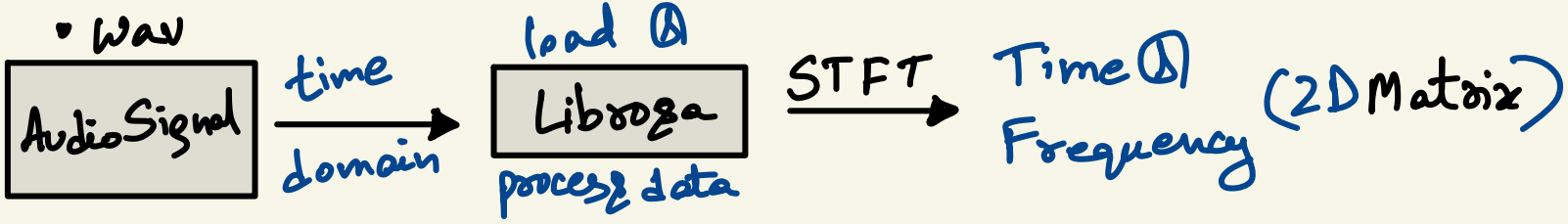
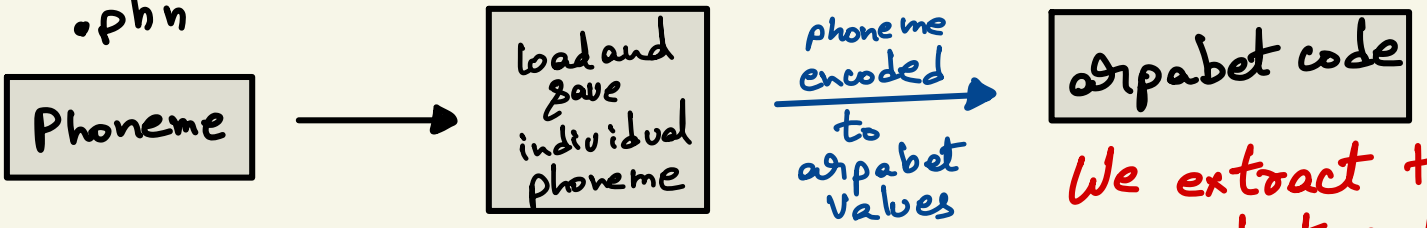




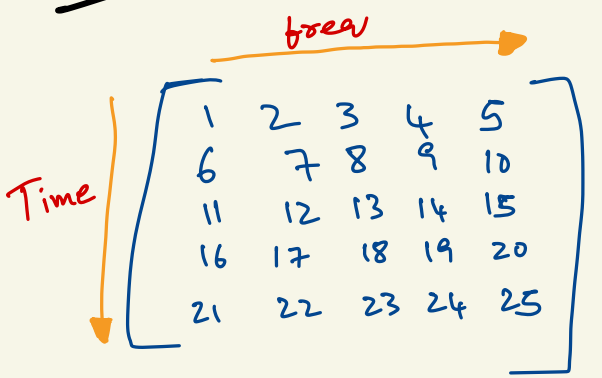
We extract the arpabet code

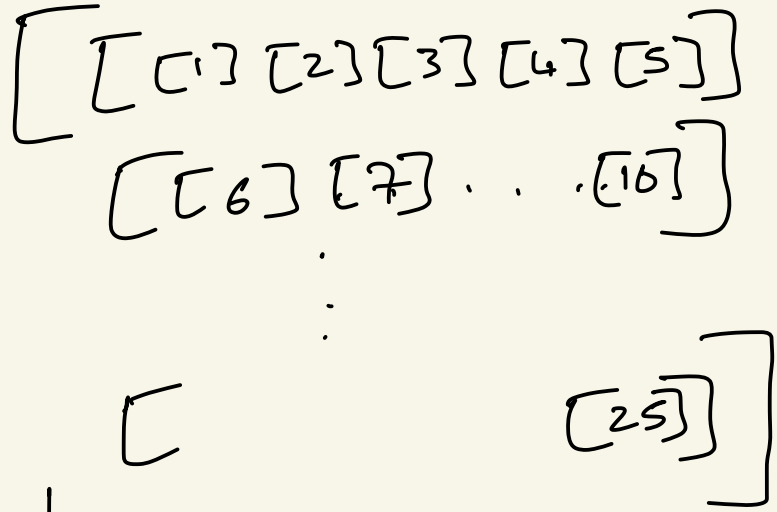
For all the .wav files 2D matrix stored we calculate length of the column & pad Zero's

