An Assessment of SEA Model Quality

Mark J. Moeller, **Robert S. Thomas** and **Robert E. Powell**, Ford Motor Company, Dearborn, MI, USA

Statistical Energy Analysis (SEA) models are routinely being adopted in up-front automotive sound package design. SEA models serve two important functions. First they provide a means of assessing noise and vibration performance relative to absolute targets. Secondly, they are used to assess various alternative designs or changes required to meet targets. This article addresses how to objectively evaluate both the absolute and relative predictive capability of SEA models. The absolute prediction is assessed using a hypothesis test to determine membership of the analytical prediction relative to a set of test data. The relative prediction is assessed using hardware-designed experiments to estimate design sensitivities. Both have been found useful to drive model improvement efforts. Being able to objectively document model capability also improves the credibility of SEA model predictions and the design information they deliver.

Competitive cost pressures are forcing design/manufacturing concerns to move from hardware prototype based design strategies to Computer Aided Engineering (CAE) based alternatives. Statistical Energy Analysis (SEA) has been widely adopted as a means of assessing high frequency noise, vibration and harshness (NVH) concerns in the automotive industry. This article addresses how to objectively evaluate the predictive capability concerns assessing the feasibility of using SEA models for this purpose.

The quality of CAE models is becoming increasingly important as availability of test information decreases. Several activities have grappled with this issue. The most extreme example would be the nuclear weapons arena, which is being forced to move to a completely CAE based process due to the voluntary nuclear test ban in place. There are active discussions in the literature on determining the quality of many modeling technologies. Computational Fluid Dynamics has an especially active discussion ongoing as evidenced by Roache.[1] Structural safety has similar model quality discussions that can be found in the NAFEMS SAFESA material[2,3] for finite element analysis. Han[4] discusses crash safety model quality. Some organizations have adapted ISO 9001 as a quality system for CAE in their organizations.

This article is a continuation of the SEA model quality discussion started by Thomas.[5] It will incorporate efforts toward adapting NVH CAE Quality Metrics, described by Moeller,[6] to high frequency models. The desired result is a framework for objectively assessing suitability to task for high frequency modeling technology. This can then be integrated into an overall quality framework for SEA modeling.

The first section develops a metric for assessment of baseline model correlation for transfer function prediction. The next section applies the metrics to current SEA model capability. SEA model design capability is examined next. The last section presents the discussion and conclusions.

Baseline SEA Model Quality Metric Development

The question addressed by baseline model quality is "Are these transfer functions (FRFs) representative of test?" A hypothesis test was shown to be useful for deterministic models in Moeller.[6] This test has been extended to statistical models here. The hypothesis test is examined here to examine the assumptions and present supporting data.

The following notation was adopted in the initial investigation. For the reference (test) set:

i = vehicle with $i = 1,\ldots,n$
j = frequency with $j = 1,\ldots,J$
k = particular FRF with $k = 1,\ldots,K$
$x_{i,j,k}$ = amplitude in dB for vehicle i, frequency j, FRF k.

In this study the number of reference vehicles n was nine. For the other vehicles in the comparison set:

j = frequency with $j = 1,\ldots,J$
k = particular FRF with $k = 1,\ldots,K$
$y_{j,k}$ = amplitude in dB for frequency j, FRF k.

Assumptions.

1. $\{x_{i,j,k}\}$ are mutually independent
2. $\{x_{i,j,k}\}$ are normally distributed with mean $m_{j,k}$ and standard deviation $s_{j,k}$
3. The standard deviations at different frequencies are equal for a specific FRF: $s_{j,k} = s_k$
4. $\{y_{j,k}\}$ and $\{x_{i,j,k}\}$ are independent of each other
5. $\{y_{j,k}\}$ are mutually independent
6. $\{y_{j,k}\}$ are normally distributed with mean $\nu_{j,k}$ and standard deviation $t_{j,k}$
7. The standard deviations at different frequencies are equal for a specific FRF: $t_{j,k} = t_k$
8. The standard deviations for the vehicles and the test vehicles are equal: $t_k = s_k$ for $k = 1,\ldots,K$

Hypotheses. The test conducted is of the null hypothesis that the FRF of a test vehicle is the same as those for the reference set versus the alternative that it is different. Given the above assumptions, this corresponds to testing the means: H_0: $\mu_{j,k} = \nu_{j,k}$ for all j and k versus H_1: $\mu_{j,k} \neq \nu j,k$ for some j and k

Test Statistic.

