THE DEVELOPMENT OF DATA CAPTURING AND GEOCODING MODULES BY SPEECH-VOICE RECOGNITION

Dekkie TCHA and Prof. Kyujon CHO, Korea

Key words: Voice Recognition, Geocoding, GPS, Totalstation, Mobil Mapping.

ABSTRACT

Men's desire for the human interface, due to the development of voice processing technology of computer, and the development of intelligent MMI (Man-Machine Interface) computer technology enabled us to operate computers with our voice without using keyboards or other input systems. Especially, by obtaining field data from the complicated surveying environment and applying the voice recognition technology to the actual surveying work, we can save a lot of working hours and costs. According to the result of this study, the real time Geo-Coding and graphic data-coding were possible by connecting the software engine which recognizes 50,000 different words and the voice recognition technology based on the super IC which recognizes 60 different words with the Total-station and the RTK-GPS.

1. INTRODUCTION

Voice recognition is a software technique that identifies the closest results in pre-entered recognition data by extracting and analyzing the voice features of people delivered on computer through mike. This is divided into speaker-dependent technique and speaker-independent technique depending on the scope of users, and into small-word recognition technique and large-word recognition technique depending on the number of words for recognition. It can be also divided into word recognition technique, consecutive voice recognition technique, dialogue recognition technique depending on the naturalness of speaking. That is, voice recognition technique is one of core technologies that will naturally and conveniently improve the interface between human and machine, but it can be used for practical purposes only when the voice recognition level is over 95%. As a result of research on this field for last ten years, intelligibility has been obtained to some extent, but more research is required in the naturalness. However, in case of Geo-Coding, which is generated in the surveying process, there will be no difficulty in obtaining field data with intelligibility rather than naturalness. It is considered that the small-word recognition technique below 500 words can sufficiently process a surveying job. This study was proceeded to obtain site surveying data more conveniently by applying this voice recognition technique to the process of capturing surveying data. Search of a digital map, support of vehicle operation, information system for the visually disabled are cases that applied voice recognition technique to the surveying fields in Korea. It is difficult to obtain the voice recognition engine developed in Korea. Thus in this study, we combined the Speech Recognition Engine 4.0 of Microsoft, capable of supporting multi language (e.g., English, Japanese, and Chinese) and recognizing universal multilateral voice, Voice Director 364 Module Chip of Sensory with recognition level of 99.5%, and PCBASIC One-chip hardware of

1

Single Board Computer type, developed by Comfile Co., Ltd. In Korea. Also, we performed modularization and programmed to apply voice recognition technology to data acquisition and graphic processing by connecting Total Station, a measuring instrument, with Real Time Kinematic GPS, using Delphi 5.0 of Insprise Co., Ltd., a software development language.

2.VOICE RECOGNITION AND COMBINATION SYSTEM

Voice recognition technique can process data capturing more promptly than the existing process by keyboards or mouse, and can be used in working environment where these devices cannot be used. In addition, since it is easy to move and control the device, and does not require repetitive key input and mouse works, it can remove the monotonous and tedious process of entering data. As it allows Voice User Interface(VUI) including telephone line and modem, several users can work on it at the same time. In particular, it is economic in saving data since it can save more data with smaller capacity than voice signal. Voice recognition system processes the signal obtained from the signal sound of voice by several forms of recognition theory. Voice recognition techniques include Furistic algorithm by Fuzzy theory, Hidden Markov Modeling by real time visual chance, linear forecast factor, vector unit couple speech, NN(Neural Network) skill which uses formal learning ability by the human neural network. In particular, Natural Speaking technology at the stage of voice recognition converts the recognized information and text information to voice information based on various languages, to be more accessible to human dialogue mode, and currently achieves multi-language interpretation. Voice recognition technique has approached the stage of maturity; Microsoft USA loads voice recognition and natural speaking technology as one of core engines of Windows 2000. Top 3 companies including AT&T, Lucent Technologies and Motorola have developed VXML(Voice Extensible Makeup Language) technology which enables to search various information over the Internet through telephone and voice recognition software, and played a pivotal role in development of web-based contents services and voice recognition applications through phone. In particular, it is expected to be applied in surveying fields by entering the data using portable devices. Even though this voice recognition technique is still limited to the English speaking countries, it is expected to bring sweeping changes to the paradigm of customer marketing and multilingual support in the future.

