

Artificial Intelligence Applications in Speech Recognition: Natural Language Processing

Manali H Savant¹, Mithali H. Savant², Vijayalakshmi N. Reddy³

^{1, 2, 3} Department of Computer Science and Engineering, Jain College of Engineering and Research, Belagavi, India

manalisavant24@gmail.com, mithalisavant444@gmail.com, v2ksbt@gmail.com

Abstract— India has a diverse list of spoken languages throughout the country; India has 22 officially recognized languages such as Hindi, Kannada, Marathi, Tamil, etc. Artificial Intelligence (AI) is branch of computer science that deals with making smart machines which are capable of performing various tasks that need less human interaction. Phonetics is the systematic study and classification of sound produced by human i.e. speech. Speech Recognition is a process of enabling a machine or a device to identify and respond to the voice produced by humans. This paper describes the Artificial Intelligence Applications in Speech Recognition is subfield of Natural Language Processing particularly for Indian native language.

Keywords— Artificial Intelligence (AI), Natural Language Processing (NLP), Phonetics, Speech Recognition.

I. INTRODUCTION

The present era is of human machine interaction which plays a vital role in various fields like Banks and Financial Institutions, Defense and Military, Education, Medical and Transportation fields, Reservation Systems, Enquiry Systems. Under developed areas and rural communities are being denied for technologies because of English that lead to spread of awareness about computer networks and communication. The best solution to Non-English user could be smart devices interacting with human in mother tongue language. India is a language diverse nation, as per 2001 census India has 1599 languages, 122 major languages and 22 official languages in which some of them are Hindi, English, Nepali, Kashmiri, Gujarati, Punjabi, Sanskrit, Bengali, Oriya, Manipuri, Marathi, Kannada, Konkani, Tamil, Telugu and Urdu [1,2,3] as per 8th Schedule. These are the naturally spoken languages in India. This paper focuses on linguistic code choice that is shift from one language to another within a single utterance, also known as Code-Switching.

Kannada [1] is a Dravidian Language spoken mainly by the people of Karnataka and the neighboring states such as Maharashtra, Andhra Pradesh, Telangana, Tamil Nadu, Goa and Kerala in Southern part of India. It is the administrative and official language of Karnataka. Kannada was the assembly language for many powerful Empires in Southern India and was written in Kannada Script in Kadamba Dynasty from the 5th century [2]. The Language uses 49 phonemic letters, divided into three groups among them Swaragalu (Vowels) are 13, Yogavaahakagalu (part-

vowel, part consonants) are 2 and Vyanjanagalu (Consonants) are 34.

Characteristics of Kannada Language are:

Method of writing: alphasyllabary in consists of consonants that has inborn vowels.

- Vowels are written as individually.
- When the consonants are together without the help of inborn vowels, they form conjunct symbol.
- *Direction of writing:* Kannada language is written from left to right in horizontal lines.

Kannada script is an abugida (alpha syllabary) of the Brahmi (Indic) script. It is a segmental, non-linear alphabet script characterized by consonants appearing with different vowel. Each alphabet is called as Akshara and each letter has its visible and audible representation of sound. Giving the visible and audible representation. Kannada alphabet [3] is popularly known as varnamale and it consists of 49 characters. In order to make the recognition system compatible to the earlier varnamale set 51 characters are considered as characters can combine to form compound characters leading to ottaksharas.

Classification of Kannada Varnamale: The 49 basic letters are classified into three categories. They are Swaragalu (vowels), Yogavaahakagalu (part vowel, part consonants) and Vyanjanagalu (consonants). Each sound has its own distinct letter, and it is pronounced the way it is spelt.

The accent comes from the first syllable. Every consonants sound has two different pronunciation. The sound with normal pronunciation (known as deergha) is used in the varnamale (aksharamale)

1. Short without the help of vowel.

(ಠ known as Hrasva)

2. Long in union with the first vowel.

(ಠ known as Deergha)

Swaragalu (Vowels): There are 13 vowels called as Swaragalu. It represents the speech sounds pronounced with the help of free passage of mouth through the oral cavity.

