Audio Engineering Society
Convention Paper

Presented at the 129th Convention
2010 November 4–7 San Francisco, CA, USA

Practical Implementation of Perceptual Rub & Buzz Distortion and Experimental Results

Steve Temme[1], Pascal Brunet[2], and Brian Fallon[3]

[1]Listen, Inc., Boston, MA, 02118, USA
stemme@listeninc.com

[21]Listen, Inc., Boston, MA, 02118, USA
pbrunet@listeninc.com

[3 1]Listen, Inc., Boston, MA, 02118, USA
bfallon@listeninc.com

ABSTRACT

In a previous paper [1], we demonstrated how an auditory perceptual model based on an ITU standard can be used to detect audible Rub & Buzz in loudspeakers using a single tone stimulus. In this paper, we discuss a practical implementation using a stepped sine sweep stimulus and present detailed experimental results on loudspeakers including comparison to human listeners and other perceptual methods.

1. INTRODUCTION

Rub & Buzz in loudspeakers is very obvious and annoying to listeners as the buzzing sound superimposes on the audio source material. It is very noticeable because it is unrelated to the music or speech being played by the loudspeaker. The buzzing is triggered mostly by bass content that produces large voice coil excursions and it may appear as loud, or even louder, than the audio source material itself. Production line detection of audible Rub & Buzz defects is therefore important for the end-user's perception of quality from the manufacturer.

Several manufacturing issues can cause Rub & Buzz, including voice coil misalignment, poorly glued parts such as the driver's spider or surround, lead wires hitting the cone or the spider, and cone breakup at high frequencies.

The typical signature of Rub & Buzz in the time domain is a periodic transient superimposed on the sine excitation (figure 1).

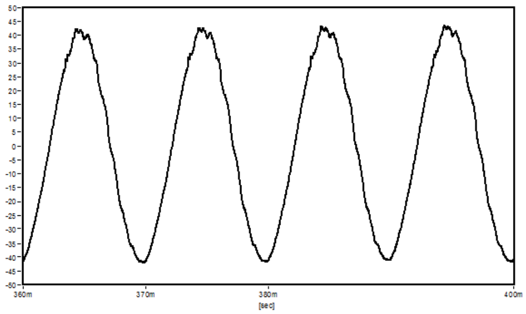

Figure 1: Typical Rub & Buzz at the positive peak in the time domain

In the frequency domain the typical Rub & Buzz signature is a very extended and slowly decreasing family of harmonics (figure 2).

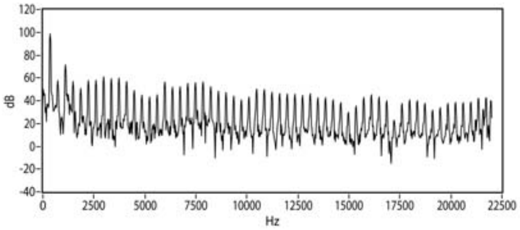

Figure 2: Rub & Buzz in the frequency domain

Many companies use trained listeners with acute hearing, often known as "Golden Ears", to detect Rub & Buzz defects on the production line. However, this manual approach has its drawbacks. There is an inherent variability among different "Golden Ears", and for each person, the detection ability depends on fatigue, focus and health condition. Additionally, there is no quantitative metric for the perceived Rub & Buzz loudness.

Objective methods based on temporal or spectral analysis offer a reliable alternative. However, the measurement algorithms that are widely used tend to identify all Rub & Buzz instances regardless of whether they are audible or not. Although these are reliable and valid techniques that offer excellent analysis of production line defects, the correlation to human perception can be poor. In the interests of increasing yields, many manufacturers now choose a model where they reject only *audible* defects. This has driven the demand for automated perceptual Rub & Buzz measurement.

In a previous paper [1], we demonstrated that an algorithm based on the principles of the ITU PEAQ standard was able, using a pure tone stimulus, to show a strong correlation between perceived loudness and a metric of distortion loudness combined with Error Harmonic Structure (EHS).

The distortion loudness encompasses all added noise and distortion components. In order to minimize the effect of background noise and only measure distortion, a measurement with a good signal to noise ratio is required. As distortion loudness is a measure of loudness, it is expressed in phons. For a given sound the loudness in phons is equal to the level in dBSPL of a 1000 Hz tone which is equally loud.

