International·Biometric·Group

Research Consulting Integration

# Lessons Learned from Comparative Biometric Testing

# International Biometric Group

- Independent biometric research, consulting and technology solutions firm
- Founded in 1996
- Offices in New York and Washington, D.C.
  - Including BiometricStore, a hands-on showroom and test facility with over 85 hardware and software solutions
- Technology-neutral and vendor-independent
  - Extensive experience across all biometric technologies
  - IBG does not resell or distribute biometric hardware or software

# What We Do

- Biometric Research
  - Provide comprehensive market research reports and analysis
  - Conduct industry's leading scenario-based system testing
- Biometric Consulting
  - Build long-term product and solution development strategies for systems integration, high-tech, and industrial firms
  - Perform feasibility studies for government agencies on large-scale biometric usage
- Biometric Technology Solutions
  - Evaluate, design and deploy custom biometric solutions for IT security, e-commerce, access control, public sector ID systems
  - Provide standards-compliant, interoperable solutions for employee, customer, and citizen authentication

# Comparative Biometric Testing

- Begun in 1999 to address the need for objective, real-world data on biometric system accuracy
- Test rounds conducted annually to gauge comparative accuracy of leading biometric technologies
- Test Sponsors have included…
  - Fidelity Investments
  - Microsoft
  - American Airlines
  - Financial Services Technology Consortium
  - Star Systems
  - Lockheed Martin
  - EDS
- Results available to test purchasers

# Purpose of Comparative Biometric Testing

- To help public/private sector institutions determine the degree to which commercially available biometric systems are suitable for deployment in logical and physical access applications

# Comparative Biometric Testing Methodology

- Scenario testing emulates real-world operations
  - Operators guide subjects through systems
  - Environment carefully controlled
  - Testing evaluates sensors and algorithms together
  - 240 Subjects; varied demographics, ages, ethnicities
- Generates false match rates (FMR), false non-match rates (FNMR), and failure to enroll rates (FTE)
  - Where possible, tests occur at low, medium, and high security
- Testing includes Primary Visit and a Return Visit (+6 weeks)
- Normalizes enrollment and matching for different systems
- Ensures that error rates can be compared for systems with dissimilar enrollment and matching processes

# Test Process

- Primary Visit
  - User attempts to verify as an imposter against 2 pre-selected enrollments
    - Attempts at low, medium, high
    - Allowed three attempts/placements
  - User attempts to enroll
    - Allowed 2 full enrollment sequences; enrollment sequence=minimum # of placements necessary to enroll plus normalized #
    - Those unable to enroll after 2 full sequences allowed additional “try harder” sequences
  - User verifies against enrollment
    - Attempts at low, medium, high
    - Allowed three attempts/placements
- Return Visit (+6 weeks)
  - Identical to Primary Visit false non-match testing

# Best Practices Compliance Statement (1/4)

- CBT compliant with recommended UK BWG Best Practices (BP) version 2.01; diverges mildly in certain respects
- Paragraph 18 of BP states:
  - …we define (a) the false match rate and (b) the false non-match rate, to be the error rates of the matching algorithm from a single attempt-template comparison…
    - CBT defines these error rates subsequent to a normalized number of attempts or attempt sequences (e.g., 3 placements)
    - Most logical and physical access applications allow a specific number of placements or a certain duration of interaction
    - IBG finds that defining error rates according to this normalized standard provides a truer reflection of real-world performance
- BP focuses on Receiver Operator Characteristic curves
  - CBT also views FTE as a critical indicator of overall system performance, and places this data on a par with FMR, FNMR

# Best Practices Compliance Statement (2/4)

- Paragraph 44 of BP states:
  - For scenario evaluations, test data must be separated in time from enrollment by an interval commensurate with “template ageing” of the target system...
    - CBT defines a consistent period of time between enrollment and subsequent true match attempts regardless of technology
    - Testing occurs at the same time for Fingerprint, Face, iris-scan, etc., although the template aging for these characteristics may vary
    - The template aging of a biometric characteristic generally does not strongly inform deployment decisions
- Paragraph 47 of BP states:
  - …the collection [of test data] must ensure that presentation and channel effects are either: 1) uniform across all volunteers; or 2) randomly varying...If the effects are held uniform across volunteers, then the same presentation/channel controls in place during enrollment must be in place for collection of test data…
    - Submission advice and direction given to subjects is greater during enrollment than during verification
    - May increase likelihood that presentation effects inform performance

