Biometric Usability, Quality, and Technical Performance Measurement Uncertainty: A review of 3 bodies of work from NIST

James L. Wayman, Ph.D., FIET

San Jose State University

"Usability"

- M. Theofanos, S. Orandi, R. Micheals, B. Stanton, N.F. Zhang, "Effects of Scanner Height on Fingerprint Capture", NISTIR 7382, 2007, http://zing.ncsl.nist.gov/biousa/docs/NISTIR_7382_Height_Study.pdf
- "Usability Testing of Ten-print Fingerprint Capture", NISTIR 7403, 2007, http://zing.ncsl.nist.gov/biousa/docs/NISTIR-7403-Ten-Print-Study-03052007.pdf
- "Does Habituation Affect Fingerprint Quality?", 2006, http://zing.ncsl.nist.gov/biousa/docs/WP302_Theofanos.pdf

“Quality”

- E. Tabassi and P. Grother, “Quality Summarization. Recommendations on Biometric Quality Summarization. across the Application Domain”, NISTIR 7422, May 2007 www.itl.nist.gov/iad/894.03/quality/reports/enterprise.pdf
- P. Grother and E. Tabassi, “Performance of biometric quality Measures” IEEE Trans on PAMI(29), pp/ 531-543 April 2007, http://www.itl.nist.gov/iad/894.03/quality/reports/qoq_pami.pdf
- E. Tabassi, C.L. Wilson, and C.I. Watson, “Fingerprint Image Quality”, NISTIR 7151, August 2004 http://www.itl.nist.gov/iad/894.03/quality/reports/ir_7151.pdf

Estimating Measurement Uncertainty

San José State UNIVERSITY

- B.N.Taylor and C.E. Kuyatt, “Guidelines for Evaluating and Expressing the Uncertainty of NIST Measurement Results”, NIST Tech. Note 1297 (1994). http://physics.nist.gov/Pubs/guidelines/TN1297/tn1297s.pdf
- “Standard and Reference laboratories: Error requirements”, ISO 5725, ISO/TC69 Application of Statistical Methods
- M. Carroll Croarkin. “Realistic Evaluation of the Precision and Accuracy of Instrument Calibration Systems”, http://nvl.nist.gov/pub/nistpubs/sp958-lide/html/129-131.html
- S.M. Stigler, “Statistics and the Question of Standards”, Journal of Research of the National Institute of Standards and TechnologyVolume 101, Number 6, November–December 1996, http://nvl.nist.gov/pub/nistpubs/jres/101/6/j6stig.pdf
- W. J. Youden, Enduring Values, Technometrics 14, 1–11 (1972)

Quality

- "..the term quality is not used … to refer to the fidelity of the sample, but instead to the utility of the sample to an automated system … This viewpoint may be distinct from the human conception of quality." – Grother, Tabassi,PAMI
- Quality is a proxy for error rate
- A single sample can have a utility for comparison, which is a sample-reference operation
- Quality metric for a dataset depends upon threshold –NISTIR 7422

Quality

Variation

- over time (to expose seasonal variation, or trends),
- for each sensor (to identify defective devices),
- at each site (to identify problem locations)
- of officials or attendants (to assess adherence to operating procedures), and
- per user basis (to identify users that consistently yield low quality samples).

-- NISTIR 7422

Useability

Fingerprint quality is impacted by:

- Scanner height
- Training conditions
- Age
- Habituation
- Supervision

--NISTIR 7382 & “Habituation”

Quality is a Proxy for Error Rates

Therefore, error rates depend upon:

- Season
- Location
- Officials and attendants
- Users
- Height
- Training
- Age
- Habituation
- Supervision

