

NAVIGATION SECURITY MODULE WITH REAL-TIME VOICE COMMAND RECOGNITION SYSTEM

Mustafa Yagimli

Okan University, Vocational School, Department of Property Protection and Security, TURKEY

Huseyin Kursat Tezer

Turkish Navy Academy, Institute of Naval Science and Engineering, Tuzla, TURKEY

ABSTRACT

The real-time voice command recognition system used for this study, aims to increase the situational awareness, therefore the safety of navigation, related especially to the close manoeuvres of warships, and the courses of commercial vessels in narrow waters. The developed system, the safety of navigation that has become especially important in precision manoeuvres, has become controllable with voice command recognition-based software. The system was observed to work with 90.6% accuracy using Mel Frequency Cepstral Coefficients (MFCC) and Dynamic Time Warping (DTW) parameters and with 85.5% accuracy using Linear Predictive Coding (LPC) and DTW parameters.

Keywords: Maritime Navigation, LPC, MFCC, DTW, Voice Command Recognition

INTRODUCTION

Due to technological development, marine traffic has increased and navigation became more demanding for the crew ([1], [2]).

In the literature, there are voice recognition studies in which different techniques are used. Zhizeng and Jinghing used the Linear Predictive Coding (LPC) reciprocal spectrum coefficient as an eigenvector, and adopt Dynamic Time Warping (DTW) to process the LPC reciprocal spectrum coefficient. In their experiment, the correctness of recognition is more than 90% [3]. Bala et al. discussed Mel Frequency Cepstral Coefficient (MFCC) and DTW modules used for voice recognition systems, which are important in improving its performance. They presented the feasibility of MFCC to extract features and DTW to compare the test patterns [4]. Washisht et al. presented a study of robust speaker recognition for regional Indian accents with MFCC and DTW [5]. Ferrando et al. developed such a system, but able to recognise predefined words under water. They suggested to use a wedding between the DTW parameter, and a multi-resolution analysis algorithm: the Mallat algorithm [6].

Unlike the analyses mentioned in this study, by using MFCC, DTW and LPC parameters, a system has been

developed that will test the compatibility of command, and its implementation between the cruise control person and the steersman.

Currently, the systematic of the course of the ship occurs between the person who gives the command and the steersman that implements it; the accuracy of the implementation is observed primarily by the person who gives the command and by the cruise control team on the bridge and the relevant staff.

The real-time voice command recognition software was designed using primarily a Graphic User Interface (GUI) in the MATLAB R2011b program. Based on a main menu, the system consists of a Training Interface that the cruise control staff on board will use to provide a command reference-bank system, and a Test Interface that will allow the system to be used in actual operations and for assessments. In this design, Linear Predictive Coding (LPC) and Mel Frequency Cepstral Coefficients (MFCC) parameters are used separately for the extraction of the voice command features; and Dynamic Time Warping (DTW) parameter is used for attribute matching. Finally, a comparison circuit is designed to evaluate the command compatibility between the cruise control person and the steersman.

The main consideration taken into account during the design process is the availability of the system in the ship environment, it is open to development and the cost is low.

PROPOSED SYSTEM

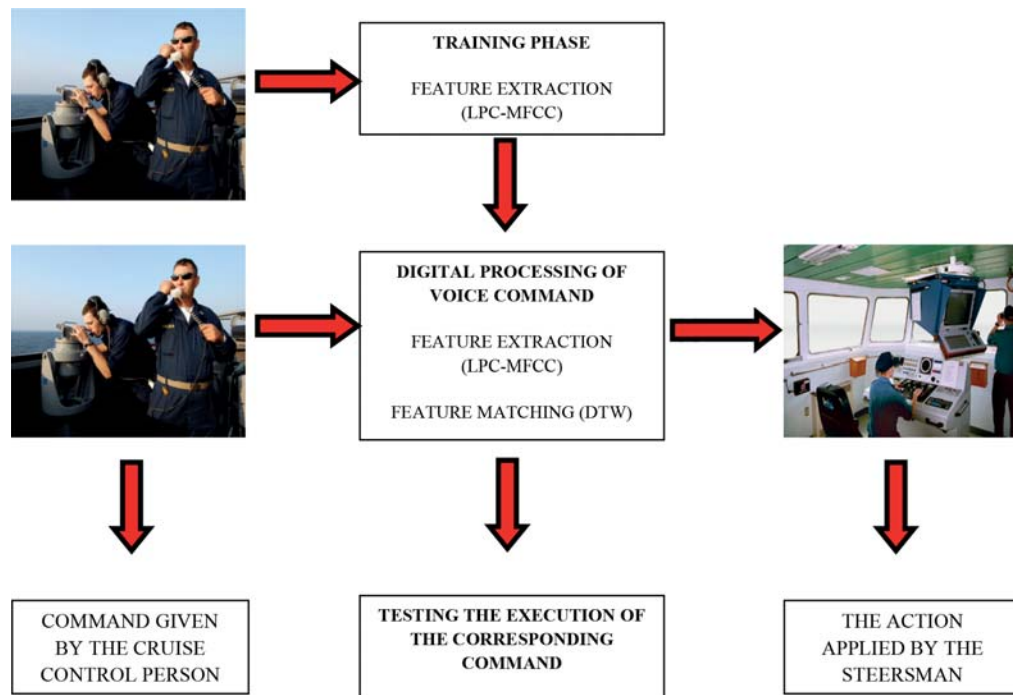


Fig. 1. Voice Command Recognition System Architecture

As described in Fig. 1, the voice command recognition algorithms include two discrete phases. While the first phase is the training phase, the second phase is the operation or the assessment phase.

