Audio Engineering Society

Convention Paper 5891

Presented at the 115th Convention

2003 October 10–13 New York, New York

Auditory Perception of Nonlinear Distortion

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

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

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

ABSTRACT

A new metric to the perception of distortion was recently proposed by Geddes and Lee (2003). Psychoacoustical data were measured, correlation and regression analysis were applied to examine the relationship and predictive value of this new metric to the subjective assessment of sound quality of nonlinear distortion. Furthermore, conventional metrics such as total harmonic distortion (THD) and intermodulation distortion (IMD) were also compared. Thirty-four listeners participated in a listening task, rating twenty-one stimuli using a 7-point scale. No significant relationships were observed when comparing the subjective ratings with TDH and IMD metrics. Significant correlation ($r=0.95$, $p<.001$) was observed between the subjective ratings and the new proposed GedLee (G_m) metric. Furthermore, robust predictive power was verified utilizing the GedLee metric. GedLee metric has demonstrated remarkable potential to quantify sound quality ratings of nonlinear distortion.

1. Background

Sound quality in audio playback systems has always been a major aspect of audio system design [1]. A significant component in the sound quality assessment of an audio system is distortion. There are two types of distortion, linear (signal level independent) and nonlinear (signal level dependent).

Linear distortion has been studied extensively, reasonably quantified and correlated with subjective perceptual assessments. Nonlinear distortion of audio systems has also been quantified by conventional metrics such as THD and IMD. However, many sound practitioners have noted that these current measures of distortion may be inadequate as a gauge of the subjective perception of an audio system distortion. This issue is exemplified by the recent advent of compressed audio, most notably MP3. It is well known that an MP3 sound transmission can have a measured THD of upwards of 50%, and yet be perceived by listeners as an acceptable quality reproduction. Consequently, the validity of utilizing such traditional metrics to define nonlinear distortion has been questioned.

Geddes [2] proposed a new approach to the quantification of distortion that is significantly different from the traditional metrics. His method is based on well-established concepts in the general theory of nonlinear systems by Schetzen [3]. While well founded in theory, this new approach did not have a simple metric associated with it, and its validity had not been established through psychoacoustic or perceptual data. Geddes and Lee [4] proposed a simple metric (the GedLee Metric - G_m) based on this new approach.

This study aimed to evaluate the psycho-acoustical relevance of this new metric (G_m) for non-linear distortion, in comparison to the traditional metrics (THD and IMD).

2. Method

2.1. Participants

Thirty-four individuals with normal-hearing sensitivity were recruited as listeners in the study. Each had pure-tone audiometric thresholds of 25 dB HL or better at audiometric frequencies from 250 to 8000 Hz [5]. The ages of these participants ranged from 19-39 years (mean = 21). They participated in a single session lasting roughly 1 to 1.5 hours. All listeners were paid for their participation.

2.2. Materials / Apparatus

The reference music was based on a 15-second excerpt from "Music of the Night" from Phantom of the Opera by Andrew Lloyd Weber that featured a male vocalist with orchestral music. This reference was chosen because it captured both musical and vocal passages at extreme levels. It is preferred over simple sine waves or multiple tones because of its complexity and tendency to be revealing of system distortion. An overall impression of the stimulus is assessed because studies have shown that sound quality percepts and spectral characteristics of a stimulus are highly correlated with the overall impression of the signal.

The music was recorded directly from the original compact disc as a wave file, referred to as the reference. Twenty-one stimuli were then simulated using MathCad. The goal was to represent a large array of distortion types so as to have data which was relevant for a wide range of nonlinearity types. Each stimulus file was generated by multiplying the input data samples of the reference by a specified nonlinear transfer function $T(x)$. The nonlinear output stimuli used in this study do not contain any frequency dependence of the nonlinearity. As discussed in Geddes [2], loudspeaker nonlinearities are usually not frequency independent. However, to even begin to study the complex nature of the frequency dependence, one must first have a robust metric independent of frequency. We will return to this discussion in the last section.

