An Efficient Isolated Speech Recognition Based on the Adaptive Rate Processing and Analysis

Saeed Mian Qaisar
Electrical and Computer Engineering Department, Effat University, Jeddah, KSA
sqaisar@effatuniversity.edu.sa

Abstract—This paper proposes a novel approach, based on the adaptive rate processing and analysis, for the isolated speech recognition. The idea is to smartly combine the event-driven signal acquisition and windowing along with adaptive rate processing, analysis and classification for realizing an effective isolated speech recognition. The incoming speech signal is digitized with an event-driven A/D converter (EDADC). The output of EDADC is windowed with an activity selection process. These windows are later on resampled uniformly with an adaptive rate interpolator. The resampled windows are de-noised with an adaptive rate filter and their spectrum are computed with an adaptive resolution short time Fourier transform (ARSTFT). Later on, the magnitude, Delta and Delta-Delta spectral coefficients are extracted. The Dynamic Time Warping (DTW) technique is employed to compare these extracted features with the reference templates. The comparison outcomes are used to make the classification decision. The system functionality is tested for a case study and results are presented. An 8.2 times reduction in acquired number of samples is achieved by the devised approach as compared to the classical one. It attains a significant computational gain and power consumption reduction of the proposed system over the counter classical ones. An average subject dependent isolated speech recognition accuracy of 96.8% is achieved. It shows that the proposed approach is a potential candidate for the automatic speech recognition applications like rehabilitation centers, smart call centers, smart homes, etc.

Index Terms—Event-Driven Processing, Speech recognition, Adaptive Resolution Analysis, Features extraction, Dynamic Time Warping, Classification.

I. INTRODUCTION

The Human beings naturally communicate via speech. Speech not only carries the message to convey but it also possesses an important information about the speaker.

The automatic speech recognition process can be traced back in 1960s when computer scientists were discovering algorithms and methods to make computers able to record, interpret and comprehend the human speech. After decades of progress, the first systems were introduced in 1980s. However, these early systems were very limited in scope and performance.

The spoken language is a primary way of communication among humans. Therefore, people expect machines and computers, which can recognize the incoming speech and can act accordingly [1]. The automatic recognition of speech involves generating a sequence of words best matches the incoming speech signal.

The main challenges, faced by any automatic speech recognition system are the confusable sounds, speaker variability and homophones words [1].

The performance of speech recognition systems is typically measured in terms of accuracy. It can be presented in terms of the Word Error Rate and the Command Success Rate [1, 2].

The recent technological advancements have revolutionized the speech based human machine interaction [3, 4]. It has diverse applications like smart cities, secure access, rehabilitation centers, criminals investigation analysis, etc. [5-7]. These systems are founded on the machine based speech comprehension and recognition. The principle is to achieve an appropriate extraction of phonetic features form the incoming speech. Later on, these extracted parameters are employed to classify the incoming speech. It allows realizing effective and robust machine based speech recognition systems [8-10].

The classical speech recognition systems are time invariant in nature [8-10]. It can lead towards a useless usage of the system resources and power consumption [11-15]. An efficient system can be realized for such signals by adjusting the acquisition, processing and transmission rates as a function of the intended signal temporal variations [13-15]. In this framework, an EDADC is employed for the intended speech signals acquisition. It works on the opportunistic sampling principle and therefore allows the devised system to overcome the downsides of the counter classical ones up to a certain extent [16]. Therefore, it promises a simplified and power efficient front-end electronics along with a real time data compression [13-16].

This work focuses on intelligently employing the event-driven and adaptive arte signal processing and analysis tools. It is in the objective to achieve an efficient automatic isolated speech recognition.

The proposed system principle is described in Section II. Section II, also discusses the employed materials and methods, used to realize the devised system. A summary of the system performance verification results is discussed in Section III. Section IV, finally concludes the article.

II. THE PROPOSED SOLUTION PRINCIPLE

Figure 1, shows the proposed solution block level diagram. It shows that the intended speech signals are acquired via an EDADC. These pulses are firstly passed through an antialiasing
band-pass filter with \([F_{\text{Cmin}}=50; \ F_{\text{Cmax}}=3.5k]\)Hz. The band limited signal \(x(t)\) is acquired with a uniform quantization based 5-Bit resolution EDADC [15, 17, 18]. The EDADC output is time windowed with the Activity Selection Algorithm (ASA) [13, 14]. The windowed signal is resampled uniformly and then de-noised with an appropriate adaptive rate filtering algorithm [14]. The de-noised signal spectrum is computed with an ARSTFT [13]. The spectral magnitude and delta coefficients are extracted. These features are later on employed by the DTW module for the intended incoming speech recognition. A detail of different system modules is provided in the following subsections.

