



INSTITUTE OF INFORMATICS  
SLOVAK ACADEMY OF SCIENCES

# Personnel authorization via voice biometrics in ATM and airport operations

Marián Trnka, Sakhia Darjaa, Milan Rusko (II SAS)

December 16, 2020



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement number 832969.



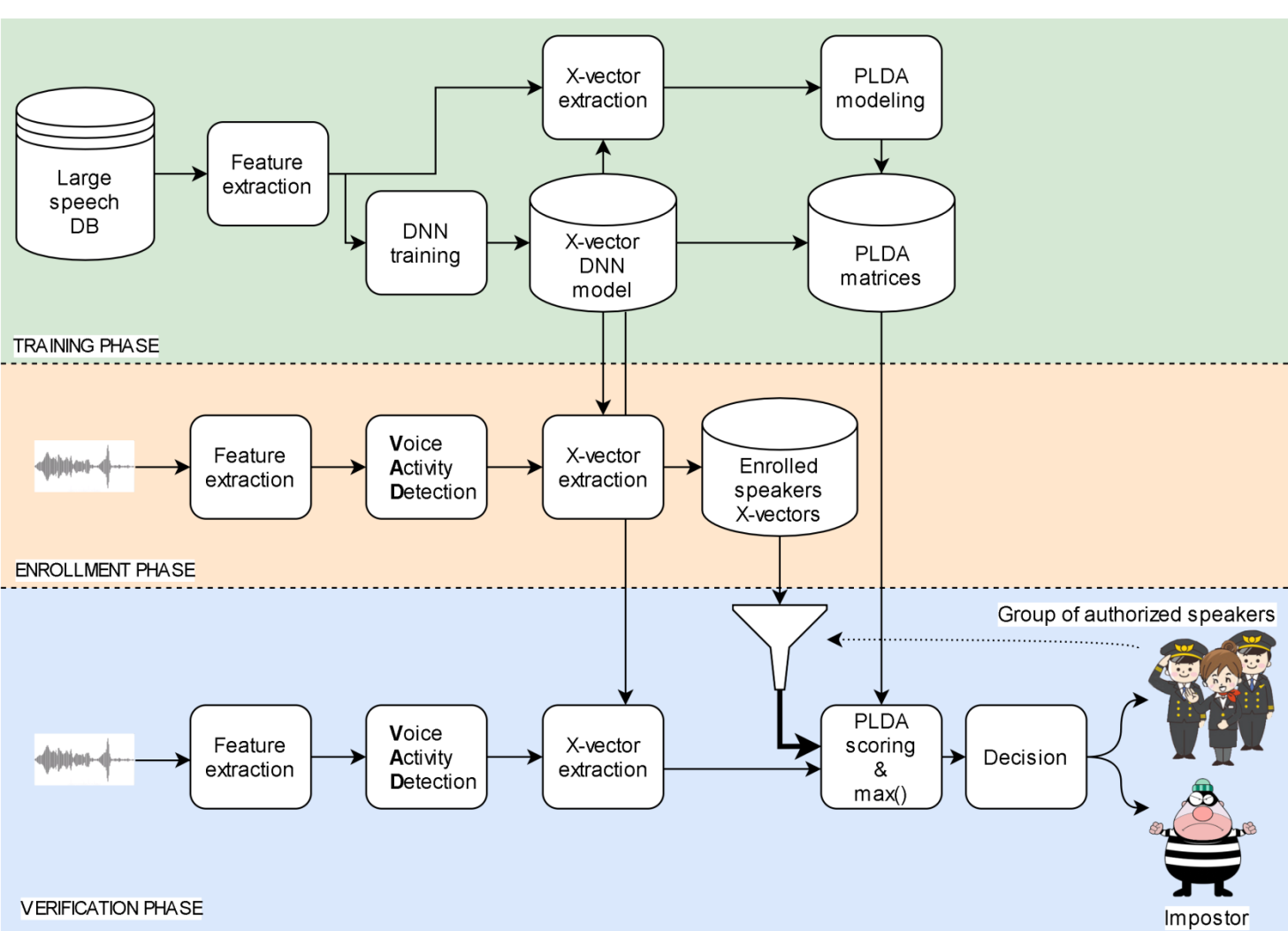
1. Introduction
2. Speaker Verification
3. SV Module in the SATIE project
4. Tests and results

