Audio Engineering Society

Convention Paper 9549

Presented at the 140th Convention
2016 June 4–7, Paris, France

Modelling the Perceptual Components of Loudspeaker Distortion

Sune L. Olsen[1,2], Finn Agerkvist[1], Ewen MacDonald[1], Tore Stegenborg-Andersen[2], and Christer P. Volk[2]

[1]*Technical University of Denmark*
[2]*DELTA SenseLab*

Correspondence should be addressed to Sune L. Olsen (olsensune@gmail.com)

ABSTRACT

While non-linear distortion in loudspeakers decreases audio quality, the perceptual consequences can vary substantially. This paper investigates the metric Rnonlin [1] which was developed to predict subjective measurements of sound quality in nonlinear systems. The generalisability of the metric in a practical setting was explored across a range of different loudspeakers and signals. Overall, the correlation of Rnonlin predictions with subjective ratings was poor. Based on further investigation, an additional normalization step is proposed, which substantially improves the ability of Rnonlin to predict the perceptual consequences of non-linear distortion.

1 Introduction

Throughout 2003 and 2004 a series of studies with the general goal of producing a model that could accurately predict the perceived quality of non-linearly distorted speech and music signals were performed [2]. These studies culminated in the development of the Rnonlin metric [1] which was shown to be able to provide high correlations (r^2 value of 0.85 and 0.89 for speech and music stimuli respectively with linear regressions) between subjective measurements of perceived distortion and the metric itself. The correlations were obtained for 10 "real" nonlinear systems, 12 artificial systems and one undistorted system.

This study, completed initially as a master thesis [3], aims to investigate whether the Rnonlin metric is generalizable across several types of loudspeaker distortions when measuring perceived non-linear distortion. The Rnonlin metric is calculated by computing, frame-by-frame, the average sum of weighted cross-correlations between a non-linearly distorted signal and a reference signal at the outputs of a gammatone filterbank that models the auditory system. Here, the metric was computed for a set of both simulated and real loudspeaker systems. The computed results were then compared with subjective ratings of perceived distortion.

2 Recordings

A set of 4 samples and 6 systems, both with distinct characteristics, were chosen and recorded/simulated at several different levels to test the generalizability of the metric. 4 of the 6 systems were real acoustic loudspeakers while 2 were simulated; a simple hardclipper and a state space model based on [4] that models

System	Description
Hardclip	Simulated system that processes the input signal so that a certain percentage of the signal is clipped
Model	A state-space model that models the non-linear compliance, inductance and force factor of a woofer
System 1	Medium-sized homebuilt no-brand speaker with an unknown price. Consists of a tweeter and a woofer.
System 2	A small active speaker in the shape of a pig, priced at 45 euro. Consists of 2 tweeters, 2 midrange and a single subwoofer facing downwards.
System 3	Older studio monitor, price unknown. Consists of a woofer with the suspension removed.
System 4	Small active PC speakers, priced at 10 euro. Consists of 2 speakers with a tweeter and midrange in each.

Table 1: The different systems and their descriptions

non-linearities arising from the nonlinear compliance, force factor and inductance of a loudspeaker. A brief description of each system can be seen in table 1.

The samples were composed of a grand piano sample, a contemporary folk song, a speech sample and a set of logarithmically spaced sines comparable to the signal used in [2]. These samples are henceforth known respectively as Piano, Warnes, Speech and Log-Spaced Sines.

To provide the assessors with the audio from the real speakers required for the psychoacoustic experiments, recordings were performed in an anechoic chamber. A laptop sent the audio through to a Fiio E07K Andes soundcard via USB, which sent it into either the speaker directly (for system 2 and the system 4 which were active) or into the NAD amplifier (for system 1 and system 3 which were passive) in the anechoic chamber. System 4 required USB power and were therefore connected to the laptop as well.

