Kannada speech recognition enquiry system for farmers

Smithashree K P,Manoj C, Madhurya M, Keerthana S, Shruthi K R
maharaja institute of technology, Mysore
affiliated to VTU, Belagavi
(E-mail: manoj1997mys@gmail.com)

***Abstract*—**India is an agricultural country. In order to provide a stimulus to the agricultural growth, the Ministry of Agriculture, Government of India set up many websites which displays information regarding problems and solutions. But most of the farmer are far behind in technology to use website. In this project, we describe the process of acquiring speech data for training an ASR system for Kannada language that will form the core of a voice interface to the webpage providing information about crops as a voice. We also make use of Text to Speech engine to convert the information to voice. NVP algorithm is used to make the voice conversion more human less robotic. Provision will be provided to display the information on a screen using a website This would need implementation of a voice interface to the website using Automatic Speech Recognition (ASR) and Text to Speech Synthesis (TTS) technologies. A revolutionary technology that converts digital text to speech, text to speech offers benefits to the illiterate farmers to solve their agricultural queries. The text to speech converts the digital text into speech using text to speech convertor. This software is good news for all people is that they no longer need to read the text to understand it. Instead, they can listen to it to understand.

Keywords*—speech recognition; kannada speech; speech conversion; automatic speech recognition; text to speech synthesis;*

I. INTRODUCTION

Language processing refers to the way human beings use words to communicate ideas and feelings, and how such communications are processed and understood. Most of the information such as agricultural related queries, news, and weather report in the internet are all available only in selected number of languages especially in English. But people who know only local languages may not be able to use these resources though the resources are available. Kannada language is a morphologically rich language and translation between a language with simple morphology like English and a language with complex morphology like Kannada is generally complex task. Speech recognition is primarily used to convert spoken words into computer text. The human needs to train the speech recognition system by storing speech patterns and vocabulary of into the system. Speech recognition systems convert speech from a recorded audio signal to text. Speech recognition software works by breaking down the audio of speech recording into individual sounds, analyzing each sound, using natural language processing to find the most probable word fit in that language, and transcribing those sounds into text. A revolutionary technology that converts digital text to speech, text to speech offers benefits to the illiterate farmers to solve their agricultural queries. The text to speech converts the digital text into speech using text to speech convertor. This software is good news for all people is that they no longer need to read the text to understand it. Instead, they can listen to it to understand.

II. EASE OF USE

A. Reach the illiterate farmers:

Without the assistance of any other people, illiterate farmers can get any information about agriculture related problems. the farmer can listen to that solution and can solve his agricultural related problems. This software will be very useful for illiterate farmers.

B. Customer retention:

The software provides the potential for personalized services based on customer preferences, and there by the potential to improve the customer experience. The Software assists the user without the need for access to information by keyboard or touch tone button pushes.

III. METHODDOLOGY

The Methodology to be followed for the project is as follows:

1)Concatenative speech synthesis techniques will be used in order to get the naturalness quality in the synthetic speech.

2)The phonemes of the Kannada language can be used as the basic unit for speech synthesis

3)To design a user interface in which the end user can write editable Hindi text to be converted into speech in the text box

4)Speech database for Kannada Language will be developed by using phoneme.

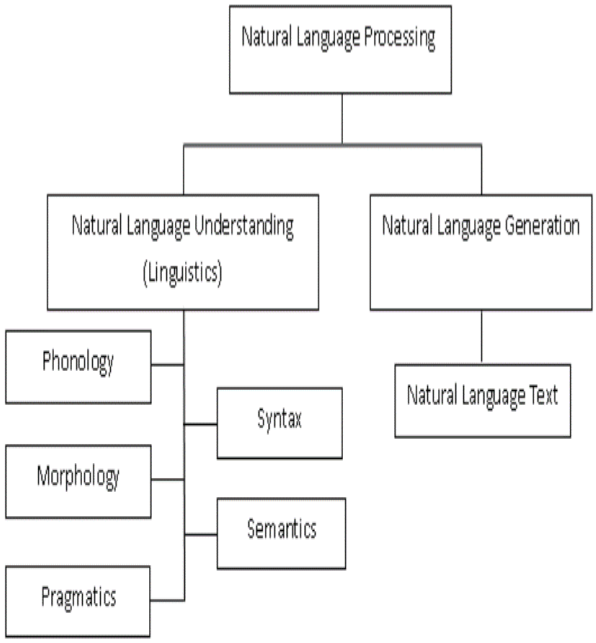
5)The input voice will be separated into Kannada Phoneme.

6)Phonemes will be searched in the database and corresponding phonemes sounds will be concatenated to generate synthesized output speech.

Natural Language Processing (NLP) is a tract of Artificial Intelligence and Linguistics, devoted to make computers understand the statements or words written in human languages. Natural language processing came into existence to ease the user's work and to satisfy the wish to communicate with the computer in natural language. Since all the users may

not be well-versed in machine specific language, NLP caters those users who do not have enough time to learn new languages or get perfection in it.

