ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 5, Issue 11, November 2017

Application based on Voice as Medicine Guide for Visually Impaired

Ansari Kasaf[1], Ayashabi Mashayak[2], Saima Shaikh[3], Patill S.C[4]

Al-Ameen Educational and Medical Foundation College of Engineering, Koregaon Bhima, Pune, Maharashtra, India[1,2,3]

Associate Professor, Department of Computer Engineering Al-Ameen Engineering College, Koregaon Bhima, Pune, Maharashtra, India[4]

ABSTRACT: This system is an application which helps the visually impaired people to find the medicine and to take it according to their Doctor's prescription. This android application is used when visually impaired people can't find correct medicines or they don't have any idea of current timings so that they can take their medicine. In this application, a reminder is set which tells the user when to take the medicines, as voice output. The pictures of the medicine strip held in the hand are captured by the inbuilt camera of the mobile. The image is processed and consequently text localization and extraction is done by which the name of the medicine is identified. A spotter section is also consolidated with this application which checks the prescription which has been already uploaded in the user's mobile, compares with the name of the medicine identified and if the medicine has to be taken at that time, then it tells the quantity of medicine to be taken to the user as voice output. On receiving the voice output from the mobile, the user intakes their medicines according to their prescription. It can also be useful for uneducated people who suffer to find which medicine must be taken. Especially elder people who are not educated usually suffer to read their medicine names on their own.

KEYWORDS: Image Processing , visually impaired, medicine identification ,Text localization and Extraction, SQLite,TTS,vision api.

I. INTRODUCTION

Among people there are some that are unable to read, either because of blindness (complete or partial), or for other reasons. Therefore, initially for this category of people, there is a need for computer-generated speech, namely the conversion of written texts into acoustic files in Albanian language .In the market, there already are different solutions to synthesize speech from text written in different languages. However, these solutions cannot be used for generating speech in other languages, because for every language, specific algorithms should be used to synthesize speech, since each language is different during the speech. Therefore, for the English language too, there is a need to design a particular algorithm and to create the system to synthesize speech in this language.

From the published works, it is known that to synthesize speech from written text, generally three different methods are used: concatenative synthesis, formant synthesis and articulatory synthesis. Each of these methods has advantages and disadvantages. Quality of synthesis is assessed by naturality andcomprehensibility of the generated speech . Naturality has to do with the likeness of the generated speech to the human speech, while, comprehensibility has to do with the clarity of the generated speech.

The advantage of the proposed system over the social platform systems is that visually impaired users can store only the desired information without having to discard irrelevant information uploaded by other people. Even doctors can keep watch on their health record of the patient so it would be easy to keep patient health record.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 5, Issue 11, November 2017

II. TEXT TO SPEECH(TTS) ALGORITHM

The speech synthesis is artificial generation of human speech from written texts. For this purpose, adequate algorithms are designed, which then through relevant programs make it possible to synthesize texts to speech. The process of converting text into speech is also known as Text-To-Speech(TTS) system [5]. In this paper are given basic principles to be used when designing a system to synthesize speech in English language1 from written texts. Currently there are solutions that enable natural speech generation for various world languages. However, unfortunately these are not universal solutions to be used for other languages too, because the volume generated for other languages is incomprehensible and unnatural. For this reason, for every language one should seek solutions that address the specifics of it, always with the aim of generating voice to suit the nature of language. Generating systems that are currently used mainly rely on the use of the concatenation method [6], during which acoustic segments of text files are joined, which are previously digitized and stored as such in a database. For English language, we consider that on the textual part of the database, as basic segments to be used are: the most frequent words, two-letters and letters [4]. However, in a particular part of the database are included various abbreviations, i.e. textual equivalents and their acoustics files, to be used also during the generation of appropriate speech. Whereas, with the aim of synthesizing the various numerical values written in the decimal system, in database were added values, respectively their corresponding sound files, whereby speech is generated for different numbers. The first part of the paper is a brief presentation of the English language [1], respectively of the alphabet used in writing the language and its most frequent words.

