The Introduction Types of Vocal Sound in Choir in Realtime Using Hankel Transformation and Macdonald Function

Wahyu Fuadi[1,] Mutammimul Ula[2], Muhammad Sadli[3]

Department of Technical Information, Malikussaleh University Lhokseumawe
Aceh Utara, INDONESIA.

[2] mutammimul.ulla@gmail.com

ABSTRACT

A group sang in the choir consists of several parts, among others soprano voice, alto, tenor and bass. Prior to the formation of choral group first performed the determination of the sound of each singer by coach or expert vocal choir. Voice recognition process carried out with samples of recorded voice input and voice samples realtime on the program. Sounds are included in the sample will produce energy weight values based on calculations using the Hankel transform and Function Macdonald. The results showed that the accuracy of the system depends on the accuracy of sound techniques trained and tested, and obtained an average percentage of successful recognition of sounds in record reached 66.87%, while the average percentage of the successful introduction of the type of voice in real time reaches 72%.

Keywords: Choir, Sopran, Alto, Tenor, Bass, Transformasi Hankel, Fungsi Macdonald

INTRODUCTION

Music has become an integral part of human life. Many hobby related to music, ranging from hobby to hobby playing instruments or singing voice though. In music theory, humans have the type of sound that varies both men and women. Type in the male voice is divided into tenor, baritone and bass. While the women are divided into soprano, mezzo-soprano and alto.[1]

Generally, a choral group performed choral music which consists of several parts of the sound. Prior to the formation of a choral group, is the determination of the type of sound each singer by coach or an expert choral vocals. This should be done in front of a choir trainer and use the piano to match the singer sung tones.[2]

Chorus or choir is a term that refers to a musical ensemble consisting of singers and music performed by the ensemble. Generally choir consists of four voice parts, ie soprano, alto, tenor, and bass, although it can be said that there is no limit to the number contained voices in the choir. In addition to the four voices, the number of the most common types of sound in the chorus is three, five, six, and eight. When singing with one voice, the chorus sing termed Unisono.[2]

Until now there is no specific software has a function to identify the type of singer's voice automatically. It is very unfortunate given in the field of speech recognition technology is growing very rapidly.

For lay people who are just learning, usually will have difficulty in determining the type of their voices. In the process of determining the type of sound, usually done manually by using the help of musical instruments, generally piano, by an expert or a vocal coach. To assist this process made a speech recognition system by utilizing digital signal processing technology. The system can distinguish four types of sound used in the choir, which soprano, alto, tenor and bass.

This system uses the Hankel transformation which aims to recognize the input voice as a sample and can normalize sound waves. It also uses the MacDonald function in speech recognition system as a combination to produce a better system performance and more accurate in the process of speech recognition.

RESEARCH METHODS

Sound Production Mechanism

Human voice production requires three elements, namely resources, sources of sound and voice modifier. This is the basis of the theory of source-filterpada speech signal production. Resources on a normal voice signals generated from the movement of the muscle compression of the lungs. Sound source, voiced and unvoiced signals, the result of each vibration of the vocal cords. Modifying the sound is articulators, which alters the shape of the vocal tract so that the characteristic frequency of the acoustic cavity through which the sound passes through what.[3]

The three main controls on the production of sound is lung (resources), the position of the vocal cords (sound source), and the shape of the vocal tract (voice modifier). Vocal tract consists of the pharynx (the connection between the esophagus to the mouth) and mouth. Nasal tract, starting from the back of the palate and ends at the nostrils.[3]

Formation of speech begins with the airflow produced by the lungs. How it works similar like a piston or a pump which is pressed to generate air pressure. At the time of vocal cordsberada in a state of tension, the air flow will cause vibration in the vocal cords and produce speech sounds called voiced sound. At the time of the vocal cords are in a state of weakness, the air flow going through a narrow area on the vocal tract and cause turbulence, resulting in a sound that is known by the unvoiced sound.[4]

Human Voice Types

Type of sound are different types of sounds that are classified using certain criteria. Classification is the process by which the sound of the human voice assessed, then be classified into certain types of sound. There are many different types of sounds based on a wide variety of classification systems. Here's vocal range in accordance with the type of voice and representation in the frequency based on scientific pitch notation.[3]

