

Voice Recognition and Authentication as a Proficient Biometric Tool and its Application in Online Exam for P.H People

Dwijen Rudrapal

Assistant Professor, Dept. of
Computer Sc. & Engineering
National Institute of Technology,
Agartala
Jirania, Tripura (W). India

Smita Das, S. Debbarma

Assistant Professor, Dept. of
Computer Sc. & Engineering
National Institute of Technology,
Agartala
Jirania, Tripura (W). India

N. Kar, N. Debbarma

Assistant Professor, Dept. of
Computer Sc. & Engineering
National Institute of Technology,
Agartala
Jirania, Tripura (W). India

ABSTRACT

Voice Recognition is a fascinating field spanning several areas of computer science and mathematics. Reliable speech recognition is a hard problem, requiring a combination of many complex techniques; however modern methods have been able to achieve an impressive degree of accuracy. On the other hand, today, most of the companies or institutes are conducting their examinations online to be a part of this best ever growing world. In this system user can give any available examination at any accessible center as per his/her choice and authority also can condense manpower and process delay overhead. This paper offers one way to conduct online examination for physically challenged people who can use their voice only to register and attend the examination. In addition, in the course of this paper it has been tried to authenticate one registered user and to make the authentication process persistent throughout the examination interlude.

Keywords

Butterfly Algorithm, Authentication, Fast Fourier Transform, Discrete Fourier transform Chebyshev's Inequality.

1. INTRODUCTION

Humans have distinctive features that distinguish one person from another. Biometric recognition [1], or simply biometrics, refers to automated methods of recognizing a person based on psychological or behavioral characteristics. The use of biometrics [2] for human identification has a lot of advantages because the features explored are part of the personal information [3] that, in most cases, is impossible to fake, share or forget like, for example passwords or PINs.

The way an individual talk is one of those distinctive features that may be used for recognition. Voice, usually considered as a form of behavioral biometric is in fact a combination of both physiological and behavioral biometrics.

A voice biometric [4] is a numerical model of the sound, pattern and rhythm of an individual's voice. A voice biometric or "voice print," is as unique to an individual as a finger or palm print. Any Authentication application that employs a voice channel during the Authentication session is able to add voice biometric authentication to the process for even higher levels of authentication and security. Voice verification technology uses the different characteristics of a person's

voice to discriminate between speakers. Speech recognition [5] allows providing input to an application with voice.

2. EXISTING ONLINE EXAMINATION SYSTEMS

The manual procedure used for conducting examination is time consuming process and error prone due to human limitations. Online Examination System is mainly used to test the Domain knowledge of the students and employees with respect to a particular technology. For complete automation, the old manual exam procedure is changed to Online Web Based Examination System [6]. All those Online Examination system finds use of mouse and keyboard only, but none of them uses voice as an essential tool. Through this paper, we have proposed a system that uses voice of user for attending the online test. Specially, the physically challenged people, those can't use keyboard and mouse for attending the online examination, will now be able to use their voice for the said test. The proposed system will also authenticate the user through verifying the voice and will find an alternative to interact with the system though out the examination.

3. OVERVIEW OF THE PROPOSED METHOD

As voice of the user is the only verifiable entity to authenticate, the proposed method creates voice templates through registration of the user at the first time and it also authenticates the user by recognition of the created voice template. After recognition and authentication, the system proceeds for next phase i.e. the question answer module. Hence, the proposed system can be separated in the following parts:

3.1 Voice Recognition

Speech recognition is the process by which a computer (or other type of machine) identifies spoken words. Basically, it means talking to your computer, and having it correctly recognized what you are saying. Voice or speech recognition [7] is the ability of a machine or program to receive and interpret dictation, or to understand and carry out spoken commands. In the proposed system only voice will be

interaction tool to a user with the system for registration and verification. For the voice recognition part the following steps have to be followed:

- i) At first, we have to provide the user details as input in the form of voice asked by system.
- ii) The system will then generate a “.wav” file and the generated file will be saved in the database for future references.
- iii) At the time of log in by the user, user needs to provide the same information given at the time of registration and the system compares the recorded voice with the one saved in database. If both match, user logs in successfully, otherwise not.

