Speech Recognition Techniques in Agriculture Sector : A Comparative Study

Somnath Hase[1], Sunil Nimbhore [2]

{hasesir2009@gmail.com[1], nimbhoress@gmail.com[2]}

Department of CS & IT Dr. Babasaheb Ambedkar Marathwada University, Aurangabad (MS), India.[1]

Abstract. Desire and need of people to communicate with a machine with the advancement of technology affect the natural language processing evolution. Communication between humans and machines is indispensable day by day. The automatic speech recognition system plays a crucial role. It is a method of communicating to machines by using speech. In the early days, farmers relied on clouds for rain. They are now looking towards Cloud Computing (CC) as a solution for cultivating crop production. With excessive use of chemicals, sudden climate change, bacterial attacks, and lack of information, the crop is damaged every year in India. In this research, various techniques related to speech recognition in field agriculture have been studied in detail. Most of the speech recognition systems are based on isolated keywords specifically designed for accessing the price of commodities. Some are designed to assist farmers based on crop symptoms. There is an agricultural voice-based speech recognition system through oral description using isolated speech databases. This study helps to compare different tools and techniques of speech recognition that are used in agriculture.

Keywords: Agriculture, speech recognition(SR), TDIL, Commodity Price(CP), speech database.

1 Introduction

Despite recent technological advancements, agriculture remains the primary source of income for a large number of people. Food, Water, Clothing, Shelter, and Oxygen are basic needs of human beings to survive on this planet. As a result, agriculture is the backbone of many countries. According to the survey, the world's population will be about 10 billion by the year 2060. However, due to population growth, the need for food grains has increased dramatically in recent years. Unfortunately, the food grain is not directly related to population growth. As a result, food production should be increased globally in the next few years[1].

Agriculture provides a livelihood for more than 50% of India's population. The vast majority of Indian farmers are small-scale in terms of land and also 78% of farms are less than five acres in size [2]. The majority of Indian farmers are uneducated, and they confront several challenges

ICIDSSD 2022, March 24-25, New Delhi, India

DOI 10.4108/eai.24-3-2022.2318772

such as a lack of water, lack of knowledge about modern farming equipment, dependence on the traditional crop, absence of storage facilities, transportation problems in the rural area. The Ministry of Agriculture, Government of India, had put up several websites that offer information on challenges and solutions in the agricultural field, to stimulate agricultural development. However, most farmers are technologically behind and unable to utilize the website [3]. The use of Information Communication Technology in the agriculture sector is the solution to enable and empower rural Indian Famers.

The Artificial Intelligence-based tool and technologies help farmers to use weather forecasts to determine what type of crop may be cultivated and when seeds should be sown. It helps a farmer to identify the nutrient deficiencies in the soil and to identify crop health conditions. Thus, the new technologies-based tool has changed traditional farming, it not only helps farmers to automate their farming but also shifts to precision cultivation for better quality and higher crop yield with fewer resources. There has been tremendous progress in speech recognition technologies in the last few years. The people are more comfortable with voice instead of keyboard and mouse to communicate to machines like a computer. The Speech recognition system is helpful for disabled people like the blind and handicapped along with people who don't know these devices.

One of the key issues in the agriculture industry is the difficulty in obtaining necessary information in the local language. The government of India launched a programme Kisan Call Centers (KCCc)" on January 21, 2004, to answer farmer queries on a telephone call with 22 local languages. The farmer faces a lot of problems with KCC like network busy, sometimes waiting time too high and so on [4]. The Government of India's Department of Information Technology project Technology Development for Indian Languages (TDIL) aims to provide a voice interface in Hindi, Bangla, Marathi, Telugu, Tamil, and Assamese languages. Seven Institutions namely IITG, IITB, IITK, IITM, IIITH, TIFR, and C-DAC Kolkata are involved in implementing the project. The work is carried out at IITB and TIFR Mumbai to develop Marathi Language Automatic Speech Recognition System. The purpose is to deploy and develop a speech-based system for farmers so that they can access various agricultural commodities prices from different districts across India by using mobile phones [5].

