NPL
National Physical Laboratory

Counting on IT

A National Measurement Newsletter

Issue 21

Contents:

Doping in sport

Guidelines for interpreting uncertainty in doping test measurement.

The outcome of a doping test on an athlete depends on the permitted concentration of a substance and the test result. Suppose the test indicates a concentration of 3.2 ng/ml. The limit set by WADA, the World Anti-Doping Agency (http://www.wada-ama.org/en/), is 2 ng/ml. It seems that the athlete has tested “positive”. But how confident are we in the measurement made, since no test is perfect? Depending on the result, a player’s career could be affected or ruined. The other side of the coin is that a result could unfairly stand.

Therefore, it is necessary to make a statement about the *quality* of the measurement, respecting the difficulty of measuring small concentrations. The WADA limit *is* small as regards accurate measurement, but a concentration appreciably greater would have a performance-enhancing effect. Is therefore 3.2 ng/ml “appreciably greater” than 2 ng/ml? WADA guidelines recognise the difficulty, stating that measurement uncertainty should be taken into account in establishing whether a substance exceeds a limit.

On the basis of the measurement and an understanding of the equipment used, the test laboratory makes a statement about the set of values likely to be taken by the unknown concentration. In the above case it was concluded that this set of values ranged from 1.7 ng/ml to 4.7 ng/ml for a 95% coverage probability. Because the lower endpoint lay *below* the limit, the interpretation, taking WADA guidelines into account, is that the result is *not positive*. Evidently, for a somewhat larger test concentration of, say, 4.2 ng/ml, and the same dispersion, the result *would* be positive.

A statement such as the above concerning a set of values within which a quantity can lie forms a major part of metrology, and is known as an expression of uncertainty associated with a measured value, for which the accepted “Bible” is the *Guide to the Expression of Uncertainty in Measurement* published by the International Standards Organisation. The smaller the uncertainty, the greater the possibility of obtaining a clearer statement of whether a concentration lies above or below a limit. The number of cases in which there is genuine doubt would be decreased. Metrologists strive to reduce measurement uncertainty for the general benefit of the community.

For more information please contact:
Maurice Cox
020 8943 6096
maurice.cox@npl.co.uk

Natural features

What is it that tells us that a product is "natural" and how can naturalness be quantified?

Many products fail in the marketplace because people don't like the look or feel of them. How can a designer predict how people will react to a product? What perceptual and cognitive mechanisms does a person employ in deciding to buy one product over another (other than looking at the price label)? For many years, metrologists have been interested in human sensory systems, particularly vision systems, and how they respond to physical stimuli. Now they are trying to answer more difficult questions relating to perception.

The Measurement of Naturalness project (MONAT) is looking at how people assess whether a product (for example: wood, stone, fabric) is 'natural' on the basis of visual and tactile information. The project has three main experimental elements: (a) measurements of a set of materials to determine physical properties such as spectral reflectance, thermal conductivity and hardness, (b) psychophysical studies in which subjects rate the naturalness of a range of samples, and (c) functional magnetic resonance imaging experiments in which the areas of brain activity are recorded while a subject views or feels a sample. On the basis of these measurements, we hope to develop a better understanding of how people perceive products as natural or synthetic [Figure 1].

A major component of the project is developing classification methods that will use a set of physical measurements of a sample to predict how people will perceive the sample as natural or not. Classification methods generally have two elements, the extraction of relevant features from the data and a decision scheme that performs the classification on the basis of which features are present or absent. In the MONAT project, we are particularly interested in perceptually relevant features, i.e., features that human sensory systems are able to detect. For example, images of natural and synthetic wood [Figure 2] may well have features at a microscopic level that indicate whether the image comes from a natural wood or a synthetic wood. An image analysis algorithm could successfully classify on the basis of these microscopic features but we know that human visual perception cannot be using these features.

Natural wood has the larger scale features that we expect to see, variation in grain, knots, changes in colour, a lack of regularity. How can feature extraction algorithms quantify the graininess, lack of regularity, etc? Computer vision science has developed a range of texture analysis algorithms for image feature extraction, such as co-occurrence matrices, Gabor filters and wavelet analysis. The MONAT project is developing these algorithms along with measures of structure and heterogeneity in order to capture those features that seem most relevant to the perception of naturalness.

MONAT is a EU funded project within the Measuring the Impossible theme of the New and Emerging Science and Technology Programme in Framework 6.

