An international key comparison of free-field hydrophone calibrations in the frequency range 1 to 500 kHz

Stephen P. Robinson, Peter M. Harris, Justin Ablitt, Gary Hayman, and Alex Thompson
National Physical Laboratory, Teddington, Middlesex, TW11 0LW, United Kingdom

A. Lee van Buren, Joseph F. Zalesak, and Robert M. Drake
USRD (Code 151), Naval Undersea Warfare Center, 1176 Howell Street,
Newport, Rhode Island 02841-1708

Alexander E. Isaev and Alexander M. Enyakov
VNIIFTRI, Mendeleevo, Moscow Region 141570, Russia

Christopher Purcell
Defence Research and Development Canada, 9 Grove Street, Dartmouth, Nova Scotia B2Y 3Z7, Canada

Zhu Houqing, Wang Yuebing, and Zhang Yue
Acoustics Laboratory, National Institute of Metrology, 18 Bei San Huan Dong Lu, Beijing 100013, China

Pierre Botha
CSIR—National Metrology Laboratory, P.O. Box 395, 0001 Pretoria, South Africa

Dieter Krüger
Wehrtechnische Dienstelle für Schiffe und Marinewaffen (WTD71), Berliner Strasse 115,
24340 Eckernförde, Germany

(Received 19 September 2005; revised 24 May 2006; accepted 24 June 2006)

A description is given of the results of a Key Comparison of primary free-field standards for underwater acoustics at frequencies from 1 to 500 kHz. This is the first such Key Comparison exercise in the field of underwater acoustic calibration and measurement. Laboratories from UK, Germany, USA, Russia, China, Canada, and South Africa participated by calibrating three reference hydrophones, with project coordination provided by the National Physical Laboratory, UK. The agreement between the results obtained from the comparison was generally encouraging, with the calibration values reported by the laboratories agreeing within quoted uncertainties over the majority of the frequency range, and the results generally lying within a ±0.5-dB band for frequencies up to 300 kHz. A discussion is given of the general sources of uncertainties in the calibrations, in particular those which are thought to have contributed to the differences in the results between laboratories. The results of the participants have been used to estimate the equivalence of national measurement standards within this field. [DOI: 10.1121/1.2228790]

PACS number(s): 43.30.Xm, 43.58.Vb, 43.30.Yj [KGF] Pages: 1366–1373

I. INTRODUCTION

International trade requires a secure technical foundation. To avoid disagreement between the measurements made in separate countries, there must be demonstrable equivalence between the measurement standards maintained by the individual countries. One method of determining the equivalence is to undertake an international comparison of measurements and calibrations. These exercises often indicate where improvements are required and the results may represent progress toward a more accurate realization of SI units.

In 1999, a major development in this field was the ratification of the Mutual Recognition Arrangement (MRA), which was prepared by the Comité International des Poids et Mesures (CIPM).[1] All member states of the Metre Convention are signatories to the MRA, which has the objective of providing for the mutual recognition of calibration certificates. This in turn requires an evaluation of the degree of equivalence of national measurement standards maintained by individual national metrology institutes, and the mechanism for determining this equivalence is through international comparisons. Such comparisons may be at a world level, in which case they are known as Key Comparisons and are organized under the auspices of the relevant consultative committee of the CIPM. The participants in a Key Comparison will typically include at least one representative from each regional metrology organization (RMO). Comparisons may also be organized within a region and fall within the auspices of the relevant RMO, such as EUROMET in Europe, or SIM in the Americas. These regional comparisons may then be linked to the Key Comparison via the common participant(s).

With the formation in 1999 of the Consultative Committee on Acoustics, Ultrasound, and Vibration (CCAUV) of the CIPM, a number of comparisons were commenced to fulfill the requirements of the MRA. For acoustics in air, there has been considerable activity to compare the calibrations of standard microphones. This work has been undertaken using Key Comparisons and regional comparisons, and this has already been extensively reported in the literature.[2–6]

For water-borne acoustics, the infrastructure for metrology is perhaps less mature than for air acoustics. However, at megahertz frequencies a Key Comparison of hydrophone comparisons has recently been completed,[7] this activity following on from less formal comparisons undertaken some years earlier.[8,9] In the field of underwater acoustics at frequencies less than 1 MHz, there have been few comparisons of hydrophone calibrations organized in the past. The first significant attempt was made by Trott in 1968, who organized a pioneering worldwide comparison.[10] Although other regional comparisons have been undertaken in the meantime,[11–14] until now there has been no attempt to repeat the exercise on a worldwide scale.

