THE CCAUV.A-K3 KEY COMPARISON OF PRESSURE RECIPROCITY CALIBRATION OF LS2P MICROPHONES: RESULTS AND ANALYSIS

Vicente Cutanda Henríquez[1], Knud Rasmussen[2] and Lars Nielsen[2].

[1)]ETS Ingenieros de Telecomunicación, Universidad de Vigo
Campus Universitario, E-36310 Vigo, Spain
Tel.: (+34) 986811949, Fax: (+34) 986812116
e-mail: vcutanda@gts.tsc.uvigo.es

[2)]Danish Fundamental Metrology
Matematiktorvet, B307, DK-2800 Lyngby, Denmark

Abstract: The CCAUV.A-K3 Key Comparison has involved 15 countries organized in two loops with two common laboratories, CENAM and DPLA. The measurements took place in 2003. This is the first CCAUV key comparison organized with more than one loop, and therefore the analysis of the results required a more elaborated method to arrive to the desired degrees of equivalence per laboratory and between any pair of laboratories. An implementation of the linear least squares has been applied to the analysis of the CCAUV.A-K3 key comparison. In this paper, the mathematical development is presented in parallel with the results of the comparison. The details of the analysis are reviewed, and suggestions are made in order to further improve comparison protocols in acoustics.

1. INTRODUCTION

The *Bureau International des Poids et Mesures* (BIPM), is an international organization dedicated to ensuring world-wide uniformity of measurements and their traceability to the International System of Units (SI).[1]

To achieve such uniformity, the National Metrology Institutes (NMIs) of the Member States conduct international comparisons of measurement standards, where the *degrees of equivalence* between measurements performed in different countries are established. The Consultative Committee of Acoustics, Ultrasound and Vibrations (CCAUV) is one of the committees within BIPM performing this activity.

Key comparisons represent the highest level of international comparisons. There are up to four key comparisons in the field of Acoustics at different stages of development within CCAUV. The present paper deals with the third of such comparisons, named CCAUV.A-K3.

The protocol for the CCAUV.A-K3 was approved during the 3rd CCAUV meeting in Paris in October 2002. The standards employed in this comparison are LS2P microphones (Laboratory Standard, half inch microphones) and the pressure sensitivities had to be measured using the reciprocity technique. Fifteen countries, listed in table 1, finally participated. The *Centro Nacional de Metrología* (CENAM) of Mexico was designated as the pilot laboratory conducting the exercise, with the technical assistance of the Danish Primary Laboratory of Acoustics (DPLA). One of the authors, V. Cutanda, was working at CENAM during the comparison and finished his contribution after leaving the institution in December 2004. The measurements were performed from January to November 2003.

In this key comparison two pairs of LS2P microphones were circulated in two loops. This meant that only two laboratories, the so-called *linking laboratories*, measured all four standards: CENAM and DPLA. The remaining 13 laboratories received only one of the two microphone pairs for calibration. This is a novelty in key comparisons organized by the CCAUV and required an analysis method of the results which is new in this metrology field. The main advantage of this procedure has been the shortening of the circulation time to nearly one half, reducing the risk of drift of the standards.

This paper presents the method used in the CCAUV.A-K3 and reviews some of the aspects and conclusions of the exercise.

The present status of the comparison is, at the time of the release of this paper, *approved for equivalence*, which means that the results have been reviewed and approved by the CCAUV members and the final report released. In other words, the exercise is finished.

***Table 1** Participants in the CCAUV.A-K3*

Participant NMI	Acronym	Country
Danish Primary Laboratory for Acoustics	DPLA	Denmark
National Physical Laboratory	NPL	United Kingdom
Physikalisch-Technische Bundesanstalt	PTB	Germany
National Metrology Institute of Japan	NMIJ	Japan
Laboratoire National d'Essais	LNE	France
Korea Research Institute of Standards and Science	KRISS	Korea
Central Office of Measures	GUM	Poland
Centro Nacional de Metrología	CENAM	Mexico
National Institute of Standards and Technology	NIST	United Status
National Institute of Metrology	NIM	China
National Measurement Laboratory	CSIRO	Australia
National Research Council	NRC	Canada
Ulusal Metrolojì Enstitüsü	UME	Turkey
Instituto Nacional de Metrologia, Normalizacão e Qualidade Industrial	INMETRO	Brazil
All-Russian Scientific and Research Institute for Physical-Technical and Radiotechnical Measurements	VNIIFTRI	Russian Federation

2. ANALYSIS METHOD

The analysis of the results should be able to produce:

i) degrees of equivalence per laboratory, that is, deviations from a key comparison reference value (KCRV) and their uncertainties
ii) degrees of equivalence between laboratories.