$$D = \sum_{k=1}^{K} D_k$$

$$D_k = \sum_{j=1}^{J} d_{j,k}$$

$$d_{j,k} = \left(\frac{n}{n+1}\right)\left(\frac{\left[y_{j,k} - \bar{x}_{\bullet,j,k}\right]^2}{s^2_{\bullet,\bullet,k}}\right)$$

where

$$x_{\bullet,j,k} = \frac{1}{n}\sum_{i=1}^{n} x_{i,j,k}$$

$$s^2_{\bullet,j,k} = \frac{1}{n-1}\left[\left(\sum_{i=1}^{n} x^2_{i,j,k}\right) - n\bar{x}^2_{\bullet,j,k}\right]$$

$$s^2_{\bullet,\bullet,k} = \frac{1}{J}\sum_{j=1}^{J} s^2_{\bullet,j,k}$$

$\bar{x}_{\bullet,j,k}$ and $s_{\bullet,j,k}$ are the sample mean and sample standard deviation, respectively, of the n vehicles in the reference set at the j^{th} frequency and the k^{th} FRF. $s_{\bullet,\bullet,k}$ is the pooled standard deviation, and it estimates the common s_k for the k^{th} FRF. $n \times J$ observations are used to estimate s_k. Because this is a large number, we will assume that the pooled standard deviation is nearly equal to σ_k. Then, under H_0, D_k has a chi-squared distribution with J degrees of freedom, and D has a chi-squared distribution with $J \times K$ degrees of freedom.

In addition to testing the comparison set to the reference, the methodology was also checked by jack-knifing the reference data, i.e. comparing each sample to the remaining eight vehicles in the reference set. For the u^{th} vehicle, the mean and standard deviations of the other eight vehicles were used in the formula for D.

Decision Rule. H_0 is rejected in favor of H_1 for large values of the chi-squared statistic. Define α, the level of significance,

Based on Paper #2001-01-1624, "An Assessment of SEA Model Quality," © 2001 SAE International, presented at the Noise & Vibration Conference, April 30-May 3, Traverse City, MI, 2001.

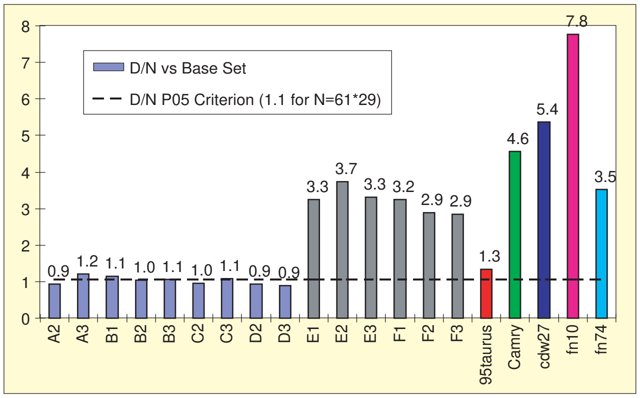

Figure 1. Quality metrics for acoustic reciprocity test, normalized D (D/N) with mid-size vehicle test as base set, 61 frequency response functions, 400-2000 Hz, 1/12 octave.

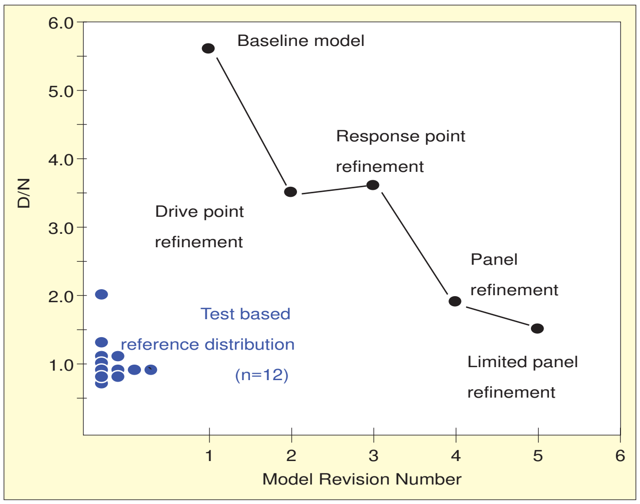

Figure 2. Tracking chart for model updating.

to be the probability of rejecting H_0 given that it is true. The typical value of α used in this study is 0.05. Then H_0 was rejected when the chi-squared statistic was larger than $(1 - \alpha) \times 100$ percentile of the chi-squared distribution.

Validation for High Frequency Data. High frequency variability data were acquired to validate the metric for use with SEA models. The data were acquired using reciprocity[7] for a frequency range from 400 Hz to 2 kHz on a series of vehicles. The source positions were driver's ear acoustic space and seat track. The response points were distributed over the attachment points to the body. Nine nominally identical vehicles were used to construct a reference set. Six nominally identical vehicles and four other vehicles were used to create a comparison set.

The data were checked by jack-knifing. One data set was compared to a new reference set constructed from the remaining eight and checked for membership. This was done for each of the nine vehicles in turn and is shown in Figure 1. Also shown in Figure 1 are the results of comparing the nine vehicle reference set to data from vehicles that were from other vehicle lines. As can be seen in the figure, good discrimination is achieved except for the vehicle that is the sister vehicle of the reference set. The reference set is significantly different than the comparison set, strongly suggesting that the baseline metric is appropriate for high frequency data and for SEA models in particular.

