JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Speech Recognition for Virtual Assistants: A Review of Techniques, Performance Metrics, and Future Trends

Sandra C
School of Computer Application
Lovely Professional University
Phagwara, Punjab, INDIA

Soumya C
School of Computer Application
Lovely Professional University
Phagwara, Punjab, INDIA

Gajanan Prajapati
School of Computer Application
Lovely Professional University
Phagwara, Punjab, INDIA

Bhawna Sharma
Asst. Professor, SCA
Lovely Professional University
Phagwara, Punjab, INDIA

1.Abstract

This study provides an in-depth analysis of speech recognition methods used in virtual assistants, with an emphasis on how deep learning techniques like Hidden Markov Models (HMMs) are integrated into natural language processing. We investigate the development of speech recognition algorithms, their performance measures, and new trends through a methodical investigation. Using HMMs for sequential modeling and deep learning for feature extraction and semantic understanding are important components. Assessment measures include user happiness, latency, and accuracy, and they shed light on how well different strategies work. We also talk about future directions to further increase the capabilities of virtual assistants in various situations, such as multimodal integration and robustness enhancements.

1.1. Speech Recognition Introduction

Speech recognition has long been a systems favorite because it allows machines to translate speech into speech. This is a technology driven by the need for speed and efficiency. The most important advances in speech recognition occurred i n the second half of the 20th century. The first speech recognition machine, called "Audrey", was introduced at Bell Labo ratories in 1952 and became a leader in the field of speech recognition. Unfortunately, even the most important recogni tion (such as being able to recognize only 10 voice numbers within the recording and accounting limit that works as a ch arity) is not possible due to the relationship and limited activity of the body. related to. However, it wasn't until the 1960 s that researchers first used digital computers to measure speech recognition. This requires recording the audio, convert ing it from analog to digital, and using an inexpensive electronic device to identify simple numbers on the screen. Howev er, expensive and transaction costs associated with hardware and software are still obstacles that hinder and determine progress in speech recognition. In the early 20th century, the possibility of using speech recognition technology increase d significantly due to decreasing hardware costs, decreasing microprocessor speed and capacity, significant advances in signal processing algorithms, and higher performance. Appropriate amplitude and time measurement sensors are more reliable. This is in addition to the continued growth of digital products and research and development efforts. Interest an

d growth in this field has increased as companies such as Dragon Systems and government agencies state that it is almos t impossible to create records. IBM released its first job recognition system in 1982. Then, starting in 2013, the use of sp eech recognition technology increased as businesses continued to develop new models and improve fixes. Speech recog nition, which was previously used for speech, is also used for virtual assistants. But the relationship between humans an d computers is different. Winkler said a virtual assistant is defined as "a software intermediary that allows data to comm unicate with the computer." Industry 4.0 technology and related systems can become “smarter” by creating virtual assis tants that replace the need for users to talk to normal processes. This has the advantage of allowing users to interact mo re with the platform, for example. Many fundamental technologies were created during this period, especially before th e year 2000.

This technology has increased the educational level of ASR and other activities. Compared to early successes, progress in ASR research and application in the decade before 2010 was slow and insufficient. However, in the meantime, importan t techniques such as GMM-HMM discriminant training have also been developed and applied in the real world.

1.2. The Importance of Speech Recognition for Virtual Assistants

One of the key benefits of speech recognition is the ability to communicate with digital devices and robots, including virtual assistants1. This is useful, for example, when driving or when your hands are weak and typing is not possible or possible. Speech recognition is also important in helping blind and deaf people because it allows them to better communicate and use technology. Treatment is another important benefit. Speech recognition can help speed up information, improve medical records, and improve customer service calls by translating text into text.

Productivity: Work hands-free. It facilitates and improves multitasking. : Make translation faster. It eliminates communication problems. User preferences and opinions.

1.3. Overview of Virtual Assistants

The research paper “Speech Recognition for Virtual Assistants: A Review of Current Research, Performance Evaluation, and Future Trends” provides a comprehensive assessment of standards, measures, and needs for advancement in this field. This article covers HMM, hybrid methods, and deep learning, among other speech recognition techniques.

