

# Audio System Evaluation with Music Signals

Wolfgang Klippel, Stefan Irrgang

KLIPPEL GmbH, Mendelssohnallee 30, 01309 Dresden, Germany

Synthetic test signals, such as multi-tone signals or sweeps, are mostly used for the development and end-of-line testing of components and the complete audio systems in cars. Those signals provide objective, reproducible and interpretable test results in a short time. In contrast, the customer uses the audio system to reproduce music and speech, which are non-stationary signals with complex spectral and temporal properties. This paper discusses measurement methods that can be used for assessing the performance of the audio system by using any synthetic and natural (music) stimuli. A new technique based on adaptive modeling of the linear time variant distortion is used to combine physical and perceptual evaluation of the residual nonlinear distortion.

## 1 Introduction

The comprehensive evaluation of the performance of an audio system in the final target application (in-situ) requires physical measurement of the signal distortion generated in the reproduced sound and perceptual evaluation of those signal components by listening as well.

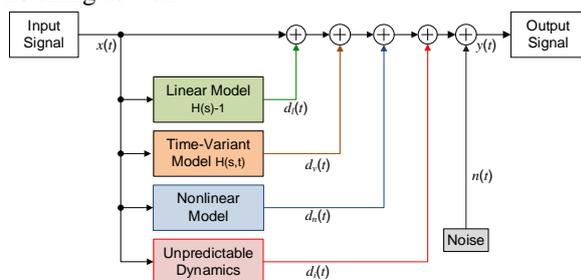


Figure 1: Black box modeling of the distortion components generated by an audio system

Figure 1 illustrates the generation of signal distortion in the output signal  $y(t)$  that is not found in the input signal  $x(t)$ . This model comprises five subsystems that generate particular distortion components. A linear system with a fixed transfer response  $H(s)-1$  generates a linear distortion component  $d_l(t)$  that corresponds to the amplitude and phase variation versus frequency. A second linear system with a time varying transfer response  $H_t(s, t)$  generates the second distortion component  $d_t(t)$  caused by voice coil heating and other slow changes of the transducer properties due to aging, fatigue and viscoelasticity of the suspension material. A nonlinear model generates regular distortion  $d_n(t)$  related to the nonlinearities inherent in the transducer. This kind of distortion  $d_n(t)$  is deterministic and predictable during the design process. A further nonlinear subsystem with unpredictable dynamics generates irregular distortion  $d_i(t)$  caused by transducer defects (e.g. coil rubbing), system defects (e.g. door panel rattling) and overload conditions (e.g. voice coil bottoming). Some of the irregular distortions are generated by random causes (e.g. loose particles) and cannot be predicted by modelling.

The separation of the distortion components in the output  $y(t)$  of a real audio device can be accomplished by using an artificial test signal that has a sparse spectrum (e.g. single tone). For example, the spectral analysis may be used for those test signals to separate nonlinear distortion components from the linear distortion found at the fundamental frequencies excited by the input signal  $x(t)$ . The spectral separation of the distortion fails for ordinary stimuli (e.g. music and speech) having a dense excitation spectrum. This paper presents alternative techniques to separate the linear signal components  $d_l(t)$  and  $d_t(t)$  from the nonlinear distortions  $d_n(t)$  and  $d_i(t)$  and shows how to use this data for objective and perceptual evaluation of the reproduced sound quality.

## 2 Correlation Analysis

Conventional correlation analysis applied to a steady-state input signal  $x(t)$  and output signal  $y(t)$  allows estimation of the linear transfer function  $H(f,n)$  and the coherence between the two signals. The coherence describes the fraction of the power in the output signal that can be explained by the input signal  $x(t)$  and a transfer function  $H(f,n)$ .

The correlation analysis can be effectively realized in the frequency domain by using FFT as illustrated in Figure 2.

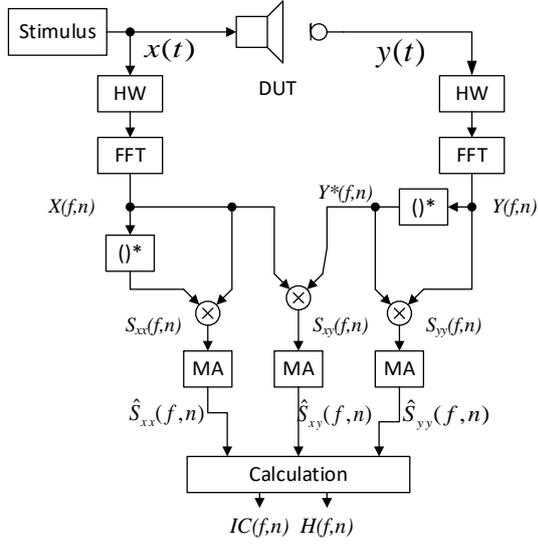


Figure 2: Random data analysis to estimate the linear transfer function  $H(f,n)$  and incoherence  $IC(f,n)$  between input and output signal

The input spectrum  $X(f,n)$  is calculated from overlapping data blocks  $x(n)$  derived from the input signal  $x(t)$  tapered by a Hanning window with block length  $T$ . In the next step the auto-spectral density function of the input signal is estimated

$$\hat{S}_{x,x}(f,n) = \frac{1}{K} \sum_{k=0}^{K-1} X(f,n-k)X^*(f,n-k) \quad (1)$$

by multiplying the input spectrum  $X$  with the conjugated value  $X^*$  and applying equally weighted averaging to the input power spectrum over  $K$  history blocks.

