International Journal of Scientific & Engineering Research Volume 9, Issue 3, March-2018
ISSN 2229-5518

Survey on Caller Identification by Voice

Arpit Thakkar, Kashmira Wankhede, Jitendra Patil

Abstract - The article studies age and gender related variations of speech characteristics of an individual to develop software for caller identification or to create his/her characteristics by analysis of voice. This security project covers speaker recognition, speaker's age, and gender and emotion detection. This complex system will include speech signal analysis, generation of pattern database and appropriate classification methods. System aims to provide authentication and identification in the areas where biometrics cannot be used. We have considered different voice features like pitch, intensity, jitter, shimmer, formants. Voice samples will be recorded and analyzed for different age groups of both the genders.

Index Terms— Caller Identification, Gender, Age, Formants, Jitter, Shimmer, Speech, Pitch, Intensity, Voice Features, Caller Authentication.

---

1 Introduction

In today's world, finding out the identity of the caller, which he or she claims to be, is an important thing. Whenever we receive a call, it is necessary for the receiver to identify the caller as the true caller. Also in the case of emergency situations, getting the correct and legit information about the caller is necessary for the responders or receivers. Whenever we get a call from a person X, it is important that we get a confirmation that the calling the calling person is X itself and not someone else under the identity of X. Software which identifies the calling person will help us to be sure that the identity claimed by the calling person is true. As discussed earlier, in case of emergency situations, getting correct information about the caller, in less time, is necessary. We don't have much time to ask for each and every detail in case of emergency situation. Also trying to authenticate or authorize the user on our own will take a lot of time. While in emergency situations, every second of delay between asking for help and sending necessary services can prove to be harmful.

Physically exhaustive type of work leads to loss of concentration and waste of time. We should be able to help the ones in emergency situations, as earlier as possible. Thus any system that supports identification of caller in efficient time span reduces the call time and allows the receiver to help the caller quickly in case of emergency situations. This system can also help Threat detection as it will help the receiver to authenticate the caller as a legit caller and not someone who's trying to fool him / her. This can be done with the help analysis of different voice characteristics.

Human voice consists of sound made by the human being using the vocal tract and mechanism for generating human voice can be divided into three parts; the lungs, the vocal folds within the larynx and the articulators. Muscles of larynx adjust the length and tension of vocal folds to fine-tune the pitch and tone. Studies show that an individual's voice features changes with age and gender. Different voice characteristics like pitch, frequency, jitter, shimmer and formant frequencies experience age with the aging.

- *Arpit Thakkar is currently pursuing Bachelor's degree program in Information Technology engineering in Mumbai University, India. E-mail: tipra9345@gmail.com*
- *Kashmira Wankhede is currently pursuing Bachelor's degree program in Information Technology engineering in Mumbai University, India. E-mail: kashmirawankhede@gmail.com*
- *Jitendra Patil is Assistant Professor and H.O.D of Information Technology engineering in college in Mumbai University, India. E-mail: jeetoo.patil@gmail.com*

1.1 Voice Parameters and it's correlation with age and gender

Pitch is the rate of vibrations of vocal folds. Sound of voice changes as the rate of vibration varies. As the number of vibrations increases, pitch increases, meaning the voice would sound higher. These Vibrations are dependent on the length and thickness of the vocal cord as well as the tightening of the muscles surrounding them. Male voice pitch tends to rise with age while female voice pitch remains same or may lower slightly.

Shimmer is the measure of variability of the peak-to-peak amplitude of the signal. Shimmer has been found to have strong correlation with age. Shimmer measures for older males are higher as compared to adult males.

Jitter is a measure of cycle-to-cycle variation of pitch period. Jitter is caused by instability in vocal fold vibrations. It correlates with hoarseness in voice. Jitter increases with age for both males and females.

As all these features changes with Age and Gender, we can find out the age group and gender of the individual with the help of his or her voice.

1.2 Overview

The aim of our work is to develop a tool or application for caller identification, by the analysis of his / her voice signal. This will require the speech signals of different users from different categories, a tool for sampling of speech signals and extraction of attribute values of speech, a database, pattern matching algorithm, a user interface, etc.

In this paper we present the advances in development of system that will identify the caller by voice characteristics. System will integrate various tools that will help in gathering the required information, collect the voice samples and extract the attributes.

