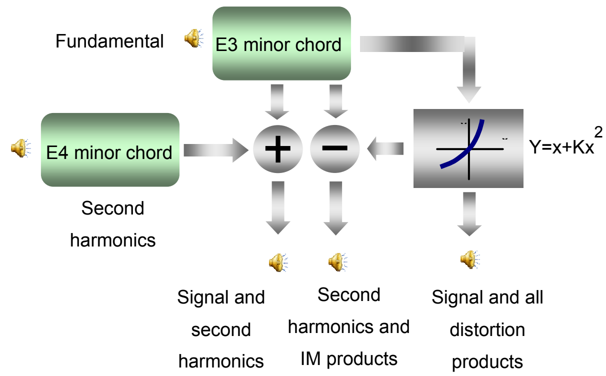

Figure 4: Demonstration of a difference between the second harmonic distortion and the second order nonlinear distortion.

2 HARMONIC DISTORTION VERSUS NONLINEAR DISTORTION

There is some misunderstanding and confusion related to the assessment of nonlinearity in audio equipment and in loudspeakers in particular. One of them is the confusion between harmonic distortion of a certain order and nonlinear distortion of a certain order. The former includes reaction to a single tonal signal only, while the latter implies nonlinear reaction to an arbitrary input signal. For example, it is sometimes believed that the second harmonic distortion may be even benign because the second harmonics are in octave consonance with the fundamental tones.

The first auralization example illustrates perceived difference between musical signal (E3 minor chord) that passed through a nonlinear system generating only second harmonic distortion and a nonlinear system characterized by the second order (quadratic) nonlinearity - Fig. 4.

In the first case the second harmonics are essentially the same *Em* chord shifted one octave up, and it does not adversely affect sound quality, whereas in case of the second order nonlinearity the sound quality significantly deteriorates and the extracted distortion signal is severely non-harmonic with pronounced low- frequency components.

Second harmonics are not irritating but they are indicative of the presence of the second-order nonlinearity. Harmonics are always accompanied by other distortion products if more than one tone is applied to a nonlinear system. If a multitone signal is applied to a DUT, typically harmonics have negligible energy compared to intermodulation products [4], [5]. In reality in a distorted musical or speech signal the distortion signal is a non-stationary one with constantly changing spectrum and waveform.

3 SINUSOIDAL TESTING SIGNAL VERSUS MUSICAL SIGNAL.

The main difference between a sinusoidal testing signal and a real musical signal is their statistical distribution – Fig. 5. Statistical distribution of a musical signal is close to that of the Gaussian noise and the peaks are rare, meanwhile the sinusoidal signal remains nearby its peak levels most of the time. The distortion measurements obtained through the use of the tonal signal may exhibit large level of harmonics or THD, but psychoacoustically it may be rather benign. In the previous works [1], [2] it was demonstrated that the hard clipping producing over 20% THD may be rather acceptable from the psychoacoustical standpoint. It is explained by the fact that the peaks of the signal that is adversely affected by the hard clipping are rare and in addition, the instantaneous distortion products are masked by the signal.

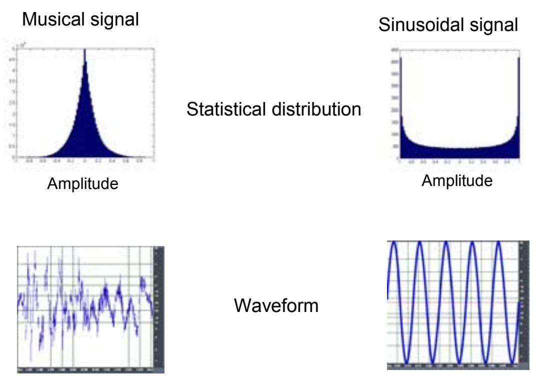

Figure 5: Difference between the musical and sinusoidal signals.

4 "RECEIVER" - HUMAN AUDITORY SYSTEM

The hearing system is a significantly more complex "receiver" that any analyzer used in distortion measurements. In this section only a brief review of the major psychoacoustic and physiological effects will be carried out.

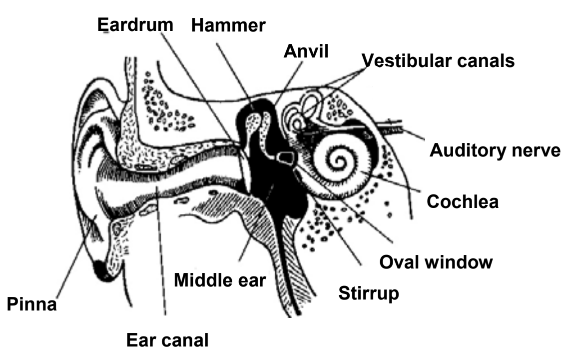

Figure 6: Schematic diagram of human hearing system.

Using analogy with a technical device, the peripheral part of the auditory system plays a role of "acoustical antenna", "diffraction filter", and "preamplifier" that "equalizes" and "amplifies" sound signal, then "microphone" converts the acoustical signal into mechanical vibration of the eardrum and three tiny bones (ossicles) in the middle ear. The ossicles provide transformation of oscillation of air particles with small forces and large displacements to oscillation of cochlea fluid with large forces and small displacements. The ossicles play the role of impedance-matching transformer with transformation ratio about 15. The middle ear also performs a function of a "protection limiter" preventing very high amplitude signals from reaching inner ear. This property is provided by the middle ear muscles' contraction, and it is called acoustic reflex. Mechanical vibrations of the ear drum and ossicles are further transformed into hydraulic wave's propagation in the cochlea. The inner ear plays the role of "spectral analyzer" and "analog-to digital converter" that turns analog signals into a sequence of electrical impulses. Continuing this analogy, hearing system also has a "nonlinear compressor" and a "signal processor" where the initial analysis of the incoming sound signal is carried out – Fig. 7.

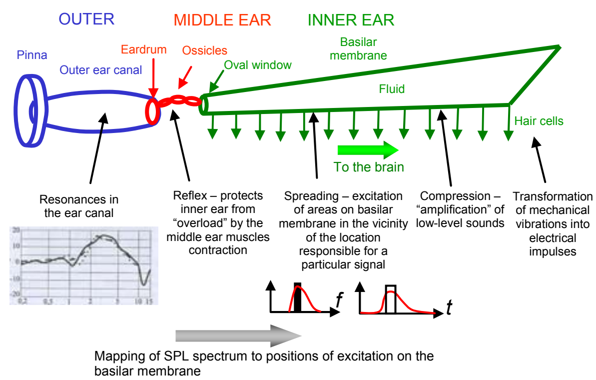

Figure 7: Simplified schematic diagram of the outer, middle, and the inner ear.