TABLE I. SWARAGALU(VOWELS) IN KANNADA

Swaragalu (vowels in Kannada)	Equivalent vowel (English)	Swaragalu (vowels in Kannada)	Equivalent vowel (English)
ಅ	A	ಋ	Rū
ಆ	Ā	ಌ	Ye
ಇ	I	ಋ	Yē
ಈ	Ī	ಋ	Ai
ಉ	U	ಋ	O
ಊ	Ū	ಋ	Ō
ಋ	Ru	ಋ	Au

Vyanjanagalu(Consonants): There are 34 consonants called as Vyanjanagalu. It represents the speech sound produced by a partial or complete obstruction of the air ways of the speech organs in mouth. The Consonants are classified into two types.

1. VargiyaVyanjanagalu (Structure Consonants)
2. AvargiyaVyanjanagalu (Unstructured Consonants).

VargiyaVyanjanagalu: The Structured Consonant are categorized based on the tongue touches the mouth palate as shown in Table II.

TABLE II. VARGIYAVYANJANAGAL IN KANNADA

Voicless aspirate	Voicless aspirate	Voiced aspirate	Voiced aspirate	Nasal
ಕ (ka)	ಖ (kha)	ಗ (ga)	ಘ (gha)	ಙ (nga)
ಚ (cha)	ಛ (chha)	ಜ (ja)	ಝ (jha)	ಞ (ña)
ಟ (ṭa)	ಠ (ṭha)	ಡ (ḍa)	ಢ (ḍha)	ಣ (ṇa)
ತ (ṭha)	ಥ (ṭhaa)	ದ (ḍa)	ಧ (ḍha)	ನ (na)
ಪ (pa)	ಫ (pha)	ಬ (ba)	ಭ (bha)	ಮ (ma)

Avargiya Vyanjanagal: The Unstructured Consonants the tongue doesn't touches the mouth palatethese consonants are called Unstructured Consonants as described in table in Table III.

TABLE III. AVARGIYAVYANJANAGALU IN KANNADA

ಯ (ya)	ಱ (ḷa)	ಷ (ṣa)	ಶ (ṣa)
ರ (ra)	ವ (va)	ಸ (sa)	ಹ (ḷ)
ಲ (ḷa)	ಳ (ḷa)	ವ (ḷa)	

Consonants Conjuncts: Kannada language is rich in conjunct i.e. consonant clusters, they are subjoined in form in Table IV.

Yogavaahakagalu(part vowel, part consonants): The Yogavaahakagalu has 2 letters:

1. Anusvara: ಅಂ (Am)
2. Visarga: ಅಃ (Aha)

TABLE IV: CONSONANTS CONJUNCTS WITH KANNADA LETTER MA

ಮ	ಮ್	ಮ್	ಮ್	ಮ್	ಮ್
	ಮ್	ಮ್	ಮ್	ಮ್	ಮ್
	ಮ್	ಮ್	ಮ್	ಮ್	ಮ್
	ಮ್	ಮ್	ಮ್	ಮ್	ಮ್
ಮ್ಯ	ಮ್	ಮ್	ಮ್	ಮ್	ಮ್

TABLE V: YOGAVAAHAKAGALU (PART VOWEL, PART CONSONANTS) IN KANNADA

Anuswara	Equivalent (English)	Anuswara	Equivalent (English)
ಅಂ	Aom	ಅಃ	(ahā)

II. PHONETIC: NATURAL LANGUAGE PROCESSING

The phonetic studies were at the 6th century BCE by Sanskrit grammarian, well-known Hindu Scholar Panini was the early investigator, whose grammar, written around 350 BCE linguistics in modern language. He described important phonetic principles, including voicing and production of sound.

Phonetics [4] is a branch of linguistics which focuses on how human's making and perceive sounds. Phoneticians - linguists who expertise in phonetic focus on properties of speech physical.

