Audio Engineering Society

Convention Paper

Presented at the 141st Convention
2016 September 29–October 2 Los Angeles, USA

This paper was peer-reviewed as a complete manuscript for presentation at this Convention. This paper is available in the AES E-Library, http://www.aes.org/e-lib. All rights reserved. Reproduction of this paper, or any portion thereof, is not permitted without direct permission from the Journal of the Audio Engineering Society.

In-Vehicle Audio System Distortion Audibility versus Level and Its Impact on Perceived Sound Quality

Steve Temme[1] and Patrick Dennis[2]

[1] *Listen, Inc., Boston, MA, 02118, USA*

[2] *Nissan Technical Center North America, Inc., Farmington Hills, Michigan, 48331, USA*

Correspondence should be addressed to Steve Temme (stemme@listeninc.com) and Pat Dennis (dennisp@nissan-usa.com)

ABSTRACT

As in-vehicle audio system output level increases, so too does audio distortion. At what level is distortion audible and how is sound quality perceived as level increases? Binaural recordings of musical excerpts played through the in-vehicle audio system at various volume levels were made in the driver's position. These were adjusted to equal loudness and played through a low distortion reference headphone. Listeners ranked both distortion audibility and perceived sound quality. The distortion at each volume level was also measured objectively using a commercial audio test system. The correlation between perceived sound quality and objective distortion measurements is discussed.

1 Introduction

In a previous paper on headphone distortion audibility [1], various headphones were equalized to have the same target response, and different musical excerpts were then played through the headphones and recorded on a head and torso simulator. These recordings were played back on a reference, low distortion headphone and both subjectively and objectively compared for distortion. Different distortion metrics, THD (Total Harmonic Distortion), IMD (Intermodulation Distortion), and NCD (Non-Coherent Distortion) were compared to subjective preferences by a panel of trained listeners performing double-blind testing. Although the listeners had a difficult time discriminating distortion preference on 3 out of the 5 headphones, they all agreed that one headphone in particular had both audible and annoying distortion. The NCD metric had the best correlation to the subjective impression of distortion. A couple of interesting findings emerged: 1) equalizing different headphones to have the same target response is challenging, 2) the trained listeners struggled to describe and rank the distortion they heard. In this paper, the focus has shifted to in-vehicle audio systems because their distortion is typically higher than headphones. The equalization and description challenges faced with headphones was eliminated by using the same in-vehicle audio system and recording at different volume levels to achieve different levels of distortion. By playing these recordings back over a very low distortion reference headphone with each recording matched to the same loudness level, it was possible to focus on the distortion audibility versus level and its impact on perceived sound quality. In addition, a subjective rating of the distortion and an indication of the frequencies at which the distortion was most audible and annoying was obtained.

2 Test Setup

2.1 Music Selection

Three test tracks were used.

Program/Artist/Track/Album	Description
BW - Jennifer Warnes/Bird on a Wire/Famous Blue Raincoat/ Cypress Records, 1986	Female Pop Vocal
CD - Steely Dan/Cousin Dupree/Two Against Nature/ Giant Records/ 2000	Male Pop Vocal
AB - Estelle w. Kanye West/ American Boy Shine/ Atlantic Records, 2008	Male/Female Hip Hop

Table 1. Test Tracks Used.

Each track was digitally transferred directly from compact disc and edited into 20s to 30s loops. If the maximum peak of the waveform in each loop was not at 0dBFS, the track was normalized to 0dBFS to ensure maximum gain. The edited tracks were then transferred to a compact disc to be played back in-vehicle [2].

Binaural recordings were made at 48 kHz, 24 bit using a Head Acoustics HMS II.6 Head and Torso simulator and BEQ II measurement front end in the driver seat of a Nissan Altima with a 6 speaker audio system. The speaker arrangement consisted of two 3" speakers on top of the instrument panel (first order crossed over at 1.2 kHz), two full range 6.5" speakers low in the front doors and two full range 6x9 speakers in the rear parcel shelf. The head unit did not have electrical distortion limiting. The recordings were made from volume step 20 to volume step 40 (maximum) in 2 step increments which translate to a 1 dB/step change from volume step 20 to 22, a 2 dB/step change from volume step 22 to 34, and a 1 dB/step change from volume 34 to 40.

The binaural recordings were loaded into Head Acoustics Artemis and exported with the HSU III ID (Independent of Direction) ear EQ applied. Adobe Audition was then used to match each recording for overall loudness as well as spectrum in the bass region. The spectrum matching offsets the effect of dynamic loudness present in the head unit of the audio system. These adjustments removed the overall level and other frequency components, leaving only the amount of distortion as the distinguishing characteristic between recordings.