This paper describes the first Key Comparison for the primary free-field standards for sound in water at frequencies between 1 and 500 kHz (designated comparison identifier: CCAUV.W-K1).[15] The exercise involved participants in seven countries each calibrating the same three hydrophones. The goal of the comparison was to compare the measurement standards of the participating nations, and express their equivalence in terms of the difference between their results and the uncertainties on those differences. To achieve this, the results of the participants have been used to calculate a reference value termed the Key Comparison reference value (KCRV) at each acoustic frequency. The degree of equivalence of national measurement standards has then been calculated from the differences of the participants' results from the KCRV.

II. THE COMPARISON

A. Organization of the comparison

The comparison had seven participating countries, each represented either by the respective National Metrology Institute (NMI), or by an organization officially designated as representing the country for this exercise. For Canada, Defence Research and Development Canada (DRDC) was nominated by the Institute for National Measurement Standards of the National Research Council; for Germany, Wehrtechnische Dienstelle für Schiffe und Marinewaffen (WTD) was nominated by the Physikalisch-Technische Budesanstalt; and for the USA, the Underwater Sound Reference Division (USRD) of the Naval Undersea Warfare Center (NUWC) was nominated by the National Institute for Standards and Technology. These were joined by the National Institute for Metrology (NIM) from China, the All Russian Institute of Physical-Technical and Radio-Technical Measurements (VNIIFTRI), the Council for Scientific and Industrial Research from South Africa (CSIR), and the National Physical Laboratory from the UK (NPL). The participants are listed in Table I along with their country and relevant regional metrology organization (RMO). These are EUROMET (the metrology organization for Europe), APMP (Asia Pacific Metrology Programme), COOMET (Coopération Métrologique), SADCMET (South African Development Community Cooperation in Measurement Traceability), and SIM (Sistema Interamericano de Metrologia). Also listed in Table I are the country codes used to identify the results in the graphical representations later in this paper. The full contact details of each participant are listed with the author affiliations.

TABLE I. Participants in the comparison.

Institute	Country	Country code	Regional metrology organization (RMO)
DRDC	Canada	CA	SIM
NIM	China	CN	APMP
WTD	Germany	DE	EUROMET
VNIIFTRI	Russia	RU	COOMET
CSIR	South Africa	ZA	SADCMET
NPL	United Kingdom	UK	EUROMET
USRD/NUWC	United States	US	SIM

The pilot laboratory for the project was NPL, which undertook the initial assessment and calibration of the hydrophones, prepared the protocol document for the comparison, and performed checks on the hydrophone sensitivities between the calibrations by participants to ensure that the hydrophone sensitivities were stable. The hydrophones were circulated to the participants in a round-robin fashion, with the devices being returned to NPL after each set of calibrations by the participants.

B. The hydrophones

The hydrophones chosen for the comparison were an H52 hydrophone manufactured by USRD-NUWC in the USA; a B&K8104 hydrophone manufactured by Brüel & Kjær in Denmark; and a TC4034 hydrophone manufactured by Reson in Denmark. The devices used for the calibration are listed in Table II along with the frequency over which calibrations were undertaken. Each participant calibrated the three hydrophones at approximately 40 discrete acoustic frequencies for each hydrophone in the range 1 to 500 kHz. These hydrophones were chosen because they were routinely in use as standard measuring hydrophones within laboratories.

TABLE II. Details of the three hydrophones used in the comparison.

Hydrophone type	Manufacturer	Frequency range (kHz)	Nominal sensitivity (1 kHz) (dB re: 1 V/μPa)	Integral preamplifier
H52	USRD	1–100	−177	Yes
8104	Brüel & Kjær	10–150	−205	No
4034	Reason	100–500	−218	No

TABLE III. The range of overall standard uncertainties quoted by participants (expressed as relative uncertainties in percent). Note that the high value for participant ZA is only applicable for the highest frequencies close to 500 kHz.

UK	DE	US	RU	CN	CA	ZA
2.5–3.9	5.0	2.0–4.6	2.2–3.5	2.7–4.1	3.6–6.3	5.9–28.8

C. Calibration methods

Participants were asked to perform an absolute measurement of the end-of-cable free-field open-circuit sensitivity for each hydrophone using their own in-house methods and procedures for the calibrations. The "left-hand" *XYZ* coordinate system suggested by IEC 565 was adopted,[16] with each participant asked to align the hydrophones such that an alignment mark on the hydrophone body pointed in a direction parallel to the direction of propagation of the incoming acoustic wave. NPL prepared and circulated a protocol document describing the measurements required.

