MEASUREMENT OF HARMONIC DISTORTION AUDIBILITY USING A SIMPLIFIED PSYCHOACOUSTIC MODEL - UPDATED

STEVE TEMME, PASCAL BRUNET[2], AND PARASTOO QARABAQI

[1] *Listen, Inc. Boston, MA 02118, USA*
stemme@listeninc.com
[2] *Listen, Inc. Boston, MA 02118, USA*
pbrunet@listeninc.com
[3] *Listen, Inc. Boston, MA 02118, USA*

A perceptual method is proposed for measuring harmonic distortion audibility. This method is similar to the CLEAR (Cepstral Loudness Enhanced Algorithm for Rub & buzz) algorithm previously proposed by the authors as a means of detecting audible Rub & Buzz which is an extreme type of distortion[1,2]. Both methods are based on the Perceptual Evaluation of Audio Quality (PEAQ) standard[3]. In the present work, in order to estimate the audibility of regular harmonic distortion, additional psychoacoustic variables are added to the CLEAR algorithm. These variables are then combined using an artificial neural network approach to derive a metric that is indicative of the overall audible harmonic distortion. Experimental results on headphones are presented to justify the accuracy of the model.

1. INTRODUCTION

Distortion caused by non-linear components of audio systems (e.g. loudspeakers, phones, and hearing aids) is a known issue to electro-acousticians. For example, a loudspeaker with a magnetic or electric field whose strength changes with diaphragm position exhibits non-linear distortion. Harmonic distortion is a special type of distortion that is specific to pure tones and results in the generation of multiples of the signal's frequency at the output. Asymmetrical non-linearities, such as in the example above, result in even order harmonic distortion products, while symmetrical non-linearities result in odd order harmonic distortion. Low order harmonics (typically harmonics 2 and 3) are especially useful for loudspeaker design. In an electromagnetic driver, the force factor of the motor, the electrical inductance of the voice coil and the stiffness of the cone suspension are nonlinear functions of the cone excursion. A strong harmonic 2 will indicate an asymmetry of these characteristics along the excursion range and/or poor centering of the voice coil in the magnetic gap. A strong harmonic 3 will reveal a symmetrical alteration at both ends of the excursion range. These measurements offer valuable insight to the designer as to what parameters to adjust (e.g. the geometry of the pole piece, the length of the voice coil, the rest position of the cone, etc.) for optimum sound. It is noticeable that harmonic components are not always audible due to the masking effect of the fundamental frequency and they are not always unpleasant [4]. For example, even order harmonics (especially powers of two multiples of the fundamental frequency) coincide with perfect octave intervals on the musical scale and can even enhance the sound. It is therefore desirable to investigate a method for measuring the perceived harmonic distortion.

Research has been underway for several decades on designing tests for measuring nonlinear distortions. Conventional objective metrics for distortion measurement such as Signal to Noise Ratio (SNR) or Total Harmonic Distortion (THD) do not show reliable correlation to the ear's perception of harmonic distortions and a standard method for measuring the audibility of harmonic distortion does not yet exist.

In recent years, methods have been proposed that take into account the psychoacoustic effects in measuring the audibility of distortion. One example is the work by Geddes and Lee [5] that incorporates the signals' masking patterns in selecting a metric for auditory perception of nonlinear distortion. This method ignores several other psychoacoustic effects and may hence be too simplistic to fully capture the perceptual properties of harmonic distortion.

The Perceptual Evaluation of Audio Quality (PEAQ) standard, which was developed by the International Telecommunication Union (ITU) for objective measurement of perceived audio quality, provides a set of tools for measuring audible distortion. The PEAQ standard has two versions; a basic version that is designed to allow for a cost efficient real time implementation, and an advanced version that helps achieve a higher accuracy at the cost of more complexity. Both versions take a reference signal

(stimulus) and a test signal (response) as input. The input signals are then fed into a Fast Fourier Transform (FFT)-based model (applicable to both versions) or a filter bank-based model (for the advanced version only). These models apply the effects of several psychoacoustic mechanisms to the signals, such as adding the outer and middle ear frequency response, transformation to pitch domain representation, adding the frequency dependent internal noise, and frequency and time-domain spreading. The spreading functions model the spectral auditory filters and forward masking effects respectively. Finally, Model Output Variables (MOVs) are generated that describe different properties of the reference and the test signals.

