

A Voice Recognition Based Game Design for More Accurate Pronunciation of English

Emre Avuçlu ^{a,*} , Murat Köklü ^b 

^a Department of Software Engineering, Aksaray University, Aksaray, Türkiye

^b Department of Computer Engineering, Technology Faculty, Selçuk University, Türkiye

ARTICLE INFO

Article history:

Received 16 November 2022

Accepted 29 November 2022

Keywords:

English pronunciation

Game programming

Technology

Voice recognition

ABSTRACT

The development of technology has made it considerably easier for people to meet a number of needs. Without technology, it is no longer possible to run certain applications. Nowadays, in many countries, the process of speaking English poses some problems in terms of different situations, such as allocating time for people. In this study, a game-based application was developed to help a person anywhere in the world pronounce English better. The application was implemented in C# programming language. The Speech.dll library was used to introduce voice commands to the system and to perform other necessary operations. Voice commands can be sent by the user via the wireless headset from anywhere in the shooting area. There is no need to wait at the computer while using the application because the developed application gives voice feedback to the user that it is right or wrong after the voice recognition process. In this application letter, word or sentence exercises can be done. The program aims to improve the level of pronunciation of people who want to improve their English-speaking skills.



This is an open access article under the CC BY-SA 4.0 license.
(<https://creativecommons.org/licenses/by-sa/4.0/>)

1. Introduction

The use of technology has become an indispensable part of today. People have constantly developed technology to benefit some applications for their own benefit. Controlling any application with a computer is very easy thanks to the software. Today, some arrangements are made with technology so that people can live more comfortably. It is possible to see examples of this in every field. From medicine to the automotive industry, computerized control, diagnostics, etc., in almost every field, are done by computerized control software. Thus, it is easier to get fast, reliable, quality results. Studies on voice, speech, speaker recognition are as follows:

Today, technology can be defined as an effective part of directing discoveries by using data sharing in the most effective way [1]. Different voice recognition algorithms have been used on MATLAB, they have used "Open", "Close", "Start" and "Stop" instruction sets [2-3]. Using PIC 16F876, attempts were made to recognize voice recognition commands under 4 different conditions [4]. It has been tried by setting up different algorithms on a phone simulation. It has been observed that the results obtained vary according to the way the sound is pronounced [5]. A voice recognition-based security system has been developed [6]. A structure named EllaVoice has been

developed by using improved dynamic time warping algorithms [7]. A program has been developed for NASA by using Mel frequency Cepstrum coefficients algorithm and dynamic time warping and hidden markov model algorithms separately [8]. A remote-controlled robot design has been made using the RS 232 connection with voice command [9]. More than 80% success has been achieved in voice recognition on the letters "a", "e" and "i" [10]. Separate tests were conducted on male and female users [11] and a voice recognition system was used [12].

A speech recognition system independent of text and speaker has been developed on the Turkish language using Artificial Intelligence techniques [13]. Turkish word recognition system has been developed [14]. By using Artificial Neural Networks and Dynamic Time Warping algorithms separately, home applications working with voice command were made [15]. A successful recognition rate was found for 10 people in the simulation environment on Matlab [16]. Mobile vehicle design was made using different voice recognition algorithms [17]. They applied music and speech recognition [18]. Class Non-Principal Component Analysis was compared with Vector Quantization algorithm [19]. They used different voice recognition algorithms [20]. Successful results were obtained in the study that performed 40 commands [21]. The numbers uttered between 0 and 9 were first perceived

* Corresponding Author: emreavuculu@aksaray.edu.tr

by the sound detection system independently of time, and then sound processing techniques were used [22]. It was controlled by voice commands of a remote-control car [23]. It has been tried to determine the English pronunciation of the numbers 0-9 [24-25]. An attempt was made to control a submarine model moving underwater with voice commands [26]. In the study, based on the calls made to the call center, 5 different emotion detections were tried to be tested by 30 people and 70% success was achieved [27]. The success rate of the system, which can make simultaneous comparisons without registration, was 67.5% [28]. A voice-controlled robot was designed using Artificial Neural Network algorithms [29]. In the literature, different studies have been carried out on the development of voice recognition and English pronunciation [30-31].

In this study, a voice recognition-based application was developed to improve English pronunciation. No matter where the purpose of this application is in the world, a program has been written that can be used by people of all ages to improve their English speaking.

