

# **USER VOICE IDENTIFICATION (URVIN)**

Aladdin Ariyaeinia, Reza Sotudeh, Chris Bailey and Fredrik Rodin

## **AMADEUS VIRTUAL RESEARCH CENTRE**

# USER VOICE IDENTIFICATION (URVIN)

## OVERVIEW

- AMADEUS Centre (Architectures Machines And Devices for Efficient Ubiquitous Systems)
- Speaker Recognition
- Challenges in Open-Set Speaker Identification
- Experiments
- Future

## AMADEUS Partners

**CYBULA**  
high performance pattern recognition systems

**dti**

**fulcrum** www.fulcrumvoicetech.com  
**voice technologies**

**alphamosaic**

**Infineon**  
technologies

**mpe**

**LEXICLE™**

**vita nuova®**

**Vulcanic**

**THE UNIVERSITY of York**

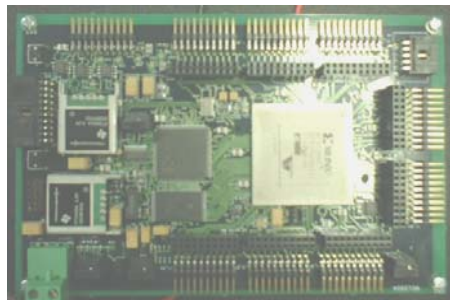
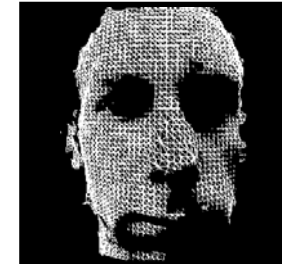
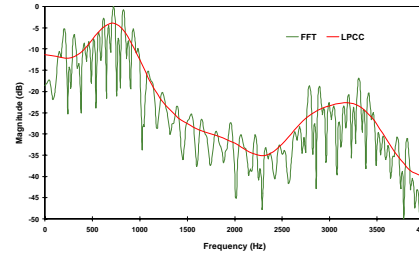
**UH** University of Hertfordshire

## AMADEUS Projects

- **UDA – Ubiquitous Digital Assistant (Software)**
- **AICP – Ambient Intelligence Co-Processor**
- **VIDEOWARE – Video Feature Extraction**
- **URVIN – Audio Feature Extraction**
- **FAR – Facial Attribute Recognition**
- **DEMOS – Digitally Enhanced Micro Operating Systems**
- *Two Additional Projects in the pipeline*

# AMADEUS Objectives

Development of Technology  
Concepts in Hardware



USE FPGA Platform V1

Prototyping Platforms & Demonstrators

Highly Integrated Systems



# URVIN

## User Voice Identification

### MOTIVATION

Requirement in ubiquitous computing applications:  
*user identification for the purpose authorisation*

Proposed approach:  
*speaker recognition (voice biometrics)*

**Aim:**

*to investigate an effective hardware system capable of extracting highly reliable speech features in the presence of background noise.*

# Speaker Recognition

## Speaker Identification:

To determine the correct speaker from a given population

The test utterance is compared with the reference models of the registered population.

## Speaker Verification:

To determine if a speaker is who he or she claims to be

The test utterance is compared only with the reference model for the claimed identity.

# Speaker Identification

**Text-Independent**

Users are not constrained to provide utterances of specific texts during recognition trials

**Text-Dependent**

Users must provide utterances of the same linguistic content for both training and recognition trials

# Speaker Identification

Training:

Training Utterances

Feature  
Extraction

Reference Model Generation

Verified Identity

Testing:

Test Utterance

Feature  
Extraction

Pattern Matching

Decision

User Identity

Reference Models

# Open-Set Speaker Identification

## Mathematical description:

$N$  speakers enrolled in the system

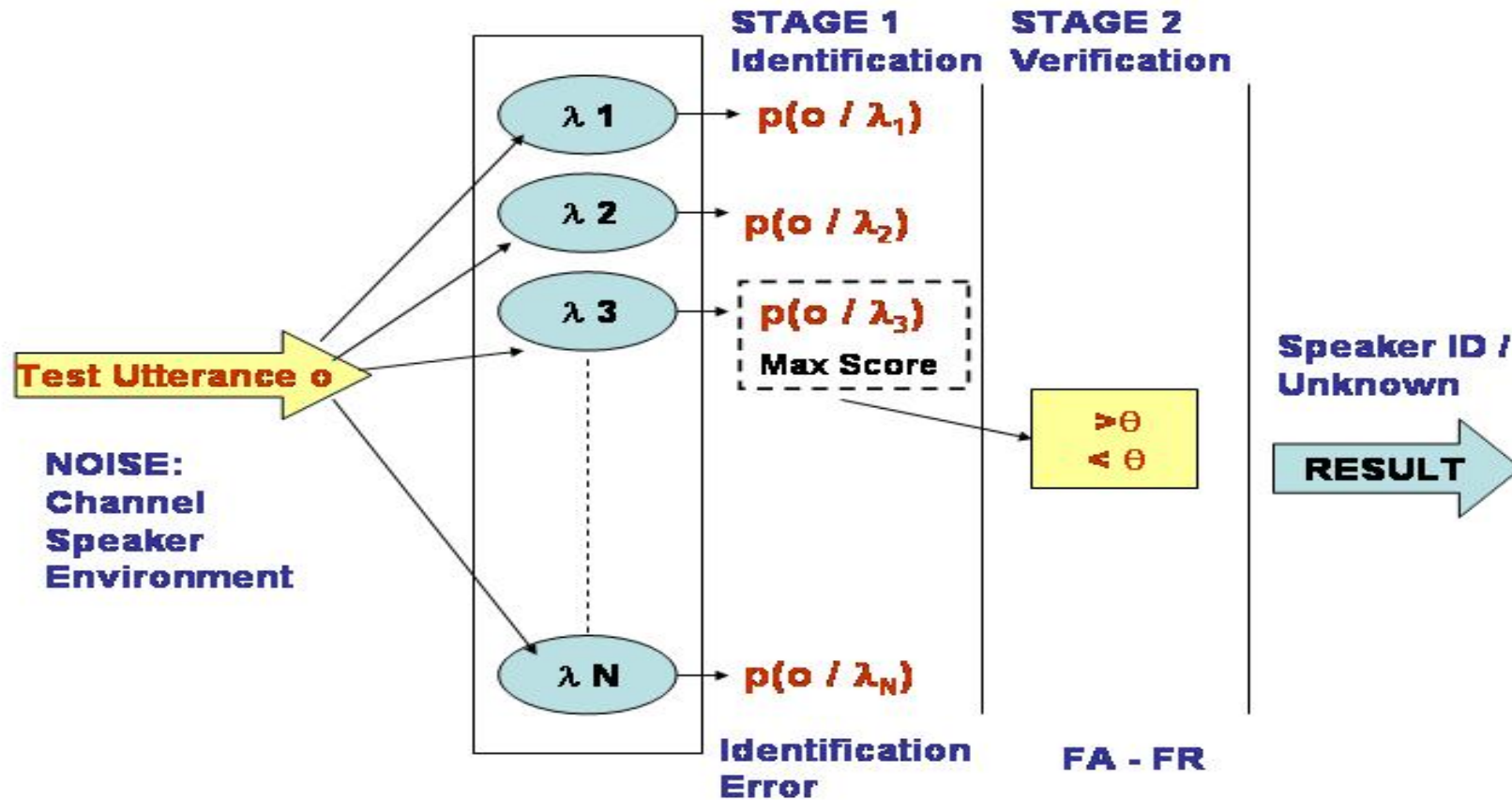
their statistical model descriptions:  $\lambda_1, \lambda_2, \dots, \lambda_N$ .

