# Biometrics

Lucía Montero Sanchis

Thursday 1st February, 2018

## Question 1 – Leading Biometric Technology

**Physiological Characteristics**

**01. Fingerprint**

**Analysis, Modeling and Interpretation of Biometric Data**
- Mathematical Tools for Biometric Signal Processing (contextual filters, Local Binary Patterns (LBPs)) and Pattern Recognition (ridge-based, MCC-based, hybrid)
- Sensing and Storage (Multispectral, Finger On The Fly )
- Representation and Feature Extraction (ridge-based, hybrid)
- Models of Features for Recognition and Classification (FingerCode, Minuta Cylinder Code (MCC), Hybrid representations)
- Enrollment and Template Creation (quality control, image enhancement)
- Biometric System Errors (independent evaluation)
- Synthetic Fingerprint Generation

  - Mathematical Tools for Biometric Signal Processing (Gabor filters) and Pattern Recognition (binarization, thinning, crossing number (minutiae detection))
  - Sensing and Storage (optical, capacitive, ultrasonic, etc.)
  - Representation and Feature Extraction (level 1, 2 (minutiae) and 3 features)
  - Models of Features for Recognition and Classification (minutia coordinates and local ridge orientation, Delaunay triangulation and triangular matching)
  - Enrollment and Template Creation
  - Biometric System Errors

**Generalities**:

- Friction skin (hairless, many sweat glands...). *Biological* (physiological) characteristic. *Applications* in security and forensics. *Characteristics*: permanent (except for desease, scarring...), unique (even with same DNA).

- *Galton's details* – It suffices if 12 points are the same between two fingerprints.

- **Lights-out identification** – system requiring minimal or zero human assistance that outputs a short candidate list.

- India Universal ID System with Biometrics – *To give the poor an identity*

- **Fake finger** (spoof attack) – Gelatin, Silicone, Latex.

**Sensing**: Optical, capacitive, ultrasonic...

- *On-line acquisition* – Optical, Capacitive, Piezoelectric scanners. Single- or multi-finger

- *Off-line acquisition* – Thermal scanner, Inked impression, Latent fingerprint.

- Others: touchless (TBS The Surround Imager; Finger On The Fly), acquisition of derma image (internal fingerprint, more reliability).

**Feature extraction**: (level 1, 2 (minutiae) and 3 features)

- Minutiae: Ridge bifurcations, endings, ... (52 types)

- Core: Uppermost point on innermost ridge

- Delta: Separating point between pattern and non-pattern areas

- Pattern class: Determined by ridge flow characteristics.

- **Feature Levels**:

  - **Level 1 – singularities (core points)**: Core and delta points. Classification (left loop, right loop, whorl, arch, tented arch). Ridges (*flow* can be described with *directional map*; line density can be described with *density map*)

    * **Average Square Gradient Method**: Gradient vector lengths are squared and their angles doubled.

    * Noisy fingerprint → *Local structure* for matching is very difficult, but *global structure* is more stable.

  - **Level 2 – Minutiae (major ridge path deviations)**: Ridge ending, Bifurcation, Valley, Lake, Independent Ridge, Point, Spur, Crossover.

  - **Level 3 – Intrinsic or innate ridge formations**: Sweat pores, incipient ridges, creases...

**Templates**: (minutia coordinates and local ridge orientation, Delaunay triangulation and triangular matching)

- **Segmentation** (Isolate foreground from background)

- **Normalisation** (Mean 0 and variance 1, to standardise image intensity values)

- **Orientation image** (Computation of gradients over square-meshed grid)

- **Frequency image** (Ridge frequency)

- **Automatic Minutiae Detection**

  1. Binarization

  2. Thinning

  3. Crossing number (Detection of minutiae)

- **Fingerprint enhancement – Gabor filtering**: Contextual filters (characteristics change according to local context), context defined by local ridge orientation and frequency. Efficiently removes undesired noise preserving ridges and valleys.

- **Fourier Analysis** – Interpreting the spectrum; alternate representation of the signal, providing more information. Filtering and convolution become trivial.

  - Energy map

  - Frequency map

  - Orientation map

  - Fourier Domain Based Enhancement

- **Pre-alignment**:

  - Absolute pre-alignment – w.r.t. core and delta points. First align usign global structure, then use loca lstru for point-to-point matching.

  - Relative pre-alignment – superimposing singularities, correlating orientation images, correlating ridge features.

- **Match Score Generation**

- **Triangular matching** – Delaunay triangulation creates a set of triangles such that triangles don't contain any other minutiae points.

**Matching (comparison)**:

- **Fingerprint Matching**:

  - *Correlation-based* – Correlation between pixels computed for different alignments. Non-linear distortion and missalignments makes the correlation very difficult. Direct application of 2D correlation is very expensive computationally.

  - *Minutiae-based* – Find the alignment that results in maximum number of minutiae pairings. Minutia Cylinder Code (MCC). Time consuming.

    * MCC is computationally fast, easy bit-based implementation, invariant to rotation and translation.

  - *Ridge feature-based* – size and shape of fingerprint, number type and position of singularities, sweat pores, local orientation and frequency, ridge shape... FingerCode. **Gabor filter responses**.

    * *Texture-based Representation* – Ridge pattern in fingerprint is seen as an oriented texture pattern with fixed dominant spatial frequency and orientation in a local neighborhood.

    * *Local texture patterns* – *Local Binary Patterns LBPs*

  - **Hybrid**

- **Intra-variability** – Variations in impressions of one same finger. Overlap, displacement, rotation, non-linear distortion (because finger is 3D), pressure, skin condition, noise, *feature extraction errors*.

  - *Feature extraction errors*: Aggresive enhancement, low-quality images...

- **Local Clarity Score (LCS)** – computes the block wise clarity of ridge and valleys by applying linear regression to determine a gray-level threshold, classifying pixels as ridge or valley.

  - Minutiae + Local Correlation

  - Minutiae + Local Binary Pattern (LBP) Histogram

**02. Face (2D and 3D)**

### Analysis, Modeling and Interpretation of Biometric Data
- Mathematical Tools for Biometric Signal Processing (2D DCT, Local Texture Patterns (LTPs)) and Pattern Recognition (Linear Subspace Analysis - Principal Component Analysis (PCA), Singular Value Decomposition (SVD))
- Sensing and Storage (photo-camera, video-camera, etc.)
- Representation and Feature Extraction (geometric and appearance (local and global features) based)
- Models of Features for Recognition and Classification (GMM, histograms of LTPs, PCA)
- Enrollment and Template Creation
- Biometric System Errors

**Generalities**:

- Three levels of details:
  1. General face geometry, global skin color – <30 *interpupilary distance* (IPD)
  2. Localized face information – 30-75 IPD
  3. Unstructured micro level features: scars, freckles, moles, skin discoloration

- Complications: Facial expression, point of view, illumination, inter-class similarity, automatic face detection, appearance changes (quick, slow).