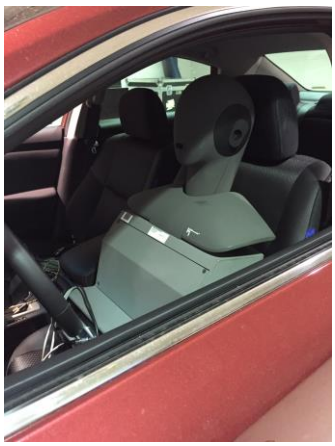

Figure 1. Binaural Recording Setup

2.2 Playback

Playback was conducted through a Head Acoustics PEQ V programmable equalizer, Head Acoustics HPS IV headphone amplifier and Stax SR303 headphones. The system is calibrated for aurally accurate playback of files recorded with the Head Acoustics BEQ II and HMS II.6.

2.3 Test Administration

A custom test GUI was written in MAX/MSP to administer the test. It was based on ITU-R BS.1116-1 Methods for Subjective Assessment of

Small Impairments in Audio Systems [3]. This is a method to evaluate impairments in audio systems using a double-blind triple-stimulus with hidden reference system. Three audio tracks, A – the reference file as well as B and C, one of which is the reference and the other is the impairment file, are played back simultaneously. The listener may freely switch between the tracks to compare distortion. The participant is asked to identify which file, B or C has the impairment and then rate the severity of the impairment on a 5 point scale:

1: Inaudible
2: Slightly audible
3: Audible
4: Noticeably distorted
5: Very distorted

The participant was then asked to describe the difference they heard using either a drop-down menu or a free form description field for both the reference and impaired files. Finally they were asked to identify in which portion of the music the difference was heard: bass, midrange and/or treble. No time limit was given to complete the evaluation, each participant was able to take as much time as they needed. Volume setting data was also collected.

Prior to testing the participant was taken through the GUI interface to get comfortable with the controls and the tasks to be completed. Examples of each audio track were also played so that the participant could familiarize themselves with the program material and impairments to be listened for. An information sheet was also available during the test if the participant felt they needed a reminder of the test setup and tasks at hand.

The song and track presentation order were randomized along with which file (reference or impaired) was assigned to B or C.

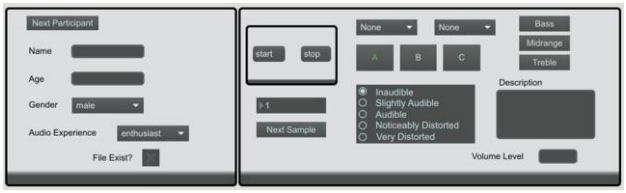

Figure 2. GUI

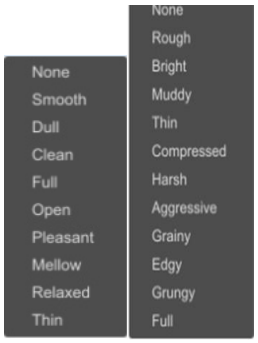

Figure 3. Reference and Impairment Descriptors

2.3. Test Participants

A total of 28 listeners participated in the study. Each participant was asked their level of audio experience from "Expert" to "Rarely Listen". There was no qualification for Expert since no test such as "Harman How to Listen" was administered to verify the assessment. The listeners consisted of:

Males: 26 Females: 2
Age Range: 19 – 61 Median Age: 35.5

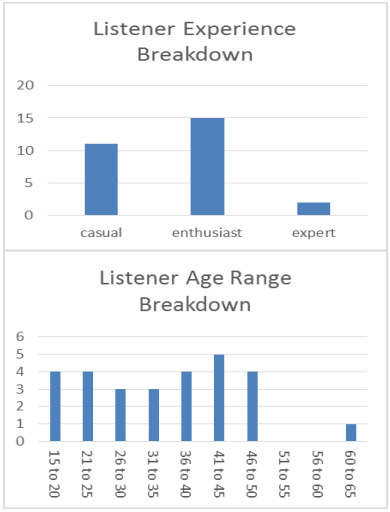

Figure 4. Listener breakdown by experience and age

3 Listening Results

3.1 Statistical Analysis

ANOVA and T-test

Analysis of Variance (ANOVA) and T-Tests were performed to discriminate the participants' level of perception of distortion severity. Before the analysis was done the severity data was normalized according to the formula provided in the ITU-R BS.1116-1 method for testing for audio impairments. The following analysis was performed:

Single Factor ANOVA Tests

1. Fixed independent variable was song (3 levels) and the dependent variable was severity level
2. Fixed independent variable was song (3 levels) and age range (7 levels) and the dependent variable was severity level
3. Fixed independent variable was track (6 levels for AB, 4 levels for BW & CD) and the dependent variable was severity level
4. Fixed independent variables were song (3 levels) and experience (2 levels) and dependent variable was severity.

