Journal of Navigation and Port Research International Edition
Vol.35, No.7 pp. 551~556, 2011 (ISSN-1598-5725)
DOI : http://dx.doi.org/10.5394/KINPR.2011.35.7.551

Implementation of Real-time Wheel Order Recognition System Based on the Predictive Parameters for Speaker's Intention

*Serng-Bae Moon**, † *Seung-Hwan Jun*

*, † *Div. of Navigation Science, Korea Maritime University, Busan 606-791, Republic of Korea*

Abstract : *In this paper new enhanced post-process predicting the speaker's intention was suggested to implement the real-time control module for ship's autopilot using speech recognition algorithm. The parameter was developed to predict the likeliest wheel order based on the previous order and expected to increase the recognition rate more than pre-recognition process depending on the universal speech recognition algorithms. The values of parameter were assessed by five certified deck officers being good at conning vessel. And the entire wheel order recognition process were programmed to TMS320C5416 DSP so that the system could recognize the speaker's orders and control the autopilot in real-time. We conducted some experiments to verify the usefulness of suggested module. As a result, we have confirmed that the post-recognition process module could make good enough accuracy in recognition capabilities to realize the autopilot being operated by the speech recognition system.*

Key words : *Speaker's intention, Autopilot, Wheel order, Speech recognition, TMS320C5416 DSP*

1. Introduction

New generation navigation equipments such as GPS, AIS, ECDIS have been helping the officer of the watch manage an ever-increasing amount of information efficiently and make decisions relating to navigation safety effectively. And Integrated Bridge System makes it possible that the watch officer can control the main engine RPM directly without any interference of the engineers. These high technology systems bring about the transition of watch-keeping environment in the bridge and realize the One-Man Bridge System(hereafter OMBS) that only one officer has to make a safe navigation alone without assists from other crew members, not even a quartermaster(Park, 2007).

But OMBS should not be carried out in the heavy traffic water area or at night. In the conditions the officer has some difficulties in performing lookouts and operating the autopilot manually. And he/she might be in critical situation after all.

Even though he/she sails around the safe waterway during daytime aboard on OMBS, he/she may change the heading to avoid dangerous targets manually at times. If the marine traffic is increased and the officer has so many works to do during the watch, it's possible that the OMBS may give rise to the risky situation.

For this reason, some researches recently have been conducted on the new system that can recognize the duty officer's wheel orders and control the rudder angle and ship's heading as part of endeavor to resolve it. Most researches so far studied have typically depended on the universal speech algorithms to improve the recognition rate without respect to the specific conditions such as the noise in the bridge, the structure and vocalization methods of wheel orders, etc(Park, 2007; Son and Kim, 2003). This is why the recognition rate could not get to the satisfactory level that can be applied to real autopilot system on the bridge.

In this study new parameter values for predicting the likeliest wheel order based on the previous wheel order were introduced into the post recognition process in order to enhance the recognition rate which would be the very important factor in realizing the autopilot system operated by speech recognition algorithm.

Meanwhile, the rapid progress of semiconductor technologies has made it possible that high performance micro-processor and large-storage memory chip could be available for cheap prices. And also it has put the speech recognition technology to practical and commercial use(Chung, 2007). The real-time ship's wheel order recognition system have to not conduct the automatic recognition of wheel order but also control of autopilot and

* Responsible author, msbae@hhu.ac.kr 051)410-4280
† Corresponding author, korjun@hhu.ac.kr 051)410-4245

playing of recognized order. Therefore we have loaded the entire recognition process on the TMS320C5416 DSP(Digital Signal Processor) with integrated memory and high-performance CPU and stereo codec, etc..

2. Speaker's intention predicting recognition system

2.1 Conversion of analog speech sound

The wheel orders that the user speaks out are the analog signals. They should be converted into digital signals for the speech recognition.

