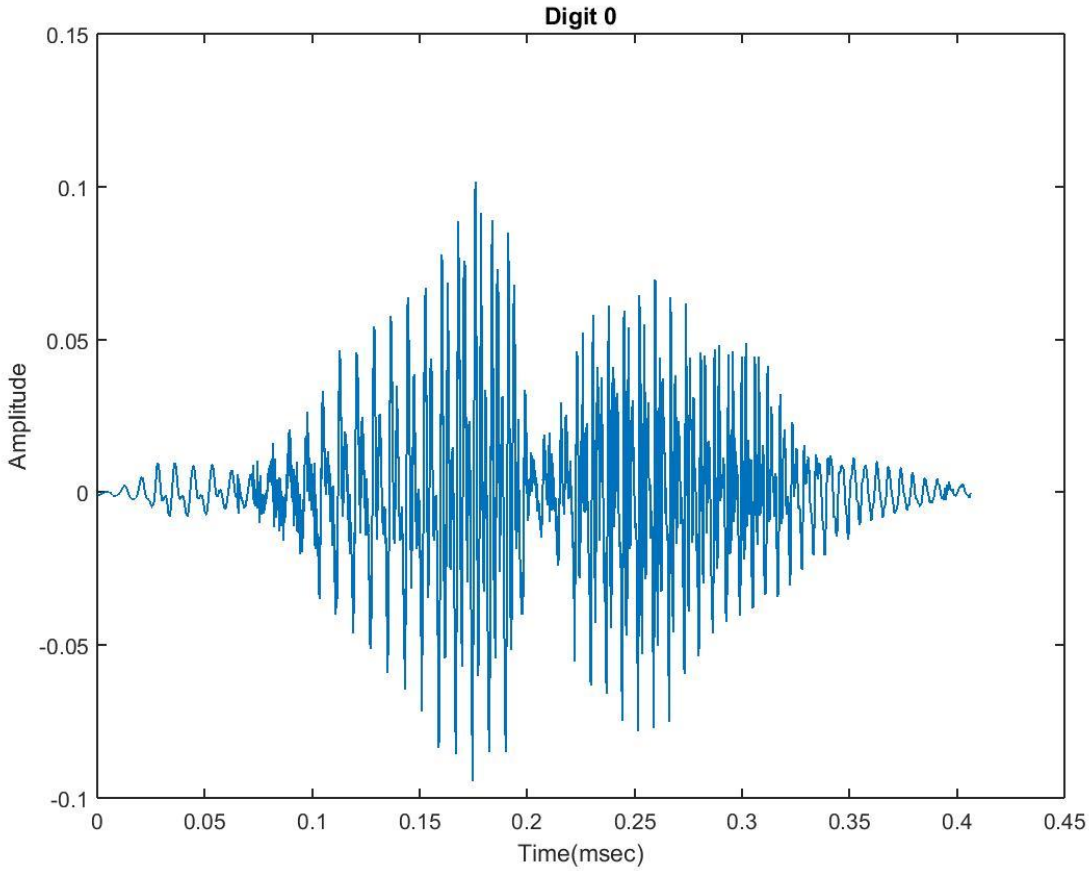
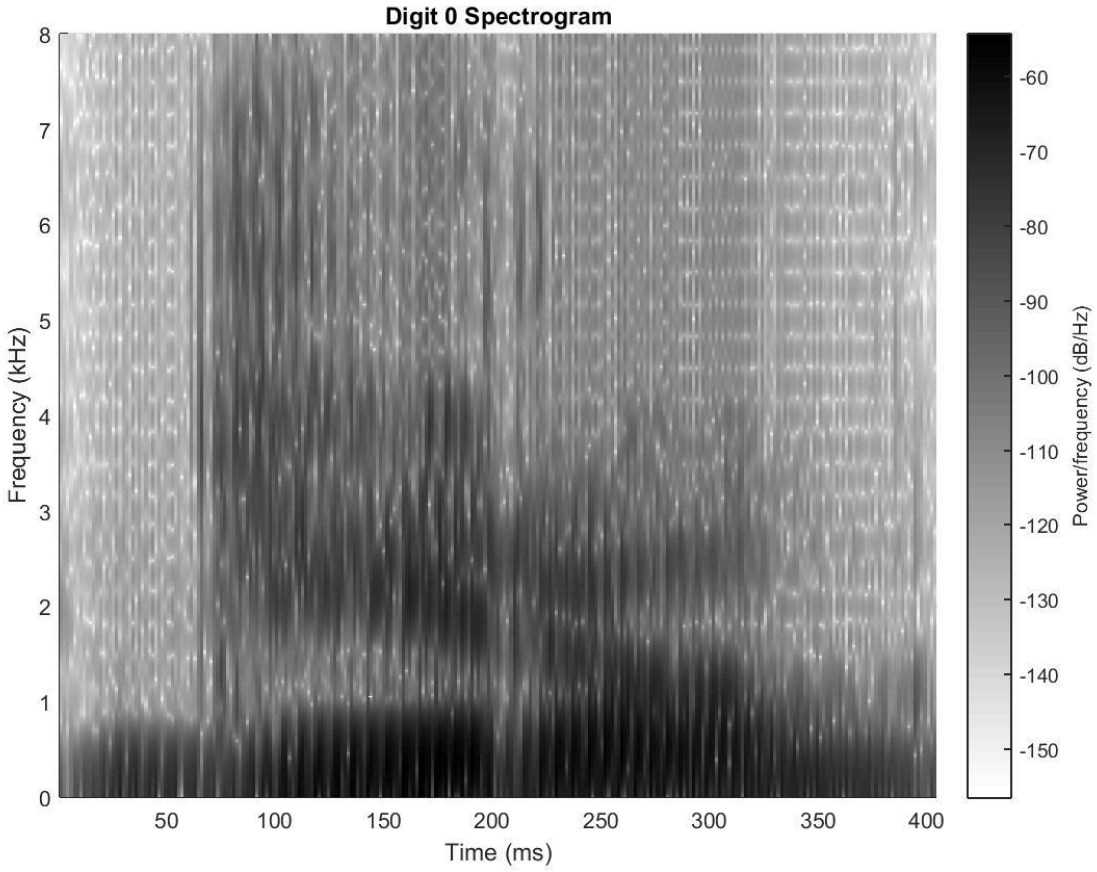
Assignment 2 : CRL 707

Akashdeep Bansal (2016ANZ8049)

