International Journal of Computer Applications (0975 – 8887)
Volume 36– No.10, December 2011

Intelligent Adaptive Feedback Assessment System in Learning Environment

G. Suvarna Kumar
C.S.E Department
M.V.G.R C.E
A.P, India

P.V.G.D Prasad Reddy
CS & SE Department
Andhra University
A.P, India

Sumit Gupta
C.S.E Department
M.V.G.R.C.E
A.P, India

Sandeep V
E.C.E Dept
GITAM University
A.P, India

ABSTRACT
Effective evaluation of tutors and learning environment is a vital educational activity for the development of educational quality and enhancement of an educational institute's improvement in terms of quality driven teaching and learning. Considering the above statement, it is mandatory to assess the learning activities and attitude of students towards learning along with changing trend from a teaching and learning perspective. In this paper an intelligent adaptive feedback-assessment (IAFA) system framework has been proposed where we provide relative feedback assessment by correlating different learning styles of students along with teacher's resource utilization. The assessment is based on the features extracted using the techniques from the fields of computer vision and speech analysis. In audio processing the perception attitude of the students is taken into account which makes us compute relative and desirable audible features like speech rate, loudness for feedback analysis. Using video processing techniques for tracking and template matching, we compute resource utilization for teaching aids like black board, projector and student interaction and compare it with the desired values. The present algorithm has been rigorously tested with several sample classes of around 60 students with different student backgrounds i.e., non-English medium/English Medium/Native English speakers, and the performance results have been reported.

General Terms

Perceptual loudness, mermelstein algorithm, Frame Differencing, Edge Detection, Template Matching, adaptability.

Keywords
PCA Algorithm, EMD, Heuristics.

1. INTRODUCTION
Teaching is a pivotal task in efficient learning. Maintaining qualified personnel in educational institutions is the central link to teaching task. A teacher should utilize the resources efficiently and effectively [1]. Evaluating a teacher is very essential. Traditional methods to evaluate a teacher such as online feedback, manual feedback have proved to be a little inefficient in terms of reaching the point of accuracy in the judgment of the teaching quality. Automatic assessment systems have evolved since many students give feedback without reading the questionnaire properly, still such assessments fail to justify their objective by being teacher-centric and neglecting the learning styles of the students. To make the system more reliable, we need an adaptive, accurate and faster system with good learning tools. In this paper, the various perceptions and learning styles of students are taken as the main frame of reference for developing and implementing the intelligent adaptive feedback-assessment (IAFA) system. The main objective behind developing an adaptive feedback analysis is to improve the reliability of any feedback system which is both student and teacher centric. In a pragmatic regular classroom, there exist students with three different English proficiencies (non-English / moderate English / professional English).For finding student desired learning styles and attitudes towards learning based on their English proficiency level. Perceptional evaluation has been conducted for students in various class rooms.

Here, both image and audio processing techniques have been used for providing adaptive feedback. We capture the frames containing shots of a teacher and track his movements, which enables us to classify his resource utilization. This classification helps us to decide how much time the teacher is allocating to the blackboard, projector and student interaction. Speech processing techniques have been used to extract the audio features such as loudness, speech rate and activeness of the speech from the video. If the audibility and rate of speech meets the desirable learning style of the student, there is a higher probability of a student understanding and grasping the contents of the lecture. Maintaining a monotonous pitch while delivering the speech can easily give the audience an impression that the speaker is not enthusiastic or confident about the topic [2].

In section II, we discuss various feature extraction methods for audio and video processing. Section III elaborates the implementation details and results of IAFA system. Finally, in section IV conclusions and future work are presented.

2. AUDIO-VISUAL FEATURE EXTRACTION METHODS

2.1 Speech Processing Techniques
In order to provide the intelligence to the IAFA system, loudness and speech rate are employed for analyzing the learning style with respect to teacher's speech

2.1.1 Measuring speech rate

Initially, speaking rate is measured quantatively by counting the number of words spoken by the teacher per minute. By identifying these gaps we count the number of words and hence, the speech rate (words /min) is calculated. Figure-1 shows a typical speech waveform in which we have marked the three regions corresponding to gaps as there the amplitude is too low. Such gaps enable us to compute the speech rate.

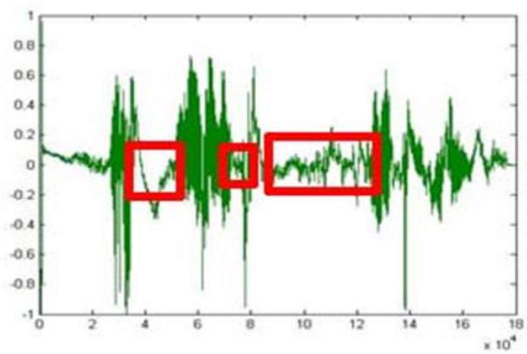

Figure-1: calculating speech rate by counting gaps.