Training phase; establishing reference-bank by providing voice command samples to the system by the control staff on board.

Testing phase; testing the compatibility of the voice command with the reference-bank and the system input, and testing the execution of the corresponding command.

VOICE COMMAND RECOGNITION SYSTEM ARCHITECTURE

The voice command recognition process starts with voice recording. The process continues with the detection of the expression, the processing of the voice, comparison and matching. Each word in the incoming audio signal is isolated and then analysed, to identify the type of excitation and resonate frequencies. These parameters are then compared with previous examples of spoken words to identify the closest match.

A main program-based sub-program was designed using the Graphic User Interface (GUI) that is available in MATLAB R2011b.

MAIN MENU

The designed interface is shown in Fig. 2. The main menu is used by the administrator to access the interface programs needed for training and testing the system.



Fig. 2. Main Menu

SYSTEM TRAINING

Fig. 3 illustrates the designed system training interface program. The interface is used in order to establish a reference-bank by introducing examples of voice commands to the system provided by the control staff on board. In shipping; starboard, port, astern, forward, mean “right”, “left”, “back”, “forward” respectively. The commands to be registered in the system (starboard-sancak-, port-iskele-, astern-tornistan- and forward-ileri) are introduced to the command bank of the system by using the ‘save’ keys. These commands have been trained in Turkish in the system.

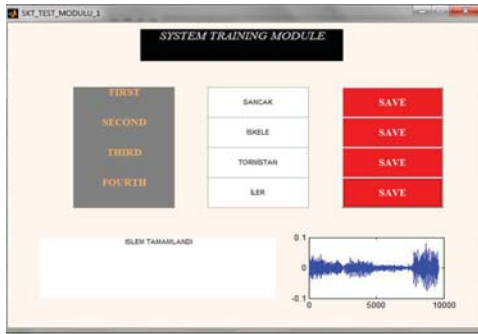


Fig. 3. System Training Interface Program

SYSTEM TESTING

Fig. 4 illustrates the designed system testing interface program. The voice commands introduced to the system are stored by using the 'save' key and by comparing the commands already registered in the reference-bank by the system; the appropriate answer is then displayed on the interface.

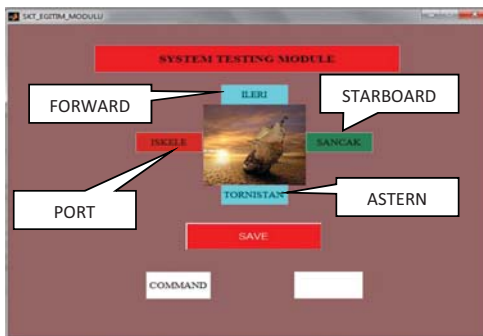


Fig. 4. System Testing Interface Program

VOICE RECOGNITION ALGORITHMS

During the training and testing phases of the system, the voice command is first recorded by means of a microphone. The analog signal is first saved, and then converted to digital ([7], [8]). Upon evaluating the stress level of the navigation environment, and by taking into account that the command of the cruise control person would be communicated after pressing the signal button on the console, the recording time in the developed system - covering this gap- has been identified as 3 seconds. Fig. 6 illustrates the sampled command signals with 16 kHz.

According to the Nyquist theorem, the sampling frequency must be at least twice the signal bandwidth. If there is no minimum sampling frequency, aliasing occurs. Having registered and sampled the signal, the digital signal should be filtered to eliminate the noise. [9]. In this study, sampling frequency is taken as 16 kHz. 16 kHz is the optimal sampling frequency. When the ambient noise of the microphone is added to the noise of the microphone, and the noise of the

computer, a considerable amount of noise will emerge. This noise can be separated from the speech signal. Fig. 7 illustrates filtered signals. A high-pass filter was used in this study.

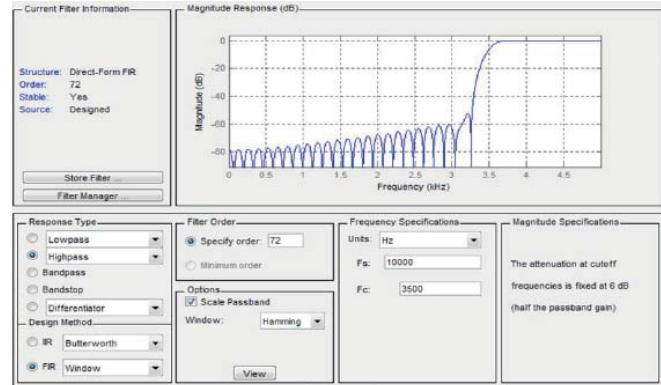


Fig. 5. High Pass Filter Design Tool

Fig. 5 illustrates the high pass filter design tool. Having recorded and sampled the signal, it should be filtered to remove noise. The digital signal must be filtered from the noise of the environment, the microphone and the computer. For this purpose, a high pass filter has been used. As seen in the Figure 5, the cut-off frequency (F_c) is selected as 3500 Hz.

There are unwanted and unused gaps at the start and finish on the existing filtered signal. These gaps are not used, and cause unnecessary processes on the computer [10]. These gaps should be deleted by determining the beginning and the end of the gaps, and removed by the system. After this process, we are left with only raw speech signals. Appropriate signal processing methods can now be used on these signals (Fig. 8).

