ABSTRACT

Customer feedback indicated that there was a difference in noise levels and frequency content between nominally identical diesel engines manufactured in two different plants. A test program was created to determine whether this perceived difference was real. This program opened the opportunity to investigate several levels of variability in SAE J-1074 noise measurements. From the data, estimates of confidence intervals were developed to predict true mean noise levels in the presence of back-to-back, install-test-remove, and sample-to-sample test variability. The data from this project can be used to estimate the number of tests required in order to achieve a given level of accuracy in estimating the true mean noise level of an engine or of an engine population. The results can also be used to design a test program that will provide a given level of confidence in determining whether there is a statistically significant difference between two engine populations.

INTRODUCTION

This project was created to determine whether a noise level difference existed between nominally identical engines built in two different plants. Many of the components used by the two plants, while nominally identical, came from different suppliers. The SAE J-1074 engine sound level measurement procedure (Ref. 1) was used to determine the noise level of engines from the two plants. The first issue facing the project was that no published information was available on the expected level of variability of the SAE J0174 noise test itself, or of diesel engine populations tested according to J1074. Without information on variability, it is difficult to know whether measured differences should be attributed to real differences between the engines or to the variability of the test itself. There is a good reason for this shortage of information: acquiring the necessary data is an expensive and time-consuming process.

Several types of variability can affect the results of SAE J0174 noise tests:
- Back-to-back test variability
- Install-test-remove cycle variability
- Variability within an engine population
- Variability between two engine populations

SAE J-1074 noise tests are also routinely used in engine development programs to make engineering decisions about the NVH impact of proposed changes or features. An understanding of the accuracy and reliability of J-1074 results is important to ensure that decisions are made based on statistically significant data. During development, many of the changes to be evaluated may result in only a small noise difference. However, several small changes can add up over time, making it important to accurately determine each small change in engine noise level. Given a sound, statistically founded understanding of these different types of variability, it is possible to answer questions such as:

1. If an engine is modified, how small of a noise difference can be reliably detected with a single test?
2. How can repeat tests before and after modification be used to improve the accuracy of and confidence in results?
3. How well do one or two sample engines represent a population?
4. How many test repeats are required to detect a real difference between two populations of engines with 90% confidence, even if the difference is as small as 0.2 dB?

A previous SAE paper examined these questions in some detail for the J-366 driveby noise test (Ref. 2). This paper will use a similar approach to evaluate the J-1074 engine noise test.
THE SAE J-1074 TEST

The SAE J-1074 procedure specifies the test environment, engine installation details, microphone positions, and more. It specifies that a sound level meter and octave filter set shall be used to measure the sound levels. As an option, tape recorders or graphic level recorders may be part of the instrumentation chain. The procedure specifies three microphone locations, on the left, front, and right sides of the engine at exhaust manifold height.

J-1074 calls for noise measurements to be made at a few important steady state engine operating conditions. The sound level meter or other instruments are set to slow response. The measurement is repeated and the reported result is the average of two measurements, provided they are within 2 dB of each other. Octave band data are only required for the microphone position giving the highest overall A-weighted level. The fact that two results are averaged together (provided that they are within 2 dB of each other) indicates that the authors of the standard expected a fair degree of test-to-test variability in the sound level measurement results.

Users in industry tend to take liberties with the J-1074 procedure. For example, sound level meters are rarely used to acquire test cell noise data today, and the remaining graphic level recorders are mostly in museum collections. It is far more common for the microphone signals to be fed into a computer-based signal analysis system. For steady state measurements, this allows averaging of results over much longer periods than the slow response setting of a sound level meter, which represents roughly a 1 second average. 20 to 60 second averaging times for steady state measurements are common today. Therefore, it has become unusual to repeat a steady state test and average the results.

Signal analysis systems also allow more information to be extracted from the microphone signals. Narrow band and 1/3-octave spectra are commonly calculated for all microphone positions. Many users also add an additional microphone position above the engine, and a few users add more microphones in other locations. Another commonly performed test, which is beyond the scope of J-1074, is to measure engine noise during a speed sweep, rather than or in addition to steady state measurements. This allows the engine to be evaluated over a wide range of operating conditions, revealing resonance issues and other noise characteristics that may not be apparent during steady state operation.