In this paper, we propose an algorithm for the measurement of Perceptual Total Harmonic Distortion (PTHD). The PTHD algorithm is based on the PEAQ standard in terms of incorporating psychoacoustic methods to model the perceptual effects of the auditory system. However, modifications have been made to the methods defined in PEAQ to tune the algorithm to better suit the measurement of perceptual distortion in loudspeakers and headphones. Audio tests are designed for measuring the perceptual harmonic distortion of loudspeakers and headphones, and test results that verify the effectiveness of the proposed algorithm are presented.

The rest of the paper is organized as follows; the PTHD algorithm is described in Section 2. Sections 3 and 4 discuss the audio tests and the artificial neural network approach used for developing the model. The experimental results are discussed in Section 5, and conclusions are summarized in Section 6.

2. THE PTHD ALGORITHM

The algorithm takes as input a reference signal and a test signal that is obtained by playing the reference signal through a Device Under Test (DUT) and recording the played signal. The reference signal is chosen to be a pure tone at a certain frequency and level. This choice enables the study of harmonic distortion that is specific to pure tone signals. In order to fully study the characteristics of each DUT, the algorithm is repeatedly run at several frequencies and levels of the tone.

The reference and test signals are framed in the time-domain and Fourier-transformed to obtain the stimulus and response spectra respectively. The signal spectra are then fed into a series of functions that model several psychoacoustic mechanisms. These functions include adding the outer and middle ear frequency response, transformation to pitch domain representation, adding the frequency dependent internal noise, and frequency-domain spreading due to spectral auditory filters. More details on the structure and the mathematical description of these functions are available in [1] and a schematic overview is shown in Figure 1.

