Designing Talking ATM System for People with Visual Impairments

Damodar Magdum[1]*, Tejaswini Patil[2], Dr. Maloji Suman[3], Dr.T.B. Mohite Patil[4]

[1]*Research Scholar, Koneru Lakshmaiah Education Foundation, Vaddeswaram (A.P., INDIA).*
[2]*Principal Technical Officer, GIST, C-DAC, Pune (M.S., INDIA).*
[3]*Professor, Koneru Lakshmaiah Education Foundation, Vaddeswaram (A.P., INDIA).*
[4]*Principal, Dr.D.Y.P.C.O.E.SN, Kolhapur (M.S., INDIA).*
*Email: damodarm@cdac.in

Abstract

In this paper, we discuss process of design and development of talking ATM for visually impaired people. Automated Teller Machine (ATM) has become vital part of our life to perform financial transactions without intervention of human banker. ATM facilitates cash withdrawal, balance check, mini statement and fund transfer. But, these banking services using ATM cannot be directly used by some set of people of society such as people with low vision, visually impaired, illiterate as lack of accessing ATM through screens. Even they can be defrauded at ATM centers. To digitally include these set of people, talking ATMs are evolved. Talking ATM provides accessibility to ATM services by providing audio component. Many ATMs employ headphone jack that facilitates user to do transaction with security. The audio information is generated either using pre-recorded speech corpus or through speech synthesis engine. The paper summarizes how ATM works, need, proposed solution of talking ATM for visually impaired users, design and development talking ATM using concatenated Text To Speech.

Keywords: Automated Teller Machine; Visually Impaired; Talking ATM; Text To Speech.

1. Introduction

ATMs are cash machines which enables bank customers to withdraw cash from electronic device without any human interaction i.e. cashier. Most of the ATM's allow cash withdrawal from machine not belonging to banks where customer does not have account through interbank networking. When a customer swipes card in a machine, account information stored in magnetic strip on card is read by card reader in machine (see Fig.1). Then machine requests for entering pin which is encrypted and send via router to base transmission station (BTS). BTS transmits request to host processor linked with bank terminal. Host processor verifies pin with bank information. If customer requests for cash withdrawal then request is forwarded to host processor through networks for approval. The network demands bank to verify customer's account for balance through private network among them. After verification, e-fund is transferred from bank to host processor. In turn host processor sends approval code to ATM. An electronic eye counts currency and money comes out of the cash dispenser. Thus, ATMs are facilitating easy access to banking sector but for users with low vision, visually impaired, people who are dyslexic, people who can't read and people not speaking English are not able to use the ATM services like normal users. In India, Reserve Bank of India (RBI) circulars and the Indian Banks' Association (IBA) procedural guidelines on inclusive banking have established a strong basis for Inclusive Banking for Disabled Persons, particularly Blind and Low-Vision customers [1]. This circular lays down standards of what is truly accessible ATMs. It mentions accessibility for wheel-chairs, Braille keypads, standardized user interface, IVRS-based menus, language support for Hindi and English, local language can be extended and security by option of screen-off.
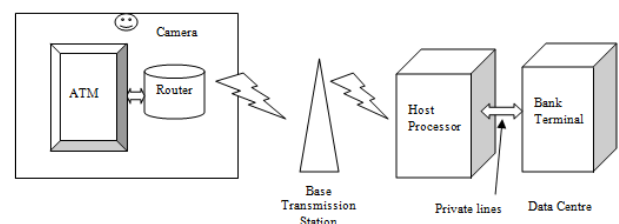

Figure 1: Network Diagram of ATM

The world's first talking ATM for the visually challenged people was an NCR machine unveiled by the Royal Bank of Canada on dated October 22, 1997 [4][14]. In dated 6 June 2012 public sector bank Union Bank of India unveiled India's first ever Talking ATM in Vastrapur, Ahmadabad, Gujarat.[14] Today there are nearly 6000 bilingual accessible ATMs across the country by various banks.

Talking ATM is advancement of existing ATMs providing audio facility to assist users having reading disorders, visually impaired people, and illiterate. Audio facility in Talking ATM can be provided using recordings as well as integrating text to speech system. Okebiro [3] discusses about assistive technologies to be provided to visually impaired people for better services. We are targeting one of them as audio assistance using text to speech. Umar Dahiru [4] suggests to interview bank management for data collection. This is needed as introducing speech as assistive technology in existing ATMs for visually impaired people should maintain consistent interface for visuals as well as for audio. Hence, data collection should consider existing ATMs visual screen messages. Moreover, requirements in perspective of need of visually impaired people must be considered. Hence, discussion with real time users need to be carried out before finalization of audio messages. The paper proposes text to speech for audio generation and instead of developing new application, it can be plugged into existing applications. Understanding of operating system of ATM is required for porting of text to speech. Even, voices used in text to speech should be platform compatible.

