Speaker Verification

Members:
Zeeshan Ali (141623)

Muhammad Khalid (141636)

Overview

- The speaker verification is the task of verifying the claimed identity of a speaker by using their voice characteristics as captured by a recording device such as a microphone.

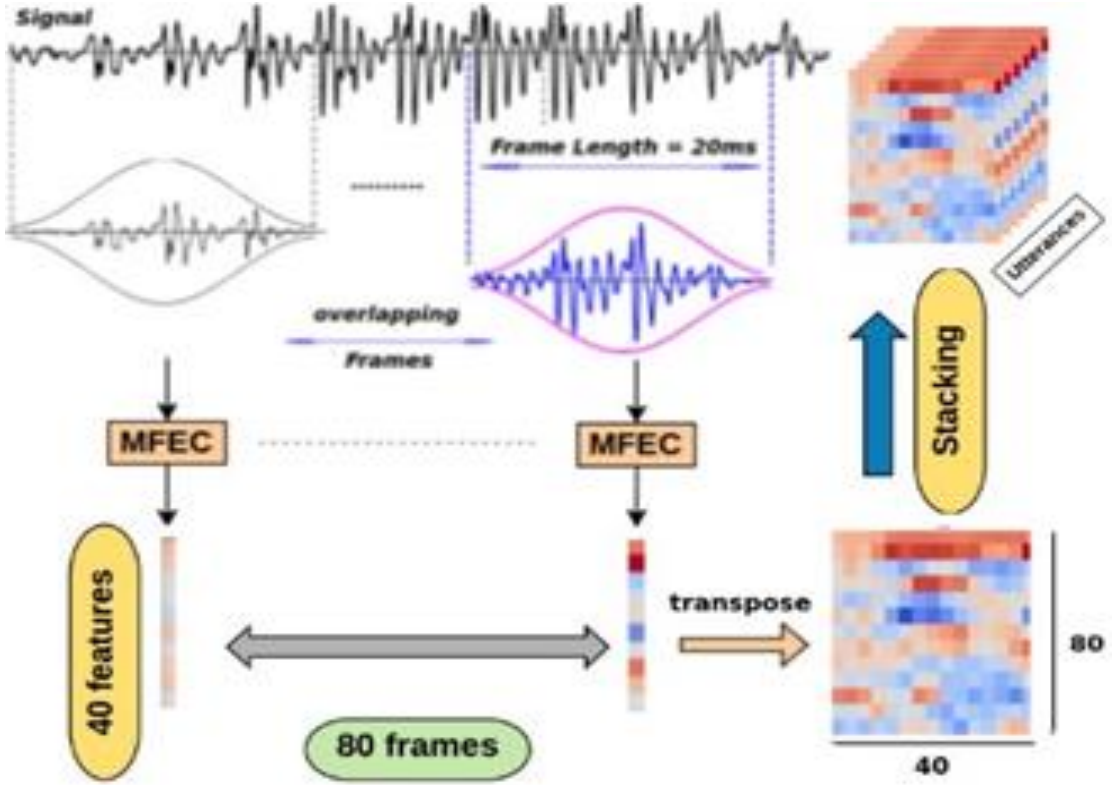
- We used CNN model to extract speaker representation as a features of a particular speaker.

