Dean&Francis

ISSN 2959-6157

The exploration of human-computer interaction technology based on intelligent voice technology

Pengbo Kou [1, *]

[1] Department of Artificial Intelligence, School of Information Engineering,
Chang'an University, Xi'an, China

*Corresponding author: pengbo_kou@chd.edu.cn

Abstract:

With the gradual development of artificial intelligence technology, intelligent voice technology has been integrated into people's lives. The human-computer interaction established by intelligent voice technology is also more wide spread, intelligent voice technology is everywhere, from the military field to the news broadcast game field, this paper describes the current implementation of intelligent voice technology and its applications in military, radio, television, games and other applications, and analyzes the development and application of intelligent voice technology and its future trends.

Keywords: Intelligent voice technology; human-computer interaction; future intelligence.

1. Introduction

As artificial intelligence technology matures, it is increasingly integrated into various fields; intelligent voice technology is a notable example. Its main core is the human voice, which, through a series of transformations, processing into a machine can understand the information to achieve the effect of human-computer interaction. Smart voice technology has been around since the 1950s, but it has been slow to catch on in the real world because of the limitations of the technology and the lack of hardware, there are many applications of intelligent speech, speech recognition, speech synthesis, speech assistant, intelligent home control are good examples, people can make these convenient English to work more efficiently, in recent years, deep neural network speech recognition technology, especially convolutional neural network (CNN) and recurrent neural network (RNN) applications, has given new life to intelligent speech technology, allowing robots to better understand people's complex languages, embedded intelligent voice technology has brought convenience to people, people will gradually get used to voice and machine dialogue. However, intelligent speech technology still faces many challenges, such as dialects, accents, noise interference, cross-national languages, and so on, that will affect the accuracy of speech recognition. It also illustrates the challenges that speech recognition will face.

2. Current implementation of intelligent speech recognition systems:

In today's era, speech recognition has been gradually reflected in all aspects, including smart furniture, Apple's Siri, virtual assistants and chat robots, virtual voice assistants for medical Lola, etc. Most have already been applied to people's lives. Voice is a

common medium, from people used to communicate by voice, to now people use voice to control and communicate with machines. Nowadays, most speech recognition systems are based on the deep neural network speech recognition framework, which basically consists of three parts: acoustic model, language model, and pronunciation model. The process of speech recognition is based on the assumption of conditional independence of acoustic and linguistic models. These modules are concatenated to convert speech signal sequences into text sequences gradually [1]. Through the system, people can use voice to send out instructions and, at the same time give machine signals, and then get feedback from the machine. Of course, a complete speech recognition system is a very complex structure, in short, a speech recognition system includes preprocessing, feature extraction, acoustic model, language model, and search algorithm modules, Figure 1 below shows the structure of the speech recognition system.