The division of a voice / vocal type is important to diketahui.Pembagian kind of voice for women is high soprano voice to voice, mezzo-soprano voice to sound moderate, and alto voice to sound kind of low. Meanwhile, the division of the type of male voice is the sound of the tenor voice for this kind of high, baritone voice to the type of sound being, and the sound of the bass to sound kind of low. The combination of the types of sounds it can produce a beautiful sound.[2]

Table 1. Frequency Allocation Each Type of Voice

Gender	Vocal Type	Vocal Range in music notation	Vocal range (Hz)	Basic Frequency (Hz)
Male	Tenor (High)	C3 - C5	130 - 523	16.35
	Bariton (middle)	F2 - F4	87 - 349	21.8
	Bass (Low)	E2 - E4	82 - 329	20.6
Female	Soprano (high)	C4 - C6	261 - 1046	16.35
	Mezzo-soprano (middle)	A3 - A5	220 - 880	27.5
	Alto (low)	F3 - F5	174 - 698	21.8

WAV

WAV is the format standardd audio Microsoft and IBM for personal computers (PCs), typically using coding PCM (Pulse Code Modulation) WAV is uncompressed data so that the entire sample were saved audio everything in hard disk. Software which can create WAV from analog sound example is Windows Sound recorder. This Audio rarely used files on the internet because of the relatively large size with a maximum limit for a WAV file is 2GB.[5]

These parameters are expressed settings used by the ADC (Analog-to-Digital Converter) at the time of the recorded audio data. As an illustration, digital audio data stored in the CD audiomemiliki characteristics of the sample rate of 44100 Hz, 16 bits per sample, and 2-channel (stereo), which means that every single second sound composed of 44100 samples, and each sample is stored in the data for 16-bit or 2 bytes. Sample rate is always expressed for each channel. So suppose a digital audio data has 2 channels with a sampling rate of 8000 samples / sec, then indeed in every second there will be 16000 samples. [6]. As explained earlier that to stream audio data using the form header PCMWAVEFORMAT.PCM structure which stands for Pulse Coded Modulation, which is a method used to convert the audio signal from analog form to digital form.[6]

Description of the fields of this WAVEFORMAT structure can be seen in Table 2.2.

Table 2(Part-I). The fields in the data structure WAVEFORMAT

Field	Explanation
wFomatTag	Indicates the type of data format and has a value of WAVE_FORMAT_PCM.
nChannels	Shows the number of channels that exist in the data *waveform audio*. For mono use one channel while for *stereo* use two channel.
nSamplesPerSec	Shows the magnitude of *sample rate* in *sample* / sec.

Table 2(Part-II). The fields in the data structure WAVEFORMAT

Field	Explanation
nAvgBytesPerSec	Shows the average data transfer rate, in *byte* per second. for example, for PCM 16-bit *stereo* in 44.1 kHz, nAvgBytesPerSec worth 176400 (2 canal * 2 byte per *sample* * 44100).
nBlockAlign	Shows the number of *byte* used for only one *sample*. for example, for PCM 16-bit *mono*, nBlockAlign will worth 2.

The structure of the data in the audio file varies depending on its audio format. WAV file data structure has a structure as shown in figure 2 & 3.
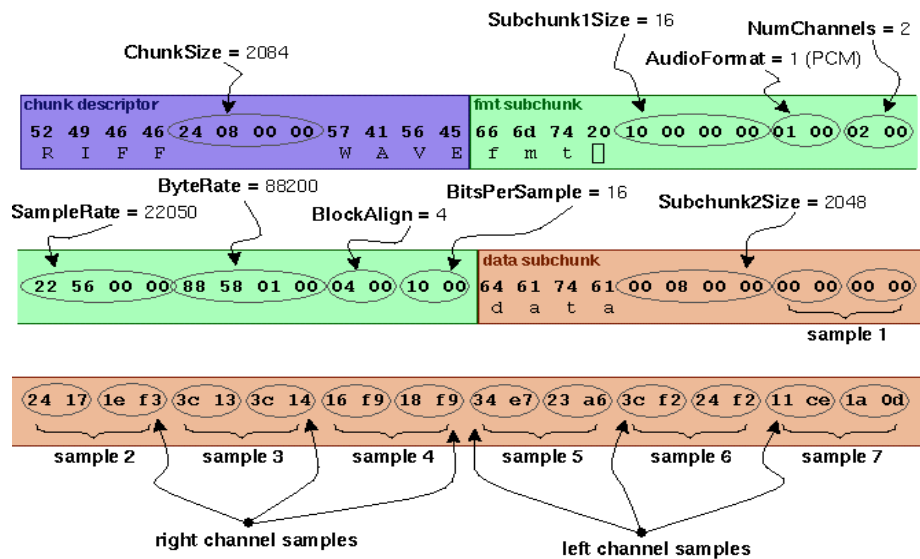

