Journal of Theoretical and Applied Information Technology
15 August 2012. Vol. 42 No.1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

INCORPORATING FUZZY INFORMATION FOR THE EVALUATION OF CONFIDENCE MEASURE IN DETECTION OF PUTATIVE ERRORS AND OOV WORDS IN ASR SYSTEM

[1]Dr.C.P. SUMATHI, [2]V. MEENAKSHI, [3]Dr.T. SANTHANAM
[1] Assoc. Prof., Dept. of Comp.Sci., S.D.N.B. Vaishnav College for Women, Chennai-33, TN, India
[2] Asst. Prof., Dept. of Comp.Sci., Govt. Arts College [Autonomous], Salem-7, TN, India
[2] Assoc. Prof., Dept. of Comp.Sci., D.G.Vaishnav College, Chennai-33, TN, India
Email : [1]santsum@hotmail.com, [2]vaimeena@yahoo.co.in

ABSTRACT

Confidence Measure (CM) in Speech Recognition System (SRS) provides the information about how much confident the recognizer in recognizing the word. Accurate recognition of Automatic Speech Recognition (ASR) is one of the most difficult problems in speech recognition today. When speech is produced in a carefully planned manner, ASR systems are successful at accurate recognition and transcription. In this work, Fuzzy Reasoning Scheme is proposed to perform the information compilation step that includes recognition related features that combines through a compilation mechanism, into a more effective way to distinguish between correct and incorrect recognition results. A novel method is described to combine different knowledge sources and estimate the confidence in a word hypothesis via Fuzzy Inference System (FIS) and measure the joint performance of recognition and confidence systems. Here the HCM (Hybrid Confidence Measure) which is a combination of Acoustic CM and Phone Duration based CM and Likelihood Score Ratio (LSR) are incorporated with Fuzzy if-then rules to add up recognition information into CM. Definitions and Algorithms are illustrated with results on the HUB4_trigram corpus. Experimental result shows higher performance of this approach compared against standard compilation methods in rejection of putative errors and detection of Out-Of-Vocabulary (OOV) words. Fuzzy Inference Systems (FIS) represent a natural way to increase the performance of Confidence Measure (CM). This approach treats the uncertainty of recognition hypotheses in terms of "possibility" contrasted to the "probability" notion of previous works. Different features are incorporated with FIS that produces better results of recognition.

Keywords: *Confidence Measures (CM), Out-Of-Vocabulary (OOV) words, Fuzzy Inference System (FIS), Hybrid Confidence Measure (HCM), Likelihood Score Ratio (LSR)*

1. INTRODUCTION

In order for speech recognition technology to be viable and useful in everyday applications (*e.g.* meeting transcription, telephone-based systems), there is a need to develop methods to improve recognition accuracy on speech recognition and obviously there comes the need to develop effective method for word and sentence level CM (Fei Huang and Watson, 2009). The objective of this research work is to improve the strength and robustness of core speech recognition technology. In this paper, two confidence measures (CMs) in speech recognition are illustrated: one based on acoustic likelihood and the other based on phone duration (Joel Pinto and Sitaram, 2005). For a decoded speech frame aligned to an HMM state, the CM based on acoustic likelihood depends on the relative position of its output likelihood value in the probability distribution of likelihood value in that particular state. The CM of whole phone is the geometric mean of CMs of all frames in it. The CM based on duration depends on the deviation of the observed duration from the expected duration of the recognized phone. The two CMs are combined to produce HCM (Hybrid Confidence Measure) which in turn has been incorporated with Likelihood Score Ratio (LSR) using FIS that generates "if-then" rules to produce a resultant CM. This CM shows significant improvement over the CM based on earlier ones.

