# Biometric Testing
# Celebrating 35+ Years!

**Dr. James L. Wayman**

**San Jose State University**

**9 December, 2003**

# This Talk Asks:

Where are we?

How did we get here?

What needs to be done?

# MMUA in the 1970s

San José State UNIVERSITY

- Beardsley, Charles T. “Is your computer insecure?”, IEEE Spectrum, Jan. 1972, pg. 67-78
- S. Reed and D. Branstad, “Controlled Accessibility Workshop Report”, NBS Tech. Note 827, May 1974
- D.E. Raphael and J.R. Young, Automated Personal Identification, SRI, International, 1974
- “Guidelines for Evaluation of Techniques for Automated Personal Identification”, Federal Information Processing Standards Publication 48, April 1977
- A. Fejfar, “Combining Techniques to Improve Security in Automated Entry Control”, 1978 Carnahan Conf. On Crime Countermeasures,, Mitre Corp. MTP-191, May 1978
- G. H. Warfel, Identification Technologies (Charles C. Thomas, Springfield, IL, 1979)

# Early Testing Documents


- L. G. Kersta, “Voiceprint Identification”, Nature,vol.196 ,Dec 29, 1962, pp.1253-1257
- J. Wegstein, “Automated Fingerprint Identification”, NBS Tech. Note 538, Aug. 1970
- O.Tosi, “Experimental Studies on the reliability of the voiceprint identification technique”, Proc.of 3rd National Symposium of Law Enforcement and Technology, 1970

# Early Testing Documents


- A.J. Goldsten, L.D. Harmon, and A.B. Lesk, "Identification of Human Faces", Proc. IEEE, 59(5), May, 1971
- L.D. Harmon, "The Recognition of Faces", Scientific American 229(5), 1973
- W. Haberman and A. Fejfar, "Automatic ID of Personnel through Speaker and Signature Verification – System Description and Testing", 1976 Carnahan Conference on Crime Countermeasures, U of KY, May, 1976

# Early Testing Documents


- A. Fejfar and J. Myers, “The Testing of 3 Automatic ID Verification Techniques for Entry Control”, 2nd International Conf. On Crime Countermeasures, Oxford, 25-29 July, 1977
- A. Fejfar, “Combining Techniques to Improve Security in Automated Entry Control”, 1978 Carnahan Conf. On Crime Countermeasures,, Mitre Corp. MTP-191, May 1978
- “Guidelines for Evaluation of Techniques for Automated Personal Identification”, NBS, FIPS Pub. 48, April 1977

# Kersta (1962)

San José State UNIVERSITY

- Human processing
- Gender segregation of database
- Implicit closed set testing
- "Some voices more difficult to identify than others"
- More data led to lower error rates
  - How many words?
  - Some words harder for speaker ID

# Kersta (1962)

"Work continues, there being questions to answer and problems to solve. Among these are:

1) Is the voice unique enough to afford infallible identification? This requires sampling of the voices of a much larger section of the population.
2) Does the voice of an adult change significantly over time? If, so how?
3) How will attempts to disguise the voice succeed?
4) How successful in disguise are voice mimics?"

# Kersta (1962)

"Work is in progress to devise a classification and indexing system for filing voiceprints. The system must afford compactness, rapid access, and reduce the population of filed prints needed for subjective identification to a relatively small number"

# Tosi (1970)

1967 review of the test methods of Kersta

"Each task of identification is characterized by the following parameters…"

- a) open or closed
- b) contemporary or non-contemporary spectrograms
- c) 9 or 6 cue words
- d) 1,2,3 utterances of each word
- e) Direct recording, telephone, telephone with 40dB white noise
- f) isolated words, fixed context, random context
- g) 10, 20, or 40 known speakers

# Wegstein (1970)

San José State UNIVERSITY

- Score matrix
  - Explicitly tests open and closed
- Open task metrics
  - M,  the percentage of misses
  - F, the percentage of false matches
- Closed task metrics
  - CS = Max over all rows {Max off-diagonal score/min diagonal score}

