Comprehensive Design and Analysis Project

Design Document

IT16160330 U.P.A.S.D. Amarasinghe

Fraud Detection Based on Human Voice and Speaking Patterns

Bachelor of Science (Honors) in Information Technology

Department of Software Engineering

Sri Lanka Institute of Information Technology

Sri Lanka

May 2019

TABLE OF CONTENT

Table of Contents

LIST OF FIGURES

1 INTRODUCTION

1.1 Purpose

This design document main purpose is to present detailed description of fraud detection based on human voice patterns. In this research component able to identify lies by analyzing voice signals of a customer. And this document also contains what kind of application delivering the end use and what kind of benefits will be receive to the end-user. Also, how we are going to improve performance and other non-functional requirements optimizing the system. This document will act as a technical guide for users and developers that uses the system.

1.2 Scope

Scope of this document covers main functionalities of this specific research component and main functional requirements and non-functional requirements. This research component is all about fraud detection by analyzing voice signals. High level architecture diagrams as well as System interfaces, software interfaces as well as hardware interfaces will be described in this document.

1.3 Definitions, Acronyms, and Abbreviations

ML – Machine Learning
GUI – Graphical User Interfaces
UI – User Interface
DD - Design Document
PC - Personal Computer
LAN - Local Area Network
WAN - Wide Area Network

1.4 Overview

The fraud detection module which is based on human voice patterns is all about detecting the fraudulent activities as early as possible when customer interacting with the cashier. Customer voice single will be preprocessing before analyzing it. Since our product deploying on a crowded environment. First voice clip has to preprocess by reducing background noises and isolating current customers voice. Then frauds detecting by comparing voice signal with psychological lie detecting facts. After analyzing the voice signal counter point user will be alerted according to result. Results is more likely suspicious level of current customer according customer's voice signal.

2 OVERALL DESCRIPTIONS

This application is all part of a larger application main application is about finding frauds earlier based on human behavioral patterns and transaction patterns. Behavioral patterns are facial expressions, body language and human voice and speaking patterns.

This design document is narrow down to only voice-based fraud detecting research component. Using ML models original sound signal will be prepressing to clean the background noises and isolate the speaker's voice. Then based on psychological lie detecting facts another ML model predict the suspicious level of the certain customer according the his/her speaking pattern. If suspicious level percentage is higher than average level user will be alerted by the system.

Main hardware equipment is the microphones for this component. Software GUI interfaces are login screen, dashboard, current customer details UI and previous customers details UI. Only two actors will be considered is this system admins (bank manager) and user (counter pointe cashiers). Cashier will be alerted about the current customer, everything else can access by admins.

All predicted details will be stored in a secured database. And in future second phase of this system that information will be using to fine tune this system by re training the ML model.

2.1 Product perspective

This product is more likely innovative product there are no other similar products to compare with this product. Polygraphs machine is one of similar device but purpose of polygraph machine is to detect lies on criminal activities. Polygraphs machine can only be able to use with single person and covered environment without any background noises. This fraud detecting product is all about finding frauds before it happens and this product is mostly focusing crowded environments. [1]

And another product is the X13-VSA is the closest publicly available product found on internet. This system only analyzing the voices stress level to detect the lies. although this system is analyzing voice with algorithms. And this product also focusing one person in a crowded environment this product also will not works. [2]

In **Error! Reference source not found.** . According to this information some required components of our product exist individually but combined product doesn't exist yet.

	Hardware Equipment	Third Party Software	Datasets (public)	Services, Libraries or Models	Technology Researches
Background Noise Reduction	**YES**	**YES**	**YES**	**YES**	**YES**
Isolating Vocals	**NO**	**YES**	**YES** **mir-1k** [3]	**YES**	**YES**
Identify the Speaker	**NO**	**NO**	**YES** **TIMIT** [4]	**YES** [5]	**YES**
Video Vocal Mapping	**NO**	**NO**	**NO**	**NO**	**YES** **Google** [6]
Voice Based Lie Detectors	**YES** **Polygraph** [1]	**YES** **X13-VSA** [2]	**NO**	**NO**	**YES**
Identify Key Points of Voice	**NO**	**SOME** **emotions**	**NO**	**NO**	**SOME** **Stress, emotions**

Table 1:Summarized table of existing resources related to voices-based fraud detection

2.1.1 System interfaces

The main application is deployer in the administrator PC, and counter places applications installed in cashier point PCs. All these applications running on windows operating systems. Database server is a MS SQL server.