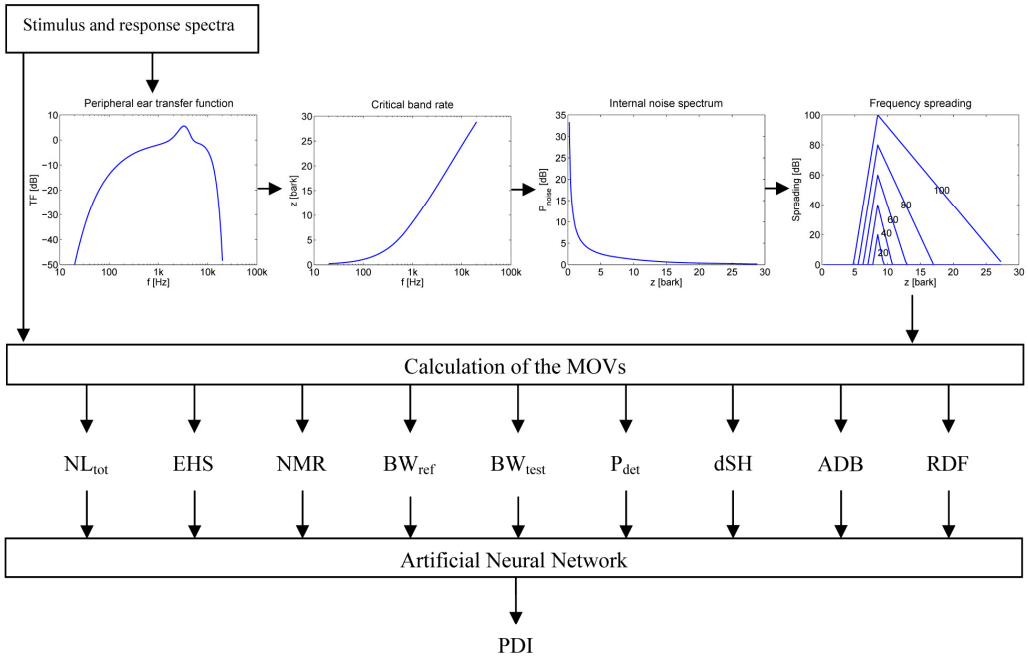

Figure 1 Block diagram of the PTHD algorithm

The resulting patterns after applying the effects of the psychoacoustic mechanisms are known as "excitation patterns". In the next step, the excitation patterns are used to calculate several variables that are referred to as Model Output Variables (MOVs). The PEAQ standard specifies 11 MOVs for its basic version. These MOVs are defined to measure different phenomena including changes in the modulation, loudness of the distortion, bandwidth, rate of occurrence of audible distortions, noise-to-mask ratio, probability of detection of the distortion and harmonic structure of the error. Since the test signals in our study are steady state sinusoids, we ignore those MOVs that describe temporal changes such as changes in the modulation. We also add a new MOV that describes the sharpness. The MOVs that are used in our study are as follows.

Total Noise Loudness (NL_{tot}) estimates the loudness of additive distortions in the presence of the masking reference signal. More details on the calculation of this MOV, which was also used in the CLEAR algorithm for Rub & Buzz detection, can be found in [1].

Error Harmonic Structure (EHS) is designed to detect extended harmonic structure of signals. As described in the CLEAR algorithm, cepstral analysis is used to calculate this MOV [1].

Noise to mask ratio (NMR) per critical band is the ratio of the noise level to the masking level of the reference signal. The noise or the error is the difference between the reference and the test signals.

$$\Theta[k] = 10\log_{10}\frac{P_{noise}[k]}{M[k]} \qquad (1)$$

where k spans the bandwidth in barks and P_{noise} and M represent the noise and masking levels respectively. The mean NMR over all critical bands is obtained as

$$NMR = 10\log_{10}\frac{1}{Z}\sum_{k=0}^{Z}\frac{P_{noise}[k]}{M[k]} \qquad (2)$$

In the PEAQ standard, a time average of the instantaneous values of the NMR is used. In the PTHD algorithm we ignore the time averaging because our signals are steady state sinusoids.

Bandwidth of the reference signal (BW_{ref}) is calculated as the frequency at which the level of the reference signal is 10 dB above the signal level at 21.6 kHz which is the practical end of the spectrum.

Bandwidth of the test signal (BW_{test}) is calculated similarly, except that the level difference is defined to be 5 dB for the test signal.

Probability of Detection (P_{det}) describes the probability that the existence of distortion is detected. The effective detection step size $s[k]$ is first calculated as an approximation to the just noticeable level difference as measured by Zwicker [6].

$$s[k] = \sum_{i=0}^{4} c_i \tilde{E}_{test}[k]^i + d_1\left(d_2 / \tilde{E}_{test}[k]\right)^g \qquad (3)$$

where $c_0 = -0.199, c_1 = 0.055, c_2 = -0.001, c_3 = 5\times10^{-6}$, $c_4 = 9\times10^{-11}$, $d_1 = 5.951$, $d_2 = 6.395$, $g = 1.713$, and $\tilde{E}_{test}$ is the excitation pattern of the test signal.

The error signal is defined as the difference between the excitation patterns of the reference and test signals

$$e[k] = \tilde{E}_{ref}[k] - \tilde{E}_{test}[k] \qquad (4)$$

The total number of steps above the threshold at frequency band k is

$$q[k] = \frac{|INT(e[k])|}{s[k]} \qquad (5)$$

where $INT(x)$ is the largest integer less than or equal to x. The total number of steps above the threshold over the entire bandwidth is

$$Q = \sum_{\forall k} q_c[k] \qquad (6)$$

This parameter will be used for the derivation of other MOVs that will be discussed later.

The probability of detection at each frequency band is defined as

$$p_{\text{det}}[k] = 1 - 2^{-(e[k]/s[k])^6} \qquad (7)$$

The total probability of detection over all bands is

$$P_{\text{det}} = 1 - \prod_{\forall k}(1 - p_{\text{det}}[k]) \qquad (8)$$

In the PEAQ standard, P_{det} is time filtered and a maximum of the filtered probability of detection over all time frames is introduced as a MOV. As explained

earlier, we ignore the temporal functions and consider the instantaneous P_{det} as a model output variable.