The speech recognition industry uses statistical models called Hidden Markov Models (HMMs) to simulate sequences of sounds in speech. The basic phonetic unit of speech is represented by the hidden state level, and the hidden Markov model predicts the sound level based on the hidden state. HMM models are trained on large amounts of speech data to estimate the probability of sound emission in a situation and the probability of the situation changing. Deep neural networks can learn the representation of words and recognize speech accurately. Deep learning methods for speech recognition include short-term memory (LSTM) networks, random neural networks (RNN), and neural networks (CNN).

Hybrid techniques combine HMM and deep learning to ensure accurate speech recognition. Hybrid methods combine the best features of deep learning and HMM methods. For example, deep learning models can learn complex representations of speech symbols, while HMM models can attempt to process speech. . The most commonly used in speech recognition is called WER, which calculates the number of errors made when recognizing a word. Accuracy measures the percentage of words recognized correctly. Processing speed, an important aspect of real-time data, is the time required to recognize speech. Language, personalization and information. Advances in NLP use machine learning algorithms to understand the context and meaning of speech. Multilingual and accented speech recognition refers to the ability of speech recognition to recognize speech in different languages and accents. Understanding context and identity requires knowledge of the language to be able to recognize context and tailor responses based on the user's preferences and past interactions. It includes deep learning, hybrid and HMM methods. This article discusses the future direction of speech recognition and the importance of performance metrics such as WER, accuracy, and processing speed. Other topics include conceptual understanding, multilingualism and language skills, NLP development and personality.

2.Techniques for Speech Recognition

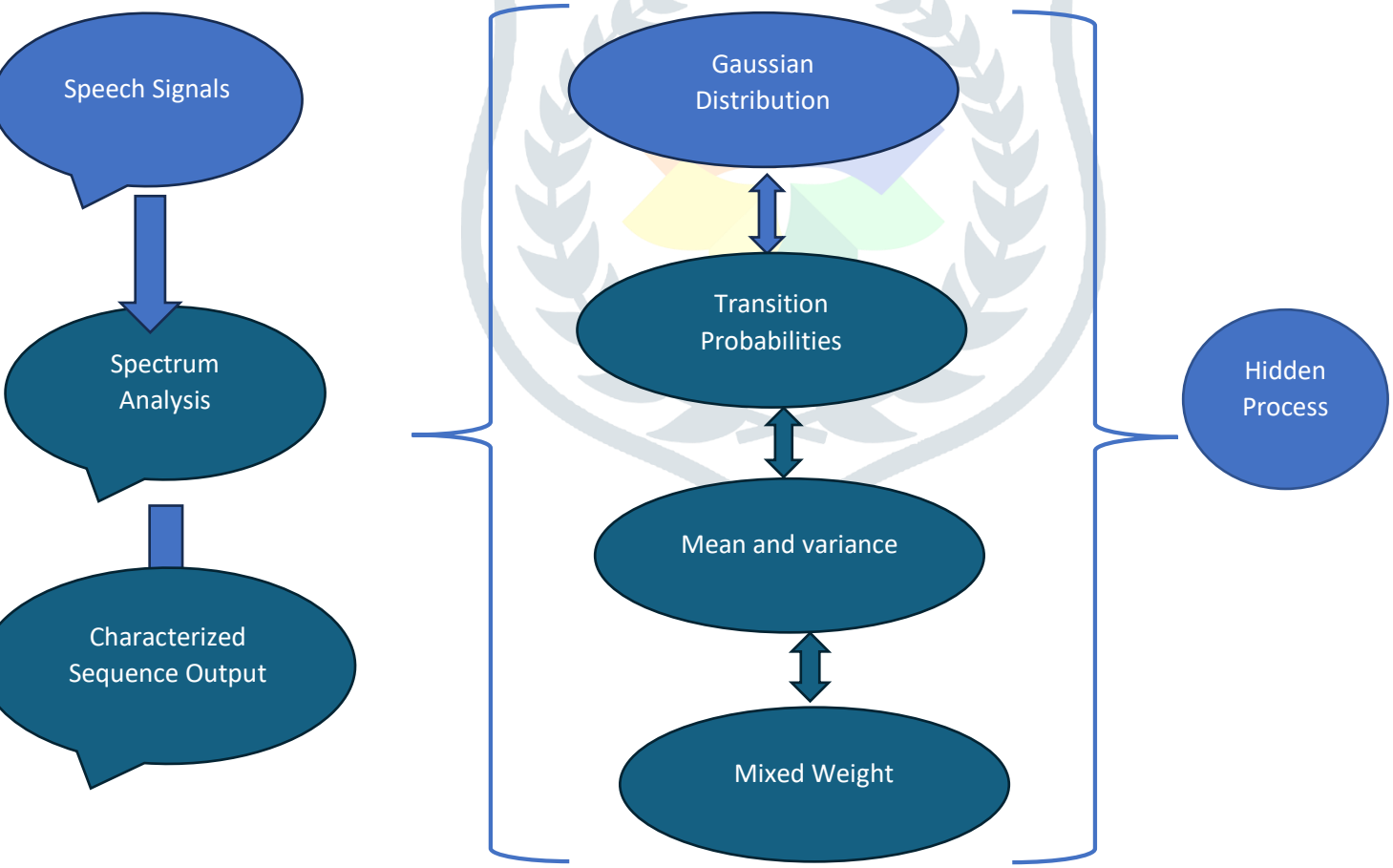
2.1. Hidden Markov Model (HMM)

