



Journal of Applied Sciences

ISSN 1812-5654

science
alert

ANSI*net*
an open access publisher
<http://ansinet.com>

Application of Bernstein and Pattern Recognition Methods for Speech Command Recognition

Halil Ibrahim Bulbul

Department of Computer Education, Gazi University, 06500 Ankara, Turkey

Abstract: In this study, a software prepared by using pattern recognition method then the mathematical function of Bernstein method is applied to this software. When compared these two softwares it is find out that the rate of recognition of the Bernstein function applied software is higher than other. But the speed of voice recognition is slow than other software. In addition to that Character sounding of a vector, which is very important in vocal commands and the pre-operations that will be done on the voicing is given with its details.

Key words: Speech recognition, pattern recognition, Bernstein function, windowing, preemphasis filter, autocorrelation function

INTRODUCTION

Speech command recognition provides important advantages for the communication between human and computer. It is so easy to obtain a speech command data. It does not require a special talent such as keyboard and the other storing data methods. By using speech command it can be too fast to transfer texts to the electronic form. It supplies free movement and the practicability of using hands to the users (Mengüşoğlu, 1999).

The first step for recognizing the speech command of computer is to transform voice signals into statistical form, to take the real character of the sound out of the signal, to make recognition procedure by inserting in formation obtained into recognition procedure. In the speech command to recognition procedure, the transfer of speech command to the computer by microphone is the first rank. By transforming the voice into digital form, it has been ready for windowing, filtering and other analysis. By these procedures noise existing in voice and refining the voice from the elements which are dependent on the person as depending on the using area certify.

The methods below are used definitely in speech recognition (Doğan, 1999).

- Pattern recognition.
- Hidden Markov model.
- Dynamic time compressing.
- Nevre webs.

There are two voice vectors which are removed especially in the pattern recognition method. The similarities and the differences of the vectors, vice versa,

are determined by using various distance measurement methods. The first step in the pattern recognition procedure is to apply some preparations to the raw voice data. In this study, as different from the normal pattern recognition procedure, after the voice has been transported preparations Bernstein function is applied to the statistical voice data obtained. The digital voice data which have been applied Bernstein function come closer each others and this situation increases the recognition performance. The purpose of this study increase the rate of recognition of processed voice data by applying Bernstein function.

BERNSTEIN METHOD

In the branches of mathematics for example approximation theory, probability theory, number theory, the solution of the integral and differential equations and the others, Bernstein Polynomials is used for important applications (Heitzinger and Selberherr, 2001; Babu *et al.*, 2002; Groetsch and King, 1973; Büyükyazıcı and Ibikli, 2004). The reason for common application is the simple structure of the Bernstein Polynomials. For any continuous function defined on $[0,1]$. Bernstein Polynomials, are defined (Doğan, 1999) (Fig. 1).

$$B_n(f; x) = \sum_{k=0}^n f\left(\frac{k}{n}\right) \binom{n}{k} x^k (1-x)^{n-k}, \quad 0 \leq x \leq 1 \quad (1)$$

And all $\epsilon > 0$ is given, for any $x \in [0,1]$, $n_0 = n_0(\epsilon)$ number can be found for satisfying the inequality

