Journal of Sound and Vibration (2000) **231**(3), 819–829
doi:10.1006/jsvi.1999.2565, available online at http://www.idealibrary.com on IDEAL®

MEASUREMENT AND ASSESSMENT OF NOISE WITHIN PASSENGER TRAINS

A. E. J. HARDY

Noise & Vibration Team, AEA Technology Rail, P.O. Box 2, London Road,
Derby DE24 8YB, U.K.

(*Received in final form* 23 *September* 1999)

Railways are becoming increasingly market-driven. Consequently, it is important that passengers are provided with a comfortable environment that reflects the operator's desired image for the service. A major factor in determining how passengers perceive the environment within trains is the level and nature of sound to which they are exposed. Unfortunately, the subject of noise within railway vehicles has had less attention in recent years, and is therefore less well developed, than external "environmental" noise. Two specific areas that merit investigation are methods for its quantification and assessment. A variety of criteria are used for assessing the noise environment within buildings, and may be considered appropriate for the quantification of internal train noise. These include "noise criteria" (NC), "preferred noise criteria" (PNC), "noise rating" (NR), and "room criterion" (RC). Recently, the automotive industry has also been using loudness level. Simple descriptors, such as the A-weighted sound level, have not been found to correlate well with perceived acoustic comfort. A complicating factor when considering internal rail vehicle noise is that its level and quality is not constant, with significant variability likely to occur over the duration of a journey. This difficulty is compounded by acoustic spatial variation within a vehicle. The paper considers the problems inherent in the quantification of noise within rail vehicles, and in the determination of the relationship between this noise and passenger response. Methods by which these problems may be overcome are discussed, drawing on real data and on long experience of study in this field.

1. INTRODUCTION

Making repeatable and meaningful noise measurements inside railway vehicles may, at first sight, seem relatively straightforward. Once the instrumentation, its location and the operating conditions have been defined, then the measurements should be consistent. The instrumentation and measurement locations are fairly simple to define; it is the operating conditions that create problems. Defining a specific speed may seem a fairly obvious approach, but it has been found that the noise levels inside some passenger coaches are not simply speed dependent. It is possible to go into great detail in specifying the track conditions, in particular the roughness of the rail, the weather conditions, etc. only to find that the resulting document creates some real practical difficulties. The result is a testing method that

0022-460X/00/130819 + 11 $35.00/0

becomes virtually impossible to apply and, consequently, fails to provide the reproducible results that were the aim of the standard in the first place.

In addition to the practical difficulties, there is also the problem of how the data might be applied. In the majority of cases, the aim is to try to assess the impact that the noise levels will have on people inside the vehicle. In many ways this makes the application of very restrictive testing standards even more inappropriate. If the measurements are made for conditions that apply only rarely, then they are not going to represent what the average passenger will experience on a typical journey. This becomes increasingly significant as the train operators become more interested in providing high levels of perceived comfort.

When determining passengers' response to noise levels the situation is further complicated by the choice of criteria. In the past, the A-weighted sound pressure level (L_A) has been used extensively. However, there is growing evidence that this may not be appropriate for a number of situations. For example, in buildings, noise criteria (NC) [1], noise rating (NR) [1] and room criteria (RC) [2] are all in use. These are all based around a series of standard curves that define the spectrum shape, and can be found in a number of textbooks. In the car industry there has been a growing interest in the use of loudness level [3]. Even in the railways, NR, RC and the B-weighted sound pressure level have all been used as an alternative to L_A.

This paper looks at the repeatability of measurements inside vehicles and whether using the subjective judgements of a group of people is an appropriate way forward.

2. MAKING MEASUREMENT

Thirty years ago, instrumentation restricted the type of measurements that could be taken. The portable instruments of the day were bulky and heavy by modern standards, and were very limited in what they could do. By comparison, modern instrumentation can carry out a wide variety of functions and is being increasingly interfaced with computers to make it more verstaile. Improvements in the way data is collected and processed creates the opportunity to consider new methods for evaluating noise measurement data.