The test program described here deviates to some degree from the standard J-1074 test in ways that are typical for many users of the test. In this test program, 4 microphones are used: left, front, right, and top. As is common, the overall levels from the 4 microphones are averaged to produce a single reported level. Speed sweeps were added to the test program, and they are the focus of the statistical analysis.

TEST PROCEDURES

For this program, all the engines were tested in the same semianechoic test cell, which has a cutoff frequency of 200 Hz. The engines were mounted on production vehicle mounts. No transmission was used, so the flywheel was directly connected to the dynamometer by a driveshaft. The flywheel housing was enclosed with an acoustical barrier material to prevent noise from the flywheel from influencing the result. Intake air was drawn from outside the cell through an acoustically treated duct, and the exhaust was routed out of the cell through a pipe wrapped with acoustic absorption and barrier material. Discharge air from the turbocharger was routed through a floor-mounted water-to-air cooler, which simulated the performance of the air-to-air aftercooler used in the vehicle. The aftercooler ductwork was wrapped with acoustical barrier and absorption material. All tests were conducted with the FEAD brackets in place, but without accessories. The test cell setup was designed to ensure that intake, exhaust, driveline, dynamometer, and engine accessory noise did not influence the results.

B&K Type 4190 ½ inch free field microphones were used with Type 2669 preamplifiers and Type 2609 amplifiers. The amplified microphone signals and an engine tachometer signal were fed into an HP VXI data acquisition system. Digital recordings of the data were made to disk using an LMS CADA-X Throughput Acquisition module. LMS CADA-X software was used for all data analysis, with Excel being used for the statistical analysis and plotting.

Sound pressure data were acquired under the following test conditions for all the engines:

- Full-load speed sweep from 700 to 3400 RPM
- No-load speed sweep from 700 to 3400 RPM
- Full-load steady state at 1000, 1500, 2000, 2500, 3000, and 3300 RPM
- Part-load steady state at 2300 RPM and 155 Nm torque
- No-load steady state at idle (680 RPM)

For each steady state test condition, a single set of data averaged over 30 seconds was acquired. The speed sweeps were carried out at a rate of approximately 60 RPM per second. Sound pressure levels were calculated at each 50 RPM increment, so the level at each speed increment represents less than 1 second of data averaging. To reduce the variability caused by such a short averaging time, each sweep was repeated 3 times and the results averaged.

The original plan was to test 6 engines from each of the two manufacturing plants. Unfortunately, 3 engines from Plant A were tested, along with 4 engines from Plant B. One engine from each plant was tested a total
of 3 times each, to evaluate the variability of the install-
test-remove cycle.

STATISTICAL ANALYSIS METHODS

Standard statistical analysis methods were used to
analyze the data. The equations are taken from Ref. 3.
To determine the average value of a set of noise levels,
the levels, which are in dB, must be logarithmically
averaged. The standard deviation of a set of levels can
then be determined by the equation

$$S_x = \sqrt{\left(\frac{x_1 - \bar{x}}{n-1}\right)^2 + \left(\frac{x_2 - \bar{x}}{n-1}\right)^2 + \cdots + \left(\frac{x_n - \bar{x}}{n-1}\right)^2}$$

(1)

Next, a confidence interval can be calculated to
determine the range of possible true values represented
by the test results. In this study, the 90% confidence
interval was calculated. Based on the limited number of
data samples available, there is 90% confidence that the
true average of the population would fall within the
confidence interval. The confidence interval, CI, is given
by the equation

$$CI = \bar{x} \pm t_{\alpha} \frac{S_x}{\sqrt{n}}$$

(2)

where $\bar{x}$ is the average of the test results, $t_{\alpha}$ is the t-
value, $S_x$ is the standard deviation of the data set, and $n$
is the number of data samples. The t-value is
dependent on the degree of freedom of the population
of data samples. The degree of freedom, $v$, is equal to
the number of data samples minus 1, or ($v = n-1$). Knowing
the number of data samples, the t-value can be taken
from a statistics table such as the one in Ref. 3. For
example, for a population of 5 data samples,

$$CI = \bar{x} \pm 2.132 \frac{S_x}{\sqrt{5}}$$

(3)