The assumptions in the baselining metric were further examined using vehicle FRF data reported in the literature by Bernhard and Kompella.[8,9] The results are discussed in Appendix A.

SEA Model Baselining to Assess SEA Model Quality

The baseline metric was used to evaluate and improve an SEA model of a midsize vehicle for structureborne sound FRFs

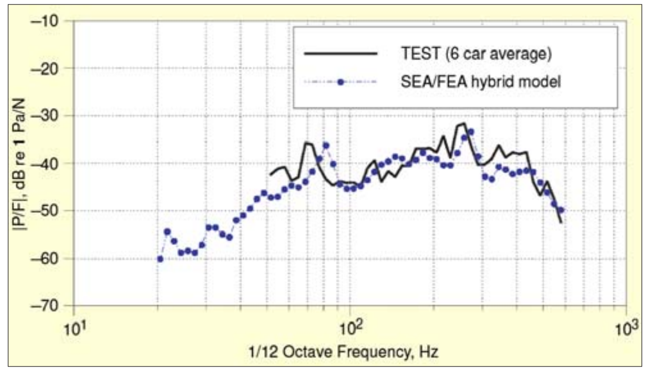

Figure 3. Best 100-400 Hz FRF after model updating.

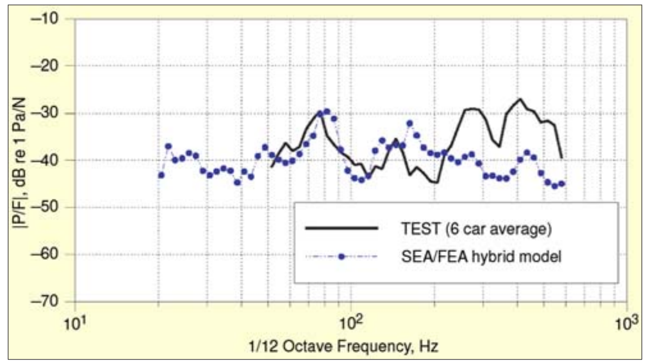

Figure 4. Worst 100-400 Hz FRF after model updating.

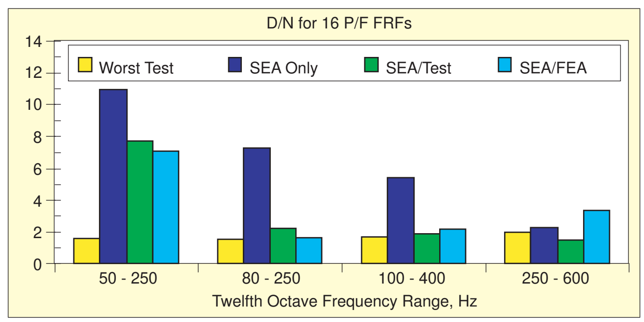

Figure 5. Frequency performance of modeling technologies.

(Figure 2). The first refinement in the model was to improve the drive points as described by Manning.[10,11] This resulted in a significant improvement because the input power to the system was then more accurate. The change to the output subsystems, in this case, did not significantly change the model correlation. A second change, which was significant, was the updating of panel subsystems. The updated SEA model was able to predict the response to within the spread of the test data, although the final model would still not be recognized as a member of the test set.

Example frequency response functions are shown in Figures 3 and 4. Shown are examples of best and worst agreement out of 16 FRFs to demonstrate the range of results.

Another application of the baseline metric is the determination of useful frequency range. Two modeling technologies are represented in Figure 5 – analytical SEA and a hybrid SEA with external information from testing or from finite element analysis (FEA) to update the SEA subsystems. Also shown in Figure 5 is the worst test from the jack knife experiment. As can be seen from the chart, the SEA model by itself is strictly a high frequency tool and it diverges from the test the lower in frequency it is used. Both the test and finite element updating of the SEA model extend its useful frequency range to lower frequencies. Also of note is the fact that the finite element updating starts to diverge as higher and higher frequencies are used. Extension of the D/N baseline metric from FRF to operating spectra is covered in Appendix B.

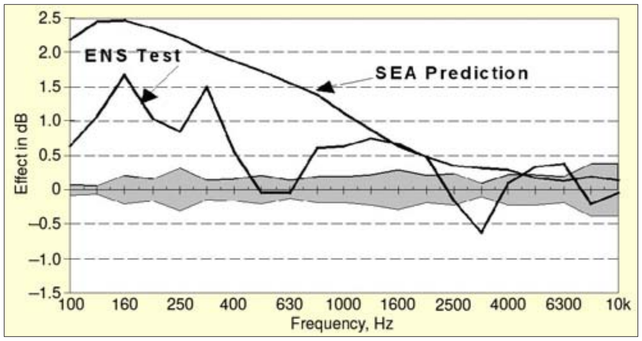

Figure 6. Engine noise simulator designed experiment, dash doubler effect at driver head.