Alternatively, the auto-spectral density can be calculated with less numerical effort as

$$\hat{S}_{x,x}(f,n) = \mu \hat{S}_{x,x}(f,n-1) + (1-\mu)X(f,n)X^*(f,n) \quad (2)$$

by using exponential weighted averaging to the input power spectrum with the smoothing factor  $\mu$ . The exponential weighted averaging effectively uses input data from a larger number of blocks where the last blocks have the largest influence on the estimated auto spectral density.

The same processing is applied to the auto-spectral density function of the output signal, given for equally weighted averaging in Eq. (3):

$$\hat{S}_{y,y}(f,n) = \frac{1}{K} \sum_{k=0}^{K-1} Y(f,n-k)Y^*(f,n-k) \quad (3)$$

The cross-spectral density function is calculated by multiplying the input spectrum  $X(f,n)$  with the conjugated output spectrum  $Y^*(f,n)$  and by averaging the cross power spectrum constantly weighted over  $K$  history blocks in Eq. (4):

$$\hat{S}_{x,y}(f,n) = \frac{1}{K} \sum_{k=0}^{K-1} X(f,n-k)Y^*(f,n-k) \quad (4)$$

The auto-spectral density functions  $\hat{S}_{x,x}$ ,  $\hat{S}_{y,y}$  and cross density functions  $\hat{S}_{x,y}$  vary over each block  $n$  (corresponding to time  $t$ ) if the input signal  $x(t)$  is not stationary.

The ratio between the estimated cross- and auto-spectrum gives the transfer function

$$H(f,n) = \frac{\hat{S}_{x,y}(f,n)}{\hat{S}_{x,x}(f,n)} \quad (5)$$

which may be updated at each block  $n$  after the time  $T/2$ .

The Coherence Function

$$\gamma(f,n) = \frac{|\hat{S}_{x,y}(f,n)|}{\sqrt{\hat{S}_{x,x}(f,n)\hat{S}_{y,y}(f,n)}} 100\% \quad (6)$$

can be estimated in percent based on the ratio between cross spectrum and the geometrical mean value of the auto-spectra.

For the evaluation of the signal distortion, it is more convenient to use the in-coherence function defined in percent as

$$IC(f,n) = 100\% - \gamma(f,n) \quad (7)$$

or decibel as:

$$IC_{dB}(f,n) = 10 \log \left( 1 - \frac{\gamma(f,n)}{100\%} \right) \text{dB} \quad (8)$$

## 2.1 Interpretation

The correlation analysis assumes that the input signal  $x(t)$  has steady-state properties and provides sufficient excitation for the activation of the nonlinearities and defects inherent in the device under test. White noise at high amplitudes generating sufficient voice coil displacement would fulfil those requirements for electro-dynamical transducers. However, some music and speech signals have non-stationary properties and would not provide sufficient energy to identify the system under test over the full audio band. At low amplitudes, the measurement noise  $n(t)$  will dominate the output signal  $y(t)$  and the incoherence value will be 100 %. The poor signal-to-noise-ratio (SNR) will generate a high variation of the estimated transfer function  $H(f,n)$  over block number  $n$ . This variation can be reduced by increasing the total number  $K$  of blocks or the smoothing factor  $\mu$  used in the averaging of the power spectra Eqs. (1) and (2), respectively. In order to cope with nonstationary properties of the stimulus, the measured data is excluded from the calculation of the auto and cross-spectral densities

$$\left. \begin{aligned} \hat{S}_{x,x}(f,n) &= \hat{S}_{x,x}(f,n-1) \\ \hat{S}_{y,y}(f,n) &= \hat{S}_{y,y}(f,n-1) \\ \hat{S}_{x,y}(f,n) &= \hat{S}_{x,y}(f,n-1) \end{aligned} \right\} \text{if } \frac{X(f,n)X^*(f,n)}{\hat{S}_{x,x}(f,n)} < L \quad (9)$$

at those frequencies  $f$  where the instantaneous input power spectrum referred to the mean value is below a defined limit  $L$ .

The block length  $T$  used in the FFT and the effective width of the time window determine the spectral resolution of the correlation analysis. The measurement of a subwoofer requires a typical block length  $T > 1$  s to get 1 Hz resolution at low frequencies. A tweeter with a typical resonance frequency  $f_s > 500$  Hz requires a smaller resolution and can be measured with a block length of  $T < 0.2$  s. The averaging of the power spectrum in the calculation of the cross spectral density  $\hat{S}_{x,y}$  in Eq. (4)

suppresses the incoherent power generated by noise  $n(t)$  and the regular and irregular nonlinear distortion  $d_n(t)$  and  $d_i(t)$ , respectively. Thus, without averaging ( $K=1$ ), the measured incoherence value is zero but rises with the increasing number of blocks  $K$  to the true value. The relative error of the incoherence measurement can be estimated as

$$e_{ic}(f) = \frac{\Delta IC(f)}{IC(f)} = \frac{100\%}{\sqrt{K}} \quad (10)$$

depending on the total number of blocks  $K$  used for averaging. This requirement determines the total measurement time

$$T_{tot} \geq \frac{K}{2} T \quad (11)$$

for overlapping blocks using a Hanning window. For example, a total measurement time  $T_{tot} = 5$  s comprising 10 blocks of data analysed with a spectral resolution of  $\Delta f = 1$  Hz, which is required to measure a woofer with a resonance frequency  $f_s < 100$  Hz, gives an relative error  $e_{ic} \approx 30\%$  that makes the measured value  $IC_{dB}$  approximately 1.5 dB lower than the true value determined with longer averaging ( $K \gg 10$ ).

