Diyala Journal of Engineering Sciences

ISSN 1999-8716
Printed in Iraq

Vol. 10, No. 3, pp. 96-106, September 2017

DESIGN A DEPENDENT VOICE COMMANDS SELECTIVE SYSTEM

Ahmed Salah Hameed
Assistant Lecturer, Computer Engineering Department, College of Engineering
University of Diyala, Iraq
ah_first86@yahoo.com

(Received: 25/9/2016; Accepted: 28/11/2016)

ABSTRACT: - Voice commands recognition is a way of understanding human speech and converting it to a communication input of computer. Based on the data used to feed the voice recognition system, systems could be classified as dependent or independent systems. In this paper, a dependent voice commands selective system is presented. The system uses the spectral subtraction algorithm for noise cancellation. A new algorithm is named Select Matching Command Algorithm SMCA is used for matching voice command identification. Six different users (3 male and 3 female) are used to test the system. Each user tests the system two times, one by using twenty four voice commands samples previously selected and other time by using another twenty four voice commands out of the twenty four preselected samples. The proposed system is able to correctly identify the matching voice commands 90.97% of the time when testing the system with the predefined voice commands. Testing the system with voice commands out of the predefined commands shows 92.36 % of no matching identification. According to the results noted from the six users, overall percent of selective accuracy of the suggested system is 91.67 %.

Keywords*:* *Voice Commands Recognition, Cross Correlation, and Spectral Subtraction.*

1- INTRODUCTION

Voice commands recognition is a way of understanding human speech and converting it to a communication input of computer devices. The speech goes in different phases of processing to be understandable by computer devices and these phases are: speech analysis, extracting feature of speech, features and speech modelling, and result testing [1]-[2]-[3]. When a voice commands recognition system is build, a training data is used to feed the system with a specific package of reference voice commands and this is the case in which the system will be known as a dependent system and it is accessible by one user. When the system is built with no specific data to be fed to the system, it will be known as an independent system and it is accessible by different users [4].

Voice commands selective system is one of the forms of “biometric recognition systems” that are widely used and developed nowadays [5]. Voice commands are used as control input or security identification for different modern devices such as personal computers and smartphones. Basically; the purpose behind voice commands selective systems is either to identify the speaker or to identify the spoken commands regardless of the speaker. The main job of this research is to design a dependent voice commands selective system that able to recognize the spoken voice commands regardless of the speakers of it.

2- VOICE AND SIGNAL PROCESSING TECHNIQUES

Human voice represents a continuous signal which is a non-stationary signal. Dealing with the voice signal in small intervals of time can make a stationary signal [6]. To deal with the whole band of frequencies of humans sound and to reduce the aliasing effects, the recorded voice signals should be sampled with sampling rate above 10000 Hz [7]. Many of signal processing Techniques used in processing of voice signals different techniques are used to process the voice signal to implement voice recognition systems. The paper is going to discuss the main techniques used in building the voice commands selective system proposed in this research.

A. Fast Fourier Transform

Fast Fourier Transform, which is abbreviated as FFT, is usually used in voice recognition systems as a technique for frequency spectrum analysis. FFT is best choice to increase the calculation of the Discrete Fourier Transform (DFT) and it is basically used to convert signal from time domain to frequency domain. Equation for FFT is [6]:

$$x(k) = \sum_{i=1}^{N} x(i).e^{\frac{-2\pi i}{N}(i-1)(k-1)} \ldots\ldots\ldots\ldots\ldots\ldots (\mathbf{1})$$

The range of k in the equation above is *0* to *N-1*. The function $x(i)$ represents the samples of signal based on i index. According to the FFT equation, the resulted function $x(k)$ is represented as array of calculated values related to each index in the time signal.

