Enhancement Electronic evaluation for Semantic Arabic Oral Exam

Eman K. Elsayed[1], Kamal A. ElDahshan[2] and AlshimaaAbdelraoof[3]

[1]Dept. Math and Computer Science, Faculty of Science (girls), Al-Azhar University, Egypt
Email:*emankaram10 @hotmail.com*
[2] Dept. Math and Computer Science, Faculty of Science, Al-AzharUniversity,Egypt
Email:*dahshan @gmail.com*
[3] Dept. Math and Computer Science Faculty of Science (girls), Al-Azhar University, Egypt
Email:*alshimaa22 @gmail.com*

Abstraction:

From the importance of knowledge in the speech, we knew the importance of oral exam. So in this paper we integrated BOW (Bag of Word), LSA(Latin Semantic Analysis), ASR (automatic speech recognition), zero crossing rate, and Ontology based approach to automate the online oral exam especially in Arabic language with take into consideration the authentication problem. Our proposal method faced many challenges in Arabic language because there isn't semanticdictionary like WordNet in English and HowNet in Chinese. Also Arabic language has complicated synonyms. Our proposal can help improving meaningfulness. Finally, the proposed method in this paper didn't forget automation thefeedbackfor determining learning disability.

Council for Innovative Research

Peer Review Research Publishing System

Journal: INTERNATIONAL JOURNAL OF COMPUTERS & TECHNOLOGY

Introduction

Online scoring exam in E-learning system provide a great benefits not only to the teacher but also to the student, the teacher will be able to finish evaluating in a few of minutes and the student can easily know his degree immediately and get a feedback for his score ,also the online scoring minimize the cost on e-learning system for example in Britain the researcher found that teacher spend about 30% of their time in evaluating and grading student answer which cost 3 billion pounds per year[1].

But till now only the types True/False, Multiple answer, Ordering, Matching and Fill in the blank are supported online scoring in popular e-learning application like Moodle and Blackboard.

Oral exam is defined as assessment of student knowledge levels through a face-to-face dialogue between the student and examiner. Although, the importance of oral type of exam is for all students specially blind one and the people who don't use keyboard for interaction with the system. There are many problems in online oral exam not solved until now, especially in Arabic Language as needing to coordinateduring the various phases of development. Also oral exam istypically take at least 20 minutes per student [2]whichcostly in terms of time.

in our method we try to solve the problems in oral test for both reading comprehension and open question by making the oral test automatically graded so we don't need any coordinator ,Also we provide a smart feedback for both student and teacher by adding comments according to the student grade and apply some measurement to the recording answer to provide teacher with information about the student reading skills, in the same time, automatic grading for oral test will save time and reduce cost, Finally due to the student pressure we apply solve defects pronunciation in pre-processing stage to overcome student pressure problems.

The paper is organized as follows. After the introduction section two for main concepts section three to present literature review,and then section four for proposal methodology, then we applied and analyzed our proposal in section five. Then we displayed the conclusion and further work directions. Finally we listed the references.

2 Main concept:

LSA(Latin Semantic Analysis):

LSA is a fully automatic statistical IR(Information Retrieval) technique that was originally designed for indexing documents and text retrieval. It is not a traditional natural language processing or artificial intelligence program; automatic dictionaries, knowledge bases, semantic networks, grammars, syntactic parsers, and it takes as its input only raw text parsed into words defined as unique character strings and separated into meaningful passages or samples such as sentences or paragraphs.[3]

Zero crossing rate:

Zero crossing rate is a measure of number of times in a given time interval or frame that the amplitude of speech signal passes through a value of zero. The rate at which zero crossing occurs is a simple measure of the frequency content of the signal. This feature is very useful for analysis and segmentation of the speech signal. [4]

ASR(Automatic Speech Recognition):

Speech Recognition is also known as Automatic Speech Recognition or computer speech recognition. It is the process of converting a speech signal to a sequence of words, by means of an algorithm implemented as a computer. Speech recognition technology has made it possible for computer to follow human voice commands and understand human languages.[5]

Smart feedback:

Feedback was developed to support teacher and to give the score more meaning for student. However, learning without supportive feedback is unlikely to produce a noticeable learning effect. Unfortunately, frequent and individualized feedback is hard under real life school conditions,so smart feedback make the e-learning exam more effective**.**

Learning disabilities:

Learning disabilities (LD) are neurological disorders. LD make difficult to acquire certain academic and social skills for learner[6]. Oral Reading Fluency (ORF)is monitoring process which using reading text quickly, accurately, with good prosody, and with comprehension to determining LD type.

