

Multi-Dialect Arabic Speech Recognition

Abbas Raza Ali

Faculty of Science and Technology

Bournemouth University

Poole BH12 5BB, United Kingdom

abbas.raza.ali@gmail.com

Abstract—This paper presents the design and development of multi-dialect automatic speech recognition for Arabic. Deep neural networks are becoming an effective tool to solve sequential data problems, particularly, adopting an end-to-end training of the system. Arabic speech recognition is a complex task because of the existence of multiple dialects, non-availability of large corpora, and missing vocalization. Thus, the first contribution of this work is the development of a large multi-dialectal corpus with either full or at least partially vocalized transcription. Additionally, the open-source corpus has been gathered from multiple sources that bring non-standard Arabic alphabets in transcription which are normalized by defining a common character-set. The second contribution is the development of a framework to train an acoustic model achieving state-of-the-art performance. The network architecture comprises of a combination of convolutional and recurrent layers. The spectrogram features of the audio data are extracted in the frequency vs time domain and fed in the network. The output frames, produced by the recurrent model, are further trained to align the audio features with its corresponding transcription sequences. The sequence alignment is performed using a beam search decoder with a tetra-gram language model. The proposed system achieved a 14% error rate which outperforms previous systems.

Index Terms—Automatic Speech Recognition, Corpus Development, Deep Neural Networks, Multi-Dialect Arabic, Recurrent Neural Networks

I. INTRODUCTION

Deep Neural Networks (DNN) has become a de-facto standard in the Computer Vision task. In recent years, they are gaining traction in speech and language processing tasks with a significant improvement in acoustic modelling [18]. Despite that human speech varies with respect to pitch, amplitude, duration, and the combination of phonemes and morphemes, yet how they tie together to create words, is the primary focus. The acoustic modeling is a contextual problem which generates sequential information. The Recurrent Neural Networks (RNN), a class of DNN, is considered as the most powerful tool for this kind of problem [12]. Instead of combining RNNs with Hidden Markov Model (HMM) and Gaussian Mixture Model (GMM), [14] introduces a concept of end-to-end automatic speech recognition (ASR) using solely RNN. Thus, end-to-end training replaces the traditional acoustic modeling approach to predict different states efficiently. Also, it makes the overall task simplified and combined with a language model (LM) gives a state-of-the-art performance.

Arabic is the 5th most widely spoken language even though it is an under resource language. Although, there are few open-source Arabic speech corpora available which can be used

to build a baseline speech recognition model. However, for the phonetically rich language like Arabic, these resources are not sufficient to build a reasonably accurate acoustic model. The key challenge in the case of the Arabic language is to obtain diacritized transcription of the audios which is not very common in open-source speech corpora. The native Arabic speakers understand the meaning of the text from the context rather than expecting diacritized text. Another challenge associated with the Arabic language is the existence of several dialects with a multitude of pronunciations [6].

This paper addresses a number of the challenges faced during the design and development of multi-dialect Arabic ASR. Typically, the DNNs require a large amount of relevant data to train a model. Hence, a large multi-dialectal corpus has been developed which is one of the contributions of this work. Additionally, the open-source corpus has been gathered from multiple sources. The corpus brings variants of character-set and their transcripts were either fully-, partially- or un-diacritized. Thus, the transcription of all the sources was normalized to a common character-set, defined in the next section, whereas automatic diacritization followed by a manual review has been performed. The automatic diacritization systems were quite effective for Arabic, Urdu [9], and a few other languages.

The rest of the paper is organized as follows. The development of speech corpus, from various sources, is discussed in Section II. Section III is devoted to elaborate on the overall approach of this work. Section IV reports the experimental results followed by their analysis. Finally, the paper is concluded in Section V.

