A Study of Sindhi Related and Arabic Speech Recognition System

Author's Details:
[1]Dil Nawaz Hakro,[1]Tuba Qureshi, [2]Intzar Lashari, ,[1]Rajota Kharwal, [1]Maryam Hameed
[1]Institute of Information and Communication Technology, University of Sindh, Jamshoro,Pakistan
[2]Institute of Business Administration, University of Sindh, Jamshoro,Pakistan
Corresponding emails: dill.nawaz@gmail.com,

Abstract:
Speech Recognition is the understanding human words by computer that was spoken by the human. These words may be the human language and changing the human language will demand different challenges for the different language which means the algorithms designed for English speech recognition cannot be employed to recognize another language such as Sindhi. It requires entirely new and separate algorithms to understand spoken words for Sindhi language. In this regard, every language and script pose different challenges related to script. This paper introduces a study related to speech recognition systems available in various language specially related to Sindhi language. An emphasis has been given to architecture of automatic speech recognition system, various challenges posed by the scripts with special attention to Sindhi and its related languages.

1. Introduction

In many current situations, speech is the most natural communicative element for a human. There are many applications which are based on the voice recognition such as giving a query in database or information retrieval system, giving a dictation, or generally giving a command to a computer or another device. It plays a vital role otherwise one's hands are required to complete the task. Today's globalised world speech recognition is much more important to communicate with the computer via text and speech (spoken word). Due to speech recognition, typing effort has been removed by dictating the text. Speech recognition is an art of work that selects the words from a trained vocabulary of words. Every spectral vector is identified by a label (phoneme), identification of word based the matching the string of label, phone machines based on label and transition probabilities and Markov chains (CM). The acoustic models were used to match the phonetic elements or phonemes. Phonemes are generated by the label of each word by an acoustic processor in the response to spoken word or input (Bahl et al.,1998). Voice recognition system are based on a micro phone translating a voice into electrical voice signal.to analyze the voice for generating a voice pattern in the form of time frequency distribution frequency analyzer were used, level of voice change sometimes too strong or too week for suitable processing. A voice recognition system consists on a voice input (Muroi et al.,1989). Speech (voice) recognition system provide a simple, efficient interaction amid others for people who have any disabilities. voice recognition system based on four stages:one is acquisition of the data, second in preprocessing, third is feature extraction and fourth is pattern recognition (Gonzalez et al.,2016). The voice recognition system main focus on to make a system more efficient to communicate with the computer, many voice recognition systems are available for other languages but there is not any voice recognition system for Sindhi language this system help for the Sindhi speaker as well as different language speaker

2. Automatic speech recognition (ASR)

The enlargement of devices in automatic speech recognition system able to translate the normal language. Voice recognition are generally describing as: 1) unique or isolated words are separated in the different pause and speech recognition system able to recognize that word and 2) incessant voice recognition in this procedure of identification of sentences is created a repeatedly in a normal method and 3) voice recognizing in sense that the system in which aim is transcription such as a database query system or a robot systems. These systems are replied correctly to a spoken word, request and instruction (Bahl et al., 1983). Automatic speech recognition is an advanced technology, and play a dynamic role in past decades.