$$|B_n(f; x) - f(x)| < \epsilon, \quad n \geq n_0(\epsilon) \quad (2)$$

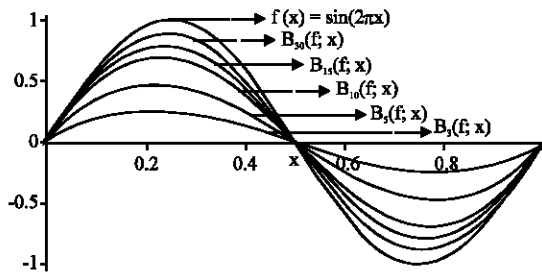


Fig. 1: Approximation of $f = \sin(2\pi x)$ by Bernstein Polynomials

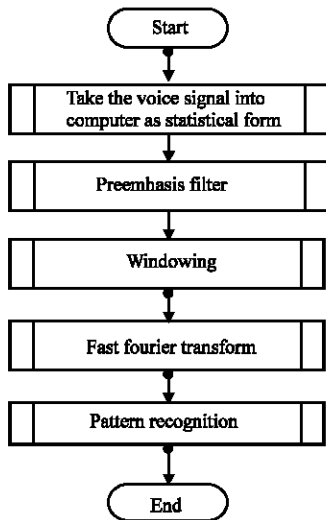


Fig. 2: Overview to the speech recognition

PREPARATION IN THE SPEECH RECOGNITION

There are a lot of pre-processes from taking voice signal occurring speech recognition to model speech recognition procedure. If we approach the recognition procedure with the preparations is used in speech recognition by a basic overview. The method obtained is seen in Fig. 2, step by stop and generally (Doğan, 1999).

PLANING THE SPEECH RECOGNITION SOFTWARE

The software developed in this study recognizes the limited words as dependent on person by applying Bernstein function to the pattern recognition method. In the software, at firstly the user forms dictionary database with the words which she/he wants to be recognized. Software applies Bernstein function to the words that speaker said and all the words in the database during

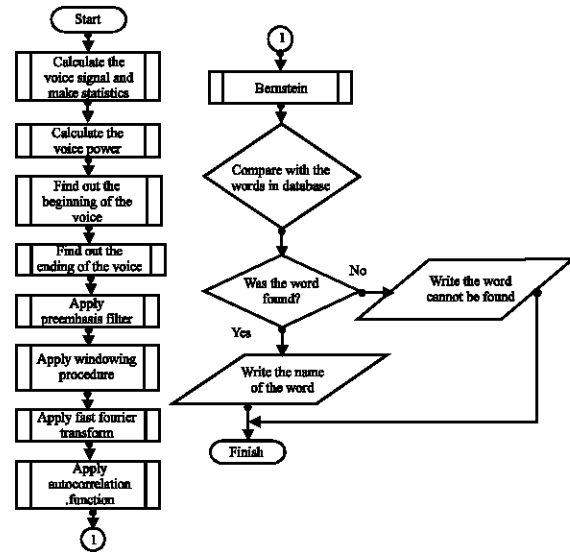


Fig. 3: General movement figure of the developed software

word recognition. Then it makes the comparison procedure and tries to the resembling proportion of all the words in the database one by one and shows the word whose measurement is the most as found word. If there is an order or a program that is wanted to be worked depending on the found word, it works it. The most general movement figure of the developed software is shown (Fig. 3).

Taken the voice into computer: The microphone is used to be taken the voice into computer, digital form. When the voice is taken into computer, the record procedure of audio 1 is used. Codes of record procedure are given below;

```

procedure TForm1.Audio1Record(Sender: TObject; LP,
RP:pointer;BufferSize: Word);
begin
    defa := defa+1;
    if defa = 1 then list1.items.add('Başla') else list1.
items.add(inttostr(buffersize));
    gercekses := lp; Veriyikutupla(lp,sonsize);
    if kutupkont.Checked = true then Veriyigoruntule
(sayisalsessinyali,image1,sonsize);
Enerji; sesyakala; end;
    
```

Voice power calculation procedure: One of the elementary representations of voice signal is to show it with its power (Ikizler, 2003). Signal evaluation method used frequently in speaking recognition systems finds the wave power which is counted as the sums of square values in every point of the wave or its density in a

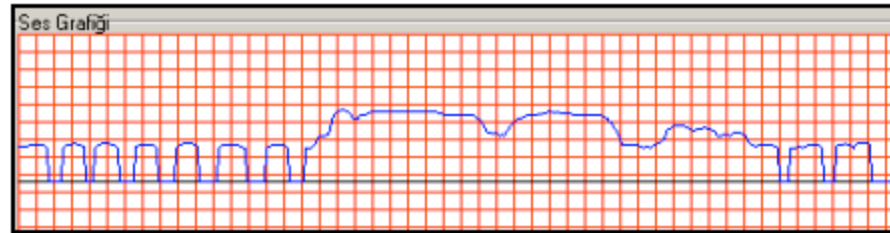


Fig. 4: The power of word save

determined time window (Akçay, 1994). This energy is calculated with short term energy formula. This energy value is used to calculate the energy that voice signal has in a determined time (Qiang and Youwei, 1998) (Fig. 4).

