

Speech recognition example

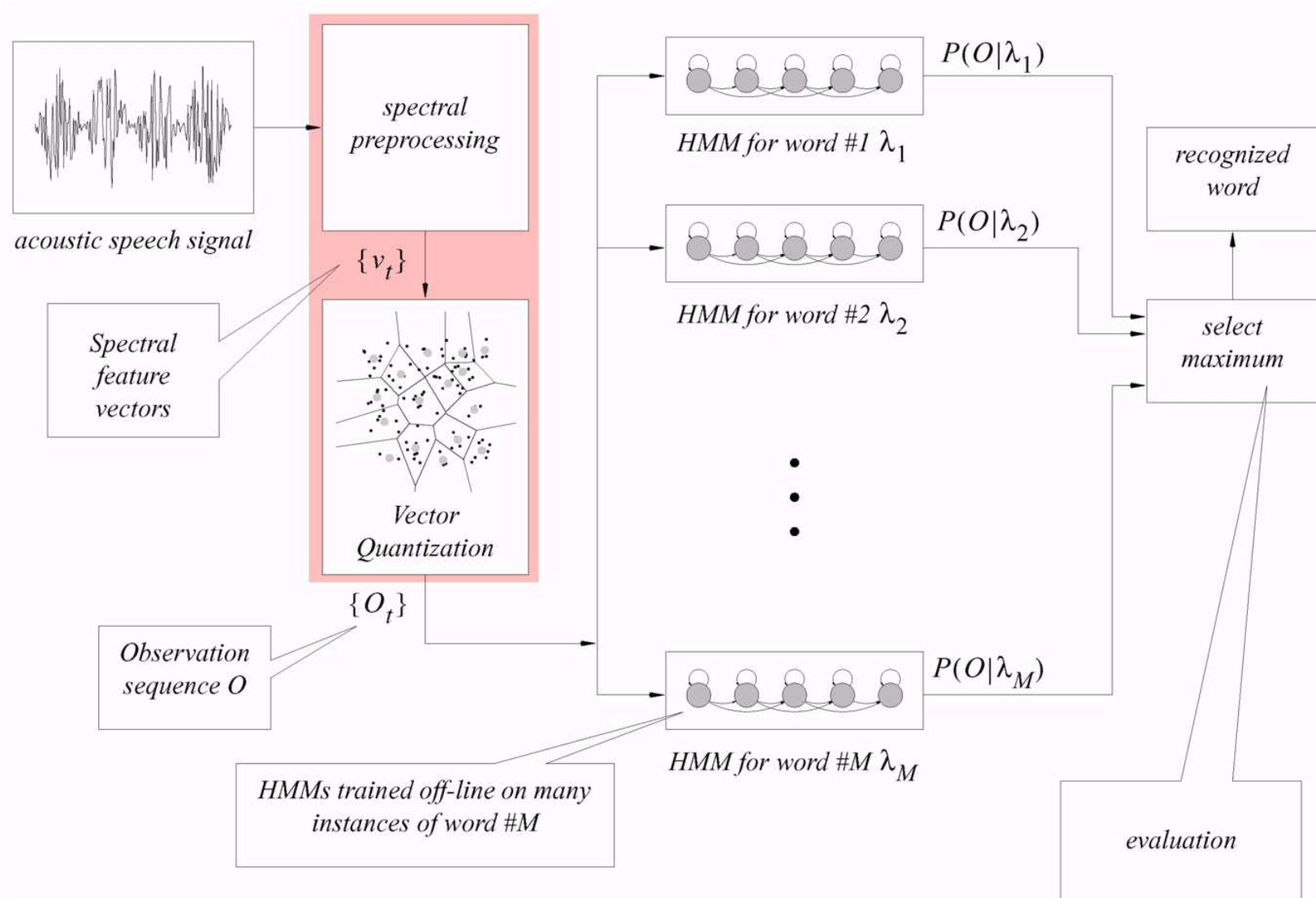
- **Isolated-word**
- **Speaker-dependent (me)**
- **Small-vocabulary:**
 - *{one, two, three, four five}*
 - *{dog, god}*
- **Simple feature extraction (FFT) based**

Sample sound files ...