In order to compare two populations, a combined degree
of freedom $v$ must be determined. In our case, with 5
ingines from Plant A and 4 from Plant B,

$$v = \frac{\left(S_{x_A}^2 / 5\right) + \left(S_{x_B}^2 / 4\right)}{\left(S_{x_A}^2 / 5\right) / 4 + \left(S_{x_B}^2 / 4\right) / 3}$$

(4)

where $S_{x_A}$ is the standard deviation of the data from
Plant A engines, and $S_{x_B}$ is the standard deviation of
Plant B engines. Next, the t-values are calculated from
the average and standard deviation of noise levels for
each engine population at each test speed using the equation

$$t = \frac{\bar{x}_A - \bar{x}_B}{\left(S_{x_A}^2 / 5 + S_{x_B}^2 / 4\right)^{1/2}}$$

(5)

where $\bar{x}_A$ and $\bar{x}_B$ are the average noise levels for
engines from Plant A and Plant B at each specific RPM
and $S_{x_A}$ and $S_{x_B}$ are the standard deviations.

The $t_{0.05}$ values corresponding to $v$ values from equation
(4) can be found in a statistics reference such as Ref. 3.
If the value of $v$ falls between two whole numbers, then
the “t-value from table” is calculated by interpolating
between the corresponding $t_{0.05}$ values from the table. If
the calculated t-value from equation (5) lies within the
range of “t-value from table” found using the v-value of
equation (4), the two populations are not statistically
different. If, however, the calculated t-value from
equation (5) is greater than the “t-value from table”
obtained from equation (4) and the table, then there is at
least 90% confidence that the two populations are
different.

BACK-TO-BACK TEST VARIABILITY

A total of 20 full-load sweeps were acquired on one
Plant B engine to determine the variability of back-to-
back test runs. The results reported here are from 4
microphone averages of the overall noise level in dB(A).
Figure 1 shows the results of the 20 back-to-back full-
load sweeps. The spread from the highest to the lowest
noise levels is about 1 dB(A) at low speeds, but the
spread decreases to about 0.6 dB(A) at higher speeds.

![Figure 1. Comparison of 20 full-load speed sweeps.](image)

The standard deviation of the 20 back-to-back full-load
sweeps is shown in Figure 2. The standard deviation is
as high as 0.5 dB at low RPM, but it declines to values
around 0.2 dB at most speeds.
Figure 2. Standard deviation from a 20 sweep average.

Using equation 2, the confidence interval for a 20 run set of back-to-back tests can be calculated. The result is shown in Figure 3. The same information is presented in a different format in Figure 4. These plots show that we can have 90% confidence that the average of 20 back-to-back test runs is within 0.1 dB(A) of the true level. Note that the confidence interval is a little wider at low speeds, where there was a larger spread in the results and thus a larger standard deviation (see Figures 1 and 2).

Figure 3. Confidence interval of the mean noise level from a 20 sweep back-to-back test at full-load.

Using the standard deviations determined for the 20 run test, the confidence intervals for a 3 run average can be determined. Data from the first 3 sweeps of the 20 run set were used to find confidence intervals for a 3 sweep average. The results are shown in Figures 5 and 6. For a 3 sweep back-to-back test, the confidence interval is about 4 to 5 times larger than for a 20 run average. The 3 sweep back-to-back confidence interval ranges between 0.25 and 0.5 dB(A), compared to 0.1 dB(A) or less for the 20 run average. This larger confidence interval is the penalty for using a smaller number of tests.

Figure 4. Confidence interval of the mean noise level calculated from the average of 20 full-load sweeps of Figure 3.

Figure 5. Confidence interval of the mean noise level from a 3 sweep back-to-back test at full-load. Compare with the 20 sweep results in Figure 3.

Figure 6. Confidence interval of the 3 sweep average data from Figure 5. At low speeds, the true mean is 90% likely to fall within about 0.5 dB of the measured mean. At higher speeds, the true mean is 90% likely to fall within 0.3 to 0.4 dB of the measured mean. Compare these confidence intervals with the 20 sweep intervals of Figure 4.

