TEST Engineering & Management

May – June 2020
ISSN: 0193-4120 Page No. 8121 - 8131

Effect of feature normalization objective improvement of over Noisy Single-channel Speech Enhancement with Neural Networks

Dr.S.China Venkateswarlu[1], Dr.N.Udaya Kumar[2],K.Chaitanya[3], A Usha Sree[4]

[1](Professor-ECE, Institute of Aeronautical Engineering, Hyderabad, Telangana, INDIA.)

[2](Professor-ECE, Marri Laxman Reddy Institute of Technology and Management, Hyderabad, Telangana, INDIA.)

[3](Asst.Professor-ECE, Institute of Aeronautical Engineering, Hyderabad, Telangana, INDIA.)

[4](Professor-ECE, Gokaraju Rangaraju Institute of Engineering and Technology, Bachupally, Hyderabad, Telangana, INDIA.)

Article Info
Volume 83
Page Number: 8121 - 8131
Publication Issue:
May - June 2020

Article History
Article Received*: 19 November 2019*
Revised*: 27 January 2020*
Accepted*: 24 February 2020*
Publication*: 18 May 2020*

Abstract:

The signal obtained from the real world environment is often corrupted by means of unwanted noise. So, it is important to effectively ensure speech quality and obtain a noiseless speech signal of higher quality by applying the optimal noise cancellation technique. The main aim is to improve the speech intelligibility and speech quality. The signal obtained from the real world environment is often corrupted by means of unwanted noise. So, it is important to effectively ensure speech quality and obtain a noiseless speech signal of higher quality.se cancellation technique. Single Channel Speech Enhancement aims to reduce noise and retain speech quality to the best extent possible from noisy speech. Our overall assumption is that our noisy speech comes from addition of a clean speech and the noise signal and there is no other assumed distortion non-linear distortion, channel distortion and reverberation. The other general assumption we make is the noise attributes typically change slower than speech .to suppress noise, to retain speech to the best extent possible, to improving perception. In this Research work, our goal is for the end-users, the human listeners who are going to listen to the enhanced clips. So that will be our research investigated speech quality performances with help of neural networks.

Keywords: *speech signal, speech corpus, neural networks, speech enhancement*

1. Introduction: The main objective of Speech Enhancement algorithms is to improve the perceptual quality of extracting speech signal from noisy speech. Noise estimation is soul component in speech enhancement techniques, as better noise estimation gives a high quality of speech extraction. In recent decades removing noise from noisy speech is challenging issue because of the spectral properties of non-stationary noise is very difficult to predict. Noise estimation is the issue in speech enhancement algorithms which is complicated since if the noise power is more than speech power, and then the speech content may be removed and treated as noise. Speech processing has its uses and applications in teleconferencing systems, speech recognition based security devices, biomedical signal processing, hearing aids, ATM machines and computers, in speech enhancement only noise signal is present and therefore it is the most complicated and interested research area in digital signal processing. Many different Algorithms have been developed by many researchers to improve the noisy speech but it is still a challenging area as the characteristics of the noise varies in a dramatic way depending upon time, noise . There are various speech enhancement techniques. Overview of the generic speech enhancement. We have the flow diagram of a classical signal processing-based speech Enhancement system. We start with our time-domain waveform signal x of t, and throughout the talk, I will be assigning x to all the noise signals. That signal goes through a short-time spectrum analysis, typically the short-time Fourier transform, to get the short-time spectral

May – June 2020
ISSN: 0193-4120 Page No. 8121 - 8131

characteristics. After that, we separate the short-time spectral features into phase and into magnitude denoted by these little blocks there into magnitude denoted by these little blocks there. One challenging aspect of single-channel enhancement is that the phase is typically very hard to recover. So that is out of scope of our talk today as well and will be leaving it as it is, the noisy phase for reconstruction. We do the majority of our work in the magnitude domain generic modules to estimate noise or estimate the gain from the magnitude of the noise spectrum. After that, we send into an estimator, will be called a gain estimator that applies basically a gain function in the frequency domain on each frame of our noisy spectra. We use that to point wise multiply to our noisy speech and use that enhance magnitude to recover the clean speech.

2. Literature Survey: There's a research work by Efren and Malah in 1984. The problem by assuming complex as TFTs of speech and noise has Gaussian distributions and are uncorrelated and they solved for the optimal solution in minimum mean squared error sense. For deep learning based approach, we actually don't have any assumptions about distributions of anything and we simply learn by stochastic gradient descent and hopefully we get to a point where its low enough for training and low enough for test. The mean squared error has a staple convergence because if you take the gradient of a square, you have a linear gradient across everywhere. STFT(Short Term Frequency Spectra Transformation) The Short Term Frequency Spectra Transformation is There's a research work by Efren and Malah in 1984 but actually in timed minutes, Robert Winner in 1947.I was going to emphasize the Gaussian. But wiener actually is based on mean square error in time domain signal, with that observation; we can rewrite this mean squared error.