In the case of human to machine interaction, less intelligible speech can be dealt with CM. CM

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

assigns a degree of confidence to the recognized words. Using CM, Automatic Speech Recognition (ASR) could identify the words which are likely to be erroneous and the application using ASR could use corrective action. Starting with a brief overview of ASR systems, a short discussion of CM is made based on log-likelihood score and its shortcomings as a confidence measure. Use of acoustic CM and phone duration based CM is also covered. Next is the discussion of incorporating FIS technique to combine the outputs of multiple recognition systems and finding out the best overall hypothesis.

Speech Recognition System (SRS)

Speech recognition systems follow the standard, two-stage pattern classification paradigm. Stage one is to extract relevant features from the observed signal, and the second stage is to make some decision based on the features.

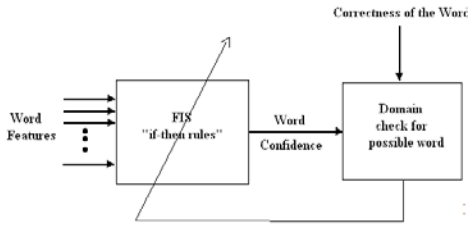

Figure 1. Block diagram in finding CM based on FIS

Fig. 1 shows the block diagram of FIS incorporated CM. Different word features are fed into the fuzzy ”if-then” block and the resultant CM is checked for the domain match. The corresponding word is checked for the correctness of the word based on the Confidence Confirmation Algorithm.

For recognition purposes, a speech utterance is modeled as a sequence of sound units. The speech pattern classification engine attempts to automatically identify the correct sequence of sound units found in the speech signal based on the observed sequence of feature vectors. Typical recognition system use the phonemes in the language as basic sound units, but other units of varying durations are possible (*e.g.* phoneme sequences, syllables, words, word compounds). CM attempts to assign trust to the hypotheses produced by speech recognition system. In SRS, various features (Meenakshi and Shanthi, 2007(a)) are considered in word level as well as utterance level to find the effective CM.

Nowadays, to certain degree, the capability to evaluate reliability of speech recognition results has been regarded as a crucial technique to increase usefulness and ‘‘intelligence’’ of an ASR system in many practical applications.

Confidence Measures

In this area, researchers have proposed to compute a score (preferably between 0 and 1), called confidence measure (CM), to indicate reliability of any recognition decision made by ASR systems(Hui Jiang, 2005),(Sangkeun Jung, et al.,2008). For example, a CM can be computed for every recognized word to indicate how likely it is correctly recognized or for an utterance to indicate how much the result for the utterance as a whole can be trusted. Despite a large amount of research efforts in the past, it is still believed that robust speech recognition and CM will remain as two most active and influential research topics in speech community for a foreseeable future. Due to importance of CM in ASR systems, it has attracted considerable research attention from major speech research groups all over the world and an excessive amount of research works have been reported in the past decade.

2. MATERIALS AND METHODS

In this paper, for each recognition hypothesis, a set of CM are computed and combined together into a confidence feature vector. The features which are utilized are chosen because, either by themselves or in conjunction with other features, they can be shown to be correlated with the correctness of a recognition hypothesis (Meenakshi.V and Shanthi.V, 2007(b)). The feature vectors for each particular hypothesis are then passed through a confidence scoring model which produces a single confidence score based on the entire feature vector. This score can then be evaluated by an FIS which produces combined CM for the hypothesis. This approach is utilized in this work for both utterance level and word level confidence scores. In this research work, two CM are proposed, one based on acoustic likelihood value and the other based on phone duration. The two CM are combined to obtain a hybrid CM.