Parallel to the information mentioned above, the following techniques have been used for extracting the attributes of the voice command given by the cruise control person and saving them in reference-bank storage, and for comparing the voice command exercised:

- Linear Prediction Coding (LPC) and Mel Frequency Cepstral Coefficients (MFCC) for Feature Extraction,
- Dynamic Time Warping (DTW) for Feature Matching.

FEATURE EXTRACTION

Feature extraction is also called signal processing front-end. The aim of feature extraction is to simplify recognition by summarizing the amount of speech data, and obtaining the acoustic properties that define speech individuality [11].

FRAMING

Framing is the process of segmenting the speech samples obtained from analog to digital conversion (ADC), into small frames with time-length within the range of 20 to 40 ms. Adjacent frames are divided by M ($M < N$). The first frame consists of N samples. The second frame begins M samples after the first frame, and overlaps it by $N - M$ samples. Each frame overlaps with two other subsequent frames. This

operation is performed throughout the entire audio signal. N is the number of samples in each frame. Typical values used are $M = 100$ and $N = 256$ ([4], [12]).

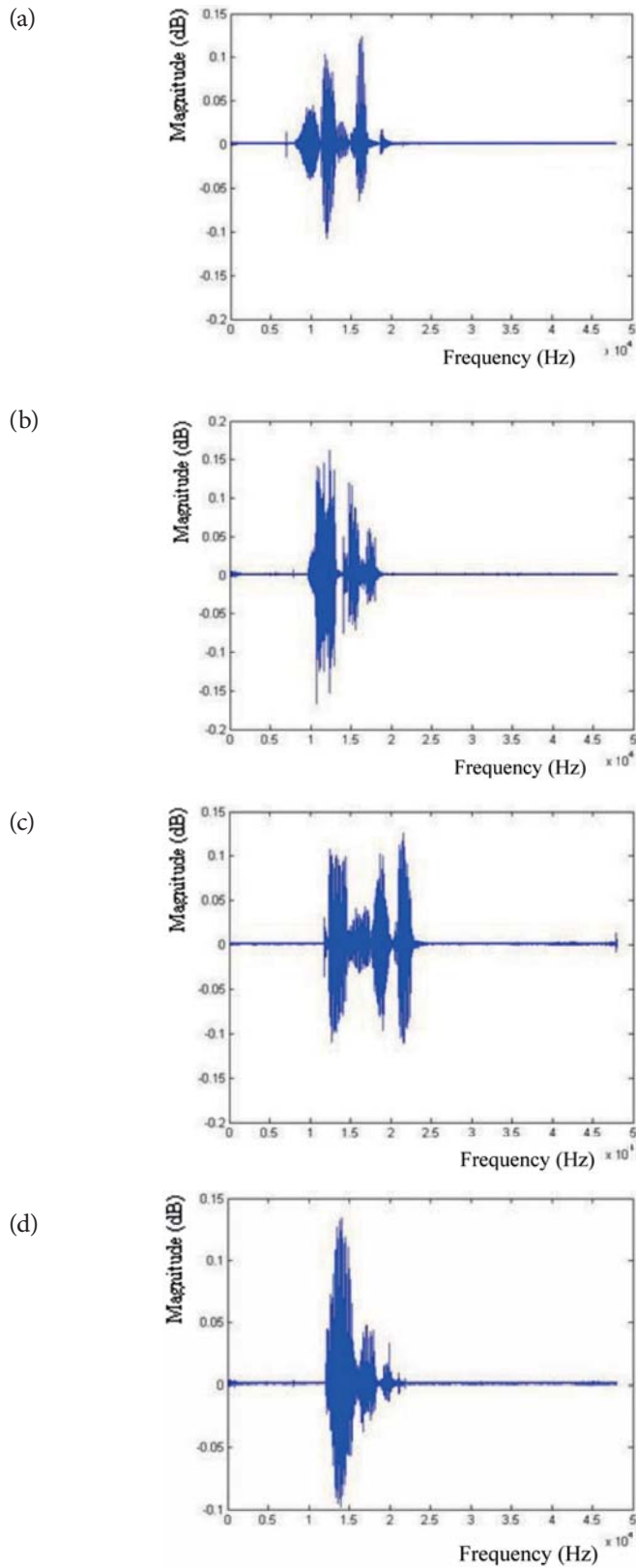


Fig. 6. Sampled command signals
(a. Starboard; b. Port; c. Astern; d. Forward)