In this system, the model of G330 headset manufactured by Logitech company was chosen to receive the wheel orders. Fig. 1 shows the headset that has a microphone and headphones. And the principal specifications of the microphone and headphones are summarized as follows(Logitech):

Fig. 1 Headset(G330 Model)

<Specification of Headphone>

- Driver: 30mm diameter; neodymium magnet
- Frequency response: 40Hz ~ 18 kHz (-10 dB)
- Impedance: 32 ohm nominal
- Sensitivity: 94 dB SPL Ref: 1 mW (-18.2 dBV = 1m W), 1 kHz

<Specification of Microphone>

- Pickup pattern: Unidirectional (cardioid)
- Frequency response: 100Hz ~ 10 kHz
- Sensitivity: -44 dBV/Pa re: 0 dB = 1Pa, 1kHz
- Test conditions: 3.0V, 2.2 Kohm

The conversion of the analog signals into digital signals was carried out by PCM3002 stereo codec embedded on TMS320VC5416 DSP starter kit(DSK). It has Analog-to-Digital Converter(ADC) and Digital-to-Analog Converter (DAC) functions with single-ended analog voltage input and output. The ADC and DAC employ delta-sigma modulation with 64-times over-sampling. The ADC include a digital decimation filter and the DAC include an 8-times over-sampling digital interpolation filter. The principal specifications of PCM3002 codec are summarized as follows(Texas Instruments):

- Monolithic 20 Bit ADC and DAC
- 16/20 Bit input/uutput data
- Software control
- Stereo ADC:
- Single-ended voltage input
- Anti-aliasing filter
- 64× over-sampling
- SNR : 90 dB

This codec was set to 8kHz sampling frequency and 16-bit quantization in consideration of the sentence length of wheel orders and the size of data memory.

2.2 Design of Speech recognition program

The speech recognition program was designed based on isolated word recognition algorithm. Because all wheel orders for recognition are composed of a small number of words. The maximum number of words is four.

1) Wheel order and each word detection module

Accurate wheel order and each word detection program is very important to achieve the high recognition rate. Because it is essential to confirm whether the wheel order is spoken or not, and distinguish the exact starting point and end point of the words from continuously converted data.

Most speech detection theory can be characterized into three classes in general. These are the frame energy detection, zero-crossing rate detection, and higher-order-spectral detection that are used for speech and noise classification. The frame energy detection algorithm is not robust at huge background noise environment, and the zero-crossing rate detection algorithm is not robust in a various type of noise sources. Even though the higher-order-spectral detection algorithm is robust in complicate noise environment, it requires a heavy computational complexity.

In this system the frame energy detection algorithm was used for the process of speech detection, taking into account the noise environment at bridge and hardware capabilities such as memory and CPU. This algorithm basically detects the existence of speech from each short-term frame data. Each frame is consisted of 100 samples at a 8kHz sampling rate.

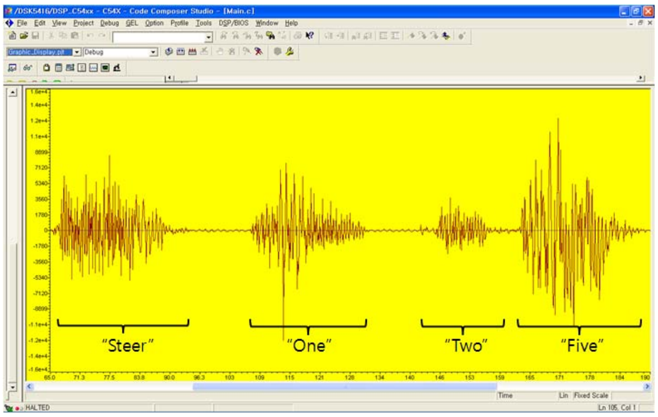

Fig. 2 Determined waveform of "Steer one two five' wheel order

The speech detection module was designed and programmed to find the starting and end points of whole wheel order sentence and each word at this sentence. If the starting point of the sentence is detected, the module will store the entire data to the detected end point. The starting point is counted at the first frame when all energy values of four continuous frames are above the threshold value. And also the end point is decided at the first frame when all energy value of four frames is lower than the threshold value.