Difference in sharpness (dSH) calculates the difference between the sharpness of the reference and test signals. Sharpness is a sensation that occurs when the sound contains mainly high frequency components; hence it can be representative of high order harmonics. Sharpness can be given as [6]

$$S = 0.11 \frac{\int_0^{24} N g(z) z dz}{\int_0^{24} N dz} \tag{9}$$

where N is the total loudness of the signal, z is the critical band rate, and $g(z)$ is a polynomial weighting function that can be approximated as

$$\mathrm{g(z)} = \sum_{\mathrm{i}=0}^{4} g_i z^i \tag{10}$$

where $g_0 = 3.5$, $g_1 = -1.81$, $g_2 = 0.1$, $g_3 = -0.0056$ and $g_4 = 0.00012$.

Average Disturbed Block (ADB) is defined as

$$ADB = \begin{cases} 0 & P_{\text{det}} < 0.5 \\ \log_{10} Q & P_{\text{det}} \geq 0.5, Q > 0 \\ -0.5 & P_{\text{det}} \geq 0.5, Q = 0 \end{cases} \tag{11}$$

where P_{det} and Q are the probability of detection and the steps above threshold as defined earlier. This formula quantifies the loudness of distortion and it is a special case of the ADB variable defined in PEAQ when there is one frame only.

Relative Disturbed Frames (RDF) is a threshold function that takes a value 1 if

$$\max_{\forall k} \Theta[k] \geq 1.5 \text{ dB} \tag{12}$$

and a value 0 otherwise. It is a special case of the RDF variable in PEAQ when only one frame exists.

Once the MOVs are calculated, they are mapped to a variable measured through listening tests that is indicative of the Perceived Distortion Index (PDI). The mapping is obtained using an artificial neural network.

3. LISTENING TESTS

Two types of tests were designed for the measurement of harmonic distortion; (a) controlled two-tone test; and (b) headphone distortion measurement test. In the first test, the fundamental and the second or the third harmonics were generated at user-defined frequencies and levels. The two tones were then played through a professional headphone that was expected to add negligible distortion to the signal. Bose's QuietComfort15 headphones were used for this test because their distortion was very low and they perfectly eliminated background noises. The signal was recorded using Brüel and Kjær's Head and Torso Simulator (HATS) and its spectrum was saved. The test was performed using SoundCheck, Listen's audio test and measurement system. SoundCheck allows signals to be played and recorded at precise sound pressure levels, and this functionality enabled the headphone-ear simulator channel to be equalized, thereby achieving the desired sound pressure spectrum with high accuracy. A custom test sequence written with SoundCheck's sequence editor was used to perform the numerous tests. The sequence automatically saved the spectra corresponding to each test in a MATLAB-friendly format.

The recorded spectra were imported into MATLAB and the MOVs described in Section 2 were calculated for each test case. The test was designed to cover several cases that target different fundamental frequencies and levels, as well as harmonic orders and distortion levels. The test signals were also played for several subjects who ranked the signals according to their perceived distortion. The reported PDI values were used as the neural network's target outputs.

The second test was designed to measure the harmonic distortion produced by a popular brand of headphones. Pure tones at different frequencies and levels were used as test signals. Signal generation and recording and MOV calculation were performed as in the first type test.

The subjects were asked to rate the perceived distortion based on the scale described in Table 1. A grade PDI=1 was assigned to signals with very weak harmonic distortion that were perceived to be undistorted. At weak distortion levels, where the signal quality was not perceived to be as good, the signal was given a grade PDI=2. At medium harmonic distortion levels, the signal was usually perceived to be pure; but a change was perceived in the pitch and timbre of the signal compared to an undistorted tone. Such distortion was rated with a grade PDI=3. At higher distortion levels, the fundamental and the harmonic were clearly distinguishable (PDI=4). When severe high order harmonics were present, an annoying buzz sound was

perceived which was usually indicative of the existence of the Rub & Buzz phenomena (PDI=5).