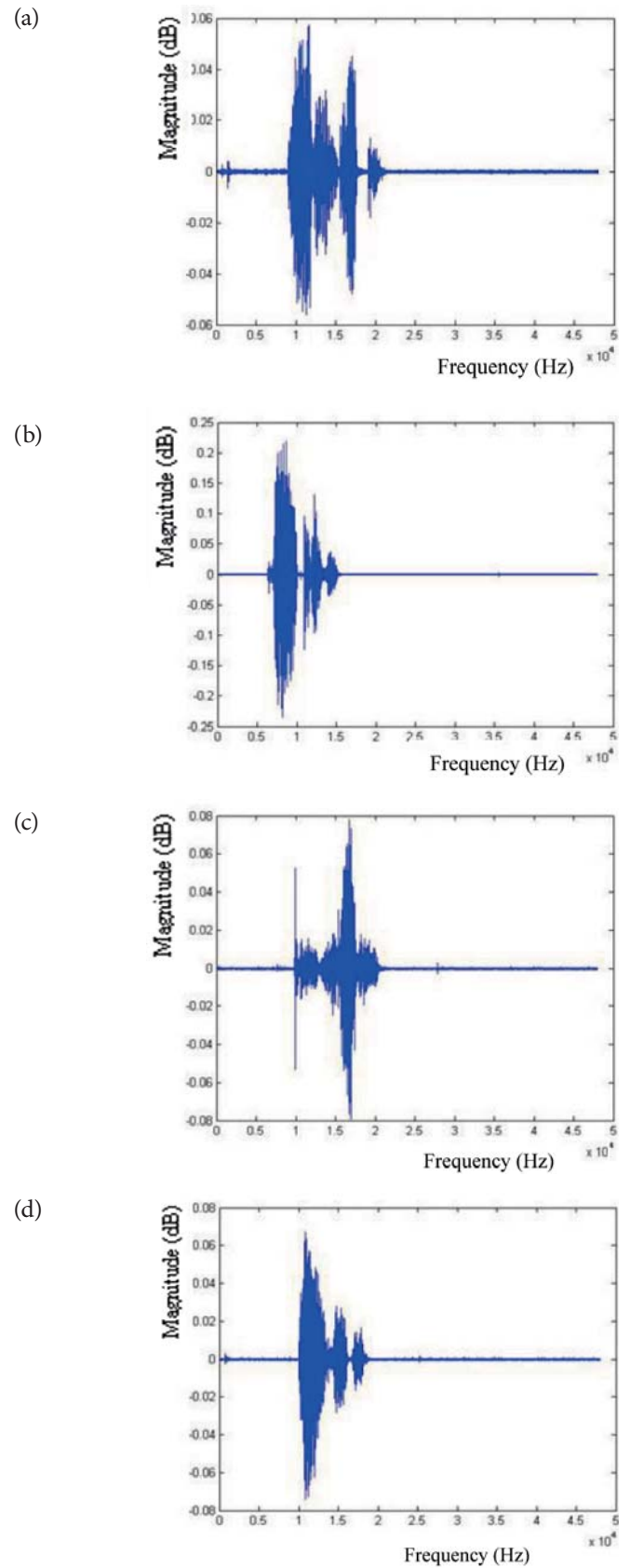


Fig. 7. Filtered signals
(a. Starboard; b. Port; c. Astern; d. Forward).

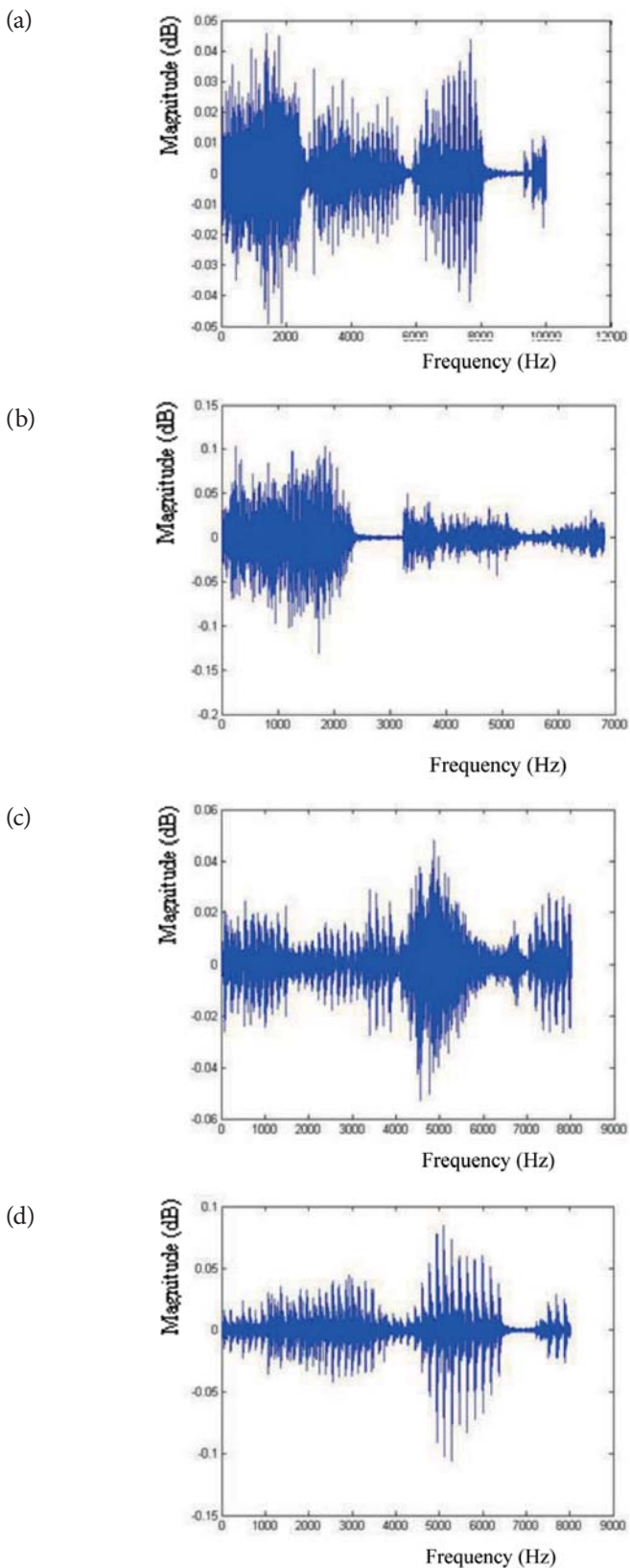
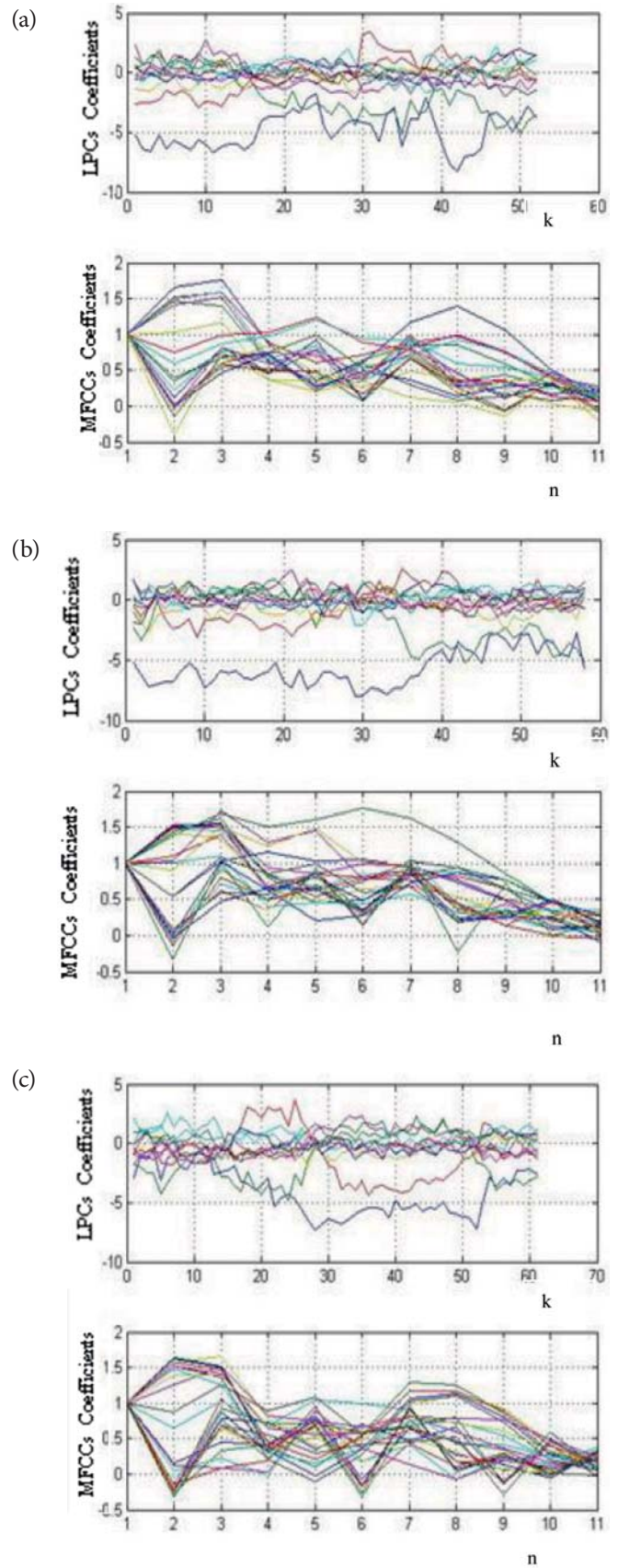