Currently, it is normal practice for measurements to be made when the train is running along track in open country and on smooth rails. Although this approach may produce consistent results, trains usually also run on rails that are not particularly smooth, and in tunnels and cuttings. Measurements in open country and on smooth rails may therefore give no real indication of the overall performance of the vehicles in service conditions. For example, in open country on good quality track, the A-weighted sound pressure level measured inside a British Railways (BR) Mk 2A coach travelling at 110 km/h is approximately the same as in a BR Mk 2D coach travelling at 145 km/h. However, when these vehicles enter a tunnel, the A-weighted sound pressure level inside the Mk 2D coach increases by approximately 5 dB, but in the Mk 2A coach the increase is between 10 and 20 dB. The main difference between these two vehicle types is that the Mk 2A has opening

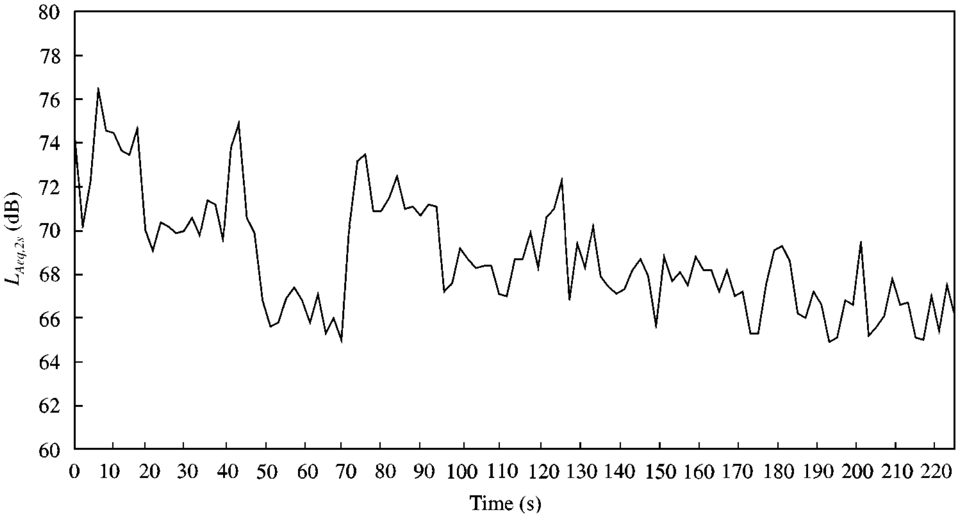

Figure 1. A-weighted sound pressure level inside a Mk 3 coach running at approximately 140 km/h.

windows while the Mk 2D is air-conditioned with fully sealed windows. When stationary, the air-conditioning equipment fitted to the Mk 2D means that it is noisier than a Mk 2A. If a choice between these two types of vehicles were made solely on noise levels within the vehicle, it would have to depend on the mixture of stationary conditions, running in open country and running in tunnels.

Parallels to this situation are found over a range of vehicles. Simply comparing noise levels measured under a fixed set of conditions is unlikely to give an accurate picture of what is likely to be encountered when a vehicle is in service.

Figure 1 shows a typical time history of the noise level inside a passenger coach. In this case, the equivalent continuous A-weighted sound pressure level (L_{Aeq}) for every 2 s was measured when the vehicle was travelling at 140 ± 8 km/h over a distance of around 9 km. It can be seen that the L_{Aeq} varies within a range of around 10 dB. Measurements on other vehicles show that this is fairly typical. An interesting feature of Figure 1 is that, although there is a tunnel between “70” and “80 s”, this does not produce the highest noise levels. Except for its passage under two bridges, the train is in open country from “0” to “70 s” and in a series of cuttings from “80” to “130 s”. After “130 s”, the train is back in open country.

Figure 1 clearly shows that the highest internal noise levels occur in open country, on a length of track that is known to be environmentally noisy. Similarly, the quietest internal levels arise on a length that is known to be quiet. Overall, Figure 1 shows that the noise level inside a passenger coach running at relatively constant speed varies significantly with time. This behaviour is typical of locomotive-hauled coaches used within the U.K.