Mechanical vibration of the ear drum is converted through the ossicles to the waves of fluid that is contained within the cochlea. The hydraulic waves excite the mechanical movement of the basilar membrane. One of the important properties of the basilar membrane is mapping of the signal's spectral content to positions of mechanical excitations on the membrane. The mapping is accompanied by spreading in time and frequency domains – Fig. 7. If the initial signal has a spectrum of a narrow rectangular band, the excitation on the basilar membrane is "spread" around its central position. Similarly, if the initial signal in time domain has a shape of a short impulse, the excitation has a shape of a "bell" rather than a short pulse. The effect of the auditory spreading lies behind concepts of critical bands and temporal and spectral masking, one of the most important properties of the auditory system responsible for the perception of distortion.

Depending on the spectral content of the initial acoustical signal different parts of the basilar membrane have different positions of maximum vibration. Parts of the membrane responsible for the analysis of the high frequency signal are located in the vicinity of the narrow and stiff entrance to the basilar membrane and the oval window, whereas the low-frequency signals are analyzed at the softer and wider apex of the membrane. The transformation of mechanical vibrations of the basilar membrane into electrical impulses is performed in hair cells attached to the basilar membrane. There are two types of the hair cells, the inner ones which perform the function of the auditory receptors and transmitters of information about external sound events to the brain, and the outer hair cells that perform feedback function and they are responsible for otoacoustic emission and the signals' amplification and compression.

The electrical impulses are directed to the part of the brain responsible for hearing. This part of the auditory system plays the role of a logical processor that decodes and discriminates sound signals, groups them in accordance to certain properties and compares them with "images" stored in memory to assess informational value of the signals, provides cognitive "decoding", and form "images" of the signals. The signals coded in the inner ear are directed to the auditory nerve – Fig. 8. The auditory nerve carries the signal into the brainstem and synapses in the cochlear nucleus. From the cochlear nucleus, the auditory information is divided into two streams similarly to the visual pathways that are split into motion and form processing. Auditory nerve fibers that go to the ventral cochlear nucleus synapse on their target cells with hand-like terminals. The ventral cochlear nucleus cells then project to a collection of nuclei in the medulla called the superior olive. In the superior olive the "fine analysis" of the timing and loudness of the sound in each ear is carried out and the decision about direction of the sound source is made.

The superior olive then projects up to the inferior colliculus through a fiber tract called the lateral lemniscus. The second stream of auditory information starts in the dorsal cochlear nucleus. This stream

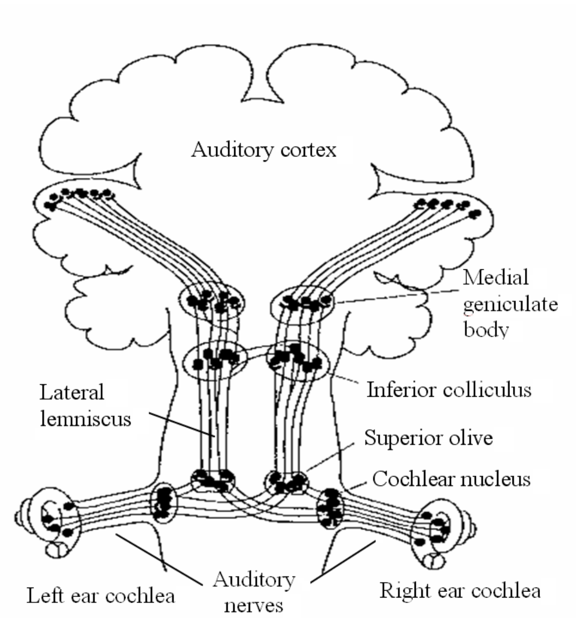

Figure 8. Simplified schematic diagram of auditory system in the brain.

analyzes the quality of sound; tiny frequency differences are analyzed and decisions about particular sound's semantic information are made. This pathway projects directly to the inferior colliculus also through the lateral lemniscus. From the inferior colliculus, both streams of information proceed to the sensory thalamus. The auditory nucleus of the thalamus is the medial geniculate nucleus. The medial geniculate projects to the primary auditory cortex located on the banks of the temporal lobes of the brain [6], [7]. This description is largely simplified and the real processing of signals by hearing system is significantly more complex.

In addition to the spreading and masking, the hearing system is characterized by other psychoacoustic properties such as critical bands, perception of loudness, perception of pitch, and many others. The hearing system possesses a number of nonlinear properties. In the middle ear they are mechanical nonlinearity of the ossicles and the acoustical reflex that works as a limiter and prevents large amplitude signal from reaching inner ear. The inner ear in its turn is characterized by the nonlinear compression that boosts low-level signal and expands the dynamic range of perceived sounds. The cochlea is characterized by the nonlinear otoacoustic emission where the ear actually radiates sound by tympanic membrane in response to short incoming pulses. In addition, the movement of fluid inside cochlea may be affected by turbulence. Also, the basilar membrane is characterized by a mechanical nonlinearity and the excitation of hair cells is a nonlinear process. The masking process which is based on the auditory filtering may be considered a nonlinear process as well. It is obvious that the auditory system is significantly more complex "device" than any analyzer used in traditional objective measurement of distortion.

Since the auditory filters, critical bands and masking play significant role in the perception of distortion, more detailed consideration of these effects is worthwhile. The concept of the auditory filters and critical bands stems from the frequency analysis of signals on the basilar membrane. Different frequencies of the signal are mapped to different positions of maximum excitation (displacement) on the basilar membrane. This effect is somewhat similar to the effect produced by an array of band-pass filters. The bandwidth of each auditory filter is called the critical band and there are 24 critical bands that cover the audio range. This corresponds to approximately 1.3 mm distance between each auditory filter on the basilar membrane. From the psychoacoustical standpoint the analysis of signals whose spectra are confined within a single critical band is different from an analysis of signals whose spectra do not belong to the same critical band. For example if the bandwidth of a noise signal is progressively increased, it has the same loudness until the bandwidth of the noise signal exceeds the bandwidth of the critical band, and a further increase of the noise bandwidth is associated with increased loudness.

