POLİTEKNİK DERGİSİ

JOURNAL of POLYTECHNIC

ISSN: 1302-0900 (PRINT), ISSN: 2147-9429 (ONLINE)
URL: http://dergipark.org.tr/politeknik

An application to control media player with voice commands

Ses komutları ile media player kontrolü için bir uygulama

***Yazar(lar) (Author(s))**: Emre AVUÇLU[1], Ayhan ÖZÇİFÇİ[2], Abdullah ELEN[3]*

ORCID[1]: 0000-0002-1622-9059

ORCID[2]: 0000-0001-7733-9959

ORCID[3]: 0000-0003-1644-0476

Bu makaleye şu şekilde atıfta bulunabilirsiniz(To cite to this article): Avuçlu E., Özçifçi A. ve Elen A., "Ses komutları ile media player kontrolü için bir uygulama", ***Politeknik Dergisi***, 23(4): 1311-1315, (2020).

Erişim linki (To link to this article): http://dergipark.org.tr/politeknik/archive

DOI: **10.2339/politeknik.646675**

An Application to Control Media Player with Voice Commands

Highlights

- *In the developed application, operations with keyboard and mouse can be done with voice commands.*
- *Voice commands can be sent with the wireless headset from anywhere in the shooting area.*

Graphical Abstract

The following Figure shows a general voice recognition process.

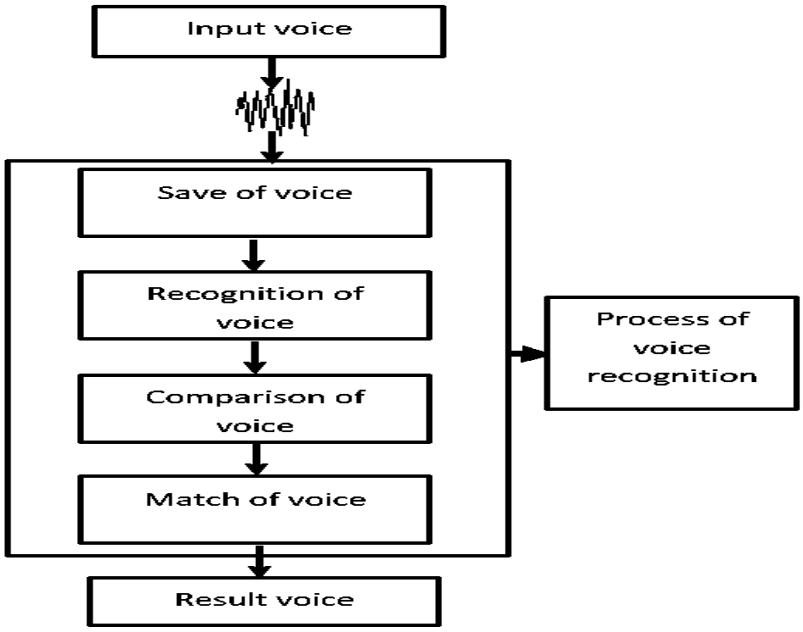

Figure. Voice recognition process

Aim

This application was developed to address the needs of people who cannot listen to music on their own due to any disability.

Design & Methodology

In order to manage the media player with voice commands, voice recognition libraries were first used.

Originality

In this study, an application that provides media player control with voice commands was developed.

Findings

In this study, test procedures were performed with 20 people. In some word tests, more than one test was performed over the same person's voice.

Conclusion

100% accurate recognition can be achieved by using short words and words with full pronunciation when making voice definitions.

Declaration of Ethical Standards

The author(s) of this article declare that the materials and methods used in this study do not require ethical committee permission and/or legal-special permission.

