Speech Based Question Recognition of Interactive Ubiquitous Teaching Robot using Supervised Classifier

Umarani S D[#1], Raviram P[*2] and Wahidabanu R S D[#3]

[#]Department of ECE, Government College of
Engineering, Salem - 636011,
Tamilnadu- 637503, INDIA,
[1]umaraviram@gmail.com
[3]rsdwb@yahoo.com

*Department of CSE,
Mahendra Engineering College,
Tiruchengode,
Tamilnadu- 637503, INDIA,
[2]drpraviram@gmail.com

***Abstract*—This paper presents the research work that aims at designing a speech based interactive ubiquitous teaching robot for quality enhancement of Higher Education ensuring better performance by making possibility of interaction between the Robot Teacher and Student/ Learner. Students/ Learner can ask technical questions for which the robot replies. Here the reply by the Robot is limited by the answers available in the database. The proposed speech based interactive teaching system has two phases; they are training and testing. A feed forward neural network is used as a supervised classifier after it is trained by means of a back propagation (BP) algorithm using the features of sound signal from the Student/ Learner. The number of queries present in the database and non database queries recognized is presented by Receiver Operating Characteristics curve (ROC). The ROC curve lies above the diagonal of the ROC plot. This indicates that the proposed method of interaction of Robot with the learners is acceptable. The percentage of question recognition resulted in 74%.**

Keyword - Feed forward neural network, Supervised Classifier, Back propagation (BP) algorithm, Teaching Robot, Speech processing, Receiver Operating Characteristics.

I. Introduction

Today, number of people going for higher education worldwide is very less. Also after getting higher degrees they are not interested in selecting teaching as their profession, they are fancied towards IT field. At the same time there is less expertise in cutting edge technologies to explain the learners about it. To overcome these difficulties emerging ubiquitous technology can be incorporated in education to support learners. Nowadays the robots are used everywhere, its application area has become wide. They are used in industries, military, medical, as a social partner to help elders and patients, as a security guard etc. With the advent of information technology, artificial intelligence and ubiquitous technology in robotics, the robots can be used in education to create a new trend in present educational systems. In educational field robots are already used as teaching tools [2, 5]. Many research works have been done to utilize robots in educational field as a teaching assistant or as a teacher and they are only in initial stage. Many countries are using robots for improving English communication skill. Robots like Robosapien [9], Robovie [4], Papero [2, 3], Tiro [7], and Dooly [8] are in elementary schools for educating English through storytelling, singing rhymes etc. But for higher education the robots are not used to teach any subject.

Web-based lecture technologies (WBLT) are there to provide flexible access to lectures for learners and it has gained popularity in many higher education institutions [1]. But they cannot clear the doubts of learners on the spot. Till date for higher education robots are not yet used as an educator or even as an assistant for teacher. Based on these constraints this research has been done to design an interactive ubiquitous teaching robot for higher education and the interaction done through speech. The proposed speech based interactive teaching system consists of two sections; they are training and testing. The training phase extracts features like no of words in a question, the maximum and minimum magnitude and their corresponding indexes, the strength of the word etc., from all the questions to be answered by the robot and uses them to train the Feed forward Neural Network (FFNN) used in the proposed method. In the testing phase the above mentioned word features will be

extracted from the interrogative speech when any anonymous learner asks the question, and they will be given as input to the trained supervised classifier. Then, the trained supervised classifier will identify the answer to the question based on the features provided as its input and provide the same as its output.

The structure of the paper is organized as follows: A brief review and architecture of the proposed robotic speech recognition system is given in Section 2. The process of feature extraction and training of neural network is given in section 3. The experimental results of the proposed approach are discussed in Section 4. Finally, the conclusions are given in Section 5.

I. ARCHITECTURE

Fig. 1 shows training and testing of Interactive Teaching Robot. In this design possible question and answers of a particular subject are stored in separate databases and also a link is provided between the question and corresponding answer. During class sessions the learner can raise questions through speech and it will get recognized by the robot teacher. From the received speech signal the features are extracted. Before extraction of the features, noise present in the speech signal is removed by nullifying the magnitude value with the tolerance value 'ω'. The extracted features are given as input to the Feed forward Neural Network that provides the target value.