B. The Spectral subtraction algorithm

Spectral subtraction algorithm has the wider use in noise cancellation of voice recognition systems [8]. Based on the name of the technique, it is work in frequency domain and simply calculated after the noise is predefined or calculated to be subtracted from the noisy signal. The general description of the spectral subtraction method is shown in figure (1). Some voice processing systems are using an estimation procedure to find the noise related to the main voice signal to be used in spectral subtraction calculation. A simple and direct way to find the noise of the voice signal could be by recording a specific interval of background noise which the noise subtracted from the main voice signal. The voice signal is simply defined as:

$$\mathrm{v(n) = cv(n) + no(n)} \ldots\ldots\ldots\ldots\ldots\ldots (\mathbf{2})$$

Where $\mathrm{v(n)}$ is the noisy voice signal, $\mathrm{cv(n)}$ is the clean voice signal, and $\mathrm{no(n)}$ is the noise related to the main voice signal.

C. Cross correlations technique

Cross correlation technique is widely used in the applications build based on digital signal processing. In the implementation of voice recognition systems, the cross correlation technique is used in calculation of shifts relation of two signals, and in finding the signal that determine if one signal is match to another one or not. Based on the technique, applying a cross correlation with two identical signals generate a symmetric correlated signal. The following function is the general function used to calculate the cross correlation of two different signals [10]-[11]:

$$c[x] = \mathrm{(r * t)[x]} = \sum_{i=-\infty}^{\infty} r[i].t[i+x] \ldots\ldots\ldots\ldots\ldots\ldots (\mathbf{2})$$

In the above equation, c[x] refers to the cross correlation of the two signals r[x] and t[x], when r[x] is any signal belongs to the pool of reference signals, and t[x] is the target voice command signal. To calculate cross correlation efficiently, the number of lag represented by x in the equation is equal or less than the number of samples of r[x] and t[x] signals.

3- THE PROPOSED VOICE COMMANDS SELECTIVE SYSTEM

In this paper, a dependent voice commands selective system is proposed. One user is used to feed the system with the training data which is stored in what it is known a training data pool. Even though that one user feeds the system with training data, multiusers can test and use the system. For the proposed system, six users are used in testing the system. The training data pool of the system will be fed with a group of famous voice commands used in robotic control systems. The proposed system has three phases of processing which are: generating the reference signals pool, generating the correlated signals pool, and detecting the matching voice command signal.

Phase 1: Generating the reference signals pool

In this phase the training data of the system is recorded and processed to generate the Reference Signals pool (RS pool) of the system, as shown in figure 2. Three seconds of each training data samples with the preferred sampling frequency is recorded. For the system designed in this research the sampling frequency 12000 Hz is selected to record all the signals processed by the system. For further spectrum analysis, the samples will be converted to frequency domain using FFT method. One of the important things done in this phase is the noise cancellation of the training data samples. The spectral subtraction algorithm is used for noise cancellation in which the spectrum of background noise as average magnitude is subtracted from the original sample. To do the process of spectral subtraction, another three seconds sample of background noise is recorded and converted to frequency domain with spectrum analysis. After calculating the Spectral subtraction for each sample, the generated samples are stored in a RS pool.

Phase 2: Generating the correlated signals pool

In this phase, the target voice command signal, which is the voice command used to test the system, is recorded and processed in the same procedure used in recording and processing the training data samples. The main job done in this phase is generating the correlated signals pool (CS pool). By using cross correlation method, the processed target signal will be correlated with each signal in the RS pool generated in phase 1. This phase is end up with generating the CS pool, as shown in figure 3.

Phase 3: Detecting the matching voice command signal

This is the phase in which the matching signal to the target voice command is selected among the signals available in the CS pool. The proposed algorithm used to do the job of matching is named a Select Matching Command Algorithm (SMCA). SMCA is used to make the decision of matching as it shown in figure 4. The first two signals in the CS pool will be compared to select the more matching signal to the target signal. The selected signal between the first two signals in the CS pool will keep compared with each signal in the pool separately till it end with selecting the best matching signal. Based on the procedure of the proposed algorithm, the decision of selecting matching signal between two compared signals can take one of the following directions:

A) By comparing any two signals from CS pool, the signal that has the maximum value of cross correlation is the best matching signal to the target voice command.
B) According to cross correlation technique, two identical signals generate a symmetric correlated signal; the signal that is closed to be symmetric is the best matching signal to the target voice command.
C) When the system failed in determining the matching signal in A and B directions, it will choose one of the two signals to be compared with the next signals. The system output will be “no matching command in the pool”, when we reach the last comparison in the CS pool and no decision of matching done by A or B.