Ses Komutları ile Media Player Kontrolü İçin Bir Uygulama

Araştırma Makalesi / Research Article

Emre AVUÇLU[1*], Ayhan ÖZÇİFÇİ[2], Abdullah ELEN[3]

[1]Teknik Bilimler Meslek Yüksek Okulu, Bilgisayar Tek. Bölümü, Aksaray Üniversitesi, Türkiye

[2]Mühendislik Fakültesi, Endüstri Müh. Bölümü, Aksaray Üniversitesi, Türkiye

[3] TOBB Meslek Yüksek Okulu, Bilgisayar Tek. Bölümü, Karabük Üniversitesi, Türkiye

(Geliş/Received : 14.11.2019 ; Kabul/Accepted : 07.01.2020)

ÖZ

Günümüzde teknolojiyi kullanmak insanların hayatlarını kolaylaştırmak açısından büyük öneme sahiptir. Teknoloji ile bazı uygulamaları çalıştırmak çok kolay bir hal almıştır. Bu çalışmada ses komutları ile media player kontrolü sağlayan bir uygulama geliştirilmiştir. Herhangi bir engelinden dolayı kendi kendine müzik dinleyemeyen kişilerin ihtiyaçlarını gidermek için bu uygulama geliştirilmiştir. Uygulama C# programlama dilinde gerçekleştirilmiştir. Media player'ı ses komutları ile yönetebilmek için önce ses tanıma kütüphanelerinden faydalanılmıştır. Geliştirilen uygulama da klavye Mouse ile media player üzerinden yapılan işlemler ses komutları ile gerçekleştirilebilmektedir. Ses komutları kablosuz kulaklık ile çekim alanının olduğu bir yerden verilebilir.

Anahtar Kelimeler: Ses tanıma, media player kontrolü, engelli birey.

An Application to Control Media Player with Voice Commands

ABSTRACT

Using technology today is of great importance in terms of making people's lives easier. It has become very easy to run some applications with technology. In this study, an application that provides media player control with voice commands was developed. This application was developed to address the needs of people who cannot listen to music on their own due to any disability. The application was implemented in C# programming language. In order to manage the media player with voice commands, voice recognition libraries were first used. In the developed application, operations with keyboard and mouse can be done with voice commands. Voice commands can be sent with the wireless headset from anywhere in the shooting area.

Keywords: Voice recognition, media player control, disabled individual.

1. INTRODUCTION

Today, it is nearly not possible for people to live and carry out some operations without technology. People have developed and used technology every day for their own benefit. Today it is very easy to control any application with software. Nowadays, many applications can be controlled with software so that people can live more comfortably. People can see examples of such practices in every aspect of their lives.

To facilitate the social life of people in studies in this field in the literature: Different voice recognition algorithms and command sets were used on MATLAB [1]. With different voice recognition algorithms, "On TV", "Off TV", "Volume Up", "Volume Down" and "Channel One" command sets were tried separately for male and female users [2]. It has been tried by establishing different algorithms on a phone simulation. The results obtained were found to vary according to the way the voice is spoken [3]. Over 80% success was achieved in voice recognition on the letters "a", "e" and "i" [4]. In a different study, separate tests were performed on male and female users with different algorithms [5].

Using artificial intelligence techniques, a voice recognition system independent of text and speaker was developed on the Turkish language [6]. The syllable-based Turkish word recognition system was developed using different voice recognition algorithms [7, 8]. In the simulation environment performed on MATLAB, the successful recognition rate for 10 people was found to be 99% [9]. They performed music and speech recognition [10]. Successful results were obtained in the study which performed 40 commands [11]. It was controlled by voice commands of a remote controlled car [12]. It has been tried to determine the English pronunciation of the numbers 0-9 [13].

In this study; an application was developed to control the media player to listen to music with voice commands over the computer. The application was implemented using the SpeechRecognitionEngine Class in the System.Speech library in the C # .Net Framework. If there is a match in voice commands, the operations that can be done with mouse and keyboard are executed.

**Sorumlu Yazar (Corresponding Author)*
e-posta : emreavuclu@aksaray.edu.tr

2. MATERIAL and METHOD

The application was programmed in C# programming language. This section describes how the voice recognition process is performed.