For more information please contact Alistair Forbes:
020 8943 6348
alistair.forbes@npl.co.uk

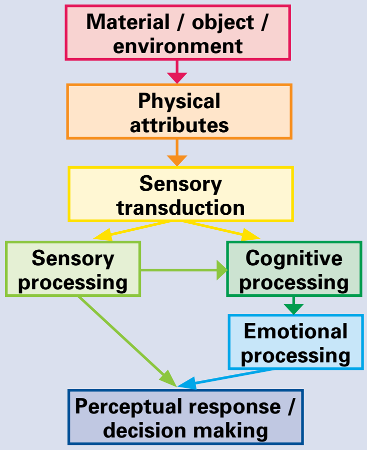


Figure 1 The Perceptual Process

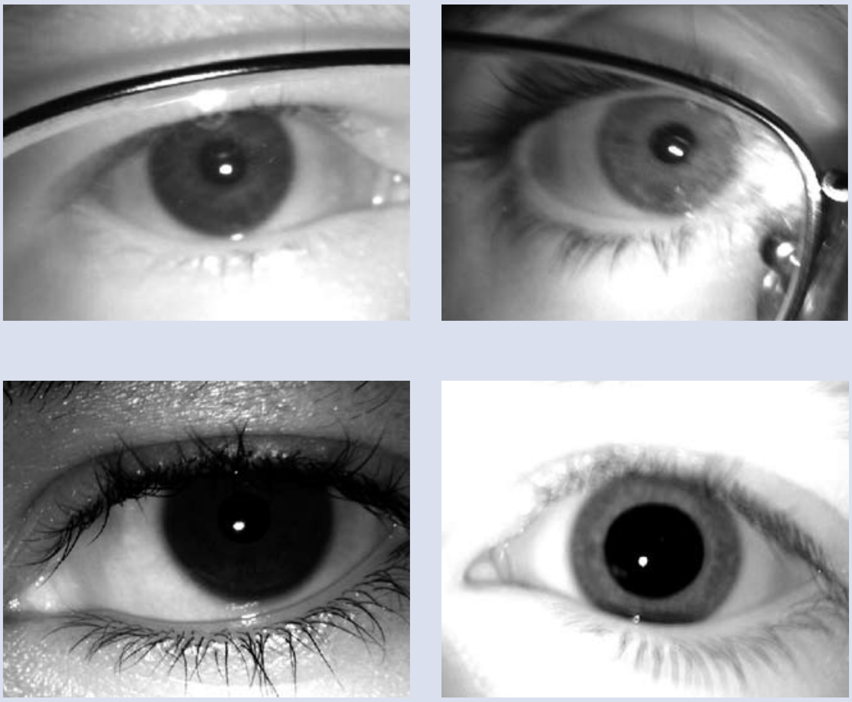
Figure 2 Wood Samples

The project partners are:

- National Physical Laboratory
- Unilever Research Port Sunlight Laboratory
- College of the Holy and Undivided Trinity of Queen Elizabeth, Dublin University
- Parc Científic de Barcelona
- Laboratoire de Physique Statistique, Centre National de la Recherche Scientifique
- Laboratoire d'Electronique et Technologie de l'Information
- Biometris, Wageningen University and Research Centre

Biometric image quality

A study of the effects of iris image quality factors on the performance of iris recognition biometric systems.



Examples of some of the poor quality eye images used in the comparison

Biometric identification technologies, such as automatic face, fingerprint and iris recognition, are being used for user authentication in an increasing variety of applications including computer login, building access control, and fast-track clearance through immigration. NPL is a recognised source of advice on the evaluation of biometrics technologies and has run several trials of these technologies under controlled conditions.

We have recently completed a project on a generic framework for data quality in biometric systems, including a study into the effects of the quality of iris images on the performance of iris recognition biometric systems. The data quality framework is described in terms of quality factors from the user (both their appearance and their behaviour), the environment, and the effects from the imaging system itself. These factors are further subdivided to cover all the aspects of data quality that can (or could) be measured.

The iris image quality study was based on two different iris recognition systems, which each calculated a number of quality measures, and a number of databases of iris images from two biometric trials run at NPL and elsewhere. The overall approach was to collect data on the quality scores of the images in the databases, and to correlate these scores against the recognition performance, in terms of the errors in localisation, and false matches and false non-matches in comparison. The results showed that different aspects of image quality were important at different stages of processing. For example, correct localisation of the iris in an image was most dependent on good iris/pupil boundary contrast quality; while accuracy of recognition was most dependent on the proportion of iris visible. Moreover for some measures of performance, some quality aspects had opposite effects on the two recognition algorithms; for example, lower values for the proportion of the iris visible reduce the chance of a false match with one algorithm, but increased this error rate for the other algorithm.