Figure 1. Example of the WAVE file structure in the form of hexa

In the WAV file structure consists of:

1. Chunk Descriptor consists of the data: 52 49 46 46 24 08 00 00 57 41 56 45.

2. Fmt subChunk consisting of data subChunk1size, audioFormat, numChannel, samplerate, byteRate and BlockAlign namely: 66 6d 74 20 10 00 00 00 01 00 02 00 22 56 00 00 ## 58 01 00 04 00 10 00.

3. Data subChunk subChunk2size that consists of data and samples, namely: 64 61 74 61 00 0##1 00 00 00 00#2 24 17 1e f3 #3 3c 13 3c 14 #4 16 f9 18 f9 #5 34 e7 23 a6 #6 3c f2 24 f2 #7 11 ce 1a 0d.[6]

Hankel Transformation

Hankel transformation is an integral transformation developed by **Hermann Hankel**. [7] General equation Hankel transformation:

$$F_v(k) = \sum_{r=0}^{N-1} f(r) \times r \times K_v(kr)$$

for r = 0, 1 ,2, 3, ..., N-1.

This equation is known denganTransformasi Hankel equivalent to the multiple Fourier transform. In general graph Hankel integral transformation tends to approach the abscissa as the increase in the value of the numerical absisnya.Perhitungan be a good alternative to calculate Hankel transformation. Because the Hankel transformation is integral form, the numerical calculation called numerical integration.[7]

Macdonald Function

MacDonald function is a modification of Cylindrical Bessel functions and function, which is made by H.M. MacDonald. [8] General equation of the MacDonald function, namely:

$$K_v(kr) = \frac{\pi}{2} \frac{I_{-v}(kr) - I_v(kr)}{\sin v \, \pi}$$

where v is an arbitrary for real number of non-integral.[8]

$$I_v(kr) = \sum_{m=0}^{\infty} \frac{\left(\frac{kr}{2}\right)^{v+2m}}{m! \, \Gamma(v+m+1)}$$

$$I_{-v}(kr) = \sum_{m=0}^{\infty} \frac{\left(\frac{kr}{2}\right)^{-v+2m}}{m! \, \Gamma(-v+m+1)}$$

Data Collection

After planning the study, which is done first is to collect reference on Signal Processing and Voice and data dibutuhkandalam making application. The data used in this study is divided into two, namely the data directly (realtime), and the data are not directly in the form of sampled sound recording soprano, alto, tenor and bass. The details are as follows:

Table 3. Details of Sample Sound Recording (Record)

Types of Voice in Vocal Choir				
Voice Sampel				Total Sampels
Sopran	Alto	Tenor	Bass	
20	20	20	20	80

Table 4. Details of Direct Voice Samples (Real Time)

Register Women Voice in Singing Technic				
Voice Sampel				Total Sampels
Sopran	Alto	Tenor	Bass	
5	5	5	5	20

For realtime voice samples pronounced directly by some respondents with the help of a microphone. Realtime voice sample was used as a comparison to see the accuracy of the voice recognition system directly to the speech recognition process is taken from a sample recording.