Q7 a & b) Time waveforms and wideband spectrum of various digits are –

Zero

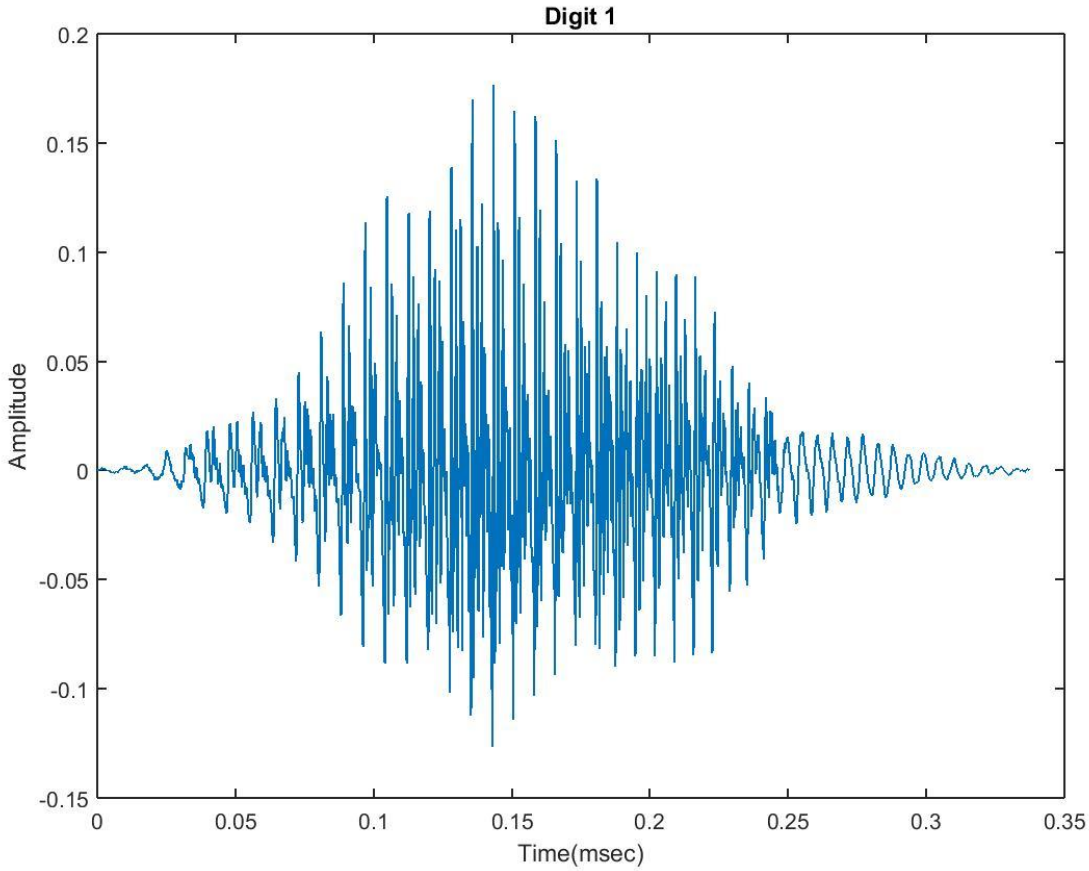


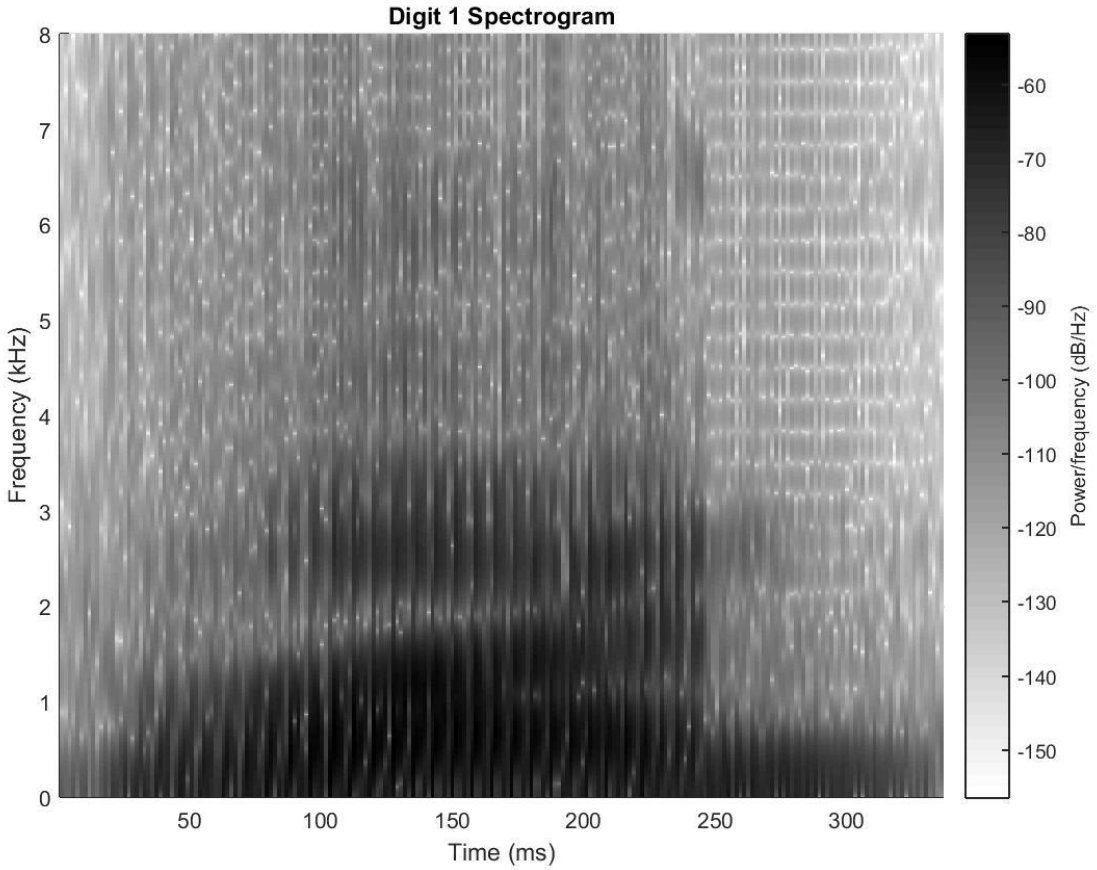


Phonetic Transcription of Zero: ˈzɪroʊ

The phoneme z is a fricative , in which energy is distributed. We can see from spectrogram plot, it seems to be lie around 0 to 55 ms. ɪ is a front vowel, which we can identify with the help of formant frequencies. So, the expected range is 55 to 160 ms. r is a semi-vowel and o and ʊ are back vowels. So, it's a bit hard to exactly figure out the separate boundaries. r is expected between 160 – 250ms. o is expected between 250 – 330ms. and ʊ is expected from 330 – 400ms, it's a release of energy.

One

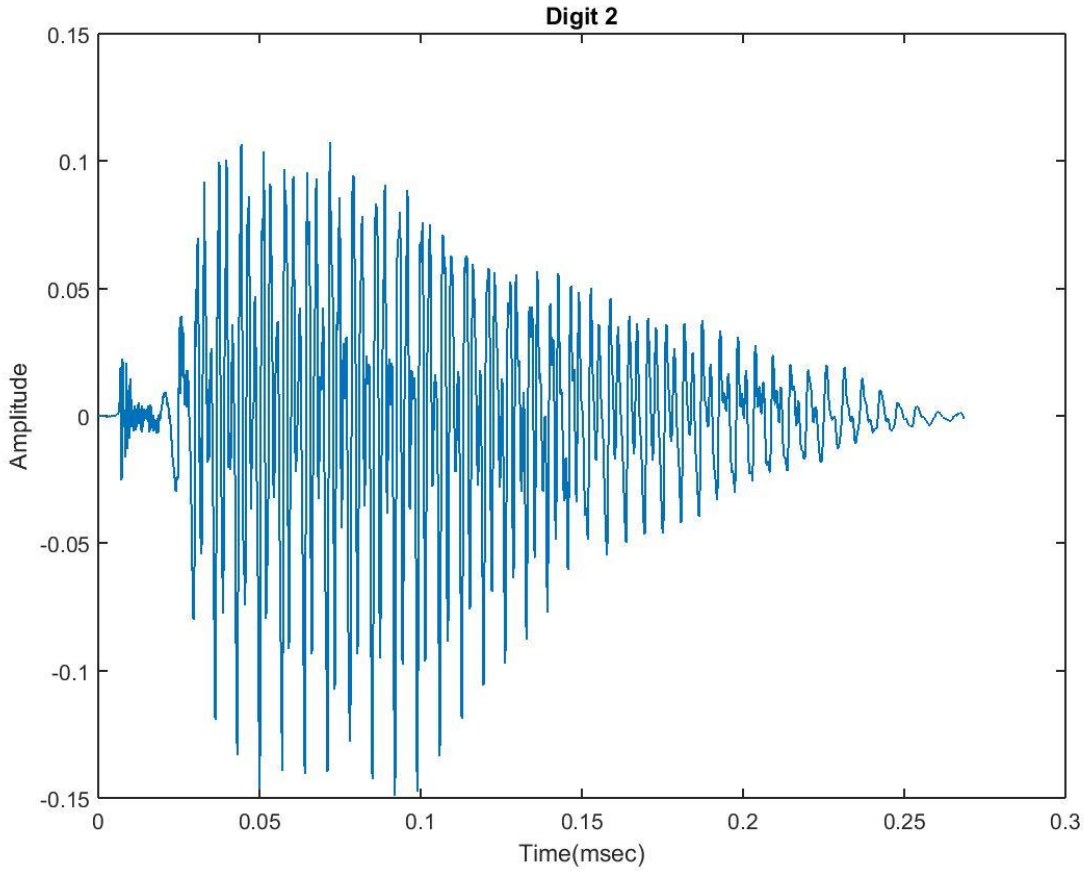

Digit 1
Amplitude
Time(msec)
0.2
0.15
0.1
0.05
0
-0.05
-0.1
-0.15
0
0.05
0.1
0.15
0.2
0.25
0.3
0.35

