

MEDICAL TRANSCRIPTION USING SPEECH RECOGNIZER

B.V.S.Mahidhar¹, D.L.Sankeerthana², K.Bharath Reddy³, G.Asha Nikhath⁴
^{1,2,3,4} UG student,

*Department of Computer Science Engineering,
SRM University, Ramapuram, Chennai – 89*

¹mahidharbvs@gmail.com, ²lathasankeerthana96@gmail.com
³bharathreddy20196@gmail.com, ⁴asha212nikath@gmail.com

R. J. Poovaraghan⁵

⁵ Assistant Professor,

*Department of Computer Science Engineering,
SRM University, Ramapuram, Chennai - 89*

Abstract— Speech recognition is a vital part in medical transcriptions. The existing speech recognition systems that run as standalone desktop applications fall short in many cases due to low accuracy rates and high processing time. This paper proposes a novel collaborative approach for the automation of speech to text conversion. The model proposed leverages the power of cloud computing and the unique nature of computing. This framework has many real time applications such as in Medical Transcription systems, IVR systems etc., The generic framework here is advantageous, because the speech models in Automatic Speech Recognizer (ASR) could be trained according to specific domain required, allowing wide usability. The proposed speech framework is used for medical transcription process which is automated by using the proposed speech framework. With this system, an android application was developed which acts as a medium between doctors and their patients. The application helps diagnose the patients using their symptoms. The doctors prescribe medicines through this application to patients for various kinds of illness. It is user friendly and secure as it can be used by anyone round the clock. The entire model is developed for a mobile cloud environment considering the characteristics of cloud delivery models.

Keywords— Speaker Recognition; Sub region Model; Model Synthesis.

1. Introduction

Speech recognition aims to recognize claimed identities of speakers, including identification and verification. It has gained great popularity in a wide range of applications including access control. After decades of research, current speaker recognition systems have achieved rather satisfactory performance.

2. Literature Survey

This model presented here follows the same idea of exploiting phonetic knowledge learned by Automatic Speech Recognizing systems. An advantage of our approach is that strong language models can be applied to offer more accurate phone labels; additionally, the acoustic classes are modeled explicitly in our proposal, which is highly flexible. Regarding the research for SUSR, it has been known that the model possesses some advantages when dealing with short utterances [1]. A widely used speech unit definition in Chinese is based on the Initial/Final structure of syllables, where the initials correspond to consonants, and the finals correspond to vowels [2]. Although language-independent speech units can be defined, e.g., through the International Phonetic Association (IPA) [3] and speech recognition systems [4], [5], language-dependent speech units generally cover the acoustic space in a better way. For the knowledge based approach, we simply follow the definition of speech unit classes described in [6].

3. Existing Method

Generally, Medicare developed to regulate the cover a range of “telehealth” services, such as the use of telecommunications and information technology (IT) to provide access to health assessment, diagnosis, intervention, consultation, supervision and information across distance. The existing speech recognition systems that run as standalone desktop applications fall short in many cases due to low accuracy rates and high processing time. The bottleneck in these systems is the lack of computation power (in terms of processing power and memory) made accessible to them.

3.1.1 Drawbacks of Existing System

- Not be used by all people.
- Not efficient.
- Speech recognition technology is also expensive an setup.
- Single person can be served at a time.

4. Proposed Method

To overcome these issues, novel speech framework proposed for medical transcription process. The patient spells the keyword then the details of the keyword will display and reduce the patient work. This process is automated by using the proposed speech framework. The entire model is developed for a mobile cloud environment considering the characteristics of cloud delivery models. Automatic Speech Recognition is gaining importance in various industries, because it reduces the manual work done by a person. For example, in a Voice Response system, ASR system can be used to automatically provide response to the user according to request.

Advantages of Proposed System

- Flexibility.
- Speed and accuracy.
- Inventory of speech units varies for different languages.

4.1 Architecture

In Architecture diagrams are the primary form of representation; they carry a design from conception to construction. Architectural design is about connections shapes configurations and orientations of physical forms they not only represent physical elements but also forces and flows. Thus arrows, lines and other symbolic representations of forces and flows appear in architectural diagrams conveying spatial characteristics.