The audio was recorded by a B&K 4190 free-field microphone with a type 2669 pre-amp situated one meter away from the center of the given system, to ensure recordings were in the far field. A connected B&K type 5935 dual microphone supply was set to gain with $+30dB$ for all levels below $90dB$, and to $+20dB$ for all levels above. From the microphone power supply, the signal was sent into a Sound Devices 744T HDD Recorder. All samples were combined into one sound file with small "clicks" added as synchronization points between samples, to ease the process of aligning signals later on. Levels were set by placing a NTI Audio MC230 microphone with a type MA220 pre-amp attached to a NTI Audio XL2 sound level meter at the recording location, and measuring the LAeq over 30 seconds of a pink noise signal with an amplitude of 0.1 rms.

Great care was taken to ensure that the influence of any kind of noise was minimal. To ensure this, measurements of background noise for each system were performed. This had to be done separately for each system since some systems had audible internal noise. The frequency response of each system overlayed with the spectra of the noise, recorded at the lowest levels that were to be presented to the assessors, can be seen in fig. 1. It is apparent from these frequency responses that the signal falls below the noise level for all systems below 30 Hz, and even 100 hz for system 4. It was therefore decided to high-pass all samples so that the influence of noise was reduced. The cutoff frequency was set to 50 Hz. While removing some influence of the noise, complete separation was not practically possible as some noise will always present.

The file for the recording of system 1 at 70 dB was unfortunately corrupted. For this reason a lower level recording at 60 dB was used instead, which provided a slightly lower signal-to-noise ratio.

The recordings and simulations were loudness equalized across all samples and systems using DELTA's Loudness Equalizer software. This software is implemented in Labview and utilizes a stationary Zwicker loudness model [5]. The loudness curve at 67 Phon was selected for the equalization.

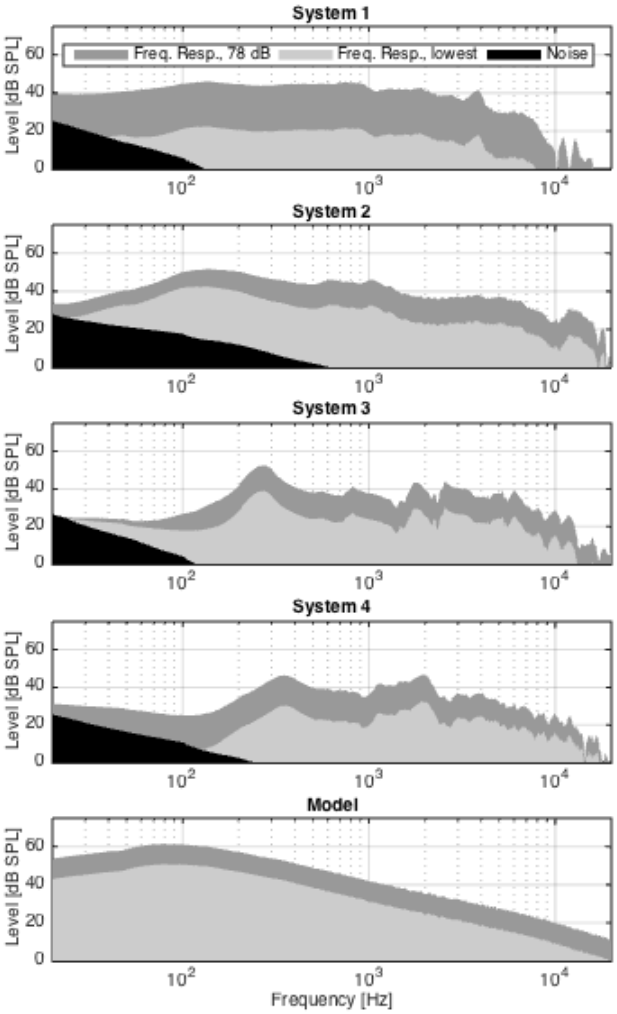

Fig. 1: The frequency response of each system, overlayed with the spectra of background noise, all measured at 70 dB(A) except for system 1 measured at 60 dB due to a corrupted 70 dB(A) recording. No noise was present in the state space model.

