DEEP NEURAL NETWORK USED FOR SPEECH SEPARATION

Bhagat Anuradha Ramnath and R.S. Pawase

Department of Electronics and Telecommunication Engineering, Amrutvahini College of Engineering, India

Abstract

Within the proposed system Deep Neural Network (DNN) is employed to get the speech features of target speaker and interfere for speech separation. This paper focuses on separating the target speech signal from the inputs during this system a regression approaches via deep neural network (DNN) for unsupervised speech separation during a single channel setting. This technique is believe a key assumption that two speakers might be well segregated if they're not too almost like one another. To demonstrate that the space between speakers of various genders is large enough to require possible separation. Then proposed DNN architecture having two outputs, from that one representing the feminine speaker group and another one is male speaker group. Finally the trained and tested DNN dataset performs the speech separation of the target speech.

Keywords:

Deep Neural Network, Regression Model, Noise Reduction, Speech Separation, Gender Mixture Detector, Speech Separator

1. INTRODUCTION

In the recent speech separation challenges the task of speech separation is extremely important in sort of applications like Automatic Speech Recognition (ASR).The speech separation is aims to separate the speech signal from the target signal and also separate the speech signal from noise, during this system formulate the matter of two mixing speakers as,

$$X_m = X_t + X_i \quad (1)$$

where,

X_m = mixed speech signal

X_t = target speech

X_i = interfering speakers

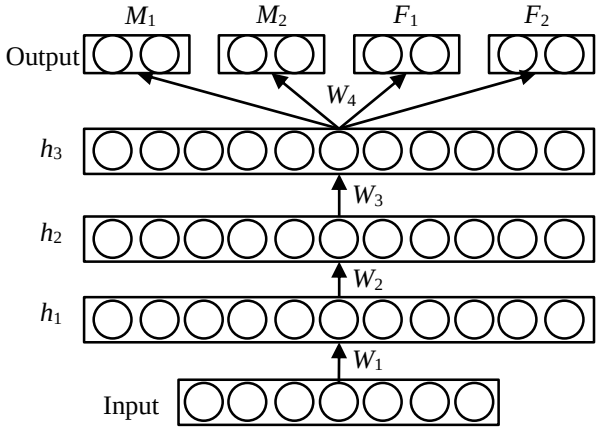
Number of various approaches are proposed in literature under various assumptions. During this study the DNN based regression model is employed to find out the complex mapping function from noise and to urge clean speech. Nonlinear DNN based regression models using multi-condition training data of various key factors within the noisy speech which incorporates speakers, noise types like female or male and signal to noise ratios (SNRS).

The DNN is employed to model the highly nonlinear mapping relationship from mixed speech to the target signal and interfering signal or noisy signals during a supervised or semi supervised mode. Within the supervised mode, we all know both the target and interfering speakers; While within the semi supervised mode, the target speaker is understood and therefore interfere is assumed to be unknown [1] [3].

In this system use the DNN approach for unsupervised speech separation of two speakers where both the speakers are unknown and relates this feasibility to some speaker distance measures i.e. the larger distance between competing speakers of mixed speakers might be separated. During this system a deep neural specification with dual output, where one representing the male speaker group and another one is representing the feminine group. In the other words DNN acts as a gender separator to segregate co-channel speech effectively [2].

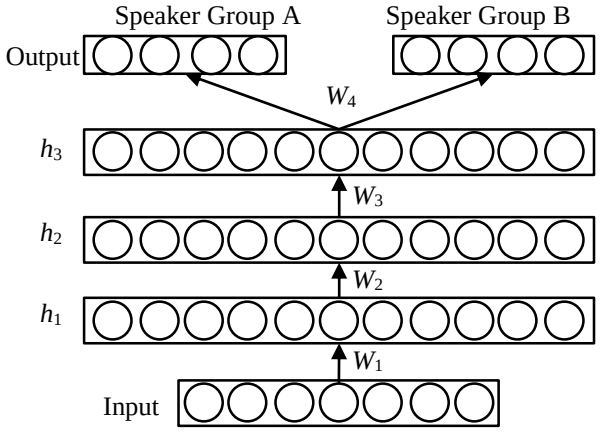
1.1 DNN BASED SPEECH SEPARATION

DNN may be a feed-forward multilayer perception with number of hidden layers. It is been widely used for the task of classification within the speech recognition, object detection and image processing. For the speech enhancement, DNN was adopted as a regression model to find out the connection between noisy and clean speech. Here the DNN architecture was applied to speech separation in unsupervised mode [4] [11]. The DNN architecture illustrated in Fig.1 and Fig.2.