2.1.2 User interfaces

In main application there are compulsory user interfaces for long and the landing page is the dash board.

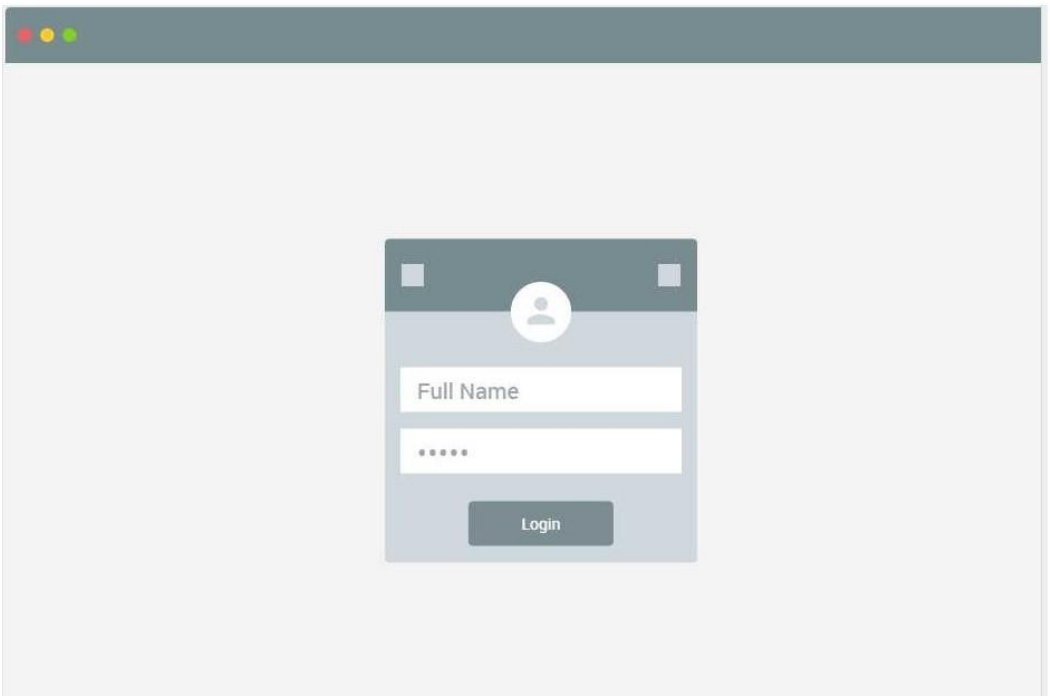