The total measurement time  $T_{tot}$  required in the correlation measurements is usually too long to follow the time varying properties of the device under test corresponding to the time varying linear distortion component  $d_i(t)$  in Figure 1. For example, the increase of the voice coil temperature by 100 K due to the electrical heating by the stimulus varies the electrical resistance  $R_e$  by 39 % and reduces the pass band sensitivity of a loudspeaker by 2.8 dB. This change occurs with a thermal time constant  $\tau_c$  of the voice coil depending on the thermal capacity  $C_{TV}$  of the coil and the thermal resistance  $R_{TV}$  between coil and pole plate. This thermal time constant  $\tau_c$ , which is typically 1 s in a micro-speaker and 10 s for a woofer, is in the same order of magnitude as the averaging time  $T_{tot}$  in the correlation analysis. Thus, the auto- and cross-spectral densities represent the device under test in the mean assuming stationary signal properties. The averaging of the interval  $0 < t < T_{tot}$  of the power spectrums assigns all time

variance of the transfer function  $H(f,t)$  to the incoherence value  $IC$ . This may be considered as an artefact of the correlation analysis when applied to a non-stationary process. However, the distortion components caused by time variance  $d_i(t)$  (see Figure 1) change only the amplitude and phase of the spectral components provided in the input signal  $x(t)$  but generates no additional spectral components (harmonics, intermodulation). Although that distortion may dominate the incoherence value  $IC$ , it degrades the reproduced sound quality much less than the nonlinear distortion  $d_n(t)$  and  $d_i(t)$ . For a shorter averaging time (smaller value of  $K$  and  $\mu$ ), this artefact will vanish, and the estimated transfer function  $H(f)$  converges to the instantaneous transfer function  $H(f,t)$ , and the incoherence value will represent the nonlinear distortion only. Unfortunately, there is a conflict with the condition defined in Eq. (11), and the correlations analysis requires a compromise between measurement error  $e_{ic}$  and the separation of the time variant distortion  $d_i(t)$ .

A noise signal  $n(t)$  with stationary properties (e.g. microphone noise) will always contribute to the incoherence value  $IC$  independent of the averaging parameters  $K$  and  $\mu$ . At very low and high frequencies where the reproduced output spectrum  $Y(f,n)$  has a low signal-signal-to-noise ratio (SNR), the incoherence value  $IC$  becomes 100 %. Increasing the level of the stimulus  $X(f)$  at those frequencies will improve the SNR( $f$ ) and reduce the incoherence  $IC$  while the higher amplitude may increase the nonlinear distortion  $d_n(t)$  and  $d_i(t)$ , which may have the tendency to increase the incoherence. Thus, multiple measurements with varying stimulus level is a practical method to separate the effect of noise and nonlinearities.

The incoherence is a power related metric and, for this reason, less sensitive for detecting the irregular nonlinear distortion  $d_i(t)$  which has a crest factor due to the random and impulsive nature of loudspeaker defects and parasitic vibration problems (e.g. door panel rattling). Although the peak value of the irregular distortion  $d_i(t)$  exceeds the peak value of the regular nonlinear distortion  $d_n(t)$ , the incoherence will only reveal the regular nonlinear distortion if this component provides a higher power to the correlation analysis.

The assumptions, requirements and particularities of the correlation analysis make the interpretation of the measured incoherence values difficult and limit their diagnostic value.

### 3 Residual Distortion Analysis

An interesting alternative to the correlation analysis is the modeling of the linear distortion with an

adaptive filter that can cope with time variant properties of the device under test.

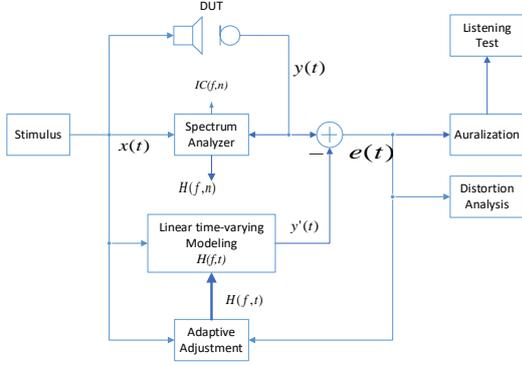


Figure 3: Residual distortion analysis based on linear time variant modelling and adaptive parameter adjustment.

