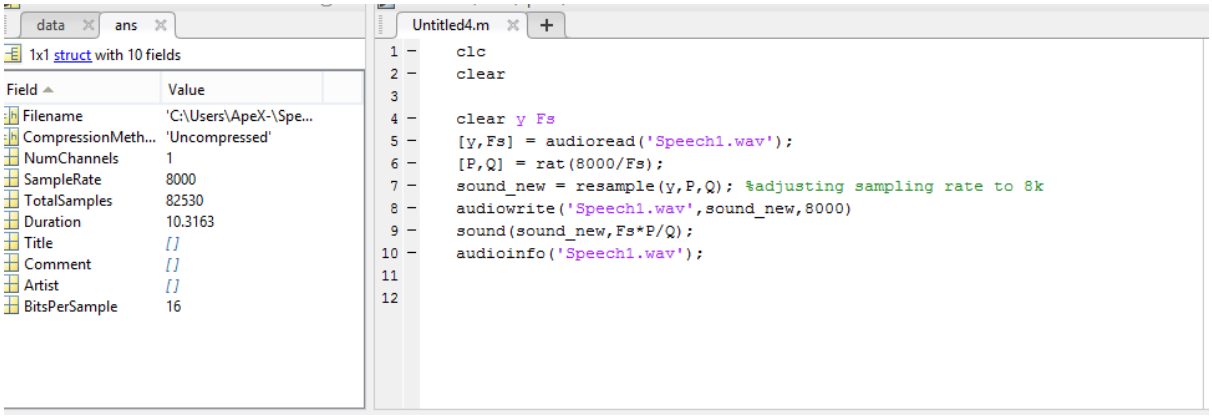
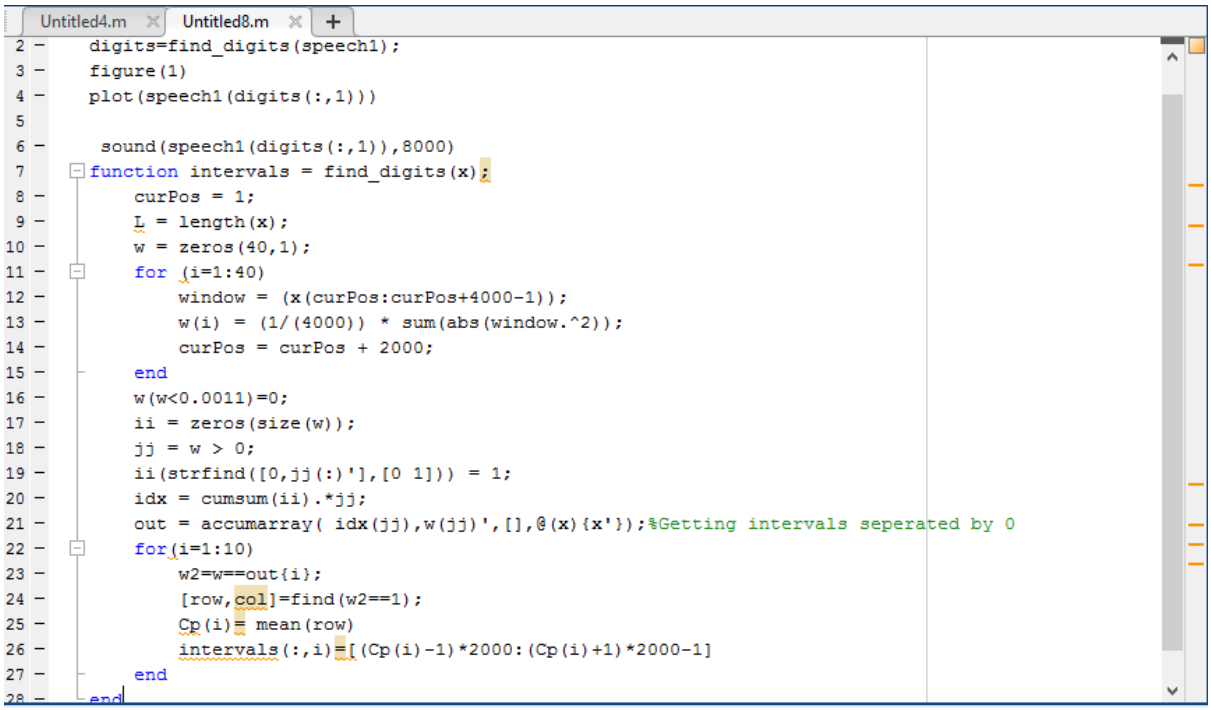
Part1:



Part2:



We calculated energy of the signal in different intervals with some overlaping to increase accuracy, then we seperated each interval by zeros(silent intervals), then we specified the interval of each digit's sound.