Itunuoluwa[11] discusses briefly different synthesis technologies and NLP, DSP module. We are used domain specific synthesis module. In this module we used domain specific sentences, syllables, words and language barakhadi for creating database. V.Manoj[12] discusses Personal Identification Number (PIN) to access user account. The paper proposes voice conversion technique which is based on cloud storage. Delivery of balance statement is based on voice response by using text to speech. We used audio component based on text to speech system. Hence even though message change audio can be generated using TTS system.

2. Challenges

Visually impaired persons has to go bank for cash withdrawal. At ATMs they need assistance to operate machine and hence they can be easily defrauded at ATM centers [2]. ATM centers difficult to locate for physically challenged people. Assistive technologies can empower visually challenged people to access ATMs. Challenge they face are [3]:

- They cannot easily locate ATM and cannot reach it.
- They cannot view ATM screens and text on keypads.
- It's difficult to locate ATM card slot, cash dispenser in the machine.
- Due to inability to see, they face problem to insert the ATM card which they can repeat doing wrongly.
- They struggle to collect their money from the dispenser.
- It is difficult to them to check different transactions, such as balance check, print statement, pin change etc.

To reach ATM centre ramp should be given for wheel chairs. Assistive technologies that can help set of peoples to access ATMs can be braille notations on keypads, near ATM card slot, cash dispenser. Also, for people having low vision screen magnifiers for zoom in zoom out can be given. Speech technologies have to be used for easy access without any human assistance. So, here we are targeting speech prompts i.e. voice aided ATM systems. Thus, speech enabled talking ATM will bring assistance, comfort and security to blind users.

3. Proposed talking ATM solution to digitally include visually impaired people

Many ATMs provide visual component as well audio component. Audio component is employed by headphone jack and set of recorded messages. The paper focuses on providing audio component using concatenated based text to speech system. Use of text to speech will benefit even though messages change, add

and delete, as audio can be generated using TTS at any time. In India ATMs are used by multilingual people. So installing ATM in three languages are considered one local language, Hindi and English. When user will insert headphone in audio jack first of all he/she will be asked to select language preference. Accordingly, screen will be displayed and audio will be played [5]. Integrating audio component using text to speech engine in existing ATMs involves designing and implementing audio component, providing interfaces to existing system which will not alter existing services. On text data collection and speech corpus generation for text to speech system, necessary things are identified as -

- Voice messages should be natural, clear, and with proper speed.
- Speech system should be able to utter all ATM domain words and generic words as well.
- System should able to speak out currency, date, time, phone numbers.
- Currency conversion should be in standard format.
- Able to speak out new messages.

3.1. Enabling Speech

Enabling deployed ATMs with talking feature should have no extra overhead of changing complete software. If user inserts headphone jack in ATM, voice messages should get enabled in ATM. Synchronization of text on screen and audio has to be handled.
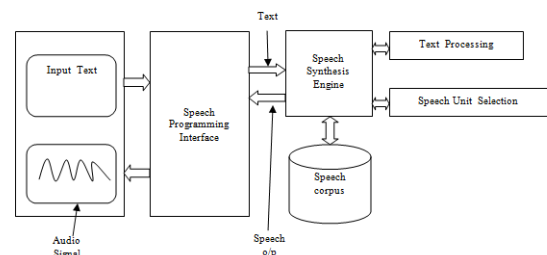

Figure 2: Design of audio component in talking ATM

3.2. Speech application interface

The aim of implementing speech application interface is enabling existing ATMs for speech i.e. integrating audio support without changing present functionality of ATM. There are standard speech interfaces to add speech support to applications. We have implemented Microsoft Speech API (SAPI) 5.1 enabled text to speech. SAPI contains many interfaces and classes for speech recognition and speech synthesis. For Text To Speech, we have overridden following methods:

Get Voices() : Populates SAPI compliant voices. Here, we have our own SAPI complaint voice for different languages.
Speak() : Returns the audio on output stream. Audio is played using this function.
Pause() : pauses audio output.
Resume() : resumes audio output [10]
To give speech enabling feature to application, application has to connect to speech interface. So, standard speech interfaces need to be implemented for accessing voice and TTS engine. Speech interface allows easy integration in application (see Fig.2). We have implemented Microsoft Speech API (SAPI) 5.1 enabled text to speech engine and voice. SAPI enables easy integration.

3.3. Text To Speech (TTS)

Text to Speech takes input text messages and outputs speech signal. Hybrid TTS in financial domain is developed based on concatenation approach for Talking ATM. As TTS engine will be deployed in ATM machine, it should have low memory footprint and high response time. TTS development has taken account of numbers in currency, date format in input text. Audio quality of

TTS output is measured by two parameters naturalness and intelligibility. Recording has been done by voice artist in professional studio. Text data for recording is collected from various banking domain sites [7], cleaned, applied optimization algorithms for minimal recording. Engine is developed in 'C' language to port on any platform.