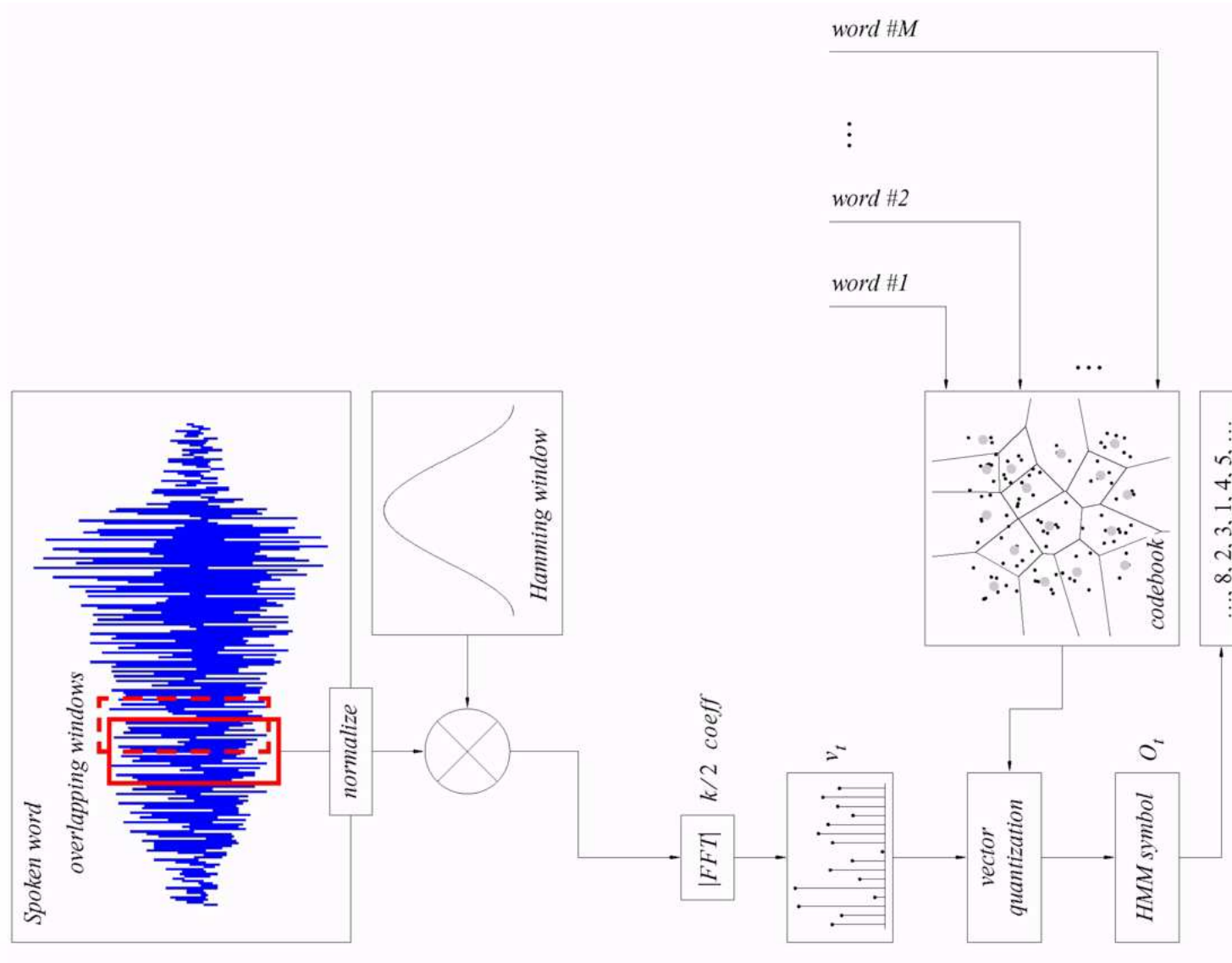
Speech recognition example: summary

- **Power-based word segmentation**
- **FFT-based feature extraction**
- **LBG VQ algorithm**
- **~80 training samples/word, 40 test samples/word**

Block diagram



Signal-to-symbol conversion

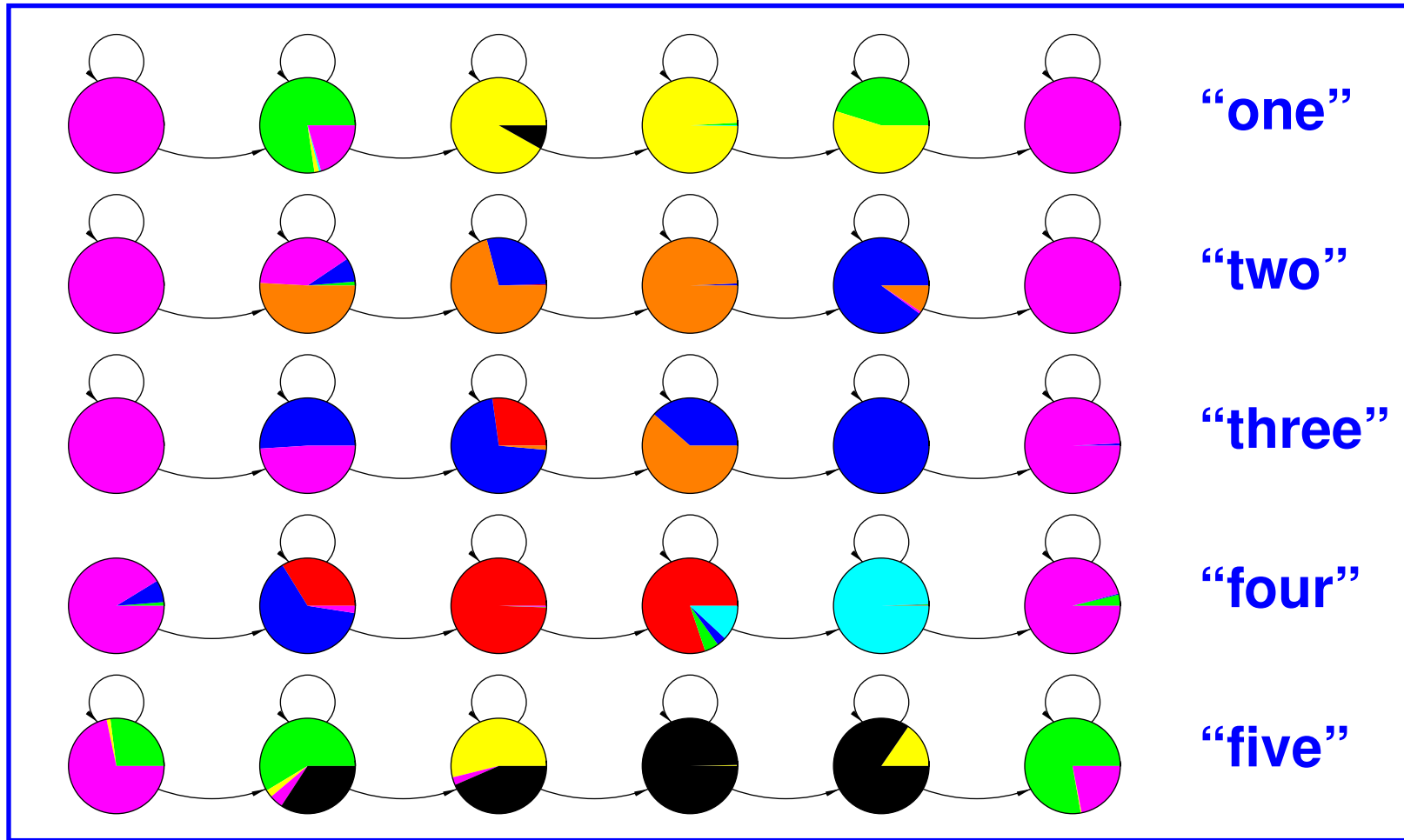


FFT magnitude prototype vectors (numbers example)