- Advantages: Low-cost, non-contact, non-intrusive, overt (user aware) or covert (user unaware), legacy databases, socially and culturally accepted, always acquires.

- Proprietary algorithms to generate templates → not-interoperable templates (interoperable *original* photo). *Receiving State* uses its own vendor algorithm to compare taken and stored facial images.

- **Machine Readable Travel Document** (**MRTD**)

- ICAO standard: 300 dpi (112 kB); not cropped or cropped from chin to crown.

- Process:
  1. Face Detection
  2. Face Normalization
  3. Feature Extraction
  4. Classification

**Sensing**:

- **Imaging challenges**:
  - **Acquisition geometry** – Necessary face detection; in-plane (1 degree of freedom) and in-depth (2 degrees of freedom) rotation; scale.
  - **Imaging conditions** – Lighting, intrinsic camera characteristics (automatic white balancing, gain control, noise reduction). *Perfect conditions*: uniform background and lightning, acceptable quality.
  - **Compression artifacts** – Image degradations due to compression for transmission/storage. JPEG and MPEG often used but not designed to preserve human face.

- **Detection** (what faces have in common, tell face from non-faces); *Localization* (find position of one existing face); **Face recognition** (what differentiates two faces)

- Overlapped detections are merged if 60% overlap.

- *Detection* and *Recognition* **approaches**:

  - **Feature/Geometry-based** (analytic: eyes, mouth...) – Matching feature constellations. *Rotated models* (one model for each rotation).
    **Haar features**: *Haar-like transform* calculates difference of intensity in neighbor regions. With low resolution images, better than purely geometrical methods. Applies Haar-like basis functions. **Viola-Jones** face detector.

  - **Appearance-based** (local and global features, analyze distributions of individual faces in face space) – Neural Networks, **PCA**, **LDA**, **HMM**, SVM, **GMM**, Graph matching.
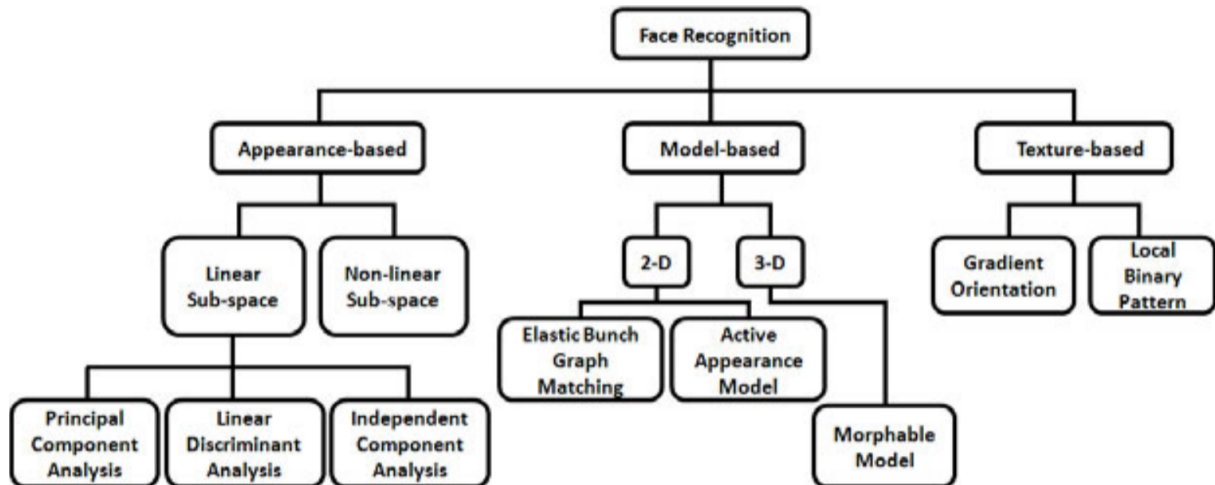


**Feature extraction**:

- **Appearance-based** Face Recognition:

  - *Local* Features (segment image and get several feature vectors):

    * **2D-DCT** based (Discrete Cosine Transform): Divide in $m$ $u$ by $v$ blocks, obtain for each $n = uv$ DCT coefficients $\Rightarrow m$ vectors of length $n$, each vector a row of matrix $M^{m,n}$. Coefficients are ordered in zig-zag pattern.
      Then **GMM**: 2D-DCT features for training; score = log-likelihood(2DDCT | model)

    * Local Texture Patterns:

      · Local Binary Patterns **LBPs**:

      $$\text{LPB}_{P,R} = \sum_{p=0}^{P-1} s(g_p - g_c)2^p$$

      with $g_c$ the gray value of the center pixel $(x_c, y_c)$; $g_p$ the gray values of $P$ equally spaced pixels on a circle of radius $R$; $s$ a thresholding function s.t. $s = 1$ if $x \geq 0$, else 0.
      Histogram intersection measure: $H(p,q) = \frac{\sum_i \min{(p_i, q_i)}}{1/2 \cdot (\sum_i q_i + \sum_i p_i)}$

      · Local Ternary Patterns **LTPs**: One ternary code can be expressed as two binary codes (one is sustracted from the other). Ternary Codes use -1, 0 and 1 (instead of 1 and 0 only)

  - *Global* Features (Holistic approach, obtain one single feature vector):

    * Principal Component Analysis **PCA**

    * Linear Discriminant Analysis **LDA**

  - **2D Model: Elastic Bunch Graph Matching** (*relational approach*, requires $\geq 2$ images)

  - **3D Morphable Model**

**Templates**:

- 

**Matching (comparison):**

-

## Iris – 31.10.2017

### Analysis, Modeling and Interpretation of Biometric Data
- Mathematical Tools for Biometric Signal Processing (Complex Gabor Wavelets (CGW)) and Pattern Recognition (Hamming Distance, Binomial Distribution, circular edge detection)
- Sensing and Storage (Proprietary camera systems)
- Representation and Feature Extraction (from Cartesian to polar system, phase quantization from CGW or DFT)
- Models of Features for Recognition and Classification (IrisCode and Hamming distance based)
- Enrollment and Template Creation
- Biometric System Errors (Very low identification errors)
- Large Scale Iris Identification Systems (IrisGuard)

**Generalities**:

- Biological. Distinctive features: arching ligaments, furrows, ridges, crypts, rings, corona, freckles, etc. Random pattern, mostly stable through life.

- **Epigenetic** (not genetically determined, as opposed to *genotypic*).

- eBorders in the United Arab States.