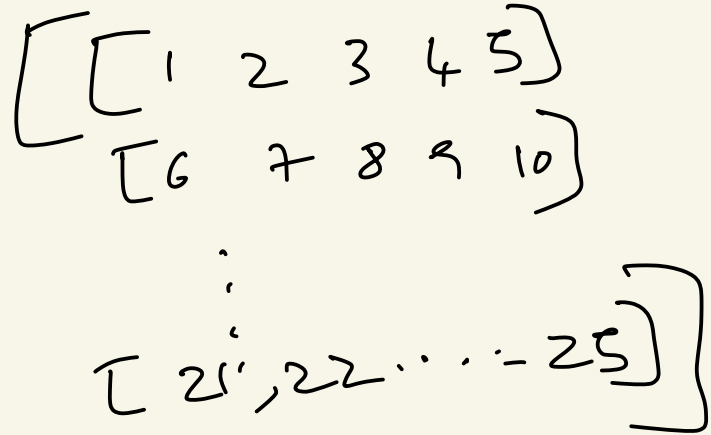
Since every row has atleast one column

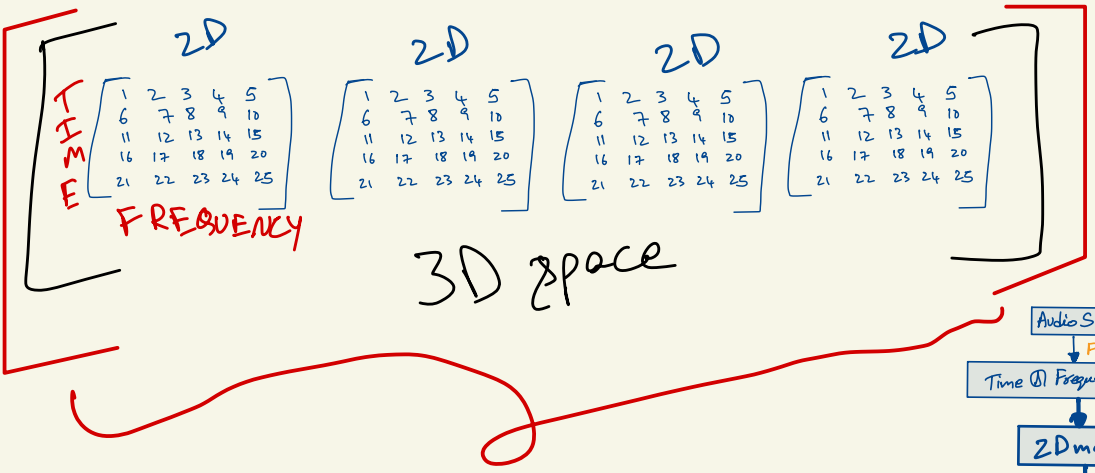
The first column is the longest one

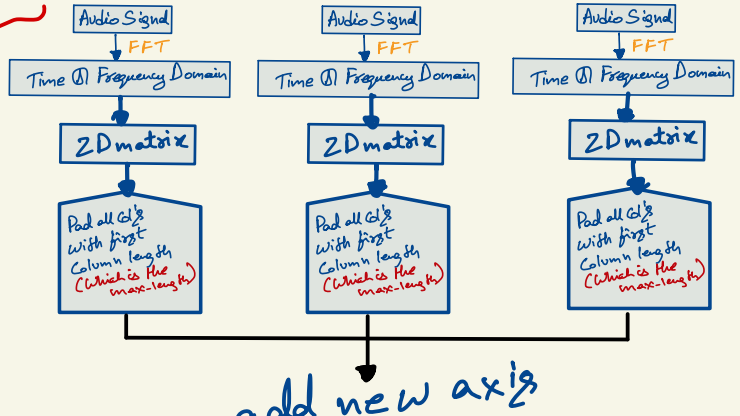
2D matrix :-

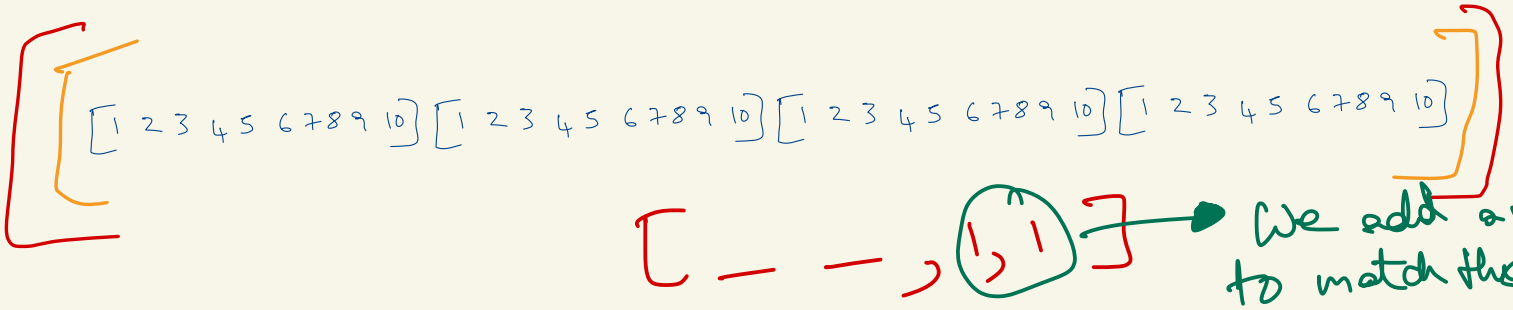

freq
Time
1 2 3 4 5
6 7 8 9 10
11 12 13 14 15
16 17 18 19 20
21 22 23 24 25


[1] [2] [3] [4] [5]
[6] [7] . . . [10]
[25]


1 2 3 4 5
6 7 8 9 10
21, 22 . . . 25


2D
2D
2D
2D
TIME
FREQUENCY
3D space
4D space


Audio Signal
FFT
Time & Frequency Domain