3 Rnonlin Analysis

Rnonlin takes into account that non-linearities influence the output of system in different frequency regions. This is achieved by modelling the auditory filters with a filterbank of gammatone filters that cover the audible range. The model is based on the idea that the correlation between the output of a given non-linear system and the input in the matching frequency region can provide information on the amount of distortion. With high correlation, the system is considered less distorted than when the correlation is low.

3.1 Method

A block diagram of the model required to calculate the Rnonlin metric is shown in fig. 2 and the steps are described below. Both are adapted from [1].

1. Two input signals are fed into the system; a signal with non-linear distortion (y) and the same signal without non-linear distortion (x) (known as the input reference).

2. The signals are filtered through a 4097 coefficient FIR filter that corrects for the transfer functions of the inner and outer ear.

3. The filtered signals are sent through a gammatone filterbank with 40 bands of width 1 ERB_N to provide ample frequency resolution. The center frequencies are spaced by their bandwidth from 50 to 19739Hz to cover the audible frequency range.

4. The output from each filter j in the filterbank is then divided into non-overlapping frames i of 30-ms length each.

5. The cross-correlation, r_{xy} between the two processed signals r_{xy} is then calculated according to eq. 1 with lags from -10 to 10ms. Here η is the lag index and L is the length of the frame.

$$R_{xy}(i,\eta,j) = \frac{\sum_{n=(i-1)L+1+\eta}^{iL+\eta} x(n,j)y(n-\eta,j)}{\sqrt{\sum_{n=(i-1)L+1+\eta}^{iL+\eta} x^2(n,j) \times \sum_{n=(i-1)L+1+\eta}^{iL+\eta} y^2(n-\eta,j)}} \quad (1)$$

6. The maximum value of the normalized cross-correlation, Xmax, is determined for each filter in each frame over all values of the lag index. The smaller value of Xmax obtained the less similar the two signals are, and, in turn, the larger the influence of distortion is.

7. To take into account the fact that filters with low output have less impact on the perception of distortion, a weighting function is applied to Xmax.