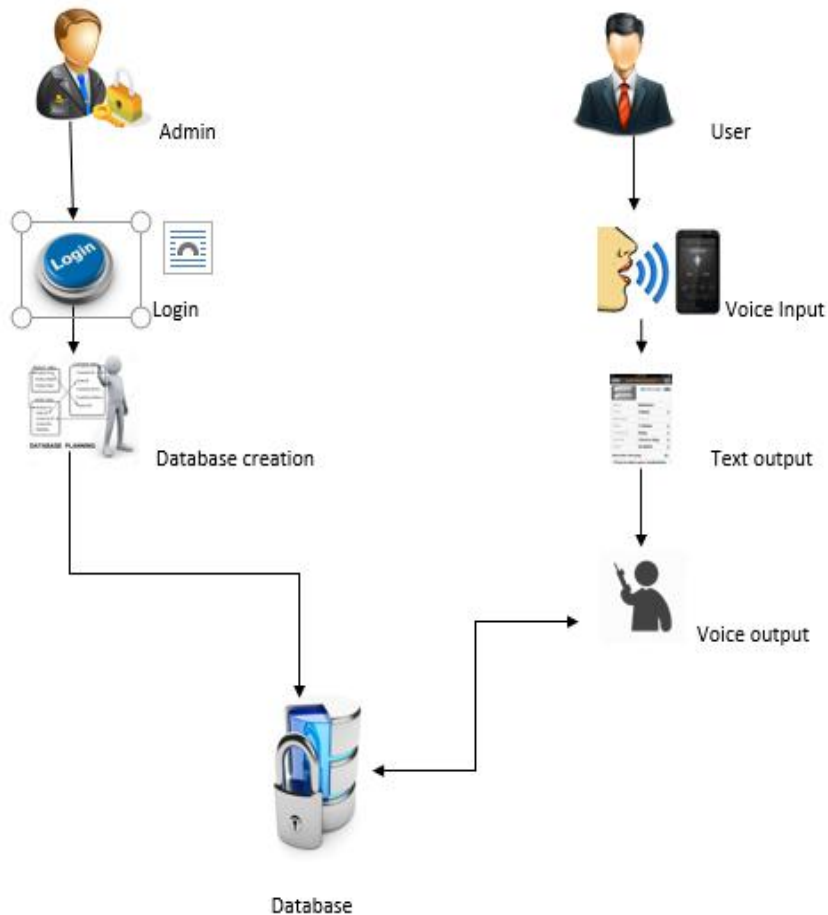


Fig 1 Architecture Diagram

The architecture explains how each module is connected to each other and path of data flow. It helps to clarify, understand and communicate ideas about structure of the system and requirements that must be supported by the system. It's a frame work which is used in system planning phase. It is the process of defining a structured solution that meets all operational and technical requirements while optimizing attributes such as security performance, and manageability.

4.1.1 Admin

Admin has the complete authority to access, update the database periodically and modify the data layout which is available to the user. The admin helps diagnose the user based on the symptoms, age etc., and prescribe them with relevant medications.

4.1.2 Validation

The admin creates a unique username and password to login and access the database. The username and password can consist of alphabets, numbers and special characters. Only the admin can login to database.

4.1.3 Database Creation

The admin creates the database using SQL Helper. He has the authority to enter, delete and update information into the database.

4.1.4 Database

Database is the collection of information that is organized so that it can be easily accessed. Database is organized in rows and columns called tables.

4.1.4 User

The user does not have to register separately to access the application. This makes it user friendly. The user enters his symptoms and age to diagnose him. He can do this process multiple times. The user can either type in the information or speak them out into the mic.

4.1.5 Voice Input- Text / Voice Output:

A speech recognizer is used to convert the voice input given by the user into a text format. The resulted prescription is in a text format which can also be converted into a voice output.

4.2 Modules:

Four modules present in the system:

- Text to Speech Engine
- Speech to Text Engine
- Patient
- Doctor



4.2.1 Doctor

Doctor will login to the system and view the list of patient details and read the disease and symptoms information. Then doctor prescribe the medicine details, update in the database which is sent to the patient via the text to speech engine.

4.2.2 Patient

After choosing their age group, the patient enters their symptoms. They can either type in their symptoms or speak it into the mic. Then analysis of symptoms is done and a prescription with relevant medication is shown in the form of text. It can also be converted to speech.

4.3.3 Text to Speech Engine

Here normal text is converted to voice message which can be heard using speakers. This module is used by patient to convert the text in prescription into speech. Text to speech conversion engine is used here.

4.3.4 Speech to Text Engine

Here speech is converted to plain text format. This engine is used by the patient to convert the symptoms which are given in the form of voice input that are converted into text. Speech to text conversion engine is used here.

5. Conclusion

Further, project can be enhanced by involving a few features. Multiple admins can login into the database. Medication and diagnosis can be done for rare diseases. Automatic Speech Recognizer can be updated with multiple languages to make it more users friendly.

References

- [1] A. Kanagasundaram, R. Vogt, D. B. Dean, S. Sridharan, and M. W. Mason, “i-vector based speaker recognition on short utterances,” in Proceedings of the 12th Annual Conference of the International Speech Communication Association. International Speech Communication Association (ISCA), 2011, pp. 2341–2344.
- [2] J.-Y. Zhang, T. F. Zheng, J. Li, C.-H. Luo, and G.-L. Zhang, “Improved context-dependent acoustic modeling for continuous chinese speech recognition.” in Proc. INTERSPEECH’01, 2001, pp. 1617–1620.
- [3] I. P. Association, Handbook of the International Phonetic Association: A guide to the use of the International Phonetic Alphabet. Cambridge University, Press, 1999.



[4] T. Schultz and A. Waibel, "Language-independent and language adaptive acoustic modeling for speech recognition," *Speech Communication*, vol. 35, no. 1, pp. 31–s51, 2001.

[5] A. Colomé, "Lexical activation in bilinguals' speech production: Language-specific or language-independent?" *Journal of memory and language*, vol. 45, no. 4, pp. 721–736, 2001.

[6] N. Fatima, X.-J. Wu, T. F. Zheng, C.-H. Zhang, and G. Wang, "A universal phoneme-set based language independent short utterance speaker recognition," in *11th National Conference on Man-Machine Speech Communication (NCMMSC'11)*, Xi'an, China, 2011, pp. 16– 18.