Acoustic Confidence Measure

The triphone is normally modeled by an *N*-state left-right HMM (Fei Huang and Watson, 2009). The output density in an HMM state is modeled as a Gaussian mixture. For state *j* in triphone *i*

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

denoted by *state*(*i, j*), the output density $b_{ij}(O)$ is given by:

$$b_{ij}(O) = \Sigma\ W_{ijk}\ N(\ \mu_{ijk}U_{ijk}\ ;\ O)$$

where Σ varies from k = 1 to M, whereas, *M* is the number of mixtures, W_{ijk} is the k^{th} mixture weight and $N(\mu,U\ ;\ O)$ is the unimodal Normal density with mean μ, covariance matrix *U* and is given by:

$$N(\mu,U;O) = (2\Pi)^{N/2}|U|^{-1/2}\exp(-1/2(O-\mu)^{T}U^{-1}(O-\mu))$$

ASR returns the recognized word sequence as well as HMM state sequence. Suppose that the t^{th} speech frame O_t is aligned to *state*(*i, j*) and has an output likelihood value of $b_{ij}(O_t)$, the new acoustic CM c^A_t for that frame is defined as:

$$C_t^A = P[B_{ij} = b_{ij}(O_t)]$$

Where B_{ij} is a single dimensional random variable denoting the output likelihood value of feature vectors that are correctly aligned to *state*(*i, j*). The CM c_t^A is the probability that the output likelihood value in *state*(*i, j*) is lesser than the observed test vector likelihood $b_{ij}(O_t)$. The CM for a phone is computed as the geometric mean of the CM of the speech frames in the phone. If a phone *p* has T_p frames, its acoustic CM, C_p^A is given by

$$C_p^A = \exp\ (1/T_p\ \Sigma\ \log\ C_t^A)$$

where Σ varies from t= 1 to T_p

Duration based Confidence Measure

Several approaches have been tried in the past to use phone duration as a feature in utterance verification (Koo., et al., 2001), and word and phone level acoustic Confidence scoring (Kamppari and Hazen, 2000). Let *D* be the discrete random variable denoting the phone duration (in terms of number of frames), $p_D(n)$ the probability mass function (pmf) of D and μD the expected duration of phone. Suppose *d* is the observed duration of the recognized phone, the new confidence measure is based on the deviation ($d' = |d-\mu D|$) of the observed duration from the expected duration as opposed to directly using $p_D(n = d)$. If the random variable $D' = |D-\mu D|$ denotes the deviation, the duration based CM can be defined as:

$$C_pD = 1 - \Sigma\ P_D(n)$$

where Σ varies from [μD - | $d - \mu D$|] and [μD + | $d - \mu D$|]. Closer the value of observed duration to its expected duration, the higher is the duration confidence measure of that phone. To evaluate the duration pmf of each triphone, the training data to its correct transcript is to be aligned and obtain the histogram of the phone duration is obtained. The histogram is then smoothened and normalized to get the pmf $p_D(n)$.

Hybrid Confidence Measures(HCM)

Weighted geometric mean is used to combine the acoustic and duration CM to obtain a hybrid phone CM.

$$HCM = \exp\ (Wa\log\ (CpA) + (1\text{-}wa)\log\ (CpD))$$

Where w_a is the acoustic CM weight factor. The word confidence measure CM is the geometric mean of the hybrid phone confidence measures of the constituent phones.

3. CONFIDENCE MEASURE USING FIS

Probability is useful when dealing with serial events that require an enumeration notion of uncertainty but is not very useful when the uncertainty is about the degree of accomplishment of a known situation(Laviolette, and Seaman, 1994). This is the case of CM where the task is to know for every single recognition hypothesis, its degree of possible correctness. In FIS the notion of "possibility" is taken as the advantageous one as opposed to "probability" notion of other approaches. The spirit of CM is to express the uncertainty of speech recognition results in gradual terms and not in frequency. Under such consideration, fuzzy logic represents natural foundations for confidence measures.