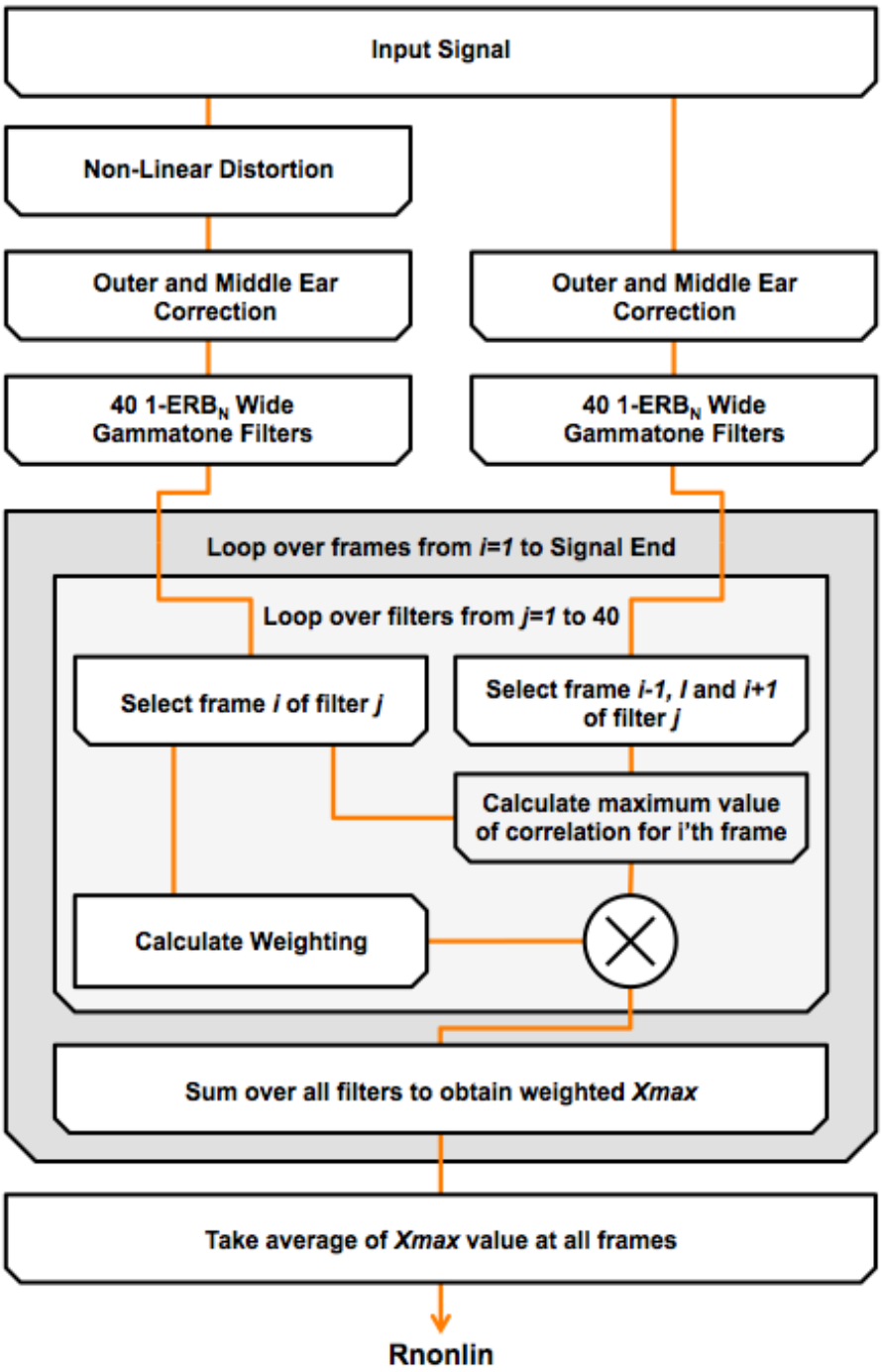

Fig. 2: Flow diagram of the process required to calculate the Rnonlin Metric adapted from [1]

To determine the weighting, the power at the output of each filter is calculated and converted to decibels as shown in eq. 2.

$$Level(i,j) = 10 \cdot log_{10}\left(\frac{1}{L}\sum_{n=(i-1)L+1}^{iL} y^2(n,j)\right) \quad (2)$$

For the specific frame the maximum value of $Level(i,j)$ is determined, known as $Level_{max}(i)$. The weighting is then found based on three cases.

- If $Level(i,j)$ of a specific filter is within 40 dB of $Level_{max}(i)$, the maximum weighting is applied to Xmax for that filter.
- If $Level(i,j)$ of a specific filter is more than 80 dB below $Level_{max}(i)$, a weighting of zero is applied to Xmax.
- If $Level(i,j)$ lies between 40 and 80 dB below $Level_{max}(i)$, the weight applied to Xmax increases linearly as $Level(i,j)$ increases.

The weights are all normalized so that they sum to unity.

8. The weighted values of Xmax are summed across filters for each frame, giving an overall measure of correlation for that frame.

9. The average correlation across the frames is then found. This number is then defined as the metric Rnonlin.

Rnonlin is susceptible to influence from linear distortion. In this study this was not a desirable trait which would be included in the metric if the distorted loudspeaker recordings were compared to the source signal directly. Therefore the input reference, in the case of the 4 real systems and the model, was the lowest level recording/simulation where the non-linear distortion was assumed to be imperceptible. The hardclip system provided no linear distortion and the input reference was therefore the source signal.

3.2 Modifications to Rnonlin

It was observed that noise, such as background and electrical noise, influenced the metric as well. This noise, when present in the non-distorted input offset the Rnonlin metric for the higher level samples. This