31.25Hz increments



Best HMMs (numbers example)

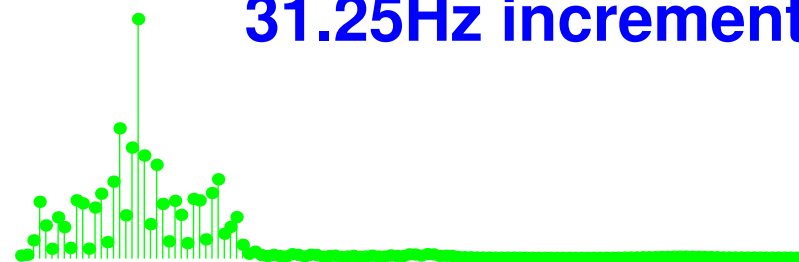


Classification error: 2% (200 test samples)

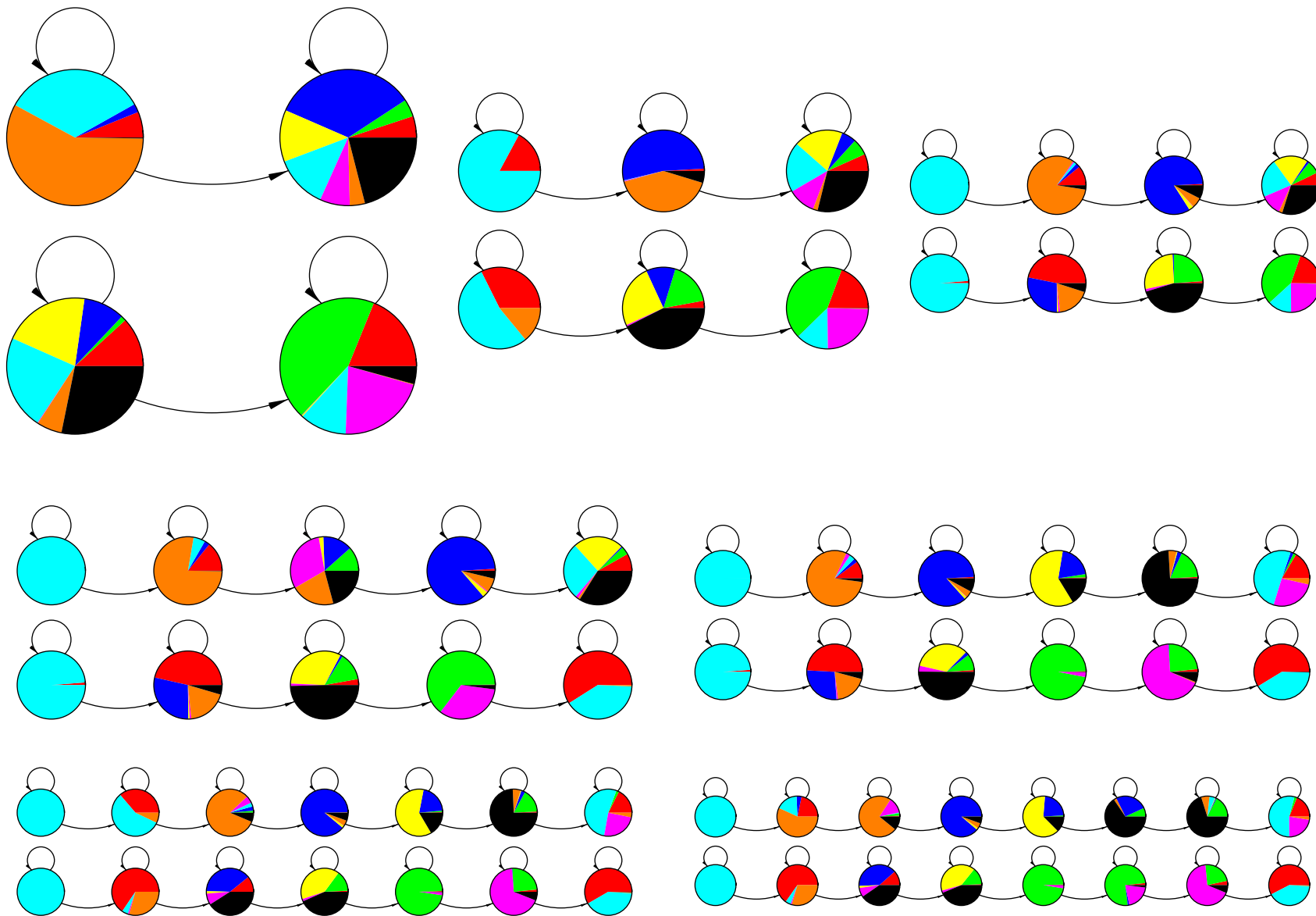
(2 one/five switches, 2 two/three switches)

FFT magnitude prototype vectors (dog/god example)

