# AN INVESTIGATION OF THE SUBJECTIVE RESPONSE OF OCCUPANTS TO INTERIOR CAR NOISE

S. M. Dickinson

Faculty of Engineering Science+ - The University of Western Ontario

London, Ontario N6A 5B9

## INTRODUCTION

The work described in this paper represents a continuation of a project initiated by the Vehicle Research Group of Ford Motor Company, England, the long term objective of which was to arrive at a means of predicting subjective response to interior car noise from objective measurements[1].*

The work conducted by Ford Motor Company involved the development of a high quality recording and replay system to allow subjective testing to be conducted in the laboratory, rather than in the field, and the subsequent proving of the experimental procedure by means of the comparison of results from subjective tests carried out in both environments. Attempted replication by the author, using the same tape recordings, showed the Ford results to be reasonably reproducible, although slight modification to the test procedure was found to be desirable. In these proving tests, recordings made inside five different vehicles were used, the frequency content and sound levels of each being different, thus considerable difficulty was experienced in attempting to use these results to isolate the effects of frequency and sound level upon subjective response.

Further subjective tests using controlled levels and spectrum shaping of the interior noise of a selected vehicle were therefore conducted. The results of these tests suggested that standard indices, such as dB(A), dB(B), dB(LIN) and PNdB, were not satisfactory for the prediction of subjective response of occupants. The results also suggested that, provided extremes were not encountered, interior car noise in which low frequencies predominated was preferable to that in which high frequencies were predominant, and that a weighting scale similar to the A-scale, but not attenuating the very low frequencies so much and attenuating the medium low frequencies more, may be applicable for rating 'annoyance' of interior car noise.

It was subsequently found that such a curve was recently proposed by Spring[2], based upon considerable subjective experimentation, and that the objective measures obtained by weighting the present author's spectrum shaped recordings according to Spring's curve were in complete rank order agreement with the corresponding subjective results.

## RECORDING/REPLAY SYSTEM

The recording system, developed by Ford Motor Company, comprised a high quality, two channel tape recorder, together with a weighting device to improve the dynamic range over the spectrum of the interior noise, recording through 1/2 inch condenser microphones located in the ears of a dummy head. The dummy head was in turn located in the position of a typical passenger's head in the vehicle under consideration. Replay was through high quality electrostatic

+ The work described in this article was carried out by the author whilst on sabbatical leave at the Institute of Sound and Vibration Research, Southampton University, England.

* Numbers in brackets designate References at the end of paper.

headphones, with account being taken of the effect of the frequency weighting device.

CONDUCT OF SUBJECTIVE EXPERIMENTS

The subjective tests conducted in the laboratory involved the use of tape recordings comprising 25 eight second samples of interior car noise, each sample separated from the next by a period of approximately four seconds. The 25 samples were made up of noise, recorded as described, from each of five different vehicles (or, in the case of the spectrum shaped tapes, five different spectra). The last 21 samples were arranged in a balanced order such that each of the five different noises was twice adjacent to each of the other noises, once preceeding and once succeeding the other noise. The last four samples were repeated at the beginning of the tape and were treated as an adjustment period for the judges (that is, the first four decisions made by each judge were discarded during analysis of the results).

The judges, all of whom were volunteers from the post-graduate students and academic, research and secretarial staff of the Institute, were fitted carefully with the headphones and seated comfortably in a room shielded from exterior noise and remote from the remainder of the playback system.

In order to gain experience and to establish the reliability of the approach to be used, the author initially conducted a series of tests aimed at replicating the results obtained by Ford Motor Company in proving their in-laboratory arrangement to be representative of the in-car situation. The Ford work involved the use of five vehicles (1. Ford Cortina 2.0L - THK500L, 2. Renault 4L, 3. Opel Rekord, 4. Vauxhaull Victor, 5. Ford Cortina 2.0L - XVX 395L) their in-car tests being conducted in these vehicles whilst being driven at 30 mph in third gear over a special tar and chip surface. Tape recordings were made in the same vehicles under the same conditions and a master tape prepared, for in-laboratory subjective tests, comprising 25 samples of the five recorded noises in the order 2, 3, 4, 5, 1, 3, 5, 2, 4, 3, 2, 1, 5, 3, 1, 4, 2, 5, 4, 1, 2, 3, 4, 5, 1. The replication tests used this original five car tape and the same playback system as that used by Ford.

