Audio Engineering Society

Convention Paper

Presented at the 127th Convention
2009 October 9–12 New York, NY, USA

Practical Measurement of Loudspeaker Distortion Using a Simplified Auditory Perceptual Model

Steve Temme [1], Pascal Brunet [2], and D. B. (Don) Keele, Jr. [3]

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

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

[3] DBK Associates and Labs, Bloomington, IN 47408, USA
dkeelejr@comcast.net

ABSTRACT

Manufacturing defects in loudspeaker production can often be identified by an increase in Rub & Buzz distortion. This type of distortion is quite noticeable because it contributes an edgy sound to the reproduction and is annoying because it often sounds separate or disembodied from the fundamental signal. The annoyance of Rub & Buzz distortion is tied intimately to human perception of sound and psychoacoustics. To properly implement automated production-line testing of loudspeaker Rub & Buzz defects, one has to model or imitate the hearing process using a sufficiently accurate perceptual model. This paper describes the results of a Rub & Buzz detection system using a simplified perceptual model based on human masking thresholds that yields excellent results.

1. INTRODUCTION

Methods to detect Rub & Buzz and other related defects in loudspeakers have been a hot topic for many years in the loudspeaker industry [1-17]. This is a testament to the difficulty of testing for these types of problems. These types of defects often do not cause major failures of the loudspeaker but may be very irritating to the person listening to it. The challenge is to detect the defects that are simply irritating psychoacoustically in addition to those which subsequently cause later in-use operational failures.

This paper discusses a practical distortion measurement method for loudspeaker Rub & Buzz defect identification based on a simplified auditory perceptual model. The simplified model is based on human masking thresholds [18-32] and follows some of the guidance in the recent ITU standard for objective measurement of perceived audio quality called PEAQ [33-37].

First, some of the typical loudspeaker manufacturing defects and the effects of these are outlined. Next, psychoacoustics and non-linear distortion perception including the new ITU standard recommendation for objectively measuring audio quality of a transmission channel (PEAQ) are explained. We then describe the use of masking curves and the use of PEAQ for measuring the audibility of harmonic and Rub & Buzz distortions. Finally, we demonstrate experimental results and describe possible future developments.

2. LOUDSPEAKER MANUFACTURING DEFECTS AND THEIR EFFECTS

Loudspeaker production and fabrication is plagued by many types of manufacturing defects that affect sound quality. Primarily, these include structural and mechanical faults which cause audible distortion effects in the acoustic output of the loudspeaker.

There are several measured types of distortion including harmonic, intermodulation, and added noise. Commonly, high-order harmonic distortions are grouped under the descriptive term of “Rub & Buzz”. These distortions are created in the loudspeaker by various mechanical defects such as the voice coil rubbing the magnet, the cone touching connection wires, etc… and are outlined in the next section.

Detecting Rub & Buzz is a critical production line measurement used to decide whether a speaker passes or fails QC inspection. Unfortunately, this added distortion must be judged in the light of how it is perceived by the listener which solidly enters the problem into the area of psychoacoustics.

The following sections describe the various mechanical defects and their resultant distortions and their effects.

2.1. Mechanical Defects:

The various mechanical defects that may occur in loudspeaker production include:

2.1.1. Voice coil misalignment (rub)

Various alignment problems and asymmetries may cause the voice coil to not be centered in the gap and therefore contact the magnet assembly. When the voice coil moves it therefore generates high-order harmonics.

2.1.2. Glue interfaces (buzz) e.g. spider and surround

Various parts of the loudspeaker are often attached with adhesive in the production process and may become detached due to problems with the adhesive itself or in its application. This may cause acoustical problems when the unattached parts flap or rub against each other.

The glue interfaces may include: surround to frame, cone to surround, cone to voice coil, voice coil to spider, and spider to magnet/frame, among others.

This will also causes higher order harmonics.

2.1.3. Lead wires hitting the cone or spider

The loudspeaker’s lead-in or litz wires transport the electric current to the voice coil from the speaker’s input terminals. Because one end of the wire is stationary and the other is moving, the unsupported central portion of the wire is prone to hitting other objects in the loudspeaker’s assembly under high excursion and therefore may generate high-order harmonics.

2.1.4. Mechanical clipping, e.g. voice coil hitting the backplate

Under high excursion, the voice coil assembly may hit the backplate, generating loud impact and impulsive sounds.

2.1.5. Loose particles

During the manufacturing process, loose particles may become trapped in the loudspeaker, resulting in a distinctive defect that is easily heard but difficult to measure [9-10]. This defect does not generate harmonics because of its random nature, and should not be confused with Rub & Buzz.

2.1.6. Defective cone and spider parts

Sometimes the loudspeaker is assembled with defective or incorrect moving parts such as the cone and spider. These problems may not generate typical rub and buzz sounds but may increase the low-order harmonic distortion.

2.2. Harmonic and Intermodulation Distortion Effects:

Most of the defects listed in the previous section will increase the harmonic distortion of the speaker’s radiated sound when a single narrow-band signal is applied to the speaker.

The signature of the resulting harmonic distortion depends on the details of the defect. Some generate low-order harmonics in the range of the 2nd to the 5th and others generate high-order harmonics that may extend up to the 50th or higher. Identifying which harmonics are associated with which defects, and establishing distortion thresholds and their signatures is very difficult.

Some of the defects may also increase so-called intermodulation distortion when two signals are applied simultaneously to the loudspeaker in different frequency ranges.

2.3. Typical Rub & Buzz Distortion Signatures:

The following graphs illustrate some of the effects of typical Rub & Buzz distortion in the time, frequency, and joint time-frequency domains.

2.3.1. In the time domain:

Figure 1 shows the waveform of a driver exhibiting typical Rub & Buzz defects when driven by a sine wave. Note the waveform distortion on the top and bottom of each sine wave cycle. Although the waveform deformation does not seem to be visually significant, the resultant distortion is quite audible.

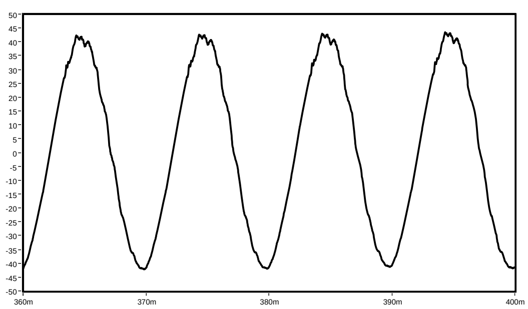

Figure 1: Acoustic output waveform of a loudspeaker exhibiting high levels of Rub & Buzz distortion when driven by a sine wave. Note the waveform anomalies on the top and bottom of each cycle of the sine wave.

2.3.2. In the frequency domain:

The Figure 2 & Figure 3 show the frequency spectrums of the acoustic output waveform generated by two loudspeakers: a "good" loudspeaker with no Rub & Buzz problems and a "bad" loudspeaker which exhibits high Rub & Buzz distortion. The spectrum of the "bad" loudspeaker exhibits high levels of high-order harmonics. Note that although the "good" loudspeaker has low Rub & Buzz distortion, its low-order total harmonic distortion (THD) is actually higher than the "bad" loudspeaker!

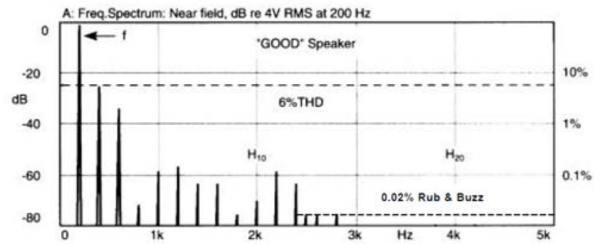

Figure 2: Waveform spectrum of the output of a "good" loudspeaker exhibiting low Rub & Buzz distortion. THD = 6 %, Rub & Buzz distortion = 0.02%.