3 Literature review:

Eman K Elsayed[7]etalproposeduse a biometric voice solution. That is to solveoral multi model authentication and to enhance "Dynamic Time Wraping" (DTW) method. That is for scoring oral short answerbasedon(Mel-Frequency Cepstral Coefficient)MFCC and (Vector Quantization)VQ algorithm with fuzzy technology to enhance the security of the online exam .the proposed method gave a good result only for short answer.

Jian Cheng etal [8] represent an automatic scoring test of facility with spoken Modern Standard Arabic (MSA) ,the research perform test yields an ability profile over four sub scores, Fluency and Pronunciation as well as Sentence Mastery and Vocabulary (content) using a total of 246 hours of speech in response to the test items was collected from

natives and learners and was transcribed by educated native speakers of Arabic. But in this reference the answer must be only short text and the system can't score the answer by meaning and also can't score any words from any delicate.

M. M. Refaat and A. A. Eweesetal [1] present an automated assessment of students' Arabic free-text answers, in this research grade Arabic essays automatically based on Latent Semantic Analysis and Cosine Similarity based in special data set and the result was hopeful so in our research we depend on Latent Semantic Analysis (LSA) and Cosine Similarity and enhance the result by adding the stemming of the word according to Arabic dictionary.

Gomaaetal in reference [9]presents an Arabic short answer scoring with effective feedback for students, in this research new data set created with 50 question and 12 answers for each question with total 600 answer the scoring measured by comparing student answer with model answer ,also the research define feedback comment .

4The methodology:

In this paper,we proposed a fully automatic solution to solve Arabic oral exam with free answer. That is by solving many problems as defects pronunciation andsynonym. We also define a new feedback algorithm to make the oral test more meaningful for both teacher and student.

Our proposal method applied on Arabic language, so we can't simply use Arabic language dictionary.

For many reasons that is as the meaning of the words dependent on the subject. We classify the Arabic dictionary words as a bag of words for all lessons in learning object for certain course. Each bag has the words and its derivative with the corresponding meaning. Where ontology based structure formed by 4 tuples (C, R, I, A), in which C is a concept, R is a relation, I is the instance and A is the axiom. Axiom is used to provide information about the classes and properties. Where in semantic e learning each chapter is class and has lessons as subclasses which have bag of words as instances.
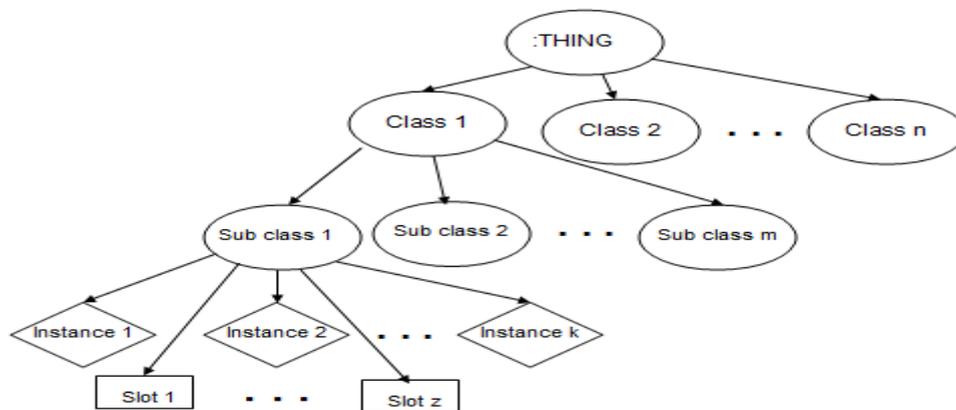
Fig1: ontology components

So Using Ontology base which define as an artificial intelligent knowledge representation. That is more useful than using database. That's for take advantage of using Ontology base in e learning systems which called new generation of e-learning systems.