- **Modules**:

  1. **Acquisition** – Get 2D image with *monochromatic CCD camera* sensitive to *NIR* (Near InfrarRed: 700-900 nm) light spectrum.
  Infrared light is preferred because it's more reflected by melanin than visible light, thus more iris texture is visible. Illumination is ANSI and Cenelec Certified safe to use. Pigmentation variations in iris are due to melanin density and are invisible in the NIR.

  2. **Segmentation** – *Localize iris* in eye image
  Edge Detection and Hough Transform – The operator is a circular edge detector, can be used to detect the iris as well as the eyelid boundaries.

  3. **Normalization** – Geometric normalization from *Cartesian* to *polar* coordinates $(r, \theta)$ with $r \in [0, 1]$ and $\theta \in [0, 2\pi]$
  This compensates pupil dilation and iris size inconsistencies but not rotational inconsistencies (this is accounted for during *matching* by shifting the iris templates in the $\theta$ direction until templates are aligned)

  4. **Encoding** – Feature extraction to produce a binary code. Invariant to pupil dilation and iris size. It's reliable and false. False Match Rate very small.
  2048 bits (256 bytes) are extracted from iris image and an equal number of masking bits to signify whether any region should be ignored in the demodulation code.
  The information extracted from the iris is described in terms of *phase* $\Rightarrow$ *insensitive* to contrast, camera gain, and illumination level (unlike correlation methods).
  Correlations within an iris (local structure is self-predicting). All IrisCode bits are equally likely to be 0 or 1 $\Rightarrow$ IrisCode have *maximum entropy* bitwise.

  5. **Comparison** – How closely the produced code matches the encoded features in the database. Fractional Hamming Distance ($HD$, fraction of bits that disagree – 10% for same eye, 45% for different)
  $$HD_{\text{raw}} = \frac{||(\text{code}_A \oplus \text{code}_B) \wedge \text{mask}_A \wedge \text{mask}_B||}{||\text{mask}_A \wedge \text{mask}_B||}$$

- Iris Code bit comparisons are Bernoulli Trials – *binomial distribution*:
  $f(x) = \frac{N!}{m!(N-m)!} p^m (1-p)^{N-m}$

- If *less iris visible* then decision criterion becomes *more demanding*:
  $HD_{\text{crit}} = 0.32 - 0.01 \cdot \log_{10} N$

- Score re-normalisation to compensate for number of bits compared:
  $HD_{\text{norm}} = 0.5 - (0.5 - HD_{\text{raw}})\sqrt{n/911}$

- Fewer (more) than 911 bits penalizes (improves) Hamming Distance. 1152 bits corresponds to 100% visibility of each iris.

- Extremely fast when computed in parallel (1M IrisCodes per second with 3 Gb)

**Sensing**:

- Techniques for improving user interface:

  - Use extremely high resolution CCD

    que es CCD?

  - Well-designed optical system to improve DOF (Depth of Field)

  - Cold mirror for user to adjust his eye

  - Auto-focus

  - Active pan/tilt camera optics

  - Facial feature detection and tracking for guiding

  - Distance sensor or image content based distance estimation

  - Dual-eye iris camera

**Matching (comparison)**:

- **Hamming Distance** – Iris Codes are matched using a XOR to find bits at which they differ. AND is used to ensure that the compared bits are uncorrupted.

- **Liveness detection** – Photonic and *spectrographic* countermeasures (tissue, fat, blood and melanin pigment spectra; *red eye* effect) and *behavioral* countermeasures.

- *Contact Lenses* and *Large difference in dilation* result in few more False Rejections.

**04. Periocular, Retina, Ear**

**Generalities**:

- 

**Sensing**:

- 

**Feature extraction**:

- 

**Templates**:

- 

**Matching (comparison)**:

-

**05. Hand Geometry**

**Generalities:**

- 

**Sensing:**

- 

**Feature extraction:**

- 

**Templates:**

- 

**Matching (comparison):**

-

**06. Palmprint and Palm Veins**

**Generalities:**

- 

**Sensing:**

- 

**Feature extraction:**

- 

**Templates:**

- 

**Matching (comparison):**

-

**Behavioral Characteristics**

**07. Voice (02, 03)**

**Generalities**:

- **Voice biometric** combines physiological and behavioral characteristics. Useful for remote-access transactions over telecommunication networks. Voice is subject to many sources of variability.

- **Disambiguation** – Voice recognition can refer to *Speaker recognition* (who is speaking) or *Speech recognition* (what is being said).

- **Perceptual Cues**: *High-level* (learned behaviors – Semantic, Dialogic, Idiolectal that depend on status, education, place of birth) vs *Low-level* (physical characteristics – Spectral, Prosodic, Phonetic that depend on anatomical structure)

**Sensing**:

- **Microphones** (e.g. in smartphones)

- **Speech signal** is real, continuous, non-stationary, 4-dimensional (4D), has finite energy. It's complex and variable over time. Sometimes periodic (pseudo-periodic) for voiced sounds; Sometimes random for fricative sounds; Sometimes impulsive in explosive phases of occlusive sounds.

- **Bandwidth** – Frequency Band of Telephone Speech: 300 Hz – 3.4 kHz

- **Coding Bands**:
    - Hi-Fi: 20 Hz – 20 kHz (sampling frequency 44.1 kHz)
    - Wideband: 50 Hz – 7 kHz (sampling frequency 16 kHz)
    - Narrow band: 300 Hz – 3.4 kHz (sampling frequency 8 kHz)

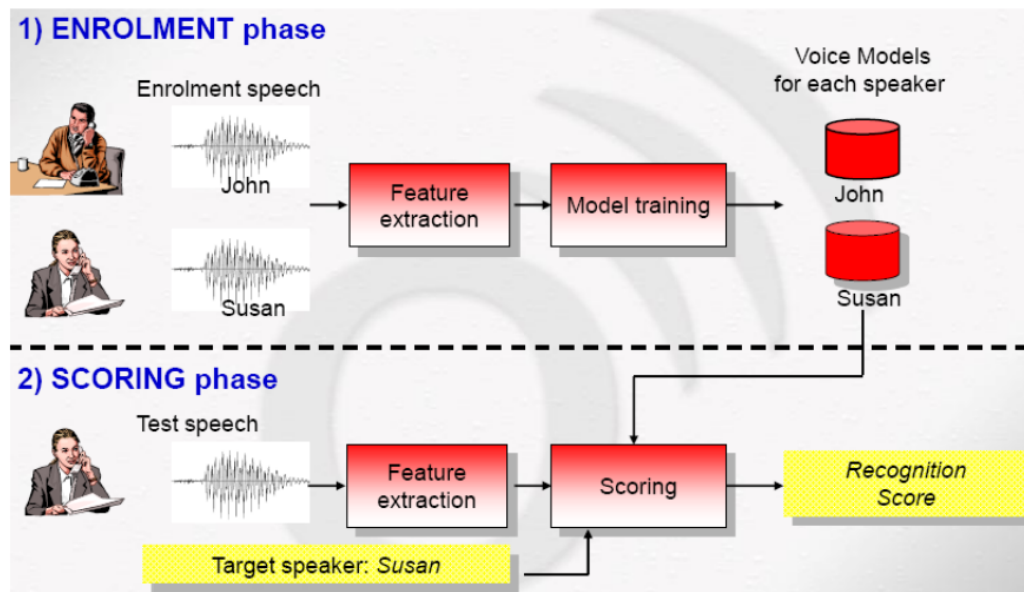- **Sampling** and Uniform **Quantization**:

**Figure 1:** Enrolment and Scoring in Speaker Recognition

**Feature extraction**: Spectral envelope, MFCC

this??