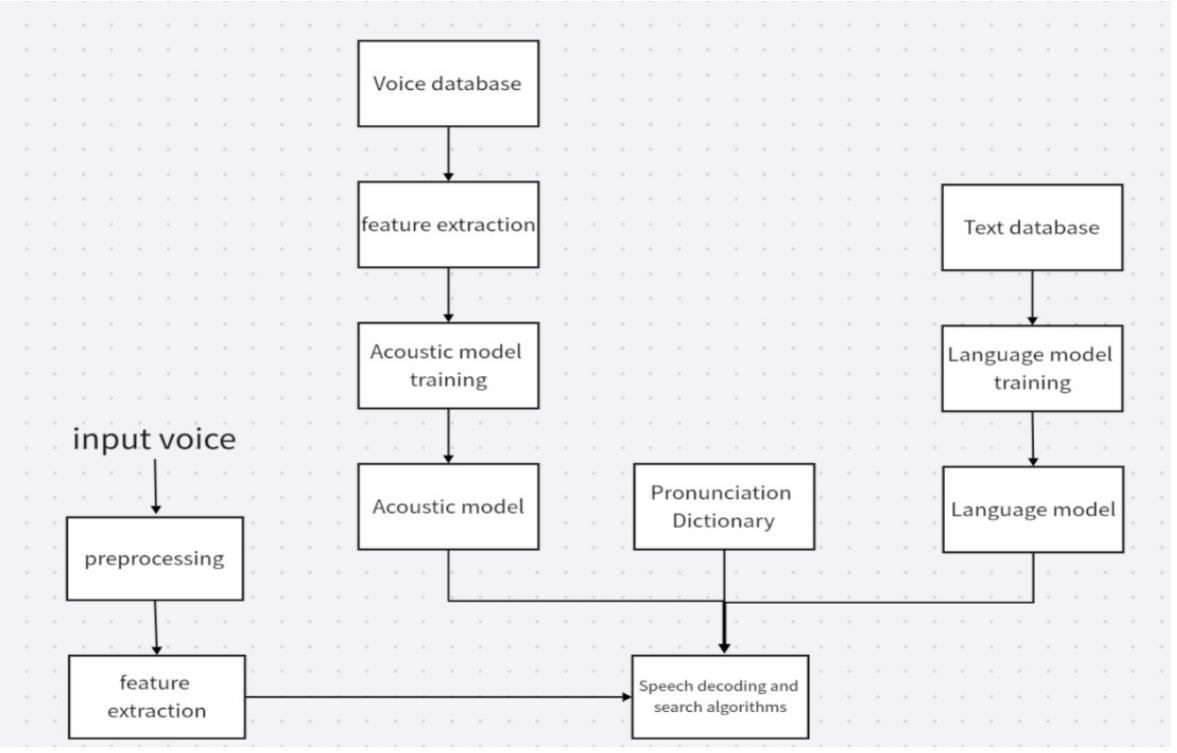

Fig. 1 Structure of speech recognition system

In general, the pre-processing operations include sampling, pre-filtering, a/D conversion, pre-weighting, frame-by-frame windowing, and endpoint detection. The digital speech signal is divided into short-time signals by signal frame. This is mainly because the speech signal is a non-stationary signal, with time-varying characteristics that are not easy to analyze. However, they generally remain largely unchanged for a short period (typically between 10 and 30 milliseconds) and remain stable for a short period, which allows them to analyze their characteristic parameters. After training the language model and the acoustic model, the module that searches the best path (that is, the most possible output word sequence) is called the search module, and then into the weighted finite state converter (also known as WFST) dynamic network (used to combine HMM state, dictionary, syntax, etc.) and finally through the Search engine algorithms such as Beam Search results, the text output, give feedback [2]. From early vocoder to rule-driven from the 1950s to 1970s, to data-driven, and finally to deep learning-driven, speech recognition systems have undergone a long period of development and evolution. Most of the speech recognition systems are based on this model.

3. The application of AI voice technology

Based on the realization of the basic speech recognition system, people gradually apply AI speech technology to various fields, including the military field, news editing field, radio and television supervision field, and vehicle-engine interaction field. In the military field, the airborne speech recognition control system is the most famous example. Table 1 shows that the US, France, and other Western countries have been using voice recognition technology to identify F-16s, rafters, typhoons, and other

ISSN 2959-6157

aircraft in the 1980s and 1990s. The pilot can aim at the target by voice-controlled fire control radar, attack the target by weapon, and resist the target by electronic countermeasure equipment. This allows them to concentrate more intently on maneuvering the aircraft, thus improving their combat effectiveness [3].

Table 1. Applications in the military field_[3]

Country	Platform	Application
America	F-16AFTI Testing Machine (1982)	The pilot controlled the weapon by voice
France	Rafale validator (1986)	The pilots control the weapons and electronics by voice
West Germany, Italy, Spain, England	Typhoon fighter (EFA)(1992)	The pilot uses a voice control system
Pakistan	J-10CE (2022)	The pilot can talk to the plane and get information

With the continuous development of artificial intelligence technology, speech recognition technology has also improved a lot. More military equipment has been equipped with intelligent speech recognition technology, and pilots can even adjust the aircraft's flight posture through voice. At the same time, operators can remotely control battlefield robots and UAVs through voice, greatly improving operational efficiency. In military security, people can also identify voiceprints through intelligent voice recognition technology, ensuring successful authentication.

In the field of news gathering and editing, an intelligent voice transliteration system has become an important tool, which can effectively improve the efficiency of journalists. By using an intelligent voice-over-writing system, journalists can instantly convert voice information into text when conducting an interview, allowing them to focus more on the content of the interview than on the recording process. In addition, the system enables the rapid and accurate conversion of audio recordings into written transcripts in the collation of interview records, which greatly reduces the time for journalists to obtain information from a large number of audio recordings. The search function of the intelligent voice-to-speech system enables reporters to quickly find interview clips on specific topics, thus improving the quality and efficiency of news reporting [4]. At the same time, the voice-over-writing system can also be applied to legal evidence, judges can use the technology to record court content, and lawyers can also use the system for favorable evidence accurate and rapid recording. In case of emergency, the voice transliteration system can also be used as a means of communication, which can reduce the misunderstanding and risk caused by unclear voice.