AS clarified above in phase 3, SMCA is used for matching voice command identification and it is build based on cross correlation algorithm. Figure 3 shows the flowchart of the suggested algorithm. Based on the flowchart, the algorithm works as following:

Step 1: Setting the main factors of the algorithm which are:

1) X represents the number of signals in CS pool which is previously selected to test the proposed system.

2) SSA represents Selected Signal Address which is used to hold the address of matching voice command signal. SSA is set to zero at the beginning of any test.
3) TSSA represents Temporary Selected Signal Address is used to hold the address of voice command when no matching decision is made by the first two directions of comparison. TSSA is also set to zero at the beginning of any test.

Step 2: The first two signals in the CS pool will be compared to select the more matching signal to the target signal. The decision of matching could be resulted by one of the three directions of comparison listed in phase 3 above. The first two directions of comparison A and B are only the way leads to update SSA factor. Any decision made based on direction C is leads to update TSSA factor without changing the value of SSA. The cause upon which we only update TSSA as a result of direction C is that the decision of direction C is only made to keep the procedure of comparison and it is not a right matching decision. As a result, each test of comparison updates one of the two factors SSA or TSSA.

Step 3: SSA or TSSA value determines which voice command signal is compared with the next signal in CS pool. Before starting the next loop of comparison, X is tested and as well as we have X greater than zero the comparison is continue with the next signals in CS pool to select the best matching voice command signal. When X is equal to zero, in other words no more signals to be compared in the CS pool, the comparison loop is done.

Step 4: SSA determines which signal is the matching voice command signal. If we have SSA equal to zero, it means no decision made by directions A and B and target voice command signal has no matching signal in the system. Any other value of SSA represents the address of matching voice command signal. The signal that holds the SSA value is the matching voice command of the target voice command.

4- TESTING THE PROPOSED SYSTEM

The system proposed in this research is a dependent system that its work depends on a specific training data. Before testing the system, twenty four voice commands samples were selected to test the system with computer simulation in MATLAB environment. The samples of voice commands used for simulation of the system are shown in table (1). To do any test to the system, the twenty four samples as well as the background noise should be recorded to feed the RS pool of the system. By recording and processing the samples of voice commands, the system get ready to be tested with desire input voice command. Six different (3 male and 3 female) users are used to test the system. Each user will test the system two times one by using the 24 voice commands of table (1) and other time by using 24 voice commands out of table 1. All the voice commands signals used in testing the system go in the procedure of the three phases clarified in section 3 of the paper. SMCA algorithm decides the matching or no matching decision of the proposed system. The percent of voice commands selective accuracy for each user is recorded as shown in table 2 and figure 5. The proposed system is able to correctly identify the voice commands 90.97% of the time when using the voice commands of table (1) in system testing while it is 92.36 % when using a voice commands out of table (1). According to the results noted from the six users, overall percent of selective accuracy of the system is 91.67 %.

5- CONCLUSIONS AND FUTURE WORK

In this research a design of a dependent voice commands selective system is presented. According to the results obtained by the proposed system, the new designed system has a significant value of matching accuracy. The matching accuracy of the system is said to be significant and new since the system is able to recognize the spoken voice commands regardless of the speakers of it. In other words, the system can be tested by any speaker and not only by the speaker of the training voice commands with total matching accuracy equal to 91.67 %.

The new design accomplish the high percent of accuracy value through several design factors i.e., by using a spectral subtraction algorithm as a noise canceller, and by using cross correlation technique with the new algorithm SMCA as matching signal detector. For future work, the design can be enhanced through the building of the SMCA with an Artificial Neural Network (ANN). Building SMCA based on ANN could give the design the best speed and the best value of matching accuracy.