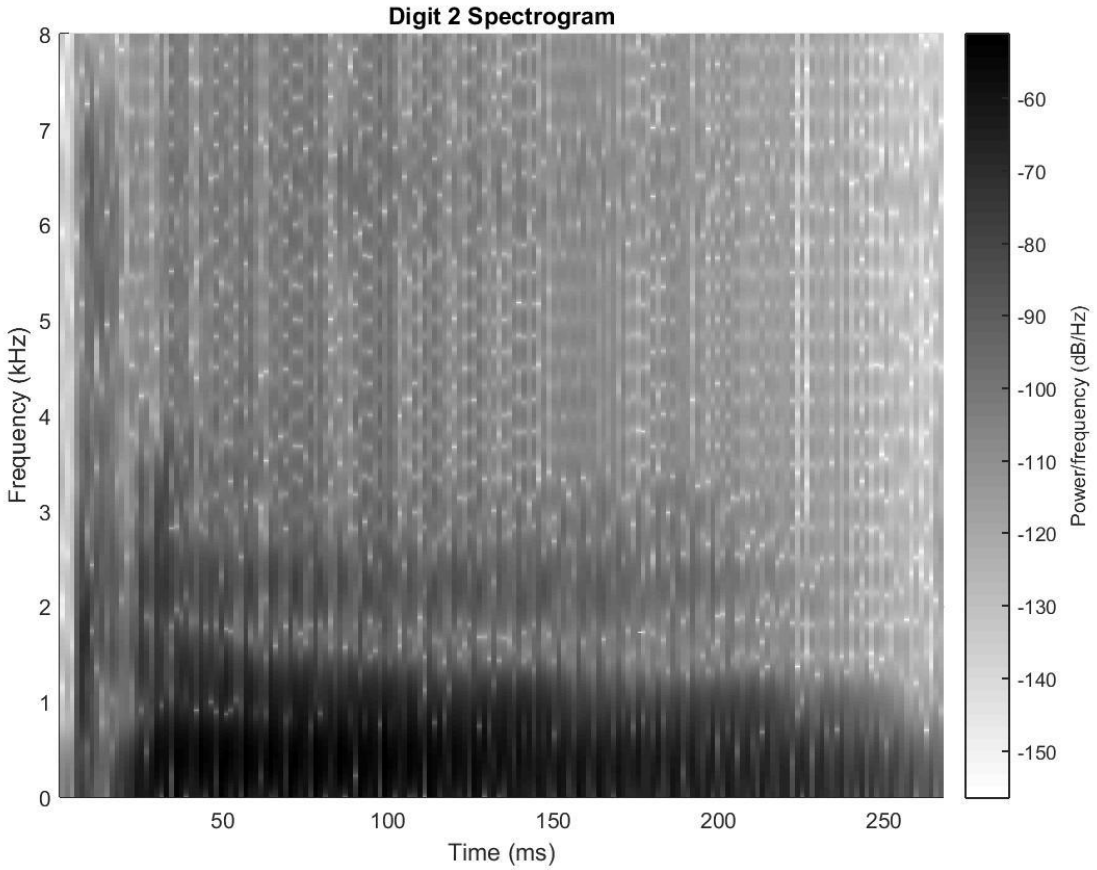


Phonetic Transcription of One: wʌn

w is a semi-vowel, which lies in 0-200ms. ʌ is a vowel. Fricative frequencies are strong in the range 200-250ms. n is a nasal sound, which lies in 250 onward.

Two

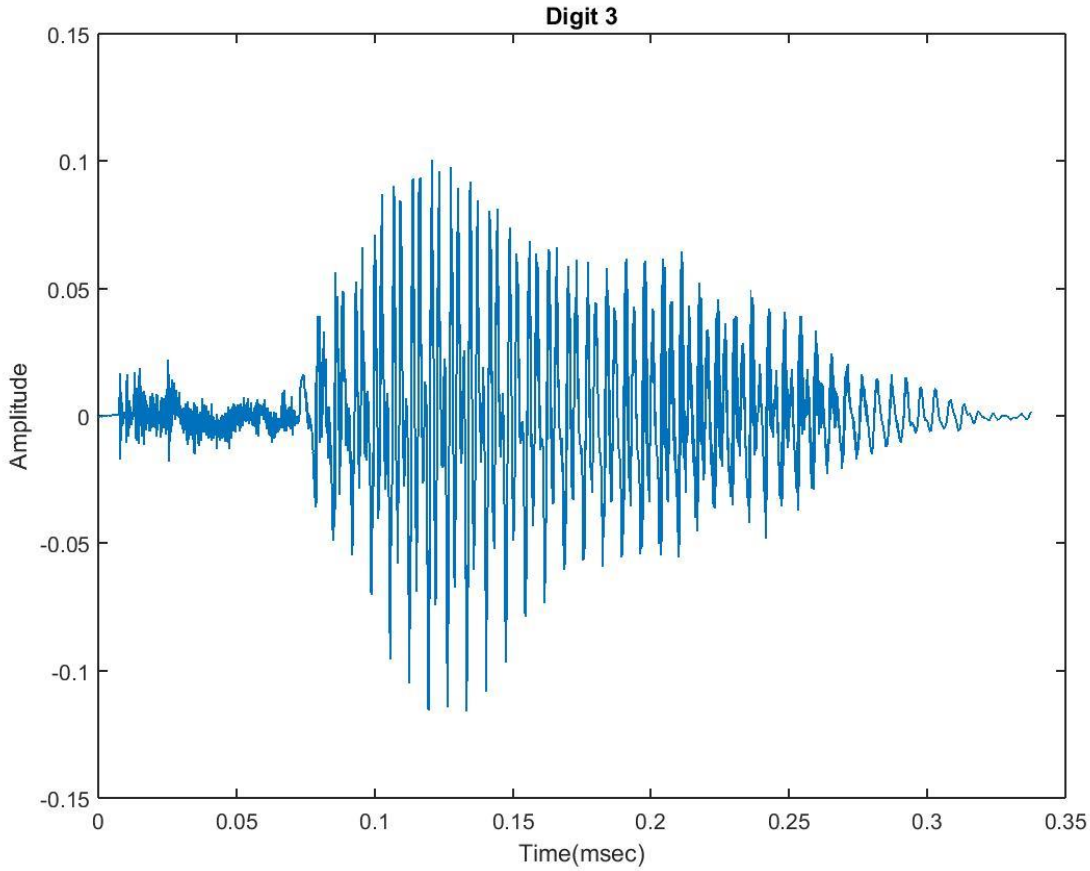

Digit 2
0.15
0.1
0.05
0
-0.05
-0.1
-0.15
Amplitude
0
0.05
0.1
0.15
0.2
0.25
0.3
Time(msec)

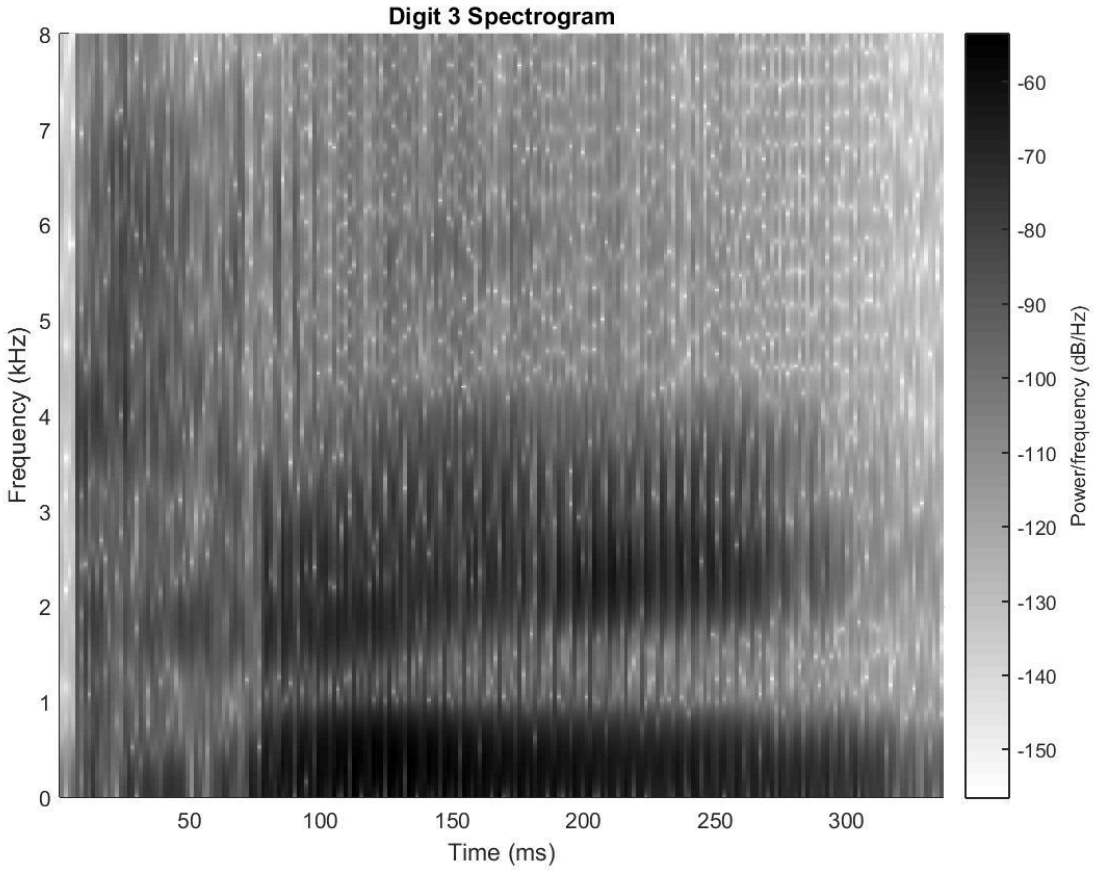


Phonetic Transcription of Two: tu

The t is a unvoiced stop and u is a vowel. In 0-30ms we have distributed energy. So t lies from 0-30ms and u is 30-260ms.