# Best Practices Compliance Statement (3/4)

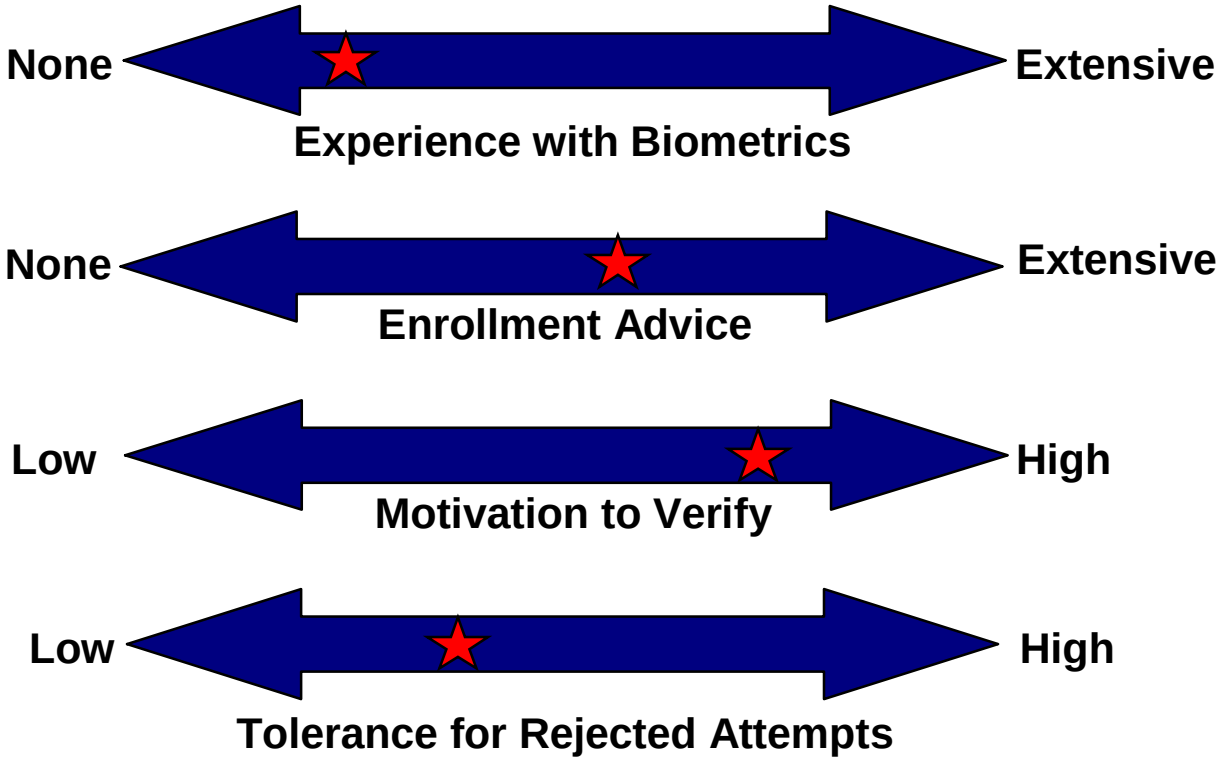
- Paragraph 48 of BP states:
  - The ‘failure to acquire’ rate measures the percentage of the population unable to give a usable sample to the system as determined by the experimenter or the quality control module.
    - CBT does not generally distinguish between FTA and algorithm-based matching errors when calculating error rates
    - From a deployer’s perspective, FTA and matching errors are difficult to separate in a meaningful fashion; as acquisition and template generation functions can be difficult to separate
- Paragraph 50 of BP states:
  - Volunteers should not be told whether the current comparison is genuine or impostor to avoid even unconscious changes in presentation.
    - CBT does inform Test Subjects of the type of match attempts (true and false) being executed
    - Placement advice and protocols are identical in each case
    - Individuals are normally aware of whether they are executing a true match or false match attempt, and may alter behavior accordingly

# Best Practices Compliance Statement (4/4)

- Paragraph 147 of BP states:
  - To improve the independence of different samples from a single volunteer, it is possible that an evaluation allows enrolment of more than one finger/hand/eye/etc. as different (sub-) identities. However, within-individual comparisons are not equivalent to between-individual comparisons, and must not be included in the set of impostor transactions.
    - Above and beyond standard FMR testing, CBT conducts an additional test for fingerprint technology consisting of middle versus index finger false match testing
    - Potentially an interesting reflection of algorithm's capabilities to discern between like samples
    - Data presented separately from main report data