The masking, in a nutshell, is a psychoacoustic suppression of a weaker sound signal by a stronger one. Masking may occur in the frequency domain where, for example, a narrow band of noise (masker) may psychoacoustically suppress a lower-level tonal signal if its frequency is close or lies within the frequency band of the masker. Masking curves are obtained through psychoacoustic experiments by applying various maskers and signals and registering audibility of the masked (or unmasked) signals [7]. Masking may also occur in time domain where a higher level short-duration signal may psychoacoustically suppress lower level signals that occur after or even before the occurrence of the masker. Fig. 9 illustrates typical frequency-domain masking curves.

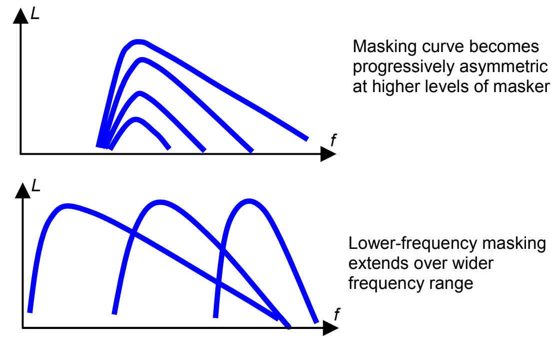

Figure 9: Frequency domain masking curves.

Two important properties of the frequency domain masking can be observed: with progressive increase of the masker's level, the masking becomes more asymmetric and has a tendency to spread over a higher frequency range. It means that the distortion products whose spectrum falls below the spectrum of the signal (e.g. difference intermodulation products) are more likely to be noticed compared to the high frequency distortion products that are more likely to be masked. The second graph showing dependence of the masking curves on the masker's frequency indicates that the lower frequency signals are better maskers than the high frequency signals.

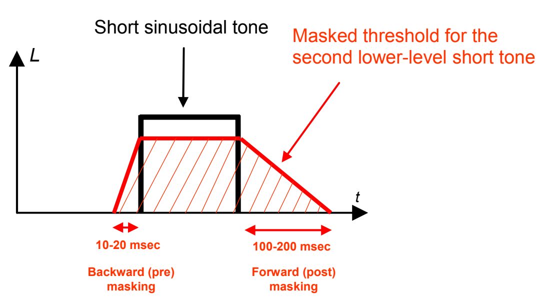

Figure 10: Time-domain masking characterized by pre and post-masking.

Fig. 10 illustrates effect of masking in time domain. In this example a short sinusoidal tone produces masking that occurs not only after the signal is ceased (forward, or post-masking), but even before the signal begins (backward or pre-masking). Fig. 11 illustrates experiment with masking described in Fastl and Zwicker book on psychoacoustics [7]. In this experiment the authors demonstrated the masked threshold of the pure tones masked by a critical band-wide noise centered at 1 kHz and having SPL=70 dB.

Three series of tone triplets are reproduced along with the noise signal. The first series is played at the level of 75 dB, the second at the level of 60 dB, and the third one at the level of 40 dB. Each series consists of six tones triplets with the frequencies 600 Hz, 800 Hz, 1000 Hz, 1300 Hz, 1700 Hz, and 2300 Hz. In the second series the third tone triplet at 1000 Hz is masked by the narrow-band noise, and in the third series the third and the fourth triplet at 1000 Hz and 1300 Hz are masked. This masking effect produces an impression that the masked tones merely disappear.

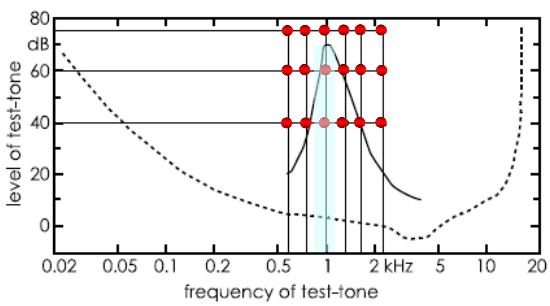

Figure 11: Experiment with a narrow band white nose masking sinusoidal tones (from Fastl and Zwicker, "Psychoacoustics: Facts and Models", Springer 2006).

The auditory masking is just as an objective effect as the visual masking; we can cover a pencil by our palm and it becomes invisible for us, however it exists. Similar effect occurs during auditory masking: some signals physically exist but we may not hear them at all, psychoacoustically they do not exist for us. Auditory masking is a foundation for numerous systems of low-bit compression schemes such as MP3, WMA, AAC, ATRAC, and many others. Components of the signal that are masked by a stronger signal are just removed from the signal and therefore its size becomes much smaller. Masking is also used in psychoacoustics-based methods of assessment of sound quality of codecs [8], [9].

5 DISTORTION AUDIBILITY AND MASKING

Two simple experiments were carried out to illustrate the audibility of nonlinear distortion and the role played by the masking. A schematic diagram of the first experiment is shown in Fig. 12. A musical signal is applied to a two-way system, and the high-pass channel of this system is adversely affected by a static nonlinearity that produces 15% THD in a sinusoidal tone. The crossover frequency was 2 kHz. Audio files could be reproduced at every stage of this system. The signal that passed the high-pass filter and the nonlinear block sounded terrible with a strong emphasis on the low-frequency components whose spectrum fell below the spectrum of the filtered signal. This perceived dominance of the low-frequency distortion components is explained by the poor masking of the lower-frequency components by the higher – frequency components due to the asymmetry of the masking curves – Fig. 8. However, adding the undistorted signal from the low-pass filter played a mitigating effect on the perception of distortion and the annoying distortion components in "tweeter" were partly masked by the undistorted signal from the "woofer".

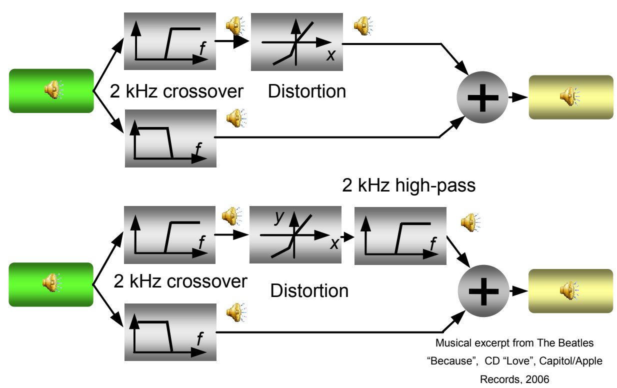

Figure 12: Schematic diagram of the experiment illustrating masking of distortion generated in the high-frequency channel of a two-way system.

