

Development of a Voice Chatbot for Payment Using Amazon Lex Service with Eyowo as the Payment Platform

Isaac Samuel¹, Fiyinfoba A. Ogunkeye²

Department of Electrical and Information Engineering, Covenant University, Ota, Nigeria

isaac.samuel@covenantuniversity.edu.ng

fiyinfoba.ogunkeye@stu.cu.edu.ng

Ayobami Olajube³, Ayokunle Awelewa⁴

Department of Electrical and Information Engineering, Covenant University, Ota, Nigeria

ayobami.olajube@covenantuniversity.edu.ng

ayokunle.awelewa@covenantuniversity.edu.ng

Abstract— Engaging in financial transactions has remained a hassle for the visually impaired due to the lack of technological products to facilitate their financial independence and inclusion. Automated teller machines (ATMs) and online banking applications do not provide any means through which the blind can engage in transactions without the need of a third party in managing their finances. This study aims at building a voice chatbot device that can be used for payment using Amazon Lex Service with Eyowo as the payment platform. The chatbot is built by leveraging on Artificial Intelligence (AI) technologies in the form of a service called Amazon Lex for configuring the bot with utterances and responses and Lambda Functions to validate the responses while carrying out the transactions by calling the Eyowo Application Programming Interface (API). A Raspberry pi single board computer is utilised as the medium of communication between the user and the chatbot. The Raspberry pi runs a script that collects voice input through a Universal Serial Bus (USB) microphone connected to it, which is sent to the Amazon Lex to be processed using Automatic Speech Recognition (ASR) and Natural Language Understanding (NLU). Then the chatbot sends back a suitable response to the user through the speaker connected to the Raspberry pi. The device brings comfort and security to the visually impaired by providing balance and cash transfers.

Keywords— Voice chatbot, Artificial Intelligence, Amazon Lex, Voice payments, Raspberry pi

I. INTRODUCTION

Artificial Intelligence (AI) has brought about some of today's most important technological advances. The emergence in AI and interaction of people with each other and the software used to connect to people has changed the way people communicate by providing an opportunity for people to talk directly to software, thus bringing forth the emergence of chatbots. Chatbots are computer programs that perform human-to-machine conversation in the accomplishment and fulfilment of a task using either auditory or textual interfaces depending on the preferred input modality [1].

Text-based chatbots can perform various tasks ranging from booking a hotel reservation to paying a bill, thus making life easier for consumers [2]. The use of text as the input modality is a limitation

since a keyboard is required at every point whenever a user (the blind) wants to communicate to the bot [3]. Voice chatbots are considered more natural and efficient as it does not just make use of the most primary form of communication of humans (speech/voice). Still, it will be providing a more integrated customer experience with immediate verbal feedback, which will allow users to multitask with ease. The use of voice chatbots adds a certain level of trust from the user to the service being used [4], thus increasing the usability and continuous operation of the service to customers [5].

Voice chatbots allow users to perform the task much quicker than typing [6]. The number of tasks and applications of voice chatbots is limitless as it can be used in almost every sector, such as payment, health, commerce, and marketing. The advancement and integration of AI have brought new opportunities for payment services and platforms to create new methods of carrying out transactions [7]. Engaging in financial transactions using voice technology is now being explored by banks and other payment platforms. As PayPal pioneered the use of voice-activated payment systems in 2016, other banks and payment services have followed suit.

With the advent of digital technology in developed nations, developing countries such as Nigeria have failed to implement these digital services in its banking sector, thereby denying and depriving blind customers the opportunity to engage independently with their finances. However, the use of voice chatbot systems stir-up questions bordering on the security measures put in place either by the use of fingerprint (biometric) authentication, a four-digit pin or by a biometric voiceprint [8], to ensure risk is reduced, and cases of fraud are prevented thereby improving the effectiveness and efficiency of the system. Improving on the current level of authentication in chatbots systems will push its adoption rate significantly higher [9].

II. MATERIALS AND METHODS

The voice chatbot system will provide an alternative means of engaging in financial transactions just by speaking, thereby fostering innovations in the Technology industry, most especially with AI in line with the United Nations Sustainable Development Goal of Industry, innovation, and infrastructure (9). This will be done by making use of two Amazon services, namely: Amazon Lex and AWS Lambda Function, alongside the use of a payment platform called Eyowo, in resolving all financial matters.

