inter·noise 83

STATISTICAL SIGNIFICANCE TESTING FOR THE PROOF OF AEROPLANE NOISE CONTROL MEASURES

H.W. Dahlen

DFVLR, Abteilung Technische Akustik im Institut für Entwurfsaerodynamik, Bienroder Weg 53, 3300 Braunschweig, Germany

INTRODUCTION

Although the flyover-noise testing procedures of aeroplanes for noise certification purposes are well established, their limited data evidence for other applications, such as comparisons, are frequently subject to misinterpretation. This is particularly the case, if technical improvements leading to "acoustical changes" are to be verified by comparative tests. Without an in-depth understanding of the inherent statistical error of flyover-noise data wrong conclusions and expensive but uneffective developments may result.

CERTIFICATION VERSUS COMPARATIVE TESTING

Obviously, the simple approach of averaging a number of n maximum flyover-noise levels L_{ij} to establish a mean

$$L_{i\cdot} = 1/n \sum_{i=1}^{n} L_{ij} \tag{1}$$

and fulfilling the certification validity requirement, that the confidence interval must not exceed a limiting value (with an error probability $\alpha = 0.1$),

$$u_{o,1} = \left[\frac{1}{n(n-1)} \sum_{j=1}^{n} (L_{ij} - L_{i\cdot})\right]^{\frac{1}{2}} \cdot t_{n-1;0,1} \quad \text{of } 1.5 \text{ dB}, \tag{2}$$

where $t_{n-1;o,1}$ is the Student-factor, does not allow a comparison of two means $L_{1\cdot}$, $L_{2\cdot}$, since necessary statistics are lacking. Within a comprehensive empirical study it had been shown, how general *precision parameters* of flyover noise can be estimated, including *repeatability*, *reproducibility* and *critical differences* [1]. In order to trace small noise level differences - occurring f.i. with changes in propeller speed, blade loading, exchange of propellers or exhaust mufflers - ,basic statistic test methods on measured noise data had been applied to prove significance of acoustical changes. Some experience obtained in testing normality of noise-level series distributions through the *Wilk-Shapiro-test* [3], and a subsequent *two-sample-t-test* for *indepent random* samples, or the nonparametric *Wilcoxon-test* for *two independent* samples, were reported in [2].

IMPROVEMENTS IN SIGNIFICANCE TESTING

Statistical evaluation of measured data from flyover noise tests always suffers from the prohibitively small sample-sizes (i.e. the number of flyovers), and the lack of equality or homogeneity of variances of the different samples (flyover test series) which are to be compared. Three methods, used individually or successively combined, may offer improvements in significance testing, specifically

1. Increasing the number of flyovers n,
2. Reducing the variability of noise-level series from successive flights by matched-pair testing of aeroplanes,
3. Using the *spectral* information contained in the noise time-history instead of the customary single number value only.

The first method achieves only a limited improvement, because the maximum number of valid flyovers is usually dictated by the stability of meteorological conditions. A change, especially by wind and atmospheric turbulence, results in large variations of the measured L_{ASmax}-level series, leading to larger variances, while the standard errors of the means diminish only with $1/\sqrt{n}$. The second remedy would yield optimum results, because different aeroplane configurations would not be tested subsequently within inherently larger time spans, but rather simultaneously. Therefore paired-, or connected-samples statistic testing could be applied, with considerable reduction of the sampling error and better resolution (*t-tests for pairs* with normal distributed pair-differences or nonparametric *Wilcoxon matched signed-rank test*). Unfortunately, the large expenditure for this kind of flight- test procedure seems only feasable for aircraft manufacturers.

Since a substantial improvement apparently cannot be achieved by the flight test procedures, we ought to concentrate our attention to the evaluation of information contained in the whole flyover noise time-history, especially the (time dependent) spectra. Use of only L_{ASmax}, as a single value descriptor of the flyover-noise, seems a waste of information. As a first step, at least one spectrum for each flyover, preferably at the time of L_{ASmax}, should be determined. It must be established however that consecutive band levels within a spectrum, taken simultaneously, are random, i.e. statistically independent.
The applied procedure for spectral analysis and subsequent statistical comparison of spectral means or averages as measures of overall levels by one-way analysis of variance (AOVONE) is discussed in the following example.