2.1 Speech processing: The term speech processing basically refers to the scientific discipline dealing the analysis and processing of speech signals in order to achieve the best performance in various practical scenarios . The field of speech processing is undergoing a rapid growth in terms of both performance and applications. This is due to the advancement in the field of microelectronics, computation and algorithm design. Speech processing still covers an extremely broad area, which relates to the following three engineering applications

2. 2 Separating Speech and Noise Objectives:
To emphasize the widowing but different techniques actually is based on mean square error in time domain signal, with that observation; we can rewrite this mean squared error. If we put in statistical form is expected value instead of actual average. If you just rewrite a little bit and we ignore the cross term there, we ended up with two, that's up by the way that's a very cores assumption that may be doesn't hold. But for the, our goal is to separate speech distortion from noise suppression and by ignoring the two terms, what we ended up is actually the mean square error between the signal enhanced. The S here is the clean signal, so it's a mean square or between the clean signal and the clean signal itself multiplied the gain function. We have a mean squared error of just noise multiplied by the gain. We first did this approximation and then we come up with this new loss function that has two separate terms. The first one is on a speech distortion and the second term is on noise suppression.. If you want to balance you do so with alpha. For the speech, we only do that for the speech active region. So we apply a simple energy-based voice activity detector , the detector is simply a thresholding [3]on the energy Accumulated from three kilohertz to 5000 hertz which is typically where speech happens. Can we get that on a plain speech, yes, yeah that's a very crude energy based VAD.

3.1 Training consideration:
So from the classical decision directed approach from Efren and Malah, we have hidden state, Priori and a Posteriori SNRs as your hidden states in deep

May – June 2020
ISSN: 0193-4120 Page No. 8121 - 8131

learning language. The hidden states from the previous estimate affect the current by a exponential smoothing process. We have this analog in the RM based approach. But what we have is a blog walks almost with hidden states that we don't know the meaning of the hidden states they carry. But we know that they are capable of learning very long temporal sequence. They are learning through back propagation through time m. we want to actually study the effect of the length of the sequence we pass in because this is just a simple pseudo recurrent Neural Network I have here. Just a simple pseudo recurrent Neural Network I have here, so this is your hidden state from previous time frame and the hidden state from the current. I have from all the way back to zero but we can control this length and see how it affects the impact, how it affects speech quality of the enhanced signal.

4. Speech Enhancement Algorithm

The block diagram of the proposed enhancement system is illustrated. The different steps of our system are detailed in the following paragraphs. The first is feature extraction and the Neural Network itself, the learning objective, and how we actually train our system. Our method will be broken into four divisions.

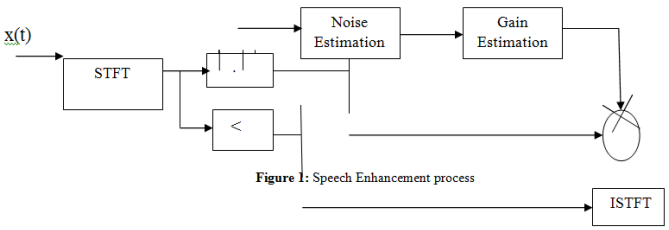

Figure 1: Speech Enhancement process

The first is feature extraction and the Neural Network itself, the learning objective, and how we actually train our system. Our method will be broken into 4 pieces.

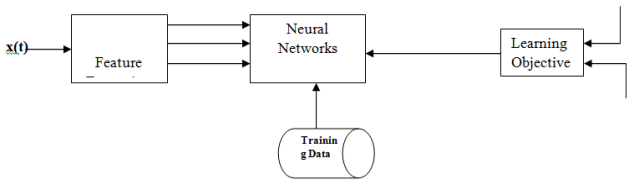

Figure 2: Speech Enhancement with Neural Networks- Training and learning data with objective

4.1 Musical Noise Problem in Spectral Subtraction Algorithm:

Spectral subtraction algorithm is very smooth to put into effect and will be having much less calculations. yet, there are some troubles found in this set of rules. It will not examine the distribution of the noise spectrum. As a result, if the distribution is uneven, spectral subtraction algorithm will produce a huge quantity of musical noise even as performing the noise reduction [4]. As this algorithm is obtained by way of subtracting the energy of the blended speech from the common power of the noise, large amount of noise is remained at the position where the noise sign is robust, and much less noise will be compensated at the region wherein the noise signal is week. those residual noises will contribute to the musical noise at last.

.

4.2 STFT (Short Term Frequency Spectra Transformation): The Short Term Frequency Spectra Transformation is there's a research work by but actually in timed minutes, Robert Winner in 1947. But wiener actually is based on mean square error in time domain signal, with that observation; we can rewrite this mean squared error.

4.2.1 Separating Speech and Noise Objectives:

To emphasize the windowing techniques but speech techniques actually is based on mean square error in time domain signal, with that observation; we can rewrite this mean squared error. If we put in statistical form is expected value instead of actual average. If you just rewrite a little bit and we ignore the cross term there, we ended up with two, that's up by the way that's a very cores assumption that may be doesn't hold. But for the, our goal is to separate speech distortion from noise suppression

May – June 2020
ISSN: 0193-4120 Page No. 8121 - 8131

and by ignoring the two terms, what we ended up is actually the mean square error between the signal enhanced.

MSE: $E[(S-S^{\wedge})^2] = E[(S-G(S+N))^2] = E[(S-GS)^2] + E[(GN)^2]$ -------(1)

Assign weighting to separated speech distortion and noise suppression terms. the S here is the clean signal, so it's a mean square or between the clean signal and the clean signal itself multiplied the gain function. We have a mean squared error of just noise multiplied by the gain.