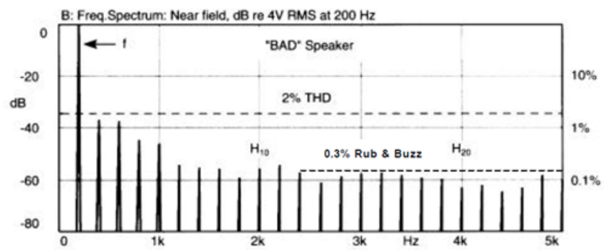

Figure 3: Waveform spectrum of the output of a "bad" loudspeaker exhibiting high Rub & Buzz distortion. Note the high level of high-order harmonics. THD = 2 %, Rub & Buzz distortion = 0.3 %. This driver exhibits a buzz due to a rubbing voice coil as a result of a bent frame.

2.3.3. In the time-frequency domain

The next graph shows a joint time-frequency map of a driver with high Rub & Buzz distortion.

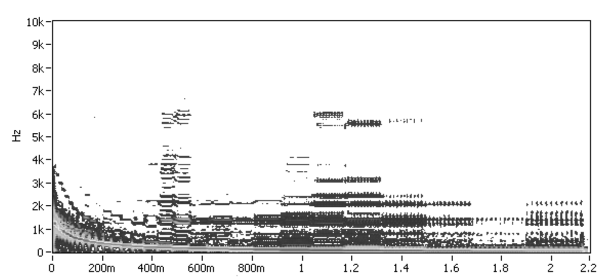

Figure 4: Joint time-frequency map of a driver with high Rub & Buzz distortion.

In this graph, the Rub & Buzz distortion can be seen as clusters of harmonics or horizontal bands "smeared" along the time axis and is recognized as a periodic disturbance.

3. PSYCHOACOUSTICS AND PEAQ

3.1. Psychoacoustic and Audio Engineering

Psychoacoustics is a science at the junction of psychology, physiology and acoustics. It is defined as "The study of the interaction of the auditory system and acoustics." (see Glossary).

Why use psychoacoustics? Every audio engineer knows that a human ear is very different from a microphone plus spectrum analyzer, yet for a long time there was little overlap between the audio industry and psycho-acoustics. It has been mostly limited to the use of dBSPL, fractional octave analysis and A-weighting curve.

The current "digital age" has brought psychoacoustics to the forefront of audio R&D. It is particularly prevalent in the areas of optimizing data transfer rates for telecommunications and for audio compression and storage.

In 1990 a demonstration by Johnston and Brandenburg at AT&T Bell Labs made a powerful impact. A noise with a specific spectral distribution was added to a music signal. Even though the SNR was only 13 dB the noise was not audible. That demonstration was later referred to as the "13 dB Miracle". Dr Karlheinz Brandenburg then continued to work on the application of psychoacoustics to music compression at the Fraunhofer Institute and became one of the main authors of MPEG.

MP3, AAC and other formats are now everyday examples of the power of applied psychoacoustics: 90% of the musical signal is thrown away without perceptible difference, and distortion can exceed 20% without annoyance.

Nowadays the telecommunications, music and broadcasting industries use perceptual auditory models to store and transmit music and voice with great success. However the loudspeaker industry has been slow to adopt psychoacoustic theory.

3.2. Current State of Psychoacoustic Theory

Following is a brief overview of the current state of knowledge in psychoacoustics.

First, from the outer to the inner ear the sound is attenuated by a band pass filter transfer characteristic [Figure 5].

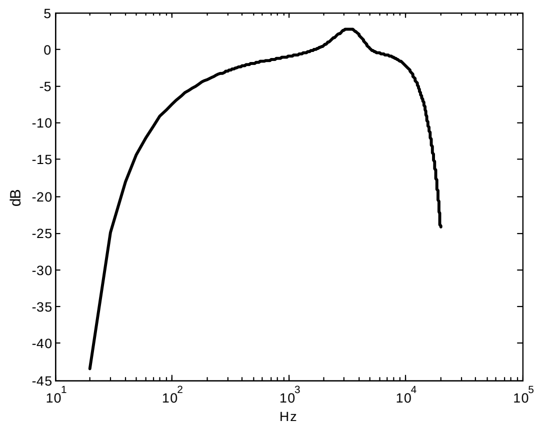

Figure 5: Outer and Middle Ear Frequency Characteristic

As the sound progresses inside the ear, noise caused by blood flow is added. This noise is greatest at low frequency [Figure 6].

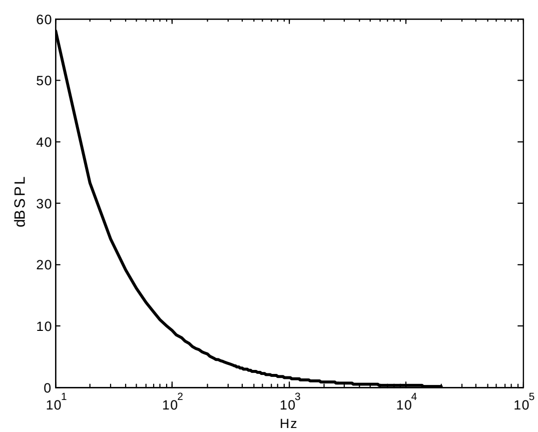

Figure 6: Internal Noise Spectrum

The combination of the transfer function and the inner noise plus some other minor effects gives the absolute threshold of hearing [Figure 7].

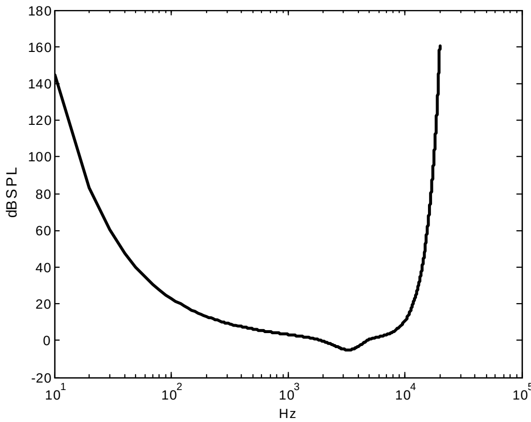

Figure 7: Absolute Threshold of Hearing

In the inner ear, the cochlea contains a series of hair cell receptors. They cover the length of the cochlea and they are divided in subgroups, each group specialized in a frequency band. These frequency bands make up a scale of 24 non-overlapping bands [29] and constitute the Bark scale [Figure 8].

The position along the scale corresponds to the pitch.

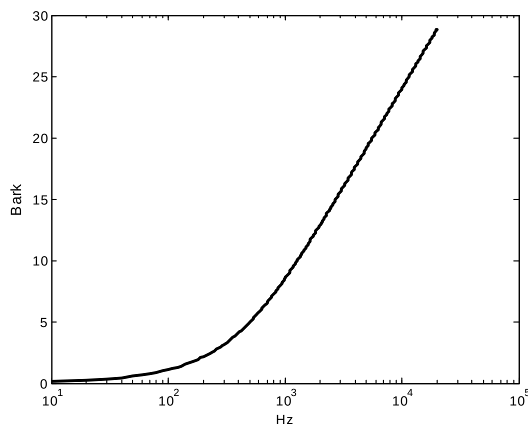

Figure 8: Bark scale vs. Frequency

Each hair cell acts as a non-linear band-pass filter. Its characteristic is triangular in shape. The shapes are nearly constant along the bark scale. The lower slope is +27dB/Bark and the upper slope varies with the sound level from -5 to -30dB. Because the filter characteristics overlap, a pure tone will excite a range of hair cells and the louder the sound the wider the range. When several components are present the different corresponding excitations add up in a non-linear way. Then the frequency content is smeared along the pitch scale.

The net effect of the smearing is that the threshold of hearing is raised for frequencies above and below a tone (the masker). Therefore a weaker adjacent tone (the maskee) can become inaudible if it is too close to the masker. E.g. in Figure 9, the tone at about 3 kHz is inaudible due to the stronger tone at about 1 kHz. The excitation pattern is shown overlaid over the two tones and exhibits only a small perturbation at the location of the maskee.