No comparable testing was done for steady state noise measurements, so no statistical evaluation of steady state back-to-back results is available.
One engine from each plant was run through 3 install-test-remove cycles to determine the variability introduced by removing and reinstalling the test engine. During each cycle, three sweeps were run and averaged to determine the engine noise level. Figures 7 and 8 show the confidence intervals for Plant A and B engines running full-load sweeps. The Plant A engines actually show a smaller confidence interval for the install-test-remove cycle than was found above for back-to-back testing. The Plant B engine gives mixed results. At some speeds, the confidence interval is quite small, while at other speeds, it is well over 0.5 dB(A).

Another set of speed sweeps was run under no-load conditions. These results can be seen in Figures 9 and 10. The confidence intervals for no-load sweeps are somewhat larger than the intervals for full-load sweeps. There is also a fairly strong fluctuation in the confidence interval as a function of engine speed. This suggests that there may be some characteristic of the engine that causes more variability under no-load conditions than under full-load conditions, especially at certain speeds.

Figure 7. Confidence intervals for 3 install-test-remove cycles of a Plant A engine under full-load sweeps.

Figure 8. Confidence interval for 3 install-test-remove cycles of a Plant B engine under full-load sweeps. Note that Plant B engines show more scatter than Plant A.

These results are not surprising, given that they suffer from a very small sample size. Only 3 install-test-remove cycles were run in each case, and only 3 sweeps were averaged to produce the mean noise level of each install-test-remove cycle. In general, these results show that the install-test-remove cycle adds little variability over the back-to-back test variability demonstrated in the previous section.

Figure 9. Confidence interval for 3 install-test-remove cycles of a Plant A engine under no-load sweeps.

Figure 10. Confidence interval for 3 install-test-remove cycles of a Plant B engine under no-load sweeps. Again, Plant B engines show more scatter than Plant A.

Steady state data were taken at six full-load conditions and one part-load condition in the install/test/remove test cycles. The confidence interval was determined for each speed condition, and the average interval for all 7 conditions was calculated. The average confidence intervals for Plant A engines was 0.38 dB, while Plant B engines had an average steady state confidence interval of 0.48 dB. The steady state standard deviations averaged 0.22 dB for Plant A engines and 0.29 for Plant B. Thus, the variability in steady state results is very comparable to the sweep results, even though the steady state results are generated from a single 30
second average, while the sweep results are generated from an average of 3 very short time samples.

**POPULATION VARIABILITY**

Five Plant A engines were tested to estimate the mean noise level of the Plant A population. Figures 11 and 12 show the confidence intervals of Plant A engines under full and no-load sweeps. Again, each engine test consists of a 3 sweep average. The confidence intervals for Plant A engines are generally 0.5 dB(A) or less, only slightly more than the intervals identified for install-test-remove cycles.

Figure 11. Confidence interval for the population of Plant A engines under full-load sweeps, based on 5 samples.

Figure 12. Confidence interval for the population of Plant A engines under no-load sweeps, based on 5 samples.

Four Plant B engines were tested to estimate the mean noise level of the Plant B population. Figures 13 and 14 show the confidence intervals of Plant B engines under full and no-load sweeps. The confidence intervals for the Plant B engines are somewhat larger than for the Plant A engines, with intervals as large as 1 dB(A) at a few speeds. Remember that the small population sizes in these samples will lead to some variability in the statistical results.

The next question is whether there is a difference between Plant A engines and Plant B engines. Figure 15 shows the difference in average levels between the two populations for full-load sweeps, while Figure 16 shows the same result for no-load sweeps. These figures do show some difference between the two populations, but the question remains whether the measured differences are statistically significant.

Figures 15 and 16 suggest that the two engine plants produce engines with different noise levels. To determine whether the differences shown in Figures 15 and 16 are statistically significant, the data would have to be evaluated in a different way.
Figure 15. Comparison between the average full-load noise levels of engines from Plants A and B.

Figure 16. Comparison between the average no-load noise levels of engines from Plants A and B.