2. Materials and Methods

The design of the application made in this study consists of a series of stages. A number of processes are carried out in the process from the recognition of voice commands to the execution of the necessary actions. The flow chart of the developed application is as shown in Figure 1.

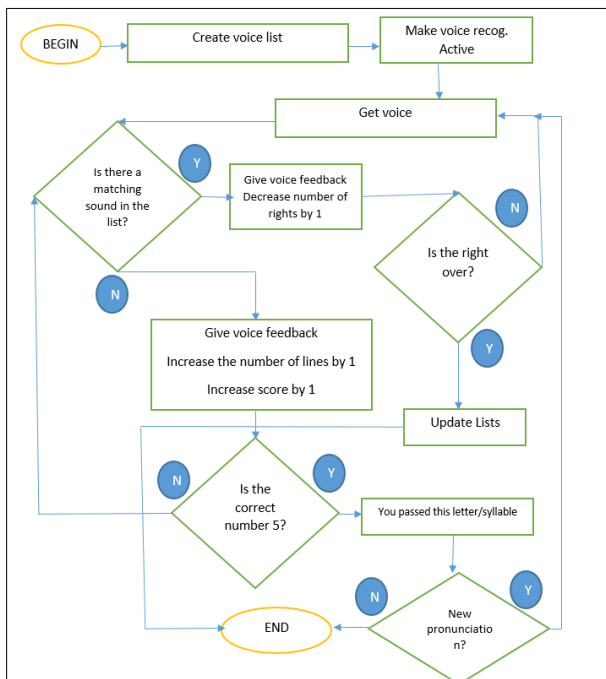


Figure 1. Flow chart of the application

The application made in this study is programmed with the C# programming language. Regarding voice recognition, the following libraries must be added to the system first. After installing the required SDK version on

the computer, the definitions required for the voice recognition related engine to work are defined as shown in the code block below.

```
SpeechSynthesizer SpeechSynth = new
SpeechSynthesizer();
PromptBuilder Builder = new PromptBuilder();
SpeechRecognitionEngine myRecognize = new
SpeechRecognitionEngine();
```

The process of defining which sounds the recognition engine will be sensitive to during the voice recognition process is defined as in the code block below. New definitions can be made to the list as letters or syllables.

```
Choices SpeechList = new Choices();
SpeechList.Culture = new
System.Globalization.CultureInfo("en-US");
SpeechList.Add(new string[] { "one", "2", "3", "a", "b", "c",
"d", "e", "F", "g", "h", "i", "j", "k", "l", "m", "n", "o", "p", "q", "r",
"s", "t", "u", "v", "w", "x", "y", "z" });
Grammar gr = new Grammar(new
GrammarBuilder(SpeechList));
```

After the necessary definitions are made, the sound begins to be heard from the outside. If the received sound is in our list, it returns the necessary answer to us. Otherwise, the catch block is executed and the system is waiting for the sound again.

```
try
{
myRecognize.RequestRecognizerUpdate();
myRecognize.LoadGrammar(gr);
myRecognize.SpeechRecognized +=
sRecognize_SpeechRecognized;
myRecognize.SetInputToDefaultAudioDevice();

myRecognize.RecognizeAsync(RecognizeMode.Multiple);
}
catch
{
pictureBox2.Visible = true;
MessageBox.Show("error", e.ToString());
return;
}
```

In the developed application, first the letter or syllable to be practiced is selected. This process is illustrated in Figure 2.

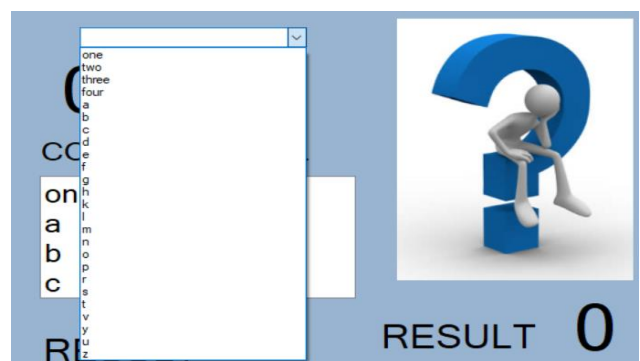


Figure 2. Word selection process in application

The selected letter or syllable is automatically set in the "Input Text" as shown in Figure 3.