2.1. Voice Recognition Process

First stage; the voice recorded in the system. Once the voice is recorded, it can go through various processes and be processed. The following Figure 1 shows a general voice recognition process.

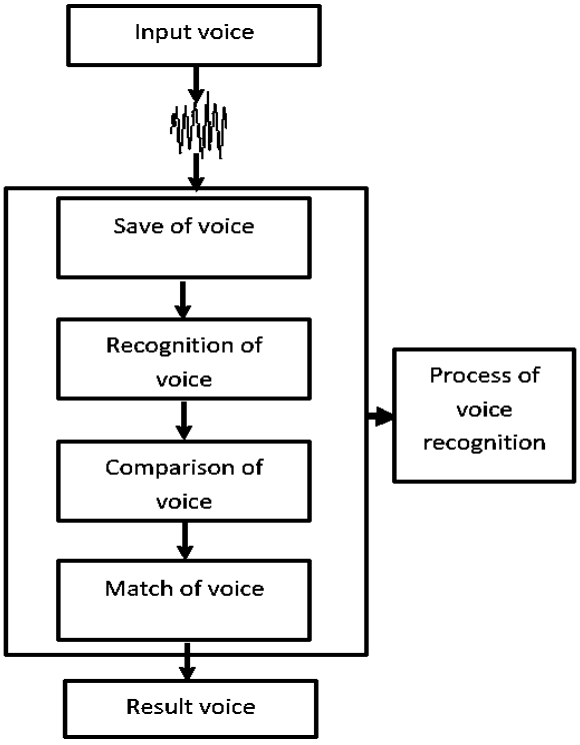

Figure 1. Voice recognition process.

The voice is digitized to perform these operations. The voice is first filtered and then sampled for digitization. Figure 2 shows an example of digitization function.

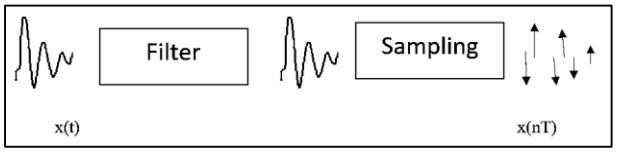

Figure 2. Example of digitization function.

Where x(t) is the analog signal, x(nT) is the digitized signal. In the digitizing stage, the filter shown in Figure 2 refers to the analog filter analog filtering and sampling are performed during the recording of the voice.

In order to use digital signal processing techniques, the analog signal must be represented as a series of numbers [14]. It utilizes the analysis and separation of voice signals to detect voice after sampling.

The voice wave that forms the sound has two important features. These properties are amplitude and frequency [15]. Frequency, while determining the soundness and quiver characteristics of voice; amplitude determines the intensity of the voice and the energy it carries. Equation 1 is given for the Total Amplitude (TG) calculation.

$$TG = \sum_{t=1}^{n} x(t) \qquad (1)$$

In this equation x (t); amplitude at time t; In other words, it expresses the energy carried by the voice wave at the moment t. If the sum of the total amplitude value calculated by this method is above a certain value, then the meaning of sound, that is, speech, is started.

Filters are used for two purposes in the processing of voice. These are the separation of the voice signal and the correction of the voice signal. Digital filters are FIR (Finite Impulse Response) filter and IIR (Infinite Impulse Response) filter. In FIR filters, the input signal forms the output y_n, which is the weighted sum of the current and previous inputs versus x_n. The mathematical expression of this filter is given by Equation 2.

$$y_n = b_0 x_n + b_1 x_{n-1} + b_2 x_{n-2} + \cdots + b_q x_{n-q} \qquad (2)$$

In this equation y_n is the result of the filter output. In IIR filters, the input signal constitutes the output y_n, which represents the weighted sum of the previous outputs, together with the weighted sums of the current and previous inputs versus x_n. In this model, together with the x_n input, the weighted sum of the previous p outputs gives the filter output y_n. After digitizing the voice, the voice is encoded and the voice recognition process is completed. The following libraries should first be added to the system for voice recognition.

```
using System.Diagnostics;
private SpeechLib.SpSharedRecoContext objRecoContext = null;
private SpeechLib.ISpeechRecoGrammar grammar = null;
private SpeechLib.ISpeechGrammarRule menuRule = null;
```

The design of the application consists of certain stages. From the recognition of voice commands to the execution of the media player, a number of operations are carried out. The flow diagram of the developed application is as in figure 3.