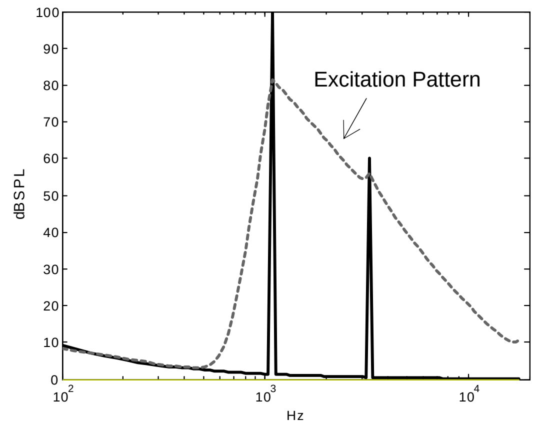

Figure 9: Frequency spreading applied to two tones. The resulting excitation pattern is the wide curve superimposed on the two discrete tones.

In addition to the frequency smearing there is also a temporal smearing. Hair cells have a time constant and need some time to adjust to sound level change. There is also a reaction delay. The time constant and reaction delay depend on frequency, level and even sound duration. In an analogous way to simultaneous masking, the temporal smearing causes the threshold of hearing to come back gradually to the absolute threshold after the sound stops. The recovery time depends on the masker level, frequency and duration, and it can go up to 150ms. Strangely, there is also a pre-masking time that can go up to 5ms.

The masking threshold is a fuzzy transition, not a hard limit. The masking threshold corresponds to a 50% chance of detection by an average person.

Because of the smearing non-linear characteristic the loudness of a sound is not equal to the sum of the loudness of its components. The loudness of a weak sound partially masked by a strong one is reduced. Overall the total loudness is non-linear function of the sound pressure level and its spectral distribution.

Loudness is measured in phons. The number of phons is the level in dBSPL of an equally loud 1000 Hz tone.

The physiological effects of the human ear create an inner representation of the sound that constitutes the information given to the brain.

Beyond that we enter the cognitive level, where the brain interprets the data and psychology intervenes. For example, added non linear components are more annoying than linear distortion. On the other hand a complex sound can mask distortion. At that level the brain selects the relevant information. It becomes a matter of taste, culture and personal background, and is the domain of subjectivity.

3.3. PEAQ

In the early 90's, speech and music codecs were proliferating, but there was no standard way to qualify them. Because codecs are non-linear and non-stationary, traditional measurement methods (Frequency Response, THDN…) do not provide good results.

In 1994 the ITU asked several institutions to work on competitive solutions to measure the audio performance of codecs [34-35]. The different methods that came out of that joint effort were then compiled in one single method called PEAQ (Perceptual Evaluation of Audio Quality).

In 1998, the first version of ITU-R Recommendation BS.1387 "Method for objective measurements of perceived audio quality" [33] was published.

Here is an extract of the front page of that recommendation

"The ITU Radiocommunication Assembly,

considering

a) that conventional objective methods (e.g. for measuring signal-to-noise ratio and distortion) are no longer adequate for measuring the perceived audio quality of systems which use low bit-rate coding schemes or which employ analog or digital signal processing;

d) that formal subjective assessment methods are not suitable for continuous monitoring of audio quality, e.g. under operational conditions;

e) that objective measurement of perceived audio quality may eventually complement or supersede conventional objective test methods in all areas of measurement;

f) that objective measurement of perceived audio quality may usefully complement subjective assessment methods;

g) that, for some applications, a method which can be implemented in real time is necessary,

recommends

that for each application listed in Annex 1 the method given in Annex 2 be used for objective measurement of perceived audio quality."

Point "e)" is significant. The goal of ITU-R BS1387 is to get to an objective audio quality measurement. Its purpose is to provide a hearing model that can emulate a subjective assessment of sound quality. Even if it is aimed primarily at codecs, such a hearing model could be applied to any kind of audio device.

Figure 10 shows a general block diagram of PEAQ.

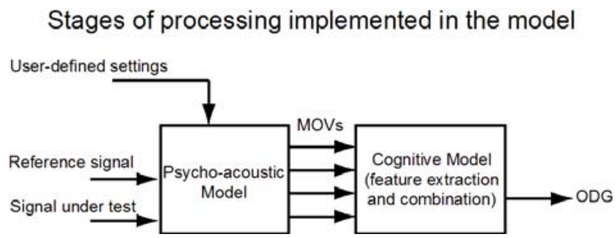

Figure 10: Block diagram of PEAQ

It is worth noting that the model produces a single metric: ODG (Objective Difference Grade). That simple index rates the perceived audio quality of the signal under test compared to the reference signal.

They are two versions of PEAQ algorithm:

- A basic version which is FFT based
- An advanced version which uses a combination of FFT and filter bank

The basic version is for real-time implementation. The advanced one is for in-depth analysis and is about 4 times more complex than the basic one.

Both algorithms transform successive time frames of the signals into "internal representations", where the loudness of the sound is distributed along a pitch scale. In other words, the model transforms a time-frequency distribution of sound pressure into a time-pitch distribution of loudness. During the process of going from physics to physiology, the sound energy is smeared along the pitch scale as well as the time scale. The smearing along pitch scale models the frequency masking and the smearing along the time scale models the temporal masking. The absolute threshold of hearing is obtained by combining an ear frequency weighting and an internal noise frequency dependent offset. The main outputs of the model are the excitation patterns and the masking thresholds as time-frequency functions.

The Model Output Variables (MOV) are:

- modulation differences between reference and test signals
- noise loudness (includes non-linear distortions)
- linear distortion
- bandwidth measurement for reference and test
- noise to mask ratio
- probability of detection and statistics of impaired frames
- harmonic structure in error signal

The cognitive model at the end condenses these MOV into a single quality index (ODG in [Figure 10]) that is a combination of weighted MOV. The weight is optimized by a neural network learning algorithm.

4. USE OF PEAQ FOR MEASURING THE AUDIBILITY OF HARMONIC DISTORTION AND RUB & BUZZ

4.1. Introduction

Testing with a sine wave input is still the standard procedure in the loudspeaker industry because harmonic analysis allows easy detection of specific fabrication defects.

We are therefore using PEAQ with a pure tone (sine wave) test signal to quantify audible distortion for production line QC.

In order to meet the speed requirements of production line we are using the basic FFT version mentioned in section 3.3.

We have further simplified the algorithm by ignoring everything related to time smearing and time modulations, assuming a steady sine response.

And finally for this paper, for a first evaluation of our approach, we have focused on only two MOV: the partial noise loudness and the error harmonic structure.

4.2. Design of our PEAQ algorithm

Figure 11 is a block-diagram of our simplified PEAQ algorithm. The inputs (stimulus and response) are both FFT spectra in dBSPL vs. Hz. The response spectrum is a measurement of the sound pressure emitted by the loudspeaker excited by a steady tone (stimulus).

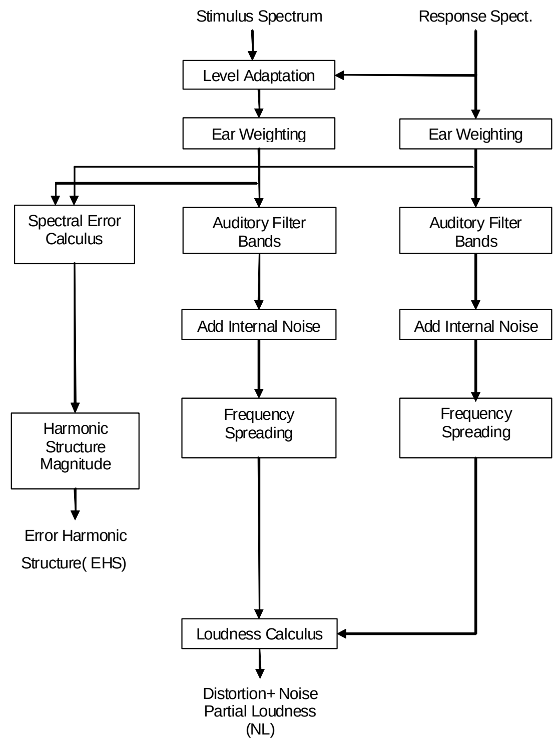

Figure 11: Overall flowchart of our PEAQ algorithm

4.3. Level Adaptation

Because we use an auditory model all input data must be in dBSPL.