The Automatic Speech Recognition system converts acoustic waveform (input) into a string of words (output). In Speech recognition, we have an acoustic observation like O= O1, O2, O3,..., On, the objective is to find the sequence of a word like W= W1, W2, W3,...., Wn with maximum posterior probability P(W|O).

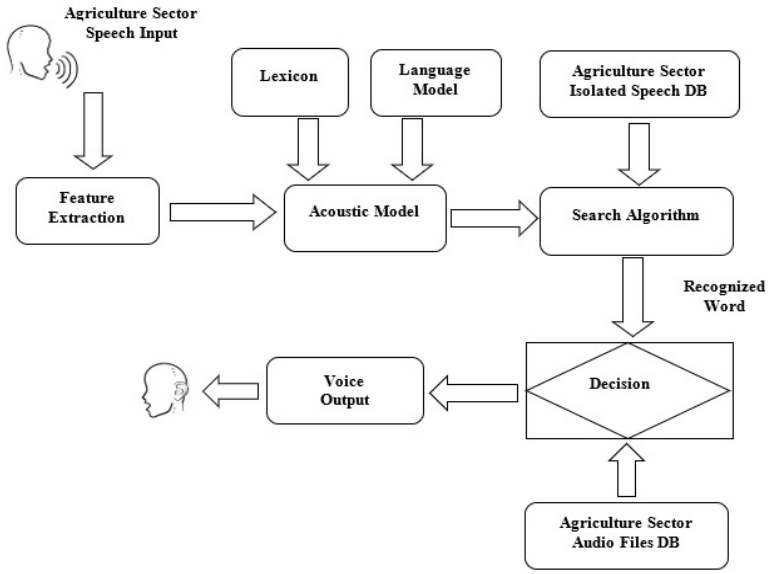

Fig. 1. Typical Speech Recognition System with input-output in Agriculture sector

Fig. 1 illustrates a typical speech recognition system with input-output in agriculture. A speech feature are extracted from the speech input. A lexicon is simply a pronunciation dictionary and a language model consists of syntax and semantics to frame a sentence. Acoustic model is developed using lexicon, language model, and features that have been extracted in the last step. With help of speech databases, a search algorithm recognized words. A decision between recognized word and audio files related to the agriculture field is made. At last voice output is generated by systems.

This paper summarizes work that has been made by various authors in the field of agriculture that uses speech recognition technologies. Section I covers the introduction part, Section II presents related work that has been made using speech recognition in the agriculture field while conclusions are drawn in section III.

2 Related Work

Wenhao Ou et al. proposed "Application of keyword-based speech recognition in the agricultural voice information system" [6] using Microsoft windows speech recognition development platform SAPI. Real-time agricultural information can be obtained using computers, telephones, a television in a modern rural area. In the proposed system farmer, speaks the name of agricultural products orally. When the recognition process completes the system plays the audio file of price information related to the product obtained by searching words from the database. Speech recognition systems are divided into two modes continuous dictation and command control mode. Due complexity of continuous dictation and recognition accuracy practical requirements, command mode is selected which shows 85 % of accuracy

with high practical value. Syntax libraries for command mode need to be designed and for generating keyword grammar library, XML format should be used.

Gautam Varma et al. demonstrated "A speech-based conversation system in Indian languages for accessing agriculture commodity prices" [7]. They developed Mandi Information System to obtain prices of various agricultural commodities such as fruits, pulses, vegetables, and spices for farmers in rural and semi-urban areas across the state of Andhra Pradesh in India. The size of the vocabulary of Mandi Information System consists of 72 commodities, 348 Market, and 23 Districts names. The system is developed by using the Telugu language where speech samples are collected from the framer over a mobile phone. The different pronunciation variation is captured by providing a list of word of commodity, market, and district name that has been uttered for recording. A total of 96 speakers have participated in data collection wherein each speaker is asked to speak near about 500 words. The system prompts for explicit confirmation, whenever a user provides speech data to MIS.