The learning object Ontology base created to contain the words and its derivative as instances in the lesson class. That organized data will minimize the time for searching and then the time of scoring.

4.1) the system starts by authentication process where the security isn't often reliable enoughin e-learning packagesso we should hence strive to make online exams secure and non-costly.In our system, we concentrated on enhancing the security system using a multi-model technique that depends on a regular login system and a biometric algorithm. This will be more appropriate for voice testing, making it ultimately and more accurate method. The algorithm works online. The proposed algorithm ameliorates MFCC (Mel-Frequency Cepstral Coefficient), which is used to isolate the unique features of a human voice. It represents the short term power spectrum of the human voice. Also in the proposal Algorithm VQ (Vector Quantization) is an important aspect of data compression. The purpose of data compression is to reduce the bitrate, in order to minimize requirements of communication channel capacity or digital storage memory, all while maintaining the requisite fidelity of the data. VQ divides a large set of vectors into multiple groups: each group is represented by its own centroid point.When learnersexceed this step the system will send the oral question to them site and receive fromthem online solutions.

4.2) Then the system convert automatically the long Arabic speech to text by ASR with HMM (Hidden Markove Model) by taking into consideration dialects in natural Arabic language. Using comparing with audio files is suitable only in short answers. But in case of long answer that is method consume time so automatic transcription systems are necessary for enabling effective scoring to the student answers.

4.3) Now, the system will solve defects pronunciation in Arabic for example bumbling; a repetition of speech. One of the defects pronunciation that may occurs in some student answers due to worrying from the exam these problems may effects in the scoring results.

4.4) Then the system automaticallyremove Stopwordsin Arabic language.

4.5) Then we have extracted the root of the main words which called "Stemming" as extract the root " استئذن "from the word" يستئذنهم" and" حى "from""يحييهم.

4.6) In Synonyms processing step,the systemextract the synonyms words in the solution automatically in the same bag of words and send this bag to the certain ontology subclass in certain lesson class which called ''lessonBOW''which resolve the synonym of each word. That is the main point to accept the semantic.

4.7) Then our proposal used LSAtoscore the prepared answers. In this step we will analysis the text depend on LSA method as presented in reference [10]. So we will convert the text asan m-dimensional vector (t_m),let$tf(d,t)$donate the frequency of term $t \in T$ where $T = \{t_1, \ldots., t_m\}$ the set of distinct terms occurring then the vector representation of a solution d is:

$$t_d = (tf(d, t_1), \ldots., tf(d, t_m)) \qquad \longrightarrow \qquad (1)$$

That is to capture terms and reflect their importance we transform basic term frequencies $tf(d,t)$ into the $tfidf$ (term frequency and inverse document frequency). "$tfidf$" the frequency of a term in a document "d" with a factor discount its importance according to its appearances in the whole document.

$$tfidf(d, t_m) = tf(d, t_m) \times \log\frac{|d|}{df(t_m)} \qquad \longrightarrow \qquad (2)$$

Then we will compare between textafter creating $tfidf$ for both model answer which select from ontology and student's answer. Then we measurethe similarity between two document using cosine similarity measureswhich is one of the most popular similarity measure used in text.

4.8)thenthe system will translate the scoring result to knowledge as a smart feedback:

In this step the method classified the smart feedback to two main parts for student and teacher. This step is fully automatic. Time is an impact factor[11] that can affect in solving oral exam and specially in reading question. So some knowledge feedback automatically will send to students. Part of the feedback which students get on their written work is the mark or grade itself and also giving a brief comment to every group of degrees present a good effect for the student.

We try to apply this comment to view automatic according to the student grade so the student can easily know the meaning of his result.

Let the average time to solve the exam is "t",then convert "t" to interval using the standard error equation the correct time answer will be in the ranget ±standard error. Then the smart feedback has one of the following cases:

case1: if the student answertime is more than t range and the answer is right, then feedback is" correct answer but with unnecessary information".

case2: if the student answertime is less than t range and answer is right, then feedback is" correct answer with ability to summarize".

case3:ifthestudent answer time is not in t range and the student hastype of defects pronunciation, then feedback is said that.

case4: finally, if the answer is wrong then the system doesn't measure time so the feedback will be " wrong answer".

