Abstract
This paper introduces the use of Graphics Processors Unit (GPU) for computing acoustic likelihoods in a speech recognition system. In addition to their high availability, GPUs provide high computing performance at low cost. We have used a NVidia GeForce 8800GTX programmed with the CUDA (Compute Unified Device Architecture) which shows the GPU as a parallel coprocessor. The acoustic likelihoods are computed as dot products, operations for which GPUs are highly efficient. The implementation in our speech recognition system shows that GPU is 5x faster than the CPU SSE-based implementation. This improvement led to a speed up of 35% on a large vocabulary task.