The overall conclusions are that measurement of quality is important to maintaining the performance of biometric systems, several relevant quality factors are not universal, and more work is needed to define and calculate a measure for such quality factors that relates to performance independent of algorithms. It is challenging, and perhaps infeasible, to develop universal quality metrics without algorithm bias. Our conclusions formed part of the UK contribution to the international standards for iris images data interchange (ISO 19794-6) and for data quality (ISO 29791-1).

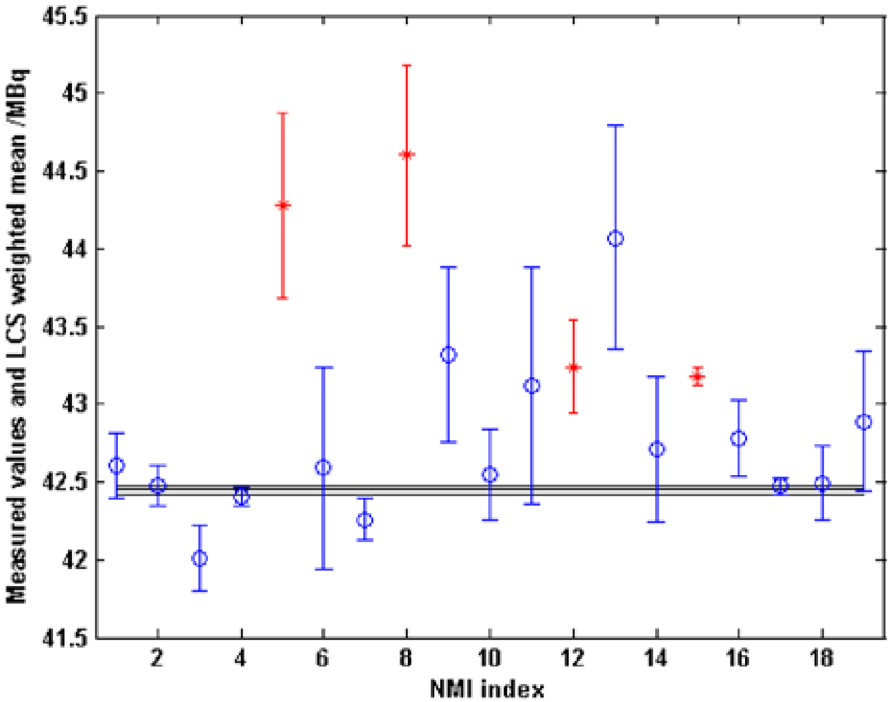
We are currently working on a project to determine how best to measure the accuracy of the imaging systems used in biometric systems based on 3D face images and will report on this in a later issue. As part of the Software Support for Metrology programme, we are starting a new project on the characterisation of image-based measurement data.

Digital image data is increasingly playing an important role in metrology, and this new project will exploit our group's experience of image systems for biometrics and the use of imaging for measurement throughout the rest of NPL. The project aims to produce better measurements from (2D) image data: measurements that are more reliable, robust and traceable by considering meta-data (data about how the image is made). The work benefits many areas of NMS and NPL science that use measurements from image data.

For more information please contact Tony Mansfield:
020 8943 7029
tony.mansfield@npl.co.uk

Key comparison data evaluation

The Software Support for Metrology team at NPL provides support on measurement data evaluation for interlaboratory and key comparisons.



Traceability of measurement at the international level is supported through a Mutual Recognition Arrangement (MRA) [1] by interlaboratory comparisons and key comparisons, which are used to establish the equivalence of measurement of national measurement institutes.

The evaluation of key comparison data is a significant task in the world of metrology because of the relevance to global trade and because such comparisons are intended to test the principal techniques in the field. Within such an evaluation is the determination of a key comparison reference value (a value agreed by all the participating laboratories) and the associated uncertainty, and the degrees of equivalence of and between national measurement standards.

The design of a key comparison can take different forms. The simplest case is where each participating laboratory provides an independent measurement of a property of a single stable artefact circulated around the participating laboratories. More complicated comparisons involve multiple artefacts, which sometimes can be unstable or inhomogeneous, also measured values having associated correlation provided by laboratories taking traceability from a common source. A challenge is to address the linking of a comparison organised by a regional metrology organisation, such as EURAMET, to a key comparison organised by a Consultative Committee of the CIPM.