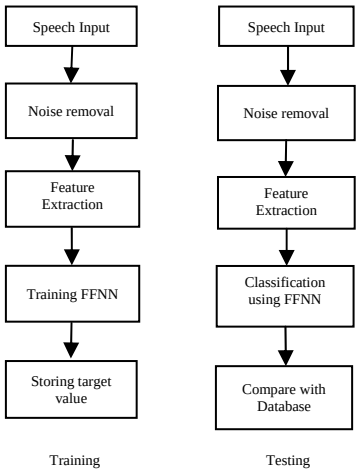

Fig.1. Training and testing of speech/Questions from learners by Interactive Teaching Robot

During recognition the output of neural network is tested with the target value if there is a match then the corresponding question is recognized and the corresponding answer file is identified through the link provided between question file and answer file. The words in the answer file are read and corresponding speech file in assembled which is sent as a reply to the learner. If there is a mismatch between question in database and the raised query then the system will send some comments to the learner and ready to get next query.

II. EXPERIMENTAL SETUP

The performance of proposed speech recognition system was evaluated using the generated database. The database contained totally 'M* N' no of speech signals where 'N = 25' is the no of questions and 'M = 30' is the no of persons. The speech signal was selected such that each had a different frequency and amplitude. The recorded continuous signal was converted into a discrete signal by sampling at a frequency of 44 KHz (Sampling interval =0.2ms) and the noise was removed by nullifying the magnitude values at a tolerance value of ω =0.5.

The total no of questions were divided into five data sets and 63% of speech signal for each question in the database were utilized for training and the remaining 37% were used for testing the proposed system. The ANN used certain time and certain epochs during the training. After removing the noise, features like total no of words in a question, strength of each word in a question, the maximum and minimum magnitude value of each word with their corresponding maximum and minimum index and the mean and standard deviation of magnitudes of each word present in the training set database were extracted for each of the questions and the neural network was trained using the extracted features.

Intelligent speech recognition enables robot to understand spoken questions. After data acquisition the signals are segmented based on each word. The speech signal for the query 'List the difference between histogram equalization and specification' is shown in Fig.2 (a) and the corresponding segmented signal for the words 'List ', 'the' ,' difference ', 'between ', 'histogram ', 'equalization ', 'and ', 'specification' is illustrated in Fig.-2(b), fig. 2(c), fig. 2(d), fig. 2(e), fig. 2(f), fig. 2(g), fig. 2(h), fig. 2(i), respectively.

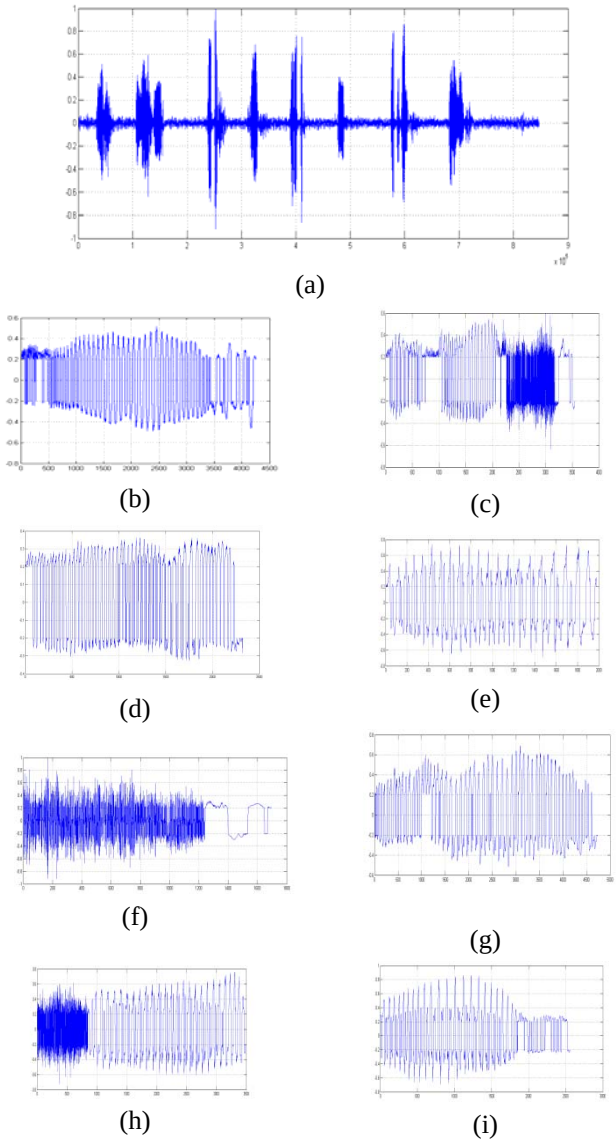

Fig. 2. Various Speech Signals. (a). The speech signal for a sample question 'List the difference between histogram equalization and specification'; The (b),(c),(d),(e),(f),(g),(h) & (i) represent the segmented speech signal of the words 'List ', 'the' ,' difference ', 'between ', 'histogram ', 'equalization ', 'and ', 'specification' respectively.