The starting and end points of each word are detected within the data. Two frames are the basic length to find out the starting and end points of each word. Fig. 2 presents the entire detected waveform of wheel order sentence 'Steer one two five' which was determined by cutting off the silent parts based on the starting and end point decided by this module.

2) Feature vector extraction and vector quantization

Feature vector sequence is used for making code book which is need in the process of word recognition using pattern matching algorithm. There are a few ways to extract the feature vectors from a data sample sequence. These are LPC(Linear Predictive Coding) Cepstrum, PLP(Perceptual Linear Prediction) Cepstrum, MFCC(Mel Frequency Cepstral Coefficient), Filter Bank Energy, etc. In this study the step of feature extraction was programed based on the LPC Cepstrum which is derived from the model of speech organs and generally applied to the speaker-dependant recognition system.

The LPC Cepstrum coefficients was obtained from the LPC coefficients which are calculated based on the Levinson-Durbin algorithm. We have as c_m coefficients(for order p)(Rabiner and Juang, 1999):

$$c_m = a_m + \sum_{k=1}^{m-1}\left(\frac{k}{m}\right)c_k a_{m-k}, \quad 1 \leqq m \leqq p \tag{1.a}$$

$$c_m = \sum_{k=m-p}^{m-1}\left(\frac{k}{m}\right)c_k a_{m-k}, \qquad p < m \tag{1.b}$$

where a_m is the LPC coefficients, p is the order of LPC and m is the order of cepstrum.

Code books used for pattern matching process are obtained by the quantization of LPC Cepstrum coefficients in two types. These art a single section code book and multi-section code book. It has known that the recognition rate of multi-section code book is better than single section code book. So the recognition system was programed that the code books were made by the type of multi-section code. The size of section was 100 samples, and overlap size was 50 samples.

The quantization of LPC Cepstrum coefficients was performed by the K-means method that is a simple unsupervised learning algorithm.

3) Pre-recognition process

The typical word speech recognition process is the pattern matching between unknown input pattern and reference patterns. It is performed by measuring the euclidean distance between the unknown input pattern and reference patterns, and detecting one of the reference patterns which provide the smallest distance. But it is common phenomenon that two given word data sequences(time-dependent) have different length in time series. In this study DTW(Dynamic Time Warping) algorithm being useful at the speaker-dependent word speech recognition system is used to enhance the recognition rate and measure similarity between two patterns(Kim, 2003). After measuring the euclidean distance by DTW, the candidate words are decided.

4) Post-recognition process

We suggested the post-recognition process which multiplies the euclidean distance of the candidate words by the weight value of predictive parameters based on the word number of wheel order sentence and relative position, and decides the final word.

The post-recognition process is given by

$$\mathrm{D_W}(n) = \min\left[D_{CW}(n,k) \cdot W_N(n,k) \cdot W_L(n,k)\right] \tag{2}$$
$$, \ 1 \leqq k \leqq 19, \ 1 \leqq n \leqq wn$$

where D_{CW} is the euclidean distance of the candidate words measured by DTW template matching program, W_N is the weight value obtained from the word number of wheel order sentence, W_L is the weight value obtained from relative position of each word, k is the whole of wheel words, $wn(1 \leq wn \leq 4)$ is the word number of wheel order sentence(Moon, 2008).

In this study we have developed another parameter that can predict the likeliest wheel order being able to follow the previous wheel order. For example, if the wheel order, "Hard port" or "Hard starboard", is spoken first, "Ease-to ten" will be selected as the next wheel order than