Figure 3 shows a block diagram of this approach where a linear time variant model with the transfer function  $H(f,t)$  is connected in parallel to the device under test and where an error signal  $e(t)$  is generated as the difference between measured and model output:

$$e(t) = y(t) - y'(t) \quad (12)$$

The transfer function of the model is permanently updated by

$$H(f,t) = H(f,t - \Delta t) + \mu_H \frac{E(f,t)X(f,t)^*}{\hat{S}_{xx}(f,n)} \quad (13)$$

using the previous estimate  $H(f, t - \Delta t)$ , the estimated auto spectral density, the spectrum of the error signal  $e(t)$  and the input signal  $x(t)$ . The smoothing factor  $\mu_H$  is small enough to avoid noise  $n(t)$  and the nonlinear distortion components  $d_n(t)$  and  $d_i(t)$  causing a significant variation of the estimated transfer function  $H(f,t)$ , while the smoothing factor  $\mu_H$  is still large enough to follow any time variant properties of the device under test. After a short adaptation time, the error signal  $e(t)$  will not comprise any time variant distortion  $d_i(t)$  and  $d_n(t)$ , which are coherent with the input signal  $x(t)$ . The adaptation of the transfer function  $H(f,t)$  in the frequency domain as shown in Eq. (13) can be realized by using an FFT applied to history data with a block's length  $T$  for any time instance  $t$ .

An interesting alternative to FFT is the usage of the wavelet transformation corresponding with a filter bank with constant relative resolution. The processing load can be reduced by updating the coefficients of an FIR filter by using standard recursive least-squares estimation.

The adaptation process ensures that the error signal  $e(t)$  is the residuum of the linear time variant

modelling that comprises the regular and irregular nonlinear distortion components and noise.

### 3.1 Interpretation of the Residuum

Straightforward signal analysis can be applied to the residuum  $e(t)$  and to the measured output signal  $y(t)$  to determine a peak value

$$e_{peak}(n) = \text{MAX}_{t=0}^T (|e(t_n - t)|) \quad (14)$$

and an RMS value

$$e_{rms}(n) = \left( \frac{1}{T} \sum_{t=0}^T e(t_n - t)^2 \right)^{1/2} \quad (15)$$

within a data block  $n$  of length  $T$ .

The ratio of RMS value of the residuum distortion  $e(t)$  referred to RMS value of the measured output signal  $y(t)$  within the same data block  $n$  gives a residuum distortion metric in percent

$$D_{res}(n) = \frac{e_{rms}(n)}{y_{rms}(n)} 100\% \quad (16)$$

or in decibel

$$D_{res}(n) = 20 \lg \left( \frac{e_{rms}(n)}{y_{rms}(n)} \right) \text{dB} \quad (17)$$

which evaluates the incoherent power in the output signal generated by noise  $n(t)$  and both nonlinear distortion components  $d_i(t)$  and  $d_n(t)$ .

The crest factor of the residuum, defined as the ratio between peak and RMS value within the block  $n$ ,

$$C_E(n) = 20 \lg \left( \frac{e_{peak}(n)}{e_{rms}(n)} \right) \text{dB} \quad (18)$$

is a strong indicator for nonlinear distortion components  $d_i(t)$  and  $d_n(t)$  having a higher crest factor than electrical measurement noise ( $> 12 \text{ dB}$ ). Irregular nonlinear distortion  $d_i(t)$  generated by loose particles, bottoming and other loudspeaker defects generate the highest crest factor value.

The coincidence of high residuum distortion metric  $D_{res}(n)$  and a high crest factor value  $C_E(n)$  in one data block  $n$  can be used to select the critical data blocks for further frequency and time analysis.

Short-term FFT, wavelet, Wigner distribution or other filter banks reveal the instantaneous spectrum  $E(f,t)$ . It is recommended to combine the spectral analysis with listening to the down-sampled and interpolated residuum:

$$e_{up}(t) = e(\lambda_{slow} t) \quad (19)$$

A slow-down factor  $\lambda_{slow} < 1$  shifts all spectral components to lower frequencies and makes the temporal structure of the residuum easier accessible to human perception.

The probability density function  $\text{pdf}(e)$  of the residuum gives further insight into the amplitude distribution. The distribution of the residuum signal

over the sound pressure output, voice coil displacement and other internal state variables gives further clues to identify the physical root cause of the regular and irregular nonlinear distortion.

The mean value of the instantaneous transfer response  $H(f,n)$  averaged over all available blocks  $n=1,2, \dots,N$

$$\bar{H}(f) = \frac{1}{N} \sum_{n=1}^N H(f,n) \quad (20)$$

describes the time invariant linear properties of the device under test corresponding to the distortion signal  $d_i(t)$ .

The time variance of the amplitude response can be evaluated as the ratio of the instantaneous transfer response  $H(f,n)$  referred to mean value  $\bar{H}(f,n)$  in decibel:

$$\Delta H(f) = 20 \lg \left( \frac{|H(f,n)|}{|\bar{H}(f)|} \right) \text{dB} \quad (21)$$

A negative value of  $\Delta H(f)$  indicates thermal compression by voice coil heating in the pass band of the loudspeaker and at frequencies below the fundamental resonance. At the resonance frequency, a positive value of  $\Delta H(f)$  may be generated by a change of the electrical damping due to the reduced  $Bl(x)$  value for high displacement  $x$ . At higher frequencies above cone break-up, nonlinear and visco-elastic effects may shift the modal resonances and change the modal shape of the mechanical vibration, which causes positive and negative variation of the transfer function.