In the second part of this experiment the distorting block was followed by another high-pass filter having the same cut-off frequency 2 kHz. It played a dramatic role in mitigating of the annoying effect of distortion products. The distortion signal still could be heard but the removal of the lower frequency distortion components significantly improved sound quality. Adding the signal that comes from the undistorted low-pass channel further improved perceived quality of the sound material. Interestingly, the measurements of THD or harmonic distortion in this two-way system before and after the insertion of the second high-pass filter would not exhibit any difference. Meanwhile application of the multitone stimulus before and after the insertion of the second high-pass filter would clearly show the difference in the spectral content of distortion products – Fig. 13 – 15. The THD metric is "blind" to the difference, lower frequency distortion components; it "sees" only harmonic components having higher frequency than the initial input sinusoidal signal.

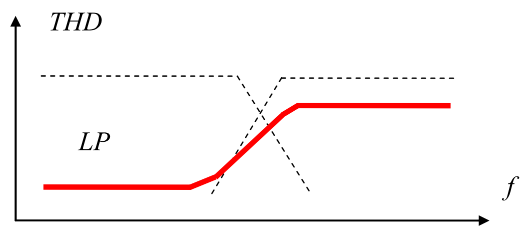

Figure 13: THD before and after connection of the second high-pass filter.

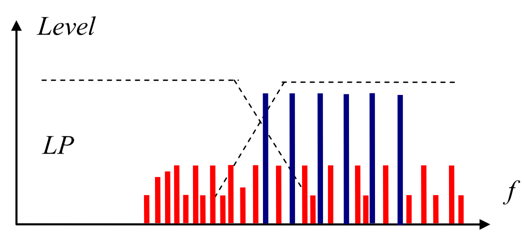

Figure 14: Reaction to multitone stimulus before connection of the second high-pass filter

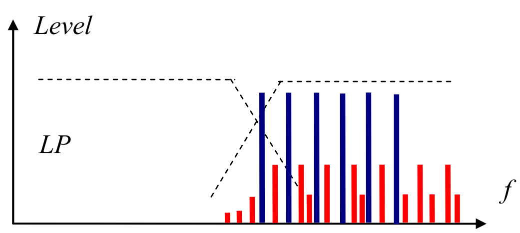

Figure 15: Reaction to multitone stimulus after connection of the second high-pass filter.

Initially the application of multitone would reveal distortion components whose spectrum is lower than the high-pass filter's cut-off frequency, and then, after the second high-pass filter is connected, these "low-frequency" distortion components would disappear. In other words, multitone measurements would have some correlation with perceived sound quality.

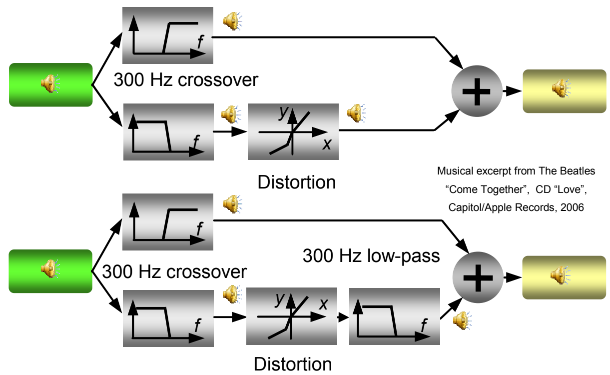

Figure 16: Schematic diagram of the experiment illustrating masking of distortion generated in the low-frequency channel of a two-way system.

In the next experiment the distorting block was placed after the low pass filter – Fig. 16. Musical signal was applied to this two-way system and the nonlinear block (similar to the one used in the previous experiment) was connected after the low-pass filter. Crossover frequency was 300 Hz. Perceived sound quality of the signal that passed the low-pass filter and the nonlinear block was very bad. In this case however, the distortion components whose spectrum was below the signal's spectrum were not pronounced simply because they were not reproduced by a physical woofer used in the experiments. Adding the undistorted signal from the high-frequency channel slightly mitigated distortion in the low-frequency channel but not to the degree that occurred in the previous experiment. This is explained by the fact that a high frequency signal is a poor masker for a low frequency signal.

At the next stage of the experiment the distorting block was followed by another low-pass filter tuned to the same cut-off frequency 300 Hz. Therefore all the distortion products whose spectrum exceeded 300 Hz were suppressed and the sound quality of the distorted signal that passed the second low-pass filter was very close to that of the signal before the distorting block. This shows a strong masking produced by the undistorted signal on the distorted one if the spectrum of the latter remains within the spectrum of the "good" signal. Another conclusion is that at low frequencies the hearing system may tolerate rather high levels of distortion. Adding the undistorted signal coming from the high-pass channel practically restored perceived sound quality of the initial signal. Interestingly, in both experiments the sources of the severe distortion remained "inside" the systems, but simple manipulations with linear devices (high-pass and low-pass filters) made it possible to significantly diminish negative psychoacoustical effect produced by the nonlinear devices. The spectra of the distortion signals were confined within the spectra of the undistorted signal and that was enough to significantly improve perceived sound quality. Masking plays significant role here.

Fig. 17 shows the waveform and spectrum of the sinusoidal signal that passed through the nonlinear block used in the experiments. The spectrum is rich with harmonics, THD corresponds to 15%. Reproduction of 440 Hz tone affected by this nonlinearity in comparison with the "pure" tone produced impression of the augmented high frequencies; the sound became louder and "edgy", but not necessarily annoying and unpleasant. This simple experiment illustrates inapplicability of pure tones for assessment of sound quality. The sonic picture obtained on musical signals is completely different from that of produced by a single tone.

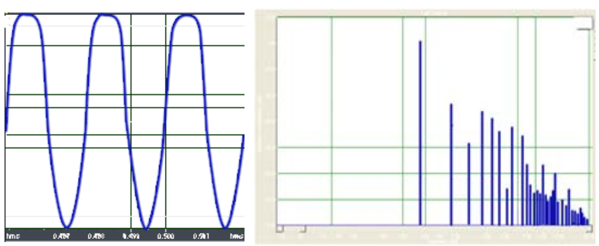

Figure 17: Waveform and spectrum of sinusoidal signal that passed through the nonlinear block.