The method of calibration used by participants was the method of three-transducer spherical-wave reciprocity. By use of this method, which is described in IEC 565,[16] a hydrophone may be calibrated absolutely by making purely electrical measurements. Most commonly, participants used laboratory tank facilities of varying sizes, the largest being $15 \times 7.5 \times 7$ m and the smallest dimension of any of the test tanks used being 4.5 m. All of the tanks had a framework or traversing system used for mounting and positioning the transducers. One participant used an open-water facility on a lake, which had a water depth of 11 m, a laboratory platform being created using a pier or pontoon-based structure from which transducers may be lowered into the water. For all participants, discrete-frequency tone-burst signals were employed, with reflections isolated from the direct-path signal by use of gating and time-windowing techniques.

D. Uncertainties

Each participant was requested to provide a value for the overall uncertainty for the calibrations assessed according to the ISO document *Guide to the Expression of Uncertainty in Measurement*.[17] These are summarized in Table III. In addition, each participant was requested to provide a breakdown of their uncertainties in two categories: type A components which are estimated by statistical methods through repeated calibrations (sometimes termed the repeatability or "random" uncertainty); and type B components which cannot be estimated by repeated calibrations (often these components can be considered to express any systematic "bias" in the calibrations). Since each participant had used a slightly different implementation of the reciprocity method, the sources of type B uncertainty and the values of the individual components varied. In general, the values quoted varied with frequency, the full range of uncertainty values being shown in the table.

TABLE IV. Variations in the results of check calibrations performed by NPL for the hydrophones. The type A uncertainty is the repeatability of the results expressed for a confidence level of 95%.

Frequency (kHz)	Overall mean (dB re: 1 V/μPa)	Maximum deviation (dB)	Minimum deviation (dB)	Mean type A uncertainty (dB)
		H52 hydrophone		
5	−177.43	0.14	−0.18	0.21
10	−177.94	0.10	−0.20	0.19
20	−178.28	0.12	−0.24	0.19
50	−177.66	0.19	−0.23	0.21
100	−179.64	0.24	−0.18	0.22
		B&K 8104 hydrophone		
10	−206.31	0.10	−0.14	0.25
20	−206.68	0.18	−0.21	0.25
50	−203.65	0.20	−0.22	0.27
100	−211.31	0.18	−0.16	0.21
150	−218.71	0.23	−0.24	0.22
		TC4034 hydrophone		
100	−218.71	0.12	−0.22	0.21
200	−218.36	0.19	−0.30	0.21
300	−216.04	0.20	−0.22	0.21
400	−219.26	0.27	−0.27	0.25
500	−230.00	0.32	−0.37	0.31

III. RESULTS

A. Hydrophone stability

As coordinating laboratory, NPL undertook "check calibrations" on each hydrophone in between the calibrations of the other participants in order to monitor the stability of the hydrophones. The check calibrations were undertaken using the same procedure for full free-field calibrations of the hydrophones at NPL, but with measurements made only at selected frequencies (to save time). For each hydrophone, the results of the eight check calibrations are presented in tabular form in Table IV showing the maximum and minimum deviation from the mean occurring in the results of the check calibrations, and the mean repeatability (type A uncertainty) for the NPL calibrations. The typical type A standard uncertainty for the check calibrations was of the order of between 0.5% and 1%, calculated from at least four repeated calibrations. To express this for a confidence level of 95%, the standard uncertainty was multiplied by a coverage factor derived from the Student's *t*-factor for the appropriate degrees of freedom. The mean value for the uncertainty of all the check measurements was calculated and is shown in the table, expressed in decibels. This allows some judgment to be made regarding the significance of the variations observed in the check calibrations.

The results of the check calibrations showed that the hydrophones may be considered stable for the purposes of the comparison exercise. The maximum variation in the check calibrations at each frequency was generally within or of the same order as the repeatability of the NPL measurements, and although the deviation slightly exceeded the uncertainty at some of the frequencies, this was not considered significant. There was some evidence that there may have been a gradual increase in the sensitivity of the H52 of

0.01 dB per month during the comparison, but no corrections were applied since this was considered a marginal variation. A comprehensive description of all of the checks on stability is given in the Key Comparison Final Report.[15]

B. Analysis methodology

To facilitate a comparison between participants, the results provided by the participants were used to derive a reference value that provides a type of "overall grand mean" of results. This reference value was termed the Key Comparison reference value (KCRV). This was calculated at each acoustic frequency using a weighted mean approach.[18] In this approach, the results of the individual participants are weighted according to the inverse of the square of the stated uncertainty. For the results, x_i, from a given device with associated uncertainties, u_i, where i is the index for a particular laboratory, w_i is the weight applied to the participant results, and $i=1,\ldots,N$, then the weighted mean, y, is evaluated from

$$y=\frac{\sum_{i=1}^{N} w_i x_i}{\sum_{i=1}^{N} w_i}, \quad w_i=\frac{1}{u_i^2}, \tag{1}$$

with the associated uncertainty on the weighted mean, $u(y)$, determined from

$$\frac{1}{u^2(y)}=\sum_{i=1}^{N}\frac{1}{u_i^2}. \tag{2}$$

This analysis was applied to the results at each frequency for each hydrophone to derive the KCRVs for each device. For this comparison, the actual values of the KCRVs have little inherent value in themselves, being merely the sensitivities of some arbitrarily chosen hydrophones. Their value is purely in their role in evaluating the degrees of equivalence.