Intelligent voice technology plays a crucial role in radio and television. It enables the automatic transfer of programs and allows regulators to use natural language processing to identify topics and analyze emotions in real-time. This technology also helps in processing user feedback to understand audience needs. In the field of radio and television supervision, intelligent voice technology is applied in various ways. Voice recognition technology automatically transmits television programs and generates instant transcripts, aiding oversight departments in understanding, recording, and analyzing content. Natural language processing helps regulatory authorities identify and analyze program content, facilitating the timely detection of regulatory violations or deviations from the main purpose. These technologies enable real-time monitoring and processing of programs. Furthermore, smart voice technology can automatically process large amounts of user feedback, allowing regulators to stay informed about audience needs [4].

In addition to these, intelligent voice systems are increasingly being used in transportation. Its application in vehicle-engine interaction systems enables drivers to use voice commands to control various functions without pressing physical keys manually [4]. Through speech recognition (ASR), natural language processing (NLP), speech synthesis (TTS), and other core technologies to complete the interaction between people and cars, this interactive mode can greatly reduce the driver's distracted behavior during the driving process, and thus effectively reduce the possibility of driver's manual operation resulting in traffic accidents [5]. The core functions include voice wake-up and control, voice print recognition, etc. People can use this technology for social interaction, map navigation, vehicle settings, etc. It can help the system analyze the voice command from the context to the semantic, better understand the needs of users, and provide users with a more personalized car experience.

4. The application of intelligent voice system in games

The ASR, NLP, and TTS technologies mentioned above are also the main support of Smart Voice technology in the game. Replica Smart NPCs is an application that defines personalized NPCs. It enables people to communicate with NPCs through voice, it is a good embodiment of intelligent voice human-computer interaction in the game, can provide players with novel experience, provide players with personalized NPC, and is the future direction of development of the game. Of course, this method of interaction is similar to ChatGPT, and interaction with virtual digital human NPCs can be achieved by using the virtual engine and CHATGPT as core components, Virtual digital human voice interactive applications can provide high-quality natural language processing capabilities. By combining multiple functional modules and technologies, an efficient, natural voice-interactive application is achieved. It contains the following key elements, as shown in Figure 2, and the steps to implement the process [6].

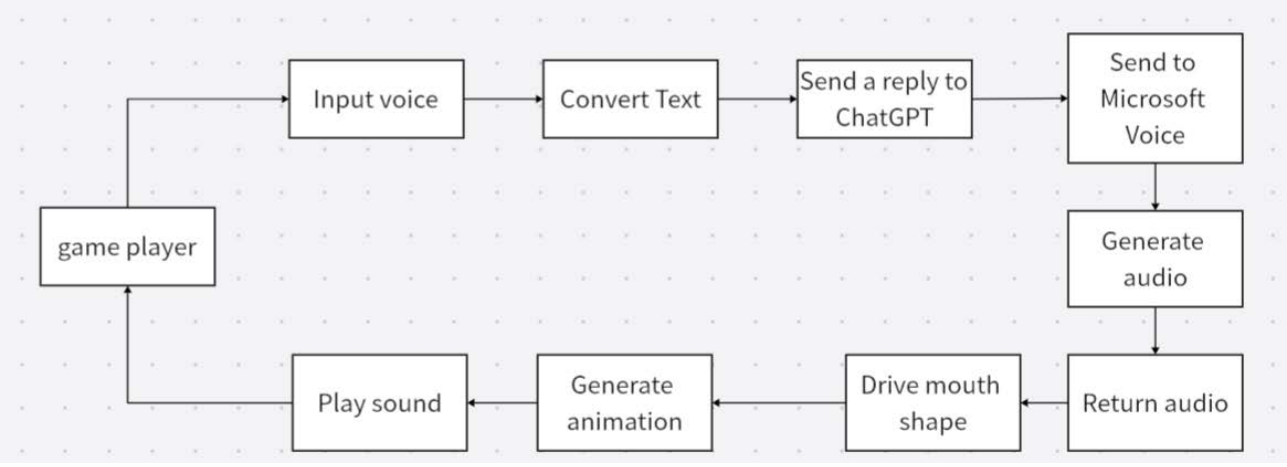

Fig. 2 Basic implementation steps

Based on the diagram in figure 2, the user's audio data is captured by the user's recording device, then converted into a format recognizable by ChatGPT. Next, corresponding feedback text is generated by CHATGPT, which will be converted into actual voice output. The voice output will be synchronized with virtual character animations for the final interaction.