II. SPEECH CORPUS

In order to train an acoustic model, a large amount of speech corpus has been gathered. The corpus, covering multiple dialects, was recorded from 450-500 Arabic speakers possessing different accents. The distribution of the gender was around 35-65 for female and male speakers respectively. Additionally, the corpus was composed of various speaking styles including a) read speech, b) answer speech (or interviews), c) descriptive speech (or documentary), and d) spontaneous speech (or conversation). However, irrespective of the speech style, the audio was segmented on average between 3 and 50 seconds which was suitable to train a good speech recognizer. The quality of the transcription varied significantly in different sources. The variations in the transcript were in the form of un-diacritized or semi-diacritized transcript. Also, the conversational speech

TABLE I
LIST OF SPEECH CORPORA USED TO BUILD ACOUSTIC MODEL

Dataset	Dialect	Duration	Prompt
Aldiri [3]	MSA	2 hours	Read speech
KACST [5]	MSA	2 hours	Read speech
Isolated Words [16]	MSA	10 hours	Read speech
Arabic News Channel [19]	MSA, Egyptian, Gulf, Levantine, North African	57 hours	Conversations, interviews, and news documentary
KSU [13]	MSA	120 hours	Read speech
MGB-2 [20]	MSA, Egyptian, Gulf, Levantine, North Africa	1,200 hours	Conversations, interviews, and news documentary
This work	MSA, Egyptian, Gulf	70 hours	Read speech

contains overlapping of phrases and the usage of non-standard Arabic alphabets [25].

The speech corpus was gathered from multiple sources which are listed in Table I. The Aldiri corpus consists of 4,740 utterances spoken by 6 speakers with the distribution of 3 males and 3 females. The KACST dataset consists of 4,573 utterances in Modern Standard Arabic (MSA) dialect. The ‘Isolated words’ data-source contains 9,992 utterances of 20 words spoken by 50 male Arabic speakers. The ‘Arabic News Channel’ corpus was composed of conversational speech style with around 27,000 utterances. There were a few challenges associated with this corpus because of its speaking style, such as overlapping speech and usage of non-standard alphabets. The non-standard alphabets that appear in the transcript were normalized to a common character-set and the overlapping phrases were omitted from the final corpus. The KSU database was recorded from more than 200 speakers belonging to different ethnic groups. The transcription of the corpus was composed of phonetically rich and balanced words, digits, sentences, and paragraphs. The audio was recorded from low and high-quality microphones and telephony mediums in different environments including office, sound-proof room, and cafeteria. The Multi-Genre Broadcast (MGB)-2 corpus was recorded from 19 different programs of an Arabic TV channel where the content covers three speaking styles including conversations, interviews, and documentaries. The MSA dialect was most dominating in the programs along with Egyptian, Gulf, Levantine, and North African.

A. Character-set

The raw Arabic corpora were gathered from multiple sources using different character-sets in their transcripts. A common character-set has been defined to normalize all the sources. Initially, for some experiments, the regular Arabic short vowels called: fatha, damma, and kasra including a diacritical-mark shadda were part of the character-set. The remaining Arabic diacritical-marks were rare in most of the data-sources, hence, they were excluded comprising of fathatan, kasratan, dammatan, and sukun. Three long vowels were part of the character-set including alef, waw, and ya. The rest of the alphabets were standard consonants of Arabic. In the final experiments, the short vowels and shadda were also excluded which were causing dataset imbalance. The Arabic character-set used in training the system is shown in Fig. 1.

B. Data Cleansing

The data cleansing and normalization become critical when developing a corpus from different sources that possess different rules. The audio files were recorded using the low and high quality of microphones with different background noise. A high-pass filter was applied to the audios with a cutoff frequency of 150 Hz to cancel the DC offset. The audio files were converted to a single-channel 16-bit 16,000 Hz sample rate encoded using Linear Pulse Code Modulation (LPCM). The long audio files were split between 2-30 seconds segments. Similarly, the transcript was normalized accordingly by segmenting long sentences into multiple phrases to synchronize them with their corresponding audios. Also, different sources bring a different combination of diacritical-marks which were normalized by, initially, limiting the text to fatha, damma, kasra, and shadda. In the later experiments, all the diacritical-marks were removed except ‘hamza above’. The ‘hamza above’ mark was used with alef, waw and ya. The transcripts with non-standard Arabic characters have been completely removed from the corpus.