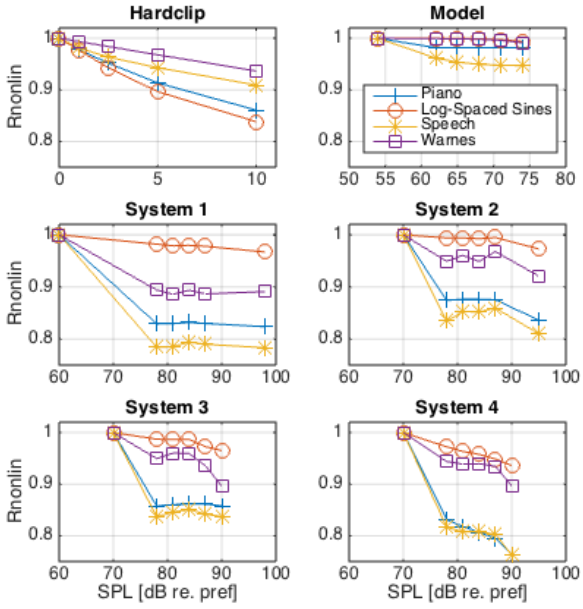

Fig. 3: Initial results obtained from running all recordings and simulated samples through the original Rnonlin model. Note that the x-axis for the hardclipping system is in percentage of sample clipped.

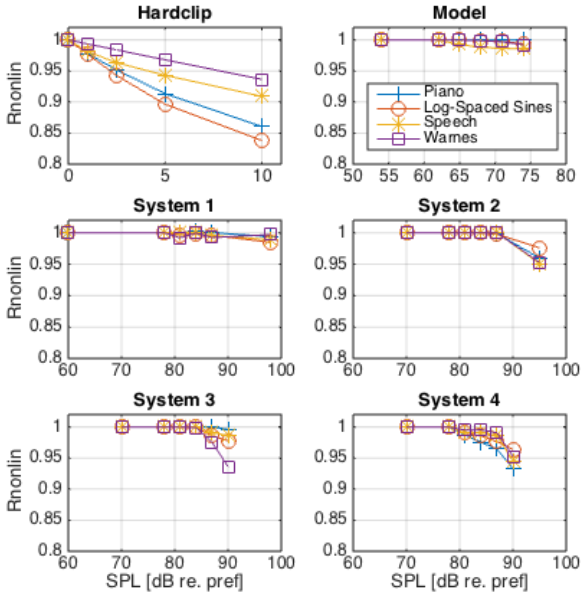

Fig. 4: Results obtained from running all recordings and simulated samples through the modified Rnonlin model.

can be seen in the results obtained when using the unmodified model shown in fig. 3.

The lowest level points for all conditions are the reference files which are always equal to one. It can be seen, especially for the speech and piano samples in the real systems, that the difference in Rnonlin value between the non-distorted input and the first level is very large. The difference between the higher levels is much less pronounced. Since the audible distortion was determined, by the assessors, to only occur at higher levels, this large difference was assumed to be due to audible noise. The same problem does not occur for the hardclipping system since no noise is present in the undistorted source file.

An attempt to remove the effect of noise was performed by assuming that the non-linear distortion only became audible when the Rnonlin value began to decrease a second time when increasing the level. Utilizing this assumption, all Rnonlin values lower than the level at which the decrease began were set to one and all higher levels were increased by the difference between 1 and the Rnonlin value of the level at which the decrease began. The results from using this modified version of the Rnonlin can be viewed in fig. 4.

4 Listening Test

A total of 12 assessors performed the listening test, with ages ranging from 20 to 46 with a median age of 30. Four different levels were used for each system, the values of which can be seen in table 2. Each condition was rated twice.

4.1 Setup

The setup for the listening test was based on the method used in [6]. Upon arrival assessors were given an introduction to the experiment where they were told to try and ignore both noise and linear distortion, and only focus on non-linear distortion. Assessors interacted with a screen where the six different systems with a specific sample and level, along with the clean reference, were presented to the assessor. The assessor was then tasked with rating the amount of non-linear distortion on a given scale ranging from 0 (not distorted) to 150 (extremely distorted). The assessor was free to listen in any order, and zoom in on specific sections if necessary. The audio was presented to the listener through a Sennheiser HD650 headset connected to a PC via a Firestone Audio Fubar III DAC.