2.1.2 Measuring loudness of speech

To measure the loudness of the speech two techniques are used:

a) Perceptional loudness test :

In this technique, the perceptual loudness is carried out by conducting subjective tests with respect to students in a class [3]. The students are asked to fill a sample form for each session.

b) Extraction of loudness measure from Hilbert envelop of LP residual :

An objective measure (η) of perceived loudness based on the abruptness of glottal closure can be derived from the speech signal. The abruptness of the glottal closure derived from the EGG signal is known to be high for loud speech compared to the soft and normal speech. When the glottal closure is abrupt, the Hilbert envelope of the LP (Linear Prediction) residual of the speech signal will have sharper peaks at the GCIs (Glottal Closure Instants) [4]. The impulse like feature of excitation can be observed in Fig. 2(c).Comparative analysis of soft, normal, and loud speech is made and reported. The sharpness of the peaks in the Hilbert envelope at the epochs is derived by computing the ratio η**.** The loudness measure defined by

$$\eta = \sigma / \mu. \quad (1)$$

Here μ denotes the mean, and σ denotes the standard deviation of the samples of the Hilbert envelope of the LP residual in a short interval (2 ms) around the GCIs. We extract GCIs using Auto correlation function. GCIs are the pitch locations and the Hilbert envelope r[n] of the LP residual e[n] is given by

$$r[n] = \sqrt{e^2[n] + e_H^2[n]} \quad (2)$$

where $e_H[n]$ denotes the Hilbert transform of e[n]. The Hilbert transform $e_H[n]$is given by

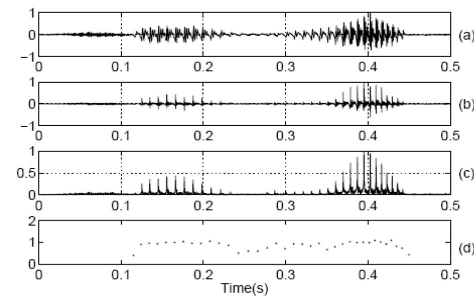

Figure-2: a)Segment of a speech signal used from collected speech database b)12th order LP residual c)Hilbert envelope of the LP residual and (d) contour of extracted from the Hilbert envelope.

$$e_H[n] = IFT(E_H(\text{w})) \quad (3)$$

where IFT denotes the inverse Fourier transform

Here *E* (w) denotes the Fourier transform of the signal $e[n]$. The sharpness of the Hilbert envelope of the LP residual can be captured by the parameter η, which is computed using a 2 ms segment of the Hilbert envelope around each epoch. The mean value of η derived from the utterances of loud speech has been observed to be greater than that of normal and soft speech. Fig. 2(d) illustrates the contour of values of η for the segment of speech signal shown in Fig. 2(a).

2.1.3 To analyze the activeness of speaker

The speaking rate is observed to evaluate the activeness of the speaker and predict whether it is a memorized or spontaneous speech. In a memorized speech, the ideas to be expressed are formally prepared and formulated before the speech production begins whereas, in a spontaneous speech, the formulation process and speech production are simultaneous. It has been observed that speaking rate is higher in memorized speech than in spontaneous speech. In this analysis, the quantitative metric for the speech rate is taken as syllables per second [4]. This metric is chosen because of the robustness in estimation of the syllable boundaries and close remembrance to the speech production characteristics [5-7]. Identification of the syllable boundaries is performed using syllable detection algorithm proposed by Mermelstein [8].

The variations of syllable rate have been analyzed to illustrate the variations of speaking rate in spontaneous and memorized speeches in a mean subtracted histogram of syllable. The standard deviation (σ**)** of spontaneous speech is higher than that of memorized speech

2.2 Image processing techniques

In this work the following are the three key steps for object tracking system:

2.2.1 Tracking an object using Frame Differencing:

The system first analyses the images, being recorded by the camera, for detection of any moving object. For this purpose The *Frame Differencing* algorithm is used, which gives the position of the moving object in the image [9]

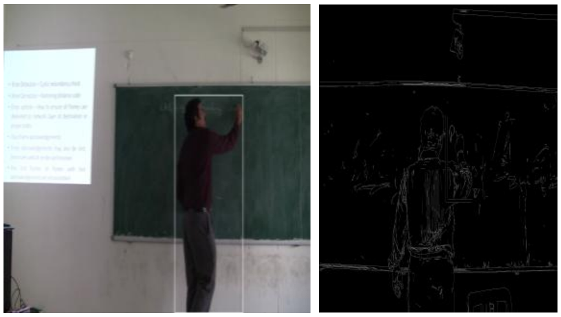

Figure-3: (a) Tracking of an object (b) Image after edge detection.