REFERENCES

1) Santosh K.Gaikwad, Bharti W.Gawali and Pravin Yannawar, "A Review on Speech Recognition Technique", International Journal of Computer Applications (0975 – 8887) Volume 10– No.3, November 2010.
2) Nidhi Desai, Kinnal Dhameliya and Vijayendra Desai, “Recognizing Voice Commands For Robot Using MFCC and DTW”, International Journal of Advanced Research in Computer and Communication Engineering, Volume 3, Issue 5, May, 2014.
3) F. G. Barbosa and W. L. S. Silva, “Automatic Voice Recognition System Based On Multiple Support Vector Machines and Mel-Frequency Cepstral Coefficients”, Natural Computation (ICNC), 2015 11th International Conference on, Zhangjiajie, 2015, pp. 665-670. doi: 10.1109/ICNC.2015.7378069
4) Preeti Saini and Parneet Kaur, “Automatic Speech Recognition: A Review”, International Journal of Engineering Trends and Technology (2231-5381) Volume 4 Issue2- 2013.
5) Wahyu Kusuma and Prince Brave, “Simulation Voice Recognition System for Controlling Robotic Applications”, Journal of Theoretical and Applied Information Technology (1992-8645) Volume 39 No. 2-2012.
6) Ovidiu Buza, Gavril Toderean and Alina Nica, “Voice Signal Processing For Speech Synthesis”, 2006 IEEE International Conference on Automation, Quality and Testing, Robotics, Cluj-Napoca, 2006, pp. 360-364.doi: 10.1109/AQTR.2006.254660.
7) Y. Elmir, Z. Elberrichi and R. Adjoudj, “Score Level Fusion Based Multimodal Biometric Identification (Fingerprint & Voice)”, Sciences of Electronics, Technologies of Information and Telecommunications (SETIT), 2012 6th International Conference on, Sousse, 2012, pp. 146-150. doi: 10.1109/SETIT.2012.6481903
8) Zhixin Chen, “Simulation of Spectral Subtraction Based Noise Reduction Method”, International Journal of Advanced Computer Science and Applications (IJACSA), Vol. 2, No.8, 2011
9) Karam M., Khazaal H., Aglan H. and Cole C., “Noise Removal in Speech Processing Using Spectral Subtraction”, Journal of Signal and Information Processing, 5, 32-41. (2014) doi: 10.4236/jsip.2014.52006.
10) Dave Hale, “An Efficient Method for Computing Local Cross-Correlations of Multi-Dimensional Signals”, CWP Project Review Report, pages 1 – 8, May 2006.
11) Douglas A. Lyon: “The Discrete Fourier Transform, Part 6: Cross-Correlation”, in Journal of Object Technology, vol. 9. No. 2, pp. 17 – 22, April 2010.

Table (1): 24 samples of voice commands used in training the proposed system.

Go	Left	Hold	Five
Back	Right	Zero	Six
On	Stop	One	Seven
Off	Start	Two	Eight
Wake	Up	Three	Nine
Sleep	Down	Four	Ten

Table (2): Selective accuracy of the proposed system.

Words used as a voice commands	Percent of voice commands selective accuracy						Overall Percent of selective accuracy
	User 1 (female)	User 2 (female)	User 3 (female)	User 4 (male)	User 5 (male)	User 6 (male)	
24 voice commands from table 1	91.67%	87.50%	95.83%	91.67%	87.50%	91.67%	90.97%
24 new voice commands out of table 1	95.83%	91.67%	91.67%	95.83%	83.33%	95.83%	92.36%
							91.67%