Part2:

Figure: 1, digit: 0

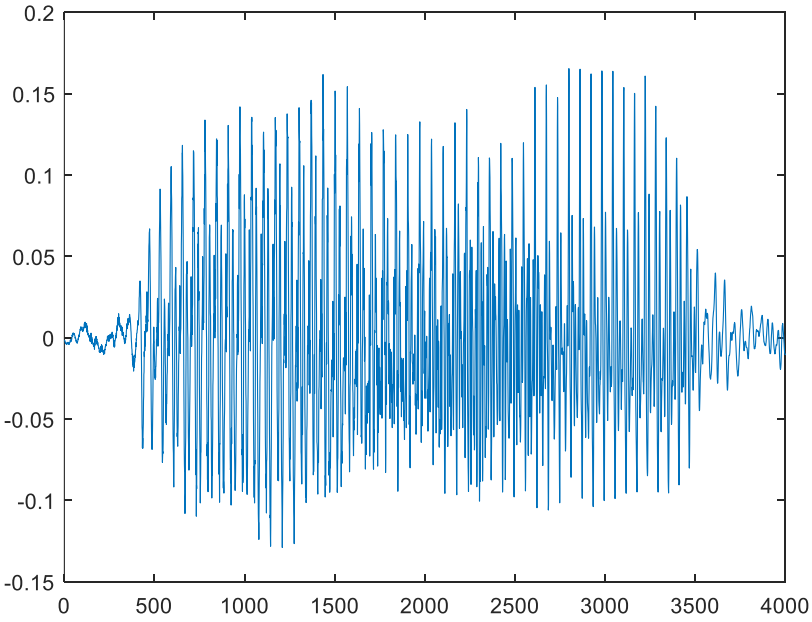


Figure2 digit: 1

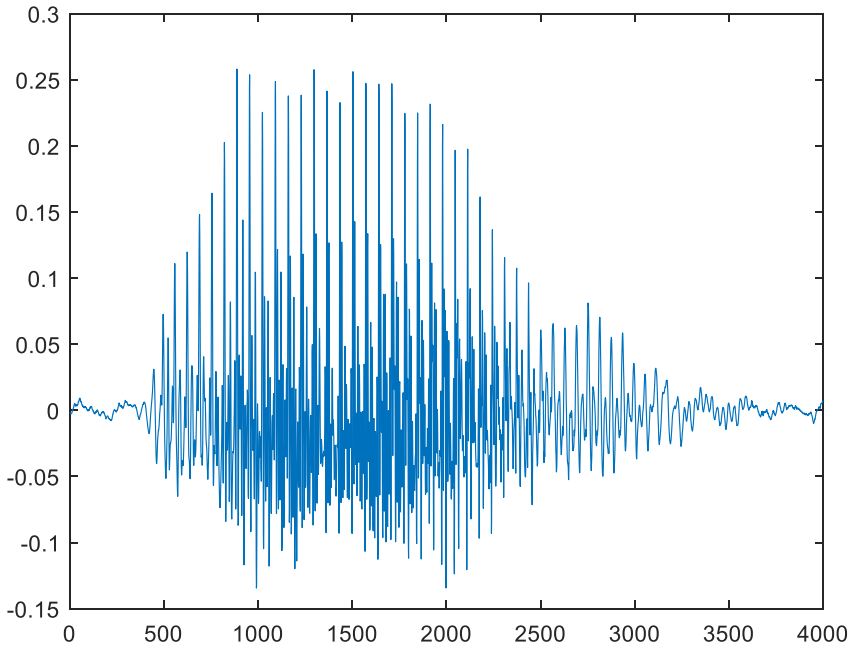


Figure 3 digit: 2