It must be remembered that the data in Figure 1 are for a vehicle travelling at a relatively constant speed. Over the full range of speeds of operating conditions, the spread of the data will be even greater. For example, measurements of the A-weighted 1 s equivalent continuous sound pressure level, made in Mk 3 Buffet

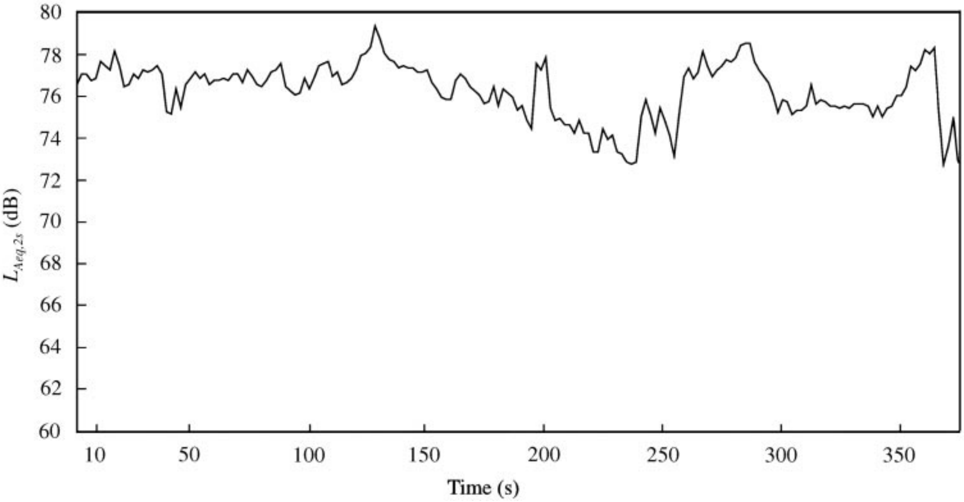

Figure 2. A-weighted sound pressure level in a class 158 DMU over a range of speeds and engine powers.

Cars over a total running time of more than 18 h show that there is a range of 20 dB between the 50 and 95 percentiles. In addition, the same measurements show that the median value of the A-weighted 1 s equivalent continuous sound pressure levels, ±0·5 dB, occurs for less than 10% of the time.

In situations where the noise level varies with time, two approaches are commonly used. One is to measure the level averaged over time in the form of an L_{Aeq}, and the other is to consider the "maximum" noise level. In some cases, both are used. The noise inside railway vehicles is in many ways an exception, as it is effectively the minimum sound level that is specified (open country, smooth rails).

The measurements presented in Figure 1 were taken over a small range of speeds. Although the noise level is some rolling stock is speed dependent, it is not universally the case. Figure 2 shows the noise levels in a Class 158 Diesel Multiple Unit as it accelerates up to 120 km/h, then slows to stop at a station, and then accelerates up to 120 km/h again. The lowest levels occur when the train is stationary, but the changes from running levels are small. Once the vehicle is moving at even a moderate speed, the levels vary by only around 3 dB, even though the route includes open country, cuttings, bridges and rails in a range of conditions. In this case, the noise experienced by the passenger will depend largely on how much time the vehicle spends at speed.

Compared with the Mk 3 coach, the noise levels in the Class 158 Diesel Multiple Unit show a relatively small spread. Typically, the median value of the A-weighted 1 s equivalent continuous sound pressure levels, ±0·5 dB, measured inside a Class 158 Diesel Multiple Unit occurs for more than 33% of the time.

In both the Mk 3 coach (air-conditioned, locomotive hauled) and the Class 158 (air-conditioned, diesel multiple unit), measuring the noise levels at full speed on good track probably has little relevance to the noise levels experienced by the passenger, as such conditions only occur rarely. In addition, it is no indication of

TABLE 1

Mean, variance and number of samples for the A-weighted sound pressure levels measured in a Mk 4 coach travelling at 200 km/h

Measurement position	Mean (dB)	Variance (dB^2)	No. samples
Centre, 3 vehicles, 1 site	67·3	3·9	6
End, 7 vehicles, 1 site	72·0	11·7	12
End, 1 vehicle, many sites	73·6	6·3	201

the overall acoustic performance of the vehicle as it is possible for two vehicles to produce similar levels under one set of conditions and totally different levels under another.

By consideration of a Mk 4 coach (air-conditioned, locomotive hauled), it was found that the main variation between measurements resulted from different measurement positions within the train. Furthermore, the average level for different vehicles and the average level for different sites were similar for the same measurement position. The results for the A-weighted sound pressure levels are given in Table 1.

The differences in the variances for the measurements made at the ends of the vehicles are probably a result of the different sample sizes. It should also be noted that the " *1 vehicle, many sites*" measurements are for a 22 km continuous stretch of track that includes rail in a range of conditions, but the " *7 vehicles, 1 site*" measurements were made on a site with "good quality" track. From these measurements it appears that there is little point in seeking out "good quality" track provided a large enough sample of data can be collected.

