



# APSIPA

Asia-Pacific Signal and Information Processing Association



# ***Speaker Verification – The present and future of voiceprint based security***

***Prof. Eliathamby Ambikairajah***

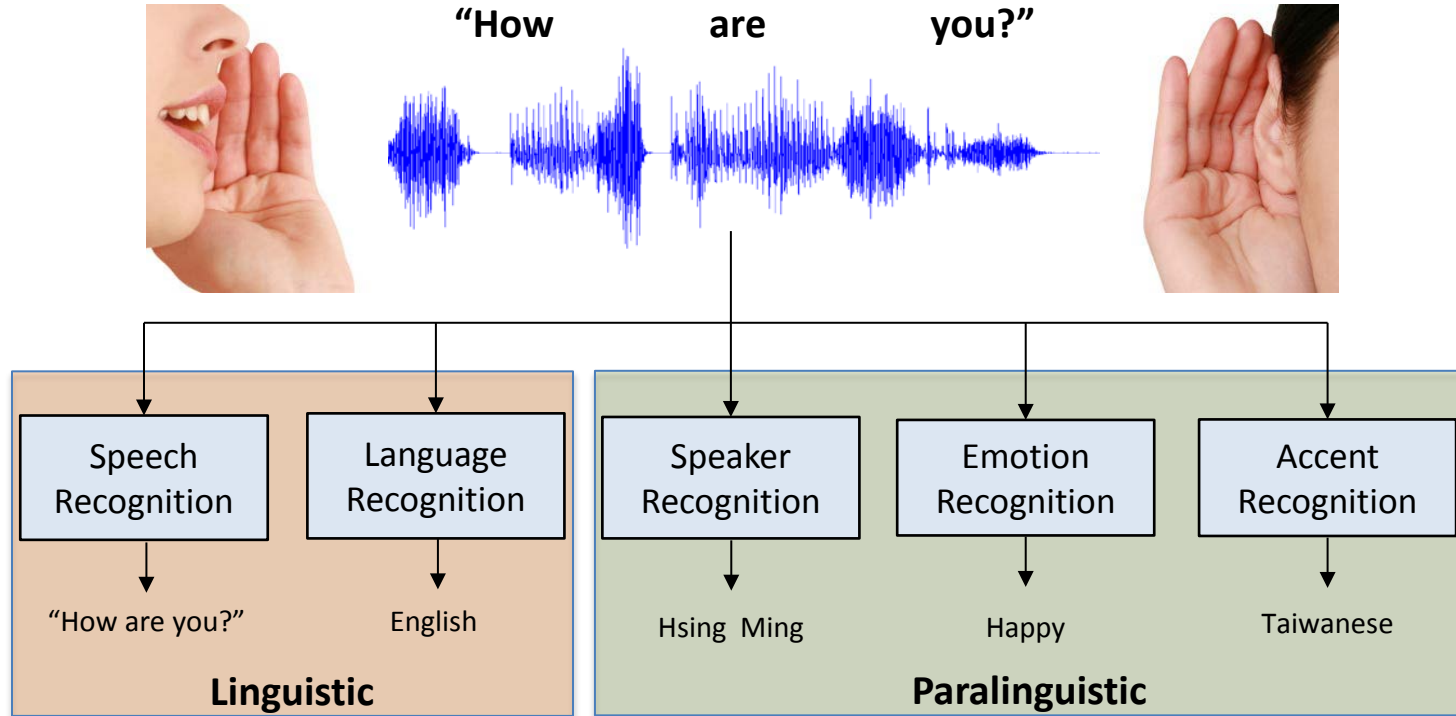
*Head of School of Electrical Engineering & Telecommunications,  
University of New South Wales, Australia*



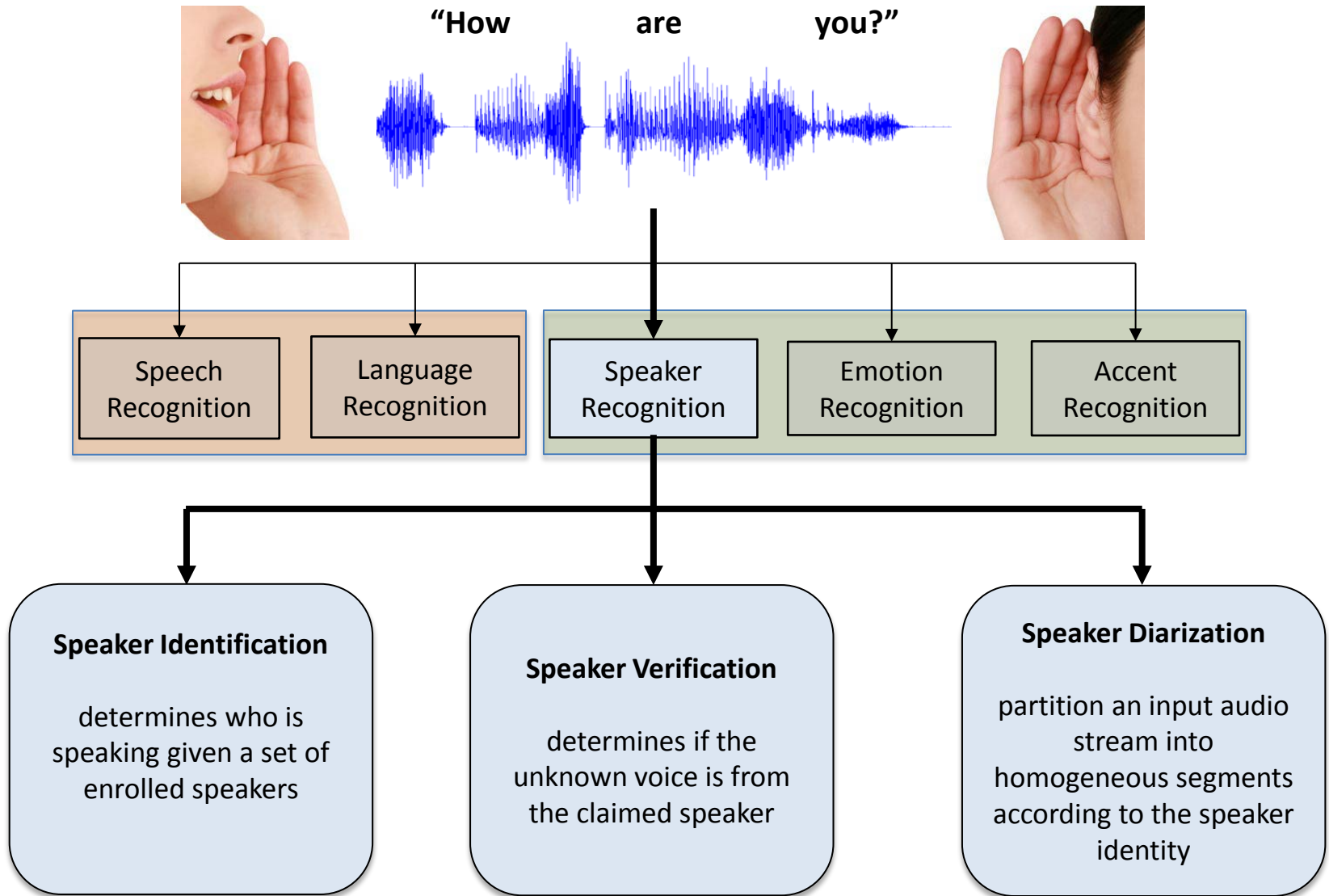
## Outline

- Introduction
- Speaker Verification Applications
- Speaker Verification System
- Performance measure
- NIST Speaker Recognition Evaluation (SRE)
- Discussion

## Introduction



- Speech conveys several types of information
  - Linguistic: message and language information
  - Paralinguistic : emotional and physiological characteristics



