# REVIEW AND PERSPECTIVES OF NATURAL LANGUAGE PROCESSING FOR SPEECH RECOGNITION

*Khin Myat Nwe Win [1], Zar Zar Hnin [2], Yin Myo Kay Khing Thaw [3]*

[1]*Faculty of Computer Science, University of Computer Studies (Mandalay), Mandalay and 05041, Myanmar*
[2]*Faculty of Computer Science, University of Computer Studies (Mandalay), Mandalay and 05041, Myanmar*
[3]*Information Technology Supporting and Maintenance Department, University of Computer Studies (Mandalay), Mandalay and 05041, Myanmar*

## ABSTRACT

***Speech Recognition system is the application of Natural Language Processing(NLP), a major area Artificial Intelligence(AI) research, that identify words and phrases in spoken language and convert them to a machine readable format and explores how computers can be used to understand and manipulates natural language speech to do useful. Speech recognition and NLP are usually used together in Automatic Speech Recognition engines that takes acoustic signal as an input and then it tries to determine which words were actually spoken, voice assistants and speech analytics tools that simply the ability of a software to recognize speech. Recently, the focus in AI applications in NLP was on knowledge representation, logical reasoning, and constraint satisfaction - first applied to semantics and later to the grammar. The paper distinguished review of NLP components followed by presenting the history and methods of NLP, state of the business perspectives of Natural Language Processing for Speech Recognition.***

***Keyword: Natural language Processing (NLP), Speech Recognition, Artificial Intelligence(AI)***

## 1. INTRODUCTION

Natural Language Processing (NLP) focuses on system development that allows computers to communicate with people using everyday language. It is a collection of techniques used to extract grammatical structure and meaning from input in order to perform a useful task as a result, natural language generation builds output based on the rules of the target language and the task at hand. NLP is useful in the tutoring systems, duplicate detection, computer supported instruction and database interface fields as it provides a pathway for increased interactivity and productivity. [1] Natural language generation system converts information from computer database into readable human language and vice versa. The applications of Natural language processing include fields of study, such as machine translation, natural language text processing and summarization, user interfaces, multilingual and cross language information retrieval (CLIR), speech recognition, artificial intelligence(AI) and expert systems. It uses the context free grammars for representing syntax of that language presents a means of dealing with spontaneous through the spotlighting addition of automatic summarization including indexing, which extracts the gist of the speech transcriptions in order to deal with Information retrieval and dialogue system issues. Some well-known application areas of NLP are Optical Character Recognition (OCR), Speech Recognition, Machine Translation, and Chatbots.

Speech recognition is an interdisciplinary subfield of computational linguistics that develops methodologies and technologies that enables the recognition and translation of spoken language into text by computers. It is also known as automatic speech recognition (ASR), computer speech recognition or speech to text (STT). When you talk, ASR takes acoustic signal as an input and then it tries to determine which words were actually spoken. The output typically consists of a word graph – a lattice that consists of word hypotheses. [3]

Speech recognition technology can be used to perform an action based on the instructions defined by the human. The human needs to train the speech recognition system by storing speech patterns and vocabulary of their language into the system. By doing so, they can essentially train the system to understand them when they speak. It incorporates knowledge and research in the linguistics, computer science, and electrical engineering fields. Anything that a person says, in a language of their choice, must be recognized by the software. The five phases of NLP involve lexical (structure) analysis, parsing, semantic analysis, discourse integration, and pragmatic analysis. [2]

## 2. METHODS OF NATURAL LANGUAGE PROCESSING

### *2.1. Natural language processing for Speech Synthesis:*

TTS synthesis makes use of NLP techniques extensively since text data is first input into the system and thus it must be processed in the first place. Text Normalization Adapts the input text so as to be synthesized. It contemplates the aspects that are normally taken for granted when reading a text. The sentence segmentation can be achieved though dealing with punctuation marks with a simple decision tree. [1]

The text-to-speech (TTS) synthesis is to convert an arbitrary input text into intelligible and natural sounding speech. TTS system includes mainly two parts: natural language processing and digital signal processing. The general block diagram of TTS system is shown in figure 1. Natural language processing contains three steps. They are text analysis, phonetic analysis and prosodic analysis. The text analysis includes segmentation, text normalization, and part of speech (POS) tagger. Phonetic conversion is to assign phonetic transcription to each word. [2]

There are two approaches in phonetic conversion. They are rule based and dictionary based approaches. Rule based is applied for unknown words whereas dictionary based is used for known words. Prosodic analysis is to determine intonation, amplitude and duration modeling of speech. It describes speaker 's emotion This is based on the text to speech conversion (TTS) in which the text data is the first input into the system. It uses high level modules for speech synthesis. It uses the sentence segmentation which deals with punctuation marks with a simple decision tree. But more confusing situations require more complex methods. Some examples of these difficulties are the period marking, the disambiguation between the capital letters. in proper names and the beginning of sentences, the abbreviations, etc. The tokenization separates the units that build up a piece of text. It normally splits the text of the sentences at white spaces and punctuation marks. This process is successfully accomplished with a parser.

