The accuracy of audio and AV recognition is higher than that of purely visual recognition. This is due to the higher dimensionality of audio and AV data. With data of higher dimensions, a larger number of Gaussians are multiplied, and thus the gap produced by the correct and incorrect HMM increases. Hence, the accuracy increases for higher dimension data. Note that curse of dimensionality is not the case here because the number of available states and the volume of training data samples remain small.