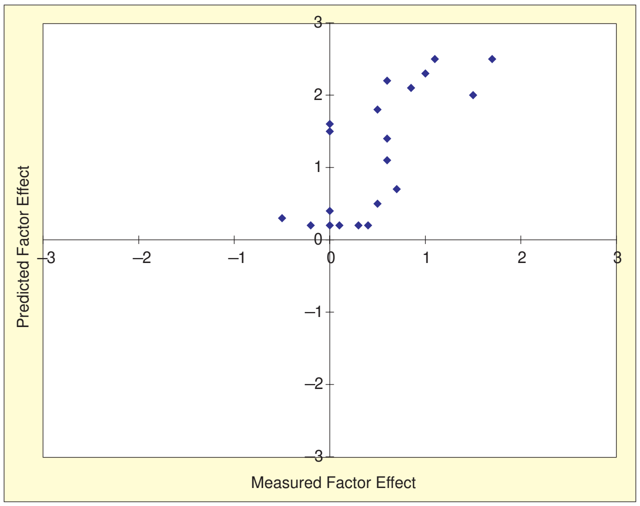

Figure 7. Factor effect plot for test versus SEA model prediction.

Assessing Design Capability

In the Thomas[5] discussion of quality of SEA models, several dimensions of model quality were defined and discussed. Under the dimension of usability were two subticks related to how well the model predicted changes in the response due to changes in design variables. The first subtick had to do with the ability of the model to predict idealized/gross changes in the hardware. The second assessed its suitability to detail design. The article went on to propose using designed experiments to validate the design capability of the SEA models.

Using designed experiments to assess SEA model design capability was reported in Pan[12] for a midsized passenger vehicle and in Huang[13] for a sport utility vehicle. The plot from Huang[13] is shown in Figure 6. This figure shows the measured effect from a hardware designed experiment and the same effect predicted by the SEA model. Also shown in the graph is the gray shaded region of uncertainty in the hardware experiment where one could not conclude that the measured effect was statistically significant.

It is possible to show these data as a scatter diagram by plotting the measured effect against the SEA effect as is done in Figure 7. The desired outcome would be a one to one correspondence between the SEA model and the physical test. That would represent the model reproducing both the direction and amplitude effects of the test. An alternative means of checking design capability is to conduct a series of one factor at a time experiments. Results achieved by Wang et al[14] are discussed in Appendix C.

Conclusions

1. The 'baseline' quality metric proposed for deterministic model assessment is valid for statistical model assessment.
2. Design assessment capability can be done for statistical models using either designed experiments or several one factor at a time experiments.
3. The proposed metric and design assessment capability were used to assess proposed model improvements and to assess applicable frequency ranges of modeling techniques.

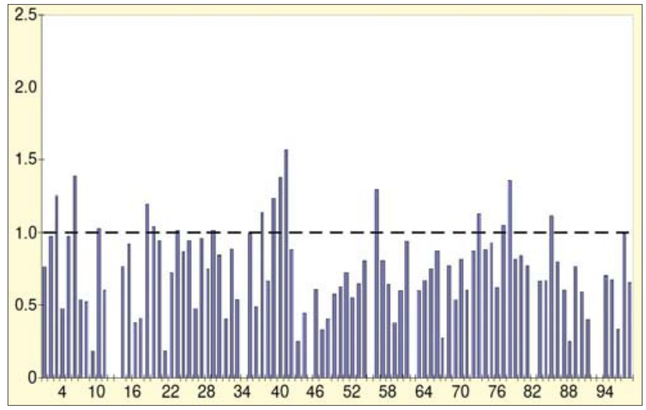

Figure A1. Preview metric values of 98 Isuzu Sport Utility Vehicles, 1 air-borne FRF, frequency range 100-500 Hz, 1/12 octave band.

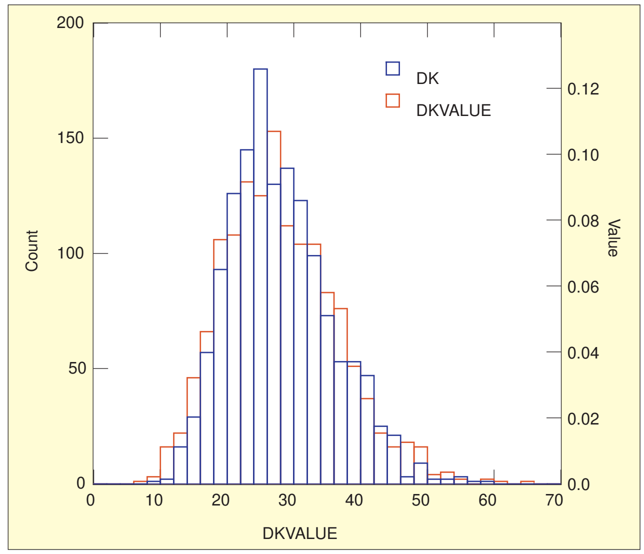

Figure A2. Structure-borne data 100-500 Hz, extreme outliers removed.

