Zhengyu Chen

# SOFTWARE FOR RECOGNIZING SPEAKER BY VOICE

*This system, based on the lightweight Flask web framework and the ECAPA-TDNN deep learning model, designs and implements a complete speaker recognition system.[1] The backend uses Flask to build a RESTful API service, with core functionalities covering four main modules: voiceprint registration, voiceprint recognition, user management, and audio comparison. The database layer uses SQLite to store user information and embedding vectors, and the frontend uses the Jinja2 template engine to dynamically render pages, displaying recognition results intuitively as a progress bar.*

## Introduction

As a biometric identification technology, speaker recognition has experienced rapid expansion in its application domains in recent years, driven by advancements in deep learning. It is now widely deployed in areas such as call center operations, criminal investigations, smart speakers, human-robot interaction, and mobile security authentication. This report details the design and implementation of a speech-based speaker recognition software system.

## I. Model Training

The system adopts the ECAPA-TDNN model as its core architecture and is trained on a large-scale speech dataset. Before training, audio data is uniformly preprocessed to extract FBank features, and online data augmentation strategies such as noise and reverberation are introduced to enhance model robustness in complex acoustic environments. The model is optimized using the AAM-Softmax loss function to improve feature discriminability, ultimately outputting fixed-dimensional speaker embedding vectors.[2]

## II. Model Evaluation

The model is evaluated on standard test sets by computing cosine similarity between speech segment pairs for speaker comparison. Length normalization and adaptive score normalization are introduced to improve scoring stability. Two core metrics, Equal Error Rate and minimum Detection Cost Function, are used to comprehensively reflect the model's recognition accuracy and misjudgment cost, with the results validating the model's effectiveness and advanced performance.

## III. System Design and Implementation

The system builds its backend services based on the Flask framework, adopting an MVC architecture. The frontend uses the Jinja2 template engine for dynamic page rendering and communicates asynchronously with the backend via AJAX. Core functions include voiceprint registration, recognition, user management, and audio comparison. The system performs unified format preprocessing on uploaded audio, extracts embedding vectors, and stores them in an SQLite database. Identity matching is achieved through cosine similarity and threshold judgment, forming a complete closed-loop speaker recognition application.

## Conclusions

A neural network-based speaker recognition system has been developed. The system supports voiceprint registration, speaker identification, user management, and audio comparison. Identity matching is achieved through neural network-extracted embedding vectors and cosine similarity. Experimental results validate the system's effectiveness, successfully bridging deep learning with practical deployment.

### *References*

1. Desplanques, B. ECAPA-TDNN: Emphasized Channel Attention, Propagation and Aggregation in TDNN Based Speaker Verification / B. Desplanques, J. Thienpondt, K. Demuynck // Interspeech. – 2020.
2. Arcface: Additive angular margin loss for deep face recognition / J. Deng, J. Guo, N. Xue [et al.] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. – 2019. – P. 4690–4699.

*Chen Zhengyu*, Master student, Department of Information Security, BSUIR, 760309916@qq.com.
*Supervisor: O. B. Zelmansky*, Associate Professor of the Department of Information Security, PhD in Engineering Sciences, Associate Professor, BSUIR, 7650772@rambler.ru.