A. AMAZON LEX

Amazon Lex is a service from Amazon Web Services (AWS) for the development of conversational interfaces called chatbots for any application using voice and text. It is designed to allow a user to build a chat interface inside the Amazon Lex console. Lex has many functionalities and features which it has to undergo before a conversation can take place. It makes use of deep learning technologies like Automatic Speech Recognition (ASR) for converting speech to text, and Natural Language Understanding (NLU) assists the bot to understand and figure out what the intentions of the user are. These technologies work together to create a Speech-Language Understanding system that helps in ensuring the applications built are dynamic and as flexible as possible. Speech recognition and comprehension of natural languages are some of the most challenging problems to address in software engineering, requiring sophisticated deep learning algorithms to be trained in vast quantities of data and infrastructure. With Amazon Lex, time spent and energy used in building the chatbot is optimised due to the incorporation of these technologies in a single service.

1) Automatic Speech Recognition (ASR)

Automatic speech recognition (also known as Speech recognition or "speech-to-text") is the mechanism by which algorithms incorporated in a system or a machine are used to translate voice signals into its corresponding sequence of words or other linguistic entities [10].

A speech recogniser is what is responsible for transforming the audio input stream into the form of a text. The databases are the acoustic model, the lexicon, and the language model [11]:

Fig.1 below shows the speech recognition process.

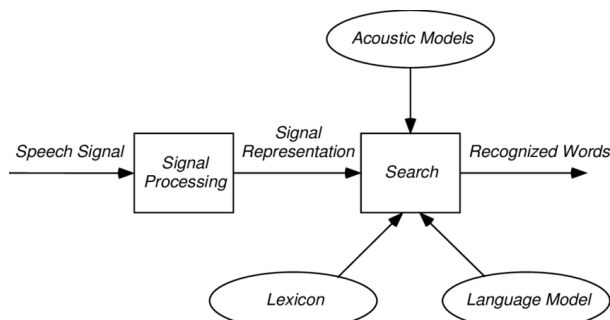


Fig. 1: Speech-to-Text Process

2) Natural Language Understanding (NLU)

NLU is a branch of Natural Language Processing. It is the understanding a computer has gained on the human language that will enable users to interact using natural sentences. NLU goes beyond recognition to determine a user's intent making use of machine learning models that can learn to identify language patterns, syntax, context, and purpose of the user. It is an artificial intelligence that recognises and understands users' text or speech.

Whatever methodology is employed, most NLU systems share specific similar components. The program includes a language lexicon and rules for parser and grammar to separate the sentences to an internal representation. It is done by translating the derived definition into a series of statements and then using testable prediction for assumptions to be drawn.

3) Speech Synthesis

Speech synthesis (also known as "text-to-speech") is the method of getting a voice output out of a given text through artificial means. While other systems turn symbolic semantic interpretations into

speech, a speech synthesis system converts any text input into a voice output. A speech synthesiser can be incorporated for this purpose by using it in software or hardware products. It is possible to generate synthesised speech by concatenating the different parts of captured voice recordings, which are held in a database. A naturally occurring speech quality is produced by a synthesiser that can be implemented into the vocal tract and other human voice characteristics.

The quality of a speech synthesiser is determined by its ability to be well understood and its resemblance to the human voice. Blind people and the ones with dyslexia can use coherent text-to-speech programs to listen to written words on a computer.

B. PAYMENT PLATFORMS

The advent of payment platforms in Nigeria has brought about significant improvement in the performance of businesses and organisations through their use in e-payments and e-commerce [12]. There are various payment platforms available in Nigeria today, such as Interswitch, eTranzact, Paypal, Eyowo, just to mention a few, and these platforms are built with APIs to enable easy integration with businesses through websites or mobile applications.

Eyowo payment platform allows users to make or receive secure payments just by using a mobile phone number. It makes use of a six (6)-digit pin as a form of authentication from users to the bank to verify the transaction to be proceeded with and completed. In this work, Eyowo will be used as the payment platform to facilitate any financial transaction to be carried out by the chatbot.

C. SYSTEM DESIGN

The design of the Voice chatbot system using the Raspberry pi single-board computer is mostly software-based. All the hardware components are ready-made peripherals that are directly attached to the Raspberry pi and configured for specific purposes. Fig. 2 below shows the block diagram of the hardware setup of the system.