Fig. 8. Processed speech signals
(a. Starboard; b. Port; c. Astern; d. Forward).



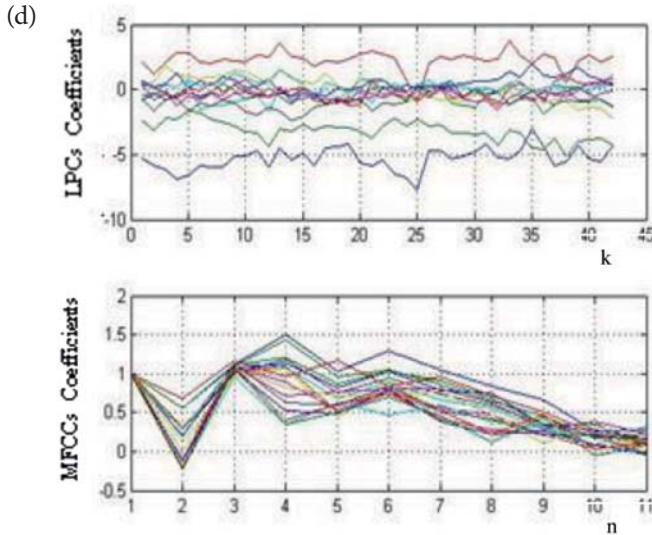


Fig. 9. LPC (below)-MFCC (above) coefficients (a. Starboard; b. Port; c. Astern; d. Forward).

Figure 9 shows the final vectorial representations of the commands' MFCC (Mel Frequency Cepstral Coefficient) and LPC (Linear Predictive Coding) applications. In this figure, the LPC coefficients are equations (3-8), and the MFCC coefficients are obtained from equations (9, 10). Figure 10 illustrates LPC coefficients. The MFCC coefficients are shown above in Figure 9. Calculation of these coefficients is given in LPC and MFCC titles.