The weighted mean has the advantage that it makes use of all of the data provided by the participants, including the quoted uncertainties as well as the sensitivity values. A test was performed to determine whether the observed value of the chi-squared statistic was significant at the 95% confidence level. The test demonstrated overall consistency of the data with the weighted mean model, demonstrating that the weighted mean was an acceptable model to use for this data set. The test was passed for all except 11 of the 117 frequency points measured, with five of the failures occurring at the five highest frequencies of calibration for the H52 hydrophone. In this range of 80 to 100 kHz, the hydrophone is being used at the limit of its useful range, and this led to exclusion of these results from the final analysis. For comparison, the KCRVs were also calculated using two other estimators: the median and the unweighted mean.[15] The results showed close agreement, with the median and unweighted mean agreeing with the weighted mean to within the uncertainties of the weighted mean at all frequencies for all hydrophones. This shows that for the data in this comparison, the values of KCRV were not highly sensitive to the choice of reference value estimator.[15] None of the results provided by participants were classified as outliers.

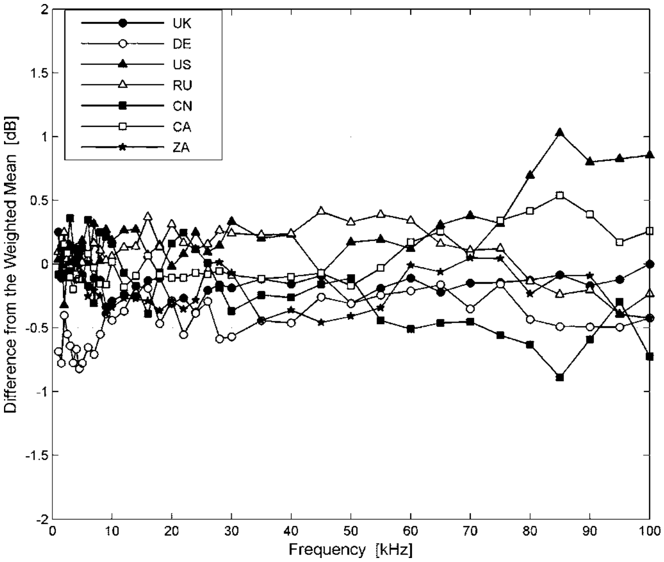

FIG. 1. The differences from the KCRV (weighted mean) for the H52 hydrophone.

C. Differences from the reference values

The results of the calibrations are presented in Fig. 1, Fig. 2, and Fig. 3 as differences from the KCRV for each hydrophone. The agreement between the results was generally encouraging, with the calibration values reported by the laboratories agreeing within quoted uncertainties over the majority of the frequency range, and the results generally lying within a ±0.5-dB band for frequencies up to 300 kHz, a factor of 2 improvement on the spread of results obtained in the 1998 EUROMET comparison.[13]

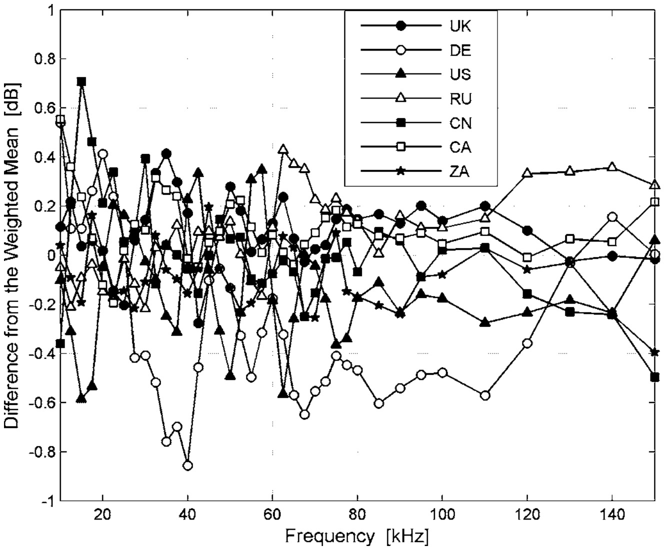

FIG. 2. The differences from the KCRV (weighted mean) for the B&K8104 hydrophone.

