© 2023 IJNRD | Volume 8, issue 11 November 2023 | ISSN: 2456-4184 | IJNRD.ORG

Unlocking Voice-Powered AI with SpeakEZ GPT

1st T Heli Singh
Dept of CSE
Chandigarh University
Punjab, India
20bcs7301@gmail.com

2nd Priyansh Yadav
Dept of CSE
Chandigarh University
Punjab, India
Priyanshy55@gmail.com

3rd Sumer Singh Rawat
Dept of CSE
Chandigarh University
Punjab, India
sumerrawat321@gmail.com

***Abstract*—Voice-powered artificial intelligence (AI) is reshaping the way we interact with technology, making it more intuitive and accessible. In this paper, we introduce SpeakEZ GPT, a groundbreaking conversational AI system that empowers users to engage with GPT-3.5 through voice commands and receive spoken responses. This innovative project bridges the gap between human communication and AI, leveraging progressive dialog acknowledgment and ordinary linguistic dispensation machinery. We delve into the development process, system architecture, and potential applications of SpeakEZ GPT, shedding light on its role in revolutionizing voice-driven AI interactions.The emergence of voice-powered AI represents a significant shift towards more intuitive and natural interactions, where users can converse with machines as if they were talking to another person. In this paper, we introduce SpeakEZ GPT, an innovative conversational AI system designed to unlock the potential of voice-powered AI by allowing users to interact with the GPT-3.5 language model through voice commands and receive spoken responses.**

***Index Terms*—Voice-Powered AI, Conversational AI, SpeakEZ GPT, Speech Recognition, Natural Language Processing, Text-to-Speech, Human-Computer Interaction, Accessibility, Virtual Assistants, Multilingual Support.**

I. Introduction

Traditional modes of human-computer interaction often require users to adapt to machine interfaces, leading to friction and inefficiency. Voice-powered AI seeks to revolutionize this paradigm, allowing users to communicate naturally with machines. SpeakEZ GPT represents a significant step towards realizing this vision. In this review, we evaluate the project's key features and contributions[1]. Artificial intelligence (AI) has transformed the method we interrelate through knowledge besides has been integrated into many aspects of our everyday life. The progression of human-computer interaction from conventional techniques, such typing and clicking, to a more intuitive and natural approach through voice-powered AI lies at the heart of this transition. Herein dissertation, we customary obtainable on an expedition to examine and assess the ground-breaking initiative known as ”SpeakEZ GPT.” The conversational AI system that this research presents enables users to interact with the GPT-3.5 language model via voice commands and hear spoken answers[2]. SpeakEZ GPT aims to transform the voice-powered AI environment and improve user experiences by spanning the breach among humanoid communication besides mechanism understanding.

˙ff

The traditional human-computer interface frequently requires people to adjust to machine interfaces, which creates a sense of complexity and separation. However, voice-powered AI aims to change this dynamic by enabling consumers to speak naturally with technology. Users should be able to communicate with AI systems in a way that is similar to speaking with a friend or coworker through the creation of an intuitive and inclusive experience. Technology might become more democratized by moving toward more organic contact, making it more usable and accessible[3].A key effort in the field of voice-powered AI is SpeakEZ GPT. To understand user intent and provide replies that resemble those of a person,it customs commanding language acknowledgement knowledge to adapt vocal arguments into manuscript. This manuscript is formerly analyzed using natural language processing (NLP) models. This fusion of technology creates an AI system that can have true conversational interactions with users by understanding not just their speech but also their intentions. SpeakEZ GPT is designed with the primary aim of enhancing user interactions with AI. Its simplicity lies in voice-based commands, fostering a more accessible and engaging experience. Users initiate conversations with a straightforward ”Hey AI,” followed by their queries or requests. This user-centric approach places technology at the service of users, ensuring a more intuitive and user-friendly interaction.Cutting-edge the subsequent pieces of this paper, we investigate profounder hooked onthe architecture, user experience, applications, and future prospects of SpeakEZ GPT[5]. As we explore the intricacies of this pioneering project, we gain valuable insights into the transformative potential of voice-powered AI in reshaping the human-technology relationship. SpeakEZ GPT stands as a beacon in this journey towards a future where technology adapts to human communication, making it an integral and seamless part of our daily lives.

