

# ANASE: LESSONS FROM 'UNRELIABLE FINDINGS'

P Brooker      Cranfield University, Cranfield, UK

## 1 INTRODUCTION

In late 2007, the ANASE (Attitudes to Noise from Aviation Sources in England) report was published. It claimed that people are increasingly annoyed by aircraft noise, and it estimated how much they would be willing to pay to get rid of it. But its quantitative 'findings were rejected as unreliable by the Department for Transport' (BBC website). The project's managers were warned in its early stages that the work would fail to deliver good value for money and not meet accepted technical/statistical standards. How and why did it fail? What were the methodological and project management failings? What are the lessons for acoustics professionals?

## 2 BACKGROUND

The ANASE project was initiated by the Department for Transport (DfT) in 2001. Its aim was twofold: to explore the relationship between aircraft noise and annoyance; and a monetary valuation of annoyance by aircraft noise (Stated Preference [SP]). The DfT said that: "...the new study underlines the UK Government's commitment to underpin our policy on aircraft noise by substantial research that commands the widest possible confidence."

The focus here is on the annoyance component of ANASE, which was intended to update the Aircraft Noise Index Study (ANIS) carried out in the early 1980s. (The ANIS study concluded that there was no better metric than the noise energy measure LAeq (Leq here) in terms of correlation between aircraft noise and community annoyance. Following consultations, the Government decided to adopt the use of Leq to describe noise, and decided that 57 Leq (16-hour period) marks the approximate onset of significant community annoyance from aircraft noise<sup>1</sup>.) ANASE's website includes the report, its technical appendices and peer reviews [<http://www.dft.gov.uk/pgr/aviation/environmentalissues/anase/>]. In particular, DfT paid two objective and knowledgeable acoustics experts to review the ANASE draft material on annoyance and noise<sup>2</sup>.

## 3 PREDICTED AND ACTUAL FAILURES OF ANASE

ANASE was in part managed through a Steering Group (SG), covering a variety of interests, including environmental organisations, but largely with a non-technical membership. The author was invited to join the SG as an expert on aircraft noise disturbance, but left it after two meetings because of major concerns about the project. These concerns were set out in a series of letters to the DfT Permanent Secretary and other officials in early 2002. In summary, these concerns were that ANASE would not be robust, nor technically reliable, nor capable of withstanding scrutiny, nor good value for money; and that it would be a source of considerable vulnerability for DfT. Thus, the ANASE outputs would be poor value for the taxpayer, residents near airports, and the aviation industry. DfT did not heed these warnings, nor the specific recommendation to use an independent expert audit team, proposed with the aim of getting the work back on the right track and hence ensuring that the results would command the widest possible confidence.

The ANASE specification said the project should take 'in the order of three years'. The actual duration was from December 2001 to November 2007, including about eighteen months of peer reviews. The time over-run was therefore about 97%.

DfT's Permanent Secretary currently (January 2008) says that the actual spend on ANASE was £1.78 million, compared with the original budget of £0.53 million. These figures do not match responses to parliamentary questions in 2007 because those answers explicitly did not include the 2007/2008 spend. The over-spend was therefore 236%.

ANASE's early results were originally intended to inform the preparation of the Air Transport White Paper, then expected to be published in late 2002. The outputs from the final report should then have had a similar role in the development of the work that has led to the 2007 DfT consultation on 'Adding Capacity at Heathrow Airport'<sup>3</sup>. This is an extremely important document, because it brings forward specific development options for consultation with residents living around the airport. ANASE's outputs failed in three important ways: annoyance, SP valuations, and confusion (its specific claims are examined in the next section).

The ANASE claims on aircraft noise annoyance are disregarded in the Consultation Document. Its analysis simply focuses on existing policy. This continues to use Leq as the annoyance metric, with DfT's standard 57 Leq contours – and the enclosed households and populations – being taken to represent the scale of the noise disturbance generated by the airport.

No use is made in the Consultation Document of the ANASE claims on SP valuations. Its cost benefit analyses instead use results from research on 'hedonic pricing', in which the cash value of disturbance is assessed by examining the reduction in the prices of houses affected by aircraft noise. These kinds of valuations were developed well before ANASE took place.