The Hidden Markov Model (HMM) forms the basis of many successful methods for acoustic modeling in speech recognition. The main element that makes this success possible is the model's ability to analyze speech events and its accuracy in speech recognition applications. Currently, HMMs form the basis of Large Vocabulary Continuous Speech Recognition (LVCSR) systems.

While the main ideas behind HMM-based LVCSR are simple, direct application of these ideas will lead to unexpected change in the workplace and will be disadvantageous due to predictable and simple assumptions. Therefore, the use of HMMs in today's systems requires expertise.

Hidden Markov Models (HMM) provide a suitable framework for developing these models because speech has a physical structure and can be recorded as a set of spectral vectors covering a variety of sounds. Markov Models

Hidden Markov Models (HMMs) are useful tools when it comes to statistical modeling of linked data such as speech, text, or DNA. They can identify dependencies and patterns in data and identify hidden states or conditions underlying observed results. convenience and efficiency. They may contain additional information such as context, grammar or pronunciation, and phonemes can be exchanged for various information such as words.



This figure shows the block diagram of the Hidden Markov Model (HMM) for speech recognition. HMM is a m odel used to **evaluate** observations resulting from **fundamental** processes. In speech recognition, speech recogn ition is the process **of observing** the properties of speech **symbols.** Speech is always **expressed with** symbols. F

ourier transform is the mathematical method used to do this. It is hidden because we cannot see it immediately f rom the flow of the conversation. For example, going from "compile" to "install" **does** not go from "compile" to "compile". In contrast to the Gaussian mixture distribution for **"basic" data,** the Gaussian mixture distribution for **"extended" data** can be **found** in many groups of acoustic variables. They represent the foundation and **con tinuation** of the pipeline. They represent the foundation and **continuation** of the pipeline. Speech is represented by a **group** of symbols. HMM then uses the acoustic properties of the **frame** to determine the probability of ea ch state at each **frame interval.** The Viterbi algorithm is then used to identify **multiple** hidden processes **resulti ng** in multiple observations. Finally, the **situation is often associated with symbols to be sorted** in the **simple process.**

2.2. Deep Learning Approaches

Deep learning is a new development in machine learning research. Deep learning uses a complex transformation of model abstractions for big data. Recent research shows that deep learning supports cognitive development in many ways.

Deep neural network (DNN) is a neural network with three or more layers. In fact, most DNNs have more than one layer. DNNs are trained on large amounts of data to identify and classify events, find patterns and connections, evaluate options, make predictions and make decisions. Deep neural networks consist of multiple layers that work together to refine and improve the predictions and decisions made by a single neural network layer. Human participation in scanning. It enables cutting-edge discoveries such as artificial intelligence and driverless cars, as well as modern products and services such as voice-activated TV, digital assistants, and credit card fraud.

Artificial neural networks, sometimes called deep learning neural networks, mimic the working of the human brain by combining input data, weights, and biases. When combined, these elements enable accurate product identification, classification, and description in the data.