The LPS Feature of mixed speech

Fig.1 DNN architecture fronted for gender mixture detector



The LPS Feature of mixed speech

Fig.2. DNN based speech separator

The Fig.1 and Fig.2 shows that the DNN architecture for male and feminine speech separation. The DNN architecture has dual outputs for both female and male speaker groups within the

current frame gives the input features of mixed speech with multiple neighboring frames. The input is mixed by different genders speech of arbitrary speakers while output refers to the separate speech segments of male and feminine speaker group. This architecture has the advantages that these are avoids the restrictions of abundant data of the target speaker which is required to develop the speaker dependent models. Also this technique is capable of providing perceptually relevant parameters. Also the proposed DNN architecture improves the continuity of estimated clean speech [5] [6].

2. MATERIAL AND METHODS

The Fig.3 shows the proposed system architecture which is predicated on DNNs for speech separation. Very firstly construct four speaker clusters M1, M2, F1 and F2 with the training speaker data from each of the four groups. Then the gender mixture detector is implemented by a DNN with four outputs which are like four speaker groups [7], [8]. Then finally the speech mixtures of various combinations are adopted to coach the set of DNN separators. Within the proposed system three DNN separators are used which incorporates MM, FF, and MF separators to hide all the possible gender combinations [7].

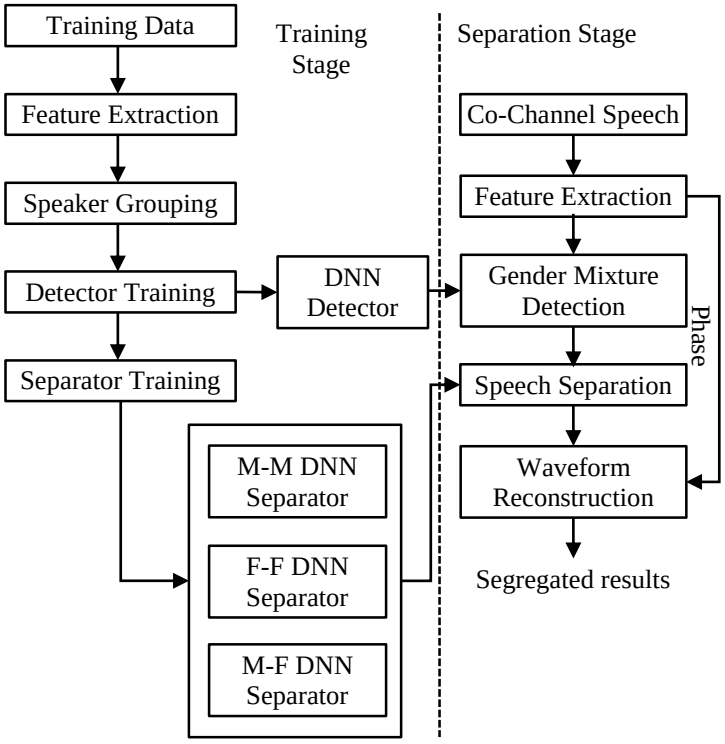


Fig.3. System Architecture

Then within the separation stage the feature extraction were done. The feature extraction stage firstly processed mixed speech by using gender mixture detector which is beneficial to work out sort of gender combinations. Then after this the speech separation is conducted with the corresponding DNN separator obtained within the training stage [14].

3. EXPERIMENTAL SETUP

For the evaluation of the method take randomly selected some different gender mixtures having a number of male female and a few of female male combinations from the entire SSC test set referred as two talker mixtures with signal to noise ratio. Then built DNN in total which were trained on different speaker groups. All the utterances of male and feminine speakers within the training set were wont to train each DNN. Then it had been evaluated on the speech mixtures of the opposite unseen male and feminine speakers.

3.1 TRAINING DATASET

Within the total SSC corpus there have been 10 Female and 10 male speakers. From that only pick small subset to coach the gender mixture detector and speech separator. 5 speakers were randomly chosen from each of the four groups that were M1, M2, F1 and F2. The F1 and F2 speaker groups were used for FF separator training while the M1and M2 speaker groups were used for MM separator training. And for the M1+M2 male group and F1+F2 female group were adopted to coach MF separator training. Then of these combinations were finally wont to train the gender mixture detector. For the noise reduction from the speech signal the utterances are taken randomly selected human voices. Noise is that the different random audio which is mixed within the training audio data. The audio signal mix with different noise signal is employed as testing dataset. At the time of testing the noise part is employed are different to make sure the testing and training data are distinct from one another.

4. RESULTS

The output of the system is to get original signal from mixed audio signal. The wave forms of all speech separators are shown below. Output of the system is to get female and male speech separately from the mixed speech for two samples shown below.