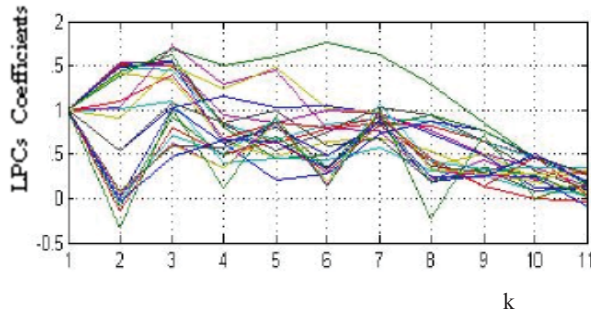


Fig. 10. LPC coefficients

HAMMING WINDOWING

The Hamming windowing is to minimize the signal's head, and the discontinuous parts at the end. The goal of Hamming windowing is to chop by using the window [4]. The Hamming window equation is given as:

If the window is defined as $w(n)$, $0 \leq n \leq N-1$ where
 N – number of samples in each frame
 $y_1(n)$ – output signal
 $x_1(n)$ – input signal
 $w(n)$ – window

$$y_1(n) = x_1(n)w(n), \quad 0 \leq n \leq N - 1 \quad (1)$$

Hamming window used in the system;

$$w(n) = 0.54 - 0.46 \cos\left(\frac{2\pi n}{N-1}\right), \quad 0 \leq n \leq N - 1 \quad (2)$$

LINEAR PREDICTIVE CODING (LPC)

The LPC method is one of the most important of the sound analysis techniques. In this method, the vowels are modelled as periodic pulses, the consonants are modelled as random pulses [13]. The obtained LPC analysis results are linear predictive model coefficients. This model is expressed by the transfer function as in Equation 3. p is the level of the LPC encoder. [14].

$$H[z] = \frac{X[z]}{Y[z]} = \frac{1}{1 - \sum_{i=1}^p a_i z^{-i}} \quad (3)$$

Equation (4) is obtained when inverse-z transform is applied to Equation (3).

$$x[n] = \sum_{i=1}^p a_i x[n - i] + e[n] \quad (4)$$

The LPC focuses on the principle that it can be approximated from a previous series of samples.

$$\dot{x}[n] = \sum_{i=1}^p a_i x[n - i] \quad (5)$$

The minimization of the sum of squares of the error is calculated to find the variation [15].

$$e[n] = x[n] - \dot{x}[n] = x[n] - \sum_{i=1}^p a_i x[n - i] \quad (6)$$

The LPC analysis results are linear prediction model coefficients [16] (Fig. 8),

$$A_m = \{a_1, \dots, a_p\} \quad (7)$$

$$a_k = a_k^{(p)}, \quad k = 1, \dots, p, \quad (8)$$

MEL FREQUENCY CEPSTRAL COEFFICIENT (MFCC)

MFCC is based on known variance of the human ear with critical frequency bandwidth [17]. MFCC has two types of filter, which are spaced linearly at low frequency below 1000 Hz., and have logarithmic spacing above 1000 Hz. A subjective pitch is present on Mel Frequency Scale to capture important characteristics of phonetic in the voice [11].

Equation 9 is a Mel equivalent of a frequency of type Hz.

$$mel(f) = 2595 * \log\left(1 + \frac{f}{700}\right) \quad (9)$$

Mel power spectrum coefficients (Fig. 9) that are the result of the last step are we can calculate the MFCC coefficients, as:

$$c_n = \sum_{k=1}^K (\log S_k) \cos \left[n \left(k - \frac{1}{2} \right) \frac{\pi}{K} \right] \quad (10)$$

$$n = 1, 2, 3, \dots, K$$

Equation 10 is called an acoustic vector.

DYNAMIC TIME WARPING (DTW)

Dynamic Time Warping (DTW) and Hidden Markov Model (HMM) use the same environmental conditions in isolated voice command recognition. HMM is a very complex algorithm. DTW is the best method for finding the shortest distance between the attribute matrix and the unknown matrix.

DTW parameter is based on dynamic programming techniques. This parameter is for measuring similarity between two time series, which may vary in time or speed ([6], [18]). This technique is also used for finding the optimal alignment between two time series, if one time series may be “warped” non-linearly by stretching or shrinking it along its time axis.

Suppose we have two time series Q and C, of length n and m respectively, where:

$$Q = q_1, q_2, \dots, q_i, \dots, q_n \quad (11)$$

$$C = c_1, c_2, \dots, c_i, \dots, c_m \quad (12)$$

To align two sequences using DTW, an n-by-m matrix where the (ith, jth) element of the matrix contains the distance $d(i, j)$ between the two points and is constructed. Then, the absolute distance between the values of two sequences is calculated using the Euclidean distance computation:

$$d(q_i, c_j) = (q_i - c_j)^2 \quad (13)$$

Each matrix element (i, j) corresponds to the alignment between the points and . Then, accumulated distance [3] is measured by:

$$D(i, j) = \min \left\{ \begin{array}{l} D(i, j - 1) \\ D(i - 1, j) \\ D(i - 1, j - 1) \end{array} \right\} + d(i, j) \quad (14)$$

COMPARISON AND DECISION-MAKING CIRCUIT DESIGN

Whereupon the command given by the cruise control person is perceived in the created system, the compatibility of the action of the steersman is compared with the command, and an error signal is aimed to be developed in case of a divergence. In this context:

- Ship Direction Commands (Starboard, Port) and Ship Direction Commands (Forward, Astern) are assessed discretely,
- The formation of an error signal has been secured for each discrete set, in case of the following:
 - A discrepancy occurs between the command given by the cruise control person and the action applied by the steersman,
 - A command is given by the cruise control person, and it is failed to be executed by the steersman.

The matrix containing the above-mentioned possibilities is presented in Tab. 1.