The Consultation Document discussion of ANASE's claims produces confusion. For example, ANASE's claim of increased annoyance over time is mentioned – a comment guaranteed to cause bewilderment, given that DfT's use of a standard Leq contour explicitly assumes no change in annoyance over time. None of the 'positive spin' on ANASE in the Document is traceable to acousticians or attitude measurement researchers. The outputs have not convinced non-UK researchers, e.g. Dr Rainer Guski, a well-known German researcher on aircraft noise annoyance: "The ANASE statement that the degree of aircraft noise annoyance has changed in comparison with the ANIS study cannot be held, because the methods of data gathering are not comparable."

## **4 WHAT WENT WRONG? – DESIGN BIASES**

What went wrong? The ANASE report generated various claims, but it is easy to demonstrate that the claims are unreliable. There are unrepairable major problems with questionnaire design and process, noise estimates, analysis techniques, and selective attempts to compare with international work. The following summarises an analysis of the problems, most of which are the product of design biases.

The study report makes a number of aircraft annoyance claims:

*Claim:* "For the same amount of aircraft noise, measured in Leq, people are more annoyed in 2005 than they were in 1982."

*Claim:* "The modelling work also showed that respondents were less sensitive to changes in sound level below 42 Leq and above 59 Leq, adding support to a logistic dose-response form. There was no threshold, or discontinuity, in the relationship between mean annoyance and Leq."

*Claim:* "The results from the attitudinal work and the SP analysis both suggest that Leq gives insufficient weight to aircraft numbers, and a relative weight of 20 appears more supportable from the evidence than a weight of 10, as implied by the Leq formulation."

These are dramatic claims, particularly given the tough DfT criterion 'commands the widest possible confidence'. Thus, they would need to be robust, technically reliable, and capable of withstanding scrutiny. But there is good evidence that they are the consequence of design biases.

The major design problem stems from 'context effects': the ANIS and ANASE questionnaires are markedly different and are implemented in markedly different circumstances. Figure 1 shows a schematic comparison of the questionnaire set-ups. Two potential context effects are worth noting:

The installation of noise playback equipment precedes ANASE, but not ANIS. Thus, ANIS is a social survey and ANASE is a combination of a social survey and a foreshadowed laboratory experiment, as, later in the interview, noises are played to respondents.

ANASE starts immediately with questions on aircraft noise annoyance, but ANIS leads up to them by asking about perceptions of the local area, and thus allows the interviewee to mention aircraft noise spontaneously.