$$\text{Sum of Square Energy: } E = \sum_{i=1}^N x(i)^2 \quad (3)$$

The procedure of capturing voice: It is very important to determine the voice directly in noisy positions. Before developing a new program about speech recognition, it needs to have an effective parting technique (Ganapathiraju *et al.*, 1996). Wrong determination of voice ending point in speech recognition has two negative effects: The first, recognition default occurs because of the determination of voice ending frontier. The second, when voice frontier is determined wrongly, calculation procedure increases because the sections without voice join the circuit (Ying *et al.*, 1993).

The procedure of establishing voice beginning and ending: To establish voice beginning and ending it is needed to look at the energy value calculated. Minimum energy threshold value and this threshold value's continuity during how many samples must be determined. This procedure is made in program sample number is determined as 100 and threshold value is determined as 40. The best values can be found by experimental method. Program looks at energy values by beginning a determined point and if 100 energy value is over level 40 in remaining, it means that voice beginning is found. After voice beginning has determined. Sesbitara sub program is called to find voice ending.

When voice ending is found energy values are looked again. The reduction of energy value under determined maximum energy threshold value in program determines that the voice has finished has determined sample number in program. These two parameters benefit from voice capturing control window in program. Default values of these parameters are 50 and 40 in other words, during 50 samples energy reduces under level 40, it means that voice ending has established.

Preemphasis filter: Voiced explanations have high energy values for low frequencies. The filter called as preemphasis is used that the compounds in low frequency don't mask those in high frequency (Artuner, 1994).

Preemphasis filter is usually explained with the relation in equality 4.

$$P(z) = 1 - \mu z^{-1} \quad (4)$$

μ squared changes between 0.9 and 0.95 (Burrows, 1996). z^{-1} represents standart delay (Doğan, 1999).

Program codes written connected with Preemphasis filter are given below.

```
Procedure preemphasisfilter (veri:pointer; size:integer;
katsayi:real);
var
preemhasisli:pointerintegerdizi; i:integer;
begin
preemhasisli := veri;
if preemhasisli[0] > 10 then preemhasisli[0] := 0;
for i := 0 to size-1 do begin
preemhasisli[i+1] := round(preemhasisli[i+1] -
Katsayi * preemhasisli[i]);
end; end;
```

Windowing: Captured voice datum must be windowed, taken of a determined part for manipulation with Fourier Analysis. This part taken must be the part which explains the smallest meaning part and be as Fourier Analysis Calculation in short time. This time is generally 30 min (Doğan, 1999). Voice signal graphs, applied raw voice signal and Hanning windowing are shown in Fig. 5.

Fast fourier transform (fft): Fast Fourier Transform (fft) divides voice signal into frequencies formed. A normal voice signal is formed by connecting voice in various frequencies such as the colors. The raw form of voice signal receives different images although human says the same word. By the way, it is difficult to perceive voice signal with its raw quality (time and amplitude). For this