4.2 Results

An overall impression of the results obtained from the listening tests can be gained through fig. 5 showing the subjective ratings for the systems and levels,

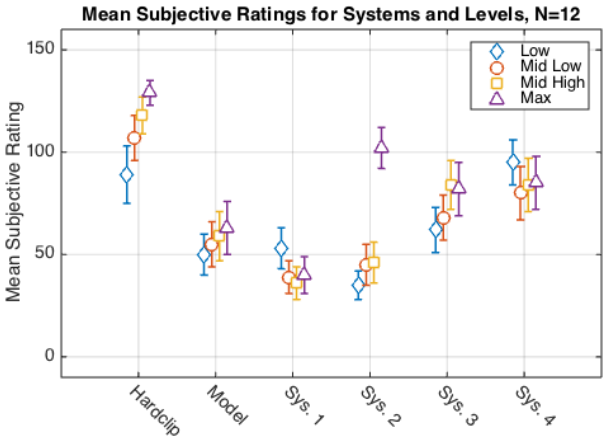

Fig. 5: Listening test results for each system and level

It can be seen that, in general, distortion is perceived to increase as the level of the system increases. In system 2 it is clear that the distortion perceived is low for the three first levels and has a sudden increase at the max level. This indicates that the range between these is where distortion becomes very audible for this system. The hardclip and model systems show a steady increase in average ratings as levels increase as well. System 3 seems to saturate at the two highest levels, with the mean rating for the maximum level being rated a bit lower than the $87dB(A)$ level. There is a bit of uncertainty visible through the overlapping confidence intervals in both the model and system 3, but in general the results are good.

4.3 Discussion

Two unexpected trends in the data can be observed: The lowest level for both System 1 and System 4 are rated higher than the other levels. This is very likely due to the fact that the lowest level of system 1 was at $60dB(A)$ instead of $70dB(A)$ and therefore the signal to noise ratio is lower. System 4 also provided an inconsistent amount of noise throughout recording, which also serves to underline the point that assessors may very well have difficulty separating noise from distortion.

The three highest mean values for system 4 and system 1 are also almost identical in mean and confidence

System	Low	MidLow	MidHigh	Max
Hardclipper	1%	2.5%	5%	10%
Model	54 dB(A)	68 dB(A)	71 dB(A)	74 dB(A)
System 1	60 dB(A)	84 dB(A)	87 dB(A)	90 dB(A)
System 2	70 dB(A)	84 dB(A)	87 dB(A)	95 dB(A)
System 3	70 dB(A)	84 dB(A)	87 dB(A)	98 dB(A)
System 4	70 dB(A)	84 dB(A)	87 dB(A)	90 dB(A)

Table 2: Different levels for the different systems used for the listening test.

intervals. This would make sense if no distortion was perceived, but system 4 is rated moderately high in distortion. If noise is less perceivable at higher levels, it seems more likely that assessors are rating distortion as high for system 4 because it simply sounds terrible. System 4 was in fact singled out and commented on by several assessors after having completed the listening test as being difficult to rate.

5 Comparison of Results

To gauge the improvement in generalizability from the original to the modified versions of Rnonlin, the subjective ratings were plotted along with the Rnonlin values and a linear regression. The results for the original Rnonlin can be seen in fig. 6, while the results for the modified Rnonlin can be seen in fig. 7.

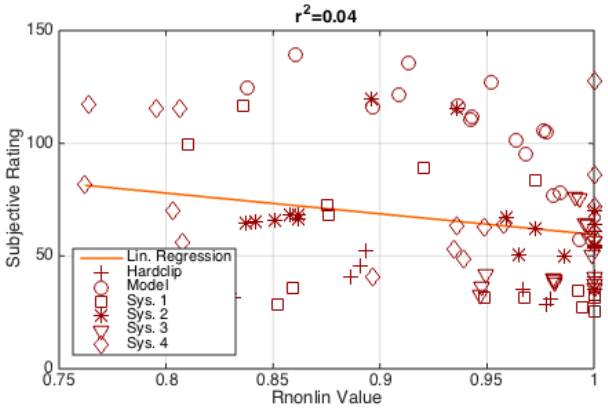

Fig. 6: Results from plotting the original Rnonlin metric values against the subjective ratings

The modification of the Rnonlin provides a vast improvement over the original, indicating that modifying Rnonlin brings the metric closer to actual perception.

