Designing Intelligent System for Arabic Instructor Performance Evaluation

Amany F. Elgamal
Dept. of Computer Teacher Preparation
Mansoura University, Mansoura, Egypt

Doaa M. Elbourhamy
Dept. of Computer Teacher Preparation
Kafr El-sheikh University, Kafr El-sheikh, Egypt

ABSTRACT

Instructor evaluation is an important field in the educational process because it develops the level of instructor which can improve the educational level of students consequently. In this work, intelligent techniques are used for instructor performance evaluation in Arabic. This study used group of Arab lecturers through YouTube websitefor testing this system. The proposed system converts instructor's speech (system inputs) to text, and then analyzes the text to extract related knowledge for instructor evaluation that depends on a set of criteria, finally provides advice to the instructor (system outputs). Experimental results demonstrate the effectiveness of the proposed system in instructor performance evaluation. The proposed system can improve reliability and efficiency of instructors' performance; provide the basis for performance improvement that will affect students' academic outcomes.

Keywords
Intelligent System; Instructor Evaluation; Arabic language; speech recognition; artificial intelligence.

1. INTRODUCTION
While the most research concentrated on improving the presentation of understudies and building up the educational program, notwithstanding every one of the components that influence the instructive procedure, there are barely any kinds of research that have been proposed for educator Performance evaluation.

Performance evaluation has been characterized as an efficient procedure of assessing an individual specialist's activity execution and viability in connection to certain pre-set up criteria and hierarchical destinations [1-2].As indicated by Keifer, evaluation is the way toward looking at and rating of a subject, in view of its significant highlights [3]. While, evaluation in instruction can be alluded to as the deliberate assurance of legitimacy, worth, and importance of a learning procedure by utilizing a few criteria against a lot of guidelines or an efficient securing and appraisal of data to give valuable input about some item [4-5].The two definitions concur that evaluation is a precise Undertaking. Also, the term 'item' or 'subject', here, could be a program, arrangement, innovation, individual, need, or action. Be that as it may, the later definition underlines getting and surveying data since all assessment work includes gathering and filtering through information, making decisions about the legitimacy of the data and of derivations got from it.As per Nakpodia, evaluation is an intercession system that has gotten huge consideration in scholarly, business and political circles for data gathering process, finding out the choice to be made, choosing related data, gathering and investigating data to report a synopsis of information helpful for leaders in determination among options [6]. Specialists at various levels have proposed and utilized wide-going ways to deal with assessing educators'

presentation. Notwithstanding, the effectiveness and reliability of old style strategies have been questionable. Thus, there was no standard strategy or modernized answer for assessing educators' presentation that catches the mind boggling nature of the craftsmanship and study of instructing and learning framework in organizations [7-11].

So Instructor Evaluation is a significant field in instructive procedure since it builds up the degree of teacher which can improve the instructive degree of understudies and in the instructive procedure when all is said in done. Ola and Pallaniappan [12] utilized coordinated displaying and smart strategy for an evaluation of teachers' execution in higher foundations of learning, and proposed a perfect count and formed a structure framework which is appropriate for anticipating educators' exhibition.

Ahmadi and abadi [13] broke down the presentation of conclusive instructor evaluation and introduced the result which is cultivated using WEKA (Waikato Environment for Knowledge Analysis) apparatus. It is a prevalent suite of AI programming written in Java, created at the University of Waikato, New Zealand. Information used as a piece of their examination were 104 records on teacher's practices in the study hall with information mining calculations such Association Rule and decision trees (j48). The rightness of rules relies upon an assortment of informational indexes and factual occasions which can change. Be that as it may, information mining devices, for example, WEKA can close assortment results that help training chiefs in colleges. These outcomes will be used by directors as a piece of basic leadership to submit new teachers and continue with picked old instructors.

Mardikyan and Badur [14] directed an investigation to comprehend the key elements influencing the showing execution of the educators and recognizing the components related with the showing execution, during the period 2004-2009. They utilized two distinct information mining systems; stepwise relapse and choice trees. The educator frames of mind are the most significant factor to clarify the teachers' showing execution, which are fundamentally estimated by the evaluation procedure. Likewise, the participation of the understudy is another significant factor that effects emphatically the presentation of the teacher. Subsequently, the teachers that pull in more understudies to the classes are evaluated even more successfully.