But fuzzy logic is not just a theoretic tool to represent uncertainty, a good share of its success is due to the several practical implementations it has. Fuzzy logic systems or Fuzzy Inference Systems (FIS) are schemes that allow to map a number of fuzzy variable inputs into a number of fuzzy outputs (Mendel, 1995). The mapping is done by a set of fuzzy rules that relates inputs with outputs in an "if … then" fashion. Inputs and outputs can be represented by means of fuzzy variables able to contain language terms and fuzzy hedges. By analyzing the histograms of each of the features that are proposed, some characteristics of them are observed and some fuzzy thresholds to separate their values when there is a correct or incorrect result are proposed. This analysis allows us to define some rules of behavior according to the correctness status of the hypotheses. The collected expert knowledge can be condensed in a fuzzy inference system. In this application, the fuzzy system can be understood as a non-linear classifier (just as a neural network) that transforms several inputs into a unique output that compiles all the information given.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

4. FEATURE COMPILATION SCHEMES

Likelihood score ratio (LSR)

Likelihood score ratio (LSR) is a feature in which the recognition hypothesis is normalized by the score of an alternative recognition network:

$$LSR = \log L(\vec{X}|\wedge_p) - \log L(\vec{X}|\wedge_a)$$

X is the vector of acoustic features related to the actual input utterance and $\wedge_p$ and $\wedge_a$ are the sets of hidden Markov models of the principal and alternative recognition networks respectively. Due to its unconstrained (and inaccurate) nature, the purpose of the alternative network is to model the unrestricted signal probability, P(X). This procedure tends to approximate Bayes law in posterior probability calculation. Because of its simplicity and its high performance, this feature has been taken as the baseline. It is customary that the compilation step of the information included in the recognition features is performed by means of a uniting tool based on the development of conditional probabilities. In such a way, Bayesian classifiers, linear discriminative analysis, decision trees and neural networks (Weintraub, et al., 1997) have been used as reasoning schemes to compile the involved features. Here in this paper, some classifiers, based on some of the schemes are compared with the performance against fuzzy systems.

Bayesian classifier (BC)

This is a rather simple classifier that maps recognition features into a confidence measure by means of a linear combination. The coefficients for such a combination are calculated from the covariance matrix of the features derived from training data. The fundamental rule in statistical speech recognition is the Baye's rule given by:

$$W_{opt} = \arg\max_w P(O/W)\, P(W)$$

The recognized word sequence *Wopt* is the one which maximizes the posterior probability $P(W \mid O)$, where $p(O \mid W)$ is the acoustic model, $P(W)$ is the language model and $p(O)$ is the unconditional acoustic likelihood of the observation sequence. While decoding, the unconditional acoustic likelihood $p(O)$ is normally omitted since it is invariant to the choice of a particular word sequence. As a result, the acoustic score $p(O \mid W)$ obtained during recognition will be unnormalized and cannot be used as a measure of confidence. Different approaches have been tried to approximate $p(O)$ to obtain the correct CM.

Neural Networks (MLP)

Neural networks have been broadly used to combine recognition features into CM (Necip Fazil, 2005). High performing results have been reported and their advantages over other combination systems have been largely discussed. Network topology is always a delicate issue. Remarkable results have been achieved with multi-layer perceptrons (MLP) when trained under a back propagation framework. Simpler configurations have been preferred instead of complicate ones since performance is quite similar. For experimentation, a feed-forward MLP with 1 hidden layer containing 4 to 6 elements are chosen, each with a hyperbolic tangent sigmoid transfer function. The input layer deals with the values of the features and the output layer deals with the CM value. The parameters of the net are adjusted in a back propagation learning phase taking examples from a training database.

5. FUZZY INFERENCE SYSTEM (FIS)

Fuzzy inference system, as a classification engine, can be equipped with expert knowledge capable to separate class elements. A Sugeno-type FIS is chosen due to its good behavior as classifier and its simplicity (Jang, 1993). The number of input variables depends on the number of features used. The output variable is the value of the CM. The fuzzy rules are designed with a "reinforcement" spirit. Likelihood score ratio is treated as the main discrimination variable and the rest of the features are employed to reinforce its values. The rules of the FIS allows to activate and deactivate the influence of the reinforcing features conveniently. The set of "if-then" rules for this fuzzy system is illustrated in Fig. 2. The consequent parts of the rules are constants, but it is proved that, even with this simple configuration, FIS performs better as uniting tool for CM.

```
If LSR = Low  and HCM = Low     then CM =    0
If LSR = Low  and HCM = MidLow  then CM =  0.05
If LSR = Low  and HCM = MidHigh then CM =  0.20
If LSR = Low  and HCM = High    then CM =  0.30
If LSR = High and HCM = Low     then CM =  0.60
If LSR = High and HCM = MidLow  then CM =  0.80
If LSR = High and HCM = MidHigh then CM =  0.90
If LSR = High and HCM = High    then CM =     1
```