If  $\mathbf{O}$  denotes the feature vector sequence extracted from the test utterance, then

$$\max_{1 \leq n \leq N} \{p(\mathbf{O} | \lambda_n)\} \geq \theta \rightarrow \mathbf{O} \in \begin{cases} \lambda_i, i = \arg \max_{1 \leq n \leq N} \{p(\mathbf{O} | \lambda_n)\} \\ \text{unknown speaker model} \end{cases}$$

# Challenges in Speaker Identification

## OPEN SET IDENTIFICATION





## Open-Set Speaker Identification: a twofold problem

1. To identify the speaker model in the set, that best matches a given test utterance.
2. To determine whether the test utterance has actually been produced by the speaker associated with the best-matched model, or by some unknown speaker outside the registered set.

## Possible Errors

- **O**, which belongs to  $\lambda_m$ , not yielding the maximum likelihood for  $\lambda_m$ .

OSIE: First Stage

- Assigning **O** to one of the speaker models in the system when it does not belong to any of them.

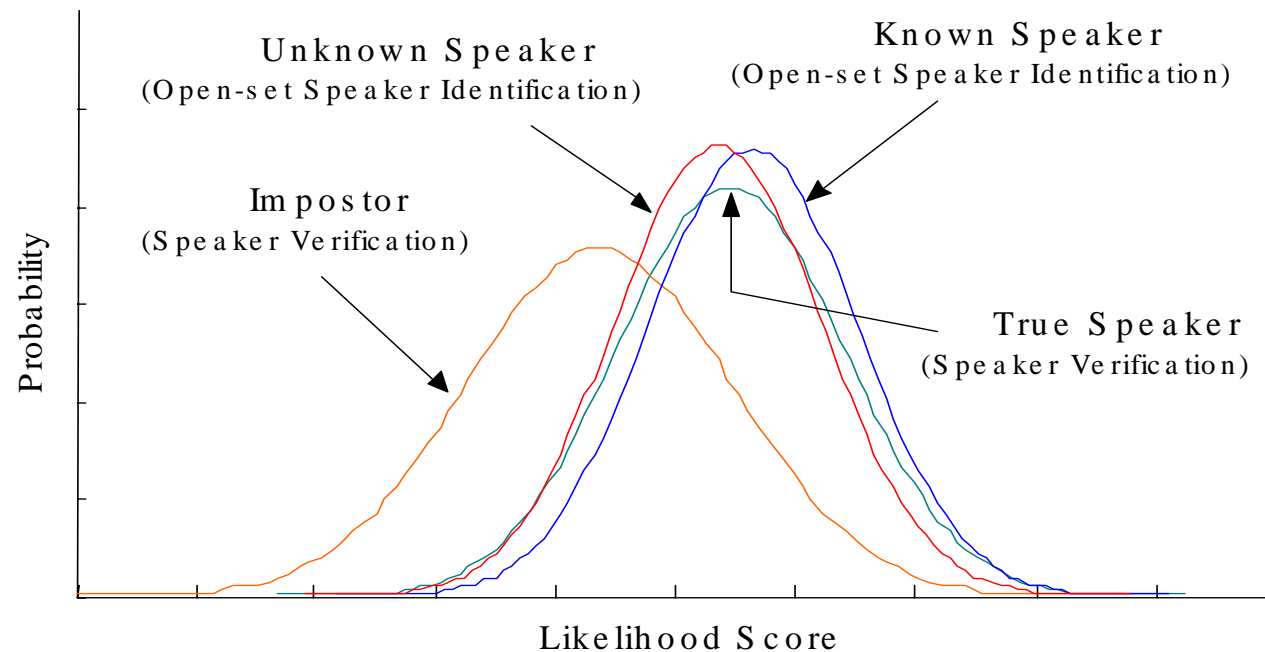
OSI-FA: Second Stage

- Declaring **O** which belongs to  $\lambda_m$ , and yields the maximum likelihood for it, as originated from an unknown speaker.