Figure: Login Page

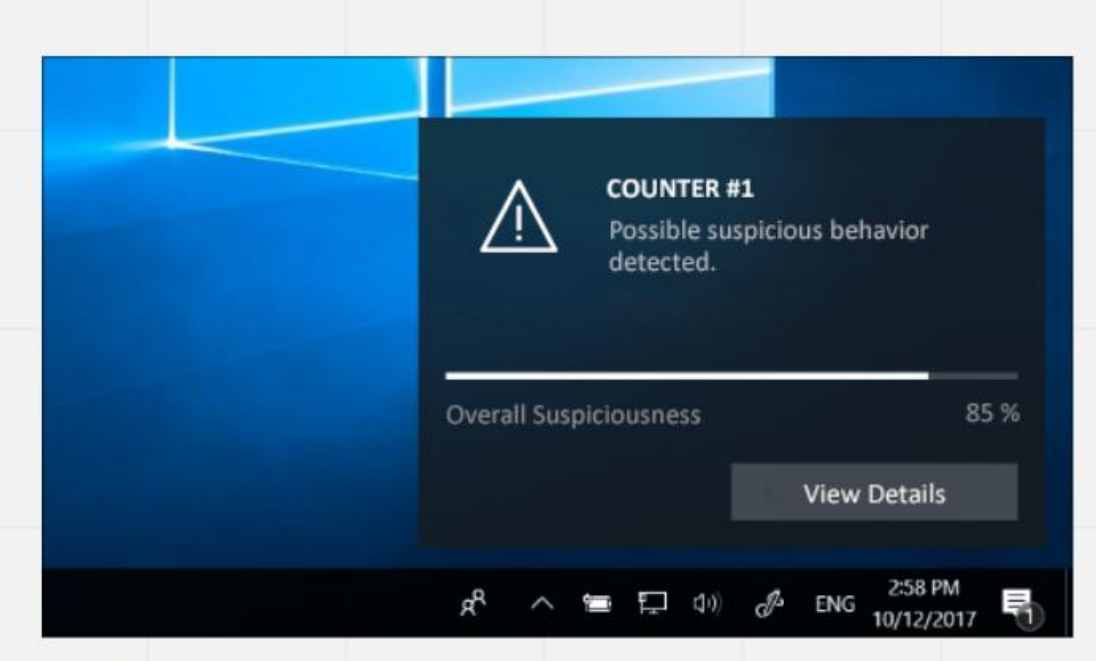

Figure 1:Windows Notification

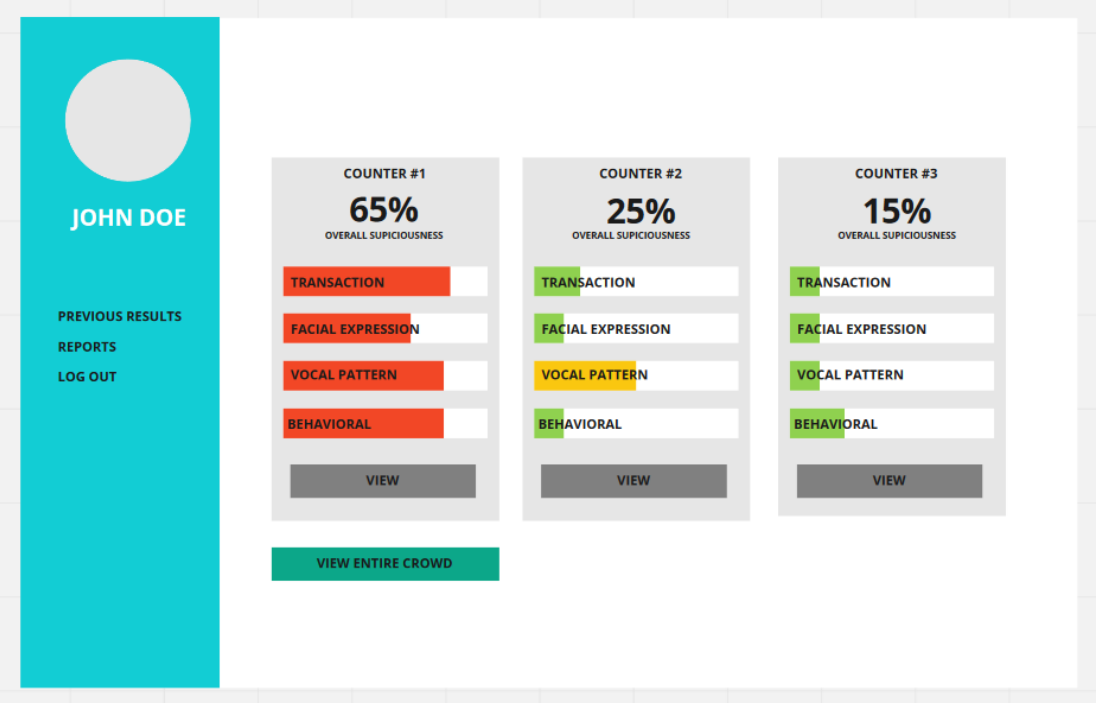

Figure 2 : Lading page dashboard UI.