After successful log in, the set of questions from database appears on the screen which is read by system itself. And the user will speak out the answer as option A, B, C or D which goes to the database where it matches with the previously stored answer for comparison. Based on the comparison result, the system keeps record of the user's score. For correct matching, at the time of registration, system stores user voice for A, B, C and D. These files are compared for matching the questions answer later.

3.2 Comparison of voice recordings.

A human can easily recognize a familiar voice. However, getting a computer to distinguish a particular voice among others is a more tricky task. Immediately, several problems arise when trying to write a voice recognition algorithm. The majority of these difficulties are due to the fact that it is almost impossible to say a word exactly the same way on two different occasions. Some factors [8] that continuously change in human speech are how fast the word is spoken, emphasizing different parts of the word, etc. Furthermore, suppose that a word could in fact be said the same way on different occasions, and then we would still be left with another major dilemma. Namely, in order to analyze two sound files in time domain, the recordings would have to be aligned just right so that both recordings would begin at precisely the same moment.

We start comparison by storing two voices in .wav files. Then we plot both signal and try to match them. Directly comparison done here by producing wav files through sampling which is done by calculating it's Fourier transform. Next we plot it's power spectra and then truncate it to form a new power spectra with differences like noise and height of peaks which is to be normalized resulting in a new power spectra. Using mathematical functions we compute and plot an average power spectrum [9] which is also normalized to compare it with two individual voices giving us desired results.

The voice comparison and analysis in our proposed system has been done in the following way illustrated with an example:

We have recorded the word “hello” twice by the same speaker. The sounds are stored in “test1.wav” and “test2.wav”. We start by reading them into MATLAB vectors Y_1 and Y_2 .

```
Y1=wavread('test1');           Y2=wavread('test2');
```

Next, we plot both signals in one figure. To do this we type

```
subplot(2,1,1);               plot(Y1)
subplot(2,1,2);               plot(Y2)
```

The result is shown in below figure :

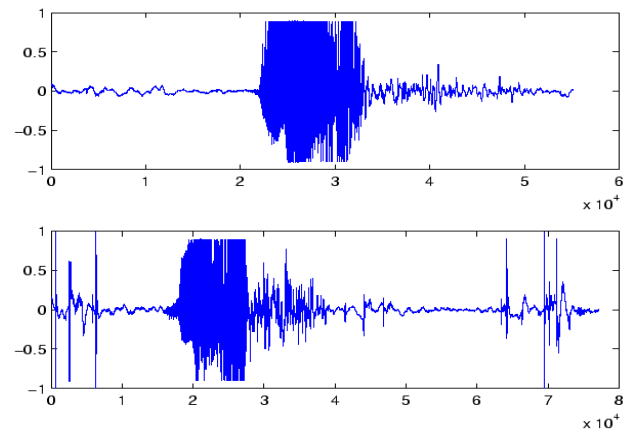


Fig 1: The two signals Y_1 and Y_2

Comparing these graphs we have seen that, the bulk of the signals are at different positions. Matching these positions is difficult. The signal Y_2 starts at approximately 16,000 and ends at 40,000, Y_1 starts at 20,000 and ends at 44,000. So we create two new vectors.

```
X1(1:24000)=Y1(20001:44000);
```

```
X2(1:24000)=Y2(16001:40000);
```

and plot them again using

```
subplot(2,1,1);               plot(X1)
subplot(2,1,2);               plot(X2)
```

The result is shown in below figure :

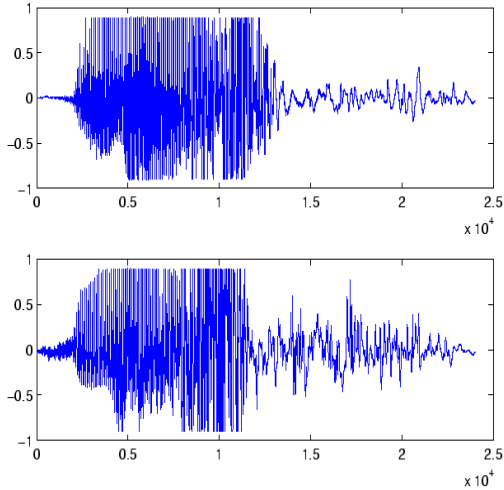


Fig 2: The two ``cleaned-up'' signals X_1 and X_2

Now the two signals match approximately and we can try to compare them directly.