II. Types of Speech Recognition

By characterizing what kinds of utterances they are able to recognize, language gratitude schemes may stand divided into a amount of separate collections. These lessons fall under the following categories:

A. *Speaker-Dependent Recognition*

We now live entirely online thanks to speech recognition, a revolutionary technology that transcribes spoken words into text or commands. Speaker-Dependent Recognition (SDR) is one of the important subcategories of speech recognition. SDR is distinguished by its personalized nature, in contrast to certain other types of speech recognition, as it has to be trained on a particular person's voice to attain maximum accuracy and performance.Speaker-Dependent As the name implies, recognition is based on the distinctive vocal traits of a particular speaker. In this method, the speech recognition system is specifically taught to comprehend and record the speech of a specific person[6]. For the system to learn to detect the unique peculiarities of that person's speech, the user must provide a set of voice samples, often reciting specified sentences or phrases.

Speaker-Dependent Recognition is a powerful tool in the realm of speech recognition, offering exceptional accuracy and a personalized user experience. While it may not be suitable for all scenarios, its ability to tailor recognition to an individual's unique voice characteristics makes it invaluable in applications where precision and personalization are paramount[7]. As technology continues to advance, Speaker-Dependent Recognition holds promise in further enhancing user interactions with voice-activated systems and devices.

B. *Speaker-Independent Recognition*

Speech recognition technology has revolutionized the way we communicate with machines and other technologies in our daily lives. Speaker-Independent Recognition (SIR) is a major category in the field of voice recognition that contrasts with Speaker-Dependent Recognition (SDR)[8]. SIR is intended to identify a wide range of voices and accents without prior training, in contrast to SDR, which depends on individualized instruction, giving it a flexible option for varied applications.Cutting-edge the circumstance of Utterer-Self-governing Gratitude, voice exercise is not necessary for this type of automated speech recognition (ASR). Instead, it is designed to comprehend and transcribe any speaker's speech, giving it a more comprehensive and all-encompassing approach. SIR is intended to be flexible and tolerant with the intention of recognizing the speech of a wide range of people[9].

Speaker-Independent Recognition is a vital component of modern speech recognition technology. Its ability to understand and transcribe the speech of diverse speakers without prior training makes it a versatile choice for applications where adaptability and inclusivity are paramount. As technology continues to advance, SIR will likely play an increasingly central role in our interactions with voice-activated systems, contributing to a more accessible and user-friendly future[10].

C. *Isolated Word Recognition*

With the use of speech recognition technology, robots can now comprehend and analyze spoken words. Isolated Word Recognition (IWR), one of the several speech recognition techniques, stands out as a fundamental and adaptable approach. IWR focuses on isolating words or brief phrases as opposed to continuous speech recognition, which transcribes whole utterances. This focused strategy has found use in a variety of fields, including interactive voice response systems and voice dialing on mobile phones.As the name implies, isolated word recognition is focused on the recognition and transcription of single words or brief sentences[11]. IWR systems are made to perform well when discrete, distinct words or instructions are given, rather than having to cope with the difficulties of continuous speech. Usually, pauses are used to provide distinct borders between these isolated words to aid in identification.

Isolated Word Recognition is a valuable approach within the broader field of speech recognition. Its ability to transcribe individual words or short phrases with high accuracy and efficiency makes it well-suited for applications that rely on discrete spoken commands or keywords[12]. As technology continues to advance, IWR will continue to play a pivotal role in enhancing the user experience in voice-activated systems and applications.