On different hands basic reason for good practice in assessment is the presence of clear and quantifiable criteria which must be reliably applied by capable (prepared and experienced) evaluators [15]. This requires the advancement of express rules about what is normal from proficient practice. These criteria are relying upon the expressions of the speaker during his talk. The words are characterized into classes, for example, (positive concepts, communication with students,

examples, jokes, content of lecture). Among the methods used to remove words from the teacher's discourse (his lecture), is the keyword spotting system.

Keyword spotting system (KWS) is an innovatively relevant issue, which assumes a critical job in sound ordering and discourse information mining applications [3]. KWS is additionally utilized for finding events of the watchword in discourse signal [4]. This issue resembles discourse acknowledgment, yet the extra flag around the expressions of intrigue must be dismissed. In light of that is viewed as keyword spotting system is to perceive the nearness of a little arrangement of pre-decided words in a persistent stream of discourse. The procedure incorporates perceiving picked watchwords in discourse articulations containing unessential (out of vocabulary) discourse and commotion [5].

Keyword spotting techniques can be separated into two classes: speaker dependent and speaker independent. speaker dependent systems, models are made for a specific speaker and in this way are not planned to work for various speakers. Speaker independent systems should be progressively nonexclusive and therefore require increasingly complex structure. Discourse catchphrase spotting frameworks will in general be considerably more confounded as a result of the enormous variety of articulations, even from a similar speaker, contingent upon the specific circumstance and temperament of the speaker [17].

In this work, the proposed system uses keyword spotting system (speaker dependent) for Arabic instructor evaluation and gives a set of advices for improving the performance of instructor. The keyword spotting system has been designed by using MFCC and Hamming Distance (HD) technique for Continuous Speech Recognition of Arabic Language.

2. METHODOLOGY

The proposed framework involves the accompanying parts: A. keyword spotting system; B. instructor evaluation by set of criteria, as appeared in Figure 1.
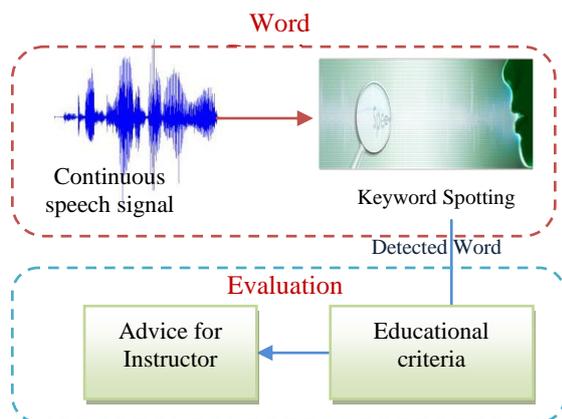

Fig 1. Block diagram of proposed system

2.1 Keyword Spotting System

The keyword spotting system used to remove the nearness of a little arrangement of pre-decided words in a ceaseless stream of discourse. The keyword spotting system comprises of two stages: (1) preparing stage; (2) testing stage. The accompanying part delineates the preparation stage in subtleties as shown in Figure 2.
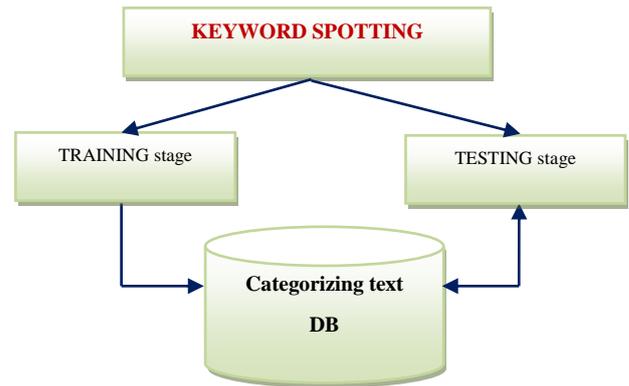