Figure `plot(abs(X1-X2));`

Produces a graph of the difference shown in figure below. As we see in the figure, this can take value up to 1.8 whereas the original signal has only values between -0.9 and 0.9.

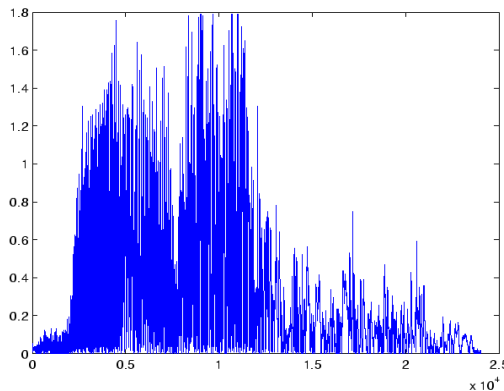


Fig 3: The difference of X_1 and X_2

The problem seems to be that we need to exactly match the two figures. Since this seems to be an impossible task, we continue this using the Fourier transform.

WAV files are produced by sampling 22050 times per second. i.e the vector Y_1 with 55,130 entries represents approximately 2.5 seconds. In order to compare the fast Fourier Transforms of two vectors, they need to be of the same length. Otherwise, we would get a slight shift in frequencies. We will base our length on 60,000, which represents approximately 3 seconds worth of signal. To compute the Fourier Transform based on this length we proceed as

`Z1=fft(Y1,60000);` `Z2=fft(Y2,60000);`

To visualize this we compute the power spectra and plot them as:

```
p1=abs(Z1).^2;      p2=abs(Z2).^2;
subplot(2,1,1);      plot(p1(1:30000))
subplot(2,1,2);      plot(p2(1:30000))
```

The results are shown in below figure:

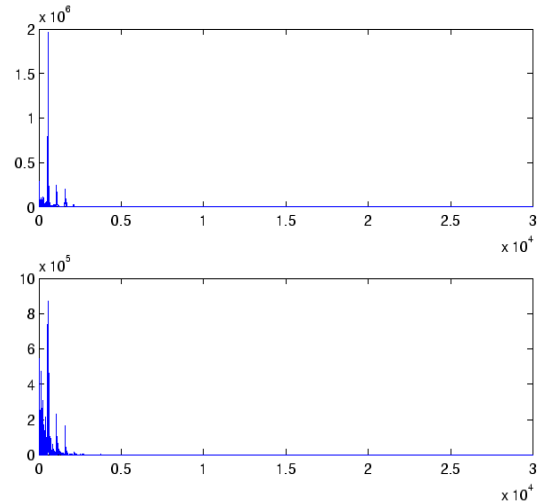


Fig 4: The power spectra p_1 and p_2

We see that in both spectra the peaks are at approximately the same locations. The index k along the X-axis represents the frequency. Since the length of the signal was approximately 3 seconds, $k=500$ corresponds to 499 cycles per 3 seconds or approximately 160Hz. We also see that everything happens for $k < 2500$, which corresponds to frequencies below about 900 Hz. We recall, that the C-major scale ranges from 261 Hz to 523 Hz, and conclude that most action of our vocal cords also happens under 900 Hz. It therefore makes sense to truncate the Fourier Transforms Z_1 and Z_2 at approximately 2500. The new power spectrum is computed and plotted below as:

```
t1=Z1(1:2500);      t2=Z2(1:2500);
q1=abs(t1).^2;      q2=abs(t2).^2;
subplot(2,1,1);      plot(q1(1:30000))
subplot(2,1,2);      plot(q2(1:30000))
```