It is not feasible to show all twenty-one examples in this paper, hence only four typical examples are shown in Figure 1. These examples demonstrate a distinct deviation from the linear function, as shown by the dotted line. While the primary focus of this study is to understand the perception of loudspeaker distortion, a wide variety of nonlinear transfer functions was applied to yield data that may be applicable to a broad array of systems. The transfer functions used in the study may not be representative of all systems (electronic, acoustical, and mechanical) since different systems tend to generate specific nonlinearities that each has unique characteristics of their own. The intention of this study was to sample a broad spectrum of nonlinearity types that would represent a variety of systems, and not be limited to just a loudspeaker, or an amplifier.

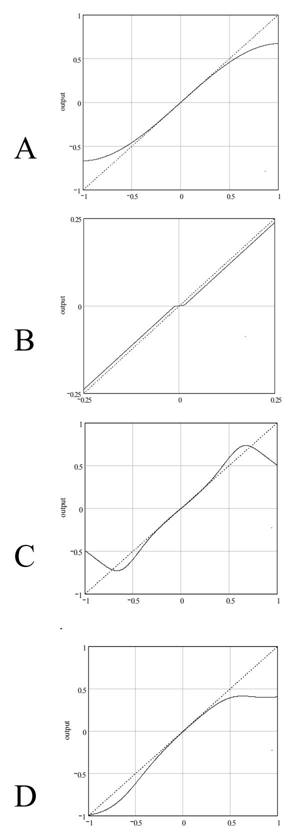


Figure 1. Example of stimuli used in the experiment. In each panel, the solid line represents the input /output transfer function, and the dotted line represents the linear function. Panel A (top) is derived from the Taylor series. Panel B is derived from the second modification. Panels C and D are derived from the Fourier series.

Three modification types were applied to the stimuli used in this study. Initially, a simple model of nonlinearity – Taylor series (*eq.1*) was used to generate gentle and gradual transfer functions, as shown in Figure 1A. Three stimuli (8, 11, 12) were generated based on this modification. Taylor series is known to have extremely slow convergence to sharp discontinuities and found to be satisfactory for only limited class of nonlinearities.

$$T(x) = \sum_n A_n x^n \qquad 1)$$

In order to generate sharp discontinuities in the slope of the transfer functions, which would produce more audible distortion than the Taylor series, a second approach (*eq.2*) was used. This approach could produce a simple discontinuity in the transfer function, as shown in Figure 1B. This method was used to generate four of the stimuli (9, 10, 15, 20).

$$\begin{aligned} T(x) = \; & x - \Delta \;\; if \;\; x > \Delta \\ & x + \Delta \;\; if \;\; x < \Delta \\ & 0 \quad otherwise \end{aligned} \qquad 2)$$

The third method used for modification used a simple Fourier series (*eq.3*), which appeared to be the most proficient for our purpose. By large, this function is the most general, converges rapidly and can yield a very broad range of nonlinearity curves. The Fourier approach provides a general purpose equation that can be used to specify various system nonlinearities in a very efficient manner. The Fouier, or nonlinearity spectral approach to nonlinearity specification will be examined in greater detail in follow up studies. This technique was used to generate fourteen of the twenty-one stimuli in this study.

$$T(x) = A_0 x + \sum_{n=1} A_n \sin(n\pi x) + \sum_{m=0} B_m \cos(m\pi x) \qquad 3)$$

Three measures, or metrics, of the nonlinearity were calculated for each of the twenty-one stimuli. These measures were Total Harmonic Distortion (THD), Inter-modulation Distortion (IMD), and the new GedLee (G_m) metric. Table 1 displays the values of these three measures for all the stimuli used in the study. The THD values were obtained by using a sinusoidal wave file that was processed by the MathCad program, and read into Spectra-Plus software. The Spectra-Plus would then read out the distortion values. The IMD values used a combination of two sinusoidal waves, 100 Hz and 6000 Hz, mixed in a ratio of 4:1 into a single waveform and then likewise input into Spectra-Plus for its calculation. Finally, the MathCad programs would calculate the G_m values directly from the transfer functions using the following equation (*eq.4)*:

$$G_m = \sqrt{\int_{-1}^{1} \left(\cos\left(\frac{x\pi}{2} \right) \right)^2 \left(\frac{\partial^2}{\partial x^2} T(x) \right)^2 dx} \qquad 4)$$

A dilemma occurred in the case of the discontinuous stimuli when calculating the G_m, because the integration would become singular at the discontinuities as the second derivative goes to infinity at that point. This problem was alleviated by using a very small smoothing of the discontinuity in order to avoid the extremely large values of the second derivative at these points.