2.2.2 Edge Detection

Edge detection [7] is a fundamental tool used in most image processing applications to obtain information from the frames as a primary step to feature extraction. This process detects outlines of an object, boundaries between them and the background in a frame. An edge-detection filter can also be used to improve the appearance of blurred or anti-aliased video streams.

In this work Canny algorithm [7] is used for an optimal edge detector, based on a set of criteria which include finding the most number of edges by minimizing the error rate, marking edges as closely as possible to the actual edges to maximize localization.

2.2.3 Template Matching

The basic problem that often occurs in image processing is to determine the position of a given pattern in an image with respect to region of interest. After determining the position of the pattern, we obtain a correlation using which we can determine the activity of the professor i.e. facing the black board or pointing hand towards LCD projector. In template matching [8], to handle translation problems on images and comparison of the intensities of the pixels, the SAD (Sum of absolute differences) measure is used.

A pixel in the search image with coordinates (x_s, y_s) has intensity $I_s(x_s,y_s)$ and a pixel in the template with coordinates (x_t, y_t) has intensity $I_t(x_t , y_t)$. Thus, the absolute difference in the pixel intensities is defined as

$$Diff(x_s, y_s, x_t, y_t) = |I_s(x_s, y_s) - I_t(x_t, y_t)| \qquad (4)$$

$$SAD(x, y).l = \sum_{i=0}^{Srows} \sum_{j=0}^{Scols} SAD(x, y) \qquad (5)$$

The mathematical representation of the concept of looping through the pixels in the search image as we translate the origin of the template at every pixel and take the SAD measure is as:

$$\sum_{i=0}^{Srows} \sum_{j=0}^{Scols} SAD(x, y) \qquad (6)$$

S_{rows} and S_{cols} denote the rows and the columns of the search image and T_{rows} and T_{cols} denote the rows and the columns of the template image, ly. In this method the lowest SAD score gives the estimate for the best position of template within the search image.

Two basic steps are performed

1. The estimation of position in the pattern is calculated using cross correlation.
2. Based on the correlation factor obtained, the position of The pattern is determined.

For continuous functions, *f* and *g*, the cross correlation [9] is defined as

$$(f * g)(t) = \text{def} \int_{-\infty}^{\infty} f * (\tau)(t + \tau) d\tau \qquad (7)$$

Where f * denotes the complex conjugate of f. Similarly, for discrete functions, the cross-correlation is defined as

$$(f * g)[n] = \text{def} \sum_{m=-\infty}^{\infty} f * [m] g[n + m] \qquad (8)$$

The cross-correlation is similar in nature to the convolution of two functions. Convolution involves reversing a signal, then shifting it and multiplying by another signal, whereas correlation only involves shifting it and multiplying (no reversing).

3. IMPLEMENTATIONALDETAILS AND RESULTS

This section focuses on describing adaptability of the IAFA system. Here, the feedback assessment of the teacher is student centric which makes it more reliable by justifying its objective.

3.1 Implementation of IAFA System with image processing

3.1.1 Image perception evaluated data material

This data plays a major role for maintaining a proper intelligence and adaptability of the system, since the teaching resource utilization is subject centric and also synchronizes with the learning style and attitude subject to the required set of students. To further understand and polish this aspect, we consider the number of subjects to be 5 namely A, B, C, D, E.

A questionnaire is designed by prioritizing the above two factors. All the 60 students are guided to complete the questionnaire in order to specify how frequently a specific resource is used such as black board, LCD Projectorstudent interaction levels with respect to the given subject. The levels are examined and scrutinized by subject experts.

The following table I represents the reference values of resource utilization for subjects A, B & C (Based on the questionnaire results.

TABLE I: Image reference values

	Board usage	LCD presentation	Student interaction
Subject A	0.50	0.20	0.30
Subject B	0.15	0.55	0.25
Subject C	0.30	0.30	0.40

In real time the total video lecture of 50 minutes is considered. The image processing techniques mentioned in the section 2.2 are used to analyze how frequent the resources are utilized while delivering the lecture. There are three key steps in analyzing frames from the video to the resource using object tracking system. Firstly, tracking the person is done using frame difference. Then, the position is detected using edge detection. Finally, the activity is tracked using template matching [11].For measuring the resource utilization, template matching plays a vital role for its calculations. According to template-matching theory [8], an incoming stimulus is compared to specific patterns (or templates) that have been stored in memory in an attempt to find a good match. When a good match is found, the stimulus is identified based on that match.

