جامعة الموصل

كلية علوم الحاسوب والرياضيات

التعرف على هوية المتكلم باستخدام التعلم الآلي

قاسم صديق محمود إسماعيل

رسالة ماجستير

علوم الحاسوب

بإشراف

د. يسرى فيصل الأرحيم

أستاذ مساعد

1443 هـ 2021 م

الملخـص

يعد التعرف على هوية المتكلم أحدى المشاكل الرئيسة في نمذجة الصوت و معالجة الكلام. ويوجد العديد من تطبيقات التعرف على هوية المتكلم منها يعتمد على المصادقة في أنظمة الأمان، إن تطبيقات التعرف على هوية المتكلم تشكل تحديًا كبيراً على نطاق واسع؛ إذ يتطلب تقانات حديثة تتميز بالبحث السريع في قاعدة بيانات الأصوات وتعتمد على الذكاء الاصطناعي لتحقيق النتائج المرجوة من النظام، تتطلب هذه الأنظمة عادةً العديد من العمليات والمراحل التي يجب أن تمر فيها الإشارة الصوتية للوصول إلى النتيجة المرجوة، فهناك فئتان من هذه الأنظمة التي تستخدم للتعرف على هوية المتكلم منها الأنظمة المستقلة عن النص أو الأنظمة التي تعتمد على النص.

في هذه الرسالة صمم نظامان لتحديد هوية المتكلم، يتميز النظامان بقدرتهم على التعرف على هوية المتكلم بأقل وقت و أقل عدد من عينات التدريب ولا يتطلب النظامان موارد كبيرة للتعرف على هوية المتكلم. يتكون كل النظام من أربعة أجزاء رئيسة: الجزء الاول بناء قواعد بيانات صوتية، اعتمد في الرسالة على قاعدتين للبيانات الصوتية الأولى قاعدة بيانات (QS-Dataset)، سجل ملفاتها الصوتية الباحث في بيئة وظروف مناسبة خالية من أي ضوضاء أو مؤثرات صوتية وتتضمن هذه القاعدة مجموعة من عينات صوتية للذكور والإناث، أما قاعدة البيانات الثانية (audioMNIST_meta) فحصل عليها من الانترنيت ولها خصائص قاعدة البيانات الأولى نفسها. وقد عولجت وهيئت الملفات بطريقة شرحت في متن البحث لاحقاً. الجزء الثاني استخلاص الخصائص: تستخلص الخصائص من قواعد البيانات الصوتية عن طريق خوارزمية معاملات درجة النغم(Mel Frequency Cepstral Coefficients) (MFCC). االجزء الثالث تحديد المصنف:

استخدام نوعين من الشبكات العصبية الشبكة بيرسبترون متعددة الطبقات (Multilayer Perceptron Neural Network (MPNN) والشبكة العصبية الالتفافية (Convolution Neural Network (CNN)) كمصنف.الجزء الأخير تحقق من النظام: التحقق من عمل و نتائج النظام عن طريق مقاييس (معدل القبول الخاطئ (False Acceptance Rat(FAR)) ومعدل الرفض الخاطئ (False rejection rate(FRR))).

أوضحت نتائج الاختبار على قواعد البيانات (QS-Dataset) و(audioMNIST_meta) قدرة شبكة (MPNN) على التعامل مع أقل عدد من البيانات التي تشمل 5 أشخاص وحققت نسبة (97% و 86% على التوالي) وأما مع البيانات الكبيرة التي تشمل 35 شخصاً كانت نتائج (77% و 57% على توالي) على عكس شبكة (CNN) كانت نتائج غير جيدة للبيانات القليلة (85% و 61% على توالي) وأما مع بيانات الكبيرة كانت النتائج (97% و88% على التوالي)، وكما اختبرت الشبكات على مستوى الأمان والتحقق من النظام عن طريق استخدام المقاييس (FAR و FRR) حيث أظهرت النتائج عالية جدا.

University of Mosul
College of Computer Sciences
and Mathematics

Speaker identification using machine learning

Qasim Sadeeq Mahmood Esmaeel

M.Sc./Thesis

Computer Sciences

Supervised by

Dr. Yusra Faisal Al-Irhayim

Assistant Professor

2021 A. D **1443 A.H**

Abstract

Speaker identification is one of the major problems in voice modeling and speech processing. There are many applications of speaker identification, including authentication in security systems. The applications of speaker identification pose a huge challenge on a large scale, as it requires modern technologies that are characterized by rapid search in the voice database and rely on artificial intelligence to achieve the desired results from the system. These systems usually require several processes and stages that the audio signal must pass through to reach the desired result. There are two categories of these systems that are used to identify the identity of the speaker, including text-independent or text-dependent systems.

In this thesis, two systems were designed to identify the speaker, as the two systems are characterized by their ability to identify the identity of the speaker in the least time and the least number of training samples, and the two systems do not require large resources to identify the identity of the speaker. The whole system consists of four main parts, Part One Building audio databases: The thesis relied on two audio databases, the first is a QS-Dataset whose audio files were recorded by the researcher in an appropriate environment and conditions free from any noise or sound effects. This base includes a set of samples The second database (audioMNIST_meta) was obtained from the Internet and has the same characteristics as the first database. The files were processed and formatted in a way that was explained in the body of the research later. Part Two feature extraction: Characteristics are extracted from audio databases by the Mel Frequency Cepstral Coefficients (MFCC) algorithm. Part Three Defining the classifier: Using two types of neural networks, the Multilayer Perceptron Neural

Networks (MPNN) and the Convolution Neural Network (CNN) as a classifier. The last part checks the system: checking the functioning and results of the system by means of a scale (False Acceptance Rate (FAR) and False rejection rate (FRR).

The results of the test on the (QS-Dataset) and (audioMNIST_meta) databases showed the ability of the (MPNN) network to deal with the least number of data that includes 5 peoples, as it achieved a percentage (97% and 86% respectively) as for the big data that includes 35 peoples, the results were (77% and 57%, respectively), in contrast to the network (CNN) the results were not good for the small data (85% and 61%, respectively) and with big data the results were (97% and 88%, respectively) as the networks were tested On the level of safety and verification, where the results showed very high.