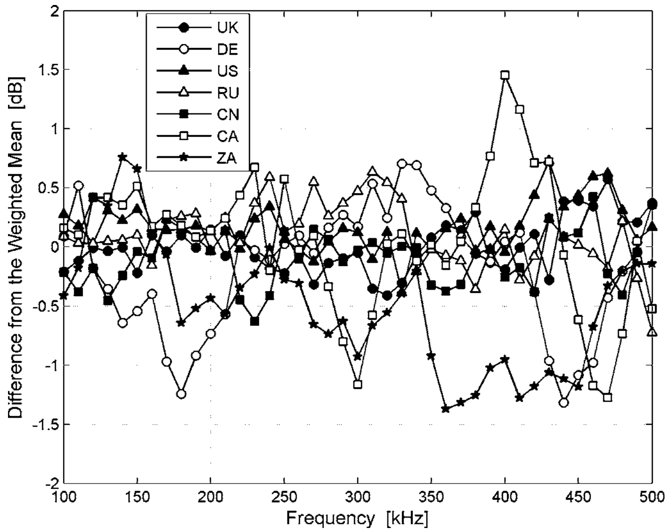

FIG. 3. The differences from the KCRV (weighted mean) for the TC4034 hydrophone.

D. Equivalence between participants

To estimate the equivalence between the calibrations of different laboratories requires that some measure of the difference in their results be calculated. This was done by calculating the degree of equivalence for each participant. For the purposes of this comparison, the term *degree of equivalence* of measurement standards is taken to mean the degree to which a standard is consistent with the reference value. The degree of equivalence of each national measurement standard is expressed quantitatively by two terms: its deviation from the reference value, and the uncertainty of this deviation (at the 95% level of confidence).

The degrees of equivalence (DoEs) of laboratory i were evaluated from

$$d_i = x_i - y, \tag{3}$$

where x_i and y are defined in Eqs. (1) and (2). The associated standard uncertainty $u(d_i)$ is determined from

$$u^2(d_i) = u^2(x_i) - u^2(y). \tag{4}$$

For frequencies at which more than one hydrophone was calibrated, an estimate of the DoE was available for each of the hydrophones used. Of the 94 distinct frequencies of measurement covered by the three devices, a calibration of more than one device had been undertaken at only 18 of these frequencies (less than 20% of the frequency points). It was necessary to combine these multiple estimates to provide a single value of the degree of equivalence at each frequency point.

To do this, values of the relative DoE were used since the reference values for each hydrophone varied appreciably in magnitude from hydrophone to hydrophone. For example, at 100 kHz the reference value for the H52 hydrophone was 1315.7 μV/Pa, whereas for the B&K8104 it was 48.2 μV/Pa, and for the TC4034 it was 11.8 μV/Pa. The relative DoE, r_{ij}, is therefore expressed as $r_{ij} = d_{ij}/y_j$, where i denotes the index of the laboratory and j the index of the hydrophone. The method used to perform the combination was to calculate the weighted mean r_i of the estimates r_{ij} with allowance made for the correlation between those estimates arising from mutual dependencies between the measurements made of the different hydrophones by the same laboratory. The estimate r_i was obtained as the solution to the generalized least-squares problem,

$$\min_{r_i}(\mathbf{r}_i - A r_i)^T V_i^{-1}(\mathbf{r}_i - A r_i), \tag{5}$$

in which, for the case of two hydrophones, $\mathbf{r}_i = (r_{i1}, r_{i2})^T$ contains the estimates for the hydrophones with associated uncertainty (covariance) matrix V_i of size 2×2, and $A = (1,1)^T$. The uncertainty budgets provided by each laboratory were used to identify those components of uncertainty that described systematic effects that were common to its calibration of the different hydrophones. These uncertainties were then used to evaluate the covariances associated with the measurements (and, hence, to derive the off-diagonal elements of covariance matrix V_i).

Figure 4 shows the combined DOEs for 8 selected frequencies out of a total of 94 frequencies in the range 1 to 500 kHz, namely: 3, 10, 50, 100, 200, and 350 kHz. Note that Canada did not undertake calibrations at frequencies less than 2 kHz, and South Africa did not undertake calibrations at frequencies less than 3 kHz, so it was not possible to calculate DOEs for these participants at frequencies below these limits.

IV. DISCUSSION

A number of factors may influence the results of calibrations and contribute to the variations in results between participants, either because of unavoidable differences in environmental conditions, or because of differences in the procedures used by the participants. Potential influences include: poor alignment, lack of acoustic far-field conditions, the length of wetting and soaking time for the hydrophones, the lack of steady-state conditions, interference from boundary reflections, the influence of noise, and electrical loading by cables/amplifiers. Other influences include the effect of the mounting or rigging used with the hydrophones, and the variation of water temperature and depth of immersion between participant calibrations.