### 3.2 Auralization of Signal Distortion

The residuum signal  $e(t)$  separated from the linear signal allows for investigation of the audibility of the regular and irregular nonlinear distortion and their impact on the rated quality of sound reproduction by generating a virtual output signal

$$y_{vir}(t) = y'(t) + S_{DIS} e(t) \quad (22)$$

using a defined scaling factor  $S_{DIS}$  [9]. For  $S_{DIS}=1$  (corresponding to the 0 dB gain), the virtual output  $y_{vir}(t)$  equals the measured output signal  $y(t)$ . A larger scaling factor  $S_{DIS} > 1$  enhances the nonlinear distortion and noise in the virtual output  $y_{vir}(t)$  while keeping the linear distortions constant. This auralization technique can be used to generate a set of wave files where the residuum is varied in 3 dB steps, which is the basis for finding the audibility threshold in double blind tests (AB comparison between distorted signal  $y_{vir}(t)$  and undistorted signal  $y'(t)$ ) and for rating the perceived quality of the sound reproduction.

## 4 Practical Automotive Application

The presented methods were applied and verified in an experimental setup. A standard mid-class car, about 5 years old, was tested in two configurations:

1. Normal Car: The car is measured as is. There are no obviously audible defects when operated at medium amplitudes. However, at higher amplitudes minor symptoms of nonlinear distortion (loudspeaker, door buzzing) can be detected by careful listening.
2. Defective Car: A tambourine was inserted to the driver side door panel door tray of the same car to simulate an additional defect generating irregular nonlinear distortion  $d_i$ . The tambourine behaves like a nonlinear parasitic resonator generating a rattling noise if the loudspeaker generates enough vibration in a narrow band at about 100 Hz.

The correlation analyses and linear residual modelling are applied to the reproduction of a typical pop music (Vangelis, Album 1492, Track 2, 5 min) in order to extract reliable symptoms of the defect. The new methods are then compared with conventional methods using synthetic stimuli.

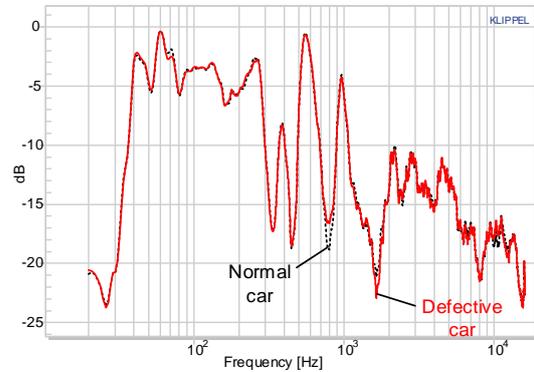


Figure 4: Measured transfer functions  $H(f)$  of the normal car (dashed curve) and defective car (solid curve)

### 4.1 Correlation Analysis

Fig. 4 shows the transfer function  $H(f)$  of the two configurations measured between the AUX input and the driver position using the audio stimulus (5 minutes) analyzed with a block length of 1 s. The difference curve is shown in Fig. 5.

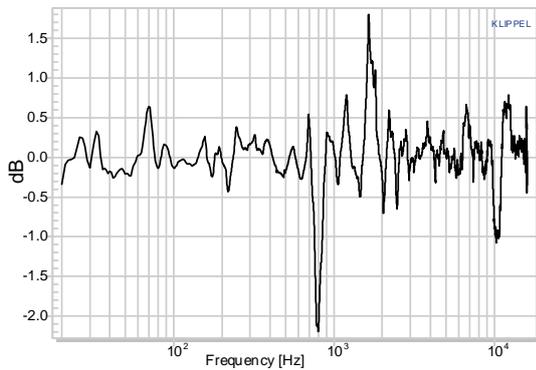


Figure 5: Difference between the transfer function  $H(f)$  of the normal and defective car

There are two higher deviations at 800 Hz and 1650 Hz caused by low SNR where acoustical cancellations at the microphone position reduce the available power for an accurate estimation. The remaining  $\pm 0.5$  dB difference between the two transfer functions do not provide any reliable clues for detecting the artificial defect (tambourine) even if the remaining setup (audio system, car interior, microphone) is identical. In practice, the microphone position and the influence of the particular car configuration (upholstery, extras, ...) may cause larger differences in the transfer function  $H(f)$  between two cars of the same series.

The measurement of the non-linear distortion components  $d_n$  and  $d_i$  provides more reliable clues for detecting the defect. The incoherence value  $IC(f)$  calculated based on the 5min music stimulus according to Eq. (7) and depicted in Fig. 6 reveals no significant difference between the normal and defective car.