Fig 2. Keyword spotting system

2.1.1 *Preparing stage*

Preparing phase includes a set of steps as shown in Figure 3, which illustrates a block diagram of this phase.
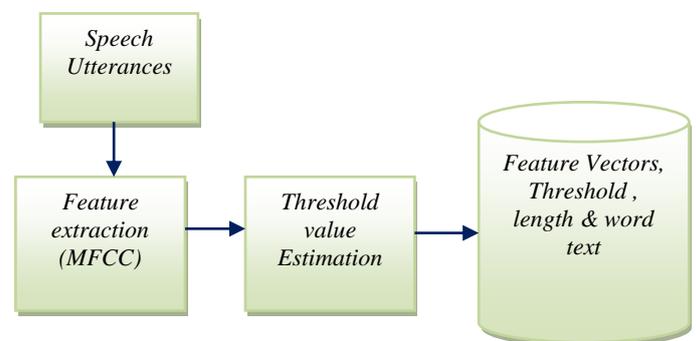

Fig 3. Block diagram of preparing stage

2.1.1.1 *Feature extraction – Mel Frequency Cepstral Coefficient (MFCC):*

The explanation of this module is to change over the speech waveform to some kind of parametric portrayal (at an impressively lower data rate) for further examination and handling. This is regularly alluded to as the sign handling front end. The discourse sign can be characterized as a moderate time shifting sign (it is called semi stationary). At the point when it was inspected over an adequately brief timeframe (somewhere in the range of 5 and 100 millisecond), we found that its qualities are genuinely stationary. Be that as it may, over significant stretches of time (on the request for 1/5 seconds or more) the sign attributes change so as to mirror the diverse discourse sounds being spoken. In this manner, the brief timeframe ghastly investigation is the most widely recognized approach to depict the speech signal [18]. Truth be told, MFCC is widely utilized in discourse acknowledgment framework predominantly for three reasons: (1) the cepstral highlights are pretty much symmetrical as a result of the Discrete Cosine Transform (DCT), cepstral mean division, (2) to kill static channel commotion, and (3) it is a lesser measure of delicate to added substance clamor than licensed Professional Clinical Counselor (LPCC) [19-25].

As shown in Figure 4, MFCC consists of six computational steps. Each step has its function and mathematical approaches as discussed briefly in the following [20-23]:
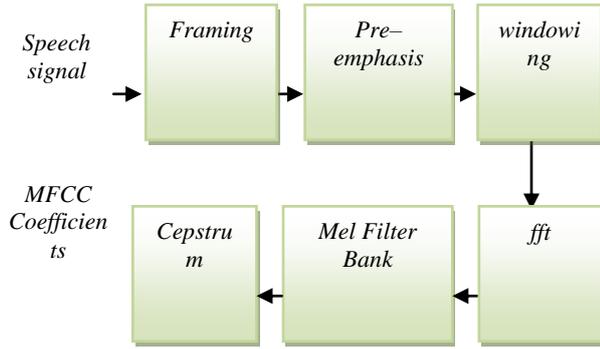
Fig 4. Block diagram of MFCC process

After the finish of highlight extraction process, which is the initial segment of the preparation stage, the information was spared in the database to be utilized in limit esteem assessed; which considered as a second piece of the preparation stage, the accompanying area shows the threshold value estimated in details.

2.1.1.2 Threshold value estimation:
Legitimate thresholding is imperative to keep up the exchange off among review and accuracy. Utilizing one threshold of the limit for all watchwords brings about in general lackluster showing. Subsequently, the threshold for every keyword is diverse [26].

The threshold is resolved for each word using 10 utterances to single word from 10 unique speakers, for example, the web, and two distinct words, for example, as and look. These words were picked due to their terms being roughly the equivalent. The objective was to decide a limit of disparity for perceiving the watchword web ward of the speaker from the above rundown of 12 expressions. This edge must yield as hardly any bogus positive and bogus negative acknowledgment scores as workable for the chose set of articulations. When the edge set up for a catchphrase, the equivalent can be utilized for recognizing its essence among different expressions [27].