But in Reading measurementOral exam;sometimes oral test used to scale reading skills and to find the students suffering from learning difficulties. Wedeveloped our method to feedback the important information by using zero crossing rates to calculate the silence time in the recording answer. Oral reading fluency measure has recently been an important thing to identifying students with learning disabilities.

Some measurement that is useful in this field is correctly read words per minute(CWPM)[12] and relative number of correctly read words (REL_ CW). CWPM is a widely used measure for oral reading proficiencyand is computed by subtracting deletion and substitution errors from the reference text and then dividing the result by the reading time in minutes. After the student reading the comprehension then the actual time is recorded and the silence time is computed by using zero crossing method:

$$ZCR_i = \sum_{n=0}^{N-1}|seg[x_i(n)] - seg[x_i(n-1)]| \qquad \longrightarrow$$

Where; N denotes the length, in samples, of the spe $_3$ segment (segx_i(n)).

So CWPM can be easily automatically computed.

REL CW uses the same formula but is not normalized by time but by the passage length, i.e., is notindicative of reading speed and measures reading accuracy only. The following figure will show the infrastructure of the proposal method figure 2.
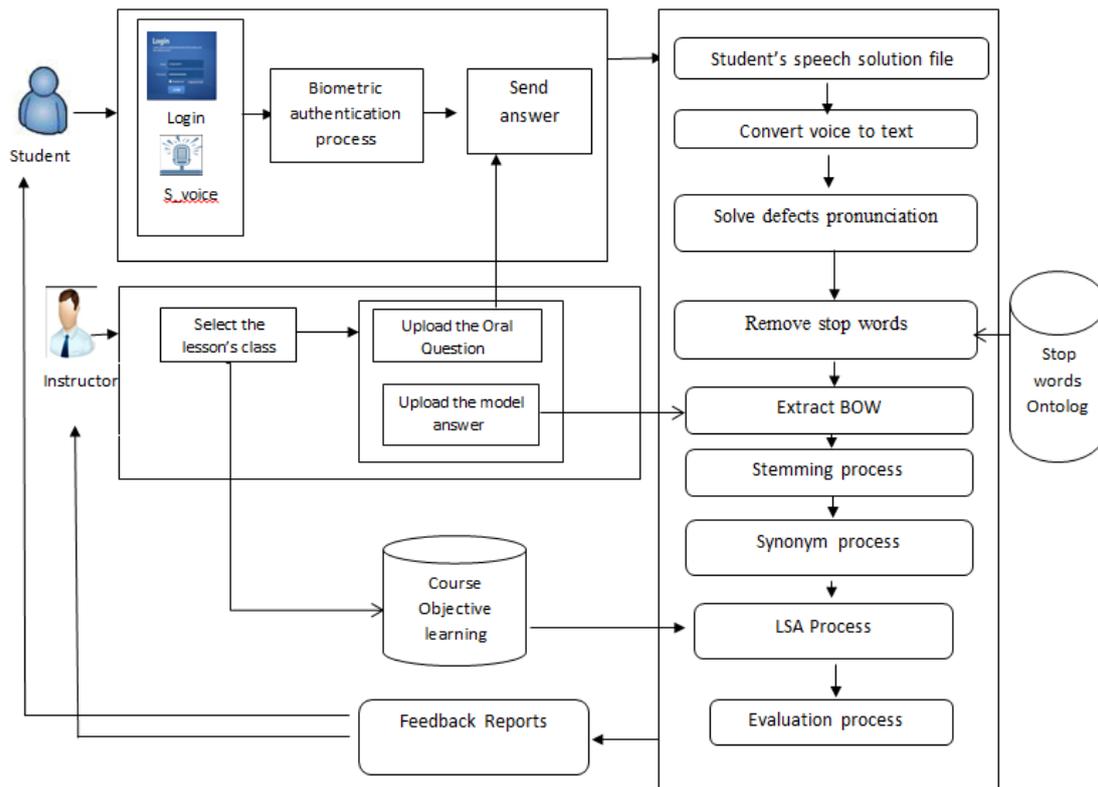
Fig2 :Proposal method infrastructures