$E[(S-S^{\wedge})^2] = E[(S-G(S+N))^2] = E[(S-GS)^2] + E[(GN)^2]$ -------(2)

Assign weighting to separated speech distortion and noise suppression terms

$L_{SN}(\phi; S,N) = \alpha \| S-GS \|^2 + (1-\alpha) \| GS \|^2$ ----------(3)

So because we are not solving for any optimal solutions in statistical sense and also we want to balance the speech distortion and noise suppression terms, We first did this approximation and then we come up with this new loss function that has two separate terms. The first one is on a speech distortion and the second term is on noise suppression. So the way to interpret this is, let's say your Enhancement system does nothing which means they are going just simply pass everything then your speech distortion[6] is zero but then you have all the error coming from the noise. If you want to balance you do so with alpha. We don't stop there we also have this observation from classical signal processing point of view that when we have a noisy signal that's almost clean then we don't want to destroy any speech content in there. So the result is we pass almost all the noisy speech on change to retain the speech quality. When there's so much noise is speech that we cannot even get a hang of where the speech is we just apply a very heavy suppression on the entire thing. A very heavy suppression on the entire thing. So that basically says, when the SNR is approaching infinity we want very little speech distortion and when SNR is approaching zero, we want very aggressive suppression on the noisy speech .

4.2.2 SNR weighted objectives:

The weighting is static, but our goal varies across different scenarios: we want little speech distortion when only speech is present SNR to infinite. We want aggressive suppression when only noise presents SNR to 0. Existing work in classical Speech processing approach. Adapt the loss of each example pair such that speech and noise by the global

SNR: $L_{SNR}(\phi; S^{(i)}, N^{(i)}) = \alpha \sigma 2 \frac{S(i)}{\sigma 2} N(i)$ $S^{(i)}$ $\| _{SA}$-$GS^{(i)}$ ^{2} + (1- α) α $\sigma 2 \frac{S(i)}{\sigma 2} N(i)$ $GN^{(i)\,2}$ –(4)

Motivated by this observation, we have another loss function built on top of the previous one. With this SNR terms multiplied to each part of the loss function there to each part of the loss function there I have to mention here that the original intention was to view this as a whole term. So this is the waiting for speech and one minus the other multiply with this parenthesis here, But this is by example. So imagine you have a batch of audio. ,still it will be more correct if you do it with waiting on the alpha by the SNR[10].

4.3 Training consideration:

The classical decision directed-Priori and a Posteriori SNRs as your hidden states in deep learning language. The hidden states from the previous estimate affect the current by a exponential smoothing process. We know that they are capable of learning very long temporal sequence. They are learning through back propagation through time m. we want to actually study the effect of the length of the sequence we pass in because this is just a simple pseudo recurrent Neural Network I have here.

May – June 2020
ISSN: 0193-4120 Page No. 8121 - 8131

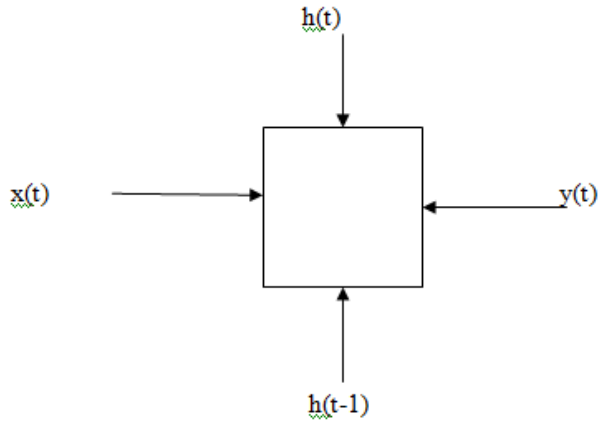

Figure 3: Training consideration

h(t) =y(t) , y(t)=f(x(t)+h(t-1))

= f(t)+h(t-1)) (f(x(t-1)+h(t-2)) h(t-2))

= g(x(t), x(t-1), ------x(0) (6)

We want to compare a small batch of long sequences to a large batch of short sequences, given the same amount of information per batch. Just a simple pseudo recurrent Neural Network I have here, so this is your hidden state from previous time frame and the hidden state from the current. As the output here and your input is x ( )t and your output some y (t) here. Let's just say your hidden state of t simply equals to your output and your output is simply a function of your input plus your previous hidden state. Then if we take the partial derivative of the output with respect to the learning parameters of the network, we see it's a function of your current instantaneous gradient multiplied by something from the previous time frames[13] and this t here. Here, I have from all the way back to zero but we can control this length and see how it affects the impact, how it affects speech quality of the enhanced signal. So we are doing this comparison as well. That's the end of our method.

Now, let's move on to evaluation clean speech. So there's no overlap to training at all. For noise, we are picking six challenging classes from the 11 in training, but we have different signals for test. Those are very challenging noise types. we have the competing talker in neighbor and we have transient noise , or the door shutting, and a airport announcement. Noise clips in the test data are not ever presented in the training set. No right we have five different combinations of SNRs from 0dB to 10dB with a 15 dB step and all clips are sampled at 16 kilo hertz's.