In all, four tests were conducted in the replication investigation, there being minor variations in each, and at least twelve judges participated in each test. The first was identical to that used by Ford and involved each judge listening to the playback of the master tape and making two judgements on each sample of noise, as indicated by the following questions:

(1) Which noise causes you less annoyance - the one you have just heard or the previous one?

(2) What rating would you give the noise you have just heard using the 1 to 10 scale given? (The scale ranged from 1. Exceptionally poor to 10. Excellent).

The judges were asked to record their replies on 25 prepared answer sheets, each of which was turned face down upon completion to avoid subsequent reference. In this way, each judge gave a numerical rating for the annoyance of each noise and, simultaneously, underwent a paired comparison test.

A second set of judges were subjected to the same test but, in order to assist the judges to imagine themselves being seated in a vehicle, an introductory tape was provided, comprising a recording (made with the same dummy head system) of a vehicle accelerating from rest to 70 mph and decelerating back to standstill.

The data from these two tests were processed using a computer program which, briefly, calculated the mean and median scores for each vehicle from the numerical ratings assigned by the judges and, from the scores, computed the rank order of the noises in terms of annoyance. In addition, a rank order was derived from the paired comparison 'table of agreements'. The mathematical basis for the analyses performed by the program may be found in Moroney[3] or Kendall[4]. Application of significance tests described by Kendall[4] showed that the probability of any of the sets of subjective results occurring from random decisions by the judges was less than 1% (the limit of the tables provided), which was also the situation for all subsequent tests conducted and reported in this article.

Examination of the processed results showed that the rank orders determined from the numerical ratings were inconsistent with each other and with the Ford in-car and laboratory tests. The paired comparison results were consistent with each other but only partly with the Ford and objective measurement results. It was recognized that replication of the Ford results had not been achieved and that further tests were necessary. In view of the general inconsistency of the numerical rating results and of comments from a number of judges - to the effect that there was insufficient time allowed for the performance of both rating and comparison tasks, the former being the more difficult to complete reliably - it was decided that all subsequent tests should involve only paired comparisons.

At this stage it was discovered that the noises presented to the judges during these first two tests had been about 4dB down on their true levels. On an annoyance basis, this should not have significantly affected the rank ordering but the inclusion of a test to check this was prompted. It was thus decided to run two further attempted replication tests, using paired comparisons only, the first having sound levels as in the previous two tests and the second having levels 4dB up (that is, as recorded), the introductory acceleration/deceleration tape being retained in both cases. The rank orders obtained from these two tests were totally consistent with each other and correlated well with the in-car and laboratory tests conducted by Ford, thus it was considered that the subjective approach adopted was satisfactory for the conduct of such tests. Comparison of both the Ford subjective rankings and those from these latter two tests with objective measurements made by Ford suggested that conventional measures such as dB(A), dB(B) and PNdB would to some extent, though not completely, allow rank order to be predicted.

In an attempt to learn more about the annoyance of interior car noise, narrow band frequency analyses were performed on the five vehicle noises recorded but this analysis proved to be of little value, since the variation in the spectra together with the variation in sound pressure levels caused interpretation of the results to be very difficult.

## SPECTRUM SHAPED TESTS

In an attempt to reduce the number of variables, two new test tapes were prepared, one for which all the sample noises had the same sound level on the linear scale (used for Test A) and the other for which all samples had the same A-weighted sound level (used for Test B). The tapes took the same form as the Ford five car test tape but, this time, the five different samples were all prepared from one channel of one sample of one of the vehicle noises on the Ford tapes. The vehicle chosen (number 2) was that having the widest spectrum of noise and, incidentally, the one consistently adjudged the most objectionable. The five different samples were obtained by shaping the spectra of the single selected channel using a B & K one-third octave spectrum shaper and re-recording on both tracks of a stereo tape, adjusting the levels as required.