# Goldstein/Harmon/Lesk (1971) and Harmon (1972)


- Closed set FR testing using human examiners
- Physiological “features”
- Reference to W.W. Bledsoe’s 1966 work on automated FR
- “correct face in 10$^{th}$ place or better in 99 out of 100 trials” (out of 255)
- Recognition improved by blurring

# Haberman and Fejfar(1976)

- Handwriting and speech
- 209 volunteers
- 2600 trials for each technology
- 100,000 cross comparisons segregated by gender
- Variation of Type I errors by person
- Male error rates lower

# Fejfar and Myers (1977)

Goals are

- Validate the contractor's reported error rates
- Identify any weaknesses
- Identify any learning curve effects(over20 attempts)
- Suggest improvements
- Establish the degree of independence between the systems

# Fejfar and Myers (1977)

| | Speaker | | Handw | | Fingerprint | |
|---|---|---|---|---|---|---|
| | FRR | FAR | FRR | FAR | FRR | FAR |
| Lab | 0.2 | 4.4 | 3.2 | 1.7 | 4.6 | 2.2 |
| Field | 1.1 | 3.3 | 1.9 | 5.6 | 6.5 | 2.3 |
| Time | 6.2 sec | | 13.5 | | 8.9 | |

# Fejfar (1978)

| | S&H | H&F | F&S | S&H&F |
|---|---|---|---|---|
| FRR | 3.0 | 8.4 | 7.6 | 9.5 |
| FAR | 0.18 | 0.13 | 0.08 | 0.0043 |
| Ver Time | 19.7 | 22.4 | 15.1 | 28.6 |
| Total Time | 32.0 | 34.7 | 27.4 | 40.9 |

# Fejfar (1978)

| | S\|\|H | H\|\|F | F\|\|S | S\|\|H\|\|F |
|---|---|---|---|---|
| FRR | 0.02 | 0.12 | 0.07 | 0.0014 |
| FAR | 8.9 | 7.9 | 5.6 | 11.2 |
| Ver Time | 10.4 | 12.8 | 8.2 | 10.7 |
| Total Time | 22.7 | 25.1 | 20.5 | 23.0 |

# FIPS PUB 48 (1977)

San José State UNIVERSITY

"Guidelines on evaluation of Techniques for Automated Personal Identification"

- Type I = false alarm rate (FAR)
- Type II = impostor pass rate (IPR)
- "Effects of allowing multiple attempts"
- "Combining personal identification methods" using AND/OR rules

# Evaluation Criteria

San José State UNIVERSITY

1. Resistance to Deceit
2. Ease of counterfeiting
3. Susceptibility to Circumvention
4. Time Required
5. User Convenience
6. Device Costs

# Evaluation Criteria

San José State UNIVERSITY

7. Interfaces
8. Ease of Updating database
9. Computer system processing req's
10. Reliability and Maintainability
11. Cost of Protecting Device
12. Cost of Distribution and Logistical Support

