Politecnico di Torino
Dipartimento di Automatica e Informatica

PhD in Computer and Control Engineering
XXXVI cycle

Supervisor
Sandro Cumani

Speaker verification and multi-modal identity recognition

PhD Candidate: *Salvatore Sarni*

1. Context

Identity verification may be necessary for both online and offline services. Example include facial recognition to unlock mobile devices and speaker verification employed by virtual assistants.

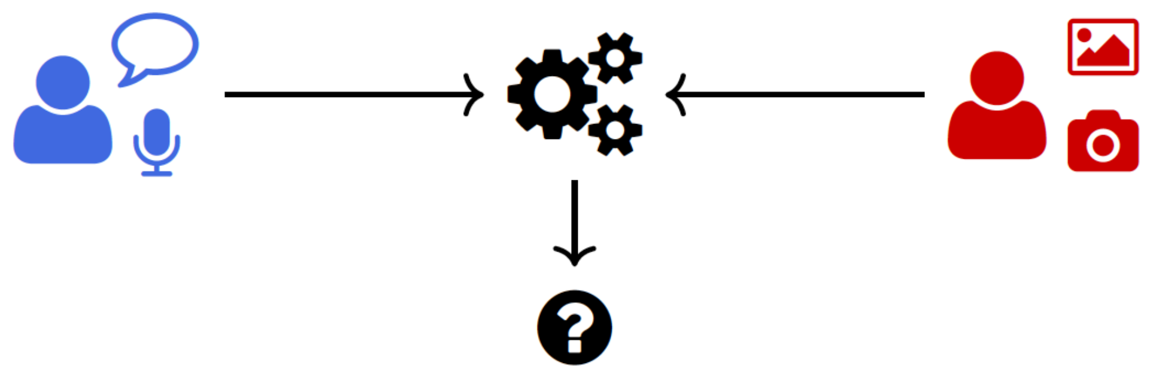
2. Goal

Design and improve speaker verification systems

- Embedding Extraction: from segments with different durations to a low-dimension and fixed representation object
- Scoring & Evaluation: backend classifiers and calibration models

3. Multi & Cross-Modal

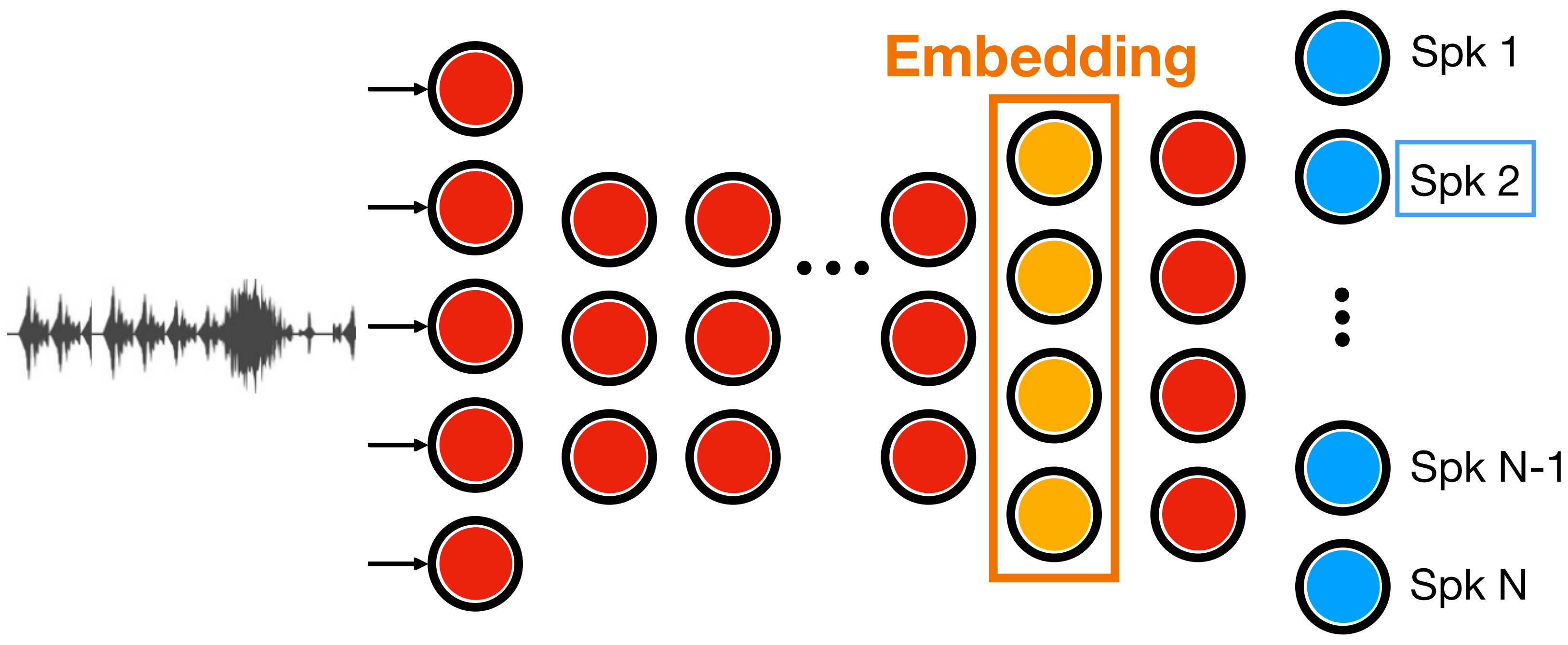
Combining faces and voices



4. Neural Networks

Learn to classify a large number of speakers.

