A Voice Processing Technology for Rural Specific Context

Zhiyong He, Zhengguang Zhang, Chunshen Zhao

A VOICE PROCESSING TECHNOLOGY FOR RURAL SPECIFIC CONTEXT

Zhiyong He [1,*], Zhengguang Zhang [1], Chunshen Zhao [1]

1 Institute of Computer Application, Sichuan University of Science and Engineering, Zigong Sichuan, P. R. China 643000

** Corresponding author, Address: Institute of Computer Application, Sichuan University of Science and Engineering, Zigong 643000, Sichuan Province, P. R. China, Tel: +86-13778520180, Fax: +86-813-5505966, Email:hzy@suse.edu.cn*

Abstract: Durian the promotion and applications of rural information, different geographical dialect voice interaction is a very complex issue. Through in-depth analysis of TTS core technologies, this paper presents the methods of intelligent segmentation, word segmentation algorithm and intelligent voice thesaurus construction in the different dialects context. And then COM based development methodology for specific context voice processing system implementation and programming method. The method has a certain reference value for the rural dialect and voice processing applications.

Keywords: voice processing, COM, specific context, rural informatization

1. INTRODUCTION

Intelligent voice processing technology is to resolve how to make computers understand natural language of mankind and can be output of natural language fluently, so that the computer has the capacity of human language. It is mainly divided into the speech synthesis technology, voice recognition technology, and voice evaluation and voice coding techniques. In the popularization of agricultural informational, China has large rural population, the larger regional language differences. People have put forward higher requirements for speech technology.

In this paper, the problem is therefore in the formation of the use of COM technology, based on the specific context and mixed multilingual text as the

background voice processing, according to the characteristics of TTS technology, a library based on the establishment of specific context voice processing technology and intelligent voice processing rules system, in order to address the specific context of multi-lingual text and mixed-voice processing problems.

2. DEVELOPMENT AND APPLICATION OF TTS TECHNOLOGY IN SPECIFIC CONTEXT

Specific context refers to time, place, occasion, object and the use of objective factors such as language, identity, ideology, personality, occupation, self-cultivation, the situation, feelings and other subjective factors posed by the use of the language environment (Yibiao Yu et al., 2002; Bo Yang et al., 2005). Specific context of speech synthesis research, including the realization of multi-lingual speech synthesis, such as: minority languages, the local dialect speech synthesis, Chinese and foreign language. TTS that is, "from text to voice." It is the use of linguistics and psychology, in the built-in support chips, and smart text to flow into a natural voice. It can convert a text file in real-time, converting a short period of time can be seconds. In its unique role in the smart controller voice, the voice of the text output to achieve smooth temperament, making the listener feel when listening to nature, there is no voice output machines and jerky sense of indifference. Common TTS system mainly includes text analysis, prosodic processing, and speech synthesis. Text analysis of the input text linguistics is analyzed, the sentence in vocabulary, grammar and semantic analysis to determine the sentence structure and every word of the phonemes. Dealing with synthetic sound quality is not only a rhythm, but also making the sound quality of voice to speak close to the voice synthesis to ensure the natural tone of words, consistency. Speech synthesis is a good means to deal with the text of the corresponding word or phrase from the speech synthesis library extracted from the linguistic description into a speech waveform. The TTS core of this intelligent speech processing system used this technical route (Weijun Shen et al., 2000). Its technical route diagram is shown in Fig.1.

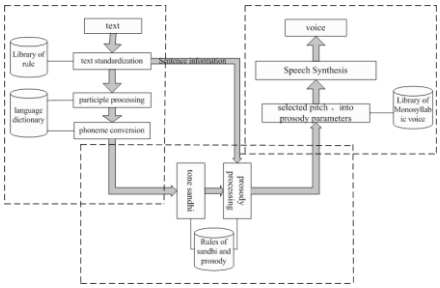

Fig.1 Technical route of TTS

For the specific context of the complexity of the situation, Clustering Algorithm uses in text analysis for word segmentation. First, all possible results are available in the stage of segmentation, part of the lexical analysis level ambiguous rule excluded. Second, excluded from the remaining ambiguous areas on the Chinese understanding of the follow-up stage, through a combination of ontological knowledge base for semantic analysis in order to rule out the ambiguity, and then the right does not understand the results of a new thesaurus word segmentation feedback to the thesaurus management system, at the same time record the frequency of its emergence, when the frequency of a standard to meet when it joined the segmentation thesaurus in order to achieve the recall of unknown words and improve the efficiency of system operation. At the same time in the understanding of understanding of the final stage is finished, before the quasi-automatic model segmentation results of the feedback module clustering words based on statistical training systems, clustering results at the same time feedback to the thesaurus management system, to improve the next segmentation accuracy and efficiency (Qiang Li et al., 2000).