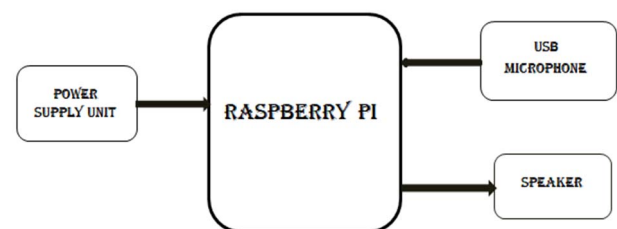


Fig. 2: Systems Block Diagram

1) Power Supply Unit

A micro Universal Serial Bus (USB) supply powers the Raspberry pi 3 through the micro USB port on the side of the Pi. The power supply from the micro USB device provides constant DC voltage to the Pi. The recommended input voltage is 5V, and the current input current is 2A. The micro USB power supply could be the Raspberry pi universal power supply or the standard USB power cords used by many smartphones today, which can be used in line with a power bank of appropriate rating. The standard power supply used for the Raspberry pi is within the range of 4.75V-5.25V (5V \pm 5%) @ 2.5A. The reason for the slight increase in voltage is to negate any voltage drop triggered by the excessive current drawn.

2) Raspberry Pi

The Raspberry Pi is a small credit-card-sized single-board computer that consists of a processor, Random Access Memory (RAM), and

networking ports for interfacing devices. The Raspberry Pi is a low-priced computer that runs on Raspbian OS based on Linux. It has 40 digital input-output pins, of which 26 are GPIO, eight pins are allocated to ground, four pins are assigned to voltage supplies, and the remaining two pins are reserved for Hardware attached on top. The Raspberry pi has an inbuilt operating system but requires a Secure Digital (SD) card for booting and long term storage. It is the core of the whole project, capable of performing multiple tasks at a time. For the implementation of this project, the Raspberry pi 3 model B of the Raspberry pi series will be used.

3) USB Microphone

The USB microphone is used as an input device for collecting voice input from the user. It comes with an inbuilt sound card and can run comfortably off the power supplied by the raspberry pi USB port. The USB microphone can be used directly without any particular configuration, but the USB microphone must be confirmed as ready for use with the raspberry pi, the following steps should be followed;

- i. Open the terminal of the Raspberry pi
- ii. Type the command "lsusb" to check the USB connection
- iii. Type the command "arecord-l" to confirm that the microphone has been accepted as a recording device.
- iv. Test the recording for five seconds by using the command "arecord -D plughw:1,0 -d 5 demo.wav".
- v. To play the recording, use the command "aplay demo.wav".

4) Digital Audio Controller (DAC)

DAC is used to run the speakers directly from the raspberry pi without the need for an additional power source. For the course of this thesis, the Adafruit I2S 3W Stereo bonnet will be used as the DAC. The stereo bonnet has a 2 * 20 header with two terminal blocks. After the header is soldered onto the Printed Circuit Board, it can be attached to the raspberry pi via the GPIO pins on the raspberry pi.

5) Speakers

The speaker is connected to the raspberry pi to give back output response from the bot to the user. The Adafruit Stereo Enclosed Speaker Set -3W 4ohm is used by attaching it to the solderless terminal in the middle of the DAC. The support for I2S DAC must be installed on the Raspberry pi using the command "curl -sS https://raw.githubusercontent.com/Raspberry-Pi-Installer-Scripts/master/i2samp.sh | bash" to enable the speakers and the DAC to work with the Raspberry pi, after which a system reboot is required'.

D. SYSTEM OPERATION

The proposed system is built using a raspberry pi connected to a USB microphone and a speaker to serve as the medium through which the user interacts with the chatbot when trying to engage in any financial transaction. The chatbot makes use of an Amazon Web Service called Amazon Lex to create and configure the bot that responds after it is spoken to through the use of some AI technologies such as NLU, Speech Recognition and Speech Synthesis. The chatbot also makes use of another Amazon Web Service called Lambda Functions. Lambda helps to ensure that all validations at each point of accepting input from the user are appropriately responded to, and it also resolves all payment transactions that the bot will carry out by calling the Eyowo API.