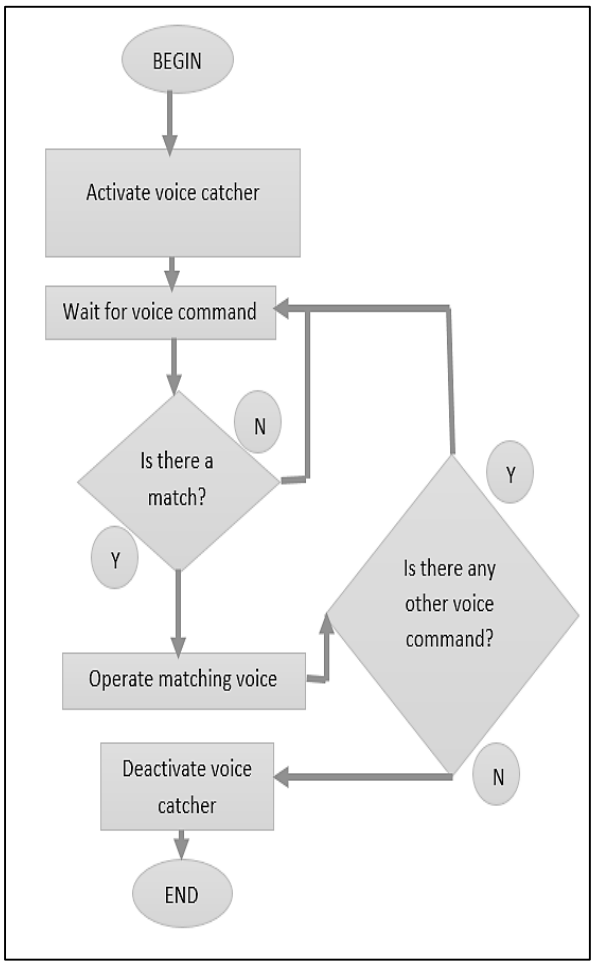

Figure 2. Flow diagram of the system.

Firstly, we need to add the media player component to our application as shown in Figure 4.

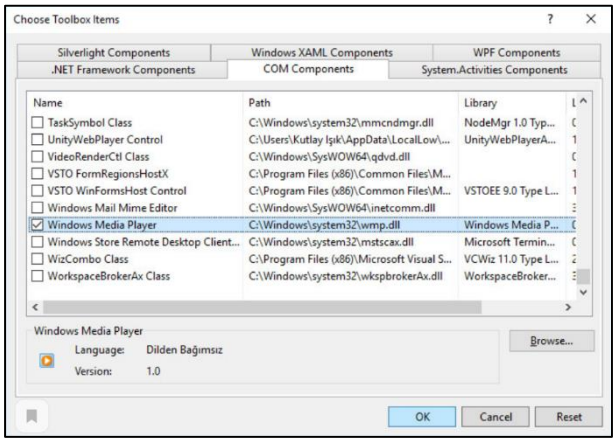

Figure 3. General structure of the system.

The general form design view of the application to be managed by voice commands is shown in Figure 5. We can activate or deactivate this application at any time.

