Query by Example

David Barbas Rebollo

Automatic Speech Recognition
January, 2021

1 Introduction

In this laboratory assessment, the task at hand was to implement an algorithm to calculate the cost between a word and all segments of an audio file. Once done, it is possible to analyze the appearence of a target word based on the associated cost of each segment.

Additionally, the whole sequence was analyzed by frames to find hidden local minimums as the global minimum is fairly easy to attain.

2 Results

Experiments were conducted with several different words to check if the results obtained correspond with the searched target word. All experiments were done using cosine distance and the top 6 results are analysed. The resutls and its analysis are described below:

2.1 0013 Música

Música is the word with 4 appearences so, it would be interesting to see how many appearences could be detected by the algorithm. Despite this fact several pronuntiations are missed, starting at 2125.99 and 2207.85, probably due to the window size. As they are near the second pronuntiation found.

Table 1: Comparision of actual appearences and predicted appearences for word 0013 Música.

Actual		Predicted		Cost
Start	End	Start	End	
2000.66	2005.01	2004.21	2005.05	0.000561
0.0	0.0	-	-	0.005628
2282.07	2283.72	2283.24	2283.79	0.251484
2282.07	2283.72	-	-	0.251484
1347.23	1355.48	-	-	0.264256
1319.62	1321.81	-	-	0.264430
753.0	754.41	-	-	0.270019

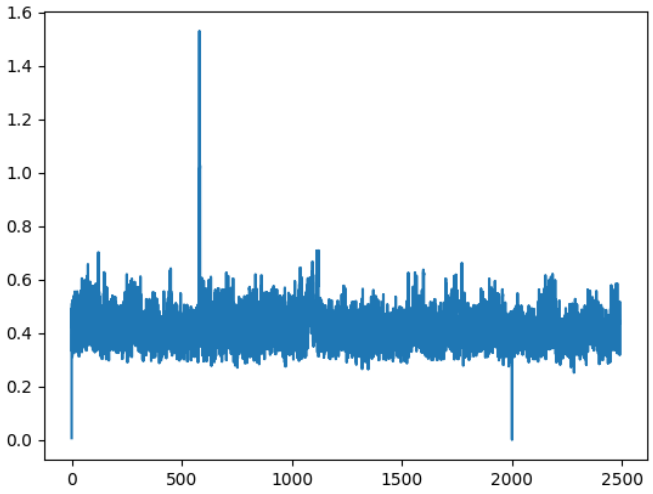

Figure 1: Costs results obatained from the implemented algorithm comparing the word 0013 Música with all segments of the audio file.

2.2 0030 Europa

Europa is the word which appeares once, however a phantom prediction appeares again at the start of the sequence. This can be seen in table 2 and figure 2. Even so, this word is still recognized.

Table 2: Comparision of actual appearences and predicted appearences for word 0030 Europa.

Actual		Predicted		Cost
Start	End	Start	End	
2038.05	2039.36	2038.82	2039.40	0.000970
0.0	0.0	-	-	0.006989
1375.11	1375.38	-	-	0.212457
896.47	896.65	-	-	0.223277
836.34	836.49	-	-	0.227647
1300.5	1300.98	-	-	0.228363

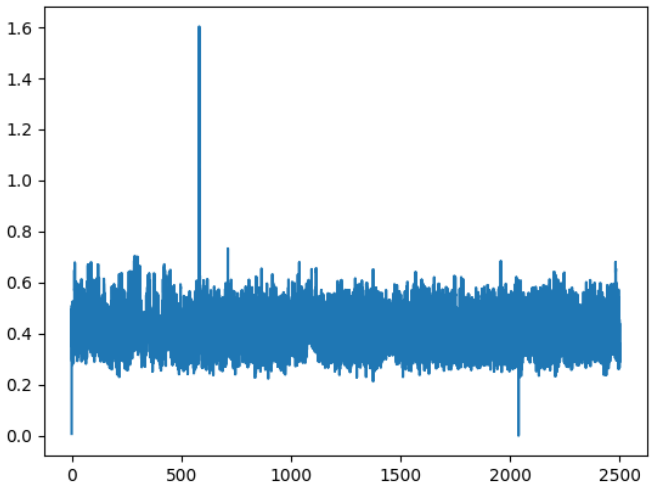

Figure 2: Costs results obatained from the implemented algorithm comparing the word 0030 Europa with all segments of the audio file.