The Hamming Distance is a number used to signify the distinction between two double strings. It is a little bit of a more extensive arrangement of recipes utilized in data investigation. In particular, Hamming's equations enable PCs to identify and address mistake individually. Hamming's augmentations to data innovation have been utilized in such advancements as modems and smaller plates. [28]: The Hamming distance between two vectors, $\mathbf{A} = [\mathbf{a_1 a_2} \dots \mathbf{a_N}]^\mathbf{T}$, and, $\mathbf{B} = [\mathbf{b_1} \ \mathbf{b_2} \dots \mathbf{b_N}]^\mathbf{T}$ of N elements, each was carried out for each pair of normalized features. TheHamming distance between A and B given by:

$$\mathbf{D_{st}} = (\mathbf{A}.\mathbf{B})/\mathbf{N} \qquad (1)$$

Evaluation criteria used in the proposed system are: examples, communication, good concepts, jokes and content of lecture. Sample of words in each category shown in table 1.

Table 1. Category of words

Category	Samples
Positive concepts	شكرا, جيد
Communication	نتعرف , ما هى على
Examples	يعنى, مثل
Jokes	نكته, مضحك

The length of each utterance was calculated using the length function and the resulting data stored in the database. Table 2, Shows the list of the used attributes.

Table 2. List of attributes

Variable Name	Data Type	Description
WordID	Number(integer)	ID of words
WordName	Text(char)	Word name of utterances
WordLength	Number(integer)	number of samples in each word signal
CategoryWordID	Number(integer)	Category words ID
ThresholdValue	Number(decimal)	Threshold value for word
CourseName	Text(char)	Course name
LectureID	Number(integer)	Lecture ID
Duration	number(decimal)	Duration of each lecture
GoodWord	Number(integer)	Number of positive word
ExampleWord	Number(integer)	Number of example word
JokeWord	Number(integer)	Number of joke word
CommunicationWord	Number(integer)	Number of communication word

2.1.2 Testing stage
This section reviews the testing phase, which consists of a set of steps, as shown in Figure 5.
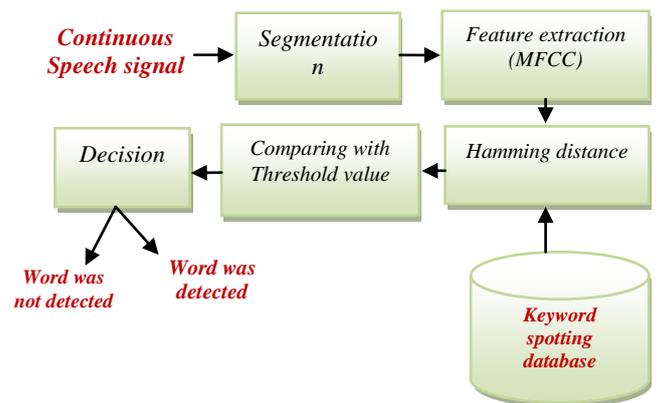

Fig 5. Block diagram of testing phase

2.1.2.1 Segmentation
One of the major problems that affect the efficiency of a speech recognizer is detecting the start and end points of voice activity. However, short-term power and zero crossing rate are commonly used as parameters for distinguishing speech/non-speech regions [29], [30].

In our proposed system, a method based on Voice activity detection (VAD) is used to compute short-term power and zero crossing rate. The recognized keywords have been used to evaluate instructor. Where, the number of words in each category is used to compute the percentage for each criterion.The percentage of each criterion was computed using the equation below:

$$Percentage = \frac{number of keywords recognized in each criterion \times 100}{default number of lecture keywords for this criterion in DB} \quad (2)$$

Moreover, the final percentage of all criteria is defined as:

$$Final\ percentage = \frac{sum of criteria elements percentage}{5} \quad (3)$$

A. Instructor performance evaluations

Modern educational organizations start developing and enhancing the educational system increasing their capability to instructors' performance evaluation. Instructor evaluation incentives the basis for school or university reforms has recently attracted considerable attention and support from researchers and policy makers [31-32]. In this section, instructor evaluation is done through a set of criteria. as shown in Figure 6.
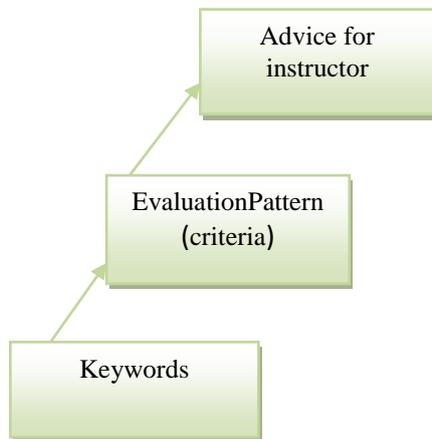