Three

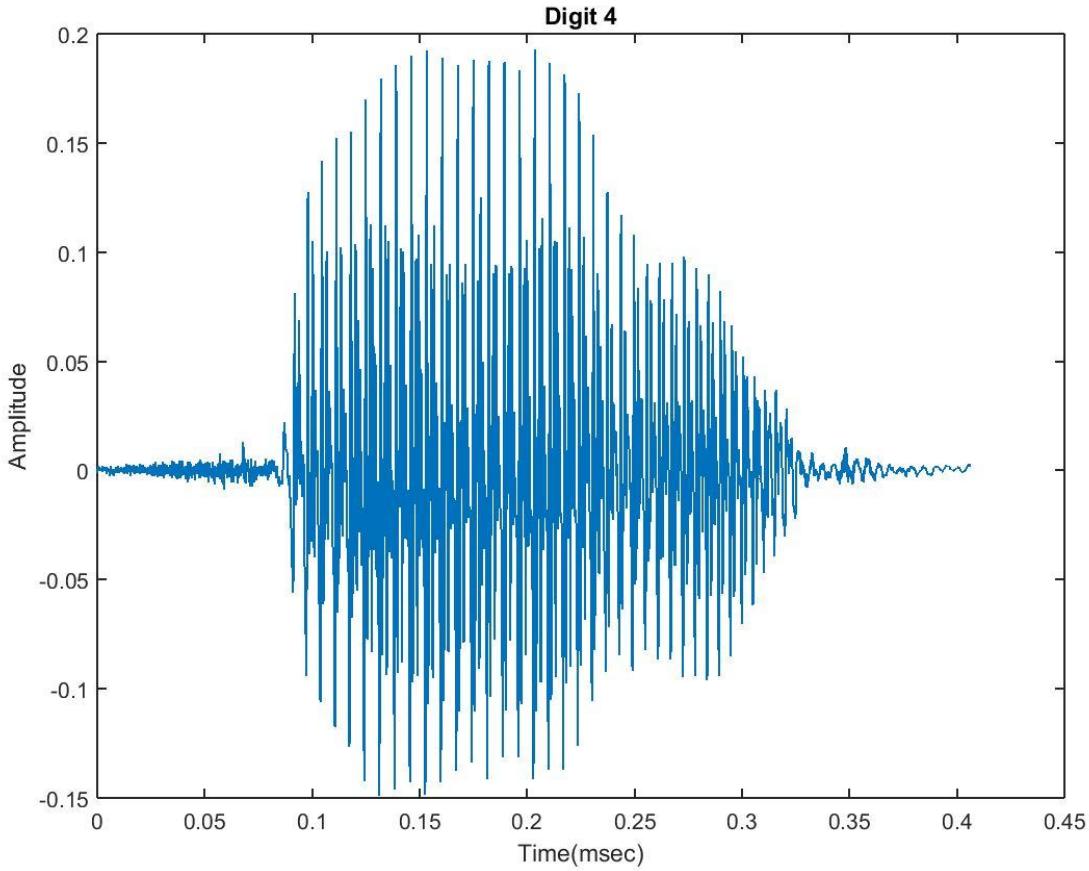

Digit 3
0.15
0.1
0.05
0
-0.05
-0.1
-0.15
Amplitude
0
0.05
0.1
0.15
0.2
0.25
0.3
0.35
Time(msec)

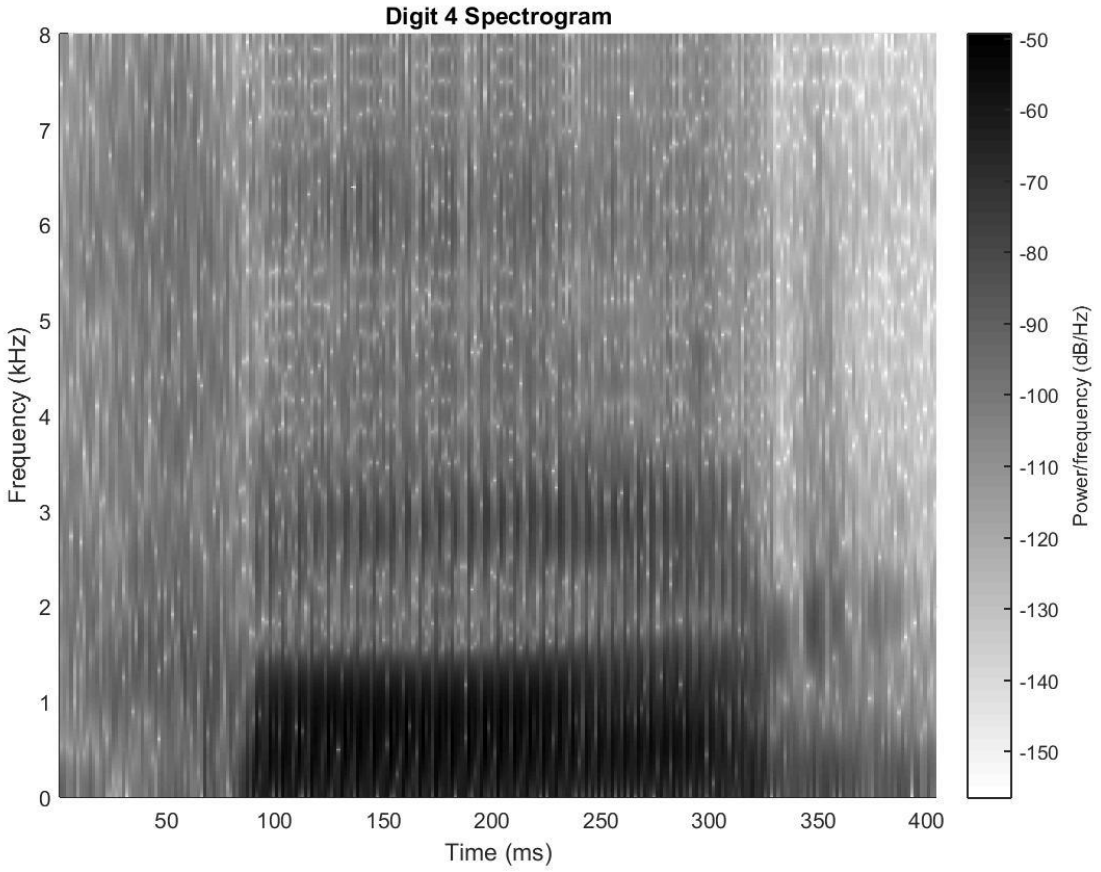


Phonetic Transcription of Three: θri

In this, θ is a fricative, r is a semi-vowel and i is a vowel. In 0-75ms energy is distributed, which is θ. r and i can be separated around 200ms with the help of the formant variation.