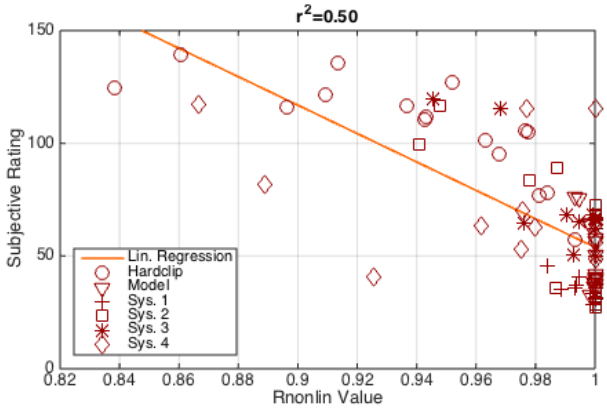

Fig. 7: Results from plotting the modified Rnonlin metric values against the subjective ratings

From looking at fig. 7 it can be observed that the system that is most unlike the others is system 4. It provides high subjective ratings, even when Rnonlin predicts low distortions. Very likely this is because system 4 exhibited fluctuating electrical noise for all levels that could not be removed. The Rnonlin level would invariably decrease due to this difference in noise and not due to an increase in distortion as intended. For this reason system 4 was excluded and the results in fig. 8 were obtained.

With the removal of system 4 a pattern also emerges. Much as is indicated in [1], a curvilinear fit would be suitable to transform the data. From looking at the data, a polynomial transformation seemed a possible solution. Several powers were tried with the best fit being achieved with eq. 3.

$$PredictedRating = Rnonlin^{23} * (-150) + 150 \quad (3)$$

Using this transform, predicted ratings were acquired and can be seen in fig. 9. While not being as high as the results in [1], this is still a vast improvement to the

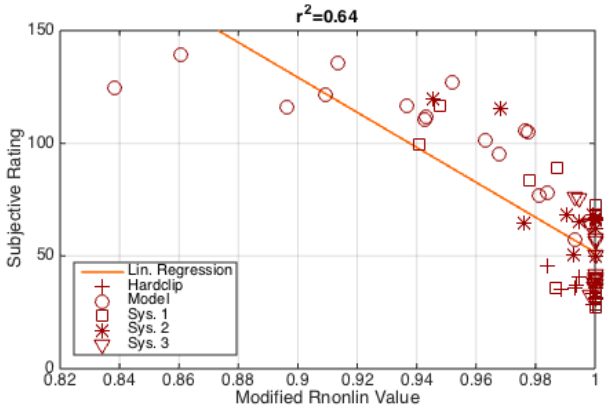

Fig. 8: Results from plotting the modified Rnonlin metric values against the subjective ratings with the removal of system 4

results obtained with the original Rnonlin, and indicate that the Rnonlin can be generalizable for several types of loudspeaker distortion if the modification is applied.

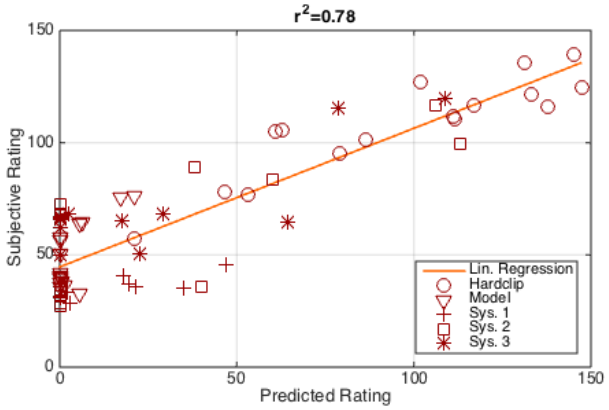

Fig. 9: Results from plotting the modified and transformed Rnonlin metric values against the subjective ratings with the removal of system 4

6 Discussion

The results from this study can be summed up in four points:

- The original version of the Rnonlin metric was not found to be generalizable for predicting the subjective rating of non-linear distortion from loudspeakers in this study.
- Choosing a low level recording of a specific system as the input reference removed the influence of linear distortion.
- Modifying the Rnonlin metric by assuming that distortion only becomes perceptible at higher levels provides significantly improved correlations with subjective measurements.
- Instability of speaker systems and the audible noise arising from this can, in some cases, render both the subjective ratings and the Rnonlin metric incompatible.