Acknowledgements

We would like to acknowledge the work done on these projects by Mr. Frank (Guoming) Wu during his tenure at Ford. Professor Peter Lenk at the University of Michigan was instrumental in developing the metric theory.

References

1. Roache, P. J., *Verification and Validation in Computational Science and Engineering*, Hermosa Publishers, ISBN 0-913478-08-3, Albuquerque, NM, 1998.
2. NAFEMS QAWG, "Quality System Supplement to ISO 9001 Relating to Finite Element Analysis in The Design and Validation of Engineering Products," Issue 1.3, 2 July 1993.
3. The SAFESA Consortium, "SAFESA Technical Manual to Structural Qualification Supported by Finite Element Analysis," published by NAFEMS, Ref. R0041.
4. Han, R. M., "Repeatability of Full-Scale Crash Tests and a Criteria for Validating Simulation Results," in Transportation Research Record No. 1528, Transportation Research Board, National Academy of Sciences, Washington, DC, January 1996.
5. Thomas, R. S., Pan, J., Moeller, M. J. and Nolan, T., "Improving SEA Models Using Quality Technology," *Noise Control Engineering Journal*, January, 1997.
6. Moeller, M. J., Thomas, R. S., Chen, S. E., Chandra, N. and Lenk, P., "NVH CAE Quality Metrics," SAE Noise and Vibration Conference, SAE Paper 1999-01-1791, Traverse City, MI, May 1999.
7. Gu, P., Thomas, R., Lee, J., Kuo, E., Geck, P. and Tao, D., "Measurement Quality of Acoustic Reciprocity Testing For Vehicle Structure Vibro-Acoustic Evaluation," Noise-Con Proceedings, Ypsilanti, MI, April, 1998.
8. Bernhard, R. J. and Kompella, M. S., "Measurement of Statistical Variation of Structural-Acoustic Characteristics of Automotive Vehicles," Purdue University Report HL 92-3, March, 1992.
9. Kompella, M. and Bernhard, R. J., "Measurement of the Statistical

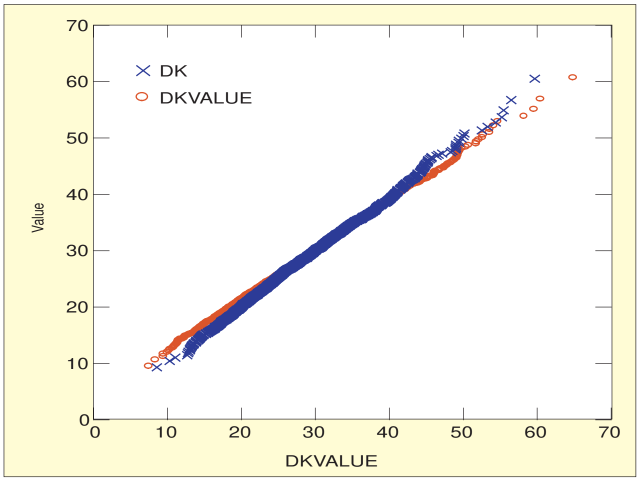

Figure A3. Structure-borne data 100-500 Hz, extreme outliers removed.

Variation of Structural-Acoustic Characteristics of Automotive Vehicles," SAE Noise and Vibration Conference, 931272, Traverse City, MI, 1993.

10. Manning, J. E., "Validation of SEA Models Using Measured Modal Power," SAE Noise and Vibration Conference, SAE Paper 1999-01-1703, Traverse City, MI, May 1999.
11. Manning, J. E., "Use of Measured Mobilities to Improve SEA Models in the Mid Frequency Range," ASME Paper DETC99/VIB-8196, Las Vegas, NV, September 1999.
12. Pan, J., Thomas, R. S. and Moeller, M. J., "Verifying Vehicle SEA Model Predictions for Airborne Noise Transmission Using Designed Experiments," inter-noise, Liverpool, UK, April 1996.
13. Huang, X., Moeller, M. J., Lee, J. and Powell, R. E., "Application of Statistical Energy Analysis (SEA) to the Development of a Light Truck Sound Package," IUTAM Symposium on Statistical Energy Analysis, July 1997.
14. Wang, D., Goetchius, G. and Onsay, T., "Validation of a SEA Model for a Minivan: Use of Ideal and Structure-Borne Sources," SAE Noise and Vibration Conference, SAE Paper 1999-01-1697, Traverse City, MI, May 1999.
15. Bharj, T., Lee, J. J., DeJong, R. and Keller, T., "Accuracy of Statistical Energy Analysis (SEA) Model to Evaluate Vehicle Noise Concerns," inter-noise, Fort Lauderdale, FL, Dec. 1999.