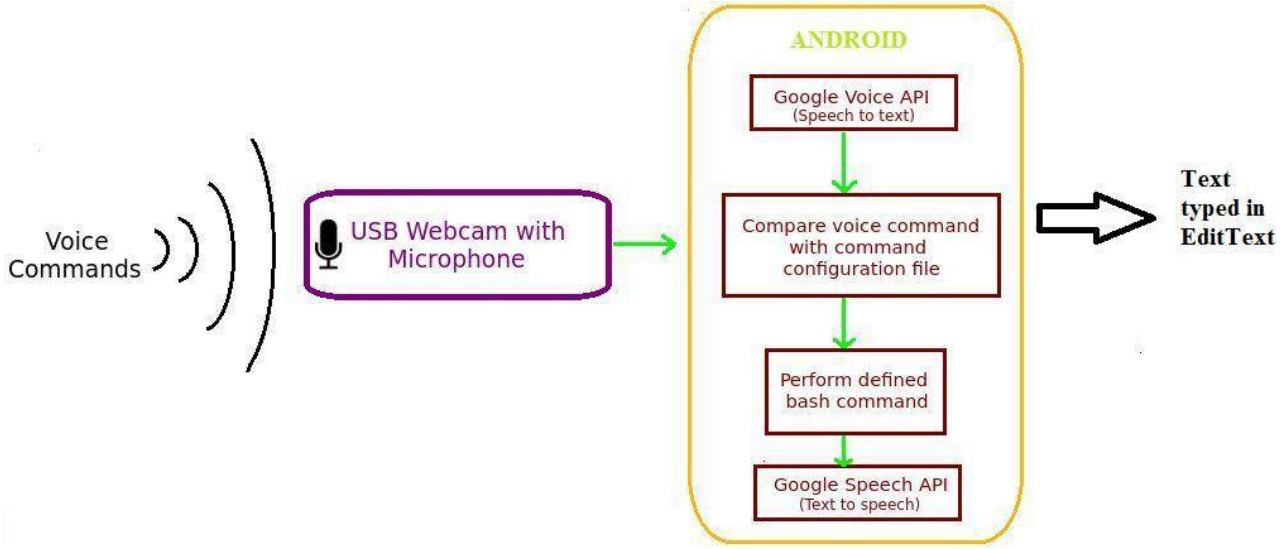

Fig..speech to text structure .