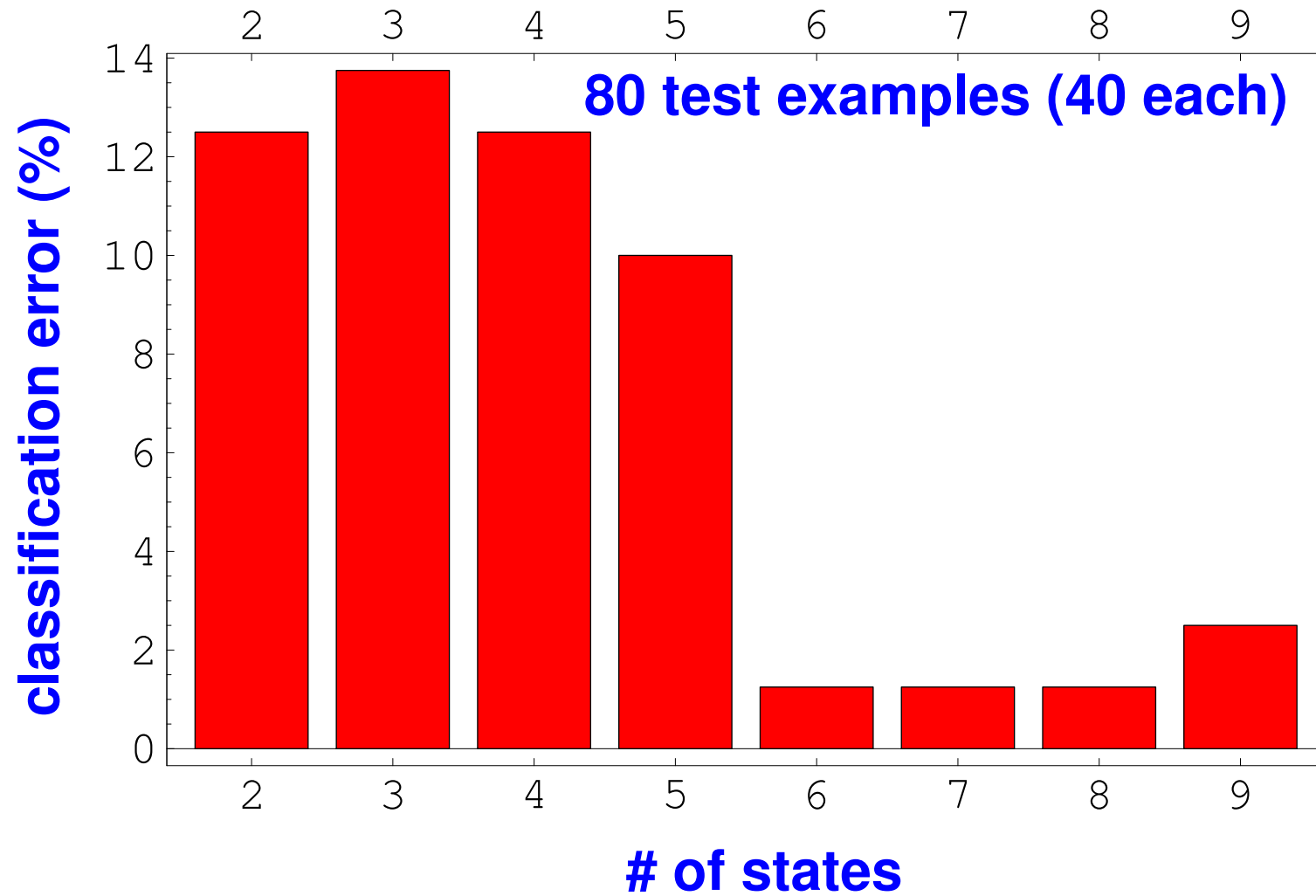
31.25Hz increments



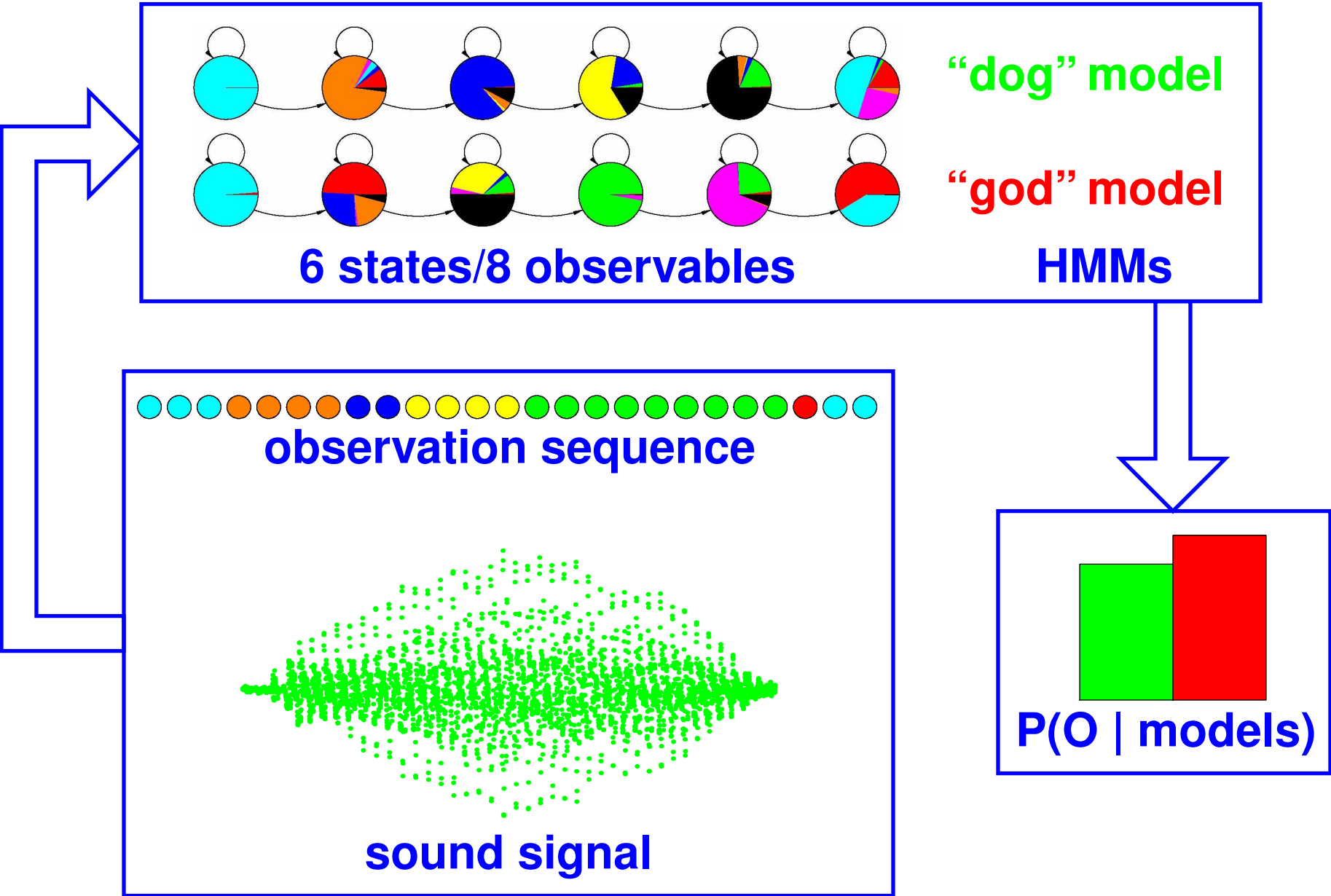
Varying the number of states (dog/god example)