Figure 17. Comparison of calculated t-values for full-load sweeps to the table t-values. Where the calculated t-value falls outside the range of table values, there is 90% or greater confidence of a difference between the two engine populations.

Figure 17 shows the calculated t-values (from equation 5) for the difference between the two populations at full-load, compared to the range in “t-values from table” determined from equation 4 and a statistics table. While Figure 15 suggests that there are noise differences between engines built in the two plants, Figure 17 shows that there is no statistically significant difference, except at speeds below 1000 RPM.

Figure 18 compares engines from the two plants during the no-load speed sweeps. Here there are statistically significant differences at many of the test speeds, showing that engines from Plant A are louder than engines from Plant B over a fairly wide range of speed at no-load. A comparison of Figures 16 and 18 shows that a difference of about 0.8 dB is required to establish statistical significance.

This same comparison can be presented in a different format, as shown in Figures 19 and 20. Here, the noise levels of each engine population are plotted as a range. The range represents the 90% confidence interval for each population. The full-load plot, Figure 19, shows that the confidence intervals of Plant A and Plant B engines almost always overlap to some degree, indicating that there is not a statistically significant difference between engines built in the two plants. On the other hand, Figure 20 shows that under no-load sweeps, Plant A engines are generally louder than Plant B engines. At most speeds, there is a gap between the two confidence intervals, indicating a statistically significant difference between the two populations.

Steady state data was taken on the two populations at six full-load conditions, one part-load condition, and at idle. Combining these results, the average steady state confidence interval for Plant A engines is 0.30 dB, and 0.58 for Plant B. The population standard deviations averaged 0.31 for Plant A and 0.49 for Plant B. The steady state confidence intervals are somewhat tighter than the results from sweep testing, which may reflect the longer averaging time of the steady state measurements. There is also a trend here, in that Plant
B engines consistently show more variability than Plant A engines. However, it must be noted that the sample size is very small, so it cannot be stated with great confidence that Plant A engines are more consistent than Plant B engines.

The confidence intervals for different types of tests are addressed.

**SUMMARY OF STATISTICAL RESULTS**

The confidence intervals for different types of tests are summarized in Figure 21. Each bar in this figure represents the average of all the different speeds tested during a sweep. Thus, the average confidence interval for 20 full-load sweeps is 0.082 dB(A), while the average confidence interval for the 3 sweep averages used throughout the program is 0.37 dB(A).

Figure 21 shows that the average full-load install-test-remove confidence interval of Plant A engines is only 0.21 dB(A), while Plant B engines have nearly twice the interval at 0.41 dB(A). The no-load sweep confidence intervals of engines from the two plants are almost identical at about 0.48 dB(A). The average confidence intervals from steady state tests are 0.38 dB(A) for Plant A and 0.48 dB(A) for Plant B. These results bracket the 3-sweep back-to-back confidence interval, so the install-test-remove cycle did not introduce significant additional variability in this project. Long years of NVH test experience indicate that in some cases, the variability due to removal and reinstallation can be significant.

Figure 21 also shows that the average population confidence interval for full-load sweeps is only slightly higher than that for install-test-remove cycles at about 0.33 dB(A) for Plant A engines and 0.59 dB(A) for Plant B engines. The steady state results are very similar, at 0.30 dB(A) for Plant A and 0.58 dB(A) for Plant B. Under no-load sweeps, the population variability is slightly less than the install-test-remove variability. Since it is impossible for the total engine population to show less variability than install-test-remove cycles with a single sample engine, this result reflects the small sample sizes used in the program and the fact that the two populations were surprisingly consistent in their noise levels. Again, long years of NVH test experience teach that a population of engines can have much greater variability than what was seen in these two populations.

NOW, THE QUESTIONS asked earlier can be addressed.

1. If an engine is modified, how small of a noise difference can be reliably detected with a single test?
In order to obtain statistical confidence that a measured difference is real, the two confidence intervals must not overlap. If a change is made and measured in a back-to-back test using a 3 speed sweep average, the measured difference must be greater than 0.8 dB(A) to be statistically significant. In other words, if the test produces a result showing a 1 dB(A) difference between the two configurations, there is 90% confidence that the true difference is between 0.2 and 1.8 dB(A).