- **Mel-Frequency Cepstral Coefficients (MFCCs)**: Mel-Frequency Cepstrum (MFC) is a representation of spectral energy of a sound on the mel scale, and is made up by the MFCCs (coefficients). In an MFC the frequency bands approximates the human auditory system's response – better representation of sound.

  - GMMs are used to capture the distribution of MFCCs in the feature space.

  - We enroll a speaker by adapting the UBM using the speaker's input.

  - Are obtained through FFT (Short-term transform), Logarithm, Deconvolution source/tract (cepstre). The acoustic vector consists of cepstrum, delta-cepstrum and delta-delta-cepstrum

  - **Short-Term Feature Extraction**:

    Window, frame, feature vector (acoustic vector)

  - **Short-term Processing**:

    window N, frame M, window and frame duration definitions

  - **Short-term DFT**

  - **Spectral envelope**

    (+spectrogram)

  - **Real Cepstrum**

  - **Dynamic features**

- Recent research on supervectors and i-vectors.

**Templates**:

- **Template** is a compact, electronic representation of a biometric sample that is created at the time of enrollment and stored in the system database for future reference and comparison. The process of creating a template and storing it in the database is called *enrollment*.

- **Speech Modalities**:

– *Text-dependent*: System knows text spoken (e.g. find fixed phrase, prompted phrase). Used for strong control over user input. Improved system performance.

– *Text-independent*: Not know text spoken (e.g. conversation). Less control over user input. More flexible and more difficult.

- **Vector Quantization**

```
explain VQ
```

| | Deterministic | Statistical |
|---|---|---|
| **Text-dependent** | Dynamic Time Warping (DTW) | Hidden Markov Model (HMM) |
| **Text-independent** | Vector Quantization (VQ) | Gaussian Mixture Model (GMM) |

**Table 1:** *Templates and Models* in Speaker Recognition

**Matching (comparison)**:

- **Comparative analysis** is a comparison between two biometrics, typically a tested sample and an enrollment (reference) template or model. The output of the comparison is a distance or score.

- **Speaker Recognition Systems**:

  – *Conventional Speaker Verification System*
  Enrollment:
  GMM training → Speaker-dependent GMM → Database

  – *SV with Verbal Information Verification*
  Automatic enrollment:
  Verbal Information Verification ⇒ Save for training → Verified pass-phrases for training → HMM training → Speaker-dependent HMM → Database

  – *Text-prompted Speaker Recognition*: Uses **speaker-specific phoneme models** as basic acoustic units. New prompted sentence every time – can't cheat with pre-recordings.

- **Prerequisites for good performance in Speaker Recognition**: Speakers must not disguise their voices; Recording conditions and signal processing techniques are known or controlled; Speech recorded in conditions similar to those of the test signal is available; Reference values for similarity measures established in similar conditions as the test signal; Decision thresholds calibrated according to reference values depending on the application.

**08. Dynamic Signature**

### Analysis, Modeling and Interpretation of Biometric Data
- Mathematical Tools for Biometric Signal Processing and Pattern Recognition (Hidden Markov Model (HMM))
- Sensing and Storage (Multivariate sensing)
- Representation and Feature Extraction (Local and Global Features)
- Models of Features for Recognition and Classification (DTW, GMM, VQ, HMM chain with discrete distribution of observation probabilities, Continuous Density HMM (CDHMM), Viterbi algorithm for maximal probability scoring),
- Enrollment and Template Creation (Baum-Welch algorithm for re-estimation of HMM chain parameters)
- Biometric System Errors

**Generalities**:

- *Behavioral* characteristic. Combines *knowledge* and *biometric*. Not *permanent* (invariant over time). Currently can be digitalized and is in ID cards.

- **Applications**: Signature forensics, Signature authentication, Signature surveillance, Digital Rights Management, Biometric cryptosystems.

- **Off-line** or **Static** – scanned from paper documents, written conventionally.

- **On-line** or **Dynamic** – written with electronic device. Dynamic information (pen tip location through time) usually available at high resolution, even if pen not in contact with paper.

**Sensing**: Multivariate sensing

- Tablets, smartphones, IKEA's and Sunrise's SignPad, UPS and SwissPost...

**Feature extraction**:

- **Basic (Local) Features:**

  1. X, Y coordinates

  2. Velocity

  3. Acceleration

  4. Pen azimuth (0 - 359 deg)

  5. Pen altitude (0 - 90 deg)

  6. Pressure

  7. First and second derivatives of feature

- **Global features (more than 150)**

  - Signature length, height, weight

  - Total signature time

  - Total pen-down and pen-up time

  - Avg., max. and min. velocity

- Pre-processing: Smoothing; Segmentation (determine beginning and end); Initial point alignment.

- Forgery: Zero-effort; Home-improved (based on static image); Over-the-shoulder (observe signing); Professional (skilled individuals). Over-the-shoulder + Home-improved combo is called *skilled*.

**Templates**

- *Models of Features for Recognition and Classification* are DTW, GMM, HMM chain with discrete distribution of observation probabilities, Continuous Density HMM (CDHMM), Viterbi algorithm for maximal probability scoring)

- *Enrolment and Template Creation*: Baum-Welch algorithm for re-estimation of HMM chain parameters.

**Matching (comparison)**:

- DET curves to compare different GMMs, HMM...

- Resistant to imposters, non-invasive, can be changed by users, fast and intuitive enrolment, fast verification, independent of native language user.

- Inconsistent signatures increase error rates, limited applications, for good accuracy a 5D pen is needed (costly), some people can't sign.