# Introduction

- Voice communication is widely used in ATM – Radio, Telephone, VoIP
- Insecure analogue radio communication is still in use
- Voice biometrics – for securing radio communication in ATM (GAMMA project 2013-2017)
- Similar concept is applied to secure communication on airports (SATIE project 2019-)
- This presentation will be about the development and testing of speaker authorization module

# SATIE Speaker Authorization task

- Real-time application
- Short utterances – typically 2 – 5 seconds
- VoIP speech in Airport simulator (*VHF - Radio channel + Noise*)
- text-independent
- ROLE - Binary decision - membership to group of the authorized persons (whitelist) or not



Training databases

**VoxCeleb1 (2017)**

spks – 1250, utts – 150k

**VoxCeleb2 (2019)**

spks – 6000, utts 1.1M

Data augmentation  
reverberation (RIR)  
noise (MUSAN)

SF: 16kHz

Mfcc dim 30

X-vector DNN embed (400)  
Snyder, et al. "X-vectors: Robust  
DNN embeddings for speaker  
recognition." ICASSP (2018).

Kaldi environment

Group verification

# Speaker verification module testing

- The basic (binary) speaker verification reliability test
- Impact of Whitelist Size
- Channel mismatch - Reliability on real radio channel
- Influence of affective speech

# Testing databases

- Librispeech / Voxforge
  - 2444 (English) speakers
- SpeechDat-E Sk
  - 1000 (Slovak) speakers
  - telephone channel
- CRISIS
  - 35 (Slovak) speakers
  - 3 levels of arousal (neutral, higher, high)

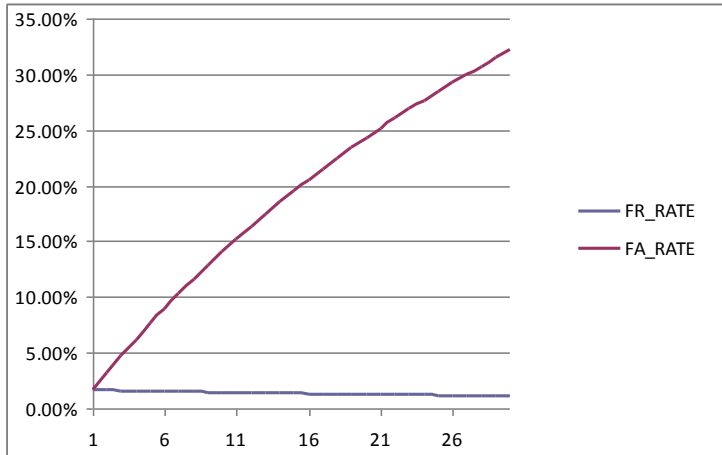
# Basic speaker verification test

DB name	No of enrollments	No of test files	No of test scores	EER
Librispeech	2444	24440	59731360	0.864%
SpeechDat	888	888	788544	0.906%
VoxForge	579	7215	3520920	1.634%