Appendix A – Metric Corroboration

The population model proposed in the baseline metric is consistent with the assumptions associated with SEA modeling. The vehicles being independent is a necessary assumption. The transducer locations being independent is reasonable as long as they are more than half a wavelength separated from each other. Dealing with broadband data, typically third octave, helps reduce serial correlation in the data. Standard deviations across frequency and amplitudes normally distributed result in the gross error being chi-square distributed. The standard deviations can be control charted versus frequency to verify that they are equal.

Bernhard and Kompella[8,9] performed a classic experiment to observe the variability of structural acoustic transfer functions of vehicles. They produced a very rich data set with four transfer functions with 57 and 99 observations for an Isuzu pickup truck and sport utility vehicle, respectively. The raw transfer function data were received from the authors' reports on Isuzu variability. These data were processed into 1/12th octave bands and used to evaluate the proposed baseline metric scheme.

The data represented a structure-borne transfer function and an airborne transfer function for nominally identical vehicles. The full data sets were used to compute the pooled standard deviations. The metric was first used to screen the data and the most extreme outliers were removed. A sample screening is shown in Figure A1. Then each car within a set was compared to every other car in the data set. A random trial with the same number of realizations as the test data set was generated for a chi-square distribution and the histograms and probability plots were generated for both data sets and compared in Figures A2 through A5. The data with the outliers removed does tend to follow a chi-square distribution.

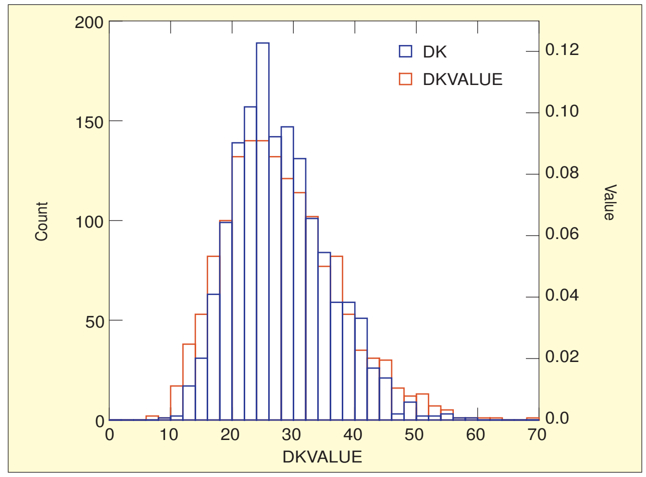

Figure A4. Air-borne data 100-500 Hz, extreme outliers removed.

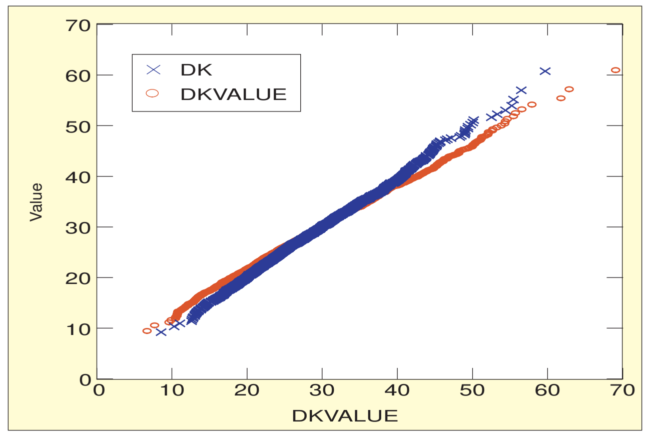

Figure A5. Air-borne data 100-500 Hz, extreme outliers removed.

Appendix B – Operating Vehicle Metric

This Appendix develops the capability to objectively assess operational vehicle data. NVH CAE vehicle system models are routinely exercised for target setting, conformance demonstration and problem resolution. Both deterministic and statistical vehicle system models are used. Developing predictive capability for road noise was one key element of NVH CAE analysis capability. It was desired to demonstrate techniques for objective assessment of NVH CAE model performance for road noise.

The ability to classify cars based on their frequency response functions has already been demonstrated in this article. This was accomplished using a hypothesis test methodology to assess whether or not a sample function belonged to a reference set. The methodology is extended to road noise here. Two experiments were run by two different test groups both in support of this project. The two tests had different but supporting roles. The first test was to collect variability data on a fleet of cars for both deterministic and statistical model validation. Both data from reference and comparison set vehicles were gathered. This allowed the metric to be checked by jack-knifing the data and then comparing the results to those of the comparison set in a manner exactly analogous to the FRF metric. The second data set was used to develop the load cases for exciting the model.