D. *Noise-Adaptive Recognition*

Clarity of speech is crucial in the field of voice recognition technology. However, background talk, traffic, equipment, and other noises are frequently present in real-world settings. Noise-Adaptive Recognition (NAR) is a particular a fix for the issue of properly detecting speech in loud environments. In spite of auditory noise, spoken words are accurately transcribed by NAR systems because they are designed to adapt to and reduce ambient noise[13].Automatic speech recognition (ASR) that can adapt to different types and amounts of background noise is known as "Noise-Adaptive Recognition." NAR systems are able to adapt, filter, and improve incoming audio to assist clear speech recognition, in contrast to traditional speech recognition systems, which may struggle in loud circumstances. The fundamental objective is to maintain a signal-to-noise ratio that is advantageous for accurate transcribing.

Noise-Adaptive Recognition is a crucial advancement in speech recognition technology, addressing the real-world challenge of noise interference. Its ability to adapt to varying noise levels and enhance the clarity of spoken words makes it invaluable in applications where precise communication is essential[15]. As NAR technology continues to evolve, it promises to improve user experiences, enhance accessibility, and increase the reliability of voice-activated systems in noisy environments.

III. AUTOMATIC SPEECH RECOGNITION SYSTEM

Recognizing speech automatically (ASR) is a remarkable technology that has revolutionized human-computer interaction by enabling machines to transcribe and understand spoken language. ASR systems, often referred to as speech recognition or voice recognition systems, have a wide range of applications, from voice assistants like Siri and Google Assistant to transcription services, customer service automation, and healthcare documentation. This article delves into the world of ASR, exploring its key components, working principles, applications, challenges, and future prospects[16]. ASR is a tool for turning spoken words into written text. In essence, it converts human speech's auditory information into a format that computers can understand and interpret. Making oral communication available to computers will enable them to correctly translate spoken words and, in certain situations, comprehend the meaning underlying those words.

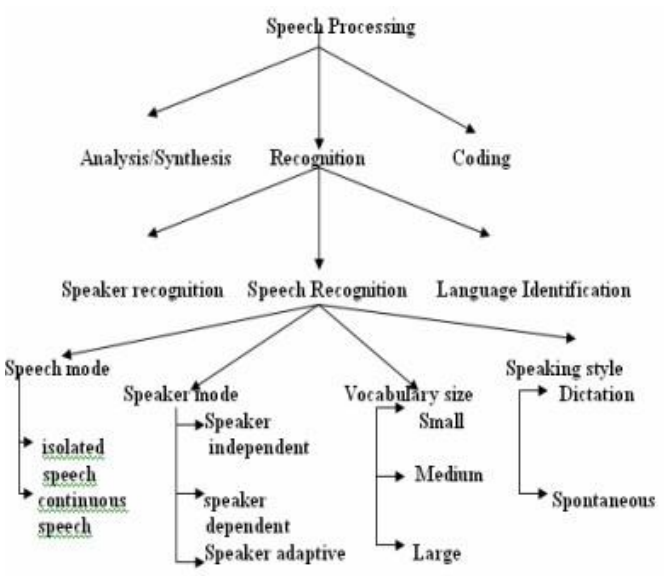

Fig. 1. Speech processing

The applications for voice processing are highlighted in the following tree structure. Systems for automatic speech recognition can be categorized as indicated in figure 2 depending on the chosen criterion. ASR systems employ a series of steps to convert spoken language into text:

Audio Input: The ASR system begins by receiving an audio input, typically in the form of recorded speech or a live audio stream.

Feature Extraction: The audio signal is processed to extract relevant acoustic Spectrograms and Mel-frequency cepstral coefficients (MFCCs) are examples of features. These features represent the sound characteristics of the speech[17].

Phoneme Recognition: The extracted features are analyzed to recognize individual phonemes. The acoustic model plays a crucial role in identifying these phonemes by matching the audio features to phonetic representations.

Language Modeling: The language model factors in the recognized phonemes and uses linguistic context to determine the most likely words and phrases. It assigns probabilities to different word sequences.

Word Transcription: The ASR system combines the results of the acoustic and language models to transcribe the spoken words into text. The transcription may contain recognized words, and sometimes, it may include special markers for uncertain or ambiguous segments.