2.1 Recognition System in terms of Software

Voice recognition engine developed by MS requires equipment to recognize voice, including mike for conversion of electric waveform signal of voice, digital processing voice card, speakers for reproduction of voice, voice recognition engine-speaking recognition engine-application soft. Voice recognition is to convert the voice signal captured from mike or phone, and to enable the function processing of computer, input of data, document works. And it requires the technique to reduce the stage of this voice recognition. This automatically recognizes the human voice with API, which can directly access to the independent hardware, and carry out interface filters to enable the user to do the works he wants. In case of notebook, these devices are integrated as one to enable the implementation without separate additional devices. Next, by processing the speaking voice, converting signal and operation system process in terms of hardware, it raises the word recognition rates through application of unique recognition

algorithm, voice recognition speed and training. When enabling the voice recognition in Windows environment, the words required in operating system and open menu-select menu-scroll-check box-select-click are set to be automatically executed without preparation of application. Accordingly, since the voice recognition for basic operation of Windows does not necessitate the preparation of separate program, it is sufficient to prepare program on the factors required for surveying only. Then, declare the procedure for this Window control as follows:

API Function : AddControl(controlHandle, 'control name', 'ctrlvoc',
      GrammarVocabulary, NullArray,0,windowHandle);

Engine used :   mscsrl.exe   - Microsoft Speech Recognition Engine V.4.0
       msttsl.exe   - Microsoft Text-to-Speech Engine V.4.0
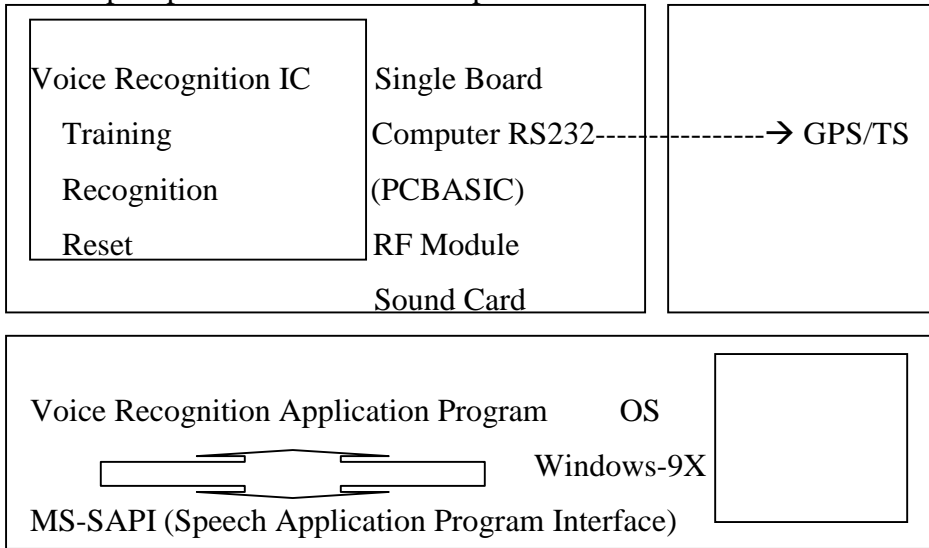   Spchapi.exe   - Microsoft Speech API 4.0

Voice Recognition IC	Single Board	
Training	Computer RS232----+----------→ GPS/TS	
Recognition	(PCBASIC)	
Reset	RF Module	
	Sound Card	