OSI-FR: Second Stage

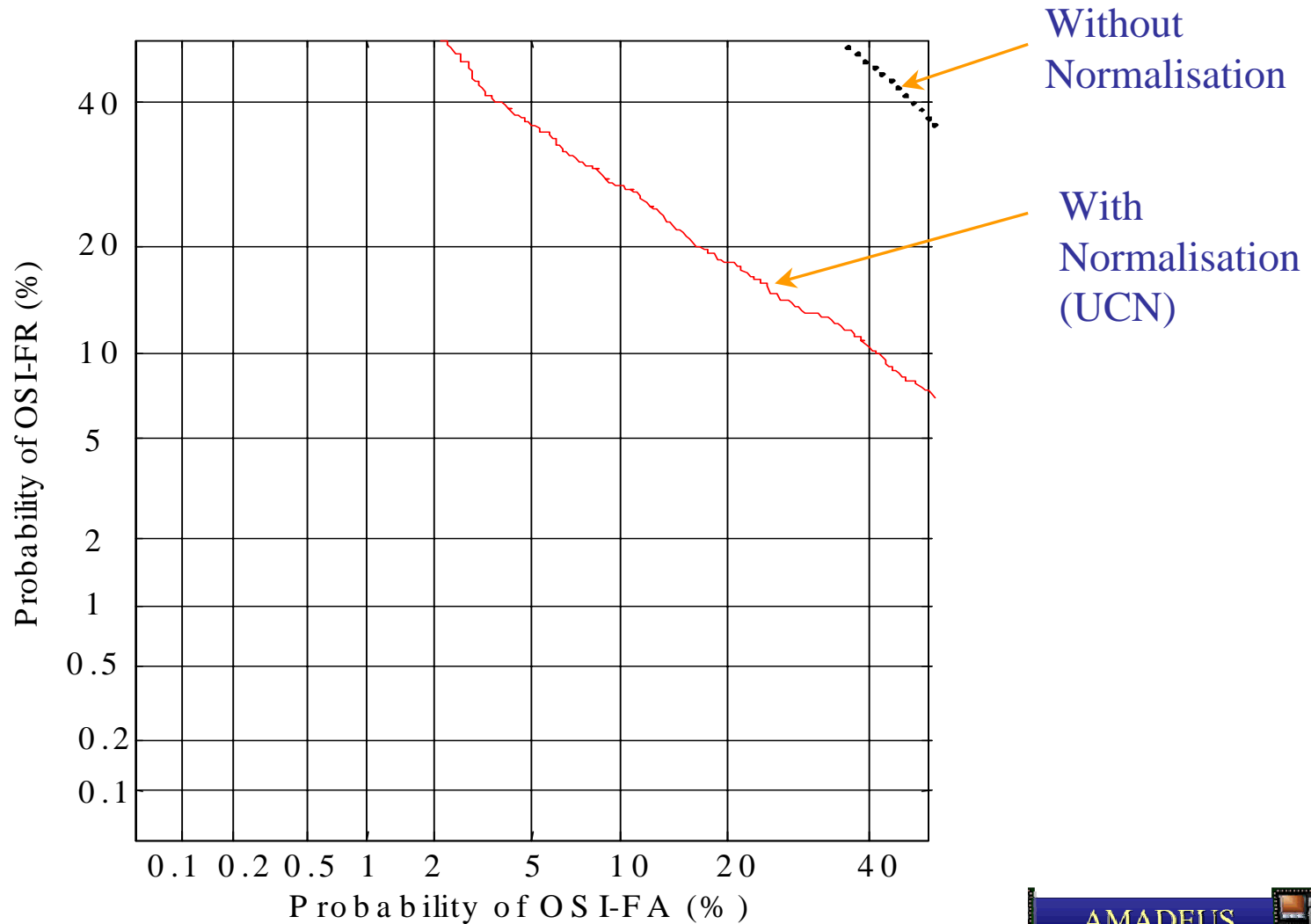
## Second Stage of OSTI-SI & Speaker Verification

The challenge in open-set identification can be viewed as a special (but unlikely) scenario in speaker verification in which each impostor targets the speaker model in the system for which he/she can achieve the highest score.