C. Development of New Corpus

In addition to the collection of open-source Arabic speech corpora, an effort has been made to design and develop a phonetically rich and balanced corpus. The development of the corpus consists of three key stages: 1) the content preparation, 2) speech recording and validation and 3) verification. The high-level steps of collecting and developing Arabic speech corpus are shown in Fig. 2. The idea was to first utilize all the available open-source resources to build a basic corpus followed by the development of a new corpus. Following are the three key areas of new corpus development:

1) *Transcription*: A phonetically rich and balanced transcription was prepared which covers most of the possibilities of Arabic phonemes and their different positions in a word. Thus, the transcript of Saudi Accented Arabic Voice Bank (SAAVB) [7] was chosen. The transcript includes 1,033 unique phrases including numbers, words, sentences, and paragraphs that were designed to differentiate the number of dialects.

2) *Corpus Recording and Validation*: The corpus recording process started with the recruitment of a large number of Arabic speakers possessing different dialects. The recruited speakers were required to read 190 phrases that were assigned to them via an automatic system. A recorded phrase was first evaluated using speech processing tools and was stored in the

Grapheme	Arpabet	Description	Grapheme	Arpabet	Description
ْ	*	hamza above	ش	SH	shīn
ء	E	hamza	ص	SS	ṣād
ا	AE E	alif	ض	DD	ḍād
آ	AE:	alif maddah	ط	TT	ṭā
إ	E	alif + hamza below	ظ	DH2	ẓā
ب	B	bā	ع	AI	'ayn
ة	T H	ta marbūta	غ	GH	ghayn
ت	T	tā	ف	F	fā
ث	TH	thā	ق	Q	qāf
ج	ZH	jīm	ك	K	kāf
ح	HH	ḥā	ل	L	lām
خ	KH	khā	م	M	mīm
د	D	dāl	ن	N	nūn
ذ	DH	dhāl	ه	H	hā
ر	R	rā	و	W UW	wāw
ز	Z	zāy	ى	AE	alif maksura
س	S	sīn	ي	Y IY	yā

Fig. 1. Arabic character-set with phonemes and description

database only if it passes the quality test. In case of failure, because of either noise, unnecessary silence, repetition, etc., the speaker was required to re-record the same phrase until it gets approved by the speech analyst. This process was repeated on all the phrases and speakers.

3) *Verification*: A detailed verification of the recording setup has been performed by the linguists before the initiation of a recording session. The verification process was not limited to the verification of recording equipment, it also checks the quality of the recorded speech. Further, prior to the recording sessions, the speakers were instructed on the usage of recording hardware and software.

III. METHODOLOGY

Arabic is a Semitic language having unique phonemes and phonetic features, multiple dialects, and a complicated morphological word structure [2]. In order to train a speech recognizer using this kind of complex data, the DNN capacity has been increased via depth. Two models were trained independently namely; acoustic model and language model, see Equation 1. The speech corpus, comprising of transcribed audios, was used to train the acoustic model. The acoustic model was used to predict phonemes, given a segment of audio features. On the other hand, an n-gram based language model predicts the next word in a sequence. It trains on text corpus comprising of transcripts of conversations.

$$\begin{aligned}
 \text{ASR} &= \arg \max_{w \in \text{vocab}} (\text{language model}) \times (\text{acoustic model}) \\
 &= \arg \max_{w \in \text{vocab}} (\text{words}) \times (\text{features} | \text{words}) \\
 &= \arg \max_{w \in \text{vocab}} (w) \times (x | w)
 \end{aligned} \tag{1}$$