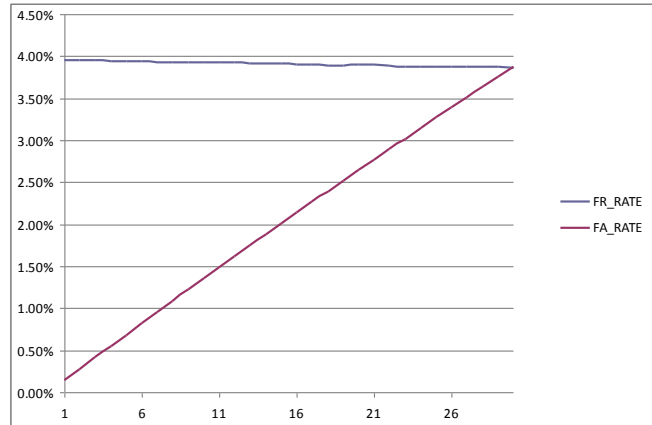


# Influence of the number of group members on group verification

## Fixed-threshold

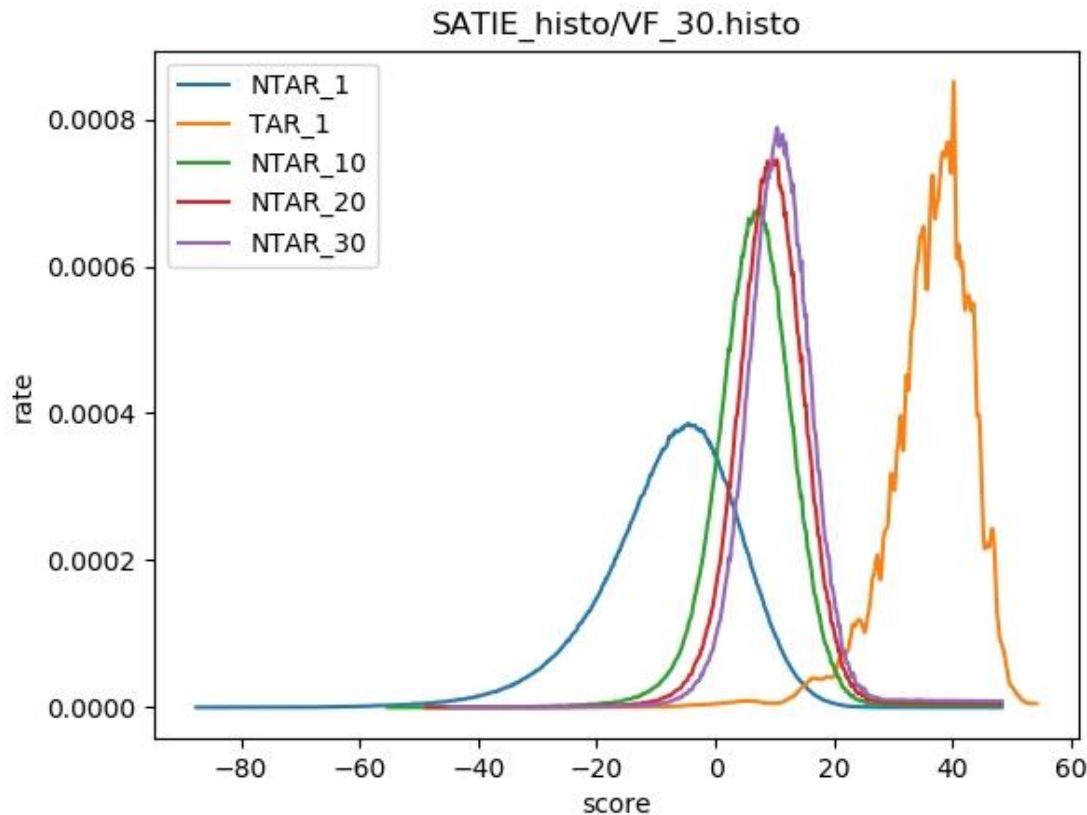


threshold set to achieve EER  
for group-size 1 person

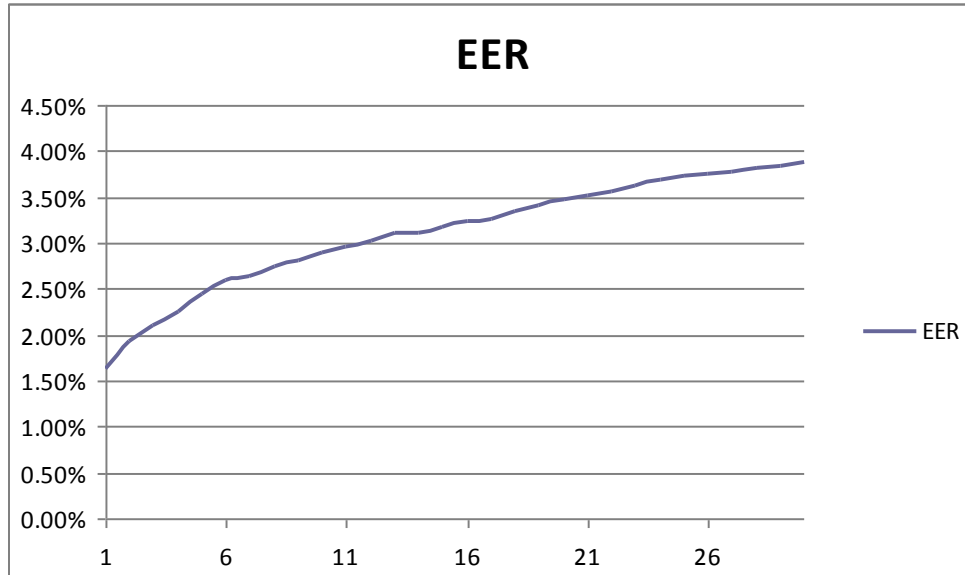


threshold set to achieve EER  
for group-size 30 persons

# Nontarget (NTAR) score distributions shifting



# Influence of the number of group members on group verification



Adaptive-threshold

GROUP SIZE	1	2	3	4	5	10	15	20	25	30
THRESH OLD	13.06	14.69	15.60	16.18	16.57	17.95	18.83	19.22	19.60	19.93
EER	1.63%	1.92%	2.10%	2.26%	2.45%	2.90%	3.17%	3.47%	3.72%	3.87%

# Channel mismatch and Radio-voice database

- Radio transmitted speech database
  - VoxCeleb1 – train and test set
- Baofeng UV5R hand-held radios
- the transmission was performed on publicly available PMR frequencies of 446.0 and 446.1 MHz (**UHF**)



# Radio mismatch test

Training data	Test data	EER
VoxCeleb1_train	VoxCeleb1_test	1.50%
RadioVoxCeleb1_train	RadioVoxCeleb1_test	2.80%
VoxCeleb1_train	RadioVoxCeleb1_test	5.90%
RadioVoxCeleb1_train	VoxCeleb1_test	3.00%
MIXED VoxCeleb1_train + RadioVoxCeleb1_train	VoxCeleb1_test	1.20%
MIXED VoxCeleb1_train + RadioVoxCeleb1_train	RadioVoxCeleb1_test	2.60%

# Style mismatch and emotionally activated (aroused) speech database

		enrollment		
		1 neutral	2 aroused	3 highly aroused
test	1 (neutral)	1.57%	2.79%	6.43%
	2 (aroused)	3.50%	1.64%	2.79%
	3 (highly aroused)	10.00%	4.29%	1.71%
AVG		5.02%	2.91%	3.64%



Thank you for your attention

**Announcement:**

The work presented at this work is funded from the project that has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 832969. The outputs of the presented papers reflect the views only of the author(s), and the European Union cannot be held responsible for any use which may be made of the information contained therein. SATIE public website: <http://satie-h2020.eu/>

The presented work is also co-funded by European Regional Development Fund, Ministry of Transport and Construction of the Slovak Republic and Ministry of Economy of the Slovak Republic, in the frame of the project Early Warning of Alzheimer (reg. No. 67/2020-2060-2230-V631) and from the VEGA agency project No. 1/0667/18, Automatic assessment of acute stress from speech.