The Application of Intelligent Speech Recognition Technology in the Comparison between Japanese and Manchu-Tungus Language

Hongfang Duan[1*], Hexin Wang[2]

duanhongfang1983@163.com[1*], wanghexin202309@163.com[2]

HeiLongJiang University of Technology, Jixi, 158100, China

Abstract: In order to understand the application of comparison between Japanese and Manchu-Tungus language, the application research of an intelligent speech recognition technology in comparison between Japanese and Manchu-Tungus language is put forward. In this paper, the development of intelligent speech recognition technology is briefly introduced, and the concept of intelligent speech interaction is introduced. Secondly, the comparison between Japanese and Manchu-Tungus language is described. It can be seen that the homology of third person pronouns and demonstrative pronouns is universal to some extent. It is not only a feature of Japanese and Manchu-Tungus language, but also a common form in Japanese dialects. Finally, the intelligent speech recognition system and its application analysis are summarized. When the voice information of the detected individual changes greatly with the change of the external environment, the system can identify the voice-related information in the shortest time, and identify the maximum probability of the state transition of the voice information. Then use the relevant sample content of the database to make a deeper analysis and judgment on the voice model of the detected person.

Keywords: speech recognition; Manchu-Tungus language; contrastive analysis

1 Introduction

Speech recognition began with the Audry system in 1950s, which is the beginning of speech recognition research. In 1959, some scholars used digital computers to recognize English vowels and isolated words, and began computer speech recognition. Speech recognition is a technology that computer converts speech signals into corresponding texts through recognition, which belongs to the category of multi-dimensional pattern recognition and intelligent computer interface. The research goal of speech recognition is to make the computer "understand" the spoken language of human beings. With the continuous development of science and technology in China, the development and application of speech recognition system are more and more extensive. In essence, speech recognition system is a type of biometric system, which is a new technology based on traditional biometric technologies such as fingerprint recognition and palmprint recognition. In the application process of this technology, it can recognize the obtained various speech signals and achieve the effect of matching and discrimination. Conceptually, speech recognition is a process of identifying and analyzing one or more speech signals, and then realizing speech matching and discrimination. The computer's "understanding" is not only to convert the spoken language into the corresponding written

ICMEIM 2023, September 08-10, Wuhan, People's Republic of China

DOI 10.4108/eai.8-9-2023.2340188

language word by word, but also to make a correct response to the requirements or inquiries contained in the spoken language, without being entangled in correctly converting all words into written language. Speech recognition technology can be applied to voice communication system, voice-controlled telephone exchange, data query, booking system, hotel medical service, banking service, computer control, industrial control and other fields. It is becoming a key and competitive technology in the fields of robot control and security system[1-2].

In speech recognition, "topology" usually refers to the topological structure of the acoustic model, which is a model structure that represents the relationship between speech signals and text. The topological structure of speech recognition is particularly important in traditional acoustic models based on Hidden Markov Models (HMMs), which describe how speech signals correspond to phonemes, syllables, or higher-level language units. This topology is used to model the transfer patterns between different speech units, thereby matching acoustic features with text. Japanese is a complex language, and the relationship between its pronunciation and text may involve different phonetic variants, high and low intonation, and a rich phoneme system. In Japanese speech recognition, topological structures need to take into account these language characteristics. More states may be needed to model different pronunciation changes and phoneme combinations for better recognition performance. When applying intelligent speech recognition technology to different languages, the design and optimization of topology is a crucial step. The phonetic characteristics and pronunciation patterns of different languages can affect the complexity and structure of topology. A suitable topology can improve recognition accuracy, while incorrect topology design may lead to performance degradation.

2 Difficulties and countermeasures of speech recognition technology

Although the development of speech recognition technology can not meet the practical requirements, it is mainly manifested in the following aspects:

(1) Adaptive problem. The poor adaptability of speech recognition system is reflected in its strong dependence on environmental conditions. There are some adaptive training methods such as cepstrum normalization technology, relative spectral transform (rasta) technology and LINLOG RASTA technology.

(2) Noise problem. When the speech recognition system is used in a noisy environment, the speaker's mood or mind changes, resulting in pronunciation distortion, pronunciation speed and tone change, resulting in Lom bad/Loud effect. The commonly used methods to suppress noise include spectral subtraction, environmental correction technology, modifying the recognizer model to make it suitable for noise without modifying the speech signal, and establishing the noise model.