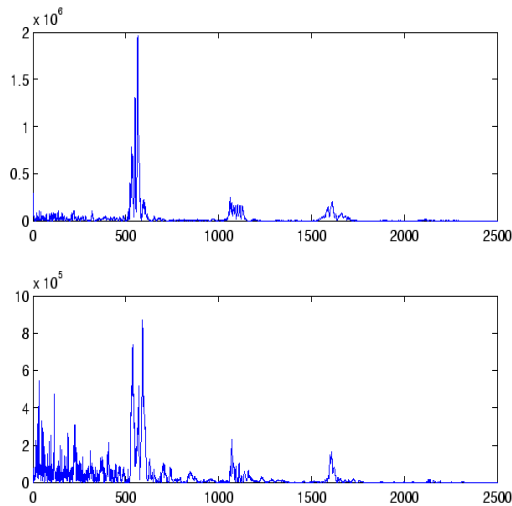


Fig 5: The truncated power spectra p_1 and p_2

We see three major peaks in the two graphs of Figure, which agree in location. However, there are two distinct differences in the two graphs. First, the second graph shows much more in the region for $k < 500$. But, these are Frequencies below 170 Hz. The limit of our hearing is near 100 Hz. This represents low frequency noise and is not necessarily part of the voice. We will filter this out.

```
t1(1:500)=zeros(500,1);
t2(1:500)=zeros(500,1);
```

```
Q1(1:500)=zeros(500,1);
Q2(1:500)=zeros(500,1);
```

The other difference is height of the peaks. This represents in some way the volume. The way around this is to normalize the spectra and plot the new power spectra as:

```
nrm1=sqrt(sum(q1));
nrm2=sqrt(sum(q2));
t1=t1/nrm1;
t2=t2/nrm2;
Q1=abs(t1).^2;
Q2=abs(t2).^2;
subplot(2,1,1)
plot(Q1(1:2500))
subplot(2,1,2)
plot(Q2(1:2500))
```

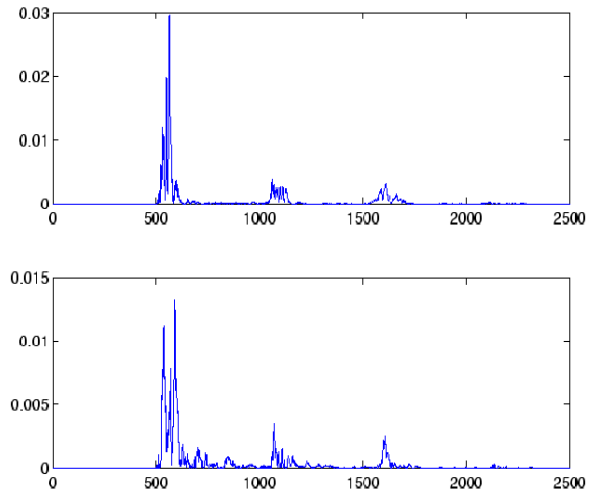


Fig 6: The cleaned up power spectra Q_1 and Q_2

Since, t_1 and t_2 are unit vectors; their inner product gives us the cosine of the angle between them. We can compute this quantity by typing

```
corr=abs(sum(t1.*conj(t2)))
```

Recall that t_1 and t_2 are complex vectors, and the answer is necessarily a complex number. Nevertheless, if this number has an absolute value close to one, these vectors are very "close" to collinear, if it is close to zero they are nearly perpendicular. The result we get is 0.0664.

Our vector space has 2500 complex dimensions, and our voice is not "on-dimensional". Attempting to solve the voice identification problem via linear algebra alone seems to be futile. Not only would we need enough recordings to span the entire range of our voice, this range will probably not form a subspace. Moreover, there will be large overlaps with other people's voices. This brings us to the use of statistics. To continue we compute the "average" spectrum of our two voice samples

```
t=(t1+t2)/2;
```

and plot its power spectrum in figure below:

```
pa=abs(t).^2;
plot(pa);
```

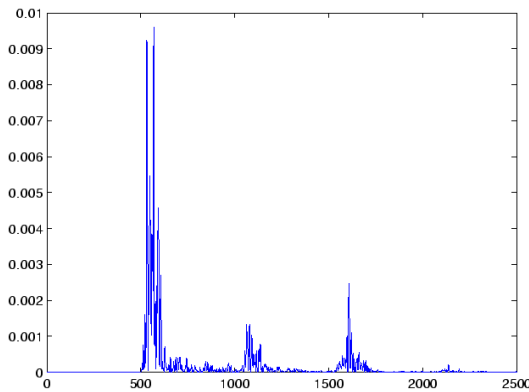


Fig 7: The average power spectrum

We see that this average is rather similar to the individual spectra. We now want to directly compare it to the individual spectra. However, the average is not a unit vector, whereas the individual spectra are. So we first normalize the average,

```
nrma=sqrt(sum(pa));
```

```
t=t/nrma;
```

and then compute the norm of the differences t_1-t and t_2-t ;

```
pd1=abs(t-t1).^2;      pd2=abs(t-t2).^2;
```

```
nd1=sqrt(sum(pd1));    nd2=sqrt(sum(pd2))
```

The differences we compute are $nd_1 = 0.7280$ and $nd_2 = 0.8077$.