Some of the influences listed above are related to the calibration methodology and were controlled by specifying the procedure for measurements very carefully in the protocol document. Each participant produced an uncertainty budget for the calibrations that included estimates for the uncertainty contributions due to these influences. However, some of the influences listed relate to environmental effects.

It is known that the mounting configuration used can affect the measured sensitivity for some hydrophones, and this may contribute to the variation in results. Ideally, the mount should not cause any reflections or reverberations in the acoustic signal, but should be rigid enough to allow precise positioning of the hydrophones. Fortunately, there was a general similarity between the mounting arrangements used by participants, with many using some form of free-flooded tube made of metal or plastic. However, the variations in the mounting almost certainly contributed to the discrepancies

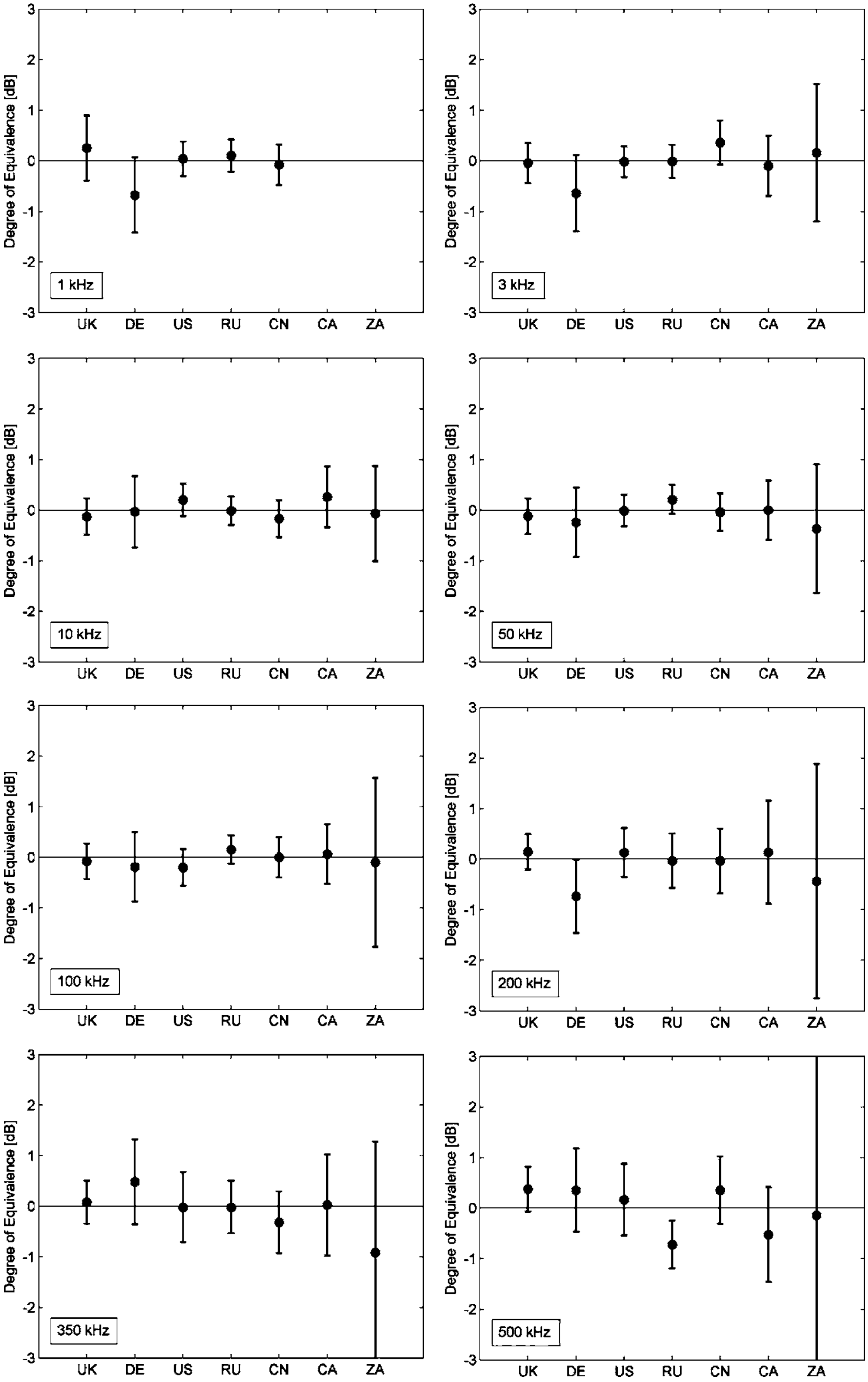

FIG. 4. Degrees of equivalence for frequencies of 1, 3, 10, 50, 100, 200, 350, and 500 kHz. Values are stated in dB and expanded uncertainties are expressed for a confidence level of 95%.

observed between the results of participants. In future comparisons, it may be better to prescribe the type of mount to be used, or perhaps circulate the mounts with the hydrophones under test.