2D matrix
Pad all 0's with first Column length (which is the max-length)
add new axis

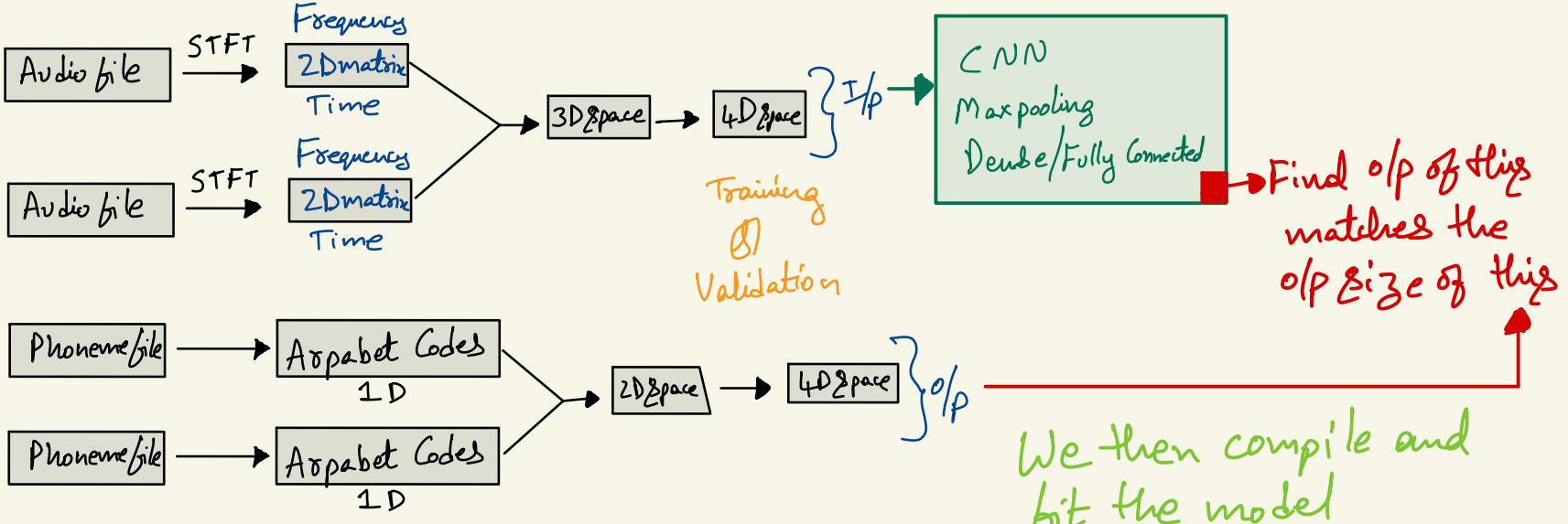

[1 2 3 4 5 6 7 8 9 10] [1 2 3 4 5 6 7 8 9 10] [1 2 3 4 5 6 7 8 9 10] [1 2 3 4 5 6 7 8 9 10]
[_ _ , 1, 1]
We add axis to match the 4D space

Audio files with amplitude

2D matrix

padding-sequence to match the column sizes

Data flow :-



We then compile and fit the model

Compile → Optimizer (Updating weights)
Loss function (check loss)
Metrics (like Accuracy)