ANOVA 1: Fixed independent variable was song (3 levels) and the dependent variable was severity level. Volume step 40 (max) was used for this comparison and the result was significant. The participants rated each of the songs as having differing severity levels of distortion. This makes sense since the tracks have differing loudness levels: AB = -1.66 dBFS, BW = -14.34 dBFS, and CD = -10.74dB FS. AB severity was rated the worse at 3.92, CD second at 3.50 and BW last with 3.06.

ANOVA 2: Fixed independent variables were song (3 levels) and age range (7 levels) and the dependent variable was severity level. As in the previous analysis, only volume step 40 was used. Only the song was significant (fig. 5).

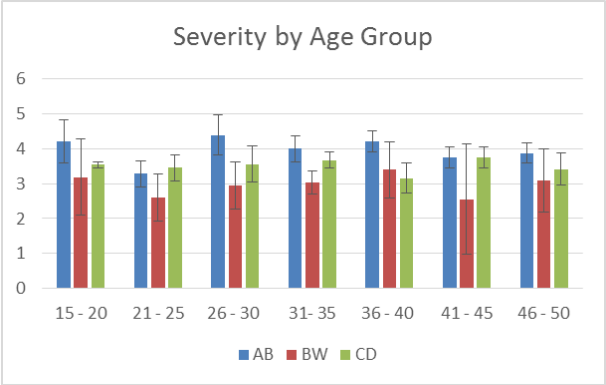

Figure 5. Severity of distortion by age group

ANOVA 3: Fixed independent variable was track (6 levels for AB, 4 levels for BW & CD) and the dependent variable was severity level. For each song tracks with > 70% correct identification were used for the analysis. All songs showed significant differences for the severity level meaning the participants were able to discriminate between differing levels of distortion within each song.

ANOVA 4: Fixed independent variables were song (3 levels) and experience (2 levels) and dependent variable was severity. Only volume step 40 was used. Analysis showed songs were significant but experience was not. Interaction between song and experience showed to be significant, with the enthusiasts rating BW less severely distorted than

the casual listeners, but a T-test comparing these two populations showed no significance.

T-tests

1. Listener Experience (Casual vs Enthusiast)
2. Track comparisons for each song

T-test 1: Listener Experience (Casual vs Enthusiast). This test was done two ways. First with all volume step 40 tracks combined and second separately for each the 3 songs. No significant difference was found for either way (fig. 6).

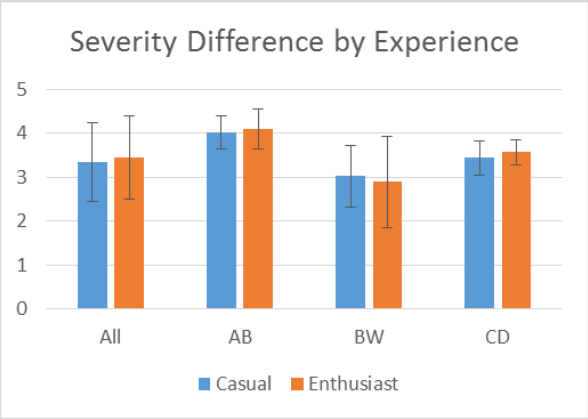

Figure 6. Severity difference by experience

T-test 2: Track comparisons for each song. For each song the tracks with > 70% correct identification were used. The analysis consisted of investigating where the severity rating difference became significant which was found in the previous ANOVA analysis (fig. 7).

AB: Tracks volume steps 30, 32, 34, 36, 38 and 40 were used. Significant differences were found between volume step 30 and 32 as well as between volume step 32 and 34. Volume step 40, 38 and 36 had no significant difference found.

CD: Tracks volume step 34, 36, 38 and 40 were used. Significant differences were found between volume step 38 and 40 as well as between volume step 34 and 36. No significant difference was found between volume step 36 and 38.

BW: Tracks volume step 34, 36, 38 and 40 were used. Significant differences were found between volume 34 and 36, between 36 and 38 and between 38 and 40.