2.3 0065 Noviembre

The 5 appearences of the word *Noviembre* which is clearly identified. However, there are 4 more, starting at 1169.26, 1472.34, 1770.85, 1893.28, which are completedly missed.

Table 3: Comparision of actual appearences and predicted appearences for word 0065 Noviembre.

Actual		Predicted		Cost
Start	End	Start	End	
558.79	559.34	558.89	559.38	0.000427
2202.62	2202.64	-	-	0.152055
483.12	483.15	-	-	0.154956
1680.61	1680.68	-	-	0.155557
398.37	398.42	-	-	0.162991
2270.3	2270.33	-	-	0.165812

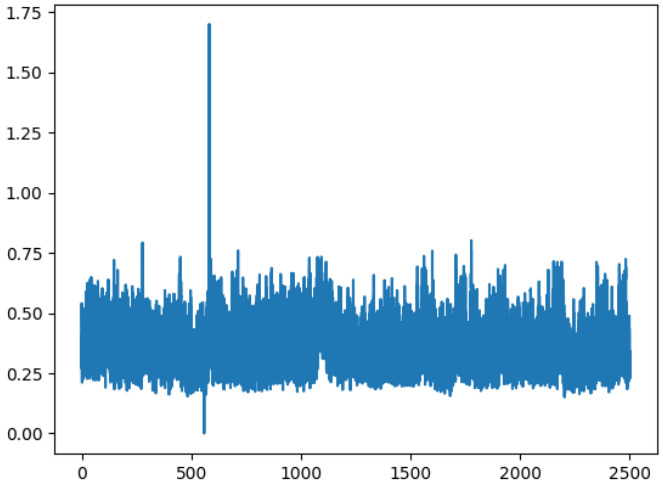

Figure 3: Costs results obatained from the implemented algorithm comparing the word 0065 Noviembre with all segments of the audio file.

2.4 0067 Oportunidad

As we can see in the results below, the algorithm peforms really well indetifying long words which appear once.

Table 4: Comparision of actual appearences and predicted appearences for word 0067 Oportunidad.

Actual		Predicted		Cost
Start	End	Start	End	
1637.29	1639.58	1638.71	1639.62	0.000497
2309.45	2310.88	-	-	0.233600
1689.25	1690.09	-	-	0.235421
2274.51	2275.12	-	-	0.235573
2208.99	2209.32	-	-	0.238933
818.03	821.97	-	-	0.240074

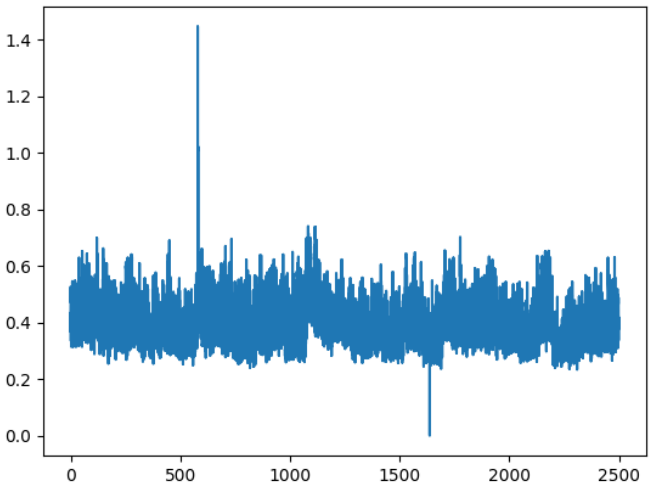

Figure 4: Costs results obatained from the implemented algorithm comparing the word 0067 Oportunidad with all segments of the audio file.