ICA 2013 Montreal
Montreal, Canada
2 - 7 June 2013

Architectural Acoustics
Session 3pAAbl: Balancing Risk and Innovation in Acoustical Consulting

3pAAbl4. A statistical analysis of acoustical measurement uncertainties for assemblies within multi-family dwellings in the United States
Wayland Dong and John LoVerde*

*Corresponding author's address: jloverde@veneklasen.com

Over the last several years, the authors have demonstrated that the uncertainties in the methods of acoustical testing are much larger than realized by professionals and lay people [LoVerde and Dong, J. Acoust. Soc. Am. 125, 2629 (2009), J. Acoust. Soc. Am. 126, 2171 (2009) J. Acoust. Soc. Am. 130, 2355 (2011)]. Acceptance of the large uncertainty immediately raises many practical questions. How much of the uncertainty is inherent in the test procedure and how much is due to differences between laboratories, installation methods, contractors, materials, etc.? Is the data hopelessly chaotic, or is there "true value" that can be obtained by suitable data processing? How many tests are required to feel confident in the characterization of an assembly? Some rules of thumb have been developed based on our experience [LoVerde and Dong, J. Acoust. Soc. Am. 131, 3319 (2012)], but these questions have not yet been systematically addressed before. A statistical analysis has been performed using a database of thousands of laboratory and field noise isolation tests. Results are presented that address these questions.

Published by the Acoustical Society of America through the American Institute of Physics
INTRODUCTION

The fields of architectural acoustics and noise control have historically have had little discussion regarding the statistical properties and uncertainties of acoustical measurements. Although such metrics as Sound Transmission Class\(^1\) or STC been defined and measured for decades, there is little reliable data on the uncertainties of the measurements. When standard deviations are available, they are typically around 2 dB even in the most controlled conditions\(^2\). While 2 dB sounds reasonably small, recall that for a normal (Gaussian) distribution, 95 percent of the values will be within \(\pm 2\) standard deviations of the mean. The total range encompassing that 95 percent therefore runs from \(-2\) to 2 standard deviations, or a spread of at least 8 dB, which is uncomfortably large given the goals of acoustical design and related analysis. The authors’ anecdotal experience is that many acoustical professionals, developers, architects, engineers and contractors have an unspoken assumption that acoustical tests “should” have results that are consistent, repeatable and reproducible to within a couple dB (corresponding to standard deviations well below 1 dB). When confronted with large uncertainties, such as two test reports on identical assemblies that differ by 6 dB, the initial reaction is that there is something wrong with one or both tests (usually related to materials and construction methods), and are unwilling to accept that the test method is that much less precise than traditionally assumed.

There are undoubtedly aspects of laboratory and field acoustical testing that can be improved to reduce the uncertainties in measurements. However, there has been little research to define and apply these potential improvements, and even if they can be identified and addressed, it seems unlikely to occur in the near future, especially within the United States. The focus, at this time, should instead be on coping with the uncertainties: quantifying the variations, analyzing the statistical behavior, and changing how assemblies are designed, if necessary. In this paper we begin to address these issues systematically and quantitatively.

AVERAGING DATA TO REDUCE VARIATION

Background

To examine the effects of averaging data with large variations, consider the transmission loss\(^3\) (TL) values of dual-glazed window assemblies. The properties of windows are of critical importance to calculating the exterior-to-interior transmission of noise from transportation and other urban sources. Such analysis is required in the United States for most residential, hospitality, and many commercial developments in urban areas, as most projects are located near busy roads, railroads, or airports. The acoustical properties of windows are often specified by a single number rating, typically STC. (This analysis is not specific to STC; other single number ratings are available but they are also calculated from the third-octave band TL values.) Traditionally the project documents will specify the STC rating, not octave or third octave TL. Given an STC rating, can a user confidently predict what the TL will be at each frequency?

Veneklasen Associates’ subsidiary testing corporation, Western Electro-Acoustic Laboratories (WEAL), has performed thousands of TL tests on windows and other glazed assemblies. For this study, we selected the subset of those tests on dual-glazed windows, and further selected for specific STC ratings as described further below.

As mentioned in the introduction, the picture is disheartening at first glance. For example, Figure 1 shows the third octave TL values for 10 randomly selected STC 37 windows. It seems there are variations of 7 or 8 dB at every frequency! Indeed, the standard deviation in TL at each frequency varies from about 1.5 to over 3 dB, regardless of window type or STC rating (see below). Assuming a normal distribution, the 95 percent range of variation in TL at each frequency is therefore between about 6 and 12 dB. This is considerably more than most anticipate and leaves a true challenge in the nature of how engineers approach and manage design and potential risk.

