Digital Receipt

This receipt acknowledges that Turnitin received your paper. Below you will find the receipt information regarding your submission.

The first page of your submissions is displayed below.

Submission author:	Ye Kyaw Thu
Assignment title:	NLP Paper
Submission title:	lang-identification-revised
File name:	Report-Paper-final.pdf
File size:	345.33K
Page count:	4
Word count:	2,790
Character count:	15,803
Submission date:	29-Jan-2021 06:33PM (UTC+0630)
Submission ID:	1496958004

Automatic Language Identification on Audio Signals

Khaing Zar Mon
ME-IST-5
Department of Information Science
University of Technology (Yatanarpon Cyber City)
khaingzarmon@utycc.edu.mm

May Phyu Khin
ME-IST-7
Department of Information Science
University of Technology (Yatanarpon Cyber City)
mayphyukhin@utycc.edu.mm

Myo Mar Thinn
ME-IST-3
Department of Information Science
University of Technology (Yatanarpon Cyber City)
myomarthinn@utycc.edu.mm

Ei Thandar Phyu
ME-IST-2
Department of Information Science
University of Technology (Yatanarpon Cyber City)
eithandarphyu@utycc.edu.mm

Nang Aeindray Kyaw
ME-IST-6
Department of Information Science
University of Technology (Yatanarpon Cyber City)
nangaeindraykyaw@utycc.edu.mm

***Abstract*—The purpose of our study is to identify which language is spoken from a speech sample, based solely on the acoustic information conveyed by the speech signal. There are many characteristics of speech that could be used to identify languages. Languages are made up of different sounds that form phonemes, so it is possible to distinguish languages based on the acoustic features present in the speech signal. We decided to study language identification of four different languages, namely English, Chinese, Myanmar (Burmese) and Shan (Tai Long). In this study, three different approaches were used to carry out the experiments. The first one is language identification from Mel Frequency Cepstral Coefficients (MFCC) using Keras Library. In the second approach, we use waveforms of raw audio signals as input to a Residual Neural Networks (ResNets). For the last approach, we use spectrograms of our audio data as input to ResNets which in turn is trained for language identification. According to our experimental results, all the approaches give promising results. The first approach, language identification with MECC using Keras, and the last approach, identification using spectrograms, give 100% testing accuracy. We expect a more useful system can be developed as more data (with more language classes) comes available in the future.**

***Index Terms*—Language Identification, Residual Neural Networks (ResNets), Mel Frequency Cepstral Coefficients (MFCC)**

I. Introduction

Recently, voice assistants have become a staple in the flagship products of many big technology companies such as Google, Apple, Amazon, and Microsoft. One challenge for intelligent assistants, like Siri or the Google Assistant, is that the language that a speaker is using needs to be preset. To improve user experience on this and similar tasks such as automated speech detection or speech to text transcription, automatic language detection is a necessary first step. Moreover, emergency call routing can be one of the applications of language identification, where the response time of a fluent native operator might be critical. The main motivation is to study language identification from audio data using different Deep Learning approaches.

Finding a dataset of audio clips in various languages sufficiently large for training a network was an initial challenge for this task. Due to the difficulty to get open-source audio corpus for various languages, we only use audio corpus with four different languages, namely English, Chinese, Myanmar (Burmese) and Shan (Tai Long). In this project, we studied the implementation and evaluation of three different language identification models and compared the results. Firstly, we extracted Mel Frequency Cepstral Coefficients (MFCC) features from our audio data and trained with Keras Library to build a language identification model. We also studied to use Residual Neural Networks (ResNets) for language identification and built two models with ResNets. Waveforms of audio data are used as input of the second model and spectrograms of audio data are used as input of the third model.

The structure of this paper is organized as follows. Related works of language identification for different languages using audio data are presented in the upcoming section. Section III describes the methodologies used in the language identification experiments. Later, in Section IV, we present corpus preparation and data pre-processing together with feature extraction techniques used for the experiments. The experimental results along with some discussions are described in Section V. Eventually, in Section VI, we conclude our works showing promising results.