**09. Gait, Typing Rhythm**

**Generalities**:

- 

**Sensing**:

- 

**Feature extraction**:

- 

**Templates**:

- 

**Matching (comparison)**:

- 

# Biological Traces

**10. DNA**

**Generalities**:

- 

**Sensing**:

- 

**Feature extraction**:

- 

**Templates**:

- 

**Matching (comparison)**:

-

# Synthetic Biometric Data Generation

**11. Synthesis of Fingerprints**

**Synthetic Fingerprint Generation**: To automatically create large databases of fingerprints, allowing to train, test, optimize and compare algorithms. Collecting fingerprints is expensive (money and time) and problematic (privacy legislation).

1. Select → Obtain:
   (a) shape parameters → **fingerprint shape**
   (b) class and singularities → **directional map** model
   (c) average density (and singularities) → **density map** model

2. **Ridge pattern** generation, to obtain **master** fingreprint
   - Gabor filters iteratively applied to an initially white image, enriched with a few random points. Orientation and frequency of the filters are locally adjusted according to directional and density maps.
   - As a result, realistic minutiae appear at random positions.

3. Add **variability** data:
   (a) Determine contact region and erosion/dilation (low-pressure or dry/high-pressure or wet)
   (b) Skin deformation
   (c) Noise and rendering
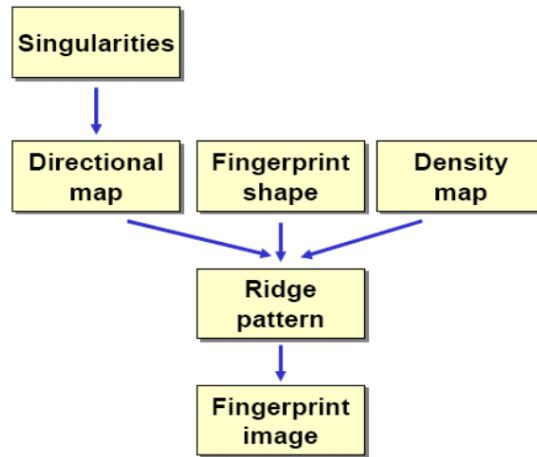   (d) Translation and rotation
   (e) Generate background



**Figure 2:** Synthetic Fingerprint Generation

# Multimodal Biometrics

**12. Multimodal Biometrics**

**Generalities:**

- 

**Sensing:**

- 

**Feature extraction:**

- 

**Templates:**

- 

**Matching (comparison):**

- 

# Miscellaneous

**13. Quality and Ageing in Classication of Biometric Data**

# Question 2 – Other Topics

## Fundamentals of Biometrics (01)

### 01. Identity and Biometrics

The role of Biometrics is to recognize a person by their body traits and link the body to an externally assigned identity.

**Biometrics**:

- *Biometrics* – automated recognition of individuals based on biological and behavioral characteristics
- *Biometry* – statistical and mathematical methods applicable to data analysis problems in the biological sciences
- *Biometric system* – automatic pattern recognition system that recognizes a person by verifying the authenticity of a specific biological and/or behavioral characteristic (biometric modality) they possess
- (forensic, judicial) *Anthropometry* – (identification of criminals by) measurement techniques of human body and its specific parts

**Identity**:

- *Identity* – whatever makes something the same or different.
- *Authentication* (identity verification) – process to link a physical person with a certain identity

People are identified by **three basic means**:

- Something they *have* – identity document or card, passport, birth certificate, token...
- Something they *know* – password, PIN, name, date of birth
- Something they *are* – human body

**Security level of each solution**:

$$\text{Know} \quad < \quad \text{Know} + \text{Have} \quad < \quad \text{Know} + \text{Are} \quad < \quad \text{Know} + \text{Have} + \text{Are}$$

**Advantages of Biometric identifiers**: *Security*; *Convenience*; *Audit trial*; *Avoid fraud*; *De-duplication*. Examples: automated comparison process occurs in seconds; can replace passwords (often forgotten, lost, or misappropriated); identity justification without paperwork.

**Ideal biometric identifier**: *Universality* (every person has it); *Uniqueness* (different for every person); *Permanence* (invariant in time); *Collectability* (measurable, practical); *Acceptability* (public has no strong objections).

**Challenge of biometrics**: *Scalability, Usability, Accuracy.*

**Identifiable biometric characteristics**: *Biological traces* (DNA, blood, saliva); *Biological (physiological) characteristics* (fingerprint, iris, retina, hand palm, hand veins, hand geometry, facial geometry); *Behavioral characteristics* (dynamic signature, gait, keystroke dynamics, lip motion); *Combined* (voice).

**Comparison of biometric techniques** (Cost and Accuracy):

| Cost: | Voice | < | Face | ≈ | Fingerprint | < | Signature | < | Hand | < | Iris |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Accuracy: | Voice | ≈ | Face | ≈ | Signature | < | Hand | < | Fingerprint | < | Iris |

**02. Recognition, Verification, Identification and Authentication (01)**

**Recognition** – used when we do not distinguish between verification, identification and authentication.

**Verification** – performs one-to-one comparison of a submitted biometric characteristic (sample) set against a specified stored biometric reference, and returns the comparison score and decision (deciding whether a sample belongs to a specified person) – *Is this person who he claims to be?*

**Identification** – performs one-to-many comparison/search to determine the identity of the user from a known set of identities – *Who is this person?*

**Authentication** – the user claims an identity and the system verifies whether the claim is genuine (link a person with a chosen identity).
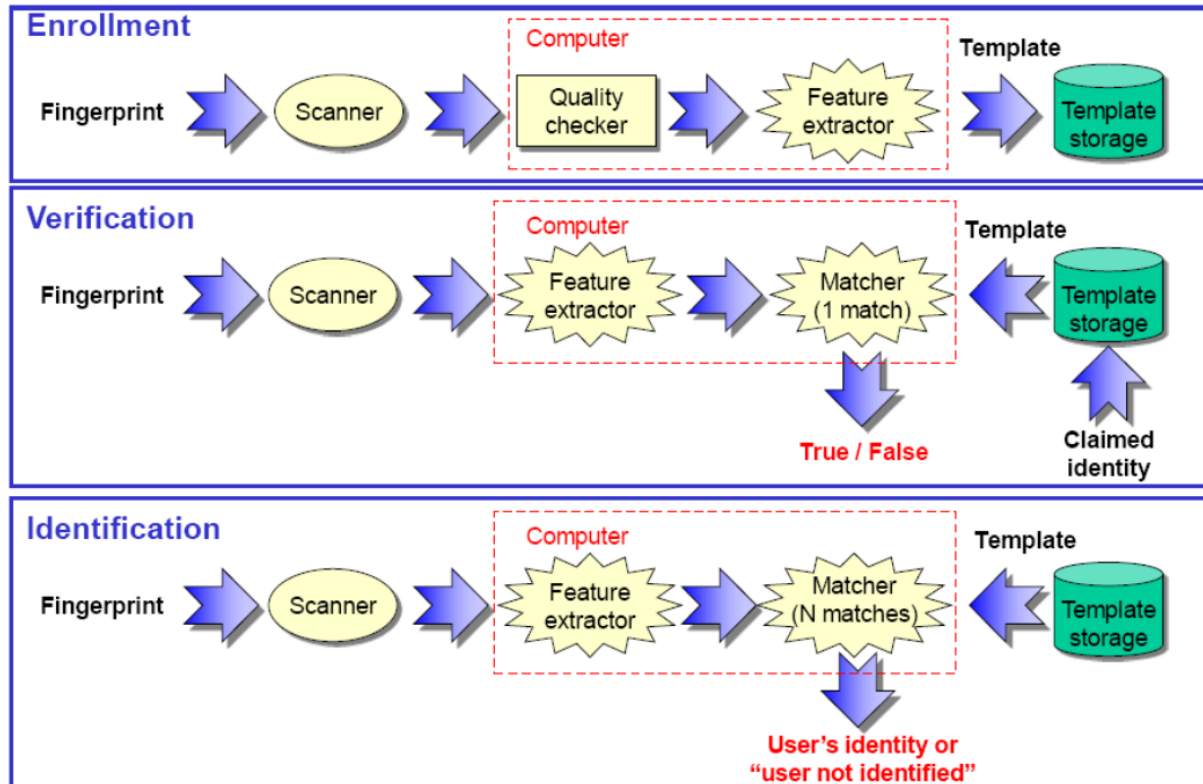


