Speaker Identification in Noisy Environments for Forensic Purposes

Armando Rodarte-Rodríguez[1], Aldonso Becerra-Sánchez[1(✉)], José I. De La Rosa-Vargas[1], Nivia I. Escalante-García[2], José E. Olvera-González[2], Emmanuel de J. Velásquez-Martínez[1], and Gustavo Zepeda-Valles[1]

[1] Universidad Autónoma de Zacatecas, Campus Siglo XXI, Carr. Zacatecas-Guadalajara Km. 6, Ejido "La Escondida", 98160 Zacatecas, Mexico
armandorodarte19@gmail.com, {a7donso,gzepeda}@uaz.edu.mx, ismaelrv@ieee.org, iemmanuelvm@gmail.com

[2] Tecnológico Nacional de México Campus Pabellón de Arteaga, Carretera a la Estación de Rincón Km. 1, Pabellón de Arteaga, 20670 Aguascalientes, Mexico
aivineg82@gmail.com, e.olvera.itp@gmail.com

Abstract. The speech is a biological or physical feature unique to each person, and this is widely used in speaker identification tasks like access control, transaction authentication, home automation applications, among others. The aim of this research is to propose a connected-words speaker recognition scheme based on a closed-set speaker-independent voice corpus in noisy environments that can be applied in contexts such as forensic purposes. Using a KDD analysis, MFCCs were used as filtering technique to extract speech features from 158 speakers, to later carry out the speaker identification process. Paper presents a performance comparison of ANN, KNN and logistic regression models, which obtained a F1 score of 98%, 98.32% and 97.75%, respectively. The results show that schemes such as KNN and ANN can achieve a similar performance in full voice files when applying the proposed KDD framework, generating robust models applied in forensic environments.

Keywords: Artificial intelligence · KDD · Prototyping · Speaker identification · Speech processing

1 Introduction

The speech is a unique biological feature in each person, caused by the differences in the organs of phonation, articulation and breathing. Thus, the particular characteristics of the speech and the way of speaking in each person are their biometric signatures [1, 2]. In this context, speaker recognition is the process of extracting information that describes the identity of people from their voice features [3, 4]. The speaker identification process is normally made up of 3 stages: pre-processing, speaker or speech feature extraction and classification [3, 5, 6]. The pre-processing stage consists in modifying the speech signal to make it suitable for feature extraction analysis [6–8]. Meantime, feature extraction is a

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2023
J. Mejia et al. (Eds.): CIMPS 2022, LNNS 576, pp. 299–312, 2023.
https://doi.org/10.1007/978-3-031-20322-0_21