With Diesel Multiple Units, such as the Class 158, the situation is slightly different. Again, the measurement position is important, although the noise level in the centre tends to be higher than at the ends. Because engine noise dominates, the noise varies much less, with the standard deviation being 1·6 dB between vehicles 0·7 dB between measurement sites. Again, this supports the concept that, provided a large number of measurements are taken, there is little to be gained in searching for a particular site.

The obvious extension to making measurements over a long continuous stretch of track is to do so over the whole journey. This raises the question of how long the journey needs to be to get a reliable measure of the noise levels. Measurements made to assess the L_{Aeq} over a typical journey indicate that they need to extend over a period of around 4 h [4]. Shorter periods can give meaningful results, but this does depend on the mixture of track and operating conditions that is likely to be met.

3. SETTING CRITERIA

Making repeatable measurements is only part of the problem. Understanding what those measurements mean is also important. In the past, the A-weighted sound pressure level (L_A) has been used extensively for measuring the noise levels in trains. However, around 25 years ago, with the introduction by British Railways of air-conditioned rolling stock, it became obvious that this was not a satisfactory way of measuring noise. Steps taken to reduce the L_A resulted in increased levels of complaints. This lead to the use of a number of criteria including the B-weighted sound pressure level, noise rating curves, etc. By the 1980s the room criteria (RC) proposed by Blazier [2] had found favour within BR. This was in part because the method was easy to use, and partially because it appeared to agree well with what was regarded as being subjectively acceptable within railway vehicles. However, this has never been validated.

With a growing interest in the concept of "passenger comfort" rather than an objective measurable quantity, an initial experiment was conducted within AEA Technology Rail to determine which, if any, of the current criteria give the best agreement with subjective assessment. This initial experiment was based in the laboratory, and so the interactions with other aspects of passenger comfort were ignored.

For these tests, five noise signatures that were typical of those found inside locomotive-hauled, air-conditioned, passenger coaches were chosen, together with a noise that had similar characteristics to these but which was expected, from experience, to be close to the ideal (e.g. having a neutral spectrum shape). The reason for only using the noise from locomotive-hauled coaches was to avoid the problems that traction equipment noise might introduce if multiple units were included. However, the same approach could be used for multiple units.

The five noise signatures were originally recorded on a digital tape recorder. On the same tape, a sample of pink noise was recorded. Each of the noise signature, along with the recording of pink noise, was then re-recorded as WAV files in a computer and edited so that all the samples were the same length. By using a miniature microphone to measure the sound pressure level inside the earpieces of the headphones worn by a subject when the pink noise sample was replayed, the frequency response and the sensitivity of the system were checked. Both earpieces were tested and the average correction thus obtained was used to adjust the test signals so that the noise levels presented in this paper accurately represent the signal heard by the subjects. As an additional check, the L_A was measured for each of the samples and found to be within $\pm 0{\cdot}5$ dB of the value calculated using the correction.

The octave band sound pressure levels for each of the sounds are shown in Figure 3, and the various relevant single number criteria are given in Table 2. It can be seen from Figure 3 that Sounds 1–4 are broadly similar in shape. Test Sound 2 is nearest to the average, so by substracting this from the other Test Sounds the difference between the spectra can be seen more clearly as presented in Figure 4. It can be seen that Test Sounds 4–6 all differ from Test Sound 2 by more than 10 dB in at least one-octave band.

The noise criteria were calculated by interpolating between the values given in the standard curves. For the RC Mk1 the amount by which the neutral spectrum is exceeded can only take a positive value, but for the RC Mk2 [5] the amount by which the standard spectrum is exceeded can take positive or negative values.