# Sandia Test Program (1983 – present)

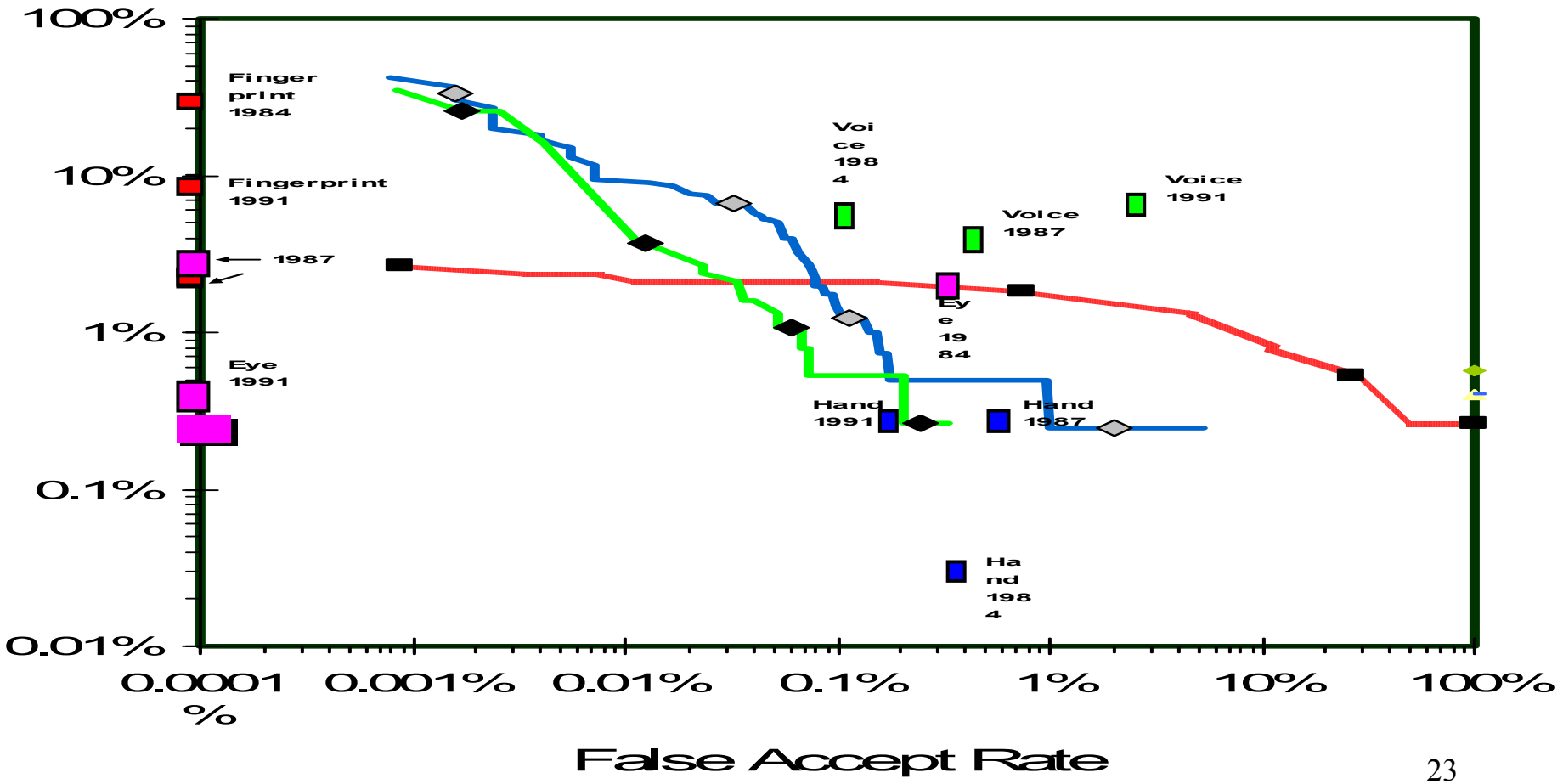

"Tighter security requirements to combat the threat of terrorism, and today's capabilities of transferring large amounts of information and funds at electronic speeds further increases the need for personal identity verification."
-- Russell L Maxwell, "The Status Of Personnel Identity Verifiers", 1985

# Sandia Test Program


Transaction time in seconds

| | SNL'84 | SNL'87 | SNL'91 | NPL2000 |
|---|---|---|---|---|
| Fingerprint | xx | 9.8 | 6.6 | 8 |
| Hand | xx | 4.4 | 5 | 10 |
| Eye | 7 | 7.5 | 7 | 12 |
| Voice | 20 | 8.8 | 19.5 | 12 |

# Sandia vs. NPL



# Fast Forward

- Doddington
- Poock
- Campbell
- Martin
- Philips
- Maltoni/Maio
- Mansfield

# Today!

- Programs
  - NIST SRE
  - FRVT
  - FVC
  - FpVTE
  - IBG
- Additional Documentation
  - “Best Practices of Testing and Report Biometric Device Performance”, www.cesg.gov.uk
  - ISO/IEC JTC1 SC37 WG5 “Testing Principles”