Four subjects performed the tests and their grades to each test case were averaged to reduce the effect of noise in the data caused by human errors and the difference in the individuals' auditory perception.

Figure 2 Audio test set-up

PDI	Description
1	Good- no perceived distortion.
2	Not as good- e.g. the signal is not as clear.
3	Distortion perceived- e.g. in the form of pitch and/or timbre change.
4	Objectionable distortion- existence of multiple tones perceived.
5	Terrible distortion- e.g. Rub & Buzz.

Table 1 Reference PDI scale for audio tests

4. ARTIFICIAL NEURAL NETWORK

Artificial neural networks are powerful modeling tools that use optimization methods to learn the behavior of complex functions by adjusting a series of model parameters. A standard two-layer artificial neural network, shown in Figure 3, is used to link the perceived distortion grades obtained from audio tests to the objective MOV measurements. The error to be minimized is the difference between a given PDI value and the network's predicted output for a corresponding set of MOV values. The error is minimized by adjusting the weights $w_x[i, j]$ and $w_y[j]$.

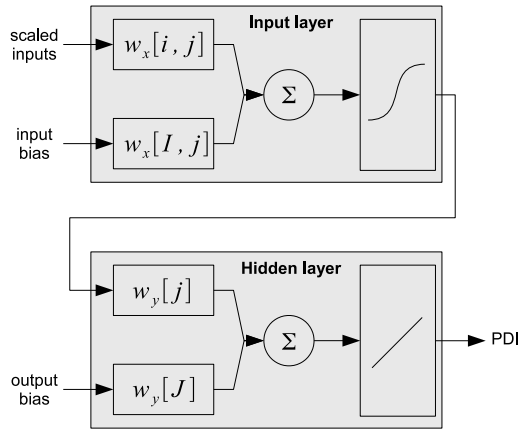

Figure 3 Architecture of the neural network

The standard neural network of Figure 3 is proven to be capable of approximating almost any function at arbitrary precision. The architecture of the network is as follows. The input layer consists of $I-1$ inputs and an input bias which are added using a weighted sum and transferred to the interval [0, 1] through a sigmoid activation function $sig(x) = 1/(1+e^{-x})$. The non-linear structure of the sigmoid function enables neural networks to model arbitrary non-linear functions. The sigmoid outputs are fed into a second layer which is known as the hidden layer. The hidden layer consists of $J-1$ hidden nodes and one output bias which are summed and transferred through a linear scaling function to generate the network output.

The architecture of the network is selected to match with the neural network of Eq. (13) used in the PEAQ standard. The network has two layers and three hidden nodes $J-1=3$. The number of the MOVs plus the input bias is $I=10$. The MOVs are denoted by $x[i]$, and $a_{\min}$ and $a_{\max}$ are scaling factors.

$$PDI = w_y[J] + \sum_{j=0}^{J-1} w_y[j] \cdot sig(z[j]) \qquad (13)$$

where

$$z[j] = w_x[I, j] + \sum_{i=0}^{I-1} w_x[i, j] \cdot \frac{x[i] - a_{\min}[i]}{a_{\max}[i] - a_{\min}[i]} \quad (14)$$

is the output of the first layer.

MATLAB's neural network toolbox is used for training the network. The toolbox provides powerful algorithms for developing various neural networks and graphically analyzing their training process and simulation results.

A sub-set of the data points is randomly selected for network training; and the remaining data are used for validation and testing of the network. Since the training data are randomly selected, the estimated weights of the neural network slightly vary each time the network training is completed. However, the weights corresponding to specific MOVs, namely the Noise Loudness, the Noise to Mask Ratio and the Error Harmonic Structure consistently obtain a relatively high overall gain.