6 LOW-ORDER NONLINEARITY VERSUS HIGH-ORDER NONLINEARITY. "LOW-LEVEL" NONLINEARITY VERSUS "HIGH-LEVEL" NONLINEARITY.

It is known that a high-order nonlinearity produces harmonics and intermodulation products whose instantaneous spectrum may protrude far away outside the instantaneous spectrum of the undistorted signal and therefore are likely not being masked and thus being audible – Fig. 18.

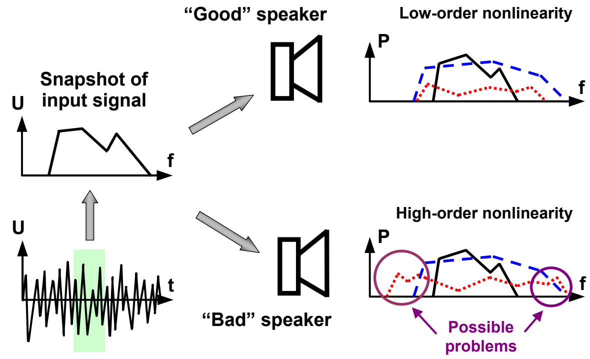

Figure 18: Low-order nonlinearity versus high-order nonlinearity. Instantaneous spectrum of the undistorted signal - solid, masking curve - dashed, instantaneous spectrum of the distortion signal - dotted.

This effect has been known for more than 70 years, although back then it was believed that it is high-order harmonics that are responsible for the audible deterioration of sound quality and the role played by other distortion products was underestimated. Nevertheless, in a document issued by the American Radio Manufacturers Association back in 1937 it was recommended to give progressively increasing weights to high-order harmonics beginning from the third one [10]. In 1950 Shorter published results of his research where he gave stronger weights to the high-order harmonics and obtained good agreement with sound quality [11]. The important role played by the high-order nonlinearity was taken into account by Geddes and Lee in their Gedd-Lee metric [12].

However, there are situations when the lower order nonlinearity exhibiting the same level of harmonic distortion as the higher order one and may produce a worse audible effect. Next experiment is focused on demonstration of this phenomenon and on corresponding explanation why it may happen. A musical signal was put through second and forth order nonlinearities. Figures 19 and 20 show corresponding "Input-Output" functions, waveforms and spectra of sinusoidal signals that passed through these nonlinearities.

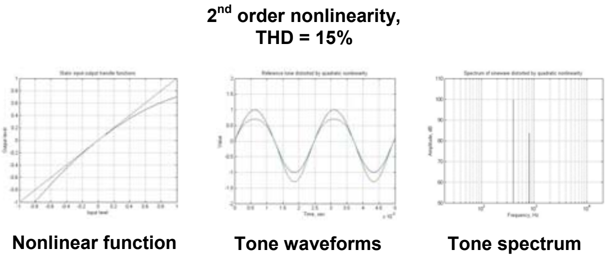

Figure 19: Input-output function of the 2nd order nonlinearity $Y(x) = x - Ax^2$, waveform and spectrum of the sinusoidal signal that passed through the 2nd order nonlinearity.

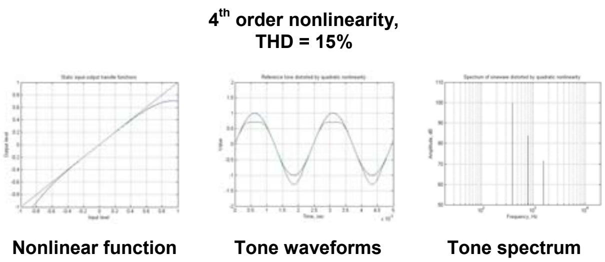

Figure 20: Input-output function of the 4th order nonlinearity $Y(x) = x - Bx^4$, waveform and spectrum of the sinusoidal signal that passed through the fourth order nonlinearity.

Both nonlinearities generated approximately 15% THD. A musical signal was put through both nonlinearities. In both cases the signal was audibly distorted, however, the signal that passed the fourth order nonlinearity sounded significantly better. This counterintuitive result has a simple explanation: the 4th order nonlinearity has a wider "linear" dynamic range, and therefore this nonlinearity starts adversely affecting signal at higher levels. To the contrary, the 2nd order nonlinearity starts distorting signals at lower levels. Taking into account statistical distribution of a musical signal (Gaussian curve) we see that the probability of a signal being distorted is higher if it passes through the 2nd order nonlinearity. This experiment is another demonstration that a nonlinearity that impairs a signal at small levels is worse than a nonlinearity that affects only at high levels of signal. This experiment comes to the similar conclusion as the experiment described in the earlier works [1] and [2] where the hard clipping producing over 20% THD sounded better than the zero-crossing distortion producing only about 3% THD.

Therefore, if the high-order nonlinearity affects signal only at high levels (hard clipping is an extreme case), it may be rather benign. Musical signals typically have Gaussian distribution, large level peaks are rare and they may be well masked by the "good loud signal". A nonlinearity that affects signal at small levels may be rather disturbing because probability of the distorted signal occurrence in this case is very high. However, the THD measured at the signal's "full swing" may be comparatively low. Figure 21 illustrates the ratio of distortion products to the input level as a function of input level. The "A" curve corresponds to a "benign" nonlinearity that generates gradually increasing distortion, however, the low-level signal remains practically unaffected. The "bad" "B" and "C" curves illustrate situations when nonlinearity distorts low-level signals.

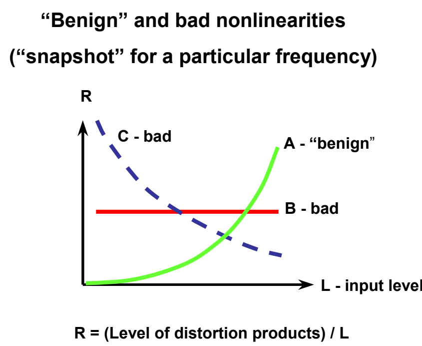

Figure 21: Ratio of distortion products to the input level as a function of input level ("snapshot" for a particular frequency).