## Speaker Identification

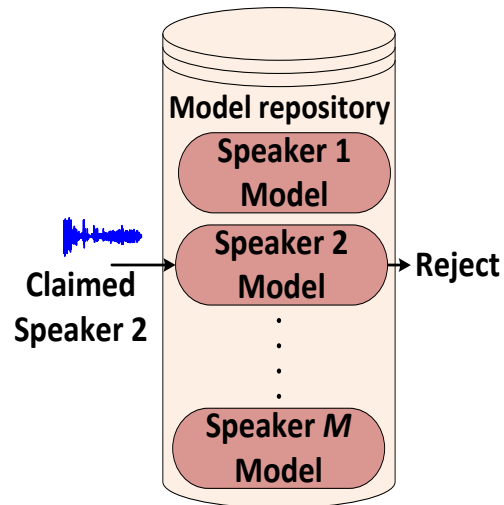
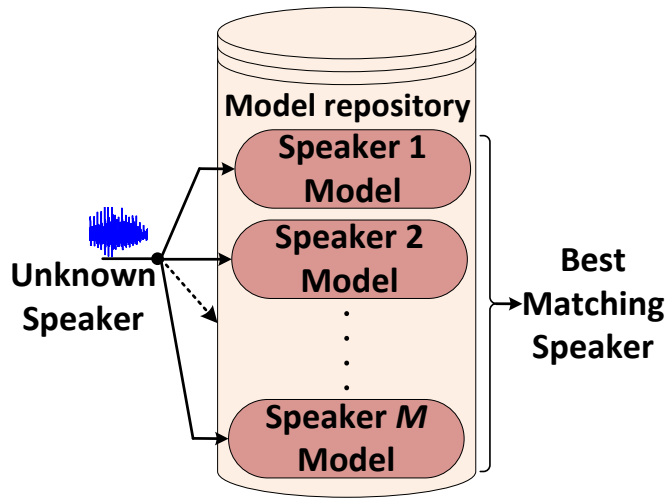
determines who is speaking given a set of enrolled speakers

## Speaker Verification

determines if the unknown voice is from the claimed speaker

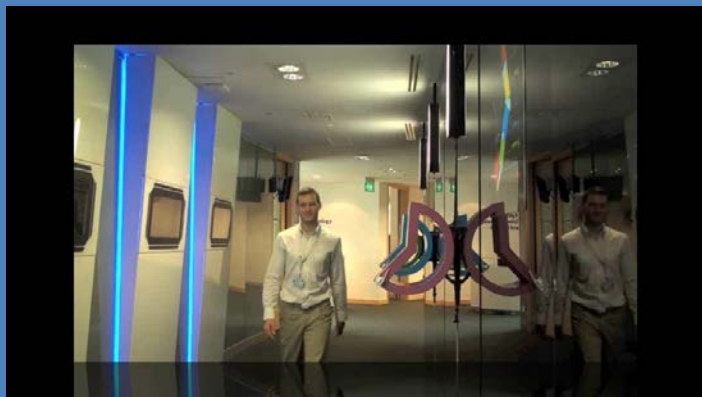
## Speaker Diarization

partition an input audio stream into homogeneous segments according to the speaker identity



## Speaker Verification Applications - Biometrics

Access control



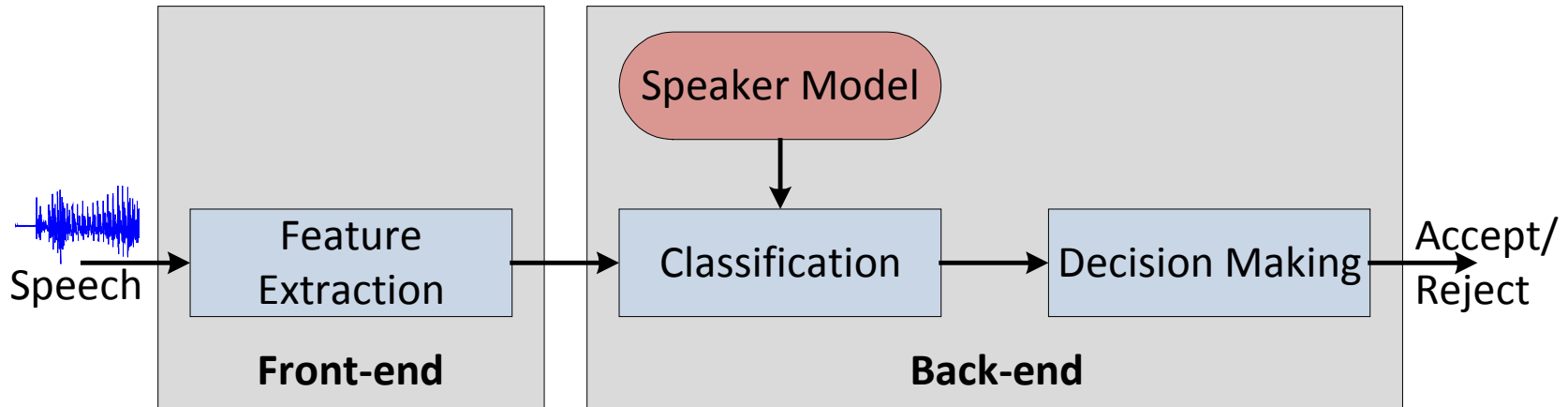
Physical facilities

Transaction authentication



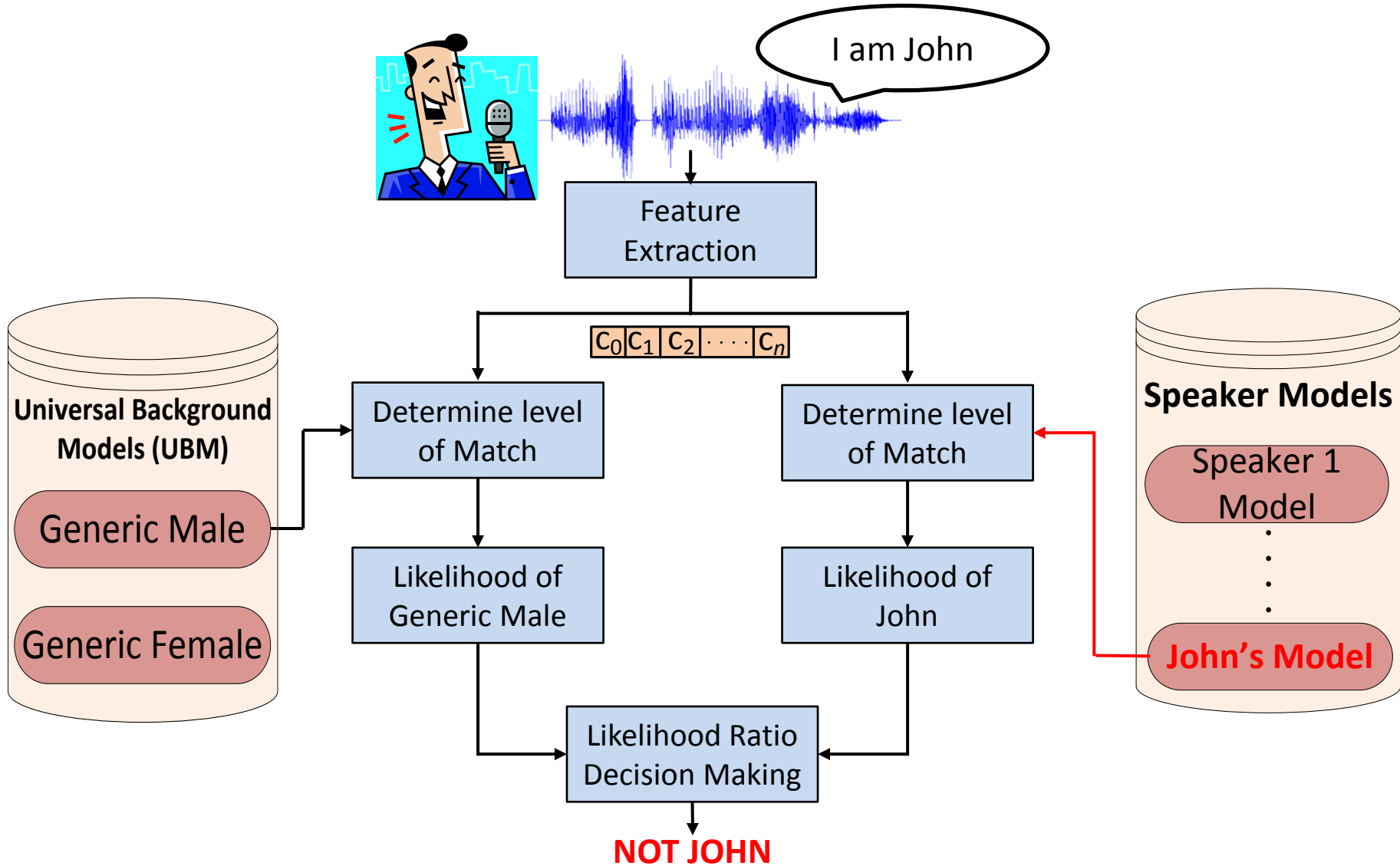
Telephone credit card purchases

## Speaker Verification System – Basic Overview



- In automatic speaker verification,
  - The front-end converts speech signal into a more convenient representation (typically a set of feature vectors)
  - The back-end compares this representation to a model of a speaker to determine how well they match

