Performance Analysis of Learning Classifiers for Spoken Digit Under Noisy Conditions
1 Syed Abbas Ali, 2 Samreen Burkiand, 3 Sara Hasan
1 Dept. of Computer & Information Systems Engineering, N.E.D University of Engineering & Technology, Karachi 2, 3 Dept. of Telecommunications Engineering, N.E.D University of Engineering & Technology, Karachi

ABSTRACT
Speech Recognition is truly challenging task due to the presence of ambient noise. Ambient noise can decrease the Intelligibility and reliability of the speech recognition system which causes misinterpretation of speech sound. Most of the practical applications of speech processing, like audio conferencing, cellular communication and the recorded speech contain a significant amount of ambient noise. In this paper, we evaluate the performance of different classifiers (such as J48, Naïve Bayes, Multiclass and Multilayer Perceptron) in term of classifiers recognition rate for recorded isolated digit and TIDIGIT corpus. The noise and speech samples taken from NOISEX-92 database and TIDIGIT corpus respectively, and recorded isolated digits (0-9) taken from real environment. The experimental results clearly show that the Multilayer Perceptron and Naïve Bayes classifiers indicate maximum recognition rate in comparison with Multiclass classifier, while J48 classifier shows minimum recognition rate for recorded isolated digits and TIDIGIT corpus.

Keywords: Learning Classifier, Recognition Rate, Ambient noise, WEKA Data Mining, Hidden Markov Model (HMM)

1. INTRODUCTION
In most of the practical applications of automatic speech recognition, the input speech is contaminated by different ambient noises, which strongly degrades the performance of speech recognizers [1]. Noise is time–varying, unpredictable and has temporal characteristics in nature. One of the significant issues in robust speech recognition is to develop a precise noise model. Most of the time ambient noise is non-stationary in nature and it originates from different environmental sources. Quality of speech signal degrades due to ambient noise which may lead inaudibility and decreases the preference of speech coding, synthesis and speech recognition system. Due to this reason the robust speech recognition systems is still a challenging task and research is being in progress. Speech recognition System can be divided into two phases. The initial phase is Feature Extraction and the next phase of recognition system is classification. Different approaches have been applied in reducing the effects of noise on the acoustic speech signal captured through microphone, results of numerous studies have established that the speech recognition system performance degrades intensely, when training condition differ from the testing condition [2-5]. This paper is an attempt to evaluate the performance of learning classifiers in term of classifiers recognition rate for recorded and TIDIGIT corpus in the presence of different noises. Rest of the paper is organized as follows. In section II, we briefly discuss the research methodology. Experimental results discussed in section III. Finally, conclusion is drawn in section IV.

2. METHODOLOGY
To elaborate the experimental framework for performance evaluation of learning classifiers, we have divided our work flow diagram as shown in Fig.1 into three different phases:

1. Feature Extraction Using MFCC (Mel-frequency cepstral coefficients)
2. Estimation of likelihood of clean and noisy isolated digit Using Hidden Markov Model (HMM)

Fig 1: Experimental Frame work

2.1 Feature Extraction Method (MFCC)
The virtues of feature extraction where the original variables are taken but processed into a smaller set to retain as much information as possible and feature selection which removes input variables that do not contribute meaningfully to model performance. The feature vector comprises of information which is suitable to recognize and differentiate speech samples which is unresponsive to speaker originality and other unsuitable factors. The Mel-Frequency Cepstral Coefficients (MFCC) feature extraction method is an important approach for feature extraction of speech sample and recent research objective is to identify performance enhancements [6, 7]. In sound processing, the Mel-frequency cepstrum (MFC) is a representation of the short-term power spectrum of a noise based on a linear cosine transform of a log power spectrum on a nonlinear Mel scale of frequency. Mel-frequency cepstral
coefficients (MFCCs) are coefficients that collectively make up an MFC. They are consequent from a type of cepstral representation of the audio clip ("spectrum-of-a-spectrum"). The differentiation between Mel-frequency cepstrum and the cepstrum is that in the MFC, the frequency bands are uniformly spaced on the Mel scale which approximates the human acoustic system's response more closely than the linearly-spaced frequency bands used in the normal cepstrum. This frequency warp can allow for better representation of speech. MFCCs are commonly consequential as follow shown in Fig.2.

- Taking Fourier transform of input signal.
- Plotting the power spectrum onto the Mel scale by using Triangular Overlapping window.
- Taking log values of power for each Mel frequencies
- Taking discrete cosine transform of Mel log power spectrum
- The amplitudes of the resulting spectrum called MFCCs.