As part of the Software Support for Metrology programme we give advice and support to the data evaluation for interlaboratory and key comparisons. For instance, we have worked recently on key comparisons organised by the CCM (Consultative Committee for Mass and Related Quantities), the CCPR (Photometry and Radiometry), the CCQM (Amount of Substance) and the CCRI (Ionizing Radiation). We also undertake work on generic aspects of key comparison data evaluation. For example, we have published papers [2, 3] on determining the largest consistent subset of comparison data. Consistency of (comparison) data is important in ensuring that the estimates of quantities determined from the data and the associated uncertainties are reliable, and therefore that those estimates and uncertainties convey useful information. Questions of consistency also arise in other applications, such as in the aggregation of data provided by different sensors (e.g., in a wireless sensor network), or by different (empirical) models for a quality of interest [3].

The graph shows the data from a key comparison of measured activity values of the radionuclide ^{75}Se. The data provided by the laboratories are shown in blue as circles with bars indicating ± one standard uncertainty, and those values excluded on statistical grounds from the largest consistent subset of the data in red as asterisks. It is important to provide plausible scientific reasons for such data exclusions.
The weighted mean of the data in the largest consistent subset, taken as the key comparison reference value, is shown as the central black horizontal line, with the upper and lower lines indicating ± one standard uncertainty associated with the weighted mean.

For more information please contact:
Peter Harris
020 8943 6961
peter.harris@npl.co.uk

[1] BIPM 1999 Mutual recognition of national measurement standards and of calibration and measurement certificates issued by national metrology institutes *Technical Report* Bureau International des Poids et Mesures, Sèvres, France

[2] Cox M G 2007 The evaluation of key comparison data: determining the largest consistent subset *Metrologia* **44** 187–200

[3] Collett M A, Cox M G, Esward T J, Harris P M and Sousa J A 2007 Aggregating measurement data influenced by common effects *Metrologia* **44** 308–318

Cutting out the noise

Quantitative approaches to digital signal processing in measurement systems: simulating a lock-in amplifier.

Introduction

The Software Support for Metrology digital signal processing project aims to ensure that NMS and industrial metrologists have a solid basis for their choices of algorithms, signal processing functions and hardware and are able to quantify the effects on their measurements of their choice of DSP methods and include appropriate terms in uncertainty budgets.

The project, which began in April 2007, is being delivered by means of two major case studies. The first study includes a simulation of a DSP system to study propagation of uncertainties through a complete measurement system.

Following consultations with NPL metrologists we decided to make the simulation of a lock-in amplifier the main focus of the first stage of our work.

Purpose of the simulation

Lock-in amplifiers are used in the laboratory to measure the amplitude and phase of a sinusoidal (AC) voltage. The instruments are particularly useful in cases when the signal-to-noise ratio is small, i.e., the signal of interest is obscured by noise many times larger than the signal. Lock-in amplifiers are widely used at NPL in areas such as optical, acoustical and electrical metrology, as well as materials characterisation.

The objective of the simulation is to help answer questions about how well an instrument can recover a signal from noise (the "accuracy" of the result provided by the instrument) and about the possible dispersion of values obtained (the "precision" of the result). A further objective is to provide quantitative information that can be used to inform the "uncertainty budget" for a measurement that depends on the use of a lock-in amplifier. The simulation allows users to investigate the performance and limitations of ideal and imperfect instruments applied to simulated and real test signals.

The basis of the simulation is a Monte Carlo calculation, in which imperfections in the instrument and the signals measured are expressed in terms of probability distributions that are used to characterise quantities that define different aspects of the measurement. The time for the simulation is much less than the time for a real measurement, and so the simulation provides a cost-effective way of designing and understanding a measurement before realising the measurement in practice.

Implementation

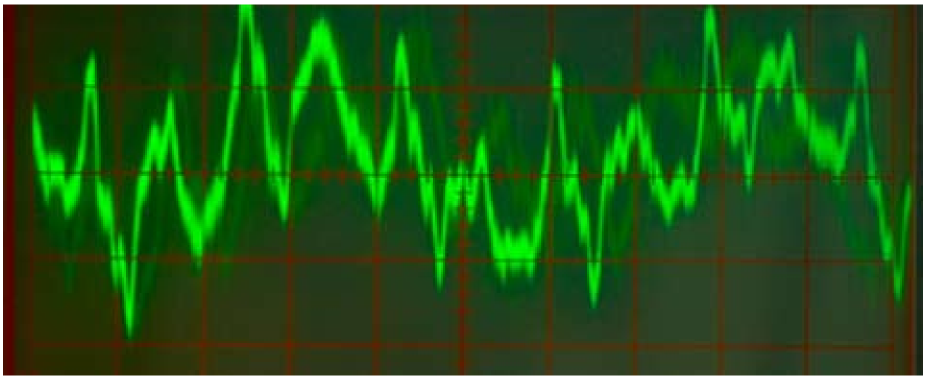
We based the simulation on the Stanford Research SR830 lock-in amplifier. However, we ensured that simulation models a generic lock-in amplifier, i.e., that it is not specific to the SR 830 but nevertheless captures many of the features of that instrument.