5. EXPERIMENTAL RESULTS

In this section, we present the audio test results for a number of the audio tests. Table 2 shows the average PDIs assigned to signals played with a test headphone. Each value is obtained by averaging the PDIs given by different listeners.

f_1 [Hz] / L_1[dBSPL]	100	200	500	1k
75	1.5	1.25	1.25	1.25
85	2.25	1.5	1.25	1.5
95	4.5	2	1.5	2

Table 2 PDIs for a test headphone

Let us focus on the first column of the table which corresponds to signals of frequency 100 Hz. The spectra of the three signals ($L = 75$, 85, and 95 dBSPL) are shown in Figure 4 and several objective grades are presented in Table 3. The figure shows that for a fundamental level of 75 dB, the second and third harmonics are 14 dB and 36 dB below the fundamental, which correspond to a 20% THD. Despite the considerable THD level, the average PDI given to this signal is 1.5 which indicates that distortion is almost inaudible. At 85 dB fundamental level, the first 8 harmonics are present, producing 42% THD. The average PDI assigned to this case is 2.25, showing borderline distortion audibility. Finally, increasing the fundamental level to 95 dB generates hundreds of harmonics and 60% THD. This is an example of a terribly distorted spectrum which receives an average subjective grade of 4.5. That spectrum clearly indicates the presence of Rub & Buzz. Note that although THD has a "correct" trend versus PDI, it is not a precise figure of merit for the estimation of distortion audibility; e.g. high THD of 42% over low order harmonics is barely audible. This result implies that low order harmonics are not annoyingly audible unless they are at the threshold of generating Rub & Buzz and are accompanied by higher order harmonics.

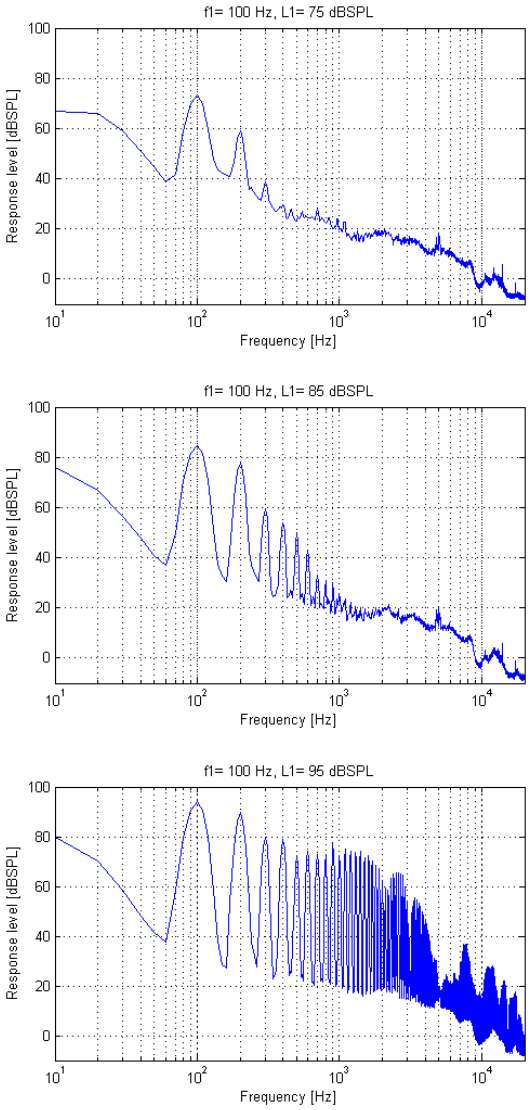

Figure 4 The spectra of good (top), borderline (middle) and bad (bottom) sounding signals

Objective grade / $L_{[dBSPL]}$	THD	NL_{tot}	NMR	EHS
75	19.70	76.30	26.06	12.00
85	41.53	86.14	26.45	14.67
95	60.11	108.20	60.49	79.00

Table 3 Objective grades for the good, borderline and bad sounding signals (f_1=100 Hz)

Three MOVs, namely Noise Loudness (NL_{tot}), Noise to Mask Ratio (NMR), and Error Harmonic Structure (EHS) out of the original 9 MOVs are presented in Table 3. These MOVs obtained the highest weights in the neural network which indicates their strong impact on the network's output. The THD is also presented as a reference. As the table shows, there is a noticeable increase in the values of the MOVs at the high level of 95 dBSPL which receives a high PDI. Let us particularly focus on the NMR which is the mean over all frequency bands of $\Theta[k]$ described by Equation (1). Figure 5 shows $\Theta[k]$ for the three signals. It is observed that its overall level is much higher at 95 dBSPL than in the other cases. The THD, however, increases linearly with the level of the fundamental and does not fairly represent the PDI.