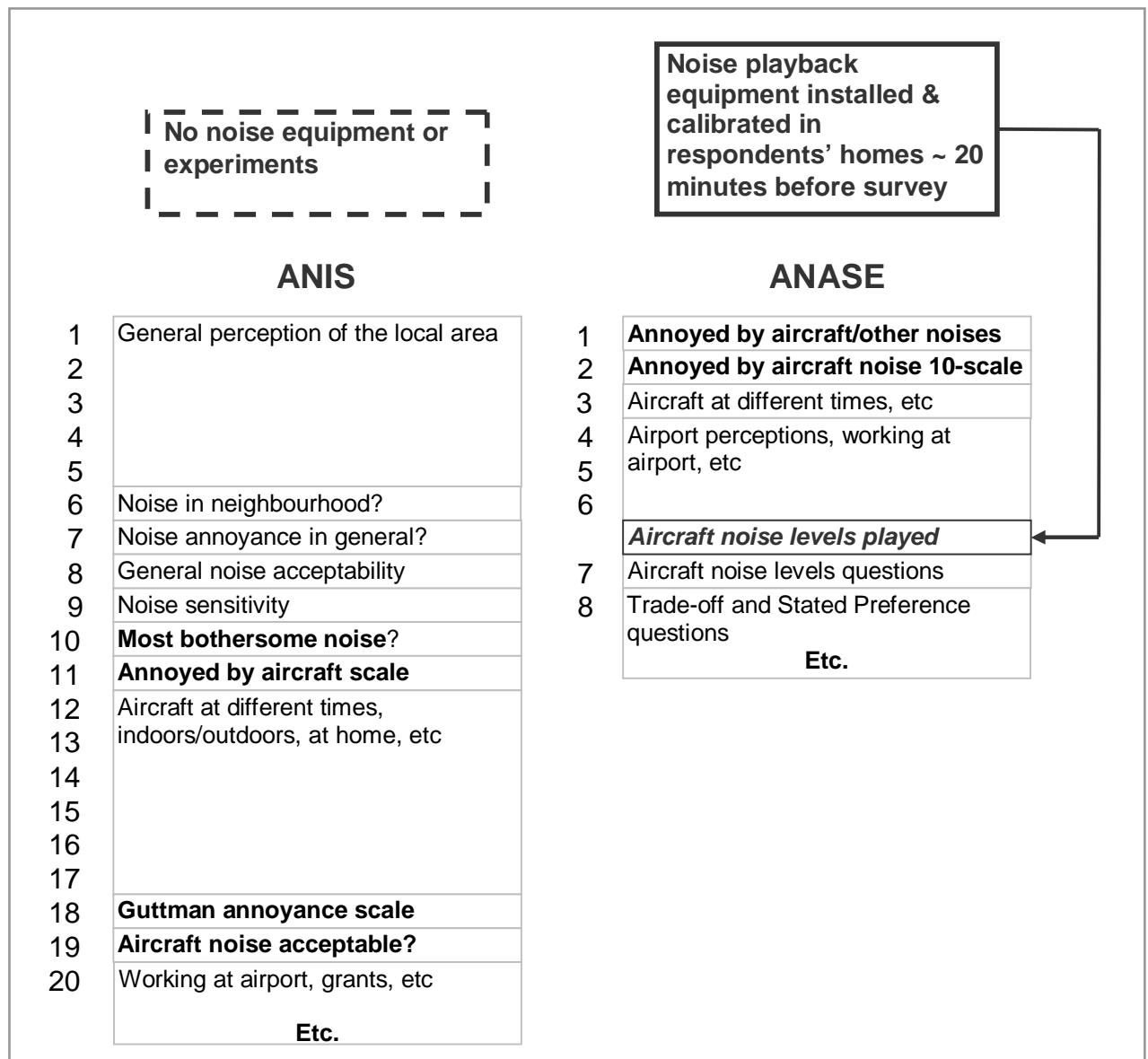


Figure 1. Key ANIS and ANASE Questionnaire context, question order and noise playback equipment differences.

ANASE used two kinds of survey sites. At one ('Full') there was the noise playback equipment of Figure 1, and at the other ('Restricted') there was no equipment. Thus, the context for the two was markedly different. If context effects are crucial in this study, then marked differences would be expected in the data from the two kinds of sites – and the data shows that they are there.

A standard international way of matching the annoyance from aircraft to people's noise exposure is to plot the percentage of survey respondents saying they are 'Highly Annoyed' against Leq or a weighted version of Leq<sup>4</sup>. The most common weighted version is Ldn (= DNL, the Day-Night Average Sound Level) used in the USA and several other countries: it is a 24-hour Leq with night flight noise levels artificially increased by 10 decibels. Figure 2 shows the '% Highly Annoyed' response for the two site types at the 27 ANASE Heathrow sites<sup>5</sup>. The Heathrow sites are selected because CAA/DfT higher accuracy Leq values for these sites are available; because it is simple to approximate internationally used DNL values (by adding 2.5 dBA to the Leq value); and to avoid airport-dependent factors. Simple linear-fit trend lines are also shown for the two sets of data.