Fig 6. Instructor evaluation system

2.1.2.2 Evaluation Pattern (criteria)

Powerful training starts with a comprehension of what individuals resemble, how they carry on and learn, and what issues they face. The region is resolved to graduate proficient grown-ups who are independent, self-trained and deep rooted students. In its initial applications, colleges and universities utilized the aftereffect of assessments to screen showing quality and to assist educators with improving their showing viability [33].In this investigation, teacher assessment is prepared by set of criteria that have expressed in [33-36]:

- Using appropriate example and illustrations (examples).
- Communicating effectively with the students (communication).
- Implementing the lesson plan effectively (content of lecture).
- Respecting and concerning with students (positive concepts).
- Using sense of humor (jokes).

Before the finish of this segment, the proposed framework Gives the educator a lot of counsel which depend on the assessment criteria to improve his presentation.

3. EXPERIMENTAL WORK

The Tests incorporates 400 sounds from two teachers (speaker dependent system), these sounds comprise of 80 catchphrases, and each word has 5 articulations from 5 unique speakers. The proposed framework is executed utilizing: MATLAB R2012a for discourse preparing and structures the Graphical User Interface (GUI), Microsoft SQL for making the database and WavePad Sound Editor for sound altering and split the sound in the preparation stage.

The system can interact with instructor through GUI to enter the lecture data in the database. The data includes lecture name, lecture duration, the number of keywords that must be present in a lecture for each criterion, according to the requirements of each lecture. The accuracy of the keywordspotting algorithm computed for each keyword using the equation below.

$$Accuracy = \frac{hits}{hits + false\ negative + false\ positive} \quad (4) \mathbf{Where}$$

Hits: a keyword is detected where it is actually present, **False negative**: a keyword is not detected even where it is present, **False positive**: a keyword is detected where it is not present [17].

The detection rate varies from keyword to keyword. Below is a table illustrates the results obtain in experiments carried out with different keywords.

Table 3. Accuracy for Different Keywords

Keyword	Accuracy
لغة HTML	95%
نتعلم	90%
هل يعنى	89%
كيف	84%
ماذا نعنى	80%
نكتبها	72%
ممتاز	70%
اساسياتها	70%
عفوا	68%
اتفضل	68%
مثل	68%
سؤال	61%

GUI of the testing phase, that enables users to select the audio file of lecture and determine course name then use test button to display the output results.The instructor was evaluated according to the evaluation criteria as shown in figure 7.

By pressing the "evaluation" button, the evaluation screen is displayed containing two sections. First section presents instructor information such as (name, lecture name). The second section presents percentages for each evaluation criterion and instructor final evaluation percentage. Also there are translate button to translate instructor evaluation to English language for the non-Arabic speaker.

Fig 7. GUI of instructor evaluation

Fig.8 show the translate screen for instructor evaluation.

Fig 8. translat of instructor evaluation screen

Fig.9 shows a comparison of two lectures for one instructor to check the veracity of the instructor evaluation. The instructor's percentage in each criterion is saved in the database.
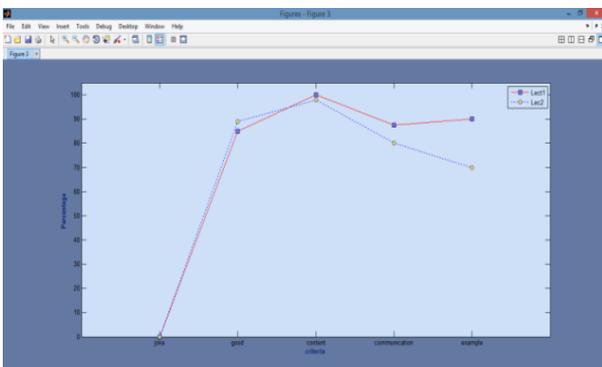

Fig 9. comparison of two lectures

After testing some lectures for each instructor, the GUI of instructor evaluation will displays "show advice" button to produce advice for instructor as illustrated in figure 10.