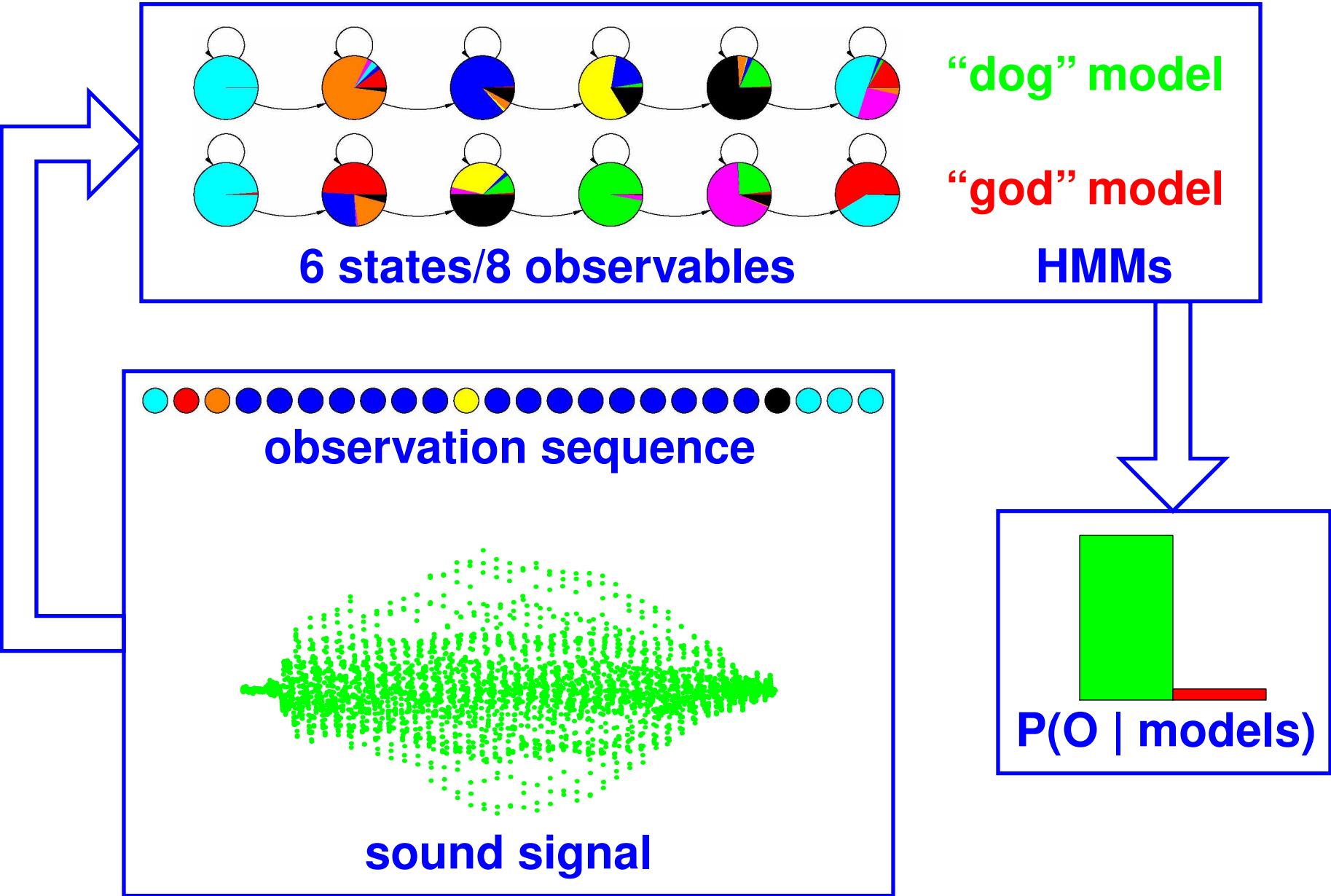
Varying the number of states (dog/god example)