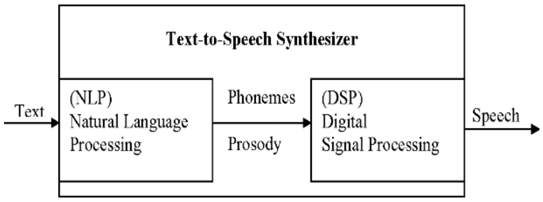

*Fig (1) Text-To-Speech (TTS) synthesizer*

### *2.2. Natural Language Processing for Speech Recognition:*

Automatic Speech Recognition (ASR) system make use of natural language processing techniques based on grammars. It uses the context free grammars for representing syntax of that language presents a means of dealing with spontaneous through the spotlighting addition of automatic summarization including indexing, which extracts the gist of the speech transcriptions in order to deal with Information retrieval and dialogue system issues. A grammar as a set of rules that determine the structure of texts written in a given language by defining its morphology and syntax. ASR takes for granted that the incoming speech utterances must be produced according to this predetermined set of rules established by the grammar of a language, as it happens for a formal language [1]

A context-free grammar (CFG) is a set of recursive rewriting rules (or productions) used to generate patterns of strings, play an important role since they are well capable of representing the syntax of that language while being efficient at the analysis (parsing) of the sentences. For this reason, such language cannot be considered natural. ASR systems assume though that a large enough grammar rule set enable any language to be taken for natural. NLP techniques are of use in ASR when modeling the language or domain of interaction in question. Through the production of an accurate set of rules for the grammar, the structures for the language are defined. These rules can either be hand-crafted or derived from the statistical analyses performed on a labelled corpus of data. The latter is generally the chosen one because of its programming flexibility at the expense of a tradeoff between the complexity of the process, the accuracy of the models and the volume of training and test data available. The literature is extensive on the data driven approaches (N-gram statistics, word lattices, etc.) bearing in mind that by definition a grammar based representation of a language is a subset of a natural language. [1]

Aiming at a flexible enough grammar to generalize the most typical sentences for an application, end up building N-gram language models. [6][7] N-gram models are widely used in statistical natural language processing. In speech recognition, phonemes and sequences of phonemes are modeled using a n-gram distribution. For parsing, words are modeled such that each n-gram is composed of n words. A means of dealing with spontaneous-speech through the spotlighting addition of automatic summarization including indexing, which extracts the gist of the speech transcriptions in order to deal with Information Retrieval (IR) and dialogue system issues. [8]

These are just a few methods of natural language processing. Once the information is extracted from unstructured text using these methods, it can be directly consumed or used in clustering exercises and machine learning models to enhance their accuracy and performance.

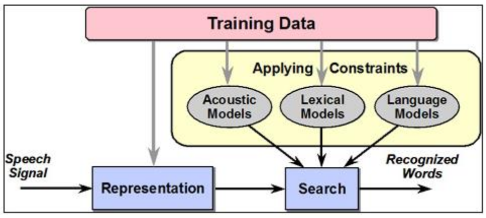

*Fig (2) Major components in a Speech Recognition System*

In speech recognition, a software application recognizes spoken words. The measurements in this application might be a set of numbers that represent the speech signal. We can segment the signal into portions that contain distinct words or phonemes. In each segment, we can represent the speech signal by the intensities or energy in different time-frequency bands. Speech recognition applications include voice user interfaces. Automatic Speech Recognition (ASR) is the process of deriving the transcription (word sequence) of an utterance, given the speech waveform. Speech understanding goes one step further, and gleans the meaning of the utterance in order to carry out the speaker's command. ASR systems facilitate a physically handicapped person to command and control a machine. Even ordinary persons would prefer a voice interface over a keyboard or mouse [10]

## 3. SPEECH RECOGNITION FOR BUSINESS PERSPECTIVES

### *3.1. Advantages of Speech Recognition for Business*

In disabilities help, one major way speech recognition technology is helping people is by allowing people with disabilities to type and operate computers. There are many people who can't type or operate a computer with their hands because of an impairment. Speech recognition technology has opened up a whole new world for these people and has allowed them to continue to participate in our highly digitized society. [9]

In spelling, another way speech recognition devices are very beneficial is because they seem to always know the correct spelling of words you're using. Many people struggle with correct spelling and the correct use of a word, especially when it comes to words like to, too, and two and whether or weather. With voice recognition technology, these people no longer struggle with this and it saves them lots of time in the process.