A. Feature Extraction

Before extraction of the features, nullifying the magnitude value with the tolerance value 'ω' has carried out for removing the noise. The total no of words in each signal is counted by checking the sampling point value '0' in the signal (i.e.) the no of zeros appearing after a numerical value is the total no of words in a question. To identify each word in a question, the no of sampling point, the strength of the word of each question, is measured. The Maximum and minimum magnitude value of each word and are extracted for each word in question and corresponding there maximum and minimum value index and are extracted. All the features which differentiate each word in the questions are extracted for training. The Neural network is trained by all the extracted features of the signal and providing the target value λ_i where i = 1,...,N, the total no of questions.

B. Training Feed forward Neural Network

In the recognition and answering phase, when the user ask a question it's all the features which are used in the training phase are extracted and given as input to the Neural Network In the proposed system the extracted features are given as input and a single output is obtained.

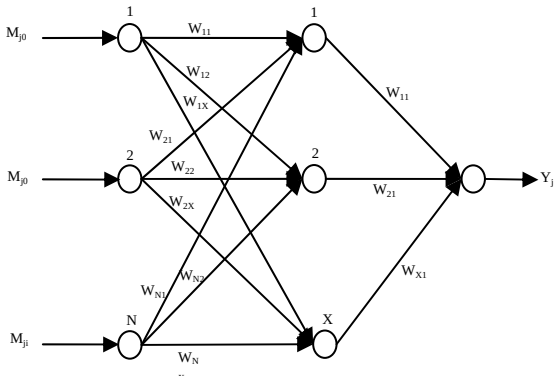

Fig. 3. Eight inputs-one output Neural Network to recognize the question

Architecture of feed forward neural network is given in Fig.3. In that $M_{j0},...,M_{jN}$ represents the input of the network, A weight represented as w_{11}, $w_{12},...,w_{1N}$ respectively is assigned to each line that connects these inputs to the neuron. y_j normally represents the threshold in artificial neuron and the following formula is used to determine the activation.

$$y_j = \alpha + \sum_{j,k=0}^{N-1} w_{jk}\, M_{jk} \qquad (1)$$

$$g(y) = \frac{1}{1+e^{-y}} \qquad (2)$$

$$g(y) = y \qquad (3)$$

The basis function (specified only for input layer) is given in Eq. (1) and the sigmoid and identity activation function, which is selected for hidden output layer are given by Eq. (2) and Eq. (3) respectively. All the remaining layers (hidden and output layer) normally use the basis function given in Eq. (1) (but with the number of hidden and output neurons respectively). The output, from ANNs is determined by giving input to all those layers. A single value is obtained by this method by combining both the weights. Finally, the out value is checked with the threshold value for recognizing the question. If $y_i \leq \lambda$, then the corresponding question is recognized and answered by the robot.

III. Results and Discussions

The performance of the Robot is evaluated using Receiver operating characteristics (ROC) of the Robot for various questions asked by the learners. Table I presents the confusion matrix values and the ROC values. The speeches of interaction of the learners with the Robot have been taken randomly with a gap of 2 to 3 hours. The output of the proposed system may be positive (answering a question correctly) or negative (answering a question incorrectly).

TABLE I

Confusion Matrix for Receiver Operating Characteristics

Instances	True Positive (TP)	False Negative (FN)	Sensitivity	False Positive (FP)	True Negative (TN)	Specificity	False Positive Rate
1	33	22	60	22	133	89.77	10.23
2	29	24	54.7	25	195	88.64	11.36
3	23	33	41	34	186	84.55	15.45
4	13	42	23.6	4 7	173	78.6	21.36
5	11	44	20	45	175	79.55	20.45
Note. *Total database queries =500, **Total non database queries=250							