The only input to the system will be the speech signals of different callers. Using this speech samples, various voice characteristics or attributes like pitch, frequency, intensity, speech rate, etc., will be calculated and then stored in database. Real time speech samples will be compared with these stored values and based upon the result of matching operation, caller profile will be displayed.

2 General System Description

The system that will be designed in this project will be able to identify the calling person and display his / her profile. An automatically generated caller profile including name, age and gender of the caller will be displayed to the receiver. The aim of the system is to identify the caller at real-time for both, emergency situations as well as threat detection. System aims to identify the caller and to authenticate and identify the caller with the help of his / her voice characteristics. People can use this application to identify the caller and to get complete and genuine information about age and gender of the calling person. system will give the age group to which the caller belongs and the gender of the caller along with the name of the caller if his/her voice samples and it's attributes are already stored in database.
Advantages of the system will be: caller identification, and, detection of age and gender of the caller.

3 Data Collection and Methodology

In this paper, we have recorded the voices of 50 males and 50 females for different age groups. 10 males and 10 females aged between 21-30 years, 10 between 31-40 years, 10 between 41-50 years, 10 aged above 50 years and 10 aged below 20 years. Voices are analyzed for voice parameters like pitch, intensity, formant frequencies and jitter and shimmer. All the voices were recorded and then the values of required voice features were extracted from all samples with the help of Praat software and were stored in database for further use. Appropriate classification algorithms will be used for classifying the voice samples according to age and for identifying the age group and gender of the real time caller.

4 Tools used

Tools available and needed for this work are namely "Praat" and "Matlab".

4.1 Praat:

Praat is digital sound analyzer. It is a free computer software package for scientific analysis of speech in phonetics. It is used for removal of noise from speech sample and calculating attributes like pitch, intensity, formant frequencies, jitter, shimmer. It can help to analyse the pre-recorded voice samples and extract attributes from them or one can directly record a voice sample using software. By opening a recorded voice samples in Praat, we will get a waveform for that particular speech signal. It gives different options for intensity, pitch, frequency; etc.one can easily get the values of those voice characteristics by selecting the option for that characteristic from the toolbar. Apart from giving a direct value for those characteristics, it also gives the value for that characteristic at the selected point (cursor point) on the waveform.

Features of Praat:

- Contains a wide range of analysis tools and algorithms.
- Contains a script language, manipulation tools, and numerical tools for optimization.
- Contains graphical processors to render results in graphical format, speech synthesis components.

4.2 Matlab:

Matlab is an abbreviation for Matrix Laboratory. It's a multi-paradigm numerical computing environment. A proprietary programming language developed by Math Works, Matlab allows matrix multiplication, plotting of functions and data, implementation of other languages including C, C++, java, FORTRAN and Python.
Matlab will be used for digital signal processing. It can be used to extract the attributes and store them directly in database.

5 Flow of Working

The system will be the result of this work will function in the following manner:
First step will be to record the voice samples of different individuals belonging to different age groups and gender. These recorded voice samples will be stored by classifying them according to age and gender. After that, the voice samples will be converted from default file format (mp3) to (.wav) for further processing. After having all the voice samples in the required file format, we will then extract the attributes like pitch, intensity, formant frequencies, jitter and shimmer from those from voice samples with the help of Praat software. Once the values of all the required voice characteristics are extracted, those values will be stored them in a database. This database will contain the values of all necessary characteristics like pitch, intensity, etc. on real-time , when receiving a call, the system will record the real-time speech signal and extract essential attributes like pitch , intensity, formant frequencies, jitter , shimmer, etc. These real-time values will be compared with the values stored in the database. On complete matching; the system will display caller profile including name, gender and the age group of the caller. For known callers, the values extracted from their pre-recorded voice samples will be already stored in database, so system will display whole profile including identity (name) of the caller. for unknown caller , system will first extract the attributes , then display caller profile including age group and gender of the caller except the age of the caller and add the new values (along with name of the caller) to the database for further use.