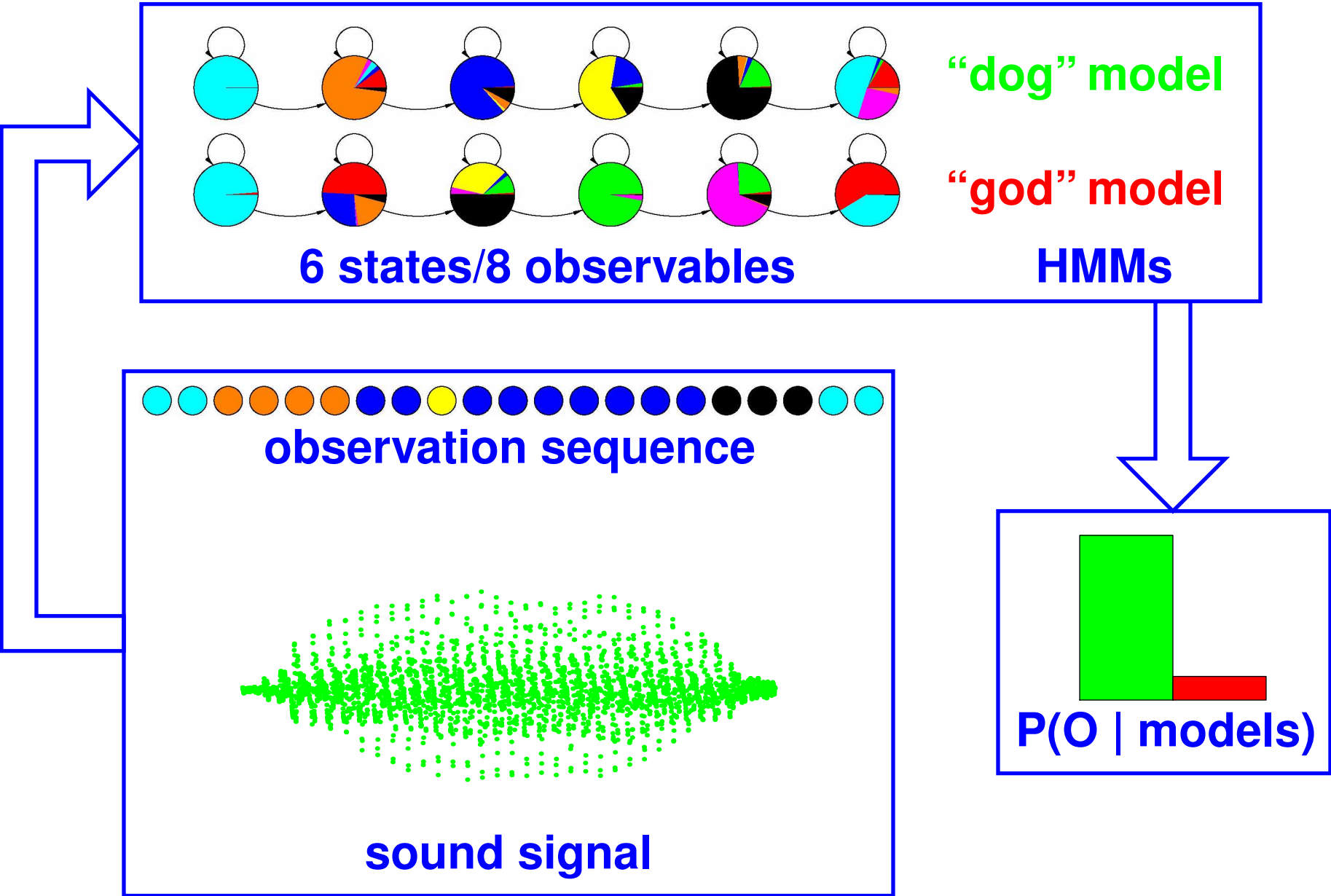
Word classification example (dog, #097): misclassified



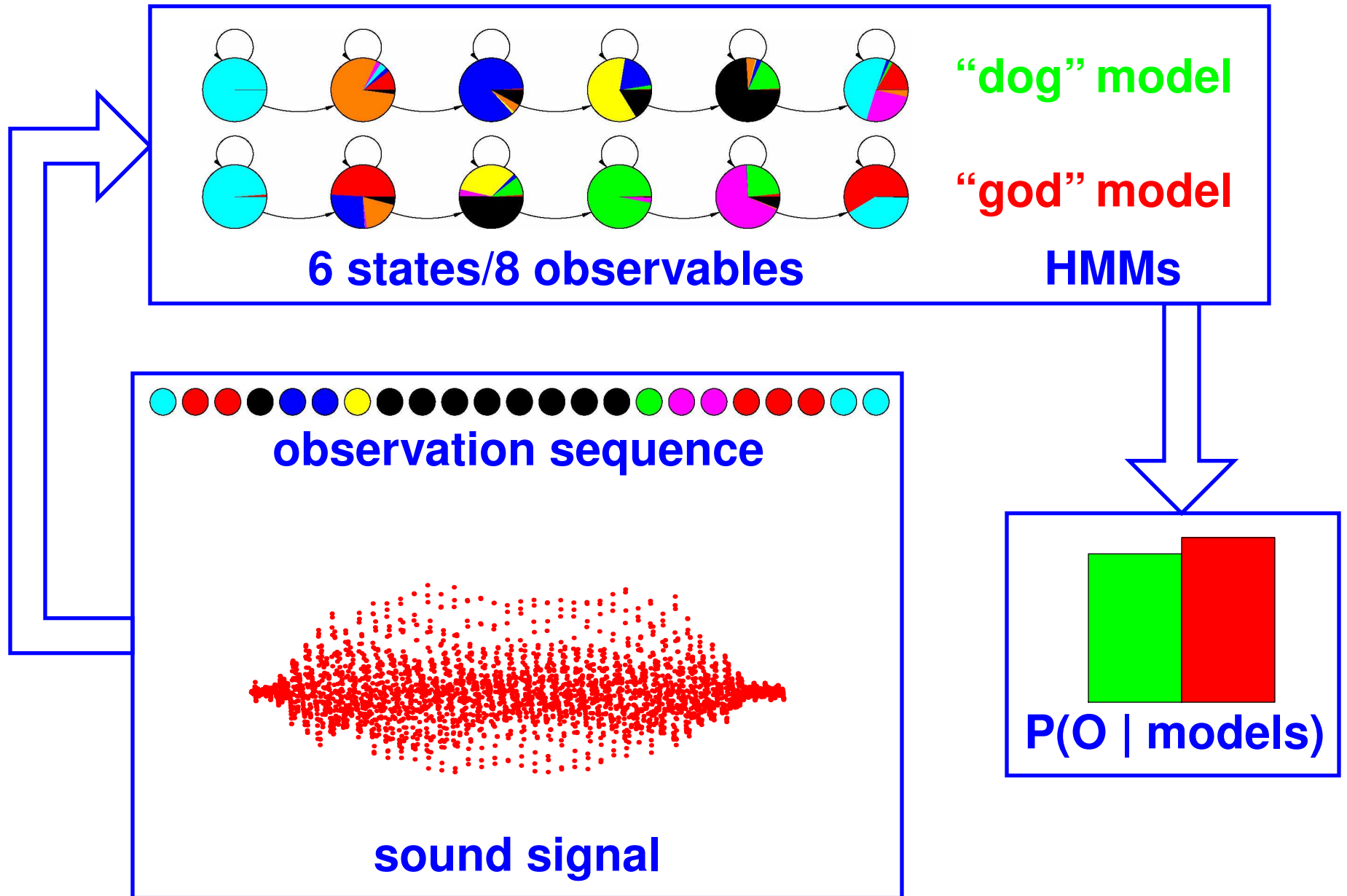
Word classification example (dog, #100): well classified