![Figure 1: The proposed system block diagram](image)

### A. The Event-Driven A/D Conversion (EDADC)

Figure 1, shows that in the studied case the band limited signal \(x(t)\) is acquired with a 5-Bit resolution, uniform quantization based, EDADC [15, 17]. The EDADC is developed on the principle of Level Crossing Sampling (LCS) [13-18]. The ADC is an essential component of any signal processing chain [11, 12, 19]. It dictates the overall system performance [19]. The Nyquist sampling and processing theory governs the functionality of traditional ADCs. The signal acquisition is performed at a constant frequency irrespective of its sporadic nature, i.e. without exploiting the local signal variations [13-18]. Consequently, design parameters of the traditional ADCs are chosen for the worst case [11, 19]. Therefore, in the case of low activity random signals like Speech, Seismic signals, etc., such ADCs are not effective [13-18]. The employment of EDADCs can treat this inadequacy up to a certain degree. The EDADCs are founded on the principle of opportunistic sampling and can adjust their sampling frequency according to the local variations of the intended signal. As a result, only the intended information is acquired. Therefore, a drastic reduction in the post processing and analysis chain, activity and power consumption is achieved. In this case, the sampling frequency is not fixed, and it is piloted by the input signal. However, the signal reconstruction is confirmed by respecting the Bernstein inequality [14]. It is realized by employing some appropriate design parameters of the employed EDADC [15].

### B. The Windowing, Resampling and Filtering

The EDADC output can be employed for further non-uniform digital processing. However, the practical system realization necessitates the finite time partitioning of the acquired data. Since a real system has limited resources like memory, processing speed, etc. [11, 12]. In this context, the ASA is employed for an effective windowing of the EDADC output. It exploits the sampling process non-uniformity to window only the relevant parts of signal. Furthermore, the characteristics of each selected signal part are analyzed in order to extract its local parameters. Later on, these extracted parameters are used to adjust the devised system parameters and processing activity accordingly [13, 14].

The output of ASA is resampled uniformly. The extracted window parameters are employed to decide the resampling frequency [11-12]. The resampler acts as a bridge between the non-uniform and the uniform signal processing domains. It allows to realize smart solutions by taking advantage of both verges [11-15]. The interpolation is required during the resampling process. The artifacts are introduced, because of interpolation, in the resampled signal as compared to the original one. In the suggested system, the resampling error is a function of the interpolation technique employed to resample the data and the EDADC resolution, \(M\), and amplitude dynamics, \(\Delta V\)[13-15]. The employed EDADC thresholds are distributed uniformly within the range, \(\Delta V\). Therefore, its quantum, \(q\), can be calculated as \(q=\Delta V/(2^M-1)\). For this case, the worst interpolation error is bounded by \(q\) [20-25].

In the devised system the Simplified Linear Interpolator (SLI) is employed for the resampling purpose. For the SLI the value of an interpolated sample \(x_{rn}\) corresponding to a resampling instant \(t_{rn}\) is set equal to the average of its prior and next non-uniform samples. The process can be mathematically expressed as: \(x_{rn}=(x_{rn-1}+x_{rn+1})/2\). For SLI the worst error per resampled observation is bounded by \(q/2\) [20-25].

The resampling frequency of each selected window is adapted by employing the selected window parameters extracted with the ASA [13, 14]. Let \(F_{rsi}\) be the resampling frequency for the \(i^{th}\) selected window \(W_i\), then its choice depends on \(F_{ref}\) and \(F_s\). Here, \(F_{ref}\) is the chosen reference sampling frequency in the system, such as it remains greater than and closest to \(F_{Nyq}=2f_{\text{max}}\). Here, \(f_{\text{max}}\) is the \(x(t)\) bandwidth. \(F_s=N/L\) is the sampling frequency for \(W_i\). Here, \(L\) and \(N\) respectively present the length in time and the number of non-uniform samples exist in \(W_i\). After resampling, there exist \(N\) uniformly placed samples for \(W_i\).

For the studied case, \(f_{\text{max}}=3.5 \text{ kHz}\). It is because of the employed \(F_{Cmax}=3.5 \text{ kHz}\). Therefore, \(F_{ref}=16 \text{ kHz}\) is chosen.

The uniformly resampled data is de-noised by applying the adaptive rate filtering approach [14]. In this case, a bank of Finite Impulse Response (FIR) filters is offline designed for a range of sampling frequencies. Later on, during online processing an appropriate filter is chosen, from this bank, during online processing. For the studied case, FIR filters cutoff frequencies are chosen as \([F_{CDmin}=300; \ F_{CDmax}=3k]\)Hz. This choice of \(F_{CDmin}\), removes any DC offset from the incoming speech segment. It improves the classification accuracy [6-10, 26]. Furthermore, it filters out the first harmonics of the pitch information from the incoming speech segment and therefore makes the classification process independent from the speaker’s gender.