The network consists of a combination of convolutional and recurrent layers followed by a fully-connected layer. The convolutional layers were limited to 2-3 with a stack of 4-5 bi-directional recurrent layers. Various sizes of network width were tried including 768 and 1024. Also, two types of RNNs have experimented which were Long short-term memory (LSTM) and Gated Recurrent Units (GRUs). A clipped Rectified Linear Unit (ReLU) function has been applied as formulated in Equation 2. The End-to-End speech recognition architecture is shown in Fig. 3.

$$\sigma(x) = \min(\max(x, 0), 20) \tag{2}$$

Three convolutional layers have been used with 32, 32 and 96 filter sizes having dimensions 41x11, 21x11 and 21x11 respectively. The strides of 2x2, 2x1 and 2x1 have been applied to down-sample the input. The regularisation is key to better performance by avoiding over-fitting when using RNN. Hence, batch normalization (BatchNorm) was used with convolutional layers whereas a dropout of 20% was used for recurrent layers. The BatchNorm also enhances the training speed of this kind of very deep networks with marginally improved accuracy. To avoid over-fitting, early stopping was also used while training the network. The sequence alignment was performed through

the Connectionist Temporal Classification (CTC) which was found to be very successful for speech and language processing tasks [24]. The training examples of varying lengths were sorted in increasing order of length since short examples was comparatively less costly.

The inference process applies normalization and extracts features of the new audio. The trained acoustic model produces phoneme probabilities over time. These probabilities were fed to the decoder, along with a language model, to compute the most likely sequence of words for the given audio input.

A. Feature Extraction

The Mel-frequency Cepstral Coefficients features (MFCC) have demonstrated promising performance for speech recognition task [27]. However, for this work, the spectrograms was extracted from the audio files which were fed as input to the convolutional layers of the network. The spectrogram is a two-dimensional representation of audio data consisting of frequency spectrum over time which can be processed as a two-dimensional image. The filter banks were computed over windows of 0.02 seconds with strides of 0.01 seconds [14].

B. Acoustic Modeling

The acoustic model was trained using more than 1,400 hours of speech data. A deep and wide network has been designed to train the acoustic model along with sequence alignment. The optimal set of hyper-parameters was found by performing a number of experiments with different values of filters and strides of convolutional layers, and the type and width of recurrent layers. The weights were initialized with Xavier-initialization [8]. A dropout of 20% has been used for recurrent layers whereas BatchNorm was used for convolutional layers. The network was trained with Adam optimizer [17] with a variable learning rate which decays by a factor of 10 on every alternative epoch with an initial value of 0.001.

The recurrent networks was trained as frame-level classifiers for speech recognition tasks [15]. These frames further require training to align the input audio with corresponding output transcription sequences. The CTC allows a sequence training without requiring any prior alignment between the input and target sequences. In this work, the CTC beam search along with a trained word-level LM has been used. The LM guides the CTC beam search decoder on domain-specific and misspelled terms. The CTC was used with a beamwidth of 512.

C. Language Model

The word-level LM predicts the upcoming word given the previous sequence of words [28]. A word-level tetra-gram LM was trained on a large Arabic corpus. The short vowels and shadda diacritical-mark was also removed from the language model corpus to synchronize with the character-set of the acoustic model. The LM has been trained over millions of unique utterances using KenLM with Kneser-Ney smoothed tetra-gram model [11]. The beam search was used to find the optimal word sequence. The trained model was quite effective for spelling and implicitly disambiguate homophones.