Processing stage in the rhythm uses feedback forms of 3-layer structure self-organizing network (Chen Zhao et al., 2002). (1) Enter the relevant context information, the language of information along a statement (sentence) → prosodic phrase (phrase) → syllable (syllable) step-by-step breakdown of information to find syllable C [consonant type C1, vowel type C2, tone type C3, in the words of the location of C4, with the former syllable coupling C5, and after the coupling syllable C6]; (2) adjacent to the former syllable Information L [vowel type L1, tone type L2]; (3) adjacent to the former syllable information N [consonant type N1, tone type N2]; (4) syllable of the prosodic phrase information where W [number of syllables W1, the location of the sentence W2, accent type W3, stress from the previous distance W4, stress from a distance after the W5, stress and stress the distance between W6]; (5) statement of information S [statement type S1 and the number of prosodic phrases S2]. More than the prosodic features (17 acoustic parameters) as the neural network input, through a multi-sample competitive template, the template matching with the established best or with the most similar to natural sounds as the output template.

TTS voice processing module is a key link. PSOLA optimization algorithm is used that is time-frequency distribution algorithm. The algorithm adjusts voice pitch and time Len of the original voice splice units in frequency domain and time domain. First the voice processing module synchronous analysis and tags keynote when the rhythm parameters arrival,

and then the pitch of short-term analysis of signal changed in frequency domain (Min Han, et al., 2004).

3. INTELLIGENT VOICE PROCESSING SYSTEM IN SPECIFIC CONTEXT

Intelligent voice system uses hierarchically structured design. The system is divided into that presentation layer, business layer, data access layer and data layer. Which the business layer and data access layer contains the recognition engine to develop a special voice interface and database interface, the data layer with special voice library can be easily loaded under various voice module, to facilitate the system's compatibility, the core processing module is designed to voice control command recognition and voice processing. Its system structure diagram is shown in Fig.2.

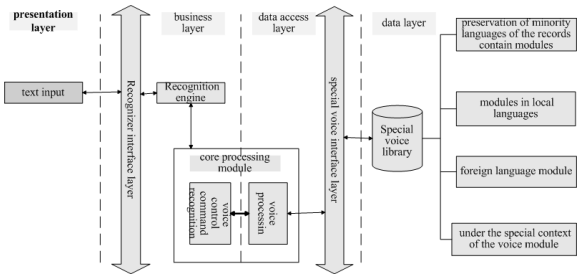

Fig.2 System structure of intelligent voice processing

The presentation layer is a simple application layer which collects text data. Recognition engine is composed of the Recognizer interface layer and the recognition engine modules. It captures packets and decodes, and then puts the results to a place designated in the data structure. The core processing module primary recognizes the data which sent from recognition engine, link data, distinct and process. These procedures include further word processing, determine errors, and call voice module. The core processing module uses voice-processing algorithms and rule base to match the rules of voice processing, find a voice module for synthesis. The special voice interface layer uses correlation analysis and sequence analysis to match and find a new voice module, and then sends the matched voice library to the core processing module by COM components methods. Special voice library contains the records kept by minority languages modules, local languages modules, foreign language module, and other specific voice context modules. It support upgrade a variety of voice expansion modules.

4. THE KEY TECHNOLOGY OF INTELLIGENT VOICE PROCESSING TECHNOLOGY

4.1 Voice processing control command recognition

Voice processing control command recognition is the core aspect of intelligent information processing. First of all, the text of the application to identify the command interface program designed to identify and initialize, then the statement of the interface objects required creating a shared text recognition engine. Second, we must create a specific context of the context of speech synthesis interface, the realization of multi-lingual speech synthesis engine function. Once the correct order of operations to be identified, the text sent to the host program identification information. A speech synthesis module interface is created, loaded and activated for application, when the application has been received information from incident immediately (Flanagan，J.L et al., 1972). Its process diagram is shown in Fig.3.