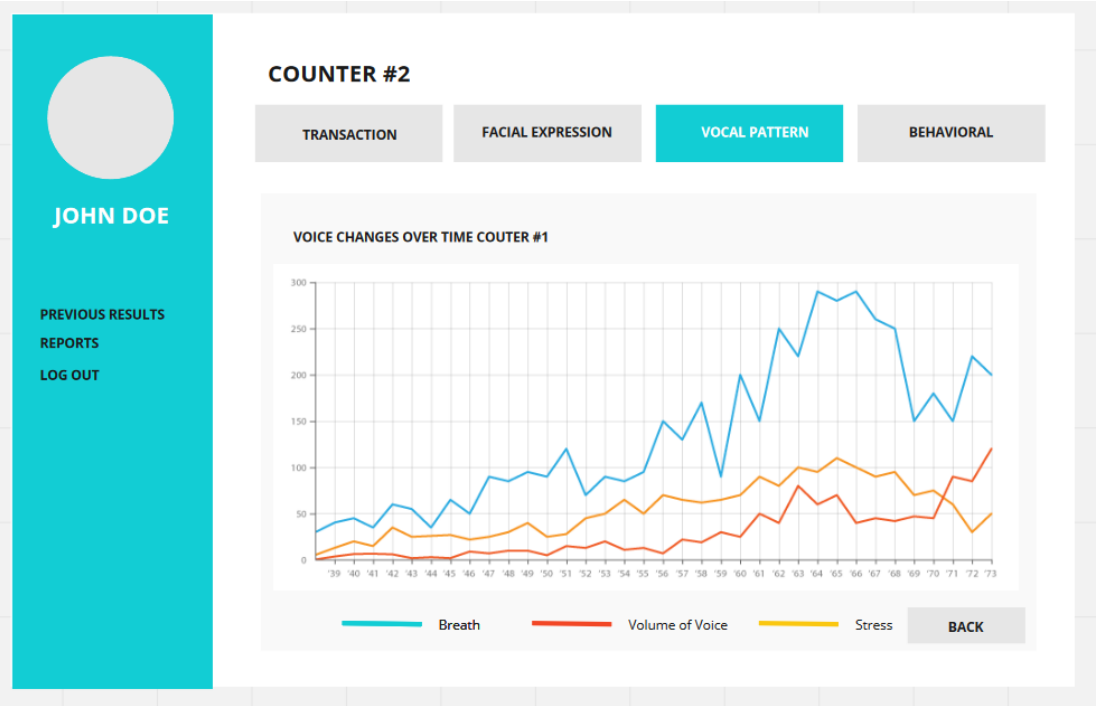

Figure 3:Voice Analyzed Details Graph

2.1.3 Hardware interfaces

There are few main required hardware interfaces involving to the this application.

Microphones – To captures the voice of the current customer.

USB Hub – To connect all microphone the main system.

2.1.4 Software interfaces

For this application all main functions handling by ML models. TensorFlow is the main ML library we are using for this component. Python is the main programing language to use for developments.

End application is based on ASP.Net framework-based C# application. MS SQL server is the database management solution.

2.1.5 Communication interfaces

In this application end user administrator UI are also able access through internet. Because of the requirement this application require communication devices such as router and a modem. Without the communication device main functionality will work perfectly.

2.1.6 Memory constraints

For server machine 8GB RAM recommended to run this main application but counter points PC dose not required that much memory allocation. Since window is the main OS, we are targeting this application 4GB RAM required for each cashier point PCs.

2.1.7 Operations

Administrator required to login to the main system. Administrator dashboard is showing the suspicious level of current customers.

Cashiers doesn’t require any login to system. Cashier point machine only alert the about the customer if customer is suspicious. They can only view current customers suspicious level only.

2.1.8 Site adaptation requirements

Most banking environment use windows-based OS. Because of that our end product also windows-based application. Cashiers will be alerted about the customer using windows push notification. Cashiers dose not required not open application they are able to do their daily schedules.

Microphones deploying on a cashier point might become an issue. For this issue we are going to use some of tiny microphone that customers hardly noticing a microphone is installed at a counter.

2.2 Product functions

Voice Recording – This product required record every details of the cashier and customer.