The stimulus spectrum level (pure tone) is an internal reference and for the calculation is scaled to match the level of the response spectrum.

4.4. Ear Frequency Weighting

The transfer function of the outer and middle ear (see Figure 5) is modeled by a frequency dependent weighting function:

$$W[k]/\mathrm{dB} = -0.6 \cdot 3.64 \cdot \left(\frac{k.dF}{1000}\right)^{-0.8} + 6.5 \cdot \mathrm{e}^{-0.6 \cdot \left(\frac{k.dF}{1000} - 3.3\right)^2} - 10^{-3} \cdot \left(\frac{k.dF}{1000}\right)^{3.6} \quad \text{(Eq. 1)}$$

where dF is the FFT resolution in Hz.

The weighting W is applied to the FFT inputs.

$$F_e[k] = \left| F_x[k] \right| \cdot 10^{\frac{W[k]}{20}} \qquad \text{(Eq. 2)}$$

F_x denotes either Stimulus and Response spectra.

4.5. Auditory Filter Bands

The auditory pitch scale is calculated from an approximation given by [33]. The pitch units are in Bark.

$$z / Bark = 7 \cdot \operatorname{arcsinh}\left(\frac{f / \mathrm{Hz}}{650}\right) \qquad \text{(Eq. 3)}$$

See Figure 8.

For each spectrum the energy is mapped along the Bark scale, in 109 auditory filters ranging from 91.7 Hz up to 17700 Hz, with a resolution of 0.25 Bark.

The outputs of this stage of processing are the energies of the auditory filters, Pe[k].

4.6. Adding Internal Noise

The internal noise energy is implemented as a frequency dependent offset P_{Thres}, and is added to the energies in each frequency group:

$$\frac{P_{Thres}[k]}{dB} = 0.4 \cdot 3.64 \cdot \left(\frac{f_c[k]}{kHz}\right)^{-0.8} \qquad \text{(Eq. 4)}$$

Where fc[k] are the center frequencies of the auditory filters (see section 4.5)..

The output of this stage of processing, Pp[k] is referred to as “Pitch patterns”.

4.7. Frequency Spreading

The *Pitch patterns* $P_p[k]$ are smeared out over frequency using a level dependent spreading function.

The spreading function is a two sided triangle in dB vs. Bark. The lower slope is always 27 dB/Bark. The upper slope is frequency and level dependent. A series of spreading functions is shown in **Error! Reference source not found.** for different levels at 1 kHz ~ 8.5 Bark.

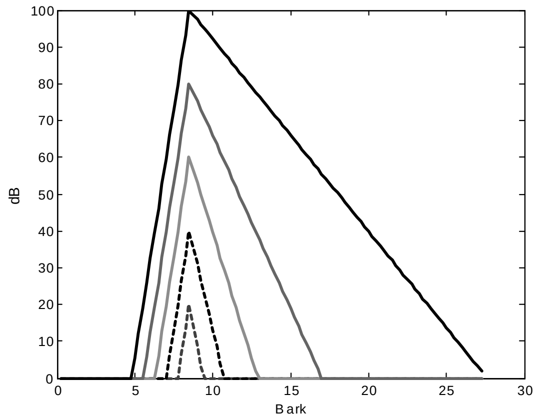

Figure 12: Different Spreading Functions at 1 kHz

The slopes are calculated according to:

$$\frac{S_u[k, L[k]]}{\mathrm{dB} / Bark} = \min(0, -24 - 230 / f / Hz + 0.2 \cdot L[k] / \mathrm{dB}) \qquad \text{(Eq. 5)}$$

$$S_l[k, L[k]] = 27 \frac{\mathrm{dB}}{Bark} \qquad \text{(Eq. 6)}$$

$$L[k] / dB = 10 \cdot \log_{10}(P_p[k])$$

The spreading is carried out independently for each frequency group *k*:

$$E[k] = \frac{1}{Norm_{SP}[k]} \left(\sum_{j=0}^{Z-1} E_{line}[j,k]^{\gamma} \right)^{\frac{1}{\gamma}} \qquad \text{(Eq. 7)}$$

γ is a compression mixing exponent ($0 < \gamma < 2$).

In PEAQ [33] its value is set to: $\gamma = 0.4$.

E_{line} is given by:

$$E_{line}[j,k] = \begin{cases} \dfrac{1}{A[j]} \cdot 10^{\frac{1}{10}\left(L[j] + res \cdot (k-j) \cdot S_l[j, L[j]]\right)} & \text{if } k < j \\ \dfrac{1}{A[j]} \cdot 10^{\frac{1}{10}\left(L[j] + res \cdot (k-j) \cdot S_u[j, L[j]]\right)} & \text{if } k \geq j \end{cases} \qquad \text{(Eq. 8)}$$

With *A[j]*, total energy of the spreading function.

The base curve $Norm_{SP}[k]$ is calculated according to the same equations but using a constant reference input $L[k] \equiv 0$.

res is the resolution of the pitch scale in Bark (0.25 in our version).

The patterns at this stage of processing, E[k], are used later on for the computation of modulation patterns and are referred to as "Excitation Patterns".

The excitation patterns are the internal representation of the signals in the hearing model.

4.8. Loudness Calculation

The specific loudness of the Signal Under Test and the Reference Signal are calculated according to the formula:

$$N[k] = const \cdot \left(\frac{E_{Thres}[k]}{s[k]} \right)^{\beta} \cdot \left[\left(1 - s[k] + \frac{s[k] \cdot E[k]}{E_{Thres}[k]} \right)^{\beta} - 1 \right]$$

(Eq. 9)

as given in [33]. The growth exponent β is set to 0.23.

The scaling constant is chosen in order to give an overall loudness of 64 sones = 100 phons for a 100 dBSPL sine tone at 1 kHz.

Threshold index *s* and excitation at threshold E_{Thres} are calculated according to:

$$E_{Thres}[k]/dB = 0.364 \cdot \left(\frac{f}{1000Hz} \right)^{-0.8} \quad \text{(Eq. 10)}$$

$$\frac{s[k]}{dB} = -2 - 2.05 \cdot atn\left(\frac{f}{4\,\text{kHz}} \right) - 0.75 \cdot atn\left(\left(\frac{f}{1\,600\,\text{Hz}} \right)^2 \right) \quad \text{(Eq. 11)}$$

with f= fc[k], filter band center frequency in Hz.

The overall loudness of the Signal Under Test and the Reference Signal is calculated as the sum across all filter channels of all specific loudness values above zero.

$$N_{total}[n] = \frac{24}{Z} \cdot \sum_{k=0}^{Z-1} \max\left(N[k,n], 0 \right) \quad \text{(Eq. 12)}$$

NOTE 1 – Due to the different peripheral ear models, the loudness calculated here is not identical to the loudness as defined in ISO 532 (Acoustics – Method for calculating loudness levels 1975).

NOTE 2 – The loudness growth exponent β is optimized for wideband signals e.g. music). It is not completely accurate for pure tones [40].

4.9. Partial Loudness Calculation

This MOV estimate the partial loudness of additive distortions in the presence of the masking Reference Signal. The formula for the partial loudness is designed to yield the specific loudness of the noise if no masker is present and to yield the ratio between noise and mask if the noise is very small compared to the masker.