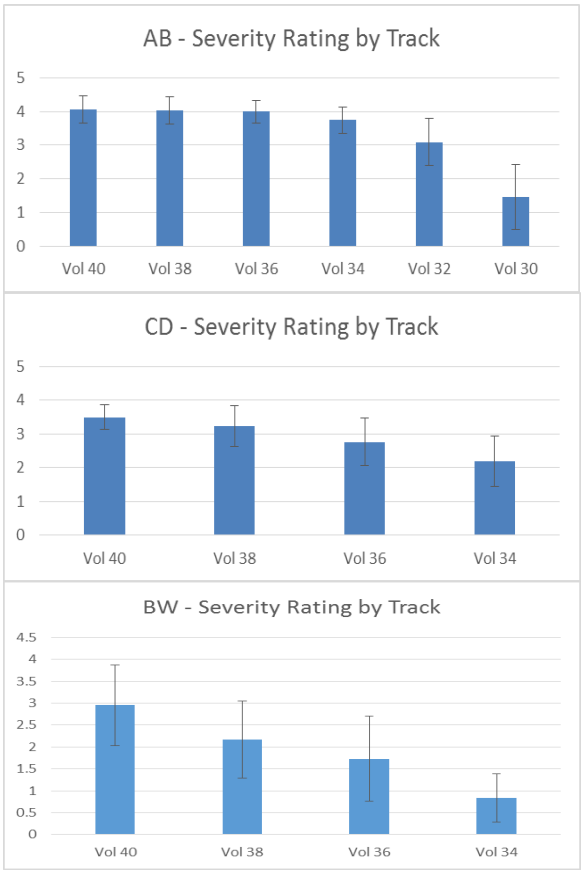

Figure 7. Severity Rating for the three different tracks, AC, CD, and BW

4 Distortion Measurements

A variety of objective distortion measurements were performed to correlate to the subjective listening tests. As pictured below in Figure 8, a single Listen SCM3 measurement microphone,

was positioned on the arm rest between the driver and passenger seats to measure the sound from the car audio system. The calibrated pressure microphone inside the car was connected to a Listen AudioConnect audio interface and microphone power supply which converts the signal from analog into digital for the SoundCheck® measurement software to analyze the measured waveforms. The digital test signals, created and played by the SoundCheck software, are passed through the audio interface for conversion to analog signals, and then amplified by the car audio head unit.

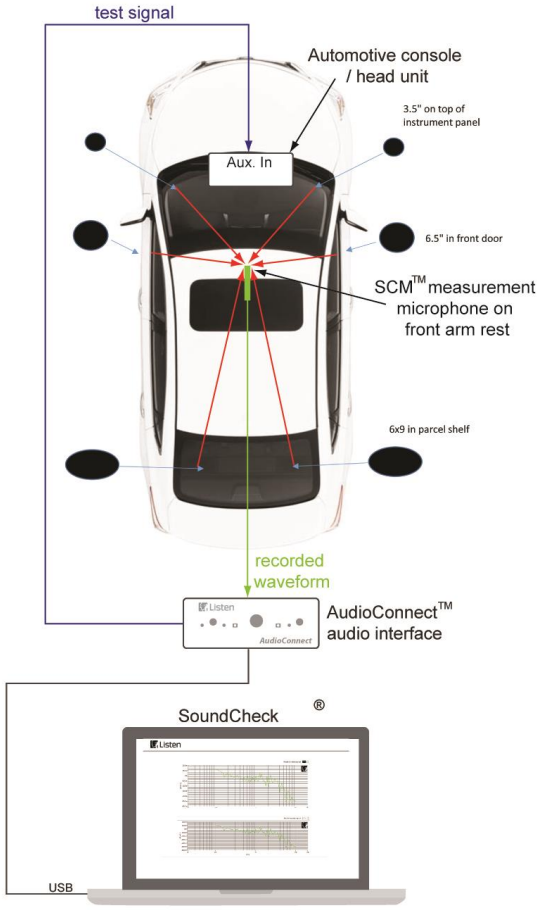

Figure 8. In-vehicle distortion measurement setup

The same binaural head with left and right ear simulator could have been used to measure the distortion but this was deemed unnecessary since both channels would either have to be reported separately or summed together in some fashion to represent what the listener is hearing. In addition, most of the distortion measurements presented are also being frequency normalized to the fundamental response and therefore are not affected by position of the microphone. For more information on frequency normalized distortion measurements, please read the author's AES paper, "How to graph distortion measurements"[12].

4.1 Test Signals for Measuring Distortion

Unlike linear measurements, non-linear distortion measurements depend heavily on the stimulus signal's spectral content and level [4]. For this reason, a variety of test signals were used: swept-stepped sine, two-tone intermodulation (one fixed and one moving tone), pink noise, and even the same music tracks used for listening as a test signal.

4.1.1 Harmonic Distortion

Traditional harmonic distortion measurements using a single test tone or sweeping tone are easy to calibrate and perform. They are also a good indicator of the type of non-linearity, e.g. asymmetric (even order harmonics), symmetric (odd order harmonics) and even Rub & Buzz (high order harmonics) [12]. Figure 9 shows the absolute fundamental frequency response for a 1/12th (R40) Octave stepped sine sweep at 1Vrms played into the Auxiliary Input jack on the car audio head unit. This was deemed to be a reasonable high input level to maximize signal to noise ratio without clipping the head unit.