This is just a close up look of the data we have on top, we have clean speech, on the bottom, we have the noise. This is our waveform as shown in below figure 9, plotted in dB and you see this is the same noisy repeated five times her from 0dB to 10 dB there. We have the speech normalize to the same level, but it's the same speech repeated five times. During training, what we did is augment data a little bit by randomly drawing a segment of waveform from any clean speech file. From noise file, we do the same and we mix them. So the SNR wouldn't change is still the five discrete SNRs.

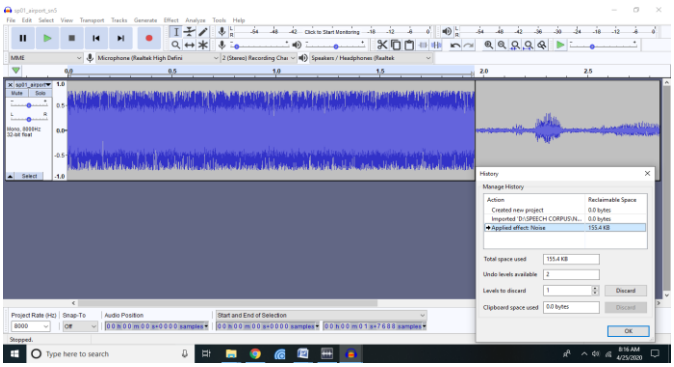

Figure 4: Evaluation and data augmentation noise repeated variations different times with Discrete SNRs assuming point wise addition

Discrete SNRs assuming point wise addition it might be even better to mix with different SNRs on the fly. So we will just draw one side and draw one speech and mix at a randomly draw SNR level, but we didn't try that, a randomly draw SNR level, but we didn't try that so that's our data and we have quite a few systems to compare. We start with a noisy unprocessed and we have the statistical based. This is the signal parsing based method developed here in MSR without training data of course. We have our processed method here with these set up and we have a recurrent Neural network which is simply our network but removing the residual connection. So we want to study how effective that is residual connection actually is everything else

May – June 2020
ISSN: 0193-4120 Page No. 8121 - 8131

stays the same. Theirs is enhancing a very crude energy and we'll hope it's a 22-band, but we have a full band 257. So what we did was took their architecture, scale up the future dimensions to match it for full band and scale up all the other dimensions within the network to accommodate this scaling difference.

The VAD No. for the proposed method it is kind of building to the learning objective because of the speech distortion. Yeah it is in the RN Noise. But we found we randomize that and it didn't change anything. Let me know the question. Finally we have oracle information plus Wiener [17] filter rule, which marks theoretically the best what we can do. So we have seven systems to compare in terms of evaluation metrics, we have 4 classical speech quality or in intelligibility measures. Single-Channel Speech Enhancement: Assumptions: Noisy Speech=Speech + Noise, Noise attributes change slower than speech .Suppress noise, Retain Speech and Improve Human or/and Machine perception

4.4 Generic Speech Enhancement Pipelines

Let's have a overview of the generic speech enhancement pipeline. On top, we have the flow diagram of a classical signal processing-based speech Enhancement system.. We do the majority of our work in the magnitude domain. You see there are some generic modules to estimate noise or estimate the gain from the magnitude of the noise spectrum. After that, we send into an estimator, will be called a gain estimator that applies basically a gain function in the frequency domain on each frame of our noisy spectra. We use that to point wise multiply to our noisy speech and use that enhance magnitude to recover the clean speech. The basic pipeline is similar in the sense that we start with our time-domain signal. We do some feature extraction which does not have to be spectral features anymore, it can be anything. But our end goal is still to estimate this time frequency gain function denoted g with a hat there. Then point wise, multiply that to the noisy magnitude to recover the clean speech, hopefully. As you can see here everything else in the middle becomes more or less a black-box because of the Neural Networks.

4.5 Deep-Learning versus classical: The first is feature extraction and the Neural Network itself, the learning objective, and how we actually train our system. Our method will be broken into 4 pieces. Speech processing versus Deep-learning for online enhancement: Before I get into the actual method, let me brief go over this short chart I picked. As you can see, we have six methods right here. The first two are from a classical signal processing-based method. The middle two are deep learning based but cannot operate in real time. The last two rows are deep –learning based and can actually operate in real-time.

4.6 Input and output sampled speech: variance normalization with FD or FI. So, why do you use the log power spectrum, so the question we got is why we use a log power spectrum is our perception is correlated on a log scale for audio. As you can see, we are actually, visually we can see the contrast if we mapped linearly the value obtained from the log scale. If I did it for just the magnitude, the contrast will be so low that you would not even visually see that different. But the magnitude actually it already contain the information for the power, right if you just double the magnitude I mean just it comes back to its power, right. Well the log power is just a non-linear compression on the linear power.

4.7 Learning Machines Noises: After the features we are getting into probably the most important part of our system which is the learning machine. The Neural Network itself. The recurrent neural network is the most natural choice for us because what the recurrent neural network does is it outputs some value for; it has a notation of time first and foremost. Then it outputs something for this time instant based on some input you obtained for the current time and also from the output you get from the previous time stamps.