**Figure 3:** Enrolment, Verification and Identification

## Analysis, Modeling and Interpretation of Biometric Data

**03. Mathematical Tools: Dynamic Time Warping (DTW) (02,03)**

- In voice: Algorithm for measuring **dissimilarity** (distance) between two temporal sequences, which may vary in speed. Given a test word $T$ and reference words $R_1, ... R_N$ (all represented by sequences of feature vectors), we choose the reference word $R_r$ with smallest distance $D(T, R_r)$.

- In voice: The recognition system is adapted to the single speaker who uttered the reference word. Limited vocabulary, without words too close phonetically. The words to be recognized are pronounced in an environment free of noise. They could be isolated in a perfect manner.

- **Non-linear time alignment**: The sequences are *warped* non-linearly in time dimension to determine a measure of their similarity independent of certain non-linear time variations.

- **Algorithm**. *Conditions*: Boundary; monotonicity; step size.
  - Recursively calculate a minimum accumulated distance for each point $(i, j)$ taking into account some local heuristic constraints and weights.
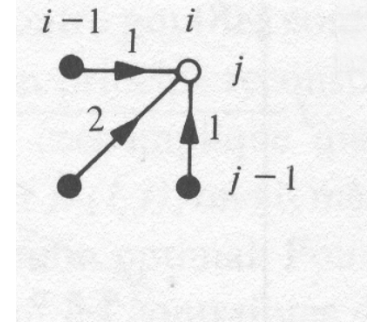
– Each grid point $(i, j)$ is associatied with a *local distance* $d(i, j)$ and an accumulated distance $D(i, j)$.
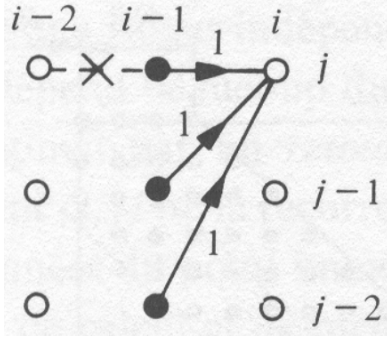


**(a)** Local constraints (A)
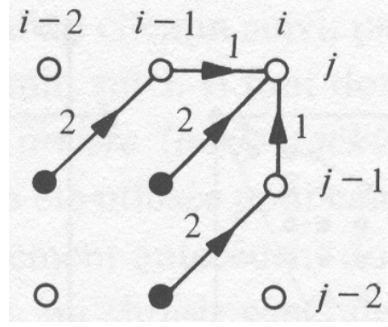


**(b)** Local constraints (B)



**(c)** Local constraints (C)

– Type C constraints: Allow a path of any shape satisfying monotonicity.

– Spectral distance:

> esto?

## 04. Mathematical Tools: Gaussian Mixture Model (GMM) (02,03)

- A Gaussian Mixture Model (GMM) is a parametric probability density function represented as a weighted sum of Gaussian component densities.

- Under the assumption that any arbitrary probability density function (PDF) can be approximated by a linear combination of uni-modal Gaussian densities, the Gaussian mixture models (GMMs) have been applied to model the distribution of a sequence of D-dimensional feature vectors.

- The sum of *mixture weights* equals 1.

- Model parameters: $\lambda = \{w_i, \mu_i, \Sigma_i\}$ (estimated with Expectation Maximization algorithm, although can also be estimated with Maximum A Posteriori estimation).

  – Expectation step: Compute a posteriori probability for component $i$

  – Maximization step: Maximize, guaranteeing a monotonic increase in model's likelihood.

- MAP estimation is used in speaker recognition to derive speaker model by adapting from a speaker independent universal background model (UBM), or to adapt a prior, general model.

- Decision is carried out using a likelihood test with $H_0$ (tested recording and speaker's model are from same source), $H_1$ (not $H_0$) and Bayes theorem ($\sigma$ is the decision threshold):

$$\frac{P(T|\lambda_0)}{P(T|\lambda_1)} > \sigma$$

- Similarity domain normalization: $\log L(X) = \log p(X|S = S_c) - \log \sum_{S \in \text{Cohort}} p(X|S \neq S_c)$

- Normalization by a general/world model: A Gaussian mixture which models the parameter distribution for free-text utterances by many speakers.

**05. Mathematical Tools: Hidden Markov Model (HMM) (02,03)**

- HMM is a statistical Markov model in which the system being modeled is assumed to be a Markov process with unobserved (i.e. hidden) states. Is doubly probabilistic finite-state machine.

- **Ergodicity**: There is transition from any state to any other state.

- **Three Basic HMM Problems**:

  1. **Decoding** – Given the observation sequence $X = [x(1), x(2), ..., x(t), ..., x(L)]$ and the word model $W = (A, B)$, how do we choose a state sequence $Q = [q(1), q(2), ..., q(t), ..., q(L+1)]$ that is optimal in some meaningful sense sense (e.g. maximal probability)?

  2. **Evaluation** – Given the observation sequence $X = [x(1), x(2), ..., x(t), ..., x(L)]$ and a word model $W = (A, B)$, how do we (efficiently) compute $P(X|W)$ (probability of the observation sequence)?