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 5, Issue 11, November 2017

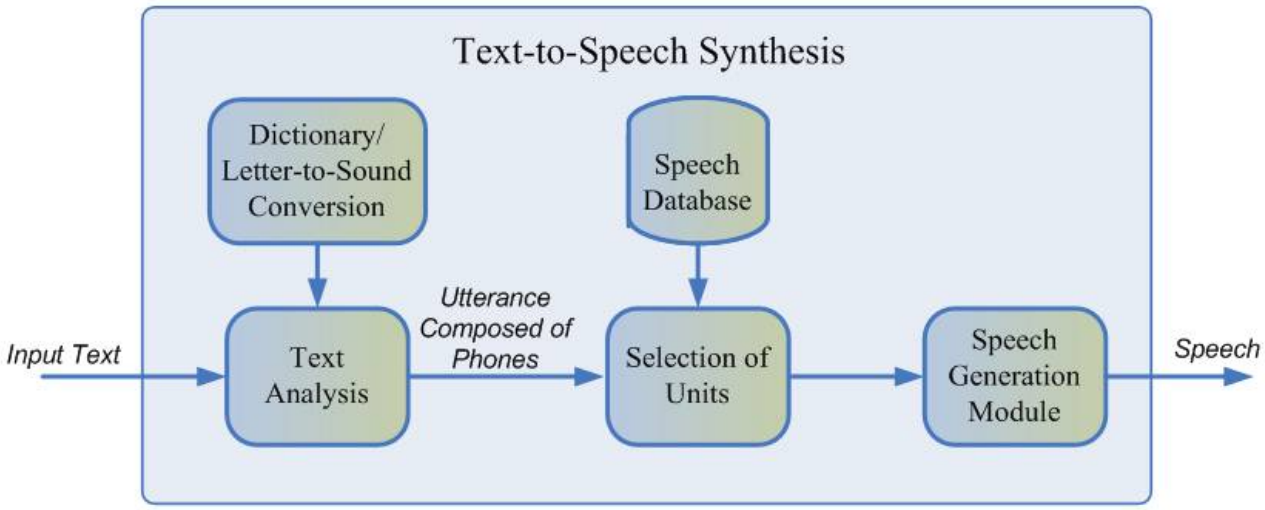

Fig.Text to Speech Synthesis

The first step involves basic operations who adopt the text.
• Initially the whole text is converted to fit so that two-symbols are replaced with single characters. • Abbreviations that are included in the database are replaced with the texts as they are
read/pronounced
. • Numbers are converted to text using words that are stored in the database. In the second step begins the process of replacement of textual segments with corresponding sound files, which will be placed in a final sound file. Here, is applied algorithm, which takes the specific words of the text and examines them under the following procedure.
• If the word is found in the database, the appropriate sound file is taken and added to the final sound file.
• Otherwise, the word is fragmented into two-letter segments and for each segment is taken appropriate sound file and is placed in the final sound file.
• Similarly, is acted with numbers and abbreviations.
• If it happens that any digraph lacks in the database, sound files of separate characters/letters are used.
We are analyzing the Albanian texts in order to find the larger units of text segments composed of two or more words, so that the relevant sound files are used, what will certainly affect the growth of naturality and clarity of the generated speech.

III. VISUALIZING THE FLOW OF DATA

The flow of data in the OCR tutorial application involves several steps:

1. An image is uploaded to Cloud Storage with text in any language (text that appears in the image itself).
2. A Cloud Function is triggered, which uses the Vision API to extract the text, and queues the text to be translated into the configured translation languages.
3. For each queued translation, a Cloud Function is triggered which uses the Translation API to translate the text and queue it to be saved to Cloud Storage.
4. For each translated text, a Cloud Function is triggered which saves the translated text to Cloud Storage.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 5, Issue 11, November 2017

It may help to visualize the steps:

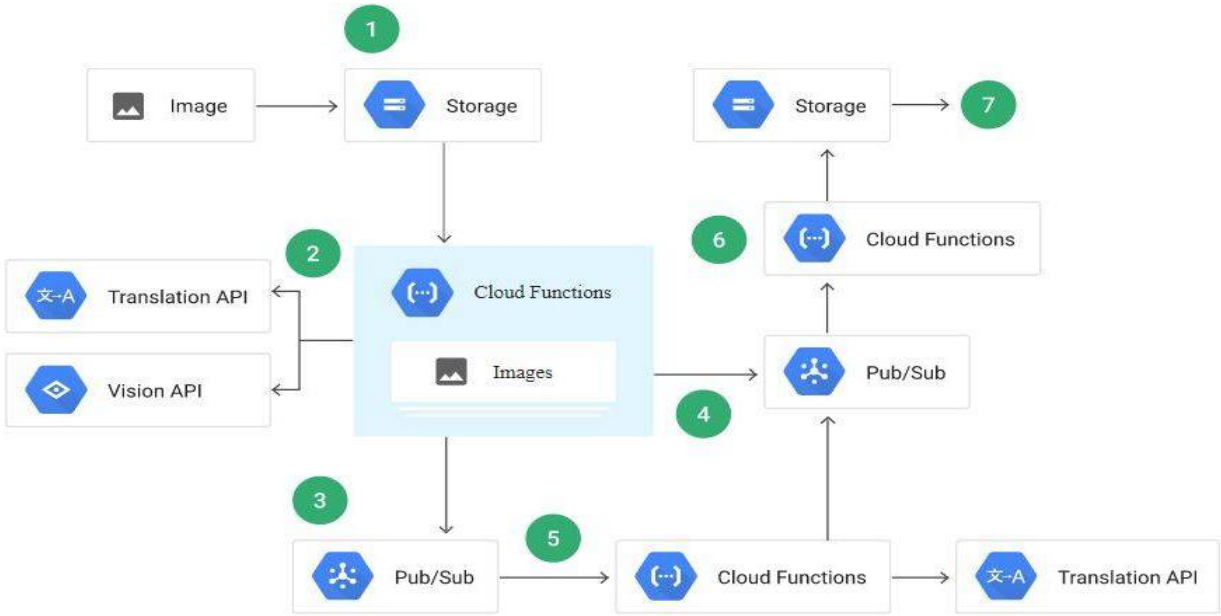

Fig.Data flow of character recognization

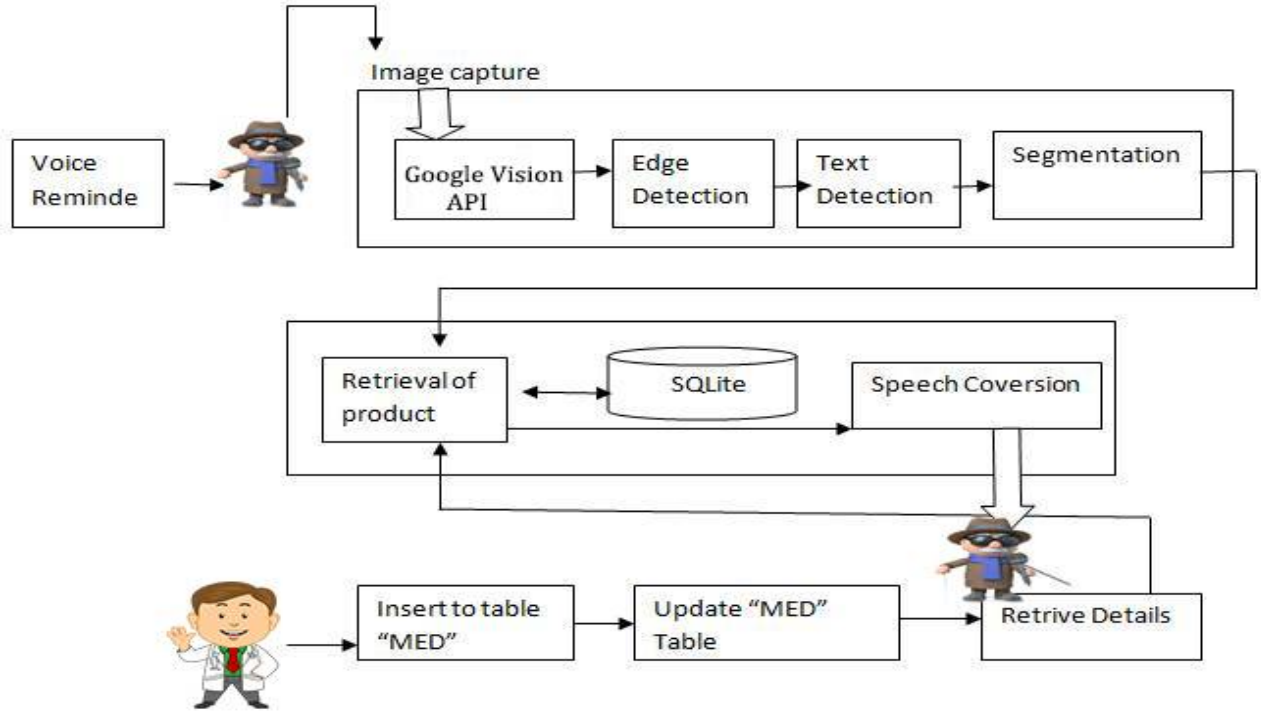

Fig.Architecture of the system

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 5, Issue 11, November 2017

IV. CONCLUSION

The present paper proposed a medicine reminder system to assist visually impaired users in recalling medicine timing and name related to doctor prescription that they have given. The medicine name were recorded as voice memos, which were able to be played back when the user takes that medicine. The medicine text matching throgh wrapper of the tablet or tonic will be done via google vision api and text to speech conversion also. The system was implemented as an application on an Android smartphone and evaluated by two experiments: image matching tests and a user study. The experimental results suggested the effectiveness of the system to help visually impaired individuals, including blind individuals, recall information about regularly taking medicines.

Let us bow to the facts : there is still a long way to HAL, the brilliant talking computer of '2001, a space odyssey'. A number of advances in the area of NLP or DSP, however, have recently boosted up the quality and naturalness of available voices, and this is likely to continue. Important issues need now be further addressed in that purpose. Among others :

How to best account for coarticulatory phenomena ? In the context of concatenation-based synthesis, this question mostly reduces to : how to derive optimized sets of segments from speech data ?

How to best formalize the relationship between syntax, semantics, pragmatics and prosody, and how to derive natural sounding intonation and duration from abstract prosodic patterns ?

A fundamental feature of speech has seldom been taken into consideration by TTS systems : its variability. Prosodic patterns, for instance, are submitted to a particular kind of variability which cannot be confused with randomness in that variations maintain some hidden coherency with each other.

How to account for speaker and speaking style effects ?