➢ Data Set

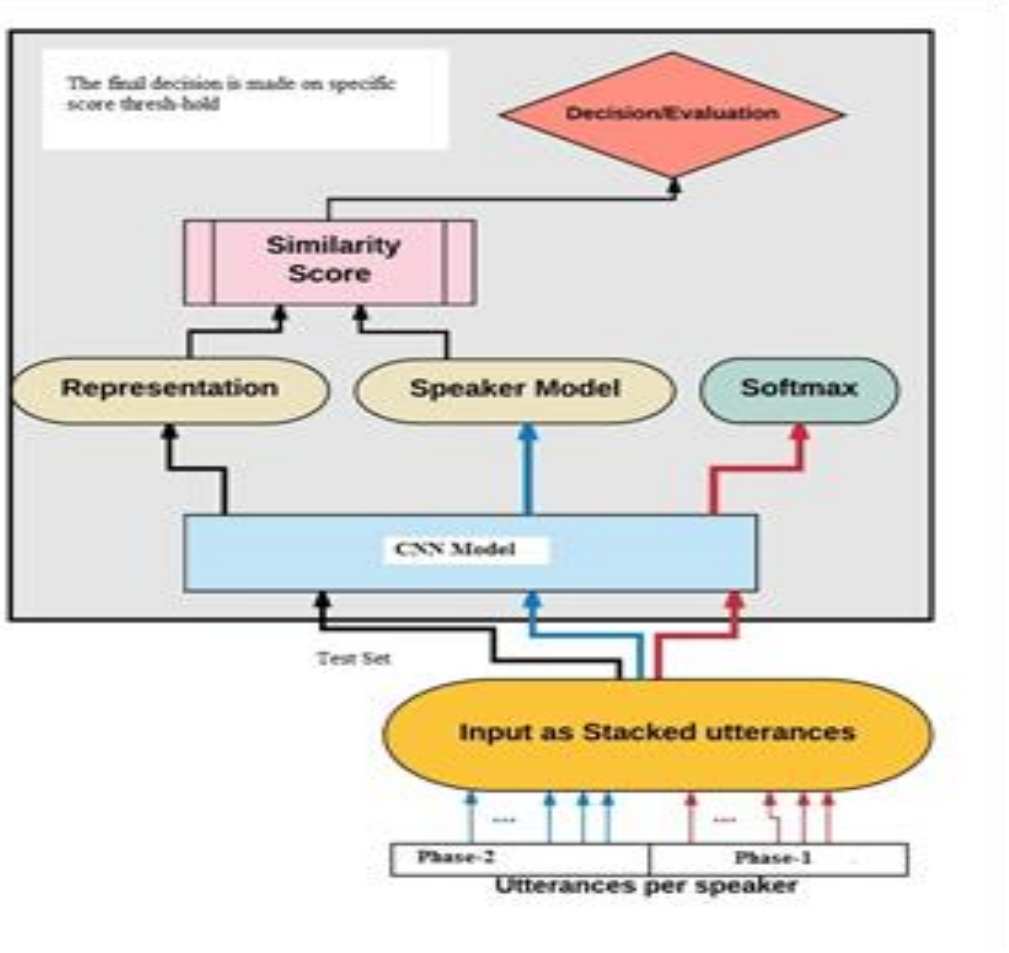
- We used KING and NKING Data which contains 51 Males' Speaker's recorded voice and each speaker's audio files are consist of 10 sessions. After getting mfcc features maps, each map is overlapped several time to get better results at utterance level. We split the data into training and testing sets with test ratio 30%.

Features generation for Model's inputs

- 0.8 seconds audio signal is used for 80 frames and 40 mfcc's coefficients



General Model Architecture



Details of Layers

Layer	Input	Output	Kernel	Stride
Conv1	80x40x20	80x40x16	16x(3x3)	1
BN, DO, ReLu, MP1	80x40x16	40x20x16	-, 0.5, -, 2x2	2
Conv2	40x20x16	40x20x32	32x(3x3)	1
BN, ReLu	40x20x32	40x20x32	-	-
Conv3	40x20x32	40x20x32	32x(3x3)	1
BN, ReLu, MP2	40x20x64	20x10x64	-, 0.5, -, 2x2	2
Conv4	20x10x64	20x10x128	128x(3x3)	1
BN, ReLu	20x10x128	20x10x128	-	-
Conv5	20x10x128	20x10x128	128x(3x3)	1
BN, ReLu, MP3	20x10x128	10x5x128	-, 0.5, -, 2x2	2
Conv6	10x5x256	10x5x256	256x(3x3)	1
BN, ReLu	10x5x256	10x5x256	-	-
Conv7	10x5x256	10x5x256	256x(3x3)	1
BN, ReLu, MP	10x5x256	5x3x256	-, 0.5, -, 2x2	2
Conv10	5x3x256	1x1x256	256(5x3)	1
FC-1	1x1x256	1x512		
FC-2	1x512	1x1024		
FC-3	1x1024	1x51		

Table-1 CNN Architecture

Results

# Utterances	Model	Accuracy
1	2D-CNN-1	48.0%
20	2D-CNN-1	27.0%
20	2D-CNN-2	60.0%
20	3D-CNN	45.0%

Conclusions

- We investigated CNN model for extracting speaker's feature that can be used for speaker verification as well as for voice recognition. We obtained better results than the previous old methods of speaker verification.
- This speaker verification model can be used for biometric authentication like application as a future work.

Questions ?