Considering a nonlinear loudspeaker, the curve "A" would most likely be produced by dependence of the $Bl(x)$ product and the suspension stiffness $K_{ms}(x)$ on the voice coil displacement, whereas the curves "B" and "C" could be possibly generated by the dependence of the voice coil inductance $L_{vc}(x, I_{vc})$ on the voice coil's position and current, especially if no measures were taken in a particular transducer to mitigate these effects by using copper rings, etc. The functions $Bl(x)$ and $K_{ms}(x)$ are typically flat at low levels of signal, so they may be considered as "soft limiters", whereas the inductance of the voice coil typically changes even at low levels of displacement depending on whether the voice coil moves inside or outside of the magnetic gap.

It is not enough to measure distortion at certain level of the testing signal. It is important to know whether the nonlinearity is affecting the signal at low levels or not. It is also important to know if the spectrum of the distortion signal “protrudes” outside the spectrum of the undistorted signal or if it remains “within” this spectrum and is therefore masked.

7 VARIOUS APPROACHES TO ASSESSMENT OF NONLINEARITY IN AUDIO.

The history of this subject can be traced back to the beginning of the 20th century and it is not the goal of this work to analyze it. The historical retrospect was carried out in [1], [2], and in [4]. Only a brief review of the major directions in assessment of nonlinearity will be given. Fig. 22 illustrates three major groups: nonlinear identification (e.g. Klippel analyzer), measurement of objective distortion (THD, harmonics, two-tone IM, multitone, coherence function, etc.), and finally methods using certain psychoacoustical principles and auditory system models.

Sometimes nonlinear identification is confused with distortion measurement. The principle difference is that nonlinear identification requires much more information about a nonlinear system (such as a loudspeaker). The result of nonlinear identification is a nonlinear dynamic model that makes it possible to simulate reaction of a loudspeaker to an arbitrary signal including audio signal. Once the model is built and its parameters are found, modeling of such characteristics as harmonic distortion or THD become a rather trivial task.

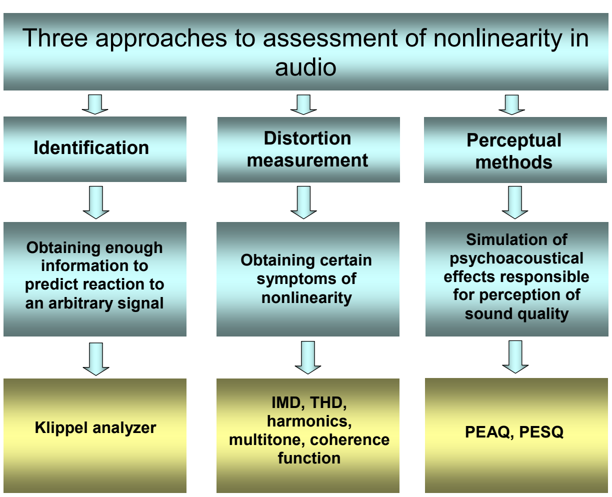

Figure 22: Three approaches to assessment of nonlinearity in audio.

For example, the Klippel analyzer uses noise signal that is applied to a real loudspeaker and simultaneously to the digital nonlinear dynamic model of a loudspeaker [3]. The difference between the measured and the simulated voice coil’s displacement and current is evaluated in a real time mode and the computer adapts the model’s parameters until this difference is minimized or brought close to zero. In this case the model obtains properties of the real loudspeaker used in the test and becomes its “representative”. Nonlinear identification does not bare relationship to psychoacoustics; however, the model could be used for nonlinear auralization and various parameters can be changed in the model giving a chance to assess influence of certain nonlinear effects on distortion audibility.

Regular measurements of distortion expressed in terms of various objective metrics such as harmonic distortion, THD, multitone, coherence function, etc. strictly speaking are not related to psychoacoustics with exception for a few methods that do not use psychoacoustical models but employ certain principles of psychoacoustics [1], [2]. Some of these methods will be addressed in more detail. Fig. 23 gives a classification of “traditional methods” based on different testing signals

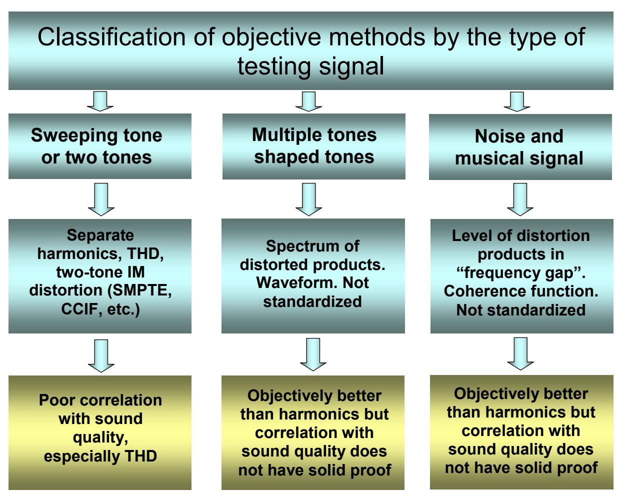

Figure 23: Classification of objective methods by the type of testing signal.

The simplest methods that have the longest history operate with sweeping tone or with two stationary or sweeping tones. Information about nonlinearity is expressed in terms of harmonics, THD, and two-tone IM products (according to recommendations of SMPTE, CCIF, etc.). These metrics, and especially THD, have a poor correlation with perceived sound quality.

The second group of methods uses multiple tones and shaped tones. Information about nonlinearity is expressed in terms of the spectrum of distortion products and waveforms of shaped tones. These methods are not standardized; they provide more objective information about nonlinearities and can reveal certain nonlinear effects not "visible" while measuring for example harmonics and especially THD, (see the experiment described in section 6) but correlation with perceived sound quality does not have a solid proof.

The third group of methods operates with noise or musical signals. The information about nonlinearity is expressed in terms of coherence or incoherence function, or as the level of a distortion signal produced by the input noise signal. Separation of the distortion signal (essentially also noise) and the input noise signal is provided by rejecting a part of the input noise's spectrum to organize a "gap" where the distortion signal is "collected". "Sweeping" the gap across the frequency range and collecting residual distortion signal makes it possible to build some kind of frequency response of distortion. These methods are reviewed in more detail in [1] and [2]. Good correlation with sound quality however has not been proven in the literature; however some of these methods seem to be promising. This subject will be addressed in the next section.