Table 1. Metrics for the three measures for each stimulus.

Stimulus	THD	IMD	G_m
1	7.20	20.20	3.90
2	19.80	57.20	1.30
3	12.50	32.00	4.80
4	9.20	16.40	2.10
5	1.00	2.10	10.40
6	8.30	10.30	1.20
7	4.20	47.00	23.40
8	9.60	22.40	0.40
9	0.10	0.30	8.50
10	0.02	0.10	12.20
11	6.70	9.40	0.70
12	17.10	18.90	1.20
13	32.60	130.00	2.20
14	27.70	46.20	3.50
15	0.06	16.00	0.36
16	51.00	126.00	2.20
17	6.00	6.00	9.40
18	15.60	38.70	1.60
19	25.40	59.20	1.47
20	20.20	50.10	0.34
21	15.50	34.90	0.86

The output stimuli were saved as 16 bit, 44.1 kHz. .wav files. The sound output was reproduced by a Turtle Beech Santa Cruz sound card. The output transducers used for the study were Etymotic ER-4 MicroPro earphones (Table 2). These earphones are designed to give the most accurate response with normal commercial recordings. They were chosen for their low distortion, natural sound character and common usage in acoustical subjective testing.

Table 2. Earphones specifications

Frequency Response (re:norminal)	20 Hz – 16 kHz (+/- 4 dB) 50 Hz – 10 kHz (+/- 2 dB)
1 kHz Sensitivity	108 dB SPL (1.0 Volt) 98 dB SPL (1mW)
Maximum Output	122 dB SPL

2.3. Procedure

Pure-tone audiometric screening was performed for all listeners immediately before testing. Individuals who did not pass the hearing screening were not allowed to participate in the study. Listeners were tested individually, seated in front of a computer. The overall loudness levels were adjusted at the beginning of the test to ensure a comfortable listening level and were unchanged for the rest of the session. Each listener rated a minimum of 63 (21 stimuli x 3 repetitions) trials and each trial would take a minimum of 15 seconds. No ratings were permitted prior to hearing the complete segment of the stimulus. Thereafter, the subject could either enter the rating on the computer; listen to a replay, or listen to and compare to the reference. After the ratings were entered into the computer, the next stimulus would continue.

Each stimulus was rated a minimum of three times on a seven-point scale, ranging from "better than reference" to "intolerable" (see Table 3). Minor changes were also available ranging from -10 to 50. Randomization was used within each repetition. After three repetitions, the standard deviation of each stimulus was calculated. An additional repetition would be initiated if the standard deviation of a stimulus across the three trials was greater than one. The program would be terminated when the standard deviation of the repetitions was less than one, or when five repetitions were completed, whichever occurred first.

Table 3. Rating scales used by subjcts

7-Point Scale	Sub-Scale
Better Than Reference	-10 to -6
Imperceptible	-5 to 4
Barely Perceptible	5 to 14
Perceptible Slightly Annoying	15 to 24
Annoying	25 to 34
Very Annoying	35 to 44
Intolerable	45 to 50

All stimuli were digitally adapted and presented to the listener via a computer using the ER-4 headphones. Precautions were taken to ensure minimal bias and minimal human errors in the study. Randomization within a repetition, test-retest reliability, double-blind test criteria, computer driven software, and optimal listening levels were applied.

3. RESULTS AND DISCUSSION

Table 4 displays the mean ratings and the standard deviation across subjects of each stimulus. The data are arranged in the ascending order referenced to the perceived ratings. In general, as perceived distortion increased, the standard deviation increased accordingly.

Table 4. Mean rating and standard deviation (SD) for each stimulus across subjects.

Stimulus	Mean rating	SD
6	0.10	0.45
8	0.11	0.43
15	0.15	0.47
11	0.15	0.41
2	0.16	0.41
21	0.21	0.39
12	0.26	0.59
4	0.27	0.48
16	0.37	0.54
20	0.41	0.44
13	0.55	0.62
14	0.58	0.60
19	0.63	0.85
18	1.16	0.72
1	1.26	0.79
7	2.10	0.65
5	2.28	0.74
10	2.31	0.91
3	2.65	0.65
9	3.28	0.78
17	4.18	1.08

The associated scatter plots for the THD, IMD, and G_m metrics and their mean subjective ratings are displayed in Figures 2, 3, and 4, respectively. The regression lines for the data are also indicated.