(a) (b)

(c)

Figure-7: Tracking the specific resource utilization: a) Board usage, b) Projector Usage and c) Student Interaction.

IAFA system determines the usage of a specific resource. It tests the blackboard usage successfully with success rate of above 80 percentages, projector usage above 75 percentage and student interaction with 90 percent success rates.

3.2 Implementation of IAFA system with respect to Speech Processing

3.2.1 Speech Material:

The perceptional evaluation test is conducted on a class of 60 students by taking the speech of 15 teachers into consideration. The teachers are given a fixed English transcript and are guided to speak in their normal, slow, fast speaking rates while maintaining their naturalist of speech. The above analysis gives 45 .wav files numbered sequentially and in addition, the same set of teachers are asked to speak in memorized and spontaneous manner.

Now, the listeners (students) are guided to listen to 45 wav files and grade the loudness and speech rate according to their learning style. The students may be from a background of non-English medium, moderate English proficiency, high proficiency in English. This observes the difference in the rating aspect.

3.2.2 Training data for IAFA system

Once the speech data has been collected, the speaking rate and loudness factor for each .wav file is calculated by the techniques mentioned in section II (A) and added to the database. Then, perception evaluation by class room students is conducted. The reference values are defined separately with respect to different learning styles of the students in table III and also on average basis in table II. The required activeness (**σ**) for each background of students is also tabulated below.

TABLE II Speech reference values on average basis

SPEECH FEATURE	REFERENCE VALUES
LOUDNESS(**η**)	0.5867
SPEECH RATE	112 (WORDS/MIN)
ACTIVENESS(**σ**)	>0.65

TABLE III Comparative reference values for three different backgrounds of students

Background	LOUDNESS (**η**)	SPEECH RATE	ACTIVENESS (**σ**)
Non-English Medium	0.6224	98	>0.65
Moderate English proficiency	0.6077	109	>0.65
Native English speakers	0.5666	123	>0.62

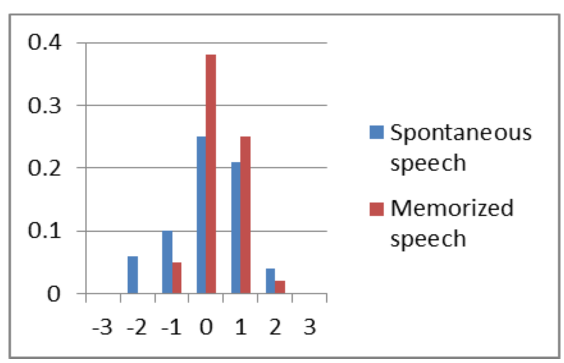

Figure-5: Mean subtracted histogram of syllable rate for spontaneous speech and memorized speech.

The figure 5 shows Bar chart to the comparisons for memorized and spontaneous speech. All the histograms are mean subtracted in order to preserve emphasis on the variations rather than on the absolute value of the syllable rate. Variations of speaking rate are observed higher in spontaneous speech than in read speech. From Fig. 5, it can be observed that the histograms corresponding to spontaneous speech are broader than the histogram corresponding to read speech. The standard deviations (**σ**) of figure (5) are 0.76 and 0.44. The standard deviation of spontaneous speech is greater than that of read speech. In our sampling, the optimum value which matches the activeness for any learning style is calculated as $\sigma > 0.65$.

The values in table II and table III assist the IAFA system for becoming more adaptive in its performance. For a general subject class with students of different learning styles equally distributed the table II values are taken into frame of reference for evaluating the teacher but considering a case where a specific set of students of a learning style (or) background are dominant, the average value is calculated with the values in table III by 'cumulative grading' for more accuracy. This gives how likely each speaker's normal rate of utterance (or slow or fast) and loudness (soft, normal and loud) meets the student's expectations leading to a better understandability.

The feedback evaluation can be extended for different ranges of classes like remedial class sessions, certification courses, Olympiad training classes, etc. For the above mentioned cases, there is variance in the dominance of specific learning style of the students. Then, the set values from the table II are considered as the frame of reference for giving the feedback result accordingly. By the above results, we can say IAFA system is adaptable to any kind of class and any learning style of the students.

TABLE IV Loudness measure from Hilbert envelope of LP residual

Faculty id	Loudness measure (η)
Teacher 1	0.6077
Teacher 6	0.5666
Teacher 8	0.6224

Table IV shows typical values of (η) obtained for some of the faculties (from among 18). The threshold value of the group of students considered is 0.5867. So, teacher 6 has to improve his delivery in terms of loudness.