The field of phonetics is divided into three types based on how human produce speech. There are two aspects in phonetics of human speech. They are:

- 1) Production – how humans make a sound
- 2) Perception- how the speech is interpreted by the human.

The phonetic [5] is of field of linguistics which enlightens on pronunciation and its speech. There are three kinds of phonetics to implement phonetic dictionary for Kannada or any other language.

1. *Articulatory*: This phonetic deals with the movement of speech organs or articulator such as vocal folds, lips, tongue position, shape, and movement as shown in Figure 1

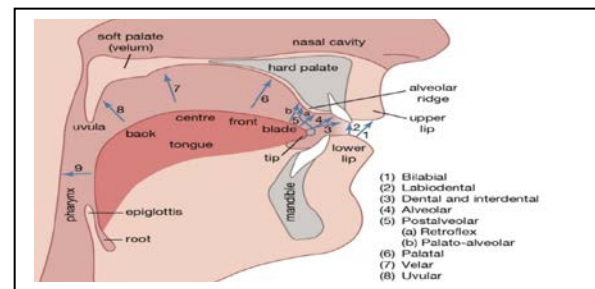


Fig 1. Places of articulation

2. *Acoustic*: This phonetics deals with physical sound waves properties of the speech such as speech harmonic structure, amplitude and sound wave frequency.

Ex: Pronunciation of sentence by the speaker, transmission to the listener.

3. *Auditory*: This phonetics deals with understanding, recognizing and categorizing the sound speech or understanding the meaning of the word.

These phonetics are interconnected by the means of sound, such as amplitude, wavelength and harmonics. Different vowels sound has a definite pattern for the production of sound. Ex: vocal folds are vibrated and the nasal passage is closed while production of Kannada vowel sound.

III. ARTIFICIAL INTELLIGENCE(AI)

Artificial Intelligence (AI): It deals [6] with human machine interaction by processors such as self- correction, reasoning and learning. Few of the applications of Artificial Intelligence includes expert systems, Speech Recognition and machine vision. Artificial Intelligence coined by John Mc Carthy an American computer scientist, in 1956 at The Dartmouth Conference where the discipline was born. The market for Artificial Intelligence Technology is flourishing, some of the variety of technologies and tools were developed are: Google Assistant, Alexa, Siri, Cortana and Eco. Some of the applications of Artificial Intelligence are discussed here.

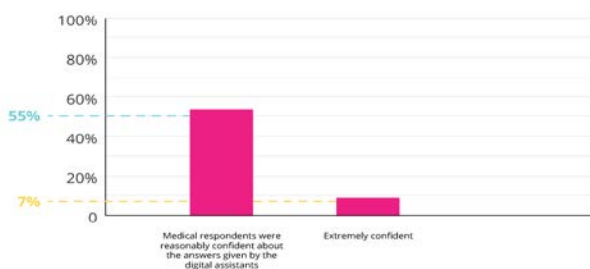
APPLICATIONS [7] OF ARTIFICIAL INTELLIGENCE (AI):

Stephen Hawking’s Speech synthesizer: Stephen Hawking, was well known English physicist, author, cosmologist, and Director of Research at the Centre for Cosmology in the University of Cambridge used speech synthesizer to interact with people. With this technology, he was able to translate text into speech. This system helped to produce the respective sounds and there was availability of word prediction.

Artificial Intelligence in Health Care: Companies like IBM’s Watson are applying machine learning for faster diagnosis and accurate results. This technology understands natural language of humans and responds to the queries asked to it. It plays vital role to assists doctors, nurses and patients for the treatment.

Benefits: Extracting and maintaining the medical records. Guiding and instructing nurses. Maintaining the data like number of patients on a floor, availability of beds in hospitals, number of emergency units and so on. The below graph 1 shown represents the survey conducted by Pediatricians in Boston Children’s Hospital.

Artificial Intelligence to help combat COVID-19 [6]: NVIDIA has introduced a platform called NVIDIA Clara Guardian to combat COVID-19 with the help of Artificial Intelligence and Speech Recognition Technology for medical assistance in smart hospitals for limiting staff exposure and monitoring. This system uses video analytics that combine speech, vision and natural language processing.