4.8 Residuals for Speech Enhancement: Global view and zoomed in view we did is simply stocking a few. In our case, just three grew layers with our residue connections. If you zoom in on each block it will look like this, except for the last layer where

May – June 2020
ISSN: 0193-4120 Page No. 8121 - 8131

we don't add this residual. The justification is that for input features we are getting something from audio which has a very high dynamic range and all the output that comes in all of the (in audible) is compressed already in the last layer before it gets transformed by a fully-connected layer. We don't want the input in dynamic range mess up with what's getting learned inside. So we don't have that. Everything else has the residual and in the end we have a fully connected layer with a sigmoid function. The outputs are again between zero and one. That is our network architecture.

so after the features we are getting into probably the most important part of our system which is the learning machine. The Neural Network itself. The recurrent neural network is the most natural choice for us because what the recurrent neural network does is it outputs some value for, it has a notation of time first and foremost. Then it outputs something for this time instant based on some input you obtained for the current time and also from the output you get from the previous time stamps. This is similar to similar to what we do with the filtering in or the classical approach we have with speech enhancement.

4.9 Evaluation data

Now, let's move on to evaluation. We have 7 hours of training data. We have 7 hours of training data. The clean stage comes from the Edinburg 15 speakers corpus, the noise comes from 14 noise types from speech corpus database and free sound. For test, we have 10 hours of test clips. The clean speech comes from the speech corpus 12 speaker's corpus. So there's no overlap to training at all. For noise, we are picking nine challenging classes from the 14 in training, but we have different signals for test. Those are very challenging noise types.

We have the competing talker in neighbor and we have transient noise such as munching, or the door shutting, and a airport announcement. Noise clips in the test data are not ever presented in the training set. No right no. yeah. We have five different combinations of SNRs from 0dB to 10dB with a 15 dB step and all clips are sampled at 16 kilo hertz's.

4.10 Evaluation Metrics: they are the scale invariant signal to distortion ratio, which is a really robust version of SNR. We have captured distance which is a distance metric in the capture domain. A capture domain is supposedly, you have a flattened channel and speech dimension. For the third short-term objective, intelligibility this is in terms of percentage. Finally, the perceptual evaluation of speech quality .Speech is always at the same level. You need augment the data with the whole input with different levels. You have a knob for your microphone, right? You can pull it down 20dB or crank it up.

5. Final Simulation Results

We use the most standard feature for a Neural Network that is the short-time Fourier transform magnitude. We also consider the short-time log power spectra with a negative 80 dB floor. What you see on the left is actually the log power spectra with a linear mapping of a color and displayed with a jet color map in MATLAB. Let's see. We have three, just for the ease of looking, we have in x-axis your time in seconds, your y-axis in frequency in kilohertz. But we have three spectrograms stack together. The top one, we have the noisy. As shown in above figure 3, I think that's with the air conditioner noise at 20 dB. In the middle, we have the clean speech signal, and on the bottom, we have this weird looking IRM or you call it the ideal ratio mass which is the ideal gain function. You plot it in 2D on a dB scale because if I plot it in between zero and one, you would not have to see the contrast. As I said before, the output we are trying to estimate is the real magnitude gaining function, the range between zero and one and some technical details about how we construct this spectrogram. We have a 16kilohertz sampling rate for our audio. We used a 32 millisecond analysis frame with a Hamming window and a 75 percent overlap. Hamming window or Hann window , Hamming window. The four percent window from the zero, yeah. Not the Hann window, the 0.46 + 0.54 times the cosine. That's Hamming. Yeah, Hamming window is from zero to one to zero. 0 to 1 to 0. Okay. Its not Hann, its Hamming, yeah. It's a raised cosine. Have you

May – June 2020
ISSN: 0193-4120 Page No. 8121 - 8131

tried different windows, different . I briefly tried out 20 millisecond 50 percent overlap and the performance went down for the network. That was a month ago, and I set it aside and never really changed my original setup. Yeah. But I think it will work. Well, the overlap might be a problem but 20 millisecond window, I think it will work.

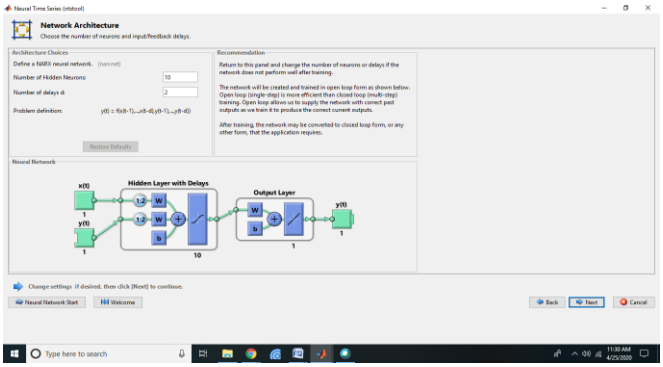

Figure 5: Network Architecture Input and output sampled speech from the training set , online mean and variance normalization with FD or FI.

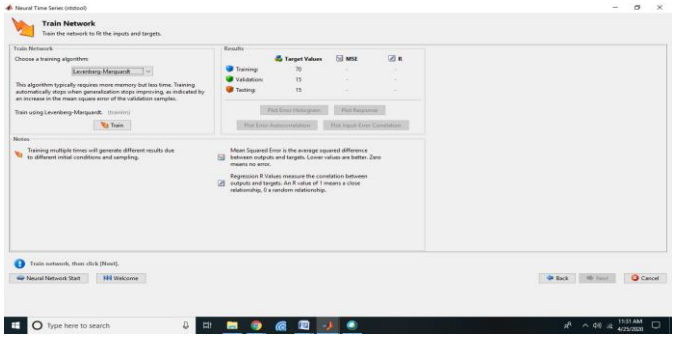

Figure 6: Network Architecture-Train Network with help of training, validation and Testing.