The partial noise loudness is calculated according to:

$$NL[k] = const \,.\, \left(E_{Thres} \right)^{\beta} \cdot \left[\left(1 + \frac{\max(E_{test} - E_{ref}, 0)}{E_{Thres} + E_{ref} \cdot \beta} \right)^{\beta} - 1 \right]$$

(Eq. 13)

where

- *Const* is a calibration factor such that NL is equal to the loudness of *Etest* when *Eref* is negligible.
- E_{Thres} is the internal noise function $P_{Thres}[k]$ as defined in section 4.6.

The excitation patterns (from section 4.7) are used as inputs.

The coefficient β, which determines the amount of masking, is calculated by:

$$\beta = \exp\left(-\alpha \cdot \frac{E_{test} - E_{ref}}{E_{ref}} \right) \quad \text{(Eq. 14)}$$

From [33] Table 11, α = 1.5

The final global value TotalNL is the overall noise loudness of NL[k].

$$TotalNL = \frac{24}{Z} \cdot \sum_{k=0}^{Z-1} Max(NL[k], 0) \quad \text{(Eq. 15)}$$

4.10. Harmonic Structure Error

A strong and extended harmonic structure in the test spectrum is a signature of Rub & Buzz [Figure 14]. The power cepstrum is used here to detect this periodicity as follows:

The spectrum is first weighted by the ear frequency response, and normalized. Then the FFT of the log magnitude spectrum (the power cepstrum) of the test signal is calculated to quantify the harmonic content.

In a spectrum, the harmonic series forms a repetitive pattern with a period equal to the fundamental frequency. Taking the FFT of the spectrum, we then get a peak situated at the inverse of the fundamental frequency. E.g. a harmonic series with a fundamental of 100 Hz, will yield a peak at 1/100= 10ms in the cepstrum. The level of that peak rises with the number and level of the successive harmonics.

The Error Harmonic Structure variable (EHS) is the magnitude of the peak in the cepstrum corresponding to the fundamental. A high value of the EHS variable indicates the presence of Rub & Buzz. In short, we could name EHS the "Buzz Factor".

5. EXPERIMENTAL RESULTS

The goal of this paper is to find a better way to correlate loudspeaker distortion measurements with perception. Since sound quality is very subjective, we first focused on distortion audibility. In other words, we want to be able to predict if the distortion we measure is audible to the average human being. In a future paper, we hope to address in more detail not just whether the distortion is audible, but how subjectively bad it sounds.

Our aim was to quantify how loud the distortion sounds irrespective of frequency, level, number of harmonics and noise, yet also identify the frequency and sound pressure level at which the perceived distortion loudness occurs so that we can be able to describe the conditions under which it takes place.

In order to quantify distortion audibility, we used masking curves to determine whether the distortion we measured was audible and our PEAQ algorithm to determine how loud in phons the distortion sounded.

We did not use the traditional measure of distortion as a percent of the fundamental linear level because the audibility changes with frequency and level due to the non-linear behavior of the human ear.

We started by measuring several 6" by 9" oval car loudspeakers with various defects.

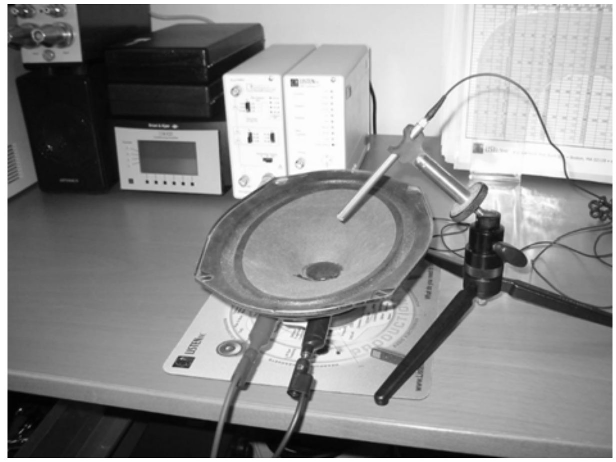

Figure 13: Test setup for measuring the spectrum of a batch of car loudspeakers using SoundCheck audio measurement system.

We had a good loudspeaker with no seriously audible distortion and several other speakers with varying levels of audible distortion. We measured the average spectrum of each unit at 100 Hz and approximately 120 dB SPL and noted the level of distortion audibility. We specifically focused on 3 units; a good unit, a borderline bad sounding unit with a rubbing voicecoil, and an extremely poor sounding unit with a badly glued spider.

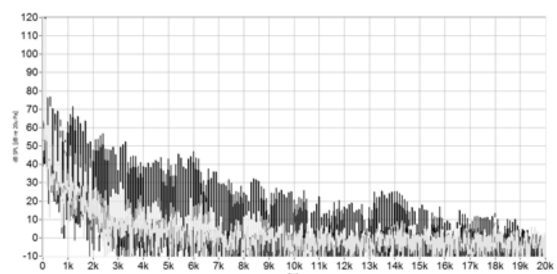

Figure 14: Spectrum of Good (bottom curve), Borderline (middle curve), and Bad (top curve) loudspeakers for 100 Hz @ 120 dBSPL

It is clear by comparing the spectrum of these three loudspeakers, that the "bad" loudspeaker has the most high order harmonics that create a buzzing sound. Interestingly, the bad loudspeaker had the lowest 2nd and 3rd harmonic distortion, whereas the "good" loudspeaker has the highest 2nd and 3rd harmonic distortion.

The slightly buzzing loudspeaker in Figure 15 has slightly more high order harmonics compared to the good unit. The 30th to 100th harmonics are over 90dB down from the fundamental level or less than 0.003% distortion! Can that possibly be audible? According to the masking curves [39] it is possible to hear these harmonics and even hear the 50th harmonic at almost -100 dB or 0.001% distortion!

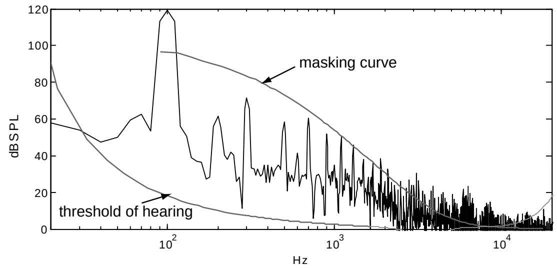

Figure 15: Spectrum of borderline loudspeaker, perceptual masking curve (top curve) for 120 dBSPL at 1 kHz, and the threshold of hearing from figure 7

The combination of the perceptual masking and the threshold of hearing shows how the human ear filters the spectrum of the loudspeaker and demonstrates that most of the harmonics, with the exception of those in the region of 5 kHz to 10 kHz, are inaudible to the average human.

We also measured the background noise as, if the background noise is higher than the masking curve and hearing threshold, it will influence the threshold of audibility. If the background noise is too high as often is the case on a production line, it will set the threshold of distortion audibility and will also influence the measurement. Even with time synchronous averaging or subtraction of the background noise with another microphone in the far field, background noise cannot easily be removed from the measurement.

The average spectrum of each loudspeaker was input into our PEAQ distortion algorithm to quantify the loudness of the distortion plus noise. Here are the results.

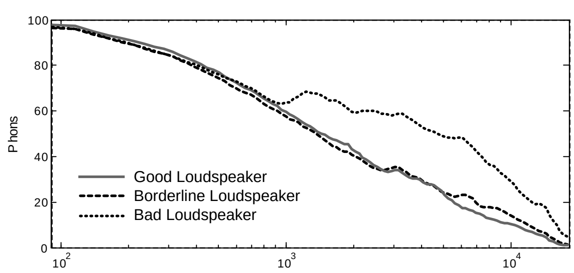

Figure 16: Perceptual Loudness curves for Good (TL = 107 phons), Borderline (TL = 109 phons), and Bad loudspeakers (TL = 112 phons)

Not too surprisingly, the loudspeaker with a lot of Rub & Buzz sounds louder than the good and borderline loudspeakers because of its high order harmonics above 1 kHz.