A language can be defined as a set of rules or set of symbols. Symbol are combined and used for conveying information or broadcasting the information. Symbols are tyrannized by the Rules. Natural Language Processing basically can be classified into two parts i.e. Natural Language Understanding and Natural Language Generation which evolves the task to understand and generate the text.



A. Phases of NLP architecture

Linguistic is the science which involves meaning of language, language context and various forms of the language. The various important terminologies of Natural Language Processing are

- 1. Phonology

Phonology is the part of Linguistics which refers to the systematic arrangement of sound. The term phonology comes from Ancient Greek and the term phono- which means voice or sound, and the suffix –logy refers to word or speech. In 1993 Nikolai Trubetzkoy stated that Phonology is "the study of sound pertaining to the system of language". Whereas Lass in 1998 wrote that phonology refers broadly with the sounds of language, concerned with the to lathe sub discipline of linguistics, whereas it could be explained as, "phonology proper is concerned with the function, behavior and organization of sounds as linguistic items. Phonology include semantic use of sound to encode meaning of any Human language.

- 2. Morphology

The different parts of the word represent the smallest units of meaning known as Morphemes. Morphology which comprise of Nature of words, are initiated by morphemes. An example of Morpheme could be, the word precancellation can be morphologically scrutinized into three separate morphemes: the prefix pre, the root cancella, and the suffix -tion. The interpretation of morpheme stays same across all the words, just to understand the meaning humans can break any unknown word into morphemes. For example, adding the suffix –ed to a verb, conveys that the action of the verb took place in the past. The words that cannot be divided and have meaning by themselves are called Lexical morpheme (e.g.: table, chair) The words (e.g. -ed, -ing, -est, -ly, -ful) that are combined with the lexical morpheme are known as Grammatical morphemes (eg. Worked, Consulting, Smallest, Likely, Use). Those grammatical morphemes that occurs in combination called bound morphemes (eg. -ed, -ing) Grammatical morphemes can be divided into bound morphemes and derivational morphemes.

- 3. Lexical

In Lexical, humans, as well as NLP systems, interpret the meaning of individual words. Sundry types of processing bestow to word-level understanding – the first of these being a part-of-speech tag to each word. In this processing, words that can act as more than one part-of-speech are assigned the most probable part-of speech tag based on the context in which they occur. At the lexical level, Semantic representations can be replaced by the words that have one meaning. In NLP system, the nature of the representation varies according to the semantic theory deployed.

- 4. Syntactic

This level emphasis to scrutinize the words in a sentence so as to uncover the grammatical structure of the sentence. Both grammar and parser are required in this level. The output of this level of processing is representation of the sentence that divulge the structural dependency relationships between the words. There are various grammars that can be impeded, and which in twirl, whack the option of a parser. Not all NLP applications require a full parse of sentences, therefore the abide challenges in parsing of prepositional phrase attachment and conjunction audit no longer impede that plea for which phrasal and clausal dependencies are adequate. Syntax conveys meaning in most languages because order and dependency contribute to connotation. For example, the two sentences: 'The cat chased the mouse.' and 'The mouse chased the cat.' differ only in terms of syntax, yet convey quite different meanings.

- 5. Semantic

In semantic most people think that meaning is determined, however, this is not it is all the levels that bestow to meaning. Semantic processing determines the possible meanings of a sentence by pivoting on the interactions among word-level meanings in the sentence. This level of processing can incorporate the semantic

disambiguation of words with multiple senses; in a cognate way to how syntactic disambiguation of words that can errand as multiple parts-of-speech is adroit at the syntactic level. For example, amongst other meanings, 'file' as a noun can mean either a binder for gathering papers, or a tool to form one's fingernails, or a line of individuals in a queue. The semantic level scrutinizes words for their dictionary elucidation, but also for the elucidation they derive from the milieu of the sentence. Semantics milieu that most words have more than one elucidation but that we can spot the appropriate one by looking at the rest of the sentence.

- 6. Discourse

While syntax and semantics travail with sentence-length units, the discourse level of NLP travail with units of text longer than a sentence i.e., it does not interpret multi sentence texts as just sequence sentences, apiece of which can be elucidated singly. Rather, discourse focuses on the properties of the text as a whole that convey meaning by making connections between component sentences. The two of the most common levels are Anaphora Resolution - Anaphora resolution is the replacing of words such as pronouns, which are semantically stranded, with the pertinent entity to which they refer. Discourse/Text Structure Recognition - Discourse/text structure recognition sways the functions of sentences in the text, which, in turn, adds to the meaningful representation of the text.

- 7. Pragmatic:

Pragmatic is concerned with the firm use of language in situations and utilizes nub over and above the nub of the text for understanding the goal and to explain how extra meaning is read into texts without literally being encoded in them. This requisite much world knowledge, including the understanding of intentions, plans, and goals. For example, the following two sentences need aspiration of the anaphoric term 'they', but this aspiration requires pragmatic or world knowledge.

IV. SYSTEM ARCHITECTURE

Natural Language Generation (NLG) is the process of producing phrases, sentences and paragraphs that are meaningful from an internal representation. It is a part of Natural Language Processing and happens in four phases: identifying the goals, planning on how goals maybe achieved by evaluating the situation and available communicative sources and realizing the plans as a text. It is opposite to Understanding.