Post-Processing: The final transcription can undergo post-processing to correct errors, improve grammar, and ensure coherence. This step aims to produce a refined and accurate transcript.

IV. Applications of Automatic Speech Recognition (ASR)

ASR technology finds extensive applications across various domains:

A. *Voice Assistants*

Voice-activated digital assistants including Echo from Amazon, Siri from Apple, and the Assistant from Google rely on ASR to understand and respond to user voice commands and queries.The way we engage with our gadgets and access information has changed as a result of voice assistants, which have emerged as a transformational force in the world of technology[18]. Our daily lives have integrated these sophisticated, voice-activated technologies, which provide an easy and natural method to interact with the digital world. Voice assistants have ushered in a new era of human-machine connection, making jobs more practical, effective, and accessible, from Siri and Google Assistant to Amazon's Alexa and beyond.

B. *Transcription Services*

ASR is used in transcription services for converting audio recordings of meetings, interviews, and speeches into written text quickly and cost-effectively. The demand for precise and effective transcribing services has never been greater than it is in today's fast-paced, information-driven society. Whether the spoken language is captured via interviews, meetings, lectures, or audio recordings, transcription services are essential for turning it into written text[19]. These services are used by a variety of sectors, including journalism, education, healthcare, and the legal sector. They make it simple and accurate for professionals to access, evaluate, and store spoken information.

C. *Customer Service Automation*

Many companies deploy ASR in their interactive voice response (IVR) systems to automate customer service interactions, reducing the need for human operators.Businesses are increasingly resorting to automation to streamline and improve their customer service operations at a time of fast technological breakthroughs and changing consumer expectations. Customer service automation, often known as CSA, is a revolutionary method of providing assistance, answering questions, and guaranteeing customer happiness. Modern resources including chatbots, self-service portals, and artificial intelligence (AI) are used to offer clients quick, dependable support all the time[20].

D. *Healthcare Documentation*

ASR is valuable in healthcare for transcribing medical dictations, improving documentation accuracy, and reducing administrative burden on healthcare professionals.Healthcare practitioners may accurately and effectively record, transmit, and manage patient information with the help of healthcare documentation, which is essential to the delivery of high-quality healthcare services. It includes the development and upkeep of comprehensive medical records that include patient histories, treatment schedules, diagnoses, and results. Healthcare documentation has changed to improve patient care, lower mistakes, and boost healthcare outcomes in an era of technology breakthroughs and the move toward electronic health records (EHRs)[21].

E. Language Translation

ASR technology plays a role in real-time language translation applications, helping users communicate across language barriers.Although language is a basic and varied form of human expression, it may also be a hindrance to efficient communication, especially in the context of our world's growing interconnectedness. People from various linguistic origins may communicate with one another thanks to language translation, a potent and indispensable instrument. It is essential for promoting global collaboration, cultural interchange, corporate growth, and intercultural understanding[22]. As the globe becomes increasingly interconnected due to globalization, language translation is more important than ever.

F. Accessibility Tools

ASR assists individuals with disabilities by providing voice-to-text capabilities for communication and interaction with devices.

G. Voice-Controlled Devices

Smartphones, smart speakers, and other voice-controlled devices utilize ASR for hands-free operation and voice commands.The way we communicate with technology has been changed by voice-controlled gadgets, which have increased the ease, convenience, and accessibility of human-computer interaction. These gadgets let users carry out a variety of activities using voice commands, from setting reminders and managing smart homes to surfing the internet and receiving information. These devices are driven by powerful speech recognition technology and artificial intelligence (AI)[23]. Voice-controlled gadgets are a testament to the continual evolution of human-machine interaction in an era of rapid technological innovation.Voice-controlled devices represent a significant advancement in the quest for intuitive human-computer interfaces. The concept of talking to machines has long been a staple of science fiction, but recent technological breakthroughs have transformed this vision into reality.