There were two frequency ranges investigated for two road surfaces at one speed each. The deterministic frequency range was from 20 to 250 Hz, and the statistical frequency range was from 100 Hz to 2 kHz. The two road surfaces were a coarse aggregate surface at 30 mph and smooth road at 50 mph.

Sample data for model validation are shown in Figure B1. These results are for Driver's Right Ear for the reference set mean and expected variation and for the comparison set. One of the most notable changes is that the standard deviation for

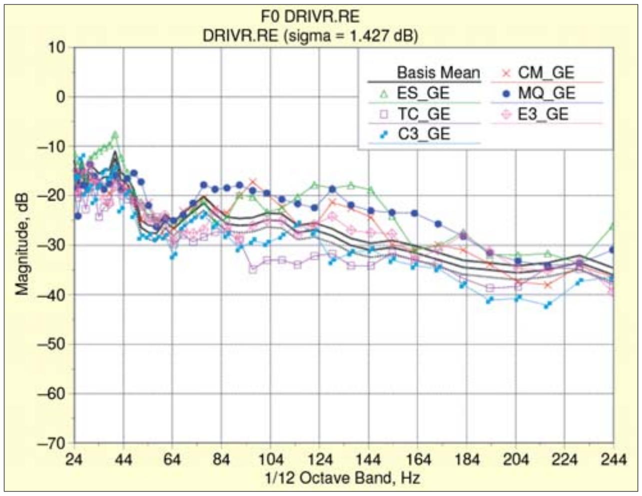

Figure B1. Low frequency road noise.

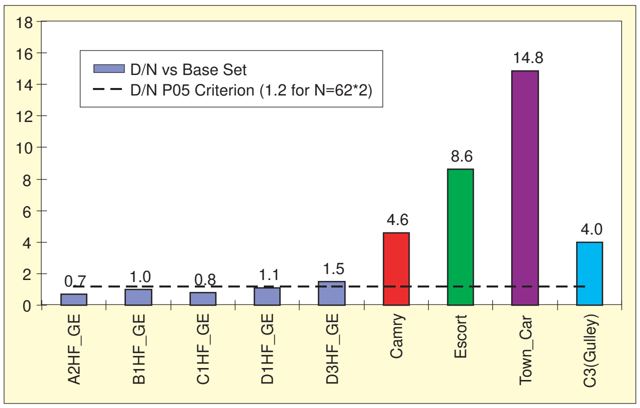

Figure B2. Quality metrics for vehicle test on Glen Eagle @ 30 mph, D/N for base set and comparison set, 2 interior responses, 63-2000 Hz, 1/12 octave.

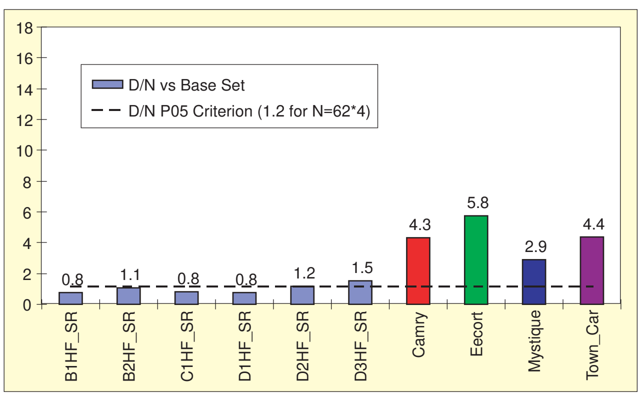

Figure B3. Quality metrics for vehicle test on smooth road @ 50 mph, D/N for base set and comparison set, 4 interior responses, 63-2000 Hz, 1/12 octave.

the response is significantly smaller than those determined for frequency response functions. In this case, the standard deviations are between 1 dB and 2 dB. It is because of this small standard deviation that it is possible to discriminate different vehicles based on their road noise signatures. From the plot in Figure B1, it is possible to see that the luxury car is the quietest and that the small car is the loudest.

The metric was adapted to handle response spectra as well as transfer functions and an effort to discriminate based on the average squared distance from the mean was checked. Prior to full analysis, the data were screened using the metric. As in any large scale experiment, there were some discrepancies found that could not be resolved. When this occurred the questionable data were either excluded from further analysis or gain corrected and included. In this case, the basis set consisted of

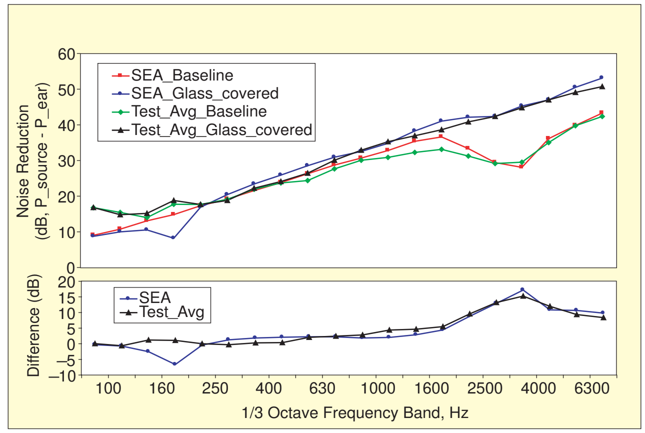

Figure C1. Effect of covering all the glass surfaces with sound barrier, source – reverberation room excitation, response – driver's head space.