Background noise reduction – Since this application is running on a most crowded environment all voice recoding contain lots of background noises and other people's voices. So, cleaning the sound clip by reducing background noises of the voice clip.

Voice Isolating – Noise reduction will not isolate single persons voice. Since that issue we are isolating current customer's voice by isolating dominant persons voice signal.

Identifying key points of lying – After pre processing the main sound clip. Product need to analyze the voice clip by comparing psychological lie detecting rules.

View Result – ML model only give percentage of suspiciousness certain customer. It should covert into graphical representation to show to user.

2.3 User characteristics

Counter User (Cashier) – High school graduated person. With basic knowledge of Computer and the windows operating system.

Administrator (Manager) – College graduated person. With basic knowledge of computers and windows operating system as well as basic psychology.

2.4 Assumptions and dependencies

This application will be a web based one. That will remove most of the dependencies found in a normal application since it is separated from the operating system, and other services.

Assume that one server can handle all the customers in a peak hour.

Assume that fraudsters are not highly trained to hide their body language and emotions.

2.5 Apportioning of requirements

All mentioned requirement is 2.2 are required to implemented in this application. Order is the same as mentioned in 2.2.

Machine learning model should be the highest priority task. It should be implemented using TensorFlow.

Then the Data should be fed to the model. After that the model will be trained.

User interface is the least prioritized task. It will be needed for the bank personnel.

specifications of this component. In the 3rd section the functional and non-functional requirements are mentioned in detail. If any major defects are found according to the requirements the testing will be done and defects will be corrected. Application will be implemented by the developers in horizontal manner and no function will be completed at the middle of the development.

Essential requirements of this system are,

- Background noise reduction.
- Isolating speakers voice.
- Identify voice-based lie detecting rules.

Desirable requirements of this system are,

- Classification of persons according to the level of suspiciousness.
- Learning from previous findings.

3 **SPECIFIC REQUIREMENTS (**FOR ORIENTED PROJECTS - DD)

3.1 External interface requirements

3.1.1 User interfaces

Login Page

Pre-determined personnel credentials are inserted in to the database beforehand. If the user is able to provide the given valid username and password, access will be granted to enter the system.

Dashboard

After logging into the system, Landing page show the current counters and customers analyzed data. User able see the current customers suspicious level and the is the customer suspicious or not. If user won't the view more details user can click the view button and then software will redirect to current customers detail view dashboard page.

Voice Analyzing Dashboard.

User allows to view more information about voice analyed information by clicking the voice-based fraud detection icon in the dashboard. In this specific interface user able check how the customer's voice based suspicious level changed over the time. Other than viewing analyzed details user cannot perform any other action in this interface.

Alert Popup

This is the most important part of the module. Alerts of this kind will be shown in the pre-selected personnel devices. The voiced analyzed details can be seen via the alert box by clicking more details button. The Contacts button will navigate the user to a screen which can be used to contact personnel which he desires. All the contact details can be pre-configured by the developers.

3.1.2 Hardware interfaces

Microphone Setup.

This proposed microphone setup will be creating with three microphones parallelly placed with the counter as in the (Error! Reference source not found.**).** All three desktop microphones (**Error! Reference source not found.**) connected to USB hub (**Error! Reference source not found.)** And the USB hub connected to PC. As in

