|  |  |  |  |
| --- | --- | --- | --- |
| **FILTER NUMBER** | **CORRESPONDING FREQUENCY(Hz)** | **FILTER NUMBER** | **CORRESPONDING FREQUENCY(Hz)** |
| 01 | 300 | 15 | 2370.1 |
| 02 | 383.42 | 16 | 2626.3 |
| 03 | 473.79 | 17 | 2903.03 |
| 04 | 571.71 | 18 | 3204.4 |
| 05 | 677.80 | 19 | 3530.1 |
| 06 | 792.74 | 20 | 3882.9 |
| 07 | 917.27 | 21 | 4265.2 |
| 08 | 1052.2 | 22 | 4679.4 |
| 09 | 1198.3 | 23 | 5128.2 |
| 10 | 1356.7 | 24 | 5614.4 |
| 11 | 1528.3 | 25 | 6141.1 |
| 12 | 1714.2 | 26 | 6711.8 |
| 13 | 1915.6 | 27 | 7330.1 |
| 14 | 2133.7 | 28 | 8000 |

VOWELS PATTERNS:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| A | E | I | O | U |
| a\_e | e\_e | i\_e | o\_e | u\_e |
| ai | ee | igh | oa | ue |
| eigh/ei | ea/ie | -y | ow/ou | ew/eu |

**PHONEMES AND THEIR CORRESPONDING FREQUENCIES:**

|  |  |  |  |
| --- | --- | --- | --- |
| **SL.NO** | **NUMBERS** | **PHONEMES** | **FREQUENCY RANGE(Hz)** |
| 01 | ZERO | ZIY  RO | 4200-6500(22,23,24,25)  200-1500(1-10) |
| 02 | ONE | W AE  N | 200-1200 &4000-5600 (1-9)&(18,19,22,23)  600-1600(9,10,11) |
| 03 | TWO | T  UW | 400-1200(1,6,7)  300-600(1,5) |
| 04 | THREE | TH  R IY | 1,12,13,21,22  1700-2800(12,13,14) |
| 05 | FOUR | F  OW  R | 11,12,13,15  100-1000(1-5)  500-1200(6,7) |
| 06 | FIVE | F  AY  V | 11,12,13,15  450-1300(3-9)  12,13 |
| 07 | SIX | S IH  K  S | 4500-6000(22,23,24)  2000(12,13)  4000-7000(21,22,23,24,25) |
| 08 | SEVEN | S  EH  V  EH  N | 4500-6000(22,23,24)  500&1500(2,3 & 11)  11,12  500&1500(2,3 & 11)  500-1500(9,10) |
| 09 | EIGHT | EY  T | 2000-2500(14,15,16)  16,17,18,19 |
| 10 | NINE | N  AY N | 600-1600(9,10,11)  13,14,15 |

MFCC

1. PRE EMPHASIS:

 Pre-emphasis refers to a system process designed to increase the magnitude of some higher frequencies with respect to the magnitude and in the lower frequencies in order to improve the overall signal-to-noise ratio by minimizing the adverse effects of such phenomena as attenuation distortion or saturation of recording media in subsequent parts of the system.

**Y[n]= X[n]-0.95X[n-1]**

The Z transform of this equation is given by:

**H(z)= 1 – (0.95/(z-1))**

Framing and Windowing

An audio signal is constantly changing, we assume that on short time scales the audio signal doesn't change much i.e. statistically stationary, the samples are constantly changing on even short time scales. so we frame the signal into 25ms frames and 10ms of overlapping.

Windowing is done using Hamming window

Each frame is multiplied with the window to make the continuous in the frame from the first to the last point. If the

signal in a frame is denoted by

Y [n], n = 0,…N-1

Then the signal after Hamming windowing is,

Y [n] × w[n]

Where w [n] is the Hamming window defined by

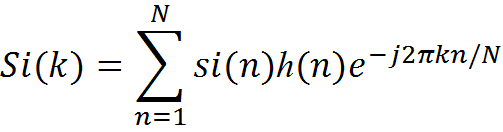


W[n]=

Hm (K)=

M(*f)*=1125 ln(1+)

M(*f)*=1125 ln(1+)



Si(k)= 1

Where K is the length of DFT

Pi(*k*)=

|  |  |  |
| --- | --- | --- |
| SL  NO | DIGIT | PHONEMES |
| 01 | ZERO | ZE\_IY\_RO |
| 02 | ONE | WAN |
| 03 | TWO | TO |
| 04 | THREE | TH\_EY\_IY |
| 05 | FOUR | F\_TO/OW |
| 06 | FIVE | F\_FA\_IVE |
| 07 | SIX | SIK\_K\_SIL\_S/SIH |
| 08 | SEVEN | SIH/S\_EHV\_(N) |
| 09 | EIGHT | EY\_SIL |
| 10 | NINE | N\_IVE\_EY/IY |