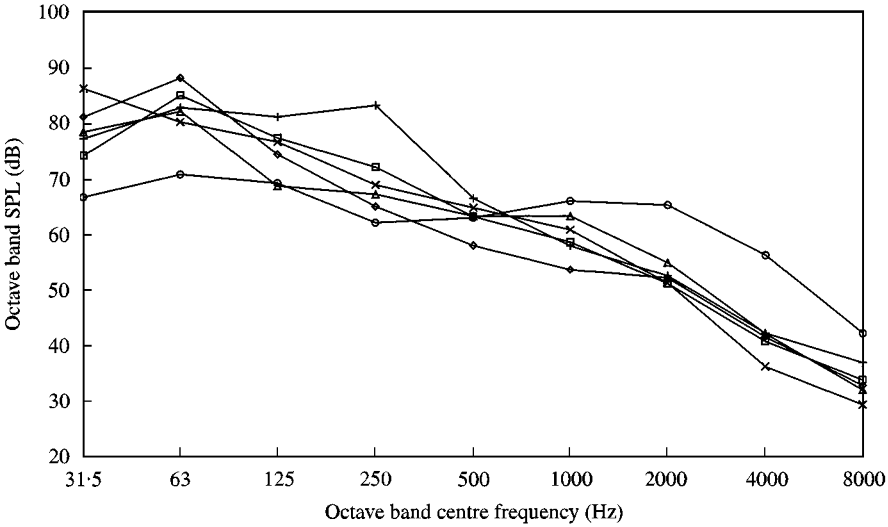

Figure 3. Octave band sound pressure levels for test sounds: –◇–, Sound 1; –□–, Sound 2; –△–, Sound 3; –×–, Sound 4; –⊖–, Sound 5; –|–, Sound 6.

TABLE 2

Various noise criteria for the test sounds

Test sound	1	2	3	4	5	6
A-weighted SPL (dB)	65·2	68·3	66·9	67·1	70·3	75·4
B-weighted SPL (dB)	79·5	78·9	75	76·8	72	83·9
C-weighted SPL (dB)	88·1	85·6	82·9	85·6	75·2	87·5
D-weighted SPL (dB)	78·2	77·9	74·3	76·1	76·1	83·3
Linear SPL (dB)	89·3	86·5	84·2	88·0	75·8	88·0
Preferred speech interference level (dB)	54·8	57·8	60·7	59·1	65·0	59·2
Exceeding RC Mk 1 at low frequency (dB)	8·6	2·5	0·0	0·0	0·0	9·3
Exceeding RC Mk 1 at high frequency (dB)	0·0	0·0	0·0	0·0	2·5	0·0
Exceeding RC Mk 2 at low frequency (dB)	7·6	− 0·5	− 2·7	1·9	− 18·5	− 1·4
Exceeding RC Mk 2 at mid frequency (dB)	1·2	3·3	− 4·1	1·2	− 10·0	8·0
Exceeding RC Mk 2 at high frequency (dB)	− 0·5	− 2·5	− 2·1	− 4·5	2·8	− 3·1
Acoustic quality index (dB)	8·1	5·8	2·0	6·4	21·3	11·1
Noise rating (dB)	67·4	64·8	63·5	63·1	67·9	76·9
Noise criteria (dB)	77·1	72·4	68·1	67·2	66·5	80·0
Loudness level (phons)	84·3	83·5	81·7	82·5	82·2	87·3

The various criteria presented in Table 2 fall into two broad categories. The weighted SPLs, the preferred speech interference level (PSIL), the noise rating (NR), the noise criteria (NC) and the loudness level all consider the magnitude of the sound in some way. The various RC criteria and the related acoustic quality index (QAI) are all based on spectrum shape. In general, the RC criteria need to be used in conjunction with the PSIL.

Table 3

The average and standard deviation for the scores from the listening tests

Sound	1	2	3	4	5	6
Average	7·83	4·00	3·50	2·98	2·61	9·17
Standard deviation	1·69	1·69	1·0	1·56	2·82	0·75

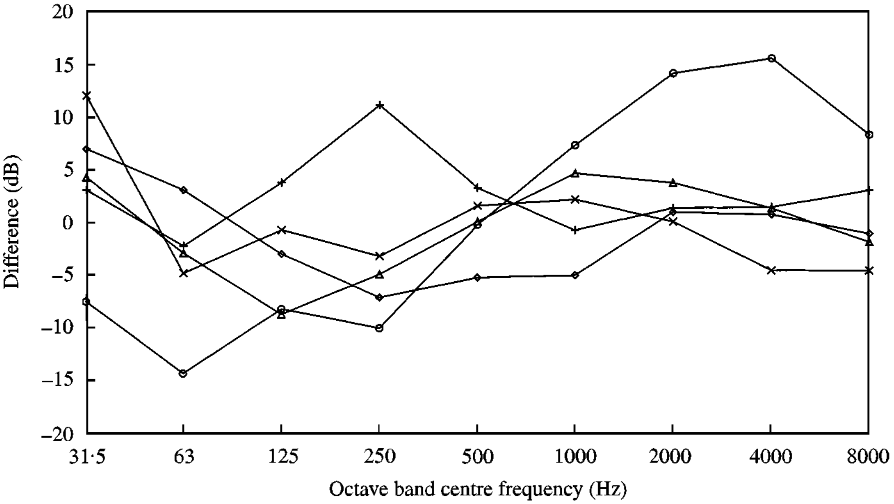

Figure 4. The difference between the octave band sound pressure levels for test sounds 1, 3, 4, 5 and 6, and test sound 2: –◇–, Sound 1; –△–, Sound 3; –×–, Sound 4; –⊖–, Sound 5; –|–, Sound 6.