Such variation raises immediate questions. Is there any useful information in these test reports, obscured by noise, or is it arbitrary and hopelessly chaotic? Is it meaningful to talk about differences in STC ratings of 3 points when the variation in each is 8 or 10 dB? If all the information we have is, for example, a dual-glazed window with a rating of STC 37, can we use that to calculate the interior noise level? How confident can be in our prediction?
The obvious way to deal with a signal with large variations is to average a large amount of data. With the assumption that there is a “true” signal that is obscured with random variation, the random contributions will cancel over a large number of independent measurements*. Table 1 shows the STC ratings selected from the WEAL testing. These ratings were tested as representative of typical ratings that required to be met within specifications for buildings. The averaged TL spectra are shown in Figure 2. The standard deviations of the set of TL values at each frequency for each window type are shown in Figure 3. Despite the apparently hopeless scatter of the spectra of the individual tests (c.f. Figure 1), the averaged spectra in Figure 2 are well-behaved and appear “textbook.” The random variations cancel and the true signal becomes apparent.

### TABLE 1. Windows studied in this analysis.

<table>
<thead>
<tr>
<th>Window STC Rating</th>
<th>Number of tests</th>
</tr>
</thead>
<tbody>
<tr>
<td>STC 30</td>
<td>104</td>
</tr>
<tr>
<td>STC 33</td>
<td>207</td>
</tr>
<tr>
<td>STC 35</td>
<td>116</td>
</tr>
<tr>
<td>STC 37</td>
<td>28</td>
</tr>
</tbody>
</table>

* It is common in science and engineering to refer to this random variation of unspecified origin as “noise” as opposed to the “signal,” as in “noisy data”, “signal to noise ratio” or “averaging to reduce the noise.” Given that “noise” of a different sort is the object of the measurement, we try to avoid this terminology as it invites misunderstanding.
Figure 2 clearly indicates that there is no significant difference between the different STC ratings at frequencies below 200 Hz. This agrees with the general anecdotal experience of most consultants who have performed a wide variety of calculations and measurements of windows. This phenomenon is attributed to the fact that the airspace between the two panes is roughly the same size for all of these specimens. This fact can easily be masked when comparing single tests of various ratings or single tests within the same STC rating (c.f. Figure 1). The averaged spectra make it impossible to miss. Further, the averaging shows that the difference in TL at higher frequencies increases linearly with the STC rating.

**Discussion**

It should be emphasized that not all of this variation is uncertainty in the measurement method. There are many different glazing constructions that can be used to achieve a given STC rating, and these variations in construction will contribute to the overall variation. It is possible that the TL at certain frequencies might be correlated with, for
example, the thickness of one of the panes of glass. Such relationships may be found in a closer examination of the data, but if they exist they have a relatively small effect. At this level of analysis, any such systematic effects are not apparent. The relative strengths of the sources of variation are unknown, but this is moot; for practical purposes, the noise appears normally distributed and can be treated as random.

To look at the data another way, we examine the distribution at the individual frequency bands. Figure 4 shows histograms of the TL for the STC 33 windows at selected frequencies. (Note that the values are for third-octave TL even though the chosen bands match octave band center frequencies; the remaining bands had the same pattern but are omitted for clarity.) Each point on the STC 33 trace in Figure 3 is the width of the curve in Figure 4 with the corresponding frequency. Figure 4 again appears to show a normal distribution of random variation about the mean at each frequency band.

To repeat, not all of this variation is (necessarily) actually random. For instance, the larger than normal variation in Figure 3 for the STC 37 windows at the lower frequencies suggests that there may be additional sources of variation that are not present for the remaining windows. It is easy to speculate on reasons for this (presence of laminated glass, glass thickness, etc.), and these relationship which may turn up on closer analysis. Or, it may be a statistical anomaly (there are considerably fewer STC 37 tests that the other ratings). Regardless, in the absence of any such information we can treat these variations as random.

Since any single test will be distributed within plus or minus 2–3 standard deviations about the mean, from the values in Figure 3 it is apparent that the variation between tests of the same rating can be much larger than the difference in STC ratings. For instance, it would be easy to find conditions where one window performed better at some frequencies than a higher-rated STC window. However, this does not indicate that the TL test does not provide useful information. Based on the above analysis, it appears that there is a reliable value of the TL at each frequency band for a dual glazed window of given STC rating, but that the uncertainties are high, and that averaging a sufficient number of tests reveals the signal. How many tests are required is discussed below.

For design and evaluation purposes, we can treat a window described only as “dual-glazed STC 33” as a window with an average TL spectrum shown in Figure 2 and with a normal distribution of noise with widths shown in Figures 3 and 4. If we use the average value as the basis of design, 50 percent of randomly selected windows will achieve that level or better. We can add a safety factor depending on how critical the application; for example, using the average minus one standard deviation as our TL, would ensure that about 85 percent of randomly selected windows would meet or exceed that TL. Based on this, the designer may have a confidence in the ability of the

---

This is not surprising. According to the central limit theorem, the mean of a sufficiently large number of independent random variables is distributed approximately normally, even if the distribution of the variables is not normal.
specimen to meet the interior calculated noise at any particular frequency and may drive the selection of the particular assembly.