Readers willing to have a deeper understanding of the problems mentioned in this paper could advantageously report to the forthcoming [Dutoit 96], which analyses DSP and NLP solutions with much more details. A number of internet sites can also be consulted, some of which propose demo programs and/or speech files.

REFERENCES

1. Portable Camera-Based Assistive Text and Product Label Reading From and Held Objects for Blind Persons Chucai Yi, Student Member, IEEE, YingliTian, Senior Member, IEEE, and Aries Arditi,2014.
2. Mobile Camera Based Text Detection and Translation Derek Ma Department of Electrical Engineering Stanford University, Qiuhau Lin Department of Electrical Engineering Stanford University, Tong Zhang Department of Mechanical Engineering Stanford University.
3. Extraction of Text in images Malik, R. Dept. Of Electr. Eng. & Computer Eng., New Jersey Inst. of Technol., Newark, NJ, USA, 1999.
4. Text Extraction from Document Images Using EdgeInformation Author(s) over, S. Nat. Inst. of Technology., Rourkela, India Arora, K. ; Mitra, S.K,2009.
5. Morphological text extraction from images Author(s): Hasan, Y.M.Y. Dept. Of Electr. Eng., Arizona State Univ., Tempe, AZ, USA Karam, L., 2000.
6. Multiscale Edge-Based Text Extraction from Complex Images Author(s) Xiaoqing Liu Dept. Of Electr. &Comput. Eng., Univ. Of Western Ontario, London, Ont. Samarabandu, J, 2006.
7. Scene text extraction in natural scene images using hierarchical feature combining and verification Author(s):Kim, K.C. Dept. of Comput. Sci., Yonsei Univ., South Korea, 2004.
8. J. Liang and D. Doermann, H.P. Li, "Camera-based analysis of text and documents: a survey," International journal on Document Analysis and Recognition IJDAR) (7), 25, pp.84-104.
9. C.Thillou and B.Gosselin, "Natural Scene Text Understanding,"In Vision Systems: Segmentation and Pattern Recognition, Ed. Croatia: I-Tech education And Publishing, 2007, pp. 307-332.
10. K...Jung, "Text information extraction in images and video: a survey," Pattern Recognition, vol. 37, no. 5, pp.977-997, 2004.
11. A.K. Jain and B. Yu, Automatic Text Location in Images and Video Frames, Pattern Recognition, 31 (1998) 2055- 2076.
12. L. Agnihotri and N. Dimitrova. Text Detection for Video Analysis. In Proc. of the International Conference on Multimedia Computing and Systems, Florence, Italy, pp. 109-113, 1999.
13. H. Li and D. Doermann. Automatic Text Tracking In Digital Videos. In Proc. of IEEE 1998 Workshop on Multimedia Signal Processing, Redondo Beach, California, USA, pp. 21-26, 1998.
14. V.Wu, R. Manmat ha and E. M. Riseman. Finding Text in Images inProc of Second ACM International Conference on Digital Libraries, Philadelphia, PA, pp.2326, 1997.
15. Shaban Demiraj. The Origin of the Albanians. Academyof Sciences of Albania, Tirana, 2008
16. Alan G. Konheim. One-Gram ProbabilityDistribution. Cryptography – A Primer, John Wiley, 1981
17. Agni Dika. Distribucioni i frekuencave të shkronjavedhe grafemave në tekstet e shkruara shqipe. Academy of Sciences and Arts of Kosova, Prishtina, 2006

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 5, Issue 11, November 2017

18. Bahri Beci, Ermira Topi. Vëzhgime për dendurinë epërdorimit të fonemave zanore e bashkëtingëllore në gjuhën e sotme letrare shqipe. Studime Filologjike 1, Tirana, 1986
19. Mentor Hamiti, Agni Dika. Learning opportunitiesthrough generating speech from written texts. In Procedia-Social and Behavioral Sciences Journal, Volume 2, Issue 2, pp. 4319-4324. ELSEVIER, Istanbul, Turkey, 2010, ISBN 1877-0428.
20. Daniel Jurafsky, James Martin. Speech and LanguageProcessing. Pearson International Edition, 2008
21. Agni Dika, Mentor Hamiti. Options for ConvertingAlbanian Written Texts into Speech. Proceedings of the
Fourth Balkan Conference in Informatics, pp. 88-93. A.T.E.I. Thessalonikis Publishing House, Thessaloniki, Greece, 2009. ISBN 978-960-287-127-0.