In this paper we introduce a new hierarchical model for the simultaneous detection of brain activation and estimation of the shape of the hemodynamic response in multi-subject fMRI studies. The proposed approach circumvents a major stumbling block in standard multi-subject fMRI data analysis, in that it both allows the shape of the hemodynamic response function to vary across region and subjects, while still providing a straightforward way to estimate population-level activation. An efficient estimation algorithm is presented, as is an inferential framework that not only allows for tests of activation, but also for tests for deviations from some canonical shape. The model is validated through simulations and application to a multi-subject fMRI study of thermal pain.