Next we want to compare this with someone else's voice. The file "test3.wav" contains the word "hello" spoken by a different voice. We repeat the steps taken for t_1 and t_2 .

```
Y3=wavread('test3');
```

```
Z3=fft(Y3,60000);
```

```
Z3(1:500)=zeros(500,1);
```

```
t3=Z3(1:2500);      p3=abs(t3).^2;
```

```
nr3=sqrt(sum(p3));  t3=t3/nr3;
```

This sequence of commands gives us a unit vector t_3 . To compare it to our average t we compute

```
pd=abs(t-t3).^2;      nd=sqrt(sum(pd))
```

The result is $nd = 1.4154$. Based on this little example we could make a decision on whether or not a voice is the same as the desired voice in the following way:

If $\|t_3 - t\| > 1$ then it is a different voice, if not it is the same voice. Here t is a unit vector in the average direction of the Fourier transforms, and t_3 is the normalized Fourier transform of the test-voice.

3.3 Authentication of voice:

For authentication of voice an efficient spectrum analysis to continuous-time signal based on DFT, the Fast Fourier Transform (FFT) method is adopted. It breaks down DFT of N -points sequence to the one of shorter sequence gradually, which will greatly improve the calculating speed of DFT.

A fast Fourier transform (FFT) is an efficient algorithm to compute the discrete Fourier transform (DFT) and its inverse. An FFT [11] is a way to compute the same result more quickly: computing a DFT of N points in the naive way, using the definition, takes $O(N^2)$ arithmetical operations, while an FFT can compute the same result in only $O(N \log N)$ operations. This huge improvement made many DFT-based algorithms practical; FFTs are of great importance to a wide variety of applications, from digital signal processing and solving partial differential equations to algorithms for quick multiplication of large integers.

The Fast Fourier Transform (FFT), a much faster mathematical algorithm. The FFT eliminates redundant calculations in the Fourier Transform and is therefore much speedier. The FFT is used in speech processing to derive spectra from complex sounds. In speech processing, the FFT converts speech, which is in the time domain, into frequency domain. Because of Fourier developed the mathematical concepts, leading to the FFT, which is essential to voiceprint technology. This lets the user to create wave forms in the frequency domain by altering the amount of energy in each frequency band (oscillation).

We have used the BUTTERFLY algorithm for its efficiency to evaluate discrete Fourier integral operations. Let N be a positive integer, which is assumed to be an integer power of 2 with no loss of generality, and a discrete Fourier integral operator (FIO) with constant amplitude is defined by:

$$u(x) = \sum_{k \in \Omega} e^{2\pi i \Phi(x,k)} f(k), \quad x \in X$$

Where $\{f(k), k \in \Omega\}$ is a given input, $\{u(x), x \in X\}$ is the output and as usual $i = \sqrt{-1}$. Here, the phase function

$\Phi(x, k)$ is assumed to be smooth in (x, k) for $k \neq 0$ and obeys an homogeneity condition of degree 1 in k , namely,

$$\Phi(x, \lambda k) = \lambda \Phi(x, k) \text{ for each } \lambda > 0.$$

In the context of Fast Fourier transform algorithms, a **butterfly** is a portion of the computation that combines the results of smaller discrete Fourier transforms (DFTs) into a larger DFT, or vice versa (breaking a larger DFT up into sub transforms).

4. PROPOSED METHODS ARCHITECTURE

The architecture for the proposed examination system is given herewith with respect to the following data flow diagram and the DFD:

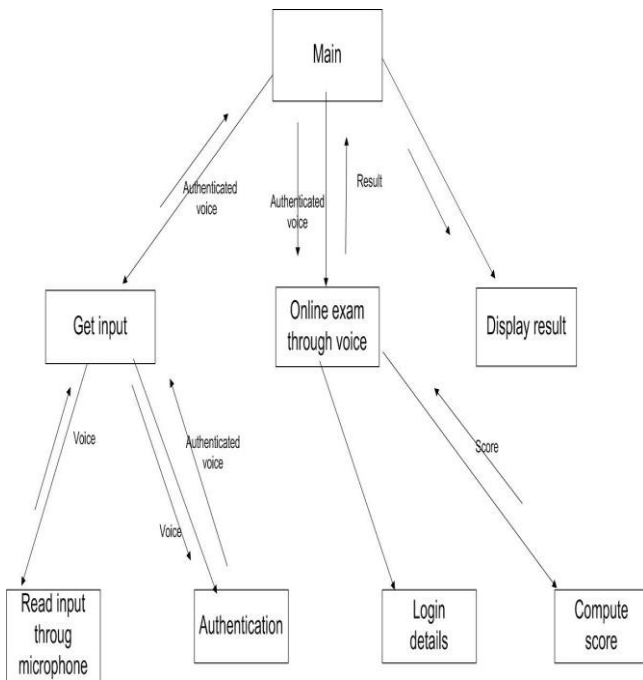


Fig 8. architecture of the proposed system

5. SIMULATION RESULT

In the paper, the full programming simulation is being done using simple Matlab Programming. Below some of the snapshots are given from the Matlab Program which controls the voice driven examination system scheme as a simulator:

After the start of the proposed system it will first registered the user through voice. Then for login it will ask the user to give requisite information. If same voice and information recognized user will logged in as shown in figure 9.

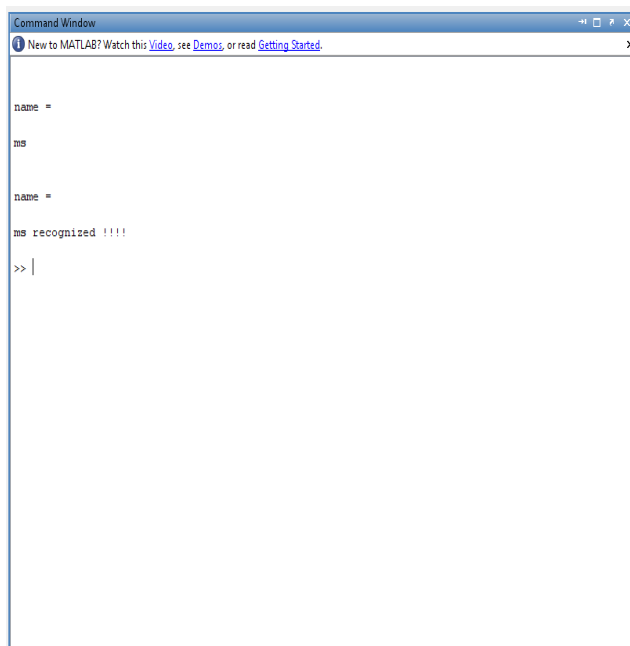


Fig 9. Screen showing user recognized login

If input voice is not matched with the template the system will not recognized and will ask again to login with voice as shown in figure 10.



Fig 10. The screen showing unrecognized user login.

For recognition the system match the frequency spectra of recording and frequency spectra of average as in the figure 11.

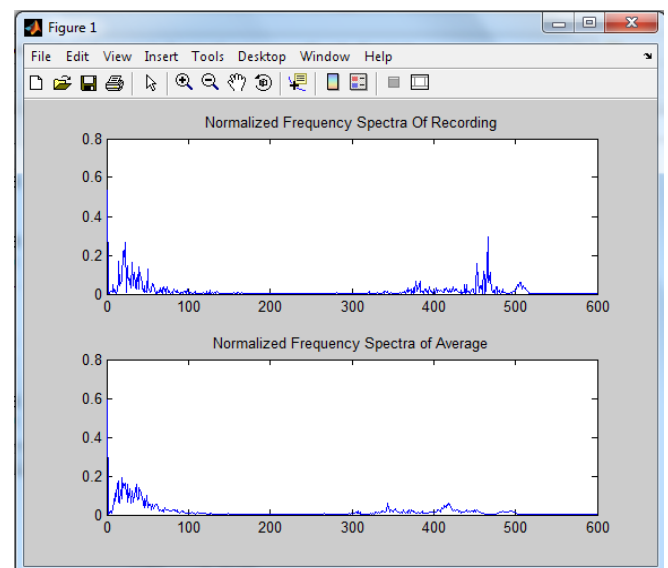


Fig 11. Spectra comparison of the user voice.

Set of questions will appear after successful login of authenticated user as next phase of the proposed system as below figure.

