1. Not using cross-validation because we didn't know the grading criteria before and the average correctness rate is very close to the result of group 6. Therefore, we only presented the result of group 6. The cross-validation process is shown below.
We used Hanqi's data set, a total of 12 samples, to save time. We divided it into six groups with each group of two samples. In the first experiment, the first group was taken as the test sample, and the rest were used as training samples. Then we took the second group as the test sample, and repeated the experiment 6 times accordingly. The average correctness rate is 52.78%.

Group	1	2	3	4	5	6
Correct rate	66.6%	33.3%	66.6%	66.6%	33.3%	50%

2. Confusion matrix

Whole data set		Prediction		
		Song A	Song B	Song C
Reality	Song A	100%	0%	0%
	Song B	0%	0%	100%
	Song C	0%	0%	100%

3. Grading criteria 4
We plot a training sequence from Song A, B and C respectively and compare them against random output sequences generated by the corresponding trained HMM using @HMM/rand.

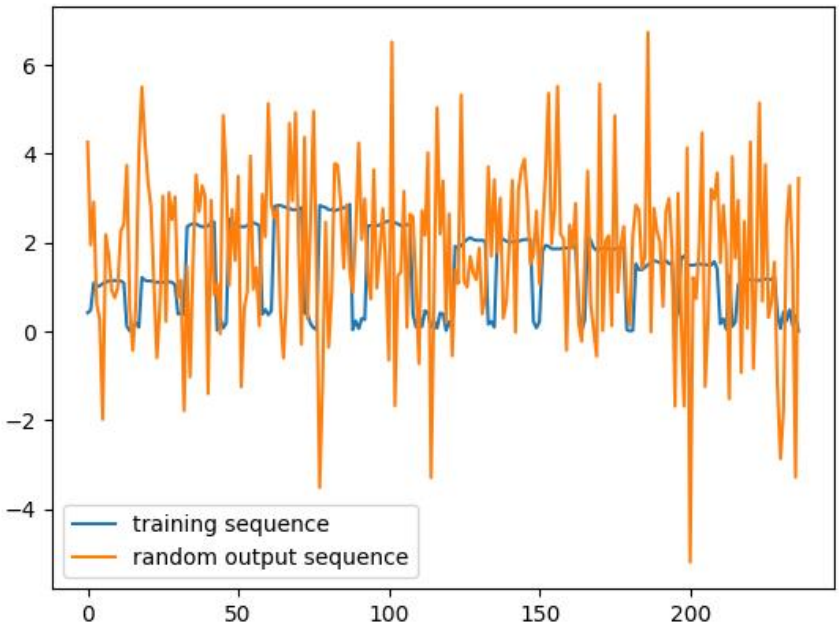

Figure 1: Song A

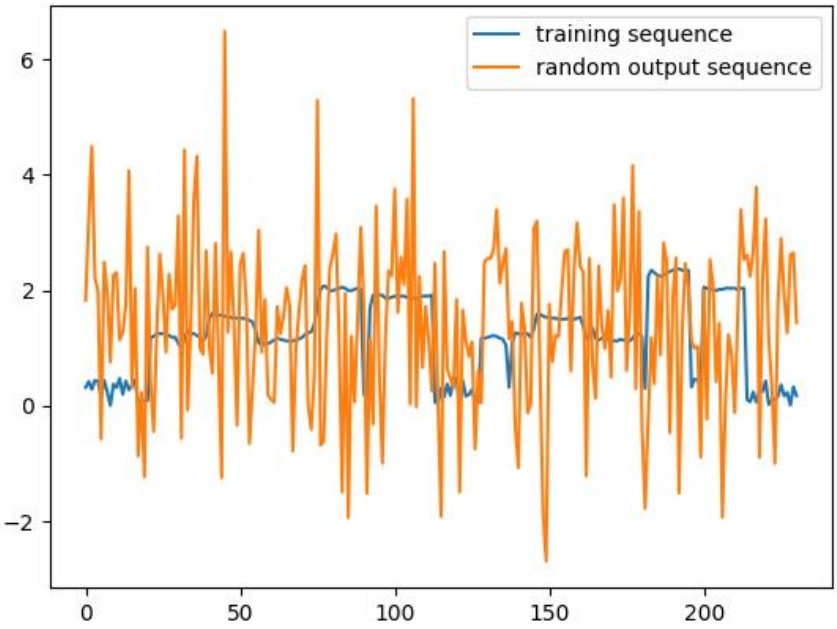

Figure 2: Song B

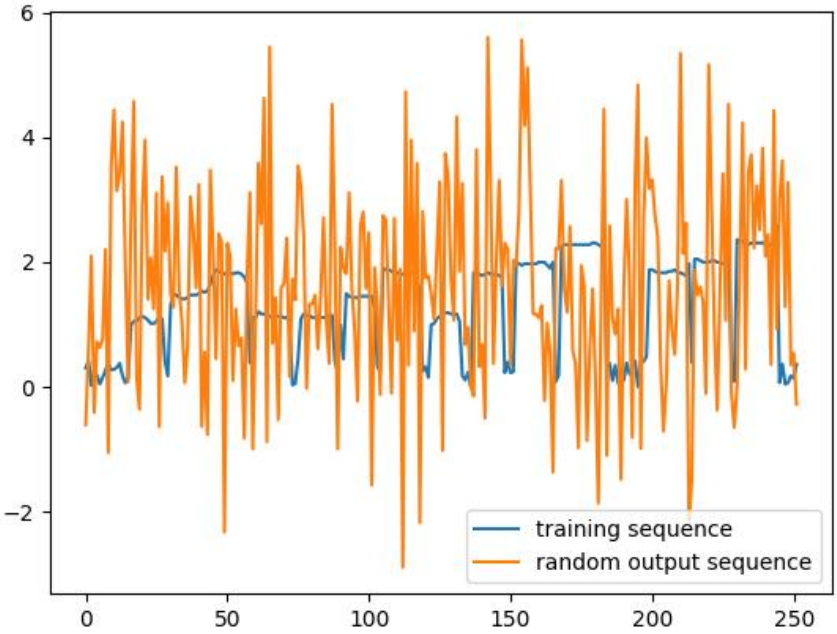

Figure 3: Song C