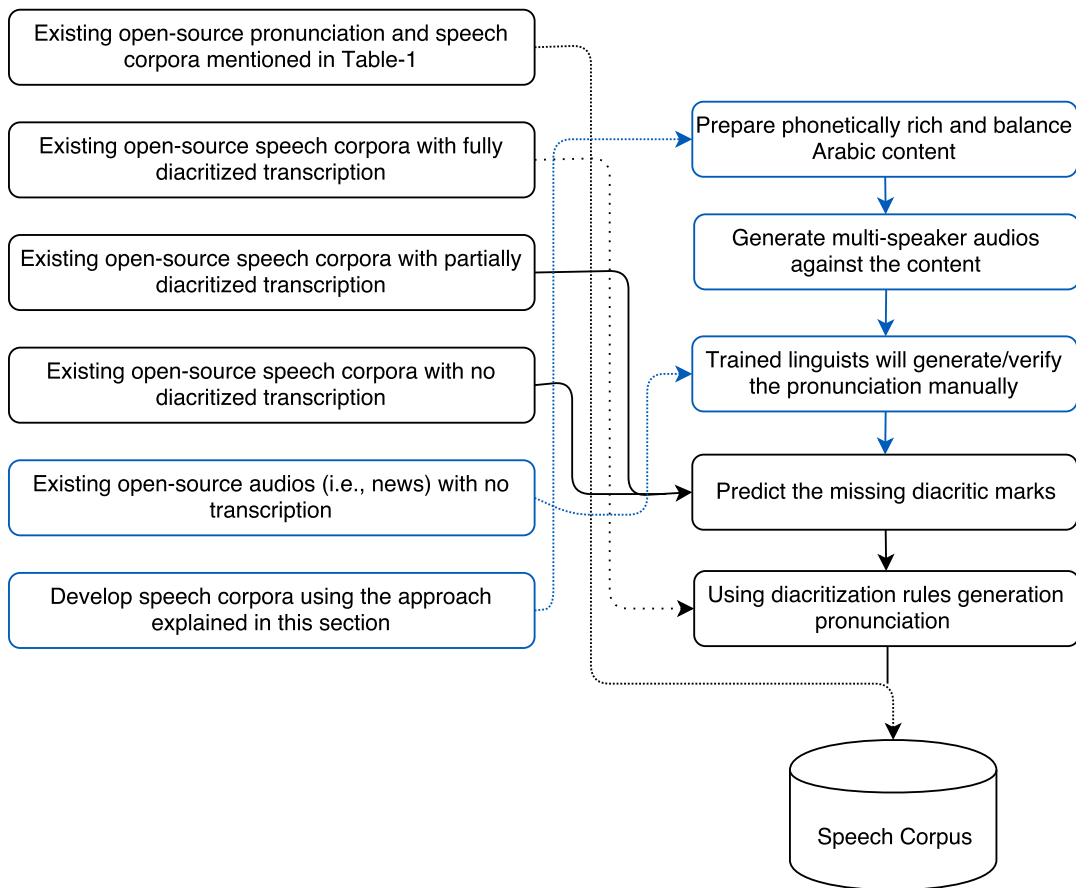


Fig. 2. Steps to build Arabic Speech Corpus

The corpus that was used to build LM was also collected from multiple sources listed in Table II. The MGB-2 has diversity in broadcast news comprising of 19 distinct programs of an Arabic TV channel. The text was crawled from the website of the channel. The Tashkeela corpus contains 75.6 million fully vocalized terms that was crawled from 97 classical and modern Arabic books. The consolidated text was normalized using the same rules which was applied to the transcription of the audio data.

TABLE II
LIST OF CORPORA USED TO TRAIN LANGUAGE MODEL

Dataset	Vocabulary	Domain
MGB-2 LM [20]	130 million	News
Tashkeela [23]	75.6 million	Islamic classical books

IV. ANALYSIS AND RESULTS

The proposed architecture was tested with different combinations of hyper-parameters using a trial-and-error approach. The corpus was divided into training, development and validation sets with a ratio of 70%, 20%, and 10% respectively. The experiments were conducted in a distributed training framework for 32-35 epochs using 10 Nvidia 1080ti GPUs. The total batch size of 80 (8 batches per GPUS) is used with

an initial learning rate of 0.001 and a momentum of 0.99 to train multiple models. There were 8 different models trained varying along 2-3 convolutional and 4-5 recurrent layers, and RNN width (768 or 1024) and type (LSTM or GRU). The models which performed well during the training phase on the development set were chosen for further evaluation. Their performance was measured using word error rate (WER) which was formulated in Equation 3.