(3) Selection of speech recognition primitives. Generally speaking, the more words to be recognized, the smaller the primitives used, the better.

(4) Endpoint detection. Endpoint detection of speech signal is the key first step of speech recognition. Research shows that even in quiet environment, more than half of the recognition errors in speech recognition system come from endpoint detectors. The key to improve endpoint detection technology is to find stable speech parameters.

(5) Other problems, such as recognition speed, refusal to recognize, and keyword detection technology, that is, removing the modal auxiliary words of and from continuous speech to obtain the real speech part to be recognized, and not responding correctly to the user's wrong input[3].

2.1 Problems faced by intelligent speech recognition technology

(1) Noise interference in the environment

There are many kinds of speech signals, and in some noisy environments, speech is difficult to be recognized. At present, the accuracy rate of speech recognition published is ninety-seven percent, which can only be achieved in a relatively quiet indoor environment. In fact, such a quiet situation rarely exists, and there is still no effective method to solve the noise interference in the environment.

(2) The rate of nonstandard speech recognition is relatively low.

Intelligent speech recognition technology has made great progress under the impetus of machine learning, but there are still some shortcomings. At present, most speech recognition technologies are aimed at Mandarin users, and it may be difficult to recognize speech mixed with dialects. However, nowadays, many people's Putonghua is not very standard, and it is more or less mixed with some local accents. This makes speech recognition software make mistakes in recognition, which is inconsistent with the expected results.

(3) Handling of fault tolerance rate

Because the publisher of speech sometimes makes mistakes, this will make the software unable to correctly identify its semantics when recognizing, which will eventually affect the accuracy. At this time, you need to manually modify or re-enter the voice. Now, some enterprises are also developing speech recognition software that can understand difficult sentences.

3 Based on the comparison between historical linguistics and linguistic typology

3.1 Analysis of Historical Linguistics

Historical linguistics pays attention to the change of language forms, and from the respective development of Japanese third-person pronouns and demonstrative pronouns, it is not difficult to see that there is no connection between Japanese third-person pronouns and demonstrative pronouns. As Mr. Wang Li mentioned in his History of Japanese Grammar: "In ancient Japanese, demonstrative pronouns and personal pronouns were closely related", and at the same time, he believed that "its" and "the" were probably used as demonstrative pronouns first and then developed into personal pronouns. In addition, some scholars believe that Japanese third-person pronouns come from demonstrative pronouns[4-5].

So, is there any connection between personal pronouns and demonstrative pronouns in Manchu-Tungusic language? Although we have no direct evidence to prove that the personal pronouns in Manchu-Tungusic language originated from demonstrative pronouns, it is not difficult to find that there is an amazing isomorphism (or the same root, such as Ewenki language) between the third personal pronouns and demonstrative pronouns in this language family, as shown in Table 1:

Table 1 Isomorphism between third person pronouns and demonstrative pronouns in Tungusic language

	Singular third person pronoun	Proximate pronoun	Distant pronoun
Manchu language	i,tere	are	tear
Xibe language	Ar, ter	er	tar
Evenki language	Tajja,tart	ery	Tart,tajja

Some scholars have pointed out in the study of Altai languages that all personal pronouns are closely related to their demonstrative pronouns. The additional elements of personal possession originate from the corresponding personal pronouns that are relatively complete and the same as pronouns, or personal pronouns with distorted truncated variants in pronunciation. Therefore, from the diachronic development and morphological changes of anaphora in Japanese and Manchu-Tungusic languages, it is not difficult to see the commonality between them, that is, they are closely related to personal pronouns, and there is a phenomenon that personal pronouns and demonstrative pronouns are isomorphic or homologous[6].