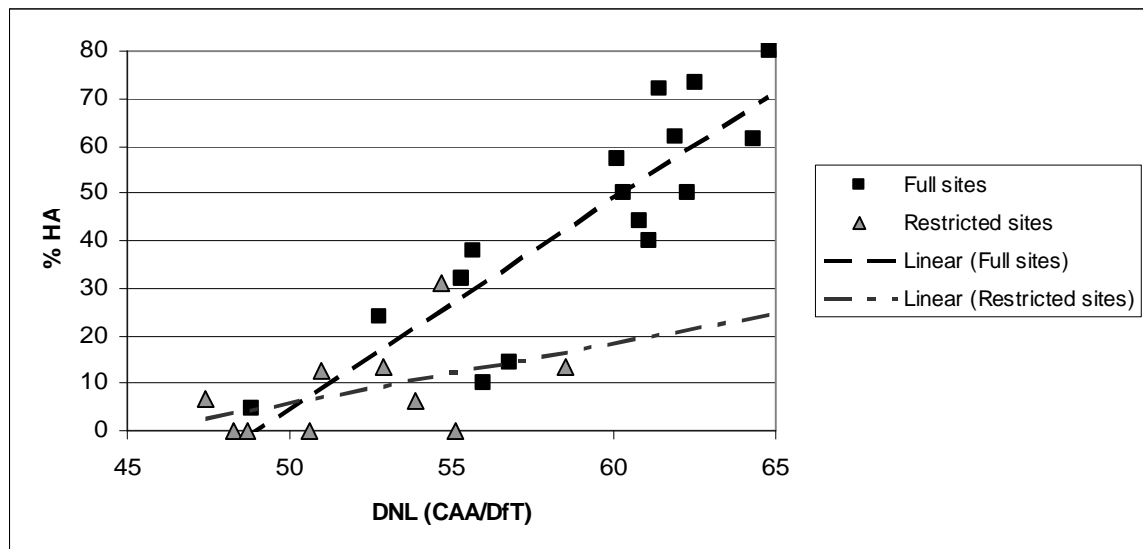


Figure 2 ANASE '% Highly Annoyed' (% HA) Heathrow results.

Figure 2 indicates that the Full and Restricted scatter plots and trend lines are very probably different – in particular the trend line slopes differ. In comparing two regression lines, the most basic hypothesis to test is the hypothesis of coincidence, i.e. if the two underlying relationships are the same. The ANASE contractors carried out statistical testing to compare Heathrow Full and Restricted data – but only at the instigation of the reviewers<sup>2</sup>. This rejected the coincidence hypothesis, finding that the differences were statistically significant (t-statistic above the standard 5% level). It is therefore unlikely that the two samples come from the same underlying population. It implies that the introduction of noise equipment changed the aircraft noise annoyance dose-response relationship, by a roughly multiplicative bias here. The ANASE contractors decided to ignore these crucial tests.

Only in circumstances when statistical testing accepts coincidence is it permissible to fit a single overall regression line to both relationships. But the ANASE statistical analysis wrongly combines Full and Restricted data sets. To ignore the statistical testing results rejecting the coincidence of the data sets is not sound practice. It removes any possible sound foundations for subsequent ANASE modelling claims about (e.g.) annoyance onsets and the weighting of the number of aircraft.

Why do the Full and Restricted data sets differ? It is not possible to offer precise reasons based on the ANASE documents, simply because the ANASE work did not investigate potential causes. One factor could be confusion between audibility/awareness of noise as compared with suffering a

degree of annoyance. The presence, and presumed intended use of the noise playback equipment, is certainly a possible strong factor.

Figure 3 shows the complete set of Full and Restricted data from ANASE (using wholly ANASE data). This again shows that there are differences between the two data sets: having noise equipment present does make a difference – showing a roughly multiplicative bias at the Full sites. The Figure also shows that ANASE Restricted sites were not wisely selected. The onus was on the ANASE contractors to select sites to be able to test effectively for Full/Restricted differences – Restricted sites at higher Leq values ('control group sites') should therefore have been included.