Several analysis strategies and recommendations have been published in the recent years concerning comparisons. For example, in [2] an adaptation of linear least squares is proposed for linking regional comparisons to an existing key comparison, maintaining the KCRV. In [3], a procedure to deal with different situations in a comparison is developed, using weighted means and, if necessary, median values.

In the present case, a general linear least squares analysis was chosen and applied for every one of the frequencies at which results were reported. In particular, the first stage of the analysis is performed according to reference [4]. This analysis assumes a linear behaviour of the measurement, which is reasonable for the case of CCAUV.A-K3.

3. CORRELATION OF THE RESULTS

The analysis by least squares implies providing an estimation of the correlation, or covariance, between different measurements. In previous CCAUV comparisons, like CCAUV.A-K1, the measurements performed by different laboratories were assumed uncorrelated, while the measurements made at the same laboratory on different standards were supposed to have full (equal to one) correlation. These hypotheses were reasonable and simplified very much the analysis.

There is no rigorous analysis of the correlation existing between measurements such as those performed in the CCAUV.A-K3. We must therefore assign values based on the available information. In principle, having no correlation between the measurements made by different laboratories is a reasonable idea: even if some of the participant laboratories employed similar setups for their calibrations, it does not show as a difference in the results.

On the other hand, we assign a high degree of correlation to the measurements performed by the same laboratory, considering it uses the same equipment, method and staff. However, we cannot assume full correlation; this would make impossible to obtain a result by least squares minimization. It was found that a correlation of 0,7 between measurements made at the same laboratory was reasonable enough and maintained the equivalence of the results.

In the future, the analysis of correlation should be improved in a similar way estimation of uncertainty has evolved. The different sources of correlation should be accounted for and studied properly.

4. LEAST SQUARES MINIMIZATION

The results are first analyzed according to the method in [4], which is briefly described here. Every laboratory reported two (four for CENAM and DPLA) values of sensitivity in dB re 1V/Pa per measured frequency, one per microphone. They are noted as y_i in (1). At several low and high frequencies, some laboratories did not report results. The uncertainties, or rather the variances, are combined with the covariances deduced in the previous section to form the covariance matrix *V(**y**)* in (1), which is symmetrical.

$$\mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} \tag{1}$$

$$\mathbf{\Sigma} = V(\mathbf{y}) = \begin{pmatrix} u^2(y_1) & u(y_1,y_2) & \cdots & u(y_1,y_n) \\ u(y_1,y_2) & u^2(y_2) & & u(y_2,y_n) \\ \vdots & & \ddots & \vdots \\ u(y_1,y_n) & u(y_2,y_n) & \cdots & u^2(y_n) \end{pmatrix}$$

The covariance matrix has in our case some off-diagonal non-zero elements, corresponding to the covariances of measurements by the same laboratory. The matrix *V(**y**)* could also contain the drift of the standards in a separate term. However, the microphones used in this comparison were sufficiently stable during the measurement period, as shown by measurements at DPLA before and after sending to every participant.

The model to be minimized is the matrix equation (2), which can be expressed as *E(**y**)* = ***Xa***. The matrix ***X*** is the design matrix relating the unknown values of the a_i parameters to the expectations of the measurement results. The a_i parameters are usually related to the values of the standards to be obtained by minimization. In our case, we have four a_i values per frequency.

$$\begin{pmatrix} E(y_1) \\ E(y_2) \\ \vdots \\ E(y_n) \end{pmatrix} = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1k} \\ x_{21} & x_{22} & & x_{2k} \\ \vdots & & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nk} \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_k \end{pmatrix} \tag{2}$$

The design matrix of this comparison is:

$$\mathbf{X} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{3}$$

The number of lines equals the number of available measurements at that frequency, 34 in most cases. By minimizing (2), we obtain estimates, $\hat{a}_i$, of the parameters using the expressions:

$$\hat{\mathbf{a}} = \mathbf{C}\mathbf{X}^T\mathbf{\Sigma}^{-1}\mathbf{y} \quad , \quad V(\mathbf{a}) = \mathbf{C} = \left(\mathbf{X}^T\mathbf{\Sigma}^{-1}\mathbf{X}\right)^{-1} \tag{4}$$

Through the design matrix, the expected values of the measurands are obtained along with their covariance matrix.

$$\hat{\mathbf{y}} = \mathbf{X}\hat{\mathbf{a}} \quad , \quad V(\hat{\mathbf{y}}) = \mathbf{X}\mathbf{C}\mathbf{X}^T \tag{5}$$

If the covariance matrix *V(**y**)* was diagonal, that is, with only non-zero elements at the diagonal, the least squares minimization would yield $\hat{\mathbf{y}}$ values corresponding to independent weighted means as described in reference [3]. Therefore no link between loops could be established.