Figure 2. THD metric versus subjective rating of all stimuli.

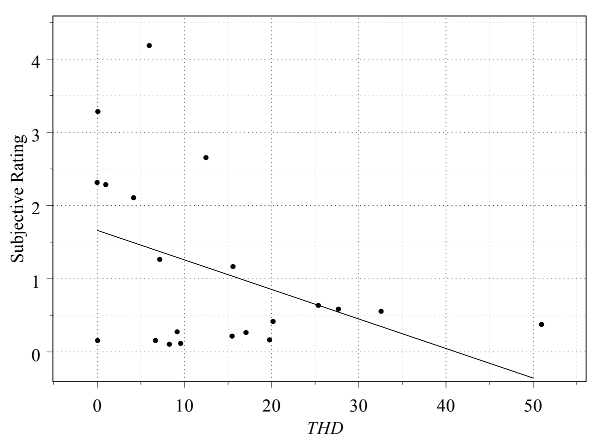


Figure 3. IMD metric versus subjective rating of all stimuli.

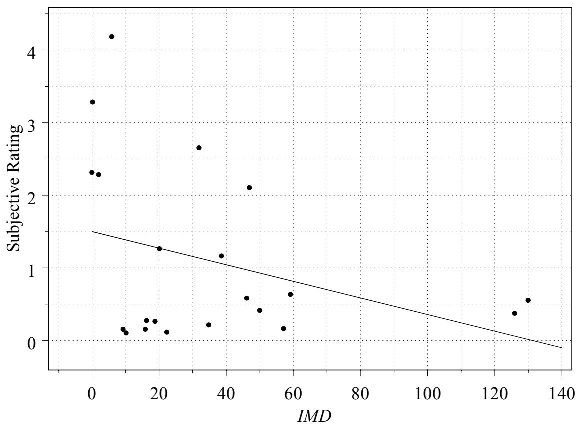
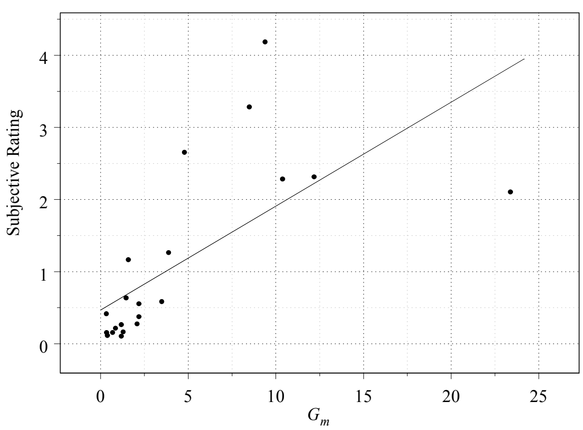


Figure 4. G_m metric versus subjective rating of all stimuli



Of primary interest in the present study was the correlation between the mean subjective ratings and the various metrics. Table 5 provides the Pearson product-moment correlation coefficients between the mean ratings across subjects of the twenty-one stimuli, and each of the three metrics. A negative weak relationship was observed with the THD (r = -.42, p=0.06) and the IMD (r=-0.35, p=0.13) metrics. These results suggested negligible predictive values when utilizing THD and IMD metric in this context. A strong positive correlation (r=0.68, p<0.001), however, was observed with the G_m metric, indicating a significant predictive relationship between the two variables. These results supported the skepticism that THD and IMD metrics were poor predictor of subjective perception of sound quality ratings. The G_m metric emerged to be a relevant predictor of subjective sound quality for nonlinear distortion.

Table 5. Correlations between predictors and mean ratings across stimulus and across subjects.