**Average STC Rating of a Given Glazing Construction**

For the above analysis, we averaged TL data while hold the STC rating fixed. Similar results are obtained from averaging the STC ratings while holding the glazing construction fixed. Figure 5 shows a figure from the authors’ presentation at the 162nd Meeting of the Acoustical Society of America in San Diego, shows the distribution of STC ratings for 632 dual-glazed windows with a total glazing thickness of ¼-inch and no laminated glass. All of these tests were performed at WEAL. The mean is 30.3 and the standard deviation is 2.65. This analysis was originally performed because a window with this description was submitted with a test report showing that it was STC 35. As can be seen from Figure 5, this result is on the upper tail (95th percentile) but it is within the distribution of results.

For the purposes of this paper, the point of Figure 5 is that the distribution is again approximately normal with a comparable width to previous distributions (c.f. Figures 3 and 4). The results of the analysis are the same regardless of which variables are used. If only the STC rating were specified, we would get a broad spread in TL as described above. If only the construction is specified, we get a similar broad spread in STC ratings.

![Figure 5. Histograms of STC ratings for 632 windows with ¼-inch thick dual glazing.](image)

**HOW MANY TESTS ARE REQUIRED?**

Given that averaging a number of noisy tests can yield accurate and useful data, how many tests do we need to perform? The answer of course depends on the width of the noise distribution (such as the standard deviations in Figures 3 and 4) and on the desired precision. The standard deviation of an average of measurements (denoted \( \sigma_{avg} \)) decreases from the standard deviation of a single measurement (\( \sigma \)) as the square root of the number of independent measurements \( n \) (assuming a normal distribution). That is,

\[
\sigma_{avg} = \frac{\sigma}{\sqrt{n}}
\]

(1)

For example, the reproducibility standard deviation for field airborne noise isolation measurements is estimated at 1.9 dB. If five measurements (\( n=5 \)) are performed, the standard deviation of the average would be 0.85 dB. In other words, if groups of 5 measurements were conducted many times, they would form a normal distribution about the “true” value with a standard deviation of 0.85 dB. Sixty-eight percent of these sets of 5 tests would be within 1\( \sigma \) = 0.85 dB of the “true” value, 95 percent would be within 2\( \sigma \) = 1.7 dB, etc.

If we wanted 95 percent confidence that our average was within 1 dB of the actual value (the presumed accuracy of the tests by many professionals in acoustics and users of the test data), we would need \( \sigma_{avg} = 0.5 \) dB. Given \( \sigma = 1.9 \) dB, Eq. (1) then gives \( n = 14.4 \), so that 15 tests are required to achieve this level of confidence.

At our presentation at the 163rd Meeting of the ASA in Hong Kong, we introduced “The Rule of Five” as a rule of thumb in evaluating tests, which is to average at least 5 independent tests to achieve an accurate result. This
reduces the uncertainty by a factor of 2.2. If our standard deviation is 1.9 dB, for a single measurement, there is
only about a 68 percent chance that the result is within 1.9 dB of the actual value. With 5 measurements, the
standard deviation is reduced by 2.2, so there is a 68 percent chance that the result is within 0.8 dB. Equivalently,
there is a 97.5 percent chance (2.2σ) that the result is within 1.9 dB, the original standard deviation.

Note that this reduction in uncertainty applies to the average of the set of \( n \) tests, not to individual tests. A small
value of \( \sigma_{\text{avg}} \) improves our confidence that the average of the test results is close to the “true” average value of the
assembly, in the sense that additional sets of \( n \) independent tests will fall within the average distribution of width
\( \sigma_{\text{avg}} \). Any single test, however, will fall within the distribution of the original width \( \sigma \).

**CONCLUSIONS**

It appears that current acoustical testing methods such as TL are accurate, at least within a single laboratory, but
there are large variations in the data. The sources of the variations are not known but are also not important for
many purposes. Averaging a sufficient number of tests results in a surprisingly clean graph, in which physically
important results (such as the constant TL around 125 Hz regardless of STC rating) are obvious.

These average values can be used for design, with the appropriate statistical mindset. Designing to the average
means that half of randomly selected windows will be satisfactory but half may not. Adjusting the average by 1 or
more standard deviations may be prudent depending on how much safety factor is required.

The standard deviation of the average decreases as the square root of the number of measurements. The number
of tests required can be calculated if the measurement uncertainty and the required precision are known. For many
situations, five is an appropriate number of measurements, corresponding to a reduction in uncertainty by a factor of
2.2, but could potentially grow to as many as 15 tests to achieve desired level of confidence. These factors should
be carefully considered when evaluating, reviewing or specifying acoustical test data.

**ACKNOWLEDGMENTS**

The authors wish to thank Western Electro-Acoustic Laboratory for their assistance.

**REFERENCES**

3. ASTM E 90-09, “Standard test method for laboratory measurement of airborne sound transmission loss of building partitions
   130, 2355 (2011).
6. J. LoVerde and W. Dong, “Coping with uncertainties in the design and evaluation of acoustical assemblies” (abstract), J.