Four

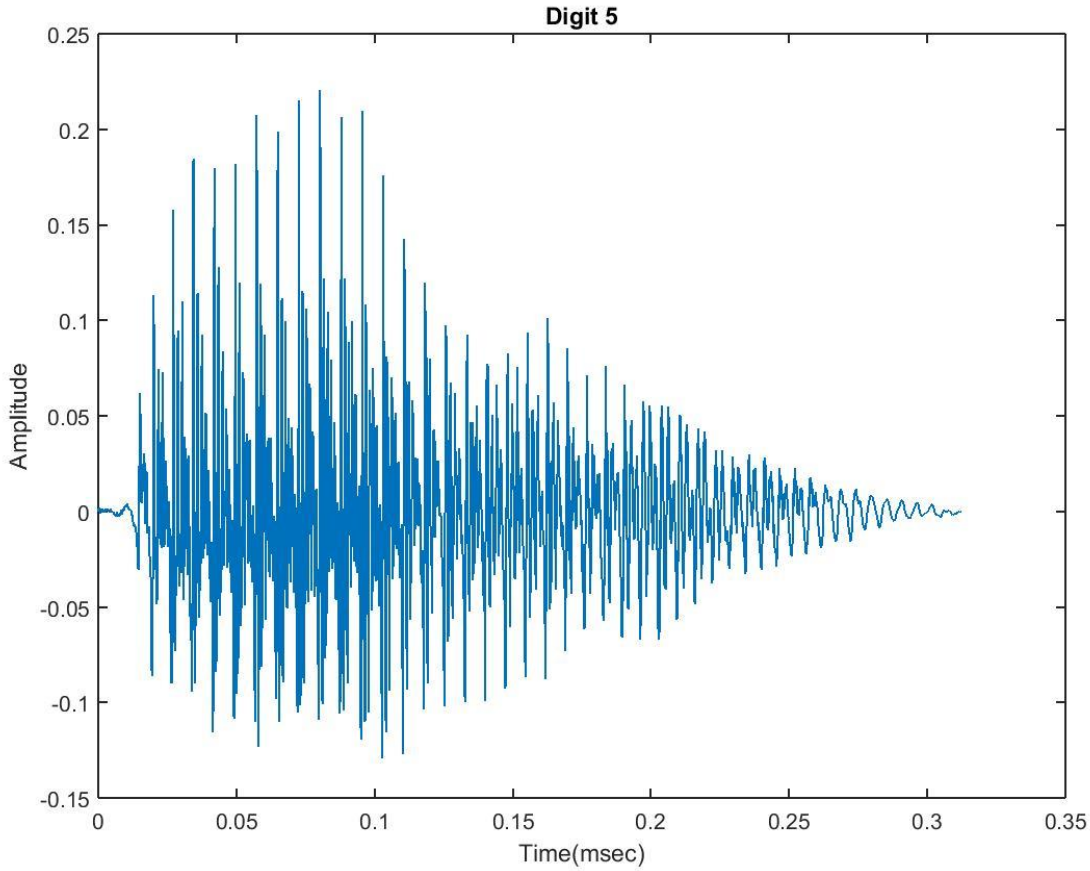

Digit 4
Amplitude
Time(msec)
0.2
0.15
0.1
0.05
0
-0.05
-0.1
-0.15
0
0.05
0.1
0.15
0.2
0.25
0.3
0.35
0.4
0.45

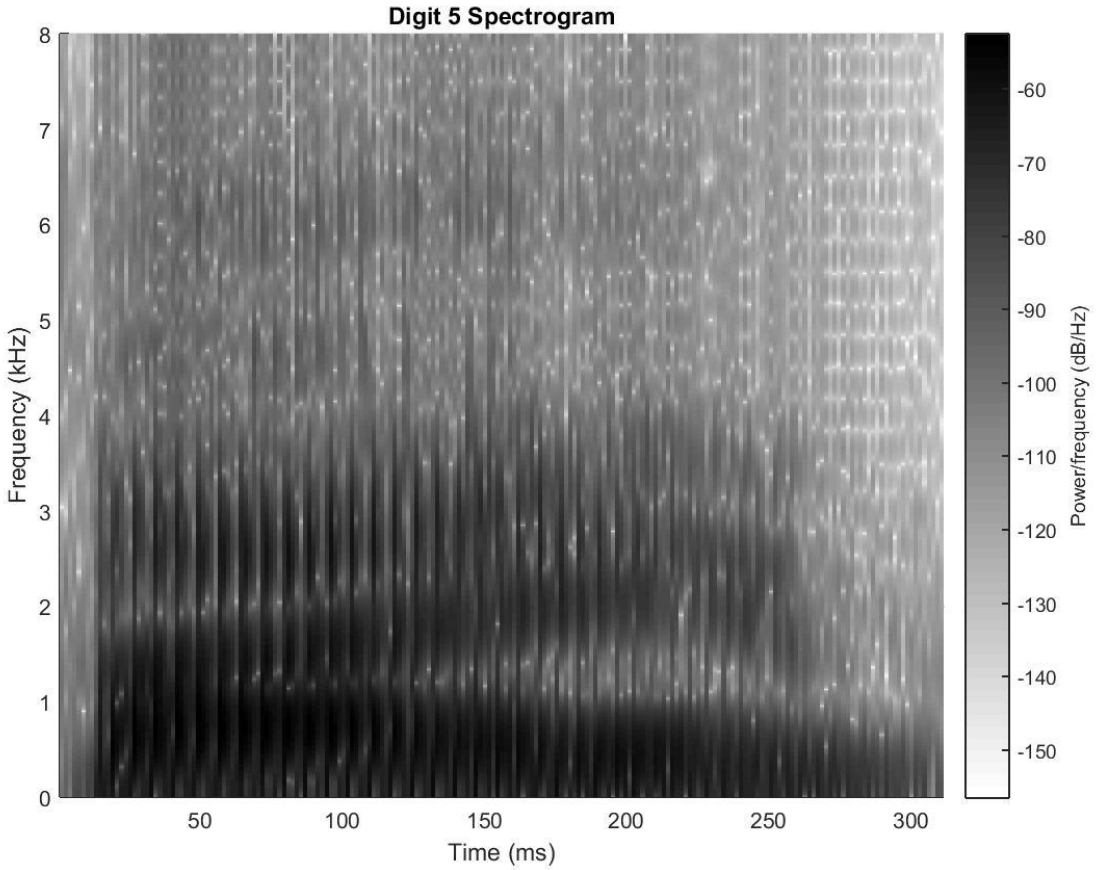


Phonetic Transcription of Four: fɔr

In this, f is a unvoiced fricative, ɔ is mid vowel and r is a semi-vowel. As there is no sharp formants from 0-80ms, which is f. ɔ and r can be separated around 300ms.

Five

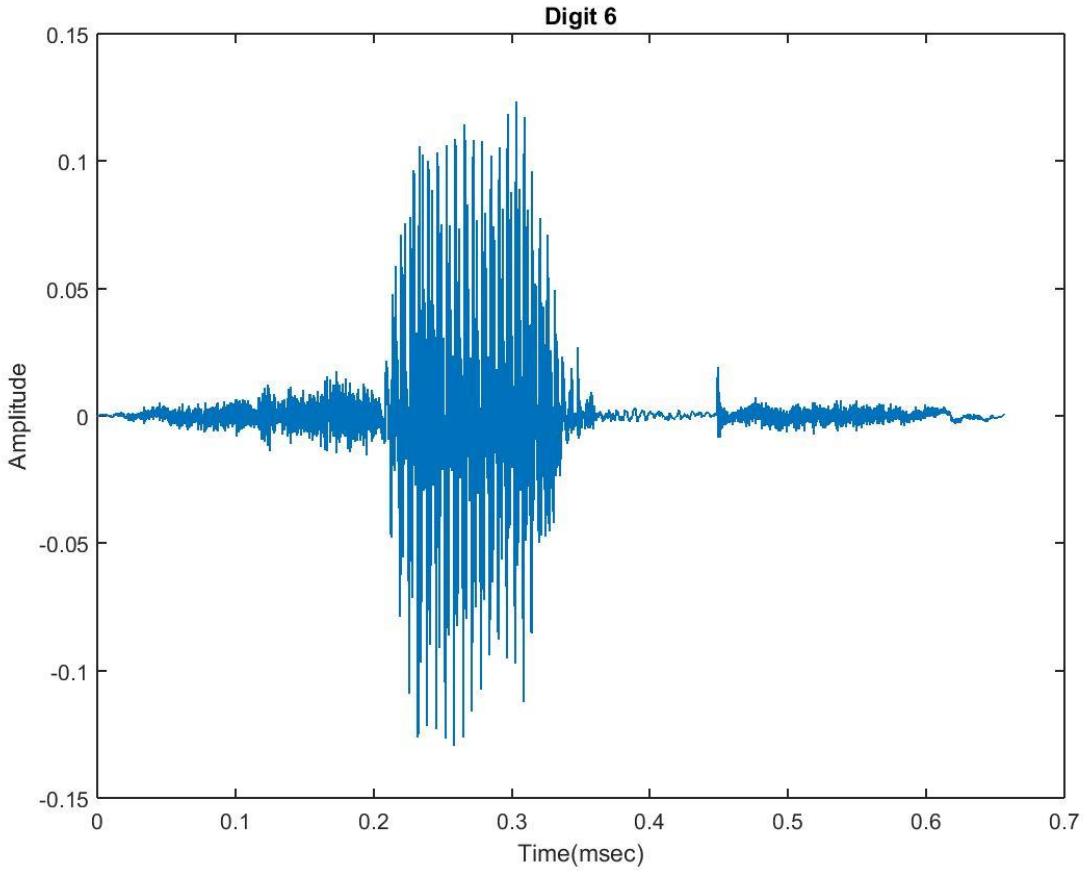

Digit 5
Amplitude
Time(msec)
0.25
0.2
0.15
0.1
0.05
0
-0.05
-0.1
-0.15
0
0.05
0.1
0.15
0.2
0.25
0.3
0.35