With regard to the first point, it is interesting to note that in the initial studies [1] the correlations were quite high. This was not the case in this study which begs the question of what the difference is between the two. The answer to this question may very well lie in the difference in the listening test paradigm, more specifically what the assessors were asked to rate. In [2] the test paradigm for [1] is described. Here it is stated that assessors were asked ”to rate the perceived quality on a 10-point scale where 10 indicates ”clean, completely undistorted” and 1 represents ”very distorted””. In this case, listeners were not tasked with ignoring linear changes to the signal or noise as they were in the listening test performed in this study. This comes across as a bit counter-intuitive when [1] states that ”.. in this paper we focus on a method for predicting the perceived effects of non-linear distortion only”. This would indicate discrimination between the linear and non-linear part of the signal, which the test paradigm does not include. It would seem that Rnonlin is influenced by noise, linear and non-linear distortion, and if the listeners are not discriminating the correlation would probably rise.

From [1] it was not clear whether the input reference was a low-level recording or the source signal file. Both were tried, but the influence of linear distortion provided very inadequate results when using the source file and it was decided this would not be used. Instead low level recordings of the systems were used which removes the influence of linear distortion, but may provide lower Rnonlin values when even the slightest amount of noise is present in the recording.

Regardless, the improvement to the correlation with the subjective measurements in this study when using the modified version of the Rnonlin still stand. However,

further testing would need to be performed to ensure the changes are generalizable for other types of systems and samples.

As mentioned in the final point, some speaker systems may simply be so inadequately designed that the Rnonlin metric fails to provide satisfactory results. While certainly presenting an problem, the implications of this are not major in a general use case since such a system would have several other priority issues that would need to be resolved before the determination of non-linear distortion became a factor.

7 Summary

Throughout this study the generalizability of the Rnonlin metric was investigated. This was by comparing subjective measurements of non-linear distortion with the metric itself for 6 loudspeaker systems and 4 samples.

The comparison between the subjective ratings and the original Rnonlin metric provided inadequate results. The r^2 value when all systems, samples and levels were compared was very close to zero.

A major leap in correlation between subjective values and the Rnonlin metric occurred when the metric was modified in an attempt to remove the effect of noise. This provided an r^2 value of 0.64. From the shape of the data it seemed that performing a simple polynomial transformation would provide an even better result. This yielded an r^2 value of 0.76.

It is recommended that the Rnonlin be tested for more loudspeakers to ensure the changes are generalizable.

References

[1] Tan, C.-T., Moore, B. C., Zacharov, N., and Mattila, V.-V., "Predicting the perceived quality of nonlinearly distorted music and speech signals," *Journal of the Audio Engineering Society*, 52(7/8), pp. 699–711, 2004.

[2] Tan, C.-T., Moore, B. C., and Zacharov, N., "The effect of nonlinear distortion on the perceived quality of music and speech signals," *Journal of the Audio Engineering Society*, 51(11), pp. 1012–1031, 2003.

[3] Olsen, S., *Modelling the Perceptual Components of Loudspeaker Distortion*, Master's thesis, Technical University of Denmark, 2015.

[4] Klippel, W., "Loudspeaker nonlinearities - Causes, parameters, symptoms," in *Audio Engineering Society - 119th Convention Fall Preprints*, volume 1, pp. 256–291, 2005.

[5] Zwicker, E. and Scharf, B., "A Model of Loudness Summation," *Psychological Review*, pp. 3–26, 1965.

[6] Pedersen, T. H., "Perceptual characteristics of audio: The sound wheel can be used to provide objective description of the sound," 2015, tech document no. 7 2015.