Fig. 3 presents the flowchart of the system, which starts by using the Raspberry pi 3 single-board computer with the aid of a USB microphone to collect the voice input from the user. The Raspberry pi runs a nodejs script that listens for a speech input from the user then sends that input to the Amazon Lex service. The users' voice

input will be converted to text by using the speech-to-text technology offered by Amazon Lex. The chatbot processes the text using NLU to determine the intent of the user that will assist it in providing a suitable response based on what was previously configured on the Amazon Web Console. Speech Synthesis is then used to convert the text response into a voice format that will be sent back to the user through a speaker connected to the Raspberry pi. The back and forth conversation between the user and the bot continues until the user can make the payment transaction or engage in any transaction of the users' choice. At this point, the conversation state is termed to be fulfilled.

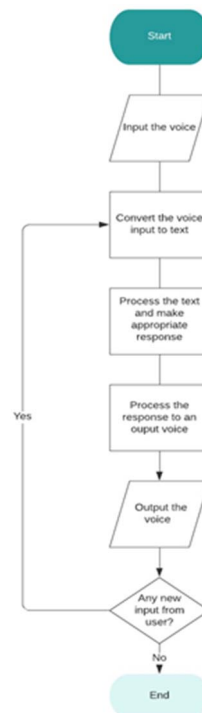


Fig. 3 Flowchart describing the voice chatbot process

III. SYSTEM IMPLEMENTATION

The voice chatbot system consists of a Raspberry pi, a speaker connected to a DAC, and a USB microphone. This device is packaged by designing a case that will contain all the components mentioned above. This design is carried out with the aid of an application called Solidworks. By the use of appropriate measurements, Fig. 4 and Fig. 5 below shows the design made using the Solidworks application.

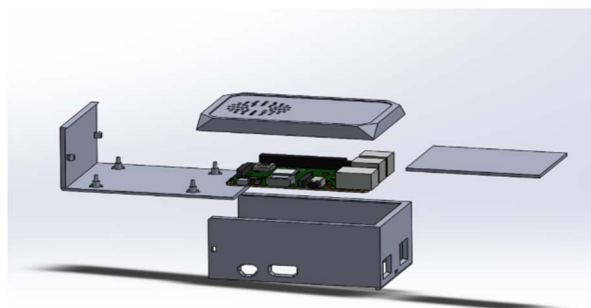


Fig. 4: Exploded view of the casing

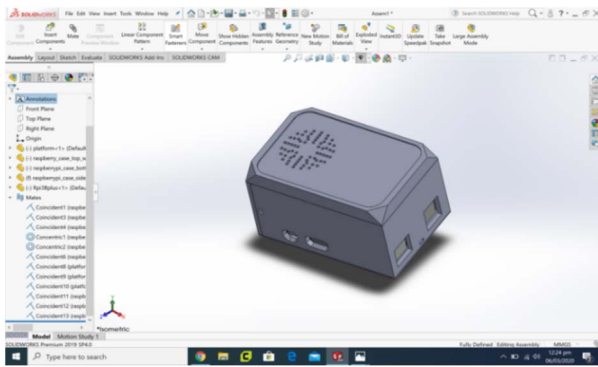


Fig. 5: Closed view of the casing with the speakers enclosed

A. PROGRAMMING THE RASPBERRY PI

The Raspberry pi was programmed via an Integrated Development Environment (IDE) using the Nodejs language. Some packages and libraries were installed to get the code running, for each component to interface with the Raspberry pi properly. Also, the Lambda function had to be correctly configured. This configuration requires the nodejs code to be written alongside the installation of specific dependencies needed for the system to function and for the payment transaction to be a success. These dependencies include:

- Axios: It is a library for making external http requests which are used to call the Eyowo API. It is installed with the command "npm install axios".
- Paystack API: It helps to resolve the users' account name as the Eyowo API is unable to do this. It is installed with the command "npm install paystack".
- Dotenv: It is a library used to load environment variables from a .env file. These environment variables include the eyowo secret keys and the paystack secret keys. It is installed from the terminal with the command "npm install dotenv".