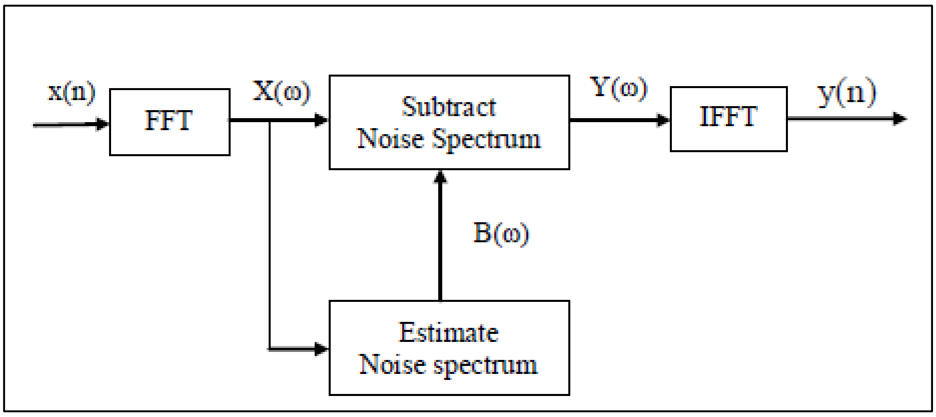

Figure (1): General description of the spectral subtraction method [8].

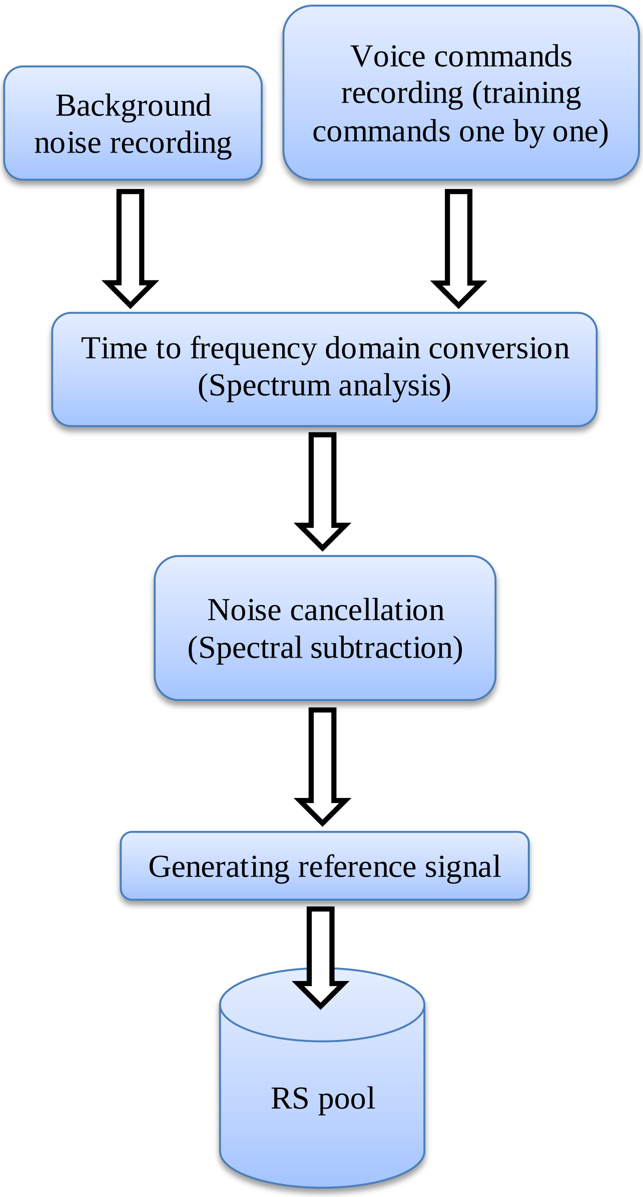
Background noise recording
Voice commands recording (training commands one by one)
Time to frequency domain conversion (Spectrum analysis)
Noise cancellation (Spectral subtraction)
Generating reference signal
RS pool

Figure (2): Generating the Reference Signals pool (RS pool) (phase 1).