Voice Recognition Application Program   OS

MS-SAPI (Speech Application Program Interface)   Windows-9X

Fig. Structure of Voice Recognition Modules

2.2 Voice Recognition in terms of Hardware

Thanks to the advanced technology on voice processing speed by neural network algorithm, it is featured to implement voce recognition with moderate price, with small memory capacity than RISC chip. Voice recognition in terms of hardware builds in the voice recognition technique in single chip. It transmits the result through voice recognition and A/D signal conversion to EEPROM/ROM/Digital I/O, after preparing separate rules by communication, makes this as event, and prepares the software program to fit the requirement of users. Since there is a limit in memory when using hardware IC chip, the user should do recognition train directly, and in this case, it is operated as speaker-dependent recognition type. This voice recognition in terms of hardware saves the pattern of voice when it succeeds in recognition through trainings repeated more than 2 times. It is based on neural network theory, and can carry out rapid search
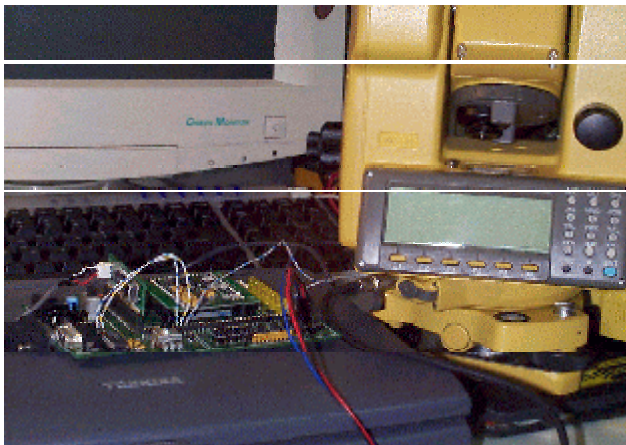
when listening for recognition. First of all, it is necessary to establish command system to be operated in software, define the word, set the training mode on IC, and execute the command by the verification and application mode of recognition repeated. In this case, it forms separate IC circuit, process the surveying performance in graphic, and connect it with notebook for operation.

2.3 Message Function

To report the result from surveying or the completed recognition to the user, it applies the message function from speaking engine, separate from character message, and applies the technique of changing the document for information on next process to voice. In this way, it delivers various messages to the surveyor. Almost all voice recognition engines contain Text-To-Speech.

2.4 Voice Training/Modeling Process

Since voice differs per person, it declares different models for several voices and selects the model similar to voice model of users, increasing the recognition rates. Since voice recognition engine supports more than 10 voice model, it can be reset whenever necessary or for lexical definition. It can determine the adoption of command based on statistical recognition rates by separate voice training. The process of voice recognition in surveying prepares the word and menu system for analysis of attribute, graphic processing and words required for orientation and operation of surveying instruments, which is necessary for surveying, and carries out repetition training to increase the recognition rates. In this case, since speaker-dependent recognition is mostly carried out, it is necessary to take care to make it as a natural speaking with proper tone and pitch. Since existing recognition engines record below 69%, surveying process should be carried out after improving the recognition rates over 90%, after training process. To enable several users to carry out works at the same time, it adopts the required mode among single mode, single series listening mode and group mode, and consecutively train the same command in the stages as shown below. 1st stage requires the ability of understanding the words and sentence structure, following the signal processing by voice recognition. But it should train the words required for surveying with artificial intelligence technique.

< Stage of training voice recognition >
1) After pressing Train key : "Say word one"   2) User : "Measure"
3) Voice Direct 364 : "Repeat"   4) User : "Measure"
5) Repeat the process from 1) to 4)   5) Save the result

3. CONFIGURATION AND OPERATION OF MODULE

To prepare the application by voice recognition, it is necessary to operate the engine to be connected with computer. In general, the engines, developed by MS and AT&T, have been widely used. In this thesis, we used voice recognition engine V4 of MS, operated it with Windows-9X/NT at the same time, controlled mike and speaker by API function, set the voice pattern of the person who talked, received or controlled the surveying performance by surveying program using serial communication, and processed and saved the performance through recognition system. Main programming

modules were configured with modules of signal conversion process by Single Board chip due to the IC operation for voice recognition, generation of code by serial communication, modules for processing graphic and receiving the GPS performance. In addition to hardware module, it passes through training process to raise the degree of understanding, compress into several words and passes through standardization process. The minimum words required for processing surveying process, and geo-features of ground facilities are the graphic recognition words for CAD screen processing, surveying words to be used in the setup an instrument point – orientation of back sight – observation and recording process, which are required in surveying, and the words for processing topology and selecting the code to classify the geo-feature. Of course it is possible to use lots of recognition words, but the user may not remember the words, if it is too much.