In Figure 6, the Neural Network calculates the subjective rating versus frequency and shows excellent correlation with our listening tests. In particular, the bad plus borderline speakers show elevated NN output levels in the 100 – 1000Hz frequency range. In developing and refining the algorithm, the weighting of the various MOVs in the neural network was varied to give the best correlation to subjective listening tests. An interesting feature of this approach is that when implemented in a test system, these weighting factors may be user-determined. This would enable user to use their own subjective measurement data to weight the MOVs. In other words, loudspeaker and headphone manufacturers can decide the weighting of particular audible distortion defects as well as the limits for any particular product. Such a system could be trained by the end user to truly reflect a human listener.

The performance of the proposed method for predicting audible distortion is compared against the THD metric. Figure 7 shows the correlation between the objective and subjective grades for both methods. Shown in the figure is also the Mean Square Error (MSE) between the subjective grade Y and the predictor $\hat{Y}$:

$$MSE = \frac{1}{N}\sum_{1}^{N}\left|Y_n - \hat{Y}_n\right|^2 \qquad (15)$$

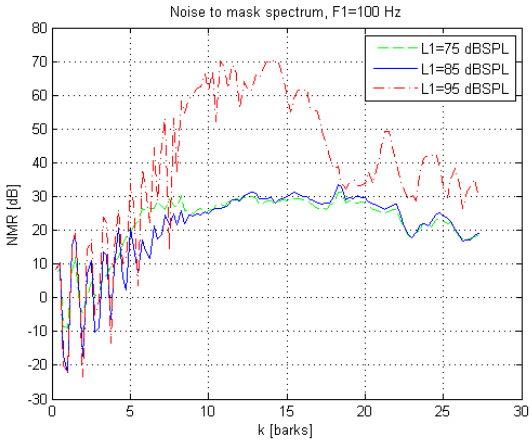

Figure 5 Noise to mask spectrum for the good, borderline and bad sounding signals

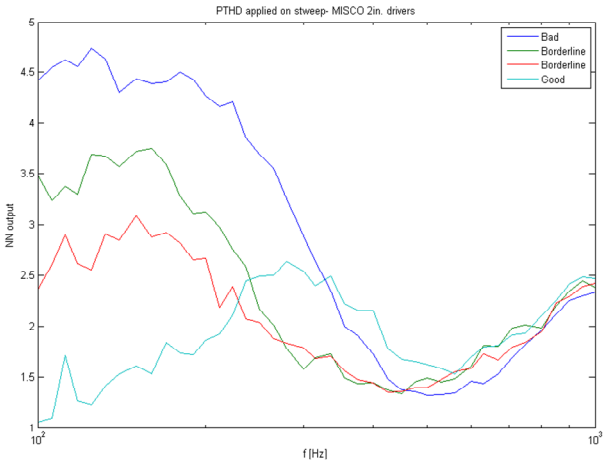

Figure 6 Neural Network(NN) output for Good, Borderline and Bad loudspeakers

For Neural Network the predictor is simply the output value: $\hat{Y} = X$, and the resulting MSE is 0.105. For THD, the value in % cannot be mapped directly to the subjective grades. A linear predictor has been derived from it: $\hat{Y} = \alpha X + 1$, with the constraint that 0% THD should correspond to a perfect grade of 1, and with α chosen has to minimize the MSE. The minimum MSE obtained is then 0.555, which is about five times bigger than the former MSE. This quantitative result is consistent with the visual inspection of the graphs.