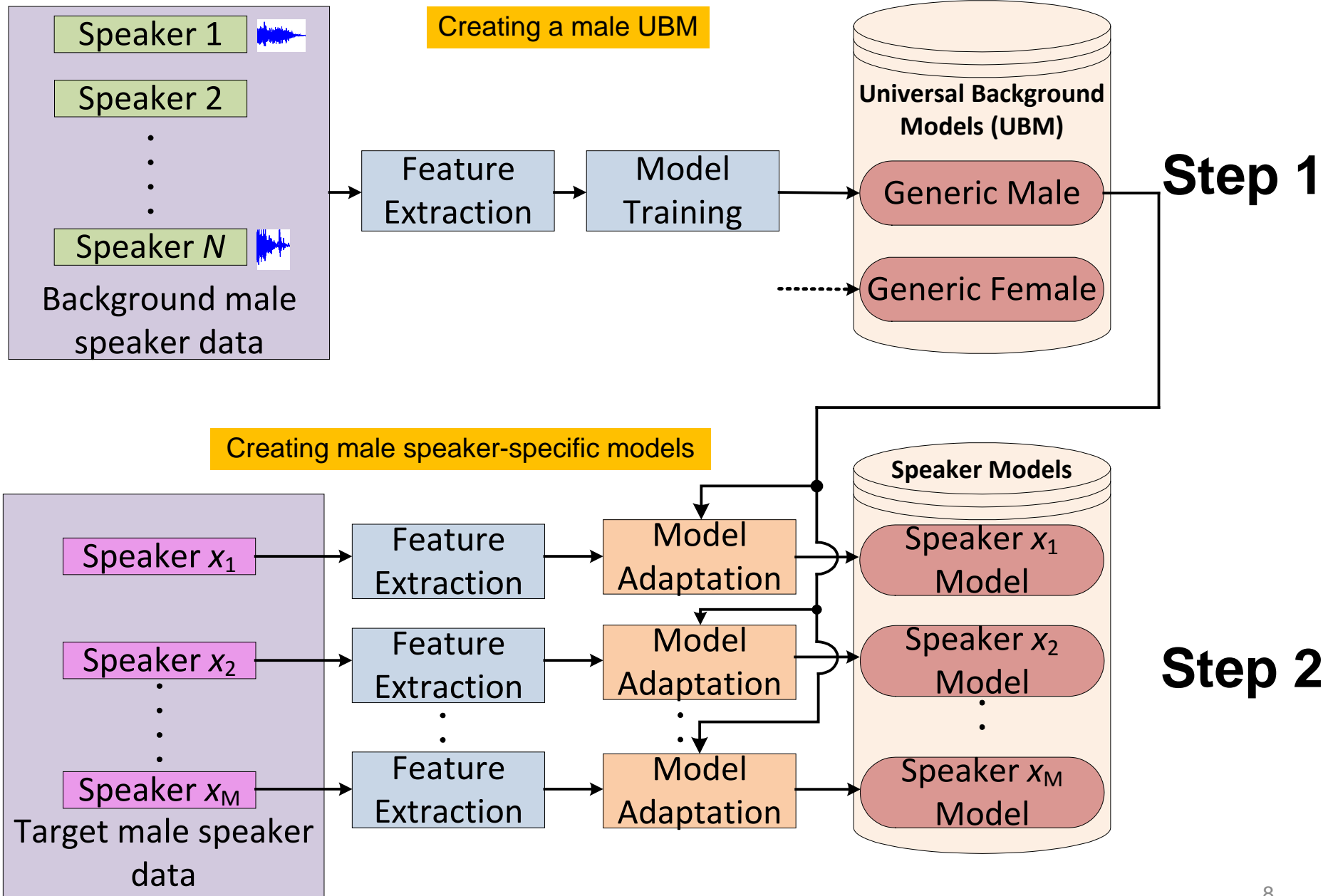
# Speaker Verification System



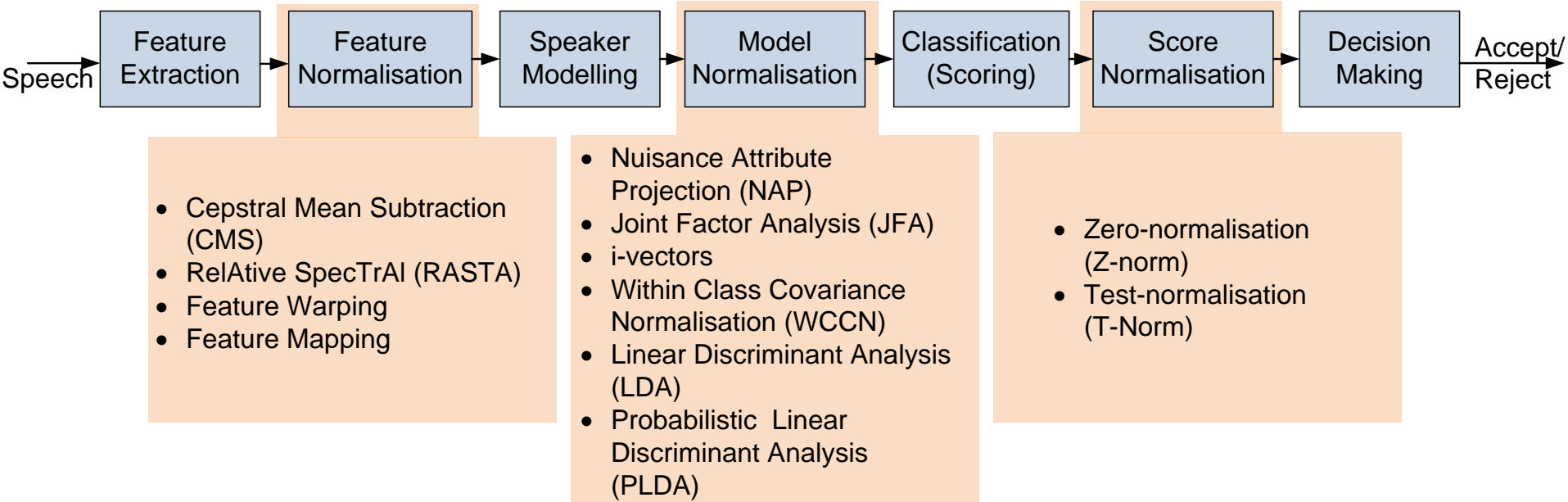
UBM: represent general, speaker independent model to be compared against a person-specific model when making an accept or reject decision.