ATM domain is limited financial domain. Development of TTS for ATM domain should consider low footprint engine, high response time, naturalness and accuracy[15-18]. TTS takes input text and do processing on it. Input text is cleaned, segmented to introduce prosody in audio output. Numbers in currency, date format are normalized and converted into phonetic string. These phonetic strings are mapped into array of speech units and speech units are concatenated, normalized to produce speech output.

3.4. Speech corpus design and development

Speech corpus design is an important step in every TTS system. The basic units of our Text to Speech system are bank domain words. Text data for recording is collected from banking and financial domain websites. Text data collected may contain many irrelevant words like, lengthy words, complicated words, other language words and un-normalized texts, symbols, mathematical notations etc. All irrelevant words should be filtered. If all these words are processed without filtering, then there may be chance that irrelevant words are part of the corpus recording script, and those words are not adding value to the system. Also invalid words are very difficult to pronounce so filtering is very important step [13]. As text corpus size is very huge so manual checking is difficult and time consuming so algorithm are developed to create phonetically balanced optimized words [8][9]. The recording script covers all the phonemes set of the banking domain. The recording script has been finalized using manually reading.
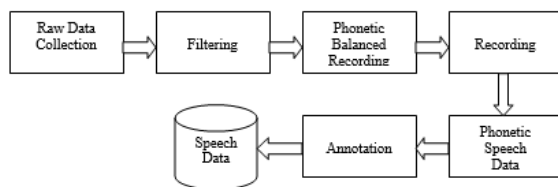

Figure 3: Speech corpus design

The final corrected list are used for recording. After recording text data, sound files are manually checked and cleaned to remove unnecessary sound. The cleaned data is transcribed and used for annotation. In annotation each word boundary is marked properly. The final annotated data is used as speech corpus for TTS (see Fig.3).

Speech corpus is optimized by chunking data into phrases. To keep corpus size small, repeated phrases are removed [6].

Examples:

- अंग्रेजी के लिए 1 दबाएँ । chunked into

  अंग्रेजी | के लिए | 1 | दबाएँ ।

- हिंदी के लिए 2 दबाएँ । chunked into

  हिंदी | के लिए | 2 | दबाएँ ।

- Hence, speech corpus contains words:

  अंग्रेजी | हिंदी | के लिए | 1 | 2| दबाएँ ।

- आपके खाते से 5000 रूपये डेबिट किये गए है। chunked into

  आपके । खाते ।से ।5000। रूपये। डेबिट। किये । गए । है।

- आपके खाते में 8000 रूपये क्रेडिट किये गए है। chunked

into

आपके । खाते । में । 8000। रूपये। क्रेडिट ।किये। गए ।है।

Hence, speech corpus contains words:

आपके | खाते | से | में | डेबिट | क्रेडिट | 5000 | 8000| किये। गए ।है।

Unknown words or words other than domain are also synthesized with hybrid TTS approach. Hybrid TTS uses high frequency words, barakhadi, syllables like units. Bank can integrate new messages in ATM. So audio support for such a messages is given by hybrid TTS. Example of messages are like festival messages, new scheme launches, etc.

4. Expected Result

1. A visually challenged person without any assistance of somebody can use talking ATM for banking transactions independently.
2. On inserting headphone in an audio jack of ATM machine, ATM is enabled in talking mode.
3. ATM is operated similar to IVRS (Interactive Voice Response system) using keypad.
4. To identify button on keypad, audio orientation of keypad buttons is provided. Beep sound is played on keypad button press.
5. All number keys can be located with reference of 5 number key with raised dot.
6. Language selection, volume level selection, screen visibility options and ATM machine orientation is given before start of the ATM operations.
7. Screen text and audio are synchronized.
8. All ATM visual messages are mapped for audio messages to maintain uniformity for all users.
9. Language standards for pronunciations of messages are considered.
10. SAPI interfaces are well implemented.
11. Multilingual audio support is given.
12. Speech corpus of text to speech based on concatenation approach is well designed and developed to generate natural speech output.
13. Multilingual support is given by providing messages in three languages. In English as universal language, in Hindi as official language of India and one local (regional) language. Some sample messages on ATM screens : English, Hindi, Marathi (Local language) are like :

- Your Account is debited with Two thousand rupees on 6th July 2016.
- आपके खाते से 6 जुलाई 2016 को दो हजार रुपए डेबिट किये गए है।
- आपल्या खात्यामधून सहा जुलैला दोन हजार रूपये वजा केले गेले आहेत.
- Your Account is credited with Two thousand rupees 8th July 2016.
- आपके खाते में 8 जुलाई 2016 को दो हजार रुपए क्रेडिट किये गए है।
- आपल्या खात्यामध्ये 8 जुलैला दोन हजार रूपये जमा केले गेले आहेत.