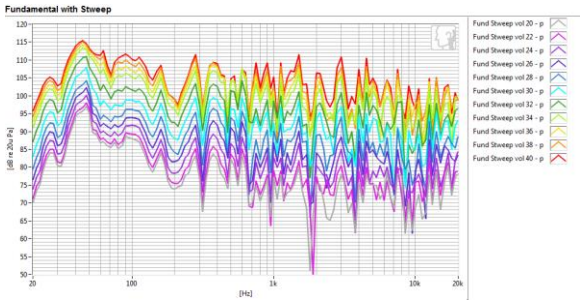

Figure 9. Fundamental vs. Volume Level

Volume levels varied from about 75dB to 110dB SPL. Figure 10 shows the Total Harmonic Distortion, frequency normalized (THD Norm) measurements at the eleven different test levels. This removes the influence of the non-flat linear response on the harmonic responses and THD curve [12].

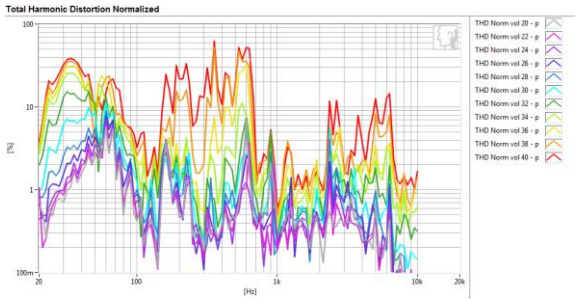

Figure 10. THD Normalized vs. Level

It was no surprise that the total harmonic distortion increases with level, one would expect this, and also expect that at some point the distortion would become clearly audible and probably unpleasant sounding.

There were high levels of THD at volume 32 and above (>95dB), especially at volume 38 & 40 which are approaching 60% THD at over 100dB SPL. This was very audible and mostly due to Buzz, Squeak and Rattle (BSR). This is clearly visible in the Perceptual Rub & Buzz graph below, Figure 11, which represents how the human ear perceives high order harmonic distortion (Rub & Buzz) [4]. It was hard to tell how much of the BSR was coming from the speaker or enclosure vibrations e.g. door panels, center console or rear parcel shelf, as at these volumes everything was vibrating.

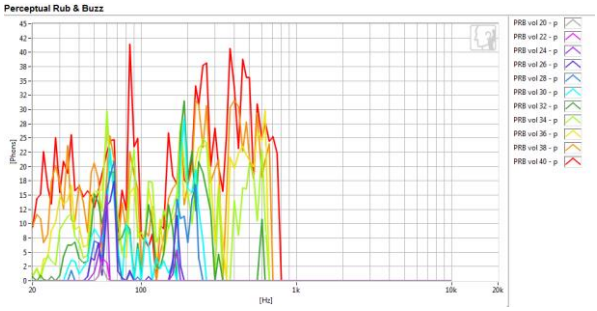

Figure 11. Perceptual Rub & Buzz vs. Level

At volume 34 (<100dB SPL) and under, the Rub & Buzz decreased significantly and could not be heard.

4.1.2 IM Distortion

An alternative to harmonic distortion (which does not reveal intermodulation products and is very different from music) is to use two test tones and measure intermodulation distortion. System theory shows that more than two tones are needed to fully characterize a non-linear system [5]. System nonlinearities cause intermodulation distortion (IMD) to occur due to amplitude and/or frequency modulation of the higher frequency components by the lower frequency components. The sum and difference components arising in two-tone interaction distortion have no harmonic musical

relationships and can therefore be quite annoying.

The test procedure fixed one tone at 60Hz and 0.5Vrms and swept the other tone from 100 to 20kHz at 0.5Vrms (1Vrms summed in phase).

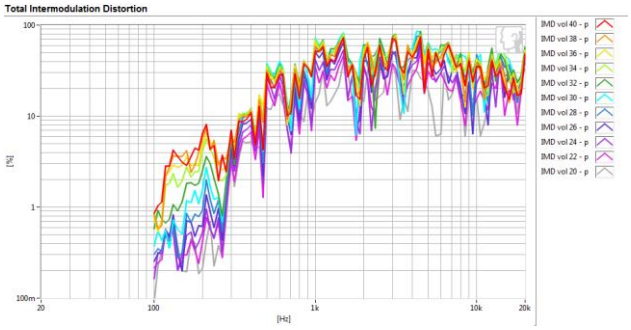

Figure 12. Total IMD vs. Level

If this was a multi-way car audio loudspeaker system with electrical crossovers, one would not expect to see much IM distortion since the low frequency fixed tone would be played out of the low frequency driver and the high frequency swept tone would mostly be played out of the midrange drivers. However, the measurements demonstrate a fair amount of IM distortion centered above1kHz. This is probably because four of the six loudspeakers are full range and trying to reproduce the low and high frequencies simultaneously.

