Introduction
The goal of this study is to attribute the remaining incorrectly transcribed versions of the letters that John Mills wrote to his son to either Typist-1 or Typist-2. We focus on the additional letters we received for this lab that lack their original, uncorrupted text.

We now have access to a total of twenty-four corrupted letters. Of them, only twelve of have their corresponding original versions. We are given that Typist-1 typed letters 2, 14, and 17, and Typist-2 typed letters 12, 13, and 19 in addition to the data we have from our last study that says Typist-1 typed letters 1, 8, 16 and Typist-2 typed letters 4, 9, and 18. As there are six letters that both lack attribution and original versions, we will attempt to attribute these letters, letter 5, 6, 20, 21, 23, and 24 to one of the two typists and eventually attempt to restore them to their original versions.

We will begin by training on a single letter from lab 7 that is attributed to a specific typist to create a sensor model. We can then calculate Jayness evidence for the typist of the other two attributed letters without using their original versions. We will then repeat the same process by training these letters using the model we built in the last lab. This model takes into account the original versions of each letter. Doing so allows us to compare the values of Jayness evidence in these two cases and assess how the unavailability of the original text effects our predictions. In order to understand how reliable our predictions are, we will train models for both typists on the six letters that are attributed and have original versions, and use this information as evidence to confidently say how often our predictions are correct. Finally, we hope to give attribution to these new six letters that lack both attribution and originals by testing them on our reliable model.

Attribution

Test A: There are three ways to choose a letter by Typist-1, three ways to choose a letter by Typist-2, and $2 * 2 = 4$ ways to choose the third letter, giving us a total of 36 different configurations. We trained the language model by adjusting our transition model according to how many times each character was preceded by another in the original, uncorrupted letters. That is, we used the six attributed letters (three by each typist) that have their original versions to train on a single letter at a time for each typist and calculate Jaynes's evidence for the other four labeled letters. The major difference in this part of the lab and the previous one was that we did not use the original versions of these four letters while calculating the evidence.

Out of the 36 predictions, 20 of them were correct. We notice that this method of using a language model and sensor model to predict the typist of each letter is slightly better than using the original versions to build a model by training on just a single letter. However, this model is not as accurate as the one where we trained on two letters by each typist. The scatter-plot below shows this data.

Letter	Evidence	Attribution
5	9.1544932	Typist-1
6	-10.91376	Typist-2
20	-124.9891	Typist-2
21	-50.15179	Typist-2
23	1.991903	Typist-1
24	-30.16881	Typist-2

Conclusion

As shown in the results above, we have attributed letters 5 and 23 to Typist-1, and letters 6, 20, 21, and 24 to Typist-2. Since the accuracy of using our language-model is around 55%, we can confidently say that this classification will work better than random chance. However, this is only slightly better than randomly attributing the letters to one of the two typists. We also notice that the magnitude of the evidence of letter 23 at around 2 decibels is a lot smaller than that of the other letters, making us the least confident about this attribution. We believe that having a larger set of uncorrupted, attributed letters to train our model on will result in greater accuracy of our language and sensor models.