V. Artificial Intelligence ChatBot with Speech Recognition

A new age of human-machine contact has begun with the creation and incorporation of voice recognition technologies into AI chatbots. These cutting-edge chatbots, which combine artificial intelligence with natural language processing, allow users to converse verbally, extending the range and practicality of AI-driven interactions[24]. AI chatbots with speech recognition are revolutionizing how we engage with technology and get information in this era of seamless and intuitive communication.The emergence of AI chatbots with speech recognition is the result of significant progress in two key areas:

Artificial Intelligence (AI): Chatbots can now comprehend NLP (natural language processing) and machine learning thanks to AI, process, and respond to human language with remarkable accuracy[25]. These advancements have made it possible for chatbots to engage in natural and context-aware conversations.

Speech Recognition: Speech recognition technology, also known as voice recognition software (ASR), has evolved to the point where it can accurately transcribe spoken language into text. This technology is a fundamental component of AI chatbots with speech recognition, as it allows the chatbot to understand user voice commands[26]. Convolutional Neural Networks (CNNs), which build a deep convolutional neural network with input, convolutional, pooling, fully connected, and output layers, are the first technique for learning a real multilayer neural network structure. As a specific instance, convolutional neural networks are utilized to cut down on the amount of spatial connection parameters and speed up general forward BP training. Before the advent of unsupervised layer-by-layer greedy neural network initial training, it was nearly impossible to train deep neural networks with BP algorithms.

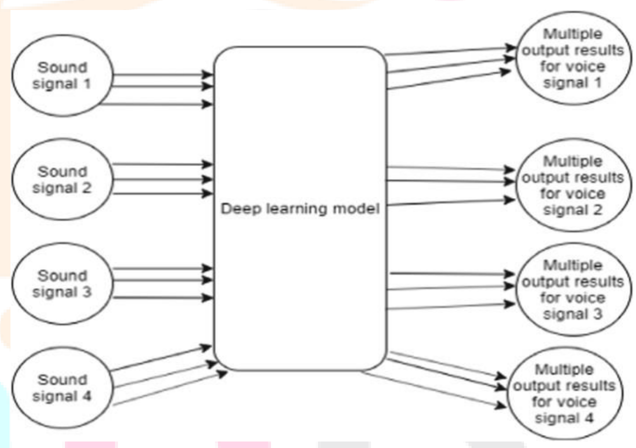

Fig. 2. Overview of the Model

The RNN-T is employed in the input data through the acoustic model output, as shown in Fig. 2. There is not a one-to-one relationship, but one input result might have several outputs. The language model regulates the quantity of output outcomes.

VI. Challenges and Future Advancements

Language is a critical procedure of message for together humans and computers. Language gratitude is widely applicable in the fields of processer discipline, medication, and additional sciences. A actual-while language recognizer's development may be hampered by a hostile environment or by the human body's structure as a whole.Underneath is a conversation of approximately of the main problems Language Acknowledgement Arrangements antagonize [26]. Homophones Homophones are disputes that comprehensive the similar after spoken then take distinct meanings(for example, ”There” and ”Their,” ”Be” and ”Bee”). It is exceedingly challenging for speech recognition systems to determine which word is the intended one at the word level. According to recent findings, due to varying dialects, speech recognition systems often reach 94 to 99 percent accuracy.

Intensive Use of Computer Power- The computer’s CPU must work very hard to track the arithmetical representations obligatory aimed at speech gratitude. One of the causes of this is the requirement to keep track of each step of the word recognition process in case the system wants to go back and find the correct term[27].

Recognition of Punctuation Marks- It has been noted that when voice is converted to text on a screen, punctuation marks are not recognized in their original form; valid words are. The difficulty of the tempo that determines these punctuation marks is handled with a variety of solutions. However, a better option is still to come[28].

Accent- The speaking accent varies depending on societal and individual circumstances (such as physiological and cultural factors)[29]. In fact, performance suffers when accented speech and non-native speech are recognized compared to native speech. Studies have demonstrated that people’s accents change while chatting to friends and parents.