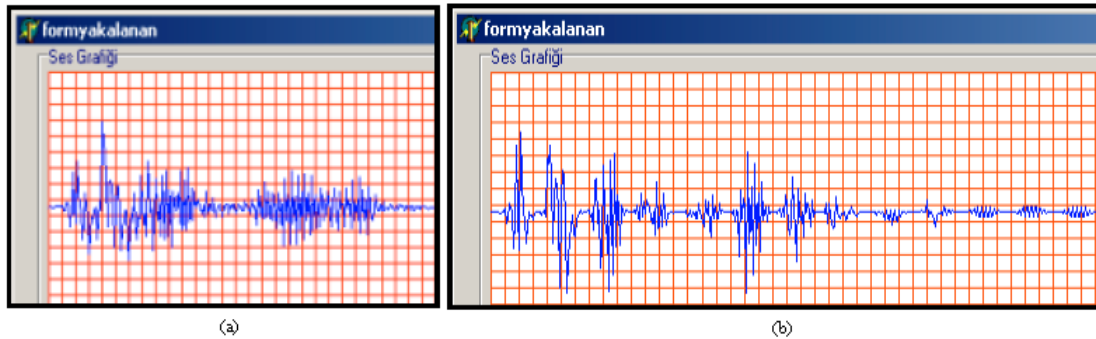


Fig. 5: (a) Statistical signal and (b) signal applied hanning windowing

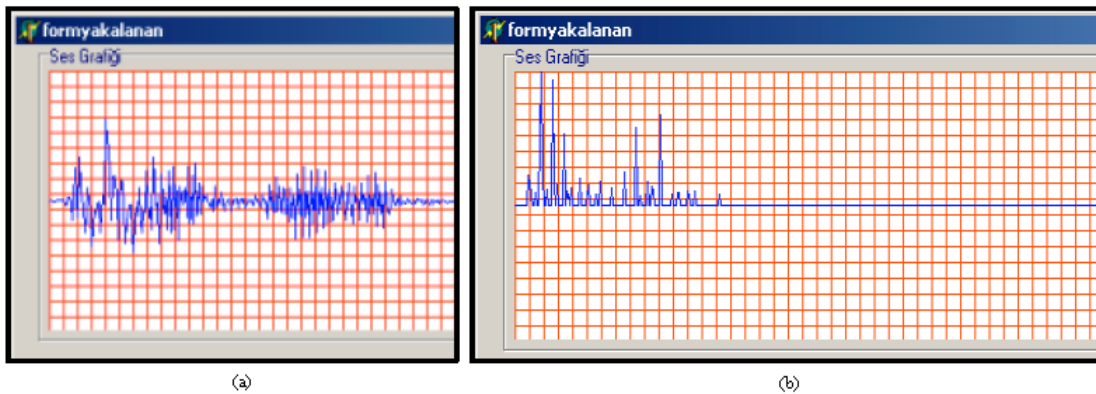


Fig. 6: (a) Raw voice signal and (b) voice signal applied FFT

reason, voice signal is manipulated in various forms and is tried to be recognized by this way. FFT is one of these techniques (Aydin, 2005) (Fig. 6).

Autocorrelation function: Autocorrelation function is a function that generally discovers main frequencies and makes recognition easy (Doğan, 1999). Discovering character frequency in noisy environments is the key technique to correct voice. A lot of developed speaking systems, the rightness of discovering character frequency depend on the quality of voice directly after the correction procedure. Autocorrelation function supplies the best performance in noisy environments (Kobayashi and Shimamura, 2000).

In the developed program about speech recognition, Blackman-Tukey autocorrelation distance is used. And Blackman-Tukey autocorrelation distance is as in equality 5 (Dendrinos and Carayannis, 1998) (Fig. 7).

$$R(k) = (1/(N-k)) \sum_{i=0}^{N-1-k} x(i).x(i+k) \quad (5)$$

Recognition: It is the section where the comparison of manipulated voice of the speaker has spoken with the source signals has been saved in database by manipulating before. In these sections as different from normal pattern recognition, Bernstein function is applied to all reference signals in database and to voice signal the speaker has spoken. Then comparison procedure is made and is tried to be found the closest reference signal. If finding proportion is lower than 85% it means that the word hasn't found. Subprogram applies Bernstein function to voice data is given below.

```

Procedure Tformcath.bernstein(size:integer);
Var pencere,pencsay,penbuy,tuti,x,ne:integer;
iyerine:integer;k,i:integer;fe,ge:array[0..5000] of integer;
begin
for k:= 0 to size do begin
    fe[k]:= fftalan[k]; ge[k]:= fftalanokunan[k]; topl[k]:= 0;
    topg[k]:= 0;
end;
penbuy:= 256;pencere:= size div penbuy;
for pencsay:= 0 to pencere-2 do begin