A five-second computer wave file for each of the six recordings was produced and then organized so that the sounds were presented as a series of paired comparisons with replications (that is both "A compared with B" and "B compared with A"). This gave a total of 30 pairs of recordings. The pairs were then played monaurally through headphones to each of the 20 subjects who were asked to select which sample from each pair of sounds was the least acceptable. The subjects ranged in age from 23 to 57 y and comprised 14 men and 6 women. The subjects were all volunteers from AEA Technology Rail (25%) and the local community (75%). None of the subjects was aware of any problems with their hearing.

When a sound was judged the least acceptable, it was given a score of 1. The total score for each sound and each subject was then calculated by summing the number of times each sound was judged "least acceptable". The total score for any sound can lie between 10 and 0. Table 3 shows the average and the standard deviation for the scores averaged for all the subjects.

It can be seen that Sound 6 is judged to be the least acceptable followed by Sound 1. Based on the average score, Sound 5 appears to be the most acceptable but the standard deviation indicates a wide variation in response. Figures 3 and 4 indicate

TABLE 4

Correlation coefficients for the average and raw scores, for the various noise criteria

	Average	Raw
A-weighted SPL (dB)	0·41	0·39
B-weighted SPL (dB)	0·87	0·71
C-weighted SPL (dB)	0·66	0·51
D-weighted SPL (dB)	0·85	0·72
Linear SPL (dB)	0·58	0·43
Preferred speech interference level (dB)	– 0·58	– 0·43
Exceeding RC Mk 1 at low frequency (dB)	0·99	0·81
Exceeding RC Mk 1 at high frequency (dB)	0·42	– 0·30
Exceeding RC Mk 2 at low frequency (dB)	0·49	0·35
Exceeding RC Mk 2 at mid frequency (dB)	0·70	0·57
Exceeding RC Mk 2 at high frequency (dB)	– 0·16	– 0·09
Acoustic quality index (dB)	– 0·04	0·01
Noise rating (dB)	0·78	0·68
Noise criteria (dB)	0·97	0·81
Loudness level (phons)	0·92	0·77

that Sound 5 has an unusual spectrum shape compared with the other Sounds. Sound 4 has the next lowest average score and with a smaller standard deviation. It is worth remembering that Sound 4 was chosen because it was thought to have a neutral spectrum. What is particularly interesting is that Sounds 1 and 6 have the lowest and highest A-weighted SPLs respectively.

The correlation coefficients for average score (averaged for all the subjects) and raw score (the score for each subject), for the various criteria, are given in Table 4.

From these results, it is clear that the A-weighted SPL is not a particularly useful indicator of the subjective assessment of acceptability. The most useful criterion appears to be the amount by which the RC Mk 1 is exceeded at low frequencies. The noise criteria and the loudness level appear to be next most useful. The acoustic quality index (QAI) appears to be independent of the subjective assessment.

It should be remembered that the RC should be used in conjunction with the PSIL. Yet, interestingly, the PSIL does not correlate particularly well with the scores. Additionally, the score appears to increase as the PSIL decreases. This is unexpected, as experience has shown that the manner in which sound interferes with speech is an effective way of making subjective assessments. The problem could lie with the nature of the tests. During the tests, the subjects were not involved in conversations. In fact, the tests were designed to avoid this type of distraction. Consequently, speech interference is not a significant issue for the subjects of the tests but may be important to train passengers. However, the reason could also be a result of the dominance of the amount by which the RC Mk 1 is exceeded at low frequencies. For the test sounds, it was found that the PSIL had a relatively large negative correlation with the amount by which the RC Mk 1 is exceeded at low frequencies (correlation coefficient of $-0{\cdot}62$).

