

Preface

This book describes novel approaches to improve automatic speech recognition for dialectal Arabic. Since the existing dialectal Arabic speech resources, that are available for the task of training speech recognition systems, are very sparse and are lacking quality, we describe how existing Modern Standard Arabic (MSA) speech resources can be applied to dialectal Arabic speech recognition. Our assumption is that MSA is always a second language for all Arabic speakers, and in most cases we can identify the original dialect of a speaker even though he is speaking MSA. Hence, an acoustic model trained with a sufficient number of MSA speakers from different origins will implicitly model the acoustic features for the different Arabic dialects and in this case, it can be called dialect-independent acoustic modeling.

In this work, Egyptian Colloquial Arabic (ECA) has been chosen as a typical Arabic dialect. ECA is the first ranked Arabic dialect in terms of number of speakers. A high quality ECA speech corpus with accurate phonetic transcriptions has been collected. MSA acoustic models were trained using news broadcast speech data. In fact, MSA and the different Arabic dialects do not share the same phoneme set. Therefore, in order to cross-lingually use MSA in dialectal Arabic speech recognition, we propose phoneme sets normalization. We have normalized the phoneme sets for MSA and ECA. After phoneme sets normalization, we have applied state-of-the-art acoustic model adaptation techniques like Maximum Likelihood Linear Regression (MLLR) and Maximum A-Posteriori (MAP) to adapt existing phonemic MSA acoustic models with a small amount of ECA speech data. Speech recognition results indicated a significant increase in recognition accuracy compared to a baseline model trained with only ECA data. Best results were obtained when combining MLLR and MAP.

Since for dialectal Arabic in general, it is hard to phonetically transcribe large amounts of speech data, we have studied the use of grapheme-based acoustic models where the phonetic transcription is approximated to be the word letters rather than the exact phonemes sequence. A large number of Gaussians in the Gaussian mixture model is used to implicitly model missing vowels. Since dialectal Arabic is mainly spoken and not formally written, the graphemic form usually does not match the actual spelling as in MSA. A graphemic MSA acoustic model was therefore used

to force align and to choose the correct ECA spelling from a set of automatically generated spelling variants lexicon. Afterwards, MLLR and MAP acoustic model adaptations are applied to adapt the graphemic MSA acoustic model with a small amount of dialectal data. In the case of graphemic adaptation, we are still able to observe a significant improvement in speech recognition accuracy compared to the baseline. In order to prove a consistent improvement across different Arabic dialects, the presented approach was re-applied on Levantine Colloquial Arabic (LCA).

Finally, we have also shown how to use the Arabic Chat Alphabet (ACA) to transcribe dialectal speech data. ACA transcriptions include short vowels that are omitted in traditional Arabic orthography. So our assumption was that ACA transcriptions are closer to the actual phonetic transcription rather than graphemic transcriptions. ACA-based and phoneme-based acoustic models performed similar but significantly outperformed the grapheme-based acoustic models.

Novel Techniques for Dialectal Arabic Speech
Recognition

Elmahdy, M.; Gruhn, R.; Minker, W.

2012, XXII, 110 p., Hardcover

ISBN: 978-1-4614-1905-1