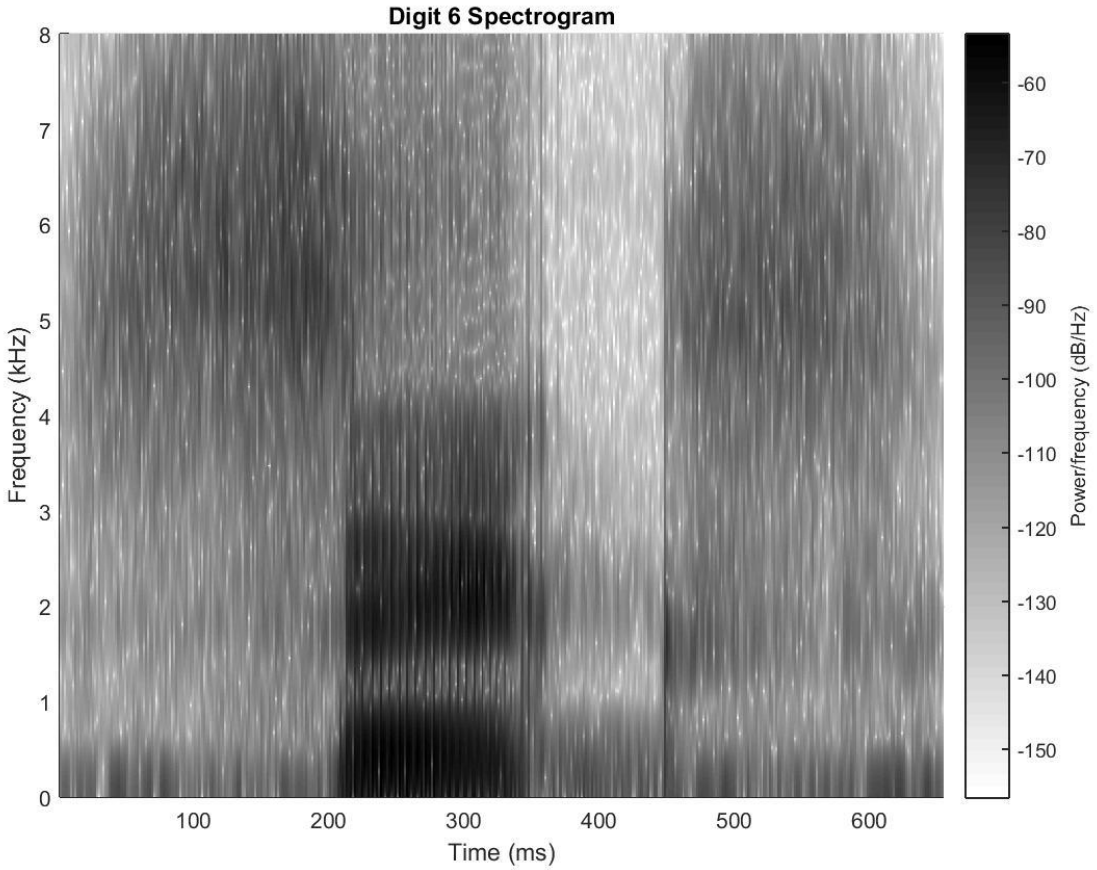


Phonetic Transcription of Five: faɪv

In this, f is 0-10ms. a is from 10-100ms. ɪ is around 100-200ms and v is 200ms onward.

Six

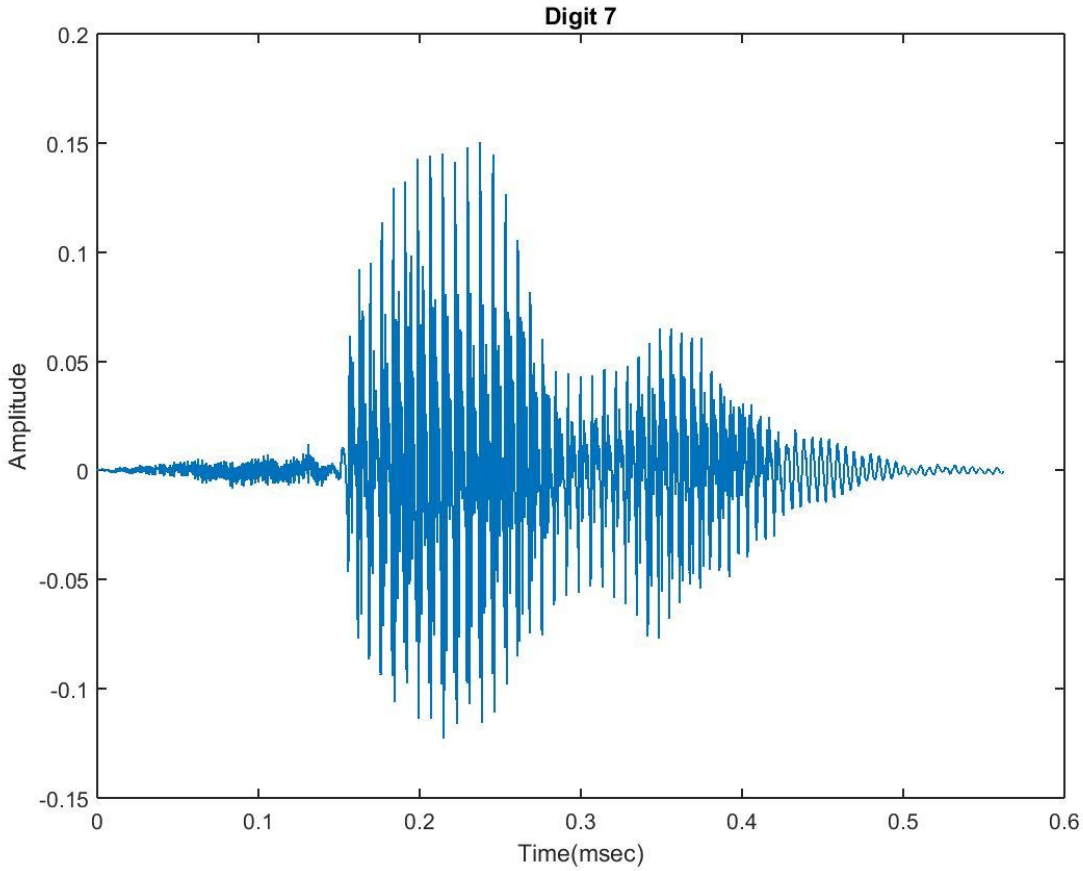

Digit 6
0.15
0.1
0.05
0
-0.05
-0.1
-0.15
Amplitude
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
Time(msec)

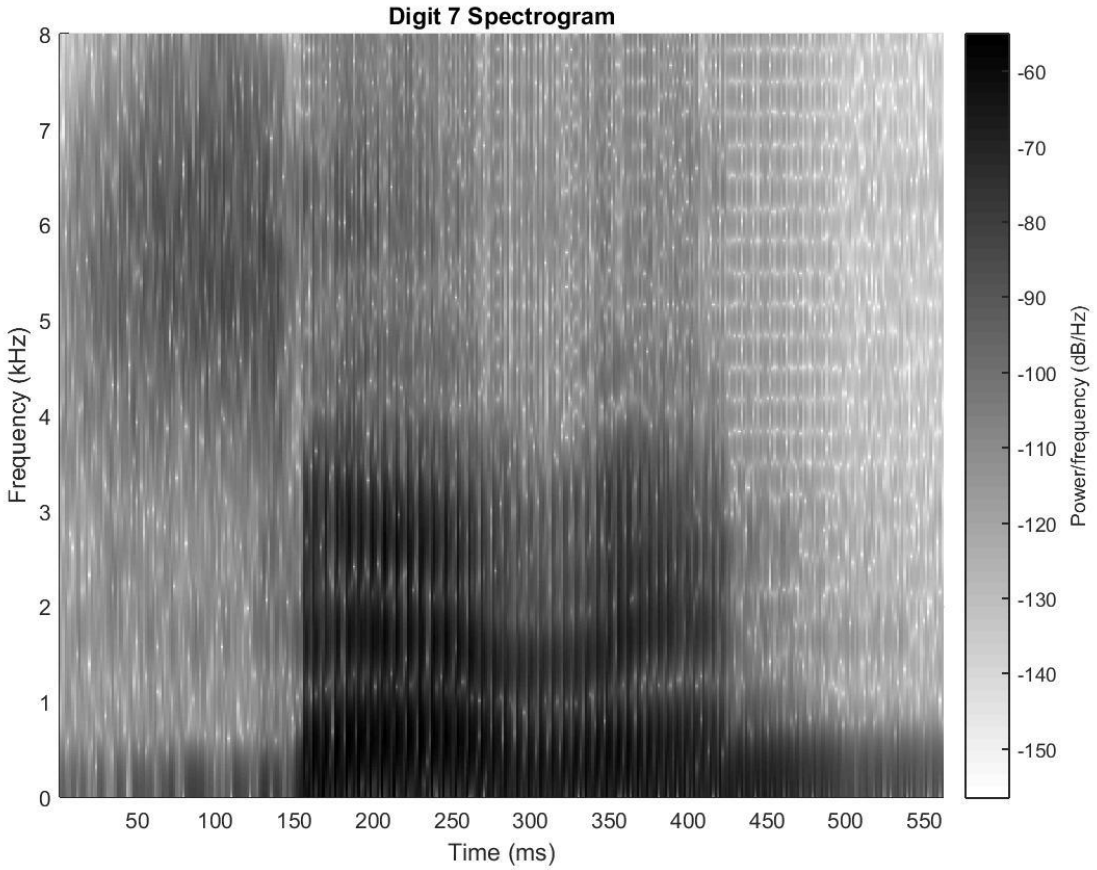


Phonetic Transcription of Six: sɪks

In this, s is from 0-200ms, ɪ is 200-380ms, k is till 450ms and s is 450ms onward.