5.1 Activity Diagram

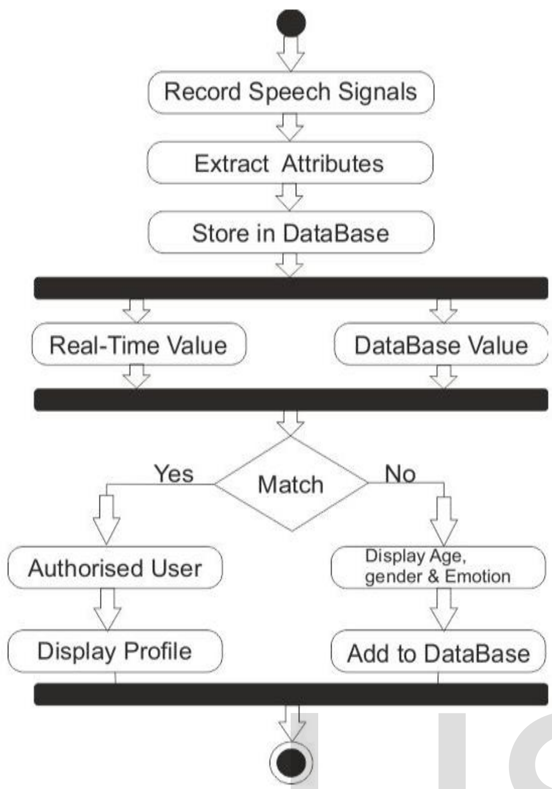

Fig.1. The figure above shows the activity diagram for the implementation of the system.

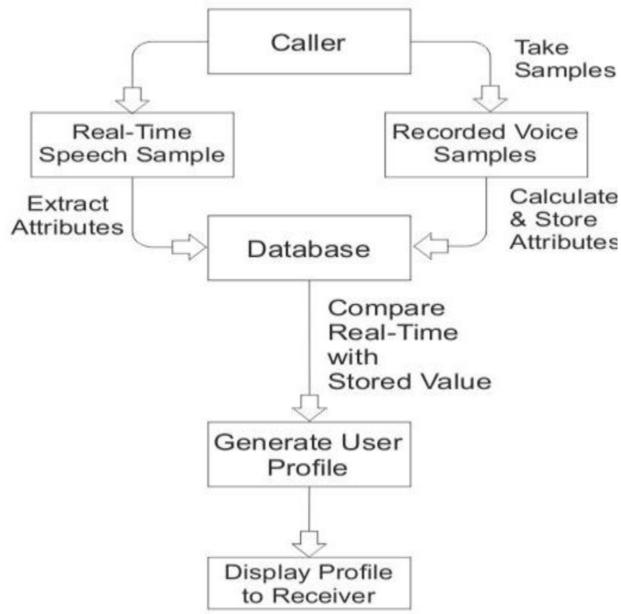

Fig.2. The figure above shows the dataflow diagram for the implementation of the system.

The first activity will be recording the voice samples. Second and third activity would be to extract and store the attributes in database. After that, the real-time values and the database values will be compared to find a match. On the basis of result obtained after matching operation, if a match is found, caller profile including age and gender along with the identity (name) of the caller will be displayed. If a match is not found, all information i.e. age and gender, except the name will be displayed and caller information will be added to the database for further reference.

5.2 Dataflow Diagram

The main objects of the system would be; caller, real-time voice samples, database voice samples and database. Speech samples of callers can be classified as real-time samples and database samples. For pre-recorded samples, first the voice will be recorded and attributes would be extracted and stored in database for real-time use. For real-time voice samples, attributes will be extracted directly at that moment. Both pre-recorded (stored) and real-time values will be compared for generating and displaying caller profile.

5 Conclusion

Based on preliminary data analysis, we can conclude that, a 30 sec long speech stored in database is enough to identify the calling person in 5 seconds using the proposed system. Automatic generation of caller profile is based on speech signal analysis. The effects of the proposed system are still unclear but the proposed concept can be a step towards reduction in frauds or identity spoofing over a call or to get information about the caller's true gender and age group.

References

[1] M. Witkowski, M. Igras, J. Grzybowska, P. Jaciow, J. Galka and M. Ziolko. Caller Identification by Voice. XXII Annual Pacific Voice Conference (PVC), IEEE - 2014.

[2] Das, B., Mandal, S., Mitra, P. et al. Int J Speech Technol (2013) 16: 19.

[3] Najiya Abdulrahiman and Ranju K.V. Text Dependent Speaker Recognition. International Journal of Electrical Communications - 2013.

[4] R. Rajeswara Rao, A. Nagesh, Kamakshi Prasad and K. Ephraim Babu. TextDependent Speaker Recognition System for Indian Languages. International Journal of Computer Science and Network Security - 2007.

[5] Lawrence R. Rabiner and Ronald W. Schafer. Introduction to Digital Speech Processing. Foundations and Trends in Signal Processing, Vol. 1: No. 1-2,pp 1-194, December - 2007.