Figure 2. Set of fuzzy rules for FIS of two features

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

6. DATABASE

Experimental framework

SPHINX-4 engine was trained, for both the language model and the lexicon modules as in architecture SPHINX-4. SPHINX-4 engine is trained in order to develop acoustic models. Training is based on an HMM model that is built on statistical information and random variables. The major contribution is mainly using the open source SPHINX-4 model in speech recognition. The system is fine-tuned and data are refined for training and validation. Optimum values for number of Gaussian mixtures distributions and number of states in HMM's have been found according to specified performance measures. Optimum values for confidence scores are found for the training data. Experimental work has been carried out to test the proposed CM using open source speech recognition toolkit Sphinx-Train and Sphinx-4 Decoder and the relevant results are discussed.

Results

Experiments were conducted on speaker independent continuous speech recognition task. The vocabulary consisted of Hub4 corpus of trigram language model for training the ASR as well as evaluating the performance of the CM. The training set consisted utterances from 225 speakers and the test set consisted of 1250 utterances from 223 speakers. Mel Frequency Cepstral Coefficients (MFCCs) were used as features for speech recognition. The speech signal sampled at 16 kHz is frame blocked with a window length of 20 msec and frame shift of 10 msec. The 13-dimension MFCC vector, delta coefficients and delta-delta coefficients form a 39-dimensional feature vector. Triphone is used as the basic speech modeling unit, modeled by a 5-state left-right HMM. The output density in each state is modeled as mixture of 4 Gaussians.

Confidence Measure Results

Table 1 compares Mean Square Error (MSE) and Classification Error Rate (CER) computed from the priors only and from the posteriors outputted by the neural network. The decrease in MSE from priors to posteriors indicate that the average estimation of the word confidence (MSE from priors)$^{1/2}$ – (MSE from posteriors)$^{1/2}$ improved by which is roughly at 14% relative improvement on the test data. For the same dataset, CER decreased by 44%. Table 1 and Table 2 represent the training and testing set respectively.

Table 1: MSE and CER for priors only and for the FIS outputs on the Hub4 corpus – Training Set

	MSE	CER
From Priors	0.2354	47.52%
From Posteriors	0.1739	26.25%

Table 2: MSE and CER for priors only and for the FIS outputs on the Hub4 corpus – Testing Set

	MSE	CER
From Priors	0.2362	48.02%
From Posteriors	0.1775	26.89%

To evaluate the performance of confidence metrics, hypothesized words are compared against the true transcription of the utterance with each hypothesized word being classified as *correct* or *incorrect*. The confidence scores for each word are then compared against a confidence threshold and the hypothesized words are either *accepted* or *rejected*. The threshold can be varied to control the tradeoff between false alarms (incorrect words that are accepted) and detections (correct words that are accepted).

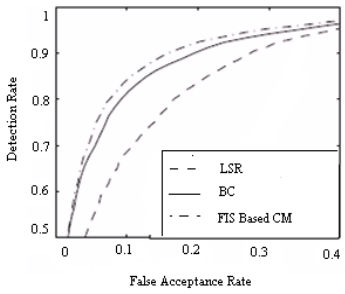

Fig.3: ROC curve for CM based on FIS, BC and LSR

By varying the confidence score threshold, a Receiver Operating Characteristic (ROC) curve can be plotted. Fig 3 illustrates ROC curve for CM based on FIS, BC and LSR.