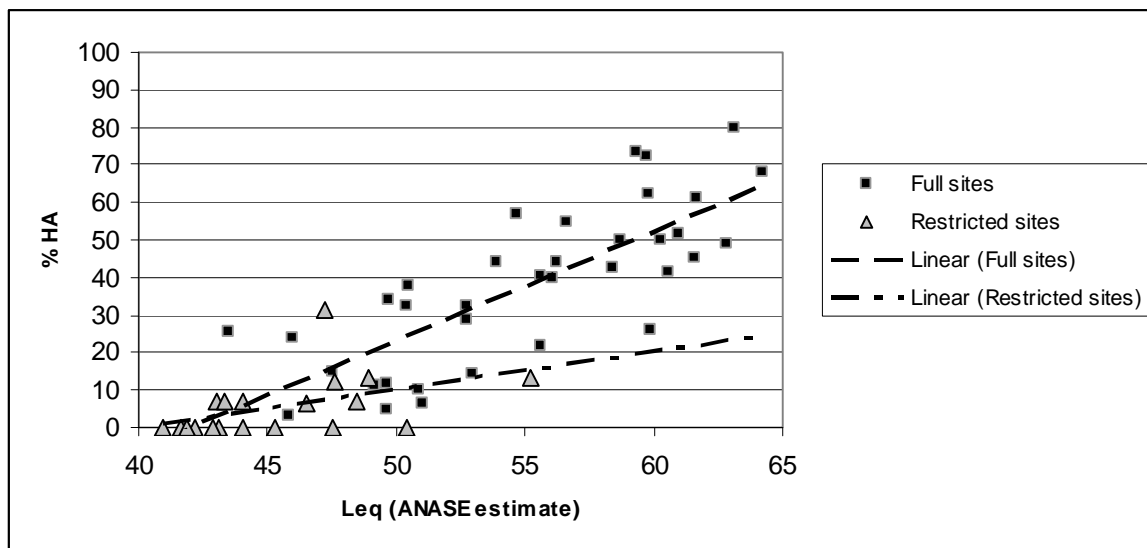


Figure 3. ANASE Full and Restricted sites % Highly Annoyed (%HA).  
[Note Site R17 excluded – as in ANASE analyses.]

ANASE data for Full sites are markedly out of line with the results of reputable international and previous UK work<sup>5</sup>. ANASE Restricted site data are broadly consistent with international and ANIS results, although could under-estimate annoyance because of the differences in the experimental/questionnaire contexts – compare Figure 1.

The conclusion is that the introduction of noise equipment changes the aircraft noise annoyance dose-response relationship by a roughly multiplicative bias factor. Thus, the ANASE results are not comparable with previous ANIS work, and the ANASE contractors' claims – e.g. increased annoyance over time, additional aircraft number effects – are invalid because they mostly derive from the biased data.

These ANASE failures would lead directly to the many problems with the SP valuations in ANASE identified by the peer reviewers. To quote one peer reviewer (Bateman): "the absolute value of such reductions [*i.e.* SP valuations] is, by the authors admission, implausibly high." The high valuations are simply explained by reference to the Figures above. Monetary valuations of noise via SP or other pricing techniques are intended to 'crystallise' people's annoyance or disturbance. All other things being equal, a bias in measuring annoyance would therefore imply a similar degree of bias in subsequent monetary valuations. The multiplicative bias on annoyance found here – from the slopes of trend lines on '% Highly Annoyed' – is roughly a factor of three. Thus, ANASE SP valuations would be expected to be about three times the values that would be obtained from aircraft noise SP studies without such biases. This can only be a rough figure – in practice, weightings and non-linear mappings probably complicate matters. But this evidence of design bias straightforwardly explains why the SP valuations are implausible (note that that the peer reviewers did not offer simple economic explanations to resolve ANASE's problems).

## 5 LESSONS ABOUT CONTRACTS

The comments here about contracts have to be general ones, because of the legal position about public comments on a specific Government contract. Contracts to carry out research studies often go wrong to some degree. To have a good chance of succeeding, and hence to provide assurance that public funds are not being mismanaged, a contract generally needs to meet some obvious positive criteria. Obvious key ingredients are:

Competent project managers

Competent contractors

Good quality technical advice

The word 'competent' means the appropriate combination of intelligence, integrity and track record. 'Good quality technical advice' means people who know the technical subject, who probably have carried out similar projects in the past themselves, and who persevere in getting answers to key project questions.

Looked at from the negative viewpoint, it is vital for the project managers to detect preventable failures. The contract may not deliver according to the specification; the work may be of poor quality; the outputs may be late and/or over-budget. Examples of the nature and causes of problems, summarised from a variety of governmental contracting guidance material, are

Improper award of contract – e.g. de-scoping specification for a particular contractor, biased contract evaluation process;

Contract not delivered properly – e.g. ineffective monitoring, no effective auditing programme;

Contract cost over-runs – e.g. 'lowballing' by contractor, Ineffective monitoring.

'Lowballing' means the contractor underquotes when bidding for the contract, hoping that the quality of supervision and monitoring will offer potentially large claims for extras and over-runs.