IV. SYSTEM IMPLEMENTATION

The chatbot is built to perform two primary financial services:

- To get a users' account balance
- To make a bank transfer

Before the voice chatbot is tested using the raspberry pi, a mock test is carried out using the test editor provided by amazon in the console as shown in Fig. 6

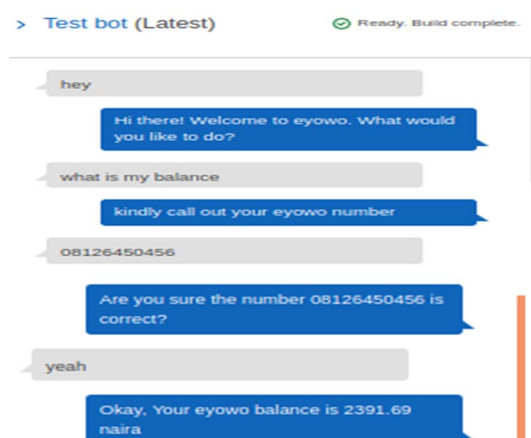


Fig. 6: Mock test showing chatbot giving a users' account balance

After a successful mock test, the voice chatbot is tested using the Raspberry pi. The Raspberry Pi is powered up. After the boot process is complete, the nodejs script for integration with Amazon Lex is executed. After a couple of back and forth interactions between the user and the bot, the transcript, and logs of the results obtained are as follows:

CONVERSATION A: User interacts with the bot to get account balance

User: what is my balance?

Bot: please, what is your eyowo number?

User: 08126450456

Bot: Are you sure the number, 08126450456 is correct?

User: yes

Bot: okay. Your eyowo balance is 2391.69 naira.

As seen in the conversation above, the user successfully gets his account balance which correlates to the conversation logs shown in Fig. 7 below.

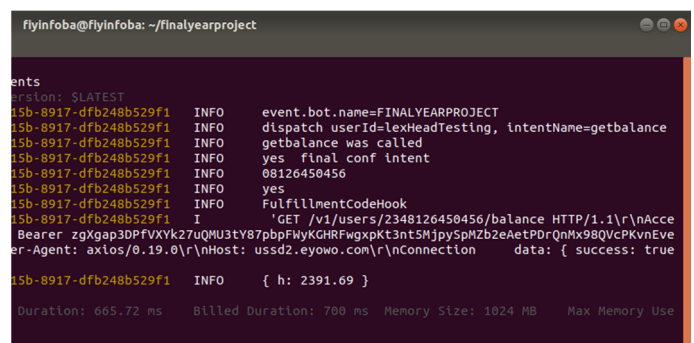


Fig. 7: Logs showing the account balance success

CONVERSATION B: User interacts with the bot to make a bank transfer

User: send 500 to my brothers' bank account

Bot: kindly call out your eyowo number

User: 08126450456

Bot: Are you sure the number, 08176450456 is correct?

User: no

Bot: kindly call out your phone number again

User: 08126450456

Bot: Are you sure the number, 08126450456 is correct?

User: yes

Bot: what bank does your brother use?

User: access bank

Bot: Are you sure your brother bank is access bank?

User: yes

Bot: kindly call out your brother account number

User: 0720357187

Bot: Are you sure the number, 03120357187 is correct?

User: no

Bot: please call out the account number again

User: 0720357187

Bot: Are you sure the number, 0720357187 is correct?

User: yes

Bot: are you sure you want to send 500 naira to your bro?

User: yes

Bot: kindly input your six-digit eyowo pin
 User: 219616
 Bot: Are you sure the pin is 219616?
 User: yes
 Bot: you are about to send 500 naira to your brother. Should I proceed with this?
 User: yes
 Bot: Great! You have successfully sent 500 naira.

As seen in the conversation above, the user successfully makes a bank transfer.

The transaction is confirmed successful through the receipt in Fig. 8 below, which shows a debit was made to the user while crediting the users' brother.