Speed of Speech- Segments of continuous, fast-moving voice signals are challenging for speech recognition algorithms to distinguish. Speaking at different rates depends on the context and any physical strain[30]. Phoneme reduction, temporal expansions, and compression can all alter pronunciation depending on how quickly someone speaks.

VII. CONCLUSION

Reviewing many facets of Platforms can employ technology that recognizes voices to produce systems that recognize speech is the goal of this essay. The study begins by describing what actually caused Speech Recognition Systems to originate. It then goes on to detail how the conversion of speech occurs in distributed real-time systems. This study briefly compares the models and algorithms that were and are now being used to create various Speech Recognition Systems, and it examines the developments that have occurred since the creation of conventional Speech Recognition Systems. Voice Activity Detector, the AURORA framework, and other tools and frameworks have all been created to help Speech Recognition Systems overcome their difficulties. The world is likely to witness considerably improved language comprehension by robots in the next years because to technological advancements in computation that have carried language gratitude knowledge where position wherever things are lot improved than they stood in the past. A significant development in voice-powered AI, SpeakEZ GPT makes technology more receptive to human speech. It is a critical step toward a future in which robots smoothly integrate into our daily lives by delivering natural speech interactions with AI models. SpeakEZ GPT is a cutting-edge technology that is reshaping the field of natural human-computer interactions as voice-powered AI advances. It is significant because it holds the key to a future in which technology is a natural extension of human talents.

REFERENCES

[1] VOICE RECOGNITION SYSTEM: SPEECH-TO-TEXT Prerana Das, Kakali Acharjee, Pranab Das and Vijay Prasad JAFS—ISSN 2395-5554 (Print)—ISSN 2395-5562 (Online)—Vol 1(2)—November 2015.

[2] Volume 5 Issue V, May 2017 IC Value: 45.98 ISSN: 2321-9653 International Journal for Research in Applied Science Engineering Technology (IJRASET) ©IJRASET: All Rights are Reserved 262 Voice RecognitionTechnique: A Review Nisha.

[3] (IJCSIS) International Journal of Computer Science and Information Security, Vol. 6, No. 3, 2009 Speech Recognition by Machine: A Review M.A.Anusuya and S.K.Katti.

[4] Speech Recognition: A Complete Perspective Ashok Kumar, Vikas Mittal international Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-7 Issue-6C, April 2019.

[5] R. Nicole, “Title of paper with only first word capitalized,” J. Name Stand. Abbrev., in press.

[6] Bennett, I.M., Babu, B.R., Morkhandikar, K., Gururaj, P. 2015. Distributed Real-time Speech Recognition, Naunce Communication Inc., Patent No. US 9,076,448.

[7] Bennett, I.M., Babu, B.R., Morkhandikar, K., Gururaj, P. 2014. Speech Recognition System Interactive Agent, Naunce Communication Inc., Patent No. US 8,762,152 B2, 24 June 2014.

[8] Petkar, H. 2016. A review of Challenges in Automatic Speech Recognition”, International Journal of Computer Applications (0975- 8887), 151(3), 23-29.

[9] Ramírez, J., Górriz, J.M., Segura, J.C. 2007. Voice Activity Detection, Fundamentals and Speech Recognition Systems Robustness, University of Granada, Spain.

[10] Hirsch, H.G., Pearce, D. 2000. The Aurora Experimental framework for the performance evaluation of Speech Recognition System under noisy conditions, ASR-2000 Automatic Speech Recognition: Challenges for the new Millennium Paris, France, 181-188.

[11] Chadha, N., Gangwar, R.C., Bedi, R. 2015. Current Challenges and Applications of Speech Recognition Process using Natural Language Processing: A Survey, International Journal of Computer Applications (0975-8887), 131(11), 28-31.

[12] Sadaoki Furui, 50 years of Progress in speech and Speaker Recognition Research, ECTI Transactions on Computer and Information Technology,Vol.1. No.2 November 2005

[13] K.H.Davis, R.Biddulph, and S.Balashek, Automatic Recognition of spoken Digits, J.Acoust.Soc.Am., 24(6):637-642,1952.