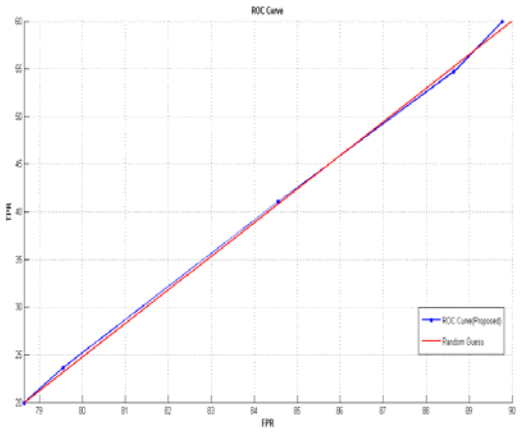

Fig. 4. Receiver Operating Characteristics

Fig. 4 represents the Receiver Operating Characteristics graph of the proposed recognition system. The red straight line along the diagonal represents the completely random guess of the recognition system. The blue line represents the ROC curve of the proposed technology. The diagonal divides the ROC space. The points above the diagonal represent good recognition results, while points below the line represent poor results. In figure-6, the three points are above the diagonal , a point is along the diagonal and a point is below the diagonal with a light deviation of 0.29.

IV. Conclusion

In this work, an attempt has been made to implement the interactive ubiquitous teaching robot. The teaching robot recognized the questions asked by the learners and answered based on the answers available in the database. In the training phase the neural network is trained using the features extracted from all the questions the robot is expected to answer. Then, in the testing phase the robot answers particular questions put before it by extracting features from the asked questions and providing it as input to the neural network for recognition. The performance of the robot is based on the decoding capabilities of the received speech and the amount of response by the robot in answering the learners. Confusion matrix have been evaluated and the Receiver Operating Characteristics(ROC) curve indicates that the proposed system is acceptable as the False positive rate versus True positive rate curve lies above the diagonal which is an acceptable region. The proposed supervised classifier method improves the robot teaching process, and many of the schools and colleges are likely to use this proposed method. This process is efficiently performed by the robot and it provides an effective learning means for learners. The percentage of question recognition resulted in 74%.

References

[1] Greg Preston, Rob Phillips, Maree Gosper, Margot McNeill, Karen Woo, and David Green, “Web-based lecture technologies: Highlighting the changing nature of teaching and learning”, Australasian Journal of Educational Technology, vol. 26, pp. 717-728, 2010.

[2] J.Han, M. Jo, S. Park, and S. Kim, “The Educational Use of Home Robots for Children”, in Proceedings of IEEE International Workshop on Robot and Human Interactive Communication ROMAN, 2005, vol. 13, pp.378-383.

[3] J-H. Han, M-H. Jo, V. Jones, and J-H. Jo, “Comparative study on the educational use of home robots for children”, Journal of Information Processing Systems, vol. 4, pp. 159-168, 2008.

[4] Hiroshi Ishiguro, Tetsuo Ono, Michita Imai, Takeshi Maeda, Takayuki Kanda, and Ryohei Nakatsu, “Robovie: an interactive humanoid robot”, International Journal of Industrial Robot, vol. 28, pp. 498 – 504, 2001.

[5] T. Kanda, T. Hirano, D. Eaton, and H. Ishiguro, “Interactive Robots as Social Partners and Peer Tutors for Children: A Field Trial”, International Journal of Human-Computer Interaction, vol. 19, pp. 61-84, 2004.

[6] Renals Steve, McKelvie David, and McInnes Fergus, “A comparative study of continuous speech recognition using neural networks and hidden Markov models”, in Proceedings IEEE International Conference on Acoustics, Speech and Signal Processing ICASSP, 1991, vol. 1, pp. 369-372.

[7] Vicki Jones, Jun Jo, and Philippe Martin, “Future Schools and How Technology can be used to support Millennial and Generation-Z Students”, in International Conference Proceedings of Ubiquitous Information Technology, 2007, pp. 886-891.

[8] Young-Jun Lee, Kyoung Kim, Heon-Chang Yu,Bo-Kyung Kye, and Beom-Seok Ko, “A Study on the Educational Application of Intelligent Robots and Appropriate Functionalities of Educational Robots”, International Journal of KERIS, vol. 5(2), pp. 22-25, 2008.

[9] Zhen-Jia You, Chi-Yuh Shen, Chih-Wei Chang, Baw-Jhiune Liu, and Gwo-Dong Chen, “A Robot as a Teaching Assistant in an English Class”, in Proceedings of the Sixth IEEE International Conference on Advanced Learning Technologies ICALT'06, 2006, pp. 87-91.