PROOF OF SIGNIFICANCE FOR A SMALL MEAN NOISE-LEVEL DIFFERENCE

The case of one propeller-driven aeroplane is considered, which made

four passes each at two barely different propeller speed settings of 2500 and 2550 RPM. Fig. 1 shows the measured L_{ASmax}-levels of both series. Obviously, the resulting difference of the means of both series (groups) is very small, hardly exceeding 1 dB. A subsequent t-test for independent samples normally distributed, yields only marginal significance at an error probability level α=0.05. The figure also shows that the highest ranked noise level of the first series falls well into the range of the second series. One should, however, be cautious to assume this value to be an outlier, because in this case one mean value out of a total of only four would have to be discarded. To clarify these findings, third-octave band analyses was performed on the data. The A-weighted spectra were triggered at the time of L_{ASmax}. An averaging time of 2 seconds, equivalent to the time constant "slow" of a sound level meter, was chosen. (It should be mentioned that these settings are not necessarily an optimum for the following statistical investigations!)
Fig. 2 shows the resulting two groups of four spectra each. An inspection simply by eye-averaging over a dozen frequency bands and four spectra for both series, resp., would not yield any enlightenment. Therefore a *one-way analysis of variance* (AOVONE) for the comparison of the

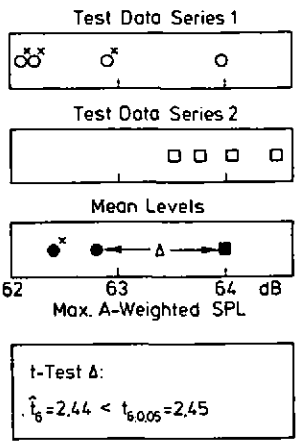

Fig.1 Dot Diagram of Two Measured L_{ASmax}-level Series (□ o), their Mean Levels (■ ●) and Significance Test of the Difference of Means Δ.

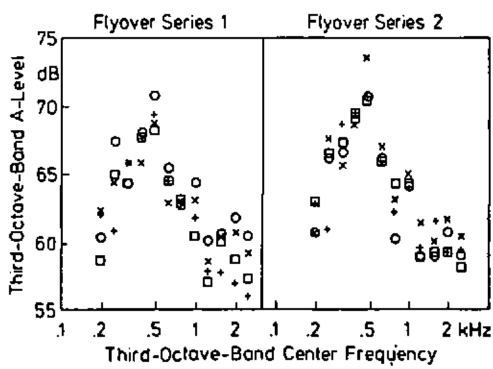

Fig.2 Variability of A-weighted Third-Octave-Band Spectra within Flyover Noise Series Obscuring Mean Overall Level Difference Δ Between Series.

individual means (and groups in a further stage) is applied, and the means or averages taken as a measure for the spectral sums because of equal sample sizes (number of bands). This AOVONE tests, whether observed differences between means are random (and therefore accidental)

or whether the sample-means indicate different populations.
The following steps are necessary for the evaluation, whereby the first three are preparational:

1. Arrangement of the spectra for both test series in columns, ranked with respect to the magnitude of their means, thus forming two groups.
2. Linear transformation of the columns by one common average spectrum in order to reduce systematic variability due to the spectrum shape.
3. Testing the columns with respect to randomness with a Neuman-Moore *mean-square successive difference test* [3].
4. Proof of equality of spectral means within each group (test-series) by AOVONE.
5. If rejection of equality occurs, then an *assessment of linear contrasts* after Scheffé [3] determines whether the diverging mean is to be discarded.
6. A final AOVONE over all means tests homogeneity. If rejected, a *least significant difference (LSD) test* proves the existance of groups [3].

The preliminary AOVONE for both series separately showed equality of means for the second flyover-series spectra, but rejection occurred for the first set. A subsequent assessment of linear contrasts indentified the fourth ranked spectrum of the first series to be the cause for disturbance of homogeneity which could be restored after discarding this flyover spectrum. A further AOVONE concerning the total remaining seven spectra led to a rejection of homogeneity of the seven means and a final *modified LSD-test* led to forming two homogeneous groups of means, the first series' three, and the remaining four, now however on a highly significant error probability level α=0.01.

CONCLUSIONS

It has been demonstrated, that flyover noise measurements can be evaluated for significance testing by frequency analysis and appropriate analysis of variance techniques, not withstanding that only a very small number of overflights were available. This is of course only possible at the expense of additional processing. The procedure may seem to be somewhat tedious, but the availibility of real-time analyzers and computers would make this statistical test procedure routine. With more appropriately processed spectral data, constant confidence limit averaging, and use of more information from the noise signal f.i., reliable statements with very high precision can be made from field noise-tests, whereby the scatter of the original data should not be discouraging.

REFERENCES

[ ] H. Dahlen and H. Heller, "Repeatability and Reproducibility of Flyover Noise Measurements of Propeller-driven Aircraft - When are Acoustical Changes Significant?", Proc. Inter Noise '81, Amsterdam, Vol.2, 1019-24, (1981).

[2] H. Dahlen, "On More Appropriate Comparative Aeroplane Noise Testing Procedures", Fortschritte der Akustik FASE/DAGA '82, Göttingen, DPG Bad Honnef, 553-56, (1982).

[3] L. Sachs, "Applied Statistics", Springer New York-Heidelberg, (1982).