between redundant states. As we show, however, by incorporating an additional self-transition bias, it is possible to make use of Dirichlet process mixtures for the emission distributions.

An important reason for the popularity of the classical HMM is its computational tractability. In particular, marginal probabilities and samples can be obtained from the HMM via an efficient dynamic programming algorithm known as the forward–backward algorithm [Rabiner (1989)]. We show that this algorithm also plays an important role in computationally efficient inference for our generalized HDP-HMM. Using a truncated approximation to the full Bayesian nonparametric model, we develop a blocked Gibbs sampler which leverages forward–backward recursions to jointly resample the state and emission assignments for all observations.

The paper is organized as follows. In Section 2 we begin by summarizing related prior work on the speaker diarization task and analyzing the key characteristics of the data set we examine in Section 8. In Section 3 we provide some basic background on Dirichlet processes. Then, in Section 4 we overview the hierarchical Dirichlet process, and in Section 5 discuss how it applies to HMMs and can be extended to account for state persistence. An efficient Gibbs sampler is also described in this section. In Section 7 we treat the case of nonparametric emission distributions. We discuss our application to speaker diarization in Section 8. A list of notational conventions can be found in the Supplementary Material [Fox et al. (2010)].

2. The speaker diarization task. There is a vast literature on the speaker diarization task, and in this section we simply aim to provide an overview of the most common techniques. We refer the interested reader to Tranter and Reynolds (2006) for a more thorough exposition on the subject.

Classical speaker diarization techniques typically employ a two-stage procedure that first segments the audio (or features thereof) using one of a variety of changepoint algorithms. The inferred segments are then regrouped into a set of speaker labels via a clustering algorithm. For example, Reynolds and Torres-Carrasquillo (2004) propose a changepoint detection method based on the Bayesian Information Criterion (BIC). Specifically, a penalized likelihood ratio test is used to compare whether the data within a fixed window are better modeled via a single Gaussian or two Gaussians. The window gradually grows at each test until a changepoint is inferred, at which point the window is reinitialized at the inferred changepoint. An alternative changepoint detection technique, first proposed in Siegler et al. (1997), uses fixed length windows and computes the symmetric Kullback–Leibler (KL) divergence between a pair of Gaussians each fit by the data in their respective windows. A post-processing step then sets the changepoints equal to the peaks of the computed KL that exceed a predetermined threshold. In order to group the inferred