3.2 Analysis of Linguistic Typology

As mentioned above, typology studies the forms and functions of different languages that have been customized or regularized. The purpose of the study is to first understand the differences and similarities between different languages, and then classify the different and similar phenomena into systems, so the emphasis is on the synchronic forms of languages. Then, from the perspective of language typology, does the similarity between Manchu and Chinese anaphora and personal pronouns only exist in Japanese and Manchu-Tonggu languages? In fact, it is not special that personal pronouns and demonstrative pronouns are homologous in Japanese and Manchu-Tungusic languages. It has been proved in many languages that demonstrative pronouns come into being before third-person pronouns, which are directly derived from demonstrative pronouns. In fact, there are many phenomena in other languages where the third person pronoun and demonstrative pronoun are isomorphic or homologous, such as Altaic languages, South Asian languages, Indo-European languages and so on. Some scholars have pointed out that Japanese and Korean third-person pronouns and demonstrative pronouns are isomorphic so far, while others have mentioned that French, Russian, English and other third-person pronouns and demonstrative pronouns are homologous. It can be seen that the homology of third person pronouns and demonstrative pronouns is universal, which is not only a feature of Japanese and Manchu-Tungusic languages, but also a common form in Japanese dialects and Indo-European languages. Table 2 shows the corresponding relationship between the third person pronoun and the far-reaching pronoun in some languages and dialects[7-8].

Table 2 Correspondence between third person pronouns and distant referential pronouns in some languages and dialects

	Third person pronaun	Distant pronoun	Remote pronoun
ancient Chinese	I,tere	ere	tere

Central Shaaroi	Er,ter	er	ter
Lanzhou dialect in Gansu prowinoe	Tart,tajja	Eri,ejje	Tart,tajja
Shanxi dialect	er	ter	the
Wa language	the	this	the

4 Based on intelligent speech recognition system and application analysis

4.1 Extraction of characteristic parameters

The extraction of feature parameters is one of the key technologies in the development of artificial intelligence. Biological speech has specific characteristics, and these speech parameters can be extracted for induction and analysis. However, the characteristics of biological speech will also change with the change of external environment. In order to avoid errors, the following three methods can be used to extract feature parameters.

First, speech spectrum technology is used for extraction. Speech spectrum technology mainly focuses on the basic vocal organs of organisms, such as trachea, vocal tract and nasal cavity, and uses the basic vocal organ of human body to extract relevant parameters, and then classifies the extracted parameters. In the process of voice comparison, these parameters can be used to find the special physiological structure of the speaker, so as to locate the speaker quickly. Second, linear prediction extraction. In speech recognition system, the extracted speech samples belong to the "past" voice, but what needs to be matched is the "current" voice content. Therefore, it is necessary to use mathematical modeling analysis to complete the matching process, and realize the predictive operation process through linear prediction extraction, which is very simple and convenient, and generally can be realized without the participation of many parameters. Third, wavelet feature extraction and analysis. This extraction method is mainly accomplished by wavelet technology. The advantage of wavelet technology is that it can accept the change of resolution. However, in the application process, this technology requires the stable intersection of speech parameters, and at the same time, it is also compatible with time-frequency domain. In the current application process, wavelet technology is quite mature, and the combination with artificial intelligence technology is increasingly close. The application flow of intelligent speech recognition system is shown in Figure 1.

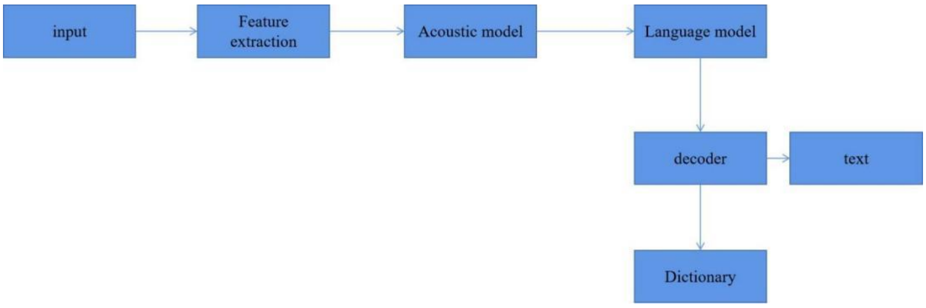

Figure 1 Application flow chart of speech recognition system of artificial intelligence