Components of NLG are as follows:

Speaker and Generator – To generate a text we need to have a speaker or an application and a generator or a program that renders the application's intentions into fluent phrase relevant to the situation.

Components and Levels of Representation -The process of language generation involves the following interweaved tasks. Content selection: Information should be selected and included in the set. Depending on how this information is parsed into representational units, parts of the units may have to be removed while some others may be added by default. Textual Organization: The information must be textually organized according the grammar, it must be ordered both sequentially and in terms of linguistic relations like modifications. Linguistic Resources: To support the information's realization, linguistic resources must be chosen. In the end these resources will come down to choices of particular words, idioms, syntactic constructs etc. Realization: The selected and organized resources must be realized as an actual text or voice output.

Application or Speaker – This is only for maintaining the model of the situation. Here the speaker just initiates the process doesn't take part in the language generation. It stores the history, structures the content that is potentially relevant and deploys a representation of what it actually knows. All these forms the situation, while selecting subset of propositions that speaker has. The only requirement is the speaker has to make sense of the situation.

V. RESULTS AND DISCUSSIONS

we are conceptualizing a voice recognising portal for kannada speaking farmers which can show the information about crops. The benefits of such system can be reaped if farmers are able to access this information easily. Presently, illiteracy, ignorance, lack of knowledge of English, computers and internet are acting as hurdles in accessing the information from the website. On the other hand, if this information can be made available just by speaking over a mic it would foster a large number of farmers. This would need implementation of a voice interface to the website using Automatic Speech Recognition (ASR) and Text to Speech Synthesis (TTS) technologies.

VI. CONCLUSION

For speech acquisition we make use of Mic which will be interface with the microprocessor. When the farmer speaks to the mic the analogue signal of his voice will be converted into digital signal and will be processed for speech recognition by ASR engine and keyword will be matched with the database. When keyword is matched MP will fetch the data from the database which will be converted into speech by TTS engine. This speech will be relayed to speaker so that the farmer will be able to hear it. At the same time this data will be displayed as text on a webpage. Speech data acquisition is the first step towards building speech recognition system. The accuracy of recognition depends on the speech data used to train the system. This paper describes the measures taken for collecting apt speech data from Kannada Speaking farmers, to develop a robust speech recognition system as a part of voice interface for agricultural information retrieval.

ACKNOWLEDGMENT

We are indeed grateful to many groups of people who have helped us with various aspects of this study. We want to thank Ms. Smithashree K P, Assistant professor, Dept. of ISE MIT, Mysore for guiding us. Her knowledge and experience about various analytical techniques and ongoing trends influenced us in overcoming many hurdles.

REFERENCES

[1] Mrs. S. D. Suryawanshi, Mrs. R. R. Itkarkar, Mr. D. T. Mane, "High Quality Text to Speech Synthesizer using Phonetic Integration" , International Journal of Advanced Research in Electronics and Communication Engineering (IJARECE) Volume 3, Issue 2, February 2014.

[2] J. Sangeetha,S. Jothilakshmi, S. Sindhuja, V. Ramalingam, "Text to Speech synthesis system for Tamil", International Conference on Information Systems and Computing (ICISC-2013), INDIA.

[3] Jia Member, IEEE, Shen Zhang, Fanbo Meng, Yongxin Wang, and Lianhong Cai, Member, IEEE," Emotional Audio-Visual Speech Synthesis Based on PAD", IEEE transactions on audio, speech, and language processing, vol. 19, no. 3, march 2011.

[4] Maninder Singh1, Karun Verma2," Text to Speech Synthesis for numerals into Punjabi language" International Journal of Computational Linguistics and Natural Language Processing Vol 2 Issue 7 July 2013 ISSN 2279 – 0756.

[5] n.swetha 2k..anuradha," Text-to-speech conversion", International Journal of Advanced Trends in Computer Science and Engineering, Vol.2 , No.6, Pages : 269-278 (2013).

[6] Swati Ahlawat, Rajiv Dahiya," A Novel Approach of Text to Speech Conversion Under Android Environment", IJCSMS International Journal of Computer Science & Management Studies, Vol. 13, Issue 05, July 2013.

[7] D.Sasirekha, E.Chandra," Text to Speech: A Simple Tutorial", International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307, Volume-2, Issue-1, March 2012.

[8] Miss. Priyanka V. Mhamunkar, Mr. Krishna S. Bansode and Prof. Laxman S. Naik," Android Application to get Word Meaning through Voice", International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) Volume 2, Issue 2, February 2013.

[9] H. Hon, A. Acero, X. Huang, J. Liu, and M. Plumpe," Automatic Generation of Synthesis Units for Trainable Text-to-Speech Systems".

[10]Lakshmi Sahu and Avinash Dhole,"Hindi &Telugu textto-Speech Synthesis(TTS) and inter-language text Conversion", International Journal of Scientific and Research Publications, Volume 2, Issue 4, April 2012 1 ISSN 2250-3153.

MANOJ C