Figure 4:Desktop Microphones

Figure 5: Lavalier Microphones

Figure 6: Multi USB Hub

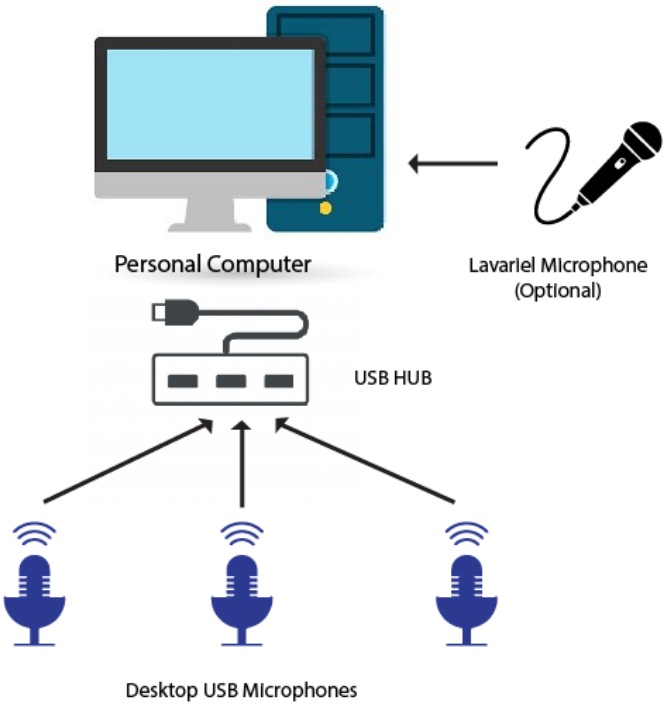

Figure 7. Microphone setup high-level diagram

3.1.3 Software interfaces

For this application all main functions handling by ML models. TensorFlow is the main ML library we are using for this component. Python is the main programing language to use for developments.

End application is based on ASP.Net framework-based C# application. MS SQL server is the database management solution.

3.1.4 Communication interfaces

In this application end user administrator UI are also able access through internet. Because of the requirement this application require communication devices such as router and a modem. Without the communication device main functionality will work perfectly.

3.2 Architectural Design <For Embedded System related Projects >

3.2.1 High level Architectural Design

Main system is to analyze the human facial expressions, behavioral patterns, voice patterns, background environment and human interaction with the environment and identify the possible fraudulent anomalies.

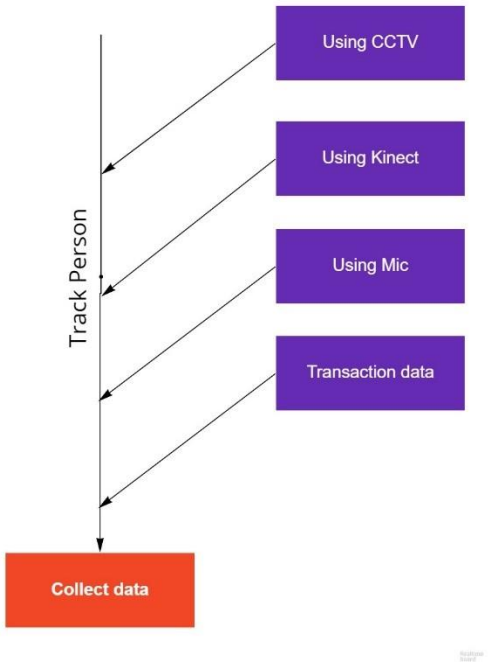

Figure 8: Data collecting process

Like human able identify frauds, lies, deceits or scams by another person's voice or speaking patterns This product is able to identify lies and fraudulent activities. Mainly there are two types of methods to identify frauds using the voice. Those types are, Language Based & Sounds Based. Since language-based type depend on specific language it cannot be implemented for Sri Lanka, because identifying meaning of Sinhala speeches are quite bigger domain. So, the target is to identifies deceits with sounds based.

Sound based frauds detecting also focused on key points and rules based (Based on psychological studies) Key points of sound-based frauds detecting are changes of breathing, changes voice frequencies, stammer, stutter, frequently pausing, repeating,

volume and the tone of voice, vocal expression of emotions (screaming, yelling, whining and crying) etc.

Since this product will be working on crowded environment before analyzing the sound clip we have to preprocess the sound clip by cancelling the background noises and identify isolated customer voice with fine details of the sound. Then trained neural network model able to analyze the sound clip and able to detect lies, this part of the solution can state a prediction for suspiciousness of certain customer by their voice. Overall flow shown in Figure 9: Voice Capture