Overall System Scheme

The scheme of the overall system in the choir in realtime using the Hankel transformation and macdonald function is as follows:
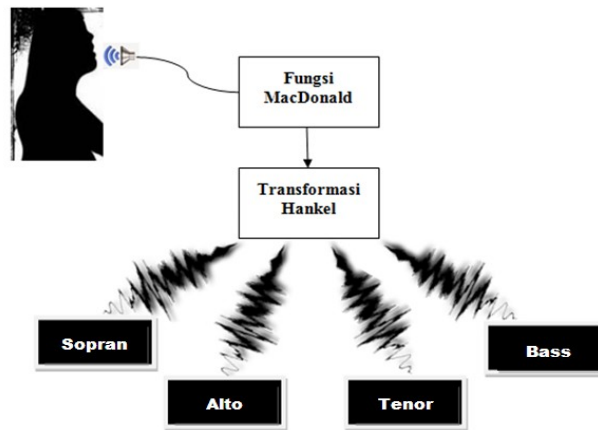

Figure 2. Schematic Overall System

Voice Training Scheme Using Macdonald And Transformation Hankel Functions

The voice training using Hankel transformation and macdonald function is as follows:
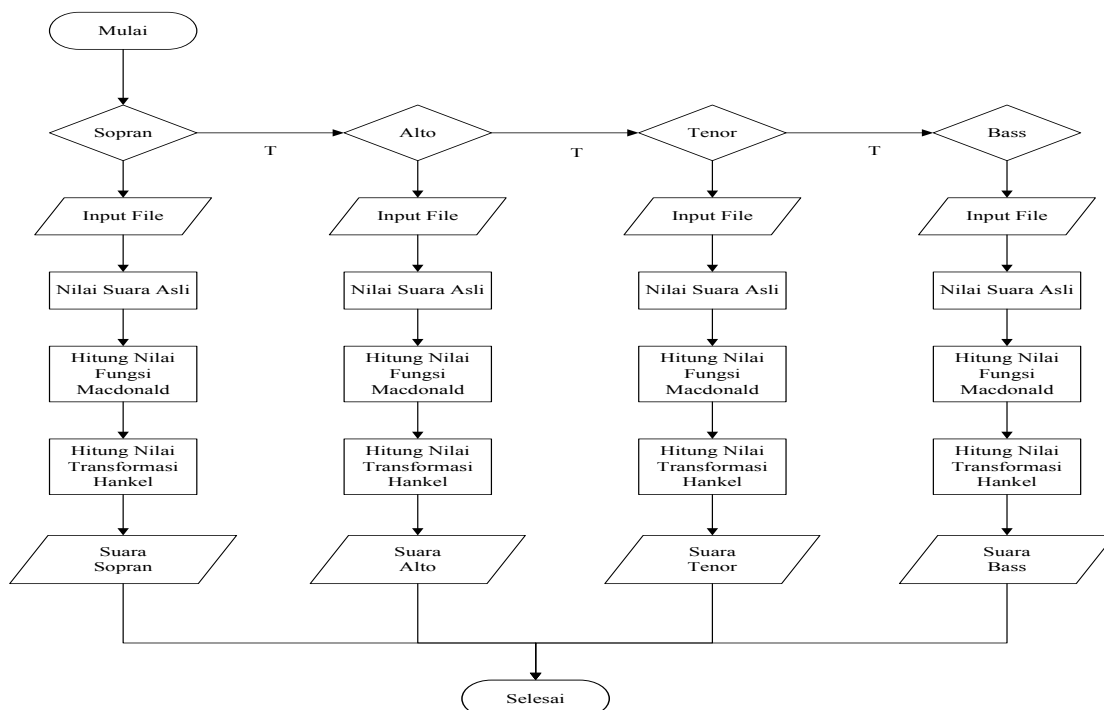
Figure 3. Voice Training Scheme Using Functions Macdonald and Hankel Transform Testing Scheme Sounds Using Functions Macdonald and Hankel Transform
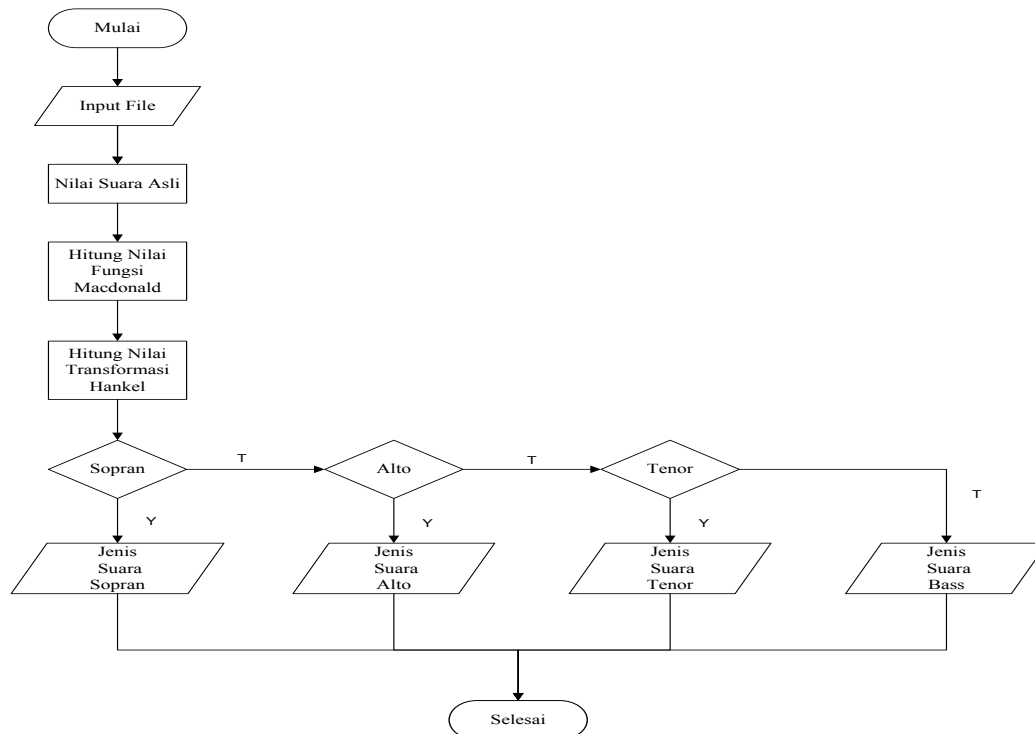