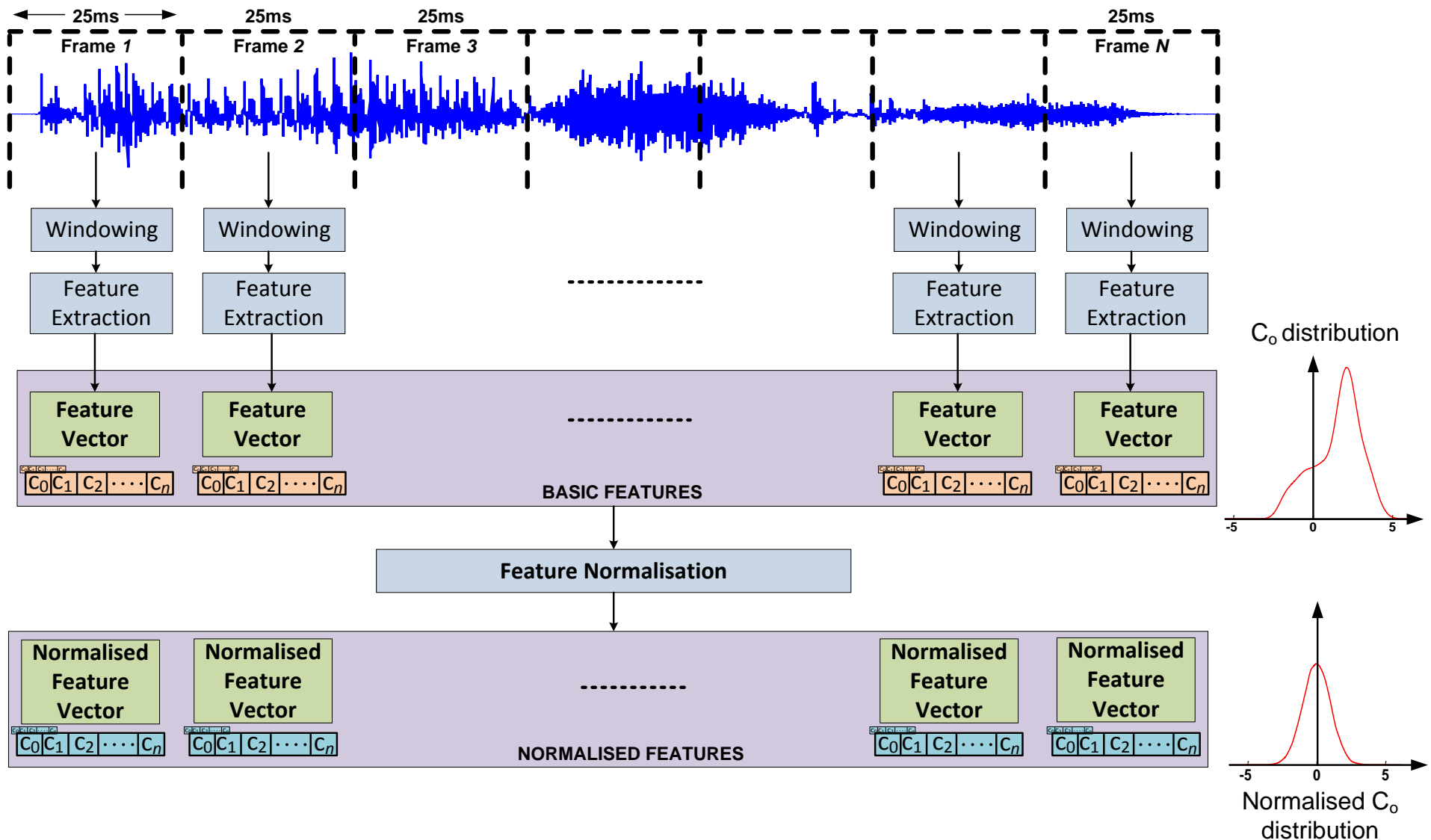
# Speaker Verification System – Speaker Enrolment