4.1.3 Non-Coherent Distortion using Pink Noise

Pink noise is a great test signal for fast frequency response measurements. It is also a more rigorous test signal for assessing system non-linearities because it produces many frequency components simultaneously and excites both harmonic and intermodulation distortion products. Statistically, it is closer to music than a sine wave or two tones. [9].

The Coherence function is a normalized cross-correlation measurement that determines the degree to which the system output is linearly related to the system input. The signal-to-distortion ratio（SDR）, where the distortion term includes all nonlinear effects and noise in the system can be computed from the coherence. The coherence can be computed for an arbitrary input signal, and is independent of the linear system response once the bias in the estimation procedure has been removed. The non-coherence is simply 100% minus the coherence to calculate distortion in percent. [6]

A 12th Octave Pink Noise from 20 to 20 kHz at 0.5Vrms (2.5Vpeak) was played and the non-coherent distortion measured.

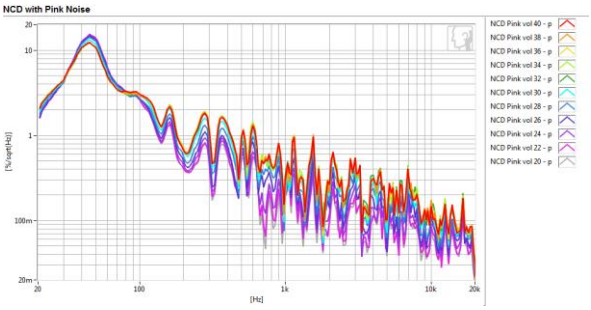

Figure 13. NCD Distortion (Pink Noise) vs. Level

The non-coherent distortion with pink noise consistently increases with volume level above 100Hz. Since pink noise has a higher crest factor than sine waves, it is not too surprising that the distortion is not as high as total harmonic distortion.

4.1.4 Non-Coherent Distortion using Music as the Test Signal

With dual channel analysis, non-coherent distortion can be measured using any stimulus signal including music [8]. In addition to pink noise, non-coherent distortion was measured using the same music selections as in the listening test (Fig. 14, 15 & 17). The music needs to have a broad frequency spectrum in order to get a good signal to noise ratio, and to avoid measuring noise instead of distortion. Although most of the pieces had plenty of low to mid frequencies, many did not have much high frequency content, so data above 10 kHz is not too meaningful.

Non-coherent distortion was measured on the car

audio system playing a 40 second excerpt of Bird on a Wire (BW) by Jennifer Warnes. The average sound pressure level for volume level 30 was 94dBC (fig. 14).

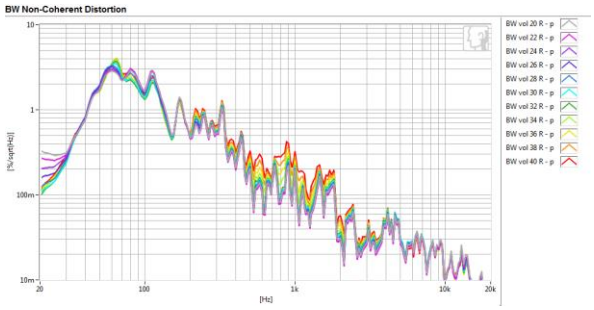

Figure 14. NCD Distortion (Bird on a Wire) vs. Level

Interestingly, the distortion increased with level at 60Hz (around resonance) and between 200 – 2 kHz, the distortion went down from 70 - 200Hz. This is probably because the speaker response is starting to compress the signal at low frequencies and move the energy to higher frequencies as Rub & Buzz. Again, there was a jump in distortion level at volume level 34 and above.

Non-coherent distortion was then measured on the car audio system playing a 30 second excerpt of Cousin Dupree (CD) by Steely Dan. The average sound pressure level for volume level 30 was 98dBC (fig. 15).

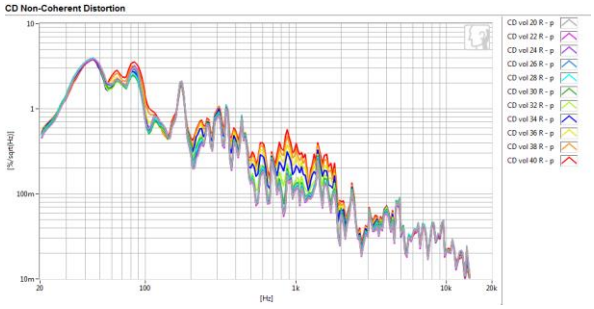

Figure 15. NCD Distortion (Cousin Dupree) vs. Level

This time the distortion increased from 60 – 120Hz but stayed about the same from 120-180Hz. Again, there was a jump in distortion level, especially at volume level 34 and above from 500 – 2 kHz.