Figure 4. Schematic Sound Testing Using Function Transformation Macdonald and Hankel

System Implementation

The following is a translation of the implementation of the Hankel transform and function Macdonald that contained in the program listing:

```
procedure TForm4.H1Click(Sender: TObject);
Var N,u,x            : integer;
   spektrum : real;
   tau_t,tau_x        : REAL;
begin
   N:=cacah+1;
   N_spektrum:=0;
   T_spektrum:=0;
   for u:=1 to N-1 do
   for x:=1 to N-1 do
begin
   tau_x:=0.01;
```

```
    fungsi_gamma:=0.0;
    gamma_macdonald:=0.0;
    hankel_macdonald:=0.0;
      repeat
        tau_t:=0.01;
      repeat
        fungsi_gamma:=fungsi_gamma+power(tau_x,tau_t-1)*exp(-tau_x);
        tau_t:=tau_t+0.01;
      until
      tau_t>=0.1;
      tau_x:=tau_x+0.01;
      until tau_x>=0.1;
Memo2.Lines.Add(floattoStr(fungsi_gamma));
// TRANSFORMASI HANKEL
 gamma_macdonald:=power(u/2,1+2*x)/x*fungsi_gamma;
 hankel_macdonald := arrayfx[x]* gamma_macdonald*x;
//spektrum
 spektrum:=abs(Hankel_macdonald);
 N_spektrum:=N_spektrum+1;
 T_spektrum:=T_spektrum+spektrum ;
 Memo3.Lines.Add(floattoStr(spektrum));
end;
end;
```

For the system to recognize the type of sound which is inputted as a sample, it first has to do training on any type of noise in order to get the value of the energy possessed by each type of sound. The following is a listing for the training of the four types of sounds sampled:

```
procedure TForm1.RadioButton1Click(Sender: TObject);
begin
    ciri_1:=T_spektrum/N_spektrum;
    namafile := 'ciri_1.txt';
    assignfile(fsinyal,namafile);
    rewrite(fsinyal);
```

```
    write(fsinyal, floattoStr(Ciri_1)+' ');
    closefile(fsinyal);
end;
procedure TForm1.RadioButton2Click(Sender: TObject);
begin
    ciri_3:=T_spektrum/N_spektrum;
    namafile := 'ciri_3.txt';
    assignfile(fsinyal,namafile);
    rewrite(fsinyal);
    write(fsinyal, floattoStr(Ciri_3)+' ');
    closefile(fsinyal);
end;
procedure TForm1.RadioButton4Click(Sender: TObject);
begin
    ciri_4:=T_spektrum/N_spektrum;
    namafile := 'ciri_4.txt';
    assignfile(fsinyal,namafile);
    rewrite(fsinyal);
    write(fsinyal, floattoStr(Ciri_4)+' ');
    closefile(fsinyal);
end;
```