To develop and improve classification or prediction, deep neural networks consist of many layers of connections where each layer builds on top of other layers. Transmission time is used to account for changes in the network. The input and output processes of deep neural networks are visual processes. After processing the data in the input process, the deep learning model creates a final prediction or classification in the output process.

The model is trained in the next process called backpropagation; This process first calculates the forecast error using techniques such as gradient descent. Then work backwards through each layer to adjust the weight and deflection of the work. When forward propagation and backpropagation are used together, neural networks can perform error prediction and correction. At the same time, the accuracy of the algorithm also increased.

Deep Learning Methodologies

The tremendous progress in deep learning in recent years has enabled computers to gradually acquire speech recognition ability. Speech recognition is important in blind and disabled assistance applications, as well as in customer service and education. Speech recognition combines various computer science techniques to recognize speech patterns. By recognizing speech patterns, computers can distinguish the different commands they are trained to perform.

The network structure of the input signal makes RNN particularly suitable for speech processing. They can try different time patterns that are difficult for their neural architecture to understand. At first, RNNs were combined with speech processing, which could be solved using Hidden Markov Models (HMMs), but this approach was flawed because HMMs required knowledge of specific operations and acknowledged the limitations of the freedom state. The fact that it creates power and discrimination makes it widely used in communication. It can be divided into one-way networks and two-way networks;

Speech recognition also uses Transformers, the deepest learning found in many NLP applications. They are particularly suitable for language learning because they can capture long-term relationships in spoken language. This process involves deriving hierarchical representations from raw speech that can improve speech processing. In many ways. However, it can be difficult to implement and may require special skills and equipment. Word Error Rate (WER) is the percentage of words recognized by the ASR model and is often used to measure the effectiveness of ASR solutions. While a WER value of 5% is

possible for scientific research on some materials, a WER value of 10 to 20% is considered suitable for practical use. Learning (DL) solutions as part of the concept Deep learning is currently booming due to the significant advancement in neural networks in recent years. Many complex artificial intelligence applications, from chatbots to customer service to image and product recognition in stores, are made with the help of deep learning (DL). The evolution of higher education standards in recent years has made deep learning (DL) attractive to many organizations. But that doesn't mean deep learning can solve all machine learning problems. The ability to drill down to manage complex problems requires discovering underlying patterns in data and understanding every interaction among various interactions. Deep learning algorithms can identify hidden patterns in data, combine them, and create better decision rules. It is a very good tool for obscure words.

2.3. Hybrid Approaches

Hidden Markov Models (HMMs) Using Discrete Learning Hidden Markov models have long been used to model physiological interactions in speech recognition. However, their performance can be improved by combining deep learning techniques such as Recurrent Neural Networks (RNN) or Convolutional Neural Networks (CNN) to learn the image body from raw data.

Temporal modeling: RNNs are particularly suitable for temporal data processing because they can capture temporal dynamics and preserve internal states. In speech recognition, RNNs can faithfully reproduce components of speech signals, including communication, language, and telephone associations.

Long-term dependencies: RNNs can detect long-term dependencies in speech, which are important for understanding context and generating accurate predictions. This is especially true for transformations such as gated repetitive units (GRU) and short time series (LSTM). Writing. >

Feature extraction: CNNs are good at extracting features from spectrograms or other time-frequency representations of speech because they are good at identifying spatial patterns in objects. They can record local features such as patterns, speech changes, and spectral features. Too high. CNN is suitable for real-time speech recognition applications due to its parallelism. and deformation with naturally elastic materials. CNNs act as specialized front-ends in such topologies, and network modeling and decision-making are handled by recurrent networks or networks. complementary quality. For example, a popular method is to use CNN to extract features at the end of the original audio spectrogram, then feed the quality features back into RNN for modeling and decision-making purposes. Additionally, the listening process is often included so that the model can focus on relevant parts of the input sequence when making decisions. Using deep neural networks instead of Gaussian mixture models in the acoustic modeling phase of the speech recognition pipeline. It has been proven that improving the structure of the relationship between documents can increase the recognition of truth. It is important for speaking skills. The hybrid method combines a computational language model with a neural network-based language model to benefit from both approaches. Neural language models are best at capturing spatial and semantic information, but computational language models are better at handling unfamiliar or unfamiliar parts of a word. Various feature extraction algorithms are used to extract audio signals. For example, spectrogram, Mel Frequency Cepstrum Coefficients (MFCC), and Perceptual Linear Prediction (PLP) features can be used to capture temporal and spectral features of the expression.