2.1 Architecture of ASR

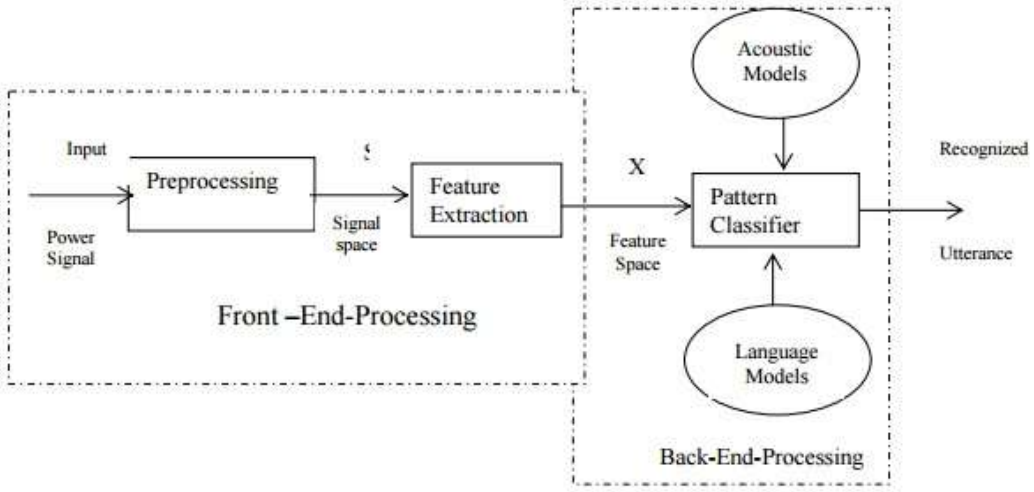

Fig:1 architecture of ASR

In automatic speech recognition system spoken words recognized by a computer which is spoken by a human by using a device such as microphone or telephone and then decode into a printed form of text translating of spoken words into text is strong challenges due to high viability of singles for examples there will be a many different speaker which have a different pronunciation, speak in different styles, different manner, different emotional states and a different accent (Ranjeet et al.,2016). Rosti, (2004) proposed in their research that Automatic speech recognition is the techniques that able to record human speech into written text, these types of system are based on the linguistic framework, auditory framework, dictionary framework, hunt framework

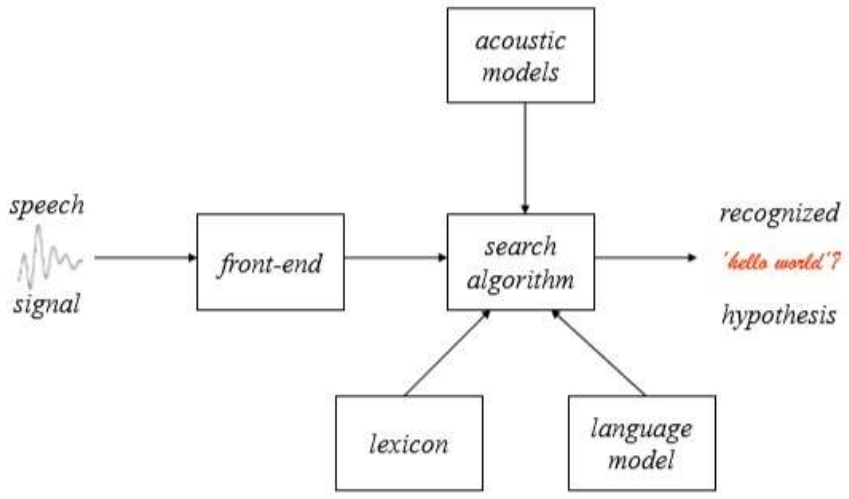

Fig 2: A real speech recognition system

In he main part of automatic speech recognition system is the feature extraction as shown in fig:2 in which is used to processed incoming signals that able to drive meaningful characteristics, and then acoustic model are assembled from the extracted features composed with text records of speech files of database, after the feature

extraction the language model is used to capture the statistical properties of language by assessing of following phones or segment, word, in a sequence of speech and lexicon contains all words, phones and other symbol. When a user speaks a word then it goes to the lexicon with the help of lexicon the search algorithm is used to finding the proper word or phones that have been spoken by the speaker

3 Speech recognition

The procedure of mining text transcription is called speech recognition (Besacier et al., 2014).Speech recognition system is assimilate into a diverse application, to better know the speech recognition system process, there is a two main types of speech recognition 1) speaker dependent these type of system is uncompromising in use , these system established for only a single operator.2) speaker independent these type of system is more convenient in use, developed for a multiple speakers(Mencattini et al., 2014).