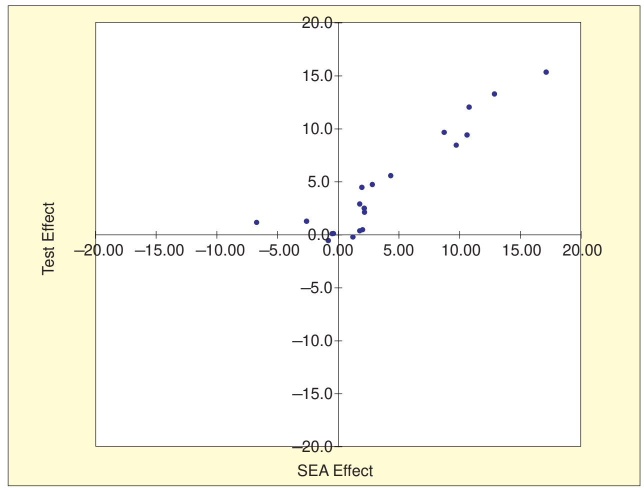

Figure C2. Cross plot of measured effect versus SEA predicted effect.

seven midsize vehicles tested by an outside vendor. Two events were investigated – Glen Eagles 30 mph and Smooth Road 50 mph. The comparison set included tests of midsize replacement, Camry (CM), Escort (ES), Mystique (MQ) and Town Car (TC), and CAE NVH Lab testing of another midsize vehicle with special processing to remove impact events. Because of the different data handling, the CAE NVH Lab tested vehicle was not directly comparable to the fleet data.

There were 11 interior response locations – two for the driver's right ear and rear seat centerline sound, three for steering wheel X, Y, & Z vibration, and six for inboard and outboard seat track triaxial vibration. Data were acquired from 20 to 250 Hz in 1/12 octave bands. The standard deviations were found to be lower than for FRFs – 1.5 dB for sound, 3.1 dB steering wheel vibration and 1.7 dB seat track vibration.

Statistical Model Validation Data. The same reference set vehicles and comparison set vehicles were used for the statistical model validation as for the deterministic model validation. The transducer line up and frequency ranges were different. There were four interior response locations – two for the driver's right ear and rear seat centerline sound and two for inboard and outboard seat track triaxial vibration. The frequency range was 63 to 2000 Hz in 1/12 octave bands. The estimated standard deviations were 1.54/1.58 dB sound (coarse aggregate/smooth road) and 2.14/2.32 dB seat track vibration (coarse aggregate/ smooth road).

As can be seen in Figure B2, the classification based on road noise responses based on a hypothesis test can determine membership in a set. As before for transfer function data, there is a larger than expected rate of misclassification of the reference set vehicles based on a chi-square parent population assumption. However, there is still good discrimination among vehicles in the comparison set. The low variance of the response improves the ability to discriminate among vehicles.

The results for a smooth road are shown in Figure B3. The same trends are evident in the data; a higher than expected misclassification rate but a good capability for discrimination.

Appendix C – One-Factor-at-a-Time Results

The designed experiment method of determining factor effects is not unique nor widely adopted. Wang[14] chose to accomplish a similar model verification using one factor at a time experiments in a reverberation room. The Noise Reduction between the interior of the car and reverberation room was measured for various configurations. The baseline was measured and then various treatments were applied to the roof, fixed glass and doors, and the noise reduction was remeasured. The authors graciously provided their data for analysis here. The test and SEA data after treatment was subtracted from the corresponding data before treatment, resulting in the difference chart in Figure C1. The test effects are plotted versus the SEA effects in Figure C2. The SEA data are clearly predictive, achieving a similar level of validation as achieved in the designed experiment of Figure 7.

Bharj[15] presented a similar one factor at a time study detailing the predictability of an SEA model. The model baseline correlation was demonstrated and then design effects were evaluated. The two design effects shown in that paper were an outer tunnel absorber and a front fender insulator.

In both cases, designed experiments and large one factor at a time experiments, it can be seen from the effect plots that the SEA models are capable of predicting the effects of physical changes to the hardware. When the model suggested a particular design variable was effective at changing the response, that predicted response was observed in the test. S/V

The authors may be contacted at: mark.j.moeller@boeing.com, rthoma11@ford.com and rpowell7@ford.com.