![Fig 2: Mel-Frequency Cepstral Coefficients](image)

### 2.2 Likelihood Estimation (HMM)

HMM is doubly stochastic process with an underlying stochastic process that is not noticeable, but can only be observed through another set of stochastic processes that produce sequence of observed symbols [8]. HMM is characterized by following elements as shown in Fig.3. Where N is the number of state, V is the number of distinct observation symbol per state M, a_{ij} state transition probability, B is the observation symbol probability distribution in state j and \( \pi_i \) is the first state distribution vector.

![Fig 3: Elements of HMM](image)

The likelihood of a HMM can be expressed in relatively general framework. Let \( \theta \) be all model parameters sets and let \( P(y_t) \) is a diagonal matrix with the conditional probabilities \( b_j(y_t) = P(Y_t = y_t | S_t = j) \), \( j = 1, \ldots, m \) on the main diagonal. The Likelihood of a HMM can be written as \( L(\theta) = P(Y_0 = y_0, \ldots, Y_{\tau-1} = y_{\tau-1}) \). This form of the likelihood has several appealing properties. For example, stationary as well as non-stationary models can be handled and a (local) maximum can be found by numerical procedures such as Newton-type algorithms or via the so called EM-algorithm [10].

### 2.3 Learning Classifiers

We used J48, Naïve Bayes, Multiclass and Multilayer Perceptron classifiers using WEKA Data Mining software to evaluate the classifiers recognition rates. In order to compare the results obtained from recorded isolated digit and TIDIGIT corpus, different environmental noises merged in a clean recorded isolated digit to deal with different noisy digit samples. We made use of HMM tool to estimate likelihood of recorded isolated digit (0-9) samples including both clean and noisy condition. Different learning classifiers applied on the likelihood estimates obtained from HMM to make comparison between clean isolated digit classifier recognition rates with the noisy digit samples and the entire process repeated for the pre-conditioned TIDIGIT corpus to analyze how much result differ from the recorded isolated digit in term of classifier recognition rate.

### 3. EXPERIMENTAL RESULTS

Experimental frame work was divided into two phases to evaluate the performance of learning classifiers in the presence of ambient noises. First phase of experiments comprise of feature extraction using MFCC and estimation of likelihood of isolated digit taken from TIDIGIT corpus and real environment with and without noise (Clean isolated digit) using HMM tool box in MATLAB. The Experiments were done using MATLAB tool version 10.0 and speech processing toolbox.
speech database used in the experiments taken from TIDIGIT corpus which contains the isolated digits in English language[11]. The TIDIGITS corpus comprises of more than 25 thousand digit sequences spoken by over 300 women, children and men. The data was collected in quite environment and digitized at 20 KHz. In this experiment, we used 10 utterances of each isolated digit (0-9) in an approximately clean environment with sampling frequency of 8KHz. Different types of noises such as pink noise; white noise; babble noise and brown noise were obtained from NOISEX-92 noise-in-speech database [12]. For the recording specification of recorded isolated digit corpus, ITU recommendations based standardized procedure was adopted for speech corpora development. The recording has been done in standard recording environment having SNR>45dB. We made use of Microsoft Windows 7 built-in sound recorder to record the 10 utterances of each isolated digit (0-9). The recording format is Mono, 32 bit PCM with sampling rate of 8000Hz using microphone with impedance of 32 Ω, Max Input power=40mW, Drive Unit=30mm, Plug Type=3.5MM, Frequency Response=20Hz ~ 20 KHz. In the second phase of experiment, we made use of different learning classifiers (such as J48, Naive Bayes, Multiclass and Multilayer Perceptron) using WEKA Data mining software to evaluate the classifiers recognition rate. After a model has been processed by using the training set (Consist of recorded and TIDIGIT isolated corpus), we examined the model by making estimate against the test set. We divided the sample withdrawal wizard into two sets: one with 30 percent of the source data, for testing the shape, and one by 70 percent of the source data, for training the mockup. We configured the wizard to set a maximum number of training cases. When we identified both a maximum percentage and a minimum number of cases, we made use of smaller values of the two restrictions as the size of the test set to ensure that the size of our test set stay consistent even if more training data is added to the model. The following steps have been performed to analyze the experimental outcomes in WEKA Data mining software.