In enhanced speed, voice recognition technology is also an amazing time-saver for people who do not type very well or fast. There are many people who are not great at typing and struggle to learn proper keyboard hand placement but have great ideas. Typing can be a painful process for these people and many get discouraged because they think much faster than they type. For these people, speech recognition devices can help them get their ideas on paper just as fast as they think of them. [9]

In specialization, another way speech recognition technology is improving businesses is through its ability for people to add notes to files. In the medical sector, Voice Command Technology (VCT) is helping physicians file notation directly into a patient's Electronic Health Record (EHR). VCT in the medical sector has a built-in comprehensive medical vocabulary allowing physicians to accurately add critical notes to a patient's files. [9]

### *3.2. Disadvantages of Speech Recognition for Business*

In training, one major con of voice recognition technology is having to train them to recognize individual voices when you first purchase them.

Voice recognition devices take time to learn voices and the way different people speak. These devices require a lot of time and patience and they're still vulnerable to mistakes, even after you've effectively taught it your voice and speaking style. For many speech recognition devices, even after a long training period, many people find that they still speak in an unnatural way and over-enunciate words. [9]

In limited vocabulary, while voice recognition technology recognizes most words in the English language, it still struggles to recognize names and slang words. With the limited vocabulary of speech recognition devices, it might not be worth the purchase if you're continually having to go back over your work and fix many mistakes. [9]

In delays, voice recognition devices have been designed to help you speed up your work, but they are prone to mistakes and mishaps. These devices often take a bit to register what's being said, which can be frustrating and interrupt your thought flow. Having many frequent pauses can easily put you in a bad mood and when it glitches, it can force you to have to abandon the technology to get your work done. [9]

## 4. CONCLUSION

Speech Recognition are the emerging scope of security and authentication for the future. Now-a-days text and image passwords are prone to attacks. In case of the most commonly used text passwords, users are required to handle different passwords for emails, internet banking, etc. Hence they tend to choose passwords such that they are easy to remember. But they are vulnerable in case of hackers. In case of image passwords, they are vulnerable to shoulder surfing and other hacking techniques. Advances in speech technology have created a large interest in the practical application of speech recognition. Therefore, this system provides the users with the appropriate and efficient method of authentication system based on voice recognition.

## REFERENCES

[1] Alpa Reshamwala, Dhirendra Mishra, Prajakta Pawar," REVIEW ON NATURAL LANGUAGE PROCESSING" , IRACST – Engineering Science and Technology: An International Journal (ESTIJ), ISSN: 2250-3498, Vol.3, No.1, February 2013.

[2]. L. R. Bahl, P. F. Brown, P. V. de Souza, and R. L. Mercer, "A tree based statistical language model for natural language speech recognition," in Acoustics, Speech and Signal Processing, IEEE Transactions on, vol. 37, Issue 7, (Yorktown Heights, NY,USA), pp. 1001–1008, July 1989.

[3].https://en.wikipedia.org/wiki/Speech_recognition

[4] Y.-Y. Wang, M. Mahajan, and X. Huang, "A unified context-free grammar and n-gram model for spoken language processing," in IEEE International Conference on Acoustics, Speech, and Signal Processing, vol. III, (Istanbul, Turkey), pp. 1639–1642, Institute of Electrical and Electronics Engineers, Inc., 2000

[5] L. Zhou and D. Zhang, "NLPIR: a theoretical framework for applying natural language processing to information retrieval," J. Am. Soc. Inf. Sci. Technol., vol. 54, no. 2, pp. 115–123, 2003

[6] P.Clarkson and R. Rosenfeld, "Statistical language modeling using the cmu-cambridge toolkit," in Proceedings EUROSPEECH (N. F.G. Kokkinakis and E. Dermatas, eds.), vol. 1, (Rhodes, Greece), pp. 2707–2710, September 1997.

[7] J. Tejedor, R. Garca, M. Fernndez, F. J. LpezColino, F. Perdrix, J. A. Macas, R. M. Gil, M. Oliva, D. Moya, J. Cols, , and P. Castells, "Ontology-based retrieval of human speech," in Database and Expert Systems Applications, 2007. DEXA '07. 18th International Conference on, (Regensburg, Germany), pp. 485– 489, September 2007.

[8] J. R. Bellegarda, "Statistical language model adaptation: Review and perspectives," vol. 42, no. 1, pp. 93–108, 2004.

[9]https://www.techfunnel.com/information-technology/7-pros-and-cons-of-using-speech-recognition-in-business/

[10] B.H. Juang, and S. Furui, "Automatic Recognition and Understanding of Spoken Language–A First Step Toward Natural HumanMachine Communication, Proc. IEEE, 88, No. 8, 2000, pp. 1142-1165.f