     - The Baum-Welch algorithm (**forward**): $\alpha_j(t) = \sum_i \alpha_i(t-1) \cdot \alpha_{ij} \cdot B_{ij}(X(t))$

     - The Baum-Welch algorithm (**backward**): $\beta_i(t) = \sum_j B_{ij}(X(t+1)) \cdot \alpha_{ij} \cdot \beta_j(t+1)$

     - **Total probability**: $P(X|W) = \alpha(L, q_F) = \beta(0, q|I)$

  3. **Training** – How do we adjust the model parameters $W = (A, B, \pi)$ to maximize $P(X|W)$? Algorithm de Baum-Welch (forward-backward):

     - Calculate *all forward-backward* probabilities for all states $q_i$

     - Calculate posterior probability of transitions $\gamma_{ij}$, from state $i$ to state $j$, conditioned on the observation sequence and the model.

     - Obtain a new estimate $a_{ij} = \gamma_{ij}(X)/\gamma_i$

     - If the value of the total probability has not improved compared to the previous iteration, the re-estimation has converged.

- Continuous Density HMM (CD-HMM) – Parametric approach: Continues probability density functions (ergodic and left-right models).

- *Viterbi Algorithm*: By induction, find the path that leads to a *Max. Likelihood* considering the best likelihood at the previous step and the transitions from it.

**06. Mathematical Tools: Principal Component Analysis (PCA) – Karhunen-Loeve transformation**

- *Seeks directions that are efficient for representing the data, maximize determinant of total scatter* – Find set of parameters s.t. most variability in the data is compressed in first parameters. The transformed PCA parameters are orthogonal. It diagonalizes the covariance matrix → diagonal elements are the variances of the transformed PCA parameters.

- Advantages and disadvantages:

  - *Advantages*: Completely decorrelates; packs the most variance in the fewest number of transform coefficients; minimizes MSE between reconstructed and original data; minimizes the total entropy of the data.

  - *Disadvantages*: Not fast implementation and computationally costly to get eigenvalues and eigenvectors; not a fixed transform – needs to be generated for each type of data statistic; problems for different illumination, pose, expression.

- To obtain $K$ eigenfaces:

  1. Let $X_{DM}$ the matrix that has features per rows ($X_{DM}$ has been mean-centered, $X = X_0 - \mu$).

  2. Compute covariance (scatter) matrix $C_{DD} = X \cdot X^T$ (or $C_{MM} = X^T \cdot X$, if $D \gg M$)

  3. SVD: Compute eigenvectors and Eigenvalues of chosen $C$

  4. Choose $K$ largest eigenvalues

5. Form $W_{DK}$ with $K$ columns of eigenvectors

6. Transform data/features by projecting onto face space: $X_{\text{PCA}} = W^T \cdot X$

7. Extra stuff: $E_{DD} = W^T C_{DD} W$, $E_{MM} = V^T C_{MM} V$ (both $E$ are the same, only adding 0 to get correct dimension); $X = W \cdot E \cdot V^T$

- To recognize a face $r$:

  1. Subtract average face from it $r_m = r - m$

  2. Compute its projection onto the face space $x_{PCA} = W^T \cdot r_m$

  3. Compute its difference w.r.t all known faces

  4. Reconstruct the face from eigenfaces: $r_{PCA} = W \cdot x_{PCA}$

  5. Distance between the face and its reconstruction: $|r_m - r_{PCA}|_2^2$ (depending on value, $r$ can be not a face, a new face or a known face)

## 07. Mathematical Tools: Linear Discriminant Analysis (LDA)

- *Seeks directions that are efficient for discrimination between the data*: maximizes ratio between determinant of *between-class scatter* and determinant of *within-class scatter*.

- Problems: Small databases; the face to classify must be in database.

- Compute Fisherfaces:

  1. Compute average of all faces ($m$)

  2. Compute average face of each person $x_i = 1/2 \cdot (a_i + b_i)$

  3. Substract average person face for training faces ($a_i^m = a_i$-$x_i$, $b_i^m = b_i$-$x_i$)

  4. Build **between-class scatter matrix** $S_B = \sum_i 2 \cdot |x_i - m|_2^2$

  5. Build scatter matrices $S_i = a_i^m \cdot (a_i^m)^T + b_i^m \cdot (b_i^m)^T$

  6. Build **within-class scatter matrix** $S_W = \sum_i S_i$

  7. We are seeking the matrix $W$ maximizing the ratio between *between-class variance* and *within-class variance* is maximized:
  $$J(W) = \frac{|W^T S_B W|}{|W^T S_W W|}$$

  8. Project faces onto LDA-space: $x_{LDA} = W^T x$

- To classify a face:

  1. Project it onto LDA space

  2. Run a nearest-neighbor classifier

- **Comparison with PCA**:

| LDA | PCA |
|---|---|
| *Fisherfaces* | *Eigenfaces* |
| Discrimination | Uncorrelated representation |
| Maximize *inter* to *intra* class variability | Lower dimensional space |
| $N^2 \Rightarrow P - 1$, $P$ nb. classes | $N^2 \Rightarrow K$ |
| Works with different illumination | Problems with different illumination |
| Same intra-class variability for all classes | Verify if it's a face |

- **Combination of PCA and LDA** – First apply PCA to reduce dimensionality, then apply LDA:

$$W_{\text{Fisher}} = \arg\max_W \frac{W^T W_{\text{PCA}}^T S_B W_{\text{PCA}} W}{W^T W_{\text{PCA}}^T S_W W_{\text{PCA}} W}$$