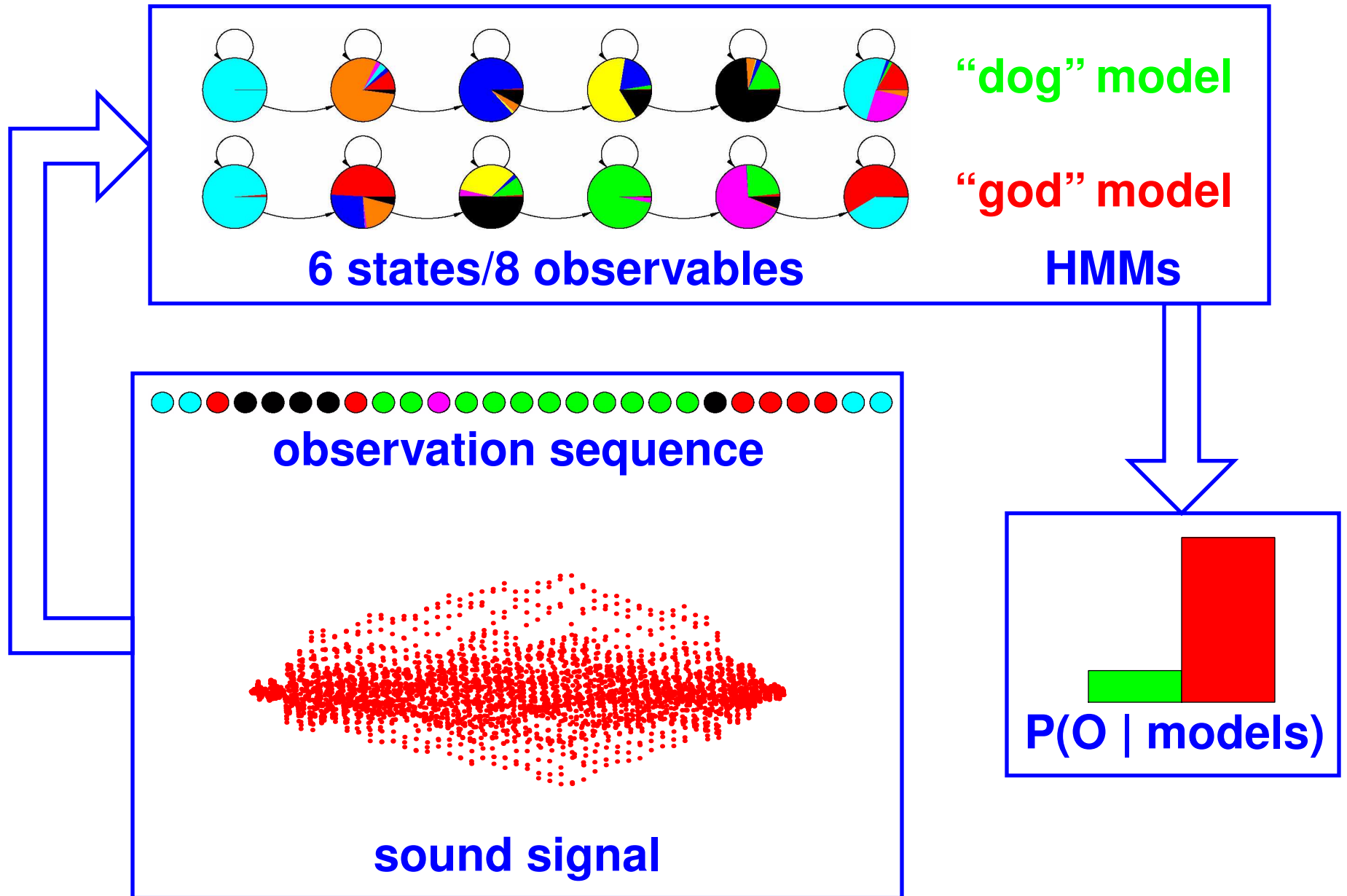
Word classification example (dog, #113): well classified