[14] D.B.Fry, Theoritical Aspects of Mechanical speech Recognition, and P.Denes, The design and Operation of the Mechanical Speech Recognizer at Universtiy College London, J.British Inst. Radio Engr., 19:4,211-299,1959.

[15] T.Sakai and S.Doshita, The phonetic typewriter, information processing 1962, Proc.IFIP Congress, 1962.

[16] D.R.Reddy, An Approach to Computer Speech Recognition by Direct Analysis of the Speech Wave, Tech.Report No.C549, Computer Science Dept., Stanford Univ., September 1966.

[17] V.M.Velichko and N.G.Zagoruyko, Automatic Recognition of 200 words, Int.J.Man-Machine Studies,2:223,June 1970

[18] H.Sakoe and S.Chiba,Dynamic Programming Algorithm Optimization for Spoken Word Recognition,IEEE Trans.Acoustics, Speech, Signal Proc.,ASSP-26(1):43- 49,February 1978.

[19] F.Itakura, Minimum Prediction Residula Applied to Speech Recognition,IEEE Trans.Acoustics, Speech,Signal Proc., ASSP-23(1):67-72,February 1975.

[20] C.C.Tappert,N.R.Dixon, A.S.Rabinowitz, and W.D.Chapman, Automatic Recognition of Continuous Speech Utilizing Dynamic Segmentation, Dual Classification, Sequential Decoding and Error Recover, Rome Air Dev.Cen, Rome, NY,Tech.Report TR-71- 146,1971.

[21] .F.Jelinek, L.R.Bahl, and R.L.Mercer, Design of a Lingusistic Statistical Decoder for the Recognition of Continuous Speech, IEEE Trans.Information Theory,IT- 21:250-256,1975.

[22] .L.R.Rabiner, S.E.Levinson, A.E.Rosenberg, and J.G.Wilpon, Speaker Independent Recognition of Isolated Words Using Clustering Techniques, IEEE Trans. Acoustics, Speech, Signal Proc., ASSP-27:336-349,August 1979.

[23] H.Sakoe, Two Level DP Matching A Dynamic Programming Based Pattern Matching Algorithm for Connected Word Recognition, IEEE Trans. Acoustics, Speech, Signal Proc., ASSP-27:588-595, December 1979.

[24] C.S.Myers and L.R.Rabiner, A Level Building Dynaimc Time Warping Algorithm for Connected Word Recognition, IEEE Trans. Acoustics, Speech Signal Proc., ASSP-29:284-297,April 1981

[25] C.H.Lee and L.R.Rabiner, A Frame Synchronous Network Search Algorithm for Connected Word Recognition, IEEE Trans. Acoustics, Speech, Signal Proc., 37(11):1649-1658, November 1989.

[26] C. J. Leggetter and P. C. Woodland, Maximum likelihood linear regression forspeaker pp.adaptation of continuous density hidden Markov models, Computer Speech and Language, 9, 171-185, 1995.

[27] Jerome R. Bellegarda,Large Vocabulary Speech Recognition with Multispan statistical Language Models, 76 IEEE Transactions On Speech Audio Processing, Vol. 8, No. 1, January 2000.

[28] Eduardo Lleida Et.al. Utterance Verification In Decoding And Training Procedures, IEEE Transactions On Speech And Audio Processing, Vol. 8, No. 2, March 2000.

[29] Lawrence K. Saul and Mazin G. Rahim, Maximum Likelihood and Minimum Classification Error Factor Analysis for Automatic Speech Recognition, IEEE Transactions On Speech And Audio Processing, Vol. 8, No. 2, March 2000.

[30] Jun Huang and Yunxin Zhao,A DCT-Based Fast SignalSubspace Technique for Robust Speech Recognition,IEEE Transactions On Speech And Audio Processing,Vol. 8, No. 6, November 2000.