The unmodified spectrum of the base sample and the five frequency 'windows' used are shown in Figure 1, the numbers 1 to 5 being subsequently adopted as identifiers for the resulting shaped samples.

Twenty judges, most of whom had been involved in the earlier tests, were persuaded to assist in the conduct of Tests A and B and the tests were run consecutively, each judge undergoing both tests at one sitting. In order to establish whether the order in which the tests were taken affected the result and, if so, to minimize such an effect, ten judges took the tests in order A, B and ten in order B, A. The order of presentation of the samples was 1, 3, 2, 5, 4, 1, 2, 4, 3, 5, 2, 3, 1, 4, 2, 1, 5, 3, 4, 5, 1, 3, 2, 5, 4 and no introductory tape was provided.

The results of Tests A and B are summarised in Tables 1 and 2 respectively, where objective measurements derived from the tapes are also given. Table 1 shows complete correlation to exist between the rank order obtained from the subjective test and that which would be predicted using the objective measures dB(A), dB(B) or PNdB for the noise samples of Test A. This is to be expected since, in order to keep all linear levels the same, it was necessary to considerably attenuate the signals having high low-frequency content, thus the apparent loudnesses were significantly affected. The judgment of annoyance thus became based essentially upon loudness alone. It is evident that the order of presentation of the tests did not affect the decisions of the judges for Test A.

The results for Test B, Table 2, are rather more interesting. It appears possible that the order of presentation of the tests did affect the subjective ranking in this test, in as much as the order of noises 1 and 5 are reversed depending upon whether Test A or B preceeded the other. Examination of the Tables of Agreement and Tables of Preference from which these results came suggested that noises 1 and 5 were ranked quite differently and that this discrepancy was not likely to be due solely to inter-judge variations. The combined result showed noises 1 and 5 to be ranked about equal last. With respect to the objective measurements, the PNdB values calculated from the tape were all between 82 and 83, thus it was not considered reasonable to extract a rank order from these results. The order suggested by the dB(B) and dB(lin) values does not agree at all with the subjective ranking and are thus considered to be unsatisfactory for response prediction.

It should be noted that the noises 1 and 5 were chosen so that each was just on the limit of sounding like the interior noise of a motor vehicle, and consequently would not be likely to be encountered in practice under the supposed conditions of the test. (Noise 1 would be approached by driving over loose chippings and Noise 5 by driving at speed with one or more windows open.) Noises 2, 3 and 4, however, were considered to be quite representative of noises possible under the test conditions. These three were ranked quite consistently by the judges in the order shown in the table and consistently as more acceptable than noises 1 and 5. This would suggest that there is a preference for low frequency noise, provided that 'buffeting' is not experienced.

The fact that a fairly consistent rank ordering was obtained even with dB(A) and, incidentally, PNdB held constant suggests that these two measures are not ideal for subjective response prediction but are considerably better than dB(B) or dB(lin), which gave essentially inverse rank ordering.

## CONCLUDING REMARKS

From the meagre results obtained thus far, it did not seem reasonable to attempt to quantify suggestions but, since it appeared that if extremes were not reached (such as in Noise 5) low frequency noise was more acceptable than high, then a weighting scale similar to the A-scale but not attenuating the very low frequencies so much (for example up to about 50 Hz) and attenuating medium low frequencies more may be applicable for rating 'annoyance' of interior car noise.

Subsequent to the completion of this work and arriving at the above conclusion, it was found that such a curve had been proposed by Spring[2] - referred to as the "Computer Tested Car" curve - arrived at by optimizing the frequency weighting used in objective measures of interior car noise used in extensive subjective experiments. Using this curve, the characteristics of which are presented in Table 3, the values tabulated under dB(CTC) in Tables 1 and 2 were obtained from the tapes of Tests A and B. Virtually complete agreement is seen to exist between the subjective and objectively predicted rank orders, suggesting that employment of the CTC weighting curve would give a good indication of the likely subjective response of occupants to interior noise of European cars. The author hopes to conduct a similar study involving typical North American cars.