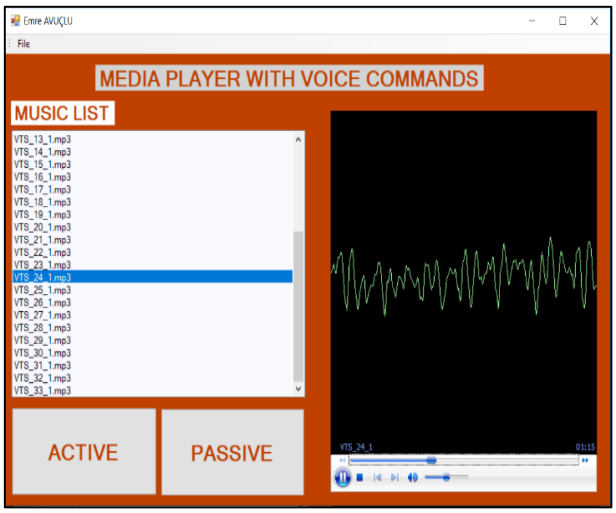

Figure 4. Interface of the developed application.

We need to include the following library in our system first.

```
using System.Speech.Recognition;
```

Voice detection can be performed with the methods in the "*System.Speech*" library in the .Net Framework. The following code blocks are used in the system for feedback after voice recognition.

```
SpeechSynthesizer Speech = new SpeechSynthesizer();
PromptBuilder Builder = new PromptBuilder();
SpeechRecognitionEngine Recognition = new SpeechRecognitionEngine();
```

First, the "*player*" command is given to start the application. This starts the application.

```
if (avuclu.Text == "player")
{
var mediaPlayer = "C:\\Program Files\\Windows Media Player\\wmplayer.exe";
System.Diagnostics.Process.Start(mediaPlayer);
}
```

The code block required to activate or deactivate the application is as follows.

```
avuclu.Text = Result.PhraseInfo.GetText(0, -1, true); // activate
objRecoContext = null; // deactivate
```

The definitions and their use for controlling the media player with voice commands are shown in Table 1 below.

Table 1. Commands and functions.

Voice commad	Feature
Player	Open the media player
Open	Add mp3 list to media player
Active	Media player active
Passive	Media player passive
Play	Mp3 Play
Pause	Mp3 Pause
Next	Mp3 Next in list
Previous	Mp3 previous in list
Stop	Mp3 Stop

After verification of the required definitions and voice command, the data transmission process is executed with the following code block.

```
avuclu.Text = Result.PhraseInfo.GetText(0, -1, true);
if (recog.Text == "play")
    {
 axWindowsMediaPlayer1.Ctlcontrols.play();
 SpeechSynth.Speak("play");
    }
if (avuclu.Text == "pause")
    {
axWindowsMediaPlayer1.Ctlcontrols.pause();
SpeechSynth.Speak("pause");
    }
if (avuclu.Text == " next")
    {
axWindowsMediaPlayer1.Ctlcontrols.next();
SpeechSynth.Speak("next");
    }
if (avuclu.Text == "previous")
    {
axWindowsMediaPlayer1.Ctlcontrols.previous();
SpeechSynth.Speak("previous");
    }
if (avuclu.Text == "stop")
    {
axWindowsMediaPlayer1.Ctlcontrols.stop();
SpeechSynth.Speak("stop");
    }
```

3. CONCLUSION

As the pronunciation of the voice command becomes more difficult and the number of letters in it increases, the level of accurate voice recognition decreases. 100% accurate recognition can be achieved by using short words and words with full pronunciation when making voice definitions. With misrecognition, the voice command performs the function linked to the nearest voice command. No action can be taken with inability to identify. In this study, test procedures were performed with 20 people. In some word tests, more than one test was performed over the same person's voice. The following in Table 2 shows the results of the experimental studies.

Table 2. Experimental results.

Words	Number of trials	Accurate recognition	False recognition	Error rate
Player	20	18	2	%10
Open	20	17	2	%15
Active	15	14	1	%6,66
Passive	15	13	2	%13,33
Play	10	10	0	%0
Pause	12	10	1	%16,66
Next	10	10	0	%0
Previous	25	18	5	%28
Stop	25	23	1	%8

As can be seen from the results, it was more difficult to identify words with a high number of words and difficult to pronounce. The application has a coding that can do everything we do about daily media player with voice commands. It is thought that the application will be useful for people who cannot use the computer for any reason (bedridden, elderly, disabled, etc.). Specially developed to facilitate the lives of the visually impaired. With the application you can meet your daily music listening needs without being connected to anyone.

