
Abstract

We describe word graph generation in terms of transducer composition, and show that a simple modification to a Viterbi search avoids the usual assumptions of word-pair or phone-pair approximations when the search space is represented with a transducer detailed down to the level of HMM transitions. On a 20,000-word French language dictation task, this graph generation method increases recognition time by only 20%. The word graphs produced can be further reduced in size by applying automata minimization, and this operation can be done faster than real-time. When the resulting graphs are rescored using larger acoustic and language models, recognition rate remains near-optimal for word graph densities as low as 8 words per spoken word.