Fig 10. GUI of instructor evaluation with "النصائح" button

by pressing "show advice" button, a set of advice for the instructor is displayed, also we can select another instructor for displaying his advice. Fig.11, shows advice GUI of each instructor

Fig 11. GUI of advice

Also in fig. 8 we can press "show advice" button, a set of advice for the instructor is displayed but by English language.

Fig 12. GUI of advice by English language

4. CONCLUSION

In this paper, the proposed system for Arabic instructor performance evaluation was produced. The system recognizes keywords from continuous speech; classify keywords and evaluates instructor by a set of evaluation criteria. The system provides the instructor with a set of advice for improving his performance.On the other hand, using Hamming Distance (HD) technique has reduced the actual processing time in the test phase.

It is worthy to say that evaluation will optimize students' academic outcomes and improve the standard of education. Consequently, this will contribute to the achievement of the goals and objectives whichare defined in the vision and mission of the new education reform agenda.

5. REFERENCES

[1] S. Mardikyan, B.Badur,."Analyzing Teaching Performance of Instructors Using Data Mining Techniques", Journal of Informatics in Education, Vol. 10, No. 2, PP.245–257, 2011.

[2] A. Doleh, D. Weir,."Dimensions of Performance Appraisal Systems in Jordanian Private and Public Organizations", International Journal of Human Resource Management, 18(1), PP.75-84, 2007.

[3] K.Keifer,."A Definition of Evaluation", Colorado State University, 2016. Copyright © 1993-2016. Available online at:

[4] http://writing.colostate.edu/about/copyright.cfm

[5] P. Bramley,."Evaluating Instructional Design", 2015., Retrieved from: http://www.nwlink.com/~donclark/hrd/sat6.html

[6] W. Trochim,."Introduction to Evaluation". Web Center for Social Research Method, 2006. Retrieved from http://www.socialresearchmethods.net/kb/intreval.htm.

[7] E.D. Nakpodia,."A Critique of the Methods of Evaluating the Competency of Lecturers in Nigerian Tertiary Institutions", African Journal of Education and Technology, Vol. 1, No. 1,pp. 53-59,2011.

[8] S. Glazerman, et al.."Evaluating Teachers: The Important Role of Value"- Added. The Brookings Brown Center Task Group on Teacher Quality, Pg 2, 2010.

[9] I. A. Archibong, M.E. Nja,." Towards Improved Teaching Effectiveness in Nigerian Public Universities: Instrument Design and Validation". Journal of the Higher Education Studies, Canada. Vol. 1, No. 2, 2011.

[10] R.L. Rothstein, et al. "Problems with the Use of Student Test Scores to Evaluate Teachers". A Publication of Economic Policy Institute, 2010. Retrieved from: http://www.epi.org/publication/bp278/

[11] O.K.Chaudhari, et al.."Soft Computing Model for Academic Performance of Teachers Using Fuzzy Logic". British Journal of Applied Science & Technology 2(2): PP.213-226, 2012.

[12] A. Ola, andS. Pallaniappan, "A data mining model for evaluation of instructors' performance in higher institutions of learning using machine learning algorithms", International Journal of Conceptions on Computing and Information Technology, Vol. 1, due 2; ISSN: 2345 – 9808, 2013.

[13] F. Ahmadi and S. Abadi, "Data Mining in Teacher Evaluation System using WEKA", International Journal of Computer Applications, Vol. 63 – No.10 , 2013.

[14] S. Mardikyan and B. Badur, "Analyzing Teaching Performance of Instructors Using Data Mining Techniques, Informatics in Education", Vilnius University ,Vol. 10, No. 2, pp.245–257, 2011.

[15] J. H. Stronge,."Teacher Performance Evaluation Program Handbook", Fairfax County Public Schools Department of Human Resources, 2012.

[16] P. Karmacharya, "Design of Keyword Spotting System Based on Segmental Time Warping of Quantized Features", Master thesis, the Temple University Graduate Board, 2012.

[17] S. Kumar and M. Rao ,"Design Of An Automatic Speaker Recognition System Using MFCC, Vector Quantization And LBG Algorithm", International Journal on Computer Science and Engineering (IJCSE), ISSN: 0975-3397 ,Vol. 3 No. 8, 2011.