There is a general trend for increasing spread in the results as the frequency is increased from 200 to 500 kHz. This was expected and is mostly likely due to the increased difficulty of undertaking free-field calibrations at higher frequencies. This occurs because the devices become more directional, leading to more difficulties with alignment; in addition, devices used beyond their main resonance can become less reciprocal in their behavior, a necessary condition of the calibration method; finally, noise can become issue if the transmitting response of the device rolls off at very high frequencies.

For this comparison, the depths of immersion during calibrations by different participants ranged between 1.8 and 4.0 m and the water temperatures ranged from 14.0 °C to 21.1 °C. For the range of depths employed by the participants in this comparison, it is highly unlikely that the depth variation has significantly influenced the results.

However, the water temperature is likely to have affected the results and contributed to some of the discrepancies observed. The majority of the measurements by participants were undertaken within a ±2 °C range centered around 19.0 °C, with the exception of two participants who undertook a portion of their measurements at lower temperatures, leading to an extreme case of a ±3.5 °C variation in the temperature (for a limited portion of the data on the hydrophones). Some data were available on the typical variation in response of the hydrophone models used here from measurements made at NPL[19] and at USRD,[20] and extra measurements were made at NPL during this work.[15] This extensive evaluation showed that the TC4034 and the H52 hydrophones were remarkably stable. For these hydrophones, even taking the most extreme variation in water temperature, the worst case variation in the responses of the hydrophones at any frequency was only 0.2 dB for the H52 hydrophone, and 0.25 dB for the TC4034 hydrophone. However, the results for the variation in response of the B&K8104 with temperature showed less stability. The results were highly complex, depending strongly on the acoustic frequency, and showing some evidence of nonlinear behavior with temperature. For this hydrophone, taking the most extreme variation in water temperature, the worst case variation of response at any frequency was 0.5 dB. However, this a pessimistic estimate based on a worst case—for the majority of the results the effect is likely to be much smaller. The participants' calibration data were not corrected for temperature because the participants felt that the variation of sensitivity with temperature was not known accurately enough, this being the first time such data have been measured as a function of frequency. In future comparisons of this type, it is recommended that particular attention be given to both controlling the temperature during calibrations (so that it falls within a narrow range for all participants), and to choosing hydrophones which are stable with temperature. The results of the work to measure the variation in response with temperature have been published elsewhere[15,19,20] and may be the subject of further publications.

Although there was some general agreement in the value of the overall uncertainty, the values ascribed to the components of uncertainty within the uncertainty budgets of participants varied significantly. In the view of the participants, further work to identify and quantify these uncertainties is justified.

V. CONCLUSION

The comparison has been extremely valuable, with much confidence gained in the performance of primary standards. This is the first such Key Comparison exercise in the field of underwater acoustic calibration and measurement, with expert laboratories from UK, Germany, USA, Russia, China, Canada, and South Africa participating. The results obtained represent the current state of the art for hydrophone calibrations in the frequency range 1 to 500 kHz. The calibration values reported by the laboratories agreed within quoted uncertainties over the majority of the frequency range, and the results generally lay within a ±0.5-dB band for frequencies up to 300 kHz, a factor of 2 improvement on the spread of results obtained in the 1998 EUROMET comparison. At higher frequencies, the spread of results increases to almost ±1 dB at 500 kHz. The uncertainties quoted by participants were typically of the order of 0.5 dB when expressed for a confidence level of 95%, increasing somewhat at high frequencies. Although most participants quoted similar uncertainties, the values of individual components varied considerably. The effect of water temperature variation is likely to have contributed to some of the discrepancies observed, particularly for the B&K8104 hydrophone.

The results demonstrate equivalence (within the quoted uncertainties) between the measurement standards maintained by the individual countries. The results represent a significant step forward in establishing the infrastructure for metrology in underwater acoustics, a field where metrology is relatively immature compared to airborne sound. The coverage of the comparison may be extended by conducting regional comparisons which include as a participant one of the laboratories taking part in the Key Comparison. These regional comparisons may then be linked to the Key Comparison via the common participant(s).