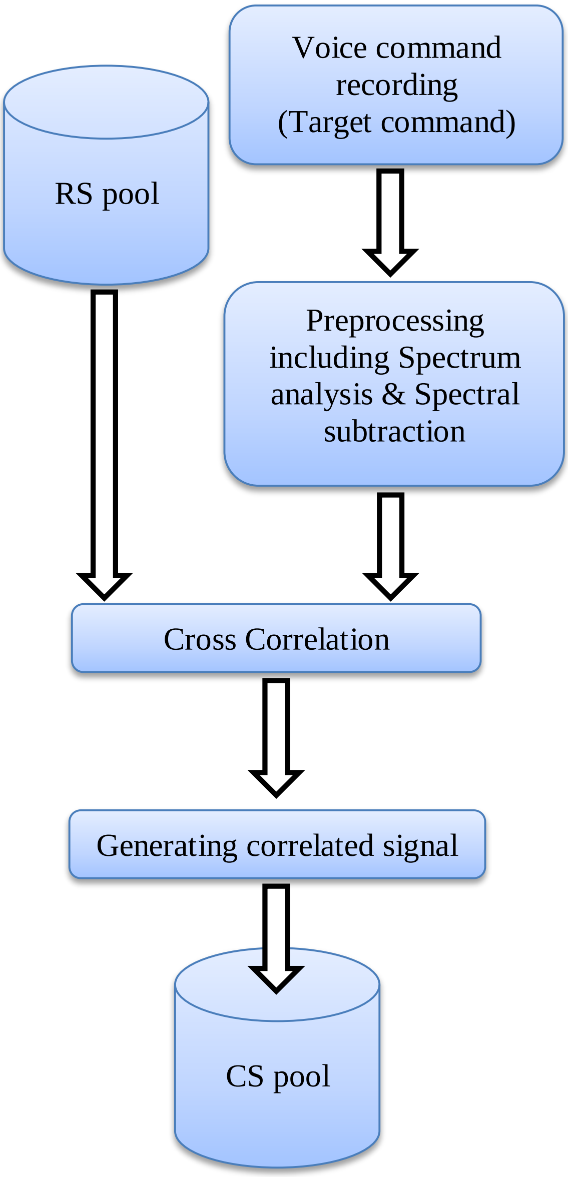
Voice command recording (Target command)
RS pool
Preprocessing including Spectrum analysis & Spectral subtraction
Cross Correlation
Generating correlated signal
CS pool

Figure (3): Generating the Correlated Signals pool (CS pool) (phase 2).