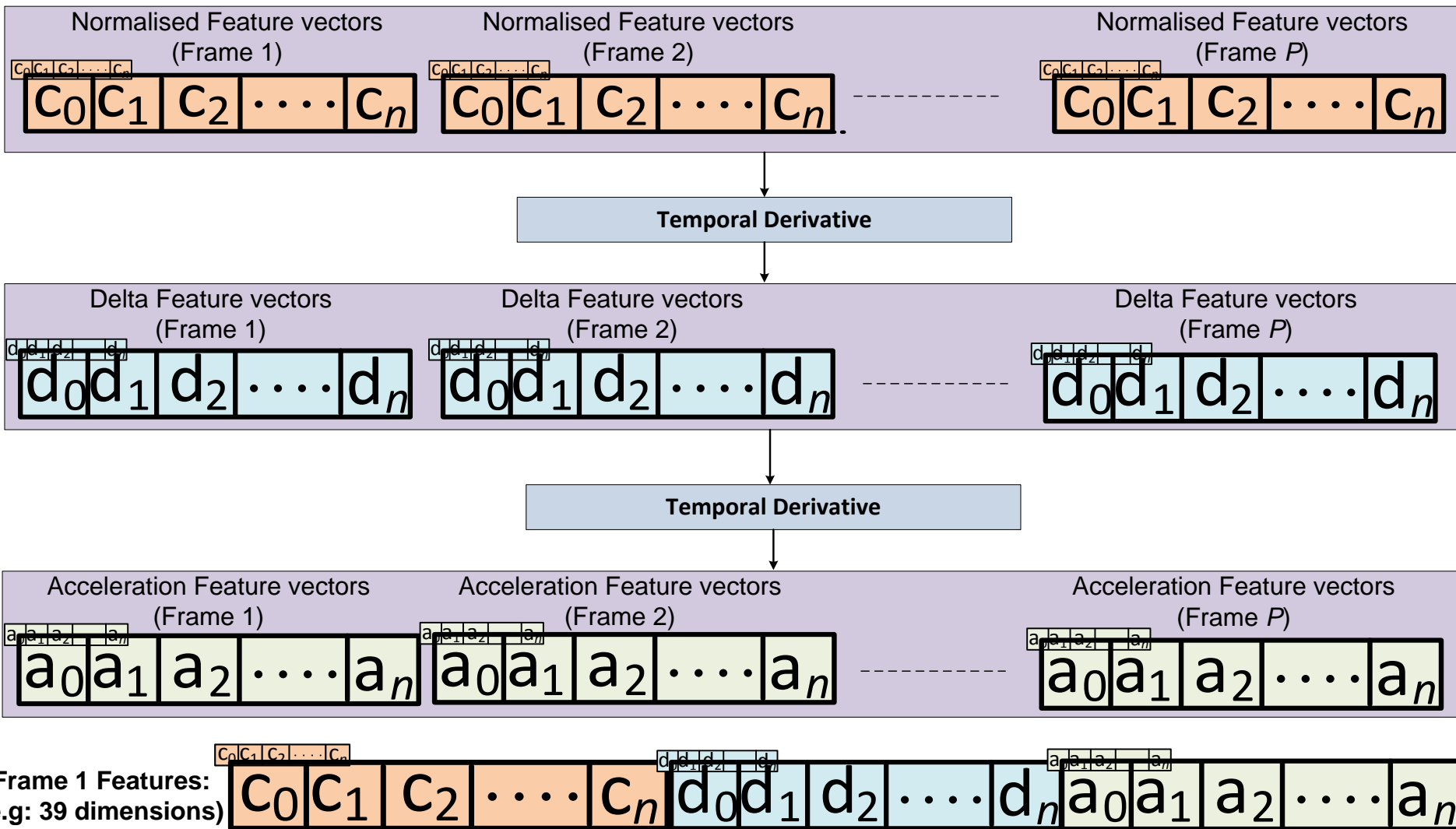


## Detailed Speaker Verification System

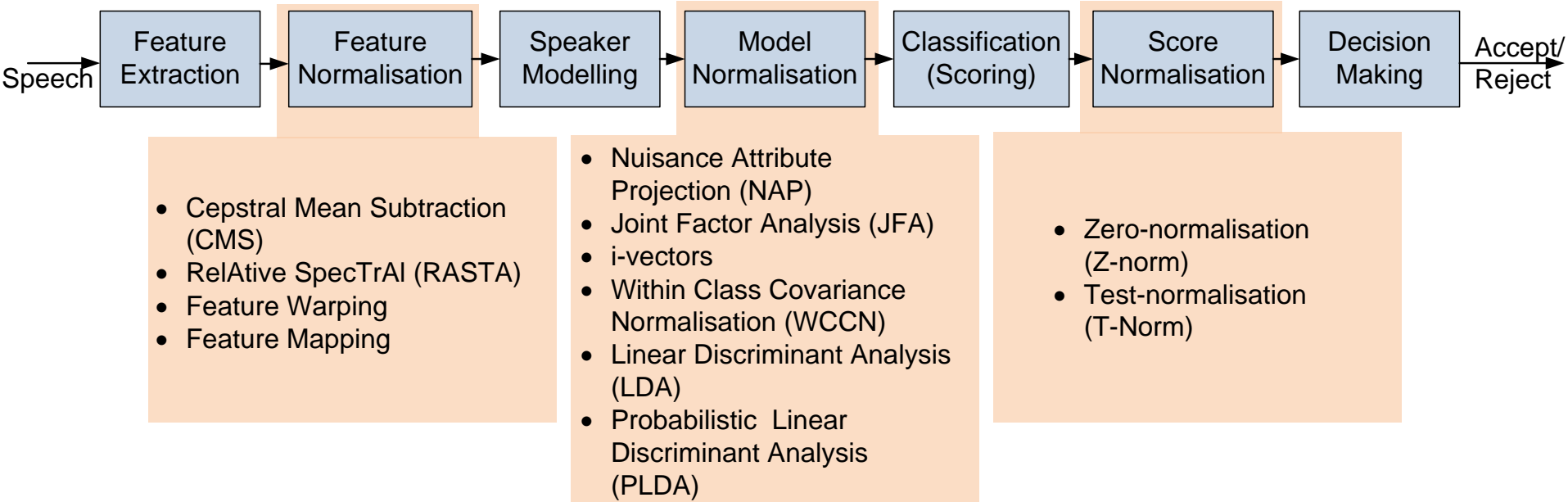


# Front-end: Feature Extraction

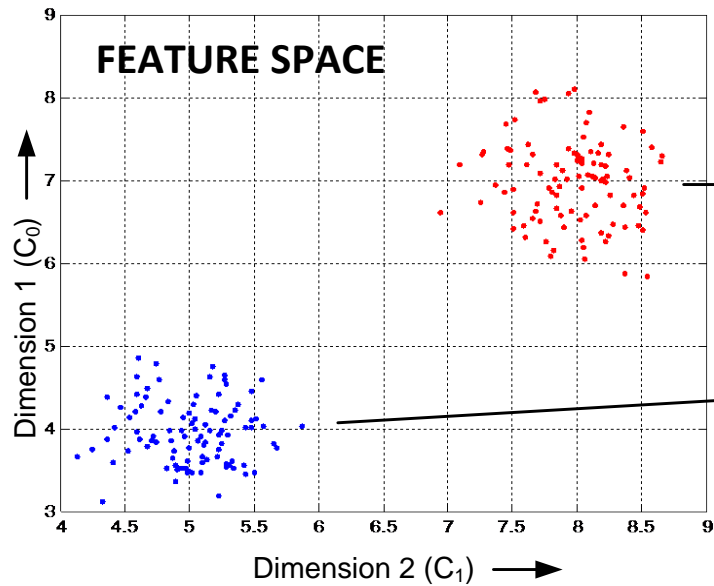




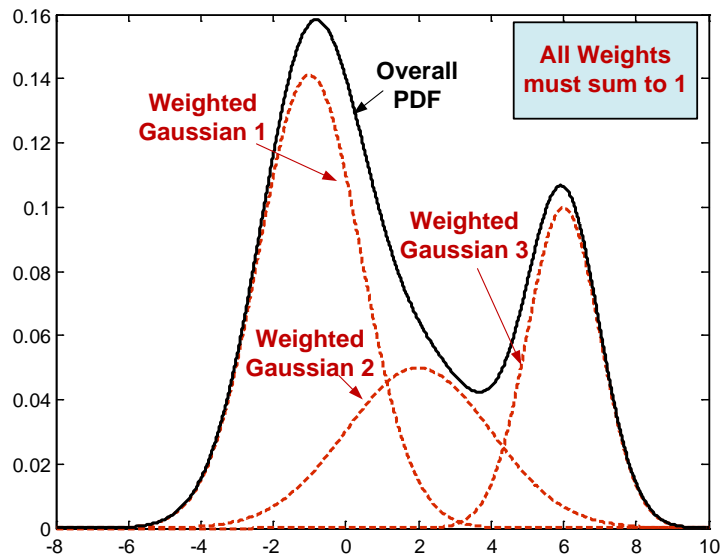
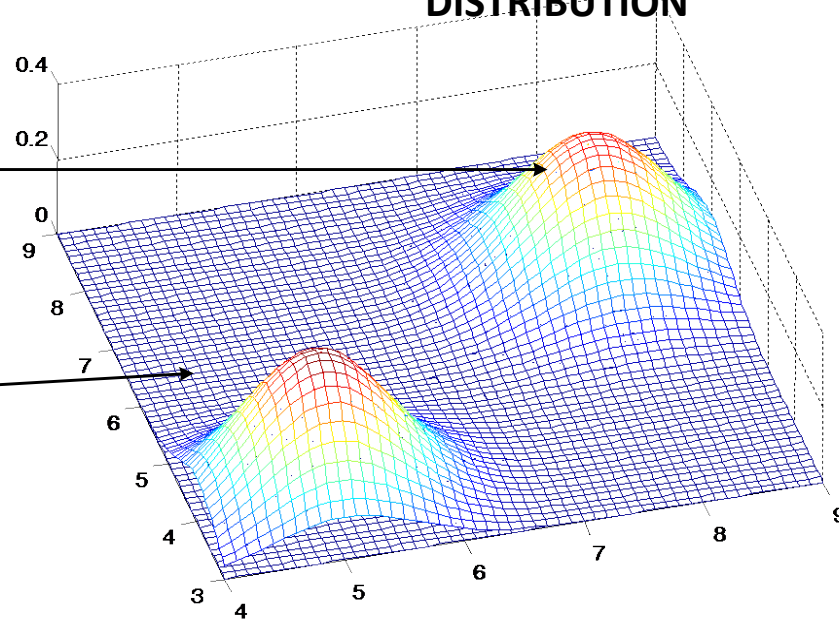
## Detailed Speaker Verification System



# Speaker Modelling



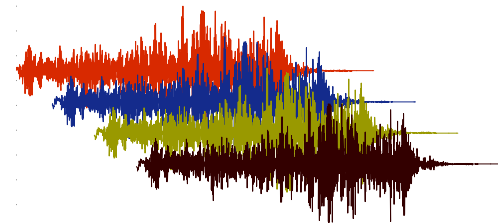
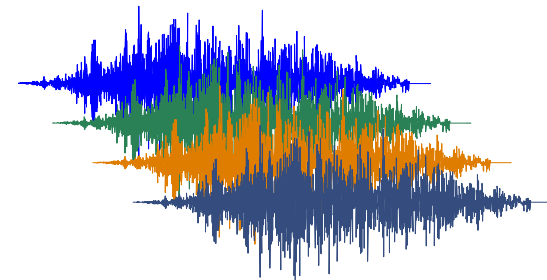
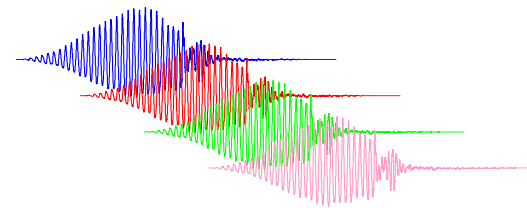
**MODELLING PROBABILITY DISTRIBUTION**

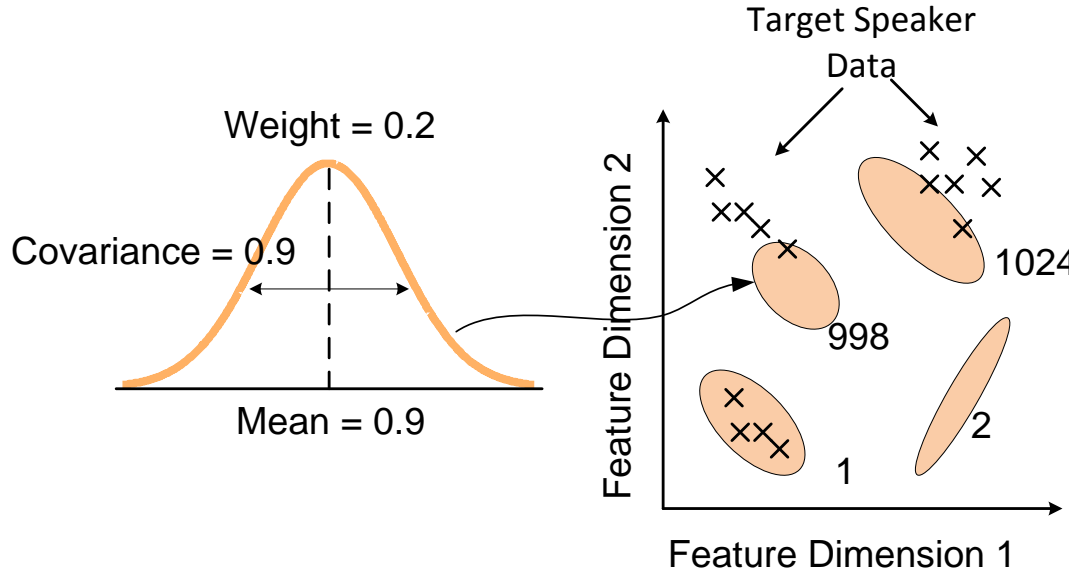


- ✓ Probability density function approximated by 3-component Gaussian mixture models
- ✓ Each Gaussian mixture consist of a mean ( $\mu$ ), covariance ( $\Sigma$ ) and weight ( $w$ )