which he defines speech recognition is the structure of a system for recording signals to a string of words. Speech recognition is the Voice or it is the capability of a machine or program to recognise and carry out spoken commands. The main purpose of speech recognition is to make a system able to communicate with the computer. To design a well-organized user interface communication is too good. There are many speech recognition systems of other languages as every language has their own rules and regulation. No any speech recognition system for the Sindhi language has been found to authors knowledge (Jurafsky,2000). In the speech recognition system, the word sequences play a dynamic role because computer testing to match the sound with word sequences because language model helps to differentiate words and phrase that are same in sound. There are many applications where the language model is used such as machine translation, speech recognition, handwriting recognition and others Isolated word recognition is used suggested as wordlist grammar the wordlist grammar is suggested for isolated word recognition. “Wordlist “is called a simplest type of model language (Kobashikawa et al., 2014).

3.1 Block Diagram of Speech Recognition

Automatic voice recognition system is composed of modules in which analog signal is captured through a high quality, disturbance, background noise unidirectional microphone in wav form and transformed into a digitized format. (Ghai and Singh., 2012)

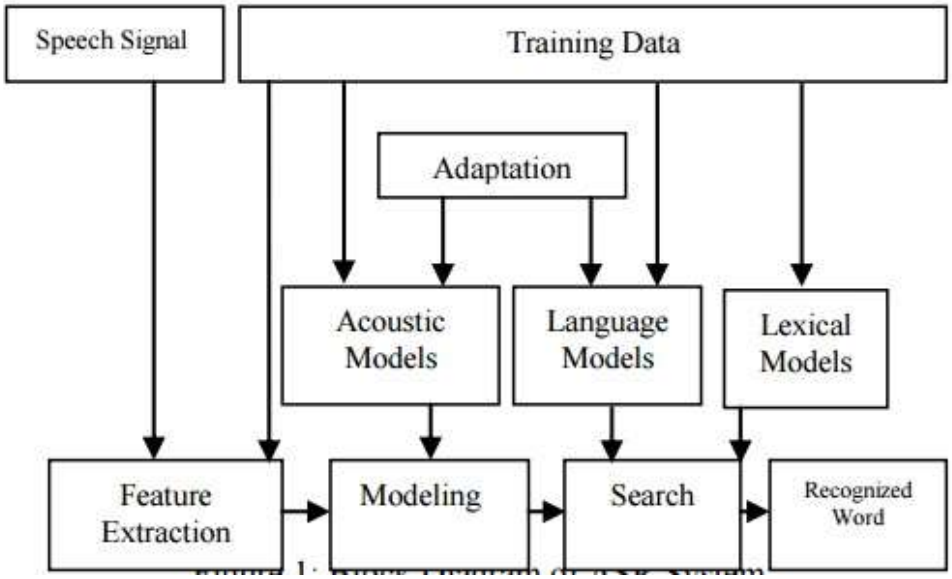

Fig:3 blocked diagram of SR

4 Sindhi language

In the subcontinent of north-west of India many languages were found, Sindhi language is correlated with the Hindi and Urdu language Sindhi language is an Indo-Aryan language The nature of the Sindhi words is polymorphemic better to cite a paper. Sindhi characters are like an Arabic, Hindi, Urdu words. The Sindhi language is ranked the third most spoken languages in Pakistan and in India Sindhi language is documented as the official language (Mahar and Memnon,2010). There are thousands of spoken languages being spoken by throughout the world. Sindhi is spoken by 40 million of people in Sindh province of Pakistan and as well as in the numerous states of the India. In India, the Sindhi language is written in both scripts, Devanagari and Persian-Arabic but in Pakistan, Sindhi is written in only Persian- Arabic script because both areas deeply share the same vocabulary. The Sindhi language consists of 52, alphabet letters (Leghari and Rahman,2010). Sindhi scripts were written in many forms in olden days Sindhi script holds a very rich and historical background. There are three main types of Sindhi script which have been used. 52 letters are contained by the Sindhi script and that letters are associated with the source languages.28 alphabet is containing by the Arabic language, Pashto (44), Urdu language have a (39) character and Persian (32). The Sindhi language have a rich background and it wrote in numerous scripts in different regions such as Shikarpuri, Chowki, Khuda Abad. The nature of Sindhi language words is polymorphemic just like a Urdu, Arabic, Hindi languages words Rahman (2009).