Finally, non-coherent distortion was measured on the car audio system playing a 40 second excerpt of American Boy (AB) by Estelle (fig. 16). The average sound pressure level for volume level 30 was a huge 105dBC. This can be explained by looking at the time waveform of the original song (fig. 17) - it looks like a square wave with a very small dynamic range).

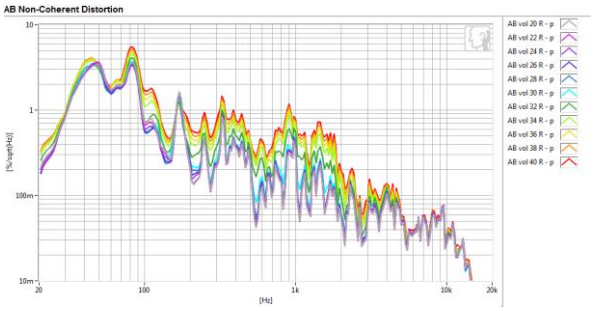

Figure 16. NCD Distortion (American Boy) vs. Level

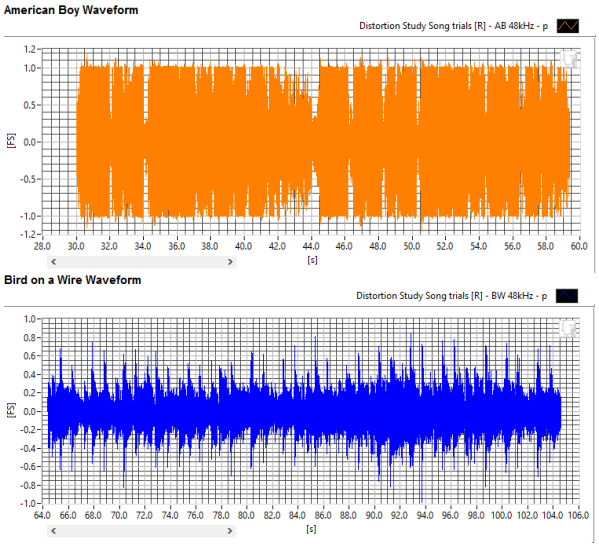

Figure 17. Time Waveform for American Boy compared to Bird on a Wire Songs

NCD distortion measurement produced similar distortion curve shapes as with the previous songs but the distortion jumps at volume level 24 and above. This is probably due to the higher recording level.

5 Discussion

This section focuses on a discussion on the correlations between the objective and subjective measurements of nonlinear distortion in the car audio system.

5.1 Correlations between Objective and Subjective Measurements

The results from the listening tests presented in section 2 indicated that listeners were able discern audible distortion above a given threshold for different songs. For American Boy, that threshold was quite low, from volume level 30 to 32 which corresponds to about 105dBC. For Cousin Dupree, the threshold was higher, volume level 32 which corresponds to 100dBC. For Bird on a Wire, the threshold was yet even higher, from volume level 36 to 38, which corresponds to about 100dBC. It therefore appears that at volume levels of about 100dBC or higher, distortion is clearly audible. But why is American Boy's threshold higher than Cousin Dupree and Bird on a Wire? Is it because American Boy sounds distorted to begin with, as can be seen in its time waveform that looks like a square wave (Figure 16)? Is there a masking affect from the recording obscuring the loudspeaker system distortion?

5.1.1 THD Measurements

The THD measurements (Fig. 10) indicate that harmonic distortion, especially high harmonic distortion jumped above volume 32 (~100dB SPL). This is consistent with the subjective listening tests. It also indicates high levels of distortion throughout the entire frequency range, in particular, 20 -80Hz (around speaker resonance), 150 - 700Hz and 2k - 7kHz. Listeners indicated that they heard most of the distortion in the bass and midrange frequencies.

In summary, there appears to be good positive correlations between the amount of THD measured in the car speakers and distortion audibility.

5.1.2 IM Distortion Measurements

The IM distortion (Fig. 12), increases at volume level 26. The IM distortion level is greatest above 1kHz, in excess of 50%. This corresponds with listeners' comment about audible distortion in the midrange but not as well with the threshold of audible distortion which most listeners heard at a higher threshold level.

5.1.3 Non-Coherent Distortion based on Pink Noise

The pink noise distortion measurements (Fig. 13) show a consistent increase with volume level above 100Hz. The fact that distortion is not increasing as much in low frequencies indicates that it is spreading to higher frequencies. This should indicate that it will be more audible since it is far apart in frequency and will not be masked

5.1.4 Non-Coherent Distortion based on Music

Figure 7, 8, & 10 shows the non-coherent distortion measured for each music track. For American Boy, the jump in distortion level at volume 30 corresponds well with listeners' perceived level of distortion. For Cousin Dupree, the jump in distortion level at volume 32 corresponds well with listeners' perceived level of distortion. For Bird on a Wire, the jump in distortion level at volume 36 corresponds well with listeners' perceived level of distortion.