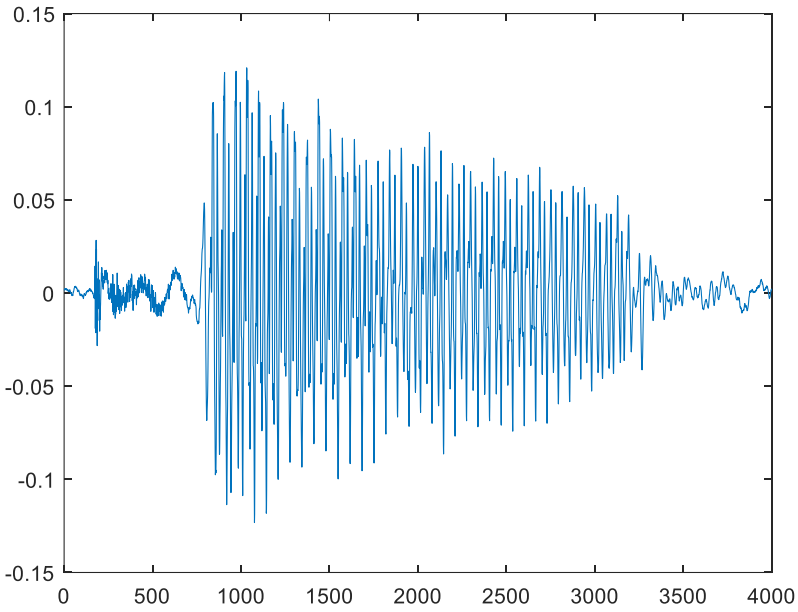


Figure: 4 digit:3

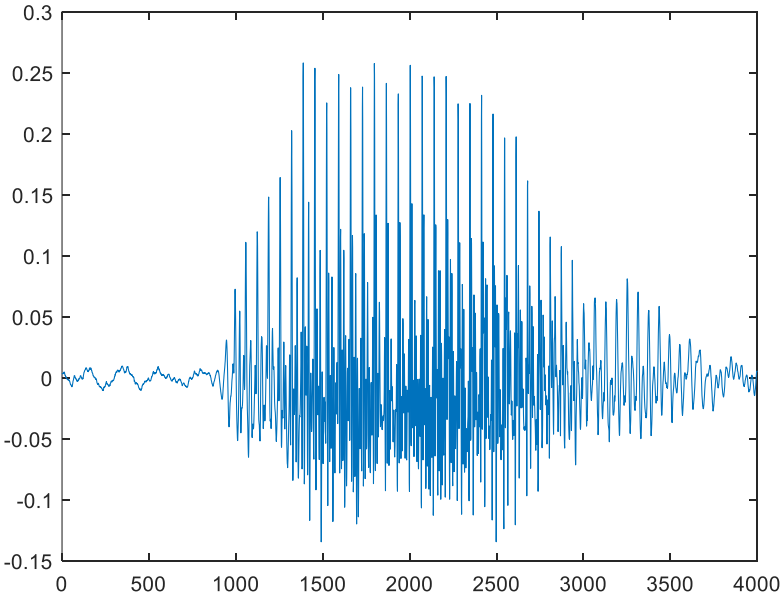


Figure 5 digit: 4

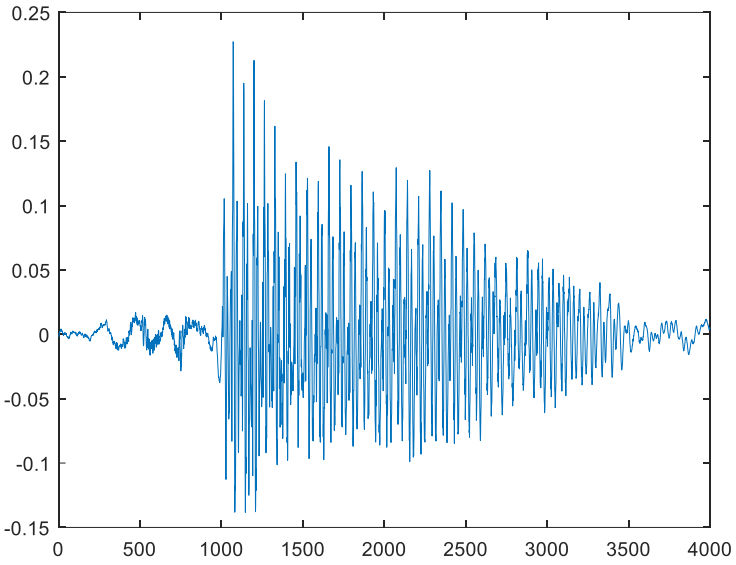