Group process: group process is combined from several basic models to improve overall performance. Speech recognition can be used at various stages of the production line, including acoustic modelling, language modeling and decision-making, to integrate multiple sources of information and reduce the experience of errors. Algorithms can leverage other types of data, including gestures, faces, and lips. The hybrid system integrates multimodal data using techniques such as early fusion (combining methods in the concept phase) and late fusion (combining methods in the decision phase) to improve truth knowing, especially in noisy or difficult environments.

Transform learning: Transfer learning technology replaces the pre-training model for activities related to or registered for the specific purpose of language learning. Transfer learning can reduce the need for multiple oral articles and improve performance, especially in limited situations. Transfer of learning leverages information from related activities or big data.

Bringing together the best of various methods, these combinations help overcome speech recognition problems and create systems that improve accuracy and accuracy in many places and applications.

3. Performance Metrics for Speech Recognition

3.1.Word Error Rate (WER)

Word Error Rate (WER) is a popular metric used to measure the effectiveness of speech recognition or machine translation. The fact that the length of a recognized string of words will differ from the use of a (said to be correct) string of words often makes performance difficult to measure. . A lower WER for speech-to-text means more accurate speech recognition. For example, a WER of 20% means the list is 80% accurate.

$$WER = \frac{S + D + I}{N} = \frac{S + D + I}{S + D + C}$$

Where

S is the change number, (change word)

¡ time, (dropped words)

I is the number of additions, (increased words)

C is the number of correct words,

N is the number of references (N = S +D + C)

Dynamic programming, Used to match the recommendation with the reference sequence to calculate WER. Substitutions, deletions, and insertions must be recorded and summed to calculate WER. To overcome these issues, it is necessary to write the best training materials that use classroom jargon, accents, language, and job-specific terminology. Training the ASR model using this data can reduce WER. If the speaker is not registering correctly, more errors will occur in the decision-making process. Therefore, proper records must be kept. Below are some errors the decoder receives to identify the type of error. REF (reference) represents the text used by the sphinx speech recognition system and HYP (hypothesis) represents the hypothesis from the sphinx speech recognition system. This change reduces speech recognition performance. REF: CHITTOORKI velle train peyremit HYP: CHITTOORKU velle train peyremit SENTENCE Correct = 75.0% 3 False = 25.0% 1 Show the word CHITTOORKU instead of CHITTOORKI. This type of variation is due to confusion between two words as the distance between their numbers is very small. Due to this uncertainty, 75% accuracy was achieved. Reference: YASHWANTHPUR EXPRESS KAACHIGOODAKU EPPUDU VASTHUNDHI HYP: AVUTHUNDHI VAITING VELLE VELLE VAITING E VUNTUNDHI VELLE VUNDHI VAITING Sentence True = 0.0% 0 False = 200.YEU3 1000.0%. DU ==> VELLE VUNDHI From the above, it can be seen that more than one word is recognized in the position of a word. In this case, the error rate increases significantly. Many attachments will appear in this. In this case, the tongue will be erased and therefore the activity of the body will decrease. REF: RIJERVESHAN ELA CHEYAALI HYP: SABARI MAYL SENTENCE Correction = 0.0% 0 Error = 100.0% 3 ELA CHEYAALI ==> MAYL

Type 4: This type of error is due to the formation of new words or level given in the result. The number of new words or words given is not enough. These errors usually occur in extended sentences containing very large keywords. The decoder sometimes cannot match the predicted message.

3.2.Accuracy

Word Error Rate (WER) represents the number of errors in text relative to standard resolution and is often used to measure the accuracy of ASR. Reducing the word error rate (WER) of data using new processing technologies such as end-to-end (E2E) models or self-monitoring is one of the main goals of ASR research and structure. The errors of some tests are so small that they even exceed the errors obtained from the dictionary. But even if the claim is high, this may not be a good measure of the overall impact of ASR. Overfitting is a risk associated with training and optimization of specific data, and the results may not be as reliable as the performance of new data. Various sources highlight the inconsistency of ASR in the literature: ASR presents racial discrimination or different facts to different speakers and serves different roles for men and women speaking. These tests do not address engine accuracy; Rather, they attempt to show that ASR contains biases that often influence speakers from anonymous individuals. Basic information is not available.