The mathematics underlying the operation of a lock-in amplifier is straightforward. A phase sensitive detector multiplies the signal under investigation with a reference signal. The output of the phase sensitive detector is simply the product of two sine waves, which consists of two signal components, one at a frequency representing the sum of the frequency of interest and the reference signal frequency, and the other representing the difference between the frequency of interest and the reference signal frequency. If the reference frequency and the frequency of the signal of interest are the same, one of the phase-sensitive detector outputs will be a DC component that can be isolated using a low pass filter. A second phase-sensitive detector performs the same multiplication on the signal of interest and a 90° phase-shifted copy of the reference signal and the output is once again low-pass filtered. The filtered outputs of the two phase-sensitive detectors are then used to calculate the amplitude and phase of the frequency component of interest in the signal under investigation.

The simulation is written in LabVIEW™. On the LabVIEW front panel the user enters parameters such as the amplitude, phase and frequency of a test signal, its noise amplitude and jitter noise amplitude and all associated standard uncertainties, followed by the characteristics and standard uncertainties of the internal oscillator that generates the reference signals for the phase sensitive detectors. The user also sets the sampling frequency, number of bits and acquisition time for the analogue-to-digital converter and finally inputs the time constant (and its uncertainty) for the filter, and the required number of Monte Carlo iterations. The simulation can also accept user-generated data files, so that it is possible to investigate the properties of experimentally obtained, as well as simulated signals. The outputs of the simulation are the amplitude and phase of the frequency component of interest in the input signal, together with their uncertainties. We also display histograms of these results. In addition, the input signal and the output of the low pass filter are displayed for each cycle of the Monte Carlo analysis.

An executable version of the software is available for download from the NPL website (http://www.npl.co.uk/server.php?show=ConWebDoc.2644). We are interested in hearing of examples of use of the software that has led metrologists to amend or improve experimental uncertainty budgets.

For more information please contact:
Trevor Esward
020 8943 6883
trevor.esward@npl.co.uk



If you would like further information on any aspect of
Mathematics and Scientific Computing Group (MCSG), please contact:
Tel: +44 20 8943 7050 | Fax: +44 20 8943 7091| E-mail: ssfm@npl.co.uk
Enabling Metrology Division
National Physical Laboratory | Teddington | Middlesex | United Kingdom | TW11 0LW
Helpline: 020 8943 6880 | Fax: 020 8943 6458 | E-mail: enquiry@npl.co.uk

Software guide for measurement

Software is an intrinsic part of measurement. It is used in instruments to control experiments, store and process measurement data, analyse and display results and to implement many mathematical techniques. Some innovations in measurement have been enabled through the use of software for simulations or complex analysis. For example the international temperature scale ITS90 requires the processing of high order polynomials and can only be implemented using software. Given this reliance, improvements in the quality of software and reduced cost of its development are vital to the effective delivery of metrology.

However, due to the increasing complexity and dependency on software, there are considerable concerns over its quality. A study by NIST in 2000 stated that *"Software bugs, or errors, are so prevalent and so detrimental that they cost the U.S. economy an estimated $59.5 billion annually"*. There is every reason to believe that Europe suffers in a similar way. NPL's recent audits of some instrument manufacturers, based on Software Support for Metrology Best Practice Guide 1, Validation of Software in Measurement Systems (BPG1), and several examinations of measurement software carried out by the PTB's Software Testing Laboratory, have indicated that software engineering techniques are not widely used.

A software guide that has been developed and accepted by leading NMI's would be more widely used and effective in the measurement community. For this reason NPL and PTB are currently collaborating on a guide to enable:

- developers of measurement software to know what they have to do to produce fit-for-purpose software;

and

- assessors of measurement software to confirm that the developed software is fit-for-purpose.

For more information please contact Graeme Parkin:
020 8943 7104
graeme.parkin@npl.co.uk

Forthcoming training courses at NPL

Understanding and evaluating measurement uncertainty

16 – 17 September 2008

A one and a half day training course on uncertainty evaluation applied to difficult measurement problems.

Developing advanced scientific engineering spreadsheet applications

12 – 13 November 2008

A two day training course to enable scientists and engineers to develop advanced spreadsheet applications which meet the requirements for accuracy critical applications.

For more information on these courses, including registration, please see http://www.npl.co.uk/server.php?show=nav.1211