Journal of Theoretical and Applied Information Technology
15 August 2012. Vol. 42 No.1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

7. DISCUSSION

To evaluate the performance of the CM, recognized words are compared against the correct transcription of the utterance and each word is classified as *correct* or *wrong*. The confidence measures of hypothesized digits are compared against a threshold to either *accept* or *reject* the digit. Receiver Operator Characteristics (ROC) - plot of the detection rate versus the false acceptance rate - is plotted by varying the confidence threshold between 0 and 1. A confidence measure is good if it has higher detection rates at lower false acceptance rate. The proposed confidence measures are tested for its efficiency in detecting putative errors (erroneous but in-vocabulary words) as well as OOV words. The OOV word modeling approach operates during the recognition search process by allowing the recognizer itself to hypothesize a generic OOV word model (Hazen and Bazzi, 2001) as an alternative to a known word. On the other hand, this confidence scoring approach is applied as a post-processing technique after the recognition search is already complete. A natural way to combine both methods is to enable OOV word detection during recognition and then utilize confidence scoring on the hypothesized known words (excluding the OOV word hypotheses) after recognition is complete. Table 3 compares the performance of different CM in rejecting the putative errors. The proposed CM have out-performed the baseline CM. Also, the hybrid CM (*wa* = 0.8) based on acoustic likelihood as well as phone duration has performed better than the CM based only on acoustic likelihoods (*wa* = 1.0).

Table 3: Performance of the CMs in rejecting putative errors. Detection rates for false acceptance rates of 30, 20 and 10%

Confidence Measure	Word Detection Rate		
	30%	20%	10%
LSR (Baseline) CM	89.3	80.4	65.1
Bayesian Classifier	91.5	89.5	80.4
FIS incorporated CM	93.2	90.7	82.1

To evaluate the performance of the proposed CM in detecting the OOV, recognition errors in the following manner are simulated. Table 4 shows the overall performance of the CM in detecting OOV. The result clearly indicates that the FIS method is having an edge over the other. Performance can also be measured in terms of a Figure Of Merit (FOM), which measures the performance of a system at or around a particular operating point on the curve. In this system it is desirable to maintain a high detection rate at the expense of increased false alarms. To capture this condition the FOM measures the area under the ROC curve in the range of .8 to 1.0 for correct acceptances.

Table 4: Performance of the CMs in rejecting OOV. Detection rates for false acceptance rates of 30%, 20% and 10%

Confidence Measure	Word Detection Rate		
	30%	20%	10%
LSR (Baseline) CM	90.2	84.4	65.8
Bayesian Classifier	93.8	90.5	80.7
FIS incorporated CM	95.2	92.1	84.1

This area is then normalized by the total area in this range to produce an FOM whose optimal value is 1. The goodness of the CM is given by FOM which is the area under ROC curve. From Fig. 4, it is seen that FIS incorporated CM produces the better performance .

CONCLUSION

In this paper, FIS represents a natural and effective approach to measure the CM of SRS. The way it handles uncertainty, in place of OOV occurances, results are better consistent that the

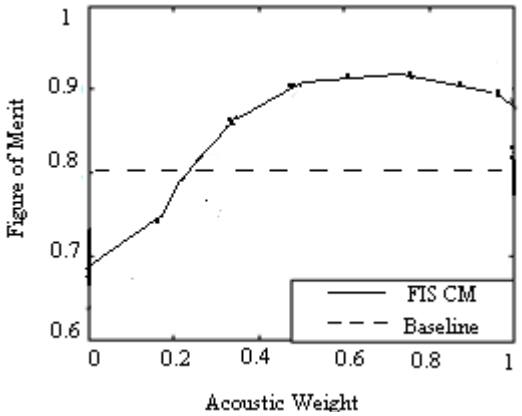

Figure 4. FOM of the FIS incorporated CM as against Baseline CM

way probability theory does. The confidence in a word hypothesis is defined as the posterior probability that the word is correct. Classification Error Rate and Mean Squar Error have been used as the criteria to measure the performance of a word confidence. Word confidence are estimated with FIS that combines various knowledge sources relative to the words and to the hypotheses. The combination of several features significantly improved our confidence estimates. Investigation