## Database for creating UBM (example)

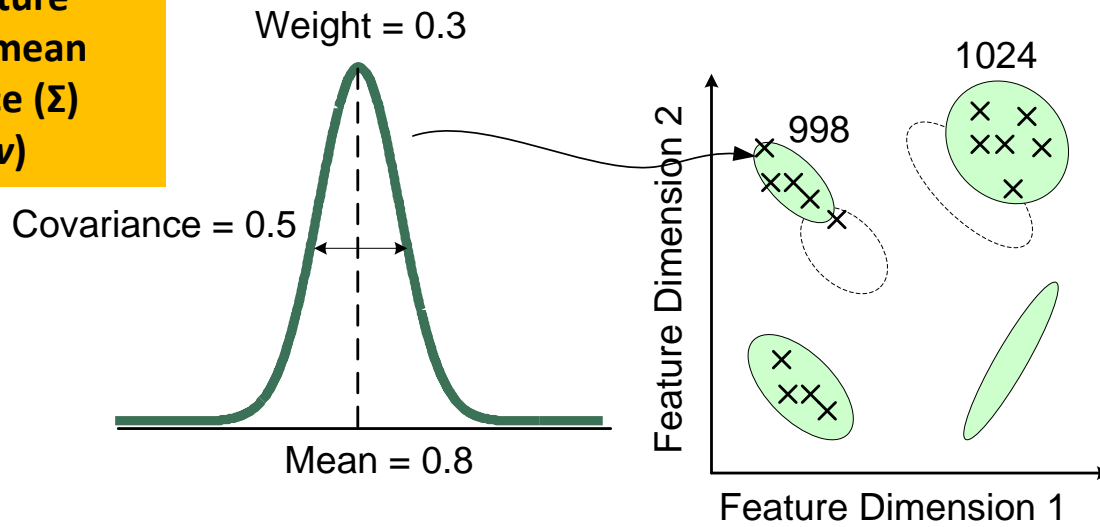
- Training set
  - 56 male speakers (each speaker consists of 2 minutes of active speech) for creating the UBM
- Target set
  - 20 male speakers (each speaker consists of 2 minutes of active speech) for speaker-specific model
- Test set
  - 250 male utterances (each speaker has many test utterances) with the known identity





Universal Background Model (UBM) consists of 1024 Gaussian mixtures

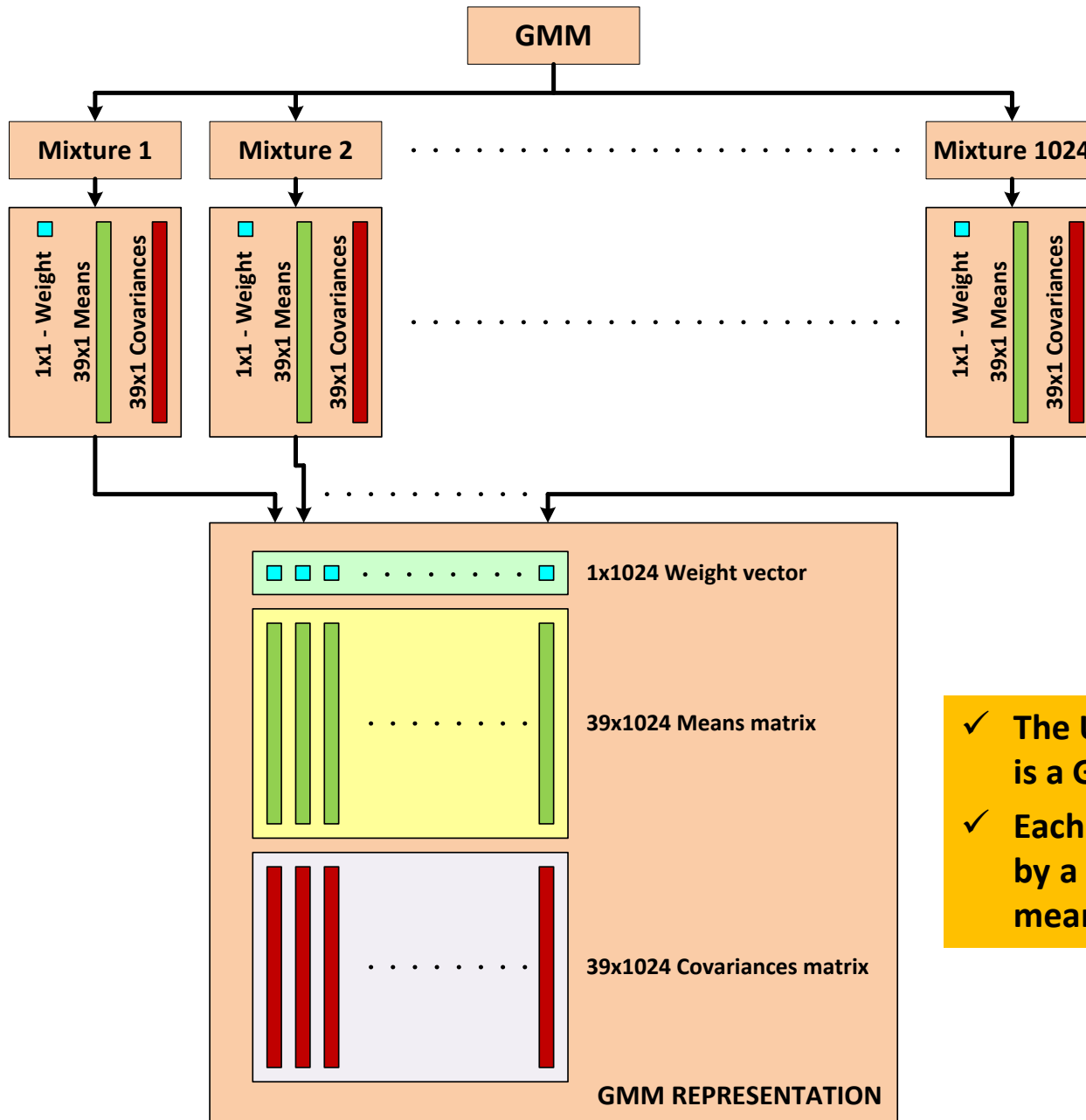
✓ Gaussian mixture consists of a mean ( $\mu$ ), covariance ( $\Sigma$ ) and weight ( $w$ )