Finally, the perceptual methods widely used in assessment of codecs' sound quality gradually have found its way in assessment of loudspeakers sound quality – Fig. 24. In general, these methods could be divided into "semi-perceptual" and strictly perceptual ones. The former do not use explicit psychoacoustical models, but some commonly accepted psychoacoustical principles such as the stronger audibility of distortion signal whose spectrum is below the spectrum of the undistorted signal, a critical role played by the nonlinearity of high order and a nonlinearity that distorts signal at low levels. Review of these methods is given in [1] and [2].

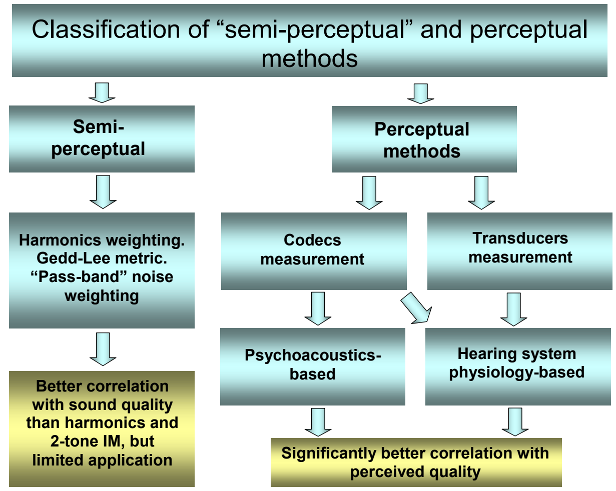

Figure 24: Classification of "semi-perceptual" and perceptual methods.

The "true" perceptual methods came from various schemes targeting assessment of sound quality in codecs and they evolved into two standards, one for assessment of a musical signal (PEAQ) and for speech signal (PESQ) [8], [9]. In general these methods fall into two major categories, one uses psychoacoustical information about masking effects, and the other one is based on physiological model of the hearing system that include various effects such as nonlinear compression, perception of loudness, etc. The latter approach was used recently for the assessment of sound quality in transducers [13] – [15]. These methods are also reviewed in [1] and [2]. Decent correlation with perceived sound quality has been reported by the authors.

8 POSSIBLE FUTURE DEVELOPMENTS

8.1 Perceptual methods

Application of these methods for assessment of transducers' sound quality show promise, but they seem to require more research. One of the major problems mentioned by the authors of [13] – [15] is the problem of distinguishing between the linear and nonlinear distortion in the model. Codecs basically have flat frequency response and this problem is not critical there. The nature of the nonlinearity in codecs is very different from that of transducers. Nonlinear effects in transducers actually add more information to the signal by "injecting" dynamic distortion signal not presented in the initial signal whereas the low-bit compression schemes withdraw information from the initial signal. From the cognitive standpoint, removing information from a signal is less disturbing that adding it. In transducers major nonlinear effects are a function of the voice coil displacement and current, i.e. the generation of distortion signals depends on the level of the input signal and nonlinear parameters of the loudspeaker. In the low-bit compression schemes generation of distortion is based on certain psychoacoustical assumptions about the audibility or inaudibility of distorted signal.

8.2 Semi-perceptual methods

These methods have existed in various forms for decades, such as the recommendations to give weighting coefficients to the measured higher order harmonics [10], [11]. Another example of more recent semi-perceptual approach is the Gedd-Lee metric that distinguishes the high-order nonlinearity and the nonlinearity that affects the signal at low levels 12]. The methods that use a noise signal seem to be promising. One of the methods that was developed over 50 years ago is worth mentioning in the context of the semi-perceptual methods is [16]. Fig. 25 shows schematic diagram of the method.

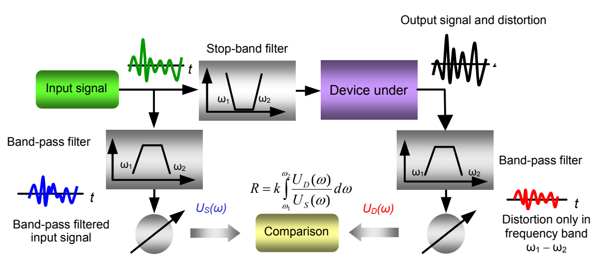

Figure 25: Dynamic method of measuring nonlinearity in audio equipment.

The method is based on using noise or a musical signal. The input signal is passed through the stop-band filter and a narrow-band notch in the spectrum of the input signal is set up. This signal is applied to the DUT and the output signal containing distortion that was "collected" in the narrow-band notch is directed to the band-pass filter with identical cut-off frequencies. Simultaneously the input signal is applied to the band-pass filter with the same cut-off frequencies. Afterwards, average values of the distortion signal and the input signal at the exit of the band pass filter are calculated and the ratio between these two values is calculated as well. This ratio is considered a metric of distortion corresponding to the "tuning" of the stop-band and band-pass filter. Afterwards, the filters are tuned to a new central frequency and the process is repeated consequentially until the entire audio frequency range is covered. The testing results in derivation of a frequency-dependent function *R(f)* . This function is indicative of the frequency-dependent level of dynamically measured distortion. Since the level of the averaged input signal is in the denominator of *R(f)* metric, the value of *R(f)* will be high if the distortion was generated at low level of input signal. This differentiates *R(f)* from other metrics obtained in other methods that use noise or musical signal.

The advantage of this method is its capability of detecting distortion generated at low levels of input signal and the capability of measuring nonlinear dynamic "black boxes" without knowing the nature of their nonlinearity. The disadvantage is the incapability of distinguishing nonlinear products of different orders, and in particular, differentiating "low-frequency" nonlinear products from more benign "high frequency" distortion products.

Using the aforementioned approach, many alternative methods could be designed to differentiate not only the "low-level" nonlinearity from a more benign "high-level" one, but also detect "low-frequency" distortion products and differentiate them from the "high-frequency" ones. A schematic diagram of a possible alternative method is shown on Fig. 26.

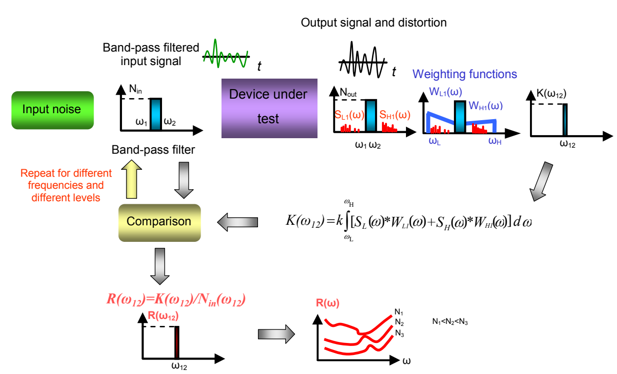

Figure 26: Schematic diagram of alternative method using weighting functions.