Graph1. Survey of Pediatrician conducted by Boston Children’s Hospital

Artificial Intelligence in Business [7] and Marketing: For highly repetitive tasks performed by humans in marketing, they have introduced Robotic process automation. E-companies and websites have launched Chat Bots to provide faster and standard services for customers. This includes voice search that helps customers and marketers to interact with each other. This helps marketers to analyze the customer requirements and present trends. With the Speech Recognition technology, marketers can analyze customer’s voice pattern, accent and vocabulary that helps them to extract customer information such as age, address and location. In upcoming years brands such as Amazon, Flipkart, Myntra can optimize their profits with help of voice search.

Artificial Intelligence in Autonomous Vehicles: Self-driving cars require sensors to understand and interpret the atmosphere around them and a brain to collect, store, process and take the right action depending upon the information gathered. Artificial Intelligence has various application for the vehicle and most important among them are:

- Directing the car based on traffic condition to find the shortest route.
- Directing the car to fuel or gas stations if it is shortage of fuel.
- Passenger can communicate with the speech recognition device that is present in car.

Artificial Intelligence in Workplace: In Workplaces, Speech Recognition Technology have been implemented to increase the efficiency of task. Example: In Office,

- Searching and Inserting files or documents or reports in computer systems.
- Creating tables and graphs with help of data.
- Requesting for printing the documents.
- Making video conferencing.
- Recording time.

Artificial Intelligence in Banking[8]: The main objective of Speech Recognition Technology in financial industry and banking sector is to reduce the friction for the customer and reduce human customer service with the help of voice activated banking such as requesting information regarding expenditure, transaction history, making payment and so on without opening mobile or other devices. This is possible with the personalized banking assistant which would improve banking standard and customer satisfaction.

Virtual Agent [9] [10]: It is one of the efficient artificial intelligence machines or assistant that serves as online representative for customer service in various platforms. Ex; Louise: it has intelligent conversation with users; perform adequate non-verbal behavior and responds to their queries.

Deep Learning: It is a platform of machines learning consisting of multiple abstraction layers with artificial neural networks, is used for classification applications and pattern recognition.

Machine Learning: It provides various algorithms, User application Interface Development and training tool kits as

well as computing power to design and deployed models into applications for user friendly simulation.

Robotics Process Automation: As Robot doesn't tire and have huge storage space, hence it is used, where human is unable to easily execute the task, as it performs the same task within the fraction of second.

Text analytics and Natural Language Processing: NLP uses and it supports text analytics for understanding the meaning and structure of sentence and its sentiment. Text analytics helps in Security, Fraud detection and etc.

IV. DIGITAL ASSISTANT: SPEECH RECOGNITION TECHNOLOGY

Speech Recognition: The process in which the speech of human is translated into machine understandable language or format is called as speech recognition. It is used in application such as personal assistants, digital assistants, voice response systems, mobile applications and so on. With the developing technology in Artificial Intelligence in Speech Recognition that are used in voice- controlled assistants are playing the significant role for upgrading the technology in the 21st century. With this technology people can interact with cars, homes and device like Google Assistant, Alexa, Siri, Cortana and Eco.

There are many Digital Assistants are developed to help the people to perform their tasks and also to respond their queries by providing access to the information from data warehouse in different digital sources [11]. These Digital Assistants will help to solve real timeproblems some speech recognition Digital Assistants are: 1) Amazon's Alexa, 2) Apple's Siri, 3) Google's Google Assistant, 4) Microsoft's Cortana and so on.

Smart Personal Assistants: In digital era, the technology that converts voice-to-text for basic conversion has become an interface that controls the new generation of personal assistants such as Google and Siri. It helps to set reminders and browse internet.

Voice-to-text: Smart phones have a standardized feature to translate voice-to-text by recording a phrase or a sentence or by pressing a button we can start interacting with the device. Artificial Neural Network Technology is been used by Google for voice search and Microsoft also have developed this type of system that transcribe conversion.