5) Implementation:

Framework that will be used to design and develop the system is NetBeans IDE 7.2.1, MATLAB R2012a is furthermore the system will be implemented using JSP with MYSQL database language.Also protégé OWL editor used to create learning object ontology and connect ontology to java. Thefirst stage will be the Authentication implementation using MATLAB Builder™ JA which enables us to create Java classes from MATLAB programs. The recentphases developed by JSP. This implementation is tested on Windows 7 64bit operating system with an Intel® core2TM i7 CPU (2.20GHz) and 8GB of RAM.

Dataset

We took a sample of twenty students in primary school in Egypt. After Arabic Tafseer Exam, we choose twenty answers which cover different cases of solutions.The sample of Open Oral question was " ما هى اداب الزيارة؟"the full mark was tenequivalents to 100% and the model answer uploaded from instructor site wherethe average rate was near fifty words.

Then we will apply our proposal method on the data set as the following steps:

First, the student must exceed theauthenticationstep where the student starts the onlineexamby two authentication methods login insert the password and record his voice as a biometric technique.

Fig3: the biometric authentication step

Then the student goes to the exam page,after insert the answer audio record. The system converts the voice to text automatically by Google ASR. That is to solve some noisy problems and text from manydialects[13]
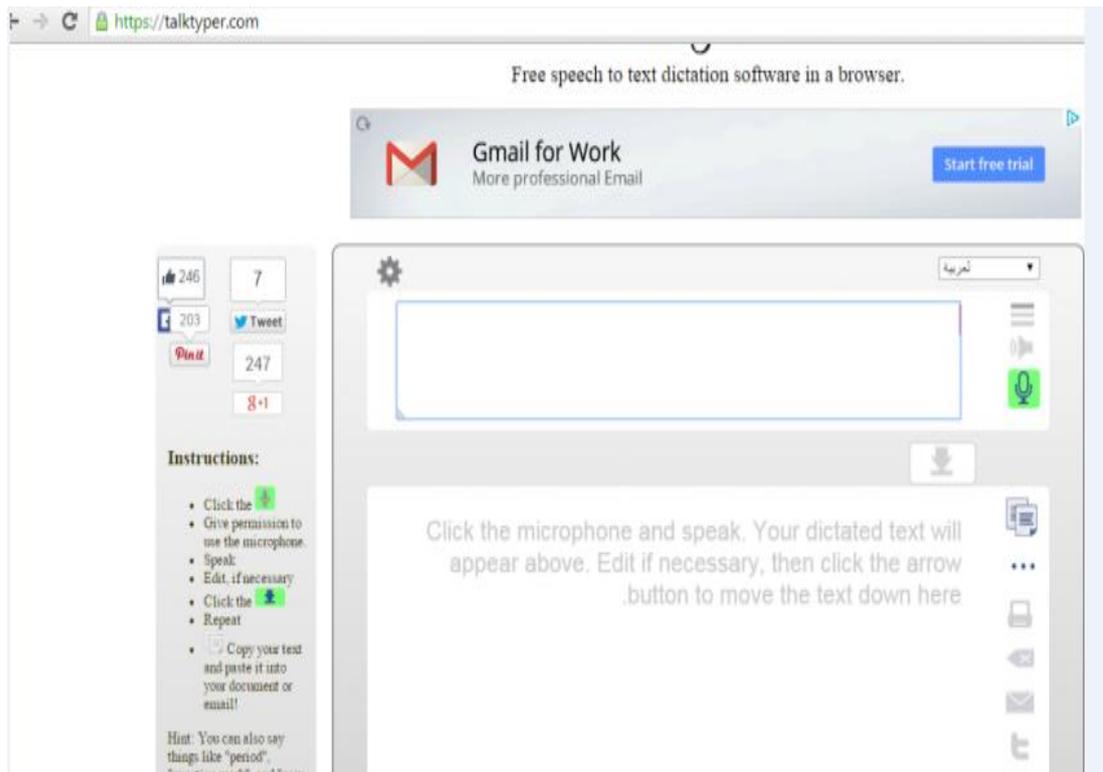
Fig4: Google ASR