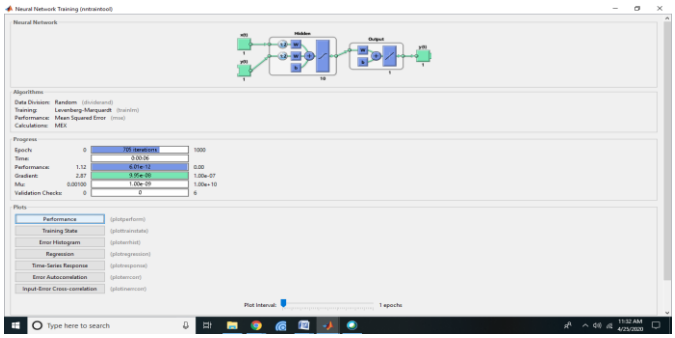

Figure 7: Network Architecture-Train Network –x(t), y(t) output variations in terms of y(t):Progress : Epoch value is 705 iterations, Time:0.00.06, performance 1.12: 6.01e^{-12}, Gradient: 2.87-9.95e^{-08}. Mu: 0.00100, 1.00e^{-08}.and validation checks..

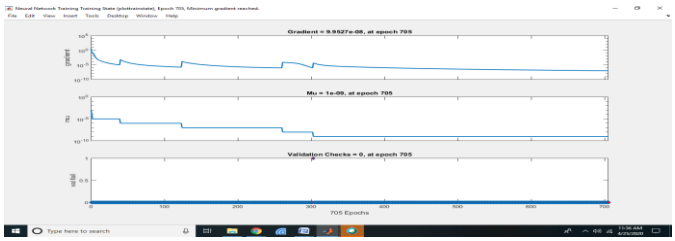

Figure 8: Network Architecture-Train Network: Gradient: 2.87-9.95e^{-08}. Mu:0.00100, 1.00e^{-08}.and validation checks..

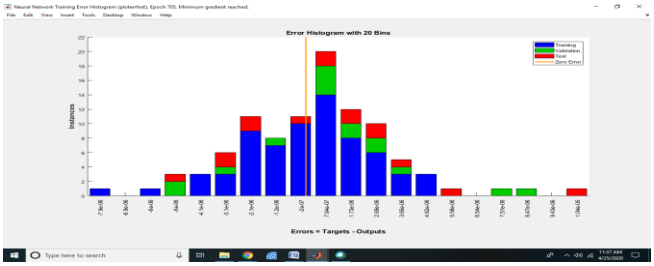

Figure 9: Network Architecture-Train Network –x(t), y(t) output variations in terms of y(t): error Histogram , error targets results.

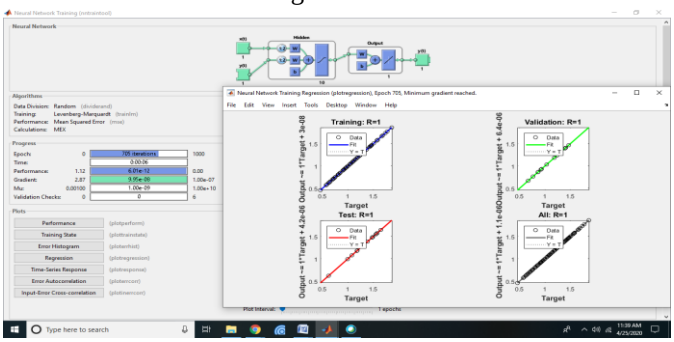

Figure 10: Network Architecture-Train Network –x(t), y(t) output variations in terms of y(t):Progress : Epoch value is 705 iterations, Time:0.00.06, performance 1.12: 6.01e^{-12}, Gradient: 2.87-9.95e^{-08}. Mu: 0.00100, 1.00e^{-08}.and validation checks..in terms or training's=1 and validations.

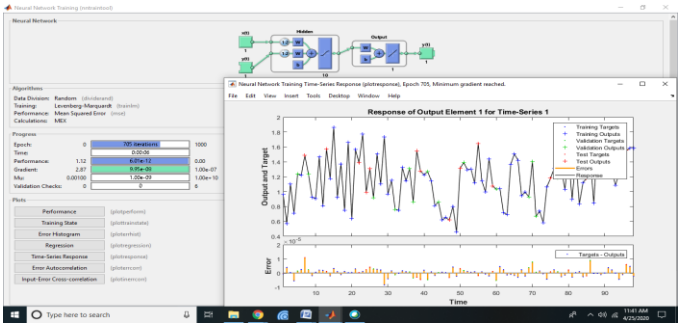

Figure 11: Network Architecture-Train Network –x(t), y(t) output variations in terms of y(t):Progress : Epoch value is 705 iterations, Time:0.00.06, performance 1.12: 6.01e^{-12}, Gradient: 2.87-9.95e^{-08}. Mu: 0.00100, 1.00e^{-08}.and validation checks..in terms Responses of output element1 for time series 1.