- Different Architecture: TDNN, ResNet, ECAPA, Conformer



5. Backend

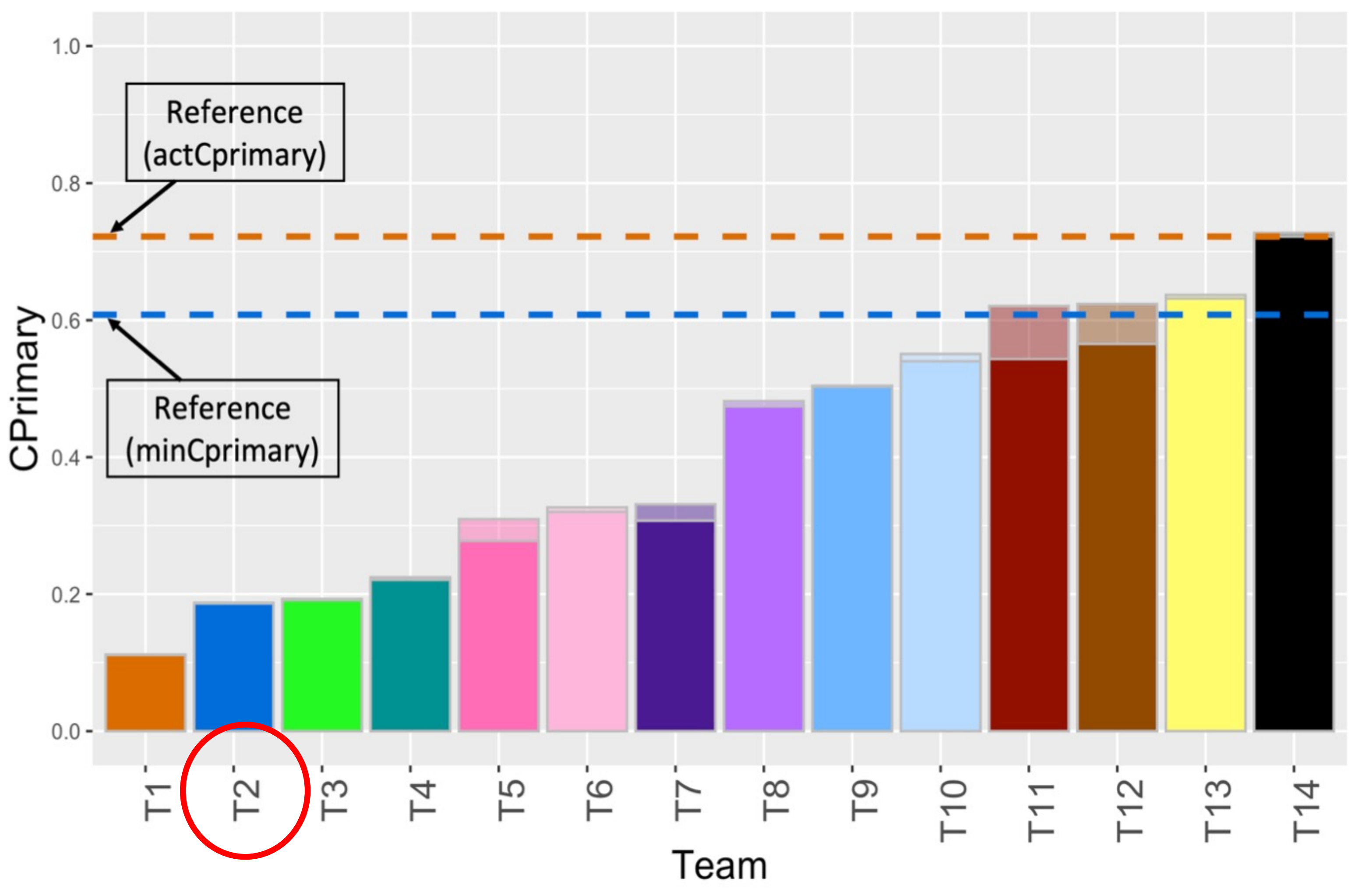
Given two embeddings, do they share the same identity?

- Score: from distance to probability, PLDA
- Thresholds define performance, which one is optimal?
- Calibration & Normalization: side information, such as the duration, is utilized to improve the performance of both discriminative and generative models[2,3]

6. Language Recognition

NIST Language Recognition Evaluation 2022 language detection challenge. Fixed condition track with low-resource test languages.[1]

- Speaker SoA architecture adapted, CNN block and early stages fusion
- Custom training of the backend



7. References

1. Sarni, S., Cumani, S., Siniscalchi, S.M., & Bottino, A. (2023). Description and analysis of the KPT system for NIST Language Recognition Evaluation 2022. *Interpseech* 2023.
2. Cumani, S. & Sarni S. "The Distributions of Uncalibrated Speaker Verification Scores: A Generative Model for Domain Mismatch and Trial-Dependent Calibration." IEEE/ACM Transactions on Audio, Speech, and Language Processing 31 (2023): 2204-2219.
3. Cumani, S. & Sarni, S. (2022). Impostor score statistics as quality measures for the calibration of speaker verification systems. In *Proc. The Speaker and Language Recognition Workshop* (Odyssey 2022) (pp. 25-32)