3.3. Processing Speed

The time required for a system to translate spoken words into written text is called speech recognition speed. It is an important part of speech recognition as it directly affects the user experience. While slow motion can be slow and frustrating, speeding up can make the camera more efficient and effective.

Real Time (RTF) is the ratio of time required to recognize speech to speech time and is often used to measure the speed of the speech recognition process. When RTF 1.0, the system can process speech immediately;

For example, Google Cloud's Speech-to-Text API often analyzes audio faster than time; this is an average of 0.5 RTF for 30 seconds of audio. This means users experience because it only takes 15 seconds for the system to produce 30 seconds of sound.

However, it is important to remember that many variables can affect operating speed, including background noise, volume, loudness, noise and sound quality.

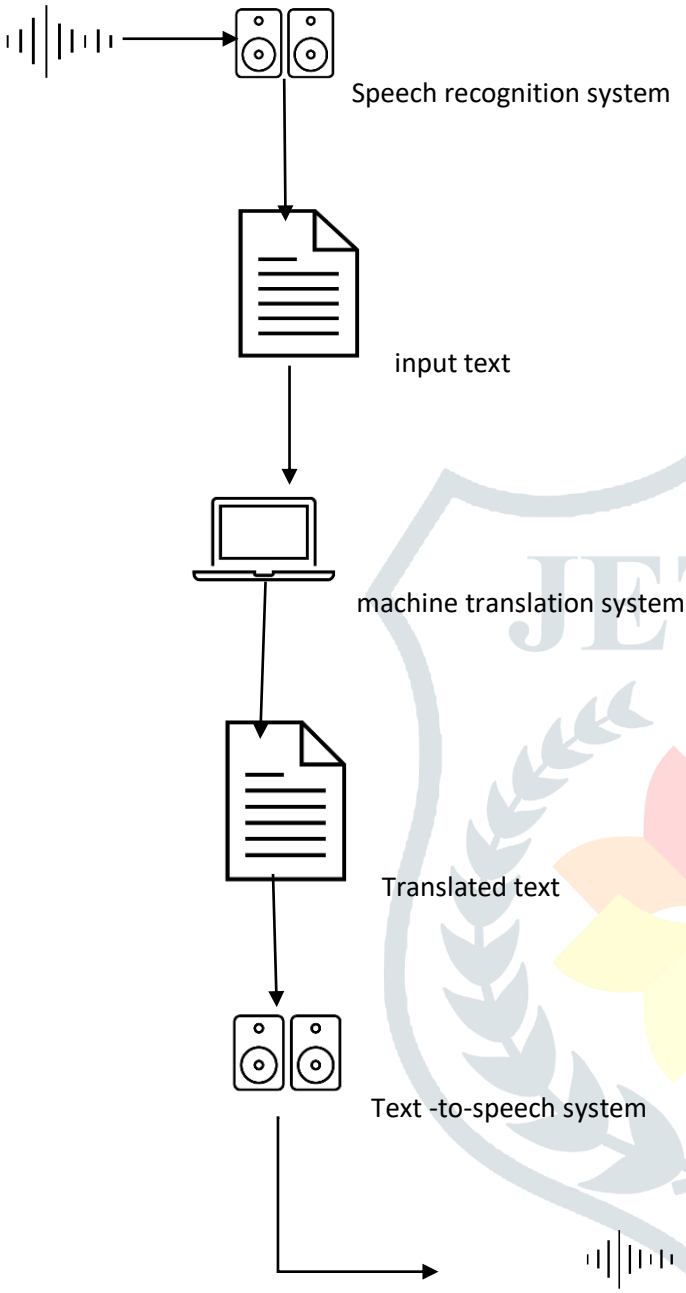
These changes may affect the accuracy of speech recognition and therefore processing speed.