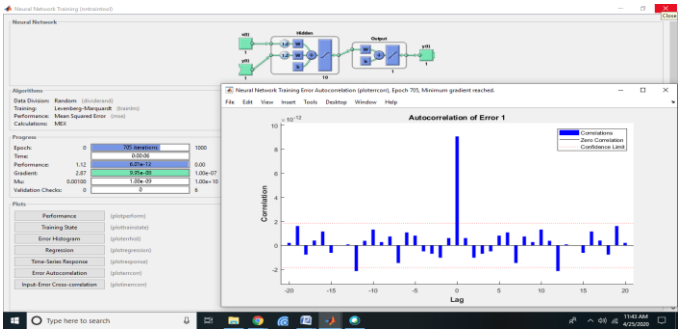

Figure 12: Network Architecture-Train Network –x(t), y(t) output variations in terms of y(t):Progress : Epoch value is 705 iterations, Time:0.00.06, performance 1.12: 6.01e^{-12}, Gradient: 2.87-9.95e^{-}

May – June 2020
ISSN: 0193-4120 Page No. 8121 - 8131

^{08}. Mu: 0.00100, 1.00e^{-08}.and validation checks..in terms Autocorrelations of Error-1

6. Conclusion and Future scope

As shown in above figure, so the first four bars are based on our short-term spectral amplitude. The next four bars are based on long spectra. Here we have the original spectra after global normalization, after online frequency-dependent normalization, and after online frequency independent normalization and the same for long spectra. Using exact same network architecture only difference is the feature. Normalization, the green loud spectrum, global normalization is actually performance reached. This is based on mean-squared error only. The speech distortion, weighted off. We mentioned that the clean speech is roughly the same. We have either but in reality this is valid if you have a microphone or you are roughly the same distance from the microphone. But if you are using a local microphone you can be half a meter or five meters away. You only have a 20 dB difference in the voice level. That dynamic range is where the normalization would tremendously help. So this is of not much value. Any conclusion here is that's true. Speech is always at the same level. That's true. we need augment the data with the whole input with different levels. We have a knob for your microphone, can pull it down 20dB or crank it up. You don't know at which time you get the audio. So you need to augment it. This is the dynamic range which normalization battles.

ACKNOWLEDGEMENT

The work is carried out through the research facility at the Department of Electronics & Communication Engineering IARE-Hyderabad.and from JNTUH. The Authors also would like to thank the authorities of JNTU Hyderabad, Institute of Aeronautical Engineering (IARE), Dundigal, Hyderabad and MLRITM, Hyderabad for encouraging this research work.

References:

[1] P. C. Loizou, Speech Enhancement: Theory and Practice, 1st ed. CRC, 2007.

[2] S. F. Boll, "Suppression of acoustic noise in speech using spectral subtraction," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-27, pp. 113–120, 1979.

[3] M. Berouti, R. Schwartz, and J. Makhoul, "Enhancement of speech corrupted by acoustic noise," *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, pp. 208–211, 1979.

[4] Y.Hu & P.Loizou, "Evaluation of objective measures for speech enhancement".IEEE Trans. Audio speech Lang. process, Vol .16. No.1, pp.229- 238, Jan-2008.

[5] P. Krishnamoorthy, "An Overview of Subjective and Objective Quality Measures for Noisy Speech Enhancement Algorithms", IETE technical review, vol 28, issue 4, jul-aug 2011, pp 292-301.

[6] IEEE Subcommittee, IEEE recommended practice for speech quality measurements, IEEE Trans. on Audio Electro acoustics, Vol. 17, Issue 3, pp. 225-246, September 1969.

[7] The NOIZEUS database: http://www.utdallas.edu/~loizou/speech/ noise

[8] K. S. Riedel and A. Sidorenko, "Minimum bias multiple taper spectral estimation," *IEEE Trans. Signal Processing*, vol. 43, pp. 188–195, 1995.

[9] P. Moulin, "Wavelet thresholding techniques for power spectrum estimation," *IEEE Trans. Signal Processing*, vol. 42, pp. 3126–3136, Dec. 1994.

[10] A. T. Walden, D. B. Percival, and E. J. McCoy, "Spectrum estimation by wavelet thresholding of multitaper estimators," *IEEE Trans. Signal Processing*, vol. 46, pp. 3153–3165, 1998.

[11] D. L. Donoho, "De-noising by soft-thresholding," *IEEE Trans. Inform. Theory*, vol. 41, pp. 613–627, May 1995.

[12] D. L. Donoho and I. M. Johnstone, "Adapting to unknown smoothness via wavelet thrinkage," *J. Amer. Statist. Assoc.*, vol. 90, pp. 1200–1224, 1995.

[13] S. G. Mallat, "A theory for multiresolution signal decomposition: The wavelet presentation," *IEEE Trans. Pattern Anal. Machine Intell.*, vol.11, pp. 674–693, July 1989.

[14] Y. Hu and P. C. Loizou, "Ageneralized subspace approach for enhancing speech corrupted by

May – June 2020
ISSN: 0193-4120 Page No. 8121 - 8131

colored noise," *IEEE Trans. Speech Audio Processing*, vol. 11, pp. 334–341, July 2003.

[15] Y. Ephraim and H. L. V. Trees, "A signal subspace approach for speech enhancement," *IEEE Trans. Speech Audio Processing*, vol. 3, pp. 251–266, 1995.