## 6 METHODOLOGICAL LESSONS

### 6.1 Get independent expert technical advice

There is a right and a wrong time to get independent technical advice about a project. The wrong time is after the fieldwork and analyses have been done, i.e. to have an ANASE-type peer review by experts. The right time for expert input is very early in the project, when pilot studies are being designed and analysed, so that the work goes in the right direction and avoids obvious traps. It is vital to find genuine experts and take their advice seriously: they may well not be right all the time, but their questions have to be answered. [This is the 'Frobisher lesson'. Martin Frobisher, the 16<sup>th</sup> Century English explorer, went on three voyages to northern Canada, bringing back increasingly large amounts of gold ore – 1100 tons on the last trip. But it was Fool's Gold – iron pyrites. Frobisher should have taken advice from a reliable metallurgist at the outset.]

### 6.2 Do the right annoyance study

ANASE's specification for assessing annoyance was largely a reasonable one, but failed to address the key policy decisions required for proposed Heathrow developments relating to 'mixed mode' operations<sup>3</sup>. Heathrow currently operates in 'segregated mode', with landings on one parallel east-west runway and departures on the other one. About 75% of the time the airport is operated westerly, i.e. with flights going towards the west, and 25% easterly. On westerly operations, there is runway alternation: each runway is typically used for takeoffs about half the time, morning or afternoon. There is no alternation for easterly operations, as takeoffs only take place on the

southern runway; northern runway takeoffs are not generally allowed because of the 'Cranford agreement', named after an area to the east of the northern runway.

The proposal to expand Heathrow's operations eliminates segregated mode operations, alternation, and the Cranford agreement. Instead, Heathrow would operate in mixed mode for both runways, for both easterly and westerly operations. Mixed mode means mixing departures (D) and arrivals (A) on the same runway, i.e. at peak hours a sequence A-D-A-D-A-D-... This would change both the total number of annual flights and the noise exposure patterns around the airport. For example, people in Stanwell Moor, living underneath westerly departure routeings from the southern runway, are currently exposed to a westerly mode Leq of about 74, contributed almost entirely by an Leq of about 77 from southern runway operations, and an easterly mode Leq of less than 60. With mixed mode, the westerly and easterly Leq values in projected scenarios (e.g. for 2015<sup>6</sup>) might both be about 74 [2015 traffic would have reduced average noise levels but increased numbers of flights]; but without periods of respite, either from day to day or during each day. How would these changes in noise exposure at different times affect people's disturbance?

The ANASE specification included no examinations of mixed mode/alternation/Cranford agreement. But these elements play a crucial role in the consultation on Heathrow's development: more than half of the Consultation Document's Response form<sup>2</sup> deals with these issues. The inference is that research specifications should try to meet the likely policy needs. Trying to understand the probable annoyance effects of a change to mixed mode would no doubt have been a difficult task. This is because of the very different diurnal and day-to-day noise exposure patterns around the airport, and the presence of confounding factors (in particular, work connections with the airport strongly affect responses – these connections are much more likely to the east of Heathrow because of the better road/underground communication links). However, it would have been a very worthwhile study if it ensured that policy decisions were based on an accurate reflection of potential disturbance.

### **6.3 Professionally attitude testing/noise estimation/statistical modelling**

A professional approach is vital. This does not mean anything very esoteric or technically complex. The need is to do the basic things sensibly, whilst focusing on the goal of meeting the project specification. There are several reputable textbooks and accessible governmental guidance documents about attitude testing and accepted social science methodology<sup>5</sup>. Equally, there are established multiple regression and basic statistical testing techniques<sup>7</sup>.

In designing social survey based studies, the aim must be to eliminate or reduce the potential for serious technical/statistical biases, and then in database modelling and analyses to try to uncover possible biases. Proper account must be taken of statistical testing results: they must not be disregarded. It is not good scientific practice to generate complex models – 'rewriting attitudinal acoustics' – without first taking proper care to check if straightforward reasons, in this case design bias, explain the observed data. If ANASE had been properly comparable with previous studies, the devastating laboratory experiment problems noted above would have been detected quickly. The alarm bells would have sounded that there was something very dubious in the study design.