Figure 3. Pronunciation listening process

Before starting to pronounce this letter or syllable, the user listens to how this sound is produced by pressing the "Speaking voice" button. The code block that performs this operation is as follows.

```
private void button1_Click(object sender,
System.EventArgs e)
{
    Builder.ClearContent();
    Builder.AppendText(textBox1.Text);
    SpeechSynth.Speak(Builder);
}
```

After the listening process is finished, the "Test Voice" button is pressed. The word "One" begins to pronounce. If he pronounces the word "one" correctly 5 times in a row without making any mistakes, the word "one" is added to the "correct" list. If it is not pronounced correctly 5 times in a row, this time the word "one" is added to the "fail" list. The user earns points for each word included in the "correct" list.

A small high-pitched sound is given as a feedback to inform the user after the correct pronunciation. After mispronunciation, a different treble sound is given to the user as a feedback. So the user knows how to behave. After adding the necessary libraries for the audio feedback process, the following code block is run.

```
axWindowsMediaPlayer1.URL =
"C:\ProgramFiles\Sound\correct.mp3";
axWindowsMediaPlayer1.URL =
"C:\ProgramFiles\Sound\fail.mp3";
```

If the voice recognition engine cannot recognize the voice command given from the outside, it shows the voice it defines as close in the text box. Thus, the user is given both audio and visual feedback.

Correct and incorrectly pronounced words obtained during the trial process are shown in Table 1.

Table 1. Expressions tested in application

Trials	Correct	Fail
one	one	
a	a	
b	b	
c	c	
Three		Three
f		f
e		e
i		i

When we look at Table 1, it is seen that the words that are difficult to express are more in the list of incorrect expressions.

3. Conclusions

In this study, an application was developed for people anywhere in the world to pronounce English words and letters correctly. The Speech.dll library was used to perform the voice recognition process in the application developed in the C# programming language. In the feedback process, the AxInterop.WMPLib.dll library was used to run the necessary codes for voice commands. With the use of the developed application, children will reach a more permanent and impressive pronunciation level at an early age. Developed in a game style, the application will both interest children and help them improve their English pronunciation. The application has been tested on different children and adults and it has been observed that the pronunciation levels have improved. As a result, the English speaking pronunciation of each person who will use this application will increase.

References

- [1] İnternet: Teknoloji, <https://tr.wikipedia.org/wiki/Portal:Teknoloji>, 2016.
- [2] Karakaş, M., Computer Based Control Using Voice Input, Yüksek Lisans Tezi, Dokuz Eylül Üniversitesi, 2010.
- [3] Muda, L., Begam, M., Elamvazuthi, I., "Voice Recognition Using Mel Frequency Cepstral Coefficient and Dynamic Time WarpingTechniques", Journal of Computing, Cilt 2, 2010.
- [4] Fezari, M., Salah, M.B., "A Voice Command System for Autonomous Robots Guidance", IEEE AMC, 2006.
- [5] Baygün, M.K., Yaldır, A.K., "Linear Predictive Coding ve Dynamic Time Warping Teknikleri Kullanılarak Ses Tanıma Sistemi Geliştirilmesi", Pamukkale Üniversitesi, 2009.
- [6] Bala, A., Kumar, A., Birla, N., "Voice Command Recognition System Based on MFCC and DTW", International Journal Of Engineering Science and Technology, Cilt 2, 7335-7342, 2010.
- [7] Hrnčar, M., "Voice Command Control for Mobile Robots", Department of Control and Information Systems Faculty of Electrical Engineering University of Zilina, 2000.
- [8] Price, J., Eydgahi, A., "Design of Matlab Based Automatic Speaker Recognition Systems", 9th International Conference on Engineering Education T4J-1, 2006.
- [9] Zhizeng, L., Jinghing, Z., "Speech Recognition and Its Application in Voice-based Robot Control System", Proceedings Of International Conference On Intelligent Mechatronics, 2004.
- [10] Öztürk, B., Çakar, T., Gerçek Zamanlı Ses Tanıma, Bitirme Projesi, İstanbul Üniversitesi Mühendislik Fakültesi Elektrik/Elektronik Mühendisliği Bölümü, 2007.
- [11] Demirci, M. D., Bilgisayar Destekli Ses Tanıma Sistemi Tasarımı, Yüksek Lisans Tezi, İstanbul Üniversitesi Fen Bilimleri Enstitüsü, 2005.