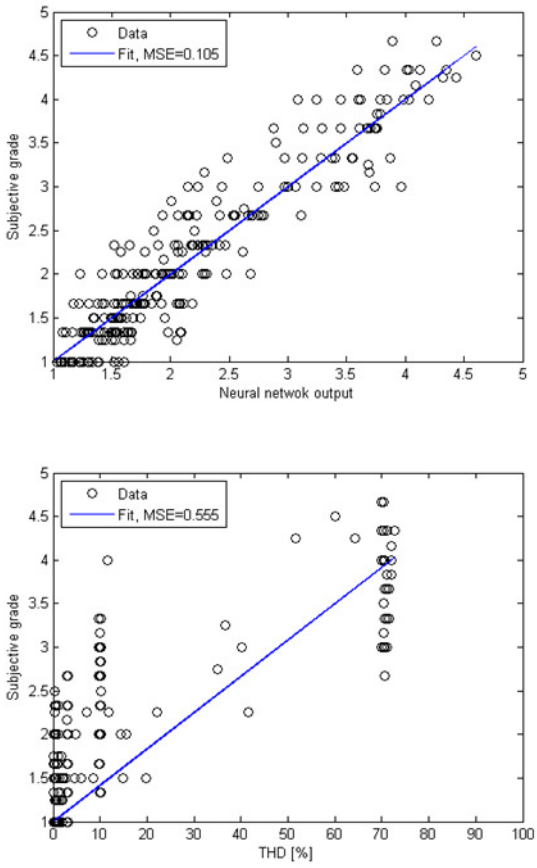

Figure 7 Goodness of fit of the proposed method (top) and the THD metric (bottom)

6. CONCLUSIONS AND FUTURE WORK

An objective method was proposed for measuring the perceptual harmonic distortion of audio systems. The method uses a framework similar to the PEAQ standard. The signal framing scheme, the stimulus types and the measured signal qualities (MOVs) used for training the model are specifically designed to measure the harmonic distortion audibility and are thus different from PEAQ. Subjective tests, so far, confirm the effectiveness of the proposed method.

The tests showed that even though significant levels of low order harmonic distortion in headphones, e.g. 42% THD, could be difficult to hear, they indicated that the transducer was at its physical limits and the onset of Rub & Buzz distortion which is both very audible and annoying sounding. At slightly higher levels, these headphones would rub and buzz.

Since PTHD compounds more psychoacoustic criteria than our previous perceptual Rub & Buzz algorithm, we believe that it presents a notable improvement and should prove to be more reliable to detect audible Rub & Buzz.

There are several planned extensions of this work:

- Perform further listening tests and research on other devices to investigate if most electro-acoustic devices and audio electronics follow the same trends.
- Allow end-users to come up with their own MOV weightings according to their application.
- Applying the perceptual method to non-harmonic distortions such as air leaks, loose particles, loose wires and etc.
- Using more complex stimuli such as two-tones, narrow band noise, and ultimately, real speech and music signals to measure audible distortion in any device.

7. REFERENCES

[1] S. Temme, P. Brunet, and D. B. (Don) Keele, “Practical Measurement of Loudspeaker Distortion Using a Simplified Auditory Perceptual Model,” Presented at the AES 127th Convention (October 2009), Paper 7905

[2] S. Temme, P. Brunet, and B. Fallon, “Practical Implementation of Perceptual Rub & Buzz Distortion and Experimental Results,” Presented at the AES 129th Convention (November 2010), Paper 8173

[3] ITU-R Recommendation BS.1387 (PEAQ)

[4] Temme, Steve, "Audio Distortion Measurements," Bruel & Kjaer Application Note, Bruel & Kjaer (May 1992)

[5] E. R. Geddes and L. W. Lee, “Auditory Perception of Nonlinear Distortion – Theory,” Presented at the AES 115th Convention, (October 2003), Paper 5890

[6] H. Fastl and E. Zwicker, “Psychoacoustics: Facts and Models,” 3rd Edition, Springer-Verlag (2007)