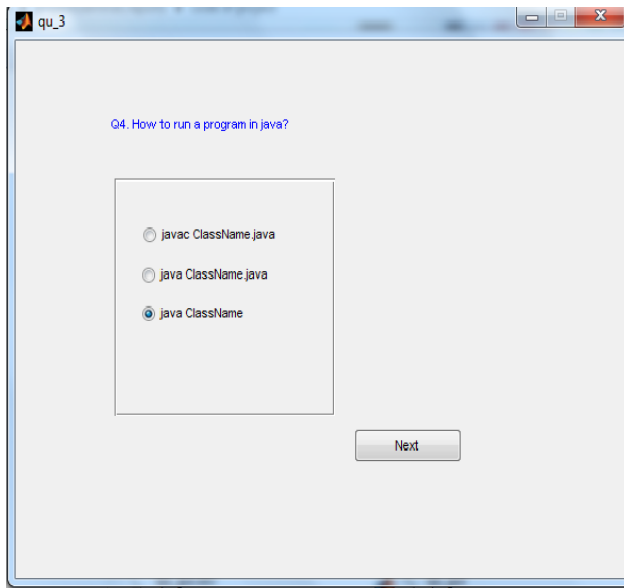


Fig 12. MCQ question screen for online test.

6. CONCLUSION

The proposed system is developed to conduct an Online Examination using voice which will be helpful for people who are blinds and the people who don't use keyboard for interaction with the system.

The proposed method will help for voice recognition where we take voice as input through microphone and then register for Online examination. Then at the time of logging user will logged in to the online examination system through voice authentication, which uses FFT for comparing input voice of the user with the template voice. Next the user will proceed with the questions displayed and read by the system.

7. FUTURE WORKS

In future we are going to do other parts of the proposed system, which is continues authentication of the user throughout the online examination period. Keystroke will be vehicle of verifying the user authentication either randomly matching with user keystroke template or as per the number of times for making the system optimized.

8. REFERENCES

- [1] Anil K. Jain, Arun Ross, Salil Prabhakar "An Introduction to Biometric Recognition" IEEE Transactions on circuit and systems for video technology, Vol 14, No. 1, January 2004.
- [2] N.K. Ratha, J.H. Connell, and R.M. Bolle, "A Biometrics-Based Secure Authentication System," Proc. IEEE Workshop Automatic Identification Advanced Technologies, pp. 70-73, Oct. 1999.
- [3] B. Schneier, "The Uses and Abuses of Biometrics," Comm. ACM, vol. 42, no. 8, p. 136, Aug. 1999.
- [4] C. Kim and R. M. Stern, "Feature extraction for robust speech recognition based on maximizing the sharpness of the power distribution and on power flooring," *IEEE International Conference on Acoustics, Speech, and Signal Processing*, March 2010, Dallas, Texas.
- [5] B. Raj, M. L. Seltzer, and R. M. Stern, "Robust Speech Recognition: The Case for Restoring Missing Features," *Proc. of the Workshop on Consistent and Reliable Acoustic Cues*, September, 2001, Aalborg, Denmark.
- [6] Liang Zhang, Yue-ting Zhuang, Zhen-ming Yuan, Guo-hua Zhan, "A Web-Based Examination and Evaluation System for Computer Education" Proceedings of the Sixth International Conference on Advanced Learning Technologies (ICALT'06).
- [7] Squires, B., Sammut, C.: Automatic speaker recognition: An application of machine learning. In: Proceeding of the 12th International Conference on Machine Learning (1995)
- [8] Barbu, T.: Speech-dependent voice recognition system using a nonlinear metric. *International Journal of Applied Mathematics* 18(4), 501–514 (2005)
- [9] K. Kumar, C. Kim, and R. M. Stern, Delta-spectral cepstral coefficients for robust speech recognition, *IEEE International Conference on Acoustics, Speech, and Signal Processing*, May 2011, Prague, Czech Republic.
- [10] Barbu, T., Costin, M.: Comparing Various Automatic Speaker Recognition Approaches. In: Proceedings of Symposium of Electronics and Telecommunications, ETC 2004, 6th edn. (2004)
- [11] B. N.S. Babu, "Performance of an FFT based Voice coding system in Quiet and noisy environments", *IEEE Transactions on acoustics, speech and signal processing*, Vol ASSP – 31, No. 5, Oct, 83.