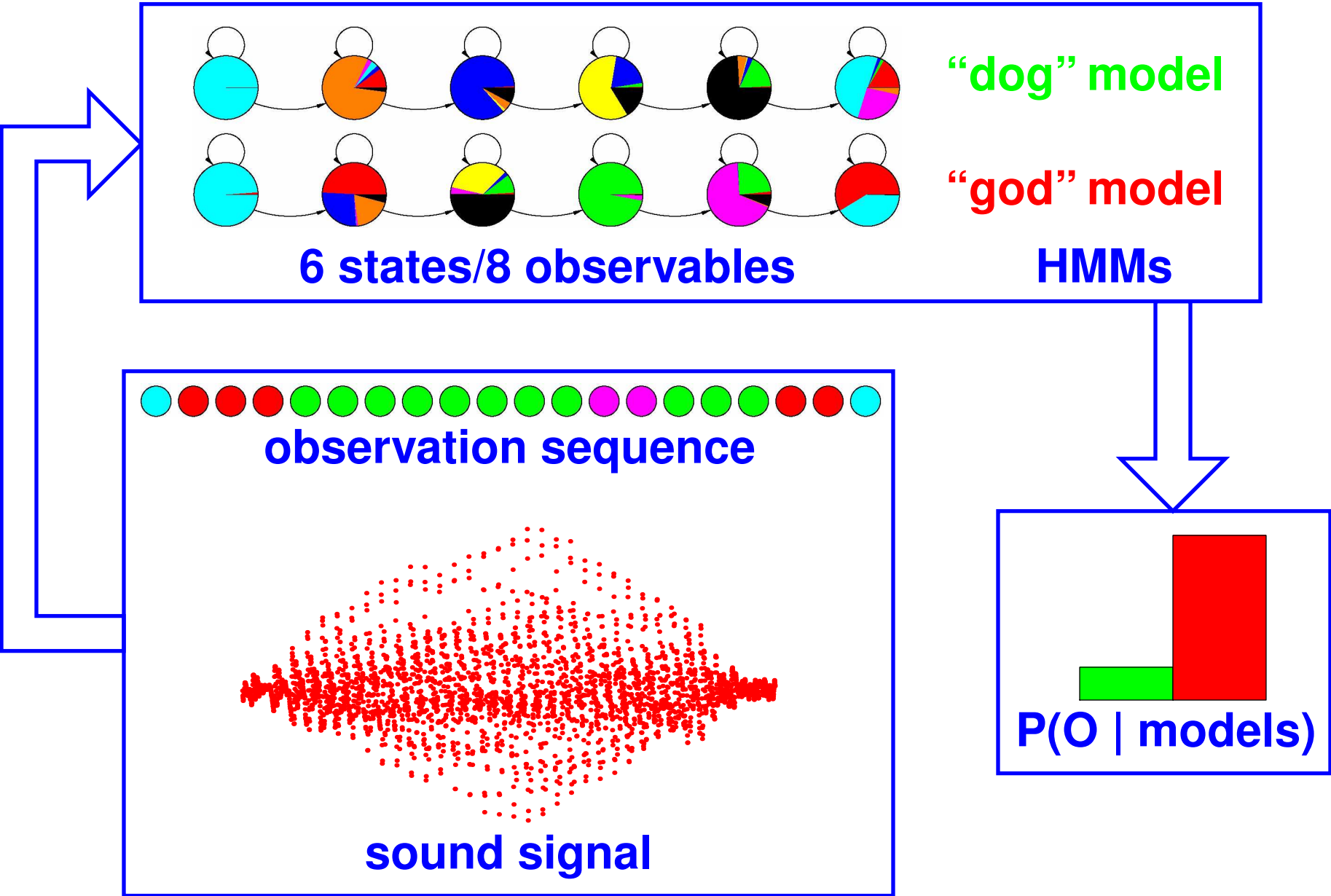
Word classification example (god, #112): poorly classified



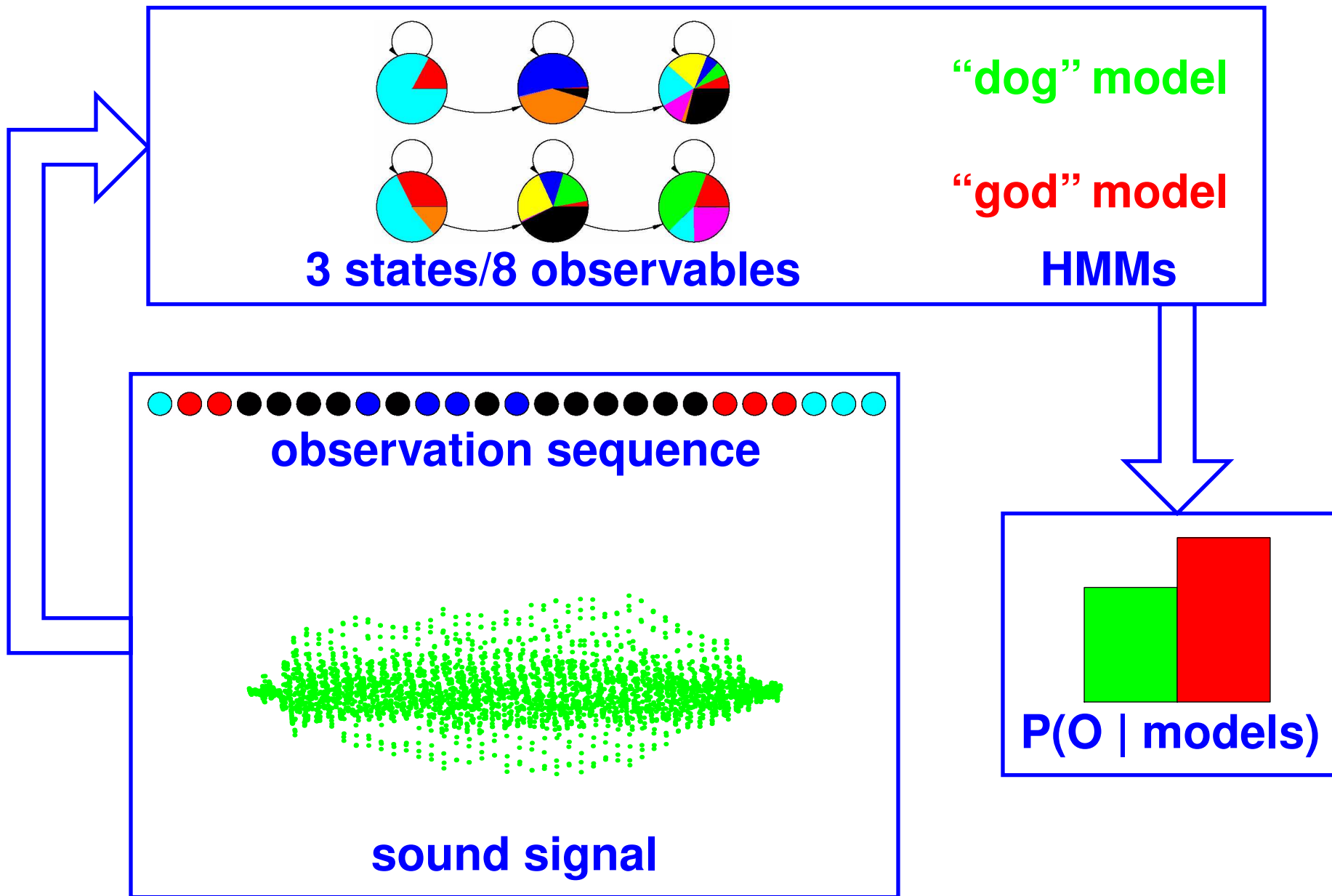
Word classification example (god, #107): well classified



Word classification example (god, #090): well classified



Importance of temporal structure (dog, #099): misclassified



Importance of temporal structure (dog, #099): well classified