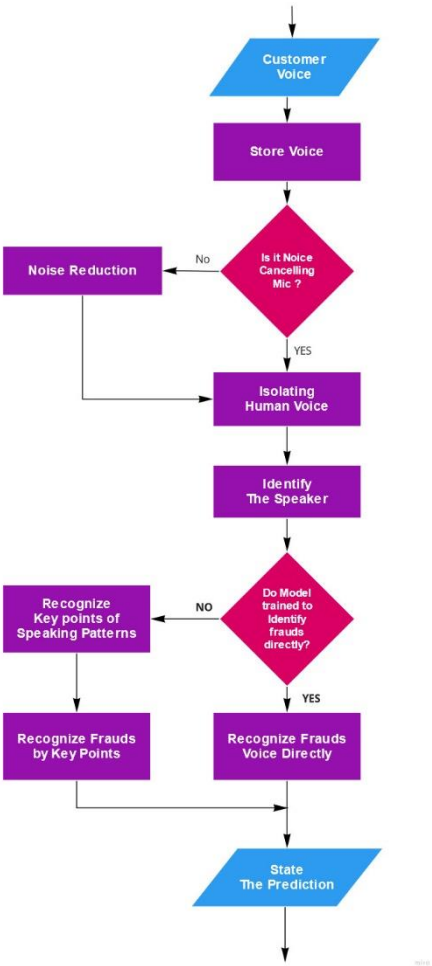

Figure 9: Voice Capture

3.2.2 Hardware and software requirements with justification

3.2.3 Cost Benefit Analysis for the proposed solution

There will be costs for hardware devices which is needed to identify customers and their various expressions. But this hardware will be crucial for the research because they are a must for this research. So, the cost of them is negligible compared to the benefit we get.

There are lot of cost benefits of using an automated system to detect the frauds,

Since this system will learn from its' past mistakes this after sometime this will very adaptive to the environment that this is deployed in and since this is an automated process it'll run 24/7 regardless of any interference.

The human error will be minimum and not susceptible to bribes there will be somewhat initial cost if the site does not have an already working CCTV camera system but it'll be more cost efficient that hiring agents to watch over the footage.

Can connect between multiple sites if there is a WAN connection between sites so we can communicate between them to identify large scale frauds.

3.3 Performance requirements

3.4 Design constraints

Using the industry best practices may make the experience more pleasant for the users. Navigation, UI designs is important to be looked at. Because the bank personnel will not be IT professionals, this machine learning model-based solution should be introduced as easy to use through a user-friendly web application. So, this application should be developed as if any person can use it irrelevant of their literacy levels.

3.5 Software system attributes

3.5.1 Reliability

Reliability is factor we cannot guarantee until the product is finalize. Since this kind product never has done before cannot state reliability of this product. Also

3.5.2 Availability

This system mostly works on LAN network. Because of that this system availability has high state of availability, Bank systems also runs on LAN network our system will not break down unless LAN network breakdown.

3.5.3 Security

Since most of detecting algorithms stored on ML model these algorithms cannot be bypass or access by other third parties. Final results are securely stored on MS SQL server. Only administrator allow to access the previous results.

3.5.4 Maintainability

For this project we have to only update the ML model to increase accuracy of this system. Other than that part there are no other maintainability requirement for this system.

4 REFERENCES

[1] E. Wikipedia, "Polygraph," [Online]. Available: https://en.wikipedia.org/wiki/Polygraph. [Accessed 8 3 2019].

[2] X.-V. Team, "X13-VSA Voice Lie Detector," [Online]. Available: https://lie-detection.com/. [Accessed 8 3 2019].

[3] M. I. R. lab, "MIR-1K Dataset," [Online]. Available: https://sites.google.com/site/unvoicedsoundseparation/mir-1k.

[4] L. F. L. W. M. F. J. G. F. D. S. P. N. L. D. V. Z. John S. Garofolo, "TIMIT Acoustic-Phonetic Continuous Speech Corpus," [Online]. Available: https://catalog.ldc.upenn.edu/LDC93S1. [Accessed 8 March 2018].

[5] GitHub, "Speaker Recoganition," [Online]. Available: https://github.com/topics/speaker-recognition. [Accessed 8 March 2018].

[6] M. Rubinstein, "Looking to Listen: Stand-up," 11 April 2018. [Online]. Available: https://www.youtube.com/watch?v=NzZDnRni-8A. [Accessed 8 March 2019].