In summary, among the distortion metrics we evaluated in this study, non-coherent distortion based on music correlates significantly better with listeners' preference ratings than THD, IM and NCD with Pink Noise measurements. It should not

be surprising that this distortion metric produced the highest correlations since it used the same test signal (music) as used in the listening tests. We suspect that even better predictions of audible distortion in music could be achieved if a psychoacoustic model was applied to take into account the masking properties of the music on audibility of distortion. This will be the topic of a future study.

6 Conclusions

The goal of this research project was to determine if there was a consistent distortion level versus frequency at which distortion became audible, regardless of the musical source material. If we take into account the difference in level of the music source material, this car audio system sounds noticeably distorted with most music above 100dBC in the frequency range of 200 – 2kHz and above 0.2% NCD. American Boy, however, with its highly compressed and distorted recording, made it more difficult to discern distortion threshold because it sounded distorted to begin with. In hindsight, we should have taken more care to match the playback levels of the different musical excerpts. Probably matching the long term RMS levels would be sufficient since loudspeakers tend not to demonstrate hard clipping like a transistor amplifier.

Most likely, at the higher volume levels, the car audio system is being pushed to its physical limits and creating lots of high order distortion or Rub & Buzz which is quite audible and annoying. Most of the listeners identified the distortion as coming from the bass and midrange region. This is not surprising since the speaker cone excursion is greatest at resonance, distortion, especially odd order harmonics will be greatest.

We identified that of the 4 distortion metrics measured, non-coherent distortion using music as a stimulus showed the best correlation to human perception.

Further tests are needed to investigate consistency with other car audio systems and other loudspeaker systems including headphones and home speaker systems to confirm that we can use the same non-coherent distortion metric.

Lastly, further research into adapting non-coherent distortion to a perceptual model needs to be investigated. This may allow a common distortion metric across all sound reproduction systems to predict distortion audibility and subjective quality for human hearing.

7 References

[1] S. Temme, S. Olive, S. Tatarunis, T. Welti, and E. McMullin, “The Correlation Between Distortion Audibility and Listener Preference in Headphones”, presented at the 137th Audio Eng. Soc. Convention. Preprint (2014 Oct.)

[2] P. Dennis, “In-Vehicle Audio System Sound Quality Preference Study”, presented at the 139th Audio Eng. Soc. Convention. Preprint 9393 (2015 Oct.)

[3] ITU-R BS.1116-1 Methods for the Subjective Assessment of Small Impairments in Audio Systems Including Multi-Channel Sound Systems

[4] S. Temme, P. Brunet, and Parastoo Qarabaqi, “Measurement of Harmonic Distortion Audibility Using a Simplified Psychoacoustic Model,” presented at the 133rd Audio Eng. Soc. Convention. Preprint 8704 (2012 Oct.).

[5] S. Temme, "Why and How to Measure Distortion in Electroacoustic Transducers", presented at the AES 11th Conference on Audio Test and Measurement, Portland, Oregon (29-31 May, 1992).

[6] J. Kates, “Cross-correlation procedures for measuring noise and distortion in AGC hearing aids” J. Acoust. Soc. Am., Vol. 107, No. 6, June 2000.

[7] M. Schetzen, "The Volterra & Wiener Theories of Nonlinear Systems", John Wiley, New York, 1980.

[8] S. Temme, P. Brunet, and D. B. (Don) Keele, "Practical Measurement of Loudspeaker Distortion Using a Simplified Auditory Perceptual Model," presented at the AES 127th Convention, preprint 7905, (2009 October).

[9] E. Czerwinsky, A. Voishvillo & Al., "Multitone Testing of Sound System Components- Some Results and Conclusions- Part 1: History and Theory", JAES, Vol. 49, No, 11, (2001 November).

[10] E. Czerwinsky, A. Voishvillo & Al., "Multitone Testing of Sound System Components- Some Results and Conclusions- Part 2: Modeling and Application", JAES, Vol. 49, No, 12, (2001 Dec.).

[11] S. Temme and P. Brunet, "A New Method for Measuring Distortion using a Multitone Stimulus and Non-Coherence," presented at the AES 121st Convention, preprint 6877, (2006 Oct.).

[12] S. Temme, "How to Graph Distortion Measurements" presented at the 94th Audio Eng., Soc. Convention, preprint 3602 (1993 March).

[13] A. Voishvillo, A. Terekhov, & Al., "Graphing, Interpretation, and Comparison of Results of Loudspeaker Nonlinear Distortion Measurements", JAES, Vol. 52, No. 4, (2004 April).