Despite the rapid development of intelligent speech interaction systems, there are still many problems that have not been solved. The first challenge that speech recognition systems face is the problem of recognition in harsh environments. Specifically, in long-distance, noisy, and other complex use of the environment, all kinds of noise, reverberation, and even the insertion of other people's speech may lead to voice signal aliasing and pollution, thus reducing the accuracy of speech recognition [1]. For this reason, games involving this technology have environmental and geographic limitations that tend to reduce the number of players, so the robustness of the speech recognition model needs to be improved. The use of the technology is still very limited today, and games involving the technology also have voice overlap issues. Especially when more than one person is speaking at the same time, the effect of speech recognition will be greatly reduced. The speech-mixing problem, also known as the"Cocktail party" problem, is not difficult for people with hearing choices, because they can focus on one person's conversation without paying attention to background noise or other people's conversations. However, speech recognition systems must recognize and distinguish the content of different speakers in such highly overlapping audio. This is also the future development and progress of speech recognition system trends and direction. At the same time, because the technology is already available to the world, one should also take into account the global nature of the game, that is, the diversity of languages, as cultural exchanges between different countries continue to increase, the multilingual speaking style is more and more used in daily communication and formal meeting, and the Chinese-English speaking style is the most representative. Another problem in current speech recognition technology is language mixing. This is because traditional speech recognition schemes use separate modeling units for different languages [1]. Therefore, how to effectively integrate and distinguish these modeling units, how to deal with the text data and voice data in Chinese-English mixed scenes, and so on. Is something that people have to think about and pay attention to. Another important issue is the identification of domain-specific terms. The accuracy of the specialized vocabulary depends largely on the coverage of

ISSN 2959-6157

the training corpus of the language model, because of the sparsity of the training corpus. In addition, the frequency of professional words is usually significantly lower than that of general-domain words, so it is very difficult to identify professional words into general-purpose words with similar pronunciation [1]. This is a major challenge for the game's intelligent speech recognition, which needs to be overcome slowly in the future.

5. Future trends of intelligent speech systems

With the development of intelligent speech technology, there is still much room for improvement, such as accurate recognition in complex environments and mixed language environments, and a deeper understanding of people's semantics. In People's field of vision, virtual characters can provide accurate and efficient voice interaction experience through intelligent voice technology. Interaction between virtual characters and players has broken down traditional boundaries, allowing cross-domain collaboration and promoting industries as diverse as entertainment, education, film and television, and digital media. Voice interaction technology, based on the CHATGPT language model, will gradually become an important driving force to promote the popularization of virtual digital applications_[6]. In the future, intelligent voice interaction technology will become more natural and smooth, and users can even talk to machines as if they were real people. In the coming days, intelligent voice systems will have to further integrate with Internet of Things technology, and enhance the deepening of natural language processing technology and computer vision. Although speech recognition systems can not function in any population, field, or resource, if researchers can gradually solve these problems and integrate them more closely with real products, this will meet the needs of more industries and solve more practical problems.

6. Conclusion

With the maturity of artificial intelligence technology, AI voice interaction technology has been embodied in various fields and played a very important role at the same time, Smart voice technology in several areas of flexible application, but also analysis of smart voice technology today's problems and challenges, as well as smart voice technology in the future direction of development. In the future life, intelligent voice technology will be more and more common, bringing more convenience and surprise to people's lives.

References

[1] Liu Qingfeng, Gao Jianqing, Wan Genshun. "Research Progress and Challenges in Speech Recognition Technology" [J]. Frontiers of Data and Computing Development, 2019, 1(06): 26-36.

[2] Ma Han, Tang roubing, Zhang Yi, Zhang Qiaoling. A survey of speech recognition research [J] Computer Systems applications, 2022,31(01) : 1-10.

[3] Xue-bao Wang, Yong-tao Tang, qing-bo Wang, Michael Tong. Application Analysis of artificial intelligence speech recognition technology in foreign military field [J] Computer Literacy and technology, 2024,20(05) : 21-23.

[4] XING R F, XU J F.Research on the application of intelligent voice technology in broadcasting and network audio-visual industry[J].Audio Engineering, 2024, 48(6): 56 - 58.

[5] DU C F.Application of intelligent speech recognition technology in vehicle-machine interaction system[J].Video Engineering, 2024, 48(4): 217-219.

[6] YAN T.Virtual digital person voice interaction application based on ChatGPT language model[J].Video Engineering, 2023, 47 (8): 182-186.