# User Population

- Defining the user population



# Comparative Biometric Testing History

| Round I (7/1999) | Round II (2/2001) | Round III (8/2001) |
|---|---|---|
| ST Microelectronics (Fingerprint) | Advanced Precision (Fingerprint) | SecuGen (Fingerprint) |
| DigitalPersona (Fingerprint) | American Biometric Company (Fingerprint) | Sagem Morpho (Fingerprint) |
| Mytec (Fingerprint) | AuthenTec (Fingerprint) | BES (Fingerprint) |
| Identix (Fingerprint) | BioPassword (Keystroke-Scan) | Precise Biometrics (Fingerprint) |
| Veridicom (Fingerprint) | CyberSIGN (Signature-Scan) | Identix (Fingerprint) |
| Sony (Fingerprint) | Digital Persona (Fingerprint) | Iridian (Iris recognition) |
| Identicator (Fingerprint) | Ethentica (Fingerprint) | AcSys Biometrics (Facial Recognition) |
| American Biometric Company (Fingerprint) | Nuance (Voice-Scan) | Viisage (Facial Recognition) |
| Visionics (Facial Recognition) | Precise Biometrics (Fingerprint) | Sony (Fingerprint) |
| Miros (Facial Recognition) | Visionics (Facial Recognition) | |

# Round IV Systems (2002)

| Vendor | Technology | Acquisition Device | Algorithm/SW |
|---|---|---|---|
| Bioscrypt | Fingerprint | AuthenTec AF-S2 / Lifeview<br>*silicon sensor* | Bioscrypt<br>"Core Connect" |
| DigitalPersona | Fingerprint | U.are.U 4000<br>*optical sensor* | DigitalPersona |
| Fujitsu | Fingerprint | MBF200<br>*silicon sensor* | Fujitsu<br>MyMinutia |
| Identix | Fingerprint | DFR-200<br>*optical sensor* | Identix |
| Indivos | Fingerprint | U.are.U 2000<br>*optical sensor* | Bioscrypt |
| Siemens | Fingerprint | ID Mouse Professional<br>*silicon sensor* | Siemens<br>Pro Demo |
| STMicro | Fingerprint | TouchChip<br>*silicon sensor* | TouchChip<br>PerfectMatch |
| Ultra-Scan | Fingerprint | Series 700 ID Station<br>*ultrasonic sensor* | Ultra-Scan<br>ID Express Demo |
| Recognition Systems, Inc. | Hand Geometry | HandKey II | Recognition Systems |
| Iridian | Iris Recognition | LG2200 | Iridian PrivateID –<br>Recognition Demo |
| Nuance | Voice-scan | Panasonic KX-TS-5B<br>telephone | Nuance<br>Verifier 3.0 |