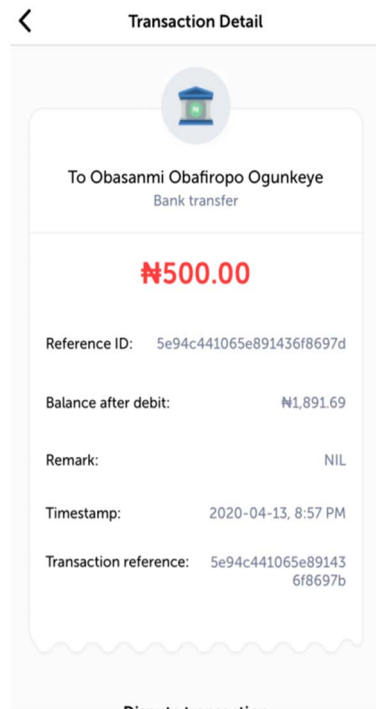


Fig. 8: Receipt showing debit made to the user

CONVERSATION C: User interacts with the bot to get account balance

User: what is my balance?
 Bot: please, what is your eyowo number?
 User: 08126450456
 Bot: Are you sure the number, 08226450456 is correct?
 User: no
 Bot: please call out the eyowo number again
 User: 08126450456
 Bot: Are you sure the number, 08126450456 is correct?
 User: yes
 Bot: okay. Your eyowo balance is 1880.94 naira.

As seen in the conversation above, the user successfully gets his account balance. The balance gotten is 500 naira less than the

previous balance gotten in Conversation A. This confirms that the transfer made in Conversation B was a success.

V. CONCLUSION

This work would bring comfort and security to not just the visually impaired, but the nation as a whole. The voice chatbot system will be used as an alternative form of payment whereby users can engage in financial transactions such as sending of money just by speaking. It would also reduce the dependence on using ATM cards for transactions or visiting ATMs to make transactions.

According to the results obtained from the tests carried out, it can be seen that implementing the voice chatbot device is beneficial in numerous ways. The system can further be improved by incorporating unique voice prints for each user in place of a six-digit pin when making payments or transactions. By doing this, the device would be able to detect fluctuations in frequency between different voices, thereby improving its security. Accommodation of other Nigerian languages besides English such as Yoruba, Igbo, and Hausa can also be implemented to make it more diverse and user friendly as any user would be able to speak comfortably in their mother tongue. Furthermore, to prepare for a rounded system, more services can be incorporated into the voice chatbot system, such as buying airtime and data, bill payments, and cardless withdrawal.

ACKNOWLEDGMENT

The authors would like to express special thanks and gratitude to Covenant University for her support.

REFERENCES

- [1] J. Zamora, "I'm Sorry, Dave, I'm Afraid I Can't Do That: Chatbot Perception and Expectations," in *Proceedings of the 5th International Conference on Human Agent Interaction*, 2017, pp. 253-260.
- [2] A. Vandana¹, D. J. Rani, V. A. Goud, M. C. K. K. Reddy, and B. V. R. Murthy, "A STUDY OF CHATBOTS THROUGH ARTIFICIAL INTELLIGENCE."
- [3] P. A. Angga, W. E. Fachri, A. Eleanita, and R. D. Agushinta, "Design of chatbot with 3D avatar, voice interface, and facial expression," in *2015 International Conference on Science in Information Technology (ICSITech)*, 2015, pp. 326-330.
- [4] R. Burri, "Improving user trust towards conversational chatbot interfaces with voice output," ed, 2018.
- [5] L. J. Hoffman, K. Lawson-Jenkins, and J. Blum, "Trust beyond security: an expanded trust model," *Communications of the ACM*, vol. 49, pp. 95-101, 2006.
- [6] M.-H. Nguyen. (2017, The latest market research,trends and landscape in the growing AI Chatbot Industry. Available: <https://www.businessinsider.de/chatbot-market-stats-trends-size-ecosystem-research-2017-10/?r=US&IR=T>
- [7] A. Fawthrop, "Voice payments set to drive a new era of commerce for Amazon Pay," 2019.
- [8] "Payments pioneer launches Voice Pay transactions," *Biometric Technology Today*, vol. 15, pp. 2-3, 2007/05/01/ 2007.
- [9] "The Rise of Voice Payment Technology in Banking," ed, 2019.
- [10] M. Anusuya and S. Katti, "Speech Recognition by Machine, A Review," *International Journal of Computer Science and Information Security*, vol. 6, 01/13 2010.
- [11] Microsoft, *How Speech Recognition Works (Microsoft.Speech)*, 2012.
- [12] M. A. Tehrani, A. A. Amidian, J. Mohammadi, and H. Rabiee, "A survey of system platforms for mobile payment," in *2010 International Conference on Management of e-Commerce and e-Government*, 2010, pp. 376-381.