Tejas Godambe et al. in "Speech Data Acquisition for voice-based Agricultural Information Retrieval" [8] describes the process of collecting speech data to train an ASR system for the Marathi language. The author stated that the Department of Information Technology has initiated a Technology Development for Indian Language program nationwide to develop voice interfaces in Indian languages like Hindi, Bangla, Marathi, Telugu, Tamil, and Assamese with a consortium of seven institutions. The data collection is carried out by Uniphore with training and testing phase in 34 districts of Maharashtra from 1500 farmers, with each district having 45 farmers. Along with speech data collection, Uniphore volunteers also maintained metadata entry sheets which are helpful to develop metadata-specific models. Agricultural data collection is divided into four blocks i.e. directed questions are useful for the collection of natural speech, the prompted questions are used to collect the names of a commodity that are traded rarely, picture prompted question helps to collect the local name of commodities, and spontaneous response while general questions are used for future work and analysis of speech.

Anchan Mohan et al. presented a paper on "Acoustic Modelling for speech recognition in Indian Language in an agricultural commodities task domain" [9]. A multi-lingual system is configured with data of Marathi and Hindi which are linguistically similar languages. A Subspace Gaussian Mixture Model is used to train the system. They use the cross-corpus acoustic normalization method of speaker adaptive training variant. A context-dependent states effects of sharing "similar" Marathi language with Hindi speech recognition performance is also introduced.

Jinpu XU et al. presented a paper on "An Approach of Agricultural price information collection based on speech recognition" [10]. They collected speech corpus under operating environment, selected triphone modeling as the decode unit for training Hidden Markov Model. The problem of insufficiency in training samples is solved by using decision tree-based clustering of states. To make a model more precise, a mixture of Gaussian components is used. Cepstral means normalization along with Cepstral variance normalization are used to reduce the discrepancy that may arise among testing and training environments. Open-source software like HTK speech recognition toolkit along with Microsoft VC++ is used in experiments. 7680 utterances of male and female are collected from two markets in Beijing in form of "product + price". For example, "Fresh Shrimp, 23.0". The author extracts 39-dimension MFCC features, 13 static coefficients with its logarithmic energy, also with its first and second-order derivatives. The proposed model shows a 95.04% recognition rate for males and 97.62% for females with CMVN modeling.

Pukhraj P. Shrishrimal et al. developed “Marathi isolated words speech database for agriculture purpose” [11] which categorized into fertilizers, vegetables, fruits, grains, diseases, pesticides, types of equipment, and cash crops. They created a text corpus by visiting several websites that have information related to agriculture in Marathi. The speech samples are collected from the native speakers by visiting 10 villages, with each village consisting of 10 speakers out of which 5 are males and 5 are females. Each speaker speaks 100 words with 3 utterances resulting in a total database of size 30,000. They used Sennheiser PC350 and PC360 for recording with PRAAT Software, 16000Hz sampling frequency, 16 bits mono audio format. The recorded files are saved in .wav format. The Spectral subtraction noise removal technique is used which is the most commonly used speech signal enhancement technique for enhancing speech samples.

Santosh Gaikwad et al. proposed “Speech Recognition for Agriculture based Interactive Voice Response System”[12]. Now a day, modern systems include speech recognition that allows speaking commands rather than using keypad selection. They created the database for 5 crops, with a total of 35 speakers out of which 17 are females and 18 are males in the age group of 20 to 30. The database is further divided into three parts i.e. crop name, symptoms on the crop, and name of disease and its solution. In the proposed system call will be connected to AGRO IVRS, which prompts the user to tell crop name and symptoms on a crop. The database will be searched and the name of the crop disease along with its solution in the form of voice will be given to the user. The AGRO IVRS system when tested provides a 91% overall system recognition rate.

Smita Magre et al. “Designed and Developed Automatic Speech Recognition System of Isolated word for Agricultural Purpose” [13] using MFCC and RASTA. MFCC feature extraction technique is applied on 30,000 utterances with 100 words spoken by 100 speakers each with 3 utterances including male and female. MFCC is a broadly and commonly used technique that consists of different steps like Pre-emphasis, Framing, Hamming Windowing, Fast Fourier Transform, Mel Filter Bank Processing, and Discrete Cosine Transform. After MFCC, RASTA (Relative Spectral) Processing is applied on input samples because MFCC values are not robust in a noisy environment, whereas RASTA finds out features in noisy data. Dynamic Time Warping is used for the recognition of isolated words which checks whether two spoken words represent equal phrases or not. A table of the distance matrix of various Ayurvedic plants like Ruai, Yerand, Tulas, Dhatura, Khorphad, Limb, Ashoka, Lavang using Dynamic Time Warping is presented.