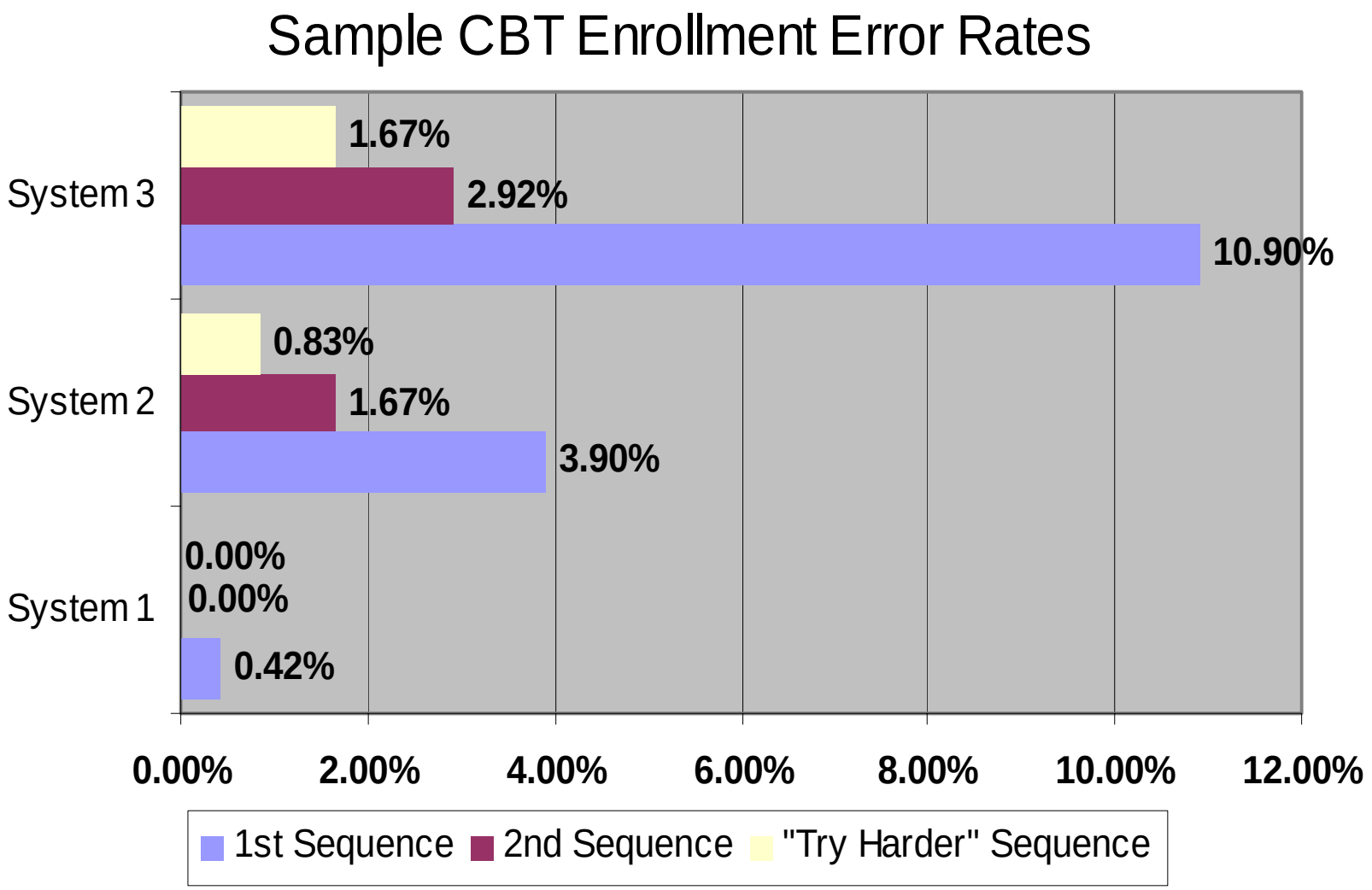
# Trends and Findings: 1999-2002

- Enrollment rates remain an issue
- Performance over time remains an issue
- False matching rates falling
- Variance within technologies
- Vendor threshold management an issue
- Systems much more reliable
- Vendor claims and test results not always aligned

# Enrollment Still an Issue

- FTE rates range from 0.00% to 11.97%
  - Even after addition instruction and attempts, 3 systems had FTE over 2%
- One of the most important findings of scenario testing
  - Assumption: that enrollment attempts must be limited to some degree
- Enrollment rates must be included in test report, as many systems trade higher enrollment rates for lower FNMR