Target speaker model consists of 1024 Gaussian mixtures

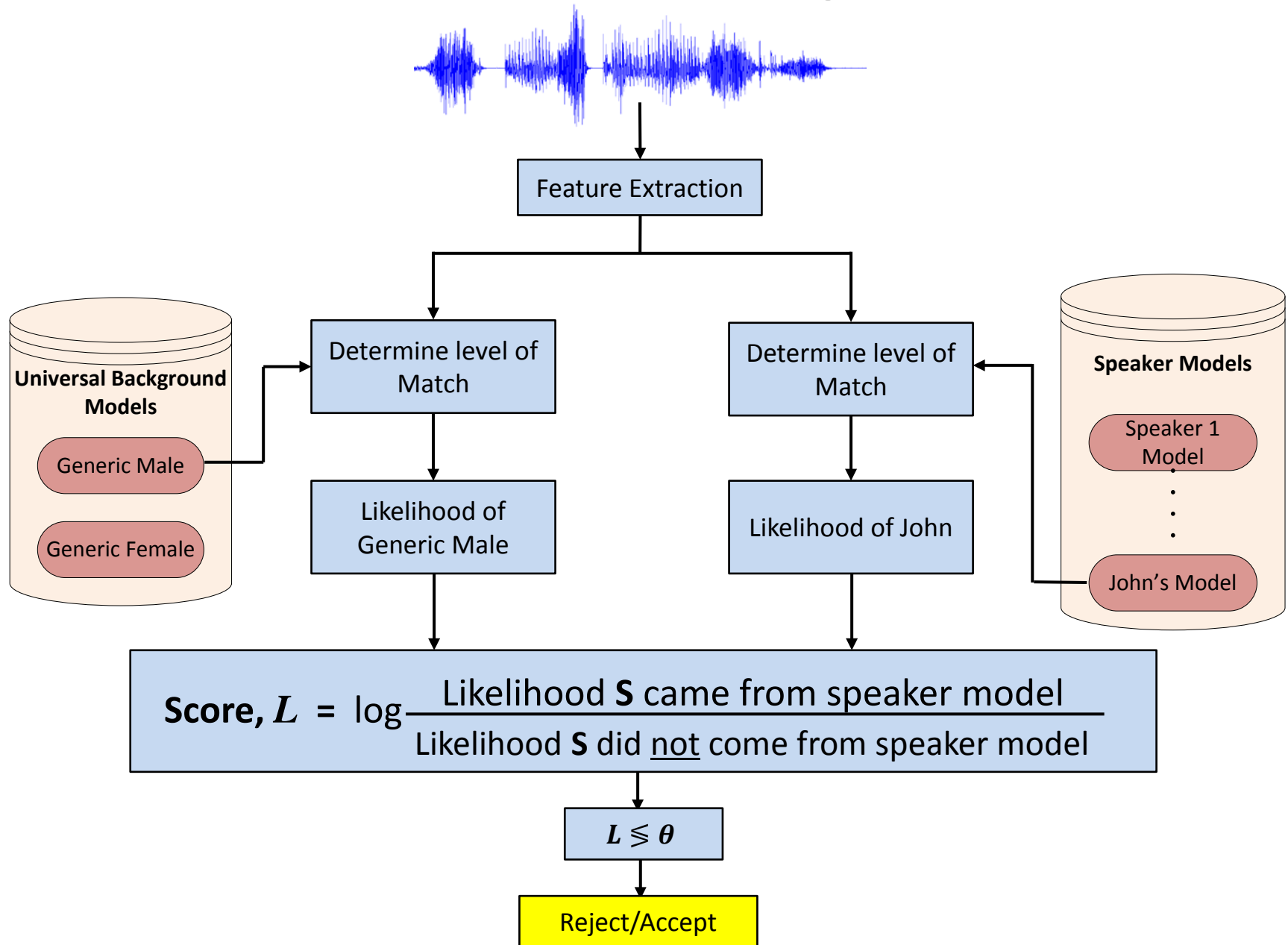


# Representing GMMs

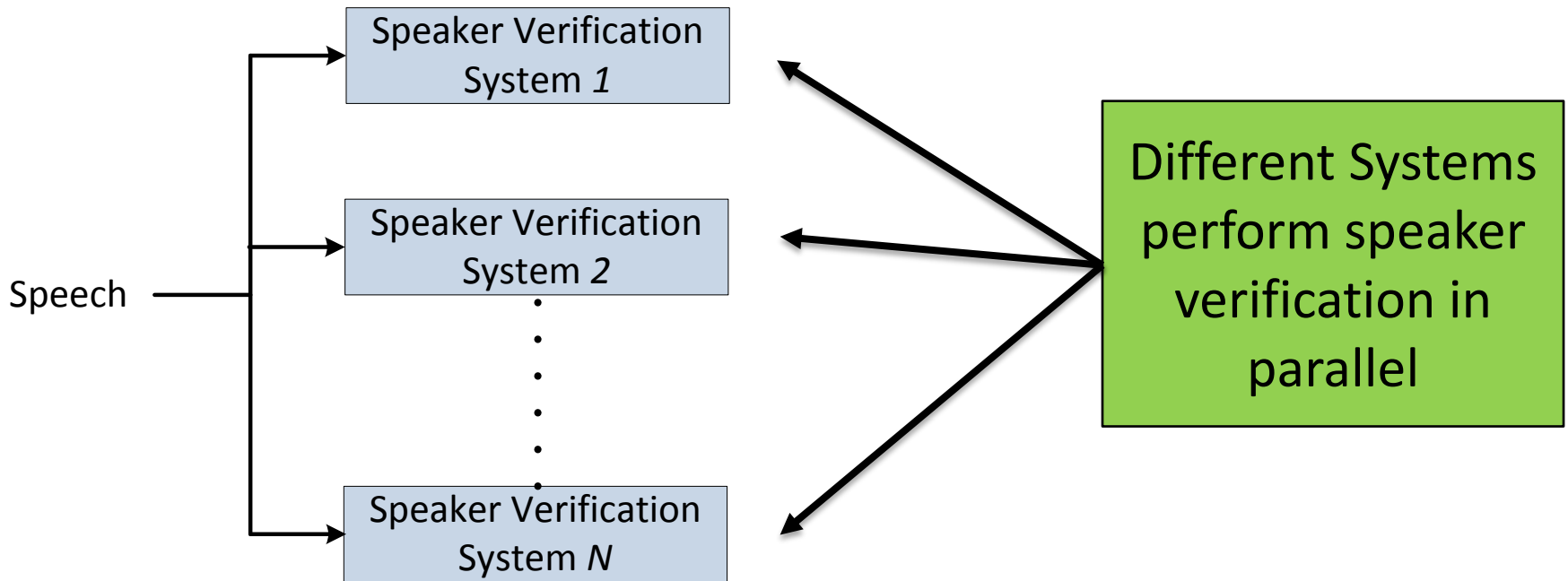


- ✓ The UBM and each speaker model is a GMM
- ✓ Each of them will be represented by a vector of weights, a matrix of means and a matrix of covariances