```

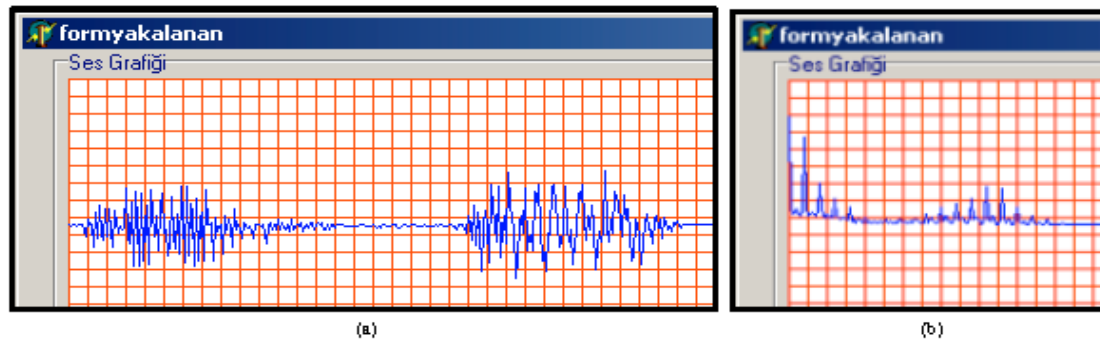


Fig. 7: (a) Raw voice signal and (b) signal applied autocorrelation function

```
for x: = 0 to 100 do begin
    iyerine: = 0;
    for i: = pencsay*penbuy to pencsay*penbuy+penbuy-
        1 do begin
        topl[i]: = topl[i]+binomial (penbuy, iyerine)* ustal
            ((x/penbuy), iyerine)*ustal ((1-(x/penbuy)), penbuy-
            iyerine)*fe[i];
        topg[i]: = topg[i] +binomial (penbuy, iyerine)* ustal
            ((x/penbuy), iyerine)*ustal((1-(x/penbuy)),penbuy-
            iyerine)*ge[i];
        inc(iyerine); tuti: =i; end; end;end;end;
```

There are various distance measurements that can be used when calculating the resemblance of voice signal. In the program that we have developed, Oclid measurement is used. Oclid measurement is the most used method. It is to find the geometric measurement between two points in vertical coordinate system. Oclid measurement is defined as (Akçay, 1994).

$$d(X,Y)_{Oclid} = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (6)$$

The voice signal which is going to be defined compares with the reference signal in database by using Oclid measurement. In this procedure, Oclid measurement, for every window, is calculated by dividing of two voice signals into the window eight times bigger than normal window number. If calculated Oclid value is smaller than the reception threshold value determined in program, found window number is increased one value more. The procedure is made for every window. At last resembling percentage is calculated by the formula of

$$\text{Resembling}(\%) = \frac{\text{Found window No.}}{\text{Total window No.}} \times 100 \quad (7)$$

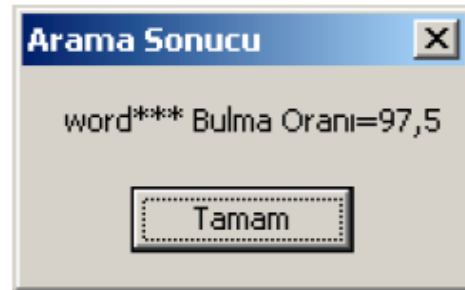


Fig. 8: The resembling proportions of voices

After these procedures, program gives if the voice has identified or not, if it has identified in how percentage proportion it has identified and which word has identified. If it wants an order dependent on the found word can be worked. For example, by the word, the word program can be worked. Besides it lists that resembles in how percentage proportions for every references registered in database. In Fig. 8, the recognition percentage of the word by the program is described as:

CONCLUSIONS

In this study, a software about speech command recognition has been developed and Bernstein function that is a mathematical function to pattern recognition method has been applied. Consequence of the study, recognition proportion of voiced explanations of software as depending on person has appeared much higher than pattern recognition method applied without Bernstein method. Besides when the words are sometimes mixed each others in normal pattern recognition method, the words are not mixed each others in pattern recognition method applied with Bernstein function together.