Jinpu Xu et al. developed “Agricultural price information acquisition using noise-robust Mandarin auto speech recognition” [14]. The database consists of “name+price” form, 142 types of agricultural price speech sentences such as “cabbage 1.5 ¥”, with recording in standard Mandarin Chinese Language. A training set with help of speakers from two markets in Beijing, with 20 subjects including 10 male and 10 female for a total of 40 speakers is collected. The testing is carried out over another six people with 3 males and 3 females with spoken 100 sentences for 600 recorded sentences. The recording of voice samples is collected using a mobile phone like Millet, Samsung, Huawei. The model is trained using continuously mixed density HMM. A 39-dimension MFCC feature vector is extracted using HMM with the left to right,

non-jump structure. However, due to fact that the Triphone model simulates the co-articulation phenomena better, its performance is superior to the monophonic model. The problem of insufficient training of sample is solved using a state clustering method which is based on a decision tree. The Cepstral Mean Normalization (CMN) is used to eliminate channel noise and also to reduce the error rate by 30% approximately. The agricultural product price recognition rate with Cepstral mean-variance normalization (CMVN) increased to 95.34 for the male sample and 97.62% for female samples.

Girish Hegde et al. in "A Study on Agriculture Commodities Price Prediction and Forecasting" [15] made comparative studies on prediction and forecasting of agriculture commodities price. The author stated that there are several problems faced by a farmer while using Kisan Call Centers (KCCs)" like Network busy, Farm Tel Advisor (FTAs) are not available, due to busy network or sometimes they are not responding, Information provided is not in the local dialect, sometimes KCC officer delivers information in high speed and hence farmers are not able to understand it. The author has studied near about 17 research papers related to the agricultural domain that used different tools and techniques for predicting agricultural commodity prices. The author surveyed research already done and stated the methodology, application, limitation of each framework. They stated that Government predicts agricultural commodities prices using short arrival and historical data, to protect the farmer from the price crash or control the inflation.

Table I represents the work that has been made in the agriculture sector with different methodologies & techniques of speech recognition.

Table I. Existing literature work using speech recognition in agriculture

Sr. No.	Author	Title	Year	Methodology/ Techniques	Application	Limitation/ Future Scope
1	Wenhao Ou et al.	Ref. [6]	2010	SAPI development platforms of Microsoft Windows with Command mode, SAPI Framework, Database module, Operation & Monitor module, Speech Recognition module.	Access agricultural voice-based speech recognition system through the oral description.	The speech database is not explained, Also Metadata is not specified which is helpful to develop metadata specific model.
2	Gautam Varma et al.	Ref. [7]	2011	Context-dependent tri-phone HMM model built with 8 Gaussian mixtures per state using Sphinx recognition with appropriate NLU, DM, NLG, and TTS	Access Agricultural commodity price.	No separation between male and female speech database. Also, speech database is not defined in terms of commodities, market district
3	Tejas Godambe et al.	Ref. [8]	2011	Data collection is outsourced with firm Uniphore. Speech samples are collected over a phone line, Two DASs for two dedicated phone lines are set up at IIT Bombay and TIFR Mumbai respectively.	Obtain the price of commodities over the phone.	Only 45 farmers per district. Data collection needs to be improved.