Seven

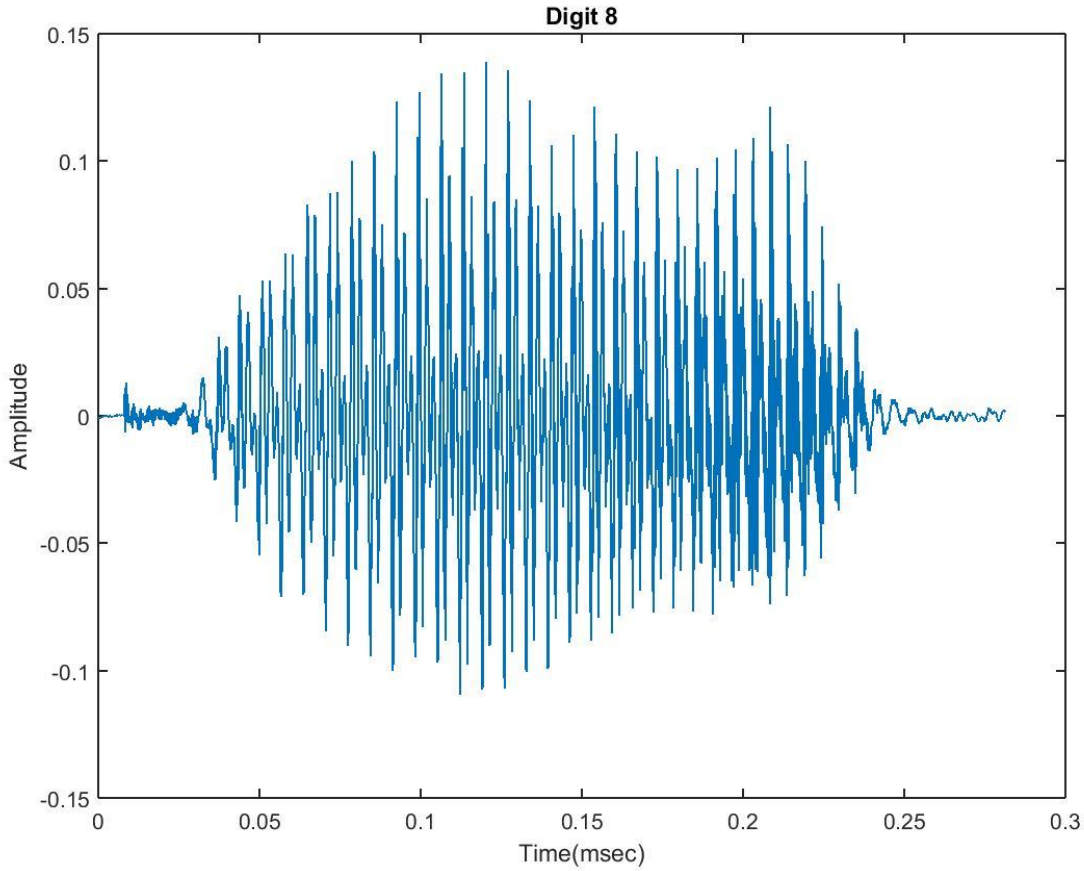

Digit 7
Amplitude
Time(msec)
0.2
0.15
0.1
0.05
0
-0.05
-0.1
-0.15
0
0.1
0.2
0.3
0.4
0.5
0.6

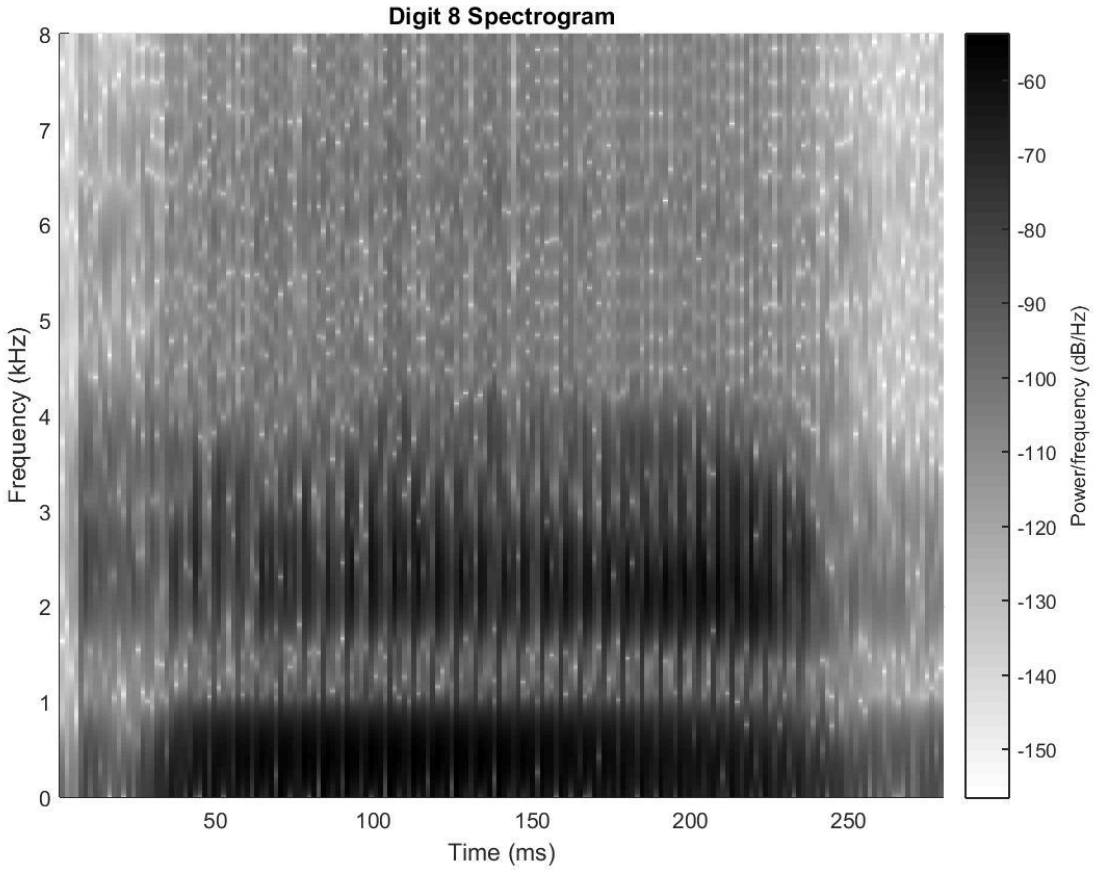


Phonetic Transcription of Seven: ˈsɛvən

In this, s is 0-160ms, ɛ is 150-250ms, v is 250-350ms, ə is 350-400ms and n is 400ms onward.