4. APPLICATION AND ANALYSIS OF VOICE RECOGNITION

Geo-feature can be largely divided into surveying activities and geo classification information. Surveying software, surveying standard of each country, surveying organization have its unique coding system. The management of this geo-feature often uses unique methods given the efficiency of program, the features of works at organization. To unify the diversity, it converts the external file using Look-Up Table to change the definition between codes. The efficiency is high only when adopting the minimum number of recognition words required for voice recognition through analysis per feature of code. This coding data vocalizes error and warning message, the message on accuracy of surveying, work confirmation message, CAD screen operation, graphic processing. This voice recognition module uses the method of using sound cards in Notebook in terms of software, and the method of forming separate hardware and IC, and transmitting the commands by serial communication. Software module receives the data from communication port of GPS receiver, converts the GPS performance, the longitude and latitude coordinates of NMEA(National Marine Electronics Association) to national coordinates, and processes the works of classifying the attribute, Tag and event work on surveying point, screen display of receiving coordinates. It consists of the preparation of the words required for surveying and voice recognition, inspection of recognition level, Speech-to-Text training and the process of inspecting and processing the exactness. By operating this software, it can provide the security measures by speaker-dependant, easy and convenient working environment, by avoiding the limit of simple key entering works.

4.1 Application of Geo Coding on Total Station

For surveying works, install instrument point where is the standard point of surveying, enter the coordinate of back sight point, and generate the operation code. Then, calculate the coordinates of instrument height and back sight point, prism height and reference orientation angle, and call the voice recognition command to set the instrument installation and back orientation. Then, collimate on all points, execute the command by voice recognition words without keyboard work, and enter the classification code of surveying point. Then, execute graphic processing and topology processing. In case of graphic processing, use separate prefix or attribute, using surveying performance and graphic narrator. Almost all uses prefix or narrator (Ex: X,Z,-,GPS, TS, STN,BK,SS---), but direct graphic processing at site is more effective.

4.2 Application of RTK(Real Time Kinematic)GPS

Since GPS provides location information in real time, convert it to regional coordinates using the parameter with current regional coordinates. In particular, determine whether there is a consecutive surveying in accuracy from the information, deliver the message to user, and finally capture it in accordance with coding process. After back processing on this, this can be converted to the data that user wants. The relations among these works pass through the process as shown below.

- Determine the coordinates of reference station of RTK GPS and RTCM broadcasting by Radio
- Calibration of Rover station by RTCM
- Extract the location information from NMEA and determine the parameter of regional coordinate conversion and use the conversion of local coordinates
- Track the graphic display of location information (GPS location),screen or CAD process by voice recognition,coding process of topography by voice recognition
- Post-process of observation data

Accordingly, the voice recognition system is applied with both total station and GPS at the same time, it may make the worker feel tedious, depending on the performance of hardware, when searching surveying point and processing the topology if several persons survey the same region. However, since most daily work amount is below 1,000 points, it does not cause any big problems.

4.3 Case of Application

Since voice recognition engine enables to carry out the interface function between computer and human interface, it applied this to two aspects of structured data entering and graphic work process in the complicated site data capturing process that accompanies the surveying. For the former, connect it with surveying CAD program indoors, and track the surveying points and other points through voices in a short time, including movement, expansion, reduction and scroll of screen. Since it saves the process of key entering process from the computer works, it makes the site surveying works easy. In particular, it can process the separate works in coding process by GPS. For this graphic processing, present the existing CAD DXF sketch or data of existing graphic pattern on the screen, and then, process the topology point of surveying relation in coding works by voice. For preparation of program, first determine the conversion variable of GPS coordinates in system, receive the data of $GPGGA form of original GPS land surveying coordinates, the method of using regional coordinate performance converted by direct parameter at controller in NMEA-0183(National Marine Electronics Association) sentence, adopt the oval and reflection method in PC, calculate the parameter and convert it. Display the antenna location due to real time movement using this coordinate, import the existing digital map data and screen processing function. Then operate voice recognition engine, and carry out coding per geo classification, topography and operation of machine related to surveying work and graphic processing(CAD) by combination and recognition of voice.

