N phonemes and λ is the HMM parameter set. M is the dimensionality of the observations and T is the length of the sentence. $\top$ denotes the transpose. Explicit duration modeling is used in hidden semi-Markov model (HSMM) for HTS proposed by Yoshimura et al. [9]. The likelihood is decomposed into two parts

$$\begin{aligned} \hat{\boldsymbol{o}} &= \arg\max_{\boldsymbol{o}} \sum_{all\,\boldsymbol{q}} p(\boldsymbol{o} \mid \lambda, \boldsymbol{q})\, p(\boldsymbol{q} \mid \lambda, \boldsymbol{b}) \\ &\approx \arg\max_{\boldsymbol{o}} p(\boldsymbol{o} \mid \lambda, \hat{\boldsymbol{q}})\, p(\hat{\boldsymbol{q}} \mid \lambda, \boldsymbol{b}) \end{aligned} \tag{2}$$

where $\hat{\boldsymbol{q}}$ is the optimal sequence of Gaussian distributions predicted by the duration model independent of $\boldsymbol{o}$ [3]. The search for all possible $\boldsymbol{q}$ is intractable. Therefore, (2) is