4	Anchan Mohan et al.	Ref. [9]	2013	Two mono-lingual ASR Conventional CD HMM and SGMM are compared. Cross- corpus acoustic normalization method of SAT variant is used.	Present the effect of sharing a "similar" context-dependent state of Marathi Language on Hindi Speech Recognition.	Already developed a small subset of data used.
5	Jinpu XU et al.	Ref. [10]	2014	Triphone Modelling as decode unit to train HMM, Decision tree-based clustering, CMN, CVN method are used,39-dimension MFCC feature with 13 static coefficients plus its logarithmic energy.	Corpus specially designed for agricultural price information collection.	The model suggests female participants for the price acquisition.
6	Pukhraj Shrishrimal et al.	Ref. [11]	2014	100 isolated word Marathi speech database using Sennheiser PC350 and PC360. Spectral Subtraction noise removal technique for speech signal enhancement.	Developed database for agriculture purposes.	Phonetic variation and also spontaneous response from the farmer are not collected. Speech dataset to be improved.
7	Santosh Gaikwad et al.	Ref. [12]	2014	AGRO Interactive Voice Response System with computer telephonic card for data collection. 10 X 10 rooms without noisy sound with sampling frequency 11025Hz room temperature.	Provide agricultural assistance to farmers based on crop symptoms.	Only 5 crops, each with 2 symptoms are considered. Dataset needs to be improved. The system takes a lot of time for recognition.
8	Smita Magre et al.	Ref. [13]	2014	MFCC feature extraction technique along with RASTA (Relative Spectral) processing. DTW is used for recognition of isolated word	Developed the Marathi Language-based isolated word database and speech recognition system.	Only a few isolated words are considered. Dataset to be improved.
9	Jinpu Xu et al.	Ref. [14]	2018	Model is based on HMM, Context-dependent triphone modeling, State clustering based on a decision tree, Gaussian Mixture Components, CMN, CMVN	Acquire agricultural product price using robust continuous Mandarin speech recognition.	Speech samples are collected from only two farmer's markets. Speech dataset to be improved.
10	Girish Hegde et al.	Ref. [15]	2020	Different tools and techniques for predicting agricultural commodity prices. KCC, Farm Tel Advisor is not available.	Studies different models or frameworks to predict and recommend the price for farmers' produce.	Limited models and frameworks are studied.

3 Conclusion

The majority of people in India are employed in agriculture. There is a need to put greater focus on agricultural growth. Farmers in India continue to seed depending on intuition experience. The use of Speech recognition technology in the agricultural sector improves crop yield, decreases staff duplication, and increases work efficiency. In this study, the survey on previous research and the frameworks of speech recognition techniques in the agriculture sector is carried out. In most cases, a limited set of speech databases is used, in some cases only using few isolated word speech recognition systems are implemented. A speech recognition system must be robust, there must be a minimum mismatch between training and testing environment. It is hoped that this study will be useful for researchers who use speech recognition techniques in the agriculture field.

References

[1] S. V, "Internet of Things (IoT) based Smart Agriculture in India: An Overview," J. ISMAC, vol. 3, no. 1, pp. 1–15, 2021, doi: 10.36548/jismac.2021.1.001.

[2] N. Patel, D. Chittamuru, A. Jain, P. Dave, and T. S. Parikh, "Avaaj Otalo - A field study of an interactive voice forum for small farmers in rural India," Conf. Hum. Factors Comput. Syst. - Proc., vol. 2, no. May 2014, pp. 733–742, 2010, DOI: 10.1145/1753326.1753434.

[3] Anjani, Pooja, V. Ranjitha, M. Acharya, and G. Kamath, "Kannada Speech Recognition Enquiry System For Farmers," Int. J. Eng. Res. Technol., vol. 7, no. 08, pp. 1–5, 2019.

[4] S. Kavitha and N. Anandaraja, "Farmers Awareness on Kisan Call Centre Services in the Telangana State of India," MULTILOGIC Sci., vol. IX, no. XXX, pp. 64–66, 2019, [Online]. Available: www.ycjournal.net.

[5] M. A. Kumar, B. Premjith, S. Singh, S. Rajendran, and K. P. Soman, "An overview of the shared task on machine translation in Indian languages (MTIL)-2017," J. Intell. Syst., vol. 28, no. 3, pp. 455–464, 2019, DOI: 10.1515/jisys-2018-0024.

[6] W. Ou, W. Gao, Z. Li, S. Zhang, and Q. Wang, "Application of keywords speech recognition in agricultural voice information system," 2nd Int. Conf. Comput. Intell. Nat. Comput. CINC 2010, vol. 2, pp. 197–200, 2010, DOI: 10.1109/CINC.2010.5643755.

[7] G. V. Mantena, S. Rajendran, B. Rambabu, S. V. Gangashetty, B. Yegnanarayana, and K. Prahallad, "A speech-based conversation system for accessing agriculture commodity prices in Indian languages," Jt. Work. Hands-free Speech Commun. Microphone Arrays, HSCMA'11, pp. 153–154, 2011, DOI: 10.1109/HSCMA.2011.5942384.