By subtracting the least squares estimates of the standards from the actual measurements, and calculating the variances and covariances of the result, we obtain the differences and their covariance matrix.

$$\mathbf{y}-\hat{\mathbf{y}}=\mathbf{y}-\mathbf{X}\hat{\mathbf{a}} \quad , \quad V(\mathbf{y}-\hat{\mathbf{y}})=V(\mathbf{y})-V(\hat{\mathbf{y}}) \tag{6}$$

In (6) we have one difference per measurement and frequency. To obtain the degrees of equivalence per laboratory that we set as a goal in section 2, we need to combine these differences. At first sight, it seems reasonable to apply a second least squares minimization using $\mathbf{y}-\hat{\mathbf{y}}$ and $V(\mathbf{y}-\hat{\mathbf{y}})$ as inputs. The new design matrix would combine all the measurements by any given laboratory intro a single degree of equivalence.

However, there is a fundamental difficulty: the $V(\mathbf{y}-\hat{\mathbf{y}})$ matrix is singular, and cannot be inverted. Its rank deficiency is a result of subtracting the mean. To overcome this difficulty, one of the authors (L. Nielsen) suggested taking simple averages of the measurement deviations of each laboratory. The procedure uses an averaging matrix ***A***, with as many rows as measurements (up to 34) and as many columns as laboratories measuring at that particular frequency (up to 15). A column will have zero elements except at the places corresponding to the measurements they made, where the elements will be ½ (for those laboratories measuring two standards) or ¼ (for CENAM and DPLA, measuring all four standards). In this way we obtain the averaged differences per laboratory $\mathbf{D_{ii}}$ and their covariance matrix $V(\mathbf{D_{ii}})$:

$$\mathbf{D_{ii}}=\mathbf{A}^T\cdot(\mathbf{y}-\hat{\mathbf{y}}) \quad , \quad V(\mathbf{D_{ii}})=\mathbf{A}^T\cdot V(\mathbf{y}-\hat{\mathbf{y}})\cdot\mathbf{A} \tag{7}$$

Equation (7) gives the desired degrees of equivalence per laboratory. We take differences for every pair:

$$D_{ij}=D_{ii}-D_{jj} \quad , \quad V(D_{ij})=u_{ii}+u_{jj}-u_{ij}-u_{ji} \tag{8}$$

The u_{ij} are elements of the $V(\mathbf{D_{ii}})$ matrix in (7). The D_{ij} and their variances $V(D_{ij})$ in (8) are the inter-laboratory degrees of equivalence.

It is important to notice the absence of a single Key Comparison Reference Value (KCRV). Only the differences are obtained, and the only references are the estimated values of the four standards.

5. EQUIVALENCE AND CONSISTENCY

The least squares method described in [4] includes a test for mutual consistency of the results, taking into account the stated measurement uncertainties and correlation coefficients. The χ^2 estimator is expressed as:

$$\chi^2=(\mathbf{y}-\hat{\mathbf{y}})^T\,\Sigma^{-1}(\mathbf{y}-\hat{\mathbf{y}}) \tag{9}$$

This estimator should be distributed with a mean of *n-k*, number of measurements (up to 34) minus number of standards (four), and a variance of *2(n-k)*. If the observed value in (9) differs too much from the expected value, the results are not considered equivalent.

In the case of the CCAUV.A-K3, the probability $P\{\chi^2(\nu)>\chi^2_{obs}\}$ was only below 5% in the case of the highest frequency, 31,5 kHz, if the correlation between measurements by the same laboratory was 0,7. This is not surprising, given the difficulties to measure half inch microphones by pressure reciprocity at such a high frequency.