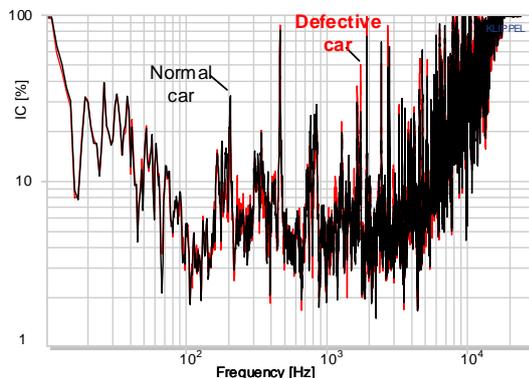


Figure 6: Incoherence  $IC(f)$  versus frequency  $f$  calculated based on a music stimulus recorded in a normal and defective car

The statistical calculation of the auto and cross spectral densities is less sensitive for a defect

comprising a nonlinear resonator because the regular and irregular nonlinear distortion generated provides much less incoherent power than time-varying distortion and the measurement noise. Thus, the incoherence analysis is well suited for identifying ambient noise at the lower and higher frequencies, where the music contains low spectral power.

Replacing the music stimulus by an artificial stimulus with stationary properties (e.g. pink noise) would improve the activation of the nonlinearities and would improve the sensitivity of the incoherence for regular nonlinear distortion  $d_n$ . However, the main problem is that the incoherence analysis evaluates the power (RMS value) of the incoherent signal part and cannot reveal any information about the fine structure (e.g. peak value) of the distortion signal. The linear residuum provides the distortions  $d_n$  and  $d_i$  that allow more advance analysis to be applied in the frequency and time domain, providing further valuable information for defect detection and root cause analysis of defects with impulsive properties.

## 4.2 Analysis of the Residuum

Figure 7 shows the measured signal and the modeled signal generated by adaptive filtering in a selected time frame (block 19). Both signals are almost identical, and the residuum is much smaller.

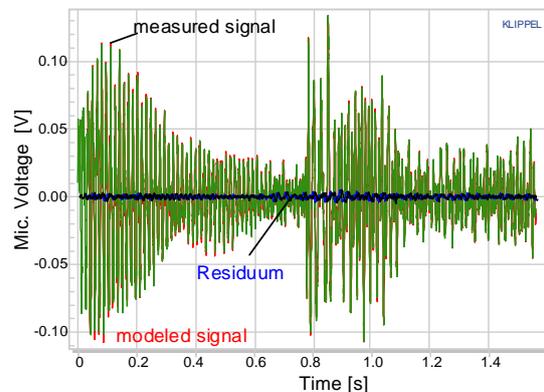


Figure 7: The reproduced music signal (measured and modeled) and the residuum of a selected block 19 where the crest factor of the residuum is high.

The residuum shows impulses and high peak values in the displayed time frame, which is a valuable clue of defects and nonlinearities. Block 19 has been selected from the complete music stimulus (32 blocks) by searching for the maximum of the crest factor of the residuum recorded in the defective car as shown in Figure 8.

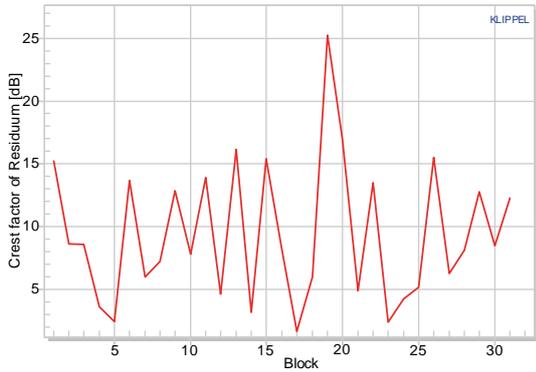


Figure 8: Crest factor of the residuum measured in the defective car over all time blocks of the music stimulus.

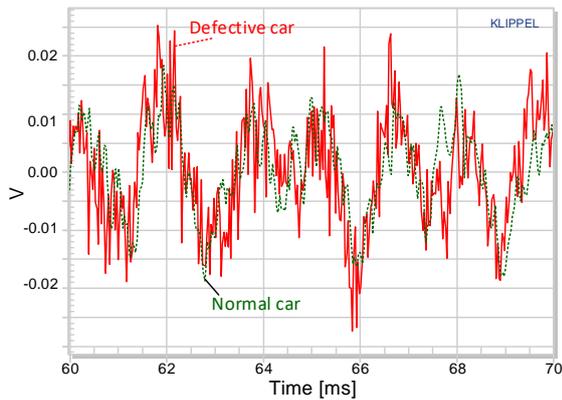


Figure 9: Detail of residuum signal  $e(t)$  of the music signal reproduced in the normal and defective car at the instance generating maximum crest factor.

Figure 9 compares the waveform of the residuum calculated from the music reproduced in the normal and defective car at the critical time section in block 19 where the high peak value in the residuum has been found. This comparison shows that the regular nonlinearity distortion  $d_n$  which is generated by the same loudspeaker nonlinearities in the normal and defective car are a dominant part in both residuum signals. The residuum of the defective comprises additional high frequency components which generate additional peaks in Figure 9. The crest factor of the residuum is a very helpful measure to find the critical parts in a longer recording and to perform a deeper analysis and careful listening to those parts.

### 4.3 Time Frequency Analysis

The time frequency analysis is a very powerful technique for distinguishing between the regular and irregular nonlinear distortion  $d_n(t)$  and  $d_i(t)$  found in the residuum  $e(t)$  and to reveal valuable clues for the

root cause of the defect. Most of this information would be not visible by applying the time frequency analysis to the measured signal  $y(t)$  because the linear signal components are usually much larger than the nonlinear distortions  $d_n(t)$  and  $d_i(t)$ .