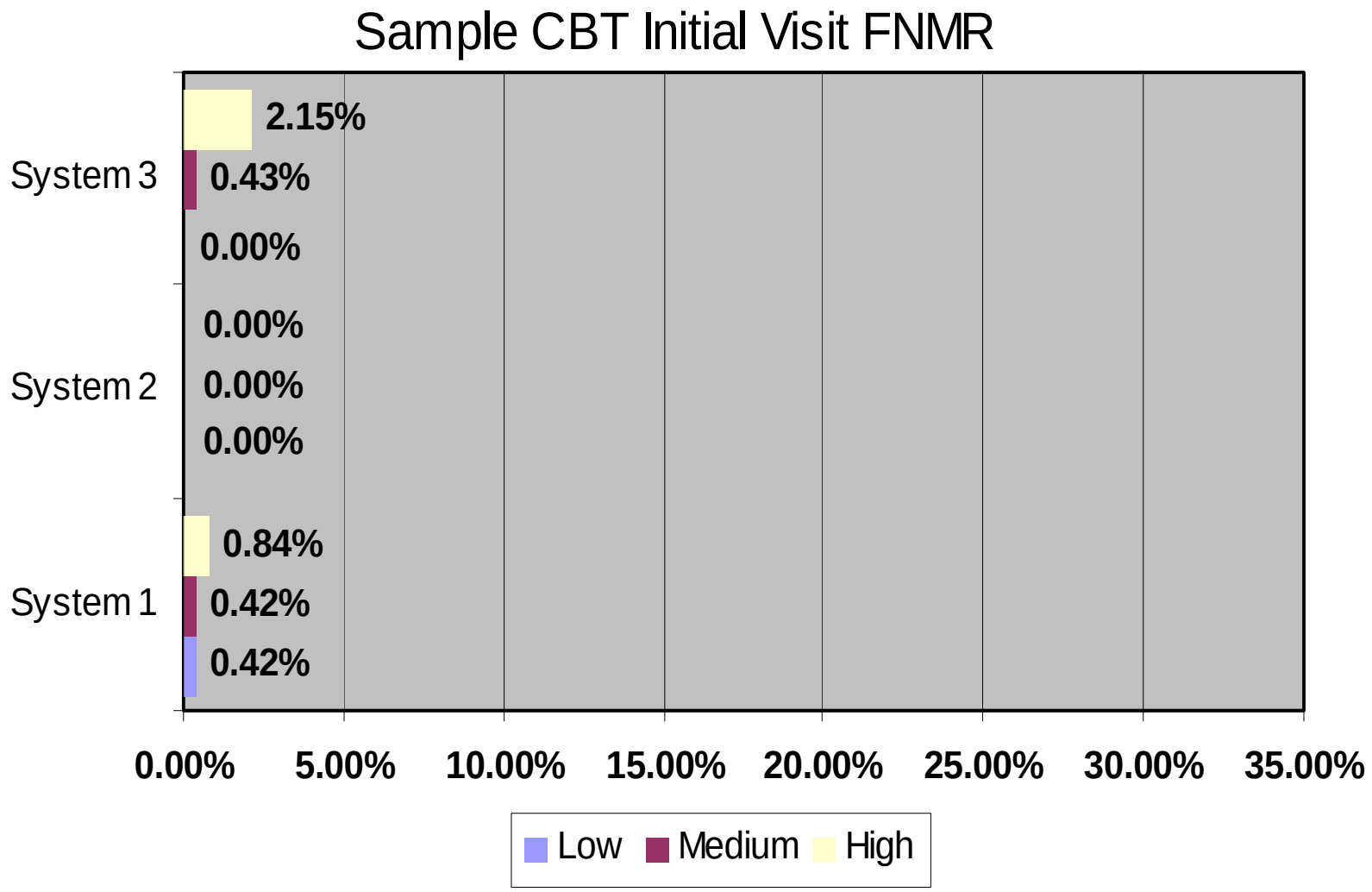
# Sample Test Data: FTE Rates



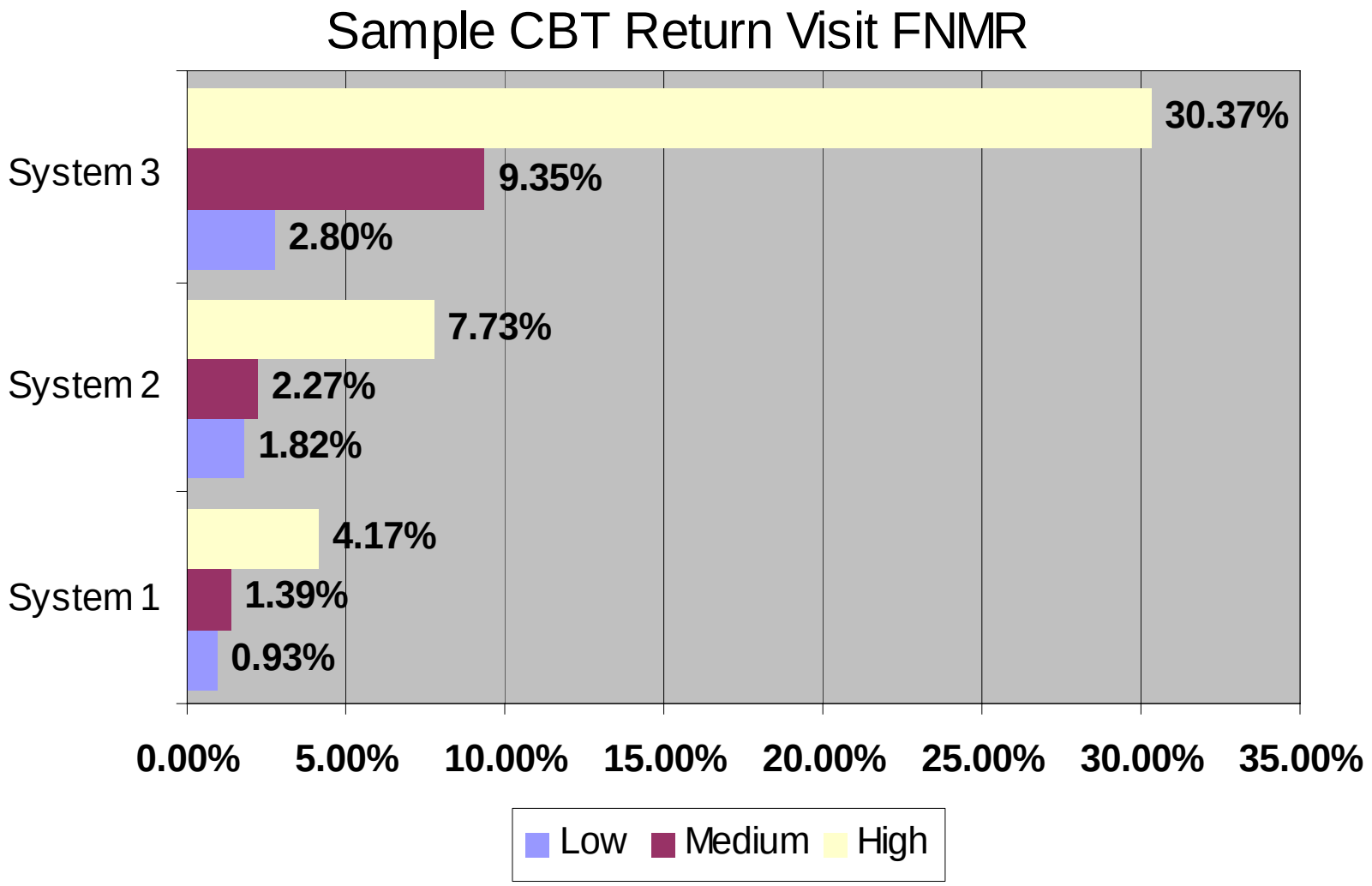
# Performance Over Time

- False Non-Match Rates over time are inevitably higher than first visit FNMR
- FNMR increases ranged from 0.51% to 8.92% at medium level
  - Worst deterioration: 14-fold increase in FNMR
- Would these rates decrease with constant usage?