In summary, speech rate information is an important factor that directly affects user experience. While slow motion can be slow and frustrating, speeding up can make the camera more efficient and effective. Current factors are often used to measure performance and can be affected by many factors such as background noise, volume, noise, noise, and sound quality.

4. Future trends in Speech Recognition for Virtual Assistance

4.1. Advances in Natural Language Processing (NLP)

One of the most important advances in machine learning today is Natural Language Processing (NLP). Natural language processing (NLP) is the study of using technology to process natural or human text and speech. Let's better understand what we mean by natural language and natural language processing. language processing. Being able to read, interpret, understand and understand human language is the ultimate goal of natural language processing (NLP). Data collection or data collection is more common in NLP activities; However, raw data is often useless in NLP applications. Engineers must first convert the original data into a machine-readable format.



Achievements of NLP:

Transformer-based model:

The creation of transformer-based model is one of the most important advances in NLP. Transformer was able to overcome the shortcomings of traditional neural networks (RNN) and convolutional neural networks (CNN) by adding methods for self-monitoring. This allows the model to complete the text step by step, providing a more accurate and effective interpretation. One of the largest AI models to date is the 175 billion language model that OpenAI released in 2020. Chatbots, content creation and creative writing. The left and right sides of the word reveal meaning and context. This bidirectional machine replaces NLP tasks such as text classification, query answering, and sentiment analysis by producing state-of-the-art results. In contrast, recent advances allow NLP to process and understand a wide range of data, including speech, images, and text. Many applications such as speech-to-text, question-answering, and graphics demonstrate the potential of multilingual processing. help. Optimizing pre-trained models like BERT and GPT-3 for specific tasks requires less data and more training time than training them from scratch. Thanks to this revolutionary learning

process, NLP is widely available and easy to use, even for developers without a background. This model, also from Google AI, outperformed BERT in January 2019.

PyTorch-Transformers

The team at Hugging Face achieved this success by creating PyTorch Transformers in July 2019. With this tool, we can implement BERT, XLNET and TransformerXL models with just a few lines of Python code. Natural Language Processing or NLP. Natural language processing (NLP) has made significant progress over the years with the development of Transformer-based models such as BERT and GPT-3, as well as deep learning. In addition to changing the way people interact with technology, these developments have created new opportunities in many industries, including healthcare, banking, education and customer service.

4.2. Multilingual and Accented Speech Recognition

Using traditional speech recognition tools to recognize speech in a single language. However, due to increasing globalization and speaker diversity, the demand for multilingualism continues to increase. Multilingual speech recognition systems can understand and record speech in multiple languages. These systems often use speech recognition to recognize speech and adjust speech patterns accordingly. Native language learning software can record speech effectively. The goal of speech recognition is to create a system that can recognize and record speech in different languages. This requires collecting a lot of training data, such as speakers using different words and different phonemes, and then modifying the language model to better include different words. Language: Personal Assistant. To do this, you can use language recognition models that can identify words from spoken words. Once the language is detected, the assistant can switch to relevant language patterns for further processing. This requires the creation and use of standard languages for all supported languages. These models must be able to understand and respond to user queries in multiple languages. Speech This requires training in speech recognition using data from speakers with different accents. Adaptive strategies can also be used to improve the ability of cognitive models to cope with speech patterns and voice differences. code conversion). To solve this problem, the personal assistant must be able to control language changes and easily change language patterns. It will get better over time. This includes adapting to new sounds heard in nature, developing language-based questions, and learning from user feedback. Standard speak for privacy concerns. Individual service providers must adhere to strict privacy policies and ensure that user information is kept secure and transparent.

Performance Optimization: Effective usage and optimization must occur instantly and ensure low latency, especially in the context of cloud-based self-services. This requires the use of effective language structures, ease of speech recognition, and the ubiquitous use of high-speed devices. Examples include providing feedback in the user's preferred language, creating user interfaces that facilitate multilingual communication, and providing specific language and instructions. Provide personalized service and access to global users. However, in order to provide reliable and effective service to customers with different languages and languages, language, cultural differences and difficulties must be carefully evaluated.

4.3. Understanding context and identity

Speech Analysis: Speech recognition analyzes the speech as well as examines the content of the speech. This takes into account previous transactions, the user's location, time of day and other relevant information.