**08. Enrollment and Template Creation**

**09. Verification and Identification System Errors**

**10. Evaluation of Biometric Systems (01,03)**

**Performance evaluation**:

- **Evaluation factors**: Speech quality, Speech modality, Speech duration, Speaker population.
- **FMR** and **FNMR**:
  - False Match Rate **FMR** – proportion of impostor attempt samples falsely declared to match the compared nonself template (number of impostor FMs / number of impostor attempts)
  - False Non-Match Rate **FNMR** – proportion of genuine attempt samples falsely declared not to match the template of the same characteristic from the same user submitting the sample (number of genuine FNMs / number of genuine attempts)
  - **Calculation**: Create biometric templates using training data set. Define a test set with genuine and impostor trials. Run test and group genuine and impostor scores. Choose threshold value T and calculate FMR(T) and FNMR(T).
- **FAR** and **FRR**:
  - False Accept Rate **FAR** – Proportion of imposters accepted (security breaches)
    **FAR**=FMRx(1-FTA)
  - False Reject Rate **FRR** – Proportion of genuine users rejected (inconvenience)
    **FRR**=FTA+FNMRx(1-FTA)
- **DET** and **ROC** curves:
  - Detection Error Tradeoff **DET** – can be computed from distributions of scores with a variable threshold. FAR vs FRR
  - Receiver Operating Characteristic **ROC** – Correct Accept Rate as function of False Accept Rate (FAR)
- **CAR** and **CRR**:
  - Convenience – Correct Accept Rate **CAR**=1-FRR
  - Security – Correct Reject Rate **CRR**=1-FAR
- **FTA** and **FTE**:
  - Failure to Acquire Rate **FTA** – Proportion that cannot be verified (does not process a certain biometric)
  - Failure to Enroll Rate **FTE** – Proportion that cannot be enrolled (system fails to complete the enrolment process, due to bad quality)
- **Performance measures for identification**:
  - Correct Identification Rate (**CIR**) – proportion of identification transactions by users in the system s.t. the user's identifier is among the ones returned.
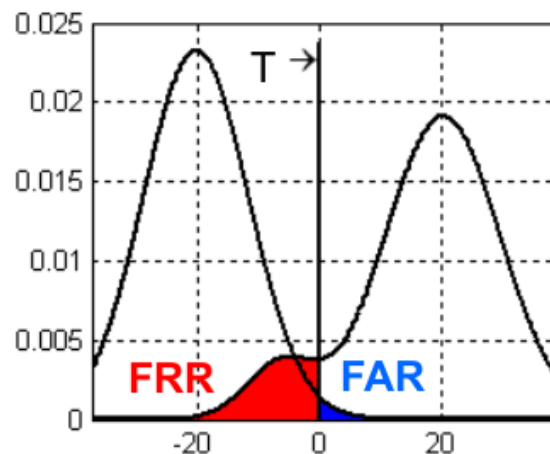  - Cumulative Match Characteristic (**CMC**) – Identification rate as function of K.

**Figure 5:** Threshold T, FRR and FAR

## Miscellaneous

**11. Forensic Automatic Speaker Recognition (03)**



- **Forensic Biometrics** – Challenge is to automate forensic biometric methods. Applications of biometric principles and methods to the investigation of criminal activities: to demonstrate the existence of a crime and determine the author. *Forensic* means the use of science or technology in the investigation and establishment of facts or evidence in the court of law.

- **Existing systems and databases**:
  - Automatic Fingerprint Identification System (AFIS) and fingerprints databases
  - DNA sequencers and DNA databases
  - Challenge: Automatic Biometric Identification System (ABIS) and databases for voice, face...

- **Speaker Identification Integrated Project (SIIP)** – Aims to develop technology to rapidly identify suspects' voices and isolate conversations of interest in a wide range of cases (kidnapping, ransom or terrorist calls...)

- **Forensic Speaker Recognition (FSR)**:

– **Aural-perceptual methods**: earwitnesses, line-ups

  – **Visual methods and *voiceprint?***: visual comparison of spectrograms of linguistically identical utterances (utterly misleading!)

  – **Aural-instrumental methods**: analytical acoustic approach combined with an auditory phonetic analysis

  – **Automatic methods**:

    * *Speaker verification* – not adequate

    * *Speaker identification* – not adequate

    * *Voice as biometric evidence* (How to measure biometric evidence?)

- **Automatic Speaker Recognition (ASR)**:

  – FASR $\neq$ Speaker Verification

  – $H_0$ ($H_1$) – speaker's model $\lambda_0$ and the tested recording $T$ have same (different) source.

$$\frac{P(H_0)}{P(H_1)} \cdot \frac{P(T|H_0)}{P(T|H_1)} = \frac{P(H_0|T)}{P(H_1|T)} \qquad \frac{P(T|\lambda_0)}{P(T|\lambda_1)} > \sigma \qquad (\sigma - \text{Decision threshold})$$

- See Table 1.

- **Forensic Automatic Speaker Recognition (FASR)**:

  – **Forensic specifity** – Role of *forensic science* is to testify to the worth of the evidence *quantitatively* (if possible). Forensic science provides *opinion* to help investigators and courts of law answer important questions. Up to the *judge/jury* to use this information in deliberations and decision.

  – **Evaluative forensic science opinion** – opinion of evidential weight, based upon case specific propositions (hypotheses) and clear conditioning information (framework of circumstances) that is provided for use as evidence in court. Is based upon the estimation of a likelihood ratio (in relation with Bayesian interpretation of evidence).

- **Bayesian Interpretation of Biometric Evidence** ($H_0 \equiv$ suspected speaker is source of the recording). Via Bayes Rule, we use the data to update prior beliefs about unknowns. See Figure 6. Freedom of: choosing evidence evaluation and its value; formulating propositions; choosing automatic speaker recognition method.



**Figure 6:** Odds form of Bayes' Theorem: Bayesian Interpretation of Forensic Evidence

- **Measures**:

  1. **Biometric Evidence** – Quantified degree of *similarity* between the speaker dependent features extracted from the trace and the extracted from recorded speech of a suspect (model).

    – **FASR – Univariate (Scoring) Method** – See Figure 1. The *score* is used together with the distributions of *between-sources variability* and the *within-source variability* to reach a decision.

2. **Strength of Evidence – Likelihood Ratio** – A likelihood ratio $LR = P(E|H_0)/P(E|H_1)$ of 9.16 means that it is 9.16 times more likely to observe the score (E) given $H_0$ than given $H_1$. If $LR > 1$ then $H_0$, else $H_1$.

    – **FASR – Multivariate (direct) Method** – E is the multivariate feature representation of trace evidence.

3. **Evaluation of the Strength of Evidence** – (similar idea to Figure 5) *Principle*: Estimation and comparison of likelihood ratios that can be obtained from same speaker and different speaker trials. $H_0$ true $\rightarrow$suspected person recording and questioned recording are from same speaker.

    – Tippett plots I to obtain *Probability of Misleading Evidence (accuracy)* PME$H_0$ and PME$H_1$.

    – Tippett plots II to obtain EPP (Equal Proportion Probability), PD (Probabilistic Distance) of case $LR_{case}$ to PME$H_0$.

    – Empirical Cross-Entropy (ECE) and Log-Likelihood Cost (CLLR)

    these two?

12. **Forensic Biometrics (Fingerprints, Face, DNA, Ear, Gait)**

- **Fingerprints**:

$$LR = \frac{p(\text{evidence}|H_p)}{p(\text{evidence}|H_d)}$$

with $H_p$ (suspect left the fingerprint), $H_d$ (someone else left the fingerprint), numerator (variability of minutiae configurations due to distortion and clarity), denominator (variability between minutiae from different sources). Delaunay triangulation, MCC and Local Quality Measures (embedding quality measures), ridge quality maps are used.

13. **Biometric Standards**

14. **Securing Biometric Data and Biometric Encryption**

15. **Biometrics in Identity Documents**

16. **Privacy and Legal Issues**