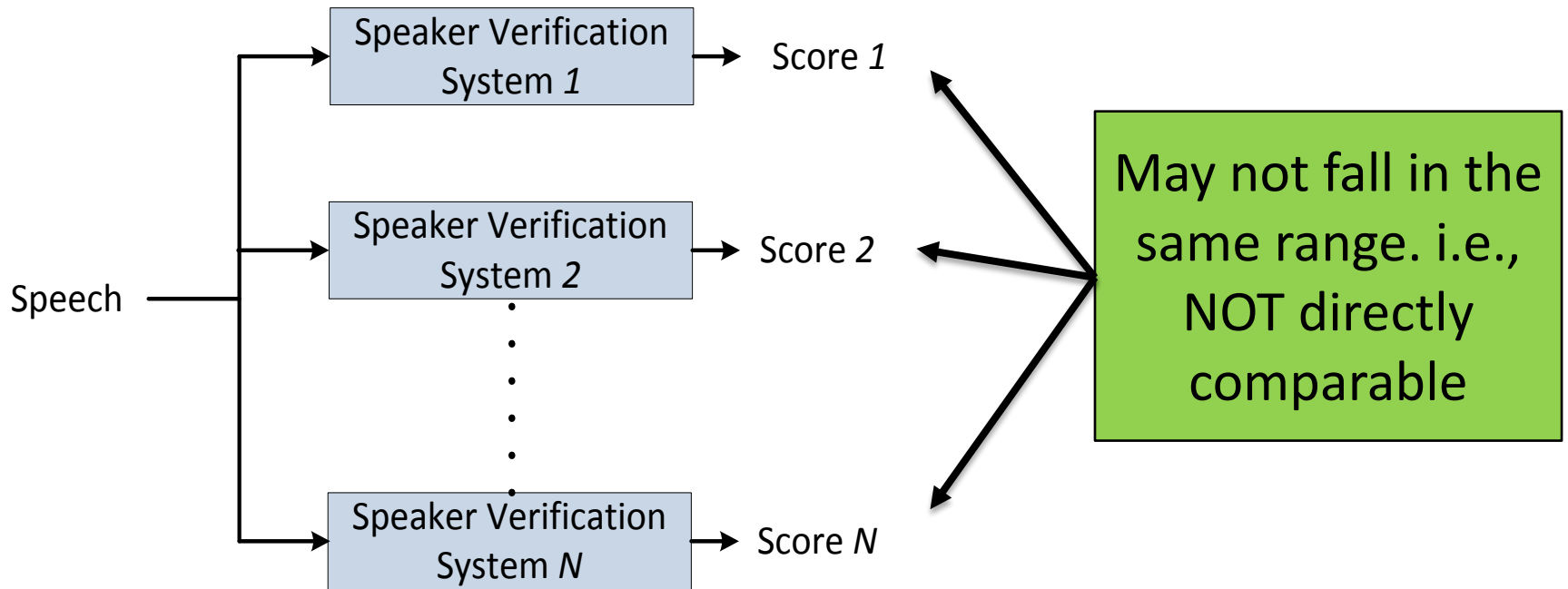
# Decision Making



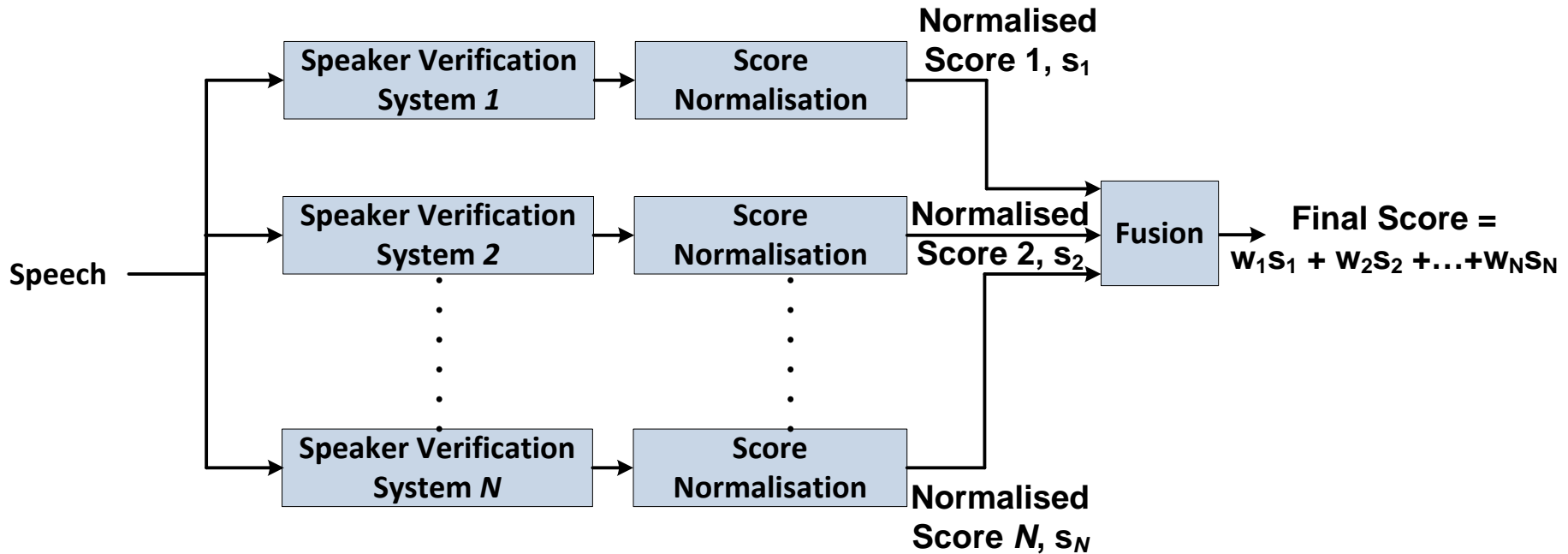
## Score Normalisation



## Score Normalisation



## Fusion



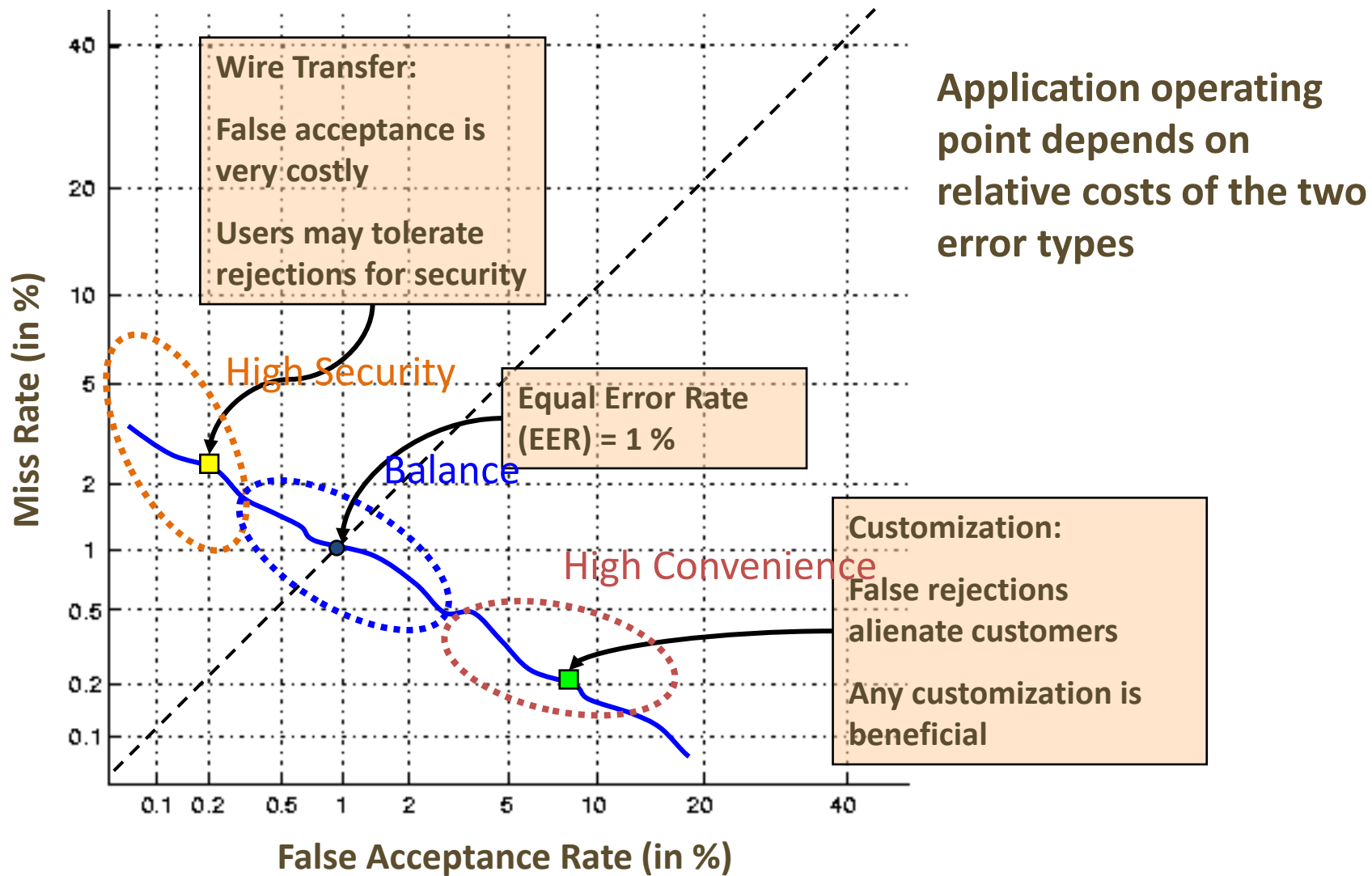
**Final score will be a weighted sum of score from each system**

# Performance measure

- Types of error:
  - Misses: valid identity is rejected
    - Probability of miss: ratio of the number of falsely rejected speaker tests to the total number of correct speaker trials.
  - False alarms: invalid identity is accepted
    - Probability of false alarm: ratio of the number of falsely accepted speaker tests to the total number of impostor trials

<b>TRUE SPEAKER</b>	<b>CORRECT DECISION</b>	<b>MISS</b>
<b>IMPOSTER</b>	<b>FALSE ACCEPTANCE</b>	<b>CORRECT DECISION</b>
	<b>ACCEPT CLAIM</b>	<b>REJECT CLAIM</b>

# Performance measure - Detection error trade-off (DET) curve





# APSIPA

Asia-Pacific Signal and Information Processing Association