Applying a wavelet analysis to the residuum signals (length 1.5s) reveals the clear defect at high frequencies in the first 200ms.

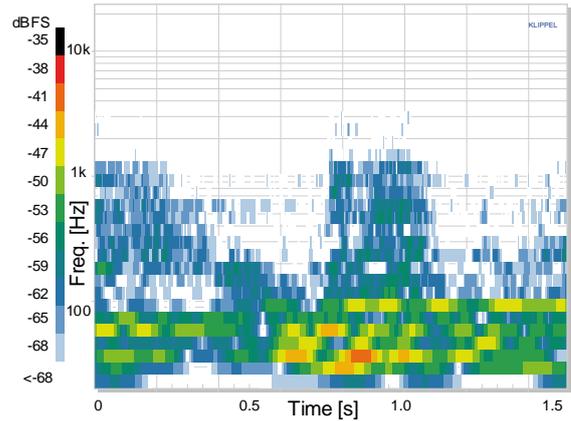


Figure 10: Time-frequency spectrum based on wavelet analysis applied to the residuum measured in the normal car

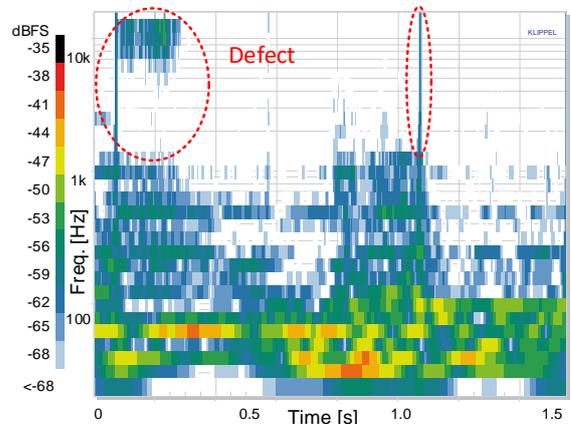


Figure 11: Time-frequency spectrum based on wavelet analysis applied to the residuum measured in the defective car for the same time block as shown in Figure 10

Figure 10 and Figure 11 show the sonograph of the residuum measured in the normal and defective car, respectively. The spectral components found at frequencies below 1 kHz are very similar in the normal and defective car and can be interpreted as the regular nonlinear distortion generated by the loudspeaker system.

At higher frequencies, the defect generates irregular distortion  $d_i$  which appear as two wide band clicks at

50ms and at 1.2s and a ringing over 100 ms at frequencies above 10 kHz.

The time frequency method shall be complemented with listening to the critical instants of the residuum. The masking from the linear response is removed and the defect symptoms are much more audible.

#### 4.4 QC EOL Measurements

In contrast to the measurements with music in the previous sections, the results with typical EOL test methods are presented for comparison. State-of-the-art high speed EOL tests are sinusoidal chirps with a logarithmic relationship between frequency and time. This property simplifies the separation of the irregular distortion from the linear components by high-pass filtering using a cut-off frequency that is ten times higher than the instantaneous excitation frequency. The sweep time was 1s using the same RMS level as for the music based tests.

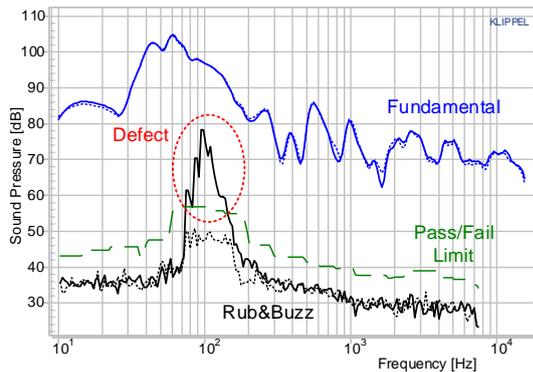


Figure 12: Frequency response of the SPL fundamental component and rub&buzz characteristics measured in normal car (dotted) and defective car (solid) by using an EOL-test system

Figure 12 shows the measured frequency response of fundamental component compared to the SPL level of the high-pass filtered distortion (rub&buzz). The EOL measurements confirm the results of the correlation measurements with music stimulus that the defect generates no significant differences in the fundamental response. The dotted rub&buzz curve measured in the normal car also reveals irregular distortion at 100Hz, which agrees with the distortion detected by careful listening in the normal car. The dotted rub&buzz curve was shifted by 6dB to higher values to generate a typical limit for PASS/FAIL decisions. Figure 12 shows the clear violation (excess of 20dB above the limit!) of the defective car, indicating a severe problem. The stationary properties of the sinusoidal chirp as well as the much higher probability of high stimulus amplitudes give the best condition for activating the defect and generating

high amplitudes of the irregular nonlinear distortion in the measured microphone signal.