The filters bank is offline designed by employing the Parks-
McClellan algorithm for a range of sampling frequencies, $F_R$, between 6kHz to 16kHz and with a uniform frequency step of 1kHz [11]. A summary of the employed filters bank is shown on Table 1. Each filter is designed for the cut-off frequencies of $\{F_{CD_{min}}=300\; Hz; \; F_{CD_{max}}=3k\}$. Therefore, it is found [1]

Yes

Yes

Yes

Frsi < Fr

No

Fsi < Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi = Fr

Frsi =Fr
between features of word under test and the reference templates [26].

III. RESULTS AND DISCUSSION

In order to demonstrate the proposed system functionality, a case study is conducted. In this context an isolated speech words database is used. The employed database is composed of 5 auditory commands: Forward, Reverse, Left, Right and Center. These are pronounced by 10 different speakers including 5 males and 5 females. The recording is carried out in a noise controlled environment. Each pronounced word is recorded for a duration of 3.0 Seconds. Each speaker 30 times pronounced every command.

Each digitized word is processed and analyzed in order to extract its features. Later on, these features are employed for the classification purpose. The process is clear from Figure 1.

An example of the auditory command ‘Forward’ is considered for the illustration purpose. The digitized word ‘Forward’, acquired in the classical case is shown on Figure 3-a. In classical case, a 16-Bit resolution A/D converter is used. It acquires the signal at a rate of 16k samples per second (SPS). The signal acquired by using the EDADC and selected with the ASA is shown on Figure 3-b. The uniformly resampled and de-noised window is shown on Figure 3-c.

Figure 3: The signal obtained at the output of classical ADC (3-a), the signal obtained at the output of EDADC and ASA (3-b), the uniformly resampled and de-noised window (3-c).

In this case, the superior bound on the selected windows lengths is chosen as $L_{ref}$ = 1 seconds[13, 14]. Parameters of selected windows, obtained respectively for an utterance of each intended auditory command, are summarized in Table 2.

Table 2: SUMMARY OF THE SELECTED WINDOWS PARAMETERS

<table>
<thead>
<tr>
<th>Auditory Command</th>
<th>$L'$ (Sec.)</th>
<th>$N'$ (Samp.)</th>
<th>$F_{sr}'$ (Hz)</th>
<th>$F_{sr}$ (Hz)</th>
<th>$N_{s}'$ (Samp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Forward</td>
<td>0.5772</td>
<td>7742</td>
<td>13413</td>
<td>13000</td>
<td>7504</td>
</tr>
<tr>
<td>Backward</td>
<td>0.5808</td>
<td>4454</td>
<td>7668</td>
<td>8000</td>
<td>4646</td>
</tr>
<tr>
<td>Left</td>
<td>0.4783</td>
<td>5398</td>
<td>11285</td>
<td>11000</td>
<td>5261</td>
</tr>
<tr>
<td>Right</td>
<td>0.4605</td>
<td>6176</td>
<td>13411</td>
<td>13000</td>
<td>5987</td>
</tr>
<tr>
<td>Center</td>
<td>0.5003</td>
<td>6749</td>
<td>13490</td>
<td>13000</td>
<td>6504</td>
</tr>
</tbody>
</table>

Table 2 displays the appealing features of the devised solution. The sampling frequency is adjusted as a function of the $x(t)$ local variations. It is achieved due to the smart features of the EDADC and the ASA. The process is clear from values of $F_{sr}'$. The adjustment of the resampling frequency for each selected window improves to the computational gain of the proposed technique. It avoids the unnecessary interpolations during the resampling process as clear from values of $F_{sr}$. $P$ demonstrates that how the unnecessary operations are reduced during the online filtering process for $W$. $N_{s}'$ exhibits that how the adjustment of $F_{sr}$ avoids the processing of unnecessary samples during the online filtering and spectral computation processes. $L'$ shows the dynamic feature of the ASA which is to correlate the window function length with the $x(t)$ local variations. Adjustment of $L'$ renders an adaptive time-frequency resolution of the ARSTFT. The phenomenon is clear from the values of $\Delta f'$ and $\Delta f$ in Table 2.

In the classical case, $F_{sr}$ is chosen equal to $F_{ref}$. The $x(t)$ is constantly sampled at 16kHz regardless of its local variations. It results in an acquisition of unnecessary samples. Besides, the windowing process is not able to select only the active parts of the sampled signal. Therefore, it causes the system to process...
unnecessary samples and as a consequence results in an increased computational activity than the devised solution. In the classical case, \( L_{\text{ref}}=1\) second remains static and is not able to adapt with the signal local variations. In this studied example the fixed \( N_{\text{ref}}=16000 \) is obtained for each window. It results in a fixed \( \Delta t=1\) second and \( \Delta f=1\) Hz for all windows.