$$\text{WER} = \frac{\text{substitutions (S)} + \text{deletions (D)} + \text{insertions (I)}}{\text{total words in the reference (N)}} \quad (3)$$

TABLE III
COMPARISON OF DIFFERENT ARRANGEMENTS OF CONVOLUTIONAL AND RECURRENT LAYERS ALONG WITH NETWORK WIDTH AND RNN TYPE

Architecture	Neurons	RNN Type	WER (%)
2-CNN, 4-RNN	768	GRU	14.07
2-CNN, 4-RNN	768	LSTM	14.67
2-CNN, 4-RNN	1024	GRU	15.32
2-CNN, 4-RNN	1024	LSTM	15.54
3-CNN, 5-RNN	768	GRU	14.88
3-CNN, 5-RNN	768	LSTM	15.30
3-CNN, 5-RNN	1024	GRU	14.23
3-CNN, 5-RNN	1024	LSTM	14.61

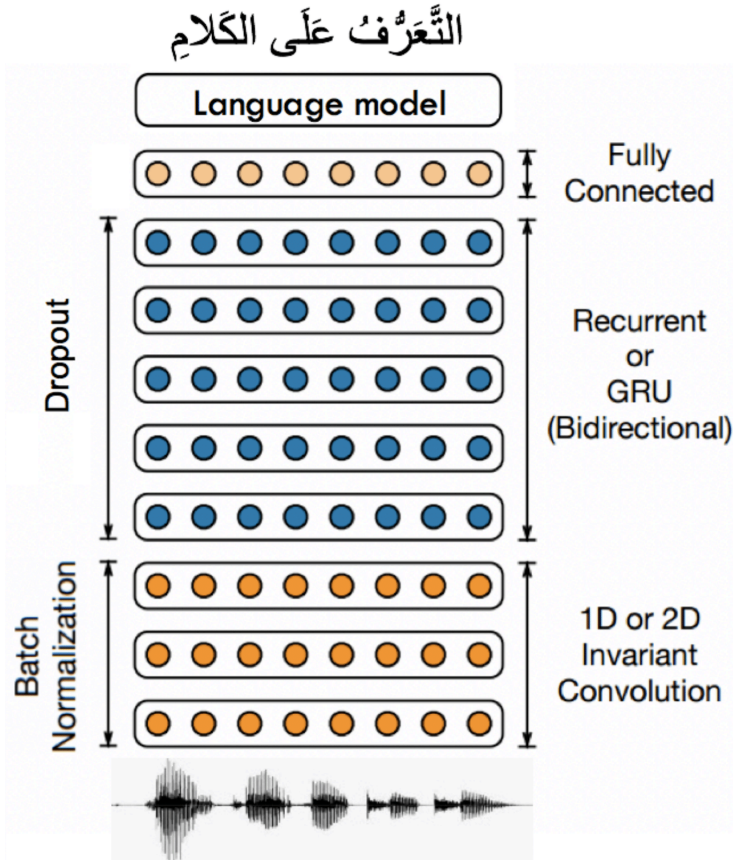


Fig. 3. The architecture of end-to-end Arabic speech recognition. In the experiments 2-3 convolution and 4-5 recurrent layers have been tested.

The outcome of the acoustic model was decoded using the CTC beam search with LM. The final WER of the different models was reported in Table III. The model with 2 convolutional layers followed by 4 bi-directional recurrent layers, composed of GRU cell and 768 neurons, significantly outperformed the rest of the variants. The statistical significance was computed using a paired t-test of the top-performing architecture’s WER with the rest of the WERs. Further, the high-performing model was using a fully-connected layer followed by a CTC beam search decoder with LM.