Aircraft noise estimation from computer models is an inherently complex process, particularly because of variability in atmospheric conditions along the propagation path, so estimates need to be matched against appropriate field data collections (e.g. the large-scale programme in ANIS) and current best practice<sup>1,2,8</sup>.

### **6.4 Produce the right specification**

The ANASE specification might have generated worthwhile project outputs if efforts had been made to prevent the SP element of the study distorting the annoyance survey, which in turn distorted the SP results. Acoustics professionals know that there are major methodological problems in combining social survey and laboratory experiments. The noise research literature has examples

showing that laboratory experiments – noise playback equipment in this case – change people's disturbance reactions.

It is puzzling why DfT focused so much on SP valuations in the specification, given that hedonic pricing methods are generally used in valuing transport noise<sup>9,10</sup> – and indeed are used in the current Heathrow Consultation document. To quote from the ANASE Executive Summary:

“Overall, therefore, we do not think that the valuations from either [ANASE SP] method are safe, and it will probably be necessary to rely on sources based on Hedonic Pricing.

SP's methodological uncertainties were known to respected researchers in this field, e.g. the peer reviewers, well before ANASE was commissioned.

The implications are that the two kinds of exercise, annoyance and SP, should not be combined unless there is confidence that bias/distortion effects are eliminated or controlled; and that very large-scale SP studies to aid policy decisions should not be carried out until there is good agreement amongst expert researchers that they will produce good quality results.

## 7 CONCLUSION

ANASE's problems were predicted, and its failure to produce cost-effective outputs to help policymakers was preventable. The study was intended to be 'substantial research that commands the widest possible confidence', but it failed to achieve that central goal or to have any worthwhile impact on the current Heathrow Consultation process – ANASE's claims added nothing but confusion. Nevertheless, there are some valuable lessons to be learned from ANASE's failings regarding aircraft noise study specifications, contracts, design methodology and data analysis.

## 8 REFERENCES

- 1 P. Brooker, The UK Aircraft Noise Index Study [ANIS]: 20 Years On. Acoustics Bulletin. May/June, 10-16. (2004). <https://dspace.lib.cranfield.ac.uk/handle/1826/1004>
- 2 P. Havelock & Turner, S. W., Attitudes to Noise from Aviation Sources in England: Non SP Peer Review. Environmental Research & Consultancy, CAA; Bureau Veritas. (2007). <http://www.dft.gov.uk/pgr/aviation/environmentalissues/Anase/nonsppeerreview.pdf>
- 3 DfT. Adding capacity at Heathrow airport - Public consultation. UK. (2007). <http://www.dft.gov.uk/consultations/open/heathrowconsultation/>
- 4 S. Fidell & Silvati, L., Parsimonious alternative to regression analysis for characterizing prevalence rates of aircraft noise annoyance. Noise Control Engineering Journal, Vol 5(2), March/April, 56-68. (2004).
- 5 P. Brooker, ANASE: Unreliable – Owing to Design-Induced Biases. Acoustics Bulletin. Jan/Feb, 26-31. (2008).
- 6 D. P. Rhodes & Beaton, D., Revised Future Aircraft Noise Exposure Estimates for Heathrow Airport. ERCD Report 0705 Prepared by the Civil Aviation Authority on behalf of the Department for Transport, London. (2007).
- 7 N. R. Draper & Smith, H., Applied Regression Analysis. Wiley-Interscience; 3<sup>rd</sup> edition (1998).
- 8 P. Brooker, Critchley, J. B., Monkman, D. J. & Richmond, C., United Kingdom Aircraft Noise Index Study (ANIS): Main Report DR Report 8402, for CAA on behalf of the Department of Transport, CAA, London. (1985).
- 9 P. Brooker, Aircraft Noise: Annoyance, House Prices and Valuation. Acoustics Bulletin. May/June, 29-32. (2006).
- 10 B. Pearce & Pearce, D., Setting Environmental Taxes for Aircraft: A Case Study of the UK. CSERGE Working Paper GEC 2000-26. [http://www.uea.ac.uk/env/cserge/pub/wp/gec/gec\\_2000\\_26.pdf](http://www.uea.ac.uk/env/cserge/pub/wp/gec/gec_2000_26.pdf)