## ACKNOWLEDGEMENTS

The author would like to thank the Vehicle Research Group, Ford Motor Company, for their cooperation and assistance in the conduct of this project and to thank the Institute of Sound and Vibration Research for the privilege of spending a sabbatical year at the Institute, and for the use of their facilities. The assistance of the numerous judges is also gratefully acknowledged, and particular thanks are due to M. Oldman, Lecturer at the Institute, for his advice and encouragement.

## REFERENCES

1. A.V. PHILLIPS and D. CHRISTIE, "Techniques of Vehicle Interior Noise Evaluation", Eighth International Congress on Acoustics, London, 1974, p.29.

2. H.W. SPRING, "Characterization of Acoustic Comfort in Motor Vehicles", Interkeller Gap, 1975.

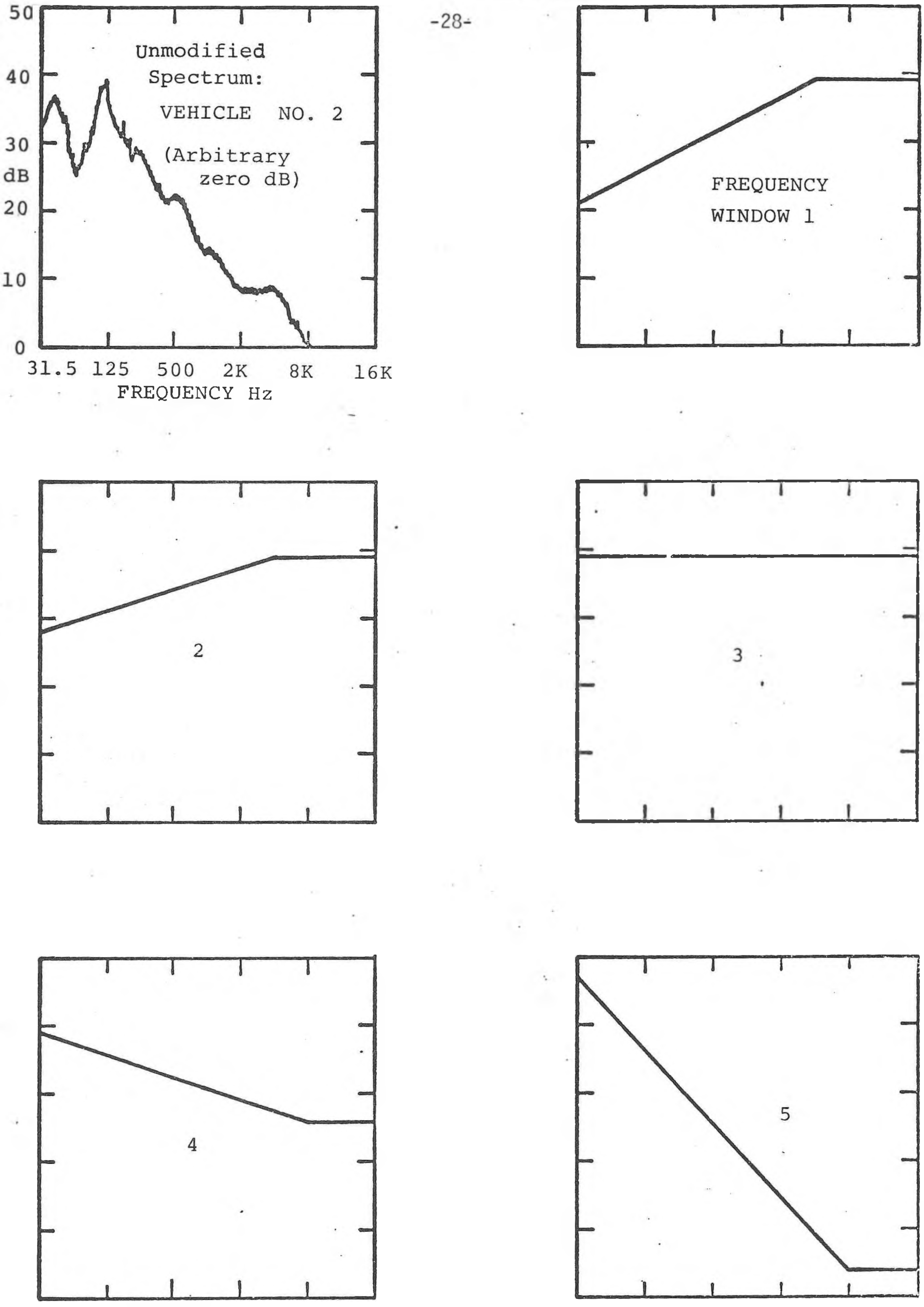


FIGURE 1 SPECTRUM SHAPED SAMPLES

Table 1 Test A - Five Spectra, All Linear Levels Equal

| | Order of Test Presentation A:B | B:A | All Judges | Objective Measurements and Predicted Rank Order dB (A) | dB (B) | dB(lin) | PN dB | dB(CTC) |
|---|---|---|---|---|---|---|---|---|
| Least ↑ Annoyance | 5 | 5 | 5 | (59) 5 | (73) 5 | 82 | (72) 5 | (61) 5 |
| | 4 | 4 | .4 | (64) 4 | (75) 4 | 82 | (78) 4 | (64) 4 |
| | 3 | 3 | 3 | (68) 3 | (76) 3 | 82 | (82) 3 | (67) 3 |
| | 2 | 2 | 2 | (71) 2 | (78) 2 | 82 | (85) 2 | (71) 2 |