Table 1 The average probabilities of predictive parameter

1st order / 2nd order	Midships	Port five	Port ten	Port fifteen	Port twenty	Port Twenty five	Hard port	Starboard five	Starboard ten	Starboard fifteen
Midships	0.08	1.00	1.00	1.00	1.00	0.95	0.93	1.00	1.00	1.00
Port five	1.00	0.08	1.00	1.00	1.00	0.95	0.95	0.63	0.80	0.78
Port ten	1.00	0.95	0.08	1.00	1.00	1.00	1.00	0.73	0.28	0.75
Port fifteen	0.98	0.95	0.90	0.08	0.95	0.93	0.93	0.70	0.70	0.23
Port twenty	0.88	0.93	1.00	1.00	0.08	0.93	0.95	0.65	0.73	0.70
Port Twenty five	0.60	0.83	0.75	0.95	0.95	0.08	0.88	0.58	0.50	0.63
Hard port	0.65	0.73	0.78	0.98	0.98	0.95	0.08	0.50	0.53	0.65
Starboard five	1.00	0.60	0.68	0.60	0.63	0.55	0.55	0.38	0.88	0.83
Starboard ten	1.00	0.58	0.65	0.58	0.60	0.53	0.55	0.80	0.45	0.83
Starboard fifteen	0.98	0.55	0.55	0.55	0.53	0.43	0.48	0.80	0.75	0.40
Starboard twenty	0.88	0.48	0.53	0.50	0.48	0.43	0.48	0.75	0.80	0.78
Starboard twenty five	0.70	0.45	0.45	0.43	0.28	0.38	0.40	0.68	0.68	0.73
Hard Starboard	0.65	0.38	0.38	0.35	0.35	0.48	0.35	0.58	0.63	0.68
Ease-to five	0.00	0.00	1.00	1.00	0.98	0.98	0.98	0.00	1.00	1.00
Ease-to ten	0.00	0.00	0.00	1.00	1.00	1.00	1.00	0.00	0.00	1.00
Ease-to fifteen	0.00	0.00	0.00	0.00	0.90	0.78	0.90	0.00	0.00	0.00
Ease-to twenty	0.00	0.00	0.00	0.00	0.00	0.73	0.80	0.00	0.00	0.00
Steady	1.00	1.00	1.00	0.98	0.90	0.85	0.88	1.00	1.00	0.98
Steady as she goes	1.00	1.00	1.00	0.98	0.90	0.85	0.88	1.00	1.00	0.98
Steer one two five	1.00	1.00	1.00	0.98	0.93	0.90	0.90	1.00	1.00	0.98

1st order / 2nd order	Starboard twenty	Starboard twenty five	Hard starboard	Ease-to five	Ease-to ten	Ease-to fifteen	Ease-to twenty	Steady	Steady as she goes	Steer one two five
Midships	1.00	0.95	0.93	1.00	1.00	1.00	0.98	0.98	0.98	0.95
Port five	0.75	0.70	0.70	0.35	0.98	0.95	0.70	1.00	1.00	1.00
Port ten	0.73	0.70	0.70	1.00	0.33	0.93	0.93	1.00	1.00	1.00
Port fifteen	0.68	0.60	0.60	0.93	0.83	0.28	0.80	0.93	0.93	0.95
Port twenty	0.18	0.60	0.60	0.95	0.95	1.00	0.25	0.90	0.90	1.00
Port Twenty five	0.60	0.15	0.55	0.78	0.85	0.93	0.90	0.83	0.83	0.83
Hard port	0.65	0.60	0.15	0.80	0.93	0.98	0.98	0.80	0.80	0.80
Starboard five	0.88	0.78	0.78	0.35	0.98	0.95	0.95	1.00	1.00	1.00
Starboard ten	0.88	0.83	0.85	1.00	0.33	0.93	0.93	1.00	1.00	1.00
Starboard fifteen	0.80	0.75	0.80	0.93	0.83	0.33	0.83	0.93	0.93	0.95
Starboard twenty	0.38	0.75	0.83	0.95	0.95	0.95	0.25	0.90	0.90	1.00
Starboard twenty five	0.60	0.30	0.73	0.78	0.85	0.88	0.90	0.83	0.83	0.83
Hard Starboard	0.68	0.80	0.28	0.80	0.93	0.98	0.98	0.80	0.80	0.80
Ease-to five	0.98	0.98	0.98	0.08	0.98	0.73	0.98	0.00	0.00	0.00
Ease-to ten	1.00	1.00	1.00	0.25	0.08	1.00	1.00	0.00	0.00	0.00
Ease-to fifteen	0.90	0.78	0.90	0.00	0.00	0.08	0.90	0.00	0.00	0.00
Ease-to twenty	0.00	0.73	0.80	0.00	0.00	0.00	0.08	0.00	0.00	0.00
Steady	0.90	0.85	0.88	1.00	1.00	0.98	0.85	0.08	0.25	0.25
Steady as she goes	0.90	0.85	0.88	1.00	1.00	0.98	0.85	0.25	0.08	0.25
Steer one two five	0.93	0.90	0.90	1.00	1.00	0.98	0.90	1.00	1.00	1.00