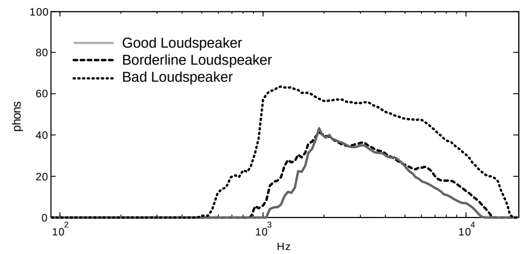

Figure 17: Partial Loudness curves for Good (NL = 66 phons), Borderline (NL = 69 phons), and Bad loudspeakers (NL = 93 phons)

The partial loudness indicates that the borderline and bad loudspeakers have more distortion and noise than the good loudspeaker. In particular, the bad loudspeaker is 27 phons higher than the good loudspeaker. The loudness in phons is the same as the level in dB SPL at 1 kHz, so a partial loudness of 93 phons is quite loud! The borderline loudspeaker with the slightly rubbing voice coil also has a discernibly higher partial loudness than the good loudspeaker.

Another output of our PEAQ algorithm is the Error Harmonic Structure (EHS) derived from the power Cepstrum which indicates if there are a lot of harmonics in the measurement. This helps separate out the noise from the harmonics in the measurement spectrums. A loudspeaker with a lot of Rub & Buzz will have a higher EHS at the reciprocal of the excitation frequency.

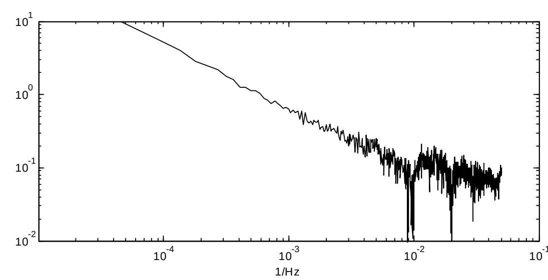

Figure 18: Power Cepstrum of Good loudspeaker (EHS = 0.13)

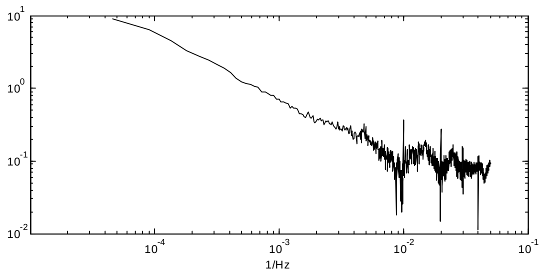

Figure 19: Power Cepstrum of Borderline loudspeaker (EHS = 0.37)

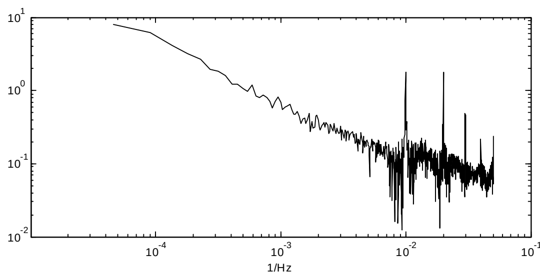

Figure 20: Power Cepstrum of Bad loudspeaker (EHS = 1.8). Note the sharp peaks that indicate the strong harmonic structure of the Rub & Buzz generated sounds.

The spikes at 1/100 Hz (10ms) indicate the strength of harmonic family associated with the 100 Hz excitation frequency. Again, it is pretty clear that the bad loudspeaker has a strong harmonic structure due to its high audible level of Rub & Buzz. The borderline loudspeaker with its just audible Rub & Buzz has a much lower Harmonic Structure but still noticeable compared to the good loudspeaker.

In order to further emphasize the differences between the three loudspeakers, we multiplied the Harmonic Structure overall level by the Partial Loudness overall level. We listened to a batch of good, bad and borderline loudspeakers and rated their sound quality subjectively from 1(best sounding) to 5 (worst sounding). We then measured these same loudspeakers and calculated the Partial Loudness, Harmonic Structure, and plotted the results [Figure 21]. It is clear that there is a strong correlation between our new PEAQ algorithm and auditory perception.

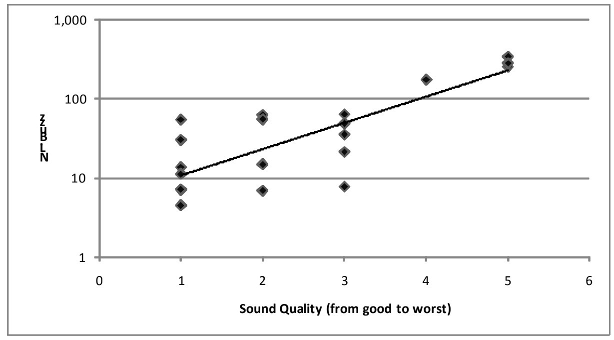

Figure 21: Sound Quality vs. Partial Loudness times Harmonic Structure with best fit trend line. Each marker represents a different loudspeaker/test condition.

It can be seen that there is a very strong correlation between our PEAQ measurements of distortion audibility (Harmonic Structure x Partial Loudness) and our subjective impressions of the loudspeaker sound quality.

6. FUTURE DEVELOPMENTS

There are several areas for future extension of this work:

- Although it can be seen that there is a very strong correlation between our PEAQ measurements of distortion audibility and our subjective impressions of the loudspeaker sound quality, further research is needed to compare this new method with existing Rub & Buzzbuzz test and measurement methods.
- The NL and EHS are just two Model Output Variables of the PEAQ algorithm. PEAQ also can output the detection probability to better understand the masking thresholds, the noise to mask ratio (distortion like variable), and the weighted sum of MOV to get a Quality Index that rates the global severity of distortion. There is much potential for comparing these other MOVs to the subjective impressions of the loudspeaker sound quality.
- In addition, the Cepstrum curves could be modified to remove the DC component so that if there are no harmonics, the curve will be flat and not ramp up toward zero time.
- The algorithm could be tuned for pure tone measurement and the calibration reviewed.

- Similar algorithms could be applied to lower order harmonic distortion and other types of distortion.
- Perceived distortion levels could be calculated using other types of signal.
- Further work could be done exploring the use of the loudness unit phon and the significance of its magnitude relative to dB SPL which is more familiar to users of audio test equipment.

7. CONCLUSION

The new Rub & Buzz testing algorithm based on our PEAQ model demonstrates a strong correlation between perceived sound quality and partial loudness x harmonic structure and shows promise as a novel method for production line testing of Rub & Buzz. There are several additional areas that need to be explored to further validate this method and broaden its range of applications.

8. REFERENCES

[1] F. J. Tang and P. Skytte, "Determination of Production-Related Defects in the Manufacture of Acoustic Transducers," Presented at the 73rd Convention of the Aud. Eng. Soc., Preprint 1989 (1983 Mar.).

[2] G. G. Groeper, M. A, Blanchard, T. Brummett, and J. Bailey, "A Reliable Method of Loudspeaker Rub and Buzz Testing Using Automated FFT Response and Distortion Techniques," Presented at the 91st Convention of the Aud. Eng. Soc., Preprint 3169 (1991 Oct.).

[3] D. B. Keele, Jr. and D. Schwing, "Loudspeaker Production Testing Using the Techron TEF System 20 TDS Analyzer and Host PC," Presented at the 11th International Conference of the Aud. Eng. Soc., (1992 Oct.).

[4] S. Temme, "Why and How to Measure Distortion in Electroacoustic Transducers," Presented at the 11th International Conference of the Aud. Eng. Soc., (1992 Oct.).

[5] S. Temme, "Audio Distortion Measurements," Bruel & Kjaer Application Note (1992 May).

[6] M. Davy, D. C. Manuel, "A New Nonstationary Test Procedure for Improved Loudspeaker Fault Detection," J. Audio Eng. Soc., vol. 50, no. 6, pp. 458-469 (2002 Jun.).