The loudness of different teachers has been trained and tested with SVM classifier and the results were accurate in determining the different levels of loudness.

4. CONCLUSIONS AND FUTURE WORK

In this paper, a novel intelligent adaptive feedback system has been proposed which takes into consideration the student background and proficiency level. Several reference values have been computed that can be employed for providing real time feedback to the tutors depending on the student responses. The proposed system framework makes the assessment both teaching and learning centric justifying the objective of a feedback. There are many things that can add more intelligence like variation in instantaneous fundamental frequency F_0, variation of strength of excitation. In image processing, even the gestures of the students can be taken into consideration. These kinds of systems will be very helpful, because they eliminate the need of manual feedback which may not be effective and reliable. This project can be applied in real time to ameliorate the standards of education in schools and colleges. A low embedded hardware prototype can made implementing the above algorithm.

5. REFERENCES

[1] GuJinling,"Design of the whole teaching quality evaluation system of higher colleges and universities[J]". China Scienceand Technology Information .2007 (8) :240-242.

[2] Http://members.tripod.com/teaching_is_reaching/delivering_your _speech.htm

[3] An adaptive procedure for categorical loudness scaling,Thomas Brand and Volker Hohmann 2002 Acoustical Society of America.

[4] G.Seshadri and B.Yegnanarayana, "Perceived loudness of speech based on characteristics of excitation source," J.Acoust.Soc.Am., vol. 126,no. 4,pp. 2061-2071, Oct. 2009.

[5] E. Shriberg, A. Stolcke, D. Hakkani-Tur, and G. Tur, "Prosody-based automatic segmentation of speech into sentences and topics," Speech Commun., vol. 32, no.1-2, pp. 127-154,2000.

[6] M. Siegler, "Measuring amd compensating for the effects of speech rate in large vocabulary continuous speech recognition ," MS Thesis, Carnegie Melon Univ., Pittsburgh, PA, USA, 1995.

[7] N. Morgan and E.Fosler-Lussier, "Combining multiple estimators of speaking rate," in Proc. Int. Conf. Acoust., Speech, Signal Process., vol. 2, Seattle, Washington, USA, May 1998, pp. 729-732.

[8] P.Mermelstein, "Automatic segmentation of speech into syllabic units,"J.Acoust.Soc.Am., vol. 58,no. 4,pp. 880-883, 1975.

[9] JAIN, R. AND NAGEL, H. 1979, ―On the analysis of accumulative difference pictures from image sequences of real world scenes. IEEE Trans. PattAnaly. Mach. Intell. 1, 2, 206-214.

[10] William McIlhagga, "The Canny Edge Detector Revisited" in Int J Computer Vision (2011) 91:251-261.

[11] Template matching- http://www.pages.drexel.edu/~weg22/templatematch.html

[12] Cross Correlation - http://en.wikipedia.org/wiki/Cross Correlation. EDWARD R. DOUGHERTY (1993) Mathematical Morphology in Image Processing. CRC Press.

6. ACKNOWLEDGMENT

Our thanks to the experts who have contributed towards development of this paper.

6. AUTHORS PROFILE

Prasad Reddy P.V.G.D is currently the Rector (Pro-Vice Chancellor) of Andhra University. He completed Phd in the year 1993. He has **24 years** of teaching &research experience which includes more than **10** years of administrative experience with Andhra University at various senior capacities. He has so far successfully guided **12** Ph.D's, **18** M.Phil's and **259** M.Tech Projects in the field of Computer Engineering , and has to his credit, more than **104** Research papers which includes **42** publications in the referred international Journals of repute, **2** Patents filed and he has presented **19** publications in International Conferences across the world.

Sumit Gupta is currently the HOD of the CSE department in MVGR College of engineering. He Completed his Phd from NUS Singapore in Image processing and 3D Visualization research area. He has 8 years teaching experience, 11 years R & D and 3 years industry experience. He has 39 publications which include 8 national journals, 11 international journals, 11 national conferences and 9 international conferences. He was a member of professional bodies like CSI and ACM.

Suvarna Kumar G is currently working as the Senior Assistant Professor in CSE Department of MVGR College of Engineering. His Research area is Speech and video processing. He published 6 international journals and attended 2 conferences in Singapore. He has 9 years teaching experience, 2 years R &D and 1 year industrial experience. He is a member of ACM, National Science Congress, IAENG and IACSIT.

SandeepVazrapu is pursuing his 4^{th} year in Bachelors of technology in the department of electronics and communications engineering in GITAM University, Andhra Pradesh, India.