	Correlation	*P* value
THD	r = - 0.423	$p = 0.06$
IMD	r = - 0.345	$p = 0.13$
G_m	r = 0.68	$p = 0.001$

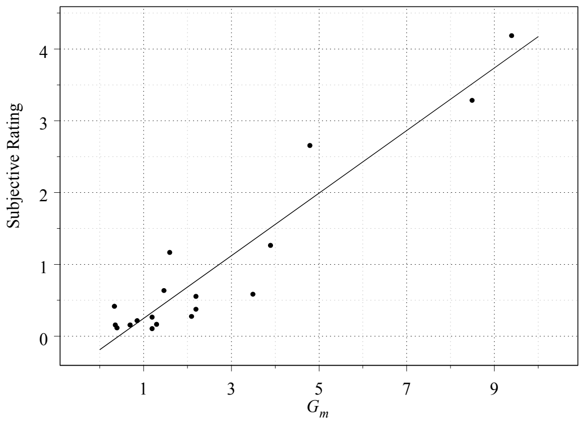
Visual inspection of the data in Figure 4 indicated large variances occurred as the G_m value increased. While the G_m metric correlates well with overall subjective ratings, its variance is large as G_m value approaches 10. Subsequently, stimuli that had G_m values of greater than 10 (# 5, 7 & 10) were excluded from the analysis and correlation coefficients were reestablished with eighteen stimuli. Table 6 displays the mean ratings and the G_m metric of all the stimuli. The shaded stimuli are the ones that were excluded from the final analysis. Using the eighteen stimuli, the GedLee metric showed significant improvement when compared to the full set (r = 0.95, p<.001). These results suggested that G_m metric is optimum when used to predict subjective ratings of nonlinear distortion at low and intermediate levels. It is, however, less appropriate when relate to extreme distortion, but certainly no worse than the others metrics studied.

Table 6. GedLee metric values and the mean ratings for all the stimuli. The shaded stimuli were excluded in the final analysis.

Stimulus	G_m	Mean
1	3.90	1.26
2	1.30	0.16
3	4.80	2.65
4	2.10	0.27
5	10.40	2.28
6	1.20	0.10
7	23.40	2.10
8	0.40	0.11
9	8.50	3.28
10	12.20	2.31
11	0.70	0.15
12	1.20	0.26
13	2.20	0.55
14	3.50	0.58
15	0.36	0.15
16	2.20	0.37
17	9.40	4.18
18	1.60	1.16
19	1.47	0.63
20	0.34	0.41
21	0.86	0.21

A linear regression was calculated for the GedLee metrics considering only the eighteen stimuli. A significant regression equation was found ($F(1,16) = 138.8$, $p<.001$), with an R^2 of 0.9. These results are shown in Figure 5.

Figure 5. G_m metric versus subjective rating of typical distortion levels (eighteen stimuli).



In general, these results suggest that the GedLee metric surpasses the traditionally metrics in both its correlation and predictive value in quantifying sound quality ratings of nonlinear distortion. The results indicate that systems where $G_m < 1.0$ can be expected to yield subjective ratings of "imperceptible" and that levsl of $G_m < 3.0$ can be expected to yield subjective ratings of "barely perceptible but not annoying". Unlike THD or IMD values, these expectations can be given with a very high degree of confidence.

4. SUMMARY

It can be argued that differences might occur with the use of other passages or other signals. It is our firm belief that while there might be differences in the results, the overall conclusions of this paper are not signal dependent. The ratings, correlation numbers and regression slopes might change slightly but we doubt if conclusions that are different than those that we have draw here would be arrived at by the use of different stimuli.

Future research will focus on efficient ways to measure G_m as well as ways to deal with values of G_m that have a frequency-dependence. In loudspeakers, for example, the G_m values, like THD and IMD, will virtually always be frequency dependent and dealing with this frequency dependence presents some interesting issues regarding masking, etc. The main point to be made, however, is that now that we have a metric with a high degree of stability and predictability we can begin to do a whole array of subjective studies of distortion mechanisms that were heretofore impossible to quantify for lack of a value yardstick with which to measure the results.

5. ACKNOWLEDGMENT

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

6. REFERENCES

[1] *"Automotive Audio,"* presented at the 112th Convention of the Audio Engineering Society, Munich, (May 2002)

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

[3] Schetzen, *The Volterra and Wiener Theories of Nonlinear Systems,*Krieger, 1980.

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

[5] American National Standards Institute, *Specification for Audiometers (S3.6-1989*), New York,1989.