Amazon's Alexa: It is a personal assistant that responds to voice instructions to set reminders, respond to the questions, to create a list, online ordering.

Amazon's Eco: It is a smart speaker which is integrated with Alexa and uses voice instruction.

Microsoft'sCortana: This Artificial Intelligence Assistance which is preloaded is used in Microsoft smart phones and in computers windows.

Glimpse into the future of Speech Recognition: Digital Assistants plays a vital role in bridging up the gap between the Smart homes and Humans. Google home was launched by Google in October 2016, which was turned out to be the competitor for Amazon's Alexa and Eco that had deep integration with Google products like Google Assistant,

Google Play Music, YouTube and Nest. With the help of voice instructions given by the user in Natural Language, user can interact with assistant and receive live updates of news, sports, weather forecast and finance; play music; ask questions; set reminders and book appointment

Example: Mark Zuckerberg, the CEO of Facebook launched a server called Jarvis, which is an emulation of Artificial Intelligence Assistant in Iron Man Films by Robert Downey. With the help of Jarvis, Mark was able to connect infinite home devices to recognize friends and family at the door step and let them in; play music and so on. To instruct Jarvis, Facebook- Messenger Bot was built to give text commands and Speech Recognition App was built to give voice commands as shown in fig 2.

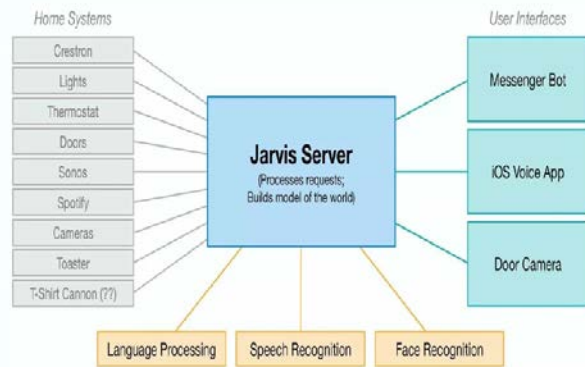


Fig 2. Jarvis Server

V. CONCLUSION

This research paper illustrates the insight of the phonetics particularly for Kannada syllable and its articulation. The place and movement ofvocal folds will help to create phonetics dictionary for Natural Language Processing in Speech Recognition Technology using Artificial Intelligence algorithms. It also focuses the applications of Digital Assistant such as Speech Recognition Technology which have higher scope in Healthcare, Banking, Business, Marketing, Workplace and etc.

REFERENCES

- [1] <https://en.wikipedia.org/wiki/Kannada>
- [2] https://en.wikipedia.org/wiki/Kadamba_dynasty
- [3] <https://omniglot.com/writing/kannada.htm>
- [4] Mallamma V. Reddy et al Phonetic Dictionary for Natural Language Processing: Kannada Int. Journal of Engineering Research and Applications ISSN: 2248-9622, Vol. 4, Issue 7(Version 3), July 2014, pp.01-04
- [5] <https://en.wikipedia.org/wiki/Phonetics>
- [6] <https://www.valluriorg.com/blog/artificial-intelligence-and-its-applications/>
- [7] <https://www.getsmarter.com/blog/market-rends/applications-of-speech-recognition/>
- [8] <https://healthitanalytics.com/news/artificial-intelligence-genomics-tools-to-help-combat-covid-19>
- [9] Minh Khue Phan Tran, Philippe Robert, François Bremond. A Virtual Agent for enhancing performance and engagement of older people with dementia in Serious Games. Workshop Artificial Compagnon-Affect-Interaction 2016, Jun 2016, Brest, France. fihal-01369878f
- [10] Grigore, Elena Corina (et al.), Talk to Me: Verbal Communication Improves Perceptions of Friendship and Social Presence in Human-Robot Interaction DOI:10.1007/9783-319-47665-0-5 PP:51-63 2016
- [11] <https://emerj.com/ai-sector-overviews/everyday-examples-of-ai/>