4.1Sindhi syntax

In the Sindhi language, the writing style is much important as compared to the Urdu language. In Sindhi script. Dots play a Significant role of this language because a single dot is responsible for making a word and that word contain a perfect meaning including and excluding more dots can make another character. In this respect, Sindhi establishes the largest extension of the original Arabic script (Hakro et al.,2014). The official language of Sindh province is Sindhi, a predictable view of Sindhi language is a 34.4 million of people in Pakistan, the Sindhi language is the third most spoken language (Ismaili et al.,2014).

Sindhi language, as very limited work, has been reported in the literature in Sindhi speech recognition. In this modern era speech recognition is one of the ground-breaking technologies.

In a Sindhi computing, Substantial work has been done, but there no such work is done on the Sindhi speech recognition (Bhurgari,2010). The Sindhi language have a variant of writing style, according to the standard of Unicode character should store in sequence of manner, if the user want to write a word Sindhi. (Sanjrani et al.,2016) they proposed in their research that the main challenges and problems of Sindhi language are that this language contain bidirectional, cursiveness, the composite set of dots, different structure and shape of the character, and a large character set, character size, compound words, context sensitivity, the pronunciation of words

5 Data Dictionary

Rare Data dictionaries are existing such as Sindhi to English and Sindhi to Sindhi in a computer form (Bhatti, et al., 2014b, Hakro, et al., 2014 and Shah, et al., 2011).

6. General problem of automatic speech recognition

In an automatic speech recognition system, there are so many parameters are including that affects the accuracy of the recognition system such as dependent or independent speaker, isolated or connected word recognition, person's vocabulary, acoustic modeling, language modeling, environment, transducers, perplexity and etc. the problems of automatic speech recognition system including given input speech pronunciation by one person in several times, noisy environment, conflicting between training and testing without complete recognition (Arora & Singh., 2012).

6.1 Human comprehension of speech compared to ASR

According to the speaker perception the speaker like to learn or get more knowledge by hearing an audio not the reading of books, papers, speaker avoid writing several papers on a single topic. Todays the nature of everyone is that they want to do work with voice or hand free system because this is more user friendly, to learn more and get more knowledge To overcome this problem and improving the prediction we construct a statistical model for grammatical structure. However, we are still facing problems to model the knowledge of speaker as well as world knowledge, of course, we cannot construct a model of world knowledge but the question is arising how can we overcome from this problem to measure the human comprehension in the AS

6.2 Noise problem

The main purpose of automatic speech recognition(ASR), is to make an application or system with the high accuracy, and remove the all background noise of spoken word, without fulfilling this condition, the overall results of speech recognizers are highly effective (Gong, 1995; Cole et al., 1995; Torre et al., 2000). a speech is vocalised in sound environments such as ticking clock, playing songs in somewhere in another room. Communication of persons in the background, horns of the vehicles, and these unwanted types of sounds is usually called noise. In ASR to improve the accuracy of speech recognition, we must distinguish all sounds and strainer out the unwanted and disturbing noise from the input signal. Alternative kinds of unwanted sound are called echo effect, unwanted sound appear in the microphone at the time of speaking therefore it seems in the microphone a few milliseconds later (Forsberg, 2003). There are so many conventional techniques are available for improving recognition speech robustness such as eliminating or reducing the mismatches for instance by intensification of the noisy speech, by applying statistical methods for given speech units in the noisy environment simply by training in different noisy conditions (Furui, S., 1997). There are so two important approaches to robustness of speech recognition including "recognized domain approaches" (Varga and Moore 1990; Gales and Young 1996), and the second one is the acoustic recognition model is enhanced or retrained by recognize of noisy speech, feature domain approaches and distorted speech (Boll 1979; Deng et al. 2000; Attias et al. 2001; Frayed al. 2001), while distorted speech and noisy feature are the first denoised and then fed into system of speech recognition whose acoustic model is trained and clean the speech.