[8] K. Godambe, T., & Samudravijaya, "Speech Data Acquisition for Voice-based Agricultural Information Retrieval," proceeding 39th All India DLA Conf. Univ. India., 2011.

[9] A. Mohan, R. Rose, S. H. Ghalehjegh, and S. Umesh, "Acoustic modeling for speech recognition in Indian languages in an agricultural commodities task domain," Speech Commun., vol. 56, no. 1, pp. 167–180, 2014, DOI: 10.1016/j.specom.2013.07.005.

[10] J. Xu, Y. Zhu, H. Liu, and J. Zhao, "An approach of agricultural price information collection based on speech recognition," 10th Int. Conf. Nat. Comput. ICNC 2014, pp. 893–898, 2014, DOI: 10.1109/ICNC.2014.6975957.

[11] P. P. Shrishrimal, R. R. Deshmukh, and V. B. Waghmare, "Marathi Isolated Words Speech Database for Agriculture Purpose," Int. J. Eng. Innov. Res., vol. 3, no. 3, pp. 248–253, 2014.

[12] Santosh Gaikwad, B. Gawali, and S. Mehrotra, “Speech Recognition for Agriculture based Interactive Voice Response System,” Int. J. Mod. Eng. Res., vol. 4, no. 2, pp. 112–115, 2014, [Online]. Available: http://www.ijmer.com/papers/Vol4_Issue2/Version-1/L04201112115.pdf.
[13] S. B. Magre and R. R. Deshmukh, “Design and Development of Automatic Speech Recognition of Isolated Marathi Words for Agricultural Purpose,” IOSR J. Comput. Eng., vol. 16, no. 3, pp. 79–85, 2014, DOI: 10.9790/0661-16377985.
[14] J. Xu, Y. Zhu, P. Xu, and D. Ma, “Agricultural price information acquisition using noise-robust Mandarin auto speech recognition,” Int. J. Speech Technol., vol. 21, no. 3, pp. 681–688, 2018, DOI: 10.1007/s10772-018-9532-7.
[15] G. Hegde, V. R. Hulipalled, and J. B. Simha, “A Study On Agriculture Commodities Price Prediction and Forecasting,” Int. Conf. Smart Technol. Comput. Electr. Electron. (ICSTCEE 2020), pp. 316–321, 2020.
[16] M. A. Anusuya and S. K. Katti, “Speech Recognition by Machine: A Review,” Int. J. Comput. Sci. Inf. Secure., vol. 6, no. 3, pp. 181–205, 2009, [Online]. Available: http://sites.google.com/site/ijcsis/.
[17] S. S. Nimbhore, G. D. Ramteke, and R. J. Ramteke, “Pitch estimation of Marathi spoken numbers in various speech signals,” Int. Conf. Commun. Signal Process. ICCSP 2013 - Proc., pp. 405–409, 2013, doi: 10.1109/iccsp.2013.6577084.
[18] L. Rabiner and B.-H. Juang, Fundamentals of Speech Recognition. 1993.
[19] S. Furui, “Fifty years of progress in speech and speaker recognition,” J. Acoust. Soc. Am., vol. 116, no. 4, pp. 2497–2498, 2004, doi: 10.1121/1.4784967.
[20] S. K. Gaikwad, B. W. Gawali, and P. Yannawar, “A Review on Speech Recognition Technique,” Int. J. Comput. Appl., vol. 10, no. 3, pp. 16–24, 2010, DOI: 10.5120/1462-1976.
[21] L. R. Rabiner and R. W. Schafer, Introduction to Digital Speech Processing, vol. 1, no. 1–2. 2007.
[22] S. Hase and S. Nimbhore, "Speech Recognition: A Concise Significance," International Conference on Computational Intelligence and Computing Applications (ICCICA), pp. 1-7,2021,doi: 10.1109/ICCICA52458.2021.9697255.
[23] S Hase and S Nimbhore, "Keyword Based Speech Recognition Technique Using Python for C #,". Sci Technol Dev. X(II):453–457.,2021.