| Most ↓ | 1 | 1 | 1 | (72) 1 | (78) 1 | 82 | (86.5) 1 | (74) 1 |
| Coef. of Agr.[t] | 0.519 | 0.533 | 0.533 | | | | | |
| Deg. of Freed. | 11.7 | | 10.8 | | | | | |

Table 2 Test B - Five Spectra, All A-Weighted Levels Equal

| | Order of Test Presentation A:B | B:A | All Judges | Objective Measurements and Predicted Rank Order dB (A) | dB (B) | dB(lin) | PN dB | dB(CTC) |
|---|---|---|---|---|---|---|---|---|
| Least ↑ Annoyance | 4 | 4 | 4 | 68 | (73) 1 | (77) 1 | | (67.4) 4 |
| | 3 | 3 | 3 | 68 | (76) 2 | (80) 2 | 82 | (67.8) 3 |
| | 2 | 2 | 2 | 68 | (77) 3 | (83) 3 | to | (68.6) 2 |
| | 5 | 1 | 5 | 68 | (78) 4 | (88) 4 | 83 | (69.8) 1 |
| Most ↓ | 1 | 5 | 1 | 68 | (81) 5 | (93) 5 | | (69.8) 5 |
| Coef. of Agr.[t] | 0.172 | 0.126 | 0.155 | | | | | |
| Deg. of Freed. | 11.7 | | 10.8 | | | | | |

Table 3 Characteristics of the C.T.C. Curve[2]

| | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Frequency (Hz) | 25 | 31 | 40 | 50 | 63 | 80 | 100 | 125 | 160 | 200 | 250 | 315 | 400 | 500 |
| Weighting (dB) | -31.5 | -30 | -28 | -27.5 | -25 | -23 | -20 | -17 | -15 | -13 | -11 | -10 | -9 | -8 |
| Frequency (Hz) | 630 | 800 | 1000 | 1250 | 1600 | 2000 | 2500 | 3150 | 4000 | 5000 | 6300 | 8000 | 10000 | |
| Weighting (dB) | -6 | -4 | 0 | +4 | +6 | +6 | -6.5 | +6 | +7 | +6 | +6 | +4 | +2 | |