[7] M. Davy, H. Cottereau, and C. Doncarli, "Loudspeaker Fault Detection Using Time-Frequency Representations," Presented at the 2001 IEEE Conference on Acoustics, Speech, and Signal Processing, (2001 May).

[8] W. Klippel, "Measurement of Impulsive Distortion, Rub and Buzz and other Disturbances," Presented at the 114th Convention of the Aud. Eng. Soc., Paper 5734 (2003 Mar.).

[9] P. Brunet, E. Chakroff, and S. Temme, "Loose Particle Detection in Loudspeakers," Presented at the 115th Convention of the Aud. Eng. Soc., Paper 5883 (2003 Oct.).

[10] P. Brunet and S. Temme, "Enhancements for Loose Particle Detection in Loudspeakers," Presented at the 116th Convention of the Aud. Eng. Soc., Paper 6163 (2004 May).

[11] J. Anthony, R. Celmer, D. Foley, T. Pagliaro, B. Sachwald, and S. Thompson, "Higher Order Harmonic Signature Analysis for Loudspeaker Defect Detection," Presented at the 117th Convention of the Aud. Eng. Soc., Paper 6251 (2004 Oct.).

[12] W. S. Galway, M. A. LaBruzzo, R. D. Celmer, and D. Foley, "Loudspeaker Defect Analysis Using Ultrasonic Harmonic Characterization," J. Acoust Soc. Am., vol. 119, no. 5, pp. 3272-3276 (2006 May).

[13] S. Irrgang, W. Klippel, and U. Seidel, "Loudspeaker Testing at the Production Line," Presented at the 120th Convention of the Aud. Eng. Soc., Paper 6845 (2006 May).

[14] A. Voishvillo, "Assessment of Nonlinearity in Transducers and Sound Systems – from THD to Perceptual Models", Presented at the 121st Convention of the Aud. Eng. Soc., Paper 6910 (2006 Oct.).

[15] A Voishvillo, "Measurements and Perception of Nonlinear Distortion – Comparing Numbers and Sound Quality", Presented at the 123rd Convention of the Aud. Eng. Soc., Paper 7174 (2007 Oct.).

[16] L. Du and Y. Ji, "A New Method for Loudspeaker Small Noise Vibration Detection Based on Wavelet Package," Presented at the 2008 IEEE Conference

on Audio, Language and Image Processing, (2008 Jul.).

[17] A. Farina, "Silence Sweep: a Novel Method for Measuring Electro-Acoustical Devices," Presented at the 126th Convention of the Aud. Eng. Soc. (2009 May).

[18] H. Fletcher, "Speech and Hearing in Communication," 2nd Edition of 1958, Republished by the Acous. Soc. Am. (Aug. 1995).

[19] A. Small, Jr., "Pure Tone Masking," J. Acoust. Soc. Am., vol. 31, Issue 12, pp. 1619-1625 (December 1959

[20] E. Zwicker, "Masking and Psychological Excitation as Consequences of the Ear's Frequency Analysis," section 7 from R.Plomp and G. F. Smoorenburg Editors, "Frequency Analysis and Periodicity Detection in Hearing," A.W. Sijthoff – Leiden – The Netherlands (1970).

[21] P. Lindsey and D. Norman, "Human information processing: An Introduction to Psychology," Lindsey-Norman (Jun. 1972).

[22] E. Zwicker and K. Schorn, "Psychoacoustical tuning curves in audiology," Int. J. of Audiology, vol.17, no 2, pp. 120-140 (1978 Jan.).

[23] B. Moore, "Psychophysical Tuning Curves Measured in Simultaneous and Forward Masking," J. Acoust. Soc. Am., vol. 63, no. 2 (1978 Feb.).

[24] M. R. Schroeder et al, "Optimizing digital speech coders by exploiting masking properties of the human ear", J. Acoust. Soc. Am. Volume 64, Issue S1, pp. S139-S139 (November 1978).

[25] E. Zwicker and U. Zwicker, "Audio Engineering and Psychoacoustics: "Matching Signals to the Final Receiver, the Human Auditory System," J. Audio Eng. Soc., vol. 39, no. 2, pp. 115-126 (1991 Mar.).

[26] H. Fastl, "The Psychoacoustics of Sound Quality Evaluation," Acustica, vol.83, 754-764, (1997).

[27] S. Van de Par, et al., "A New Psychoacoustic Masking Model For Audio Coding Application," Presented at the 2002 IEEE Conference on Acoustics, Speech, and Signal Processing, vol. 2, pp.1805-1808 (2002 May).

[28] T. Wysocki, M. Darnell, and B. Honary, "Advanced Signal Processing for Communication Systems," Kluwer Academic Publishers (2002).

[29] M. Bosi and R. Goldberg, "Introduction to Digital Audio Coding and Standards," Kluwer Academic Publishers (2003).

[30] B. C. J. Moore and C. T. Tan, "Measuring and Predicting the Perceived Quality of Music and Speech Subjected to Combined Linear and Nonlinear Distortion," J. Aud. Eng. Soc., vol. 52, no. 12, pp. 1228-1244 (2004 Dec.).

[31] H. Fastl and E. Zwicker, "Psychoacoustics: Facts and Models," Springer-Verlag (2007).

[32] T. Rossing, "Springer Handbook of Acoustics," Springer-Verlag (2007).

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

[34] C. Colomes, C. Schmidmer, T. Thiede, and W. Treurniet, "Perceptual Quality Assessment for Digital Audio: PEAQ-The New ITU Standard for Objective Measurement of the Perceived Audio Quality," Presented at the AES 17th Int. Conf. (Sept 1999).

[35] T. Thiede, et al., "PEAQ- The ITU Standard for Objective Measurement of Perceived Audio Quality", J. Audio. Eng. Soc., vol.48, no 1/2, pp. 3-29, (2000 Jan./Feb.).

[36] E. Benjamin, "Evaluating Digital Audio Artifacts with PEAQ," Presented at the 113th Convention of the Aud. Eng. Soc., Paper 5711 (2002 Oct.).

[37] A. Magadum, "Evaluation of PEAQ for the Quality Measurement of Perceptual Audio Encoders," Presented at the AES 29th Int. Conf. (Sep. 2006).

[38] S.Temme et al, "Challenges of MP3 Player Testing", Presented at the AES 122nd Convention. (May. 2007).

[39] L.D. Fielder, E.M. Benjamin , "Subwoofer Performance for Accurate Reproduction of Music", J. Aud. Eng. Soc., Vol. 36, No. 6, pp. 443-456 (1988 June).

[40] P. Kabal, "An Examination and Interpretation of ITU-R BS.1387: Perceived Audio Quality", Dept. of Electrical & Computer Engineering, McGill University, Version2: 2003-12-08

9. GLOSSARY

This glossary is provided to clarify some of the terms used in this paper and provide a convenient reference list of topics.

Glossary Word Index :

1. Bark
2. Codec (coder/decoder)
3. Critical Bands
4. Crossover Distortion:
5. Distortion
6. Frequency Spreading
7. Harmonics
8. ITU (International Telecommunication Union)
9. Loose Particle Distortion
10. Maskee
11. Masker
12. Masking
13. Masking Temporal
14. Masking Threshold
15. Model Output Variables (MOV)
16. Noise
17. PEAQ
18. Perceptual Coding
19. Phon
20. Psychoacoustic model
21. Psychoacoustics
22. Rub & Buzz Distortion
23. Sone
24. Spreading Function
25. Subjective Testing
26. Threshold

1. Bark:

The Bark is the standard unit corresponding to one critical band width of human hearing. This unit of bandwidth represents the standard unit of bandwidth expressed in human auditory terms, corresponding to a fixed length on the human cochlea. It is approximately equal to 100 Hz at low frequencies and 1/3 octave at higher frequencies, above approximately 700 Hz.

