Homework 2

Problem 1 (columnar transposition)
Consider the ciphertext given in Figure 1, which was intercepted by our signals intelligence units as it was transmitted from an enemy camp. Our agent inside recently radioed the following message:

> "Ciphertext was produced by a keyword single-columnar transposition cipher applied to an English plaintext. Specifically, the plaintext was read row-by-row into a rectangle of unknown dimensions; the columns of this rectangle were permuted according to a keyword; and the letters were read out row-by-row. For convenience, the ciphertext was arranged in five-letter blocks: these blocks have no other significance. The message completely fills the encryption rectangle and the number of columns in this rectangle is either 8 or 10. … *(message unintelligible due to enemy jamming with loud music).*"

a) What evidence suggests that the encryption process is transposition?

b) Apply the *vowel-count method* to evaluate each of the two candidate rectangle sizes. Specifically, for each candidate rectangle size do the following. Write the ciphertext row-by-row into the candidate rectangle. Let r be the number of rows in the encryption rectangle. For each row $1 \leq j \leq r$, compute f_j to be the number of vowels ('a', 'e', 'i', 'o', 'u') that appear in row j. Finally, compute the sample variance $\hat{\sigma}_f^2$ of the observed vowel frequencies; that is, compute

$$\hat{\sigma}_f^2 = \frac{1}{r-1}\sum_{j=1}^{r}(f_j - \hat{\mu}_f)^2, \tag{1}$$

where

$$\hat{\mu}_f = \frac{1}{r}\sum_{j=1}^{r} f_j \tag{2}$$

is the sample mean of the observed vowel frequencies. Which of the two candidate rectangles has the lowest sample variance of vowel frequencies?

c) After determining the rectangle dimensions, we must determine how the columns were permuted. One powerful technique is to determine which columns "fit together nicely." To carry out this technique, it is helpful to measure how well two columns fit together. Let the two columns be denoted by $X = (x_1, x_2, \ldots, x_r)$ and $Y = (y_1, y_2, \ldots, y_r)$. Thus, x_1 and y_1 are the letters in row 1 of columns X and Y, respectively.

Develop one or more quantitative scoring functions that measure how well two columns fit together. (*Hint:* consider the bigram probabilities $\mathrm{Prob}[x_j y_j]$ or the conditional bigram probabilities $\mathrm{Prob}[x_j y_j | x_j]$. Also, see pages 76–77 of Sinkov.) Precisely define and intuitively explain each of your scoring functions. Does it make more sense to use the unconditional bigram probabilities or the conditional bigram probabilities? Explain.

d) According to your scoring function, which column fits best next to column 1? How well does your scoring function work?

e) What other techniques are useful for anagraming the columns once the dimensions of the encryption rectangle have been determined?

f) What is the plaintext? Explain how you arrived at your solution.

```
eudod lhmal hfdla lhode himra eueah crnpa mtgaa
aivms adpec yratt nreie idhat gnnen slhwl yyeaa
efnmu oveis iprts sfoat nwnge niiti iwtni ihghk
ovhpy sjiee epdle edopv ivany ooils arahm slenn
```

Figure 1: Ciphertext produced by a single-columnar transposition cipher.

Problem 2 (vowel-count method)
Describe the *vowel-count method* for solving columnar transposition ciphers, and explain why it works. In your essay, include as much relevant quantitative analysis as possible.