Eight

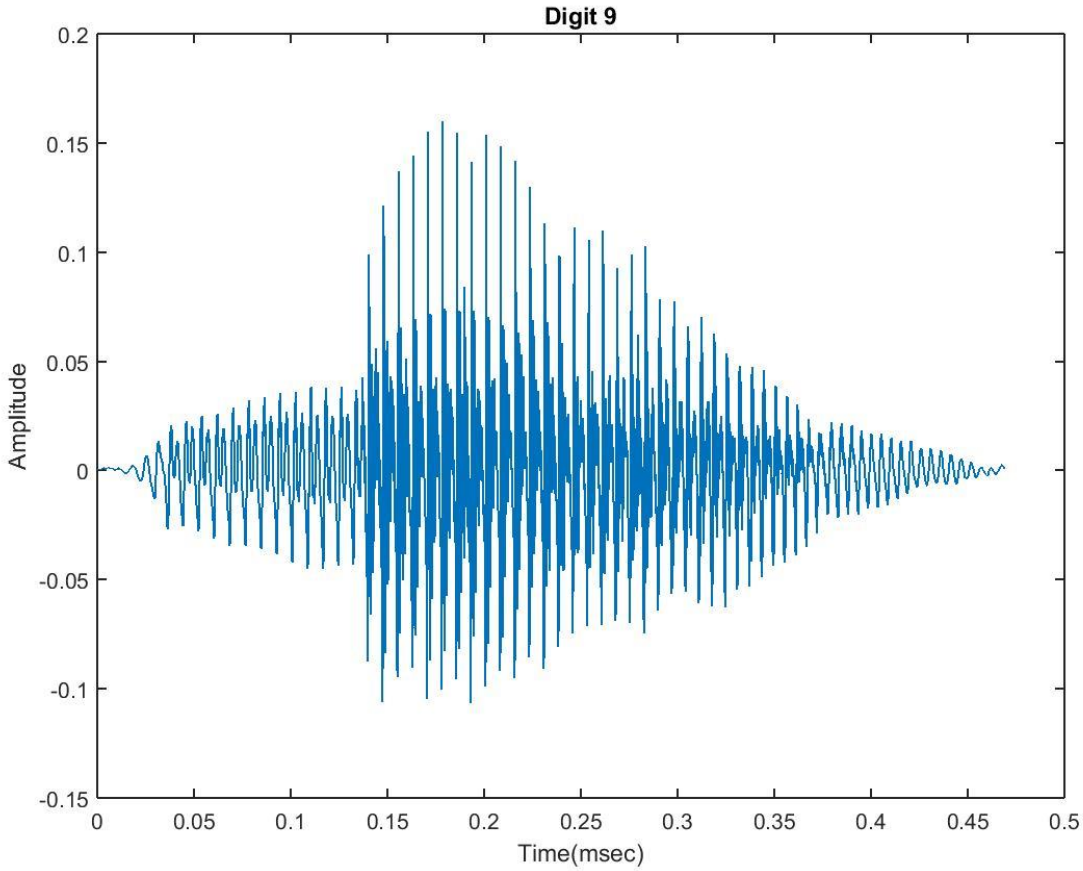

Digit 8
Amplitude
Time(msec)
0.15
0.1
0.05
0
-0.05
-0.1
-0.15
0
0.05
0.1
0.15
0.2
0.25
0.3

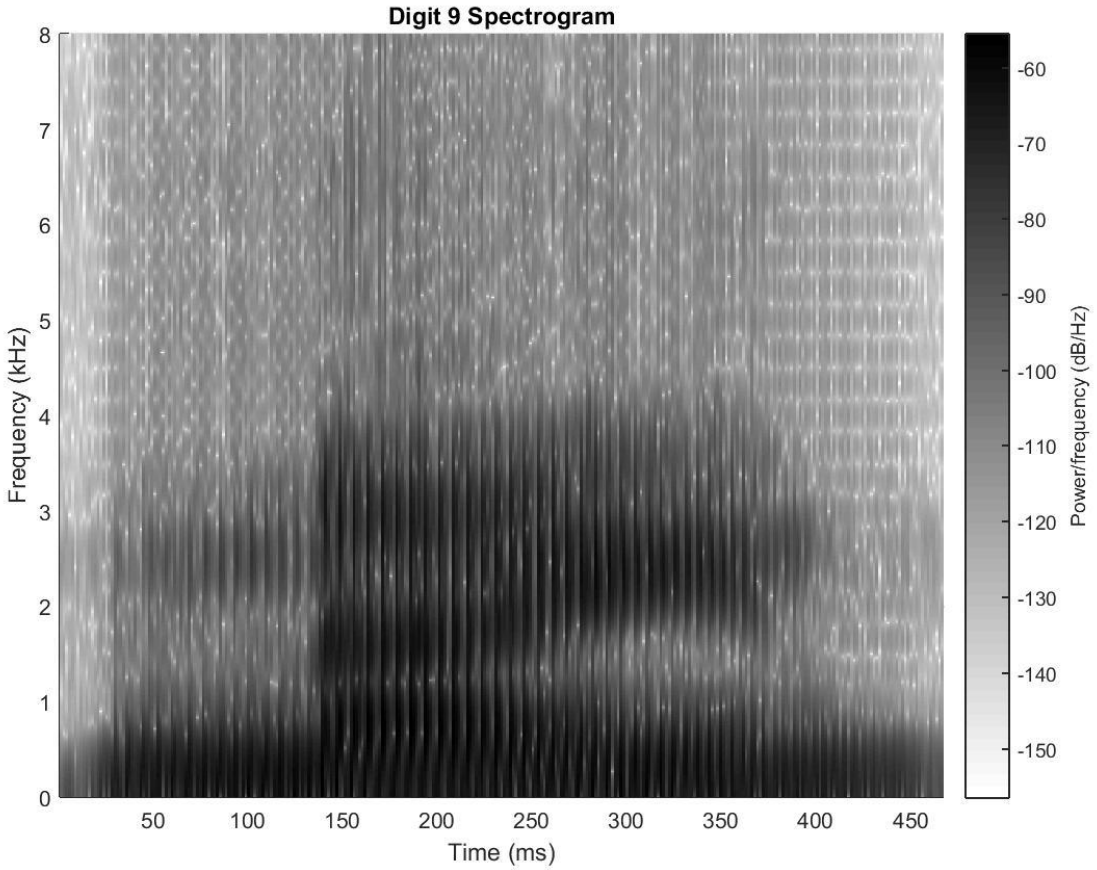


Phonetic Transcription of Eight: eɪt

In this, e and ɪ are both vowels. So its hard to exactly find separate boundary. But t can be easily identified, it is 240ms onward.

Nine

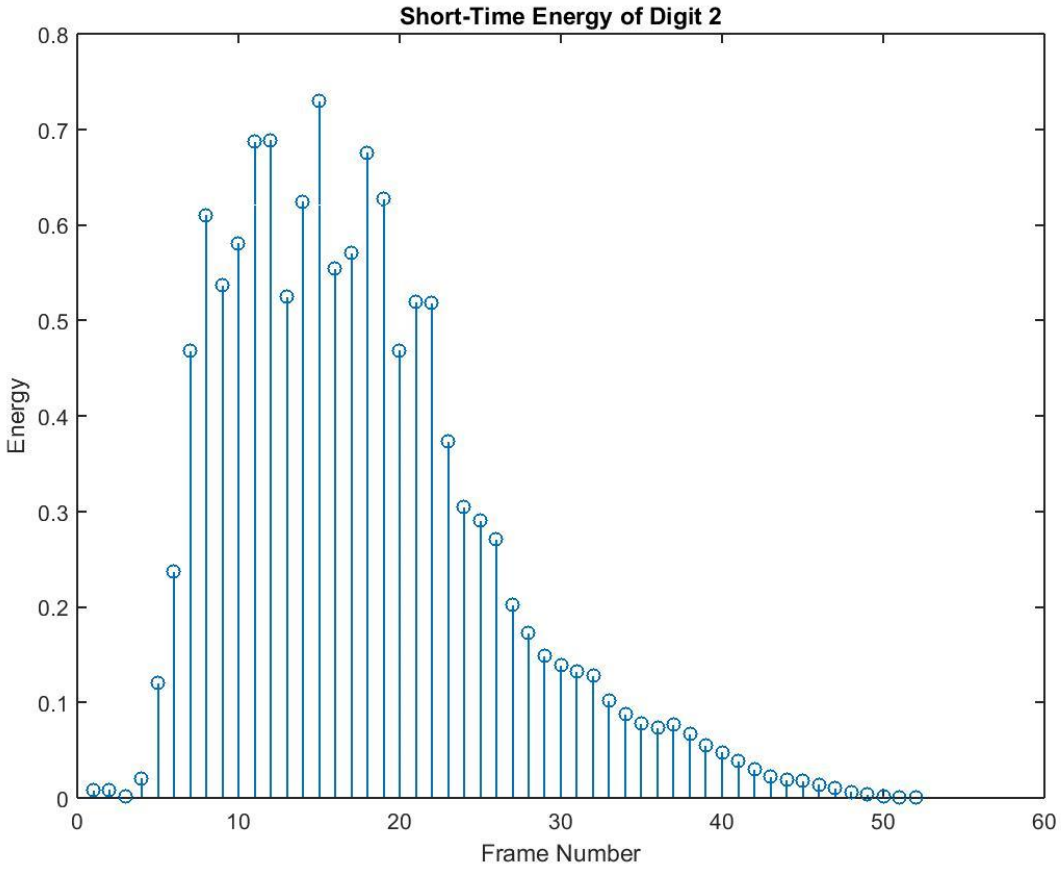

Digit 9
Amplitude
Time(msec)
0.2
0.15
0.1
0.05
0
-0.05
-0.1
-0.15
0
0.05
0.1
0.15
0.2
0.25
0.3
0.35
0.4
0.45
0.5



Phonetic Transcription of Nine: naɪn

In this, n is from 0-150ms, On the basis of formant variation around 270ms, we can separate a and ɪ. n is around 350ms onward.

c)



The voiced region have higher energy than unvoiced region. If we set a threshold of about 0.1 short-time energy, then we can say Frame 1 to 4 and 33 to 55 corresponds to the unvoiced region. Whereas, Frame 5 to 32 corresponds to the voiced region.

d) Algo for unknown digit identification

We can identify the location of various vowels, fricative and consonant on the basis of energy distribution and formants variation and then we can match with the templates of various digits.