4.2 Pattern matching recognition process analysis

After feature extraction, further depth analysis is needed and accurate matching process is achieved. Pattern matching recognition is essentially a comparative operation process, which compares the speech feature parameters that have not been recognized with the speech feature parameters in the database, and presents the final comparison results to people according to the similarity, and the distance between the similarities is presented to people in the form of tables or tree diagrams. In this process, the similarity distance has a range, and a value of similarity distance needs to be defined in the process of recognition. At present, there are two models for pattern matching recognition. First, the vectorized model. In the application of this model, it is necessary to vectorize the speech parameters first, and then vectorize the speech features of the detected individuals. In the specific application process, the feature parameter processing of the individual voice characteristics of the detected person is a voice vector that can be expressed as personal information, and then the corresponding voice standard is put forward. Second, build a randomized model. In daily life, although phonetic features are unique physiological features of each individual, a person's voice behaves differently in different environments and different States, and the uncertainty of human voice in the range and probability of change is very strong. Therefore, in the process of research, a stochastic model should be established in combination with the actual needs. In the process of constructing the random model, the phonetic parameters of the detected individuals are classified into the database, and the model of phonetic parameters is established accordingly. And in the construction of this model, we need to take into account the factors such as the probability of transfer, the efficiency and probability of transmission, etc. In order to make the randomized model more reliable and accurate in the application process, it is necessary to obtain the matrix of state transition probability and the matrix of symbol output probability in the training process. When the phonetic information of the detected individual changes greatly with the change of the external environment, the system can identify the phonetic information in the shortest time, and identify the maximum probability of the state transition of the phonetic information, and then use the relevant sample content of the database to make a deeper analysis and judgment on the phonetic model of the detected person[9-10].

5 Conclusion

In the modern society with the rapid development of intelligent technology, the combination of artificial intelligence speech recognition system can recognize and process speech signals efficiently, and extract speech feature data with obvious characteristics through the calculation process of computer, so as to realize the positioning effect. In the process of research and application in recent years, the recognition efficiency of intelligent voice system has been further improved, and its fault tolerance rate has been rising, which has a very broad application prospect. Moreover, from the technical characteristics, the development difficulty of intelligent speech recognition technology is obviously lower than that of traditional speech recognition system, which can further promote the integration and development of speech system and intelligent system and expand the function and compatibility of speech recognition system in the future development process.

Acknowledgments. Heilongjiang Provincial Undergraduate Basic Research Funds Project "A Comparative Study between Manchu-Tungus languages and Japanese Languages from the Perspective of Historical Comparative Linguistics"

References

[1] Zhu, W. , Jin, H. , Chen, J. , Luo, L. , Wang, J. , & Lu, Q. , et al. (2022). A hybrid acoustic model based on pdp coding for resolving articulation differences in low-resource speech recognition. Applied Acoustics, 1(9)2, 108601-.

[2] Austin, S. . (2022). How ai-driven speech recognition can address the challenges of today's radiology patient care. Database and network journal74(2), 52.

[3] Soky, K. , Mimura, M. , Kawahara, T. , Chu, C. , Li, S. , & Ding, C. , et al. (2021). Trieccc: trilingual corpus of the extraordinary chambers in the courts of cambodia for speech recognition and translation studies. International Journal of Asian Language Processing, 31(03n04).

[4] Poluboina, V. , Pulikala, A. , & Muthu, A. N. P. . (2022). Contribution of frequency compressed temporal fine structure cues to the speech recognition in noise: an implication in cochlear implant signal processing. Applied Acoustics, 1(8)9, 108616-.

[5] Kurokawa, M. , Kasai, T. , Sugino, A. , Okada, Y. , & Kobayashi, R. . (2022). Investigation of the key factors that affect drip loss in japanese strawberry cultivars as a result of freezing and thawing. International Journal of Refrigeration74(134-), 134.

[6] Mizuguchi, H. , Maeda, Y. , Nishimura, K. , Shinkura, H. , & Kushibiki, S. . (2021). Effects of wood kraft pulp feeding on feed digestibility and rumen fermentation of japanese black steer in the middle fattening stage. Animal Science Journal, 92(1)74.

[7] Yang, S. . (2023). Application of intelligent voice technology and sensor network in production and operation management vr intelligent teaching system. International Journal of Reliability, Quality and Safety Engineering, 30(01).

[8] Tanribilir, R. N. . (2021). Analysing antecedence of an intelligent voice assistant use intention and behaviour. F1000 Research, 1(0), 496.

[9] Sharevski, F. , Jachim, P. , Treebridge, P. , Li, A. , & Adadevoh, C. . (2021). Meet malexa, alexa's malicious twin: malware-induced misperception through intelligent voice assistants. International Journal of Human-Computer Studies, 149(3), 102604.

[10] Correia Loureiro, S. M. , Japutra, A. , Molinillo, S. , & Bilro, R. G. . (2021). Stand by me: analyzing the tourist-intelligent voice assistant relationship quality. International journal of contemporary hospitality management7(11), 33.