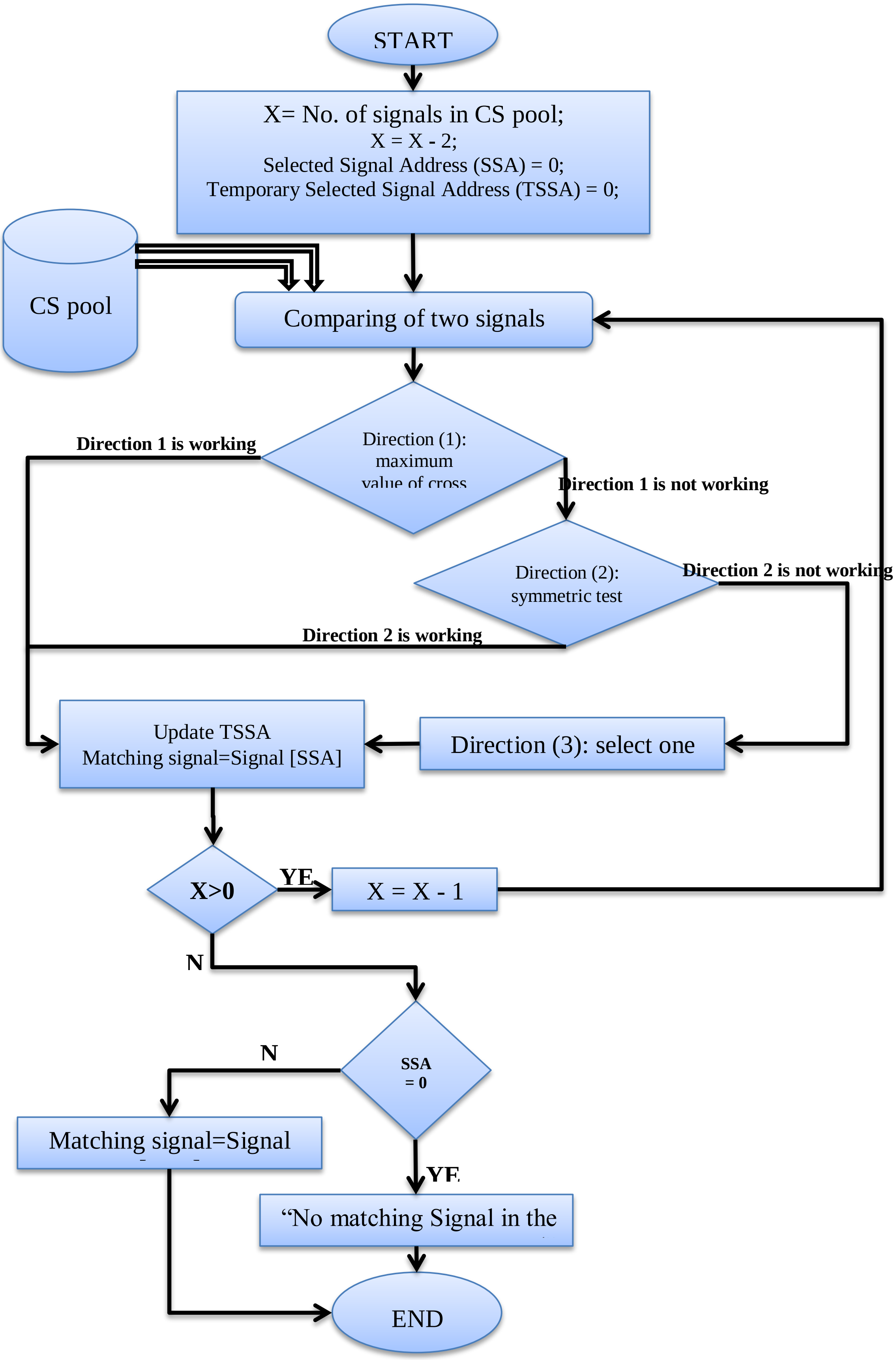
START
X= No. of signals in CS pool;
X = X - 2;
Selected Signal Address (SSA) = 0;
Temporary Selected Signal Address (TSSA) = 0;
CS pool
Comparing of two signals
Direction 1 is working
Direction (1):
maximum
value of cross
Direction 1 is not working
Direction (2):
symmetric test
Direction 2 is not working
Direction 2 is working
Update TSSA
Matching signal=Signal [SSA]
Direction (3): select one
X>0
YE
X = X - 1
N
N
SSA
= 0
Matching signal=Signal
YE
"No matching Signal in the
END

Figure (4): Flowchart of SMCA (Phase 3).