The EHS is a metric that measures the extent of the harmonic family due to nonlinear distortion. It is obtained by applying an FFT on the dB-spectrum. This technique is known as cepstral analysis and is popular in vibration analysis. EHS is simply the maximum magnitude of the Cepstrum. EHS efficiently detects the typical Rub & Buzz extended harmonic signature.

In this paper we explain how this principle has been refined to offer an accurate perceptual Rub & Buzz technique for production line use, with a stepped sine stimulus, easy limit setting and good immunity to background noise. We also demonstrate how this method results in a single absolute measurement of Rub & Buzz loudness that can easily be correlated to human audibility. Finally, we share test results and discuss future development potential.

2. THE CLEAR ALGORITHM FOR RUB & BUZZ ANALYSIS AND ITS PRACTICAL IMPLEMENTATION

The principles of psychoacoustics and how the ITU PEAQ model relies on these, is well documented in [1]. Here we therefore focus on the practical implementation.

For production line applications, there are 3 important factors in addition to accuracy: speed, noise immunity and ease of use. Most production lines use a stepped sine wave stimulus as this is used for many other end of line tests (e.g. frequency response, impedance, THD etc,) It therefore makes sense for perceptual Rub & Buzz to use the same test signal so that it can be carried out in parallel rather than sequentially to other tests. Noise immunity is also important as production lines are subject to both steady state and transient background noise. Ease of use is also critical. It is notoriously hard to set valid limits with traditional Rub & Buzz measurement methods, and considerable operator expertise is required. The CLEAR algorithm has been designed to display the Rub & Buzz measurement in absolute terms that can be easily correlated to human listening. This makes limit setting extremely simple as the limit can be set as a flat line across the entire frequency spectrum based on the phon level that is deemed to be unacceptable. This limit is absolute and not relative to a reference.

Figure 3 shows a basic flow chart of the CLEAR algorithm, and the steps are explained in more detail below.

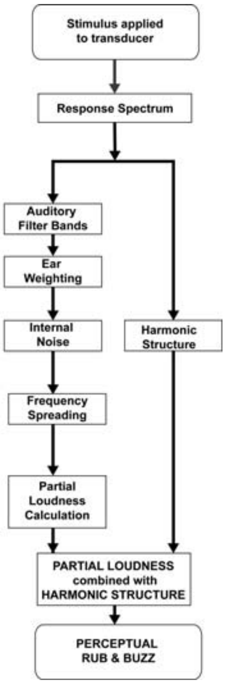

Figure 3: Basic Flow Chart of the CLEAR Algorithm

As the flow chart shows, the stimulus, a stepped sine wave (the test signal of choice for most production line loudspeaker measurements) is played and the response captured. The response signal then has two separate analyses performed, the results of which are combined at the end.

On one side, first, auditory filter bands are applied to the response signal to convert the FFT spectrum (constant bandwidth) to a Bark scale (auditory filter bands) (figure 4). This replicates the way the ear filters sound.

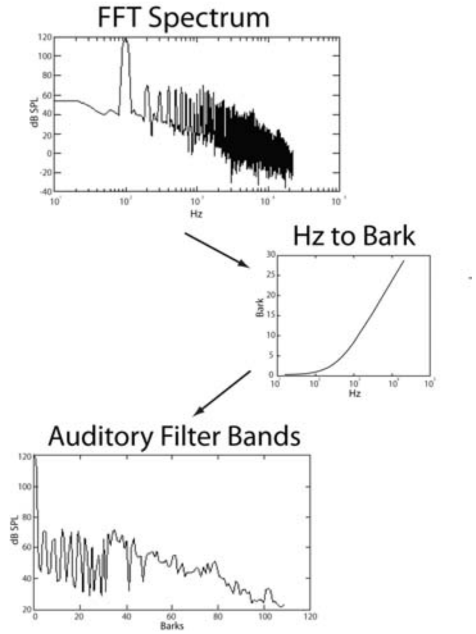

Figure 4: Auditory Filter Bands

Next, an ear weighting filter (figure 5) compensates for the transfer function of the outer to inner ear, and the internal noise of the ear (noise floor due to blood flow) is added. Together these model the frequency response of the ear.