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

are made on HCM which incorporates acoustic likelihood and another based on the duration of recognized phone. LSR is also taken as another feature to score CM. Finally FIS rule play the role based on "if-then" rules which ultimately produces CM which is having an edge over the Baseline techniques. In detection of putative errors and OOV, this method produces some marginal improvement. FIS have been efficiently used to compile features related to the recognition process into a more discriminative CM. From ROC it is clear that when compared with Baye's classifier, FIS produces stable behavior and is able to produce high detection rates while properly rejecting false alarms.

REFERENCES:

[1]. Fei Huang and Watson. T.J., 2009. "Confidence Measure for Word Alignment" *Proceedings of the 47th Annual Meeting of the ACL and the 4th IJCNLP of the AFNLP*, pages 932–940, Suntec, Singapore.

[2]. Joel Pinto and Sitaram. R.N.V, 2005, "Confidence Measures in Speech Recognition based on Probability Distribution of Likelihoods", *HP Laboratories India*.

[3]. Meenakshi.V and Shanthi.V, 2007(a) "Multiple Hypotheses in Confidence Scoring Measures for Spoken Dialog System", *Asian Journal Information Technology 6(5): 647- 651, Medwell Journals.*

[4]. Hui Jiang, 2005, "Confidence measures for speech recognition: A survey" *Elsevier, Speech Communication*, pages 455-470.

[5]. Sangkeun Jung, Cheongjae Lee and Gary Geunbae Lee, 2008, "Using Utterance and Semantic Level Confidence for Interactive Spoken Dialog Clarification" *Journal of Computing Science and Engineering*, Vol. 2, No. 1, Pages 1-25.

[6]. Meenakshi.V and Shanthi,V, 2007(b), "Evaluating Words of Discriminative Power in Automatic Speech Recognition System", *Information Technology Journal, ISSN 1812-5638 Asian Network for Scientific Information.*

[7]. Koo, M-W., Lee, C.-H., and Juang, B.-H., 2001, "Speech Recognition and Utterance Verification Based on Generalized Confidence Score", *IEEE Trans. Speech and Audio Proc*., 9(8): 821–832..

[8]. Kamppari, S.O. and Hazen, T.J., 2000, "Word and Phone Level Acoustic Confidence Scoring", *ICASSP*, 3:1799–1802.

[9]. Laviolette. M and Seaman Jr. J.W., 1994, "The efficacy of fuzzy representations of uncertainty",IEEE Transactions on Fuzzy Systems, 2(1):4–15.

[10]. Mendel. J.M.,,1995, "Fuzzy logic systems for engineering: a tutorial". *Proceedings of the IEEE*, 83(3):345–377.

[11]. Weintraub M, Beaufays.F. et al., 1997, " Neural – network based measure of confidence for word recognition. In *Proceedings of 1997 ICASSP*, volume II, pages 887–890, Munich.

[12]. Necip Fazil Ayan, Bonnie J. Dorr, and Christof Monz. 2005. "Neuralign: Combining word alignments using neural networks", In *Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing*, pages 65–72, Vancouver, British Columbia, Canada, October. Association for Computational Linguistics.

[13]. Jang. J.S.R, 1993 "Anfis: Adaptative-network-based fuzzy inference system", *IEEE Transactions on systems, man and cybernetics*, 23(3):665–685.

[14]. Timothy J. Hazen and Issam Bazzi, 2001, **"**A Comparison and Combination of Methods for OOV Word Detection and Word Confidence Scoring", *Proc. ICASSP2001*, May 7-11, 2001, Salt Lake City,IEEE.