If the student hasdefects pronunciation, the system automatically solved them. Then the system will delete the repeated words and the output result stored in new matrixfigure (5-b).Then the system automatically remove Stopwords (noise words) in Arabic Languagefigure(5-c). The Stop words database we used is found in reference[14]. We convert the stop words data base to ontology. Then the system extracts rootsfor each word as in figure (5-d).
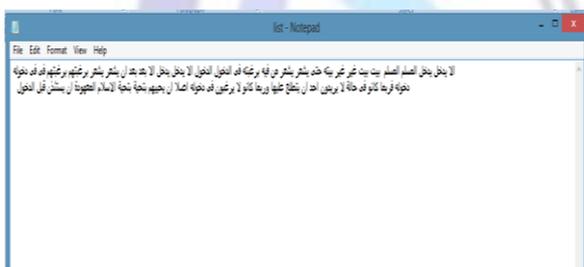
5-a)student's answer Example
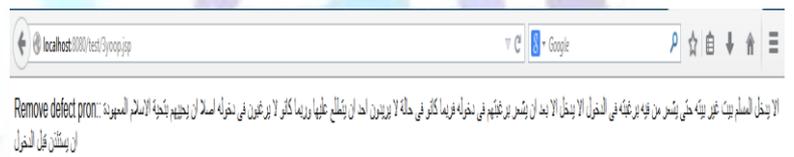
5-b) Removing defect pronunciation

5-c) student's answer without noisy word
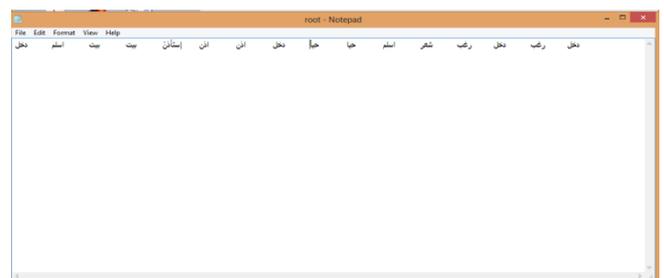
5-d) Extracting roots from student's answer

Fig 5: Snapshot of Student's answer processes

Now in synonym process, the system automatic connects to our Ontology in order to match the BOW of the solution and the BOW of the lesson. That is to collect the synonyms

Now, the system will use LSA technique to compute the score. On our data set the result would be as the column two in table (1):

Also the system evaluates student from many aspects as the following way:

The system calculate total time for reading comprehension, than calculate silent time using zerocrossing rate method and then delete it from the total time to get accurate time for reading. Then by using the number of all words and the corrected words in each answer. Then the system will calculate WCPM rate and Accuracy Fluency Rat.

Where, WCPM= words read correctly/total time in second

REL CW= words read correctly /total words

 The following figure (6) presents that part of the evaluation.

time is ::8.133151927437641 total words:48 Corrected words:44

WCPM=90%

REL CW=91.6%

Fig6: feedback example

After scoring the smart feedback will work individually for each student sothe system will send automatically the individual analysis of result which are students need to know. It will be like Figure(7):

Your Scoring:9.203084729103026

Your time is 20 minutes

the actual time is :18 minutes

correct answer with ability to summarize.
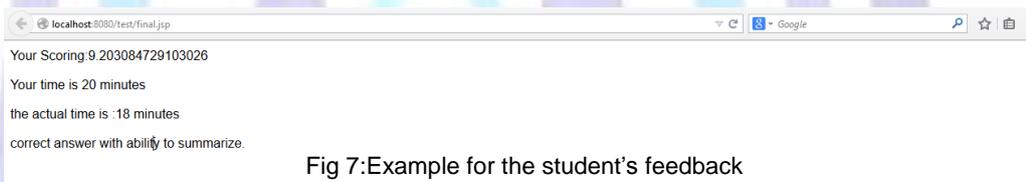
Fig 7:Example for the student's feedback

Analysis of result:

For the same data set, we compared among our proposal semantic method, manual and evaluate by using similarity roots method.The table (1) presents the comparison result. Also the graphsdisplay the comparison graph 1 and graph 2.