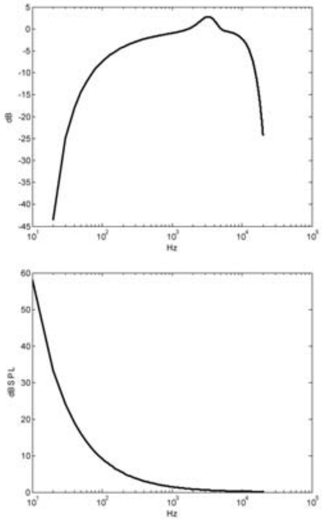

Figure 5: Ear Weighting Filters and Internal Noise

A frequency spreading function is then applied (figure 6). This is a simplified mathematical representation of auditory masking curves and this is how the algorithm mimics the psychoacoustic filters of the ear in hearing Rub & Buzz defects. These masking curves change with frequency and level.

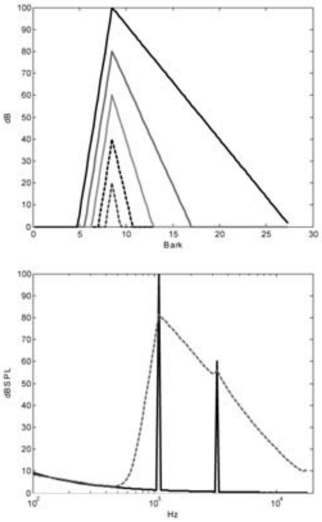

Figure 6: Frequency Spreading Functions

Finally, the fundamental and its masking effects are subtracted out from the result for the response signal to give the distortion of the speaker plus noise (figure 7). This is summed over the frequency range to give the perceptual partial loudness (in phons) for a single tone of the input signal.

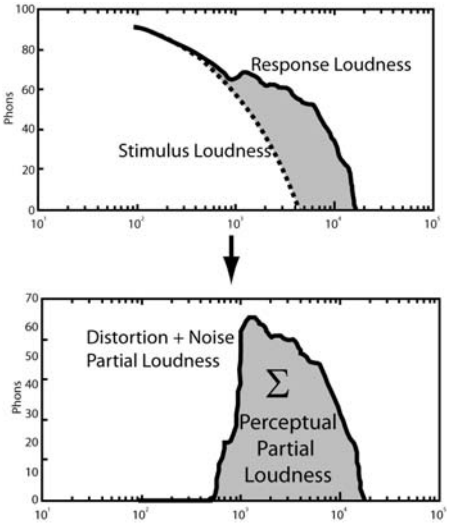

Figure 7: Perceptual Partial Loudness

On the other side of the flow chart, you can see that a harmonic structure analysis is carried out (figure 8). The harmonic structure of the response is quantified using the power cepstrum (a cepstrum is a spectrum of a log spectrum). A strong and extended harmonic structure is a signature of Rub & Buzz. It is this cepstral analysis from which the name CLEAR (Cepstral Loudness Enhanced Algorithm for Rub & Buzz detection) is derived.

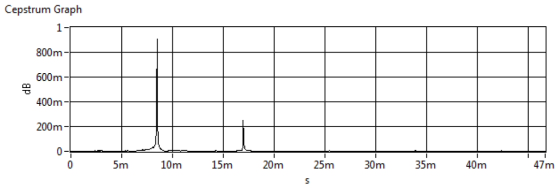

Figure 8: Cepstral Analysis of response showing strong harmonic series

In the final step, the result of the harmonic analysis (a percentage measurement) is combined with the perceptual distortion for each frequency. This accentuates the Rub & Buzz, making it easier to identify and set limits. The name 'CLEAR' (Cepstral Loudness Enhanced Algorithm for Rub & Buzz) that has been assigned to this algorithm derives from the harmonic structure analysis.

More details on the specifics of these steps, and the equations used are explained in [1].

The above description outlines the steps for any given frequency. At the end of a production line where loudspeakers or drivers are inspected, it is necessary to test the full frequency range of operation in search of defects. A sine sweep is the usual approach and is multipurpose; with one sweep, a full range of loudspeaker characteristics including frequency response, polarity, THD, loose particles, and impedance are checked.