Figure 6 digit: 5

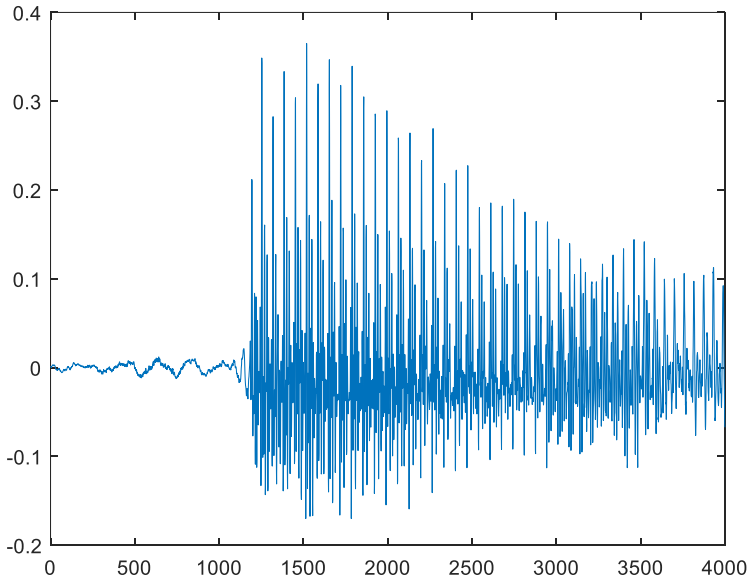


Figure 7 digit: 6

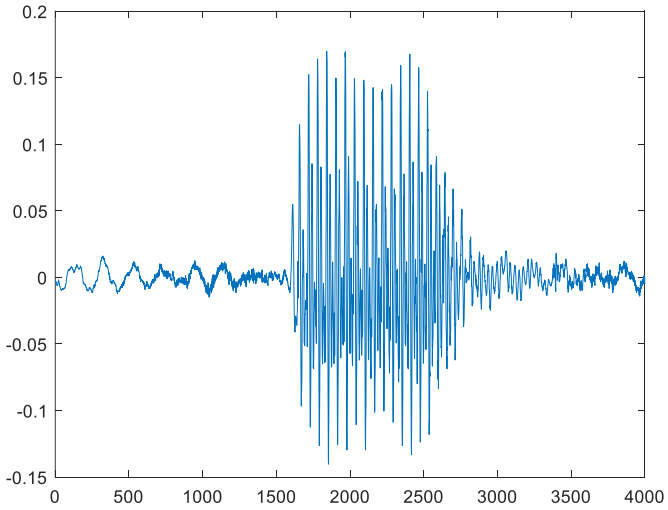


Figure 8 digit: 7

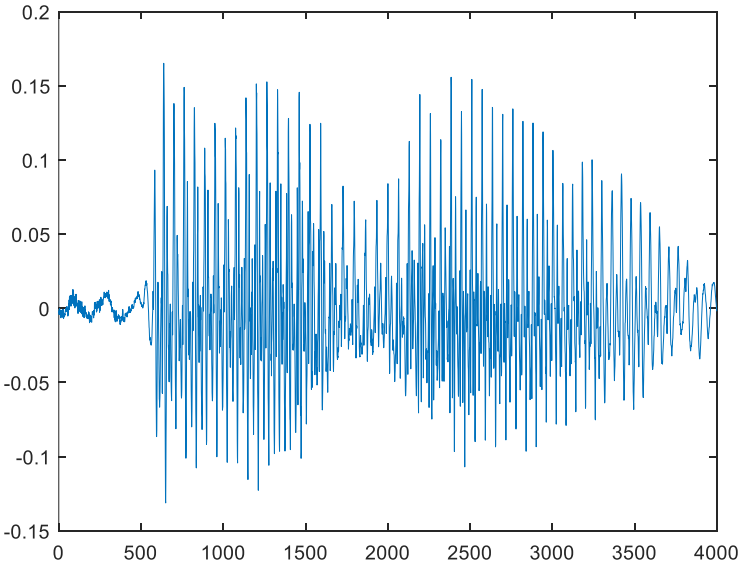