"Midships" order sentence. But "Ease-to ten" can be substituted by "Port ten" or "Starboard ten". It means that there is no fixed methods in using the wheel orders. In other words: what kind of wheel order is going to come after the previous wheel order is variable depending on an individual preference and inclination of the officer in charge of conning the vessel.

Five certificated deck officers who have wide experience in navigation at sea were requested to determine the estimated probability values in a range of between 0.0 and 1.0. The higher value means that the second order can be likely to come after the first order.

Table 1 shows the average of the values determined by five deck officers. The final post-recognition process including the probabilities is given by

$$F_W(n) = \min [D_W(n) \cdot P_A(n, k)] \tag{3}$$

where P_A is the average probability values of predictive parameter.

3. Experiments and evaluations

3.1 Experimental environment

TI(Texas instrument) has been developing a various type of DSPs. In this study the program for the real-time wheel order recognition system based on the predictive parameters of the speaker's intention was developed to be operated through TMS320C5416 DSP in consideration of the number of recognition words, the size of memory, and the length of program, etc. TMS320C5416 DSP is the fixed-point processor and consisted of high-performance and power efficiency CPU, on-chip RAM, ROM, and a various peripheral devices.

The recognition program has been designed and tested by TMS320VC5416 DSK(DSP Starter Kit) that was a low-cost development platform designed to be able to design the power-efficient applications based on TMS320C5416 DSP. The kit has some performance-enhancing features such as USB communications and true plug-and-play functionality(Texas Instruments).

Twenty wheel orders were selected for the experiments. They were consisted of nineteen rudder angle designating orders and one ship's heading designating order. Even though three hundred sixty ship's heading designating orders could be used for the conning vessel, only one sample order was selected to test the usefulness of suggested parameter.

Table 2 Scenarios for experiments

	1st scenario	2nd scenario	3rd scenario
1	Hard port	Port twenty	Steer one two five
2	Ease-to fifteen	Port fifteen	Hard starboard
3	Port five	Port five	Starboard twenty
4	Midships	Starboard five	Starboard ten
5	Starboard twenty	Steady as she goes	Midships
6	Ease-to ten	Starboard twenty five	Port fifteen
7	Midships	Ease-to ten	Port twenty five
8	Port ten	Midships	Port twenty
9	Ease-to five	Port twenty	Port ten
10	Steady	Hard port	Midships
11	Hard starboard	Ease-to twenty	Starboard fifteen
12	Starboard fifteen	Midships	Starboard twenty five
13	Starboard five	Starboard ten	Ease-to ten
14	Steady as she goes	Ease-to five	Hard starboard
15	Steer one two five	Steady	Ease-to twenty

Table 3 Result of recognition experiments

(unit : percentage)

Wheel orders	Pre-recognition process	Post-recognition process
Midships	94.4	100.0
Port five	75.0	100.0
Port ten	83.3	100.0
Port fifteen	75.0	91.7
Port twenty	66.7	91.7
Port twenty five	50.0	83.3
Hard port	83.3	100.0
Starboard five	75.0	91.7
Starboard ten	66.7	83.3
Starboard fifteen	75.0	91.7
Starboard twenty	83.3	83.3
Starboard twenty five	83.3	91.7
Hard Starboard	88.9	94.4
Ease-to five	66.7	91.7
Ease-to ten	58.3	83.3
Ease-to fifteen	83.3	91.7
Ease-to twenty	66.7	83.3
Steady	75.0	100.0
Steady as she goes	83.3	100.0
Steer one two five	75.0	91.7
Average	75.41	92.22

If the wheel orders are randomly spoken without consideration of the common conning methods, the parameters predicting the speaker's intention can not enhance the recognition rate. So three conning scenarios in Table 2 were drawn up to check whether the post-recognition process works better than the pre-recognition process. The scenarios based on the expert knowledge and experience of conning vessel were made up so that all wheel orders were spoken at least two times.