5. Conclusion and Future Work

People with visual impairment or low vision and illiterate people faces barriers in accessing electronic machines. Government is also mandating banking sectors to provide services to them. Moreover, some guidelines are set by Government for standard interfaces on talking ATMs. Thus by installation of such ATMs will provide visual as well as audio access to people. Multilingual support enables access to various language speakers along with

native speakers. Thus, ATM service of cash withdrawal can be in future extended for various services like printing passbooks and depositing cheques.

Acknowledgement

References

[1] Dr. Sam Taraporevala, "Rbi Circular On Accessible Atms 5 Years On, A Retrospect And Prospect" .Author,"Title Of The Paper", *Proceedings Of The Conference Name*, Vol.X, No.X, (200x), Pp:Xx-Yy, Http://Dx.Doi.Org/10.1109/Mmm.2013.2248651

[2] Bátiz-Lazo, B.: Emergence And Evolution Of Atm Networks In The Uk, 1967−2000. Business History 51(1), 1–27 (2009).

[3] Okebiro Jared Omari, Omariba Bosire Zachary, "Investigating Atm System Accessibility For People With Visual Impairments". Iosr Journal Of Computer Engineering Volume 15, Issue 5 (Nov. - Dec. 2013).

[4] Umar Dahiru, Sangeetha Elango, Riyaz Ahamed A.H, "System Design For Talking Atm For The Blind Users". International Journal Of Information System And Engineering, Vol 1, No1,April 2015.

[5] Bátiz-Lazo, B., Reid, R.: Evidence From The Patent Record On The Development Of Cash Dispensers And Atm Technology. In: Ieee History Of Telecommunications Conference, Paris (2008).

[6] Markéta Jůzováand, Daniel Tihelka, "Minimum Text Corpus Selection For Limited Domain Speech Synthesis". International Conference On Text, Speech, And Dialogue, Pp 398-407 (2014).

[7] Https://Www.Sbi.Co.In/Index-Hindi_Latest.Htm.

[8] Alan W Black,Kevin A. Lenzo, "Limited Domain Synthesis ". Unpublished.

[9] Charles T. Hemphill, John J. Godfrey, George R. Doddington, "The Atis Spoken Language Systems Pilot Corpus" Texas Instruments Incorporated.

[10] Https://Msdn.Microsoft.Com/En-Us/Library/Ee125663(V=Vs.85).Aspx.

[11] Itunuoluwa Isewon, Jelili Oyelade, Olufunke Oladipupo, "Design And Implementation Of Text To Speech Conversion For Visually Impaired People" International Journal Of Applied Information Systems (Ijais),Foundation Of Computer Science Fcs, New York, Usa Volume 7 No. 2, April 2014.

[12] V. Manoj Kumar, S.Prathap, B.Gowtham,G.Sriram, "Speaking Atm Using Speech Synthesis" International Journal Of Contemporary Research In Computer Science And Technology (Ijcrcst) Volume 2, Issue 3. March 2016.

[13] Damodar Magdum,Manisha Shukla Dubey, Tejaswini Patil, Ronak Shah, Swapnil Belhe, Mahesh Kulkarni, "Methodology For Designing And Creating Hindi Speech Corpus " Ieee Xplore,Klu,Vijaywada,Spaces-2015.

[14] Https://En.Wikipedia.Org/Wiki/Talking_Atm.

[15] Kishore, P.V.V., Kishore, S.R.C. And Prasad, M.V.D., 2013. Conglomeration Of Hand Shapes And Texture Information For Recognizing Gestures Of Indian Sign Language Using Feed Forward Neural Networks. International Journal Of Engineering And Technology, 5(5), Pp. 3742-3756.

[16] Ramkiran, D.S., Madhav, B.T.P., Prasanth, A.M., Harsha, N.S., Vardhan, V., Avinash, K., Chaitanya, M.N. And Nagasai, U.S., 2015. Novel Compact Asymmetrical Fractal Aperture Notch Band Antenna. Leonardo Electronic Journal Of Practices And Technologies, 14(27), Pp. 1-12.

[17] Karthik, G.V.S., Fathima, S.Y., Rahman, M.Z.U., Ahamed, S.R. And Lay-Ekuakille, A., 2013. Efficient Signal Conditioning Techniques For Brain Activity In Remote Health Monitoring Network. Ieee Sensors Journal, 13(9), Pp. 3273-3283.

[18] Kishore, P.V.V., Prasad, M.V.D., Prasad, C.R. And Rahul, R., 2015. 4-Camera Model For Sign Language Recognition Using Elliptical Fourier Descriptors And Ann, International Conference On Signal Processing And Communication Engineering Systems - Proceedings Of Spaces 2015, In Association With Ieee 2015, Pp. 34-38.