An overall reduction in the acquired number of samples, achieved by the proposed solution compared to the counter classical one, is 8.2 times. It promises a noticeable decrease in the computational complexity and the power consumption of the proposed solution compared to the classical one.

In the studied case, 1500 sound waves are recorded for 5 considered commands. These 1500 utterances are tactfully employed in order to prepare the training and testing data sets. Later on, the 10-fold cross-validation process is employed while measuring the classification precession.

The obtained subject dependent percentage recognition accuracies are summarized in Table 3.

**Table 3: Subject dependent accuracy for the 5 class isolated speech recognition**

<table>
<thead>
<tr>
<th>Auditory Command</th>
<th>Classification Accuracy (%)</th>
<th>Average Classification Accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Forward</td>
<td>96.6</td>
<td>96.8</td>
</tr>
<tr>
<td>Reverse</td>
<td>97.4</td>
<td></td>
</tr>
<tr>
<td>Left</td>
<td>96.1</td>
<td></td>
</tr>
<tr>
<td>Right</td>
<td>97.1</td>
<td></td>
</tr>
<tr>
<td>Center</td>
<td>96.8</td>
<td></td>
</tr>
</tbody>
</table>

Table 3, shows that the obtained subject dependent auditory commands classification accuracies are respectively 96.6% for ‘Forward’, 97.4% for ‘Reverse’, 96.1% for ‘Left’, 97.1% for Right and 96.8% for ‘Center’. It results into an average subject dependent isolated word recognition accuracy of 96.8%.

The summary of obtained subject independent percentage recognition accuracies is presented in Table 4. It shows that the percentage recognition accuracy is respectively 93.3% for ‘Forward’, 92.8% for ‘Reverse’, 91.7% for ‘Left’, 94.1% for ‘Right’ and 92.9% for ‘Centre’. It results into an average subject independent isolated word recognition accuracy of 95%

**Table 4: Subject independent accuracy for the 5 class isolated speech recognition**

<table>
<thead>
<tr>
<th>Auditory Command</th>
<th>Classification Accuracy (%)</th>
<th>Average Classification Accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Forward</td>
<td>93.3</td>
<td>93</td>
</tr>
<tr>
<td>Reverse</td>
<td>92.8</td>
<td></td>
</tr>
<tr>
<td>Left</td>
<td>91.7</td>
<td></td>
</tr>
<tr>
<td>Right</td>
<td>94.1</td>
<td></td>
</tr>
<tr>
<td>Center</td>
<td>92.9</td>
<td></td>
</tr>
</tbody>
</table>

IV. CONCLUSION

A novel, event-driven signal acquisition and adaptive rate signal processing and analysis based method is devised for the Arabic isolated speech recognition. It is shown that how the employment of EDADC and ASA have avoided the unnecessary samples to process. It results into an 8.2 times reduction in the number of samples, obtained by the proposed system as compared to the traditional one. It assures that the devised system will lead towards a drastic computational complexity reduction as compared to the counter classic ones. It will lead towards the design and development of low power and efficient automatic speech recognition systems. The output of ASA is resampled uniformly and de-noised respectively by using the adaptive rate SLI and adaptive rate FIR filtering. The EDADC and ASA allows these modules to focus and process only on the important signal parts at an adaptable processing rate. It improves the system computational complexity and processing efficiency. The de-noised signal spectrum is computed by using the ARSTFT. It provides an efficient and adaptive resolution time-frequency representation of the de-noised windows. Later on, the spectral magnitude, Delta and Delta-Delta coefficients are extracted for each window and are employed for the classification purpose.

It has been shown that the proposed speech recognition module achieves an average subject dependent accuracy of 96.8%. The average subject independent isolated word recognition accuracy is 93%. It expresses that the average subject dependent recognition accuracy is 3.8% higher than the average subject independent recognition accuracy. It concludes that the proposed system is a better candidate for mono user applications as compared to the poly user ones.

The devised system performance depends on the type of employed interpolator, de-noising stage, parameters extraction module and the classification algorithm. A study on the devised system performance while employing the high order interpolators like polynomial, spline, etc. is a future work. Moreover, study the system performance while employing other robust classification techniques like Rotation forest, State Vector Machine, Artificial Neural Networks, Random forest, etc. is another future task. Employment of ensemble classifiers can further improve the system classification accuracy but at a cost of an increased processing load. Exploring this approach is another research axis.

V. FUNDING

This project is funded by the Effat University with the decision number UC#7/28Feb 2018/10.2-44f.

VI. ACKNOWLEDGEMENT

Author is thankful to anonymous reviewers for their valuable feedback.
REFERENCES

[21] Qaisar, S. M., 2011. Electronic management system for rechargeable battery has measuring circuit measuring parameter determining variation of parameter transmitting data to electronic processing unit if variation is higher than predetermined threshold.