6.3 Body Language

Speech is not the only one to interact for human being but body signals play a vital role in interaction namely waving hands, moving eyes, expression of feelings, body gesture, posture, head motions etc. this information is not available in automatic speech recognition system.

6.4 Spoken language is unskilled than written language

There are a lot of variations of expressions find between used in spoken language and written language. Spoken words is less complicated as compared to written words. The main characteristic of spoken words is to remove the unwanted noise and listen the clear spoken word. Written language is one-way communication as compared to spoken language which is totally different and in bi directional communication, the reason is that we provide feedback in only case when we understand the signals from the receiver, so the speech is the type of dialogue. Another big issue is disfluencies in speech such as repetition, expressions, hesitations, slips of tongues, changes of subjects during an utterance etc. in ASR we must analyse and address these differences (Forsberg, M., 2003).

6.5 Variability of Channel

To altered the acoustic wave is the main feature of variability, to record a best voice without the noise, use the best quality of microphone that will not disturb the content of acoustic wave (Forsberg, M., 2003).

6.6 Speaker variability

All speakers have their own speciality in their voices because all human beings have their own personality traits and according to needs and personality. Voice is not the only one factor which makes speakers differentiating but other characteristics also matters for a speaker. Following is the list of these variations.

6.1.1 Realization

The repetition of the similar words multiple times result in different speech signals. In the case that a user want to pronounce the previous speech multiple times so that the similar results can be obtained then even though the same results cannot be achieved as there will always be a different result of your speech produced by the acoustic model (Forsberg, M., 2003).

6.1.2 Speaking style

All human beings have a different style of speaking. Every human has a different way to express their personality. They do not use only their own vocabulary but they possess and use their special approach to pronounce words and emphasis and surely they possess specific style. Different speakers have different speaking styles that vary in a different situation; we do not have same speaking style in the bank, as talking with friends, as with our parents, as communicating with teachers. Every person has a different way to express their emotions via speech. We are speaking in different styles in different situations such as sad, happy, excited, exhausted, stressed, disappointed, frustrated and others. If someone is not happy then he or she speaks slow and the voice will be a little bit low and will speak in high voice (loud) in happy mood along with a smile on the face (Forsberg, M., 2003).

61.1.3 Continue speech

There is no natural pause between the word boundaries during the speech. Mostly the natural pauses seem on a synthetic level like afterwards an expression or a sentence. Conversion of words from signals is a potential problem and there is an only method to face this challenge, which is the applying a fixed gap or a pause in between of the speeches. The availability of lengthy utterances may make it inefficient (Forsberg, M., 2003).

4 Specific problem of Automatic Speech Recognition

There are many speech recognition systems of other languages as every language has their own rules and regulation.

4.1 Implementing SR system interface for Indian language

(Aggarwal and Dave, 2008) presents the development of Automatic Speech Recognition system for Indian languages, after so many trials to fulfil the challenge of the implementing Speech recognition system interface for Indian languages by preparing various algorithms and executed them by high-level language VC++, Speech in and text0out interface is also implemented into it. The analysis consists of the estimation of the system using in the room, and the speech captured by microphone and sound blaster. The frequency of speech signal is 16000 Hz and the size of speech sample was 8 bits. Therefore 10.1917 dB was used to detect the words. Hidden Markov model is used to recognise the isolated Hindi words. During the speech recognition, so many different words are hypothesised in contrast to speech signal. To measure the probabilistic of a given word, the word is fractured into constituent phones and the phones is collapsed by HMMs. The combine probabilistic of phones are applied by Acoustic model. To implement all these process successfully two main important steps is need to follow the first one is transcript preparation and the second is dictionary preparation. At the time of transcript preparation a text file is arranged in which the complete vocabulary of the constructed ASR system is reported in Unicode. On the other hand dictionary brings the pronunciation for the words used in a language model. The

words pronounced by speaker broken it into a chain of sequence of sub words used by acoustic model. Dictionary is a file that produce a mapping by grapheme to phenome for a given word by speaker.