2. Codec (coder/decoder):

A generic term applied to, among other things, lossy and lossless audio compression technologies implemented in hardware or software. Encoded data can be wrapped in a file format appropriate for the data, or decoded from such a file format. For example, the MP3 file format is a wrapper that can hold perceptually-encoded audio data.

3. Critical Bands:

The frequency resolving power of the auditory system can be considered as the result of bandpass filters. Such filters have been measured extensively by masking techniques and have become known as critical bands. Critical bands can be centered on any frequency, and their width varies with frequency. In psychoacoustics, a critical band is the maximum bandwidth of noise which is perceived by humans to be the same loudness as a sine wave of the same power at band center.

4. Crossover Distortion:

A characteristic type of distortion produced in an amplifier’s push-pull output stage if improperly biased such that only the peaks of low-level signals drive the amplifier into normal amplification ranges. A “dead band” input amplitude range may consequently exist, with signals in the “dead band” not producing output.

5. Distortion:

A difference, typically unintentional and undesired, between the signals on the input and output of an audio device. Commonly measured types of distortion include harmonic distortion, intermodulation distortion, quantization distortion, and jitter. Intentional differences between input and output signals, such as level or equalization differences, are not described as distortion.

6. Frequency Spreading:

An internal operation of the PEAQ process that smears out the computed data in the frequency domain that mimics the masked hearing thresholds of human hearing.

7. Harmonics:

Also called overtones, these are vibrations at frequencies that are multiples of the fundamentals. Harmonics extend without limit beyond the audible range. They are characterized as even-order and odd-order harmonics. A second-order harmonic is two times the frequency of the fundamental; a third order is three times the fundamental; a fourth order is four times the fundamental; and so forth. Each even-order harmonic second, fourth, sixth, etc.-is one octave or multiples of one octave higher than the fundamental; these even-order overtones are therefore musically related to the fundamental. Odd-order harmonics, on the other hand third, fifth, seventh, and up-create a series of notes that are not related to any octave overtones and therefore may have an unpleasant sound. Audio systems that emphasize odd-order harmonics tend to have a harsh, hard quality.

8. ITU (International Telecommunication Union):

The ITU is a world-wide organization within which governments and private sector coordinate the establishment and operation of telecommunication networks and services; it is responsible for the regulation, standardization, coordination and development of international telecommunications as well as the harmonization of national policies.

The ITU goal is to foster and facilitate the global development of telecommunications for the universal benefit of mankind, through the rule of law, mutual consent and cooperative action.

9. Loose Particle Distortion:

A sound generated by loose particles that are trapped in the loudspeaker during the manufacturing process. The sound is easily heard but difficult to measure because of the random nature of the sound generated by the particles bouncing around inside the loudspeaker.

10. Masker:

The higher level signal in a masking process that masks lower level signals and therefore prevents them from being heard.

11. Maskee:

The lower level signal in a masking process that may not be heard in the presence of a higher level signal. The maskee signal may be higher or lower in frequency than the masker signal.

12. Masking:

1) The amount (or the process) by which the threshold of audibility for one sound is raised by the presence of another (masking) sound. In other words, a property of the human auditory system by which an audio signal cannot be perceived in the presence of another audio signal. 2) The interference of one sound by another; the interfering sound is called the masking sound. Masking is considered to be undesirable if it interferes with the audibility of desired sounds, or it may be used to beneficial effect in some forms of environmental noise control and in noise reduction or perceptual audio coding systems.

13. Masking Temporal:

The psychoacoustic effect in time where a strong signal causes weaker signals occurring just before or just after the strong signal to be inaudible.

14. Masking Threshold:

A function in frequency and time below which an audio signal cannot be perceived by the human auditory system.

15. Model Output Variables (MOV)

The MOVs are intermediate output values of the perceptual measurement method. These variables are based on basic psycho-acoustical findings and are used to determine the final audio quality index.

16. Noise:

Undesired energy or data components in a communication channel included with the signal that the channel is carrying.

17. PEAQ:

Perceptual Evaluation of Audio Quality – (Perceptual Evaluation of Audio Quality) is a standardized algorithm for objectively measuring perceived audio quality, developed in 1994-1998 by a joint venture of experts within Task Group 6Q of the International Telecommunication Union (ITU-R). It was originally relased as ITU-R Recommendation BS.1387 in 1998 and last updated in 2001. It utilizes software to simulate perceptual properties of the human ear and then, integrate multiple model output variables (MOV) into a single metric. PEAQ characterizes the perceived audio quality as subjects would do in a listening test according to ITU-R BS.1116. PEAQ results principally model mean opinion scores (MOS) that cover a scale from 1 (bad) to 5 (excellent).

18. Perceptual Coding:

Lossy compression that takes advantage of limitations in human perception. In perceptual coding, audio data is selectively removed based on how unlikely it is that a listener will notice the removal. MP3 and MPEG-2 AAC are popular examples of perceptual coding.

19. Phon:

The phon is a unit of perceived loudness level for pure tones. The purpose of the phon scale is to compensate for the effect of frequency on the perceived loudness of tones. By definition, 1 phon is equal to 1 dBSPL at a frequency of 1 kHz. The equal-loudness contours are a way of mapping the dBSPL of a pure tone to the perceived loudness level in phons. These are now defined in the international standard ISO 226:2003, and the research on which this document is based concluded that earlier Fletcher–Munson curves and Robinson-Dadson curves were in error.

20. Psychoacoustic Model:

A mathematical model of the masking behavior of the human auditory system.

21. Psychoacoustics:

The study of the interaction of the auditory system and acoustics or the study of the perception of sound. The development of perceptual coding techniques relies on psychoacoustics. Psychoacoustics is the study of human hearing and how it is influenced by the brain. In lossy audio codecs, psychoacoustic principles are applied to determine which audio data are less critical to the ear and therefore may be discarded to reduce file size.

22. Rub & Buzz Distortion:

A variety of distortions and noises created in a loudspeaker, mostly due to mechanical defects such as voice coil rubbing the magnet, cone touching connection wires, etc.

23. Sone

The sone is a unit of perceived or subjective loudness. The sone is equivalent to 40 phons, which is defined as the loudness level NL of a 1 kHz tone at 40 dB SPL. The number of sones to a phon was chosen so that a doubling of the number of sones sounds to the human ear like a doubling of the loudness, which also corresponds to increasing the sound pressure level by approximately 10 dB.

24. Spreading Function:

A function that describes the frequency spread of masking effects in the PEAQ process.

25. Subjective Testing:

Using human subjects to judge the performance of a system. Subjective testing is especially useful when testing systems that include components such as perceptual audio coders. Traditional audio measurement techniques, such as signal-to-noise and distortion measurements, are often not compatible with way perceptual audio coders work and therefore cannot characterize their performance in a manner that can be compared with other coders, or with traditional analog systems.

26. Threshold:

The point at which a stimulus is just strong enough to be perceived or produce a response.

9.1. Glossary Acknowledgements:

Most of the entries in this glossary were copied from sources on the internet without permission. Thanks go to the following sources:

a) GNU Ware: http://www.gnuware.com/icecast/appendix_b.html

b) Apple Developer Connection:

http://developer.apple.com/documentation/MusicAudio/Reference/CoreAudioGlossary/Glossary/core_audio_glossary.html

c) Stereo Sound Book:

http://www.stereosoundbook.com/pages/glossary.html#D

d) Audio Precision:

http://ap.com/library/glossary

e) Owens Corning Corp:

http://www.owenscorning.com/around/sound/glossary.asp

f) Wikipedia:

http://en.wikipedia.org/wiki/PEAQ

g) Soren Bech Book: "Perceptual Audio Evaluation-Theory, Method and Application":

http://www.amazon.com/Perceptual-Audio-Evaluation-Theory-Application/dp/0470869232/ref=sr_1_1?ie=UTF8&qid=1248970950&sr=8-1#

h) Radio Magazine Online:

http://radiomagonline.com/mag/glossary/