The method is using a band of the noise signal or a multitone stimulus. This signal is applied to the nonlinear DUT and the output noise signal will have a spectrum that extends above and below the spectrum of the initial signal. Part of the output signal's spectrum that extends below the spectrum of the signal is given a weighting function that increases towards low frequencies, therefore emphasizing audibility of the low-frequency distortion products. Similarly, the part of the output signal whose spectrum extends above the spectrum of the input signal is also given its own weighting function that increases towards high frequencies. Then both weighted spectra are integrated over a wider frequency range, summed, weighted again by a single number and then the obtained value is divided by a number representing the level of the input signal. The obtained value is plotted against the central frequency of the input signal. Then the noise frequency band is tuned to a new central frequency and the process is repeated again until the entire frequency range of the DUT is covered. Afterwards the whole process is repeated at different level of the input signal.

Weighting the level of the distortion products by the level of the input signal would reveal distortion generated at the low levels of the input signal. Repetition of the measurement process at several different input levels and overlapping obtained response curves or even plotting a three-dimensional surface of distortion would provide psychoacoustically relevant information.

9 CONCLUSION

In this work a review of various approaches to assessment of nonlinearities in audio equipment and in loudspeakers in particular was performed. The basic differences between nonlinear identification (e.g. the Klippel analyzer), traditional methods of objective distortion measurement (THD, harmonics, two-tone IM, multitone, coherence function, etc.) and perceptual methods (typically used for assessment of codecs' sound quality) were examined. Nonlinear identification and "objective" methods do not strongly correlate with psychoacoustics, whereas the perceptual methods employ various models of the human auditory system. In the last several years attempts to use perceptual methods for assessment of nonlinearity in transducers were reported [13] – [15], and the authors of these works demonstrated good agreement with the perceived sound quality.

There is a "gray area" between "objective" methods and perceptual methods. This area is sparsely populated by methods that do not operate explicitly with the models of the human auditory system but use some commonly acknowledged psychoacoustical observations such as the critical role played by higher order nonlinearities, strong audibility of nonlinear distortion generated at low levels of input signal, and the dominating role played by the distortion products whose instantaneous spectrum falls below the spectrum of the initial "good' signal. These "semi-perceptual" methods, being significantly simpler than the full-fledged perceptual methods promise to provide better results than the traditional "objective" methods. Two methods were considered and described in this work; one of them is dated back to the early 1950s and the other one is proposed by the author.

Several simple experiments based on nonlinear auralization demonstrated such effects as a difference between the second harmonic distortion and second order distortion, the critical role played by the masking on audibility of nonlinear distortion, audible differences produced by the lower order (second) and higher order (fourth) nonlinearities, and the dominating audibility of lower spectrum "difference" distortion products A strong masking produced by the undistorted signal on the distorted one whose spectrum is confined within the spectrum of the "good" signal was demonstrated as well. It was also shown that at low frequencies the human auditory system may tolerate higher levels of distortion.

It is believed that there is a great potential for development of both "true" perceptual methods and "semi-perceptual" methods for assessment of psychoacoustically relevant nonlinear distortion in loudspeakers. Application of these psychoacoustically "smart" methods may change the way we look at the traditionally measured distortion and on the way we design loudspeakers.

10 AKNOWLEDGEMENTS

The author is thankful to Dr. A. Dobrucki from Wroclaw University, Poland for letting the author know about the Wolf's publication on dynamic measurement of nonlinear distortion using noise and musical signal and nicely sending the author a copy of this article.

11 REFERENCES

[1] A. Voishvillo, "Assessment of Nonlinearity in Transducers and Sound Systems – from THD to Perceptual Models", presented at the 121st AES Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts)*, vol. 54, No 12, p. 1260 (2006, December), preprint 6910.

[2] A. Voishvillo, "Objective and Perceptual Methods of Assessment of Nonlinearity in Loudspeakers", tutorial presented at the ALMA Winter 2007 Symposium, Las Vegas, January 7th, 2007. http://www.almainternational.org/2007_symposium_training_papers.php#tutorials

[3] W. Klippel, "Loudspeaker Nonlinearities – Causes, Parameters, Symptoms", *J. Audio Eng. Soc.*, vol. 54, No. 10, pp. 907-939 (2006, October).

[4] E. Czerwinski, A. Voishvillo, S. Alexandrov, and A. Terekhov, "Multitone Testing of Sound Systems – Some Results and Conclusions, Part 1: History and Theory", *J. Audio Eng. Soc.*, vol. 49, No. 12, pp. 1011 – 1027, (2001, November).

[5] 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, No. 4, pp. 332 – 357 (2004, April).

[6] "Auditory and Vestibular Pathways", http://thalamus.wustl.edu/course/audvest.html

[7] H. Fastl and E. Zwicker, "Psychoacoustics – Facts and Models", Springer, 2006, third edition.

[8] T. Thiede et. al. "PEAQ-The ITU Standard for Objective Measurement of Perceived Audio Quality", *J. Audio Eng. Soc.*, vol. 48, No. 1/2, pp. 3 – 29 (2000, January/February).

[9] A. Rix, M. Hollier, A. Hekstra, and J. Beerends, "Perceptual Evaluation of Speech Quality (PESQ) The New ITU Standard for End-to-End Speech Quality Assessment", Part 1 - *J. Audio Eng. Soc.*, vol. 50, No. 10, pp. 755-764 (2002, October), Part 2 - vol. 50, No. 10, pp. 765-778 (2002, October)

[10] "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.

[11] D. Shorter, "The Influence of High Order Products on Nonlinear Distortion", *Electron. Eng*. vol. 22, pp. 152 – 153 (1950).

[12] 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, No. 12, p. 1226 (2003, December.), preprint 5890.

[13] 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, No. 11, pp. 1012 – 1031 (2003, November).

[14] 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, No. 7/8, pp. 699 – 711 (2004, July/August).

[15] 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,No. 12, pp. 1228 – 1244 (2004, December).

[16] V. Wolf, "Dynamic Method of Nonlinear Distortion Investigation" (in Russian), *Radiotechnika*, vol. 8, No. 2, pp. 27-37, 1953.