Table1: the results of root method, our proposal semantic method and manual method

Grade	Roots method	Our proposal method	Human method
Grade A	89.4	89.7	90.0
Grade B	98.9	98.9	100.0
Grade C	95.0	99.2	100.0
Grade D	98.2	98.5	100.0

Graph1: display the comparison among methods
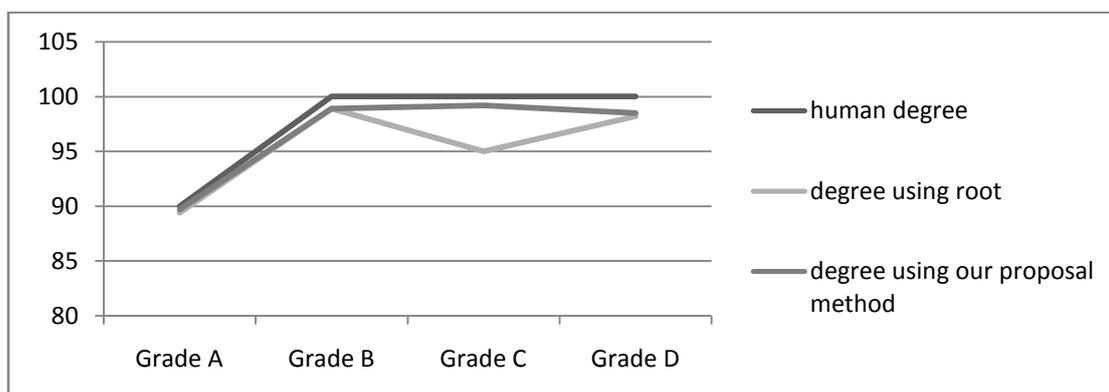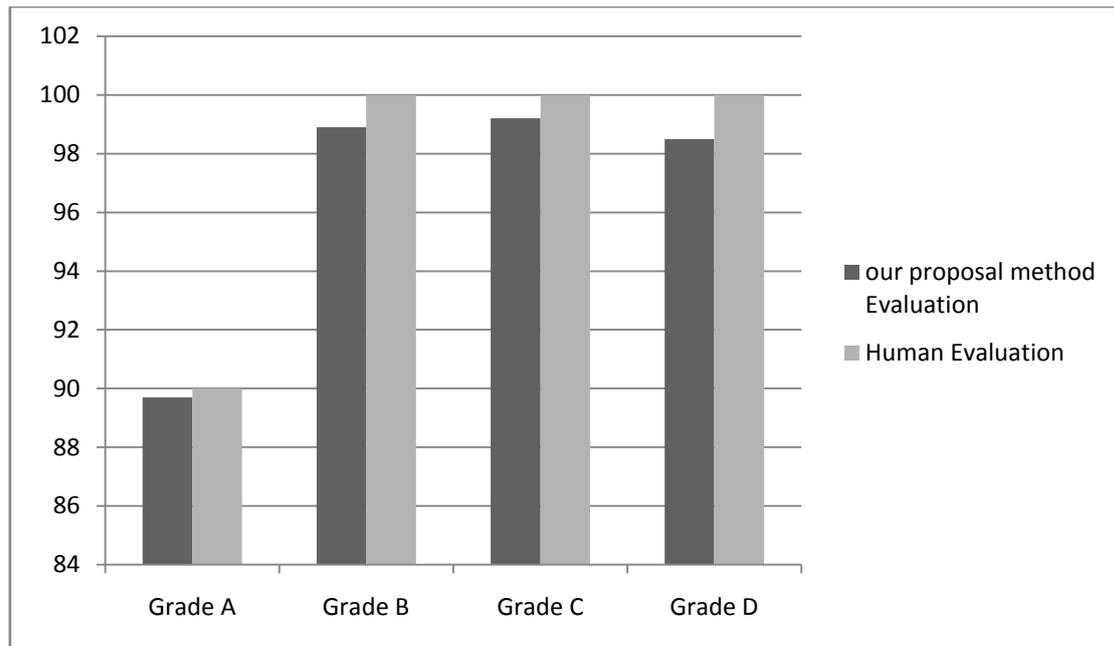
Graph2: the comparison result between our proposal method and human method for scoring

Also to calculate the accuracy of the system we will use precision, recall and F measurements. Where precision=the number of corrected document evaluated by the system/the number of document graded by the system. And

Recall=number of document corrected by the system /number of document corrected by human.