In this study, media player control was provided to listen to music by remote voice commands. Voice commands can be sent from any point with a wireless or wired headset. Media player was managed with voice commands without using mouse and keyboard. In addition, this study will enable people with disabilities, elderly or bedridden patients to meet their listening needs. The application developed in C # using the Speech.dll library was tested with different voice commands.

ACKNOWLEDGEMENT

This study was supported by Aksaray University Scientific Research Projects Coordinatorship, Aksaray, Turkey. Project Number: 2018-061.

DECLARATION OF ETHICAL STANDARDS

The author(s) of this article declare that the materials and methods used in this study do not require ethical committee permission and/or legal-special permission.

REFERENCES

[1]. Karakaş M., "Computer Based Control Using Voice Input", ***Master Thesis***, Dokuz Eylül University, (2010).

[2]. Muda L., Begam M., Elamvazuthi I., "Voice Recognition Using Mel Frequency Cepstral Coefficient and Dynamic Time WarpingTechniques", ***Journal of Computing***, 2(3):138-143, (2010).

[3]. Baygün M. K., Yaldır A. K., "Linear Predictive Coding ve Dynamic Time Warping Teknikleri Kullanılarak Ses Tanıma Sistemi Geliştirilmesi", ***Pamukkale University,*** (2009).

[4]. Öztürk B., Çakar T., "Gerçek Zamanlı Ses Tanıma", ***Graduation Project,*** Istanbul University Faculty of Engineering Department of Electrical/Electronics Engineering, (2007).

[5]. Demirci M. D.," Bilgisayar Destekli Ses Tanıma Sistemi Tasarımı", ***Master Thesis***, Istanbul University Institute of Science and Technolog, (2005).

[6]. "Bilgisayar destekli bir dil programı" [Online]. Available: https://docplayer.biz.tr/16256021-Bilgisayar-destekli-bir-dil-programi-turkce-konusma-tanima-sistemi.html [Accessed: 19-Sep-2019].

[7]. Aşlıyan R., Günel K., Yakhno T., "Dinamik Zaman Bükmesi Yöntemiyle Hece Tabanlı Konuşma Tanıma Sistemi", Çanakkale Onsekiz Mart University, ***Academic Informatics,*** (2008).

[8]. Meral O., "Doğrusal Öngörülü Kodlama ve Adaptif Algoritma Tabanlı Konuşmacı Tanıma", ***Master Thesis***, Istanbul University Institute of Science and Technology, (2008).

[9]. Dede G., Sazlı,M.H., "Biyometrik Sistemlerin Örüntü Tanıma Perspektifinden İncelenmesi ve Ses Tanıma Modülü Simülasyonu", ***Institute of Defense Sciences***.

[10]. Bolat B., Küçük Ü., Yıldırım T., "Aktif Öğrenen PNN ile Konuşma/Müzik Sınıflandırma", ***Akıllı Sistemlerde Yenilikler ve Uygulamaları Sempozyumu, (***2004).

[11]. 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.,*** 19(1): (2011).

[12]. Leechor P., Pornpanomchai C., Sukklay P., "Operation of a Radio Controlled Car by Voice Commands", ***2nd International Conference on Mechanica and Electronics Engineering (ICMEE 2010),*** (2010).

[13]. 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 (ICCCE 2010),*** (2010).

[14]. Rabiner L., and Schafer R. W., "Digital Processing of Speech Signals", ***Prentice Hall PTR***, 512, (1978).

[15]. Huang X., Acero A., and Hon H. W., "Spoken Language Processing: A Guide to Theory", Algorithm and System Development (1st Ed.), ***Prentice Hall PTR***, New Jersey, 980, (2001)