[31] Hui Jiang et.al., A Minimax Search Algorithm for Robust Continuous Speech Recognition,IEEE Transactions On Speech And Audio Processing, Vol. 8, No. 6, November 2000

[32] Wolfgang Reichl and Wu Chou, Robust Decision Tree State Tying for Continuous Speech Recognition, IEEE Transactions On Speech And Audio Processing, Vol. 8, No. 5, September 2000.

[33] George Saon and Mukund Padmanabhan, Data-Driven Approach to DesigningCompound Words for Continuous Speech Recognition, IEEE Transactions On Speech And Audio Processing, Vol. 9, No. 4, May 2001.

[34] Aravind Ganapathiraju et.al.,Syllable-Based Large Vocabulary Continuous Speech Recognition, IEEE Transactions On Speech And Audio Processing, Vol. 9, No. 4,May 2001.

[35] Alain Biem et.al, An Application of Discriminative Feature Extraction to Filter-Bank-Based Speech Recognition, IEEE Transactions On Speech And Audio Processing, Vol. 9, No. 2, February 2001

[36] Montri Karnjanadecha and Stephen A. Zahorian,Signal Modeling for High-Performance Robust Isolated Word Recognition, IEEE Transactions On Speech And Audio Processing, Vol. 9, No. 6, September 2001.

[37] Jen-Tzung Chien,Linear Regression Based Bayesian Predictive Classification for Speech Recognition, IEEE Transactions On Speech And Audio Processing, Vol. 11, No. 1, January 2003.

[38] Mohamed Afify and Olivier Siohan,Sequential Estimation With Optimal Forgetting for Robust Speech Recognition , IEEE Transactions On Speech And Audio Processing, Vol. 12, No. 1, January 2004.

[39] Mohamed Afify, Feng Liu, Hui Jiang, A New Verification-Based Fast-Match for Large Vocabulary Continuous Speech Recognition, IEEE Transactions On Speech And Audio Processing, Vol. 13, No. 4, July 2005.

[40] Rita Singh, Bhiksha Raj et. al.,Automatic Generation of Subword Units for Speech Recognition Systems, IEEE Transactions On Speech And Audio Processing, Vol. 10, No. 2, February 2002.

[41] Sadaoki Furui, Tomohisa Ichiba et.al., Cluster-based Modeling for Ubiquitous Speech Recognition, Department of Computer Science Tokyo Institute of Technology,Interspeech 2005.

[42] Michael K.Brown et.al, Training set Design for Connected Speech Recognition, IEEE Transaction on Signal Processing, Vol.39.No.6.June 1991.

[43] Chia-Yu et.al.,Histogram based quantization for Roboust and/or Distributed speech recognition, IEEE Transactions On Audio, Speech And Language Processing, Vol.16,No.1 Jan 2008.

[44] Giulua Garau Stere rebaks, Combining spectral representation for large vocabulary continuous speech recognition, IEEE Transactions On Audio, Speech And Language Processing,Vol.16,No.1 Jan 2008.

[45] Kevin M.Indrebo et.al,Minimum mean squared error estimation of mel-frequency cepstral co-efficients using a Novel Distortion model, IEEE Transactions On Audio,Speech And Language Processing,Vol.16,No.1 Jan 2008.

[46]

[47] Xiong Xiao,Normalisaton of the speech modulation spectra for robust speech recognition, IEEE transactions on Audio, Speech and Language Processing, Vol.16,No.1 Jan 2008.

[48] Bertrand Mesot, Switching Linear Dynamical Systems for Noise Robust speech recognition, IEEE transactions on Audio, speech and Language processing, Vol.15,No.6, May 2007.

[49] Soundararajan et.al.,Transforming Binary un- Certainties for Robust Speech Recognition, IEEE Transactions on Audio, Speech and Language processing, Vol.15,No.6, July 2007.

[50] Frank Wessel and Hermann Ney , Unsupervised Training of Acoustic Models for Large Vocabulary Continuous Speech Recognition, IEEE Transactions On SpeechAnd Audio Processing, Vol. 13, No. 1, January 2005