2. How can repeat tests before and after modification be used to improve the accuracy of and confidence in results?

If 20 sweeps are averaged for each configuration, differences as small as 0.2 dB(A) can be considered statistically significant. A measured difference of 1 dB(A) would be 90% likely to represent a true difference of 0.8 to 1.2 dB(A), which is a significant improvement over the 3 sweep average. Thus, additional tests can greatly improve the confidence level and/or the sensitivity of the results.

3. How well do one or two sample engines represent a population?

Based on an average of the two populations studied here, the standard deviation of an engine population is about 0.4, when each engine is tested with a 3 sweep average. As a result, if only one engine from each population is tested, the difference would have to be greater than 1.6 dB(A) to have 90% confidence that one engine population is louder than the other.

4. How many test repeats are required to detect a real difference between two populations of engines with 90% confidence, even if the difference is as small as 0.2 dB?

If the requirement is to measure the difference between two populations to an accuracy of 0.2 dB(A), then the number of engines to be tested would be impractical. 13 engines of each population would have to be tested to get the confidence interval for each population down to +/- 0.2 dB(A). Even this would only allow the detection of a statistically significant difference if the two population means differ by at least 0.4 dB(A). In the present study, where 5 engines from Plant A are compared to 4 engines from Plant B, there is confidence of a statistically significant difference between the two populations only when the measured difference is about 0.8 dB(A) or more. A measured difference of 0.8 dB(A) means that there is 90% confidence that the true difference is between 0 and 1.6 dB(A). This is a pretty wide window.

**CONCLUSION**

It is common in NVH development to rely on tests of a single engine to represent the population, and to use a single back-to-back or install-test-remove cycle to determine the noise effect of a proposed change. The results presented here show the limitations of using a single test or a small number of tests. In many cases, engineers make decisions based on a small measured difference that is statistically insignificant. Program cost and timing constraints often leave little choice. However, it is important for development engineers and program managers to become more aware of the risks they are taking in these situations. A good understanding of what it takes to achieve statistically significant results will help avoid mistakes and wasted effort. Unfortunately, there is often no substitute for more extensive testing to achieve the desired confidence in a result.

This project showed that extending the traditional SAE J-1074 test from steady state conditions to speed sweeps has little effect on test variability, provided that the results of at least 3 sweeps are averaged. Therefore, there is no reason for NVH engineers not to use speed sweeps to evaluate changes.

If J-1074 data from a single A-B comparison test is used to evaluate the effect of a change such as a new valve cover, the difference must be greater than 0.8 dB to be more than 90% confident that there is a real difference. A measured 1 dB difference can indicate a true difference of anywhere between 0.2 and 1.8 dB, which covers the range from insignificant to substantial. If a single test shows an improvement of 0.5 dB, the true difference may be anywhere between 0.3 dB louder to 1.3 dB quieter. In this case, there is a significant risk of making the wrong decision – in direction, not just in magnitude.

When using J-1074 tests to make program decisions, repeat testing is usually essential if mistakes are to be avoided. Repeat tests can be used to detect much smaller differences, or to improve the accuracy of the estimated difference.

Using data presented in this paper, engine developers can design test programs to meet desired levels of accuracy and selectivity. However, care should be taken to measure repeatability in each test environment and for each new population. As the fuel economy sticker on every new car warns, “Your results may differ.”

**REFERENCES**

CONTACT

Thomas E. Reinhart is the Program Manager for Powertrain NVH at Roush Industries, Inc.

Phone: 734-779-7448

E-mail: terein@roushind.com

Roush Industries, Inc.
Noise & Vibration Engineering
11953 Market Street
Livonia, MI 48150
http://www.roushind.com/

Arun Sampath was a Project Engineer at Roush Industries, Inc. He is now a Senior NVH Engineer at Collins & Aikman, Inc.

Kalyan S. Bagga is the NVH Manager at the International Truck and Engine Corporation

George W. Leistensnider is a Chief Engineer at the International Truck and Engine Corporation