For the testing process used if-else function to compare the results of the energy value obtained:

```
//akhir loading
uji_spektrum:=T_spektrum/N_spektrum;
if uji_spektrum>ciri_1 then hasil_uji_1:=ciri_1/uji_spektrum
            else hasil_uji_1:=uji_spektrum/ciri_1;
if uji_spektrum>ciri_2 then hasil_uji_2:=ciri_2/uji_spektrum
            else hasil_uji_2:=uji_spektrum/ciri_2;
if uji_spektrum>ciri_3 then hasil_uji_3:=ciri_3/uji_spektrum
            else hasil_uji_3:=uji_spektrum/ciri_3;
if uji_spektrum>ciri_4 then hasil_uji_4:=ciri_4/uji_spektrum
```

```
            else hasil_uji_4:=uji_spektrum/ciri_4;
edit3.Text:=floattostr(hasil_uji_1);
edit4.Text:=floattostr(hasil_uji_2);
edit5.Text:=floattostr(hasil_uji_3);
edit6.Text:=floattostr(hasil_uji_4);
end;
end.
```

RESULT AND DISCUSSION

System Analysis

Computational calculation depends on the role of the translation of the transformation Hankeldan MacDonaldke function in code / listings that correct.If wrong then the program does not run as expected. Some reasons to use Hankeldan transformation MacDonald function is as follows:

1.   Has the Gamma function that can convert analog signals into digital signals.

2.   With the combination of these two methods can give better decisions and accurate, which is in the process of execution of the application, the final decision in the form of voice recognition results can be obtained optimally.

User Interface

The voice recognition system consists of three forms, which form the veranda or the main form, the form of training and testing form, both the real-time voice recognition and voice recognition records.

Form veranda is the main form as the initial view of a third form provided on this application. In this form consists of two buttons Real Time option is to choose the voice recognition inputted directly and Record to choose the voice recognition process using a recorded voice samples.

Figure 5. Main Display Voice Recognition System

As a first step in the application user training by using one sample for each type of sound, so the program will show the value of the original signal, the value calculation of the Hankel transform and Macdonald functions, as well as graphics or sound waves from the inputted sound samples. Here is one example of the training undertaken:
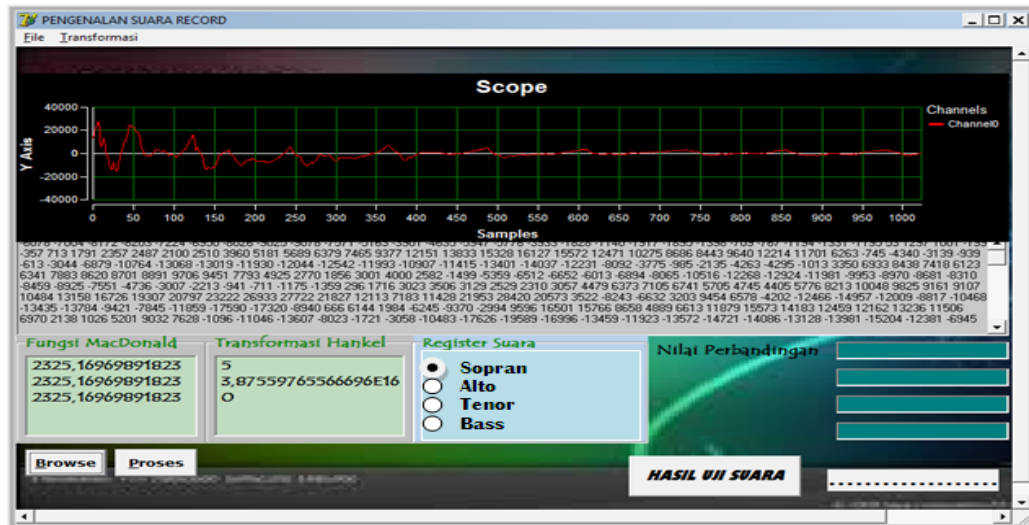
Gambar 6. Proses Pelatihan Suara (*Record*)