Despite speech recognition software applied Bernstein function recognizes voiced explanations in a very good level program velocity becomes slow and replying period increases because of increasing.

In developed speech recognition software if resembling proportion is over 85% it means that the voice has recognized. In this software, according to the trial results done separately for single illustration the lowest recognition percentage has appeared 100%. Besides when the illustration in database increases, recognition percentage and resembling proportion increases, recognition performance too.

REFERENCES

- Akçay, B., 1994. Yapay Sinir Ağları İle Türkçe Konuşma Tanıma. Yayınlanmamış Yüksek Lisans Tezi, Hacettepe Üniversitesi Fen Bilimleri Enstitüsü, Ankara.
- Artuner, H., 1994. Bir Türkçe Fonem Kümeleme Sistemi Tasarımı ve Gerçekleştirimi. Yayınlanmamış Doktora Tezi, Hacettepe Üniversitesi Fen Bilimleri Enstitüsü, Ankara.
- Aydın, Ö., 2005. Yapay Sinir Ağlarını Kullanarak Bir Ses Tanıma Sistemi Geliştirilmesi. Yayınlanmamış Yüksek Lisans Tezi, Trakya Üniversitesi Fen Bilimleri Enstitüsü, Edirne.
- Babu, G.J., A. Canty and Y. Chaubey, 2002. Application of bernstein polynomials for smooth estimation of a distribution and density function. *J. Stat. Planning Inference*, 105: 377-392.
- Burrows, L.T., 1996. Speech processing with linear and neural network models. Unpublished Thesis, Cambridge University, England.
- Büyükyazıcı, I. and E. Ibikli, 2004. Some properties of generalized bernstein polynomials of two variables. *Applied Math. Comput.*, 156: 367-380.
- Dendrinos, M. and G. Carayannis, 1998. Spectrum analysis using a new autocorrelation measure. *Acoustics, Speech and Signal Processing*, 4: 2468-2471.
- Doğan, S., 1999. Pc Ortamında Sesli Komutları Tanıma. Yayınlanmamış Yüksek Lisans Tezi, Marmara Üniversitesi Fen Bilimleri Enstitüsü, İstanbul.
- Ganapathiraju, A., L. Webster, J. Trimble, K. Bush and P. Kornman, 1996. Comparison of energy-based endpoint detectors for speech signal processing. Bringing together education. *Sci. Technol. Proc. IEEE*, pp: 500-503.
- Groetsch, C.W. and J.T. King, 1973. The Bernstein polynomials and finite differences. *Math. Magazine*, 46 : 280-282.
- Heitzinger, C. and S. Selberherr, 2001. Optimization for TCAD purposes using Bernstein polynomials. *Proceeding SISPAD Conf.*, pp: 420-423.
- Ikizler, N., 2003. Türkçe’de konuşmacıdan bağımsız hece tanıma sistemi. Yayınlanmamış Doktora Tezi, Karadeniz Teknik Üniversitesi Fen Bilimleri Enstitüsü, Trabzon.
- Kobayashi, H. and T. Shimamura, 2000. A weighted autocorrelation method for pitch extraction of noisy speech. *Acoustics Speech and Signal Processing IEEE Int. Conf.*, 3: 1307-1310.
- Mengüşoğlu, E., 1999. Bir Türkçe Sesli İfade Tanıma Sisteminin Kural Tabanlı Tasarımı Ve Gerçekleştirimi. Yayınlanmamış Yüksek Lisans Tezi, Hacettepe Üniversitesi Fen Bilimleri Enstitüsü, Ankara.
- Qiang, H. and Z. Youwei, 1998. On prefiltering and endpoint detection of speech signal. *Signal Processing Proceedings, 4th Int. Conference on ICSP*, 1: 749-752.
- Ying, G.S., C.D. Mitchell and L.H. Jamieson, 1993. Endpoint detection of isolated utterances based on a modified teager energy measurement. *Acoustics, Speech and Signal Processing, Int. Conference on ICASSP 93 IEEE*, 2: 732-735.