- Coding Processing of GPS Data
  Since GPS data consists of Point Number, Time, Position, Quality, Attribute, Topology, Reference information, the process of operation codes of surveying work such as Totalstation is not required. That is, it does not require weather information for compensation, information on prism height and machine height, setup process of orientation angle on machine point and back sight. Surveying time, point name, coding and [String] process are required.
- LUT(Look Up Table) code processing
  The method of code is divided into character code and number code. It is easy to recognize the character code in operating and preparing the program, but it has the complicated entering process and hardware processing. Convert the definition using LUT to convert the unique code of organizations and software for office job, software for site data capturing, and site code system conversion.
- Neutral file processing
  Since structure of field observation file is obtained using unique forms of hardware, it is to process this in surveying software for office job.

4.4 Result of Experiment

As a result of applying voice recognition in site surveying, the length of movement among machine points in total station is long, and the daily work amount is about 700 points. In case of GPS, since it enables to survey about 2,000 points, the distance among surveying points are relatively short, and the configuration of equipment is complicated, it is more effective to use voice recognition technique. Hardware IC shows prompt response speed of 1-2 seconds, demonstrating the fact that recognition rate is over 90%. However, it was difficult since the recognition rate in software has 2-3 second latency, and the recognition rate is below 90%. Accordingly, it is more effective to use by hardware IC rather than to use software recognition method. However, it still has problem in that the available number of words to be recognized in hardware is only 60 words at maximum.

5. CONCLUSION

There is a limit in computer technology in natural sketch in terms of voice recognition. Since lots of constraints are given in personal features, surrounding environment and computer environment, voice recognition cannot be perfectly commercialized and launched into the market. However, it is expected that the technology transcending the human real time will be emerged due to development of computer environment and processing algorithm. The use of recognition technology in surveying areas will be essential. Korean recognition engine has been developed in several institutes, and in particular, Korean technology has developed voice recognition IC, and is expected to advance the days of application in surveying areas. Despite voice recognition in surveying, voice recognition in capturing the site data using few word recognition has many merits. In case of speech, since there are lots of stop times for more than 2 persons in particular in total-station, and warning message, confirmation and information message, the application of voice recognition system has no significant meaning. However, since voice recording system and recognition system can be applied to observation of real time GPS and preparation of geo-featuring in moving object, it is

highly efficient. In particular, IC chip is slightly superior to software type in speed and performance of recognition, but hardware has weak performance in processing. By applying this voice recognition system, following conclusion can be drawn.

1. To use hardware voice recognition IC is much efficient in speed than voice recognition engine of MS.
2. Enable basic geo-coding processing by 25 few words
3. Has work efficiency in current surveying method than coding process.
4. Advantageous in measuring and coding in bad weather conditions compared to existing method.

Since voice recognition can be designed by optimizing the new menu system by user definition, it is possible to resolve the difficult surveying environment. At present, voice recognition engine by hardware IC was built in product and is in the process of commercialization. Since new technology on voice recognition system due to the development of computer hardware and software can be applied to surveying areas in the future, voice recognition technology should be applied and developed to resolve the problem in observation and surveying process on more effective geo objects with standardization and systematization on this.

REFERENCES

1. http://www.emind.co.kr
2. http://members.tripod.lycos.co.kr/rgs1109/it/1215message.html
3. Neural Networks And Their Use,Sensory, Inc., catalogue,pp-1
4. RTCM Recommended Standards for Differential GNSS(Global Navigation Satellite Systems) Service Version 2.2,RTCM Special Committee No.104, 1998.1.15
5. Voice Direct 364 Data Book Manuan,Sensory,Inc.,2000

CONTACT

Dekkie Tcha
Instructor
Department of Civil Engineering
Kyunggi University
Sangok4 dong
Kyungnam APT
Byupyung Gu
Incheon
KOREA
Tel. + 82 31 249 9718
E-mail: tcha@hananet.net

Professor KyuJon Cho
Kyonggi University, Seoul
KOREA
Vice President of Korean Association of Surveying and Mapping
Vice President of Korea Confederation of Surveyors
Tel. + 82 31 249 9708
Email: kyujon_cho@hotmail.com