Figure: 9 sound: 8

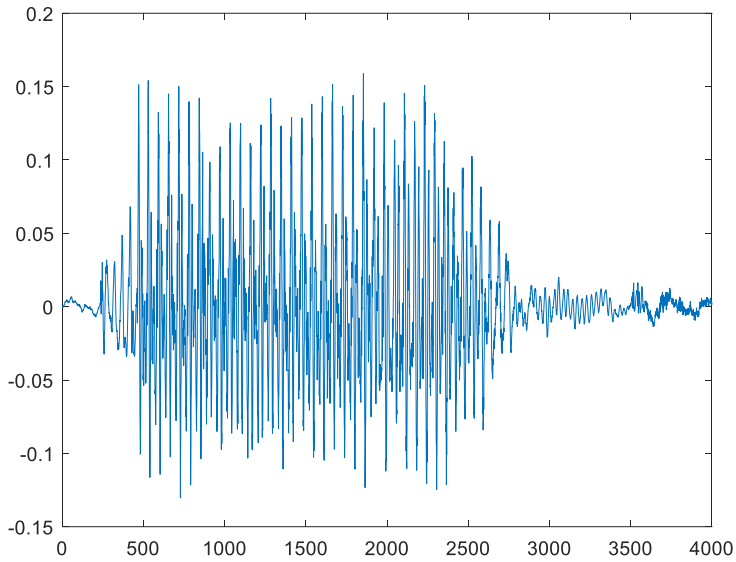
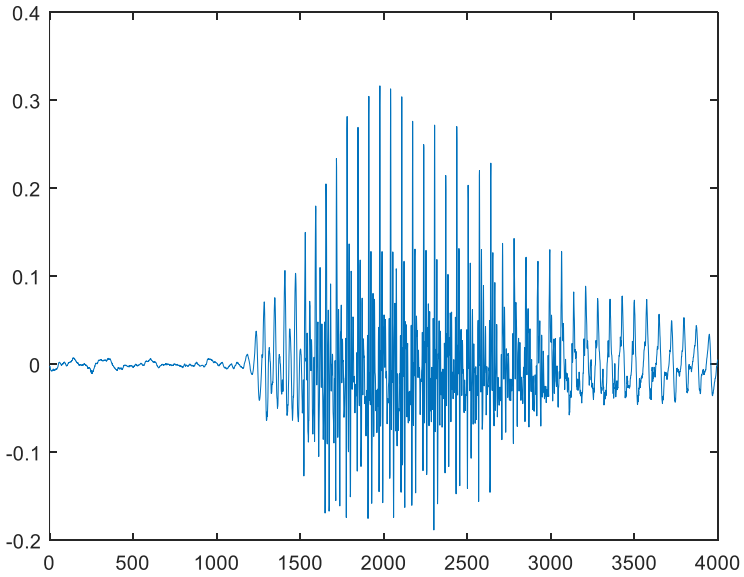
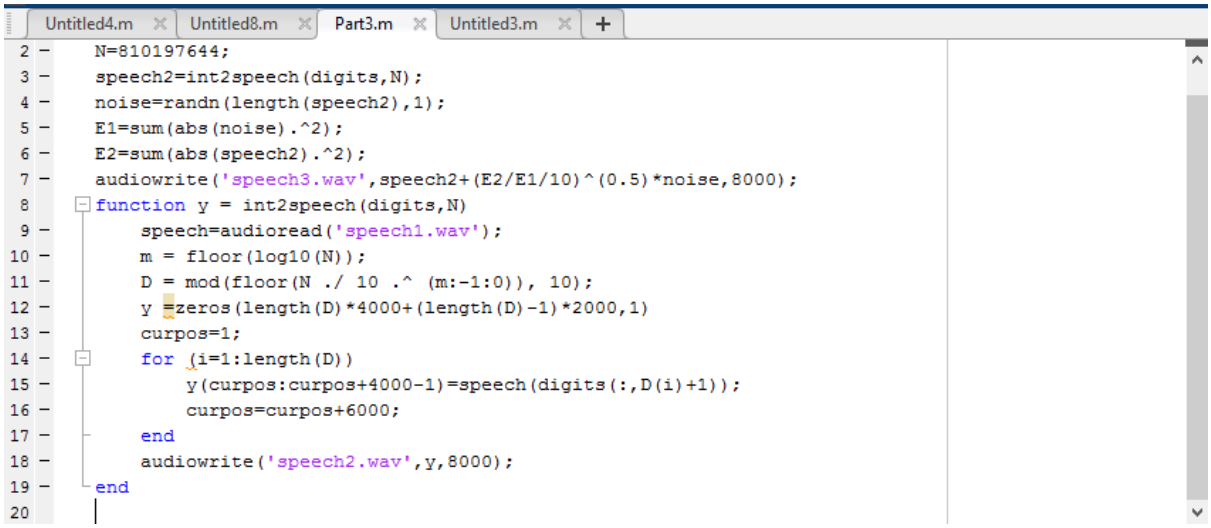