Simply, for these tests the amount by which the RC Mk 1 is exceeded at low frequencies so dominates the judgements made by the subjects that the PSIL is effectively lost. This is not universally the case as the correlation coefficient between the PSIL and the amount by which the RC Mk 1 is exceeded at low frequencies for the spectra used in Figure 1 is only −0·28.

A number of specifications produced by the former British Rail have used the RC Mk 1 and the PSIL to set limits for internal noise. Experience with this approach showed that, for most modern railway vehicles operating in the United Kingdom, the amount by which the RC Mk 1 was exceeded at low frequencies and the magnitude of the PSIL were normally the major noise problem. The difficulty with this approach is that it does not produce a single indicator of the performance. To see if a single indicator could be produced from listening tests, a multiple linear regression analysis of the raw data produces the following equation for combining the PSIL with the amount by which the RC Mk 1 is exceeded at low frequencies.

$$\text{RC (Combined)} = 0{\cdot}188\ \text{PSIL} + 0{\cdot}812\ \text{RC Mk 1 LF}.$$

The combined RC's for the Test Sounds calculated using this formula correlate better with the scores than with any of the other criteria (0·98 for the average Scores and 0·82 for the Raw Scores). This supports the position taken by British Rail in the 1980s that the RC (now known as RC Mk 1) was the best criterion for measuring noise inside railway vehicles. The findings also support the view that, once the level is below a certain threshold, the shape of the spectrum is probably more important than the overall sound level.

4. CONCLUSIONS

From the findings discussed in this paper there is a significant amount of evidence that the current methods of measuring and evaluating the noise inside the railway vehicles are probably flawed. The basic problems with the measurements are:

- "Good quality" track in "open country" is hard to define and difficult to locate.
- Testing only on "Good quality" track in "open country" does not represent fully the noise levels experienced by the passengers when travelling on the trains.
- Testing only on "Good quality" track in "open country" may fail to identify noise problems that occur under other conditions.
- Noise levels vary with measurement position.

To obtain consistent and meaningful results it is necessary to measure over typical journeys. Current indications are that a combined journey of around 4 h should be adequate. With modern automatic data collection this is not a particularly serious problem. It may be possible that a shorter time could be used, although further experimental work would be necessary to confirm this.

Using the A-weighted SPL appears not to be a very useful way of measuring the subjective acceptability of the noise within trains. From the data available from this initial study, it appears that the simple RC Mk 1 gives the best agreement, particularly when it is combined with the PSIL. However, it must be recognized that the tests were deliberately limited to a few samples and should not, at this stage, be considered to be conclusive. For example, the strong correlation between the PSIL and the amount by which the RC Mk 1 is exceeded at low frequencies could have lead to misleading conclusions. However, the findings do confirm that, once the PSIL is below a certain threshold, the spectrum shape is likely to be the dominant factor.

The listening tests show that it is possible to evaluate the noise inside a vehicle in this way. However, the variability in the responses to Sound 5 indicate that care needs to be taken if sound of very different characters are compared. This could become particularly significant when trying to compare the noise levels in locomotive-hauled stock with those inside diesel multiple units.

Although the tests reported in this paper were relatively simple, they do support the view that the acoustic environment inside a vehicle is judged in a complex manner. The judgement is likely to be made more complex by the presence of other factors that influence passenger comfort. For example, it is likely that if one particular aspect of the passenger environment is particularly poor then the overall judgement will be dominated by this single issue. The optimum level of comfort will be achieved when the subjective contribution from all areas is approximately the same. To achieve this would clearly require significant development. However, this paper demonstrates a way forward for the internal noise.

REFERENCES

1. D. Templeton (editor), *Acoustics in the Built Environment* 140–141. Butterworth Architecture, ISBN 0 7506 0538 3, 1993 (This publication contains a brief summary of the various criteria).
2. W. E. Blazier 1981 *Noise Control Engineering Journal* **16**, 64–73. Revised noise criteria for application in the acoustical design on HVAC systems.
3. L. L. Beranek 1971 *Noise and Vibration Control* 561–564. New York: McGraw-Hill Book Company.
4. Ride, Vibration and Noise Environment Inside Railway Vehicles; Railtrack Group Standard GM/RT 2160, Issue 1, September 1994.
5. G. Leventhall 1998 *Noise and Vibration Worldwide* **29**, 26–31. ASRAE's New Method for Assessing Air Conditioning Noise in Buildings.