F-measurement=2*precision*recall/precision+recall. The following table (2) presents the results of these measurements.

Table 2 : the result of accuracy measurements

Precision	386.3/400=96.5%
Recall	386.3/390=99%
F-measurement	97.4%

Conclusion

Inthis paper we introduce a new solution for evaluating Arabic oral exam. Our proposalintegrated BOW (Bag of Word), LSA(Latin Semantic Analysis), ASR (automatic speech recognition), zero crossing rate, and Ontology based approach to automate the online oral exam especially in Arabic language with take into consideration the authentication problem. Our proposal accuracy presents in table 3. Also the result of applying our proposal is better than applying the root method by approximately 1.2 % in our data set sample. Our proposal result is nearly as manual evaluation method. In our method, the smart feedback presents automatically some defects pronunciation. In the futurewe aim to enhance our method in evaluating Arabic oral exam to evaluate also sentence structure to add more types of Arabic oral test. Also we are going to ameliorate the system by Ontology of Arabic Dictionary like WordNet and HowNet. Finally, we will suggest anew solution to evaluate some defects pronunciation.

REFERENCE:

[1]M MRefaat, A AEwees, M MEisa, Ab A Sallam,"Automated assessment of student arabic free text answer", IJICIS, Vol.12, No. 1, JANUARY 2012

[2]CBSLEARNINGLAB, Oralexaminationandassessment, Leth Andersen, Hanne &Tofteskov, Jens: Eksamen&Eksamensformer, bedømmelseogbetydning, Samfundslitteratur, chapter 7, 2008.

[3]Md. Monjurul Islam, A. S. M. LatifulHoque," Automated Essay Scoring Using GeneralizedLatent Semantic Analysis",JOURNAL OF COMPUTERS, VOL. 7, NO. 3, MARCH 2012

[4]Poonam Sharma, "Automatic Identification of Silence, Voiced and Unvoiced Chunks in Speech", Thapar University,India,2012

[5]M.A.Anusuya, S.K.Katti," Speech Recognition by Machine: A Review",IJCSIS) International Journal of Computer Science and Information Security,Vol. 6, No. 3, 2009

[6]The National Center for Learning Disabilities (NCLD)http://www.ncld.org/. Access at Sep. 2014.

[7]EmanK.Elsayed, Kamal A.Eldahshan,AlshimaaAbdelraof,"Oral multi model authentication and scoring algorithm",the sixth international conference on intelligent computing and information systems(ICICIs),Cairo,Egypt,2013.

[8]Jian Cheng, JaradBernstein,UlrikePado and Masanori Suzuki, "Automatic Assessment of Spoken Modern Standard Arabic", proceedings of the NAACL HLT workshop on innovative use of NLP for building educational application ,June 2009.

[9]Wael Hassan Gomaa and AlyAlyFahmy,"Arabic short Answer Scoring with Effective Feedback for Students", International journal for computer application ,volume 86 no 2,January 2014.

[10]HananeFroud, AbdelmonaimeLachkar and Said AlaouiOuatik,"Arabic text summarization based on LSA to enhance Arabic document clustering",International Journal of Data Mining & Knowledge Management Process (IJDKP)- 2013

[11]TrangThiThuy Nguyen," The Impact Of Background Knowledge And Time Constraint on Reading Comprehension Of Vietnamese Learners Of English Asa Second Language",Southern Illinois University Carbondale,2012.

[12]Klaus Zechner, KeelanEvanini, Cara Laitusis,"Using Automatic Speech Recognition to Assess the Reading Proficiency of a Diverse Sample of Middle School Students",2011.

[13]FadiBiadsy, Pedro J. Moreno, Martin Jansche,"Google'S Cross-Dialect Arabic Voice Search", IEEE, 2012.

[14] http://arabicstopwords.sourceforge.net/