For practical purposes, the CLEAR algorithm is integrated with Listen's flagship audio test product, SoundCheck. SoundCheck uses a flexible stepped sine, where different frequency series, number of cycles and duration per step can be setup independently, and combined. After playing the stimulus and recording the waveform from the measurement microphone, the proprietary HarmonicTrak algorithm aligns the acoustic response waveform to the stimulus with 1-sample accuracy and makes a windowed FFT of each sine step. Knowing the stimulus frequency for each sine step, the algorithm retrieves the response fundamental and harmonics levels. Because the spectrum is available for each sine step, it is easy to incorporate the CLEAR algorithm above into the overall HarmonicTrak analysis. The end result is a perceptual Rub & Buzz (PRB) value for each stimulus step resulting in a curve of PRB versus stimulus frequency (figure 9). A good loudspeaker will exhibit a relatively flat curve (phons versus frequency). Conversely, a bad unit will show high values at certain frequencies, usually around resonance. Pass/Fail limits can be easily set simply by specifying a maximum PRB loudness level in phons, above which a speaker will be rejected.

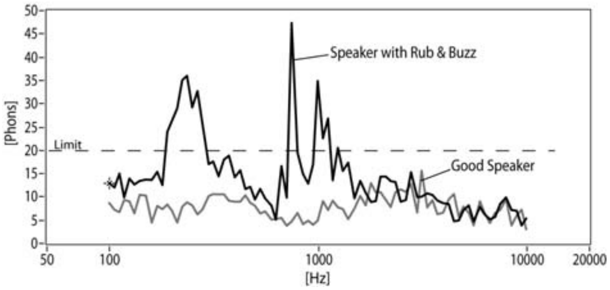

Figure 9: PRB values vs. stimulus frequencies. This speaker shows Rub & Buzz around 225, 750 and 1000 Hz.

3. EXPERIMENTAL PROCESS AND RESULTS

In order to prove the practical validity of the new Perceptual Rub & Buzz algorithm (CLEAR), we performed a series of experiments. These compared the new test method to both subjective listening tests as well as to other automated Rub & Buzz detection methods in a simulated production setting with noise.

For comparison to conventional Rub & Buzz measurement methods, we used the Normalized Rub & Buzz [3] option in SoundCheck, which was the first automated Rub & Buzz detection algorithm to be implemented in a commercial test platform. This algorithm essentially power sums selected high order harmonics using a proprietary FFT based algorithm equivalent to a parallel bank of tracking filters, and subtracts out the frequency response of the speakers. To compare the new algorithm to alternative perceptual methods we used a well known competitive perceptual distortion measurement product.

The tests were conducted on a sample batch of fifty 2" loudspeakers. The units had been pre-sorted by the manufacturer and included both good and bad sounding speakers with an array of distortion (Rub & Buzz) defects. They were all given serial numbers so that we could track the data.

In the interest of maintaining impartiality, we enlisted a third party, to both conduct the evaluations and deliver the data.

The first round of the testing was a subjective listening evaluation. Ten different people listened to all 50 speakers and rated them on a scale of 1-5 (1 being best) in terms of audible Rub & Buzz distortion. The listeners were provided with a guideline as to how the speakers should be rated. A reference unit that had no audible buzzing defect was chosen and played for the listeners as an example of a '1'. A frequency log sweep was used as the source signal with a range of 100 Hz to 2 kHz. Listeners were also given the option of using a manual oscillator to center in on a particular frequency.

The listening scores were averaged across the ten listeners, generating an overall listening score for each individual speaker.

The next step was to objectively measure each of the speakers using the three previously mentioned algorithms (CLEAR, SoundCheck's Normalized Rub & Buzz and a competitive perceptual Rub & Buzz system). The speakers were tested in the near-field with a ¼" reference microphone. A frequency range of 100 Hz to 10 kHz was used with a test level that generated roughly 100 dBSPL at 1 kHz for all measurements.

The Perceptual Rub & Buzz and the Normalized Rub & Buzz were measured simultaneously in SoundCheck from a single stepped sine sweep. Two curves of distortion (one in phons for PRB and one in % for Rub & Buzz) versus frequency were stored.