**Training Set:**
- Load the complete data sample for training
- Select the remove percentage value of filter in preprocessor unit
- Set the correct percentage for splitting purpose
- Selected values submitted to an application filter
- Collect the generated data

**Test Set:**
- Load the full data sample for testing
- Choose the remove percentage value of filter, if not chosen yet.
- Set the true value of invert selection property
- Selected values submitted to filter
- Collect the generated data

### Table 1: Classifier’s Recognition Rate for TIDIGIT Database

<table>
<thead>
<tr>
<th></th>
<th>J48 (%)</th>
<th>Naive Bayes (%)</th>
<th>Multi Class (%)</th>
<th>Multilayer Perceptron (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clean</td>
<td>83.3333 %</td>
<td>100 %</td>
<td>100 %</td>
<td>100 %</td>
</tr>
<tr>
<td>Babble</td>
<td>7.6923 %</td>
<td>92.3077 %</td>
<td>84.6154 %</td>
<td>92.3077 %</td>
</tr>
<tr>
<td>Pink</td>
<td>8.3333 %</td>
<td>100 %</td>
<td>91.6667 %</td>
<td>100 %</td>
</tr>
<tr>
<td>White</td>
<td>7.6923 %</td>
<td>92.3077 %</td>
<td>84.6154 %</td>
<td>92.3077 %</td>
</tr>
<tr>
<td>Brown</td>
<td>7.6923 %</td>
<td>92.3077 %</td>
<td>84.6154 %</td>
<td>92.3077 %</td>
</tr>
</tbody>
</table>

### Table 2: Classifier’s Recognition Rate for Recorded Digit

<table>
<thead>
<tr>
<th></th>
<th>J48 (%)</th>
<th>Naive Bayes (%)</th>
<th>Multi Class (%)</th>
<th>Multilayer Perceptron (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clean</td>
<td>70 %</td>
<td>100 %</td>
<td>100 %</td>
<td>100 %</td>
</tr>
<tr>
<td>Babble</td>
<td>9.0909 %</td>
<td>90.9091 %</td>
<td>0 %</td>
<td>90.9091 %</td>
</tr>
<tr>
<td>Pink</td>
<td>9.0909 %</td>
<td>90.9091 %</td>
<td>18.1818 %</td>
<td>90.9091 %</td>
</tr>
<tr>
<td>White</td>
<td>9.0909 %</td>
<td>90.9091 %</td>
<td>9.0909 %</td>
<td>90.9091 %</td>
</tr>
<tr>
<td>Brown</td>
<td>0 %</td>
<td>90.9091 %</td>
<td>0.9090 %</td>
<td>90.9091 %</td>
</tr>
</tbody>
</table>

Table 1 and 2, provide the performance evaluation in term of classifier recognition rate for recorded and TIDIGIT corpus in the presence of different noises. The values clearly show that the classifier recognition rates for pre-conditioned TIDIGIT corpus are far better than recorded isolated digit taken from real environment. For TIDIGIT corpus, Multilayer Perceptron displayed maximum recognition rate as compared to Naive Bayes and Multiclass classifiers, whereas j48 classifier indicated minimum recognition rate. In case of recorded isolated digit corpus, Naive Bayes and Multilayer Perceptron classifiers revealed maximum recognition rate than that of Multiclass classifier, while j48 classifier again demonstrated minimum classifier recognition rate.

**4. CONCLUSION**

In this paper, we evaluated the performance of different classifiers in term of classifiers recognition rate for recorded and pre-conditioned isolated digit corpus. Experiments have been performed with TIDIGIT corpus, NOISEX-92 noise-in-speech database and recorded isolated digit taken from real environment. We used J48, Naive Bayes, and Multiclass and Multilayer perceptron classifiers and evaluated the performance of classifiers in term of recognition rate using WEKA Data Mining tool in the presence of Babble, Pink, White and Brown noises. Experimental results evident that the Naive Bayes and Multilayer Perceptron classifiers show maximum recognition rate as compared to Multiclass classifier, while j48 classifier portray minimum recognition rate for recorded isolated digits and TIDIGIT corpus.
REFERENCES


AUTHOR PROFILES

Syed Abbas Ali, received M.Engg. Degree from NED University of Engineering & Technology, Karachi. His areas of interest are Machine Learning and Automatic Speech Recognition.

Samreen Burki, received B.E. Degree from NED University of Engineering & Technology, Karachi. Her areas of interest are Digital Signal Processing and Automatic Speech Recognition.

Sara Hasan, received B.E. Degree from NED University of Engineering & Technology, Karachi. Her areas of interest are Digital Signal Processing and Automatic Speech Recognition.