[16] S. China Venkateswarlu, ASR Reddy, K. S. Prasad, "Speech Enhancement using Boll's Spectral Subtraction Method based on Gaussian Window", Global Journal of Researches in Engineering: F, Electrical and Electronics Engineering, Volume 14 Issue 6, Version 1.0, pp.9-20,Year 2014,Type: Double Blind Peer Reviewed International Research Journal, Publisher: Global Journals Inc. (USA), Online ISSN: 2249-4596 & Print ISSN: 0975-5861.

[17] S.China Venkateswarlu, A.Karthik and .K.Naveen Kumar "Implementation of Area optimized Low power Multiplication and Accumulation", International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, pp. 2928-2932 ,Volume-9, Issue-1, November 2019. Published By: Blue Eyes Intelligence Engineering & Sciences Publication, Retrieval Number:

[18] S.China Venkateswarlu, A.Karthik and .K.Naveen Kumar "Performance on Speech Enhancement in Objective Quality Measures Using Hybrid Wavelet Thresholding", International Journal of Engineering and Advanced Technology (IJEAT), Volume 6, Issue 8(2019) ,pp.3523-3533 published on August 2019. DOI: 10.35940/ijeat.F9343.088619

[19] S.China Venkateswarlu, A.Subba Rami Reddy and .K.Satya Prasad "Speech Enhancement in Terms of Objective Quality Measures with Effect of Adjustable Window Shape Parameter on Speech Enhancement Techniques", Cromejournals based –Journal of Signal and Image Research published on July 2016.

[20] S.China Venkateswarlu, A.Subba Rami Reddy & K.Satya Prasad, "Speech Enhancement in terms of Objective Quality Measures Based on Wavelet Hybrid Thresholding the Multi-taper Spectrum", published in IJAREEIE, Volume5, Issue 1, pp.201-219, January 2016, DOI:10.15662/IJAREEIE.2015.0501036

[21]A.Subba Rami Reddy,V.Harika, S.China Venkateswarlu, "Telugu Speech Enhancement in Terms of Objective Quality Measures Using Discrete Wavelet Transform using Hybrid Thresholding", published in IJAREEIE,Volume3,Issue8,August 2014,ISSN(online): 2278-8875,ISSN(Print):2320-3765

[22] S.China Venkateswarlu, A.Subba Rami Reddy,K.Satya Prasad, "Speech Enhancement using Bolls Spectral Subtraction Method Based on Gaussian Window" published in Global Journal of Researches in Engineering –GJRE Volume 14 Issue 6 Version 1.0.pp.no.9-19.September2014.

Dr.S. China Venkateswarlu, B.Tech, M.Tech, Ph.D.,MISTE, CSI, IAENG, SDIWC, ISRD, IFERP.,working as a Professor in Dept. of ECE at Institute of Aeronautical Engineering(Autonomous), Dundigal, Hyderabad. INDIA, Worked in various Engineering College and have 21 years of Teaching, R&D and Industrial experience. Research in the area of Signal Processing / Speech Processing. Image video processing. He has presented 13 National, 12 International Conference Papers are Presented throw EDAs. He has published 15 research papers in National and International Journals. He has reviewed text book on Digital Signal Processing. He has published a text book on Speech Enhancement Techniques.

Dr. N. Udaya Kumar received the B.Tech degree in Electronics and Communication Engineering from JNTU, Hyderabad, India, in 2004, M.Tech. degree in Embedded Systems from JNTU, Anantapur and the Ph.D. degree in Digital Image Processing from Sri Venkateswara University Tirupati, India, in 2008 and 2019, respectively. He served as an Assistant Professor from 2014 to 2016 in Nagole Institute of Technology and Science, Hyderabad, and from 2016 to 2019 in Vaishnavi Institute of Technology,Tirupati. Presently

May – June 2020
ISSN: 0193-4120 Page No. 8121 - 8131

working as Associate Professor in the Department of ECE at Marri Laxman Reddy Institute of Technology and Management, Hyderabad, Telangana, INDIA. His research interests include Digital Image Processing, Embedded systems. As of today, he has published 8 research papers in various international journals and conferences. He is a member of MIE,IAENG and IFERP.

K.Chaitanya, M.Tech(Communication and Signal Processing) , Gayatri Vidhya Parishad College of Engineering (Autonomous), B.Tech (ECE) at Gokul institute of technology and sciences. Presently working as assistant professor in the Department of ECE at Institute of Aeronautical Engineering (Autonomous). His research interests include Digital Image Processing, Embedded systems. As of today, he has published 8 research papers in various international journals and conferences. He is a member of MIE,IAENG and IFERP.

A. Ushasree is currently working as Assistant Professor at Gokaraju Rangaraju Institute of Engineering and Technology, Bachupally, Hyderabad, Telangana, India and pursuing Ph.D in the area of Antenna Designing from the KL University, Vijayawada, A.P, India. She received M.Tech degree in Embedded Systems from the Annamacharya Institute of Technology & Science, Anantapur, A.P, India and B.Tech degree from Adhiparasakthi college of engineering for Women, Kalavai, Vellore, India in Electronics and Communication Engineering. Her research interests are Antenna Designing, Wireless Communication and Image Processing.