The speakers were then tested on the competitive system. This algorithm produces 'Steepness' curves for 6 different bands. These bands correspond to filters whose center frequencies are approximately 2500, 3500, 4800, 7000, 10000 and 14000 Hz. These 6 curves of Steepness (Pa/s) versus frequency were stored for each unit.

The method for measuring frequency response and distortion on the competitive system is somewhat different than that of SoundCheck. Instead of measuring all parameters in a single sweep, this system first requires a stepped sine sweep for measuring frequency response, and then a frequency log sweep for measuring distortion. The result of the dual stimulus is that the total test time was 2 seconds longer than the SoundCheck test, even when using the same stimulus frequency range.

3.1. Measured distortion versus listening results

Since our goal was to correlate objective data against the subjective listening scores, we needed to calculate a single numerical distortion value for each of the three algorithms, rather than using the 'pass/fail compared to a golden unit' metric that would be used in practice. For the Perceptual Rub & Buzz and the Normalized Rub & Buzz we simply extracted the maximum value versus stimulus frequency. This was representative of the highest level of distortion measured for each speaker. The competitive perceptual distortion data, however, was more complicated, as there were 6 curves for each unit to factor in. As the low frequencies demonstrated the largest variation between good and bad units, we calculated the maximum value of each band below 2 kHz and then chose the maximum of these 6 values.

The following graphs show the data from each of the three algorithms, compared to the listening scores. The trend lines (best linear fit) are shown with their correlation coefficient R^2.

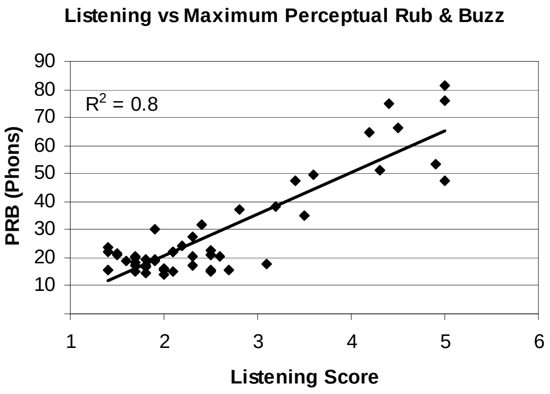

Figure 3: The correlation between Perceptual Rub & Buzz and listener score (a listener score of 1 represents a good-sounding speaker, and 5 a bad-sounding speaker)

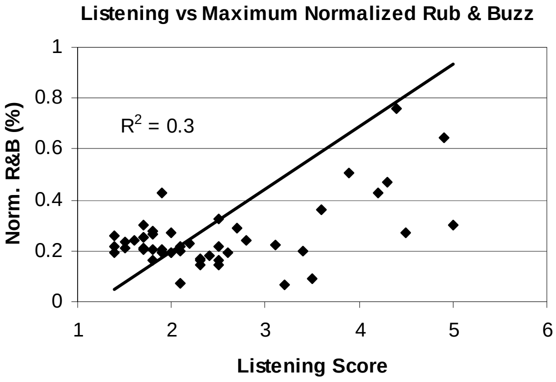

Figure 4: The correlation between Normalized Rub & Buzz and listener score (a listener score of 1 represents a good-sounding speaker, and 5 a bad-sounding speaker)

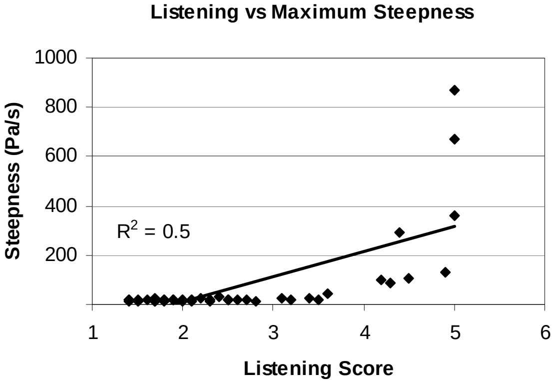

Figure 5: The correlation between maximum steepness and listener score (a listener score of 1 represents a good-sounding speaker, and 5 a bad-sounding speaker)