4.2 An Example of Dictionary

Aggarwal and Dave, (2008) designed a common interface for training, recording and for testing purpose. This interface shows in figure 1

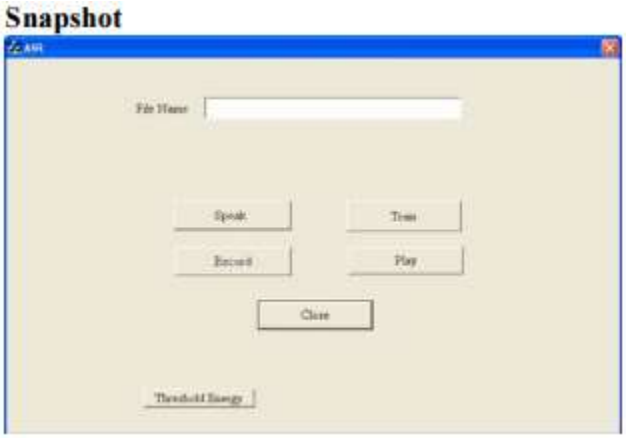

Figure 4: Interface for ASR Functioning

To record speech samples given by speaker, click the record button of the given interface. After this, the train button is pressed to do statistical modelling and to recognise the word speak button is pressed to speech a word we want to recognise. Note down that the word we want to recognise is must be already available in the dictionary. After the recording and training, testing is performed. At the time of testing it figure out the right word math and shows the accuracy rate after matching of each word. Two hundred isolated words are recording by speakers in Hindi language and trained them by different number of times. randomly fifty words are choosing for testing is made and the outcome are as given below in figure 2.

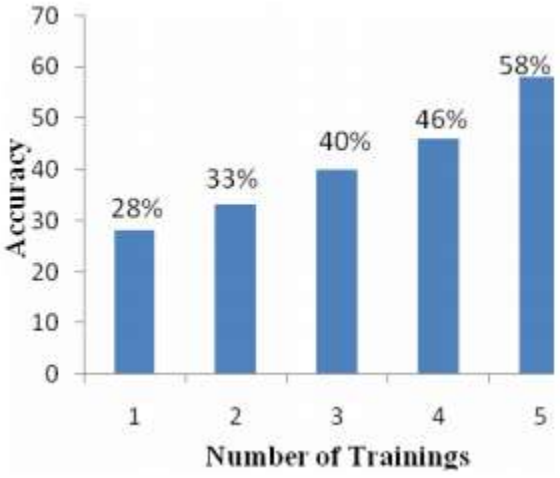

Figure 5: Accuracy vs. No. of Training

5. Summary

Greatest challenges of the modern science are to create a natural human source to communicate with the computer. Talking in front of computer via speech is an easiest way to command rather than sitting and writing via keyboard. We are working on the speech recognition system of Sindhi language and the research will help

the persons with the disability of hands and persons who dislike typing. Speech recognition is one of the ground-breaking technologies. Every language has a different challenge; this paper present the speech recognition system in Sindhi language

6. Future work

In a modern technology speech recognition is increasing day by day. There are too many works done on Sindhi computing, there is much need to develop a system that recognize the Sindhi speech or a spoken word to help the user write in own language or control machines via speech. In this research computer, will be able to recognize and write a few words of Sindhi language and make a system able to control a computer with spoken words that are trained, few words or sentence will be used to handle the window through the voice commands. Sindhi voice recognition system focus is to satisfy the user of other language or a Sindhi speaker.