Semantic analysis: Using natural language processing (NLP) technology, the system can interpret spoken words, including thoughts, emotions, and specific commands. Content that will enable user interaction.

Personalization:

User Profiles: All users' favorite languages, accents, voice patterns, and frequently used commands can be stored in user profiles where speech-related information can be generated for them.

Adaptive learning: Over time, these machines improve their accuracy and understanding by constantly learning and changing in response to human interactions. Thanks to this revolutionary learning process, the system can now identify each user's unique voice and preferences.

Response: Thanks to personalization, the system can give a response based on the user's personal information. For example, it may teach them specific content or modify the language based on their experiences and interests. Correlation is achieved using machine learning techniques such as deep neural networks.

Updates can be used to adapt them to changes in user behavior and language usage patterns. This allows the system to learn from mistakes and adapt itself accordingly.

Correcting errors: By correcting misconceptions, users can provide important information, thereby increasing accuracy and reducing the risk of making mistakes in the future. More reliable information.

Content Fusion: The system can improve overall performance and content awareness by integrating data from multiple sources.

Conclusion

As a result, this paper provides an extensive analysis of speech recognition techniques in virtual assistants, with an emphasis on the incorporation of deep learning approaches such as Hidden Markov Models (HMMs) into natural language processing pipelines. This study sheds light on the effectiveness of different approaches by exploring the development of voice recognition algorithms and their performance indicators, such as accuracy, latency, and user satisfaction. It emphasizes how important HMMs are to enhancing the capabilities of virtual assistants by highlighting their function in sequential modeling and deep learning for feature extraction and semantic comprehension. Additionally, the future directions debate emphasizes how multimodal integration and robustness improvements can be used to further increase the usefulness of virtual assistants in a variety of scenarios.

References:

[1].https://mi.eng.cam.ac.uk/~mjfg/mjfg_NOW.pdf

[2].https://www.sciencedirect.com/science/article/pii/S0895717710001597#:~:text=Hidden%20Markov%20model%20(HMM)%20is,in%20practical%20speech%20recognition%20systems.

[3].https://web.archive.org/web/20210228141444id_/https://irojournals.com/iroiip/V2/I4/05.pdf

[4]. https://arxiv.org/pdf/2305.00359.pdf

[5]. https://www.sciencedirect.com/science/article/abs/pii/S1566253523001859

[6].https://www.irjmets.com/uploadedfiles/paper/volume3/issue_5_may_2021/11395/1628083463.pdf

[7].https://ieeexplore.ieee.org/document/9770703

[8].https://www.atlantis-press.com/article/125977784.pdf

[9].https://www.aiperspectives.com/speech-recognition/#101_The_speech_signal

[10].https://www.aiperspectives.com/natural-language-processing/

[11].https://medium.com/@soukaina./advancements-in-natural-language-processing-nlp-and-future-expectations-33bec2a42d14#:~:text=NLP%20advancements%20have%20significantly%20improved,translating%20text%20between%20multiple%20languages

[12].https://www.clari.com/blog/word-error-rate/

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

[14].https://www.softobotics.com/blogs/advancements-in-speech-recognition-unleashing-the-power-of-nlp/

[15].https://www.ibm.com/topics/deep-learning

[16].https://inclusioncloud.com/insights/blog/whats-new-nlp-latest-ai-advancements/#:~:text=NLP%20is%20enhancing%20BI%20tools,machine%20learning%20to%20automate%20insights

[17].https://www.rev.com/blog/resources/what-is-wer-what-does-word-error-rate-mean

[18].https://www.rev.ai/jobs/speech-to-text

[19].https://www.sciencedirect.com/science/article/abs/pii/S0169716116300463

[20].https://www.educative.io/blog/what-is-natural-language-processing

[21].https://dl.acm.org/doi/10.1145/3636513#:~:text=The%20accuracy%20of%20ASR%20is,compared%20to%20a%20sample%20solution

[22].https://www.kardome.com/blog-posts/difference-speech-and-voice-recognition#what-is-speech-recognition

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

[24].https://transkriptor.com/speech-recognition/

[25].https://cloud.google.com/speech-to-text/docs/speech-to-text-requests

[26].https://www.ibm.com/topics/speech-recognition

[27].https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2017.01308/full