The Perceptual Rub & Buzz (CLEAR) algorithm in SoundCheck produced the best correlation to the listening scores. For the particular speakers under test, units that scored well in listening exhibited low distortion values in the 15 – 25 phons range. Speakers that were considered borderline by the listeners measured around 25 – 40 phons. Speakers that were rated bad in the listening test typically measured 40 – 75 phons depending on severity of the buzzing issue. This linear correlation proves that the hearing models used in the PRB algorithm provide a strong relation to human perception of distortion. The ability to discern not only good from bad, but also borderline units, can be extremely beneficial both for engineers and in a production setting. Naturally the levels that are deemed good, bad and borderline will depend on the speakers themselves, and the manufacturer's own tolerance for distortion.

Normalized Rub & Buzz did not correlate to the listening scores as well as the Perceptual Rub & Buzz algorithm. Several units that were considered good in the listening test measured higher with normalized rub & buzz and would likely fail a production limit. This is not a flaw in the conventional Rub & Buzz measurement methodology, but rather because Normalized Rub & Buzz is designed to measure all instances of Rub & Buzz, whether audible or not as they can be indicative of production line defects. Since the CLEAR algorithm is designed to measure only *audible* faults, it should be expected that a certain number of units with definite but inaudible faults would fail Normalized Rub & Buzz but pass Perceptual Rub & Buzz.

The competitive perceptual algorithm's data produced a more binary result than the other two analysis methods. The good units measured very low in steepness, and the bad units measured very high, but there seemed to be very little in between. This makes it challenging to determine if a unit is a borderline failure and also provides little indication how good or how bad a unit is. For example, units that scored between 3 and 4 in the listening test, which exhibit rising distortion values in the PRB algorithm, did not measure much higher in steepness than the good sounding units.

The data shows that setting limits for a production test is much easier with the CLEAR Perceptual Rub & Buzz algorithm than the other two methods. For a given product, the engineer simply needs to decide what level of distortion is audible, and therefore unacceptable, test the unit to measure the distortion in phons, and then set the limit accordingly. The limit will typically be a flat line, a consistent value across the frequency band. This simplicity provides confidence that if the distortion rises above the perceptual limit at any frequency, it will be audible.

3.2. Noise Immunity Test Procedure

Rub & Buzz testing is most often performed in a production environment and is typically subjected to a moderate level of background noise. It is therefore important to evaluate the three algorithms' ability to deliver usable and accurate distortion data in a noisy environment.

Typical factory noise can be broken down into two basic components: steady state noise (e.g. air conditioning, conveyor noise), which is constant and tends to be dominated by low frequencies, and transient noise (e.g. stamping, air compressors, speech), which is random, and can include many different frequencies.

To simulate these noise conditions, artificial signals were played through a separate speaker while performing the same measurements described earlier. For steady state noise a pink noise stimulus with an equalization applied to heavily weight the low frequencies was used. For the transient noise, a series of tightly band-limited noise bursts, centered on varying frequencies was generated. These bursts were spaced in time to occur randomly during the test.

The noise signals were each calibrated to produce 72 dBSPL at the measurement microphone. This level represents a particularly noisy, but not uncommon, factory environment and simulates a test condition where acoustic isolation has not been implemented.

Each speaker was tested three times. The first trial was performed with no noise present to serve as a baseline. The second had steady state noise playing during the test, and the third was subject to transient noise. As before, the SoundCheck system was used to measure both Perceptual Rub & Buzz as well as Normalized Rub & Buzz, and the measurements were also made on the competitive system. In each case, the data was logged separately for each of the three trials.

3.3. Noise Immunity Results

In order to quantify the relative change in measured distortion when environmental noise was present, the steady state noise results were compared to the baseline results with no noise, and the amount by which the distortion changed was calculated on a percentage basis. The calculation was performed for each frequency point in the distortion curve. This was repeated for the data with transient noise present.

The next step was to average the data of all 50 speakers to produce two curves for each algorithm: percentage error (steady state noise) versus frequency and percentage error (transient noise) versus frequency (figures 14 and 15). It should be noted that for the competitive system we averaged the data across the 6 distortion bands.