Table IV provides a comparison of the proposed system with the prior research. The performance of several DNN variants, such as Feed-forward, CNN, LSTM, Time Delayed Neural Networks (TDNN) and GMM have been evaluated.

TABLE IV
COMPARISON OF THE PROPOSED APPROACH WITH EXISTING SYSTEMS

Research	Acoustic/Lang. Model	Speech Corpus	WER %
[4]	TDNN/3-gram	MGB-2	20.5
[22]	RNN/4-gram	MGB-2	18.3
[25]	RNN/3-gram	MGB2	14.70
[26]	RNN/4-gram	Nemlar ¹ & NetDC ²	14.42
This work	CNN, RNN/4-gram	MGB-2 & custom	14.07

¹http://catalog.elra.info/product_info.php?products_id=874

²http://catalog.elra.info/product_info.php?products_id=13

V. CONCLUSIONS

An end-to-end automatic speech recognition system for multi-dialectal Arabic language has been presented in this paper. The key contributions include the development of large multi-dialectal corpus and an end-to-end speech recognition model with accuracy close to the state-of-the-art for the Arabic language. This accuracy is the result of numerous trials to explore various network architectures and their hyper-parameters, accompanied by several techniques including sorting of training samples, BatchNorm, early stopping and decoding with LM.

Two key contributions are claimed in this work. The first contribution is the development of a large multi-dialectal Arabic speech corpus consisting of 1,400 hours of speech data. The corpus has been gathered from multiple sources and normalized on a common character-set. The second contribution is the development of a framework to train the Arabic speech recognition acoustic model with close to state-of-the-art accuracy. The system has achieved an overall WER of 14%.

REFERENCES

- [1] M. Boudraa, B. Boudraa, and B. Guerin, “Twenty Lists of Ten Arabic Sentences for Assessment,” in *ACUSTICA, ACTA-ACUSTICA*, vol. 86, no. 5, 1998.

