Appendix

A. Choice of Probabilistic Models

ETM-HMMs are not the only ways to incorporate unlabeled data directly into the training of HMMs. Another possibility is an extension of the static model based on the mixture of experts (MoE) framework [108]. This method has been used by Miller and Uyar [26]. In contrast to the MoE framework, we call our approach the tied-mixture (TM) framework.

On one hand, the probabilistic model of an ETM-HMM, a mixture of HMMs within our TM framework, is written as follows:

$$\begin{aligned}
& P(X|\Theta) \\
= \; & \sum_{y} P(y) \sum_{S} \sum_{M} p(X, S, M|y, \Theta) \\
= \; & \sum_{Y} P(y) \sum_{S} \sum_{M} P(S|y, \Theta) P(M|S, y, \Theta) p(X|S, M, y, \Theta) \\
\approx \; & \sum_{Y} P(y) \sum_{S} \sum_{M} P(S|y, \Theta) P(M|S, \Theta) p(X|S, M, y, \Theta) \\
\approx \; & \sum_{Y} P(y) \sum_{S} \sum_{M} P(S|y, \Theta) P(M|S, \Theta) p(X|M, \Theta), \qquad (4.1)
\end{aligned}$$

where $P(y)$ is the class prior, $P(S|y, \Theta)$ is the state transition probability, $P(M|S, \Theta)$ is the state-conditional (i.e., class conditional) mixture coefficient, and $p(X|M, \Theta)$ is the distribution represented by a Gaussian. The last transformation means Gaussians are tied over classes. The second-last transformation means that the mixture coefficient depends on y, not directly but indirectly, through S, which depends on y.

On the other hand, the probabilistic model of the mixture of HMMs within the MoE framework is written as follows:

$$\begin{aligned}
& P(X|\Theta) \\
= \; & \sum_{y} \sum_{S} \sum_{M} p(X, y, S, M|\Theta) \\
= \; & \sum_{y} \sum_{S} \sum_{M} P(y|S, M, X, \Theta) P(S, M|\Theta) p(X|S, M, \Theta) \\
\approx \; & \sum_{y} \sum_{S} \sum_{M} P(y|S, M, X, \Theta) P(S|\Theta) P(M|S, \Theta) p(X|M, \Theta) \\
\approx \; & \sum_{y} \sum_{S} \sum_{M} P(y|S, M, \Theta) P(S|\Theta) P(M|S, \Theta) p(X|M, \Theta), \qquad (4.2)
\end{aligned}$$

where $P(y|S, M, \Theta)$ is the stochastic class selector, $P(S|\Theta)P(M|S, \Theta)$ is the gating function, and $p(X|M, \Theta)$ is the local committee (expert) represented by a Gaussian. The last transformation assumes the independence of the class selectors from feature vectors. The second-last transformation means that the output depends only on M and not on S.

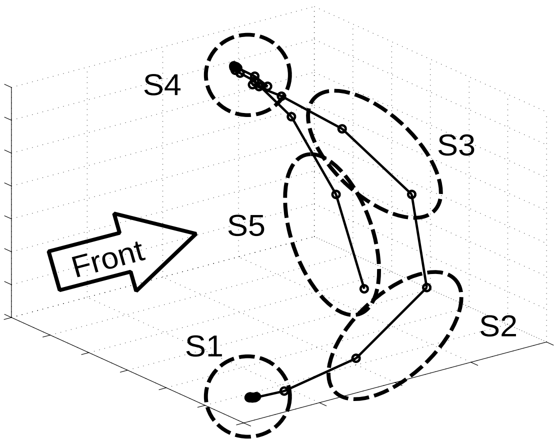

Figure 4.1. A trajectory of the right hand while signing “aisatsu” in JSL, which consists of 29 sampling points. Its five states are conceptualized by S1,$\cdots$,S5. The first state corresponds to the initial position of the hand (around the chest). In the second state, the hand is pushed forward. In the third state, it is raised. In the fourth state, it stays in front of the face, and in the fifth state, it returns to the initial position.

In the TM framework given by (4.1), the primitives of phenomena (or state) remain interpretable; $P(S|y, \Theta)$ in (4.1) corresponds to a particular stationary process of class y phenomenon. For example, a sign in sign language is viewed as a sequence of primitive hand movements (class-dependent state sequences). An example of such primitives of Japanese Sign Language (JSL) signs used in our experiment in Section 2 is shown in Fig. 4.1. This interpretability of the states is sometimes considered to be an interesting feature of HMMs in an application such as gesture understanding [109]. In contrast, $P(S|\Theta)$ in (4.2) is difficult to interpret since it is mixed over different classes. We choose to use the TM model because of this differences.