The scenarios were spoken six times one by one. To demonstrate the superiority of post-recognition process, the recognition program was designed to measure and display the euclidean distances between input pattern and reference patterns in each recognition process.

3.2 Result of experiments

The recognition rates were not calculated with each of all words compositing the wheel order sentences. In other words, all words of each wheel order should be recognized correctly. For this reason, the recognition rates of the sentences were lower than those of the words.

In the experiments of wheel order recognition, the pre-recognition process being applied by only DTW algorithm recorded about 75.41% of recognition rate. But the post-recognition process recorded about 92.22% of recognition rate. Therefore the predictive parameter of speaker's intention made progress in recognition capabilities with about 16.81%.

Specially 'Port twenty five' wheel order showed the lowest recognition rate among the results of pre-recognition process. But it was improved by the post-recognition process from about 50% to 83.3%. In the result of post-recognition process all orders showed more than 90% rate except 'Port twenty five', 'Starboard ten', 'Starboard twenty', 'Ease-to ten', and 'Ease-to twenty'.

4. Conclusions

In this study we have suggested the final step of post-recognition process that was developed to enhance the recognition rate of wheel order recognition system. The process multiplies the euclidean distance measured on the pre-recognition process by the predictive parameter values for the speaker's intention. The values were assessed by five certified deck officers being good at conning vessel.

The entire wheel order recognition process were programmed to TMS320VC5416 DSK so that the system could recognize the speaker's wheel orders and control the autopilot in real time. We conducted some experiments to verify the usefulness of suggested module.

As a result, we have confirmed that the post-recognition process module predicting the speaker's intention could make good enough accuracy in recognition capabilities to realize the autopilot being operated by the speech recognition system. If the parameter values are supplemented by the experiments for the various ship handling cases, the accuracy of wheel order recognition will be improved.

References

[1] Ann, T. O., Byun, Y. K., Kim, S. H.(1990), Isolated Word Recognition by HMM using Multisection MSVQ, Journal of Institute of Electronics Engineers, Vol. 27, No. 9, pp.158-159

[2] Chung, I. J.(2007), Implementation of a Single-chip Speech Recognizer Using the TMS320C2000 DSPs, Speech Sciences, Vol. 14, No. 4, pp.157-167

[3] Cole R., Hirschman L.(1995), The Challenge of Spoken Language Systems : Research Directions for the Nineties, IEEE trans. on Speech and Audio Processing, Vol. 3, No. 4, pp.1-18

[4] Kim, H. L.(2003), The technology of Speech Recognition, Journal of the Korean Multimedia, Vol. 7, No. 2, pp.16-20

[5] Logitech, http://www.logitech.com/en-gb/gaming/headsets /devices

[6] Moon, S. B.(2008), Enhancement of Ship's Wheel Order Recognition System using Speaker's Intention Predictive Parameters, Journal of the Korean Society of Marine Engineering, Vol. 32, No. 5, pp.791-796

[7] Park, G. K.(2007), Ship automation and marine safety information system, Journal of Institute of Electronics Engineers, Vol. 34, No. 11, pp.1262-1270

[8] Rabiner L. R., Juang B. H.(1999), Fundamentals of Speech Recognition, Prentice Hall, pp.112-116

[9] Son, N. S., Kim, S. Y.(2002), Development of Voice Information System for Safe Navigation in Marine Simulator, Journal of the Korean Society for Marine Environmental Engineering, Vol. 5, No. 3, pp.28-31

[10] Texas Instruments, http://www.ti.com/tool/tmdsdsk5416

Received 3 August 2011
Revised 20 September 2011
Accepted 20 September 2011