- [2] S. Selouani, and J. Caelen, "Arabic phonetic features recognition using modular connectionist architectures," in Proceedings of the IEEE Interactive Voice Technology for Communication, IVTTA, 1998.
- [3] B. Al-Diri, A. Sharieh, and T. Hudaib, "A time delay neural network architecture for efficient modeling of long temporal contexts," in Interspeech, 2015.
- [4] V. Peddinti, D. Povey, and S. Khudanpur, "A time delay neural network architecture for efficient modeling of long temporal contexts," in Interspeech, 2015.
- [5] M. A. Mansour, "KACST arabic phonetics database," Riyadh, Kingdom of Saudi Arabia, 2004.
- [6] A. Sankaranarayanan, S. Bangalore, and S. S. Narayanan, "Automatic diacritization of Arabic transcripts for automatic speech recognition," in Proceedings of the International Conference on Natural Language Processing (ICON), 2005.
- [7] M. Alghamdi, F. Alhargan, M. Alkanhal, A. Alkhairy, M. Eldesouki, and A. Alenazi, "Saudi Accented Arabic Voice Bank," in Journal of King Saud University, Computer and Information Sciences, vol. 20, pages 45–64, 2008.
- [8] X. Glorot, and Y. Bengio, "Understanding the difficulty of training deep feed-forward neural networks," in Proceedings of the 20th International Conference on Artificial Intelligence and Statistics (PMLR), 2010.
- [9] A. R. Ali, and S. Hussain, "Automatic Diacritization for Urdu," in Proceedings of the Conference on Language and Technology (CLT), 2010.
- [10] M. A. M. Abushariah, R. N. Ainon, R. Zainuddin, M. Elshafei, and O. O. Khalifa, "Phonetically rich and balanced text and speech corpora for Arabic language," in Journal Language Resources and Evaluation, vol. 46, no. 4, pages 601–634, 2012.
- [11] K. Heafield, I. Pouzyrevsky, J. H. Clark, and P. Koehn, "Scalable modified Kneser-Ney language model estimation," in Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics, 2013.
- [12] A. Graves, A. Mohamed, and G. Hinton, "Speech Recognition with Deep Recurrent Neural Networks," in Computing Research Repository (CoRR), vol. abs/1303.5778, 2013.
- [13] M. Alsulaiman, Z. Ali, G. Muhammed, M. Bencherif, and A. Mahmood, "KSU Speech Database: Text Selection, Recording and Verification," in European Modelling Symposium, 2013.
- [14] A. Hannun, C. Case, J. Casper, and B. Catanzaro, "Deep Speech: Scaling up end-to-end speech recognition," in Computing Research Repository (CoRR), vol. abs/1412.5567, 2014.
- [15] A. Graves, and N. Jaitly, "Towards end-to-end speech recognition with recurrent neural networks," in Proceedings of the 31st International Conference on International Conference on Machine Learning (ICML), 2014.
- [16] A. Alalshkumbarak, and L. S. Smith, "On Improving the Classification Capability of Reservoir Computing for Arabic Speech Recognition," in 24th International Conference on Artificial Neural Networks, 2014.
- [17] D. P. Kingma, and J. B. Adam, "A method for stochastic optimization," International Conference on Learning Representations (ICLR), 2015.
- [18] D. Amodei, S. Ananthanarayanan, R. Anubhai, and J. Bai, "Deep Speech 2: End-to-End Speech Recognition in English and Mandarin," in Proceedings of Machine Learning Research (PMLR), vol. 48, 2015.
- [19] S. Wray, and A. Ali, "Crowdsource a little to label a lot: Labeling a Speech Corpus of Dialectal Arabic," in Interspeech, 2015.
- [20] S. Khurana, and A. Ali, "QCRI advanced transcription system (QATS) for the Arabic Multi-Dialect Broadcast media recognition: MGB-2 challenge," in IEEE Spoken Language Technology Workshop (SLT), 2016.
- [21] A. Ahmed, Y. Hifny, K. Shaalan, and S. L. Toral, "Lexicon Free Arabic Speech Recognition Recipe," in International Conference on Advanced Intelligent Systems and Informatics, vol. 12, pages 147–159, 2016.
- [22] T. AlHanai, W. N. Hsu, and J. Glass, "Development of the MIT ASR system for the 2016 Arabic Multi-genre Broadcast Challenge," in IEEE Spoken Language Technology Workshop (SLT), 2016.
- [23] T. Zerrouki, and A. Balla, "Tashkeela: Novel corpus of Arabic vocalized texts, data for auto-diacritization systems," in Data in Brief, vol. 11, 2017.
- [24] M. Najafian, W. N. Hsu, A. Ali, and J. Glass, "Automatic speech recognition of Arabic multi-genre broadcast media," in IEEE Automatic Speech Recognition and Understanding Workshop (ASRU), 2017.
- [25] A. Ali, P. Bell, J. Glass, Y. Messaoui, H. Mubarak, S. Renals, and Y. Zhang, "The MGB-2 Challenge: Arabic Multi-Dialect Broadcast Media Recognition," in Computing Research Repository (CoRR), vol. abs/1609.05625, 2017.
- [26] M. A. Menacer, O. Mella, and D. Fohr, "An enhanced automatic speech recognition system for Arabic," in the 3rd Arabic Natural Language Processing Workshop (EACL), 2017.
- [27] L. Bouchakour, and M. Debyeche, "Improving Continuous Arabic Speech Recognition over Mobile Networks DSR and NSR Using MFCCs Features Transformed," in International Journal of Circuits, Systems and Signal Processing, vol. 12, 2018.
- [28] A. R. Ali, "Cognitive Computing to Optimize IT Services," in IEEE 17th International Conference on Cognitive Informatics and Cognitive Computing (ICCI*CC), Berkeley, CA, 2018, vol. 1, pages 54–60.