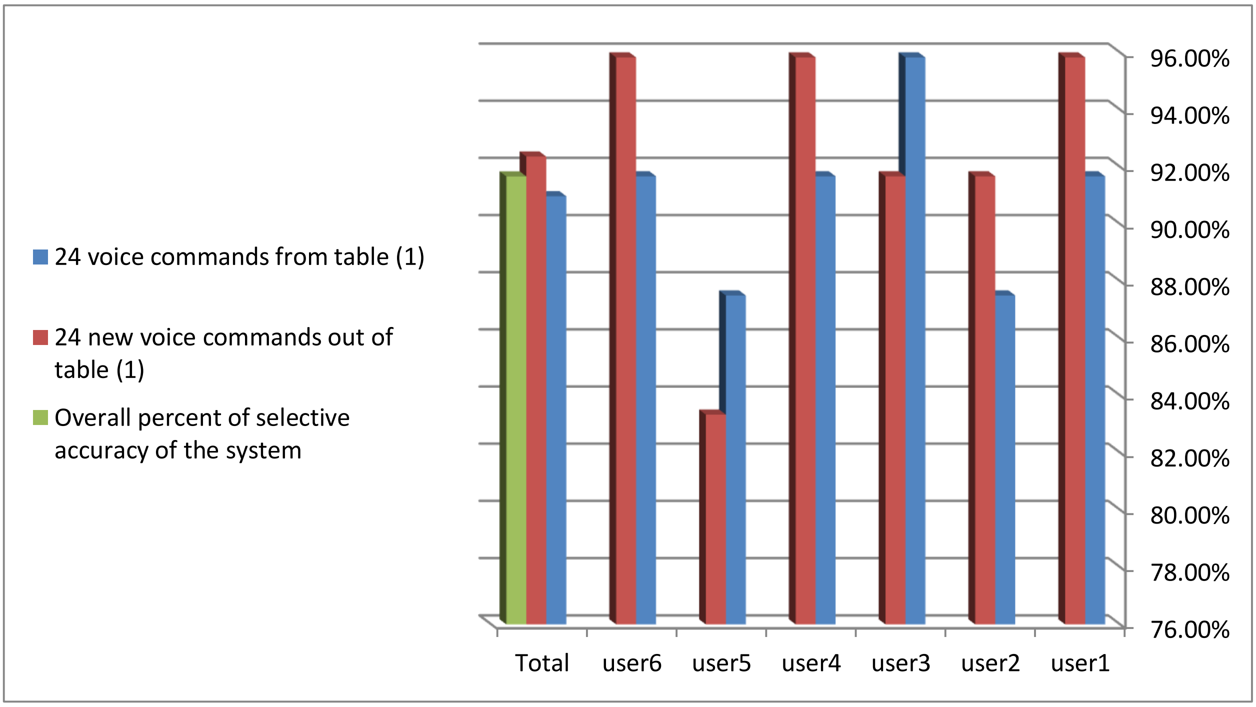

Figure (5): Efficiency chart of the proposed system.

تصميم نظام معتمد لاختيار الاوامر الصوتية

احمد صلاح حميد
مدرس مساعد/جامعة ديالى/كلية الهندسة/قسم هندسة الحاسوب

الخلاصة

إن تمييز الأوامر الصوتية يعتبر وسيلة لفهم الكلام البشري وتحويله إلى مدخلات اتصال بجهاز الكمبيوتر. و وفقا للبيانات المستخدمة لتغذية نظام التعرف على الصوت يمكن تصنيف الأنظمة إلى أنظمة تابعة أو انظمة مستقلة. في هذه الورقة يتم تقديم نظام معتمد لاختيار الأوامر الصوتية. يستخدم النظام المقترح خوارزمية الطرح الطيفية لإلغاء الضجيج. النظام ايضا يستخدم خوارزمية مطابقة جديدة تدعى SMCA لمطابقة هوية الأوامر الصوتية. تم اختيار (3 ذكور و 3 إناث) من المستخدمين لاختبار النظام ومع كل مستخدم يختبر النظام مرتين, واحدة باستخدام أربع وعشرين من الأوامر الصوتية من عينات تم اختيارها مسبقا واختبار اخر باستخدام اوامر صوتية اخرى عددها أربع وعشرين ايضا وهي غير الأربع والعشرين عينة المحددة مسبقا. النظام المقترح قادر على التحديد و بشكل صحيح مطابقة الأوامر الصوتية بنسبة 90.97٪ من الوقت عند اختبار النظام مع الأوامر الصوتية المسبق تحديدها. اختبار النظام مع اوامر صوتية جديدة غير الأوامر الصوتية المحددة مسبقا أظهر نسبة تطابق تساوي 92.36٪. وفقا للنتائج المسجلة من المستخدمين الستة فان الدقة الشاملة لاختيار الاوامر الصوتية في النظام المقترح هي 91.67٪.

الكلمات الدالة: تمييز الاوامر الصوتية، الارتباط المتقاطع، و الطرح الطيفي.