Below are the graphs comparing the three algorithms' ability to measure in the presence of high noise levels.

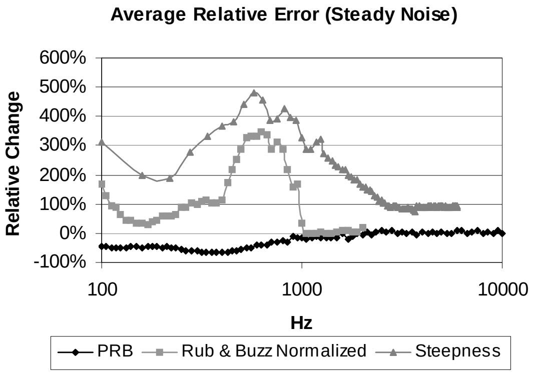

Figure 6: The percentage by which the distortion measurement changed when steady state background noise was present

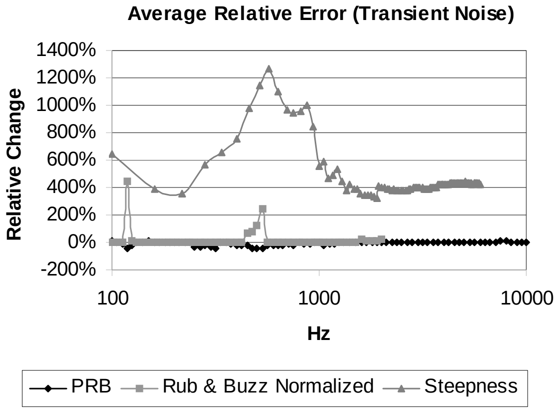

Figure 7: The percentage by which the distortion measurement changed when transient background noise was present

SoundCheck's CLEAR PRB algorithm fared best in this comparison, showing the greatest immunity to both steady state and transient noise. The measured low frequency distortion does decrease slightly in the presence of steady state noise. This is due to the noise masking some of the offending harmonics and thus making it harder to distinguish noise from distortion. The algorithm shows remarkable immunity to transient noise, which is often a big problem for traditional Rub & Buzz test methods.

The PRB algorithm performed better than the Traditional Rub & Buzz algorithm under conditions with both steady state and transient noise because the presence of additional noise in the measurement causes the level of upper order harmonics to appear artificially higher, and thus raises the measured distortion. Transients tend to present a more significant problem if they consist of higher frequencies and occur during the low frequency part of the sweep.

The competitive perceptual algorithm performed the worst in the noise immunity comparison, particularly with transient noise. The distortion data varied widely when background noise was introduced, making it extremely difficult to discern good units from bad. The steepness, which is the derivative of the acoustic pressure, is by nature a noise sensitive quantity.

While all of the measurement methods would benefit from noise isolation in this simulation, the PRB algorithm would require the least, and thus would provide the most accurate data and the greatest degree of flexibility in terms of placement on a factory floor.

4. CONCLUSION AND FUTURE DEVELOPMENTS

The experimental results prove that the CLEAR Perceptual Rub & Buzz measurements correlate better to a listener's experience than both simple high-order harmonic based Rub & Buzz measurements and the competitive perceptual distortion technique. This strong correlation means that users are able to easily set limits for the PRB algorithm with confidence that it will reject units with audible rub & buzz defects.

The data also proves that under noisy conditions the perceptual Rub & Buzz algorithm correlates better to result obtained under ideal conditions, making it an ideal choice for demanding production environments.

These encouraging results show that the perceptual approach is valid. This approach could be extended in the future to other kinds of test signals and distortion types:

- Use of a two tone test signal to perceptually measure intermodulation distortion for the detection of Rub & Buzz in tweeters and horns.

- Use of narrow-band noise as a test signal to measure Rub & Buzz distortion in telephony and other signal processing-intensive applications.

- Use of continuous Log Sweep test signal (e.g. Farina sweep) for faster tests.

5. 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), Preprint 7905

[2] Recommendation ITU-R BS.1387-1 "Method for objective measurements of perceived audio quality", (1998-2001)

[3] S. Temme, "How to Graph Distortion Measurements," Presented at the AES 94th Convention (March 1993), Preprint 3602