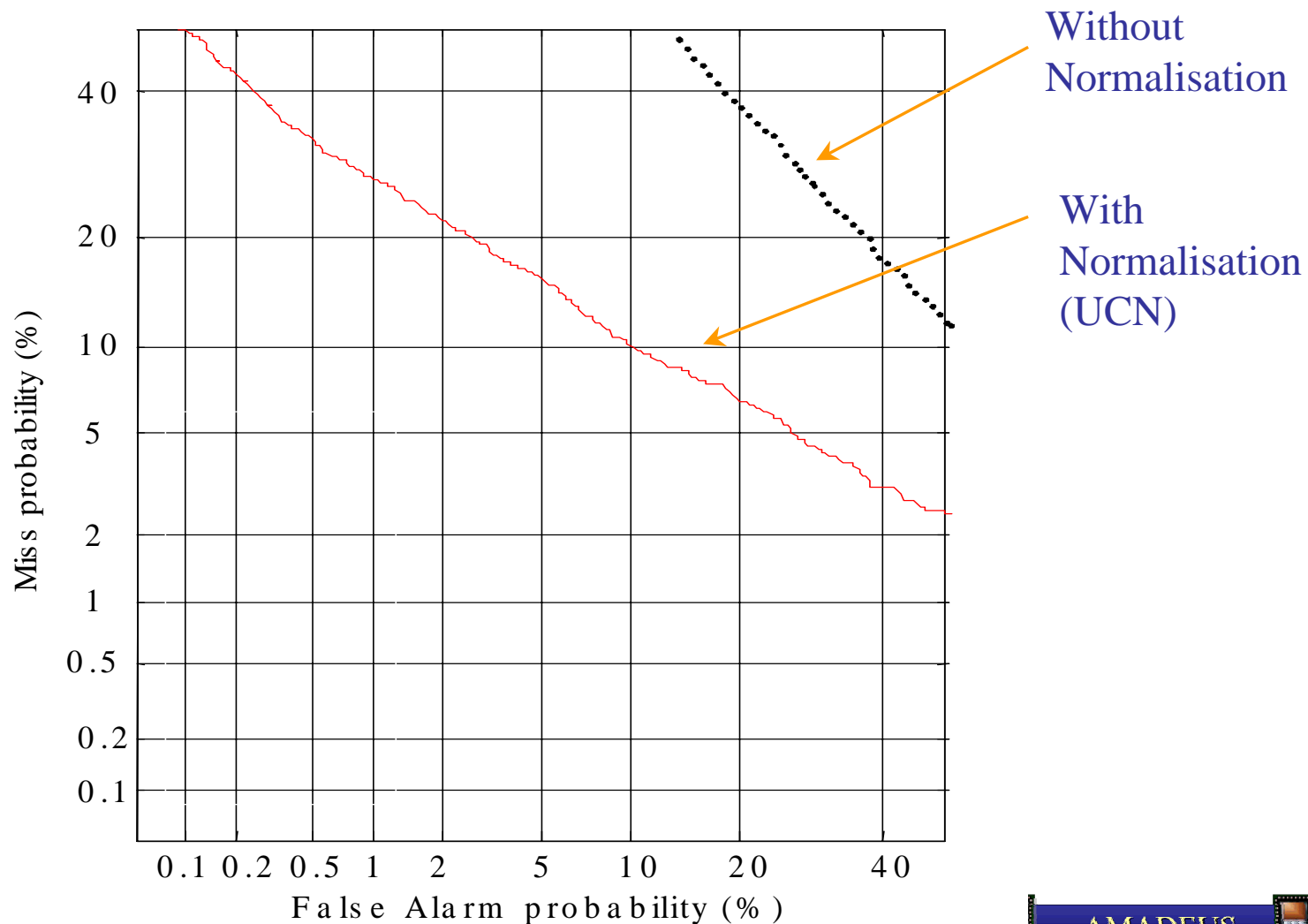
# Experimental Comparison

Results for the second stage of OSI in terms of DET curves.



# Experimental Comparison

Results for Speaker Verification in terms of DET curves.



## Current and Future Work

- Development of algorithms for the extraction of reliable speech features.
  - effective methods for speech enhancement, and
  - the feature extraction techniques which help maximise discrimination amongst speakers.
- Development of a feature extraction engine
  - Generic IP core
  - Facilitating integration with other SOC components into complete systems