# A Taxonomy of Technical Testing


- Form
  - Open/Closed
- Purpose
  - Real world(s) prediction
  - Across device comparison
  - Across time comparison
- Setting
  - Technology/Scenario/Operational
- Level
  - Black/Gray/White

# Any Combination Possible

- FERET
  - Closed, across device, technology, white
- FRVT
  - Open/closed, across device, technology, white
- NIST SRE
  - Open, across device, technology, white
- IBG
  - Open, across device, scenario, gray
- NPL
  - Open, real world, scenario, white
- SmartGate/INSPASS
  - Open, real world, operational/scenario, white

# Technology, Scenario, Operational


- Technology: testing the algorithms
- Scenario: testing the human-machine interface
- Operational: testing mob behavior

# The Campbell Questions to the BC

- Why do scenario tests fail to predict operational performance?
  - Answer: Scenario does not model behavior of target users in operational environment
- Sandia ’91 vs. Kirkland AFB
  - Techies supervised, habituated, and indoors vs. general users unsupervised, non-habituated, and outdoors

# Comments on "Best Practices"


- "Real worlds(s)" performance estimation
- Version 1.0 stolen directly from Doddington/Martin, NIST SRE 1999 protocol
  - No authorship claimed
  - "Folk wisdom"
  - "Testing Standard" not realizable
- Version 2.0 was a reorganization of version 1.0 with new statistics contributed by Mansfield.
- Community input requested for version 3.
  - What are the fundamental principles of testing?

# Some Fundamental (but often violated) Principles


- Separate "system training" and testing databases
  - Genuines and impostors cannot be from set used to create basis vectors
- Artificial images are phony
  - You are not God and don't know how people are made
  - Unfair positive bias to systems making similar assumptions
- Use of volunteers biases data

# Some Fundamental (but often violated) Principles


- "Background Databases" (not UBM) should not augment data for impostor testing
  - Not matching data collected under different conditions is a "no brainer".
- "Unknown impostors"– this is tough!!!
  - Impostors must not be in the UBM
  - Must not be in the system training data
  - Must not be in genuine training group if models are dependent
  - Multiple passes through the data allow for learning impostors

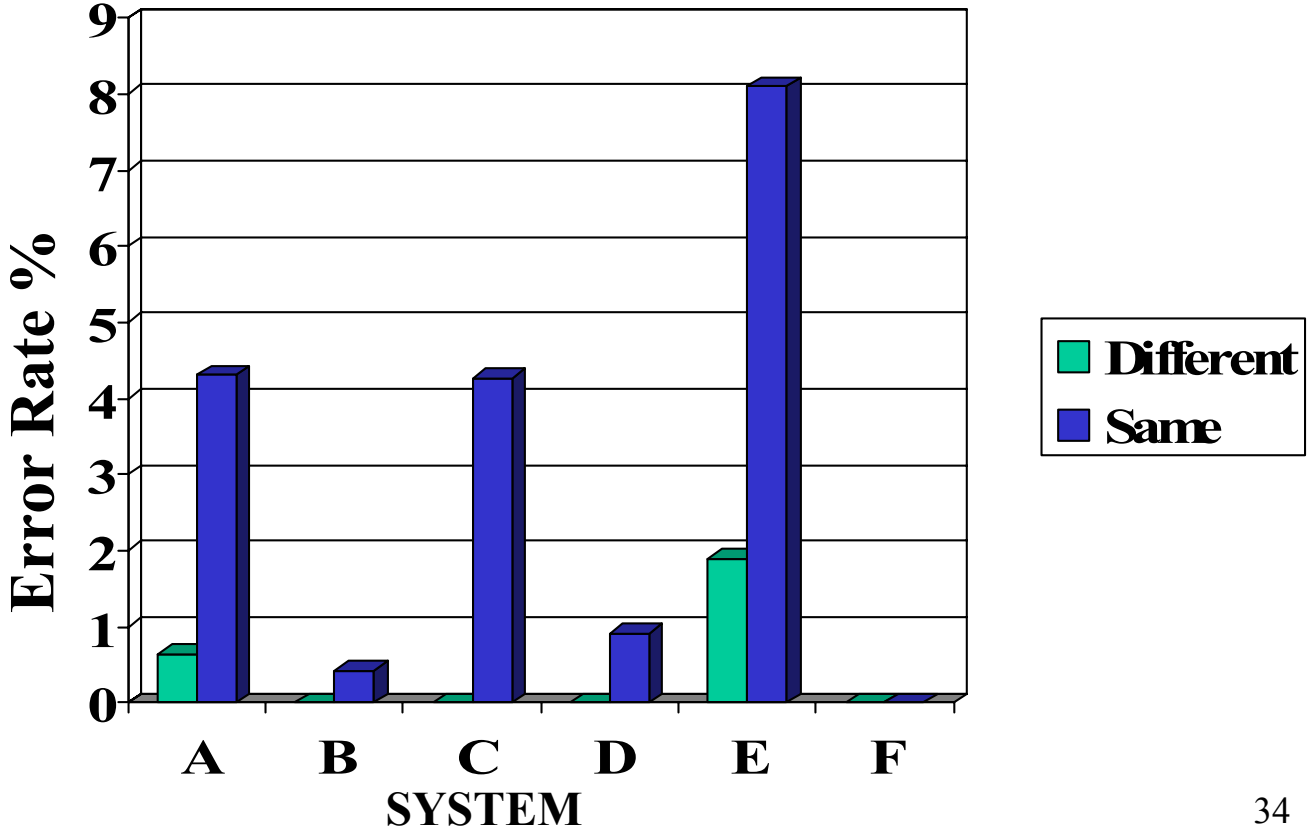
# Some Fundamental (but often violated) Principles