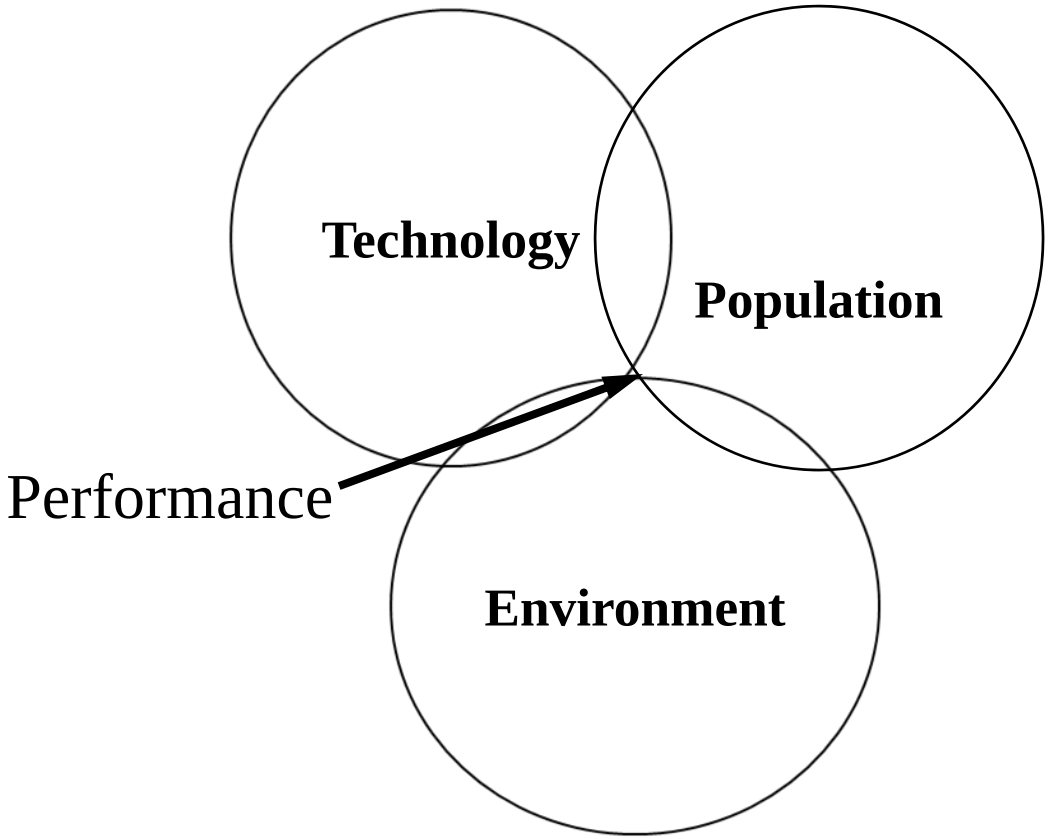
-- NIST

My Conjectures:

DET and throughput depend upon:

- Operational threshold
- Acoustic noise level
- Temperature and humidity
- Data subject stress level
- Phase of the moon

Performance



"In short, we are not… measuring anything"

- "a measurement operation must have attained what is known in industrial quality control language as a state of statistical control . . . before it can be regarded in any logical sense as measuring anything at all." – Churchill Eisenhart, Chief of the Statistical, Engineering Laboratory (SEL), Applied Mathematics Division, National Bureau of Standards (1947-1963) -- as quoted in NIST 129-131
- "Incapability of control implies that the results of measurement are not to be trusted as an indication of the physical property at hand – in short, we are not in any verifiable sense measuring anything" – R.B. Murphy, *On the Meaning of Precision and Accuracy*, Materials Research and Standards, ASTM (1961)

Two Types of Uncertainty

- Random
  - Number of test subjects
  - Number of samples per subject
  - Correlations between samples and comparisons
- Systematic
  - As previously listed

Two Types of Uncertainty Estimates for Each Uncertainty Type

San José State UNIVERSITY

- Type A:
  - "may be based on any valid statistical method for treating data. Examples are calculating the standard deviation of the mean of a series of independent observations"
  - This is a "frequentist" approach
- Type B:
  - "evaluation of standard uncertainty is usually based on scientific judgment using all the relevant information available"
  - This is a "personalist" (Baysian) approach

Applying this Philosophy to Biometrics

San José State UNIVERSITY

- In performance testing, we have not attained statistical control of systematic error attributable to “usability” (human) factors.
- Hence, we are not measuring anything at all
- The uncertainty in our measurements is both random and systematic, but the latter dominates
- Uncertainty is estimated by both frequentist and personalist methods

The Questions We Ask

Bogus:

- What is the accuracy of fingerprinting?
- How big should a test be for statistical significance?
- How do we certify technologies?

New:

- What are the expected error rates of airline passengers arriving on a 10-hour flight when the index fingerprint scanner height is 42” and instruction is by poster?
- How comprehensive should the test program be such that the experimenters will feel comfortable in estimating uncertainty.
- How do we certify populations and training methods by application?

Biometric Consortium 1995 Challenge:

External consistency:

- Why are laboratory results not a good predictor of “real-world(s)” performance?
- 2007 Answer: Systematic uncertainty attributable to uncontrolled variability in human factors

Summarizing Comments

- All biometrics are behavioral
- Uncertainty in error rate estimates depends upon the comprehensiveness of the test, not the number of transactions
- Devices cannot be certified, but people can be certified with devices
- Improving biometrics will require emphasis on human factors not the purely “technical”
- Biometrics does not start with usability, it is usability.