Tab. 1. Command Probability Matrix

		CRUISE CONTROL PERSON			
		S	P	A	F
STEERSMAN	S	0	1	-	-
	P	1	0	-	-
	A	-	-	0	1
	F	-	-	1	0

S: Starboard, P: Port, A: Astern, F: Forward
 '1' = Alarm, '0' = Proper, '-' : Discrete

By taking into account the Command Probability Matrix specified in Table-1, a circuit was designed on an Electronics Workbench using the MATLAB R2011b program for the simulation between the GUI and the ship’s rudder/throttle controlled by the steersman; a schematic diagram of the designed circuit is shown in Fig. 11. It is ensured that the ships are routed to the desired routes with Figure 11.

The integrated circuit (IC) 7486 consists of four 2-input Exclusive-OR (Ex-OR) gates in the comparison circuit. At this gate, output is 0 when the inputs are the same, and output 1 when the inputs are different [8]. The truth table of the Ex-OR gate is shown in Table 2.

Tab. 2. The truth table of the Ex-Or gate

INPUTS		OUTPUTS
A	B	F
0	0	0
0	1	1
1	0	1
1	1	0

The integrated circuit (IC) 7432 consists of four 2-input OR gates in the comparison circuit. The output of this gate is the sum of the inputs. The truth table of the OR gate is shown in Table 3.

Tab. 3. The truth table of the OR gate

INPUTS		OUTPUTS
A	B	F
0	0	0
0	1	1
1	0	1
1	1	1

When Fig. 11 is analyzed, the A-H switches are fed from a single source, and execute different possibilities simultaneously. Signals coming from the steersman, the steering that the commands are executed from and the throttle will all be connected to the same gauge.

The comparison circuit compares the signal from the DTW, and the signals from the steersman. Only two of these are available in hardware.

In order for the ship to navigate on a fixed route and speed, a design will be created so that the corrections applied by the steersman will not be accepted as inputs. The acceptable ranges for these corrections are as follows;

- Route revisions $\pm 3\%$;
- Speed revisions ± 0.5 nautical miles

Another issue pointed out in the schematic diagram is that the S key is placed right beyond the voltage source. This is an indication that, in actual cases, the system will be activated if, and only, with the command given by the person who has the control of the vessel. In this way, especially in rough waters, even if a command is not given by the cruise control person, the system will not give any error signals when necessary adjustments are made by the steersman in order to navigate steadily route and speed.

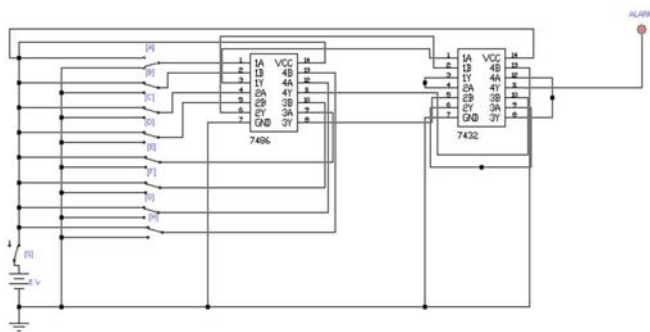


Fig. 11. Comparison Circuit Schematic Diagram

All error signals are connected to the ALARM indicator led with the 7432 circuit. A single error signal will be enough to activate the error signals.

RESULTS

In this study, by using voice recognition algorithms, a system has been developed that will test the compatibility of command and its implementation between the cruise control

person and the steersman. The voice command recognition system has been developed by using the MATLAB R2011b program with LPC/MFCC and DTW parameters on a GUI interface, and the comparison circuit has been designed in a manner to regard the actual operations which is displayed in schematics by using the Electronics Workbench program.

By means of the “Test Phase” feature added to the system software, the voice command recognition software for LPC and MFCC parameters has been tested separately, and the results are presented in Table 4 and Table 5. A total of 100 tests have been carried out with the participation of volunteers of both sexes; (The first and fourth users are male, others are female). The tests have been conducted in a house located on a street with busy-traffic, in front of an open window.

Tab. 4. MFCC-DTW Combinations Test Results

	SUCCESS RATE			
	USER1	USER2	USER3	USER4
STARBOARD	92%	96%	100%	92%
PORT	80%	84%	88%	84%
ASTERN	96%	96%	92%	96%
FORWARD	84%	80%	84%	88%
STOP	92%	92%	96%	100%
AVERAGE	88.8%	89.6%	92%	92%

Tab. 5. LPC-DTW Combinations Test Results

	SUCCESS RATE			
	USER1	USER2	USER3	USER4
STARBOARD	80%	84%	90%	88%
PORT	84%	76%	88%	84%
ASTERN	92%	88%	92%	84%
FORWARD	88%	84%	92%	80%
STOP	80%	84%	80%	92%
AVERAGE	84.8%	83.2%	88.4%	85.6%

When Tables 4 and 5 are examined, the best results are obtained from the combination of MFCC-DTW.

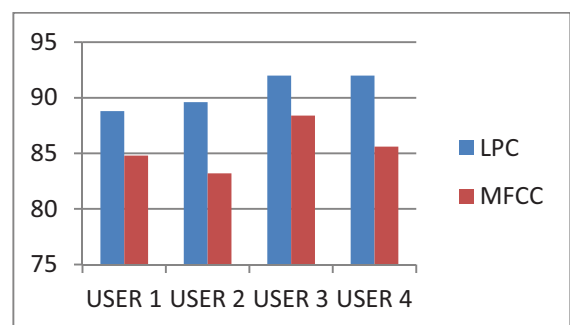


Fig. 12. LPC-MFCC Parameters Test Results

According to Table 4, average of the users is 90.6%, in the MFCC-DTW combination. According to Table 5, average of the users is 85.5%, in the LPC-DTW combination. The x-axis shows the users in Fig. 12. The y-axis shows percentages of the test result. When analyzing the phases of the test results, it has been observed that the MFCC-DTW combination results are more successful in real-time voice recognition.