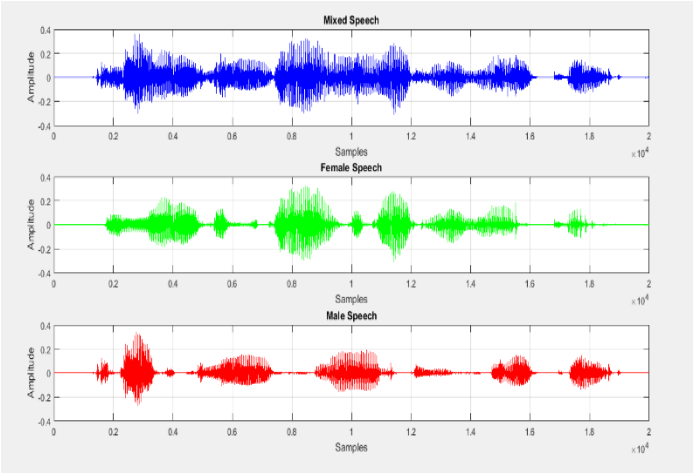


Fig.5. Separation of speech

5. CONCLUSION

The proposed system may be a DNN based gender mixture detection framework for unsupervised speech separation which is motivated by the analysis of speaker dissimilarities. Here the

importance of DNN based detector and therefore the comparison of various gender mixture combinations are conducted. The proposed system is that the demonstration of applying the deep learning technology to unsupervised speech separation which challenging open problem.

During this study the DNN module is employed which has multiple hidden layer which is nothing but the deep network and therefore the speech is separated by using DNN architecture. The DNN architecture shows the higher results of unsupervised speech separation.

REFERENCES

[1] N. Dave, "Feature Extraction Methods LPC, PLP and MFCC in Speech Recognition", *International Journal for Advance Research in Engineering and Technology*, Vol. 1, No. 6, pp. 1-4, 2013.

[2] Y. Xu, J. Du, L.R. Dai and C.H. Lee, "An Experimental Study on Speech Enhancement based on Deep Neural Network", *IEEE Signal Processing Letters*, Vol. 21, No. 5, pp. 65-68, 2014.

[3] Po Sen Huang, Minje Kim, Hasegawa-Johnson and Paris Smaragdis, "Deep Learning for Monaural Speech Separation", *Proceedings of IEEE International Conference on Acoustics. Speech and Signal Processing*, pp. 1562-1566, 2014.

[4] Y. Xu, J. Du, L.R.Dai and C.H. Lee, "A Regression Approach to Speech Enhancement Based on Deep Neural Network", *IEEE/ACM Transaction on Audio Speech, Language Processing*, Vol. 23, No. 1, pp. 7-19, 2015.

[5] Naoya Hashimoto, Kazuhiro Nakadai and Tetsuya Ogata, "Sound Source Separation for Robot Audition using Deep Learning", *Proceedings of IEEE International Conference on Humanoid Robots*, pp. 389-394, 2015.

[6] J. Du, Y. Tu and L.R. Dai, "A Regression Approach to Single Channel Speech Separation Via High-Resolution Deep Neural Network", *IEEE/ACM Transaction on Audio Speech, Language Processing*, Vol. 24, No. 8, pp. 1424-1437, 2016.

[7] P. Prithvi and T. Kishor Kumar, "Comparative Analysis of MFCC, LFCC, RASTA-PLP", *International Journal of Scientific Engineering and Research*, Vol. 4, No. 5, pp. 1-4, 2016.

[8] S. Donald, Yuxuan Wang and De Liang Wang, "Complex Ratio Masking for Monaural Speech Separation", *IEEE/ACM Transaction on Audio Speech, Language Processing*, Vol. 24, No. 3, pp. 483-492, 2016.

[9] Wang Yannan, Jun Du, Li Rong Dai and Chin-Hui Lee, "A Gender Mixture Detection Approach to Unsupervised Single Channel Speech Separation based on Deep Neural Network", *IEEE/ACM Transaction on Audio Speech, Language Processing*, Vol. 25, No.7, pp. 1535-1546, 2017.

[10] Yuzhou Liu and Wang De Liang, "Speaker-Dependent Multipatch Tracking using Deep Neural Networks", *Journal of Acoustical Society of America*, Vol. 141, No. 2, pp. 710-721, 2017.

[11] S. Xia, H. Li and X. Zhang, "Using Optimal Ratio Mask as Training Target for Supervised Speech Separation", *Proceedings of IEEE International Conference on Signal and Information*, pp. 163-166, 2017.

[12] Zhuo Chen and Nima Mesgarani, "Speaker Independent Speech Separation with Deep Attractor Network", *IEEE/ACM Transaction on Audio Speech, Language Processing*, Vol. 26, No. 4, pp. 787-796, 2018.

[13] De Liang, Fellow Wang and Jitong Chen, "Supervised Speech Separation based on Deep Learning: An Overview", *IEEE/ACM Transaction on Audio Speech, Language Processing*, Vol. 25, No. 7, pp. 2329-2340, 2018.

[14] Shreya Sose, Swapnil Mali and S.P. Mahajan, "Sound Source Separation using Neural Network", *Proceedings of IEEE International Conference on Computing, Communication and Networking Technologies*, pp. 1-14, 2019.