7. References

Aggarwal, R. K., & Dave, M. (2008, January). Implementing a Speech Recognition System Interface for Indian Languages. In IJCNLP (pp. 105-112).

Arora, S. J., & Singh, R. P. (2012). Automatic speech recognition: a review.*International Journal of Computer Applications*, *60*(9).

Attias, H., Platt, J. C., Acero, A., & Deng, L. (2001). Speech denoising and dereverberation using probabilistic models. Advances in neural information processing systems, 758-764.

Bahl, L. R.; DeGennaro, S. V.; Mercer, R. L. & others (1988), 'Speech recognition system', Google Patents, US Patent 4,718,094.

Bahl, Lalit R., Frederick Jelinek, and Robert L. Mercer. "A maximum likelihood approach to continuous speech recognition." IEEE transactions on pattern analysis and machine intelligence 2 (1983): 179-190

Besacier, L.; Barnard, E.; Karpov, A. & Schultz, T. (2014), 'Automatic speech recognition for under-resourced languages: A survey ', Speech Communication 56, 85 - 100.

Bhatti, Z., I. A. Ismaili, D. N. Hakro, and A. Waqas, (2014b). Unicode Based Bilingual Sindhi-English Pictorial Dictionary for Children. American Journal of Software Engineering, 2(1), 1-7. DOI: 10.12691/ajse-2-1-

Bhurgari, A.M., "Enabling Pakistani Languages through Unicode", Retrieved from http://download. Microsoft. com/download/1/4/2/142aef9 f-1a74-4a24-b1f4-782 d48d41a6d/PakLang.pdf on 20th August, 2010

Boll, S. (1979). Suppression of acoustic noise in speech using spectral subtraction. IEEE Transactions on acoustics, speech, and signal processing, 27(2), 113-120.

Cole, R., Hirschman, L., Atlas, L., Beckman, M., Biermann, A., Bush, M., ... & Hermansky, H. (1995). The challenge of spoken language systems: Research directions for the nineties. *IEEE Transactions on Speech and Audio Processing*, *3*(1), 1-21.

D. N. HAKRO, I. A. ISMAILI, A. Z. TALlB, Z. BHATTI, G. N. MOJAI , Issues and Challenges in Sindhi OCR, Sindh University Research Journal (Science Series), 2014

De La Torre, A., Fohr, D., & Haton, J. P. (2000, October). Compensation of noise effects for robust speech recognition in car environments. In*INTERSPEECH* (pp. 730-733).

Deng, L., Acero, A., Plumpe, M., & Huang, X. (2000, October). Large-vocabulary speech recognition under adverse acoustic environments. In INTERSPEECH (pp. 806-809).

Forsberg, M. (2003). Why is speech recognition difficult. The *Chalmers University of Technology*.

Frey, B. J., Deng, L., Acero, A., & Kristjansson, T. T. (2001, September). ALGONQUIN: iterating Laplace's method to remove multiple types of acoustic distortion for robust speech recognition. In INTERSPEECH (pp. 901-904).

Furui, S. (1997, April). Recent advances in robust speech recognition. In ESCA-NATO Workshop on Robust speech recognition for unknown communication channels (pp. 11-20).

Gales, M. J., & Young, S. J. (1996). Robust continuous speech recognition using a parallel model combination. IEEE Transactions on Speech and Audio Processing, 4(5), 352-359.

Gales, M. J., & Young, S. J. (1996). Robust continuous speech recognition using a parallel model combination. IEEE Transactions on Speech and Audio Processing, 4(5), 352-359.

Ghai, W., & Singh, N. (2012). Analysis of automatic speech recognition systems for indo-aryan languages: Punjabi a case study. Int J Soft Comput Eng, 2(1), 379-385.

Gong, Y. (1995). Speech recognition in noisy environments: A survey.*Speech communication*, *16*(3), 261-291.

