Hybrid pronunciation modeling for Arabic large vocabulary speech recognition

Mohamed Elmahdy, Mark Hasegawa-Johnson, Eiman Mustafawi

Qatar University, Doha, QATAR; University of Illinois at Urbana-Champaign, Urbana, UNITED STATES

melmahdy@ieee.org

Arabic is a morphologically rich language. This morphological complexity results in a high out-of-vocabulary rate. That is why a lookup table for pronunciation modeling is not efficient for large vocabulary tasks. In previous research, graphemic modeling was proposed by approximating pronunciation modeling to be graphemes rather than actual phonemes. In this research, we have proposed a hybrid acoustic and pronunciation modeling approach for Arabic large vocabulary speech recognition tasks. The proposed approach benefits from both phonemic and graphemic modeling techniques, where two acoustic models are fused together. The hybrid approach also benefits from both vocalized and non-vocalized Arabic resources, which is useful because the amount of non-vocalized resources is always higher than vocalized ones. Two speech recognition baseline systems were built: phonemic and graphemic. The two baseline acoustic models were combined after two independent trainings to create a hybrid model. Pronunciation modeling was also hybrid by generating graphemic variants in addition to phonemic variants. Three techniques are proposed for pronunciation modeling: Hybrid-Or, Hybrid-And, and Hybrid-Top(n). In Hybrid-Or, either graphemic or phonemic modeling is applied for any given word. In Hybrid-And, a graphemic pronunciation is always generated in addition to existing phonemic pronunciations. Hybrid-Top(n) is a mixture of Hybrid-Or and Hybrid-And by applying Hybrid-Or on the top n high frequency words. Experiments were conducted in the large vocabulary news broadcast speech domain with a vocabulary size of 250K. The proposed hybrid approach has shown a relative reduction in WER of 8.8% to 12.6% depending on pronunciation modeling settings and the supervision in the baseline systems. In large vocabulary speech domains, acoustic and pronunciation modeling is a common problem among all Arabic colloquial varieties. Thus, for future work, the proposed approach is currently being extended and evaluated with different Arabic colloquial varieties (e.g. Qatari, Egyptian, Levantine, etc.). Moreover, the proposed technique can be applied with other morphologically rich languages like Turkish, Finnish, Korean, etc. This work was funded by a grant from the Qatar National Research Fund under its National Priorities Research Program (NPRP) award number NPRP 09-410-1-069. Reported experimental work was performed at Qatar University in collaboration with University of Illinois.