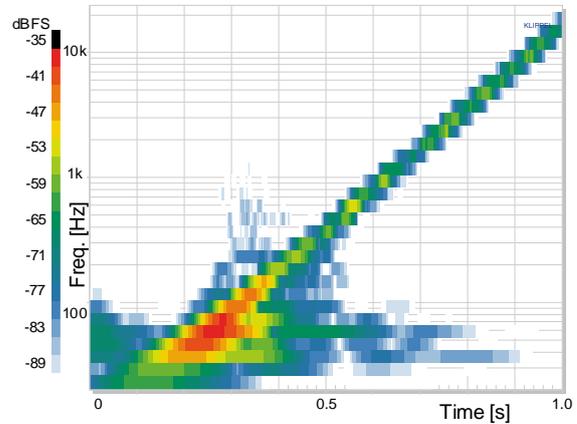


Figure 13: Time Frequency Analysis of the logarithmic chirp reproduced in the Normal Car

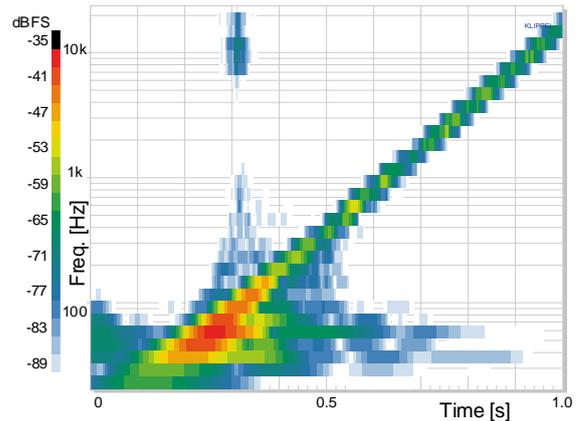


Figure 14: Time Frequency Analysis of the logarithmic chirp reproduced in the Defective Car

Although the chirp and other synthetic test signals used in EOL testing allow for detecting most defects in a shorter time, the measurement of the residuum may be necessary if a customer complains about a defect that generates audible symptoms only for a particular audio signal.

## 5 Conclusions

Excessive regular distortion (e.g. mechanical limiting of coil movement due to motor instabilities) and irregular distortion (e.g. parasitic vibration of door panels) have the strongest impact on the quality of the reproduced sound in cars. Standard measurement techniques based on synthetic test signals are available for the development and quality control at the end of the manufacturing process. However,

customers use complex audio signals that are non-stationary and have complex temporal and spectral properties.

The evaluation of the audio system in the final target application is important to both, design and manufacturing of products that provide the best performance at minimum cost to the customer. Some defects occur in the later product life due to aging, material fatigue, overload situations and other influences.

An in-situ measurement using music signals cannot be realized by traditional analysis methods for the following reasons:

- Most defects are a nonlinear process which needs sufficient excitation to generate significant symptoms, that are found as nonlinear distortions  $d_n$  and  $d_i$  in the reproduced sound.
- The non-stationary properties of the music signal require new techniques which exclude measured signals with low SNR from the analysis.
- The separation of the distortion components cannot be accomplished by a spectral analysis for stimuli having a dense spectrum.
- Some defects generate irregular, nonlinear distortion  $d_i$  with low energy but a high peak value.

Correlation techniques based on spectral cross- and auto power analysis are less suitable for non-stationary signals like music and cannot cope with time varying properties of the audio device (e.g. heating). Adaptive linear modeling of the device under test can identify the time-varying linear transfer function and can separate linear output from the residual signal, which reveals the distortions generated by the defect. Slowly time varying processes (e.g. heating) can be considered as well. The crest factor of the residuum can be used to select the most critical frame in a longer recording for a detailed analysis.

Auralization techniques can be applied to the residual signal and the linear modeled output to assess the audibility of the distortion and to rate the quality of the sound reproduction.

Time-frequency analysis can be efficiently applied to analyze the residuum and to help to identify the root cause of field complaints.

The measurement setup and data acquisition is very simple: only an in-situ recording as well as the original music stimulus is required to do the evaluation of audio systems.

The proposed method could be implemented as a web based service and can help deal with customer problems in any repair shop or even for private use.

## References

- [1] W. Klippel, "End-Of-Line Testing," Assembling Line, Theory and Practice, Chapter 10, INTECH 2011.
- [2] W. Klippel, "Sound Quality of Audio Systems" Script for Lecture at Dresden University of Technology, Dresden, 2017
- [3] L. Cohen, "Time Frequency Analysis", Prentice Hall PTR, 1995
- [4] S. Haykin, "Adaptive Filter Theory", 2<sup>nd</sup> ed., Prentice Hall, 1991
- [5] J.S. Bendat, A.G. Piersol: "Random Data", 4<sup>th</sup> ed., Wiley & Sons, Inc., 2010
- [6] S. Irrgang, et. al., "Assessment of Harshness Caused by Rattling Car Interiors," Automotive Acoustics, Research & Development, ATZ extra, September 2016, pp. 10 -15.
- [7] W. Klippel, "Loudspeaker Nonlinearities – Causes Parameters, Symptoms," J. Audio Eng. Soc. 54, no. 10, pp 907 – 939 (oct. 2006).
- [8] S. Irrgang, et. al. "In-situ Detection of Defects in Car Audio Systems," presented at the 48<sup>th</sup> Intern. Conf.: Automotive Sound Systems, September 21, 2012
- [9] W. Klippel, Auralization of Signal Distortion in Audio Systems, Part 1: Generic Modeling," presented at the 51<sup>st</sup> Intern. Conf. of the Audio Eng. Soc. Helsinki, Finland, 2013, August 22-24.