If we assume equivalence, it is possible to test consistency of every result and between pairs of results with the estimators:

$$d_i=\frac{\mathbf{y}_i-\hat{\mathbf{y}}_i}{u(\mathbf{y}_i-\hat{\mathbf{y}}_i)} \qquad d_{ij}=\frac{d_i-d_j}{u(d_i-d_j)} \tag{10}$$

which are distributed as N(0,1). Therefore, if $|d_i|>2$ or $|d_{ij}|>2$, the result is considered as outlying. Again, the CCAUV.A-K3 results are consistent with the exception of some measurements at 31,5 kHz.

6. RESULTS

6.1. Key comparison results

The results of the key comparison are included in the Final Report of the comparison, which is public through reference [5]. We include in figure 1 the degrees of equivalence at 1 kHz.

The results are rather grouped within a few hundredths of a decibel, with no outlying cases. This is a quite satisfactory outcome.

6.2. Recalculated results

During the 4th CCAUV meeting in September 2004, it was decided to run an exercise within the CCAUV.A-K3 key comparison. The laboratories would recalculate their results using the same set of microphone parameters, instead of those they estimated independently. The idea was to remove the uncertainty contribution of the parameter estimation. The results are not concluding: some laboratories fall closer to the expected value and others depart from it. However, it is believed that the results can be useful to isolate uncertainty contributions; therefore the analysis is left to the laboratories.

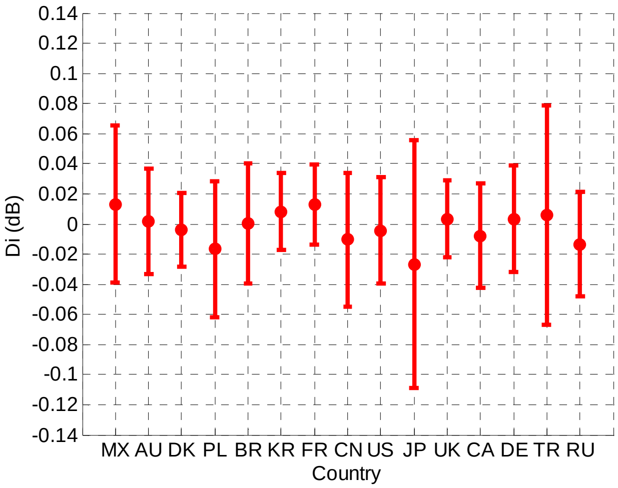

Fig. 1 *Degrees of equivalence per laboratory at 1kHz (dB).*

7. DISCUSSION

There is a balance between a shorter measurement period, with less chance of variation in the standards or measurement systems, and the difficulties for linking two loops and establish inter-laboratory degrees of equivalence. There are two factors that are important in the analysis:

- The correlation between measurements by the same laboratory must be estimated. In the present case, a general estimate of 0,7 was used. In the future, it is advisable to work out better estimations. There has been some discussion about this topic among participants, in particular regarding the use of similar equipment by some laboratories. This and other effects should be studied.

- The role of the linking laboratories is of particular relevance. Special care should be taken to improve the repeatability of their measurements, since many of the inter-laboratory degrees of equivalence are directly dependant on the link between loops they provide. Additionally, there should be as many linking laboratories as possible, in balance with time restrictions.

8. CONCLUSIONS

The use of two loops for the circulation of standards in a key comparison is a novelty in the CCAUV. On the one hand, the analysis of the results is more elaborated than a single-loop scheme. On the other hand, the time needed for the measurements is reduced to nearly one half, thus diminishing the risk of drifts in the standards and changes on the measurement systems.

Regarding the results of the comparison, they are remarkably grouped and no clear outlying cases are found. Some possible improvements to the initial protocol, like the use of an extra decimal place or the request for complete measurement descriptions, have been suggested. It is therefore expected that this comparison can serve as a reference for future comparisons at all levels.

REFERENCES

[1] Bureau International des Poids et Mesures (BIPM): www.bipm.org.

[2] C.M. Sutton. Analysis and linking of international measurement comparisons. Metrologia 41, 272–277, 2004.

[3] M.G. Cox. The evaluation of key comparison data. Metrologia 39, 589-595, 2002.

[4] L. Nielsen. Evaluation of measurement intercomparisons by the method of least squares. Danish Institute of Fundamental Metrology, DFM-99-R39, 1999.

[5] V. Cutanda Henríquez and K. Rasmussen. Final report on the key comparison CCAUV.A-K3. Metrologia 43, Technical Supplement, 2006.