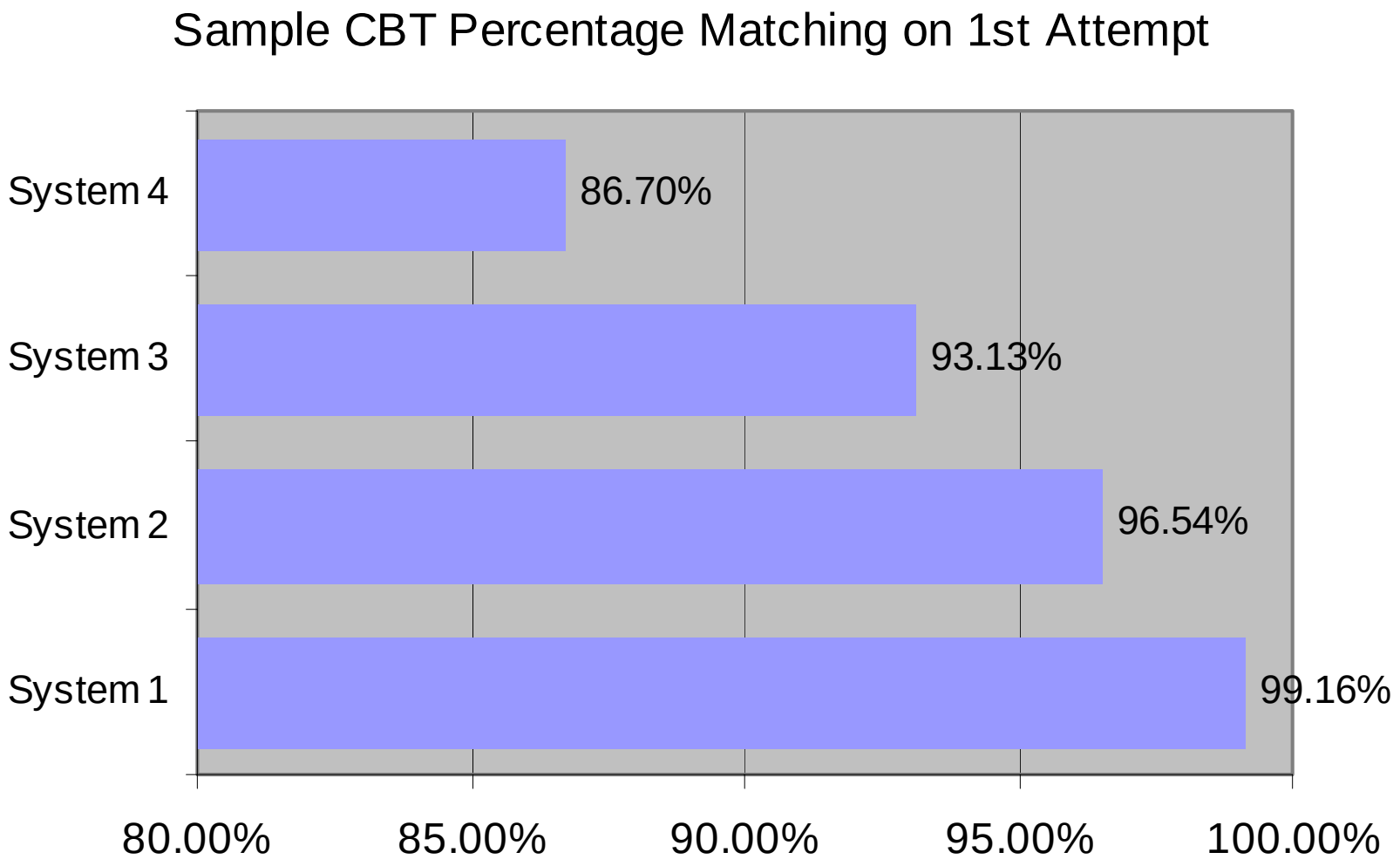
# Sample Test Data: Initial Visit FNMR



# Sample Test Data: Return Visit FNMR



# Sample Test Data: True Match Attempts





# False Match Rates Falling

- First rounds of tests (1999) showed several high-single digit FMR
- More recent rounds, many systems show next to no FMR
  - We interpret this data as being consistent with technologies' maturation
- Is implementing all vs. all false match testing in order to increase data a valid approach?

# Variance within Technologies

- Within fingerprint and voice-scan, performance varies widely
  - Fingerprint: from near perfect (<3 errors in entire test) to 33% error rates
  - Voice-scan: certain systems with very low FMR and FNMR; others with very high FNMR
- Facial Recognition performance does not vary as widely
  - Low FMR
  - High FNMR, particularly over time
  - Some FTE

# Vendor Threshold Management

- Many vendors provide thresholds not at all representative of high, medium, and low security/convenience
  - Many systems heavy skewed toward high security, therefore show high FNMR with no FMR
  - Adjustment of thresholds would have improved overall performance
- A slight trend away from providing adjustable thresholds on the part of vendors
  - Uncertain whether this may indicate a preference for one global setting

# System Reliability

- Systems much more ‘professional’ than when testing began
- Then
  - Devices crumbled under fingerprint placement
  - Multiple weeks dedicated to installation, configuration
  - Machines highly unstable (crash, lockup)
  - Devices caught on fire
- Now
  - Polished appearance and device operations
  - Simpler installation
  - Intuitive usage

# Vendor Claims vs. Test Results

- Performance as tested almost always worse than performance claimed
- Partially a function of CBT methodology's focus on replicating real-world operations
  - Non-acclimated users
  - Placement and presentation factors incorporated (as opposed to testing static data)
  - Incorporates substantial imposter data, not present in operational testing
- By normalizing test methods, CBT attempts to provide a level playing field

# Going Forward

- Many difficult-to-quantify parameters impact testing
- Testing should accommodate deployer interests and requirements, and be reflective of realistic operations
- There is still a great deal of confusion on what test data means when applied to real-world systems
- Best Practices a very valuable tool in understanding test conception and execution