After training on a number of sound samples, we doing further testing process for the introduction of any type of noise. This stage serves as the output and the climax of the whole process. After the energy values obtained using the Hankel transform method and function MacDonald in the testing phase, the introduction process is then performed for each type of sound. From these calculations, the energy values obtained to facilitate the introduction of several types of registers voice sound samples entered into the program either in real time or from recordings sound samples. Furthermore, these values will be processed in order to obtain results which indicate the level of success of the introduction of the type of voice registers. Here is one example of the testing process is done:
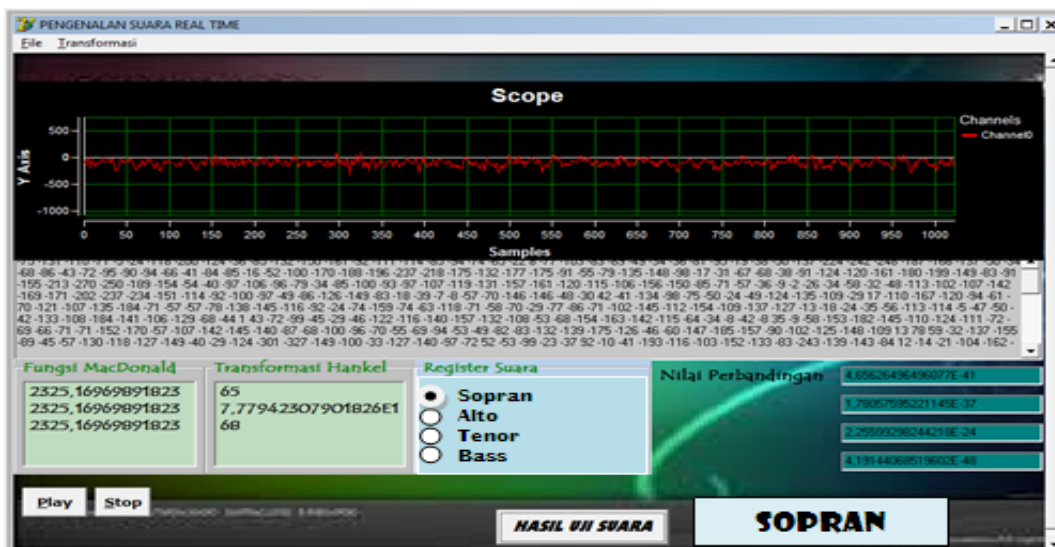

Figure 7. Process Testing Sound (Real Time) Voice Recognition System Performance

The following table combines the sound test results of four types of sound in a single table, with as many as one-time training for each type of sound and using two different training samples with the same sound sample testing of 20 samples of sound. It is only applied to the speech recognition process in the record. For more details we can see in the following table:

Table 5. Performance System To Record Voice Recognition Process

sound samples	amount of test				Correct identification						wrong identification					
	S	A	T	B	S	A	T	B	Tot	%	S	A	T	B	Tot	%
train_1	20	20	20	20	16	12	12	14	54	68	4	8	8	6	26	33
train_2	20	20	20	20	13	12	13	15	53	66	7	8	7	5	27	34

Explanations : S = *Sopran,* A = *Alto,* T = *Tenor,* B = *Bass*

To process in real-time speech recognition, voice recognition results are taken from several different respondents, with as many as one-time training for each type of sound and the use of the training samples for each respondent with sound sample testing as many as 5 voice samples for each type of sound. This is done only to see the results of the comparison between the performances of speech recognition systems in real time with the voice recognition in the record. For more details we can see in the following table:

Table 6. Performance Systems For Real Time Voice Recognition Process

Sound of Samples	Amount of Test				Correct Identification						Wrong Identification					
	S	A	T	B	S	A	T	B	Tot	%	S	A	T	B	Tot	%
train_1	5	5	5	5	5	3	3	4	15	75	0	2	2	1	5	25
train_2	5	5	5	5	5	4	3	3	15	75	0	1	2	2	5	25
train_3	5	5	5	5	5	3	4	3	15	75	0	2	1	2	5	25
train_4	5	5	5	5	3	2	4	3	12	60	2	3	1	2	8	40
train_5	5	5	5	5	3	3	5	3	14	70	2	2	0	2	6	30

Explanations : S = *Sopran,* A = *Alto,* T = *Tenor,* B = *Bass*

Description Table 6 shows that the results of the execution of the program will get the recognition success rate for this type of noise from some sample input. Column 'Voice Samples Pelatihan'memuat some samples of different sounds that serve as training. The column 'Total Testing' too, contains the number of tests that dilakukanuser, then there is the 'Identification True', which contains data about the number of successful pengenalansuara accompanied by a percentage. The latter is the 'Identification of One', the opposite of the 'Identification True', which contains data on the number of voice recognition does not work and the percentage.

From the results of research conducted on several samples of the voice, speech recognition results are done in realtime on average have a higher percentage than the introduction of the sample suaradengan recordings, although the difference was not significant. This is because the sound is inputted in real time clearer and clearer sound than the sound of the recording.

This means sound recording device also affects the performance of the system in the process of speech recognition. But it cannot be concluded that the results of real-time speech recognition entirely more accurate than the results recorded voice recognition.

CONCLUSIONS

The conclusions of this study are as follows:

1. The numbers of samples that are trained not affect the accuracy of the system. The accuracy of the system depends on the accuracy of sound techniques trained and tested. The more accurate sound samples tested and trained, the more accurate the results of the speech recognition process are done.

2. From the value of the training results obtained differences in the frequency of each type of sound making it easier to capture the results of speech recognition. And the voice is used as training samples has significantly high percentage of success of voice recognition.

3. Based on the performance of the system, the process of speech recognition in real time is more effective than the speech recognition process in the record.

4. The method applied in this application reaches the average percentage of the successful introduction of this kind of records sound reaches 66.87%, while the average percentage of the successful introduction of the type of voice in real time reached 72%.

REFERENCES

[1].   Inung Wijayanto, Reni Dwifebrianti. *Jenis Tipe Jangkauan Suara Pada Pria Dan Wanita Menggunakan Metoda Mel-Frequency Cepstral Coefficient Dan Jaringan Syaraf Tiruan Backpropagation.*Fakultas Elektro dan Komunikasi, Insitut Teknologi Telkom Bandung.

[2].   Indrajit Prawira Yudha, Iwan Iwut T, Ratri Dwi Atmaja. *Identifikasi Jenis Suara Manusia Berdasarkan Jangkauan Vokal Menggunakan Jaringan Saraf Tiruan Backpropagation.*Fakultas Elektro dan Komunikasi Institut Teknologi Telkom.

[3].   Maazouzi, F. d. H. B. (2010). *Singing Voice Classification in Commercial Music Productions*. University of Annaba, Algeria.

[4].   Sethares, W. A. (2004). *Tuning, Timbre, Spectrum, Scale: Second Edition. Springer*. Amerika Serikat.

[5].   Prahallad, K. Speech *Technology: A Practical Introduction, topic: Spectrogram, Cepstrum and MelFrequency Analysis*. Carnegie Mellon University & International Institute of Information Technology Hyderabad.

[6].   McLoughlin, I.   (2009). *Applied Speech and Audio Processing: with Matlab Examples*. Cambridge University Press. New York.

[7].   Lismanto. (2010). *Integrasi Numerik dari Transformasi Hankel Menggunakan Metode Kuadratur Gauss*. Tesis. Fakultas Matematika dan Ilmu Pengetahuan, Program Studi Magister Matematika, Universitas Indonesia.

[8].   Pagurova. (2011). *Encyclopedia of Mathematics: Macdonald Function*. *http://www.encyclopediaofmath.org/index.php/Macdonal_function*. 2 Mei 2014,14.20 WIB.