Figure10 digit: 9



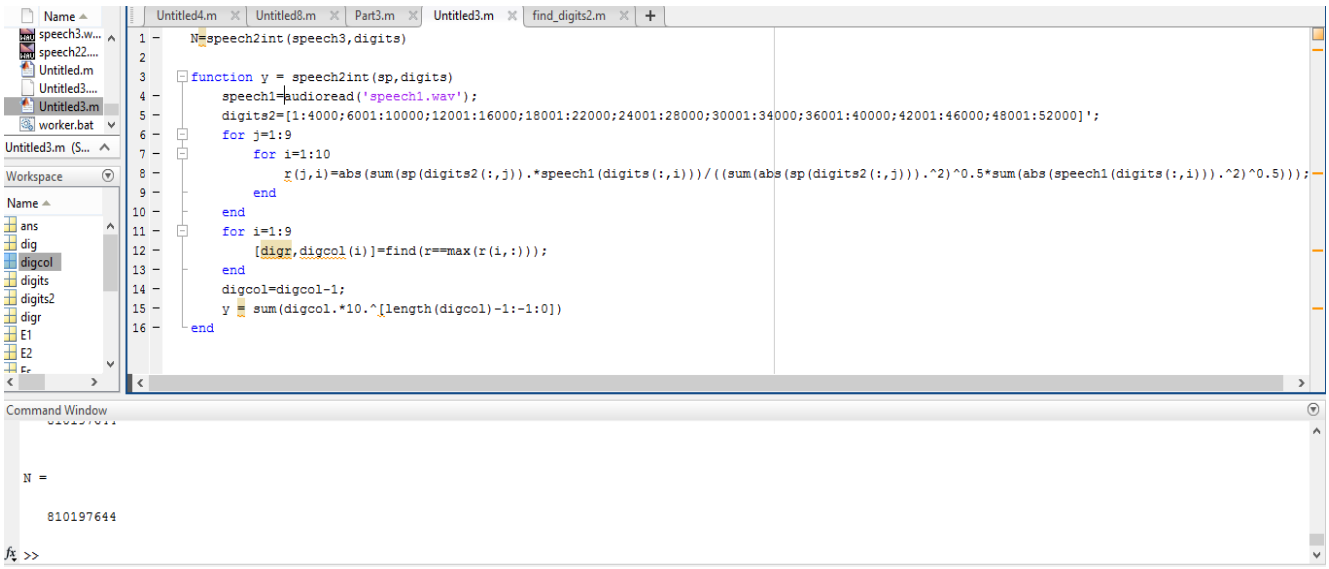
Part3:



We specify a new vector corresponding to sounds of pronunciations of N digits with 0.25s pauses in between.

Then we add some noise with a coefficient so that the energy of signal is 10 times energy of the noise.

Part4:



Codes, and the result N is perfectly correct.

First we specify intervals of digits pronunciations, then we build up the correlation matrix using 2 'for' loops, after that we determine at what number there is max correlation between the input voice and the one we used as our reference. That's the digit pronounced in the input voice, then we concatenate the digits and it gives us the output.