[18] K. Darabkh, A. Khalifeh, B. Bathech and S. Sabah, "Efficient DTW-Based Speech Recognition System for Isolated Words of Arabic Language", World Academy of Science, Engineering and Technology, Vol. 7, 2013.

[19] L. Muda, M. Begam and I. Elamvazuthi, "Voice Recognition Algorithms using Mel Frequency Cepstral Coefficient (MFCC) and Dynamic Time Warping (DTW) Techniques", Journal of Computing, Vol. 2, ISSUE 3, March 2010, ISSN 2151-9617.

[20] D. Harjani, M. Jethwani2 and M. Roja, "Speaker Recognition System using MFCC and Vector Quantization Approach", IJSRD - International Journal for Scientific Research & Development, Vol. 1, Issue , 2013.

[21] A. Thakur and N. Sahayam, "Speech Recognition Using Euclidean Distance", International Journal of Emerging Technology and Advanced Engineering, Vol. 3, Issue 3, 2013.

[22] N. Kaberpanthi, A. Datar, "Speaker Independent Speech Recognition using MFCC with Cubic-Log Compression and VQ Analysis", International Journal of Computer Applications , Vol. 95– No.26 , 2014.

[23] N. Singh, R. Khan and R. Shree, "MFCC and Prosodic Feature Extraction Techniques: A Comparative Study", International Journal of Computer Applications (0975 – 8887), Vol. 54– No.1, 2012.

[24] E. Chandra, K. Manikandan, M. Sivasankar, "A Proportional Study on Feature Extraction Method in Automatic Speech Recognition System", International Journal of Innovative Reseach in Electrical, Electronics, Instrumentation and Control Engineering,Vol. 2, Issue 1, 2014.

[25] A. Sangwan, J. Hansen, "Keyword Recognition with Phone Confusion Networks and Phonological Feature based Keyword Threshold Detection", Center for Robust Speech Systems (CRSS), Department of Electrical Engineering, the University of Texas at Dallas, Richardson, Texas, U.S.A, 2010.

[26] K. Gopalan, T. Chu and X. Miao, "An Utterance Recognition Technique for Keyword Spotting by Fusion

of Bark Energy and MFCC Features". the Air Force Research Laboratory, Rome, NY, U.S.A, 2009.

[27] R. Fooprateepsiri, W. Kurutach, "A Fusion of Trace Transform and Hamming Distance with Multiresolution Technique for Improved Accuracy Approach Face Based Identification", JOURNAL OF INFORMATION SCIENCE AND TECHNOLOGY , VOL 2 ,ISSUE 2 , 2011.

[28] M. Alam, P. Kenny, P. Ouellet, T. Stafylakis and P. Dumouchel, "Supervised/Unsupervised Voice Activity Detectors for Text dependent Speaker Recognition on the RSR2015 Corpus", 1Centre de recherché informatique de Montréal, Montréal, Canada, 2015.

[29] M. Nilsson and M. Ejnarsson, "Speech Recognition using Hidden Markov Model (performance evaluation in noisy environment)", Master Thesis, Department of Telecommunications and Signal Processing, Belkinge Institute of Technology, Ronneby, Sweden, 2002.

[30] P. Veeramuthu and R. Periasamy, "Application of Higher Education System for Predicting Student Using Data mining Techniques", International Journal of Innovative Research in Advanced Engineering (IJIRAE), Vol. 1 Issue 5, 2014.

[31] A. Pal and S. Pal, "Evaluation of Teacher's Performance: A Data Mining Approach", International Journal of Computer Science and Mobile Computing, IJCSMC, Vol. 2, Issue. 12, PP.359 – 369, 2013.

[32] C. Adams, K. Nolan, "Teacher Performance Evaluation: Building Tomorrow Today", Jackson Public School District Earl Watkins, Ph.D, 2011.

[33] C. Leong and Y. Lee, "Mining sentiments in SMS texts for teaching evaluation", Expert Systems with Applications 39, PP.2584–2589, 2012.

[34] J. Stronge, "Teacher Performance Evaluation Program Handbook", Fairfax County Public Schools Department of Human Resources, 2012.