The results available in the database on the BIPM website also include tabulated data on the bilateral degrees of equivalence between the countries along with associated uncertainties. The data generated by the comparison are available on the BIPM website at ⟨http://kcdb.bipm.org/appendixB⟩.

ACKNOWLEDGMENTS

NPL acknowledges the support of the National Measurement System Directorate of the UK Department of Trade and Industry for the financial support of this work. Of course, the comparison would not have been possible without the effort of participants, and the authors would like to express gratitude to all those who took part. In addition, the support and guidance of the CCAUV (in particular Dr. Penny Allisy-Roberts, Secretary, CCAUV) and the BIPM is gratefully acknowledged.

[1] BIPM, *Mutual Recognition of National Measurement Standards and of Calibration and Measurement Certificates Issued by National Metrology Institutes* (Bureau International des Poids et Mesures, Paris, October 1999).

[2] R. G. Barham, "A BIPM/CIPM key comparison on microphone calibration—defining the state of the art," J. Acoust. Soc. Am. **112**, 2340 (2002).

[3] R. G. Barham, "Key Comparison: Report on key comparison CCAUV.A-K1," Metrologia **40**, 9002 (2003).

[4] V. Nedzelnitsky, "Issues concerning international comparison of free-field calibrations of acoustical standards," J. Acoust. Soc. Am. **112**, 2340 (2002).

[5] G. S. K. Wong and L. Wu, "Microphone interlaboratory comparison in the Americas," J. Acoust. Soc. Am. **112**, 2340 (2002).

[6] G. S. K. Wong and L. Wu, "Interlaboratory comparison of microphone calibration," J. Acoust. Soc. Am. **115**, 680–682 (2004).

[7] B. Zeqiri and N. D. Lee, "A BIPM/CIPM key comparison covering the calibration of ultrasonic hydrophones over the frequency range 1 MHz to 15 MHz," J. Acoust. Soc. Am. **112**, 2342 (2002).

[8] R. C. Preston and S. P. Robinson, "European comparison of ultrasonic hydrophone calibrations," Metrologia **36**(4), 345–349 (1999).

[9] R. C. Preston, D. R. Bacon, S. S. Corbett III, G. R. Harris, P. A. Lewin, J. A. McGregor, W. D. O'Brien Jr, and T. L. Szabo, "Interlaboratory comparison of hydrophone calibrations," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **35**, 206–213 (1988).

[10] W. J. Trott, "International standardization in underwater sound measurements," Acustica **20**, 169–181 (1968).

[11] S. P. Robinson, R. C. Preston, and G. J. Green, "An intercomparison of hydrophone calibrations within Europe," J. Acoust. Soc. Am. **103**(4), 2755 (1998).

[12] S. P. Robinson, R. C. Preston, and G. J. Green, "An intercomparison of hydrophone calibrations within Europe," in Proceedings of the 16th International Congress on Acoustics and 135th Meeting of the Acoustical Society of America, Seattle, WA, **1**, pp. 59–60 (1998).

[13] S. P. Robinson, G. J. Green, R. C. Preston, L. Peirlinckx, L. P. Kofoed, C. skodborg, A. Roy, Y. Mori, A. Brenner, D. Krüger, S. Buogo, G. Cannelli, L. Troiano, C. Runborg, and G. Gooch, "International comparison of free-field hydrophone calibrations in the frequency range 10–315 kHz," Metrologia, **36**, 287–296 (1999).

[14] A. M. Enyakov, S. M. Likhatchev, V. A. Platonov, W. J. Yuan, Y. B. Wang, and J. Q. Li, "A Russian-Chinese international comparison of hydrophone calibration methods," Metrologia **36**, 297–303 (1999).

[15] S. P. Robinson, "Final Report for Key Comparison CCAUV.W-K1: Calibration of hydrophones in the frequency range from 1 kHz to 500 kHz," NPL Report DQL-AC 009 (2004).

[16] IEC 565:1977, "The calibration of hydrophones," (International Electrotechnical Commission, Geneva, 1977).

[17] BIPM, IEC, IFCC, ISO, IUPAC, and OIML, *Guide to the Expression of Uncertainty in Measurement*, 2nd ed. (International Organization for Standardization, Geneva, (1995).

[18] M. G. Cox, "The evaluation of key comparison data," Metrologia **39**, 589–595 (2002).

[19] G. A. Beamiss, S. P. Robinson, G. Hayman, and T. J. Esward, "Determination of the variation in free-field hydrophone response with temperature and depth," Acust. Acta Acust. 88, 799–802 (2002).

[20] A. L. Van Buren, R. M. Drake, and A. E. Paolero, "Temperature dependence of the sensitivity of hydrophone standards used in international comparisons," Metrologia, **36**(4), 281–285 (1999).