San José State UNIVERSITY

- Test data must be "unseen" (sequestered)
  - What to do about overtraining on non-sequestered elements?
  - How do we assess performance improvement over time?
- Test data must be edited
  - All editing must be subjective and before output is known

# Other Reasonable Approaches

IBG Comparative Testing



# Multimodal Testing

San José State UNIVERSITY

- Correlation, Correlation, Correlation
- 3 Levels
  - Decision
  - Score
  - Feature
- Can correlation be understood except at the feature level?
- Is GMM the model for multimodal in the feature space?
- What would Fukunaga say?

# ISO/IEC JTC1 SC37

- Working Group 1: Vocabulary Harmonization – Canada
- Working Group 2: Technical Interfaces – Korea
- Working Group 3: Data Interchange Formats -- Germany
- Working Group 4:Application Profiles – U.S.
- Working Group 5: Test and Reporting – U.K.
- Working Group 6:  Societal and Cross-jurisdictional Aspects – Italy

# ISO/IEC JTC1 SC37 WG5

- Chair – Bob Carter (UK)
- Part 1 “Testing Principles” editor – Tony Mansfield
- US Head of Delegation – Cathy Tilton
- US Head Rep to WG 5 – Ron Sutton
- Proposed Part 1 Committee Draft to be submitted by UK at Feb. 2004 meeting

# UK "Testing Principles" Submission

San José State UNIVERSITY

- Re-write of "Best Practices" version 2.01
- "Best Practices" evolution will continue
  - Consideration of closed set testing
    - Large-scale systems combine closed and open set principles
  - Operational testing issues
  - Closing zoo and returning animals to their natural habitat

# Conclusions

San José State UNIVERSITY

- Lots of work remaining
- Beyond i.i.d.
  - Correlation, correlation, correlation
- Beyond priors
  - Huge user/government confusion regarding prior, posterior and likelihood
- Who will be the next George Doddington?
- Who will be the Fukunaga of MMUA?

# Closed/Open Set

San José State UNIVERSITY

- Closed set
  - Independent variable = true match within k/N
  - Dependent variables: Threshold and prob (true match within k/N)
- Open set
  - Independent variable= threshold
  - Dependent variables: FMR and FNMR