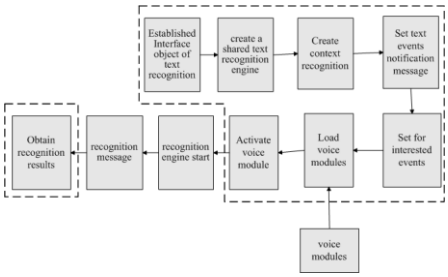

*Fig.*3 Voice processing command recognition

The main text input text data collection is a simple application layer. Including a variety of formats such as text input: text,. Doc files. Recognition engine by the Recognizer interface layer and the recognition engine modules, the interface is crawling information packet capture and decode analysis, and the results of the analysis to a place designated in the data structure. Is mainly responsible for the core processing engine will be sent to identify the data, connect data identification, handling. Include: further word processing, to determine errors, such as voice calling module. Core processing uses voice processing algorithms and rule base to match the rules of voice processing, direct find a voice module for synthesis. Treasury special voice interface layer through the use of correlation analysis and sequence analysis of the match to find a new voice module, method of use of COM components to match the voice to the core processing module library. Special voice library: preservation of minority languages of the records contains modules, modules in local languages, foreign language module,

under the specific context of the voice module, to support the expansion of a variety of voice module upgrades. (Randy Abernethy et al., 2000; Ash Rofail et al., 2005)

4.2 The development of COM components

COM (Component Object Model)(Microsoft Co., 2009) is a new software development technology; it is helpful to improve computer industry's software manufacture more in line with the behavior of human beings. Under COM framework, various kinds of components with specific function will be developed, and the components could be combined together as needed to compose a complex application system. Application or component system could be formed by the combination of multi COM objects. At the same time, component could be unloaded or replaced during its running, there is no need to re-link or compile the application. (Flanagan，J.L et al., 1972).

When the application started, the voice processing system creates a process to execute new application. Component program is realized as DLL (Dynamic Link Library). When there was language text input, the core processing module calls component program, loads the required component program to its process, and then connects to specific voice module in special voice library through the special voice library interface layer. After the establishment of communication between processing requirements and component program, the interface pointer received by core processing program will point to component program's vtable directly. The vtable pointed by demand module interface pointer includes member function's address, and then business layer could call service provided by interface directly. For the component outside of process, component program and presentation layer are not in the same process space, the communication between component program and client program must pass through the process border, therefore the component outside of process must be processed with dynamic DLL firstly, then the parameters and other call information will be assembled into a packet and passed to component program. When the component program received the data packet, it will unpack the packet, read out the parameter information, call the actual interface and send the result collected back to the core processing program at last. Then a function all has been complete.

5. CONCLUDING REMARKS

The design proposed in this paper, combined COM with voice intelligent processing technology. It is helpful to develop new ideas in future in many fields, such as multi language interaction, translation, search engine and etc. it could promoted information technology to develop better and faster.

Voice processing technology for specific context could be used in various fields, such as keyboard input, optical scanning, handwriting recognition, web-based database, PDA, home appliances, digital products and etc. it is helpful to solve the shortcomings of traditional applications and overcome the communication obstacles between human and machines. It would play a greater role in various fields in the future. The adoption of this technology will play a positive role in the rural information promotion at the same time.

REFERENCES

Yibiao Yu, Kai Duan, Rujie Shi, Prosodic Control for Speech Synthesis of Wu-Dialect Chinese text to speech system, Communications Technology, 2002,(10):87-90(in Chinese)

Bo Yang, Yinsuo Jia, Yonghong Li, Hongzhi Yu, Researches on prosody control technique and applications in Tibetan TTS, Journal of Northwest University for Nationalities(Natural Science), 2005,26(1):69-71(in Chinese)

Weijun Shen et al., A chinese Text-to-Speech system, Computer Entineering and Applications, 2000,36(9):76-80(in Chinese)

Qiang Li, Yakang Liu, Xueyong Zhu, An Algorithm of Pitch Prediction, Journal of University of Electronic Science and Technology of China, 2000,29(5):495-498(in Chinese)

Chen Zhao, Jianhua Tao, Lianhong Cai, Rule-learning Based Prosodic Structure Prediction, Journal of Chinese Information Processing, 2002,16(5):30-37(in Chinese)

Min Han, Lan Tian, Chinese prosody stepwise synthesis based on FD & TD PSOLA, Journal of Shandong University(Engineering Science), 2004,34(6):35-37(in Chinese)

http://www.microsoft.com/com/default.mspx, 2009

Ash Rofail, Yasser Shohoud, Translator: Pan Chung, COM and COM + from entry to the master, Publishing House of Electronics Industry, 2005: 136-165(in Chinese)

Randy Abernethy, translator: Hao Wang, COM/DCOM technology insider, Publishing House of Electronics Industry, 2000:56-70(in Chinese)