- [12] Phoophuangpairoj, R., "Using Multiple HMM Recognizers and the Maximum Accuracy Method to Improve Voice Controlled Robots", International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS), December 2011.
- [13] Kirişçiöğlü, F., Karabacak, E., Çetintürk, Ç., "Bilgisayar Destekli Bir Dil Programı", Turkish Speech Recognition Platform TREN.
- [14] Aşlıyan, R., Günel, K., Yakhno, T., "Dinamik Zaman Bükmesi Yöntemiyle Hece Tabanlı Konuşma Tanıma Sistemi", Çanakkale Onsekiz Mart Üniversitesi, Akademik Bilişim, 2008.
- [15] Baygın, M., Karaköse, M., "Gerçek Zamanlı Ses Tanıma Tabanlı Akıllı Ev Uygulaması" IEEE 978-4673-0056, 2012.
- [16] Dede, G., Sazlı, M.H., "Biyometrik Sistemlerin Örüntü Tanıma Perspektifinden İncelenmesi ve Ses Tanıma Modülü Simülasyonu", Savunma Bilimleri Enstitüsü.
- [17] Edizkan, R., Tiryaki, B., Büyükcan, T., Uzun, İ., "Ses Komut Tanıma ile Gezgin Araç Kontrolü", Dumlupınar Üniversitesi, Akademik Bilişim, 2007.
- [18] Bolat, B., Küçük, Ü., Yıldırım, T., "Aktif Öğrenen PNN ile Konuşma/Müzik Sınıflandırma", Akıllı Sistemlerde Yenilikler ve Uygulamaları Sempozyumu, 2004.
- [19] Ertaş, F., Haniççi, C., Konuşmacı Tanıma Sistemi İçin Yeni Bir Sınıflandırıcı, Uludağ Üniversitesi Elektronik Mühendisliği.
- [20] Meral, O., Doğrusal Öngörülü Kodlama ve Adaptif Algoritma Tabanlı Konuşmacı Tanıma, Yüksek Lisans Tezi, İstanbul Üniversitesi Fen Bilimleri Enstitüsü, 2008.
- [21] Asyalı, M.H., Yılmaz, M., Tokmakçı, M., Sedef, K., Aksebzeci, B.H., Mittal, R., "Design and Implementation of a Voice Controlled Prosthetic Hand", Turk J. Elec. Eng. and Comp., Vol.19, No.1, 2011.
- [22] Babui, G., Kumar, H., Vanathi, P.T., "Performance Analysis of Hybrid Robust Automatic Speech Recognition System", IEEE 978-1-46731318-6, 2012.
- [23] Leechor, P., Pompanomchai, C., Sukklay, P., "Operation of a Radio Controlled Car by Voice Commands", 2010 2nd International Conference on Mechanics and Electronics Engineering (ICMEE 2010), 2010.
- [24] Abushariah, A.A.M., Gunawan, T.S., Khalifa, O.O., "English Digits Speech Based on Hidden Markov Models", International Conference on Computer and Communication Engineering (ICCC 2010), 2010.
- [25] Jiang, Z., Huang, H., Yang, S., Lu, S., Hao, Z., "Acoustic Feature Comparison of MFCC and CZT based Cepstrum for Speech Recognition", 2009 Fifth International Conference on Natural Computation, 2009.
- [26] Ferrando, F., Nouveau, G., Philip, B., Pradeilles, P., Soulenq, V., Stean, G.V. P., Courmontagne, "A Voice Recognition System for a Submarine Piloting", IEEE 1-4244-2523-5, 2009.
- [27] Petrushin, V.A., "Emotion in Speech Recognition and Application to Call Centers", Andersen Consulting, 2000.
- [28] Öztürk, N., Ünözkan, U., "Microprocessor Based Voice Recognition System Realization", IEEE 978-1-4244-6904-8/10, 2010.
- [29] Özdemircan, M.Z., Robot Control With Voice Command, Bitirme Projesi, Yıldız Teknik Üniversitesi Bilgisayar Mühendisliği Bölümü, 2008.
- [30] Ayuçlu, E., Taşdemir, Ş., An Application For Web Browser Control With Voice Commands, (26.08.2019 -30.08.2019), Yayın Yeri: 8th International Conference on Advanced Technologies (ICAT'19), 2019.
- [31] Taspınar, Y. S., Köklü, M., & Altın, M. (2020). Identification of the English Accent Spoken in Different Countries by the k-Nearest Neighbor Method. International Journal of Intelligent Systems and Applications in Engineering, 8(4), 191-194.