Gonzalez, R., Muñoz, J., Salazar, J., & Duque, N. (2016, July). Voice Recognition System to Support Learning Platforms Oriented to People with Visual Disabilities. In *International Conference on Universal Access in Human-Computer Interaction* (pp. 65-72). Springer International Publishing

Hakro, D. N., I. A. Ismaili, A. Z. Talib, Z. Bhatti, and G. N. Mojai, (2014) Issues and Challenges in Sindhi OCR. Sindh University Research Journal (Science Series). Vol. 46 (2). 143-152.

Hakro, D. N., Ismaili, I. A., Talib, A. Z., Bhatti, Z., & Mojai, G. N. (2014). Issues and challenges in Sindhi OCR. Sindh University Research Journal-SURJ (Science Series), 46(2). He, Y.; Sun, G. & Han, J. (2015), 'Spectrum enhancement with sparse coding for robust speech recognition ', Digital Signal Processing 43, 59 - 70.

Ismaili, I. A.; Bhatti, Z. & Shah, A. A. (2014), 'Design & Development of the Graphical User Interface for the Sindhi Language', *arXiv preprint arXiv:1401.1486*.

J. H. M. Daniel Jurafsky. Speech and Language Processing, An introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition. Prentice Hall, Upper Saddle River, New Jersey 07458, 2000.

Kobashikawa, S.; Asami, T.; Yamaguchi, Y.; Masataki, H. & Takahashi, S. (2014), 'Efficient data selection for speech recognition based on prior confidence estimation using speech and mono phone models ', Computer Speech & Language 28(6), 1287 - 1297.

Leghari, Mehwish, and Musee U. Rahman. "Towards Transliteration between Sindhi Scripts by using Roman Script." the Conference on Language and Technology, National Language Authority Islamabad, Pakistan. 2010.

Mahar, Javed Ahmed, and Ghulam Qadir Memon. "Rule-based part of speech tagging of Sindhi language." Signal Acquisition and Processing, 2010. ICSAP'10. International Conference on. IEEE, 2010.

Mencattini, A.; Martinelli, E.; Costantini, G.; Todisco, M.; Basile, B.; Bozzali, M. & Natale, C. D. (2014), 'Speech emotion recognition using amplitude modulation parameters and a combined feature selection procedure ', Knowledge-Based Systems 63, 68 - 81.

Muroi, T., Yasuda, S., Kawamoto, T., & Fujimoto, J. (1989). *U.S. Patent No. 4,833,713*. Washington, DC: U.S. Patent and Trademark Office

Rahman, M. U., (2009), "Sindhi Morphology and Noun Inflections", Proceedings of the Conference on Language & Technology", pp. 74-81

Ranjeet, P.; Prakash, T.; Amruta, S. & Monali, S. (2016), 'Automatic Speech Recognition System', *Imperial Journal of Interdisciplinary Research* **2**(3), 165--169.

Rosti, A.-V. I. (2004). Linear Gaussian Models for Speech Recognition. PhD thesis, University of Cambridge, Wolfson College, U

Saeeda Naz and Khizar Hayat and Muhammad Imran Razzak and Muhammad Waqas Anwar and Sajjad A. Madani and Samee U. Khan, The optical character recognition of Urdu-like cursive scripts, Pattern Recognition, Handwriting Recognition and other {PR} Applications, vol. 47, No.3, pp. 1229-1248,2014

Sanjrani, A. A.; Baber, J.; Bakhtiar, M.; Noor, W. & Khalid, M. (2016), Handwritten Optical Character Recognition system for Sindhi numerals, *in* 'Computing, Electronic and Electrical Engineering (ICE Cube), 2016 International Conference on', pp. 262--267.

Shah, Z. A., and G. M. Mashori, (2011). Oxford English-Sindhi Dictionary: A Critical Study in Lexicography. ELF Annual Research Journal, 13, 37-46.

Varga, A., & Moore, R. K. (1990, April). Hidden Markov model decomposition of speech and noise. In Acoustics, Speech, and Signal Processing, 1990. ICASSP-90., 1990 International Conference on (pp. 845-848). IEEE.