CONCLUSIONS

There is an intense ambient noise; such as meteorological conditions, ship noises, the overall bridge noises formed during navigation. Therefore, the communication console (microphone) to be used by the cruise control person should be chosen wisely.

In direct proportion to the magnitude of manoeuvres, the psychological state of the cruise control person, the data provided to the command reference-bank during the software training phase might show discrepancies [19]. Except in the case of fear, stress-related anger, mood changes such as sadness, have been observed to increase the success of the voice recognition system [20].

Due to the extreme variability observed in the sound recording in time, it is not always possible to find the closest template and determine the correct words. Therefore the suggested approach would be the calculation of the average performance of the templates obtained after records performed at different times and using this template for increasing recognition.

The system is particularly designed to ensure the safety of warships during close manoeuvres, and the courses of commercial vessels in narrow waters. The sound command recognition software of the system can be developed, especially for open sea navigation, to contain voice command property as in autopilot applications. In this context, the system can be improved together with the values given along with direction and path commands (starboard 19, speed 8 and etc.).

The way the system is developed, and with the command base to be created in English, the system can take a role as a safety mechanism for the English communication problems that arise between tugboat captains and the steersman, especially among vessels navigating in international ports.

BIBLIOGRAPHY

- Lazarowska A.: *Decision support system for collision avoidance at sea*. Polish Maritime Research, 2012 (Special Issue), pp.19-24.
- Lazarowska A.: *Swarm intelligence approach to safe ship control*. Polish Maritime Research, 2015(4), pp. 33-40.
- Zhizeng L., Jinghing Z.: *Speech recognition and its application in voiced-based robot control system*. International Conference on Intelligent Mechatronics and Automation, 0-7803-8748-1, 2004.
- Bala A., Kumar A., Birla N.: *Voice Command Recognition System Based on MFCC and DTW*. International Journal of Engineering Science and Technology Vol. 2 (12), 7335-7342, 2010.
- Vashisht D., Sharma S., Dogra L.: *Design of MFCC and DTW for Robust Speaker Recognition*. International Journal of Electrical&Electronics Engineering Vol 2 (3), 1694-2426, 2015.
- Ferrando F., Nouveau G., Philip B., Pradeilles P., Soulenq V., Van-Staen G., Courmontagne P.: *A Voice Recognition System for a Submarine Piloting*. 1-4244-2523-5/09- IEEE, 2009.
- Smith S.W.: *The Scientist's and Engineer's Guide to Digital Signal Processing*. California Technical Publishing, ISBN 0-96-601764-1, 1999.
- Yagimli M., Akar F.: *Digital Electronics*. Beta, ISBN: 9786053777526, Istanbul 2014.
- Proakis J. & Manolakis D.: *Digital Signal Processing Principles, Algorithms and Applications (3rd edition)*. New Jersey : Prentice-Hall Inc., 1996.
- Karakas M.: *Computer Based System Control Using Voice Input*. M.Sc. thesis, Dokuz Eylul University, 2010.
- Demirci M.: *Computer Aided Voice Recognition System Design*, M.Sc. thesis, Istanbul University, 2005.
- Lindasalwa M, Mumtaj B, Elamvazuthi I.: *Voice Recognition Algorithms using Mel Frequency Cepstral Coefficient (MFCC) and Dynamic Time Warping (DTW) Techniques*. Journal of Computing, Volume 2, Issue 3, pp.138-143, ISSN 2151-9617, March 2010.
- Rabiner L. R., Shafer R. W.: *Digital Processing of Speech Signals*. Prentice Hall Inc., September 1978.
- Huang X., Acero A, Hon H.W.: *Spoken Language Processing: A Guide to Theory, Algorithm and System Development (1st Ed.)*. Prentice Hall PTR, ISBN 0-13-022616-5, 2001.
- Lipeika A., Lipeikien J., Telksnys L.: *Development of Isolated Word Speech Recognition System*. INFORMATICA, Vol. 13, No. 1, 37-46 Institute of Mathematics and Informatics, 2002.
- Juang B.H., Wang D.Y., Gray A.H.: *Distortion performance of vector quantization for LPC voice coding*. IEEE Trans.

on Acoustic Speech and Signal Processing, ASSP-30 (2), 294–304.

17. Jiang Z., Huang H., Yang S., Lu S., Hao Z.: *Acoustic Feature Comparison of MFCC and CZT-based Cepstrum for Speech Recognition*. IEEE 2009 Fifth International Conference on Natural Computation, 978-0-7695-3736-8/09, 2009.
18. Price J., Eydgahi A.: *Design of Matlab®-Based Automatic Speaker Recognition Systems*. 9th International Conference on Engineering Education-T4J-1, July 2006.
19. Phoophuangpairroj R.: *Using Multiple HMM Recognizers and the Maximum Accuracy Method to Improve Voice-Controlled Robots*. 2011 International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS) December 7-9, 2011.
20. Petrushin V.A.: *Emotion in Speech Recognition and Application to Call Centres*. Andersen Consulting, 3773 Willow Rd., 2009.

CONTACT WITH THE AUTHOR

Mustafa Yagimli

e-mail: mustafa.yagimli@okan.edu.tr

Okan University
Department of Property Protection and Security
34722, Kadikoy, Istanbul
tel.: +905054206650

TURKEY